МЕТОДЫ КОМПЬЮТЕРНОЙ ОПТИКИ

Под редакцией
В. А. Сойфера

Издание второе, исправленное

Допущено Министерством образования Российской Федерации в качестве учебника для студентов высших учебных заведений, обучающихся по направлению подготовки 511600 «Прикладные математика и физика»

Самарский Государственный аэрокосмический университет
№ 708126
Научно-техническая библиотека

Москва
Физматлит
2003

УДК 535.42
ББК 22.343
М54

КОЛЛЕКТИВ АВТОРОВ

А.В. Волков, Д.Л. Головашкин, Л.Л. Досколович, Н.Л. Казанский, В.В. Котляр, В.С. Павельев, Р.В. Скиданов, В.А. Сойфер, В.С. Соловьев, Г.В. Успленьев, С.И. Харитонов, С.Н. Хонина

Методы компьютерной оптики/Под ред. В.А. Сойфера: Учеб. для вузов. — 2-е изд., испр. — М.: ФИЗМАТЛИТ, 2003. — 688 с. — ISBN 5-9221-0434-9.

Излагаются основы компьютерного синтеза дифракционных оптических элементов (ДОЭ) с широкими функциональными возможностями. Обсуждаются методы получения зонированных пластинок со сложным профилем зон. Значительное внимание уделено математическим моделям и методам расчета ДОЭ: геометро-оптическому расчёту, итеративным и градиентным алгоритмам, строгому электромагнитному подходу к расчёту ДОЭ. Рассмотрены различные типы ДОЭ: фокусаторы, моданы, формирователи лазерных пучков с инвариантными свойствами, многопорядковые дифракционные решетки, аксиконы и многофокусные линзы. Все эти ДОЭ находят применение в задачах фокусировки лазерного излучения, в лазерных системах с волоконной и интегральной оптикой, а также в задачах оптической обработки информации. Освещены проблемы дискретизации и квантования в дифракционной оптике и особенности применения различных технологий создания фазового микрорельефа.

Для студентов старших курсов специальностей: прикладные математика и физика, информатика и оптика, а также для аспирантов и специалистов, работающих в соответствующих областях.

Табл. 59. Ил. 511.

ISBN 5–9221–0434–9

© ФИЗМАТЛИТ, 2003
© Коллектив авторов, 2003

Оглавление

Список основных обозначений 8

Предисловие 9

Глава 1. Введение в компьютерную оптику

1.1. Функциональные возможности зонированных дифракционных оптических элементов (ДОЭ) 11
1.2. Границы зон и фазовые функции оптических элементов 15
1.3. ДОЭ, реализуемые методами цифровой голографии 20
1.4. Геометрооптический расчет ДОЭ 27
1.5. Дискретизация и квантование фазы ДОЭ 33
1.6. Компьютерное проектирование ДОЭ 43
1.7. Заключение 46
Литература к главе 1 47

Глава 2. Итеративные методы расчета ДОЭ

2.1. Введение 49
2.2. Алгоритм уменьшения ошибки 50
2.3. Алгоритм входа-выхода 52
2.4. Адаптивно-аддитивный алгоритм 53
2.5. Адаптивно-мультипликативный алгоритм 60
2.6. Адаптивно-регуляризационный алгоритм 64
2.7. Градиентный алгоритм расчета фазы ДОЭ 67
2.8. Применение итеративных алгоритмов для расчета ДОЭ 71
2.8.1. Расчет ДОЭ, фокусирующих в радиально-симметричные области фурье-спектра 71
2.8.2. Расчет дифракционных аксиконов, формирующих осевые световые отрезки 75
2.8.3. Расчет радиально симметричных ДОЭ с квантованной фазой 77
2.8.4. Многопорядковые фазовые бинарные дифракционные решетки . . . 81
2.8.5. Многоуровневые фазовые дифракционные решетки 87
2.8.6. Фазовые ДОЭ, фокусирующие в объем и на поверхность тел вращения 93
2.8.7. Фокусировка гауссова пучка в квадрат 103
2.8.8. Фокусировка в кольцо 110
2.8.9. Композиционные ДОЭ, формирующие контурные изображения . . . 116
2.8.10. Квантованные ДОЭ для фокусировки в заданную двумерную область 122
2.8.11. Квантованные ДОЭ для формирования амплитудно-фазовых распределений 134

2.9. Заключение 137
Литература к главе 2 138

Глава 3. Электромагнитный подход к расчету дифракционных оптических элементов
3.1. Дифракция на отражающих решетках со ступенчатым профилем . . 142
3.2. Дифракция на отражающих решетках с непрерывным профилем . . . 149
3.3. Дифракция на пропускающих диэлектрических решетках 158
3.4. Градиентные методы решения обратной задачи расчета дифракционных решеток 175
3.4.1. Расчет отражающих решеток со ступенчатым профилем . 175
3.4.2. Расчет диэлектрических бинарных решеток 178
3.4.3. Расчет отражающих решеток с непрерывным профилем в приближении Рэлея 188
3.5. Дифракция на отражающих двумерных структурах 192
3.5.1. Дифракция света на криволинейной зоне 192
3.5.2. Дифракция на двумерной отражающей бинарной решетке . . 199
3.5.3. Дифракция на двумерных пропускающих диэлектрических структурах 201
3.6. Градиентный метод синтеза ДОЭ 210
3.7. Асимптотический анализ дифракции на зонированных структурах . . 212
3.7.1. Решение задачи дифракции на одномерных ДОЭ в скалярном приближении 212
3.7.2. Решение задачи дифракции одномерных ДОЭ в рамках электромагнитной теории 219
3.8. Моделирование распространения электромагнитного излучения методом конечных разностей 223
3.9. Анализ прохождения электромагнитного импульса через антиотражающую структуру 231
3.10. Заключение 235
Литература к главе 3 236

Глава 4. Технология создания ДОЭ
4.1. Типы фазовых микрорельефов и способы их получения 238
4.2. Изготовление ДОЭ методами фотолитографии 243
4.2.1. Изготовление фотошаблонов 243
4.2.2. Экспонирование и проявление пленок фоторезиста 250
4.2.3. Технология формирования микрорельефа ДОЭ 253
4.3. Изготовление ДОЭ методами электронно-лучевой литографии . . . 262
4.4. Формирование кусочно-непрерывного микрорельефа 268
4.4.1. Методы формирования кусочно-непрерывного микрорельефа . . . 268
4.4.2. Механизм рельефообразования на ЖФПК 270
4.4.3. Определение оптимальной оптической плотности фотошаблона при рельефной записи на слоях ЖФПК 273
4.4.4. Изготовление опытных образцов оптических элементов видимого диапазона 275
4.5. Технология травления 278
4.6. Формирование дифракционного рельефа с помощью лазерного микроструктурирования алмазных пленок 281
4.7. Копирование микрорельефа ДОЭ 287
4.8. Автоматизация экспериментальных исследований и технологических испытаний ДОЭ 289

4.8.1. Операции и оборудование, используемые для контроля процесса создания ДОЭ . . . 289
4.8.2. Сканирующий зондовый микроскоп в исследовании ДОЭ . . . 291
4.8.3. Средства автоматизации испытаний ДОЭ . . . 293
4.9. Примеры синтеза ДОЭ и использования программных комплексов . . 296
4.9.1. Программное обеспечение по компьютерной оптике . . . 296
4.9.2. Примеры синтеза ДОЭ . . . 300
4.10. Заключение . . . 305
Литература к главе 4 . . . 306

Глава 5. Фокусаторы
5.1. Введение . . . 310
5.2. Геометрооптический расчет фокусаторов в линию . . . 311
5.3. Расчет и исследование геометрооптических фокусаторов . . . 319
5.3.1. Дифракционная линза . . . 320
5.3.2. Дифракционная цилиндрическая линза . . . 325
5.3.3. Фокусатор в кольцо . . . 327
5.3.4. Фокусатор в полукольцо . . . 334
5.3.5. Фокусатор в поперечный отрезок . . . 336
5.3.6. Составной фокусатор в крест . . . 345
5.3.7. Фокусатор в продольный отрезок . . . 347
5.4. Фокусаторы в двумерную область. Метод согласованных прямоугольников . . . 350
5.5. Многофокусные фокусаторы . . . 356
5.5.1. Многофокусные бинарные зонные пластинки . . . 364
5.5.2. ДОЭ с нелинейно комбинированными фазами . . . 368
5.6. Дифракционные многофокусные линзы . . . 370
5.7. Двухпорядковые фокусаторы . . . 375
5.8. Расчет спектральных фокусаторов . . . 378
5.8.1. Цветоделительные решетки . . . 378
5.8.2. Спектральные фокусаторы в набор одинаковых фокальных областей . . . 381
5.8.3. Спектральные фокусаторы в различные фокальные области . . . 384
5.8.4. Расчет квантованных спектральных ДОЭ . . . 386
5.9. Заключение . . . 391
Литература к главе 5 . . . 391

Глава 6. Селекция мод лазерного излучения
6.1. Моды лазерного излучения . . . 394
6.1.1. Модовые пучки в скалярном приближении . . . 394
6.1.2. Возбуждение мод в оптических волокнах и резонаторах . . . 397
6.1.3. Метод комплексного эйконала . . . 399
6.1.4. Амплитудно-фазовые соотношения для модовых пучков в свободном пространстве . . . 402
6.1.5. Гауссовы и бесселевы моды . . . 408
6.1.6. Гауссовы моды в пассивных резонаторах и градиентных волокнах . . . 413
6.2. Формирование и селекция мод лазерного излучения с помощью ДОЭ 414
6.2.1. Постановка задачи синтеза ДОЭ, согласованных с модами лазерного излучения . . . 414
6.2.2. Методы расчета фазовых моданов . . . 417
6.2.3. Расчет одномодовых моданов . . . 417
6.2.4. Построение итеративной процедуры расчета одномодового модана . 419

6.2.5. Быстрый расчет ДОЭ, формирующего заданное одномодовое распределение радиальных мод . . . 422
6.2.6. Расчет ДОЭ, формирующих группу мод лазерного излучения . . 428
6.2.7. Организация многоканальной связи в идеальной линзоподобной среде с минимальными потерями энергии . . . 434
6.2.8. Расчет ДОЭ для анализа поперечно-модового состава пучков когерентного излучения . . . 440
6.2.9. Результаты экспериментального исследования фундаментальных свойств гауссовых мод с помощью ДОЭ . . . 443
6.2.10. Натурное исследование возможности уплотнения каналов оптической связи с помощью селективного возбуждения мод Гаусса–Эрмита . . . 449
6.2.11. Расчет ДОЭ, согласованных с модами градиентных волокон с непараболическим профилем . . . 453
6.3. Применение ДОЭ в системах сбора, передачи и хранения информации 455
6.3.1. Повышение пропускной способности систем оптической связи . . . 455
6.3.2. Волоконно-оптические датчики . . . 458
6.3.3. Экспериментальное исследование датчика микроперемещений, основанного на использовании модана . . . 460
6.3.4. Применение ДОЭ для коллимации излучения полупроводникового лазера . . . 463
6.3.5. Дифракционные делители пучка . . . 464
6.4. Заключение . . . 465
Литература к главе 6 . . . 466

Глава 7. Световые пучки с периодическими свойствами
7.1. Некоторые типы световых пучков с периодическими свойствами . . . 469
7.2. Фазовые формирователи световых полей с продольной периодичностью . . . 475
7.3. Алгоритм расчета ДОЭ для генерации вращающихся многомодовых пучков Бесселя . . . 484
7.4. Формирование пары вращающихся бездифракционных пучков с помощью бинарного фазового ДОЭ . . . 488
7.5. ДОЭ для формирования многомодовых пучков Гаусса–Лагерра . . . 495
7.6. Вращение световых многомодовых пучков Гаусса–Лагерра в свободном пространстве и волокне . . . 504
7.7. Формирование вращающихся пучков Гаусса–Лагерра с помощью фазовой бинарной дифракционной оптики . . . 509
7.8. Обобщенные эрмитовы световые пучки в свободном пространстве . . 517
7.9. Формирование мод Гаусса–Эрмита с помощью бинарных ДОЭ . . . 525
7.10. Самовоспроизведение многомодовых пучков Гаусса–Эрмита . . . 532
7.11. Заключение . . . 538
Литература к главе 7 . . . 539

Глава 8. Коррекция волновых фронтов
8.1. Проблема создания волновых фронтов . . . 541
8.2. Оптические схемы с ДОЭ для анализа асферических поверхностей . 542
8.3. Расчет плоского компенсатора . . . 544
8.4. Спектральные свойства компенсаторов . . . 545
8.5. Характеристика точности эталонного волнового фронта . . . 547
8.6. Влияние дискретизации и квантования фазовой функции компенсатора на точность эталонного волнового фронта . . . 550

8.7. Формирование волновых фронтов с малым относительным отверстием 552
8.8. Осесимметричные компенсаторы 554
8.9. Формирование волновых фронтов высших порядков 556
8.10. Формирование неосесимметричных волновых фронтов 558
8.11. Формирование внеосевых сегментов волновых фронтов вращения . . 560
8.12. Формирование волновых фронтов с заданным распределением интенсивности 564
8.13. Итеративные алгоритмы для расчета ДОЭ, формирующих заданные распределения фазы 568
8.14. Практическое использование 576
8.15. Заключение 577
Литература к главе 8 577

Глава 9. Применение ДОЭ в светотехнических устройствах
9.1. Перспективы использования в светотехнических устройствах 579
9.2. Средства проектирования светотехнических устройств с ДОЭ . . . 580
9.3. Моделирование устройств с квантованными ДОЭ 584
9.4. Проектирование автомобильных фар с ДОЭ 585
9.5. Проектирование компланарных осветителей 587
9.6. Результаты испытаний светотехнических устройств с ДОЭ 592
9.7. Заключение 595
Литература к главе 9 596

Глава 10. Оптическая обработка информации с применением ДОЭ
10.1. Оптическое формирование признаков изображения 597
10.2. Разложение светового поля по ортогональному базису 599
10.2.1. Оптимальный базис Карунена–Лоэва 599
10.2.2. ДОЭ для разложения функции интенсивности по базису Адамара . . 618
10.2.3. ДОЭ для разложения амплитуды волны с угловыми гармониками 622
10.2.4. ДОЭ для разложения пучка по базису Цернике 629
10.3. Оптическое построение поля направлений и поля пространственных частот 637
10.3.1. Оптическое распознавание дактилограмм 647
10.3.2. Оптическая расшифровка интерферограмм. 667
10.4. Оптическое выполнение преобразования Хоу–Радона 673
10.5. Заключение 676
Литература к главе 10 681

Заключение 684

СПИСОК ОСНОВНЫХ ОБОЗНАЧЕНИЙ

k	—	волновое число
λ	—	длина волны
ω	—	циклическая частота волны
$\varphi(x,y)$,	—	фаза дифракционного оптического элемента (ДОЭ)
$\psi(\mathbf{u})$	—	эйконал светового поля или оптического элемента
$\Gamma(\xi)$	—	слой рельефа фокусатора
$I(x,y)$	—	интенсивность волны
$F(x,y)$	—	комплексная амплитуда волны
δ	—	среднеквадратичная ошибка
E	—	энергетическая эффективность ДОЭ
f	—	фокусное расстояние тонкой линзы
δF	—	вариация функции или функционала
$\mathcal{F}$, $\mathcal{F}^{-1}$,	—	прямое и обратное фурье-преобразования
int	—	целая часть числа
$\lvert A\rvert$	—	модуль комплексного числа A
$\arg A$	—	фаза комплексного числа A
A^*	—	комплексное сопряжение числа A
Re, Im	—	действительная и мнимая части комплексного числа
(x,y,z)	—	декартовы координаты
(r,φ)	—	полярные координаты
$\delta(x)$	—	обобщенная дельта функция Дирака
δ_{mn}	—	символ Кронекера
d	—	период дифракционной решетки
$\mathbf{E}$, $\mathbf{H}$	—	векторы напряженности электрического и магнитного полей
$\mathbf{D}$, $\mathbf{B}$	—	векторы электрического смещения и магнитной индукции
$\mathbf{j}$	—	вектор плотности электрического тока
$[\mathbf{a},\mathbf{b}]$	—	векторное произведение векторов
$(\mathbf{a},\mathbf{b})$	—	скалярное произведение векторов
T_n	—	коэффициент пропускания дифракционной решетки
R_n	—	коэффициент отражения дифракционной решетки
c	—	скорость света
$h(x,y)$	—	высота микрорельефа ДОЭ
∇^2	—	оператор Лапласа
ψ_{pl}	—	мода лазерного излучения
$L_n(x)$	—	многочлен Лагерра
$L_n^m(x)$	—	обобщенный многочлен Лагерра
$H_p(x)$	—	многочлен Эрмита
$H_n^m(x)$	—	обобщенный многочлен Эрмита
$J_m(x)$	—	функция Бесселя
$N_m(x)$	—	функция Неймана
S/N	—	отношение сигнал/шум
$\mathrm{had}_m(x)$	—	функция Адамара
$\mathrm{rad}_m(x)$	—	функция Радемахера

ПРЕДИСЛОВИЕ

Предметом рассмотрения этой книги является новая область знаний, терминология в которой еще до конца не установилась. Разные авторы используют различные названия: «компьютерная оптика», «оптика, синтезированная компьютером», «бинарная оптика», «плоская оптика», «дифракционная оптика» и т.д. В любом случае речь идет об оптических элементах, различные отличительные особенности которых раскрываются перечисленными названиями.

Внешне такой оптический элемент представляет собой пропускающую или отражающую пластинку с тонким фазовым микрорельефом, рассчитанным в рамках теории дифракции. Первым представителем этого класса оптических элементов является дифракционная решетка, созданная более двухсот лет тому назад. Следующим по хронологии представителем указанного класса оптических элементов является зонная пластинка. Эти представители дифракционных оптических элементов (ДОЭ) имели бинарное амплитудное или фазовое пропускание. Если дифракционные решетки нашли широкое применение в приборостроении, то зонные пластинки, в основном, использовались в учебном лабораторном практикуме по оптике.

Настоящую революцию в создании ДОЭ произвело применение компьютеров. В 70-х годах компьютеры стали применять для синтеза голограмм математически заданных объектов. При этом была решена ключевая проблема кодирования — записи на физическую среду комплексных функций и получения соответствующих амплитудно-фазовых транспарантов. Это послужило толчком к компьютерному синтезу бинарных амплитудно-фазовых и фазовых пространственных фильтров, для записи которых на физическую среду применялись графопостроители, а затем и фотопостроители. Прогресс в области микроэлектроники и лазерной техники, приведший к появлению в конце 80-х годов прецизионных фотопостроителей и электронных литографов, сделал практически реализуемой задачу создания ДОЭ со сложным профилем зон. Возникло широкое поле деятельности для физиков, оптиков, прикладных математиков, электроников, технологов, специалистов по автоматизации. Появились ДОЭ с уникальными характеристиками, не достижимыми в рамках традиционной оптики, например, фокусаторы лазерного излучения и моданы.

Изложение построено так, что для первоначального знакомства с предметом можно ограничиться чтением главы 1, а далее читать отдельные главы по выбору, в соответствии с профессиональными интересами.

Книга написана сотрудниками Института систем обработки изображений РАН: глава 1 — В.А. Сойфером, глава 2 — Л.Л. Досколовичем, В.В. Котляром, В.А. Сойфером, разделы 3.1–3.4 — Л.Л. Досколовичем, разделы 3.5–3.7 — С.И. Харитоновым, раздел 3.8 — Д.Л. Головашкиным, раздел 3.9 — Д.Л. Головашкиным и В.С. Павельевым, глава 4 — А.В. Волковым, Н.Л. Казанским, В.А. Сойфером, Г.В. Успленьевым, за исключением раздела 4.4, который написал В.С. Соловьев, и 4.6, написанного Д.Л. Головашкиным и В.С. Павельевым, глава 5 — Л.Л. Досколовичем, Н.Л. Казанским, В.А. Сойфером, глава 6 — В.С. Павельевым, В.А. Сойфером, глава 7 —

В.В. Котляром, В.А. Сойфером, С.Н. Хониной, глава 8 — Н.Л. Казанским, В.В. Котляром, В.А. Сойфером, глава 9 — Н.Л. Казанским, глава 10 — В.В. Котляром, В.А. Сойфером, Р.В. Скидановым, С.Н. Хониной.

Авторы выражают благодарность:

— сотрудникам лаборатории дифракционной оптики и лаборатории лазерных измерений Института систем обработки изображений РАН и персонально Волотовскому С.Г. и Серафимовичу П.Г. за проведение большого объёма компьютерных и натурных экспериментов;

— сотрудникам Института общей физики РАН профессору В.И. Конову и В.В. Кононенко за материалы, любезно предоставленные для пункта 4.6 данной книги.

— российско-американской программе «Фундаментальные исследования и высшее образование» («BRHE»).

— зарубежным партнёрам из Исследовательского центра ФИАТ (группе доктора Перло П.), Института прикладной оптики Фридрих–Шиллер Университета, Йена, Германия (группе профессора Коваршика Р. и доктору Дюпарре М.), университета Йонсуу, Финляндия (группе профессора Турунена Я.) за помощь в изготовлении образцов ДОЭ и совместное проведение экспериментальных исследований.

— за оформительскую работу Каркану А.В., Козлову М.Д., Котляр М.И., Смагину С.В. и Тахтарову Я.Е.

Книга базируется на оригинальных научных работах, опубликованных за последние 20 лет членами авторского коллектива; соавторами ряда работ были: академик Прохоров А.М., профессор Сисакян И.Н., профессор Голуб М.А., Данилов В.А., доцент Карпеев С.В.

ГЛАВА 1

ВВЕДЕНИЕ В КОМПЬЮТЕРНУЮ ОПТИКУ

1.1. Функциональные возможности зонированных дифракционных оптических элементов (ДОЭ)

Первые ДОЭ — дифракционные решетки — были созданы более двух веков тому назад. Амплитудная одномерная дифракционная решетка представляет собой плоский транспарант, на котором чередуются светлые и темные параллельные полоски (штрихи и щели). Полоски имеют одинаковую ширину. На рис. 1.1 показан фрагмент одномерной амплитудной дифракционной решетки, а на рис. 1.2 показана функция пропускания света вдоль решетки по оси x.

Если осветить такую решетку монохроматическим пучком света с длиной волны λ, падающим нормально к плоскости решетки, то в результате дифракции света на периодической структуре щелей будет образовано множество пучков света, выходящих под различными углами α^p и соответствующих различным порядкам дифракции. Угол α^p зависит от периода решетки и в предположении малости углов определяется по формуле

$$\alpha^p = p\frac{\lambda}{d}, \tag{1.1}$$

где $p = 0, \pm 1, \pm 2, \ldots$, d — период решетки.

Интенсивность света убывает с ростом абсолютной величины p. Как следует из теории дифракционных решеток, при равной ширине штриха и щели характер убывания определяется формулой

$$I(\alpha^p) \sim \left(\frac{\sin(a\alpha^p)}{a\alpha^p}\right)^2 \approx \frac{1}{(2p+1)^2}, \tag{1.2}$$

где $a = \pi d/2\lambda$.

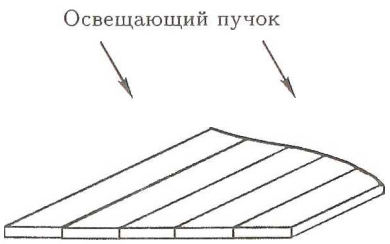


Рис. 1.1. Фрагмент бинарной амплитудной дифракционной решетки

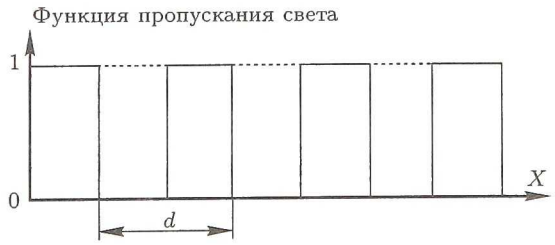


Рис. 1.2. Функция пропускания амплитудной дифракционной решетки

Из-за того, что значительная часть света поглощается амплитудным транспарантом, дифракционная эффективность в первом порядке не превышает 10%. У одномерной (1-D) фазовой дифракционной решетки этот показатель приблизительно в 4 раза выше. Фрагмент одномерной фазовой дифракционной решетки схематично показан на рис. 1.3.

Функция пропускания в данном случае является чисто фазовой, периодической с периодом d и скачкообразно меняющейся на π (см рис. 1.4). Функционально фазовая и амплитудная дифракционные решетки одинаковы.

Естественным развитием одномерных бинарных амплитудных и фазовых дифракционных решеток являются соответствующие двумерные (2-D) радиально-симметричные решетки (зонные пластинки).

На рис. 1.5 показан центральный фрагмент зонной пластинки Рэлея–Сорэ, а на рис. 1.6 — функция пропускания в зависимости от радиуса.

Поясним понятие зоны. Зона — это ограниченная область ДОЭ, на которой функция пропускания света претерпевает однократное изменение от минимального до максимального значения. Для амплитудной дифракционной решетки зона представляет собой сочетание темной и светлой полос в пределах одного периода решетки (рис. 1.1). Граница зон при этом — прямая линия. Для зонной пластинки Рэлея–Сорэ [1] зона представляет собой совокупность темного и светлого колец переменной толщины (рис. 1.5).

Границы зон при этом — окружности переменного диаметра. Далее будут рассмотрены ДОЭ с более сложными зонами и, соответственно, границами зон.

Радиусы окружности меняются пропорционально квадратным корням из последовательных целых чисел p:

$$\rho_p = \sqrt{p\lambda f}, \tag{1.3}$$

где f — фокусное расстояние.

Фрагмент 2-D фазовой зонной пластинки и соответствующая фаза функции пропускания изображены на рис. 1.7 и 1.8.

Зонная пластинка выполняет функцию фокусировки света и ведет себя как линза с множеством фокусов $f_{-2}, f_{-1}, f_0, f_1, f_2$, соответствующих различным порядкам дифракции и расположенных на оси распространения света:

$$f_n = \frac{f}{2n+1}, \quad n = 0, \pm 1, \pm 2, \ldots \tag{1.4}$$

Нормированная интенсивность света убывает с ростом номера порядка по закону

$$I_{2n+1} = \frac{4}{\pi^2}\frac{1}{(2n+1)^2}. \tag{1.5}$$



Рис. 1.3. Фрагмент бинарной фазовой дифракционной решетки

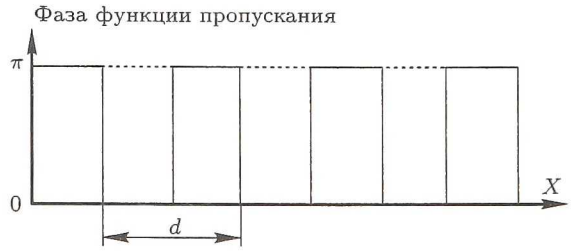


Рис. 1.4. Фаза функции пропускания фазовой дифракционной решетки



Рис. 1.5. Центральный фрагмент зонной пластинки Рэлея–Сорэ

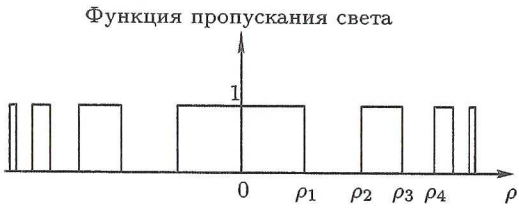


Рис. 1.6. Функция пропускания амплитудной зонной пластинки

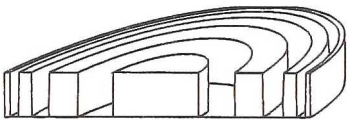


Рис. 1.7. Фрагмент бинарной фазовой зонной пластинки

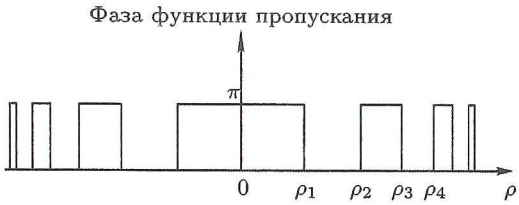


Рис. 1.8. Фаза функции пропускания бинарной фазовой зонной пластинки

Совершенствование «базовых» ДОЭ, какими являются 1-D дифракционные решетки, возможно за счет изменения профиля зон. При этом получаются так называемые решетки «с блеском». Например, можно рассмотреть амплитудную 1-D дифракционную решетку, функция пропускания которой меняется по закону $\cos^2(2\pi x/d)$.

В теории, при дифракции света на такой решетке, помимо нулевого образуются только ± 1 порядки дифракции. На практике, с учетом погрешностей изготовления ДОЭ, дело обстоит гораздо сложнее.

Для 2-D дифракционных решеток естественным является вопрос построения зон, отличных от периодических щелеобразных и кольцеобразных, и достижения на этой основе новых функциональных возможностей ДОЭ. Несмотря на кажущуюся очевидность такой постановки вопроса, с момента создания зонной пластинки он не поднимался в течение многих лет. Причина состояла в отсутствии практической возможности реализации зонных пластинок с варьируемым характером зон. Такая возможность появилась в связи с созданием компьютеров и технических средств машинной графики, в частности, графопостроителей и фотопостроителей.

В 1967 г. были предложены 2-D дифракционные решетки с варьируемыми зонами Френеля, получаемые на основе муаровых эффектов в результате суперпозиции пары синтезированных масок [2]. Синтезированные маски содержат периодические по одной из координат (например, по X) бинарные изображения, описываемые специально подобранной алгебраической функцией $\varphi(x, y)$. Варьируя вид и параметры функции $\varphi(x, y)$, а также параметры относительно сдвига Δx и поворота масок, можно получить множество амплитудных 2-D дифракционных решеток, каждой из которых будет соответствовать своя картина дифракции. К сожалению, такая процедура содержит регулярный метод синтеза только для цилиндрических и сферических зон Френеля. (Пример цилиндрической амплитудной зонной пластинки приведен на рис. 1.9). Результат фокусировки представляет собой прямую фокальную линию. Имеется также возможность построения конических зонных пластинок путем оптической регистрации результата физической суперпозиции цилиндрической и плоской волн. Результат фокусировки представляет собой наклонную прямую линию в плоскости, параллельной зонной пластинке.

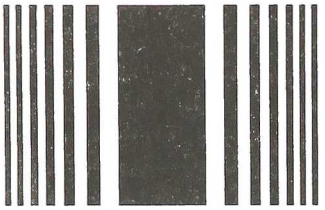


Рис. 1.9. Бинарная цилиндрическая зонная пластинка

Рассмотренные примеры демонстрируют возможность получения зонных пластинок с различной формой зон. В то же время представляет интерес построение 2-D бинарных образов зон Френеля для более сложных случаев. Например, для зонной пластинки, фокусирующей в продольный или поперечный отрезок заданной длины, кольцо или какую-либо другую геометрическую фигуру. ДОЭ такого рода получили название «фокусаторы». Рассмотрим для примера фокусатор в кольцо. Соответствующий образ зон Френеля решетки может быть получен путем комбинации 1-D дифракционной решетки и зонной пластинки. Возьмем достаточно узкий сегмент 1-D дифракционной решетки (рис. 1.10), который ведет себя так же, как целая дифракционная решетка, т.е. отклоняет входной монохроматический пучок на определенный угол в плоскости (рассматривается 1-й порядок дифракции).

Вращая сегмент вокруг центра, как показано на рис. 1.10, получим ДОЭ, зоны Френеля которого отображаются системой равноотстоящих концентрических черно-белых колец одинаковой ширины (рис. 1.11). Назовем его дифракционным аксиконом. Нетрудно сообразить, что такой ДОЭ будет отклонять входной пучок на определенный телесный угол в пространстве. Для того, чтобы собрать дифрагированные пучки в фокальной плоскости, применим зонную пластинку, причем дифракционный аксикон и зонная пластинка могут быть совмещены в одной плоскости и записаны на одной подложке. При этом в плоскости ДОЭ образуется система концентрических неравномерно отстоящих от центра колец переменной ширины (рис. 1.12). Таким образом, мы синтезировали амплитудный ДОЭ, фокусирующий в кольцо.

Следуя описанной методике, можно построить различные фокусирующие ДОЭ. Например, записывая на одной подложке две скрещенные цилиндрические линзы с разными фокусными расстояниями, можно получить ДОЭ, фокусирующий в отрезок, лежащий в плоскости, перпендикулярной оси распространения (рис. 1.13).

Видимо, нет особой необходимости говорить, что и фокусатор в кольцо и фокусатор в поперечный отрезок могут быть реализованы как фазовые ДОЭ. Достичь этого в простейшем случае можно путем фотографического отбеливания соответствующих амплитудных масок (фотошаблонов).

По существу в двух только что рассмотренных примерах речь идет о фазовой модуляции, при которой одна бинарная функция, описывающая дифракционную решетку, умножается на другую. Однако, процесс модуляции параметров дифракционной решетки может осуществляться и при более широком классе модулирующих функций. В частности, осуществляя модуляцию радиусов колец зонной пластинки, можно построить фокусатор в соосный отрезок определенной длины, который



Рис. 1.10. Узкий фрагмент 1-D дифракционной решетки

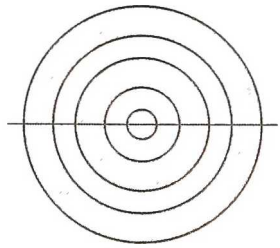


Рис. 1.11. Аксикон

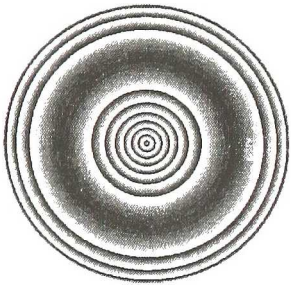


Рис. 1.12. Фокусатор в кольцо

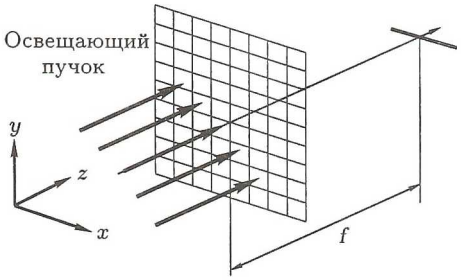


Рис. 1.13. Результат фокусировки 2-мя скрещенными цилиндрическими линзами

фактически будет представлять собой зонную пластинку с протяженной продольной аберрацией. Идея получения соответствующего радиально-симметричного ДОЭ состоит в том, чтобы его периферийная часть соответствовала зонной пластинке с фокусным расстоянием F_1, а центральная часть — зонной пластинке с расстоянием $F_2 > F_1$. Между центральной и периферийной частями должны быть записаны кольца, соответствующие зонным пластинкам, фокусное расстояние которых уменьшается от F_2 к F_1. В результате будет получен фокусатор в соосный отрезок длины $l = F_2 - F_1$. Фотошаблон такого фокусатора показан на рис. 1.14, а на рис. 1.15 показан процесс фокусировки.

Общие методы получения модулированных дифракционных решеток, обладающих заданными функциональными свойствами, требуют решения обратных задач теории дифракции и будут рассмотрены в последующих главах. Заметим лишь, что решение обратной задачи дифракции осуществляется относительно фазовой функции ДОЭ, который выполняет требуемое преобразование светового пучка. Эта фазовая функция записывается на оптическую среду в виде регулярной зонированной структуры с фазовым микрорельефом, в результате чего и получается ДОЭ.

1.2. Границы зон и фазовые функции оптических элементов

Приведение функции к интервалу. Фазовой функцией $\varphi(u, v)$ оптического элемента называется функция, описывающая зависимость от пространственных координат величины набега фазы, создаваемой оптическим элементом в точке (u, v). В ряде случаев границы зон ДОЭ могут быть получены, исходя из фазовых функций соответствующих традиционных (рефракционных) оптических элементов.

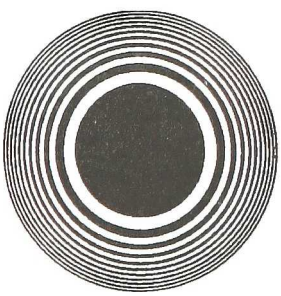


Рис. 1.14. Фокусатор в соосный отрезок

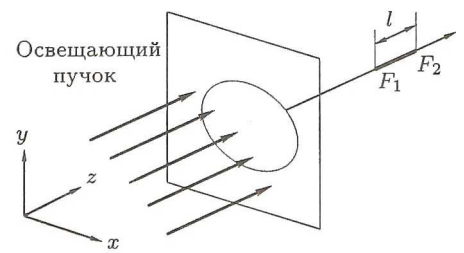


Рис. 1.15. Результат фокусировки в соосный отрезок

Традиционный оптический элемент характеризуется гладкой фазовой функцией $\varphi(\mathbf{u})$, где $\mathbf{u} = (u, v)$ — поперечные координаты точки, лежащей в плоскости оптического элемента. При освещении пучком с комплексной амплитудой $W_0(\mathbf{u})$ непосредственно за оптическим элементом образуется поле с амплитудой

$$W(\mathbf{u}) = \exp[i\varphi(\mathbf{u})]W_0(\mathbf{u}). \tag{1.6}$$

На рис. 1.16 изображена фазовая функция сферической линзы.

Толщина традиционного оптического элемента H может составлять десятки тысяч длин волн. При этом значения фазы φ лежат в интервале от 0 до тысяч единиц 2π. Изменения же фазы на рассмотренных ДОЭ лежат в диапазоне от 0 до 2π.

С учетом того, что комплексная экспонента в выражении (1.6) является периодической функцией с периодом 2π, фаза может быть приведена к интервалу $[0, 2\pi]$. Приведение к интервалу можно выполнить по формуле

$$\Phi = \mathrm{mod}_{2\pi}\varphi, \tag{1.7}$$

где $\mathrm{mod}_{2\pi}\varphi = \varphi - 2\pi j$ при $j2\pi \leqslant \varphi \leqslant (j+1)2\pi, \quad j = 0, \pm 1, \pm 2, \ldots$

Из практических соображений может оказаться целесообразным приводить фазу к интервалу $[0,\ 2\pi m)$ по формулам

$$\Phi = \mathrm{mod}_{2\pi m}\varphi, \tag{1.8}$$

где обычно $m \sim 1 - 10^2$. Эквивалентность исходной гладкой фазы φ и приведенной к интервалу кусочной фазы Φ выражается соотношением

$$T \equiv \exp(i\Phi) = \exp(i\varphi). \tag{1.9}$$

Плоская сферическая линза. Уравнение фазовой функции сферической линзы в параксиальном приближении имеет вид

$$\varphi(u, v) = \varphi(r) = -k\frac{u^2 + v^2}{2f}, \tag{1.10}$$

где $r \leqslant D/2$, $k = 2\pi/\lambda$ — волновое число; f — фокусное расстояние; D — диаметр линзы; $r = \sqrt{u^2 + v^2}$.

Приведение фазовой функции (1.10) к интервалу $[0,2\pi)$ показано на рис. 1.17. Если материал линзы имеет коэффициент преломления n, то максимальная высота рельефа имеет порядок длины волны и составляет

$$h_{\max} = \frac{\lambda}{n-1}.$$

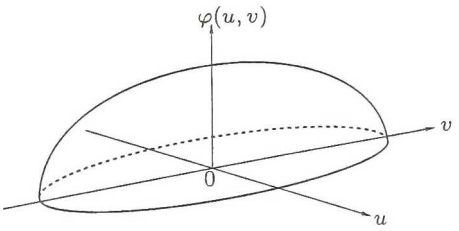


Рис. 1.16. Фазовая функция сферической линзы

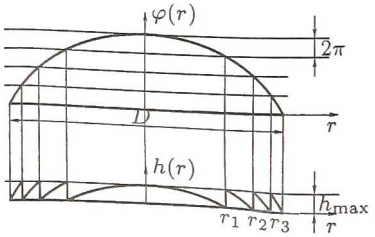


Рис. 1.17. Приведение фазовой функции линзы к интервалу $[0, 2\pi)$

Радиусы зон Френеля можно найти из соотношения

$$\varphi(r_j) = -2\pi j, \tag{1.11}$$

откуда следует, что

$$r_j = \sqrt{2\lambda f j}. \tag{1.12}$$

Число полных зон F на линзе определяется из условия $r_{j_0} \leqslant D/2$ и удовлетворяет соотношению

$$F = \left]\frac{D^2}{8\lambda f}\right[, \tag{1.13}$$

где $]\cdot[$ — означают целую часть числа с округлением в меньшую сторону.

Ширина зон плоской сферической линзы $\Delta_j = r_j - r_{j-1}$, $j = \overline{1,F}$, является переменной и уменьшается к периферии линзы. Характерным параметром является ширина Δ самой узкой, в данном случае последней, периферийной зоны, определяющей требования к технологическому оборудованию. В более общем случае ДОЭ с фазовой функцией $\varphi(u,v)$ можно определить ширину наиболее узкой зоны микрорельефа с максимальной фазовой высотой $2\pi m$. Используя в пределах этой зоны линейную аппроксимацию фазовой функции, получим выражение

$$\Delta = \frac{2\pi m}{\max|\nabla_\perp \varphi|}, \tag{1.14}$$

где max берется вдоль поверхности ДОЭ, $\nabla_\perp\varphi$ — градиент функции $\varphi(u,v)$.

Сравним между собой уравнения радиусов колец зонной пластинки Рэлея–Сорэ (1.3) и сферической плоской линзы (1.12). Видим, что радиусы зон на линзе в $\sqrt{2}$ раз больше радиусов колец на зонной пластинке. Соответственно, при заданном диаметре число колец на зонной пластинке в 2 раза больше, т.е. каждому кольцу сферической плоской линзы соответствует 2 кольца зонной пластины: темное и светлое. Плоская линза имеет непрерывный (в пределах зоны) фазовый рельеф, а фазовая зонная пластинка имеет аппроксимирующий его бинарный фазовый рельеф (см. рис. 1.18).

Если подходить к созданию ДОЭ с общих позиций дискретной аппроксимации непрерывной фазовой функции, можно продолжить процедуру ступенчатой аппроксимации в пределах зоны Френеля исходной непрерывной фазовой функции. На рис. 1.19 показан пример квантования фазы плоской сферической линзы Френеля на 4 уровня. На рис. 1.20 показана соответствующая фазовая функция.

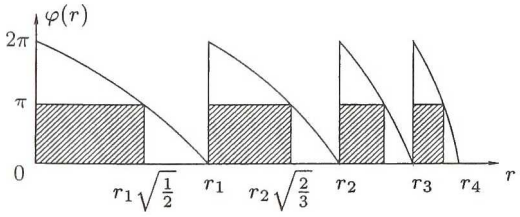


Рис. 1.18. Фазовая функция линзы Френеля и бинарной зонной пластинки (штриховые области)

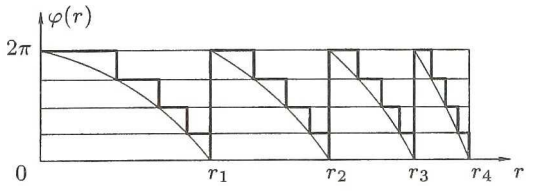


Рис. 1.19. Ступенчатая аппроксимация радиального сечения фазовой функции сферической линзы

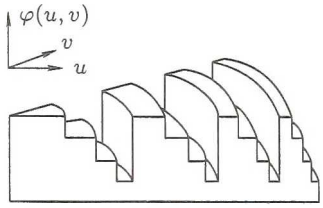


Рис. 1.20. Ступенчатая аппроксимация фазовой функции сферической линзы

Для практического получения амплитудной маски с N градациями яркости можно использовать $(N-1)$ бинарных амплитудных масок, осуществляя последовательный фотопроцесс. Существует также возможность уменьшить число бинарных масок до $n = \log_2 N$.

Нетрудно рассчитать радиусы колец, соответствующих градациям фазы в различных зонах. Знание их необходимо для расчета бинарных амплитудных масок, однако они не несут такого ясного физического смысла, как в случае зонной пластинки. Отметим также, что уравнение колец зонной пластинки Рэлея–Сорэ невозможно получить с помощью равномерной дискретизации фазовой функции плоской сферической линзы.

Плоская цилиндрическая линза. Рассмотрим цилиндрическую линзу (см. рис. 1.21), с фазовой функцией

$$\varphi(u) = -k\frac{u^2}{2f}, \quad |u| \leqslant \frac{D}{2}. \tag{1.15}$$

Приведение фазовой функции к интервалу $[0, 2\pi)$ аналогично проведенному выше для сферической линзы. Границы зон в данном случае — прямые линии, а расстояния между ними определяются формулой (1.12) при замене r_j на u_j. На рис. 1.22 приведен фотошаблон плоской цилиндрической линзы.

Отражательная зонная пластинка. Рассмотрим отражательную зонную пластинку, для работы с пучком, падающим под углом α (рис. 1.23 и рис. 1.24).

При нормальном падении максимальная высота рельефа отражательной зонной пластины должна составить $\lambda/2$ для того, чтобы набег фазы при прямом и обратном ходе составил 2π, при наклонном падении $h_{\max} = \lambda/(2\cos\alpha)$.

Уравнение границ зон отражательной зонной пластинки при нормальном падении в точности будет соответствовать уравнению границ зон пропускающей пластинки (высота зон отличается в $2/(n-1)$ раза).

Уравнение границ зон при наклонном падении можно получить путем простых рассуждений. Наклонное падение пучка эквивалентно введению поворотного зеркала. Спроецировав кольца зонной пластинки Рэлея–Сорэ на это зеркало, мы получим уравнение краев зон отражательной зонной пластинки при наклонном падении под углом α. Это будут эллипсы с коэффициентом растяжения по оси u, равном $\cos\alpha$ (рис. 1.25).

Высота рельефа отражательного ДОЭ, фокусирующего в точку под углом α, в приближении плоских волн, определяется из уравнения

$$h(u, v) = \frac{\lambda}{2\cos\alpha}\frac{1}{2\pi} \bmod {}_{2\pi}\varphi(u\cos\alpha, v). \tag{1.16}$$

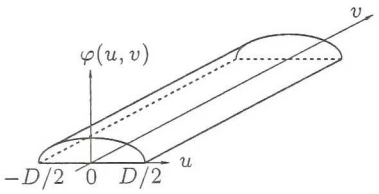


Рис. 1.21. Цилиндрическая линза

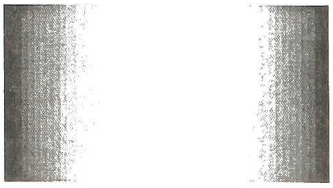


Рис. 1.22. Фотошаблон для цилиндрической линзы

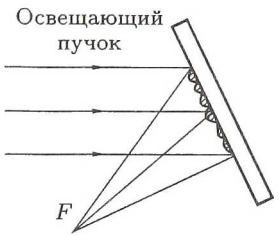


Рис. 1.23. Отражательная зонная пластинка

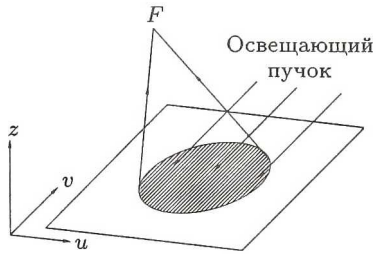


Рис. 1.24. Фокусировка в точку под углом

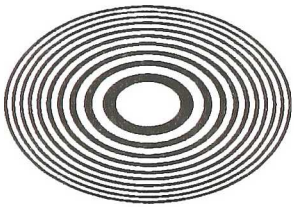


Рис. 1.25. Фотошаблон отражательной зонной пластинки с углом наклона $\alpha = 45°$

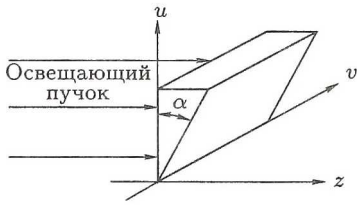


Рис. 1.26. Призма

Плоская призма. Призма с углом α (рис. 1.26) обеспечивает фазовый сдвиг, линейно зависящий от координаты, и характеризуется фазовой функцией

$$\varphi(u) = \frac{2\pi}{\lambda} u\,(n-1)\,\mathrm{tg}\,\alpha. \tag{1.17}$$

Приводя фазовую функцию $\varphi(u)$ к интервалу $[0,2\pi)$ (рис. 1.27), получим 1-D дифракционную решетку с блеском (рис. 1.28). Максимальная высота рельефа $h_{\max} = \lambda/(n-1)$, а уравнение высоты микрорельефа имеет вид

$$h(u) = \frac{\lambda}{n-1}\frac{1}{2\pi} \bmod {}_{2\pi}\varphi(u). \tag{1.18}$$

Различным углам отклонения β соответствуют различные периоды решетки:

$$d = \frac{\lambda}{\sin\beta}, \tag{1.19}$$

где $\sin\beta = (n-1)\,\mathrm{tg}\,\alpha$.

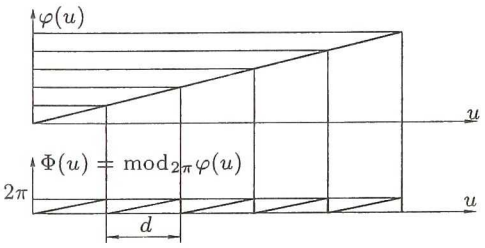


Рис. 1.27. Фазовая функция дифракционной решетки с блеском

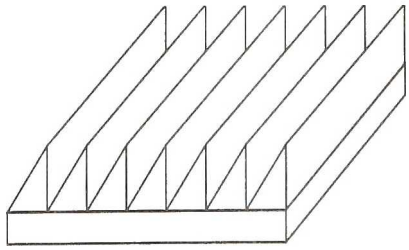


Рис. 1.28. Дифракционная решетка с блеском

Бинарная 1-D амплитудная дифракционная решетка получается при замене линейно-меняющейся фазовой функции в пределах одного периода на двоичную функцию, принимающую значения $\pm\pi/2$ (рис. 1.29).

Комбинированные и сегментированные ДОЭ. Если два ДОЭ, расположенных вплотную друг за другом, имеют фазовые функции φ_1 и φ_2, соответственно, то они могут быть заменены комбинированным ДОЭ с фазовой функцией $\varphi = \varphi_1 + \varphi_2$.

В качестве примера рассмотрим 2 скрещенные цилиндрические линзы с различными фокусными расстояниями:

$$\varphi_1(u,v) = -k\frac{u^2}{2f_1}, \quad \varphi_2(u,v) = -k\frac{v^2}{2f_2}. \tag{1.20}$$

Фотошаблон комбинированного ДОЭ, соответствующего фазовой функции

$$\varphi(u,v) = -k\left(\frac{u^2}{2f_1} + \frac{v^2}{2f_2}\right), \tag{1.21}$$

приведен на рис. 1.30. Границы краев зон являются эллиптическими. ДОЭ осуществляет фокусировку в отрезок. Плоскость сегментированного ДОЭ в пределах апертуры разбивается на 2 или большее число сегментов (рис. 1.31), а в каждом из сегментов записывается зонная пластинка, соответствующая определенной фазовой функции. Например, так можно получить фокусатор в букву «R» (рис. 1.32).

1.3. ДОЭ, реализуемые методами цифровой голографии

Голограммы Фурье и Френеля. Методы цифровой голографии начали разрабатываться в конце 60-х годов в связи с широким проникновением в оптику компьютеров и средств машинной графики. Появление компьютеров дало реальную возможность для численного расчета амплитудно-фазовых характеристик светового поля в плоскости элемента, исходя из характеристик восстанавливаемого объекта. Развитие средств машинной графики предоставило возможность для записи закодированных значений рассчитанной функции пропускания голографического элемента на физическом носителе. Методы цифровой голографии открыли возможность синтезировать голограммы объектов, заданных математически. В настоящее время количество публикаций по цифровой голографии исчисляется тысячами. Основополагающими работами следует назвать работы [3–9].

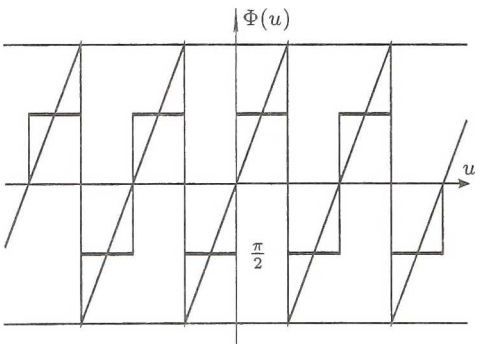


Рис. 1.29. Сравнение бинарной дифракционной решетки и решетки с блеском

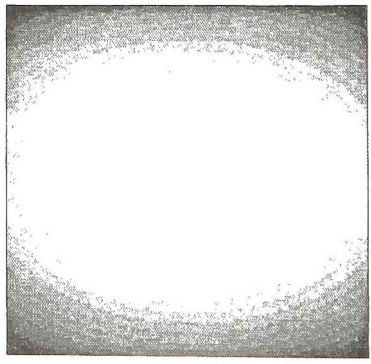


Рис. 1.30. Две скрещенные цилиндрические линзы

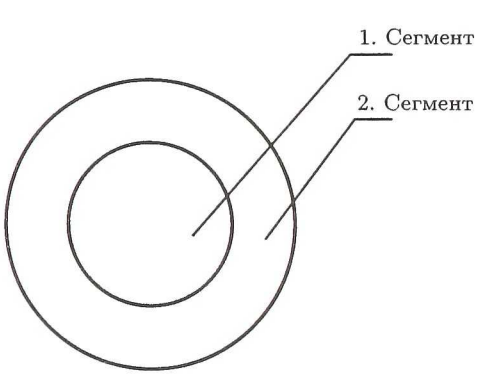


Рис. 1.31. Сегментация ДОЭ



Рис. 1.32. Фотошаблон фокусатора в букву «R»

Как известно, распространение световой волны в пространстве в параксиальном приближении описывается преобразованием Френеля. Если при этом выполняется условие дальней зоны, то преобразование Френеля сводится к преобразованию Фурье. Знание оператора, описывающего распространение световой волны в пространстве и характеристики восстанавливаемого объекта, позволяет рассчитать комплексную амплитуду световой волны в плоскости голографического элемента:

$$T(\mathbf{u}) = a_T(\mathbf{u}) \exp\left[i\varphi_T(\mathbf{u})\right], \tag{1.22}$$

где $|T(\mathbf{u})| \equiv a_T(\mathbf{u}) \leqslant 1$. Дискретизация функции пропускания $T(\mathbf{u})$ приводит к появлению соответствующей матрицы отсчетов поля («пикселов»):

$$\begin{gathered} T_{mn} \equiv T(m,n) = a_T(m,n) \exp\left(i\varphi_T\,(m,n)\right) = a_T(u,v) \exp\left(i\varphi_T\,(u,v)\right)\big|_{\substack{u=m\delta u \\ v=n\delta v}}, \\ T = \{T_{mn}\}, \quad m = \overline{1, N_u}, \quad n = \overline{1, N_v}. \end{gathered} \tag{1.23}$$

Если затем рассчитанные и определенным образом закодированные отсчеты нанести на физический носитель, то при освещении восстанавливающей волной в выходной плоскости можно получить изображение голографированного объекта. На рис. 1.33 и 1.34, соответственно, представлены схемы формирования изображения голограммами Фурье и Френеля. Под методами цифровой голографии понимается комплекс подходов к кодированию отсчетов T_{mn} и фиксации их на физическом носителе, то есть к синтезу цифровых голограмм с функцией пропускания (1.22). Амплитудно-фазовый характер (1.22) привел к появлению трех основных групп существующих типов цифровых голограмм.

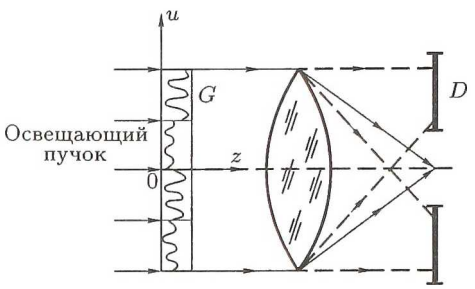


Рис. 1.33. Кодирование для голограмм Фурье

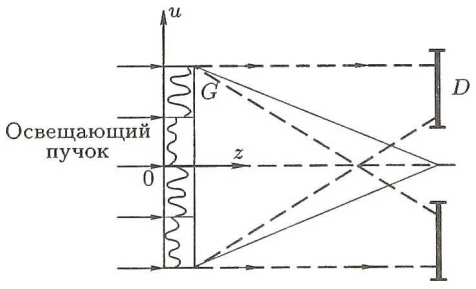


Рис. 1.34. Кодирование для голограмм Френеля

Амплитудные голограммы. Кодированные отсчеты T_{mn} записываются на амплитудном носителе.

Фазовые голограммы. Кодированные отсчеты T_{mn} записываются на фазовом носителе.

Комбинированные голограммы. Амплитудная информация $a_T(m,n)$ наносится на носитель с амплитудным пропусканием, фазовая — $\varphi_T(m,n)$ записывается на носителе с фазовым пропусканием, после записи носители совмещаются.

В настоящее время разработано большое количество методов кодирования амплитудно-фазовых характеристик как в амплитуду, так и в фазу. Наибольшее распространение получили фазовые голограммы, так как при их использовании, из-за прозрачности фазовой среды, удается создавать заданное изображение при сравнительно небольших потерях энергии.

Основным подходом к приведению поля в плоскости голограммы к чисто амплитудному или чисто фазовому является введение в плоскости восстановления объекта дополнительно вспомогательных элементов изображения. Эти элементы пространственно отделены от основного изображения и поэтому не влияют на его качество. Они лишь перераспределяют энергию пучка, уменьшая яркость восстанавливаемого изображения. Процесс синтеза и восстановления фазовой голограммы с использованием вспомогательных элементов показан на рис. 1.35, 1.36.

Кодирование функции пропускания с введением вспомогательных элементов может осуществляться аналитически (например, введением в фазу или амплитуду несущей частоты, что приводит к формированию вспомогательных дифракционных порядков) или в ходе итерационного процесса.

Если исходный объект задан лишь его амплитудным (фазовым) распределением, то для приведения функции пропускания голограммы к чисто амплитудному или фазовому виду может использоваться, соответственно, фазовое (амплитудное) распределение на объекте.

Дадим краткий обзор существующих методов кодирования комплексной функции пропускания.

Амплитудные голограммы. Исторически первым методом расчета амплитудной голограммы является метод Ломана [2], основанный на использовании свойств нерегулярной амплитудной дифракционной решетки. По Ломану любая голограмма может быть интерпретирована как искаженная дифракционная решетка. Таким образом, сутью цифровой голографии является расчет этих искажений, исходя из характеристик голографируемого объекта. Голограмма Ломана представляет собой набор ячеек вида рис. 1.37, каждая из которых соответствует одному пикселу матрицы T. Если рассматривать цифровую голограмму как искаженную решетку, то ясно, что при смещении отверстия в ячейке относительно центрального положения фаза падающей световой волны будет изменяться на пропорциональную величину.

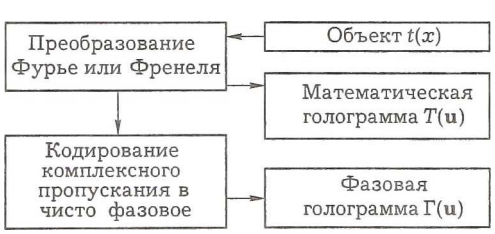


Рис. 1.35. Синтез фазовых голограмм

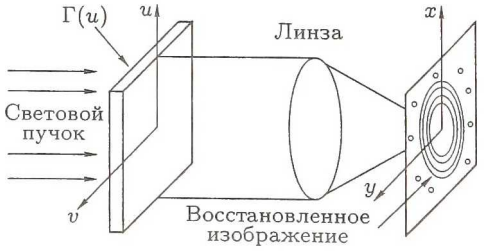


Рис. 1.36. Восстановление фазовой голограммы

Поэтому смещение выбирается следующим образом:

$$C_{u,v} = \frac{\varphi_T(u,v)}{2\pi}(\Delta b - b) + \frac{b}{2}. \qquad (1.24)$$

При изменении размеров отверстия в каждой ячейке будет изменяться амплитуда падающей волны. Размер отверстия определяется следующим образом:

$$h_{u,v} = \frac{a_T(u,v)}{a_{T_{\max}}}\Delta a. \qquad (1.25)$$

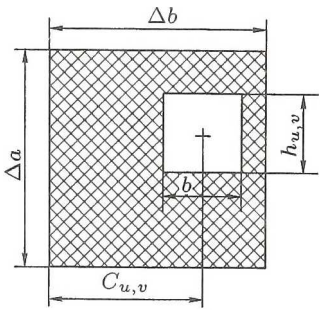


Рис. 1.37. Ячейка голограммы Ломана

Идейно близкими к методу Ломана являются подходы Ли [6] и Бэркхардта [10]. Подход Ли заключается в том, что каждому отсчету T_{mn} комплексной матрицы T ставится в соответствие две пары вещественных отсчетов $T^1_{m,n}$, $T^2_{m,n}$, $T^3_{m,n}$ и $T^4_{m,n}$, значения которых фиксируются в ячейках двух соседних строк амплитудной голограммы:

$$T^1_{m,n} = \begin{cases} T^r_{m,n}, & T^r_{m,n} \geqslant 0, \\ 0, & T^r_{m,n} < 0, \end{cases} \quad T^2_{m,n} = \begin{cases} -T^r_{m,n}, & T^r_{m,n} < 0, \\ 0, & T^r_{m,n} \geqslant 0, \end{cases}$$
$$T^3_{m,n} = \begin{cases} T^i_{m,n}, & T^i_{m,n} \geqslant 0, \\ 0, & T^i_{m,n} < 0, \end{cases} \quad T^4_{m,n} = \begin{cases} -T^i_{m,n}, & T^i_{m,n} < 0, \\ 0, & T^i_{m,n} < 0, \end{cases} \qquad (1.26)$$

где

$$T^r_{m,n} = \mathrm{Re}\,(T_{mn}), \quad T^i_{m,n} = \mathrm{Im}\,(T_{mn}). \qquad (1.27)$$

Подход Бэркхардта [10] заключается в том, что каждому комплексному отсчету T_{mn} ставится в соответствие три вещественных отсчета, фиксируемых в соседних ячейках голограммы:

$$T^1_{m,n} = a_T(m,n)\sin(\varphi_T(m,n)), \quad T^2_{m,n} = a_T(m,n)\cos\left(\frac{\pi}{6} - \varphi_T(m,n)\right),$$
$$T^3_{m,n} = a_T(m,n)\cos\left(\frac{\pi}{6} + \varphi_T(m,n)\right). \qquad (1.28)$$

Общим недостатком методов Ломана, Ли и Бэркхардта является низкая энергетическая эффективность, обусловленная амплитудным характером голограммы и возникновением высших дифракционных порядков вследствие неявного введения несущей.

Фазовые голограммы. Революционным открытием в области синтеза фазовых голограмм явилось создание в 1969 году киноформа [3]. Киноформ представляет собой тонкую фазовую пластинку с постоянным амплитудным пропусканием и фазовым пропусканием, определяемым фазой $\varphi_T(u,v)$ комплексной функции пропускания. На рис. 1.38 представлен фотошаблон киноформа, формирующего крестообразное изображение.

Теоретические исследования показали, что игнорирование амплитудной информации в плоскости голограммы приводит к тому, что около 80% энергии поля в плоскости восстановленного изображения приходится на полезную информацию об объекте. Остальные 20% энергии принадлежат помехам, снижающим качество восстановленного объекта. Для синтеза фазовых голограмм с более высоким качеством восстановления Кирком и Джонсом [4] был предложен метод сведения комплексной

функции пропускания к чисто фазовой, основанный на введении в фазу рассчитываемого элемента синусоидальной несущей. Полезное изображение восстанавливается в «рабочем» дифракционном порядке, а остальные дифракционные порядки, так же как и в методе Ломана, используются в качестве вспомогательных элементов. На рис. 1.39 приведено изображение объекта, восстановленное с голограммы Кирка–Джонса (КД). Отчетливо видны дифракционный рабочий порядок (в центре) и вспомогательные дифракционные порядки.

Впоследствии было предложено обобщение метода КД на случай использования периодической несущей произвольной формы. В этом случае функция комплексного пропускания синтезируемого элемента выбирается в виде

$$T(\mathbf{u}) = \exp[i\varphi(\mathbf{u})], \tag{1.29}$$

$$\varphi(\mathbf{u}) = \varphi_T(\mathbf{u}) + \varphi_{\max}\beta a(\mathbf{u}) q_0[\boldsymbol{\nu}\mathbf{u} + \alpha(\mathbf{u})], \tag{1.30}$$

где $\mathbf{u} = (u, v)$ — вектор, q_0 — произвольная периодическая функция с периодом $2l$, принимающая значения $q_0(t) \in [-l, l]$; $\varphi_{\max} \in [0, 2\pi]$ — максимальный сдвиг фаз на высоте микрорельефа; $\boldsymbol{\nu} = (\nu_1, \nu_2)$ — вектор несущей частоты; $a(\mathbf{u})$ — функция модуляции высоты микрорельефа, нелинейно связанная с $a_T(\mathbf{u})$ см. (1.25) в каждой точке $\mathbf{u}$; $\beta \in [0, 1]$ — коэффициент глубины модуляции.

Оптический элемент (1.30) представляет собой нерегулярную дифракционную решетку, переходящую в регулярную при $\varphi_T \equiv 0$, $\alpha \equiv 0$, $a \equiv 1$. Изменение любого из перечисленных параметров приводит к перераспределению светового потока между различными порядками дифракции, что эквивалентно направлению части мощности светового пучка в «вспомогательные» элементы. При этом в полезный «рабочий» дифракционный порядок идет различная доля энергии из различных ячеек, что эквивалентно неявному введению амплитудного пропускания наряду с фазовым, но при использовании чисто фазовой структуры.

Уравнение (1.30) для фазового пропускания позволяет строить различные методы кодирования. Для любой периодической функции q_0 может быть определен конкретный вид нелинейной зависимости нормированного распределения амплитуды от функции модуляции высоты микрорельефа:

$$a_T(\mathbf{u}) = B(\beta a(\mathbf{u})). \tag{1.31}$$

На рис. 1.40 приведены дифракционные решетки, соответствующие наиболее часто используемым несущим. В табл. 1.1 приведены расчетные формулы для различных типов базовых решеток.

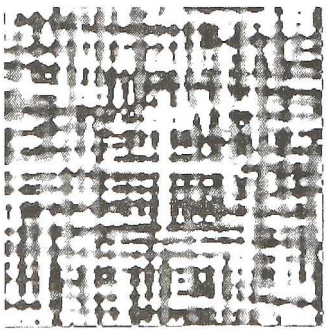


Рис. 1.38. Киноформ

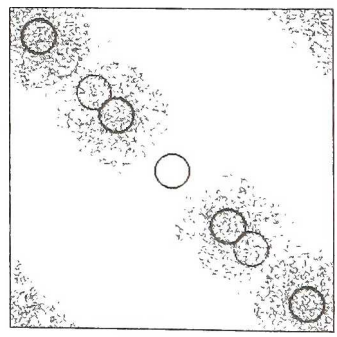


Рис. 1.39. Изображение, восстановленное с голограммы Кирка–Джонса

Таблица 1.1. Расчетные формулы для различных типов решеток

Тип базовой решетки $q_0(t)$	$\varphi_{\max}$	$B(\beta\alpha)$ для n-го дифракционного порядка
Синусоидальная; $\frac{1}{2}\cos(2\pi t)$	$\sim \pi$	$i^n J_n(\frac{1}{2}\varphi_{\max}\beta a)$
Прямоугольная; 0,5 при $l \le t < 0{,}5 + l$ $-0{,}5$ при $0{,}5 + l \leqslant t < l + 1$ $l = 0, \pm 1, \pm 2, \ldots$	π	$\cos(\frac{\pi}{2}\beta a)$ при $n = 0$ $\frac{2}{\pi n}\sin(\frac{\pi}{2}\beta a)$ при $n = \pm 1, \pm 3, \ldots$ 0 при $n = \pm 2, \pm 4, \ldots$
Пилообразная; $(t - 0{,}5 - l)$ при $l \leqslant t < l + 1$ $l = 0, \pm 1, \pm 2, \ldots$	2π	$(-1)^n \operatorname{sinc}(\beta a - n)$

Заметим, что в частном случае $n = 0$, $\boldsymbol{\nu}_1 = (\nu_1, 0)$ синусоидальной несущей получаем те же уравнения, что в методе КД [4]. В другом частном случае прямоугольной несущей приходим к методу синтетических коэффициентов [7].

Кодирование амплитудного распределения в фазовое может осуществляться не только за счет модуляции высоты микрорельефа, но и за счет модуляции других параметров несущей (рис. 1.41–1.43) [11].

Недостатком фазовых голограмм, использующих введение несущей, так же как и голограмм Ломана, Ли и Бэркхардта, является снижение энергетической эффективности из-за возникновения высших дифракционных порядков.

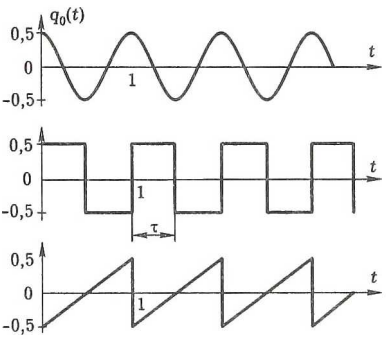


Рис. 1.40. Виды дифракционных решеток, с разной формой периода несущей

Комбинированные голограммы. Достоинством комбинированных голограмм является то, что они создают единственный рабочий порядок, а также то, что они не создают помех, связанных с потерей амплитудной информации. Однако, из-за технологической сложности изготовления и совмещения двух подложек, а также вследствие энергетических потерь, связанных с передачей амплитудного распределения, такие голограммы широкого распространения не нашли.

Итеративный подход к расчету голограмм получил широкое распространение [12–15]. Развитие данного подхода связано с постоянным повышением быстродействия средств вычислительной техники, а также с высоким качеством голограмм, рассчитанных с помощью данного подхода. В настоящее время имеется большое количество методов расчета как амплитудных, так и фазовых голограмм, развитых в рамках данного подхода. Хорошо исследованы вопросы, связанные с условиями сходимости и применимости итеративных процедур. На рис. 1.44 представлена в общем виде схема организации итеративного процесса расчета функции пропускания ДОЭ.

Математическим обоснованием построения итеративного процесса расчета ДОЭ служит, как правило, либо последовательное построение проекций на замкнутые множества, соответствующие ограничениям, накладываемым на функцию пропускания элемента в фокальной плоскости и плоскости элемента [14], либо широко известная градиентная процедура поиска функции [13].

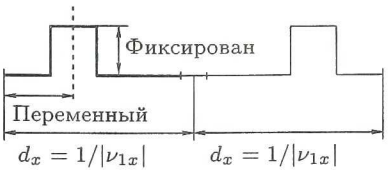


Рис. 1.41. Модуляция смещения центра штрихов дифракционного микрорельефа

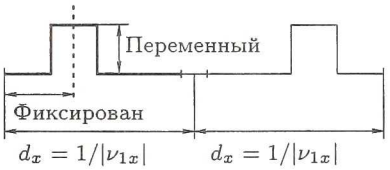


Рис. 1.42. Модуляция высоты штрихов дифракционного микрорельефа

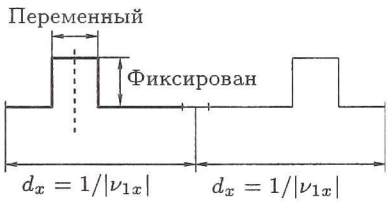


Рис. 1.43. Модуляция «скважности» штрихов дифракционного микрорельефа

В наиболее простом и часто используемом случае, когда объект задан своим амплитудным распределением, для расчета голографического элемента может быть применена итеративная процедура, известная как процедура уменьшения ошибки Фьенапа [12–13]. Заданное амплитудное распределение дополняют какой-либо стохастической или специально рассчитанной детерминированной фазой и осуществляют пересчет в плоскость голограммы. Получившееся амплитудное распределение в плоскости голограммы заменяют на амплитудное распределение освещающего пучка и осуществляют обратный пересчет. Фаза восстановленного объекта является дополнением исходного амплитудного распределения на следующей итерации. Фьенапом было показано, что такая процедура обладает свойством невозрастания ошибки.

Отметим, что расчет киноформа можно интерпретировать как результат работы нулевой итерации итеративного расчета фазовой голограммы объекта, заданного амплитудным распределением.

В случае, если исходный объект задан амплитудно-фазовым распределением, итеративная процедура может использоваться для сведения комплексной функции пропускания к фазовой или амплитудной путем итеративного формирования вспомогательных элементов в плоскости восстановления объекта.

Голографические элементы, рассчитанные с помощью применения итеративного подхода, обладают высокой энергетической эффективностью и относительно небольшой ошибкой формирования заданного распределения. К недостаткам итеративного подхода можно отнести большие вычислительные затраты, а также то, что после 10–30 итераций начинается стагнация, когда дальнейшее увеличение числа итераций не приводит к заметному снижению ошибки формирования заданного распределения. Общим недостатком фазовых голографических элементов, рассчитанных с помощью итеративного подхода и киноформа, являются технологические проблемы изготовления, возникающие из-за крайне нерегулярной фазовой структуры этих элементов.

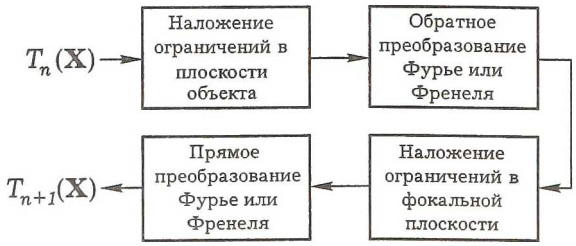


Рис. 1.44. Схема организации итеративного процесса

1.4. Геометрооптический расчет ДОЭ

Лучевая постановка задачи расчета ДОЭ. В однородной среде световые лучи являются прямыми линиями. Расстояние между двумя точками на луче, умноженное на показатель преломления среды, называется оптической длиной пути. Функция оптической длины пути в зависимости от координат точки луча называется эйконалом. Фазой называется аргумент комплексной функции, описывающей любую из проекций электрического или магнитного векторов электромагнитной волны. Геометрическое место точек равного эйконала называется геометрическим волновым фронтом. Пучок лучей, выходящих из малой области на одном волновом фронте и входящих в соответствующую малую область другого волнового фронта, называется лучевой трубкой. Вдоль лучевой трубки поток интенсивности (произведение интенсивности на площадь световой трубки) сохраняется. В рамках геометрической оптики задача фокусировки лазерного излучения эквивалентна поиску функции отображения (или преобразования) координат (u, v) в координаты (x, y), разделенных расстоянием f. Это отображение строится с помощью прямых световых лучей, соединяющих между собой точки обеих плоскостей. Так как луч перпендикулярен волновому фронту, то, зная ход лучей между двумя плоскостями, можно однозначно найти уравнение волнового фронта $W(x, y, z) = \text{const}$.

Оптический элемент представляет собой поверхность, на которой эйконал терпит разрыв. Величина этого разрыва называется эйконалом оптического элемента. В общем случае величина его зависит не только от высоты микрорельефа оптического элемента, но и от эйконала падающей световой волны.

В качестве примера рассмотрим нахождение эйконала оптического элемента, представляющего собой область, заполненную материалом с показателем преломления n и ограниченную поверхностями $z = 0$ и $z = h(x, y)$ (рис. 1.45).

Не ограничивая общности, рассмотрим случай падения плоской волны на оптический элемент со стороны плоскости $z = 0$.

Рассмотрим две опорные плоскости — входную $z = 0$ и выходную $z = L$, где $L = \max h(x, y)$. Найдем эйконал в плоскости $z = L$. Пусть (u, v) — координаты точки входа луча, (x, y) — координаты выхода луча.

Эйконал в выходной плоскости (см. рис. 1.45) равен

$$W(u + N_x\ell, v + N_y\ell) = nh(u, v) + \ell, \quad L = h(u, v) + N_z\ell, \tag{1.32}$$

где (N_x, N_y, N_z) — координаты единичного вектора $\mathbf{N}$, направленного из точки пересечения луча с выходной поверхностью, ограничивающей оптический элемент, в точку выхода луча, ℓ — длина вектора, направленного из точки пересечения луча с поверхностью оптического элемента в точку (x, y).

Разлагая выражение для W в ряд Тейлора, с точностью до линейных членов получаем выражение

$$W(u, v) + \frac{\partial W}{\partial u} N_x\ell + \frac{\partial W}{\partial v} N_y\ell = nh(u, v) + \ell. \tag{1.33}$$

Учитывая, что $\ell = L - h(u, v)/N_z$, получаем окончательное выражение для эйконала во входной плоскости

$$W(u, v) = nh(u, v) + \frac{L - h(u, v)}{N_z} - \left(\frac{\partial W}{\partial u} N_x + \frac{\partial W}{\partial v} N_y\right) \frac{L - h(u, v)}{N_z}. \tag{1.34}$$

В параксиальном приближении $N_z \approx 1$, $N_x \dfrac{\partial W}{\partial u} \ll 1$, $N_y \dfrac{\partial W}{\partial v} \ll 1$.

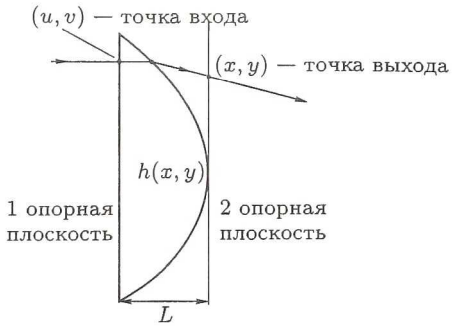


Рис. 1.45. К определению эйконала

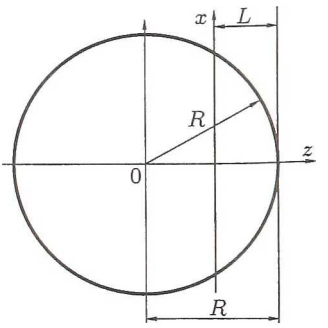


Рис. 1.46. Сферическая плоско-выпуклая линза

С учетом последних соотношений выражение для эйконала во 2-ой опорной плоскости приобретает простой вид

$$W(u,v) = L + (n-1)h(u,v). \tag{1.35}$$

Величина эйконала оптического элемента равна

$$\psi(u,v) = (n-1)h(u,v). \tag{1.36}$$

В качестве примера рассмотрим сферическую плоско-выпуклую линзу (рис. 1.46):

$$\begin{aligned} h(x,y) &= L - R + \sqrt{R^2 - x^2 - y^2}, \\ \psi(u,v) &= C + (n-1)\sqrt{R^2 - u^2 - v^2}, \end{aligned} \tag{1.37}$$

где R — радиус кривизны линзы, L — толщина линзы. Для тонкой линзы

$$\sqrt{R^2 - u^2 - v^2} \approx R - \frac{u^2 + v^2}{2R}, \quad (L \approx 0) \tag{1.38}$$

выражение для эйконала приобретает вид

$$\psi(u,v) = C - \frac{u^2 + v^2}{2f}, \tag{1.39}$$

где $f = R/(n-1)$ — фокусное расстояние тонкой сферической линзы.

Нетрудно видеть, что отображение круга в точку осуществляется с помощью сферического волнового фронта, который формируется обычной сферической линзой. Отображение прямоугольника в отрезок прямой осуществляется с помощью цилиндрического волнового фронта, который формирует цилиндрическая линза. В общем случае отображение некоторой области с координатами (u, v) в заданную область с координатами (x, y) осуществляется с помощью ДОЭ, который можно рассматривать как асферическую линзу с аберрациями.

Вывод уравнений наклонов для расчета ДОЭ. На рис. 1.47 показаны углы α и β, с помощью которых описывается единичный вектор $\mathbf{N}$, направленный из произвольной точки плоскости ДОЭ с координатами (u, v) в соответствующую точку плоскости фокусировки с координатами (x, y). Координаты данного единичного вектора будут равны $N_u = \sin\alpha\sin\beta$, $N_v = \sin\alpha\cos\beta$, $N_z = \cos\alpha$.

Выражая координаты вектора $\mathbf{N}$ через длину S отрезка, которым соединены точки (u, v) и (x, y), и через координаты проекции точки (x, y) на плоскость (u, v), получаем

$$N_u = \frac{x-u}{S}, \quad N_v = \frac{y-v}{S}, \quad N_z = \frac{z}{S}, \quad S = \sqrt{(x-u)^2 + (y-v)^2 + z^2}.$$

Пусть поверхность волнового фронта задана в неявном виде:

$$W(u, v, z) = 0, \tag{1.40}$$

в этом случае вектор нормали имеет вид

$$N_u = \frac{\partial W}{\partial x} \Big/ P, \quad N_v = -\frac{\partial W}{\partial y} \Big/ P, \quad N_z = \frac{\partial W}{\partial z} \Big/ P, \tag{1.41}$$

где $P = \sqrt{\left(\dfrac{\partial W}{\partial x}\right)^2 + \left(\dfrac{\partial W}{\partial y}\right)^2 + \left(\dfrac{\partial W}{\partial z}\right)^2}$.

В нашем случае функция $W(x, y, z)$ является эйконалом и должна удовлетворять уравнению эйконала:

$$\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial z}\right)^2 = 1. \tag{1.42}$$

Произведение показателя преломления $n(x, y, z)$ на малое расстояние вдоль луча $\mathrm{d}S$ равно приращению оптической длины пути или приращению эйконала $\mathrm{d}W(x, y, z)$ (см. рис. 1.48):

$$\mathrm{d}W(x, y, z) = n(x, y, z)\,\mathrm{d}S. \tag{1.43}$$

Из (1.43) следует, что

$$n = \frac{\mathrm{d}W}{\mathrm{d}S} = \nabla W \cdot \mathbf{N}, \tag{1.44}$$

где $\mathbf{N} = (N_x, N_y, N_z)$ — единичный вектор, направленный вдоль луча.

В случае однородной оптической среды: $n(u, v, z) = 1$, получим уравнение эйконала в виде

$$|\nabla W|^2 = 1. \tag{1.45}$$

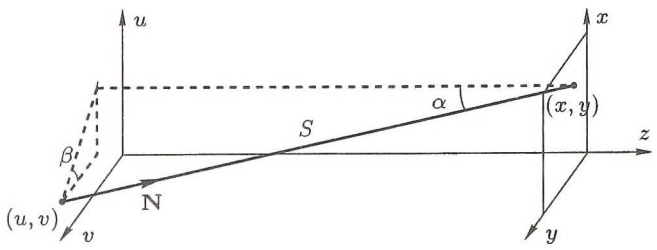


Рис. 1.47. Оптическая схема хода лучей из плоскости (u, v) в плоскость (x, y)

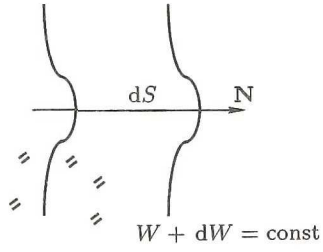


Рис. 1.48. К выводу уравнения эйконала

Получим уравнения наклонов, которые позволяет находить функцию фазы ДОЭ:

$$\begin{aligned} &\varphi(u,v) = kW(x,y,z)|_{x=u,y=v,z=0}, \\ &\frac{\partial\varphi}{\partial u} = \frac{k(x-u)}{S}, \quad \frac{\partial\varphi}{\partial v} = \frac{k(y-v)}{S}, \\ &\frac{\partial\varphi}{\partial z} = \sqrt{k^2 - \left(\frac{\partial\varphi}{\partial u}\right)^2 + \left(\frac{\partial\varphi}{\partial v}\right)^2} = \frac{kz}{S}. \end{aligned} \tag{1.46}$$

Геометрооптический расчет фокусаторов. Ставится задача фокусировки падающего пучка с интенсивностью $I_0(\mathbf{u})$ и фазой $\psi_0(\mathbf{u})$ в фокальную кривую, заданную параметрическим уравнением $\mathbf{x} = \mathbf{X}(\xi)$, $\xi \in [0, \ell]$ и линейной плотностью энергии $I(\xi)$ на кривой (рис. 1.49).

В работах по расчету фокусаторов [16–18] функция $I(\xi)$ называется линейной плотностью. Примеры функций линейной плотности при фокусировке в отрезок прямой показаны на рис. 1.50.

Расчет фокусаторов основан на поиске лучевого соответствия $\xi = \xi(\mathbf{u})$ между точками фокальной кривой и точками апертуры фокусатора с последующим восстановлением фазы фокусатора из уравнения наклонов (1.46). Поскольку фокусатор двумерен, а фокальная кривая одномерна, то существует одномерное множество точек $\Gamma(\xi)$ фокусатора, направляющих излучение в одну и ту же точку $\mathbf{X}(\xi)$ кривой. Назовем это множество точек слоем. Лучевая трубка здесь является листообразной, проходит между точками ξ и $\xi + d\xi$ фокальной кривой и в области $D(\xi, \xi + d\xi)$ между $\Gamma(\xi)$ и $\Gamma(\xi + d\xi)$ (рис. 1.49*а*). В теоретических работах по расчету фокусаторов показано, что слои представляют собой семейство гипербол. В параксиальном приближении гиперболы вырождаются в семейство прямых [17]

$$\mathbf{u}\mathbf{X}'(\xi) = p(\xi), \tag{1.47}$$

перпендикулярных касательным к фокальной кривой. Функция $p(\xi)$ в (1.47) является расстоянием от слоя до начала координат в плоскости фокусатора и определяет распределение энергии вдоль фокальной кривой.

Для заданной линейной плотности функция $p(\xi)$ ищется из закона сохранения энергии. Записывая закон сохранения светового потока в интегральной форме для части фокальной кривой, заключенной между точками $\mathbf{X}(0)$ и $\mathbf{X}(\xi)$ и соответствующей части апертуры фокусатора $D(0, \xi)$ (функция $D(0, \xi)$ однозначно определяет функцию $p(\xi)$), заключенного между слоями $\Gamma(0)$ и $\Gamma(\xi)$ (см. рис. 1.49*б*), получаем

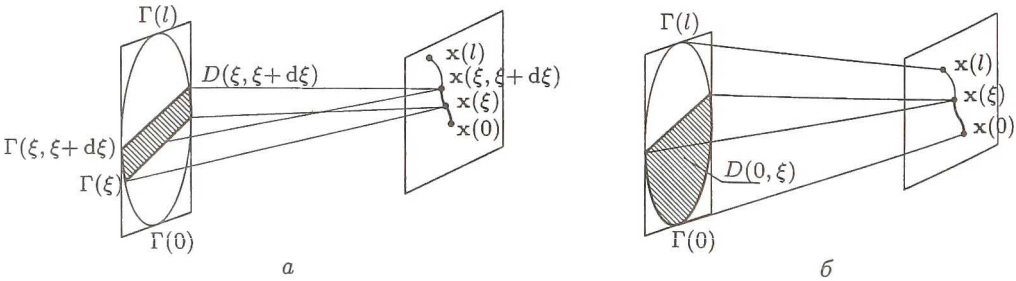
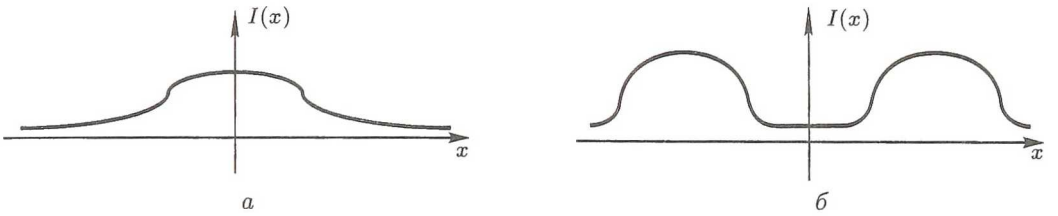


Рис. 1.49. Фокусировка в фокальную прямую



а *б*

Рис. 1.50. Примеры функций линейной плотности при фокусировке в отрезок

для определения функции $p(\xi)$ следующее уравнение:

$$\int_0^{\xi} I(t)dt = \int\limits_{D(0,\xi)} I_0(\mathbf{u})\,\mathrm{d}^2\mathbf{u}. \tag{1.48}$$

Подставляя полученную функцию $p(\xi)$ в уравнение слоя и снова решая полученное алгебраическое уравнение относительно ξ , находим соответствие $\xi = \xi(\mathbf{u})$ между точками на фокальной кривой и точками на фокусаторе. Подставляя это соответствие в уравнение наклонов

$$\nabla_\perp\varphi = \frac{k\left(\mathbf{X}(\xi(\mathbf{u})) - \mathbf{u}\right)}{\sqrt{\left[f^2 + \left(\mathbf{X}(\xi(\mathbf{u})) - \mathbf{u}\right)^2\right]}} \cong \frac{k}{f}\left[\mathbf{X}(\xi(\mathbf{u})) - \mathbf{u}\right], \tag{1.49}$$

можно найти $\nabla_\perp\varphi(\mathbf{u})$, а затем восстановить $\varphi(\mathbf{u})$ методами теории потенциала:

$$\varphi(\mathbf{u}) = \varphi(\mathbf{u}_0) + \int_{\mathbf{u}_0}^{\mathbf{u}} \nabla_\perp\varphi\,\mathrm{d}\mathbf{u}, \quad \nabla_\perp = \left(\frac{\partial}{\partial u}, \frac{\partial}{\partial v}\right), \tag{1.50}$$

где интегрирование ведется по произвольной кривой, соединяющей точку $\mathbf{u}$ с некоторой начальной точкой $\mathbf{u}_0$.

Рассмотрим несколько примеров [19–24].

Пример 1.1. Фокусировка в отрезок $|x| \leqslant \ell/2$ оси X:

$$\mathbf{X}(\xi) = (x, 0); \quad \mathbf{X}'(\xi) = (1, 0). \tag{1.51}$$

В этом случае слоями $\Gamma(\xi)$ являются отрезки прямых $u = p(\xi)$, параллельные $0v$.

В простейшем случае фокусировки пучка прямоугольного сечения $D = [-A, A] \times$ $\times\,[-B, B]$ с плоским фронтом и постоянной в D интенсивностью $I_0(\mathbf{u}) = (4AB)^{-1}$ в отрезок (1.51) с постоянной линейной плотностью, формулы (1.48)–(1.50) дают фазу фокусатора в виде

$$\varphi(\mathbf{u}) = k\left\{-\frac{u^2}{2f_0}\left(1 - \frac{\ell}{2A}\right) - \frac{v^2}{2f_0}\right\}, \quad \mathbf{u} = (u, v). \tag{1.52}$$

Фаза (1.52) описывает, как и следовало ожидать из физического смысла задачи, астигматический пучок с фокусами $f_0/(1 - \ell/2A)$ и f_0. Фотошаблон, соответствующий фазовой функции (1.52), показан на рис. 1.51.

Пример 1.2. В более сложном случае фокусировки гауссова пучка прямоугольного сечения в отрезок (1.51), фазовая функция фокусатора может быть получена из (1.48)–(1.50) в виде

$$\varphi(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \frac{k\ell a}{\sqrt{2}f\,\mathrm{erf}\left(\dfrac{A}{a\sqrt{2}}\right)} \times \\ \times \left\{\frac{u}{a\sqrt{2}}\,\mathrm{erf}\left(\frac{u}{a\sqrt{2}}\right) - \frac{1}{\sqrt{\pi}}\left[1 - \exp\left(-\frac{u^2}{2a^2}\right)\right]\right\}, \quad (1.53)$$

где

$$\mathrm{erf}(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t,$$

a — параметр гауссового пучка.

Более сложный характер фазовой функции обусловлен сочетанием астигматизма с дифракционным перераспределением энергии от центра к периферии пучка.

Геометрооптический расчет компенсаторов волновых фронтов. Компенсаторы предназначены для преобразования формы волнового фронта освещающего пучка [25–28], например, преобразования сферического волнового фронта в асферический. Геометрия формирования волнового фронта компенсатором показана на рис. 1.52.

Зададим волновой фронт σ уравнением $z = f(\mathbf{x})$. Введем в рассмотрение трехмерный нормированный градиент поверхности, определяемой функцией $f(\mathbf{x})$, в виде

$$\mathbf{N} = \frac{(-\nabla_\perp f, 1)}{\sqrt{1 + |\nabla_\perp f|^2}}. \quad (1.54)$$

Исходя из решения задачи, вид $\varphi(\mathbf{u})$ фазовой функции компенсатора нужно подобрать так, чтобы

$$\psi_0(\mathbf{u}) + \frac{\varphi(\mathbf{u})}{k} + L = \frac{\varphi_0}{k}, \quad (1.55)$$

где φ_0 — постоянная фаза, L — расстояние между точкой компенсатора и точкой волнового фронта, отсчитанное по лучу, $k\psi_0$ — фаза освещающего пучка. Величину

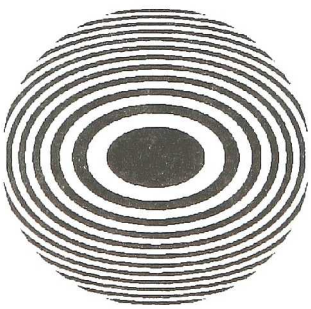


Рис. 1.51. Фотошаблон фокусатора в отрезок

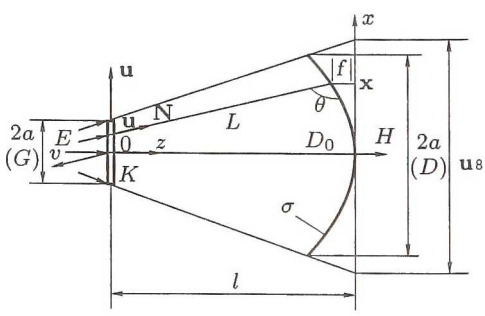


Рис. 1.52. Геометрия формирования волнового фронта компенсатором

L согласно (1.54) можно выразить в виде

$$L = [\ell + f(\mathbf{x})]\sqrt{1 + |\nabla_\perp f(\mathbf{x})|^2}. \tag{1.56}$$

Из (1.55)–(1.56) получаем уравнение фазовой функции компенсатора:

$$\varphi(\mathbf{u}) = \varphi_0 - k\left\{\psi_0(\mathbf{u}) + [\ell + f(\mathbf{x})]\sqrt{1 + |\nabla_\perp f(\mathbf{x})|^2}\right\}, \tag{1.57}$$

и уравнение соответствия координат точки $\mathbf{u}$ на компенсаторе и точки $\mathbf{x}$ на поверхности σ:

$$\mathbf{u} = \mathbf{x} + [\ell + f(\mathbf{x})]\,\nabla_\perp f(\mathbf{x}). \tag{1.58}$$

Для осесимметричных компенсаторов, формирующих волновые фронты вращения, соотношения (1.57)–(1.58) выглядят следующим образом:

$$\varphi(r) = \varphi_0 - k\left\{\psi_0(r) + [\ell + f(\rho)]\sqrt{1 + [f'(\rho)]^2}\right\}, \tag{1.59}$$

$$r = \rho + [\ell + f(\rho)]\,f'(\rho), \tag{1.60}$$

где $\ell = -f\left(\dfrac{D}{2}\right) - \dfrac{D-d}{2f'(D/2)}$, $\rho = \sqrt{x^2+y^2}$, $r = \sqrt{u^2+v^2}$, $z = f(\rho)$, $0 \leqslant \rho \leqslant D/2$, d — диаметр компенсатора.

На рис. 1.53 приведен фотошаблон компенсатора, преобразующего сферический волновой фронт в параболический волновой фронт.

1.5. Дискретизация и квантование фазы ДОЭ

Модель дискретизации и квантования фазы при создании ДОЭ. Широкие возможности дифракционной оптики ограничены характеристиками разрешения устройств расчета и генерации дифракционного микрорельефа ДОЭ. Дискретизация по аргументам и квантование по уровням функции фазового пропускания, приводит к отклонению характеристик ДОЭ от расчетных [29–34].

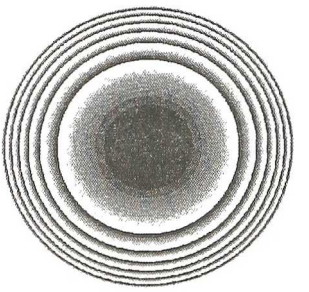


Рис. 1.53. Фотошаблон компенсатора сфера-параболоид

Здесь рассматривается технология изготовления синтезированных ДОЭ с помощью растровых генераторов изображений, управляемых компьютером. Вначале формулируется математическая модель дискретизации и квантования. Далее рассматриваются свойства возмущений дискретизации, квантования, и производится оценка погрешности фазы непосредственно во входной плоскости синтезированного ДОЭ. И, наконец, оценивается структура поля в выходной плоскости, параллельной плоскости синтезированного оптического элемента.

Все основные этапы синтеза фазовых ДОЭ с помощью компьютера так или иначе связаны с дискретизацией и квантованием фазы. Гладкая фаза $\varphi(u,v) = \varphi(\mathbf{u})$ рассчитывается в виде матрицы отсчетов (пикселов), приводится к интервалу $[0, 2\pi m)$, в результате чего образуется приведенная фаза

$\Phi(u,v) \equiv \Phi(\mathbf{u})$ (1.8), значения которой затем квантуются по M уровням и подаются для управления генератором фотошаблонов, отображающим ее в виде вариаций плотности почернения: полутоновых или бинарных. Далее следует технологический процесс формирования микрорельефа: полутоновые фотошаблоны могут быть использованы для формирования непрерывного дифракционно-фазового микрорельефа; набор бинарных фотошаблонов позволяет литографически получить ступенчатый фазовый микрорельеф.

Возмущения дискретизации и квантования являются принципиально неустранимыми для оптических элементов, синтезируемых на компьютере, и подлежат оценке.

Рассмотрим модель дискретизации оптических элементов, изготавливаемых на растровом генераторе изображений или на растровом литографе.

Синтезированные на сканирующем растровом генераторе фотошаблоны являются набранными из множества J непересекающихся ячеек разрешения $G_{\mathbf{j}}$, в объединении составляющих рабочую область фотошаблона (см. рис. 1.54 и 1.55), где $\mathbf{j} = (j_1, j_2)$ — двумерный индекс-номер ячейки разрешения с центром $\mathbf{u}_{\mathbf{j}}$. Размер ячеек $G_{\mathbf{j}}$ определяет разрешение фотошаблонов и формируемых с них оптических элементов.

Важнейшей характеристикой растрового генератора является функция распределения энергии воздействия в сканирующем пятне $E_{\mathbf{j}}(\mathbf{u})$. Эта функция должна иметь протяженность, в точности равную размеру ячейки разрешения. В противном случае будут иметь место эффекты взаимного наложения фазы в соседних ячейках.

Рассмотрим наиболее интересные частные случаи растровых генераторов. При использовании прямоугольно растровых генераторов фотошаблонов с равномерным прямоугольным пятном размера $\delta_1 \times \delta_2$ и полем $2a \times 2b$ (см. рис. 1.54) можно пользоваться следующей моделью дискретизации:

$$G_{\mathbf{j}} = \left\{ (u,v) : u_{j_1} - \frac{\delta_1}{2} \leqslant u \leqslant u_{j_1} + \frac{\delta_1}{2}, \qquad v_{j_2} - \frac{\delta_2}{2} \leqslant v \leqslant v_{j_2} + \frac{\delta_2}{2} \right\}, \tag{1.61}$$

$$\mathbf{u}_{\mathbf{j}} = (u_{j_1}, v_{j_2}), \quad j_1 = \overline{1, N_1}, \quad j_2 = \overline{1, N_2},$$

где числа u_{j_1}, v_{j_2} образуют арифметическую прогрессию с шагом δ_1, δ_2, соответственно, $N_1\delta_1 = 2a$, $N_2\delta_2 = 2b$. В этом случае

$$E_{\mathbf{j}}(u,v) = \operatorname{rect}\left(\frac{u - u_{j_1}}{\delta_1}\right) \operatorname{rect}\left(\frac{v - v_{j_2}}{\delta_2}\right), \tag{1.62}$$

где

$$\operatorname{rect}(t) = \begin{cases} 1, & |t| < 1/2, \\ 0, & |t| \geqslant 1/2. \end{cases}$$

При этом в пределах прямоугольного поля $2a \times 2b$ имеется $N_1 \times N_2$ возможных позиций пятна, лишь часть $|J|$ из которых соответствует области G (кругу).

Прямоугольно-растровые генераторы позволяют создавать фотошаблоны произвольных ДОЭ, и их фазовая функция выражается интерполяционной формулой

$$\Phi(\mathbf{u}) = \sum_{\mathbf{j} \in J} \Phi_{\mathbf{j}} E_{\mathbf{j}}(\mathbf{u}), \tag{1.63}$$

где $\Phi_{\mathbf{j}} \equiv \Phi(\mathbf{u})|_{\mathbf{u}=\mathbf{u}_{\mathbf{j}}}$.

Соответственно функция пропускания ДОЭ имеет вид

$$T(\mathbf{u}) = \exp\left[i\Phi(\mathbf{u})\right]. \tag{1.64}$$

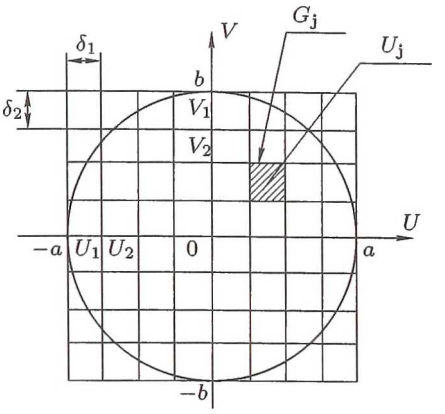


Рис. 1.54. Прямоугольная растровая структура фотошаблона

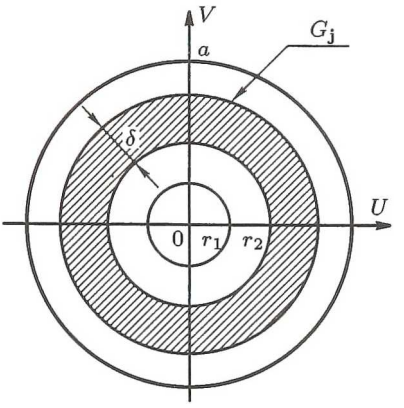


Рис. 1.55. Кольцевая растровая структура фотошаблона

При использовании кольцевых растровых генераторов фотошаблонов (рис. 1.55) область G представляет собой круг радиуса a, ячейки G_j нумеруются одним индексом $j = \overline{1, N}$ и являются кольцами:

$$G_j = \left\{ \mathbf{u} : r_j - \frac{\delta}{2} \leqslant |\mathbf{u}| \leqslant r_j + \frac{\delta}{2} \right\},$$
$$E_j(\mathbf{u}) \equiv E_j(r) = \operatorname{rect}\left(\frac{r - r_j}{\delta} \right). \tag{1.65}$$

Числа r_j образуют арифметическую прогрессию с шагом δ, $r = |\mathbf{u}|$. Кольцевые генераторы ориентированы на изготовление фотошаблонов радиально-симметричных оптических элементов с функцией пропускания $T(\mathbf{u}) = T(r)$, $r = \sqrt{u^2 + v^2}$, и фазовой функцией $\Phi(\mathbf{u}) = \Phi(r) = \sum_{j=1}^{N} \Phi_j E_j(r)$.

Рассмотрим далее модель квантования значений фазы. При создании фотошаблонов отсчеты фазы (пикселы) Φ_j подвергаются квантованию по M эквидистантным уровням с постоянным шагом,

$$h_m = \frac{2\pi m}{M}, \tag{1.66}$$

и преобразуются в отсчеты $\hat{\Phi}_j$, $j \in J$, квантованной фазы (рис. 1.57)

$$\hat{\Phi}_j = \hat{\Phi}(\Phi_j). \tag{1.67}$$

Квантованные значения $\hat{\Phi}_j$ являются управляющими сигналами растрового генератора фотошаблонов. С учетом квантования фазы по уровням фазовая функция и функция пропускания записываются, соответственно, в виде

$$\hat{\Phi}(\mathbf{u}) = \sum_{j \in J} \hat{\Phi}_j E_j(\mathbf{u}), \tag{1.68}$$

$$\hat{T}(\mathbf{u}) = \exp\left[i\hat{\Phi}(\mathbf{u}) \right]. \tag{1.69}$$

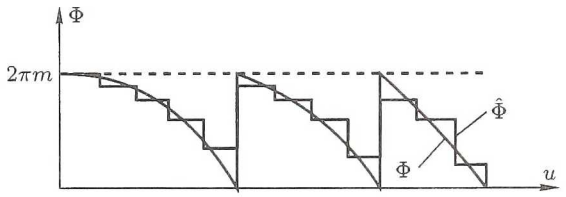


Рис. 1.56. Дискретизация (квантование) фазы

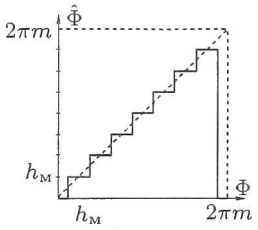


Рис. 1.57. Равномерное квантование фазы

Оценка погрешности дискретизации и квантования фазы в плоскости ДОЭ. Введем в рассмотрение погрешность дискретизации и квантования

$$\Delta\Phi(\mathbf{u}) = \Phi(\mathbf{u}) - \hat{\Phi}(\mathbf{u}). \tag{1.70}$$

Из рис. 1.56 и предыдущего материала видно, что величина ее зависит от шага дискретизации по аргументам δ_1, δ_2 и от числа уровней квантования M. Таким образом, следует различать *погрешность дискретизации* и *погрешность квантования*.

В ряде случаев погрешность дискретизации практически отсутствует. В частности, если радиально-симметричный оптический элемент изготавливается с помощью кольцевого растрового генератора, то при правильном подборе радиуса колец погрешность дискретизации удается свести к пренебрежимо малой величине.

В любом случае погрешность дискретизации стремятся свести к минимуму, однако ограничения здесь диктуются разрешением генератора фотошаблонов. В случае, когда разрешающая способность имеющегося генератора фотошаблонов не позволяет получать достаточно малую погрешность квантования, следует отказаться от его использования. Приведем соображения по выбору шага дискретизации. Рассмотрим непрерывную функцию $f(u)$, $-a \leqslant u \leqslant a$, подлежащую дискретизации с шагом Δ:

$$f(n) = f(u)|_{u=n\Delta}, \quad n = 0, 1, \ldots, N, \quad N = [(2a)/\Delta]. \tag{1.71}$$

Из теории сигналов известно: для обратимой дискретизация (по функции $f(n)$ можно восстановить функцию $f(u)$) необходимо выполнить условие

$$\Delta < 1/(2F), \tag{1.72}$$

где F — ширина спектра функции $f(u)$. Неравенство (1.72) является качественным в выборе шага дискретизации. Дадим дополнительные количественные оценки. Для этого рассмотрим фазовую функцию $\varphi(u, v)$ и наложим достаточно очевидное *условие допустимости дискретизации* в виде

$$\left| \delta_1 \frac{\partial\varphi(u,v)}{\partial u} + \delta_2 \frac{\partial\varphi(u,v)}{\partial v} \right| \ll 2\pi, \tag{1.73}$$

или сокращенно $|\nabla_\delta \varphi| \ll 2\pi$. Оценим погрешность кусочно-постоянной аппроксимации фазовой функции $\varphi(\mathbf{u})$ (см. рис. 1.51). В каждой ячейке $G_\mathbf{j}$ функция φ может быть разложена в ряд Маклорена и аппроксимирована выражением

$$\varphi(u, v) = \varphi(u_{j_1}, v_{j_2}) + (u - u_{j_1})\varphi_u + (v - v_{j_2})\varphi_v, \tag{1.74}$$

где $\varphi_u = \partial\varphi(\mathbf{u})/\partial u|_{\mathbf{u}=\mathbf{u}_j}$; $\varphi_v = \partial\varphi(\mathbf{u})/\partial v|_{\mathbf{u}=\mathbf{u}_j}$. Предложенные соотношения позволяют выразить погрешность дискретизации в виде

$$\Delta\varphi_\mathbf{j}(\mathbf{u}) = -(\mathbf{u} - \mathbf{u}_\mathbf{j})\nabla_\perp\varphi(\mathbf{u}_\mathbf{j}). \tag{1.75}$$

Заметим, что в силу условия (1.73) имеет место неравенство

$$|\Delta\varphi_{\mathbf{j}}(\mathbf{u})| \ll 2\pi. \tag{1.76}$$

Влияние дискретизации может быть охарактеризовано максимальной или среднеквадратичной погрешностью (флуктуацией) фазы:

$$\Delta\varphi_{\max} = \max_{\mathbf{u}\in G} |\Delta\varphi(\mathbf{u})|\,, \tag{1.77}$$

$$\Delta\varphi_s = \left[\frac{1}{|G|}\int\limits_G |\Delta\varphi(\mathbf{u})|^2\,\mathrm{d}^2u\right]^{1/2}, \tag{1.78}$$

где $|G|$ — площадь области G.

Оценим величины (1.77) и (1.78) для прямоугольного сканирующего пятна (1.62) и гладкой функции $\varphi(u,v)$. Подставляя представление (1.74), (1.75) в уравнение (1.77), последовательно получаем

$$\Delta\,\varphi_{\max} = \max_{\mathbf{j}\in J}\max_{\mathbf{u}\in G_{\mathbf{j}}} |\Delta\,\varphi_{\mathbf{j}}(\mathbf{u})| \leqslant \max_{\mathbf{j}\in J}\max_{\mathbf{u}\in G_{\mathbf{j}}} |(\mathbf{u}-\mathbf{u}_{\mathbf{j}})\nabla_{\perp}\varphi(\mathbf{u}_{\mathbf{j}})|\,. \tag{1.79}$$

Поскольку максимум линейной функции

$$(u-u_{j_1})\varphi_u + (v-v_{j_2})\varphi_v, \tag{1.80}$$

на прямоугольнике $G_{\mathbf{j}}$ (1.61) достигается при

$$u-u_{j_1} = \frac{\delta_1}{2}\operatorname{sgn}(\varphi_u); \quad v-v_{j_2} = \frac{\delta_2}{2}\operatorname{sgn}(\varphi_v); \quad \operatorname{sgn}(t) = \begin{cases} 1, & t\geqslant 0,\\ 0, & t<0, \end{cases} \tag{1.81}$$

и составляет

$$\frac{1}{2}\left(\delta_1\,|\varphi_u| + \delta_2\,|\varphi_v|\right), \tag{1.82}$$

то из (1.79) получается удобная оценка максимальной погрешности дискретизации фазы:

$$\Delta\varphi_{\max} \leqslant \frac{1}{2}\max_{\mathbf{u}\in G}|\nabla_\delta\varphi(\mathbf{u})|_+\,, \tag{1.83}$$

где символ $|\cdot|_+$ означает сумму модулей компонентов вектора. Аналогично для $\Delta\varphi_s$ имеем

$$|G|\cdot(\Delta\varphi_s)^2 = \sum_{\mathbf{j}\in J}\int\limits_{G_{\mathbf{j}}} [\Delta\varphi_{\mathbf{j}}(\mathbf{u})]^2\,\mathrm{d}^2\mathbf{u}. \tag{1.84}$$

Выполняя простейшее интегрирование с учетом уравнения (1.65) и соотношений

$$\begin{gathered}\int\limits_{G_{\mathbf{j}}} (\mathbf{u}-\mathbf{u}_{\mathbf{j}})\,\mathrm{d}^2\mathbf{u} = 0,\\ \int\limits_{G_{\mathbf{j}}} (u-u_{j_1})^2\,\mathrm{d}^2\mathbf{u} = \frac{\delta_1^3\delta_2}{12}, \qquad \int\limits_{G_{\mathbf{j}}} (v-v_{j_2})^2\,\mathrm{d}^2\mathbf{u} = \frac{\delta_1\delta_2^3}{12}\end{gathered} \tag{1.85}$$

получаем

$$|G| \cdot (\Delta\varphi_s)^2 = \frac{1}{12} \sum_{j \in J} [\nabla_\delta \varphi(\mathbf{u}_j)]^2 \, \delta_1 \delta_2. \tag{1.86}$$

Заменяя сумму соответствующим интегралом, получаем удобную *оценку среднеквадратичной погрешности дискретизации фазы*:

$$\Delta\varphi_s = \left\{ \frac{1}{12\,|G|} \int\limits_G [\nabla_\delta \varphi(\mathbf{u})]^2 \, \mathrm{d}^2\mathbf{u} \right\}^{1/2}. \tag{1.87}$$

Оценки (1.83), (1.87) позволяют подсчитать погрешности дискретизации $\Delta\varphi_s$ непосредственно по параметрам дискретизации δ_1, δ_2 и первым производным фазы. Интересный частный случай соответствует «цилиндрической оптике», с $\varphi(\mathbf{u}) \equiv \varphi(u)$, $|u| \leqslant a$. В этом случае формулы примут вид

$$\Delta\varphi_{\max} \leqslant \frac{\delta}{2} \max_{|u| \leqslant a} |\varphi'(\mathbf{u})|, \quad \varphi' = \frac{\mathrm{d}\varphi}{\mathrm{d}u}, \tag{1.88}$$

$$\Delta\varphi_s = \left\{ \frac{\delta^2}{24a} \int\limits_{-a}^{a} [\varphi'(\mathbf{u})]^2 \, \mathrm{d}u \right\}^{1/2}. \tag{1.89}$$

Другой важный частный случай соответствует радиально-симметричным фазовым функциям $\varphi(\mathbf{u}) \equiv \varphi(r)$, $r = |\mathbf{u}| \leqslant a$. Тогда

$$\Delta\varphi_{\max} \leqslant \frac{\delta}{2} \max_{0 \leqslant r \leqslant a} |\varphi'(r)|, \quad \varphi' = \frac{\mathrm{d}\varphi}{\mathrm{d}r}, \tag{1.90}$$

$$\Delta\varphi_s = \left\{ \frac{\delta^2}{6a^2} \int\limits_{0}^{a} [\varphi'(r)]^2 \, r \, \mathrm{d}r \right\}^{1/2}. \tag{1.91}$$

Оценки (1.83), (1.88)–(1.91) дают непосредственную связь возмущений дискретизации с производными гладкой фазовой функции.

В качестве примера на рис. 1.58 приведен график среднеквадратичного отклонения $\Delta\varphi_s$ при дискретизации фазы,

$$\varphi(r) = -k \frac{1 + (n-1)^{-2}}{8 f_0^3} r^4, \tag{1.92}$$

с параметрами $\lambda = 0{,}63$ мкм, $m = 1$, $n = 1{,}5$, разрешение δ[мкм] является варьируемым параметром.

Из рис. 1.58 видно, что при разрешении генератора фотошаблона $\delta = 6$ мкм удается реализовать оптический элемент (1.92) с погрешностью дискретизации фазы, не превышающей 5% в диапазоне изменения параметра «относительное отверстие», ДОЭ $0 \leqslant 2a/f_0 \leqslant 0{,}8$, а при разрешении $\delta = 24$ мкм соответствующий диапазон сужается в 2 раза.

Перейдем к рассмотрению эффектов квантования. Следует различать ситуацию бинарного квантования ($M = 2$) и примыкающего к ней «грубого» квантования с малым числом уровней $M = 3, 4$, а также квантования с большим числом уровней $M \gg 2$. В случае бинарного квантования фаза ДОЭ принимает два значения: 0 и π (отсюда термин «бинарная оптика»).

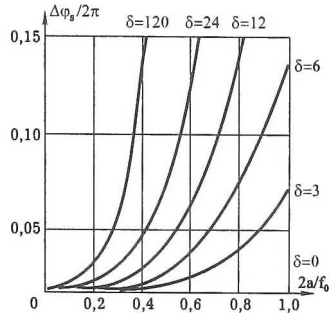


Рис. 1.58. Среднеквадратичное отклонение при дискретизации фазы

ДОЭ работает как фазовая дифракционная решетка (или зонная пластинка), т.е. дает большое количество дифракционных порядков. Рабочим обычно является первый порядок дифракции, в который попадает до 40% энергии, что и определяет предельную дифракционную эффективность ДОЭ в данном случае. При большом числе уровней квантования можно ввести в рассмотрение *шум квантования* $\Delta\varphi_k$. Из рис. 1.57 видно, что при шаге квантования (1.66) максимальное значение погрешности квантования фазы

$$\Delta\varphi^k_{\max} = \frac{\pi m}{M}. \tag{1.93}$$

Полагая, что $\Delta\varphi_л$ принимает равновероятные значения в диапазоне $[-\pi m/M, \pi m/M]$, нетрудно вычислить среднеквадратичное значение погрешности квантования фазы:

$$\Delta\varphi^k_s = \frac{1}{2\pi m/M} \int\limits_{-\pi m/M}^{\pi m/M} \left(\Delta\varphi^k_k\right)^2 \mathrm{d}\left(\Delta\varphi^k_k\right) = \frac{1}{3}\left(\frac{\pi m}{M}\right)^2. \tag{1.94}$$

При больших значениях M величина среднеквадратичной погрешности квантования фазы составляет доли процента. В этом случае эффект квантования фазы может быть сведен к появлению в плоскости ДОЭ дополнительного фазового транспаранта со случайной по полю апертуры фазой, которая варьируется в узком диапазоне $[-\pi m/M, \pi m/M]$, и описывается моделью случайного фазового диффузора.

Анализ влияния дискретизации и квантования фазовой функции на характеристики ДОЭ. Для анализа влияния дискретизации и квантования фазовой функции на характеристики ДОЭ следует рассмотреть дифракцию света на синтезированном оптическом элементе.

Рассмотрим монохроматическую волну

$$W(u, v, z, t) = f(u, v) \exp\left[-i\omega(t - \frac{z}{c})\right], \tag{1.95}$$

распространяющуюся в направлении z со скоростью c. Пусть в плоскости $z = 0$ (рис. 1.59) размещен ДОЭ с пропусканием $T(u, v)$ в апертуре G. Задачей является

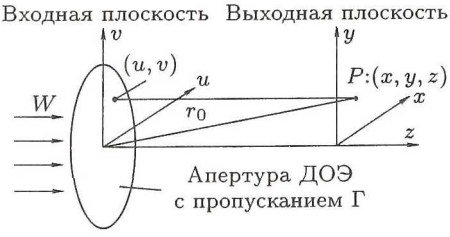


Рис. 1.59. Дифракция света на ДОЭ

определение дифракционного поля, $S(x,y,z,t) = g(x,y,z)\exp(-i\omega t)$, на больших расстояниях от плоскости ДОЭ. Пусть r_0 — расстояние от начала координат до точки $\mathrm{P}(x,y,z)$, а $\alpha = x/r_0, \beta = y/r_0$ — направляющие косинусы линии OP. Тогда, если r_0 велико по сравнению с размером области G, то поле $g(x,y,z)$ в точке P выражается соотношением

$$g(x,y,z) = Ae^{ikr_0}\iint\limits_G T(u,v)\exp[-ik(\alpha u+\beta v)]\,\mathrm{d}u\,\mathrm{d}v, \tag{1.96}$$

где A — постоянная. Полученный результат можно выразить через двумерное преобразование Фурье функции $T(u,v)$:

$$g(\omega_x,\omega_y) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\tilde{T}(u,v)\exp[-i(\omega_x u+\omega_y v)]\,\mathrm{d}u\,\mathrm{d}v, \tag{1.97}$$

где

$$\tilde{T}(u,v) = Ae^{ikr_0}T(u,v). \tag{1.98}$$

Таким образом, на поверхности сферы с центром в начале координат комплексная амплитуда дифракционного поля пропорциональна преобразованию Фурье функции пропускания $T(u,v)$. С учетом этого факта проведем качественный анализ влияния дискретизации и квантования на характеристики ДОЭ.

В случае ДОЭ, функция пропускания которого подверглась дискретизации с прямоугольным растром (1.62), выражение для поля в выходной плоскости имеет вид

$$g(x,y,z) = g(\omega_x,\omega_y) = \\ = A\frac{e^{ikR_0}}{R_0}\sum_{j_1=0}^{N_1-1}\sum_{j_2=0}^{N_2-1}T(u_{j_1},v_{j_2})\int\limits_{u_{j_1}}^{u_{j_1+1}}\int\limits_{v_{j_2}}^{v_{j_2+1}}e^{-i\omega_x u}e^{-i\omega_y v}\,\mathrm{d}u\,\mathrm{d}v, \tag{1.99}$$

где $u_{j_1} = -A_1 + j_1(2A_1/N_1)$, $v_{j_2} = -A_2 + j_2(2A_2/N_2)$; $\omega_x = k\alpha$, $\omega_y = k\beta$; $2A_1$, $2A_2$ — размеры оптического элемента; N_1, N_2 — число отсчетов по осям координат. После несложного преобразования предыдущее выражение приобретает вид

$$g(\omega_x,\omega_y) = A\frac{e^{ikR_0}}{R_0}\exp\left[-i\omega_x\left(-A_1+\frac{A_1}{N_1}\right) - \right. \\ \left. - i\omega_y\left(-A_2+\frac{A_2}{N_2}\right)\right]P(\omega_x,\omega_y)Q(\omega_x,\omega_y), \tag{1.100}$$

где

$$Q(\omega_x,\omega_y) = \sum_{j_1=0}^{N_1-1}\sum_{j_2=0}^{N_2-1}T(u_{j_1},v_{j_2})e^{-ij_1\omega_x 2A_1/N_1}e^{-ij_2\omega_y 2A_2/N_2}, \tag{1.101}$$

$$P(\omega_x,\omega_y) = 4\frac{A_1A_2}{N_1N_2}\operatorname{sinc}\left[\frac{\omega_x A_1}{N_1}\right]\operatorname{sinc}\left[\frac{\omega_y A_2}{N_2}\right]. \tag{1.102}$$

Функция $Q(\omega_x,\omega_y)$ (1.101) представляет собой двумерную периодическую функцию с периодом по осям $\pi N_1/A_1$ и $\pi N_2/A_2$, соответственно. Эта функция описывает

хорошо известный эффект размножения спектров, возникающий при дискретизации исходного сигнала. При увеличении N_1, N_2 — числа точек дискретизации исходного ДОЭ, расстояние между различными порядками возрастает. На практике число отсчетов необходимо выбирать таким образом, чтобы различные порядки в спектре были пространственно разнесены и не перекрывали друг друга. Функция $P(\omega_x, \omega_y)$ по своей форме описывает дифракцию света на прямоугольном растре при дискретизации ДОЭ. Наличие этой функции уменьшает долю энергии, приходящуюся на дополнительные спектры. При увеличении числа точек дискретизации на ДОЭ эта доля энергии убывает. Характер убывания величины P с ростом N_1, N_2 определяется зависимостью $P \sim 1/MN$.

Рассмотрим теперь физические эффекты, возникающие при квантовании фазовой функции ДОЭ. Пусть $\varphi(u, v)$ — фазовая функция ДОЭ для фокусировки пучка с комплексной амплитудой $W(u, v)$ в область S фокальной плоскости. ДОЭ будем рассматривать как добавку к линзе с фокусным расстоянием f. Критерием качества работы квантованного ДОЭ является дифракционная эффективность

$$E = \frac{\iint\limits_S I(x, y)\,\mathrm{d}x\,\mathrm{d}y}{\iint\limits_D |W(u, v)|^2\,\mathrm{d}u\,\mathrm{d}v}, \tag{1.103}$$

где $W(u, v)$ — амплитуда освещающего пучка, $I(x, y)$ — интенсивность в фокальной области, D — апертура ДОЭ, S — область фокусировки. Величина E характеризует долю энергии освещающего пучка, фокусируемую в требуемую область фокальной плоскости.

Квантование фазовой функции ДОЭ может быть описано как результат применения нелинейного преобразования $\hat{\Phi}(\varphi)$ к фазовой функции ДОЭ $\varphi(u, v)$. Нелинейное предыскажение в случае квантования фазовой функции $\varphi(u, v)$, приведенной к интервалу 2π по M уровням, имеет вид

$$\hat{\Phi}(\varphi) = \mathrm{int}\left[\frac{\varphi}{\Delta}\right]\Delta; \qquad \Delta = 2\pi/M, \tag{1.104}$$

где $\mathrm{int}[x]$ — операция взятия целой части числа x; M — число уравней квантования фазы (рис. 1.57). В случае квантования комплексная функция пропускания ДОЭ принимает вид

$$T(u, v) = \exp\left\{i\hat{\Phi}\left[\varphi(u, v)\right]\right\}. \tag{1.105}$$

Разложим функцию пропускания в ряд Фурье по переменной $\varphi(u, v)$:

$$T(u, v) = \sum_{n=-\infty}^{n=\infty} c_n \exp\left[i\frac{n}{m}\varphi(u, v)\right], \tag{1.106}$$

где $2\pi m$ — максимальная фаза на оптическом элементе,

$$c_n = \frac{1}{2\pi m}\int\limits_0^{2\pi m} \exp\left(i\hat{\Phi}(\xi) - i\frac{n}{m}\xi\right)\mathrm{d}\xi. \tag{1.107}$$

С учетом (1.106) амплитуда поля в фокальной плоскости ДОЭ имеет вид

$$F(x, y) = \sum_{n=-\infty}^{n=\infty} c_n F_n(x, y). \tag{1.108}$$

В приближении Френеля–Кирхгофа

$$F_n(\mathbf{x}) = \frac{1}{i\lambda f}\exp\left(i\frac{k}{2f}x^2\right)\iint\limits_D W(u,v)\exp\left(i\frac{n}{m}\varphi(\mathbf{u}) - i\frac{kxu}{f}\right)\mathrm{d}^2u. \tag{1.109}$$

Согласно выше приведенным формулам ДОЭ, формирует дифракционные порядки $F_n(x,y)$, $n=-\infty,\infty$. Требуемый процесс фокусировки в рабочем (полезном) порядке дифракции m описывается величиной $F_m(x,y)$. Из общего вида фазовой функции следует, что слагаемые с другими n соответствуют расфокусированным изображениям с геометрическими размерами, в n/m раз большими.

В общем случае расфокусированные изображения $F_n(x,y)$ влияют на $F_m(x,y)$, при этом (см. главу 5) получаем следующую оценку интенсивности:

$$I(x,y) \geqslant |c_n|^2 I_n(x,y), \tag{1.110}$$

где

$$I(x,y) = |F(x,y)|^2, \quad I_n(x,y) = |F_n(x,y)|^2. \tag{1.111}$$

Для энергетической эффективности ДОЭ имеет место следующая оценка:

$$E \leqslant |c_n|^2 E_0, \tag{1.112}$$

где E_0 — энергетическая эффективность ДОЭ с непрерывной (неквантованной) фазовой функцией $\varphi(u,v)$.

Рассмотрим далее квантование фазовой функции оптического элемента с максимальной фазой, равной 2π (т.е. $m=1$). Пусть число уровней квантования равно M. В этом случае выражение для c_n принимает следующий вид:

$$c_n = \frac{1}{2\pi}\sum_{k=0}^{M-1}\int\limits_{(2\pi k)/M}^{(2\pi(k+1))/M}\exp\left(i\frac{2\pi k}{M} - in\xi\right)\mathrm{d}\xi. \tag{1.113}$$

После вычисления интеграла получаем

$$c_n = i\frac{\left(\exp\left(-\dfrac{2\pi ni}{M}\right)-1\right)}{2\pi n}\sum_{k=0}^{M-1}\left[\exp\left(-\frac{2\pi i(n-1)}{M}\right)\right]^k. \tag{1.114}$$

Используя формулу геометрической прогрессии, можно упростить полученное выражение:

$$c_n = i\frac{\left(\exp\left(-\dfrac{2\pi ni}{M}\right)-1\right)}{2\pi n}\,\frac{1-\exp\left(-2\pi i(n-1)\right)}{1-\exp\left(-\dfrac{2\pi i(n-1)}{M}\right)}. \tag{1.115}$$

Раскрывая неопределенность, получаем окончательное выражение для c_n:

$$c_n = \begin{cases} \dfrac{iM}{2\pi n}\left[\exp\left(-\dfrac{2\pi in}{M}\right)-1\right], & \text{если } \dfrac{(n-1)}{M} \text{ — целое число,} \\ 0, & \text{если } \dfrac{(n-1)}{M} \text{ — нецелое число.} \end{cases} \tag{1.116}$$

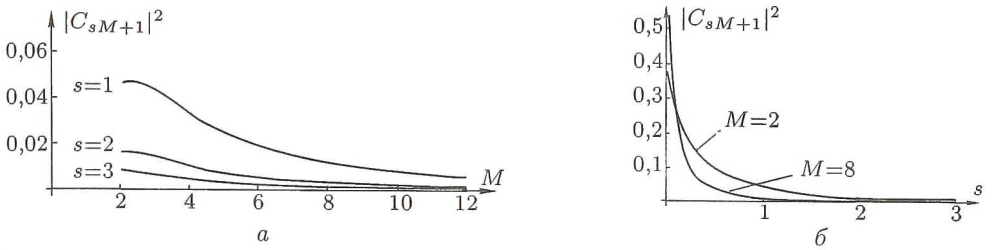


Рис. 1.60. Зависимость величины коэффициента c_{sM+1} от числа уровней квантования M и параметра s

Рассмотрим случай $n = sM + 1$ (при других n, согласно (1.116), $c_n = 0$), где s — произвольное целое число, определяющее ненулевые дифракционные порядки:

$$c_{sM+1} = \frac{i}{2\pi\,(s + 1/M)} \left[\exp\left(-\frac{2\pi i}{M}\right) - 1\right]. \qquad (1.117)$$

Как видно из рис. 1.60а, при малом числе уровней квантования (M = 2–6) существенный вес имеют первые 2–3 порядка дифракции. При увеличении числа уровней квантования фазы M все коэффициенты, кроме первого, стремятся к нулю. Это означает, что вклад побочных порядков становится очень малым.

1.6. Компьютерное проектирование ДОЭ

Главные этапы компьютерного синтеза ДОЭ. Создание ДОЭ с помощью компьютера — процесс сложный. Он включает в себя решение прямой и обратной задач теории дифракции, требует применения оптико-электронной аппаратуры регистрации изображений и связан с использованием различных технологий получения фазового рельефа. Каждый из этих этапов существенно влияет на качество изготовляемого ДОЭ и заранее трудно представить точные значения его характеристик, таких, например, как толщина линии фокусировки или дифракционная эффективность. В этой связи процесс создания ДОЭ с помощью компьютера становится итерационным: исследуя характеристики синтезируемого элемента на различных этапах его изготовления, мы вырабатываем рекомендации к их улучшению (рис. 1.61). Эта ситуация типична для процесса проектирования вообще и для автоматизированного проектирования в частности. Важную роль играет исследование полученных характеристик ДОЭ с помощью экспериментов. В компьютерной оптике требуется проводить два вида экспериментов: вычислительные и натурные, оба требуют применения ЭВМ, в первом случае по определению, а во втором в силу информативности объекта исследования и сложности обработки оптических изображений.

Вычислительный эксперимент в компьютерной оптике рассмотрим на примере синтеза фокусатора. Пусть в результате решения обратной задачи фокусировки получена фазовая функция ДОЭ и записана в виде массива чисел в памяти ЭВМ. Для того, чтобы убедиться в успешности решения обратной задачи и прогнозировать характеристики создаваемого фокусатора, не обязательно сразу изготавливать его в виде оптического элемента и проводить натурный эксперимент. Натурный эксперимент позволит исследовать характеристики оптического элемента, полученные в результате совместного действия всех этапов его изготовления, включая и технологию. Конечно мы убедимся в работоспособности (или неработоспособности) фокусатора, но провести тонкое исследование распределения интенсивности излучения

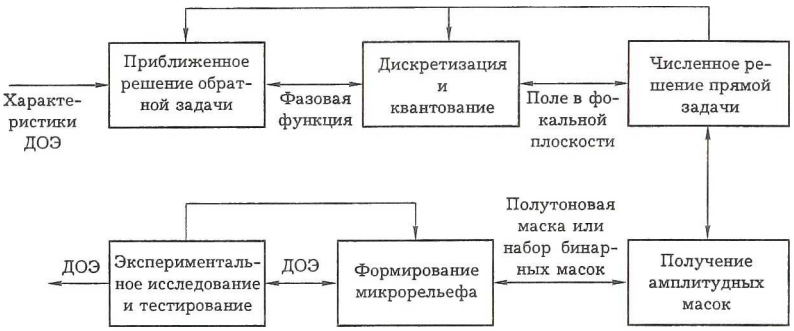


Рис. 1.61. Главные этапы компьютерного синтеза ДОЭ

в фокальной плоскости в зависимости от полученного решения обратной задачи мы не сможем. Чтобы получить интересующие нас данные, следует провести численный эксперимент с массивом чисел, описывающих синтезированную фазовую функцию фокусатора. Причем результаты следует получить в виде, привычном для оптика-экспериментатора, а это — полутоновые распределения яркости на экране монитора, графики кривых, компактные таблицы обработанных экспериментальных данных. Для решаемой задачи желательно также представить трехмерные распределения в изометрии и цвете. Именно сочетание математических методов и программ решения разнообразных задач дифракции со средствами визуализации изображений позволяет проводить вычислительные эксперименты при изготовлении ДОЭ. На рис. 1.62 показан один из результатов вычислительного эксперимента с фокусатором в полукольцо. Данный фокусатор является дополнением к линзе (т.е. перераспределяет в полукольцо сходящийся сферический пучок света) и рассчитывается в приближении геометрической оптики. Поле в фокальной области фокусатора вычисляется в дифракционном приближении с использованием интеграла Кирхгофа, как сумма аналитически рассчитываемых полей, создаваемых отдельными модулями ДОЭ линейной, кольцевой или прямоугольной формы. Вместо фокусировки в полукольцо получается фокусировка в нечеткое кольцо с большими перепадами интенсивности вдоль него.

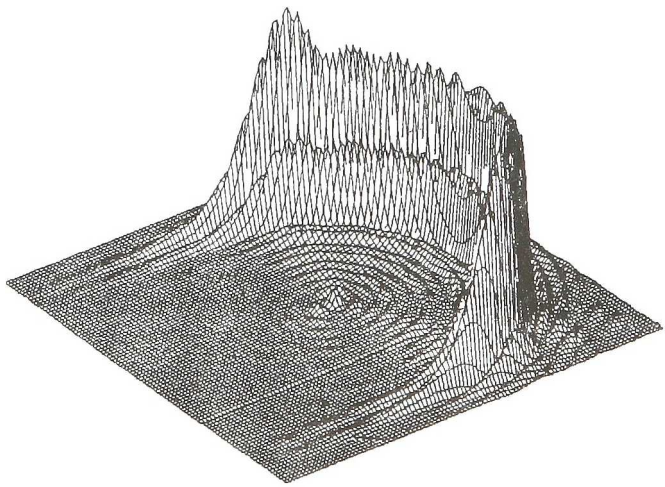


Рис. 1.62. Вычислительный эксперимент с фокусатором в полукольцо

Следующим этапом вычислительного эксперимента является исследование работы оптической системы, включающей проектируемые ДОЭ. Оптическая система при моделировании представляется в виде источников света, набора оптических сред, отражающих и преломляющих поверхностей и ДОЭ. Для моделирования используются уравнения распространения луча и расчет лучевых траекторий через оптическую систему от источника до плоскости регистрации. Чем больше лучей будет проведено через систему, тем точнее результат моделирования.

Различные подходы к формированию микрорельефа. ДОЭ представляет собой зонную пластинку с микрорельефом, отображающим фазовую функцию $\varphi(u, v)$. Существует множество автоматизированных технологий, реализующих запись фазовой функции на оптическую среду. Например, делительные машины позволяют непосредственно получать дифракционные решетки (в т.ч. с блеском), и поскольку в данном случае границы зон представляют собой прямые линии, управление делительной машиной не требует реализации сложных алгоритмов. Существует также возможность непосредственного перевода на физическую среду радиально-симметричных и даже более сложных фазовых функций с помощью прецизионных механических станков с программным управлением. Такая возможность реально используется для изготовления ДОЭ субмиллиметрового и дальнего ИК диапазона.

Внимание специалистов в области дифракционной оптики в последнее время привлекают технологии формирования микрорельефа, основанные на достижениях микроэлектроники. Определяющими в таком выборе являются следующие факторы:

— близость геометрических размеров ДОЭ и микросхем;

— близость требуемых характеристик пространственного разрешения;

— высокий уровень автоматизации и возможность реализации сложных топологий зон;

— возможность получения многих градаций фазового микрорельефа;

— наличие множества технологических участков микроэлектроники и возможность аренды оборудования;

— высокая степень надежности технологических автоматов и повторяемость результатов;

— возможность использования материалов, применяемых в микроэлектронике (субстраты, резисты и т.п).

Однако, при этом важно понимать, что применение микроэлектронных технологий — не единственный способ создания ДОЭ. Существуют и другие технологии перевода многоградационной амплитудной маски, соответствующей функции $\varphi(u, v)$ в фазовый микрорельеф.

Проведение натурных экспериментов в компьютерной оптике. Для ввода оптических изображений в цифровую память можно использовать разнообразные преобразователи оптического сигнала в электрический: телевизионные трубки на основе видиконов, фотодиодные матрицы, ПЗС-матрицы. Электрический сигнал затем должен быть преобразован в цифровой код. Основная трудность состоит в разработке цифровых методов анализа оптических сигналов. Здесь опять, как и в задаче синтеза ДОЭ, мы сталкиваемся с необходимостью обрабатывать на ЭВМ двумерные массивы чисел и решать некорректные обратные задачи. Решение осложняется тем, что зарегистрированный и записанный в память ЭВМ двумерный сигнал содержит, помимо полезной информации, различного рода погрешности возмущения и искажения. Причин для появления мешающих факторов много: это неравномерность освещения регистрируемого изображения, вибрации, колебания напряжения в регистрирующей аппаратуре, квантование непрерывного сигнала и т.п.

Создание ДОЭ предъявляет достаточно жесткие требования к компьютеру как с точки зрения расчета и хранения фазовых функций, так и с точки зрения управления генератором изображений. В состав прикладного математического обеспечения систем синтеза ДОЭ должны входить следующие разделы:

— определение фазовой функции ДОЭ (в т.ч. на основе решения обратных задач дифракции);

— решение прямой задачи дифракции для расчета результатов работы ДОЭ (вычислительный эксперимент);

— проектирование и исследование оптической системы, включающей ДОЭ;

— формирование файлов для генератора изображений;

— программное обеспечение для управления генератором изображений;

— программы ввода и обработки экспериментальных данных, получаемых при аттестации ДОЭ.

Использование компьютера в создании ДОЭ иллюстрируется рис. 1.63.

1.7. Заключение

Из материала этой главы следует, что методы и технологии компьютерной оптики позволяют создавать ДОЭ с широкими функциональными возможностями и высокой дифракционной эффективностью. Однако ДОЭ уступают традиционным оптическим элементам по максимальным размерам апертуры и величине оптической силы. Максимальные размеры апертуры ограничены тремя факторами: физическим, математическим и технологическим.

Ограничение физического характера состоит в том, что при большой апертуре оптического элемента (большой оптической силе) перестают работать геометрооптическое приближение и приближение Кирхгофа, и расчет фазовых функций ДОЭ следует осуществлять в рамках электромагнитного подхода.

Ограничения математического характера состоят в отсутствии или большой трудоемкости методов решения прямой и обратной задач дифракции волн в общей постановке. Хотя необходимо отметить существенное продвижение в теории дифракции за последние годы, стимулированное задачами нетрадиционной оптики.

Наиболее жестким фактором, ограничивающим максимальные размеры апертуры ДОЭ, является технология их реализации. По существу, следует говорить о двух технологиях:

1) вычисление и передача 2-D массивов пикселов большого размера;
2) получение амплитудных масок и фазового микрорельефа.

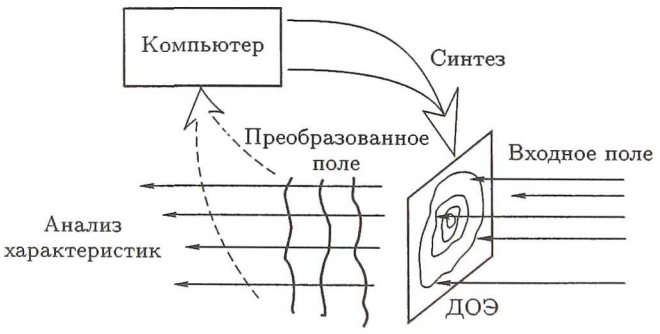


Рис. 1.63. Использование компьютера в создании ДОЭ

Рассмотрим для примера ДОЭ с минимальным размером зоны $\Delta = 0{,}5$ мкм, описываемого матрицей 10^{10} пикселов. Размеры апертуры ДОЭ при этом составят 50×50 мм. При использовании фотопостроителя, имеющего скорость передачи пикселов 25 МГц, потребуется более часа на получение амплитудной маски ДОЭ. Для получение ДОЭ с апертурой 250×250 мм необходима матрица пикселов $(1/4) \times \times 10^{12}$, для отображения которой потребуется 25 часов. При этом, исходя из постоянства минимального размера зоны, во втором случае мы получим элемент, оптическая сила которого приблизительно в 5 раз ниже, чем в первом случае. Желая сохранить оптическую силу, мы придем к необходимости получения минимального размера зоны $\Delta = 0{,}1$ мкм. При этом матрица пикселов будет иметь размер $\sim 10^{13}$, и для ее записи в виде амплитудной маски потребуется более 1000 часов. Исходя из стоимости рабочего времени на электронном литографе, мы приходим к выводу об экономической нецелесообразности создания ДОЭ с большой апертурой. Конечно, во внимание могут приниматься не только экономические соображения, но и уникальность характеристик создаваемого ДОЭ.

Литература к главе 1

1. Борн М., Вольф Э. *Основы оптики* (М.:Наука, 1973)
2. Lohmann A. W., Paris D. P. *Appl. Opt.* **6**(10) 1739 (1967)
3. Lesem L. B., Hirsh P. M., Jordan J. A. *IBM J. Res. Develop.* **13**(3) 150 (1969)
4. Kirk J. P., Jones A. L. *J. Opt. Soc. Am. A.* **61**(8) 1023 (1971)
5. Lee W. H., Binary *Appl. Opt.* **18**(21) 3661 (1979)
6. Lee W. H. *Appl. Opt.* **9**(3) 639 (1970)
7. Chu D. C., Fienup J. R. *Opt. Eng.* **13**(3) 189 (1974)
8. Gallagher N. C., Liu B. *Appl. Opt.* **12**(10) 2328 (1973)
9. Gerchberg R. W., Saxton W. O. *Optik* **35** 237 (1972)
10. Burckhardt C. B. *Appl. Opt* **9**(8) 1949 (1970)
11. Golub M. A. *Seminar on Digital image processing in medicine, remote sensing and visualisation of information* **77** (1992)
12. Fienup J. R. *Optical Engineering* **19** 297 (1980)
13. Fienup J. R. *Applied optics.* **21**(15) 2758 (1982)
14. Старк Г. *Реконструкция изображений* (М.:Мир, 1992)
15. Wyrowski F., Bryngdahl O. *Report on Progress in Physics* **54**(12) 1481 (1991)
16. Гончарский А. В. *Компьютерная оптика* **1** 19 (1987)
17. Данилов В. А., Попов В. В., Прохоров А. М., Сагателян Д. М., Сисакян И. Н., Сойфер В. А. *Письма в ЖТФ* **8**(13) 810 (1982)
18. Данилов В. А., Кулькин К. А., Сисакян И. Н. *Компьютерная оптика* **13** 3 (1993)
19. Doskolovich L. L., Khonina S. N., Kotlyar V. V., Nikolsky I. V., Soifer V A. *Optical and Quantum Electronics* **25** 801 (1993)
20. Duparre M., Golub M. A., Ludge B., Pavelyev V. S., Soifer V. A., Uspleniev G. V., Volotovskii S. G. *Applied Optics* **34**(14) 2489 (1995)
21. Golub M. A., Sisakyan I. N., Soifer V. A. *Optics and Lasers in Engineering* **15** 297 (1991)
22. Prokhorov A. M., Sisakjan I. N., Golub M. A., Soifer V. A., Karpeev S. V., Goncharsky A. V. *Optical Phase Element for Focussing Monochromatic Radiation. Патент Великобритании. Patent GB* 2185126, *приоритет* 30.04.85., *Printing* 08.07.87, *Patent publ.* 24.05.89.
23. Данилов В. А., Попов В. В. *Оптические элементы, фокусирующие когерентное излучение в произвольную фокальную линию* (М.: ФИАН СССР, 1983)
24. Голуб М. А., Карпеев С. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Письма в ЖТФ* **7**(10) 619 (1981)
25. Голуб М. А., Живописцев Е. С., Карпеев С. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Доклады АН СССР* **253**(5) 1104 (1980)

26. Ларионов Н. П., Лукин А. В., Мустафин К. С. *Оптика и спектроскопия* **32**(2) 396 (1972)
27. Сисакян И. Н., Сойфер В. А. *Компьютерная оптика* **1** 5 (1987)
28. Сойфер В. А. *Введение в дифракционную микрооптику* (Самара:СГАУ, 1996).
29. Ярославский Л. П., Мерзляков Н С *Методы цифровой голографии* (М.:Наука, 1977)
30. Кабо И. Я. *Проблемы голографии* **2** 25 (1973)
31. Мандросов В. И. *Оптика и спектроскопия* **32**(1) 174 (1972)
32. Голуб М. А. *Исследование характеристик и реализация когерентно-оптических пространственных фильтров, синтезируемых с помощью ЭВМ. Диссертация на соискание ученой степени кандидата физико-математических наук* (Куйбышев, 1981)
33. Golub M. A., Sisakyan I. N., Soifer V. A. *Proceedings SPIE* **1334** 188 (1990)
34. Кловский Д. Д., Сойфер В. А. *Обработка пространственно-временных сигналов* (М.:Связь, 1976)

ГЛАВА 2

ИТЕРАТИВНЫЕ МЕТОДЫ РАСЧЕТА ДОЭ

2.1. Введение

В задачах обработки и восстановления изображений разработаны и широко применяются итерационные (итеративные) методы. Обзор большинства из них можно найти в [1]. Итеративные методы можно разделить на несколько классов: 1) методы на основе сжимающих или нерасширяющих операторов: алгоритм Бургера–Ван Циттерта, алгоритм сверхразрешения Герчберга–Папулиса; 2) методы на основе проекций на выпуклые множества — алгоритмы Старка и Юлы; 3) итеративные алгоритмы с псевдодифференциальными операторами ограничений: алгоритм Герчберга–Секстона и алгоритм Фиенапа; 4) методы нелинейного программирования или оптимизации: алгоритм наискорейшего спуска, алгоритм сопряженных градиентов, метод модифицированных функций Лагранжа.

Применение этих методов для решения задач расчета ДОЭ требует специального исследования. Это связано с тем, что, как правило, задачи восстановления искаженных изображений сводятся к решению линейного интегрального уравнения типа свертки с аддитивным шумом. А задачи синтеза ДОЭ сводятся к решению нелинейного интегрального уравнения, нелинейность которого связана с операцией взятия модуля от комплексной амплитуды когерентного света.

В этой главе рассматриваются итеративные методы решения обратных задач скалярной теории дифракции применительно к синтезу ДОЭ. Результаты, которые получаются с помощью итеративных методов, являются квазиоптимальными, так как они приводят к достижению локального минимума функционала-критерия или целевой функции. И сами итеративные методы получаются в результате решения вариационной задачи на экстремум целевой функции. В качестве такой функции, как правило, используется среднеквадратичное отклонение заданной амплитуды светового поля в некоторой плоскости пространства от рассчитанной. Иногда вместо амплитуд сравниваются интенсивности, а вместо среднеквадратичного критерия выбираются критерии более высокого порядка. Итеративные методы, используемые в этой главе, можно разделить на две группы: параметрические и градиентные. В параметрических алгоритмах один или два параметра, от которых зависит скорость сходимости алгоритма, остаются постоянными в течение нескольких итераций. В градиентных алгоритмах (сопряженного градиента или наискорейшего спуска) оптимальное значение шага вычисляется на каждой итерации. Кроме известных однопараметрических методов расчета ДОЭ, рассматриваются также двухпараметрические алгоритмы, полученные на основе минимизации функционала-критерия с регуляризующим слагаемым.

2.2. Алгоритм уменьшения ошибки

Ниже рассматриваются алгоритмы решения нелинейного интегрального уравнения Френеля, предназначенного для расчета фазовых оптических элементов, формирующих произвольное заданное распределение интенсивности когерентного монохроматического света в некоторой плоскости, перпендикулярной оптической оси.

Изучаемые алгоритмы являются адаптивными, т.к. новая оценка искомой функции на каждой итерации выбирается не только в соответствии с требуемой функцией интенсивности, но и в зависимости от предыдущей оценки. Эти алгоритмы называются также параметрическими, т.к. скорость их сходимости зависит от выбора конкретных значений некоторых весовых или регуляризационных параметров.

В скалярной теории дифракции комплексная амплитуда волны в плоскости оптического элемента $W(u,v) = A(u,v)\exp i\varphi(u,v)$ связана с комплексной амплитудой волны $F(\xi,\eta) = B(\xi,\eta)\exp i\psi(\xi,\eta)$ в плоскости формирования требуемого распределения интенсивности $I_0(\xi,\eta)$ через интегральное преобразование [2]:

$$F(\xi,\eta) = \frac{ik}{2\pi z} e^{ikz} \iint\limits_{-\infty}^{\infty} W(u,v)\, H(u-\xi, v-\eta, z)\, \mathrm{d}u\, \mathrm{d}v, \tag{2.1}$$

где

$$H(u-\xi, v-\eta, z) = \exp\left[\frac{ik}{2z}\left\{(u-\xi)^2 + (v-\eta)^2\right\}\right] \tag{2.2}$$

— функция импульсного отклика свободного пространства в приближении Френеля, z — расстояние между ДОЭ и плоскостью наблюдения. В уравнении (2.1) комплексная амплитуда $W(u,v)$ в приближении тонкого оптического элемента (приближение транспаранта), которое не учитывает рефракцию лучей, равна произведению комплексной амплитуды $W_0(u,v)$ на собственную функцию пропускания $T(u,v)$ ДОЭ: $W(u,v) = W_0(u,v)\,T(u,v)$.

Поскольку мы предполагаем рассматривать только фазовые оптические элементы, функция пропускания ДОЭ выбрана в виде $T(u,v) = \exp[ig(u,v)]$, где $g(u,v)$ — заданная фаза ДОЭ (см. рис. 2.1). Задачу расчета фазовой функции ДОЭ $g(u,v)$ можно свести к решению нелинейного интегрального уравнения

$$I_0(\xi,\eta) = |F(\xi,\eta)|^2 = \left| \frac{k}{2\pi z} \iint\limits_{-\infty}^{\infty} A_0(u,v)\, e^{i\varphi(u,v)} H(u-\xi, v-\eta, z)\, \mathrm{d}u\, \mathrm{d}v \right|^2, \tag{2.3}$$

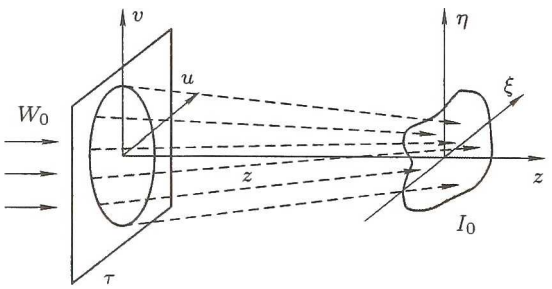


Рис. 2.1. Схема формирования изображения при помощи ДОЭ

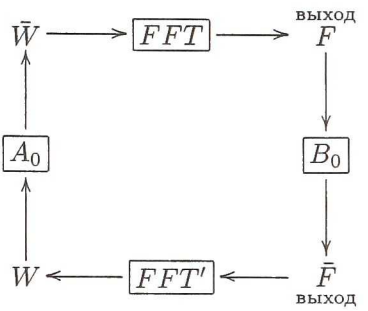


Рис. 2.2. Схема алгоритма уменьшения ошибки Герчберга–Секстона

где $I_0(\xi,\eta)$ — заданная интенсивность в области изображения, $A_0(u,v)$ — амплитуда освещающего пучка, $\varphi(u,v) = g(u,v) + g_0(u,v)$, $g_0(u,v)$ — фаза освещающего пучка. Итеративный метод расчета фазы $\varphi(u,v)$, а вместе с ней фазы $g(u,v)$, состоит в решении уравнения (2.3) методом последовательных приближений. Алгоритм Герчберга–Секстона (ГС), или алгоритм уменьшения ошибки [3], содержит следующие шаги (см. его схему на рис. 2.2):

1) выбирается начальная оценка фазы $\varphi_0(u,v)$;

2) осуществляется интегральное преобразование функции $A_0(u,v)\exp[i\varphi_0(u,v)]$ при помощи уравнения (2.1);

3) результирующая комплексная амплитуда $F(\xi,\eta)$ в плоскости формирования изображения заменяется на $\bar{F}(\xi,\eta)$ по правилу

$$\bar{F}(\xi,\eta) = B_0(\xi,\eta)\, F(\xi,\eta)\, |F(\xi,\eta)|^{-1}, \tag{2.4}$$

где $B_0(\xi,\eta) = \sqrt{I_0(\xi,\eta)}$;

4) вычисляется преобразование, обратное (2.1) относительно функции $\bar{F}(\xi,\eta)$:

$$W(u,v) = \frac{ik}{2\pi z} e^{-ikz} \iint\limits_{-\infty}^{\infty} \bar{F}(\xi,\eta)\, H^*(\xi-u,\eta-v,z)\, \mathrm{d}\xi\, \mathrm{d}\eta; \tag{2.5}$$

5) полученная комплексная амплитуда $W(u,v)$ в плоскости ДОЭ заменяется на $\bar{W}(u,v)$ по правилу

$$\bar{W}(u,v) = \begin{cases} A_0(u,v)\, W(u,v)\, |W(u,v)|^{-1}, & (u,v) \in Q, \\ 0, & (u,v) \notin Q, \end{cases} \tag{2.6}$$

где Q — форма апертуры ДОЭ;

6) переход к шагу 2.

Эта процедура повторяется до тех пор, пока ошибки δ_F и δ_W не перестанут значительно меняться:

$$\delta_F^2 = \frac{\iint\limits_{-\infty}^{\infty} [|F(\xi,\eta)| - B_0(\xi,\eta)]^2\, \mathrm{d}\xi\, \mathrm{d}\eta}{\iint\limits_{-\infty}^{\infty} B_0^2(\xi,\eta)\, \mathrm{d}\xi\, \mathrm{d}\eta}, \tag{2.7}$$

$$\delta_W^2 = \frac{\iint\limits_{-\infty}^{\infty} [|W(u,v)| - A_0(u,v)]^2\, \mathrm{d}u\, \mathrm{d}v}{\iint\limits_{-\infty}^{\infty} A_0^2(u,v)\, \mathrm{d}u\, \mathrm{d}v}. \tag{2.8}$$

Алгоритм ГС называют алгоритмом уменьшения ошибки потому, что было показано [4], что ошибки (2.7) и (2.8) с ростом числа итераций не возрастают. Кроме того, было показано, что алгоритм ГС является вариантом градиентного метода [5] или градиентного метода наискорейшего спуска [6], при помощи которого можно минимизировать функционал среднеквадратичного отклонения амплитуды восстановленного изображения от заданного значения:

$$\varepsilon_0 = \iint\limits_{-\infty}^{\infty} [|F(\xi,\eta)| - B_0(\xi,\eta)]^2\, \mathrm{d}\xi\, \mathrm{d}\eta. \tag{2.9}$$

Заметим, однако, что процесс сходимости алгоритма ГС характеризуется эффектом стагнации: в ходе нескольких начальных итераций ошибка δ_F (или δ_W) быстро уменьшается, а все последующие итерации не приводят к ее заметному уменьшению. Эффект стагнации показывает, что алгоритм достигает локального минимума функционала (2.9), и в этом смысле он является квазиоптимальным. Чтобы увеличить скорость сходимости процедуры и частично избежать эффекта стагнации, используется целый ряд адаптивных алгоритмов, в которых вводятся некоторые параметры, контролирующие скорость сходимости. Поэтому иногда эти алгоритмы называют параметрическими.

2.3. Алгоритм входа-выхода

В ряде работ Фиенапом [5, 7, 8] разработан метод входа-выхода, в котором рассматривается проблема восстановления фазы по измерению одной интенсивности в области пространственного спектра. Рассмотрим этот алгоритм.

Функции комплексной амплитуды $W(u,v)$ в области ДОЭ и $F(\xi,\eta)$ в плоскости Фурье (в задней фокальной плоскости сферической линзы), в которой формируется изображение, связаны между собой при помощи двумерного преобразования Фурье:

$$F(\xi,\eta) = \frac{ik}{2\pi f}\iint\limits_{-\infty}^{\infty} W(u,v)\exp\left[-i\frac{k}{f}(u\xi+v\eta)\right]\mathrm{d}u\,\mathrm{d}v, \tag{2.10}$$

где f — фокусное расстояние тонкой линзы, формирующей фурье-спектр.

Переход от преобразования Френеля (2.1) к преобразованию Фурье (2.10) означает, что в последнем случае следует использовать комбинацию ДОЭ плюс линза.

Итеративное решение задачи восстановления фазы светового поля при помощи алгоритма входа-выхода (одной из его модернизаций) имеет вид

$$\bar{F}_{n+1}(\xi,\eta) = \bar{F}_n(\xi,\eta) + \beta\delta F_n(\xi,\eta), \tag{2.11}$$

где β — некоторый параметр $0{,}5 \leqslant \beta \leqslant 1$,

$$\delta F_n(\xi,\eta) = \begin{cases} 0, & (\xi,\eta)\notin\gamma, \\ -F_n(\xi,\eta), & (\xi,\eta)\in\gamma, \end{cases} \tag{2.12}$$

γ — область нарушения ограничений, накладываемых на заданную функцию поля,

$$F_n(\xi,\eta) = \mathcal{F}D_A\mathcal{F}^{-1}\left\{\bar{F}_n(\xi,\eta)\right\} \tag{2.13}$$

— выходная функция, $\mathcal{F}$ и $\mathcal{F}^{-1}$ — прямое и обратное преобразования Фурье, D_A — ограничивающий оператор в области ДОЭ:

$$D_A W_n = A_0\frac{W_n}{|W_n|}. \tag{2.14}$$

Специально для решения проблемы синтеза ДОЭ предлагается выбирать приращение функции δF_n в алгоритме входа-выхода в следующем виде [8]:

$$\delta F = \left[B_0 F\,|F|^{-1} - F\right] + \left[B_0 F\,|F|^{-1} - B_0\bar{F}\,|\bar{F}|^{-1}\right], \tag{2.15}$$

где $B_0^2(\xi,\eta)$ — заданное распределение интенсивности в фурье-плоскости, $\bar{F}$ и F — входная и выходная функции или функции комплексной амплитуды в плоскости фурье-изображения после и до выполнения ограничений.

В работе [9] сообщается о новом варианте итеративного алгоритма входа-выхода для синтеза ДОЭ, в котором предполагается, что фаза независима от модуля комплексной амплитуды света $F(\xi,\eta)$. Показано, что оптимальный выбор модуля и фазы на $(n+1)$-м шаге итерации принимает вид ($\bar{F}(\xi,\eta)=\bar{F}$)

$$\begin{aligned} \left|\bar{F}_{n+1}\right| &= B_0 + \beta\left[|F_n| - B_0\right], \\ \arg\bar{F}_{n+1} &= \arg F_n + \alpha\left[\arg F_n - \arg\bar{F}_n\right]. \end{aligned} \tag{2.16}$$

На численных примерах в работе [9] также показано, что лучшие значения α и β равны 2. Заметим, что итеративный алгоритм, который использует первое уравнение в (2.16), сходен с более ранним алгоритмом, описанным в работах [10, 11]. Этот алгоритм авторами [9] назван маятниковым. На наш взгляд, он отличается от начального варианта алгоритма входа-выхода, так как явно учитывает ограничение на амплутуду B_0 в фурье-плоскости.

2.4. Адаптивно-аддитивный алгоритм

Первоначально алгоритм входа-выхода [5, 7] применялся для восстановления фазы светового поля на основе измерений одной интенсивности в плоскости пространственных частот. Такая проблема является характерной для звездной интерферометрии. Этот же алгоритм был успешно применен для расчета фазы ДОЭ [8]. Однако этот алгоритм недостаточно теоретически обоснован.

Ниже показано, что алгоритм (2.16) (только первое уравнение), который далее называется адаптивно-аддитивным алгоритмом (АА-алгоритмом), можно получить при минимизации некоторого функционала-критерия. Можно также показать, что для АА-алгоритма некоторая характерная среднеквадратичная ошибка с ростом числа итераций не возрастает.

Рассмотрим более подробно замену (2.4) для алгоритма ГС. Эта замена предполагает, что амплитуда волны $|F_n(\xi,\eta)|$ в области наблюдения, полученная на n-м итеративном шаге заменяется на заданное значение амплитуды $B_0(\xi,\eta)$. Заметим, что, хотя преобразование Фурье (Френеля) от функции с ограниченной несущей есть аналитическая целая функция экспоненциального типа, функция $B_0(\xi,\eta)$ может не быть аналитической функцией и может быть задана произвольно.

Следовательно, имеет смысл вместо замены (2.4) попробовать использовать замену, в которой обе функции (заданная и рассчитанная аналитическая) участвуют в виде линейной комбинации с разными весами [12]:

$$\left|\bar{F}_n(\xi,\eta)\right| = \left|\gamma B_0(\xi,\eta) + (1-\gamma)\left|F_n(\xi,\eta)\right|\right|, \tag{2.17}$$

В этом случае замена (2.4) в итеративном ГС-алгоритме принимает вид

$$\bar{F}_n(\xi,\eta) = \left|\bar{F}_n(\xi,\eta)\right| F_n(\xi,\eta)\left|F_n(\xi,\eta)\right|^{-1}. \tag{2.18}$$

Диапазон значений параметра γ находится из условия, что среднеквадратичное отклонение ε_0 (2.9) не увеличивается при использовании замены (2.17). То есть мы

предполагаем, что выполняется следующее условие:

$$\bar{\varepsilon}_0 = \iint\limits_{-\infty}^{\infty} \left[\left|\bar{F}_n\right| - B_0\right]^2 \mathrm{d}\xi\,\mathrm{d}\eta = \iint\limits_{-\infty}^{\infty} \left[\gamma B_0 + (1-\gamma)\left|F_n\right| - B_0\right]^2 \mathrm{d}\xi\,\mathrm{d}\eta =$$

$$= (1-\gamma)^2 \iint\limits_{-\infty}^{\infty} \left[\left|F_n\right| - B_0\right]^2 \mathrm{d}\xi\,\mathrm{d}\eta \leqslant \varepsilon_0 = \iint\limits_{-\infty}^{\infty} \left[\left|F_n\right| - B_0\right]^2 \mathrm{d}\xi\,\mathrm{d}\eta. \quad (2.19)$$

Из этого видно, что весовой коэффициент следует выбирать из условия $0 \leqslant \gamma \leqslant 2$. Заметим, что для $\gamma = 1$ замена (2.17) переходит в уравнение (2.4) алгоритма ГС, а для $\gamma = 2$ приводит к замене «зеркального отражения»:

$$\left|\bar{F}_n(\xi,\eta)\right| = \left|2B_0(\xi,\eta) - \left|F_n(\xi,\eta)\right|\right|. \quad (2.20)$$

В последнем случае ($\gamma = 2$) амплитуда $|F_n(\xi,\eta)|$ в плоскости наблюдения, рассчитанная на n-м шаге итераций, заменяется на «зеркальное отражение» относительно заданного распределения амплитуды $B_0(\xi,\eta)$.

Заметим, что оригинальная замена «зеркального отражения» в уравнении (2.20) имеет следующий вид [10, 11]:

$$\bar{I}_n(\xi,\eta) = \begin{cases} \left|2I_0(\xi,\eta) - I_n(\xi,\eta)\right|, & (\xi,\eta) \in \Omega, \\ I_n(\xi,\eta), & (\xi,\eta) \notin \Omega, \end{cases} \quad (2.21)$$

где $I_n(\xi,\eta)$ — распределение интенсивности, полученное на n-м шаге, $I_n = |F_n(\xi,\eta)|^2$, I_0 — заданное распределение интенсивности, $I_0 = B_0^2$, Ω — область задания изображения. Заметим, что если $I_n \geqslant 2I_0$, то $\bar{I}_n(\xi,\eta) = 0$.

Покажем, что замена (2.17) в АА-алгоритме минимизирует функционал невязки ε_1, который представляет собой среднеквадратичное отклонение рассчитанной амплитуды от заданной в плоскости ДОЭ:

$$\varepsilon_1 = \iint\limits_{-\infty}^{\infty} \left[A_0(u,v) - |W(u,v)|\right]^2 \mathrm{d}u\,\mathrm{d}v, \quad (2.22)$$

где $A_0(u,v)$ — амплитуда волны света, падающего на ДОЭ. Вариация функционала ε_1 относительно функции W на n-м шаге определяется соотношением

$$\delta\varepsilon_1 = 2\,\mathrm{Re}\left\{\iint\limits_{-\infty}^{\infty} \left[W_n - A_0 W_n \left|W_n\right|^{-1}\right] \delta W_n^*\,\mathrm{d}u\,\mathrm{d}v\right\}, \quad (2.23)$$

где $\mathrm{Re}\{\}$ означает действительную часть числа, * — знак комплексного сопряжения. При получении уравнения (2.23) мы использовали соотношения

$$\delta\left\{\iint\limits_{-\infty}^{\infty} A_0^2(u,v)\,\mathrm{d}u\,\mathrm{d}v\right\} = 0, \quad (2.24)$$

$$\delta\left\{\iint\limits_{-\infty}^{\infty} \left|W(u,v)\right|^2\,\mathrm{d}u\,\mathrm{d}v\right\} \neq 0, \quad (2.25)$$

$$\delta(WW^*) = W\delta W^* + W^*\delta W = 2\,\mathrm{Re}\{W\delta W^*\}. \quad (2.26)$$

Используя равенства Парсеваля

$$\iint\limits_{-\infty}^{\infty} |W(u,v)|^2 \,\mathrm{d}u\,\mathrm{d}v = \int\limits_{-\infty}^{\infty}\int \left|\bar{F}(\xi,\eta)\right|^2 \mathrm{d}\xi\,\mathrm{d}\eta, \tag{2.27}$$

$$\iint\limits_{-\infty}^{\infty} \left|\bar{W}(u,v)\right|^2 \mathrm{d}u\,\mathrm{d}v = \int\limits_{-\infty}^{\infty}\int |F(\xi,\eta)|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta, \tag{2.28}$$

где $\bar{W} = D_A W = AW|W|^{-1}$ — амплитуда волны в области ДОЭ после осуществления замены (2.6), можно показать, что выполняется следующее соотношение:

$$\mathrm{Re}\left\{\iint\limits_{-\infty}^{\infty}\left[W_n - A_0 W_n\,|W_n|^{-1}\right]\delta W_n^*\,\mathrm{d}u\,\mathrm{d}v\right\} = \\ = \mathrm{Re}\left\{\iint\limits_{-\infty}^{\infty}\left[\bar{F}_n - F_n\right]\delta\bar{F}_n^*\,\mathrm{d}\xi\,\mathrm{d}\eta\right\}, \tag{2.29}$$

где $\bar{F}_n = \mathcal{F}\{W_n\}$, $F_n = \mathcal{F}D_A\mathcal{F}^{-1}\left\{\bar{F}_n\right\} = \mathcal{F}\left\{A_0 W_n\,|W_n|^{-1}\right\}$.

Приращение $\delta\varepsilon_1$ функционала ошибки ε_1 принимает максимальное отрицательное значение на n-м шаге при следующем параметрически выраженном условии:

$$\delta\bar{F}_n = \bar{F}_{n+1} - \bar{F}_n = \gamma\left(F_n - \bar{F}_n\right), \quad \gamma > 0. \tag{2.30}$$

С учетом уравнения (2.30), вместо (2.23) мы получим

$$\delta\varepsilon_1 = -2\gamma\iint\limits_{-\infty}^{\infty}\left|\bar{F}_n - F_n\right|^2 \mathrm{d}\xi\,\mathrm{d}\eta < 0. \tag{2.31}$$

Из (2.30) следует параметрическое уравнение для расчета функции $F(\xi,\eta)$:

$$\bar{F}_{n+1} = (1-\gamma)\,\bar{F}_n + \gamma F_n = (1-\gamma)\,\bar{F}_n + \gamma\mathcal{F}D_A\mathcal{F}^{-1}\left\{\bar{F}_n\right\}. \tag{2.32}$$

Уравнение (2.32) предполагает, что на $(n+1)$-м шаге входная амплитуда представляет собой линейную комбинацию входной и выходной амплитуды на n-м шаге. Алгоритм (2.32) аналогичен алгоритму входа-выхода [5], и предыдущие рассуждения следует рассматривать как теоретическое обоснование его оптимальности. Для алгоритма (2.32) среднеквадратичная ошибка амплитуды не возрастает с ростом числа итераций при условии, что параметр γ удовлетворяет неравенству $0 < \gamma \leqslant 2$. Действительно, введя обозначения

$$F_n = B_n e^{i\varphi_n}, \quad \bar{F}_n = \bar{B}_n e^{i\bar{\varphi}_n}, \quad W_n = A_n e^{i\psi_n}, \quad \bar{W}_n = A_0 e^{i\psi_n}, \tag{2.33}$$

мы получим последовательность, в которой равенства 1, 3, 5 являются очевидными, равенства 2, 4 есть равенства Парсеваля, а неравенство представляет собой

неравенство треугольника ($|a| - |b| \leqslant |a - b|$):

$$\iint\limits_{-\infty}^{\infty} |A_0 - A_{n+1}|^2 \, du\, dv = \iint\limits_{-\infty}^{\infty} \left|A_0 e^{i\psi_{n+1}} - A_{n+1} e^{i\psi_{n+1}}\right|^2 du\, dv \leqslant$$

$$\leqslant \iint\limits_{-\infty}^{\infty} \left|A_0 e^{i\psi_n} - A_{n+1} e^{i\psi_{n+1}}\right|^2 du\, dv =$$

$$= \iint\limits_{-\infty}^{\infty} \left|B_n e^{i\varphi_n} - \left[(1-\gamma)\,\bar{B}_n e^{i\bar{\varphi}_n} + \gamma B_n e^{i\varphi_n}\right]\right|^2 d\xi\, d\eta =$$

$$= (1-\gamma)^2 \iint\limits_{-\infty}^{\infty} \left|B_n e^{i\varphi_n} - \bar{B}_n e^{i\bar{\varphi}_n}\right|^2 d\xi\, d\eta =$$

$$= (1-\gamma)^2 \iint\limits_{-\infty}^{\infty} \left|A_0 e^{i\psi_n} - A_n e^{i\psi_n}\right|^2 du\, dv = (1-\gamma)^2 \iint\limits_{-\infty}^{\infty} |A_0 - A_n|^2 \, du\, dv. \quad (2.34)$$

Из (2.34) следует, что при $0 < \gamma \leqslant 2$ имеет место релаксация среднеквадратичной ошибки:

$$\iint\limits_{-\infty}^{\infty} |A_0 - A_{n+1}|^2 \, du\, dv \leqslant \iint\limits_{-\infty}^{\infty} |A_0 - A_n|^2 \, du\, dv. \quad (2.35)$$

Отметим, что в алгоритме (2.32) учитывается, что комплексная амплитуда удовлетворяет ограничению в плоскости ДОЭ, но не учитывается, что модули входной и выходной функций комплексной амплитуды в фурье-плоскости, а именно $\bar{F}(\xi, \eta)$ и $F(\xi, \eta)$ должны стремиться к заданной функции $B_0(\xi, \eta)$. Чтобы использовать это условие в алгоритме (2.32), предположим, что входная функция на n-м шаге итерации подчиняется ограничению

$$\bar{F}_n(\xi, \eta) = D_B F(\xi, \eta) = B_0(\xi, \eta)\, F(\xi, \eta)\, |F(\xi, \eta)|^{-1} \quad (2.36)$$

и что на $(n+1)$-м шаге она подчиняется уравнению (2.32). Поэтому комбинация уравнений (2.32) и (2.36) приводит к адаптивно-аддитивному алгоритму:

$$\bar{F}_{n+1}(\xi, \eta) = (1-\gamma)\, B_0(\xi, \eta)\, \frac{F_n(\xi, \eta)}{|F_n(\xi, \eta)|} + \gamma F_n(\xi, \eta). \quad (2.37)$$

Заметим, что уравнение (2.37) отличается от (2.17) перестановкой весовых коэффициентов: $\gamma \to 1 - \gamma$.

Следует отметить также, что похожий прием был применен в работе [9], чтобы получить алгоритм (2.37) из алгоритма (2.32).

В работе [13] алгоритм (2.17) был получен как весовой алгоритм. Рассмотрим функционал невязки ε_0, но с весовой функцией:

$$\bar{\varepsilon}_0 = \iint\limits_{-\infty}^{\infty} S(\xi, \eta)\, [|F(\xi, \eta)| - B_0(\xi, \eta)]^2 \, d\xi\, d\eta, \quad (2.38)$$

где

$$S(\xi,\eta)=\begin{cases}\gamma, & (\xi,\eta)\in\Omega,\\ 0, & (\xi,\eta)\notin\Omega,\end{cases} \tag{2.39}$$

— реальная, положительно определенная весовая функция, которая задает, например, область определения синтезированного изображения, γ — постоянная.

Вариация функционала (2.38) имеет вид

$$\delta\bar{\varepsilon}_0 = 2\,\mathrm{Re}\iint\limits_{-\infty}^{\infty} S(\xi,\eta)\left[F-B_0\frac{F}{|F|}\right]\delta F^*\,\mathrm{d}\xi\,\mathrm{d}\eta. \tag{2.40}$$

Итеративное уравнение, которое позволяет получить функцию $F(\xi,\eta)$, минимизирующую функционал (2.38), следует из требования равенства нулю подынтегрального выражения в уравнении (2.40) и имеет вид

$$\bar{F}_{n+1}=F_n+S\left[B_0\frac{F_n}{|F_n|}-F_n\right]=(1-S)F_n+SB_0\frac{F_n}{|F_n|}, \tag{2.41}$$

где F_n и $\bar{F}_{n+1}$ — функции выхода на n-м шаге и входа на $(n+1)$-м шаге соответственно. При получении уравнения (2.41) использовалось нетрадиционное понимание вариации функции. Традиционное понимание вариации — это изменение амплитуды на входе и выходе от итерации к итерации: $\delta F_n = F_{n+1} - F_n$. Мы понимаем вариацию функции, как переход с выхода на вход: $\delta F_n = \bar{F}_{n+1} - F_n$. Преимущества АА-алгоритма можно проиллюстрировать на следующих примерах [14].

Пример 2.1. Рассчитывается ДОЭ, который освещается лазерным пучком с гауссовой интенсивностью

$$A_0(u,v)=\exp\left[-\frac{u^2+v^2}{r^2}\right], \tag{2.42}$$

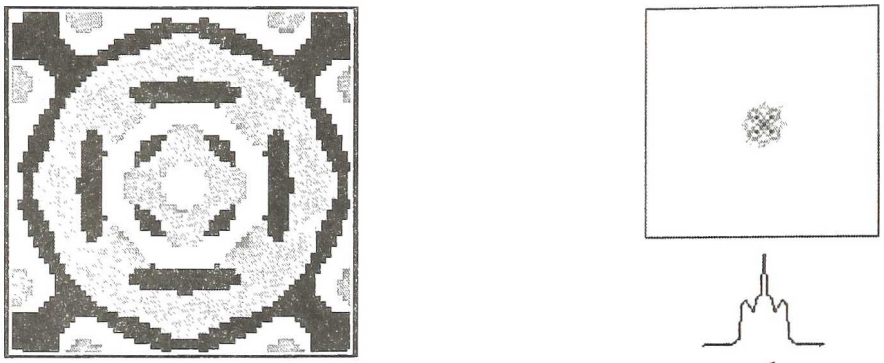
и формирует изображение «мягкого квадрата» в дальней зоне дифракции. Это означает, что заданное распределение интенсивности в плоскости наблюдения выбирается в форме супергауссовой функции:

$$B_0(\xi,\eta)=\exp\left[-\frac{\xi^{2n}+\eta^{2n}}{a^{2n}}\right],\quad n=1,2,\ldots, \tag{2.43}$$

где $2a$ — эффективная сторона квадрата, r — радиус гауссова пучка. Такая задача выравнивания интенсивности гауссова пучка актуальна для многих приложений: лазерной печати, лазерной термообработки поверхности и т.д. Подробнее эта задача рассматривается в п. 2.8.7.

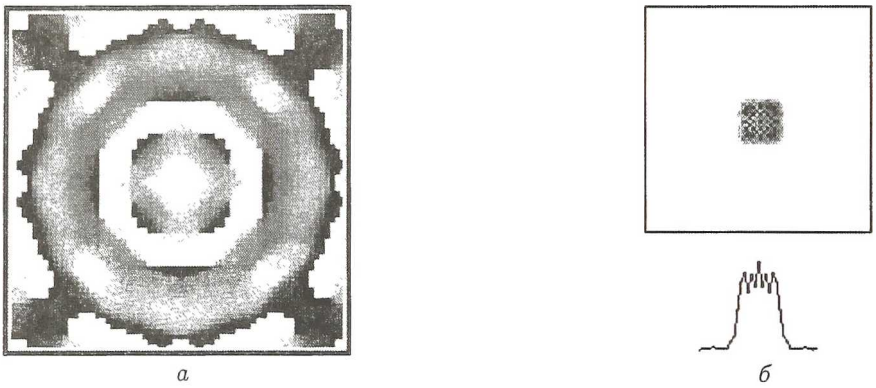
Преобразование Френеля (2.1) и обратное ему (2.5) рассчитывались при помощи алгоритма быстрого преобразования Фурье. Массив состоял из 256×256 пикселов. Интенсивность освещающего гауссова пучка вдоль границ квадратной апертуры ДОЭ составляла 10% от максимального значения в центре. Сторона изображения квадрата была равна примерно 10 диаметрам диска Эйри (минимальное дифракционное пятно). Число итераций равнялось 10.

На рис. 2.3 показаны фаза ДОЭ по модулю 2π (полутоновая, 16 уровней), рассчитанная при помощи ГС-алгоритма,изображение в дальней зоне дифракции и горизонтальное сечение распределения интенсивности изображения.



а *б*

Рис. 2.3. Расчет ДОЭ по ГС-алгоритму: фаза ДОЭ от 0 до 2π (*а*), сформированное изображение и центральное горизонтальное сечение интенсивности изображения (*б*)



а *б*

Рис. 2.4. Расчет ДОЭ по АА-алгоритму: фаза ДОЭ от 0 до 2π (*а*), сформированное изображение и центральное горизонтальное сечение интенсивности изображения (*б*)

На рис. 2.4 представлены аналогичные результаты, полученные при помощи АА-алгоритма ($\lambda = 2$).

Среднеквадратичные ошибки, рассчитанные по формуле (2.7) для приведенных случаев, были равны: 27,2% (рис.2.3) и 5,1% (рис. 2.4).

Введем понятие энергетической эффективности формирования заданного изображения. Это часть световой энергии, попадающей в заданную область определения изображения:

$$E = \frac{\iint\limits_{\Omega} |F(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y}{\iint\limits_{-\infty}^{\infty} |F(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y}, \tag{2.44}$$

где Ω — область определения изображения. В рассмотренном случае энергетическая эффективность была равна: 96,4% (рис. 2.3) и 91,5% (рис. 2.4).

Из рис. 2.3 и 2.4 видно, что фазы, полученные этими двумя методами, слабо отличаются друг от друга. Однако незначительная разница между ними приводит к существенному различию в формируемых изображениях.

Моделирование показало, что скорость сходимости итеративного алгоритма увеличивается с ростом параметра релаксации γ от 0 до 2.

Пример 2.2. Рассчитывается оптический элемент, способный эффективно определять положение источника излучения [15]. Этот элемент является частью прибора для определения координат центра поперечного сечения изображения точечного источника. Схема устройства представлена на рис. 2.5. Точечный источник света S освещает сферическую линзу, за которой помещен ДОЭ (рис. 2.5). ДОЭ рассчитывается таким образом, что изображение точечного источника имеет распределение интенсивности в форме креста. Каждая сторона изображения креста пересекает линейный фотодатчик, как показано на рис. 2.6. В качестве линейного фотодатчика, с характеристической стороной входного окна, равной 10 мм, можно использовать кремниевую полупроводниковую структуру, у которой выходной аналоговый сигнал в форме падения напряжения пропорционален расстоянию между точкой максимальной интенсивности света и исходной точкой датчика. Технические параметры такого датчика следующие: изменение напряжения на 0,1 мВ соответствует смещению максимальной интенсивности света вдоль оси датчика на 0,2 мкм.

Для расчета ДОЭ при указанных условиях использовалось параксиальное соотношение, связывающее распределение интенсивности в плоскостях источника и изображения креста[16]:

$$F(\xi,\eta) = \frac{k^2}{ab}\exp\left[i\frac{k}{2b}\left(\xi^2+\eta^2\right)\right]\times \\ \times\int\limits_{-\infty}^{\infty}\int W_o(u,v)\exp\left[i\frac{k}{2a}(u^2+v^2)\right]H\left(u+\frac{b}{a}\xi, v+\frac{b}{a}\eta\right)\mathrm{d}u\,\mathrm{d}v, \quad (2.45)$$

где a и b — это расстояния от источника до линзы и от линзы до изображения (рис. 2.5), связанные соотношением для тонкой линзы,

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{f}, \quad (2.46)$$

$H(u,v)$ — функция импульсного отклика, задаваемая в виде

$$H(u,v) = \iint\limits_{Q} P(\xi,\eta)\,T(\xi,\eta)\exp\left[-i\frac{k}{b}(u\xi+v\eta)\right]\mathrm{d}\xi\,\mathrm{d}\eta. \quad (2.47)$$

Функция $P(\xi,\eta)$ в (2.47) называется функцией зрачка линзы, Q определяет форму апертуры диафрагмы, ограничивающей линзу. Функция $T(\xi,\eta)$ — это комплексная функция пропускания ДОЭ, являющаяся чисто фазовой функцией:

$$T(\xi,\eta) = e^{i\varphi(\xi,\eta)}. \quad (2.48)$$

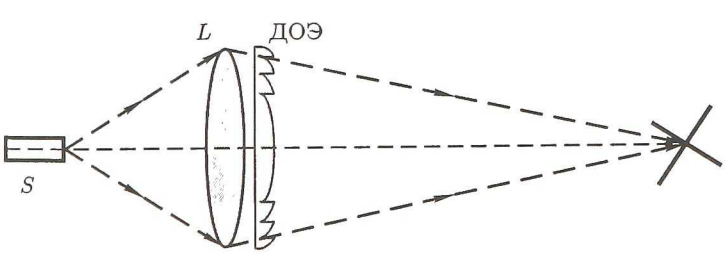


Рис. 2.5. Оптическая схема формирования крестообразного изображения

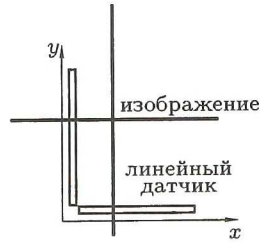


Рис. 2.6. Плоскость изображения прибора

Фаза $\varphi(\xi, \eta)$ оптического элемента находится из решения интегрального уравнения

$$I(u, v) = |H(u, v)|^2, \tag{2.49}$$

где $I(u, v)$ — заданное распределение интенсивности для функции импульсного отклика или заданная функция рассеяния. Для креста эта функция принимает вид

$$I(u, v) = \begin{cases} I_0; & -l \leqslant u \leqslant l, \quad v = 0 \text{ и } u = 0, \quad -l \leqslant v \leqslant l, \\ 0; & \text{иначе}, \end{cases} \tag{2.50}$$

где I_0 — постоянное значение, $2l$ — длина вертикального и горизонтального участков креста. Уравнение (2.49) решалось при помощи AA-алгоритма (2.17) при $\gamma = 2$.

Параметры расчета: ДОЭ был ограничен круглой диафрагмой радиусом 64 пиксела, для выполнения преобразования Фурье был выбран массив размерностью 256× ×256 пикселов, каждая сторона креста состояла из 170 пикселов, ширина была равна одному пикселу. На рис. 2.7*а* показана фаза оптического элемента, рассчитанная за 25 итераций с 5 градациями. Квадрат модуля функции импульсного отклика для комбинации «линза плюс ДОЭ» показан на рис. 2.7*б*. Горизонтальный и вертикальный профили для полученных картин фазы и функции импульсного отклика также показаны на рис. 2.7.

2.5. Адаптивно-мультипликативный алгоритм

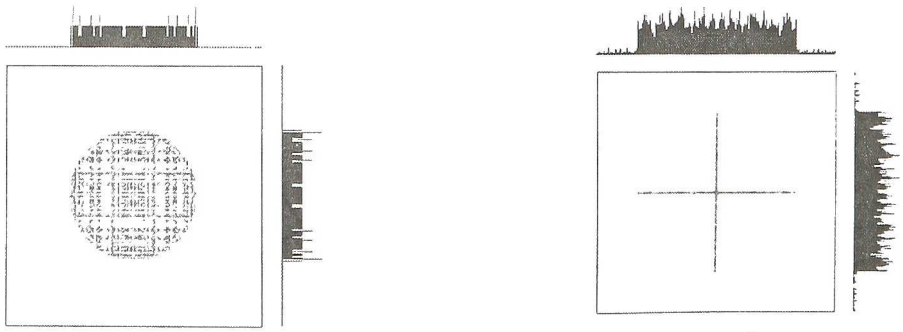
В этом разделе рассматривается еще один итеративный алгоритм для расчета ДОЭ, который также имеет скорость сходимости большую, чем в алгоритме ГС. Этот алгоритм может быть получен при помощи тождества, которое должно выполнятся в плоскости формирования изображения:

$$I_0(\xi, \eta) = F(\xi, \eta) F^*(\xi, \eta). \tag{2.51}$$

То есть заданную интенсивность $I_0(\xi, \eta)$ должно формировать световое поле с комплексной амплитудой $F(\xi, \eta)$.

Используя обратное преобразование Френеля $\hat{L}^{-1}$, получаем в плоскости ДОЭ

$$W(u, v) = \hat{L}^{-1}\left[\frac{I_0(\xi, \eta)}{F^*(\xi, \eta)}\right] = \hat{L}^{-1}\left[\frac{I_0 F}{|F|^2}\right]. \tag{2.52}$$



а *б*

Рис. 2.7. Фаза ДОЭ, 5 градаций (*а*); крест в области изображения (*б*)

Уравнение (2.52) позволяет предложить итеративную процедуру поиска фазы ДОЭ (аргумент функции $W(u,v)$), которая очень похожа на процедуру, используемую в алгоритме ГС, но вместо уравнения (2.4) использует следующую замену:

$$\bar{F}(\xi,\eta) = I_0(\xi,\eta)\, F(\xi,\eta)\, |F(\xi,\eta)|^{-2}. \tag{2.53}$$

Уравнение (2.53) являетя некорректным, т.к. возможно равенство нулю модуля $F(\xi,\eta)$ в знаменателе. Корректная формулировка задачи нахождения фазы ДОЭ должна основываться на рассмотрении некоторого функционала невязки с регуляризацией [17]. С этой целью рассмотрим функционал

$$\varepsilon_2 = \iint\limits_{-\infty}^{\infty} \left[|F|^2 - I_0\right]^2 \mathrm{d}\xi\, \mathrm{d}\eta + \alpha \iint\limits_{-\infty}^{\infty} Q\, |F|^2 \, \mathrm{d}\xi\, \mathrm{d}\eta. \tag{2.54}$$

Первый член в уравнении (2.54) — это среднеквадратичное отклонение формируемой интенсивности в плоскости наблюдения от заданной интенсивности $I_0(\xi,\eta)$.

В отличие от функционала (2.9), это слагаемое представляет собой функционал четвертого порядка относительно амплитуды светового поля. Второе слагаемое в уравнении (2.54) — стабилизирующее.

Отметим, что действительная, положительно определенная функция Q в уравнении (2.54) в методе регуляризации А.Н. Тихонова [18] задается в виде

$$Q(\xi,\eta) = \sum_{n=0}^{N} |C_n| \left(\xi^{2n} + \eta^{2n}\right), \tag{2.55}$$

где C_n — произвольные коэффициенты. Стабилизирующая константа $\alpha \geqslant 0$ определяет нижнюю границу изменения функционала ε_2. Минимизация ε_2 означает выбор самой гладкой функции из набора функций F наиболее близких по модулю к функции $\sqrt{I_0(\xi,\eta)}$. Итеративная процедура минимизации функционала (2.54) строится при помощи промежуточных функционалов, квадратичных относительно заданной функции $G_n(\xi,\eta)$:

$$\varepsilon_{2n} = \iint\limits_{-\infty}^{\infty} |F_n G_n^* - I_0|^2 \, \mathrm{d}\xi\, \mathrm{d}\eta + \alpha \iint\limits_{-\infty}^{\infty} Q\, |G_n|^2 \, \mathrm{d}\xi\, \mathrm{d}\eta, \tag{2.56}$$

где $F_n(\xi,\eta)$ — комплексная амплитуда, рассчитанная на n-м шаге итераций в плоскости наблюдения, $G_n(\xi,\eta)$ — функция, минимизирующая функционал ε_{2n}. Вариация функционала (2.56) относительно функции G_n равна

$$\delta\varepsilon_{2n} = 2\,\mathrm{Re}\left\{ \iint\limits_{-\infty}^{\infty} \left[|F_n|^2 G_n - I_0 F_n + \alpha Q G_n\right] \delta G_n^* \, \mathrm{d}\xi\, \mathrm{d}\eta \right\}. \tag{2.57}$$

Функционал (2.56) достигает минимума при условии, что вариация (2.57) становится равной 0. Как видно из (2.57), это становится возможным при условии, что

$$G_n = \frac{I_0 F_n}{|F_n|^2 + \alpha Q}. \tag{2.58}$$

Из сравнения уравнений (2.58) и (2.53) видно, что функция $G_n(\xi, \eta)$ является регулярным аналогом функции $\bar{F}(\xi, h)$. Замена (2.58) минимизирует промежуточные функционалы ε_{2n} на каждой итерации. Однако из-за сходимости итеративной процедуры разность $|F_n - G_n|$ уменьшается с ростом n, а функционалы ε_{2n} стремятся к функционалу ε_2. Следовательно, можно заключить, что в пределе замена (2.58) минимизирует исходный функционал (2.54). Приведенные рассуждения подтверждают возможность итеративного поиска фазы ДОЭ методом, сходным с алгоритмом ГС, но с использованием замены (2.58) вместо замены (2.4).

Данный алгоритм получил название адаптивно-мультипликативного (АМ) алгоритма. Его преимущества перед алгоритмом ГС видны из следующего примера.

Пример 2.3. Рассчитывается ДОЭ [17], предназначенный формировать изображение «мягкого круга» в дальней зоне дифракции. Элемент освещается лазерным пучком с гауссовым распределением интенсивности. Заданная интенсивность в плоскости наблюдения имеет вид

$$I_0(\xi, \eta) = \exp\left[-2\left\{\frac{\xi^2 + \eta^2}{a^2}\right\}^n\right], \tag{2.59}$$

где a — эффективный радиус «мягкого круга», $n = 1, 2, 3, \ldots$

Радиус a выбирался равным примерно двум радиусам диска Эйри. Массив пикселов был 256×256.

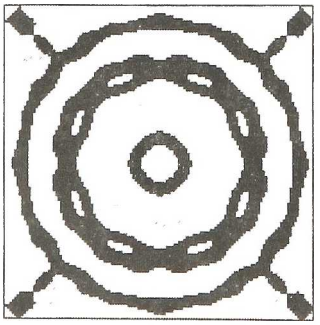
На рис. 2.8 показаны: фаза ДОЭ по модулю 2π (16 градаций), рассчитанная при помощи алгоритма ГС за 20 итераций, результирующая дифракционная картина и распределение интенсивности света в центральном сечении этой картины.

На рис. 2.9 показаны аналогичные результаты, полученные при помощи АМ-алгоритма.

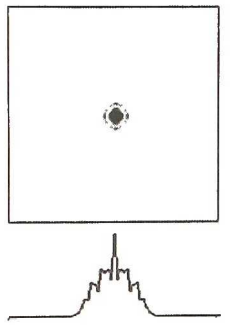
Для данного случая энергетические эффективности, рассчитанные по формуле (2.44), были равны: 97,1% (рис. 2.8), 91,8% (рис. 2.9). Среднеквадратичные ошибки, рассчитанные по формуле (2.7), были равны: 22,9% (рис. 2.9) и 3,6% (2.9).

Из сравнения рисунков 2.8 и 2.9 видно, что фазы отличаются друг от друга незначительно. Тем не менее эта незначительная разница приводит к формированию сильно различающихся дифракционных картин.

Отсутствие круговой симметрии в фазовых картинах, представленных на рис. 2.8*а* и 2.9*а* можно объяснить нарушением круговой симметрии функции $I_0(\xi, \eta)$ в уравнении (2.59) для дискретных значений переменных ξ и η, при котором окружности переходят в многоугольники.

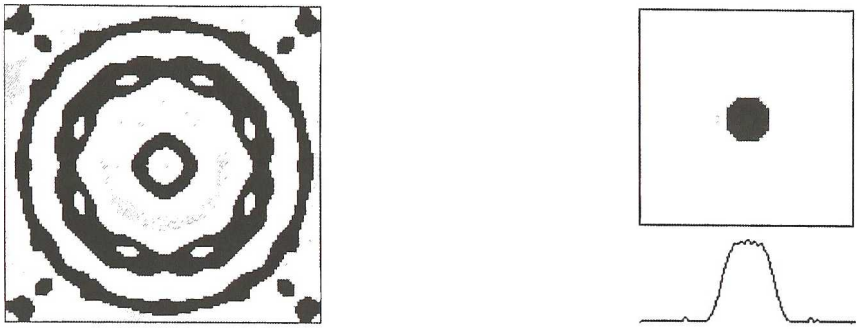


а



б

Рис. 2.8. Расчет при помощи алгоритма ГС: фаза ДОЭ от 0 до 2π (*а*), дифракционная картина и распределение интенсивности в центральном сечении (*б*)



а *б*

Рис. 2.9. Расчет при помощи АМ-алгоритма: фаза ДОЭ от 0 до 2π (*а*), дифракционная картина и распределение интенсивности в центральном сечении (*б*)

На рис 2.10 показана зависимость значения функционала ε_2 в (2.54) от числа итераций: кривая 1 — для ГС-алгоритма ($\alpha = 0$), кривая 2 — для АМ-алгоритма ($\alpha = 10^{-8}$). Из рис. 2.10 видно, что адаптивный алгоритм обладает лучшей сходимостью: кривая 2 достигает значения 1 после 4 итераций, в то время как кривая 1 достигает значения 1,2 после 20 итераций. Рис. 2.10 также выявляет недостаток АМ-алгоритма: его немонотонную сходимость. Но если брать значения функционала (2.54) через итерацию, то они будут убывать монотонно. Для всех рассмотренных примеров начальная оценка фазы выбиралась случайной, и результаты практически не зависели от ее конкретной реализации.

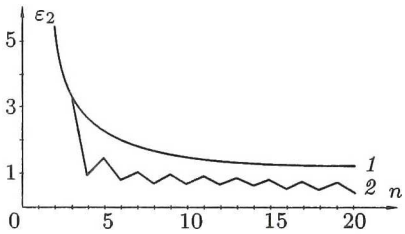


Рис. 2.10. Зависимость функционала ε_2 от числа итераций

Также оказывается трудным обосновать теоретически оптимальный выбор стабилизирующей константы α и степени полинома N для функции $Q(\xi, \eta)$, которые влияют на минимизацию функционала (2.54). Однако для ряда практических применений оказывается достаточным выбрать функцию $Q(\xi, \eta)$ как одночлен $(\xi^2 + \eta^2)^n$ при $n = 1, 2$ и найти константу α методом подбора.

Из (2.54) при $\alpha = 0$ можно получить другую версию мультипликативного алгоритма. Рассмотрим функционал ошибки, заданный выражением

$$\varepsilon_{2p} = \iint\limits_{-\infty}^{\infty} \left[B_0^2 (\xi, \eta) - |F(\xi, \eta)|^2 \right]^p \mathrm{d}\xi \, \mathrm{d}\eta, \tag{2.60}$$

где p — целое число. Вариация этого функционала относительно функции $F(\xi, \eta)$ имеет вид

$$\delta\varepsilon_{2p} = 2 \operatorname{Re} \iint p \left[B_0^2 (\xi, \eta) - |F(\xi, \eta)|^2 \right]^{p-1} F(\xi, \eta) \, \delta F^* (\xi, \eta) \, \mathrm{d}\xi \, \mathrm{d}\eta, \tag{2.61}$$

где $\delta F = \bar{F} - F$ — разность между входной $\bar{F}$ и выходной F комплексными амплитудами в изображений Фурье. Из условия равенства нулю подынтегрального выражения в (2.61) следует итеративное уравнение для нахождения входной функции $\bar{F}$, минимизирующей функционал (2.60)

$$\bar{F}_{n+1} - F_n = \gamma p \left[B_0^2 - |F_n|^2 \right]^{p-1} F_n \tag{2.62}$$

или

$$\bar{F}_{n+1} = \left\{1 + \gamma p \left[B_0^2 - |F_n|^2\right]^{p-1}\right\} F_n, \tag{2.63}$$

где γ — постоянная релаксации. Уравнение (2.63) нелинейно связывает функцию выходной амплитуды F_n на n-м шаге итераций с входной функцией $\bar{F}_{n+1}$. Заметим, что сама функция F_n рассчитывается при помощи уравнения

$$F_n = \mathcal{F} D_A \mathcal{F}^{-1}\left\{\bar{F}_n\right\}, \tag{2.64}$$

где ограничивающий оператор D_A в плоскости ДОЭ определяется уравнением (2.14).

В заключение укажем, что совместное применение АА-алгоритма, уравнения (2.37), АМ-алгоритма, и уравнения (2.58), порождает новый двухпараметрический итеративный алгоритм, который можно записать в виде

$$\bar{F}_{n+1}(\xi,\eta) = \frac{(1-\gamma) I_0 F_n(\xi,\eta)}{|F_n(\xi,\eta)|^2 + \alpha Q(\xi,\eta)} + \gamma F_n(\xi,\eta). \tag{2.65}$$

2.6. Адаптивно-регуляризационный алгоритм

Так как ДОЭ формируют, как правило, изображения, ограниченные в пространстве, то появляется свобода в выборе поведения светового поля за областью определения изображения. Эта дополнительная степень свободы используется как в ранних алгоритмах Герчберга Р. [19] и Папулиса А. [20], так и в более поздних работах Вировского Ф. [21] и Фиенапа Дж. [13], и позволяет частично устранять эффект стагнации, присущий алгоритму Герчберга–Секстона [3].

Так как рассчитываемый ДОЭ должен формировать на некотором расстоянии заданное изображение в некоторой определенной области, то световое поле за пределами этой области может быть рассмотрено как некоторый шум. Энергетический вклад светового поля за пределами выделенной области может регулироваться согласно методу регуляризации А.Н. Тихонова [18] стабилизирующим слагаемым. Сформулируем задачу расчета ДОЭ, фокусирующего падающее на него лазерное излучение в малую область фурье-плоскости, как вариационную задачу минимизации функционала-критерия (для простоты рассмотрим одномерный случай):

$$\varepsilon(\alpha) = \int_{-\infty}^{\infty} \left[|F_0(u)| - D_A |F(u)|\right]^2 \mathrm{d}u + \alpha \int_{-\infty}^{\infty} Q(u) |F(u)|^2 \mathrm{d}u, \tag{2.66}$$

где $Q(u) = \sum_{n=0}^{N} |c_n| u^{2n}$ — стабилизирующая функция, аналогичная функции (2.55),

$$D_A f = \begin{cases} f, & u \in \Omega_A, \\ 0, & u \notin \Omega_A, \end{cases}$$

— оператор пространственного ограничения, $|F_0(u)|$ — заданное распределение амплитуды в Ω_A, $F(u)$ — комплексная амплитуда волны в плоскости фурье-спектра, в которой формируется требуемое изображение $|F_0(u)|$, α — параметр регуляризации. Функционал (2.66) в отличие от функционала (2.54) квадратичный по амплитуде светового поля.

Вариация функционала (2.66) запишется в виде

$$\delta\varepsilon = 2\,\mathrm{Re}\left[\int\limits_{-\infty}^{\infty}\left(-|F_0(u)|\,\frac{F(u)}{|F(u)|} + D_A F(u) + \alpha Q(u)F(u)\right)\delta F^*(u)\,\mathrm{d}u\right]. \qquad (2.67)$$

Если выбрать вариацию δF как разность комплексных амплитуд до и после удовлетворения пространственным ограничениям, $\delta F = \bar{F} - F$, то максимальная отрицательная вариация (2.67) функционала (2.66) достигается при условии

$$\bar{F} - F = \gamma G(u) = \gamma\left(|F_0(u)|\,\frac{F(u)}{|F(u)|} - D_A F(u) - \alpha Q(u)F(u)\right), \qquad (2.68)$$

где $\gamma > 0$ — параметр адаптации алгоритма или константа релаксации. Действительно, при условии (2.68) вариация (2.67) будет иметь вид

$$\delta\varepsilon = -2\gamma\int\limits_{-\infty}^{\infty}|G(u)|^2\,\mathrm{d}u \leqslant 0. \qquad (2.69)$$

Из условия (2.68) следует двухпараметрическое итеративное уравнение расчета ДОЭ [22]:

$$\bar{F}_{n+1} = \left[1 - \gamma\left(D_A + \alpha Q(u)\right)\right]F_n + \gamma\,|F_0(u)|\,\frac{F_n(u)}{|F_n(u)|}. \qquad (2.70)$$

Заметим, что в уравнении (2.70) комплексная амплитуда на n-ом шаге итераций $F_n(u)$ является результатом удовлетворения ограничениям в плоскости ДОЭ и представима в виде

$$F_n(u) = \mathcal{F}D_B W_B \mathcal{F}^{-1}\bar{F}_n(u), \qquad (2.71)$$

где $\bar{F}_n(u)$ — комплексная амплитуда на n-ом шаге, после замены (2.70),

$$D_B f(x) = \begin{cases} f(x), & x \in \Omega_B, \\ 0, & x \notin \Omega_B, \end{cases}$$

— оператор пространственного ограничения, Ω_B — форма апертуры ДОЭ, $W_B f(x) = |f_0(x)|\,(f(x)/|f(x)|)$ — оператор присвоения заданной амплитуды $|f_0(x)|$ освещающего пучка комплексной функции $f(x)$, описывающей световое поле в плоскости ДОЭ.

Итак, уравнения (2.70) и (2.71) описывают итеративный двухпараметрический алгоритм расчета ДОЭ, формирующего заданное распределение интенсивности (изображение) в ограниченной области Ω_B пространственного спектра. Назовем его адаптивно-регуляризационным (АР) алгоритмом.

Покажем далее, что некоторые известные алгоритмы расчета ДОЭ являются частными случаями АР-алгоритма.

При следующем выборе параметров $\gamma = 1$, $\alpha = 0$, вместо уравнения (2.70) получим следующее выражение:

$$\bar{F}_{n+1} = [1 - D_A]\,F_n + |F_0(n)|\,\frac{F_n(n)}{|F_n(n)|} = \begin{cases} |F_0(n)|\,\dfrac{F_n(n)}{|F_n(n)|}, & u \in \Omega_A, \\ F_n(n), & u \notin \Omega_A. \end{cases} \qquad (2.72)$$

Алгоритм (2.72) сочетает в себе свойства алгоритма ГС и алгоритма Герчберга–Папулиса. Действительно, в соответствии с (2.72) в фурье-плоскости в заданной области Ω_A осуществляется замена рассчитанного модуля комплексной амплитуды $|F_n(u)|$ на заданный $|F_0(u)|$, а вне области Ω_A рассчитанная функция $F_n(u)$ остается без изменений. Поэтому уравнение (2.72) можно интерпретировать как алгоритм ГС, обладающий сверхразрешающим свойством алгоритма Герчберга–Папулиса. Уравнение (2.72) рассматривалось в [21] как реализация идеи дополнительной степени свободы для устранения стагнации итеративного алгоритма [3]. При выборе параметров $\gamma \neq 1$, $\alpha = 0$, вместо уравнения (2.70) получим выражение

$$\bar{F}_{n+1} = \begin{cases} \gamma |F_0(u)| \dfrac{F_n(u)}{|F_n(u)|} + (1-\gamma) F_n, & u \in \Omega_A, \\ F_n(u), & u \notin \Omega_A. \end{cases} \tag{2.73}$$

Уравнение (2.73) является аналогом алгоритма входа-выхода [19], метода ортогональных проекций [23], адаптивного [10, 11] и маятникового [19] алгоритмов, а также выражает идею «мягкого» кодирования комплексной амплитуды, высказанную в [24]. Релаксационный параметр γ в уравнении (2.73) позволяет регулировать скорость сходимости алгоритма, и при выборе $\gamma = 2$ эта скорость обычно бывает максимальной [9, 10]. При выборе параметров $\gamma = 1$, $\alpha \neq 0$, вместо уравнения (2.70) получим выражение

$$\bar{F}_{n+1} = [(1-\alpha) - D_A] F_n + |F_0(u)| \frac{F_n(u)}{|F_n(u)|} = \\ = \begin{cases} |F_0(u)| \dfrac{F_n(u)}{|F_n(u)|} - \alpha F_n, & u \in \Omega_A, \\ (1-\alpha) F_n(u), & u \notin \Omega_A. \end{cases} \tag{2.74}$$

Алгоритм (2.74) аналогичен регуляризованному алгоритму Герчберга–Папулиса, предложенному в [25] для задачи восстановления действительного объекта по зашумленному изображению. Параметр регуляризации α в (2.74) регулирует вклад рассчитанной амплитуды света $F_n(u)$ внутри и вне заданной области Ω_A.

В общем случае АР-алгоритм (2.70) можно переписать в виде

$$\bar{F}_{n+1} = \begin{cases} \gamma |F_0(u)| \dfrac{F_n(u)}{|F_n(u)|} + (1-\gamma-\gamma\alpha) F_n, & u \in \Omega_A, \\ (1-\gamma\alpha) F_n(u), & u \notin \Omega_A. \end{cases} \tag{2.75}$$

Интересно отметить, что ни при каких значениях параметров γ и α нельзя перейти от АР-алгоритма (2.75) к алгоритму сокращения ошибки [3], имеющему в данных обозначениях следующий вид:

$$\bar{F}_{n+1} = \begin{cases} |F_0(u)| \dfrac{F_n(u)}{|F_n(u)|}, & u \in \Omega_A, \\ 0, & u \notin \Omega_A. \end{cases} \tag{2.76}$$

Переход от уравнения (2.75) к уравнению (2.76) возможен только при условии рассмотрения релаксационной константы как двойной постоянной, принимающей разные значения внутри и вне заданной области:

$$\gamma = \begin{cases} (1+\alpha)^{-1}, & u \in \Omega_A, \\ \alpha^{-1}, & u \notin \Omega_A. \end{cases} \tag{2.77}$$

Пример 2.4. Преимущества алгоритма (2.70) или (2.75), по сравнению с базовым алгоритмом ГС (2.76), показаны для случая формирования в фокальной плоскости линзы распределения интенсивности вида $I(x) = I_0 \operatorname{rect}(x/a)$. Преобразование Фурье, связывающее комплексные амплитуды светового поля в плоскости ДОЭ и в фокальной плоскости линзы, вычислялось с помощью алгоритма быстрого преобразования Фурье на массиве отсчетов 256. Начальная фаза для итеративного процесса выбиралась в виде реализации псевдослучайной величины.

На рис. 2.11 показаны фаза ДОЭ, рассчитанная за 30 итераций алгоритмом (2.76), и нормированное распределение интенсивности в фокальной плоскости линзы, сформированное ДОЭ с данной фазой. Оптический элемент занимал 64 из 256 пикселов. Пунктирная линия на рис. 2.11*б* показывает требуемое распределение интенсивности, заданное на отрезке от 16 до 48 пикселов. В этот отрезок попадает 97,5% световой энергии, освещающей ДОЭ, а среднеквадратичное отклонение получившегося распределения интенсивности (рис. 2.11*б*) от постоянного значения равно 27,2%. Если использовать фазу ДОЭ, показанную на рис. 2.11*а* в качестве начальной фазы для алгоритма (2.75), то за следующие 100 итераций получится результат, показанный на рис. 2.12. При этом выбирались параметры $\gamma = 1$, $\alpha = 0{,}1$. В отрезок отсчетов от 16 до 48 пикселов попадает 92,7% от всей световой энергии, а среднеквадратичное отклонение интенсивности от постоянного значения равно 2%.

Таким образом, двухпараметрический АР-алгоритм (2.75) позволяет существенно уменьшить ошибку отклонения сформированной дифракционной картины от заданной (более чем в 10 раз в нашем примере) при несущественном снижении энергетической эффективности (на 5% в нашем примере).

2.7. Градиентный алгоритм расчета фазы ДОЭ

В работе [1] можно найти обзор алгоритмов нелинейного программирования для задач восстановления изображений. Задача сводится к минимизации целевых функционалов с учетом ограничений, накладываемых на функции, входящие в задачу. Если результирующий функционал с учетом ограничений можно представить в виде суммы линейного и квадратичных функционалов, то решение задачи находится аналитически. В противном случае требуется создавать вычислительные алгоритмы. Среди них можно выделить следующие: метод прямой оптимизации, метод градиентного спуска, метод наискорейшего спуска, метод сопряженных градиентов. Последний из перечисленных методов имеет наилучшую сходимость. Еще более быструю сходимость демонстрирует метод модифицированных функций Лагранжа,

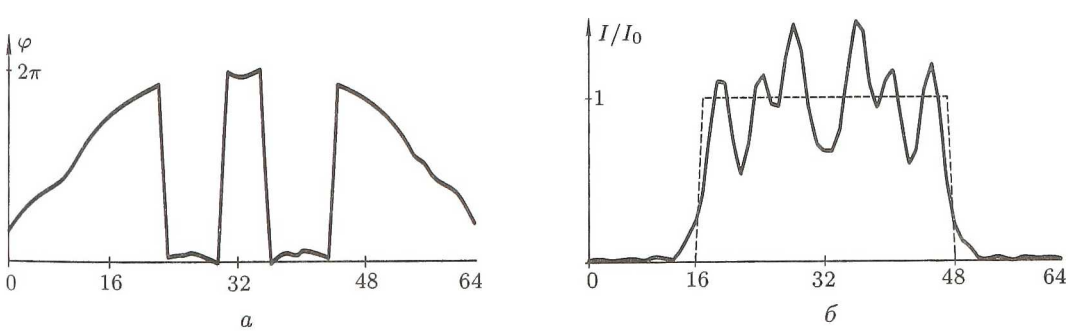
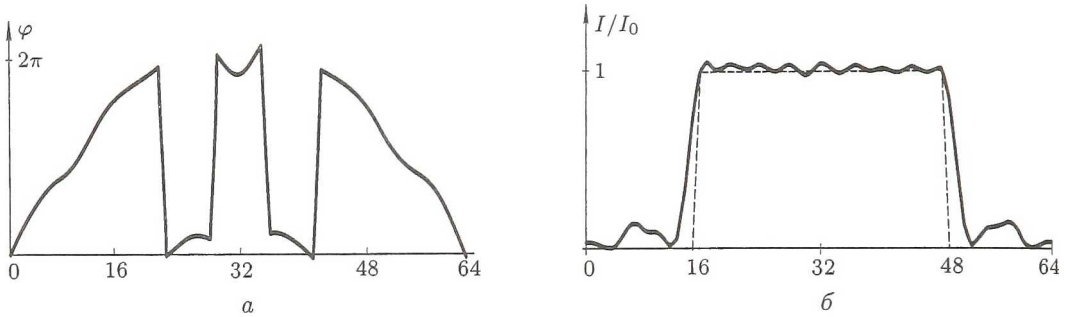


Рис. 2.11. Фаза ДОЭ, рассчитанная алгоритмом (2.76) (*а*), и распределение интенсивности в фурье-плоскости (*б*)



а *б*

Рис. 2.12. Фаза ДОЭ (*а*), рассчитанная по алгоритму (2.75), и интенсивность в фокальной плоскости (*б*)

согласно которому в функционал добавляются слагаемые, в которых функции ограничений взяты в квадрате. Однако, для выполнения одной итерации градиентных методов требуется выполнить 3–5 интегральных преобразований типа свертки или преобразования Фурье. Поэтому, чем сложнее алгоритм, тем больше требуется временных затрат при его компьютерной реализации. Применение градиентных алгоритмов в задачах расчета ДОЭ имеет ряд особенностей, некоторые из которых будут рассмотрены в этом разделе.

Итак, следуя работе [5] рассмотрим применение градиентного метода наискорейшего спуска для задачи расчета фазы ДОЭ.

Будем полагать, что переменные x и $u = 0, 1, 2, \ldots, N-1$ представляют собой массивы дискретных координат. Дискретизация переменных является необходимым шагом для проведения компьютерных расчетов. Тогда функционал (2.9) преобразуется к виду

$$\varepsilon = \sum_u \left(|G(u)| - |F(u)|\right)^2, \tag{2.78}$$

где $|F(u)|$ и $|G(u)|$ — заданная и рассчитанная амплитуды волны в плоскости пространственного спектра, а преобразование Фурье (2.10) примет вид дискретного преобразования Фурье (ДПФ):

$$G(u) = \mathcal{F}\{g(x)\} = \sum_{u=0}^{N-1} g(x) \exp\left[-i2\pi xu/N\right]. \tag{2.79}$$

Обратное ДПФ имеет вид

$$g(x) = \mathcal{F}^{-1}\{G(u)\} = \frac{1}{N} \sum_{u=0}^{N-1} G(u) \exp\left[i2\pi xu/N\right]. \tag{2.80}$$

Градиент функционала (2.78) равен сумме частных производных данного функционала от каждого неизвестного параметра p (это может быть фаза ДОЭ, амплитуда освещающего пучка и т.д.):

$$\frac{\partial \varepsilon}{\partial p} = 2 \sum_u \left[|G(u)| - |F(u)|\right] \frac{\partial |G(u)|}{\partial p} =$$

$$= \sum_u \left[G^*(u) - \frac{|F(u)|}{|G(u)|} G^*(u) \right] \frac{\partial G(u)}{\partial p} + \text{ к. с. } =$$

$$= -\sum_u G^{w^*}(u) \frac{\partial G(u)}{\partial p} + \text{ к. с. }, \quad (2.81)$$

где * — знак комплексного сопряжения, к.с. означает комплексно сопряженное предыдущее слагаемое, а функция $G^w(u)$ имеет вид

$$G^w(u) = |F(u)| \frac{G(u)}{|G(u)|} - G(u). \quad (2.82)$$

В силу линейности преобразование Фурье можно найти связь производных в плоскости ДОЭ и плоскости фурье-спектра:

$$\frac{\partial G(u)}{\partial p} = \frac{\partial}{\partial p} \left[\mathcal{F} \{ g(x) \} \right] = \mathcal{F} \left\{ \frac{\partial g(x)}{\partial p} \right\}. \quad (2.83)$$

Подставляя это выражение в (2.81) и производя некоторые преобразования, получаем

$$\frac{\partial \varepsilon}{\partial p} = -\sum_u G^{w*}(u) \mathcal{F} \left\{ \frac{\partial g(x)}{\partial p} \right\} + \text{ к. с. } =$$

$$= -2 \operatorname{Re} \left\{ \sum_x \left[\frac{\partial g(x)}{\partial p} \right]^* \mathcal{F}^{-1} \{ G^w(u) \} \right\} =$$

$$= -2 \operatorname{Re} \left\{ \sum_x \frac{\partial g(x)}{\partial p} \left(\mathcal{F}^{-1} \{ G^w(u) \} \right)^* \right\}. \quad (2.84)$$

Вычисление выражения (2.84) предусматривает: выполнение преобразования Фурье от $g(x)$, чтобы получить $G(u)$; расчет $G^w(u)$ по формуле (2.82); его комплексное сопряжение; выполнение обратного преобразования Фурье и скалярное умножение результата на $\partial g(x)/\partial p$.

Рассмотрим теперь конкретный случай, когда параметр p представляет собой отсчеты фазовой функции $\theta(x)$. Тогда имеем ($g(x) = |g(x)| \exp[i\theta(x)]$)

$$\frac{\partial g(x')}{\partial \theta(x)} = i g(x) \delta(x', x), \quad (2.85)$$

где

$$\delta(x', x) = \begin{cases} 1, & x = x', \\ 0, & x \neq x', \end{cases} \quad (2.86)$$

— функция Кронекера. Подставляя (2.85) в уравнение (2.81), получаем

$$\frac{\partial \varepsilon}{\partial \theta(x)} = -2 \operatorname{Re} \left\{ \sum_x i g(x') \delta(x', x) g^{w*}(x) \right\} = 2 \operatorname{Im} \{ g(x) g^{w*}(x) \}, \quad (2.87)$$

где

$$g^w(x) = \mathcal{F}^{-1} \{ G^w(u) \}. \quad (2.88)$$

Вычисленный градиент может быть использован в любом градиентном методе расчета фазовой функции $\theta(x)$. Например, в методе наискорейшего спуска. В этом методе, стартуя с какой-либо оценки фазы $\theta_0(x)$, каждое последующее приближение находится из итеративного уравнения

$$\theta_{n+1}(x) = \theta_n(x) + s\frac{\partial\varepsilon}{\partial\theta_n(x)} \tag{2.89}$$

(n — номер итерации) путем одномерной минимизации по s целевой функции ε, как функции от $\theta_{n+1}(x)$.

Чтобы избежать одномерной минимизации по s целевой функции ε, следуя [5], получим оценку оптимального шага s, исходя из разложения Тейлора первого порядка ε как функции от $\theta(x)$ около точки $\theta_n(x)$:

$$\varepsilon \approx \varepsilon_n + \sum_x \frac{\partial\varepsilon}{\partial\theta_n(x)}\,[\theta(x) - \theta_n(x)]. \tag{2.90}$$

Это выражение равно нулю при

$$\theta(x) - \theta_n(x) = -\varepsilon_n \frac{\partial\varepsilon}{\partial\theta_n(x)} \left\| \frac{\partial\varepsilon}{\partial\theta_n(x)} \right\|^{-2}. \tag{2.91}$$

Таким образом, при допущении (2.90) оптимальная величина шага градиентного метода записывается в виде

$$s_n = -\varepsilon_n \left\| \frac{\partial\varepsilon}{\partial\theta_n(x)} \right\|^{-2}. \tag{2.92}$$

Для увеличения скорости сходимости можно использовать и другие градиентные методы. Например, метод сопряженного градиента. Для его варианта Флетчера-Ривза каждое последующее приближение выглядит следующим образом [26]:

$$\theta_{n+1}(x) = \theta_n(x) + sd_n, \tag{2.93}$$

$$d_{n+1} = -\frac{\partial\varepsilon}{\partial\theta_{n+1}(x)} + \beta_n d_n, \tag{2.94}$$

$$\beta_n = \frac{\left\| \dfrac{\partial\varepsilon}{\partial\theta_{n+1}(x)} \right\|^2}{\left\| \dfrac{\partial\varepsilon}{\partial\theta_n(x)} \right\|^2}, \tag{2.95}$$

где

$$\left\| \frac{\partial\varepsilon}{\partial\theta_n(x)} \right\|^2 = \sum_{n=0}^{N-1} \left(\frac{\partial\varepsilon}{\partial\theta_n(x)} \right)^2 \tag{2.96}$$

— евклидова норма.

2.8. Применение итеративных алгоритмов для расчета ДОЭ

В этом разделе будут рассмотрены конкретные реализации итеративных алгоритмов, приведенных выше, для расчета различных типов ДОЭ.

2.8.1. Расчет ДОЭ, фокусирующих в радиально-симметричные области фурье-спектра

Чтобы рассчитать оптические элементы, формирующие в некоторой области пространства распределение интенсивности с круговой симметрией, оказывается полезным использовать полярные координаты (r, φ) и (ρ, θ) в плоскости ДОЭ и плоскости наблюдения соответственно:

$$\begin{aligned} u &= r\cos\varphi, \quad \xi = \rho\cos\theta, \\ v &= r\sin\varphi, \quad \eta = \rho\sin\theta. \end{aligned} \tag{2.97}$$

В этом случае интегральное преобразование Френеля (2.1) в полярных координатах будет иметь вид

$$F(\rho,\theta) = \frac{ik}{2\pi z} e^{ikz} \int_0^{\infty}\int_0^{2\pi} W(r,\varphi) \exp\left[\frac{ik}{2z}\left\{r^2+\rho^2-2r\rho\cos(\varphi-\theta)\right\}\right] r\,\mathrm{d}r\,\mathrm{d}\varphi. \tag{2.98}$$

Чтобы модуль комплексной амплитуды $F(\rho,\theta)$ в плоскости наблюдения не зависел от азимутального угла θ, ограничим наше рассмотрение набором функций, имеющих вид

$$W(r,\varphi) = g(r)\, e^{im\varphi}, \quad m = 1, 2, 3, \ldots . \tag{2.99}$$

Тогда вместо уравнения (2.98) получим

$$F(\rho,\theta) = i^{m+1} k z^{-1} e^{ikz} e^{im\theta} \int_0^{\infty} g(r) \exp\left[\frac{ik}{2z}\left(r^2+\rho^2\right)\right] J_m\left(\frac{kr\rho}{z}\right) r\,\mathrm{d}r, \tag{2.100}$$

где

$$J_m(x) = \frac{(-i)^m}{2\pi} \int_0^{2\pi} e^{i(mt + x\cos t)}\,\mathrm{d}t \tag{2.101}$$

— функция Бесселя первого рода m-го порядка.

Из уравнения (2.100) следует, что для нахождения фазы ДОЭ в рамках скалярной теории дифракции Френеля необходимо решить следующее интегральное уравнение:

$$I_0(\rho) = |F(\rho,\theta)|^2 = \left| kz^{-1} \int_0^{R} A_0(r) \exp i\varphi(r) \exp\frac{ikr^2}{2z} J_m\left(\frac{kr\rho}{z}\right) r\,\mathrm{d}r \right|^2, \tag{2.102}$$

где $I_0(\rho)$ — радиальное распределение интенсивности изображения, которое должно быть сформировано на расстоянии z от плоскости ДОЭ, $A_0(r)$ — радиальное распределение амплитуды освещающего пучка, $\varphi(r)$ искомая фазовая функция ДОЭ.

Ниже рассматриваются итеративные методы решения уравнения (2.102) для формирования заданного изображения в плоскости пространственного спектра.

Из уравнения (2.102) следует, что для итеративного поиска фазовой функции ДОЭ необходимо вычислить прямое и обратное преобразования Ханкеля:

$$H(\rho) = \frac{k}{z} \int_0^\infty h(r)\, J_m\left(\frac{k}{z} r\rho\right) r\, \mathrm{d}r, \tag{2.103}$$

$$h(r) = \frac{k}{z} \int_0^\infty H(\rho)\, J_m\left(\frac{k}{z} r\rho\right) \rho\, \mathrm{d}\rho. \tag{2.104}$$

Для быстрого расчета преобразований Ханкеля (ПХ) можно воспользоваться методом экспоненциальной замены переменных [27]. Этот метод позволяет свести ПХ к корреляции, которую можно вычислить при помощи преобразования Фурье.

Действительно, после замены переменных $r = r_0 e^x$, $\rho = \rho_0 e^y$, где r_0 и ρ_0 – постоянные, вместо (2.103) мы получим

$$\bar{H}(y) = \frac{k}{z} r_0^2 \int_{-\infty}^\infty \bar{h}(x)\, S(x+y)\, e^{2x}\, \mathrm{d}x, \tag{2.105}$$

где

$$\bar{h}(x) = h(r_0 e^x), \quad S(x+y) = J_m\left(\frac{k}{z} r_0 \rho_0 e^{x+y}\right), \quad \bar{H}(y) = H(\rho_0 e^y). \tag{2.106}$$

Функция Бесселя при $m = 0$ стремится к единице при $x \to -\infty$. Поэтому чтобы функция $S(x)$ стремилась к нулю при $x \to \pm\infty$, ее можно умножить на $\exp(x/4)$, а чтобы подынтегральное выражение в (2.105) не изменилось, следует также умножить функцию $h(x)$ на $\exp(-x/4)$.

Т.к. функция пропускания $h(r)$ ДОЭ ограничена апертурой $r \in [0, R]$, где R – радиус апертуры, то для функции $\bar{h}(x)$ при $x \to \pm\infty$ не возникает расходимостей. После переобозначения получаются выражения

$$h_1(x) = \bar{h}(x)\, r_0^2 e^{7x/4}, \quad S_1(x+y) = S(x+y) \exp\left(\frac{x+y}{4}\right), \quad H_1(y) = \bar{H}(y)\, e^{y/4}, \tag{2.107}$$

и вместо (2.103) получаем интеграл корреляции

$$H_1(y) = \frac{k}{z} \int_{-\infty}^a h_1(x)\, S_1(x+y)\, \mathrm{d}x, \quad a = \ln\frac{R}{r_0}. \tag{2.108}$$

Интеграл (2.108) можно выразить через преобразование Фурье в виде

$$H(\rho) = \frac{k}{z} \left[\frac{\rho}{\rho_0}\right]^{-1/2} \int_{-\infty}^\infty P(-w)\, U(w)\, e^{iw \ln(\rho/\rho_0)}\, \mathrm{d}w, \tag{2.109}$$

где $P(w)$ — это фурье-образ функций $h_1(x)$, $U(w)$ — фурье-образ $S_1(y)$. Аналогичным образом можно представить и обратное преобразование Ханкеля в (2.104). Отметим, что существуют и другие методы быстрого расчета ПХ [28, 29].

Возвращаясь к решению уравнения (2.102), объединим обе функции в показателе степени в одну функцию, которую следует найти:

$$\bar{\varphi}(r) = \varphi(r) + \frac{k}{2z} r^2. \tag{2.110}$$

Процедура итеративного решения уравнения (2.102) реализуется аналогично алгоритму ГС (2.4) и последовательно использует пару преобразований, состоящую из прямого и обратного ПХ. При этом осуществляется ограничивающая замена переменных как в плоскости ДОЭ, так и в плоскости наблюдения, на n-м шаге итераций. Функция $g_n(r)$, рассчитанная в плоскости ДОЭ, заменяется на функцию $\bar{g}_n(r)$ согласно правилу

$$\bar{g}_n(r) = \begin{cases} A_0(r)\, g_n(r)\, |g_n(r)|^{-1}, & r \in [0, R], \\ 0, & r \notin [0, R]. \end{cases} \tag{2.111}$$

Рассчитанная амплитуда $F_n(\rho)$ в плоскости наблюдения заменяется на $\bar{F}_n(\rho)$, при помощи соотношения

$$\bar{F}_n(\rho) = \sqrt{I_0(\rho)} F_n(\rho)\, |F_n(\rho)|^{-1}. \tag{2.112}$$

Между заменами (2.111) и (2.112) и соответствующими заменами для двумерного алгоритма уменьшения ошибки (2.4) и (2.6) нет существенной разницы. Чтобы увеличить скорость сходимости алгоритма, вместо уравнения (2.112) может быть использован АА-алгоритм (2.17) при соответствующем переходе от двумерного к радиальному случаю.

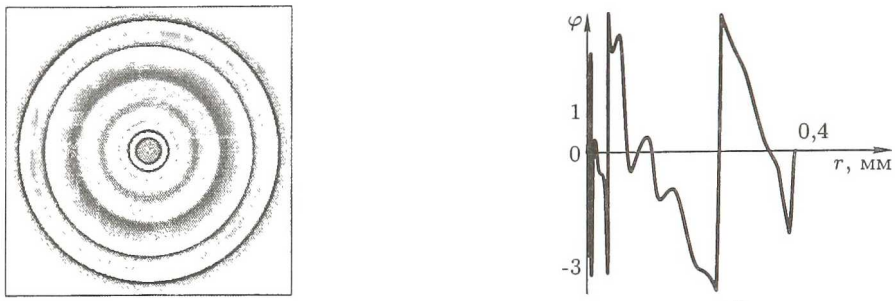
Данный алгоритм позволяет рассчитывать ДОЭ, предназначенные для формирования таких изображений, как круг, кольцо, или набор колец [29, 30].

Пример 2.5. На рис. 2.13 представлена фаза (после 22 итераций) ДОЭ, способного преобразовывать лазерный свет с гауссовым распределением интенсивности в круг с однородной интенсивностью в фокусе линзы. Интенсивность гауссова пучка на границах апертуры ДОЭ составляет 0,1 от максимального ее значения в центре пучка. Радиус ДОЭ равен 0,4 мм; число отсчетов вдоль радиуса 256, $k/f = 100$ мм$^{-2}$; радиус круга в плоскости изображения составляет 0,3 мм и примерно равен трем радиусам минимального дифракционного пятна (диск Эйри).

На рис. 2.14 показана дифракционная картина, которую формирует ДОЭ с такой фазой в фокальной плоскости. Среднеквадратичное отклонение интенсивности в сформированном круге от постоянного значения составляет 6%, а энергетическая эффективность фокусировки в круг равна 91%. Расчет осуществлялся при помощи ПХ (для $m = 0$, преобразования Фурье–Бесселя).

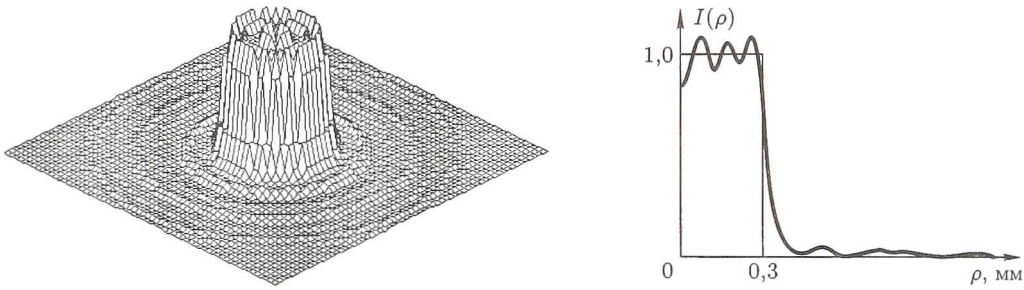
Если необходимо рассчитать ДОЭ с радиальной фазой, формирующий распределение интенсивности типа кольца, и если для такого расчета использовать ПХ нулевого порядка, то в центре кольца практически всегда будет формироваться пик интенсивности. Для его устранения надо использовать ПХ более высокого порядка.

Пример 2.6. На рис. 2.15 показан результат расчета фокусатора в широкое кольцо постоянной интенсивности. В этом случае расчет проводился при помощи преобразования Ханкеля 8-го порядка (в уравнении (2.99) $m = 8$), радиус ДОЭ равнялся



а *б*

Рис. 2.13. Фаза фокусатора в круг: фаза ДОЭ (*а*), ее радиальное сечение (*б*)



а *б*

Рис. 2.14. Сформированное изображение круга: распределение интенсивности (*а*) и ее радиальное сечение (*б*)

5 мм, радиусы кольца были: $r_1 = 0{,}1$ мм и $r_2 = 0{,}2$ мм, число пикселов вдоль радиуса составило 256.

Энергетическая эффективность фокусировки в кольцо равнялась 83%, среднеквадратичное отклонение — 8%.

Заметим, что азимутальный множитель в уравнении (2.99), в виде $e^{im\varphi}$, сам по себе представляет интерес как функция пропускания фазового оптического элемента, используя который можно осуществлять множество оптических преобразований: преобразование Ханкеля m-го порядка [31], формирование пучков Бесселя высших порядков [32], оптическое дифференцирование радиально-симметричных световых пучков [33], а также радиальный аналог преобразования Гильберта [34].



а *б*

Рис. 2.15. Фаза фокусатора в кольцо (*а*), дифракционная картина при $m = 8$ (*б*)

2.8.2. Расчет дифракционных аксиконов, формирующих осевые световые отрезки

В [35] рассмотрен новый оптический элемент — аксикон, который представляет собой конус из прозрачного материала. Такой элемент используется для формирования узкого осевого светового пучка, диаметр поперечного сечения которого обратно пропорционален углу при вершине конуса, а длина пропорциональна диаметру освещающего пучка и обратно пропорциональна углу при вершине конуса (световой пучок должен падать перпендикулярно основанию конуса). Вместе со сферической линзой аксикон используется для формирования узкого светового кольца [36].

Фазовые оптические элементы, способные в комбинации со сферической линзой формировать световое поле с увеличенной глубиной резкости, называются фокусаторами в продольный отрезок [37] и могут рассматриваться как обобщенные аксиконы [36]. Они широко примененяются для изучения разрядов в газах [38], для записи и считывания данных с оптических дисков [39], для тестирования поверхностей [40].

Ниже рассматриваются итеративные алгоритмы расчета ДОЭ (обобщенных аксиконов), которые могут формировать световые пучки заданного распределения интенсивности вдоль оптической оси.

Пусть требуется рассчитать радиальную фазовую функцию оптического элемента, который формирует осевой световой отрезок, ограниченный точками z_1 и z_2 и характеризующийся заданным распределением интенсивности вдоль оси $I_0(z)$, $z \in \in [z_1, z_2]$. В этом случае уравнение (2.100) примет вид

$$F(\rho, z) = \frac{ik}{z} e^{ikz} e^{ik\rho^2/2z} \int_0^R \exp\left[i\varphi(r) + i\frac{kr^2}{2z}\right] J_0\left(\frac{kr\rho}{z}\right) r\,\mathrm{d}r, \tag{2.113}$$

где R — радиус ДОЭ, $\varphi(r)$ — заданная фаза, $J_0(x)$ — функция Бесселя нулевого порядка. В (2.113) предполагается, что оптический элемент освещен плоской световой волной. Так как известна интенсивность только вдоль оси z, можно положить в (2.113) $\rho = 0$ и получить (опуская тривиальный сомножитель $\exp(ikz)$) уравнение

$$F(0, z) = \frac{k}{z} \int_0^R \exp\left[i\varphi(r) + i\frac{kr^2}{2z}\right] r\,\mathrm{d}r. \tag{2.114}$$

Введя обозначения $\xi = k/z$, $x = r^2/2$, перепишем уравнение (2.114) в виде

$$\frac{1}{\xi} F(\xi) = \int_0^a e^{i\varphi(x) + ix\xi}\,\mathrm{d}x, \quad a = \frac{R^2}{2}. \tag{2.115}$$

Из (2.115) видно, что заданная функция пропускания ДОЭ-аксикона $e^{i\varphi(r)}$ связана с осевой комплексной амплитудой волны $F(\xi)$ через одномерное преобразование Фурье, и проблема сводится к решению одномерного интегрального уравнения

$$\frac{1}{\xi^2} I_0(\xi) = \left| \int_0^a e^{i\varphi(x) + ix\xi}\,\mathrm{d}x \right|^2. \tag{2.116}$$

В этом случае применим алгоритм уменьшения ошибки и все его модернизации.

Если требуется рассчитать аксикон в качестве дополнения к сферической линзе, то фаза в уравнении (2.114) должна быть представлена в виде

$$\varphi(r) = \varphi_0(r) - \frac{kr^2}{2f}, \tag{2.117}$$

где f — фокусное расстояние сферической линзы. В этом случае вместо (2.114) можно записать

$$F(0,z) = \frac{k}{z}\int_0^R \exp\left[i\varphi_0(r) + i\frac{kr^2}{2}\left(\frac{1}{z} - \frac{1}{f}\right)\right] r\,\mathrm{d}r. \tag{2.118}$$

Уравнение (2.118) относительно $\varphi_0(r)$ будет эквивалентно уравнению (2.115), если использовать следующую замену переменных:

$$\xi = k\left(\frac{1}{z} - \frac{1}{f}\right), \quad x = \frac{r^2}{2}. \tag{2.119}$$

Чтобы сформировать короткие световые отрезки (при $2d = z_2 - z_1 \ll f$, где $2d$ — глубина фокуса линзы или длина заданного светового отрезка), вместо уравнения (2.118) можно использовать более простое выражение

$$F(\Delta z) = \frac{k}{f}\int_0^R \exp\left[i\varphi_0(r) + i\frac{k\Delta z}{2f^2}r^2\right] r\,\mathrm{d}r, \tag{2.120}$$

где $\Delta_z \in [-d, d]$, $f = (z_1 + z_2)/2$. После замены переменных $\xi = k\Delta z/f^2$, $x = r^2/2$ получается уравнение для поиска фазы ДОЭ:

$$I_0(\xi) = \left|\frac{k}{f}\int_0^a e^{i\varphi_0(x) + ix\xi}\,\mathrm{d}x\right|^2. \tag{2.121}$$

Уравнение (2.121) может быть решено при помощи итеративной процедуры [41, 42, 43], при этом степень близости оценочной функции $I_n(\xi) = |F_n(\xi)|^2$ к заданной функции $I_0(\xi)$ определяется из среднеквадратичного отклонения

$$\delta = \left[\int_{\xi_1}^{\xi_2} |I_n(\xi) - I_0(\xi)|^2\,\mathrm{d}\xi\right]^{1/2} \cdot \left[\int_{\xi_1}^{\xi_2} I_0^2(\xi)\,\mathrm{d}\xi\right]^{-1/2}, \quad \xi_1 = \frac{k}{z_1}, \quad \xi_2 = \frac{k}{z_2}. \tag{2.122}$$

Энергетическая эффективность формирования осевого светового отрезка находится из выражения

$$E_k = \int_0^{\rho_0} I(\rho, z_k)\,\rho\,\mathrm{d}\rho \Big/ \int_0^{\infty} I(\rho, z_k)\,\rho\,\mathrm{d}\rho, \tag{2.123}$$

где $I_0(\rho, z_k)$ — интенсивность, рассчитанная в плоскости, находящейся на расстоянии z_k от ДОЭ, а ρ_0 — радиус первого локального минимума интенсивности $I(\rho, z_k)$.

Функция E_k определяет порцию световой энергии, которая идет на формирование минимального дифракционного пятна в заданной плоскости наблюдения.

Заметим, что если заданная интенсивность $I_0(z)$ выражается в виде

$$I_0(z) = I_0 \sum_{n=1}^{N} \delta(z - z_n), \tag{2.124}$$

где $\delta(x)$ — функция Дирака, то $\varphi_0(r)$ из (2.121) будет представлять собой фазу N-фокусной линзы с N продольными фокусами равной интенсивности I_0.

Пример 2.7. На рис. 2.16*а* представлена фаза ДОЭ, рассчитанная за 100 итераций, при этом для решения (2.121) был использован алгоритм ГС. Такой ДОЭ создает световой пучок с осевым распределением интенсивности, имеющим форму прямоугольного импульса (рис. 2.16*б*). Параметры расчета: $N = 256$ — общее число пикселов; $R = 6$ мм; $f = 400$ мм — фокусное расстояние линзы, которая в окрестности своей фокальной плоскости содержит сформированный отрезок длиной $2a = 40$ мм. Длина волны 0,63 мкм. Предполагается, что освещающий световой пучок плоский. Ошибка (2.122) равна 2%. Эффективность фокусировки (2.123) изменяется вдоль отрезка сложным образом и лежит в пределах от 15% до 85%. Длина отрезка равна примерно восьми длинам Френеля.

На рис. 2.17*а* показана фаза ДОЭ в виде амплитудной маски, формирующей 10 продольных фокусов равной интенсивности (рис. 2.17*б*). Фаза многофокусной линзы приближенно представляет собой сумму фаз двух аксиконов (ДОЭ), формирующих при этом отрезки различной длины. При этом в пространстве распространяются две конических волны с различным наклоном к оптической оси. В результате интерференции этих волн и появляются продольные фокусы.

2.8.3. Расчет радиально симметричных ДОЭ с квантованной фазой

В данном пункте рассматривается итеративный алгоритм, позволяющий осуществлять расчет радиально-симметричных ДОЭ с небольшим числом уровней фазы, что дает существенное преимущество при изготовлении ДОЭ методами фото- или электронной литографии. На практике такие ДОЭ могут быть использованы для расширения минимального дифракционного пятна. Эта проблема актуальна для лазерной печати. Двукратное расширение дифракционного пятна (диска Эйри) может быть осуществлено без применения ДОЭ, путем двукратного увеличения фокальной

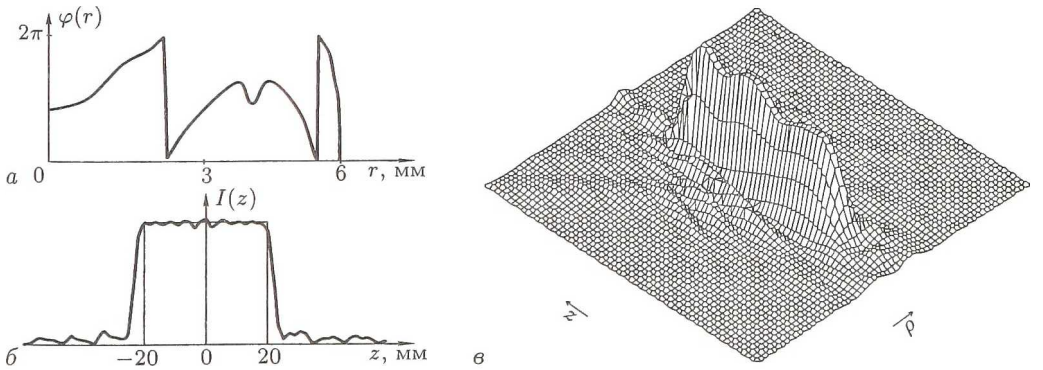


Рис. 2.16. ДОЭ, фокусирующий в осевой отрезок: радиальное сечение фазы (*а*), осевое распределение интенсивности (*б*), трехмерное распределение интенсивности (*в*)

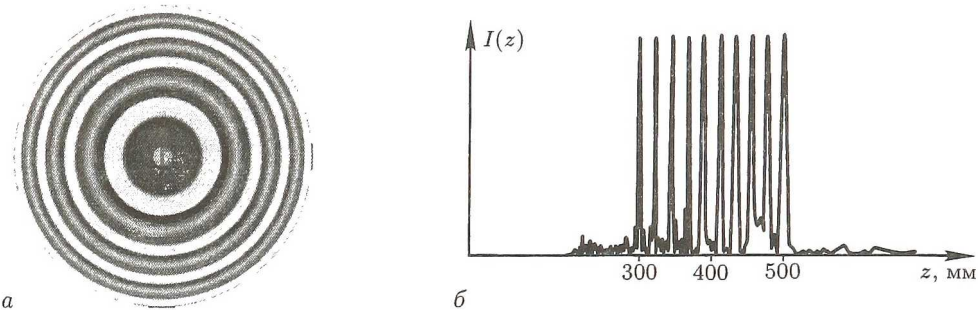


Рис. 2.17. Расчет многофокусной линзы: фаза оптического элемента (*а*), осевое распределение интенсивности (*б*)

длины линзы или путем уменьшения радиуса апертуры линзы в два раза. Однако, первый подход приводит к нежелательным изменениям в конструкции, а второй – уменьшает эффективность фокусировки в 4 раза.

Предлагается использовать комбинацию «ДОЭ+линза», которая обеспечивает более эффективное решение данной проблемы [44]. Предположим, что требуется рассчитать фазу ДОЭ $\varphi(r)$, которая зависит только от радиальной переменной, и что данный ДОЭ создает заданное распределение интенсивности $I_0(\rho)$ в фокальной плоскости линзы. В этом случае, вместо уравнения (2.102) можно записать следующее уравнение для расчета фазы $\varphi(r)$:

$$I_0(\rho) = \left| \frac{k}{f} \int A_0(r)\, e^{i\varphi(r)} J_0\left(\frac{kr\rho}{f}\right) r\, \mathrm{d}r \right|^2, \tag{2.125}$$

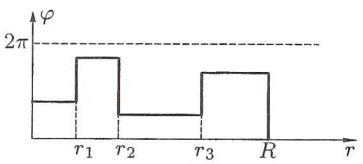


Рис. 2.18. Кусочно-постоянная фаза

где $A_0(r)$ — амплитуда освещающего пучка, f — фокусное расстояние линзы. Предположим далее, что выполняется условие $A_0(r) = 1$. Будем искать функцию $\varphi(r)$ в кусочно-постоянном виде (рис. 2.18). Тогда можно представить комплексную амплитуду волны $F(\rho)$ в фокальной плоскости линзы как сумму амплитуд, получающихся в результате дифракции плоской волны на кольцевой апертуре:

$$F(\rho) = 2\pi \sum_{n=0}^{N-1} e^{i\varphi_{n+1}} \left[r_{n+1} \frac{J_1(k\rho r_{n+1}/f)}{\rho} - r_n \frac{J_1(k\rho r_n/f)}{\rho} \right], \tag{2.126}$$

где φ_{n+1} — это значение $\varphi(r)$ при $r \in [r_n, r_{n+1}]$, $n = 0, 1, \ldots, N-1$; $r_0 = 0$, $r_N = R$, R — радиус ДОЭ, N — число точек деления радиуса ДОЭ. После соответствующего преобразования членов уравнения (2.126) получим

$$F(\rho) = \frac{2\pi}{\rho} \sum_{n=1}^{N} C_n J_1\left(\frac{\gamma_n \rho}{a}\right), \tag{2.127}$$

где

$$C_n = r_n \left[e^{i\varphi_n} - e^{i\varphi_{n+1}} \right], \quad C_N = \mathrm{R}e^{i\varphi_N}, \tag{2.128}$$

γ_n — нули функции Бесселя первого порядка: $J_1(\gamma_n) = 0$.

Из сравнения уравнений (2.126) и (2.127) следует, что точки деления r_n должны удовлетворять следующему условию:

$$r_n = \frac{\gamma_n f}{ka}, \quad a \geqslant R, \tag{2.129}$$

где a — константа, контролирующая масштаб деления ДОЭ на кольца.

Введя обозначения $\bar{F}(\rho) = (2\pi)^{-1} \rho F(\rho)$, вместо (2.127) получим

$$\bar{F}(\rho) = \sum_{n=1}^{N} C_n J_1\left(\frac{\gamma_n \rho}{a}\right). \tag{2.130}$$

Используя свойство ортогональности функции Бесселя

$$\int_0^1 J_m(\gamma_p x) J_m(\gamma_q x) x \, \mathrm{d}x = \begin{cases} [J'_m(\gamma_p)]^2/2, & p = q, \\ 0, & p \neq q, \end{cases} \tag{2.131}$$

где $J'_m(\gamma_p)$ — производная функции Бесселя в точке γ_p, коэффициенты суммы (2.130) можно представить в виде

$$C_n = 2\left[aJ'_1(\gamma_n)\right]^{-2} \int_0^a \bar{F}(\rho) J_1\left(\frac{\gamma_n \rho}{a}\right) \rho \, \mathrm{d}\rho. \tag{2.132}$$

Соотношения (2.130) и (2.132) позволяют организовать следующую итеративную процедуру для решения уравнения (2.125).

1. Начальная оценка $\bar{F}(\rho)$ комплексной амплитуды выбирается в виде

$$\bar{F}(\rho) = \sqrt{I_1(\rho)} e^{i\psi_0(\rho)}, \tag{2.133}$$

где $I_1(\rho) = (2\pi)^{-2}\rho^2 I_0(\rho)$, $I_0(\rho)$ — заданная интенсивность света в фокальной плоскости линзы, $\psi_0(\rho)$ — случайная оценка фазы.

2. На k-м шаге итераций функцию $\bar{F}_k(\rho)$ представляют как сумму (2.130) с коэффициентами $C_n^{(k)}$, которые находятся из уравнения (2.132) и подвергаются замене

$$\bar{C}_n^{(k)} = B_n e^{i\nu_{nk}}, \quad n = \overline{1, N}, \tag{2.134}$$

где

$$B_n = 2r_n \sin\left[\frac{\varphi_n - \varphi_{n+1}}{2}\right], \tag{2.135}$$

$$\left.\begin{aligned} &\varphi_n = \pi - \varphi_{n+1} + 2\nu_{nk}, \quad n = \overline{N-1, 1}, \\ &\varphi_N = \nu_{Nk}, \end{aligned}\right\} \tag{2.136}$$

где $\nu_{nk} = \arg C_n^{(k)}$. Фазы φ_n ищутся при помощи рекуррентных соотношений (2.136), вытекающих из уравнения (2.128).

3. Формируется $(k+1)$-я оценка функции $\bar{F}_{k+1}(\rho)$:

$$\bar{F}_{k+1} = \sum_{n=1}^{N} \bar{C}_n^{(k)} J_1\left(\frac{\gamma_n \rho}{a}\right). \tag{2.137}$$

4. Рассчитанная функция $\bar{F}_{k+1}(\rho)$ заменяется функцией $\hat{F}_{k+1}(\rho)$ по правилу

$$\hat{F}_{k+1}(\rho) = \sqrt{I_1(\rho)}\bar{F}_{k+1}(\rho)\left|\bar{F}_{k+1}(\rho)\right|^{-1}. \tag{2.138}$$

5. Осуществляется переход к шагу 2, и т. д.

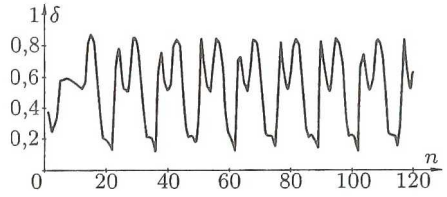


Рис. 2.19. Зависимость ошибки от числа итераций

Данный алгоритм не сходится: с ростом числа итераций ошибка изменяется квазипериодически (см. рис. 2.19 для примера 2.6).

Пример 2.8. Рассчитаем ДОЭ с квантованной фазой, который фокусирует свет в кольцо [45]. Параметры расчета следующие: $k/f = 100$ мм$^{-2}$, $R = 0{,}5$ мм — радиус ДОЭ, дискретность вдоль ρ равна 2 мкм, общее число отсчетов 256. Радиусы кольца равны $R_1 = 0{,}3$ мм, $R_2 = 0{,}5$ мм. На рис. 2.20 показан результат расчета при разбиении радиуса на несколько колец ($N = 15$) после 12 итераций. Значения энергетической эффективности фокусировки в кольцо 89%. Отметим, что в отличие от общепринятой практики [21] в данном случае квантование фазы неэквидистантное. Обычно рассчитанную непрерывную функцию фазы $\varphi(x)$ подвергают эквидистантному равномерному квантованию по N уровням:

$$\varphi_n = \left\{ \frac{2\pi}{N}n; \quad \frac{2\pi}{N}n < \varphi(x) < \frac{2\pi}{N}(n+1), \quad n = 0, 1, \ldots (N-1) \right\}.$$

В алгоритме (2.136) уже имеются N отсчетов фазы φ_n, связанные с числом колец радиального ДОЭ и они, как правило, неравномерные.

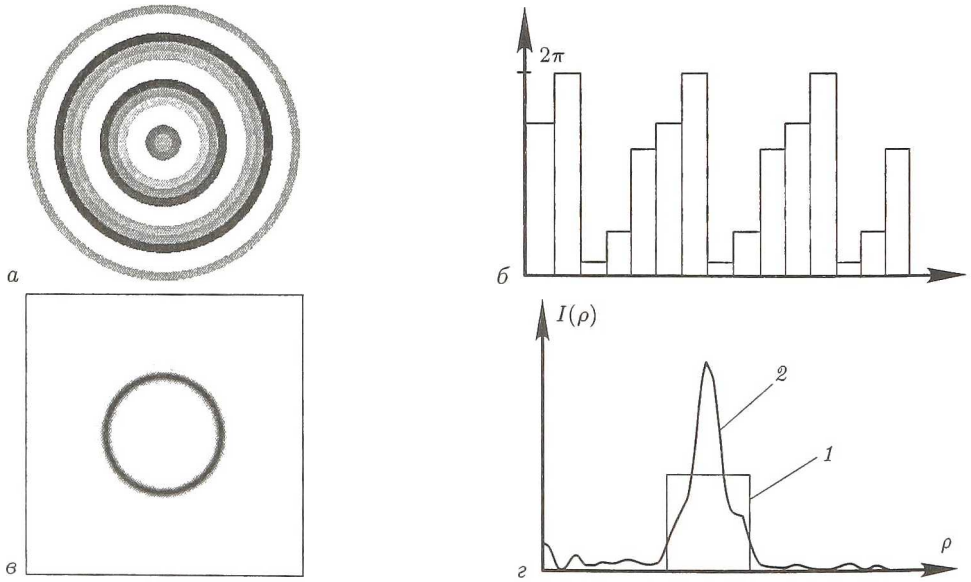


Рис. 2.20. Радиальный ДОЭ с пятью уровнями квантования фазы: фаза ДОЭ (*а*), ее радиальное сечение (*б*), дифракционная картина в фокальной плоскости линзы (*в*), ее радиальное сечение (*г*) (кривая *1* соответствует рассчитанной интенсивности, кривая *2* —заданной интенсивности

2.8.4. Многопорядковые фазовые бинарные дифракционные решетки

Многопорядковые фазовые дифракционные решетки представляют собой дифракционную структуру с периодическим фазовым рельефом и предназначены для формирования одномерного или двумерного набора плоских пучков с заданным соотношением энергии между пучками. Многопорядковые дифракционные решетки применяются в оптических устройствах мультипликации изображений, в соединителях для оптических волокон и свободного пространства, в устройствах оптической связи и обработки информации, в когерентных оптических процессорах [46–51].

Бинарные дифракционные решетки представляют особый интерес ввиду простоты их изготовления методами фотолитографии; бинарный рельеф формируется за одно химическое травление подложки.

Для расчета бинарных решеток используются, как правило, методы стохастической оптимизации, такие как прямой бинарный поиск [52, 53], алгоритм симуляции отжига [54–58] и итеративные алгоритмы [59–62]. В работах [63, 64] расчет бинарных решеток основан на решении системы нелинейных уравнений, представляющих интенсивности порядков решетки через координаты фазового профиля. При этом расчет бинарной решетки с числом порядков $2N+1$ сводится к решению нелинейной системы из N уравнений.

Для использования бинарных дифракционных решеток в качестве преобразователей лазерных пучков и формирователей пучков в сетях на оптических волокнах, является актуальным увеличение числа порядков бинарной решетки до нескольких сотен. С ростом числа порядков алгоритмы случайного поиска профиля дифракционной решетки требуют большого времени расчета и приводят к нестабильной сходимости, а методы расчета, основанные на решении системы нелинейных уравнений, становятся численно неустойчивыми. В данном пункте рассмотрен расчет многопорядковых бинарных решеток на основе градиентного алгоритма при специальном выборе начального приближения для профиля решетки. Выбор начального приближения основан на свойствах одномерного бинарного элемента для фокусировки в отрезок. Фазовый профиль бинарной решетки состоит из K прямоугольных штрихов одинаковой высоты φ но разной ширины (рис. 2.21). Координаты $x_1, \ldots, x_{2K}$ этих штрихов являются параметрами оптимизации и определяют значения интенсивности в дифракционных порядках.

Для выбора начального приближения профиля бинарной решетки, рассмотрим предварительно расчет одномерного фазового ДОЭ с апертурой d, фокусирующего при $z = f$ в отрезок с распределением интенсивности $I(\xi)$ (см. рис. 2.22). Отрезок фокусировки $\xi \in [n_1\Delta, n_2\Delta]$, где $\Delta = \lambda f/d$ — размер дифракционного пятна, может быть представлен как совокупность $(n_2 - n_1 + 1)$ дифракционных пятен. Фазовая функция $\varphi(x)(0 \leqslant x \leqslant d)$ одномерного ДОЭ может быть рассчитана аналитически в приближении геометрической оптики [65–67].

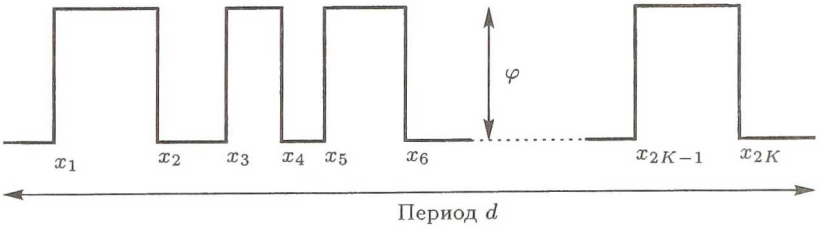


Рис. 2.21. Фазовая функция одного периода бинарной решетки

Для сходящегося цилиндрического освещающего пучка с фокусом f,

$$W_0(x) = \exp\left(-\frac{ikx^2}{2f}\right), \tag{2.139}$$

геометрооптическая функция $\varphi(x)$ имеет вид

$$\varphi(x) = \frac{2\pi}{d}(n_2 - n_1)\int_0^x \kappa(\alpha)\,\mathrm{d}\alpha + \frac{2\pi}{d}n_1 x. \tag{2.140}$$

Функция $\kappa(\alpha)$ в (2.140) ищется из решения следующего уравнения:

$$\frac{\mathrm{d}\kappa(\alpha)}{\mathrm{d}\alpha} = \left[I\left(\frac{\kappa(\alpha) - n_1\Delta}{(n_2 - n_1)\Delta}\right)\right]^{-1}, \tag{2.141}$$

с граничными условиями: $\kappa(0) = 0$, $\kappa(1) = 1$. Функция $\varphi(x)$ соответствует фазовой функции небинарной дифракционной решетки, формирующей $(n_2 - n_1 + 1)$ дифракционных порядков. Действительно, в приближении Френеля–Кирхгофа значения интенсивности в точках $\xi = \xi_j = j\Delta$ в плоскости $z = f$,

$$I(\xi_j) = \frac{1}{\lambda f}\left|\int_0^d \exp\left(i\varphi(x) - i\frac{2\pi j}{d}x\right)\mathrm{d}x\right|^2, \tag{2.142}$$

пропорциональны квадратам модулей коэффициентов Фурье в разложении функции $\exp(i\varphi(x))$. Следовательно можно рассматривать $\varphi(x)$ как фазовую функцию дифракционной решетки с периодом d и интенсивностью дифракционных порядков

$$I_j = \begin{cases} I(j\Delta), & n_1 \leqslant j \leqslant n_2, \\ 0, & \text{иначе.} \end{cases} \tag{2.143}$$

Для расчета бинарной решетки, формирующей $2N + 1$ порядков в интервале $[-N\Delta, +N\Delta]$, определим бинарную фазовую функцию в виде [68, 69]

$$\varphi_b(x) = \Phi[\bar{\varphi}(x)], \tag{2.144}$$

где $\bar{\varphi}(x)$ — непрерывная функция (2.140), рассчитанная из условия формирования $(N + 1)$ порядков в интервале $[0, N\Delta]$ и приведенная к интервалу $[0, 2\pi)$, а

$$\Phi[\eta] = \begin{cases} 0, & \eta \in [0, \pi), \\ \pi, & \eta \in [\pi, 2\pi), \end{cases} \tag{2.145}$$

— бинарная функция, описывающая квантование фазы $\bar{\varphi}(x)$ по двум уровням.

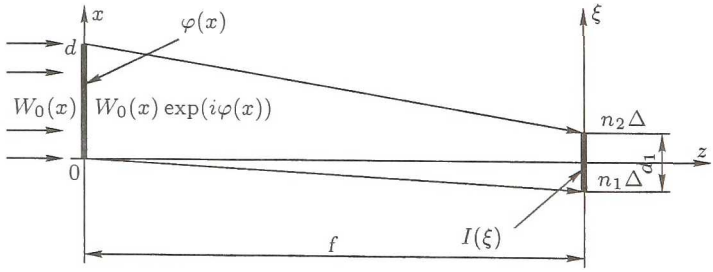


Рис. 2.22. Геометрия фокусировки в поперечный отрезок

Для описания бинарной решетки (2.144), разложим функцию $\exp(i\Phi[\eta])$ на интервале $[0, 2\pi)$ в ряд Фурье:

$$\exp\left(i\Phi\left[\eta\right]\right) = \sum_{n=-\infty}^{\infty} c_n \exp\left(in\eta\right), \tag{2.146}$$

где

$$c_n = \begin{cases} \dfrac{1-(-1)^n}{i\pi n}, & n \neq 0, \quad n = \pm 1, \pm 2, \ldots, \\ 0, & n = 0 \end{cases} \tag{2.147}$$

— коэффициенты Фурье. Полагая $\eta = \bar{\varphi}\left(x\right)$ в (2.146), представим функцию комплексного пропускания бинарной решетки с учетом 2π-периодичности в виде

$$\exp\left(i\Phi\left[\bar{\varphi}\left(x\right)\right]\right) = \sum_{n=-\infty}^{\infty} c_n \exp\left(in\varphi\left(x\right)\right). \tag{2.148}$$

Уравнение (2.148) соответствует суперпозиции пучков с фазовыми функциями

$$\varphi_n(x) = n\varphi(x). \tag{2.149}$$

Фазовые функции (2.149) порождены нелинейной операцией бинаризации (2.145) и также называются дифракционными порядками. Квадраты модулей коэффициентов Фурье (2.147) представляют распределение энергии между пучками (2.149). Поскольку $|c_1|^2 = |c_{-1}|^2 = 0{,}405$, то 81% энергии падающего пучка приходится на -1 и $+1$ дифракционные порядки. Изображения, формируемые в главных порядках -1 и $+1$, определяются функциями $-\varphi(x)$ и $+\varphi(x)$. Функции $-\varphi(x)$ и $+\varphi(x)$ являются непрерывными фазовыми функциями дифракционных решеток, формирующих по N порядков в интервалах $[-N\Delta, 0]$ и $[0, N\Delta]$, соответственно. Суммирование интервалов $[-N\Delta, 0]$ и $[0, N\Delta]$ дает $2N+1$ порядков на интервале $[-N\Delta, N\Delta]$. Бинарная решетка (2.144) может рассматриваться лишь как приближенное решение. Действительно, функция $\varphi(x)$ была рассчитана в приближении геометрической оптики, а при анализе работы решетки (2.144) были рассмотрены только два главных дифракционных порядка без учета интерференции между ними. Поэтому, даже если эффективность, даваемая приближенным решением близка к 80% ($|c_1|^2 + |c_{-1}|^2 = 0{,}81$), то среднеквадратичная ошибка формирования заданной интенсивности порядков остается в пределах 40–50%. В дальнейшем приближенное решение (2.144) будет использоваться как начальное приближение профиля для градиентного алгоритма.

Согласно уравнениям (2.140)–(2.145), координаты начального профиля решетки для формирования $2N+1$ порядков равной интенсивности, имеют вид

$$x_n = \sqrt{\frac{n}{N}}, \quad n = \overline{1, N}. \tag{2.150}$$

Аналогично, для решетки с симметричным линейным распределением интенсивности в дифракционных порядках,

$$I_n = I_0 + \alpha\left|n\right|, \quad n = -\overline{N, N}, \tag{2.151}$$

координаты начального профиля решетки определяются из следующего уравнения:

$$\frac{2N}{(b-1)(b^2-1)}\left[-3x_n\left(b^2-1\right)-2+2\left(1+\left(b^2-1\right)x_n\right)^{3/2}\right]=n,\quad n=1,2,\ldots, \tag{2.152}$$

где $b=1+\alpha N/I_0$.

Для оптимизации структуры решетки введем функцию ошибки, представляющую отличие между заданными значениями интенсивности $\hat{I}_j$ в порядках и расчетными значениями I_j в виде

$$\varepsilon(\mathbf{x})=\sum_{j=-N}^{N}\left(I_j(\mathbf{x})-\hat{I}_j\right)^2, \tag{2.153}$$

где $\mathbf{x}=(x_1,\ldots,x_{2K})$ — вектор координат границ штрихов. Без ограничения общности период решетки d будем считать равным единице. При этом интенсивности порядков, соответствующие квадратам модулей коэффициентов Фурье в разложении функции пропускания решетки, имеют вид [54, 64]

$$\begin{cases} I_j(x)=\sin^2(\varphi/2)\dfrac{\left(C_j^2+S_j^2\right)}{(\pi j)^2}, & j\neq 0,\quad j=\pm1,\pm2,\ldots \\ I_0(x)=1-4Q(1-Q)\sin^2(\varphi/2), \end{cases} \tag{2.154}$$

где

$$C_j=\sum_{n=1}^{2K}(-1)^n\cos(2\pi j x_n),\quad S_j=\sum_{n=1}^{2K}(-1)^n\sin(2\pi j x_n),\quad Q=\sum_{n=1}^{2K}(-1)^n x_n. \tag{2.155}$$

Из уравнений (2.154) и (2.155) следует, что $I_j=I_{-j}$. Вследствие симметрии, расчет бинарной решетки с числом порядков $2N+1$ требует, как минимум, $N+1$ свободных параметров; соответственно число штрихов K на периоде должно быть не менее $N/2$ [54, 64].

Для расчета профиля решетки будем использовать градиентный метод минимизации функции ошибки $\varepsilon(\mathbf{x})$. Основной проблемой, связанной с градиентными процедурами, является выбор начального профиля для получения стабильной сходимости. Как правило, при случайном выборе начального профиля, градиентные алгоритмы стагнируют при среднеквадратичной ошибке формирования заданных значений интенсивности порядков в 75–85%. Использование аналитического начального приближения (2.144), (2.150), (2.152) дает быструю и стабильную сходимость градиентных процедуры при высокой эффективности и низкой среднеквадратичной ошибке. Рассмотрим метод сопряженного градиента для минимизации функции ошибки $\varepsilon(\mathbf{x})$. Метод состоит в итерационной коррекции координат профиля решетки по правилу [70]

$$\mathbf{x}_{n+1}=\mathbf{x}_n+\mathbf{h}(\mathbf{x}_n)\,t, \tag{2.156}$$

где $\mathbf{h}(\mathbf{x}_n)=-\mathbf{g}(\mathbf{x}_n)+\gamma_n\mathbf{h}(\mathbf{x}_{n-1})$, $\mathbf{g}(\mathbf{x})=(\partial\varepsilon(\mathbf{x})/\partial x_1,\ldots,\partial\varepsilon(\mathbf{x})/\partial x_{2K})=(g_1(\mathbf{x}),\ldots$ $\ldots,g_{2K}(\mathbf{x}))$ — градиент функции $\varepsilon(\mathbf{x})$, t — шаг градиентного метода,

$$\gamma_n=-\frac{(\mathbf{g}(\mathbf{x}_n),\mathbf{g}(\mathbf{x}_n)-\mathbf{g}(\mathbf{x}_{n-1}))}{|\mathbf{g}(\mathbf{x}_{n-1})|^2}. \tag{2.157}$$

Согласно уравнениям (2.153)–(2.155) и соотношению симметрии, компоненты вектора градиента при $\varphi = \pi$ могут быть представлены в виде

$$g_n(\mathbf{x}) = 16\frac{(-1)^n}{\pi}\sum_{j=1}^{N}\frac{1}{j}\left(S_j\cos(2\pi j x_n) - C_j\sin(2\pi j x_n)\right)\times$$
$$\times\left[I_j(\mathbf{x}) - \hat{I}_j\right] + 4(-1)^n(2Q-1)\left[I_0(\mathbf{x}) - \hat{I}_0\right], \quad (2.158)$$

где S_j, C_j и Q определены в (2.155).

Для расчета шага t, представим функцию ошибки вдоль сопряженного направления как функцию от t: $\varepsilon(\mathbf{x}_n + \mathbf{h}(\mathbf{x}_n)\,t) = \varepsilon_1(t)$. Оптимальный размер шага t можно определить из разложения функции $\varepsilon_1(t)$ в ряд Тейлора до членов второго порядка, в окрестности точки $t = 0$:

$$\varepsilon_1(t) = \varepsilon_1(0) + \varepsilon_1'(0)\,t + \frac{\varepsilon_1''(0)\,t^2}{2}. \quad (2.159)$$

Далее значение шага t несложно получить из условия минимума разложения:

$$t = -\frac{\varepsilon_1'(0)}{\varepsilon_1''(0)}. \quad (2.160)$$

Аналитические формулы для производных $\varepsilon_1'(0)$ и $\varepsilon_1''(0)$ легко получить из уравнений (2.153)–(2.155).

Фазу штрихов φ можно рассматривать как дополнительный параметр в алгоритме оптимизации. Согласно уравнениям (2.154) и (2.155), значение $\varphi = \pi$ является оптимальным при расчете решетки без нулевого порядка. При расчете решетки с нулевым порядком, требуемое значение нулевого порядка может быть получено соответствующим выбором φ. В частности, при расчете решетки с равными интенсивностями порядков, определим значение φ в уравнении (2.154) в виде [52, 68]

$$\varphi = 2\arcsin\left(\sqrt{\frac{2N}{E_0 + 8Q(1-Q)N}}\right), \quad E_0 = \sum_{\substack{j=-N\\ j\neq 0}}^{N} I_j. \quad (2.161)$$

Величина E_0 является энергетической эффективностью и соответствует доле энергии, приходящейся на заданные порядки. Подставляя значение φ, определенное согласно уравнению (2.161), в (2.154) получаем, что значение нулевого порядка становится равным среднему значению интенсивности в порядках. С учетом нулевого порядка энергетическая эффективность решетки увеличивается на величину

$$\Delta E = E_0\frac{2N + 1 - E_0 - \gamma N}{E_0 - \gamma N}, \quad \gamma = 8Q(1-Q). \quad (2.162)$$

В таблице 2.1 представлены результаты расчета бинарных решеток для различного числа порядков $M = 2N+1$ равной интенсивности. Среднеквадратичная ошибка, представленная в колонке 4, имеет вид

$$\delta = \frac{1}{\bar{I}}\left[\frac{1}{M}\sum_{j=-N}^{N}\left(I_j - \bar{I}\right)^2\right]^{1/2}, \quad \bar{I} = M^{-1}\sum_{j=-N}^{N} I_j. \quad (2.163)$$

Таблица 2.1. Результаты расчета бинарных дифракционных решеток

Число порядков M	Число штрихов K	Энергетическая эффективность E(%)	Среднеквадратичная ошибка (%)
1	**2**	**3**	**4**
5	2	77,5	0,7
7	2	81,2	1,4
9	3	78,2	2,6
11	4	84,6	0,5
15	4	84,6	0,5
19	5	82,4	1,1
25	7	80,3	3,8
31	8	83,4	3,1
41	11	82,0	4,7
51	14	82,3	3,7
61	16	80,7	2,1
71	19	80,4	1,3
81	22	82,8	3,7
91	25	81,0	2,5
101	27	81,4	2,4
151	40	82,8	4,9
181	44	82,7	3,9
201	54	81,4	4,6
251	69	82,2	4,7
281	74	82,4	4,0
301	81	83,3	3,7
451	118	85,1	3,2

Высокая энергетическая эффективность E и низкая среднеквадратичная ошибка δ, полученные для любых решеток с числом порядков до 451, подтверждают высокую эффективность градиентной процедуры с аналитически рассчитанным начальным профилем (2.150). Разработанный метод можно без ограничений применять для расчета решеток с числом порядков до 1001 или более, оставаясь в пределах нескольких минут времени компьютерного расчета. Решетки с числом порядков $M > 15$ были рассчитаны при $\varphi = \pi$. В процессе расчета было получено, что при $M \leqslant 15$ оказывается более удобным формировать нулевой порядок, выбирая значение φ из уравнения (2.161). Такой выбор φ позволяет на 2–4% увеличить энергетическую эффективность для решеток с числом порядков 5–11. Шаг t для градиентного метода (2.156) рассчитывался из условия минимума квадратичной аппроксимации функции ошибки вдоль сопряженного направления. Такой выбор t значительно увеличил скорость сходимости.

Важной технологической характеристикой решетки является размер минимального штриха, определяемый как минимальное расстояние между соседними координатами. В процессе расчетов было получено, что минимальная ширина штриха $\Delta_{\min} \approx 1/M$, первоначального бинарного профиля (2.150), остается примерно такой же после итерационной коррекции (2.156). Например для решетки с числом порядков 101, отношение $\Delta_{\min}$/период равно 0,01.

На рис. 2.23 представлены рассчитанные профили для 101-порядковых решеток, формирующих постоянное и линейное распределение интенсивности в порядках дифракции. Распределение интенсивности в порядках приведены на рис.2.24. Вследствие симметрии показана только половина дифракционных порядков.

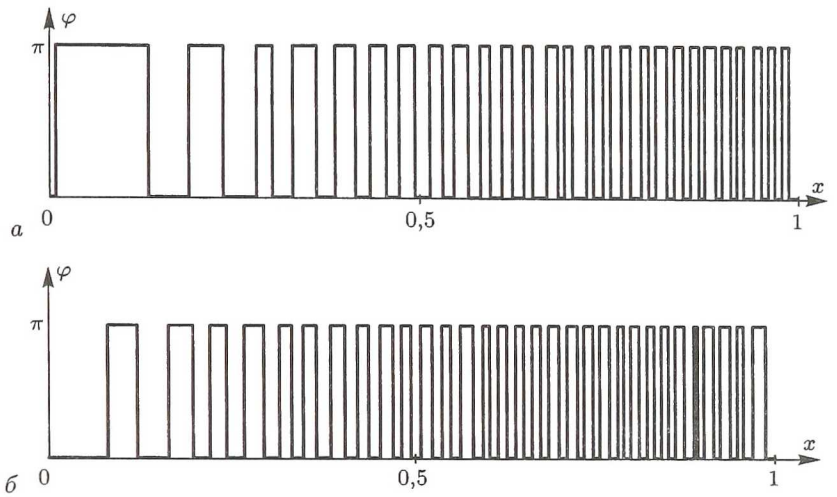


Рис. 2.23. Рассчитанный профиль 101-порядковой решетки, формирующей: постоянное распределение интенсивности (*а*), линейное распределение интенсивности (*б*)

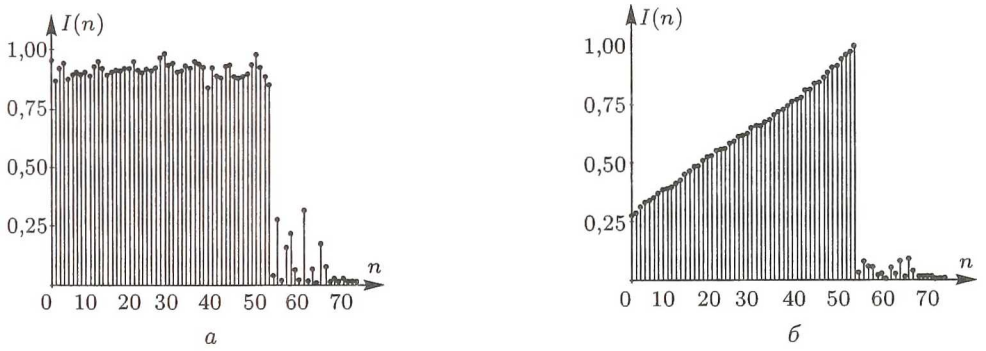


Рис. 2.24. Постоянное распределение интенсивности для 101-порядковой решетки (*а*), линейное распределение интенсивности для 101-порядковой решетки (*б*)

В заключение отметим, что предложенный метод выбора начального профиля бинарной решетки может быть обобщен на двумерный случай. В двумерном случае начальный профиль может быть получен бинаризацией фазовой функции двумерного ДОЭ, фокусирующего в прямоугольную область [71].

2.8.5. Многоуровневые фазовые дифракционные решетки

Согласно (2.154) и (2.155), бинарные дифракционные решетки позволяют сформировать только симметричные распределения интенсивности дифракционных порядков. Дифракционные решетки с непрерывной фазовой функцией позволяют создавать несимметричные и сколь угодно сложные распределения интенсивности в дифракционных порядках. Для расчета дифракционных решеток с непрерывной фазовой функцией φ чаще всего используется алгоритм Герчберга–Секстона (ГС). Для одномерных решеток, рассчитанных методом ГС, энергетическая эффективность составляет более 90%; при среднеквадратичной ошибке формирования

заданной интенсивности $\hat{I}_j$ в дифракционных порядках — 10–20% [59, 60]. Для применений дифракционных решеток в устройствах оптических вычислений и компьютерных сетях на оптических волокнах, среднеквадратичная ошибка должна быть на порядок меньше.

В работе [5] было показано, что алгоритм ГС может быть рассмотрен как градиентный метод минимизации функционала $\varepsilon(\varphi)$, представляющего квадрат ошибки:

$$\varepsilon(\varphi) = \sum_{j=-\infty}^{\infty} \left(\sqrt{I_j} - \sqrt{\hat{I}_j} \right)^2 . \tag{2.164}$$

Интенсивности дифракционных порядков в уравнении (2.164) пропорциональны квадратам модулей коэффициентов Фурье в разложении функции комплексного пропускания решетки $\exp(i\varphi(x))$:

$$I_j = \left| \frac{1}{d} \int_0^d \exp\left(i\varphi(x) - i\frac{2\pi}{d} jx \right) \mathrm{d}x \right|^2 , \tag{2.165}$$

где d — период решетки. В дальнейшем будем считать что $d = 1$.

Для алгоритма ГС характерен эффект стагнации; после нескольких начальных итераций уменьшение ошибки (2.164) происходит очень медленно. Вид итерационного алгоритма и его сходимость существенно зависят от выбора функционала ошибки. Скорость сходимости итерационного процесса можно увеличить, изменив форму функционала (2.164).

В связи с этим рассмотрим градиентный метод расчета фазовой функции $\varphi(x)$ решетки для функционала ошибки общего вида. Пусть $\varepsilon(\varphi)$ — некоторый функционал, характеризующий отличие расчетного распределения интенсивности в дифракционных порядках от требуемого распределения:

$$\varepsilon(\varphi) = \hat{\varepsilon}(\mathbf{I}, \mathbf{I}_d) , \tag{2.166}$$

где $\mathbf{I}$ и $\mathbf{I}_d$ — векторы, компоненты которых соответствуют расчетным и требуемым интенсивностям. Согласно уравнению (2.165), компоненты вектора $\mathbf{I}$ являются функционалами от фазовой функции решетки $\varphi(x)$. Задача состоит в отыскании функции $\hat{\varphi}(x)$, минимизирующей значение функционала $\varepsilon(\varphi)$. В градиентном методе минимизации функционала $\varepsilon(\varphi)$ искомая функция $\hat{\varphi}(x)$ ищется как предел последовательности функций $\varphi_n(x)$, рассчитываемых по правилу

$$\varphi_{n+1}(x) = \varphi_n(x) - t_n \varepsilon'(\varphi_n) , \tag{2.167}$$

где $\varepsilon'(\varphi_n)$ — градиент функционала на n-том шаге процесса оптимизации, t_n — шаг градиентного метода. Для вычисления $\varepsilon'(\varphi)$ в (2.167), рассмотрим приращение функционала $\varepsilon(\varphi)$, вызванное малым изменением фазы $\Delta\varphi$. Согласно (2.166), приращение функционала имеет вид

$$\Delta\varepsilon(\varphi) = \varepsilon(\varphi + \Delta\varphi) - \varepsilon(\varphi) = \sum_j \frac{\partial\hat{\varepsilon}(\mathbf{I}, \mathbf{I}_d)}{\partial I_j} \Delta I_j(\varphi). \tag{2.168}$$

Приращение интенсивности $\Delta I_j(\varphi)$ в (2.168), с точностью до членов второго порядка, может быть представлено в виде

$$\Delta I_j(\varphi) = 2\,\mathrm{Re}\left(F_j^*(\varphi)\, \Delta F_j(\varphi) \right) , \tag{2.169}$$

где

$$F_j(\varphi) = \int_0^1 \exp(i\varphi(x) - i2\pi jx)\,\mathrm{d}x \tag{2.170}$$

— комплексная амплитуда волны в j-м дифракционном порядке. Согласно (2.170), приращение $\Delta F_j(\varphi)$ имеет вид

$$\Delta F_j(\varphi) = \int_0^1 [\exp(i\varphi(x) + i\Delta\varphi(x)) - \exp(i\varphi(x))]\exp(-i2\pi jx)\,\mathrm{d}x \approx$$

$$\approx \int_0^1 i\Delta\varphi(x)\exp(i\varphi(x) - i2\pi jx)\,\mathrm{d}x. \tag{2.171}$$

Подставляя (2.169)–(2.171) в (2.168), получаем приращение функционала в виде

$$\Delta\varepsilon(\varphi) = 2\,\mathrm{Re}\left(\sum_j \frac{\partial\hat{\varepsilon}(\mathbf{I},\mathbf{I}_d)}{\partial I_j} F_j^*(\varphi)\int_0^1 i\Delta\varphi(x)\exp(i\varphi(x) - i2\pi jx)\,\mathrm{d}x\right). \tag{2.172}$$

Согласно уравнению (2.172), градиент функционала $\varepsilon'(\varphi)$ имеет вид

$$\varepsilon'(\varphi) = 2\,\mathrm{Re}\left(i\sum_j \frac{\partial\hat{\varepsilon}(\mathbf{I},\mathbf{I}_d)}{\partial I_j} F_j^*(\varphi)\exp(i\varphi(x) - i2\pi jx)\right). \tag{2.173}$$

Для удобства последующего анализа (2.173) представим в следующем виде:

$$\varepsilon'(\varphi) = 2\,\mathrm{Re}(i\exp(i\varphi(x))\,\Psi^*(x)) = -2\,|\Psi(x)|\sin(\varphi(x) - \arg(\Psi(x))), \tag{2.174}$$

где $\Psi(x) = \int_{-\infty}^{\infty} f(\xi)\exp(i2x\xi)\,d\xi$ — обратное преобразование Фурье от функции

$$f(\xi) = \sum_j \frac{\partial\hat{\varepsilon}(\mathbf{I},\mathbf{I}_d)}{\partial I_j} F_j(\varphi)\,\delta(\xi - j). \tag{2.175}$$

Расчет функции $\Psi(x)$ в (2.174) аналогичен первым трем шагам алгоритма ГС: (1) вычисление прямого преобразования Фурье для расчета значений поля F_j в дифракционных порядках, (2) замена отсчетов поля F_j на функцию $f(\xi)$, (3) вычисление обратного преобразования Фурье от функции $f(\xi)$. Отличие алгоритма ГС от градиентной процедуры состоит в виде замены (2.175). В алгоритме ГС значения F_j заменяются не функцией $f(\xi)$ (2.175), а следующими значениями:

$$\hat{F}_j = \sqrt{\hat{I}_j}\cdot F_j/|F_j|. \tag{2.176}$$

Для расчета шага t градиентного метода, представим функционал $\varepsilon(\varphi)$ вдоль направления противоположного градиенту как функцию от t:

$$\varepsilon(\varphi) = \varepsilon(\varphi - t\varepsilon'(\varphi)) = \varepsilon_1(t). \tag{2.177}$$

Значение шага t можно определить из разложения функции $\varepsilon_1(t)$ в ряд Тейлора до линейных членов в окрестности точки $t = 0$:

$$\varepsilon_1(t) = \varepsilon_1(0) + \varepsilon_1'(0)\, t. \tag{2.178}$$

Предполагая что экстремальное значение функционала $\varepsilon(\varphi)$ равно нулю, определим шаг t из условия равенства нулю разложения (2.178):

$$t_f = -\frac{\varepsilon_1(0)}{\varepsilon_1'(0)}. \tag{2.179}$$

Производную $\varepsilon_1'(0)$ в (2.179) можно найти из (2.172) и (2.173) в виде

$$\varepsilon_1'(0) = -4 \int_0^1 |\Psi(x)|^2 \sin^2(\varphi(x) - \arg(\Psi(x)))\, \mathrm{d}x. \tag{2.180}$$

Предполагая функцию $\varepsilon_1(t)$ квадратичной вдоль сопряженного направления, несложно получить оценку для оптимального значения шага t_s в виде

$$t_s = -\frac{(\varepsilon_1(0))^2}{2\varepsilon_1(t_f)\,\varepsilon_1'(0)}. \tag{2.181}$$

Для расчета шага t_s, необходим пересчет интенсивностей порядков для вычисления $\varepsilon_1(t_f)$. При стабилизации процесса сходимости функционал $\varepsilon(\varphi)$ меняется незначительно от итерации к итерации. Предполагая что $\varepsilon_1(t_f) \approx \varepsilon_1(0)$, получаем оценку для шага t_s в виде

$$t_s = \frac{t_f}{2} = \varepsilon_1(0) \left(8 \int_0^1 |\Psi(x)|^2 \sin^2(\varphi(x) - \arg(\Psi(x)))\, \mathrm{d}x \right)^{-1}. \tag{2.182}$$

Отметим, что полученные выше уравнения для градиента функционала и величины шага градиентного метода являются общими и не зависят от вида $\varepsilon(\varphi)$.

Рассмотрим связь между алгоритмом ГС и градиентной процедурой минимизации функционала (2.164). Минимизация функционала (2.164) эквивалентна максимизации следующего функционала:

$$\hat{\varepsilon}(\varphi) = 2 \sum_j \sqrt{I_j \hat{I}_j}. \tag{2.183}$$

Градиент функционала (2.183) имеет вид (2.174), если функция

$$f(\xi) = \sum_j \sqrt{\hat{I}_j} \frac{F_j(\varphi)}{|F_j(\varphi)|} \delta(\xi - j). \tag{2.184}$$

Функция (2.184) соответствует замене, выполняемой в фурье-плоскости для алгоритма ГС. Соответственно расчет функции $\Psi(x)$ в (2.174) совпадает с первыми тремя шагами алгоритма ГС. Отличие градиентной процедуры (2.167), (2.174), (2.184) от алгоритма ГС состоит в расчете нового приближения фазы $\varphi_{n+1}(x)$ на n-м шаге

итерационного процесса. Для алгоритма ГС новое приближение фазы вычисляется по правилу $\varphi_{n+1}(x) = \arg(\Psi_n(x))$, а для градиентной процедуры — по формуле (2.167). Рассмотрим ситуацию, при которой после выполнения первых итераций градиентный процесс стабилизировался, и функционал ошибки (2.183) стал изменяться медленно. В этом случае для большинства значений x в (2.174) имеем

$$\sin(\varphi_n(x) - \arg(\Psi_n(x))) \approx \varphi_n(x) - \arg(\Psi_n(x)). \tag{2.185}$$

Из уравнений (2.167) и (2.174) следует, что для градиентного метода разность $\varphi_{n+1}(x) - \varphi_n(x)$ приблизительно пропорциональна величине $\varphi_n(x) - \arg(\Psi_n(x))$. Поскольку функция $\arg(\Psi_n(x))$ является новым приближением фазы для алгоритма ГС, то можно заключить, что градиентная процедура (2.167), (2.174) и (2.184) при $t = 1$ и $|\Psi_n(x)| \approx 1$ вычисляет новую фазу $\varphi_{n+1}(x)$, близкую к той, что получается в алгоритме ГС [5].

В отличие от алгоритма ГС, градиентная процедура (2.167), (2.174) допускает выбор различных функционалов $\varepsilon(\varphi)$. Вид функционала $\varepsilon(\varphi)$ оказывает существенное влияние на сходимость итерационного процесса. Целесообразно исследовать сходимость градиентной процедуры для функционала ошибки следующего вида:

$$\varepsilon(\varphi, p) = \sum_j \left(I_j - \hat{I}_j\right)^p. \tag{2.186}$$

В отличие от функционала (2.164), функционал $\varepsilon(\varphi, p)$ (2.186) при $p > 1$ является более чувствительным к отклонению распределения интенсивности в дифракционных порядках от требуемого распределения. Градиент функционала (2.186) описывается уравнением (2.174) при функции $f(\xi)$, определенной в виде

$$f(\xi) = p \sum_j \left(I_j - \hat{I}_j\right)^{p-1} F_j \delta(\xi - j). \tag{2.187}$$

Для выбора оптимального метода расчета многопорядковых дифракционных решеток, проведем сравнение работоспособности следующих итерационных процедур: алгоритма ГС, АА-алгоритма, рассмотренного в пункте 2.4, и градиентного метода (2.167), (2.174) при функционале ошибки (2.187). Напомним, что отличие между алгоритмом ГС и АА-алгоритмом состоит только в форме замены расчетных значений комплексной амплитуды F_j в фурье-плоскости. В алгоритме ГС значения F_j заменяются на значения $\hat{F}_j$ по правилу

$$\hat{F}_j = \begin{cases} \sqrt{\hat{I}_j} \dfrac{F_j}{|F_j|}, & j \in Q, \\ 0, & j \notin Q, \end{cases} \tag{2.188}$$

где $Q = \{n_1, n_2, \ldots, n_M\}$ — индексы заданных дифракционных порядков.

В адаптивно-аддитивном алгоритме (2.37) замена (2.188) принимает вид

$$\hat{F}_j = \begin{cases} \left|\alpha\sqrt{\hat{I}_j} + (1-\alpha)\sqrt{I_j}\right| \dfrac{F_j}{|F_j|}, & j \in Q, \\ F_j, & j \notin Q, \end{cases} \tag{2.189}$$

где α — выбираемый коэффициент.

Будем оценивать работоспособность дифракционных решеток с помощью значений энергетической эффективности $E = \sum_{j \in Q} I_j$ и среднеквадратичной ошибки δ.

В таблице 2.2 приведены значения E и δ для рассчитанных дифракционных решеток с различным числом симметрично расположенных порядков равной интенсивности. Расчет производился с использованием алгоритма быстрого преобразования Фурье при числе элементов дискретизации фазовой функции $n = 64$. Для учета дифракции на растре, требуемые интенсивности порядков $\hat{I}_j$ считались не постоянными, а обратно пропорциональными значениям $f(j) = \mathrm{sinc}^2(\pi j/n)$, $j = -n/2, n/2$ [42]. В качестве начальной фазы для итерационных процедур была использована случайная фаза, равномерно распределенная в интервале $[0, 2\pi)$. В ходе расчета было получено, что для стабилизации ошибки в итерационных процедурах достаточно 80–100 итераций.

Левая часть таблицы 2.2 содержит значения E и δ для решеток, рассчитанных по алгоритму ГС (замена (2.188)). В средней части таблицы даны значения E и δ для решеток, рассчитанных по адаптивно-аддитивному алгоритму (АА-алгоритму). При расчете решеток по АА-алгоритму, первые 15–20 итераций, пока отклонение расчетных интенсивностей порядков от заданных значений уменьшается быстро, выполнялись по алгоритму ГС. После стабилизации ошибки в алгоритме ГС, осуществлялся переход к АА-алгоритму (2.189) при $\alpha = 2$. В правой части таблицы приведены значения E и δ для решеток, рассчитанных по градиентному методу (2.167), (2.174) и (2.175) для функционала (2.186) при $p = 2$. Оптимальный размер шага t_n в (2.167) определялся по формулам (2.181) и (2.182). В процессе расчетов было получено, что в ходе первых 10–20 итераций шаг t_n должен определяться по более точной формуле (2.181). Использование упрощенной формулы (2.182) для величины шага t_n приводит к осцилляции критерия (2.186). Поскольку расчет t_n, по формуле (2.181) требует дополнительного вычисления преобразования Фурье, то для упрощения расчета следует выполнять первые 10–20 итераций по алгоритму ГС. Последующее применение градиентной процедуры приводит к гладким изменениям функционала $\varepsilon(\varphi, 2)$, и использование упрощенной формулы (2.182) для расчета шага t_n обеспечивает стабильную сходимость.

Данные таблицы 2.2 показывают, что решетки, рассчитанные по алгоритму ГС, имеют среднеквадратичную ошибку $\delta = 10 - 15\%$ при энергетической эффективности $E = 90 - 99\%$. АА-алгоритм позволяет уменьшить ошибку δ до 0,1–0,2% при незначительном снижении энергетической эффективности E на 1–3%. Для решеток, рассчитанных по градиентному методу для функционала $\varepsilon(\varphi, 2)$, ошибка δ почти на порядок меньше, чем для решеток, рассчитанных по алгоритму ГС. В процессе расчетов было получено, что сходимость градиентного метода для функционала $\varepsilon(\varphi, p)$ наилучшая при $p = 2$; с ростом p скорость сходимости уменьшается.

Расчет проводился также для решеток с несимметричным расположением порядков. На рис. 2.25 показаны фазовые функции дифракционных решеток, формирующих 11 несимметричных порядков равной интенсивности. Распределения интенсивности в фурье-плоскости для этих решеток представлены на рис. 2.26. Значения δ и E составили 20,4% и 94,3% при расчете по алгоритму ГС, 0,1% и 90,2% при расчете по АА-алгоритму и 0,6% и 91,2% при расчете градиентным методом для функционала $\varepsilon(\varphi, 2)$.

Результаты расчетов (табл. 2.2 и рис. 2.26) показывают, что АА-алгоритм и градиентный метод для функционала $\varepsilon(\varphi, 2)$ работают существенно лучше алгоритма ГС. При этом сходимость АА-алгоритма (2.189) несколько лучше по сравнению

Таблица 2.2. Результаты итеративного расчета многопорядковых дифракционных решеток

Число порядков	Алгоритм ГС		АА-алгоритм		Градиентный метод для функционала (2.186) при $p = 2$	
	$E(\%)$	$\delta(\%)$	$E(\%)$	$\delta(\%)$	$E(\%)$	$\delta(\%)$
5	90,1	14,7	88,5	0,1	91,5	0,5
9	95,6	17,5	93,5	0,2	93,8	0,2
15	97,9	15,5	93,6	0,1	94,0	0,8
19	97,2	14,9	94,1	0,1	94,5	0,7
23	95,1	16,9	93,9	0,1	94,8	0,6
27	98,3	18,7	93,3	0,1	94,4	1,2
33	97,9	14,8	94,5	0,1	96,2	1,8
39	99,2	11,5	94,2	0,2	96,3	1,5
45	99,6	10,8	96,1	0,4	97,5	1,2
51	99,7	9,8	95,9	0,7	98,3	1,8

с градиентным методом. АА-алгоритм также проще в программной реализации (не требуется явного расчета градиента и величины шага на каждой итерации).

2.8.6. Фазовые ДОЭ, фокусирующие в объем и на поверхность тел вращения

Для расчета ДОЭ, фокусирующих в объемные фигуры, используют разбиение объема на N плоскостей и сведение задачи к расчету ДОЭ, формирующего плоские изображения. В [72, 73] предложены различные подходы к решению этой задачи. В [72] независимо рассчитываются амплитудно-фазовые функции $f_n(x, y)$, преобразование Френеля которых связано с заданным распределением интенсивности $I_n(\xi, \eta)$ на плоскостях $z = z_n$ уравнением

$$I_n(\xi, \eta) = \left|\mathcal{F}_n\left\{f_n(x, y)\right\}\right|^2, \tag{2.190}$$

где $\mathcal{F}_n\{\cdot\}$ — преобразование Френеля (2.1) на плоскость $z = z_n$. Это обозначение преобразования Френеля отличается от обозначения преобразования Фурье только нижним индексом n,указывающим номер плоскости (ξ, η), на которой вычисляется комплексная амплитуда волны.

Результирующая амплитудно-фазовая функция пропускания ДОЭ равна среднему арифметическому значению:

$$f(x, y) = \frac{1}{N}\sum_{n=0}^{N} f_n(x, y). \tag{2.191}$$

Для замены амплитудно-фазовой функции $f(x, y)$ на только фазовую функцию используют обычные процедуры кодирования цифровой голографии [74].

В работе [73] функции $f_n(x, y)$ рассчитываются взаимосвязанно, а не независимо, что позволяет в ходе итеративного процесса получить только фазовую функцию без дополнительных (малоэффективных) процедур кодирования.

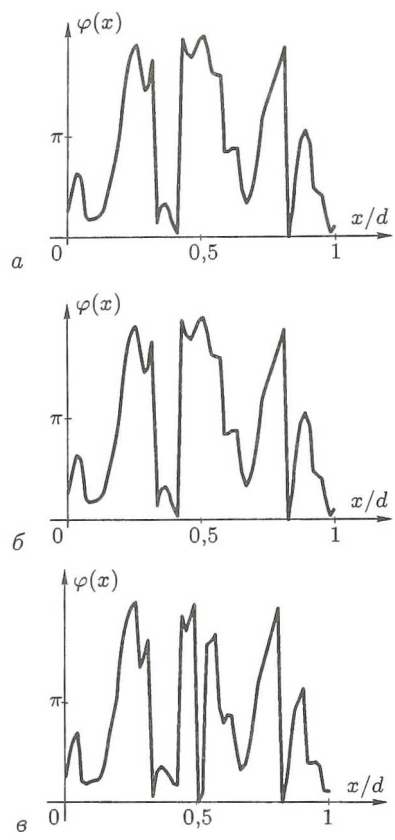


Рис. 2.25. Фазовые функции 11-порядковых решеток, рассчитанных по алгоритму ГС (*а*), по АА-алгоритму (*б*), по градиентному методу (*в*)

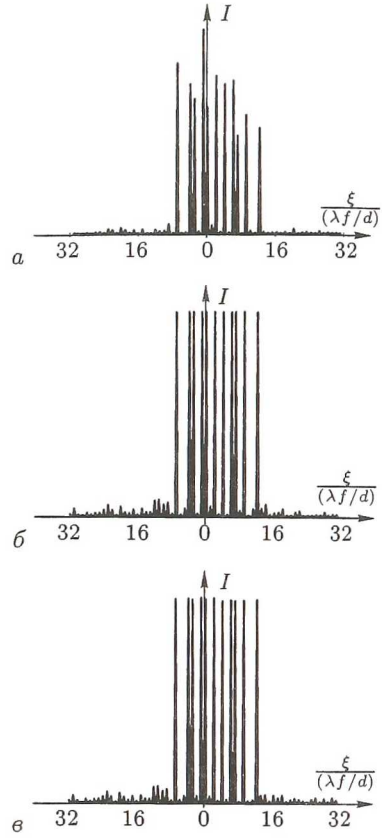


Рис. 2.26. Распределение интенсивности в Фурье-плоскости для 11-порядковых решеток, состоящих из 4 периодов: алгоритм ГС (*а*), АА-алгоритм (*б*), градиентный метод (*в*)

В [75] рассматривалась в приближении геометрической оптики задача расчета ДОЭ, фокусирующего на поверхность вращения с осью z.

В данном разделе рассматриваются итеративные методы расчета фазовых ДОЭ, фокусирующих лазерное излучение в объем и на поверхность тел вращения [76].

Рис. 2.27 показывает, что ДОЭ, который далее рассматривается как тонкий фазовый транспарант, освещается плоской волной и формирует на расстояниях вдоль оптической оси z_n, $n = \overline{1,N}$, заданные распределения интенсивности $I_n(\xi,\eta)$. Рассмотрение проводится в рамках скалярной волновой теории дифракции Кирхгофа в приближении Френеля.

В [73] описан алгоритм расчета функции пропускания ДОЭ с помощью итеративной процедуры минимизирующей функционал

$$\varepsilon = \sum_{n=1}^{N} \iint_{-\infty}^{\infty} [|F(\xi,\eta,z_n)| - A_n(\xi,\eta)]^2 \, \mathrm{d}\xi \, \mathrm{d}\eta, \tag{2.192}$$

где $A_n(\xi,\eta) = \sqrt{I_n(\xi,\eta)}$, $n = \overline{1,N}$,

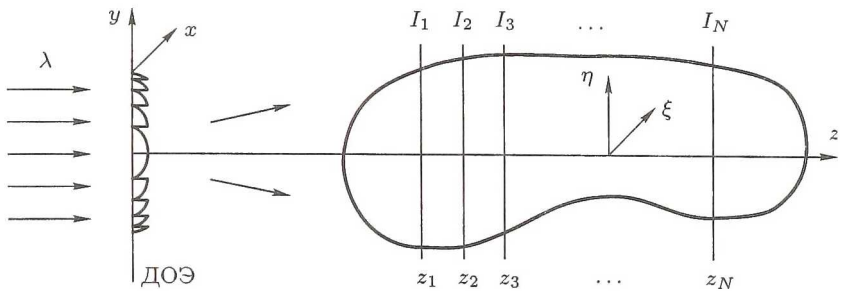


Рис. 2.27. Схема для фокусировки с помощью ДОЭ в 3-х мерную область

$$F(\xi,\eta,z_n) = \mathcal{F}_n\left\{e^{i\varphi(x,y)}\right\} = \\ = \frac{k}{z_n}\iint\limits_{\Omega}\exp\left[i\varphi(x,y) + \frac{ik}{2z_n}\left[(x-\xi)^2+(y-\eta)^2\right]\right]\mathrm{d}x\,\mathrm{d}y, \quad (2.193)$$

$\varphi(x,y)$ — искомая фаза ДОЭ, формирующего световое поле с комплексной амплитудой $F(\xi,\eta,z)$, модуль которой совпадает на заданных расстояниях z_n с требуемой амплитудой $A_n(\xi,\eta,z_n)$, Ω — площадь апертуры ДОЭ. Итеративный алгоритм, минимизирующий среднеквадратичное отклонение (2.192), имеет следующий вид:

$$\varphi_p(x,y) = \arg\left[\sum_{n=1}^{N} U_n^{(p)}(x,y)\right], \quad (2.194)$$

$$U_n^{(p)}(x,y) = \mathcal{F}_n^{-1}\left\{A_n(\xi,\eta)\exp\left[iQ_{n,p}(\xi,\eta)\right]\right\}, \quad (2.195)$$

$$Q_{n,p}(\xi,\eta) = \arg\left[\mathcal{F}_n\left\{\exp\left[i\varphi_{p-1}(x,y)\right]\right\}\right], \quad (2.196)$$

где p — номер итерации, Q_n, $p(\xi,\eta)$ — фаза волны на плоскости, отстоящей от ДОЭ на расстоянии z_n и рассчитанная на p-й итерации.

Ниже рассматриваются различные варианты итеративного алгоритма, минимизирующего функционал среднеквадратичного отклонения вида

$$\varepsilon_1 = \iint\limits_{\Omega}\left|e^{i\varphi(x,y)} - \sum_{n=1}^{N} C_n U_n(x,y)\right|^2 \mathrm{d}x\,\mathrm{d}y, \quad (2.197)$$

где C_n — комплексные весовые множители.

Далее, в зависимости от того, куда требуется сфокусировать излучение, будет меняться вид функций U_n. Этим и будут отличаться рассматриваемые задачи.

Функционал (2.192), записанный для плоскостей, отстоящих от ДОЭ на расстояниях z_n, можно с помощью равенства Парсеваля записать в виде

$$\varepsilon = \sum_{n=1}^{N}\iint\limits_{\Omega}\left|e^{i\varphi(x,y)} - U_n(x,y)\right|^2 \mathrm{d}x\,\mathrm{d}y. \quad (2.198)$$

Функции $U_n(x,y)$ в уравнениях (2.197) и (2.198) являются результатом вычисления обратного преобразования Френеля:

$$U_n(x,y) = \mathcal{F}_n^{-1}\left\{F_n(\xi,\eta)\right\}, \quad (2.199)$$

$$F_n(\xi,\eta) = A_n(\xi,\eta)\exp\left[iQ_n(\xi,\eta)\right]. \tag{2.200}$$

Если переписать функционал (2.198) в обобщенной форме,

$$\hat{\varepsilon} = \sum_{n=1}^{N} \iint\limits_{\Omega} \left| e^{i\varphi(x,y)} - C_n U_n(x,y) \right|^2 \mathrm{d}x\,\mathrm{d}y, \tag{2.201}$$

то нетрудно получить уравнение, связывающее значения обоих функционалов:

$$\varepsilon_1 = \hat{\varepsilon} + \sum_{m\neq n}^{N}\sum^{N} C_m C_n^* \iint\limits_{\Omega} U_n(x,y)U_m^*(x,y)\,\mathrm{d}x\,\mathrm{d}y - (N-1)\cdot W_0, \tag{2.202}$$

где

$$W_0 = \iint\limits_{\Omega} \left| e^{i\varphi(x,y)} \right|^2 \mathrm{d}x\,\mathrm{d}y \tag{2.203}$$

— полная энергия света в плоскости ДОЭ, Ω — площадь апертуры ДОЭ; функционалы ε_1 и $\hat{\varepsilon}$ равны с точностью до несущественной постоянной $(N-1)\cdot W_0$. Функционал общего вида (2.197) имеет произвол в выборе функций $Q_n(\xi,\eta)$.

Из уравнения (2.202) видно, что при ортогональности функций $U_n(x,y)$,

$$\iint\limits_{\Omega} U_n(x,y)U_m^*(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int\limits_{-\infty}^{\infty}\iint F_n(\xi,\eta)F_m^*(\xi,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta = W_0\delta_{mn}, \tag{2.204}$$

где δ_{mn} — символ Кронекера. Условию ортогональности (2.204) можно наверняка удовлетворить, задавая амплитудные функции $A_n(\xi,\eta)$ на плоскостях фокусировки в зоне дифракции Френеля с пространственно разделенными носителями, то есть области G_n, в которых амплитуды $A_n(\xi,\eta)$ отличные от нуля, не пересекаются между собой:

$$G_n \cap G_m = \varnothing, \quad m,n = \overline{1,N}. \tag{2.205}$$

Такое условие реализуется при расчете фазовых ДОЭ, формирующих моды Гаусса–Эрмита или Гаусса–Лагерра в различных порядках дифракции [77, 78].

Условию (2.204) можно также удовлетворить, разделив апертуру ДОЭ на N непересекающихся субапертур Ω_n, в каждой из которых задается функция $U_n(x,y)$:

$$\Omega_n \cap \Omega_m = \varnothing, \quad m,n = \overline{1,N}. \tag{2.206}$$

Разбиение апертуры ДОЭ, например, на кольцевые субапертуры применяется при расчетах аксиконов и формирователей бесселевых мод [79].

Минимизация функционала (2.197)эквивалентна поиску коэффициентов C_n, обеспечивающих выполнение следующего равенства:

$$e^{i\varphi(x,y)} = \sum_{n=1}^{N} C_n U_n(x,y), \tag{2.207}$$

где функции $U_n(x,y)$ определяются уравнениями (2.199), (2.200) и в общем случае неортогональные. Уравнение (2.207) можно рассматривать как двумерную проекцию трехмерной задачи, где функции $\psi_n(z)$ ортогональны:

$$e^{i\varphi(x,y,z)} = \sum_{n=1}^{N} C_n U_n(x,y)\psi_n(z), \tag{2.208}$$

$$\int_0^1 \psi_n(z)\psi_m^*(z)dz = \delta_{mn}. \tag{2.209}$$

Если, например, функции (2.209) выбрать в виде фурье-гармоник $\psi_n(z) = \exp[i2\pi nz]$, то коэффициенты суммы (2.208) вычисляются с помощью соотношений

$$C_n = W_0^{-1} \iint_\Omega L_n(x,y)U_n^*(x,y)\,\mathrm{d}x\,\mathrm{d}y, \tag{2.210}$$

$$L_n(x,y) = \int_0^1 \exp[i\varphi(x,y,z) - i2\pi nz]\,\mathrm{d}z. \tag{2.211}$$

С учетом уравнений (2.207)–(2.211) получим итеративный алгоритм минимизации функционала (2.197) [76]:

$$\varphi_p(x,y,z) = \arg\left[\sum_{n=1}^{N} B_n e^{iD_{n,p}} U_n(x,y)e^{i2\pi nz}\right], \tag{2.212}$$

$$D_{n,p} = \arg\left[\iint_\Omega L_{n,p}(x,y)U_n^*(x,y)\,\mathrm{d}x\,\mathrm{d}y\right], \tag{2.213}$$

$$L_{n,p}(x,y) = \int_0^1 \exp[i\varphi_{p-1}(x,y,z) - i2\pi nz]\,\mathrm{d}z, \tag{2.214}$$

где $B \geqslant 0$ — произвольные положительные числа, p — номер итерации. Искомая фазовая функция ДОЭ

$$\varphi_p(x,y) = \varphi_p(x,y,z=0). \tag{2.215}$$

Заметим, что система функций $U_n(x,y)\,\psi(z)$ не будет полной, и поэтому не всякую функцию $e^{i\varphi(x,y,z)}$ можно разложить в ряд по этой системе функций. Но так как алгоритм (2.212) — (2.214) используется для численного расчета, и число N всегда конечное, то свойство полноты базисных функций не используется.

Далее ограничим рассмотрение телами, радиально-симметричными относительно оси z. На рис. 2.28 показана оптическая схема для расчета ДОЭ, фокусирующего излучение на ряд плоскостей, являющихся сечениями тела вращения. При этом амплитуда $A_n(\xi,\eta)$ на каждой из плоскостей зависит (в полярных координатах $\xi = \rho\cdot\cos\theta$, $\eta = \rho\cdot\sin\theta$) только от радиальной переменной ρ.

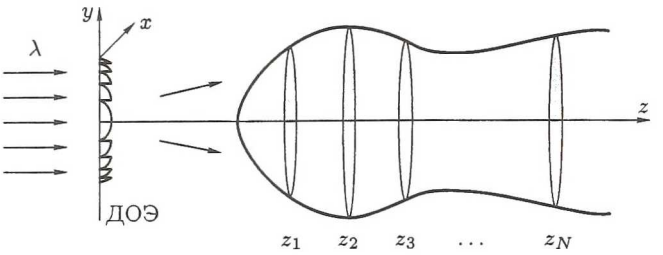


Рис. 2.28. Фокусировка в тело вращения с помощью ДОЭ

Вместо уравнений (2.208) и (2.209) можно записать, что

$$e^{i\varphi(r,\varphi)} = \sum_{n=1}^{N} C_n U_n(r) e^{in\varphi}, \tag{2.216}$$

$$C_n = (2W_0)^{-1} \int_0^R \int_0^{2\pi} e^{i\varphi(r,\varphi)} U_n^*(r) e^{in\varphi} r \,\mathrm{d}r \,\mathrm{d}\varphi, \tag{2.217}$$

$$U_n(r) = \mathcal{F}_u^{-1}\left\{A_n(\rho)e^{iQ_n(\rho)}\right\} = \\ = \frac{k}{z_n} \exp\left[-i\frac{kr^2}{z_n}\right] \int_0^\infty A_n(\rho) \exp\left[iQ_n(\rho) - i\frac{k\rho^2}{2z_n}\right] J_0\left(\frac{kr\rho}{z_n}\right)\rho \,\mathrm{d}\rho, \tag{2.218}$$

где (r, φ) — полярные координаты в плоскости ДОЭ, $J_0(x)$ — функция Бесселя нулевого порядка. Уравнение (2.218) является обратным преобразованием.

Если интенсивность внутри сечений тела вращения постоянная, то амплитуду $A_n(\rho)$ можно выбрать в виде

$$A_n(\rho) = \sqrt{I_n}\,\mathrm{circ}\left(\frac{\rho}{\rho_0}\right), \quad \mathrm{circ}\left(\frac{\rho}{\rho_0}\right) = \begin{cases} 1, & \rho \leqslant \rho_0, \\ 0, & \rho > \rho_0, \end{cases} \tag{2.219}$$

где I_n — постоянные значения интенсивности в круге на каждой плоскости.

Произвольные фазы $Q_n(\rho)$ в (2.218) можно выбрать квадратичными,

$$Q_n(\rho) = \frac{k\rho^2}{2z_n}, \tag{2.220}$$

т.к. параболическая волна является собственной функцией преобразования Френеля, а постоянные интенсивности I_n выбираются из условия сохранения энергии:

$$\rho_n^2 I_n = W_0, \quad n = \overline{1, N}. \tag{2.221}$$

Подставив выражения (2.219)–(2.220) в уравнение (2.218), получим конкретный вид функций разложения:

$$U_n(r) = 2\pi\rho_n\sqrt{I_n} r^{-1} J_1\left(\frac{kr\rho_n}{z_n}\right) \exp\left[-i\frac{kr^2}{2z_n}\right], \tag{2.222}$$

где $J_1(x)$ — функция Бесселя первого порядка.

Искомую функцию фазы ДОЭ $\varphi(r,\theta)$, фокусирующего в набор кругов, расположенных вдоль оптической оси на требуемых расстояниях z_n, с постоянными интенсивностями, с заданными радиусами, можно искать в виде суммы с не полностью определенными коэффициентами:

$$e^{i\varphi(r,\theta)} = r^{-1}\sum_{n=1}^{N} C_n J_1(\alpha_n r)\exp\left[-i\beta_n r^2 + in\theta\right], \tag{2.223}$$

где C_n — постоянные, модуль которых задается с учетом выбора интенсивностей I_n, а фаза является свободным параметром, $\alpha_n = k\rho_n/z_n$, $\beta_n = k/2z_n$.

Из уравнения (2.223) следует выражение для расчета коэффициентов:

$$C_n = W^{-1}\int_0^R\int_0^{2\pi} e^{i\varphi(r,\theta)} J_1(\alpha_n r)\exp\left[i\beta_n r^2 - in\theta\right] r^2\,\mathrm{d}r\,\mathrm{d}\theta, \tag{2.224}$$

$$W = 2\pi\int_0^R J_1^2(\alpha_n r) r\,\mathrm{d}r. \tag{2.225}$$

Уравнения (2.223)–(2.227) позволяют получить итеративный алгоритм расчета функции фазы ДОЭ, аналогичный алгоритму (2.212)–(2.214). Эти два алгоритма сходны по своей структуре и различаются только базисными функциями (2.222).

Далее рассмотрим частный случай этого алгоритма для расчета ДОЭ для фокусировки в набор осевых точек.

Нетрудно получить разложение аналогичное (2.223) для расчета многофокусных дифракционных линз, которые формируют вдоль оптической оси z заданное число фокусов на требуемых расстояниях и с заданной интенсивностью. Другие подходы к расчету ДОЭ типа многофокусных линз рассмотрены в [80, 81]. ДОЭ, рассчитанные методом (2.223), обладают свойствами голограмм. Например, малый участок такого ДОЭ способен восстановить требуемое изображение с минимальными искажениями, при прочих равных условиях.

В данном случае вместо уравнения (2.219) для фокусировки в набор кругов следует использовать уравнение для фокусировки в набор дельта-импульсов:

$$A_n(\rho) = \sqrt{I_n}\delta(\rho). \tag{2.226}$$

Тогда вместо уравнений (2.213) и (2.223) получим

$$U_n(r) = \frac{k}{z_n}\sqrt{I_n}\exp\left[-i\frac{kr^2}{2z_n}\right], \tag{2.227}$$

$$e^{i\varphi(r,\theta)} = \sum_{n=1}^{N} C_n\exp\left[-i\beta_n r^2 + in\theta\right], \quad \beta_n = k/2z_n. \tag{2.228}$$

Коэффициенты суммы (2.228) вычисляются по формуле

$$C_n = \left(\pi R^2\right)^{-1}\int_0^R\int_0^{2\pi} e^{i\varphi(r,\theta)}\exp\left[i\frac{kr^2}{2z_n} - in\theta\right] r\,\mathrm{d}r\,\mathrm{d}\theta. \tag{2.229}$$

Из уравнения (2.229) видно, что коэффициенты C_n вычисляются с помощью двумерного преобразования Фурье.

Заметим, что вместо суммы (2.228) для итеративного расчета многофокусных линз можно использовать более простое радиальное уравнение:

$$e^{i\varphi(r)} = \sum_{n=1}^{N} C_n \exp\left[-i\beta_n r^2\right] \tag{2.230}$$

Однако гауссовы экспоненты в уравнении (2.230) будут ортогональны только при условии

$$\int_0^R \exp\left[-i\left(\beta_m - \beta_n\right) r^2\right] r \, \mathrm{d}r = \delta_{mn}, \tag{2.231}$$

которое выполняется при

$$z_n = n^{-1} z_0, \quad z_0 = R^2 / 2\lambda. \tag{2.232}$$

Из (2.232) следует, что разложением (2.230) удобно пользоваться для вычисления фазы ДОЭ, если требуемые фокусы расположены только на определенных расстояниях z_n .

Ниже получен алгоритм расчета ДОЭ, фокусирующих лазерное излучение на поверхность тел вращения. Расчет данного ДОЭ, ось которого совпадает с оптической осью, был осуществлен методом геометрической оптики в [75]. Если тело вращения с осью z представить как набор поперечных сечений, то его поверхность аппроксимируется набором колец. Поэтому для расчета ДОЭ, формирующего набор световых колец с заданными радиусами ρ_n и расположенных на требуемых расстояниях z_n, можно получить уравнение аналогичное уравнению (2.223). Для этого вместо (2.219) для требуемой амплитуды на n-й плоскости запишем

$$A_n(\rho, \psi) = \sqrt{I_n} \cdot \delta(\rho - \rho_n) \cdot e^{in\psi}. \tag{2.233}$$

Подставив (2.233) в уравнение (2.218), получим, вместо соотношения (2.222), следующее выражение для комплексной амплитуды в плоскости ДОЭ:

$$U_n(r, \theta) = \frac{(-i)^{n_k}}{z_n} \sqrt{I_n} J_n\left(\frac{kr\rho_n}{z_n}\right) \exp\left[-i\frac{kr^2}{2z_n} + in\theta\right]. \tag{2.234}$$

Вместо уравнения (2.223), получим следующее уравнение:

$$e^{i\varphi(r,\theta)} = \sum_{n=1}^{N} C_n J_n(\alpha_n r) \exp\left[-i\beta_n r^2 + in\theta\right], \tag{2.235}$$

коэффициенты которого находятся по формулам, аналогичным (2.224) и (2.225):

$$C_n = W_n^{-1} \int_0^R \int_0^{2\pi} e^{i\varphi(r,\theta)} J_n(\alpha_n r) \exp\left[i\beta_n r^2 - in\theta\right] \cdot r \, \mathrm{d}r d\theta, \tag{2.236}$$

$$W_n = 2\pi \int_0^R J_n^2(\alpha_n r) r \, \mathrm{d}r,$$

где $J_n(x)$ — функция Бесселя n-го порядка.

С целью ускорения расчетов, вместо разложения (2.235), можно использовать следующее уравнение:

$$e^{i\varphi(r)} = \sum_{n=1}^{N} C_n J_0(\alpha_n r) \exp\left[-i\beta_n r^2\right]. \tag{2.237}$$

Однако при этом слагаемые в уравнении (2.237) в общем случае не ортогональны между собой. Чтобы найти условие для достижения ортогональности, можно использовать следующий справочный интеграл [82]:

$$\int_0^\infty e^{iax^2} J_\nu(bx) J_\nu(cx) x dx = \frac{i}{2a} J_\nu\left(\frac{bc}{2a}\right) e^{ih}, \quad h = \frac{b^2+c^2}{4a} - \frac{\nu\pi}{2}. \tag{2.238}$$

Для функций, входящих в уравнение (2.237), вместо (2.238) получим

$$\int_0^\infty J_0(\alpha_n r) J_0(\alpha_m r) \exp\left[-i\left(\beta_n - \beta_m\right) r^2\right] r \, \mathrm{d}r =$$

$$= \frac{-i}{2}\left(\beta_n - \beta_m\right)^{-1} J_0\left(\frac{\alpha_n \alpha_m}{2\left(\beta_n - \beta_m\right)}\right) \exp\left[-i\frac{\alpha_n^2 + \alpha_m^2}{4\left(\beta_n - \beta_m\right)}\right], \quad \beta_n > \beta_m. \tag{2.239}$$

Из (2.239) следует, что если параметры α_n и β_n удовлетворяют условию

$$\frac{\alpha_n \alpha_m}{2\left(\beta_n - \beta_m\right)} = \gamma_p; \quad n, m = \overline{1, N}, \quad p = \overline{1, N^2}, \tag{2.240}$$

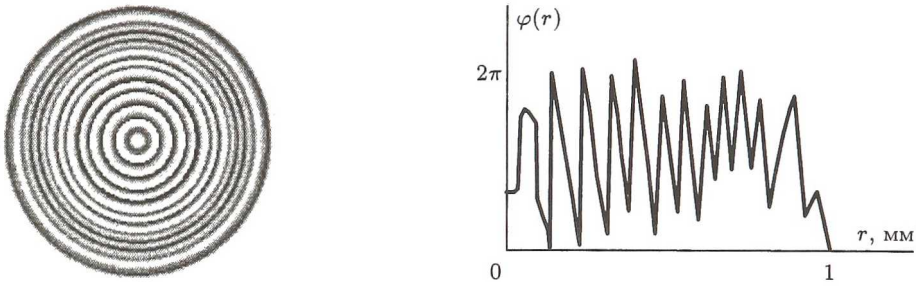
где γ_p — корни функции Бесселя ($J_0(\gamma_p) = 0$), то слагаемые в сумме (2.237) будут ортогональны и коэффициенты можно рассчитывать по формулам

$$C_n = W_n^{-1} \int_0^R e^{i\varphi(r)} J_0(\alpha_n r) \exp\left[i\beta_n r^2\right] r \, \mathrm{d}r, \tag{2.241}$$

$$W_n = 2\pi \int_0^R J_0^2(\alpha_n r) r \, \mathrm{d}r. \tag{2.242}$$

Условие (2.240) ограничивает выбор значений расстояний z_n и радиусов колец ρ_n. Однако, как показал численный эксперимент, можно получить хорошие результаты, считая слагаемые в уравнении (2.237) ортогональными и используя для расчета коэффициентов формулу (2.241).

Пример 2.9. Численные результаты, приведенные ниже, относятся к случаю фокусировки на поверхность вращения. Фаза ДОЭ при этом рассчитывается с помощью итеративного алгоритма, основанного на уравнениях (2.237) и (2.241). Расчетные параметры: $R = 1$ мм — радиус ДОЭ, $k = 10^4$ мм$^{-1}$, $n = 256$ — число отсчетов по радиальной переменной.



а *б*

Рис. 2.29. Фаза (*а*) и ее радиальное сечение (*б*) для ДОЭ, фокусирующего на поверхность расходящегося конуса

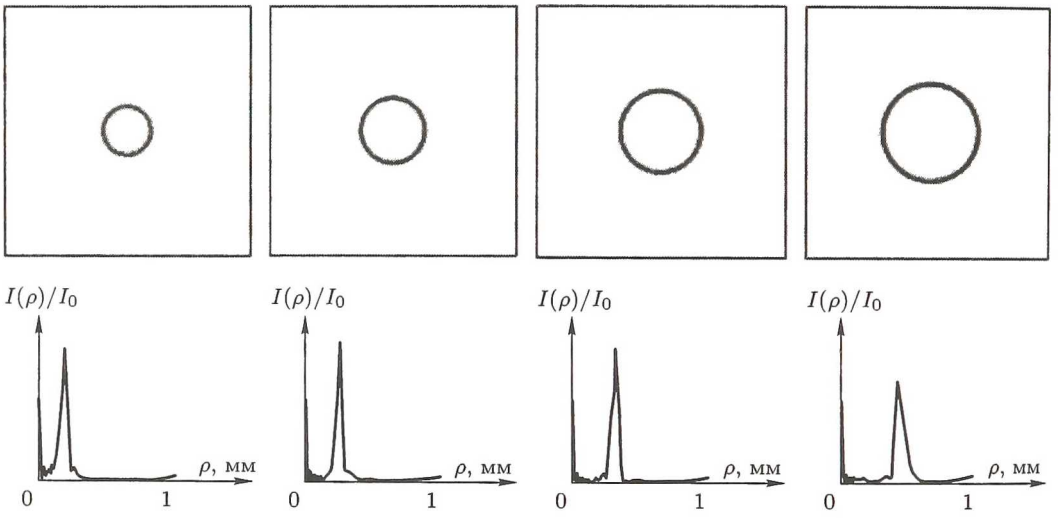


Рис. 2.30. Распределения интенсивности (верхняя строка) и их радиальные сечения (нижняя строка), сформированные ДОЭ с фазой, показанной на рис. 2.29

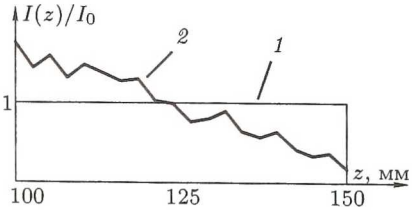


Рис. 2.31. Распределение нормированной интенсивности на поверхности расходящегося конуса

Коническая поверхность описывалась уравнением $\rho_n = \alpha\,(z_n - z_0) + \rho_0$, где $\alpha = \pm 5 \times 10^{-3}$, $z_0 = 100$ мм, $\Delta_z = 5$ мм — расстояние между плоскостями сечений конуса, $n = \overline{1,10}$. На рис. 2.29 показана полутоновая по уровню 2π фаза ДОЭ и ее радиальное сечение, рассчитанные за 10 итераций алгоритмом на основе уравнений (2.237), (2.241).

На рис. 2.30 показаны распределения интенсивности (верхняя строка) и их сечения (нижняя строка), сформированные ДОЭ с фазой, показанной на рис. 2.29, и рассчитанные на разных плоскостях вдоль оси z в диапазоне [100 мм, 140 мм] с шагом 10 мм (z растет слева направо).

На рис. 2.31 показана зависимость нормированной интенсивности на поверхности конуса (рис. 2.30, $\alpha > 0$) вдоль оси z: кривая *1* — заданное распределение интенсивности, кривая *2* — рассчитанная.

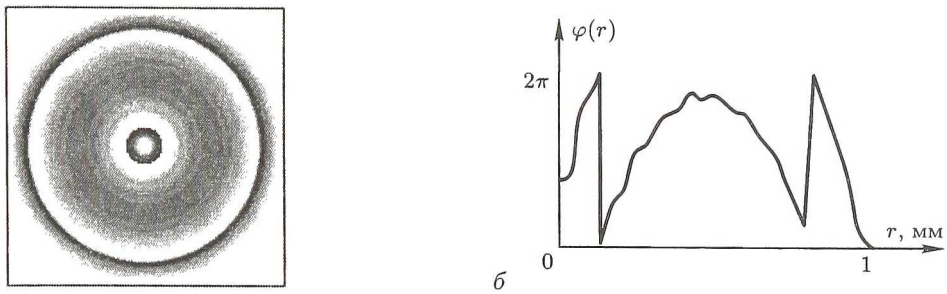


Рис. 2.32. Фаза (а) и ее радиальное сечение (б) для ДОЭ, фокусирующего на поверхность цилиндра

Пример 2.10. На рис. 2.32–2.34 приведены результаты численного эксперимента для фокусировки в цилиндр радиусом $\rho_0 = 0{,}5$ мм. На рис. 2.32 показана фаза ДОЭ и ее сечение, рассчитанные за 10 итераций по формулам (2.237) и (2.241) при $\alpha = 0$.

На рис. 2.33 даны распределения интенсивности и их сечения, рассчитанные на тех же плоскостях, что и для рис. 2.30. Расчет проводился с помощью интегрального преобразования Френеля, которое, в свою очередь, рассчитывалось с помощью алгоритма быстрого преобразования Фурье. На рис. 2.34 показана зависимость относительной интенсивности света на поверхности цилиндра от расстояния z до ДОЭ.

2.8.7. Фокусировка гауссова пучка в квадрат

Выравнивание плотности энергии в сечении когерентных световых пучков является важной задачей для применений лазеров. Для этой цели используются ДОЭ, преобразующие волновой фронт таким образом, чтобы в заданной плоскости сформировалось требуемое распределение интенсивности.

В работе [83] было предложено использовать бинарную дифракционную решетку для перераспределения энергии в пучке. Использование геометрооптического приближения для расчета ДОЭ, выравнивающего поперечное распределение интенсивности коллимированного гауссова пучка предложено в [84]. С помощью киноформной линзы в [85] экспериментально осуществлено выравнивание интенсивности гауссова пучка в дальней зоне дифракции. Итеративный подход для расчета фазового ДОЭ, преобразующего гауссово распределение интенсивности пучка в супергауссово в фокусе линзы предложен в [86]. При этом численно было показано, что предложенный алгоритм эффективно работает, если размер области фокусировки больше нескольких (4–5) минимальных дифракционных пятен.

Геометрооптическому расчету голографического фильтра для выравнивания профиля интенсивности посвящена работа [87]. Недостаток таких фильтров в низкой энергетической эффективности. В работе [88] голографический фильтр рассчитывался итеративным методом в приближении скалярной дифракции. Такой подход позволил достичь более высокой точности формирования требуемого распределения интенсивности. Численный эксперимент с фокусатором, преобразующим гауссов пучок в квадрат с постоянной интенсивностью проводился в [89]. В этой работе было, в частности, показано, что геометрооптический метод для расчета ДОЭ применим, если размер области фокусировки больше семи дифракционных пятен.

В [90] экспериментально с помощью двух голографических фильтров, осуществленно преобразование эллиптического гауссова пучка в ближней зоне дифракции в квадрат с постоянной интенсивностью и фазой. В этой же работе, и независимо

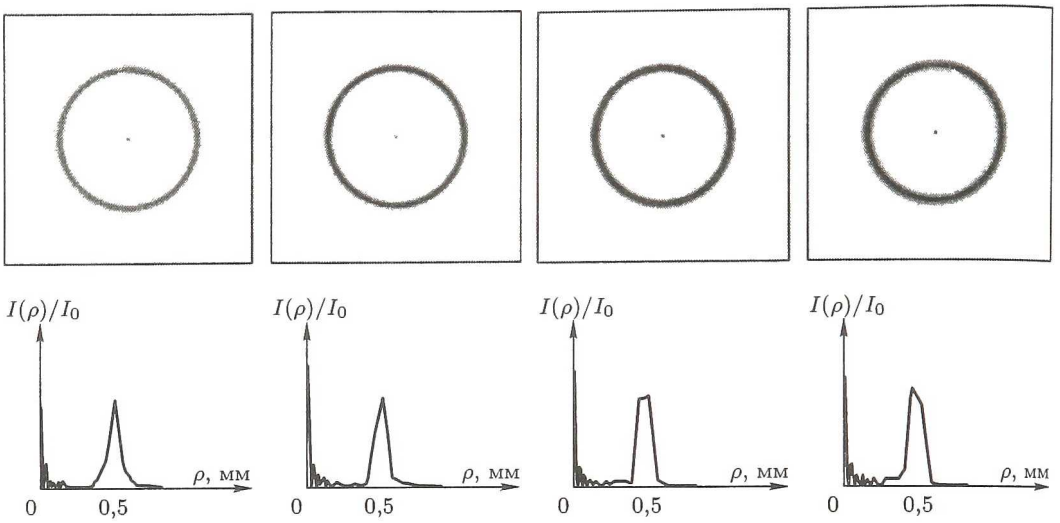


Рис. 2.33. Распределения интенсивности (верхняя строка) и радиальные сечения (нижняя строка), сформированные ДОЭ с фазой, показанной на рис. 2.32

в работе [91], получены аналитические формулы для фазы ДОЭ, фокусирующего гауссов пучок в малый квадрат с постоянной интенсивностью. В [92] приведен сравнительный численный эксперимент по расчету ДОЭ двумя способами: геометрооптическим и итеративным в приближении Френеля. Показано, что наилучшего результата по выравниванию гауссова пучка можно достичь, применяя геометрооптическое решение [90, 91] как начальное приближение для итеративного процесса.

Выравнивание профиля интенсивности пучка полупроводникового лазера с помощью бинарного фазового ДОЭ продемонстрировано в [93]. В [94] также с помощью бинарного отражающего ДОЭ с наклонным падением экспериментально показано выравнивание пучка гелий-неонового лазера в плоскости дефокусировки. Экспериментальные исследования по фокусировке гауссова пучка CO_2 лазера в прямоугольник с равномерной интенсивностью с помощью бинарного и многоградационного ДОЭ проведены, соответственно, в [95] и [96].

Во всех перечисленных работах, за исключением [85], размер сфокусированного с помощью оптических элементов прямоугольного или круглого равномерного распределения интенсивности составлял несколько минимальных дифракционных пятен. То есть полной фокусировки гауссова пучка не происходило. В [97] коротко было замечено, что выравнивать гауссов пучок в фокальной плоскости в области минимального дифракционного пятна можно с помощью бинарной фазовой пластинки с цилиндрической ступенькой.

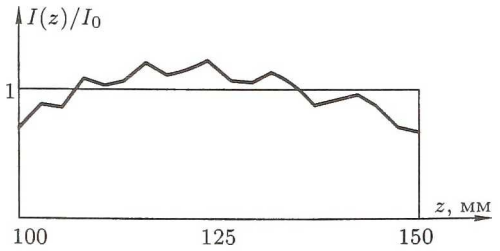


Рис. 2.34. Распределение относительной интенсивности на поверхности цилиндра

В работах [89, 90] было показано, что для ДОЭ, формирующего однородное распределение интенсивности прямоугольной формы, задаваемое соотношением

$$I(x,y)=\begin{cases}1, & |x|\leqslant d_1, \quad |y|\leqslant d_2,\\ 0, & |x|>d_1, \quad |y|>d_2,\end{cases} \tag{2.243}$$

из плоского пучка с гауссовым профилем распределения интенсивности,

$$I_0(u,v)=I_0\exp\left[-\frac{u^2+v^2}{\sigma^2}\right], \tag{2.244}$$

можно получить его фазовую функцию (при условии, что имеется линза, расположенная вплотную к ДОЭ) в виде суммы фаз $\varphi(u,v)=\varphi_1(u)+\varphi_2(v)$, слагаемые которой удовлетворяют уравнениям

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}u}=\dfrac{I_0(u)}{I(x)},\\ x=u+fk^{-1}\dfrac{\mathrm{d}\varphi(u)}{\mathrm{d}u},\end{cases} \tag{2.245}$$

где f — фокусное расстояние линзы, в фокальной плоскости которой формируется световой прямоугольник, k — волновое число, (u,v), (x,y) — координаты в плоскостях ДОЭ и фурье-спектра соответственно, $2d_1$, $2d_2$ — размеры прямоугольника, σ — параметр гауссова пучка. Первое уравнение в системе (2.245) устанавливает равенство плотностей световой энергии одномерного ДОЭ и энергии соответствующих участков прямолинейного отрезка. Второе уравнение описывает стационарные точки в параксиальном приближении. Система (2.245) подробно обсуждалась в главе 1. Дискретный вариант решения (2.245) можно записать в виде

$$\varphi_{mn}=\ln 10\left[M_x\left\{\frac{6n}{N}\operatorname{erf}\left(\frac{6n}{N}\right)+\frac{1}{\sqrt{\pi}}e^{-36n^2/N^2}\right\}+\right.$$
$$\left.+M_y\left\{\frac{6m}{N}\operatorname{erf}\left(\frac{6m}{N}\right)+\frac{1}{\sqrt{\pi}}e^{-36m^2/N^2}\right\}\right], \tag{2.246}$$

где

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t, \quad m=0,\pm1,\pm2,\ldots,\pm\frac{N}{2}, \quad n=0,\pm1,\pm2,\ldots,\pm\frac{N}{2},$$
$$M_x=\frac{\pi N_x}{6\sqrt{\ln 10}}, \quad M_y=\frac{\pi N_y}{6\sqrt{\ln 10}}, \tag{2.247}$$

M_x и M_y — количество минимальных дифракционных пятен, попадающих в прямоугольник размерами $N_x\times N_y$. В этом случае квадратная апертура ДОЭ размером $N\times N$ имеет сторону квадрата длиной 6σ: $N\Delta=6\sigma$,где Δ — расстояние между соседними пикселами в плоскости ДОЭ, а коллимированный гауссов пучок, освещающий ДОЭ, создает распределение света, задаваемое соотношением

$$I_{0mn}=\exp[-36N^{-2}(n^2+m^2)]. \tag{2.248}$$

В дальнейшем изложении для численного моделирования работы геометрооптического ДОЭ используется фаза (2.246).

Следуя работам [10, 11], кратко опишем итеративные алгоритмы для расчета ДОЭ. Пусть даны комплексная амплитуда освещающего пучка $A(u)$ и требуемое распределение интенсивности $I(x)$ в фокальной плоскости линзы. Комплексная амплитуда $F(x)$ в фокальной плоскости связана с комплексной амплитудой сразу за ДОЭ, $f(u) = a(u)\exp[i\varphi(u)]$, преобразованием Фурье:

$$F(x) = \int_{-b}^{b} f(u) \exp\left(-i\frac{k}{f}xu\right) \mathrm{d}x, \tag{2.249}$$

где $2b$ — размер апертуры ДОЭ. Уравнение для расчета ДОЭ принимает вид

$$|F(x)|^2 = I(x). \tag{2.250}$$

Чтобы найти итеративное решение уравнения (2.250) относительно фазы $\Phi(u)$ следует осуществить некоторую предварительную оценку фазы $\Phi_0(u)$ с последующим расчетом комплексной амплитуды в фокальной плоскости. В этом случае комплексная амплитуда $F_n(x)$, рассчитанная на n-ой итерации, заменяется на функцию $F_n^0(x)$ в соответствии с правилом

$$F_n^0(x) = \begin{cases} \sqrt{I_n(x)}\dfrac{F_n(x)}{|F_n(x)|}, & |x| \leqslant d, \\ \sqrt{\alpha}F_n(x), & |x| > d, \end{cases} \tag{2.251}$$

где $I_n(x) = (1+\alpha)I(x) - \alpha|F_n(x)|^2$, $I(x)$ — заданное распределение интенсивности в интервале $[-d, d]$ фокальной плоскости, α — параметр, который контролирует скорость сходимости рассчитанной интенсивности к заданной. Для $\alpha = 0$, замена (2.251) переходит в стандартную замену в алгоритме ГС [3].

Амплитуда $f_n(u)$ в плоскости ДОЭ рассчитывается с помощью обратного преобразования Фурье и заменяется на функцию $f_n^0(x)$ согласно правилу

$$f_n^0(x) = \begin{cases} A(u)\dfrac{f_n(u)}{|f_n(u)|}, & |u| \leqslant b, \\ 0, & |u| > b. \end{cases} \tag{2.252}$$

В отличие от алгоритмов условного градиента [98], в данном случае параметр α вводится непосредственно в функцию интенсивности.

Скорость сходимости интенсивности $|F_n(x)|^2$ к заданной интенсивности проверяется по среднеквадратичному отклонению

$$\delta = \left[\frac{\int_{-d}^{d}\left[I(x) - |F_n(x)|^2\right]^2 \mathrm{d}x}{\int_{-d}^{d} I^2(x)\,\mathrm{d}x}\right]^{1/2}. \tag{2.253}$$

Кроме того, вводится еще параметр E, характеризующий энергетическую эффективность фокусировки (одномерный аналог формулы (2.44)):

$$E = \frac{\int_{-d}^{d}|F_n(x)|^2\,\mathrm{d}x}{\int_{-\infty}^{\infty}|F_n(x)|^2\,\mathrm{d}x}. \tag{2.254}$$

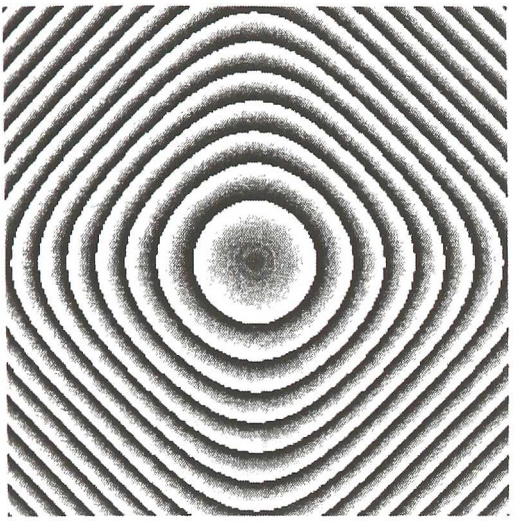


Рис. 2.35. Геометрооптическая фаза ДОЭ фокусирующего в квадрат

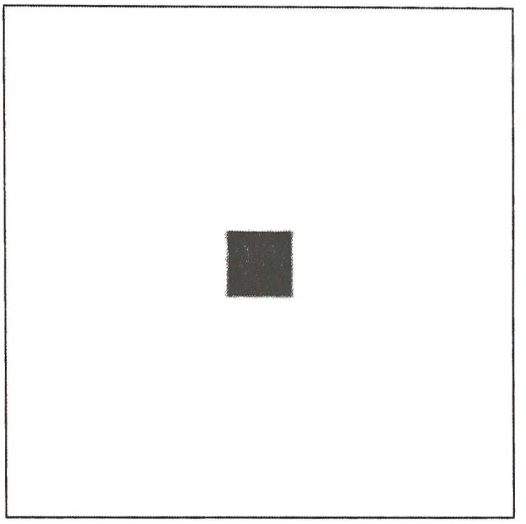


Рис. 2.36. Распределение интенсивности, полученное от геометрооптического ДОЭ в фокальной плоскости линзы

Пример 2.11. Рассмотрим ДОЭ, содержащий 32×32 пиксела и фокусирующий в квадрат, имеющий размеры, равные 10 минимальным дифракционным пятнам. Для ДОЭ, формирующего квадрат из коллимированного гауссова пучка, фаза, полученная из уравнения (2.246) и взятая по модулю 2π на массиве в 256×256 пикселов, находится в виде совокупности колец (линий равной фазы), переходящих в линии квадратного периметра (рис. 2.35). Рисунок 2.36 иллюстрирует распределение интенсивности света в фокальной плоскости линзы, рассчитанное как преобразование Фурье для амплитуды:

$$f_{mn} = \sqrt{I_{0mn}}\, e^{i\varphi_{mn}}, \tag{2.255}$$

где I_{0mn} взято из (2.248), а φ_{mn} из (2.246). Среднеквадратичное отклонение полученного распределения от равномерного распределения составляет 5%, а эффективность равна 91,6%.

Чтобы рассчитать ДОЭ, фокусирующий гауссов пучок в квадрат с однородной интенсивностью, в качестве начальной оценки выбирается геометрооптическая фазовая функция, описанная выше (см. рис. 2.35). Последующий итеративный расчет

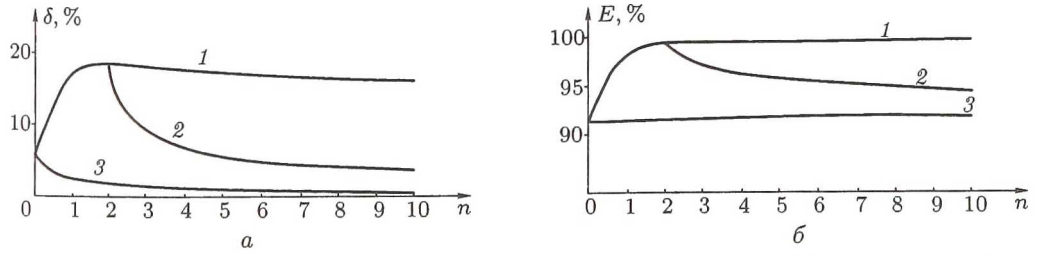


Рис. 2.37. Среднеквадратичное отклонение (*а*) и энергетическая эффективность (*б*) в зависимости от числа итераций для различных вариантов итеративного алгоритма: алгоритм Герчберга–Секстона (*1*), комбинированный (*2*), адаптивный (*3*)

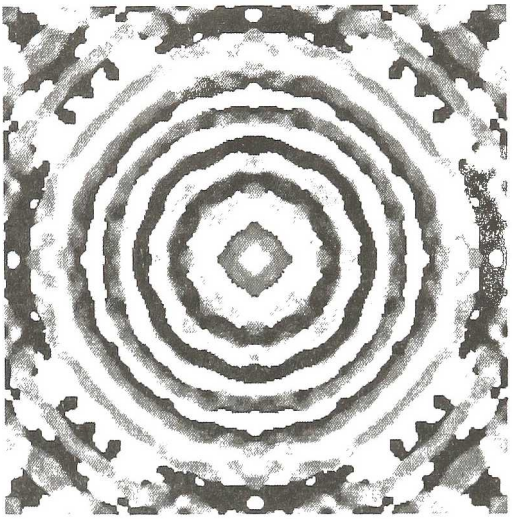


Рис. 2.38. Фаза ДОЭ, фокусирующего в квадрат, полученная после 10 итераций из случайной начальной фазы

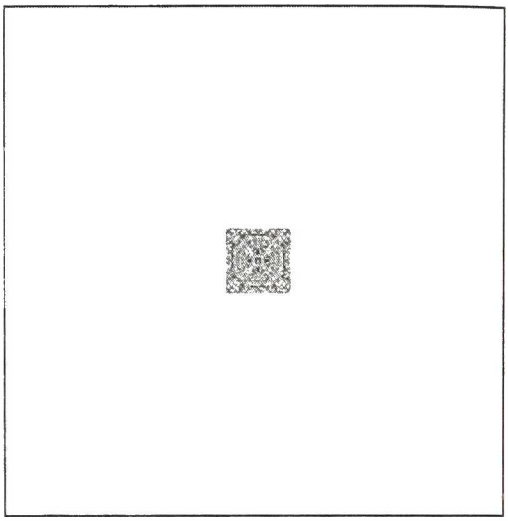


Рис. 2.39. Распределение интенсивности в фокусе линзы, полученное от ДОЭ, с фазой, показанной на рис. 2.38

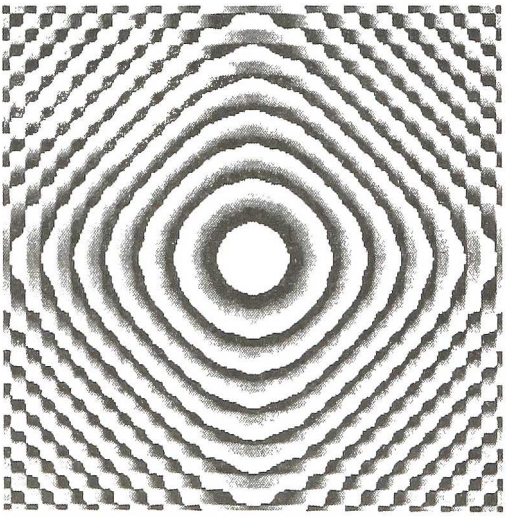


Рис. 2.40. Фаза ДОЭ, фокусирующего в квадрат, полученная после 10 итераций из первоначальной геометрооптической фазы

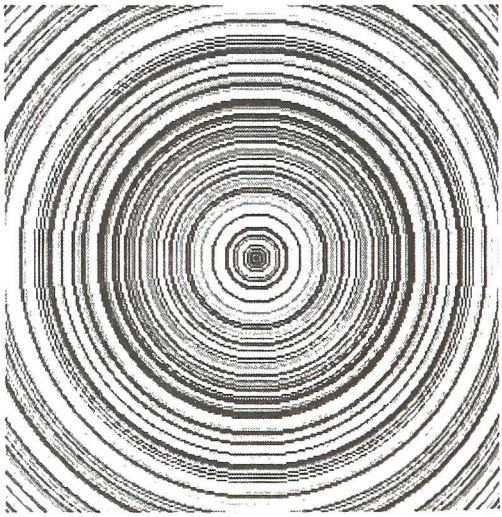


Рис. 2.41. Фаза со случайным радиусом, выбранная в качестве начальной оценки для итераций

Таблица 2.3. Сравнение ДОЭ

ТИП ДОЭ И ФАЗЫ	Энергетическая эффективность $E\%$	Относительная среднеквадратичная ошибка $\delta\%$
Геометрооптическая фаза	91,6	5,0
10 итераций алгоритма ГС на геометрооптической фазе	98,9	13,0
10 итераций адаптивно-итеративного алгоритма на геометрооптической фазе	92,2	0,1
10 комбинированных итераций на геометрооптической фазе	96,4	2,4
10 итераций на фазе со случайным радиусом	80,8	6,4

осуществлялся тремя способами. Первый подход к расчету фазы, основанный на стандартном варианте алгоритма ГС с заменой (2.251) для $\alpha = 0$, сначала дает увеличение ошибки δ на первых итерациях, а затем идет медленное уменьшение ошибки (рис. 2.37*а*, кривая *1*). В этом случае в течение 10 итераций ошибка не уменьшается ниже 13%, хотя при этом эффективность E возрастает до 98.9% (рис. 2.37*б*, кривая *1*).

Второй подход — комбинированный: первые три итерации осуществляются с заменой (2.251) для $\alpha = 0$, оставшиеся семь итераций для $\alpha = 1$ (рис. 2.37*а*, кривая *2*). Этот метод после 10 итераций приводит к формированию квадрата равномерной интенсивности, с ошибкой 2,4% и с эффективностью 96,4% (рис. 2.37*б*, кривая *2*). Видно, что этот подход более эффективен по сравнению с первым, т.к. он дает уменьшение ошибки с 13% до 2% без существенного уменьшения эффективности.

Третий подход — адаптивный, в нем замена (2.251) при $\alpha = 1$ осуществляется на каждом шаге итераций (рис. 2.37*а*, кривая *3*). При этом ошибка уменьшается монотонно и после 10 итераций становится равной 0,1%, но соответствующая эффективность падает до 92,2% (рис. 2.37*б*, кривая *3*). На рис. 2.40 показана фаза ДОЭ (по модулю 2π) рассчитанная в течение 10 итераций на основе третьего подхода из геометрооптической фазы (2.246). Во-первых, видно, что применение итеративного алгоритма не приводит к существенному изменению начальной фазы. Выраженные изменения фазы (рис. 2.35) имеются только на краях фокусатора (рис. 2.40). Во-вторых, можно сделать вывод, что алгоритм, основанный на замене (2.251), при $\alpha = 1$, дает возможность улучшить геометрооптическую фазу таким образом, что ошибка в формировании светового квадрата уменьшится более чем на порядок, при этом эффективность изменится несущественно. Для сравнения был проведен другой вычислительный эксперимент, в котором в качестве начального приближения использовалась фазовая функция со случайным радиусом (рис. 2.41), распределение интенсивности, представлено на рис. 2.39. При этом ошибка и эффективность составили, соответственно, $\delta = 6{,}4\%$ и $E = 80{,}8\%$,. Дальнейшие итерации не дали существенных изменений этих значений. Вышеприведенные данные показывают, что итеративный алгоритм, который начинается со случайной оценки фазы, приводит к нерегулярной структуре зон ДОЭ и дает значение точности и эффективности, несколько меньшие, чем при использовании геометрооптических ДОЭ.

В табл. 2.3 суммируются данные по всем ДОЭ, описанным в данном разделе. Из таблицы видны преимущества с точки зрения равномерности, эффективности, регулярности зонной структуры, которые мы получаем используя физически

обоснованное начальное приближение в виде геометрооптической фазы (сравнение строк 1–4 со строкой 5). Из сравнения строк 1 и 2 таблицы видно, что метод ГС позволяет получить увеличение энергетической эффективности на 7% при геометрооптическом приближении (строка 1), но приводит к значительной неоднородности интенсивности в квадрате (13% вместо 5%). Адаптивный метод, напротив, обеспечивает высокую степень однородности интенсивности (0,1%), но он практически не улучшает эффективность по сравнению с геометрооптической аппроксимацией (строка 3). Комбинированный метод позволяет получить довольно высокую эффективность в сочетании с приемлемой однородностью интенсивности (строка 4). Из строки 5 видно, что, не зная геометрооптического решения, но основываясь на случайной начальной фазе в оценке фазы ДОЭ и используя итеративный метод, можно получить результаты, приемлемые с точки зрения точности и эффективности формирования заданного распределения интенсивности.

Таким образом, численно показано, что при расчете ДОЭ, фокусирующих свет в малые области пространственного спектра, наилучшего результата можно добиться при использовании адаптивно-итеративного алгоритма с начальной оценкой искомой фазы ДОЭ, полученной при решении обратной задачи в геометрооптическом приближении. В ходе итераций регулярная структура зон фазы (рис. 2.47) видоизменяется, в основном, на периферийной части апертуры ДОЭ.

2.8.8. Фокусировка в кольцо

Для фокусировки когерентного света в узкое кольцо обычно используются конические аксиконы в сочетании со сферическими линзами [36] и бинарные аксиконы [99]. Если к фазовой функции, описывающей пропускание аксикона, добавить слагаемое, линейно зависящее от азимутального угла, полученный оптический элемент будет обладать новыми свойствами [100].

Ниже для фокусировки в широкое кольцо с заданным распределением интенсивности вдоль радиуса кольца, рассматривается метод расчета ДОЭ, фокусирующего в круглую область вне оси. Метод сочетает сведение задачи фокусировки в радиальную внеосевую область к одномерной задаче фокусировки в отрезок и использование итеративных способов решения одномерной задачи фокусировки в отрезок. Предложенный метод требует вычисления только двух преобразований Фурье на каждой итерации, в три раза сокращая время расчета по сравнению с итеративным расчетом при помощи преобразования Ханкеля, описанным в разделе 2.8.1.

Световое кольцо в фокальной плоскости [30] считается узким, если его ширина меньше дифракционного предела Фраунгофера для конической волны с ограниченной апертурой, дифрагирующей на ДОЭ.

Для фокусировки в узкое кольцо можно использовать оптический элемент в виде конического аксикона с комплексной функцией пропускания в виде

$$\tau(r) = \exp(i\alpha r), \tag{2.256}$$

где α — характеристика «силы» аксикона, а r — радиальная координата. Если элемент, освещается плоским монохроматическим пучком света и если за этим элементом помещается сферическая линза с фокусным расстоянием f, то в задней фокальной плоскости линзы будет формироваться световое кольцо радиусом $R = \alpha f/k$ и шириной $b = \lambda f/a$, где a — радиус апертуры аксикона. Комплексная амплитуда света в фокальной плоскости линзы задается дифракционным интегралом Фраунгофера,

выраженным в полярных координатах:

$$F_0(\rho) = \frac{k}{f}\int_0^a e^{-i\alpha r} J_0\left(\frac{k}{f} r\rho\right) r\,\mathrm{d}r, \tag{2.257}$$

где $J_0(x)$ — функция Бесселя нулевого порядка. Амплитуда пучка света на кольце рассчитывается по формуле

$$F_0(R) = \frac{k}{f}\int_0^a e^{-i\alpha r} J_0(\alpha r)\, r\,\mathrm{d}r = \\ = \frac{k}{f\alpha^2} e^{i\xi}\left\{\left[\xi J_1(\xi) - \frac{\xi^2}{3} J_2(\xi)\right] + i\frac{\xi^2}{3} J_1(\xi)\right\}, \tag{2.258}$$

где $\xi = \alpha a$, $J_1(x)$, $J_2(x)$ — функции Бесселя первого и второго порядков. Интенсивность света в центре кольца рассчитывается по формуле

$$I_0(0) = \left|\frac{k}{f}\int_0^a e^{-i\alpha r} r\,\mathrm{d}r\right|^2 = \left(\frac{ka^2}{2f}\right)^2\left[\left(\frac{\sin\nu}{\nu}\right)^2 + \left(\frac{\cos\nu}{\nu} - \frac{\sin\nu}{\nu^2}\right)^2\right], \tag{2.259}$$

где $\nu = \alpha a/2$. Как видно из последнего слагаемого в (2.259), интенсивность в центре принимает свое максимальное значение при $\alpha = 0$,

$$I_0(0)|_{\max} = \left(\frac{ka^2}{2f}\right)^2. \tag{2.260}$$

Интенсивность в центре кольца асимптотически достигает нулевого значения, когда α стремится к бесконечности. Уравнение для определения локальных экстремумов интенсивности имеет вид

$$\frac{dI_0}{d\nu} = \left(\frac{ka^2}{2f}\right)^2\left[\frac{-4}{\nu}\left(\frac{\cos\nu}{\nu} - \frac{\sin\nu}{\nu^2}\right)^2\right] = 0. \tag{2.261}$$

Решение уравнения (2.261) имеет вид

$$\nu = \operatorname{tg}\nu. \tag{2.262}$$

Энергетическая эффективность фокусировки в узкое кольцо на основе аксикона находится из соотношения

$$\varepsilon_0 = 2\pi RbI_0(R)[2\pi RbI_0(R) + \pi b^2 I_0(0)]^{-1}, \tag{2.263}$$

где $b = \lambda f/a$ — ширина кольца, практически равная радиусу центрального светового пятна, $I_0(R)$ берется из уравнения (2.258) и равняется

$$I_0(R) = \left(\frac{k}{f}\right)^2\frac{a^4}{9}\left[J_1^2(\xi) + J_2^2(\xi)\right] + \left(\frac{k}{f}\right)^2 a^4\frac{J_1(\xi)}{\xi}\left[\frac{J_1(\xi)}{\xi} - \frac{2}{3}J_2(\xi)\right]. \tag{2.264}$$

Для фокусировки в узкое кольцо также можно использовать винтовой аксикон с функцией пропускания в виде [30, 32]:

$$\tau(r,\varphi) = \exp[-i\alpha r + im\varphi], \tag{2.265}$$

где $m = 0, 1, 2, \ldots, \varphi$ — азимутальный угол; α может принимать как положительные, так и отрицательные значения. Комплексная амплитуда пучка в фокальной плоскости линзы описывается в этом случае вместо (2.257) уравнением

$$F_m(\rho,\psi) = \frac{k}{f} i^m e^{im\psi} \int_0^a e^{-i\alpha r} J_m\left(\frac{k}{f} r\rho\right) r\,\mathrm{d}r, \tag{2.266}$$

где $J_m(x)$ — функция Бесселя m-го порядка, (ρ,ψ) — полярные координаты в плоскости Фурье. Из уравнения (2.266) видно, что независимо от значений α и a интенсивность в центре кольца равна нулю, $J_m(0) = 0$ для $m > 0$. С ростом порядка m интенсивность на кольце рассчитывается по формуле ($m = 1$)

$$I_1(R) = |F_1(R,\psi)|^2 = \left(\frac{k}{f}\right)^2 \frac{a^4}{9}\left[J_1^2(\xi) + J_2^2(\xi)\right]. \tag{2.267}$$

Из сравнения уравнений (2.264) и (2.267) видно, что при использовании винтового аксикона (2.265) интенсивность света, падающего на кольцо, может быть как больше, так и меньше интенсивности, которая получается при использовании аксикона (2.256), и зависит от знака второго члена в выражении

$$I_0(R) = I_1(R) + \left(\frac{k}{f}\right)^2 a^4 \frac{J_1(\xi)}{\xi}\left[\frac{J_1(\xi)}{\xi} - \frac{2}{3}J_2(\xi)\right]. \tag{2.268}$$

Винтовой аксикон не дает значительного выигрыша в энергетической эффективности по сравнению с обычным аксиконом.

Рассмотрим приближенный метод расчета ДОЭ, фокусирующих в радиально-симметричные области. Пусть лазерный пучок с комплексной амплитудой

$$W_0(r) = \sqrt{I_0(r)}\exp(i\varphi_0(r)), \tag{2.269}$$

где $I_0(r)$ — интенсивность освещающего пучка, $\varphi_0(r)$ — фаза пучка, падает на ДОЭ с комплексной функцией пропускания $\exp(i\tilde{\varphi}(r)), r \leqslant a$. Требуется найти фазовую функцию ДОЭ $\tilde{\varphi}(r)$, обеспечивающую формирование заданного кругового распределения интенсивности $I(\rho)$, $\rho_1 \leqslant \rho \leqslant \rho_2$, в фокальной плоскости $z = f$ (см. рис. 2.42). Ниже полагается, что фазовая функция ДОЭ имеет вид

$$\tilde{\varphi}(r) = \varphi(r) - \varphi_0(r). \tag{2.270}$$

Представление (2.270) позволяет рассчитывать ДОЭ без учета фазы освещающего пучка. Комплекная амплитуда поля в фокальной плоскости ДОЭ задается интегралом Френеля–Кирхгофа, выраженным в полярных координатах:

$$F(\rho) = \frac{k}{f}\exp(i\frac{k\rho^2}{2f}) \int_0^a \sqrt{I_0(r)}\exp(i\varphi(r))\exp\left(i\frac{kr^2}{2f}\right) J_0\left(\frac{k}{f} r\rho\right) r\,\mathrm{d}r. \tag{2.271}$$

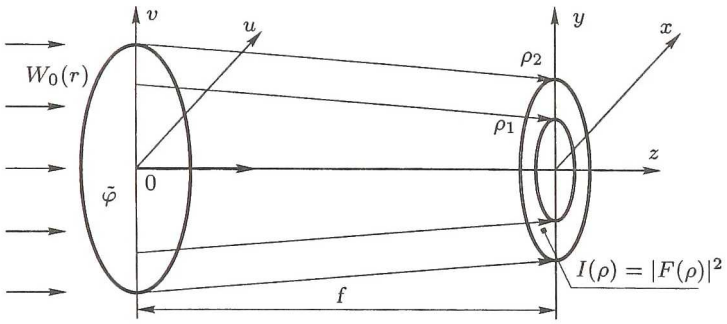


Рис. 2.42. Геометрия задачи фокусировки в кольцевую область

Используя асимптотическое представление функции Бесселя $J_0(\xi)$ при $\xi \gg 0$:

$$J_0(\xi) = \sqrt{\frac{2}{\pi\xi}} \cos\left(\xi - \frac{\pi}{4}\right), \quad \xi \to \infty, \tag{2.272}$$

получим приближение следующего вида для $F(\rho)$ при $\rho \gg 0$:

$$F(\rho) = \exp(-i\pi/4)\left(F_1(\rho)/\sqrt{\rho} + iF_2(\rho)/\sqrt{\rho}\right), \tag{2.273}$$

где

$$F_1(\rho) = \sqrt{\frac{k}{2\pi i f}} \int_0^a \sqrt{I_0(r)\, r} \exp(i\varphi(r)) \exp\left(\frac{ik(\rho - r)^2}{2f}\right) \mathrm{d}r, \tag{2.274}$$

$$F_2(\rho) = F_1(-\rho). \tag{2.275}$$

Функция $F_1(\rho)$ соответствует комплексной амплитуде волны, формируемой одномерным ДОЭ, который освещается пучком с распределением интенсивности

$$\tilde{I}(r) = I_0(r)\, r. \tag{2.276}$$

Задача фокусировки в радиальную область с распределением интенсивности $I(\rho)$, $\rho \gg 0$, может быть сведена к одномерной задаче фокусировки в отрезок. Действительно, пусть $\varphi(r)$ есть фазовая функция одномерного ДОЭ, фокусирующего пучок (2.276) в отрезок с распределением интенсивности

$$I_1(\rho) = I(\rho)\rho, \quad \rho_1 \leqslant \rho \leqslant \rho_2 \ldots \tag{2.277}$$

Заметим, что функции интенсивности $\tilde{I}(r)$ и $I_1(\rho)$ являются размерными и промежуточными функциями, которые отсутствуют в конечном результате.

Согласно уравнению (2.275) член $F_2(\rho)$ соответствует комплексной амплитуде поля при фокусировке в отрезок при $-\rho_1 < \rho < -\rho_2$ и не влияет на поле $F_1(\rho)$ при $\rho_1 \leqslant \rho \leqslant \rho_2$ (см. рис. 2.43). Следовательно, аппроксимация $F(\rho)$ примет вид

$$F(\rho) \approx \exp(-i\pi/4) F_1(\rho)/\sqrt{\rho}. \tag{2.278}$$

Соответственно, функция $|F_1(\rho)|^2/\rho$ представляющая распределение интенсивности в фокальной плоскости радиального ДОЭ превращается в заданное распределение интенсивности $I(\rho)$ для $I_1(\rho) = |F_1(\rho)|^2$, определяемое уравнением (2.277).

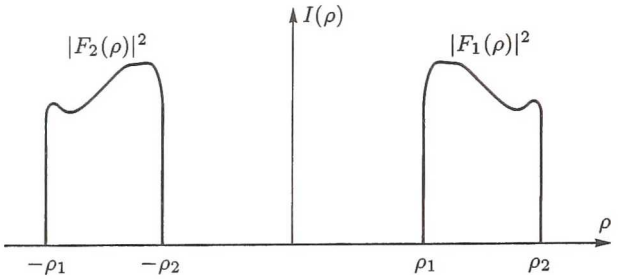


Рис. 2.43. Структура поля, описываемая членами $F_1(\rho)$ и $F_2(\rho)$ уравнения (2.275)

Следует отметить, что при расчете интеграла (2.210) при помощи метода стационарной фазы, аппроксимация (2.278) приводит к известным формулам для фазовой функции фокусатора [30, 66, 67]:

$$\varphi(r) = -\frac{kr^2}{2f} + \frac{k}{f}\int\limits_0^r \tilde{\rho}(\xi)\,\mathrm{d}\xi, \quad r \in [0, a], \tag{2.279}$$

где функция $\tilde{\rho}(\xi)$ получена из следующего соотношения:

$$\int\limits_0^r I_0(\zeta)\zeta\,\mathrm{d}\zeta = \int\limits_{\rho_1}^{\tilde{\rho}(r)} I(\xi)\xi\,\mathrm{d}\xi. \tag{2.280}$$

В работах [30, 66] уравнение (2.279) получено из уравнения эйконала, а уравнение (2.280) получено из уравнения переноса и соответствует интегральной форме закона сохранения светового потока. Например, фазовая функция ДОЭ, фокусирующего плоский пучок в широкое кольцо с равномерным распределением интенсивности, может быть получена из уравнений (2.279) и (2.280) в виде

$$\varphi(r) = -\frac{kr^2}{2f} + \frac{k\rho_1}{2fc}\left[cr\left(c^2r^2+1\right)^{1/2} + \ln\left(cr + \left(c^2r^2+1\right)^{1/2}\right)\right], \tag{2.281}$$

где $c = \left(\rho_2^2 - \rho_1^2\right)^{1/2}/(a\rho_1)$, ρ_1 и ρ_2 — внутренний и внешний радиусы кольца.

Численные оценки позволяют определить фокальную внеосевую область, в которую фокусируется излучение, как область, внутренний радиус которой в 4–5 раз больше дифракционного размера $\Delta = \lambda f/a$.

Геометрооптический метод расчета $\varphi(r)$ приводит к известным уравнениям (2.279), (2.280) для фазовой функции фокусатора. Расчет $\varphi(r)$ при помощи одномерного итеративного алгоритма соответствует новому итеративному методу расчета ДОЭ, фокусирующего в круговую внеосевую область.

Итеративный подход, описанный в пункте 2.8.1, требует вычисления шести преобразований Фурье на каждом шаге. Итеративный метод, рассмотренный в данном параграфе, требует вычисления двух преобразований Фурье на каждую итерацию. Ниже итеративный расчет, основанный на преобразовании Ханкеля, будет называться радиальным итеративным расчетом. Расчет радиального ДОЭ, основанный на итеративном расчете ДОЭ фокусирующего в отрезок, будет называться линейным итеративным расчетом. Ниже проведено сравнение между ДОЭ, фокусирующими в широкое кольцо, которые рассчитывались при помощи геометрооптического подхода при выводе $\varphi(r)$ (см. уравнение (2.279)), при помощи линейного итеративного расчета $\varphi(r)$ и радиального итеративного расчета фазы.

Таблица 2.4. Параметры E и δ для фокусатора и ДОЭ, рассчитанного при помощи линейного и радиального итеративных алгоритмов

S	Геометро-оптический метод		Линейный итеративный алгоритм		Радиальный итеративный алгоритм	
	$E(\%)$	$\delta(\%)$	$E(\%)$	$\delta(\%)$	$E(\%)$	$\delta(\%)$
8	90,4	48,0	89,1	13,1	89,2	13,3
10	89,9	39,4	92,5	11,1	92,6	11,0
12	90,7	37,4	89,4	9,7	89,5	9,6
14	90,9	40,0	91,3	9,1	91,5	9,2
16	90,2	37,7	90,4	7,8	90,7	7,7
18	89,7	34,5	90,8	7,4	90,7	7,2
20	89,7	32,4	90,1	6,9	90,2	6,6

Чтобы оценить характеристики ДОЭ, используем величины энергетической эффективности E и среднеквадратичного отклонения δ Величина

$$E = \int\limits_{\rho_1}^{\rho_2} I(\rho)\,\rho\,\mathrm{d}\rho \Big/ \int\limits_{0}^{a} I_0(r)\,r\,\mathrm{d}r, \tag{2.282}$$

характеризует долю энергии освещающего пучка, фокусируемую в заданной фокальной области. Величина

$$\delta = \frac{1}{\bar{I}} \sqrt{\frac{2}{(\rho_2^2 - \rho_1^2)} \int\limits_{\rho_1}^{\rho_2} \left[I(\rho) - \bar{I}\right]^2 \rho\,\mathrm{d}\rho}, \tag{2.283}$$

характеризует среднеквадратичное отклонение распределения интенсивности от среднего значения

$$\bar{I} = \frac{2}{(\rho_2^2 - \rho_1^2)} \int\limits_{\rho_1}^{\rho_2} I(\rho)\,\rho\,\mathrm{d}\rho. \tag{2.284}$$

Пример 2.12. Линейный и радиальный расчет $\varphi(r)$ осуществлялись при помощи АА-алгоритма (2.37), описанного в разделе 2.4. Фазовая функция (2.279) выбиралась в качестве начальной в итеративном процессе. Пересчет поля между плоскостями производился при помощи быстрого преобразования Фурье с числом пикселов $N = 256$ для следующих параметров: $\lambda = 1{,}06$ мкм, $a = 2{,}5$ мм, $f = 250$ мм. В таблице 2.4 представлены значения E и δ в зависимости от параметра $S = (\rho_2 - \rho_1)/\Delta$, $\Delta = \lambda f/a$, для $\rho_1 = 28\Delta$. Величина S характеризует ширину фокального кольца по сравнению с дифракционным размером Δ.

Из данных, таблицы 2.4 видно, что среднеквадратичная ошибка δ для ДОЭ, рассчитанных при помощи линейного итеративного алгоритма, в 3–5 раз меньше, чем для фокусатора, при этом энергетическая эффективность E меняется незначительно. Радиальный итеративный алгоритм не обеспечивает существенного улучшения E и δ по сравнению с ДОЭ, рассчитанными при помощи линейного алгоритма.

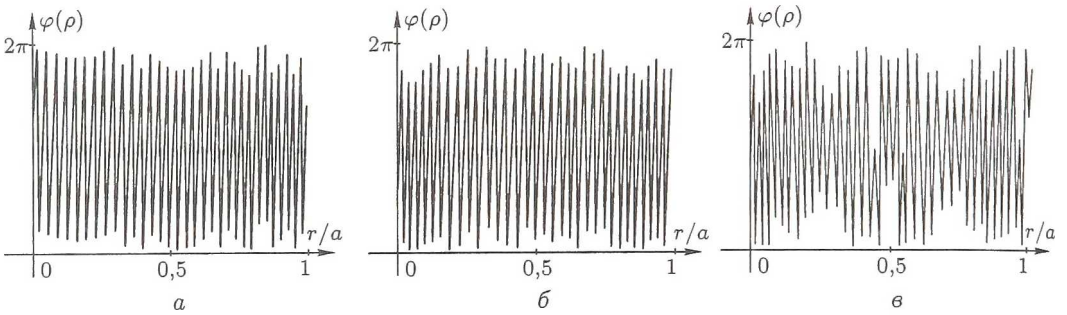


Рис. 2.44. Фазовые функции фокусатора (*а*), ДОЭ, рассчитанного при помощи линейного итеративного алгоритма (*б*), и ДОЭ, рассчитанного при помощи радиального итеративного алгоритма (*в*)

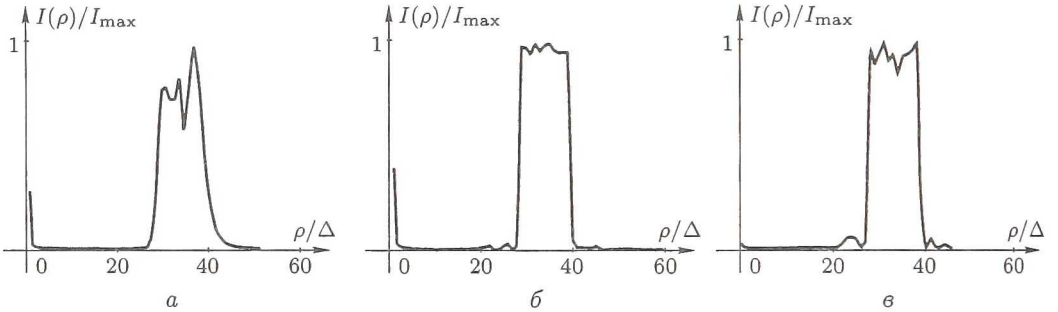


Рис. 2.45. Распределение интенсивностей в фокальной плоскости: фокусатора (*а*), ДОЭ, рассчитанного при помощи линейного итеративного алгоритма (*б*), ДОЭ, рассчитанного при помощи радиального итеративного алгоритма (*в*)

На рис. 2.44 показаны фазовые функции фокусатора и ДОЭ, рассчитанных при помощи линейного и радиального алгоритмов, соответственно, для $S = 10$. Расчетные значения распределений интенсивности на рис. 2.45 выявляют практически полное отсутствие флуктуаций интенсивности на фокальном кольце для «линейного» и «радиального» ДОЭ. Фокусатор и линейный ДОЭ создают пики интенсивности при $\rho = 0$. Наличие пиков объясняется тем, что геометрооптический метод, и линейный итеративный алгоритм не контролируют распределение интенсивности вблизи оптической оси. Для радиального ДОЭ центральный пик не появлялся.

2.8.9. Композиционные ДОЭ, формирующие контурные изображения

В [101] предложена фокусировка лазерного излучения в буквы и цифры с помощью «составных» фокусаторов. Апертура таких фокусаторов состоит из N непересекающихся субапертур, каждая из которых фокусирует свет либо в отрезок прямой линии, либо в узкое полукольцо, из которых состоят буквы. Недостатком таких «составных» ДОЭ является то, что ширина линий фокусировки определяется не размером апертуры ДОЭ, а размером субапертуры. Это приводит к уширению элементов букв и к уменьшению плотности световой энергии. Кроме того, наличие скачков фазы ДОЭ вдоль общих линий соседних субапертур приводит к рассеянию света и уменьшению энергетической эффективности.

В [76] предложен итеративный алгоритм расчета ДОЭ, фокусирующих лазерный свет в объем с заданным распределением интенсивности в N сечениях перпендикулярных к оси распространения света. Этот алгоритм основан на представлении функции пропускания ДОЭ в виде линейной комбинации комплексных амплитуд, являющихся результатом обратного преобразования Френеля от функций, описывающих световое поле в заданных сечениях объекта фокусировки.

В данном разделе рассматривается метод [102] для расчета композиционных ДОЭ. Функция пропускания такого ДОЭ представляется в виде линейной комбинации комплексных амплитуд, являющихся фурье-образами N непересекающихся изображений. Модули коэффициентов линейной комбинации характеризуют распределение световой энергии пучка между N изображениями, а аргументы коэффициентов являются свободными параметрами задачи. Фазу композиционного ДОЭ можно найти методом последовательных приближений с помощью итеративного алгоритма, аналогичного [3, 77, 79] или градиентного алгоритма наискорейшего спуска [5].

В [91] приведены полученные в лучевом приближении выражения для фазовой функции ДОЭ, фокусирующего плоскую волну света в поперечный отрезок прямой линии с постоянной интенсивностью и расположенный на заданном расстоянии. Если ограничить рассмотрение формированием с помощью ДОЭ только контурных изображений, состоящих из участков прямых линий (например, букв), то лучевое решение из [91] можно выбрать за базовое. Тогда любой отрезок в фурье-плоскости можно получить путем растяжения (сжатия), смещения и поворота базового отрезка прямой. А фаза ДОЭ, фокусирующих в произвольный отрезок, получается путем растяжения (сжатия), смещения и поворота базового решения для фазы.

Пусть плоский пучок лазерного света фокусируется ДОЭ с функцией пропускания $\exp[i\varphi(x,y)]$ в контурное изображение на расстоянии f_0 в плоскости (ξ,η), состоящее из набора отрезков длиной d_n, с центрами в точках (ξ_n,η_n) и расположенных под углами φ_n к оси ξ.

Пусть комплексная функция $f_n(\xi,\eta)$, $(\xi,\eta)\in\Omega_n$, описывает световое поле в плоскости (ξ,η). Квадрат модуля этой функции пусть задает распределение интенсивности в n-м изображении, заключенном в области Ω_n. И пусть набор таких областей не пересекается: $\Omega_n\cap\Omega_m=0$. Тогда функции $f_n(\xi,\eta)$ и их френель-образы $F_n(x,y)$ будут ортогональными:

$$\iint\limits_{-\infty}^{\infty} f_n(\xi,\eta)\,f_m^*(\xi,\eta)\,\mathrm{d}\xi\,\mathrm{d}\eta=\iint\limits_{-\infty}^{\infty} F_n(x,y)\,F_m^*(x,y)\,\mathrm{d}x\,\mathrm{d}y=Q_n\delta_{nm}, \tag{2.285}$$

$$Q_n=\iint\limits_{\Omega}|f_n(\xi,\eta)|^2\,\mathrm{d}\xi\,\mathrm{d}\eta=\iint\limits_{-\infty}^{\infty}|F_n(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y, \tag{2.286}$$

$$F_n(x,y)=\frac{ike^{ikf_0}}{2\pi f_0}\iint\limits_{\Omega_n} f_n(\xi,\eta)\exp i\frac{k}{2f_0}\left[(x-\xi)^2+(y-\eta)^2\right]\mathrm{d}\xi\,\mathrm{d}\eta. \tag{2.287}$$

Формула (2.287) является преобразованием Френеля (2.1), записанным в других обозначениях. Исходя из уравнений (2.285)–(2.287) можно сформулировать задачу расчета фазовой функции $\varphi(x,y)$ ДОЭ, фокусирующего световое поле с известной амплитудой $A_0(x,y)$ в набор непересекающихся изображений $f_n(\xi,\eta)$ в плоскости (ξ,η), расположенной на расстоянии f_0 от него:

$$A_0(x,y)e^{i\varphi(x,y)}=\sum_{n=1}^{N}C_nF_n(x,y), \tag{2.288}$$

$$C_n = Q_n^{-1} \iint\limits_W A_0(x,y) e^{i\varphi(x,y)} F_n^*(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{2.289}$$

В уравнении (2.288) квадраты модулей коэффициентов $|C_n|^2$ считаются заданными числами, пропорциональными части световой энергии, идущей на формирование n-го изображения $f_n(\xi,\eta)$. Аргументы коэффициентов $\nu_n = \arg C_n$ являются искомыми параметрами задачи, также как и функция $\varphi(x,y)$.

Вместо уравнения (2.288) для расчета композиционных ДОЭ можно исходить из функционала среднеквадратичной невязки ε:

$$\varepsilon = \iint\limits_W \left[A_0(x,y) - \left| \sum_{n=0}^{N} C_n F_n(x,y) \right| \right]^2 \mathrm{d}x\,\mathrm{d}y. \tag{2.290}$$

Решение уравнения (2.288) эквивалентно минимизации функционала (2.290). Минимум критерия можно искать с помощью градиентного метода наискорейшего спуска. В данном случае, следуя работе [5], градиентный метод можно представить в следующем виде:

$$\nu_n^{(p+1)} = \nu_n^{(p)} + t_p \frac{\partial \varepsilon}{\partial \nu_n^{(p)}}, \tag{2.291}$$

$$\frac{\partial \varepsilon}{\partial \nu_n^{(p)}} = 2 \left| C_n \bar{C}_n^{(p)} \right| \sin\left(\nu_n^{(p)} - \bar{\nu}_n^{(p)} \right), \tag{2.292}$$

$$\bar{C}_n^{(p)} = Q^{-1} \iint\limits_W R^{(p)}(x,y) F_n^*(x,y)\,\mathrm{d}x\,\mathrm{d}y, \tag{2.293}$$

$$R^{(p)}(x,y) = A_0(x,y) \frac{U^{(p)}(x,y)}{\left| U^{(p)}(x,y) \right|} - U^{(p)}(x,y), \tag{2.294}$$

$$U^{(p)}(x,y) = \sum_{n=1}^{N} |C_n|\, e^{i\nu_n^{(p)}} F_n(x,y), \tag{2.295}$$

где $|C_n|$ — произвольно заданные числа. Величина шага t_p алгоритма (2.291) на p-й итерации может быть выбрана следующим образом [5]: $t_p = \varepsilon_p \bar{\varepsilon}_p^{-1}$, $\bar{\varepsilon}_p = -\sum\limits_{n=1}^{N} \left| \partial\varepsilon \Big/ \partial \nu_n^{(p)} \right|^2$. Заметим, что так как в данном алгоритме фазы изображений $f_n(x,y)$ задаются произвольно, то у задачи (2.288) может быть множество решений, из которых можно выбирать лучшее в смысле максимальной эффективности или минимальной ошибки формирования:

$$E = \left[\sum_{n=1}^{N} |C_n|^2 \iint\limits_\Omega |g_n(\xi,\eta)|^2\,\mathrm{d}\xi\,\mathrm{d}\eta \right] \left[\iint\limits_W A_0^2(x,y)\,\mathrm{d}x\,\mathrm{d}y \right]^{-1}, \tag{2.296}$$

$$\delta = \left[\sum_{n=1}^{N} \iint\limits_\Omega \left[|g_n(\xi,\eta)| - \left| g_n^{(p)}(\xi,\eta) \right| \right]^2 \mathrm{d}\xi\,\mathrm{d}\eta \right] \left[\sum_{n=1}^{N} \iint\limits_\Omega |g_n(\xi,\eta)|^2\,\mathrm{d}\xi\,\mathrm{d}\eta \right]^{-1}, \tag{2.297}$$

где $|g_n(\xi,\eta)|$ и $\left|g_n^{(p)}(\xi,\eta)\right|$ — заданный и рассчитанный на p-й итерации модули изображений. Если изображение $g(x,y)$ состоит из N подобных элементарных изображений $g_n(x,y)$, то все изображения можно получить из одного преобразованием подобия. Например, если контурное изображение состоит из N отрезков прямых, то все функции $g_n(x,y)$, $n=\overline{1,N}$, этих отрезков могут быть получены из некоторой базовой функции $g_0(x,y)$ путем смещения T , поворота R и масштабирования (сжатия и растяжения) D:

$$g_n(\xi,\eta) = DTRg_0(\xi,\eta), \tag{2.298}$$

$$Rg_0(\xi,\eta) = g_0(\xi',\eta'), \tag{2.299}$$

$$\begin{cases} \xi' = \xi\cos\varphi_n + \eta\sin\varphi_n \\ \eta' = -\xi\sin\varphi_n + \eta\cos\varphi_n \end{cases}, \tag{2.300}$$

$$Tg_0(\xi,\eta) = g_0(\xi-\xi_n,\eta-\eta_n), \tag{2.301}$$

$$Dg_0(\xi,\eta) = Dg_0(\alpha_n\xi,\alpha_n\eta), \tag{2.302}$$

где φ_n — угол поворота, (ξ_n,η_n) — координаты центра, α_n — коэффициент масштабирования для n-го изображения.

Для контурного изображения, состоящего из набора отрезков, можно воспользоваться результатом работы [91], в которой приведено выражение для фазы ДОЭ, фокусирующего плоскую волну в отрезок прямой с постоянной интенсивностью:

$$\varphi_0(x,y) = -\frac{k}{2f_0}\left[x^2\left(1-\frac{a}{2A}\right)+y^2\right], \tag{2.303}$$

где a — длина отрезка, вытянутого вдоль оси ξ и имеющего минимальную дифракционную ширину по оси η, $2A$ — размер квадратной апертуры ДОЭ.

Считая фазовую функцию (2.303) за базовое решение нетрудно получить фазу ДОЭ, формирующего произвольный тонкий отрезок в плоскости (ξ,η):

$$\begin{gathered} \varphi_n(x,y) = -\frac{k}{2f_0}\left[A_nx^2+B_ny^2+C_nxy+2\xi_nx+2\eta_ny\right], \\ A_n = b_n\cos^2\varphi_n+\sin^2\varphi_n, \quad B_n = b_n\sin^2\varphi_n+\cos^2\varphi_n, \\ C_n = (b_n-1)\sin(2\varphi_n), \\ b_n = \left(1-\frac{a_n}{2R}\right), \end{gathered} \tag{2.304}$$

где R — радиус круглой апертуры ДОЭ. В этом случае уравнение (2.288) для расчета фазовой функции ДОЭ $\bar{\varphi}(x,y)$, фокусирующего в контур из отрезков, примет вид

$$A_0(x,y)e^{i\bar{\varphi}(x,y)} = \sum_{n=1}^{N} C_ne^{i\varphi_n(x,y)}, \tag{2.305}$$

где $\varphi_n(x,y)$ — функция, определенная уравнением (2.304). Аналитическая форма задания фазы элементарных изображений $\varphi_n(x,y)$ удобна с практической точки зрения. Действительно, решение задачи (2.305) с помощью итеративного градиентного

алгоритма (2.291)–(2.295) можно реализовать численно, используя массивы чисел небольшой размерности (например 256×256), чтобы сократить временные затраты на поиск аргументов коэффициентов C_n. Для расчета функции $\bar{\varphi}(x, y)$ из уравнения (2.305), при генерации фотошаблона, сумма, стоящая в правой части уравнения (2.305), вычисляется уже с использованием массива чисел большой размерности ($10^4 \times 10^4$ и более).

Решения (2.303) и (2.304) получены в лучевом приближении и обеспечивают высокую энергетическую эффективность фокусировки в требуемый световой отрезок, но не обеспечивают высокой точности постоянства интенсивности вдоль отрезка. Получить высокую точность фокусировки в отрезок можно с помощью скалярного дифракционного приближения. В этом случае вместо функции (2.303) следует выбирать в качестве базового решения комплексную амплитуду вида

$$f_0(\xi, \eta) = \operatorname{rect}\left(\frac{\xi}{a}\right) \operatorname{rect}\left(\frac{\eta}{b}\right), \tag{2.306}$$

из которой нетрудно получить функции, описывающие произвольный отрезок в плоскости (ξ, η):

$$f_n(\xi, \eta) = \operatorname{rect}\left(\frac{\xi \cos\varphi_n + \eta \sin\varphi_n - \xi_n}{b_n}\right) \operatorname{rect}\left(\frac{-\xi \sin\varphi_n + \eta \cos\varphi_n - \eta_n}{a}\right). \tag{2.307}$$

Тогда уравнение для расчета фазы ДОЭ, фокусирующего свет в контур из N отрезков будет иметь следующий вид (вместо (2.305)):

$$A_0(x, y) e^{i\bar{\varphi}(x,y)} = \sum_{n=1}^{N} C_n \operatorname{sinc}\left[\frac{b_n k}{2f_0}(x\cos\varphi_n + y\sin\varphi_n)\right] \times$$
$$\times \operatorname{sinc}\left[\frac{ak}{2f_0}(-x\sin\varphi_n + y\cos\varphi_n)\right] \cdot \exp\left[-i\frac{k}{f_0}(x\xi_n + y\eta_n)\right], \tag{2.308}$$

где $\operatorname{sinc}(x) = \sin(x)/x$, b_n — длина n-го отрезка, a — дифракционная ширина всех отрезков.

Пример 2.13. При численном моделировании в формуле (2.304) использовались следующие параметры: $f_0 = 100$ мм, $R = 10$ мм, x_n, y_n, $a_n \in [0, 0{,}5$ мм$]$, $\varphi_n \in [0, \pi]$, размерность дискретной матрицы 256×256.

На рис. 2.46–2.48 представлены результаты расчета ДОЭ, фокусирующих в различные контурные фигуры и результаты моделирования работы таких ДОЭ, посредством вычисления распределения интенсивности светового поля на расстоянии f_0 с помощью преобразования Френеля.

При этом в верхней части каждого рисунка показаны ДОЭ, рассчитанные по формуле (2.305), а в нижней — ДОЭ, являющиеся просто композицией угловых сегментов, рассчитанных по формуле (2.303). Из рисунков видно, что чем больше линий будет в контурном изображении, тем большее количество сегментов должен будет содержать ДОЭ. А значит, как было замечено ранее, на каждую линию изображения будет приходиться меньший участок полезной площади ДОЭ, что приводит к уширению линий и уменьшению плотности световой энергии. Этот эффект можно наблюдать и в представленных изображениях, где используется всего 3–4 линии.

Кроме того к недостаткам сегментного ДОЭ следует отнести необходимость подбора наилучшей композиции сегментов, в то время как формула (2.305) обеспечивает полный автоматизм.

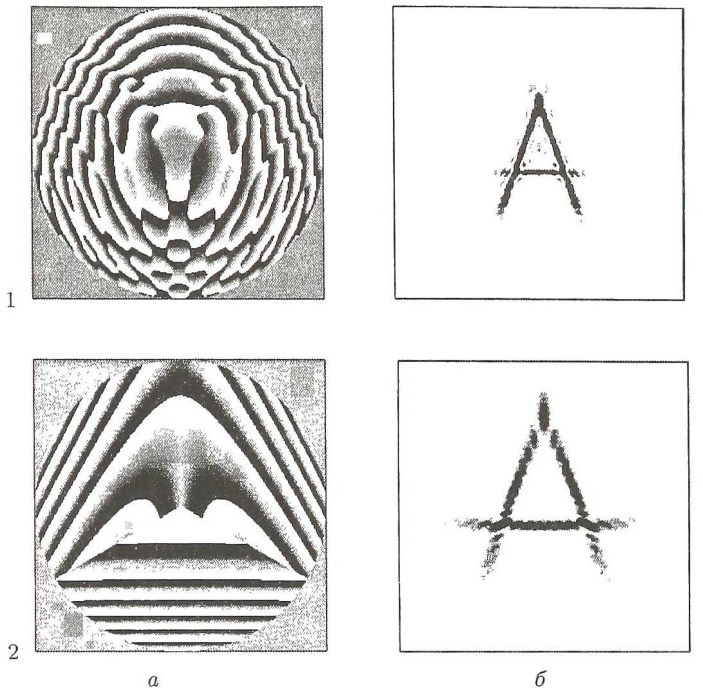


Рис. 2.46. Рассчитанные фазы ДОЭ для фокусировки в букву "А": фазы композиционного (1*а*) и составного (2*а*) ДОЭ и сформированные ими картины дифракции (1*б*) и (2*б*)

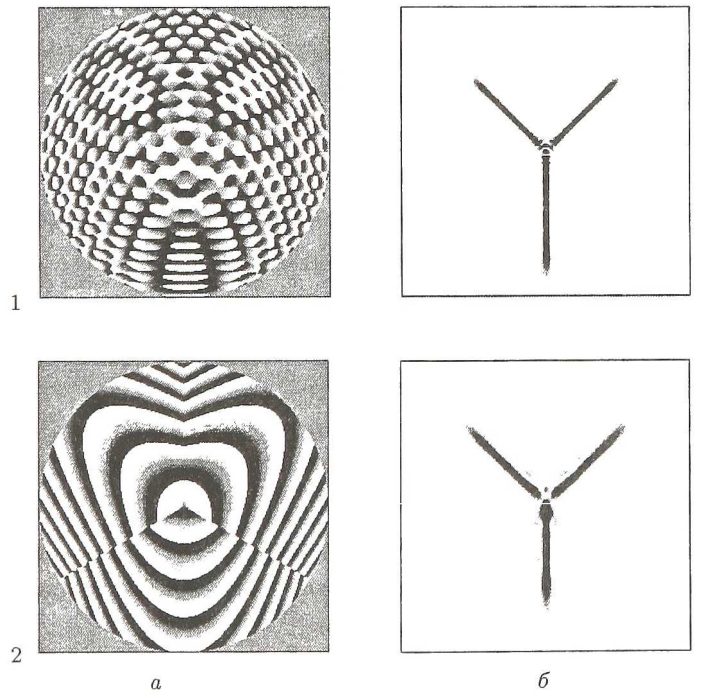


Рис. 2.47. Рассчитанные фазы ДОЭ для фокусировки в букву "Y": фазы композиционного (1*а*) и составного (2*а*) ДОЭ и сформированные ими картины дифракции (1*б*) и (2*б*)

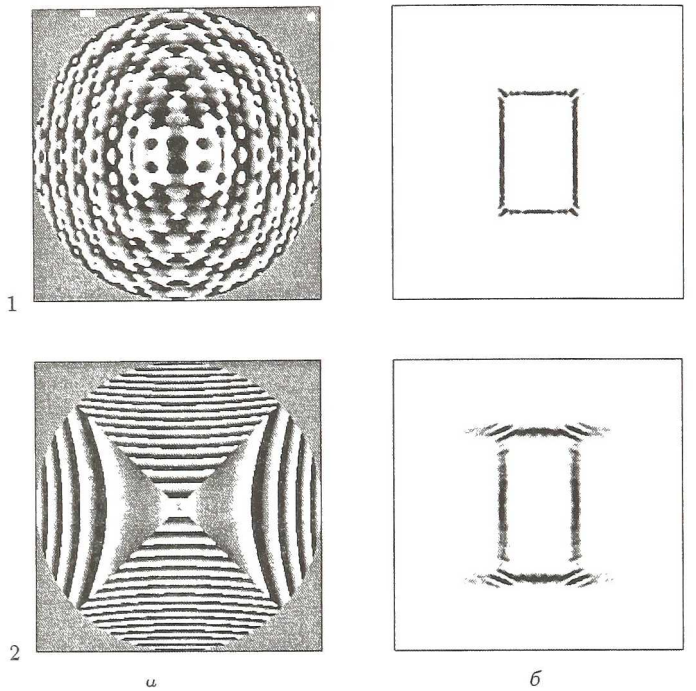


Рис. 2.48. Рассчитанные фазы ДОЭ для фокусировки в букву "О": фазы композиционного (1*а*) и составного (2*а*) ДОЭ и сформированные ими картины дифракции (1*б*) и (2*б*)

2.8.10. Квантованные ДОЭ для фокусировки в заданную двумерную область

Фазовые ДОЭ для фокусировки в заданную двумерную область с требуемым распределением интенсивности важны в задачах преобразования и выравнивания распределения интенсивности лазерного пучка.

Технология фотолитографии, традиционно используемая для изготовления ДОЭ, предусматривает квантование фазовой функции ДОЭ по M уровням. Квантованные ДОЭ с малым числом уровней квантования ($M = 3, 4$) и особенно бинарные ДОЭ ($M = 2$) представляют особый интерес вследствие простоты их изготовления методами фотолитографии. В частности, бинарный рельеф формируется за одно травление подложки.

Хорошо известные итерационные алгоритмы такие, как алгоритм уменьшения ошибки, алгоритм ГС и градиентные алгоритмы [3, 5, 7, 8, 98, 103], предназначенные для расчета ДОЭ с непрерывными фазовыми функциями, оказываются неработоспособными в задачах расчета ДОЭ с малым числом уровней квантования ($M = 2, 3, 4$). В работах Ф. Вировского [21, 104–106] представлена модификация итерационного алгоритма ГС для расчета квантованных ДОЭ при $M > 2$. Указанная модификация основана на введении в итерационный процесс операции частичного квантования и обладает достаточно хорошей сходимостью, при фокусировке во внеосевые фокальные области, даже при числе уровней квантования $M = 3, 4$. В случае расчета квантованных ДОЭ для фокусировки в области вблизи оптической оси, сходимость указанного алгоритма существенно хуже. В частности, при $M = 3, 4$, среднеквадратичная ошибка формирования заданного распределения интенсивности в области фокусировки составляет, как правило, 25–30%.

Расчет бинарных ДОЭ ($M = 2$) основан на методах, существенно отличных от быстрых итерационных алгоритмов типа алгоритма ГС. Например, для расчета многопорядковых бинарных дифракционных решеток обычно используются методы стохастической оптимизации такие, как прямой бинарный поиск и алгоритм симуляции отжига [52–58].

Градиентный метод расчета бинарных решеток, рассмотренный в пункте 2.8.4, является эффективным лишь в задачах расчета одномерных решеток. Методы стохастической оптимизации в общем двумерном случае требуют больших вычислительных затрат и оказываются малопригодными при фокусировке в области, размер которых на порядок превышает размер дифракционного пятна.

В данном пункте рассматривается итерационный метод расчета квантованных ДОЭ. Метод по сложности и вычислительным затратам эквивалентен алгоритму ГС и состоит из следующих шагов: 1) аппроксимация функции комплексного пропускания (ФКП) квантованного ДОЭ непрерывной ФКП; 2) градиентный метод для оптимизации непрерывной ФКП, 3) обратный переход к квантованной ФКП.

По сравнению с алгоритмом, приведенным в работах Ф. Вировского, предлагаемый метод более общий и пригоден для расчета как квантованных ДОЭ, так и бинарных ДОЭ. По сравнению с известными методами [52–58], ориентированными на расчет бинарных фазовых дифракционных решеток (ФДР), предлагаемый метод обладает следующими преимуществами: учет распределения интенсивности освещающего пучка, существенная двумерность, актуальная для нефакторизуемых областей фокусировки и для достижения энергетической эффективности и высокая, по сравнению с методами стохастической оптимизации, скорость сходимости.

Пусть требуется рассчитать ДОЭ с квантованной по M уровням фазовой функцией $\varphi_M(\mathbf{u})$, который при освещении пучком с распределением интенсивности $I_0(\mathbf{u})$ и сходящимся сферическим волновым фронтом (фокус f), формирует в плоскости $z = f$ область D с требуемым распределением интенсивности $I_D(\mathbf{x})$, $\mathbf{x} \in D$ (см. рис. 2.49). Распределение интенсивности $I(\mathbf{x})$ в плоскости $z = f$ в скалярном приближении связано с фазовой функцией ДОЭ $\varphi_M(\mathbf{u})$ через интеграл Френеля–Кирхгофа:

$$I(\mathbf{x}) = \left| \frac{1}{\lambda f} \int\limits_G \sqrt{I_0(\mathbf{u})} \exp\left(i\varphi_M(\mathbf{u}) - i\frac{k}{f}\mathbf{x}\mathbf{u} \right) \mathrm{d}^2\mathbf{u} \right|^2, \qquad (2.309)$$

где G — апертура ДОЭ. Уравнение (2.309) отличается от (2.3) только обозначениями.

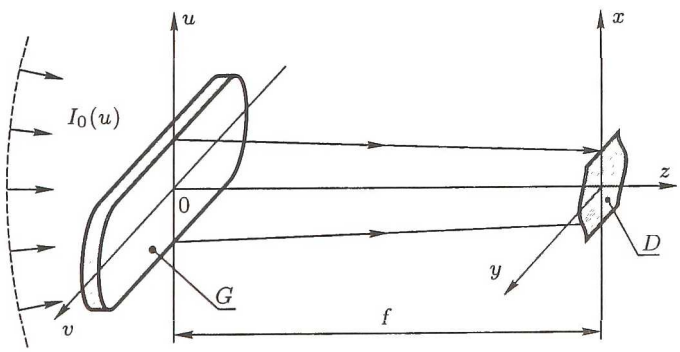


Рис. 2.49. Геометрия задачи фокусировки

ФКП ДОЭ с квантованной по M уровням фазовой функцией $\varphi_M(\mathbf{u})$, принимающей значения $\varphi_j = j(2\pi/M)$, $j = \overline{0, M-1}$, может быть представлена в виде

$$A_M(\mathbf{u}) = \exp\left(i\Phi_M[\varphi(\mathbf{u})]\right), \tag{2.310}$$

где $\varphi(\mathbf{u})$ — некоторая непрерывная функция, определенная в интервале $[0, 2\pi)$, а функция

$$\Phi_M[\xi] = \frac{2\pi}{M}\,\mathrm{int}\left(\frac{\xi M}{2\pi}\right), \tag{2.311}$$

описывает квантование по M уровням. Функция $\mathrm{int}(x)$ в (2.311) является функцией выделения целой части x. Таким образом, (2.310) представляет ФКП квантованного ДОЭ как квантованную функцию $\Phi_M[\xi]$ от непрерывного аргумента $\varphi(\mathbf{u})$.

Основная идея метода расчета квантованных ДОЭ состоит во введении непрерывной функции $\tilde{A}_M(\mathbf{u})$, аппроксимирующей разрывную ФКП $A_M(\mathbf{u})$ квантованного ДОЭ [107–110]. При этом последующее использование быстрых градиентных алгоритмов для оптимизации непрерывной функции $\tilde{A}_M(\mathbf{u})$ позволяет предложить эффективный метод расчета квантованной ФКП (2.310).

Для построения непрерывной аппроксимации $\tilde{A}_M(\mathbf{u})$ разложим функцию $\exp\left(i\Phi_M[\xi]\right)$ в ряд Фурье:

$$\exp\left(i\Phi_M[\xi]\right) = \sum_{n=-\infty}^{\infty} c_n \exp(in\xi), \tag{2.312}$$

где

$$c_n = \begin{cases} \mathrm{sinc}(\pi n/M)\exp\left(i(-\pi)/M\right), & n = 1 + pM, p = \pm 1, \pm 2, \ldots \\ 0, & \text{иначе} \end{cases} \tag{2.313}$$

— коэффициенты Фурье. Полагая $\xi = \mathrm{mod}_{2\pi}[\varphi(\mathbf{u})]$ в (2.312), представим ФКП квантованного ДОЭ с учетом 2π периодичности в виде

$$\exp\left(i\Phi_M[\varphi(\mathbf{u})]\right) = \sum_{n=-\infty}^{\infty} c_n \exp(in\varphi(\mathbf{u})). \tag{2.314}$$

Разложение (2.314) соответствует суперпозиции пучков с фазовыми функциями

$$\varphi_n(\mathbf{x}) = n\varphi(\mathbf{x}). \tag{2.315}$$

Квадраты модулей коэффициентов Фурье (2.313) представляют распределение энергии между пучками, а функции (2.315) определяют структуру изображений, формируемых в дифракционных порядках квантованного ДОЭ [111, 112].

Введем вспомогательную функцию

$$\tilde{A}_M(\mathbf{u}; L_1, L_2) = \sum_{p=-L_1}^{L_2} c_{1+pM} \exp\left(i\left(1 + pM\right)\varphi(\mathbf{u})\right). \tag{2.316}$$

С формальной точки зрения функция $\tilde{A}_M(\mathbf{u}; L_1, L_2)$ является отрезком ряда (2.316), аппроксимирующего ФКП квантованного ДОЭ. При функции $\tilde{A}_M(\mathbf{u}; L_1, L_2)$ (2.316), распределение интенсивности в плоскости $z = f$

$$I(\mathbf{x}; M) = \left| \frac{1}{\lambda f} \int\limits_G \sqrt{I_0(\mathbf{u})} \tilde{A}_M(\mathbf{u}; L_1, L_2) \exp\left(-i \frac{k}{f} \mathbf{x}\mathbf{u}\right) \mathrm{d}^2\mathbf{u} \right|^2 \tag{2.317}$$

учитывает вклад $K = L_2 - L_1 + 1$ ненулевых порядков квантованного ДОЭ (2.310) и интерференцию между ними. Введение функции $\tilde{A}_M(\mathbf{u}; L_1, L_2)$ (2.316) позволяет перейти от квантованной ФКП (2.310) к непрерывной функции. При этом (при разумном выборе параметров L_1, L_2) оказывается эффективным использование хорошо разработанных итеративных алгоритмов [3, 5, 7, 8, 98, 103], предназначенных для оптимизации непрерывных функций.

Задачу расчета квантованного ДОЭ (2.310) предлагается заменить задачей расчета ДОЭ с непрерывной ФКП (2.316) для формирования требуемого распределения интенсивности $I_D(\mathbf{x})$ при $\mathbf{x} \in D$ [107–110]. В этом случае расчет квантованного ДОЭ сводится к выполнению следующих шагов

1. Замена ФКП квантованного ДОЭ функцией $\tilde{A}_M(\mathbf{u}; L_1, L_2)$;
2. Расчет $\varphi(\mathbf{u})$ из условия формирования ДОЭ с ФКП (2.316) заданного распределения интенсивности $I_D(\mathbf{x})$, $\mathbf{x} \in D$;
3. Обратный переход от функции $\tilde{A}_M(\mathbf{u}; L_1, L_2)$ (2.316) к квантованной функции (2.310).

Для расчета $\varphi(\mathbf{u})$ на шаге 2 могут быть использованы различные итерационные алгоритмы [3, 5, 7, 8]. Рассмотрим расчет $\varphi(\mathbf{u})$ с использованием градиентных алгоритмов [98].

Введем некоторый интегральный функционал $\varepsilon(\varphi)$, характеризующий отличие распределения интенсивности (2.317), формируемого ДОЭ с непрерывной ФКП (2.316), от требуемого распределения интенсивности $I_D(\mathbf{x})$:

$$\varepsilon(\varphi) = \bar{\varepsilon}\left(\tilde{I}_M(\mathbf{x}), I_D(\mathbf{x})\right). \tag{2.318}$$

Градиентная минимизация функционала (2.318) состоит в итеративной коррекции функции $\varphi(\mathbf{x})$ по правилу

$$\varphi_{n+1}(\mathbf{u}) = \varphi_n(\mathbf{u}) - \varepsilon'(\varphi_n)\, t, \tag{2.319}$$

где индекс n — номер итерации, $\varepsilon'(\varphi)$ — градиент функционала, t— шаг градиентного метода. Проводя вычисления, аналогичные рассмотренным в пунктах 2.7 и 2.8.5, несложно получить приращение функционала $\varepsilon(\varphi)$, вызванное малым изменением $\Delta\varphi$ фазы $\varphi(\mathbf{u})$ в следующем виде:

$$\Delta\varepsilon(\varphi) = \varepsilon(\varphi + \Delta\varphi) - \varepsilon(\varphi) = 2\,\mathrm{Re}\left(\int\limits_{-\infty}^{\infty} \sqrt{I_0(\mathbf{u})} \Psi^*(\mathbf{u})(\Delta)\, \mathrm{d}^2\mathbf{u}\right), \tag{2.320}$$

где $\Delta = \Delta\tilde{A}_M(\mathbf{u}; L_1, L_2)$ — приращение ФКП ДОЭ, вызванное малым изменением фазы $\Delta\varphi$.

Расчет функции $\Psi(\mathbf{u})$ в (2.320) проводится в три этапа:

1) расчет комплексной амплитуды поля $W(x)$ в плоскости $z = f$;

2) замена функции $W(x)$ на функцию

$$W_1(\mathbf{x}) = \frac{d\bar{\varepsilon}\left(\tilde{I}_M, I_D\right)}{d\tilde{I}_M} W(\mathbf{x}); \tag{2.321}$$

3) обратный пересчет поля в плоскость $z = 0$.

Пересчет поля между плоскостями $z = 0$ и $z = f$ сводится к расчету прямого и обратного преобразований Фурье от функций $W_0(\mathbf{u}) = \tilde{A}_M(\mathbf{u}; L_1, L_2)\sqrt{I_0(\mathbf{u})}$ и $W_1(\mathbf{x})$, соответственно. Используя символы $\mathcal{F}\{\}$ и $\mathcal{F}^{-1}\{\}$ для обозначения прямого и обратного преобразований Фурье, представим функцию $\Psi(\mathbf{u})$ в виде

$$\Psi(\mathbf{x}) = \mathcal{F}^{-1}\left\{\frac{d\bar{\varepsilon}\left(\tilde{I}_M, I_D\right)}{d\tilde{I}_M}\mathcal{F}\left\{\sqrt{I_0(\mathbf{u})}\tilde{A}_M(\mathbf{u}; L_1, L_2)\right\}\right\}. \tag{2.322}$$

Интересно отметить, что расчет функции $\Psi(\mathbf{u})$ совпадает с первыми тремя шагами алгоритма ГС с точностью до замены в плоскости $z = f$.

Согласно (2.316), приращение $\Delta\tilde{A}_M(\mathbf{u}; L_1, L_2)$, вызванное малым изменением фазы $\Delta\varphi$, имеет вид

$$\Delta\tilde{A}_M(\mathbf{u}; L_1, L_2) = \sum_{p=-L_1}^{L_2} c_{1+pM} \cdot i(1+pM)\exp(i(1+pM)\varphi(\mathbf{u}))\Delta\varphi(\mathbf{u}). \tag{2.323}$$

Подставляя (2.323) в (2.320), получаем градиент функционала в виде

$$\varepsilon'(\varphi) = 2\,\mathrm{Re}\left(\sqrt{I_0(\mathbf{u})}\Psi^*(\mathbf{u})\sum_{p=-L_1}^{L_2} c_{1+pM} \cdot i(1+pM)\exp(i(1+pM)\varphi(\mathbf{u}))\right). \tag{2.324}$$

Уравнения (2.322) и (2.324) определяют градиент для непрерывного функционала (2.318), (2.317). Уравнение (2.324) при функции $\Psi(\mathbf{u})$, определенной в виде

$$\Psi(\mathbf{x}) = \mathcal{F}^{-1}\left\{\frac{d\bar{\varepsilon}(I, I_D)}{dI}\mathcal{F}\left\{\sqrt{I_0(\mathbf{u})}A_M(\mathbf{u})\right\}\right\} \tag{2.325}$$

может быть рассмотрено как градиент разрывного функционала

$$\varepsilon(\varphi) = \bar{\varepsilon}(I(\mathbf{x}), I_D(\mathbf{x})), \tag{2.326}$$

представляющего отличие распределения интенсивности $I(\mathbf{x})$ в (2.309) для квантованного ДОЭ (2.310) от требуемого распределения интенсивности $I_D(\mathbf{x})$. Действительно, функционал (2.326) является предельным случаем функционала (2.318) при $L_1 \to \infty$ и $L_2 \to \infty$. В этом случае приращение функционала $\Delta\varepsilon(\varphi)$ также представимо в виде (2.320). При этом функция $\Psi(\mathbf{u})$ в (2.320) имеет вид (2.321), а приращение Δ соответствует приращению квантованной ФКП $A_M(\mathbf{u})$, вызванному малым изменением фазы $\Delta\varphi$:

$$\Delta A_M(\mathbf{u}) = \exp(i\Phi_M[\varphi(\mathbf{u}) + \Delta\varphi(\mathbf{u})]) - \exp(i\Phi_M[\varphi(\mathbf{u})]). \tag{2.327}$$

Используя в (2.310) непрерывную аппроксимацию $\Delta A_M(\mathbf{u}) \approx \Delta\tilde{A}_M(\mathbf{u}, L_1, L_2)$ (2.323) для приращения разрывной функции $A_M(\mathbf{u})$, получаем выражение для градиента функционала в виде (2.324). Вычисление приращения $\Delta\tilde{A}_M(\mathbf{u}; L_1, L_2)$ содержит операцию дифференцирования отрезка ряда Фурье

$$S(\xi; L_1, L_2) = \sum_{p=-L_1}^{L_2} c_{1+pM} \exp(i(1+pM)\xi), \tag{2.328}$$

для функции $\exp(i\Phi_M[\xi])$. Формальное дифференцирование рядов Фурье может привести к значительным потерям в точности вычисления производной [113]. Для улучшения устойчивости градиентного метода, дифференцирование ряда (2.328) целесообразно проводить с применением методов регуляризации. Один из простых вычислительных приемов регуляризации, используемых для улучшения сходимости дифференцированного ряда Фурье, состоит в умножении коэффициентов формально дифференцированного ряда на весовые множители d_n, монотонно убывающие с ростом n. Методам выбора весовых множителей посвящено большое количество работ [113–115]. Для ряда (2.328) могут быть использованы следующие достаточно универсальные весовые множители [113]:

$$d_n = \operatorname{sinc}\left(\frac{\pi n}{m}\right), \quad m = \max(L_1, L_2) + 1. \tag{2.329}$$

При использовании множителей (2.329), приращение (2.323) принимает вид

$$\Delta\tilde{A}_M(\mathbf{u}; L_1, L_2) = \sum_{p=-L_1}^{L_2} c_{1+pM} d_{1+pM} i(1+pM) \exp(i(1+pM)\varphi(\mathbf{u}))\,\Delta\varphi(\mathbf{u}). \tag{2.330}$$

При этом градиент функционалов (2.318) и (2.326) представляется в виде

$$\varepsilon'(\varphi) = 2\,\mathrm{Re}\left(\sqrt{I_0(\mathbf{u})}\Psi^*(\mathbf{u}) \sum_{p=-L_1}^{L_2} c_{1+pM} d_{1+pM} i(1+pM) \exp(i(1+pM)\varphi(\mathbf{u}))\right). \tag{2.331}$$

Для иллюстрации влияния регуляризующих множителей (2.329) рассмотрим дифференцирование отрезка ряда Фурье для функции

$$\exp(i\Phi_2[\xi]) = -\operatorname{sgn}(\xi - \pi). \tag{2.332}$$

Функция

$$\Phi_2[\xi] = \begin{cases} 0, & \xi \in [0, \pi), \\ \pi, & \xi \in [\pi, 2\pi) \end{cases} \tag{2.333}$$

в (2.332) описывает операцию бинаризации. Согласно (2.332), производная

$$\frac{d\exp(i\Phi_2[\xi])}{d\xi}$$

равна нулю в каждой точке интервала $[0, 2\pi)$ за исключением точки $\xi = \pi$, где производной, понимаемой в классическом смысле, не существует. Обобщенная производная периодически продолженной функции (2.332) соответствует знакочередующемуся ряду из δ-функций. Полагая в (2.312) и (2.313) $M = 2$, представляем отрезок ряда Фурье для функции (2.332) при $L_1 = L_2$ в виде

$$S(\xi; L_1, L_1) = \frac{2}{\pi} \sum_{p=1}^{L_1} \frac{1}{(2p-1)} \sin((2p-1)\xi). \tag{2.334}$$

Формальное дифференцирование отрезка ряда Фурье (2.334) дает ряд

$$\frac{dS(\xi; L_1, L_1)}{d\xi} = \frac{2}{\pi} \sum_{p=1}^{L_1} \cos((2p-1)\xi), \tag{2.335}$$

который при $L_1 \to \infty$ расходится во всех точках интервала $(0, 2\pi)$ [113]. Умножая ряд (2.335) на множители (2.329), получим регуляризованную производную:

$$\frac{dS(\xi; L_1, L_2)}{d\xi} = \frac{2}{\pi} \sum_{p=1}^{L_1} \frac{\sin(\pi(2p-1)/2L_1)}{\pi(2p-1)/2L_1} \sin((2p-1)\xi). \tag{2.336}$$

Ряд (2.336), в отличие от ряда (2.335), при $L_1 \to \infty$ сходится к нулю в каждой фиксированной точке $\xi \neq \pi$ [113]. Интересно отметить, что ряд (2.336) является тригонометрическим рядом для δ-функции. Точность вычисления производной, для функции (2.332) с использованием ряда (2.335) и регуляризованного ряда (2.336) иллюстрируется на рис. 2.50. В силу симметрии относительно точки $\xi = \pi$, функции (2.335) и (2.336) рассчитаны на отрезке $[0, \pi]$. Рис. 2.50 ясно показывает, что для ряда (2.335) характерны существенно более сильные осцилляции в окрестности требуемого нулевого значения.

Оптимальное значение шага для градиентной процедуры выбирается в виде [5]

$$t = -\varepsilon_1(0)/\varepsilon_1'(0). \tag{2.337}$$

Для вычисления $\varepsilon_1'(0)$ достаточно подставить в уравнение (2.320) значение $\Delta\varphi = -(\Delta t) \cdot \varepsilon'(\varphi)$. В результате производная $\varepsilon_1'(0)$ в (2.337) примет вид

$$\varepsilon_1'(0) = -4\,\mathrm{Re}\left[I_0(\mathbf{u}) \left(\Psi^*(\mathbf{u}) \sum_{p=-L_1}^{L_2} \gamma_{1+pM} i(1+pM) \exp(i(1+pM)\varphi(\mathbf{u})) \right)^2 \right], \tag{2.338}$$

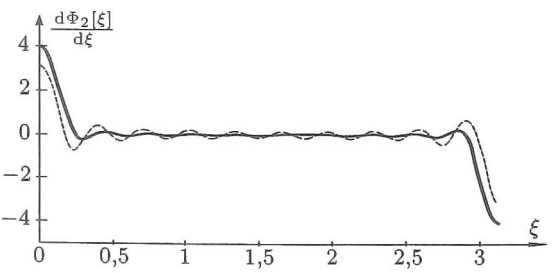


Рис. 2.50. Производные бинарной функции (2.332); пунктирная линия — ряд (2.335) при $L_1 = 10$, непрерывная линия — регуляризованный ряд (2.336) при $L_1 = 10$

где

$$\gamma_{1+pM} = c_{1+pM}, \tag{2.339}$$

при расчете без учета регуляризующих множителей (2.329), и

$$\gamma_{1+pM} = c_{1+pM} d_{1+pM}, \tag{2.340}$$

при регуляризованном расчете приращения $\Delta\tilde{A}_M(\mathbf{u}; L_1, L_2)$.

В заключение рассмотрим вопрос о выборе начального приближения для градиентной процедуры (2.319). Для расчета квантованных ДОЭ с числом уровней квантования $M > 2$ в качестве начального приближения можно использовать непрерывную функцию $\varphi_0(\mathbf{u}) \in [0, 2\pi)$, рассчитанную из условия формирования требуемого распределения интенсивности $I_D(\mathbf{x}), \mathbf{x} \in D$. Расчет непрерывной (неквантованной) функции $\varphi_0(\mathbf{u})$ не представляет сложности. Для расчета могут быть использованы быстрые итерационные алгоритмы, рассмотренные в пунктах 2.2–2.8. Расчет $\varphi_0(\mathbf{u})$ может быть также проведен аналитически в приближении геометрической оптики [37, 66, 116]. Операция квантования (2.310), (2.311) приводит к снижению эффективности фокусировки и появлению дополнительных дифракционных порядков — пучков с фазовыми функциями (2.315). Требуемая область фокусировки создается в первом порядке при $n = 1$. Паразитные порядки (2.315) при $n \neq 1$ накладываются на основное изображение и искажают его.

Доля энергии освещающего пучка $E(M)$, фокусируемая в первом порядке, соответствует квадрату модуля коэффициента Фурье c_1 в (2.313):

$$E(M) = |c_1|^2 = \operatorname{sinc}^2(\pi/M). \tag{2.341}$$

График функции $E(M)$ приведен на рис. 2.51. Согласно графику на рис. 2.51 эффекты квантования особенно существенны при малых $M = 2, 3, 4$; при $M > 4$ квантованный ДОЭ направляет уже более 85% энергии в требуемую область.

При расчете бинарных ДОЭ ($M = 2$) расчет начальной фазы $\varphi_0(\mathbf{u}) \in [0, 2\pi)$ следует модифицировать аналогично тому, как это было сделано в пункте 2.8.4 при расчете одномерных бинарных дифракционных решеток. Распределение интенсивности, формируемое бинарным ДОЭ, является центрально-симметричным; $I(\mathbf{x}) = I(-\mathbf{x})$. Поэтому область D и заданное распределение интенсивности $I_D(\mathbf{x})$ предполагаем центрально-симметричными. Представим область фокусировки D в виде объединения двух центрально-симметричных областей D_1 и D_2. Начальную фазу $\varphi_0(\mathbf{u})$ при $M = 2$ следует рассчитывать из условия фокусировки в половину области D, то есть в область D_1 или в область D_2. Действительно, пусть $\varphi_0(\mathbf{u})$ является

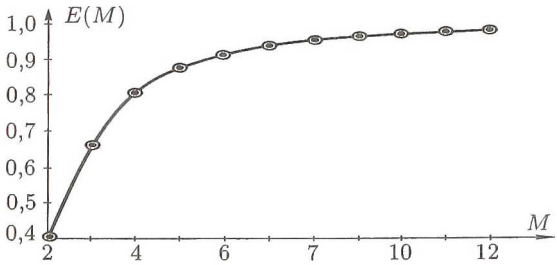


Рис. 2.51. Доля энергии освещающего пучка, фокусируемая в основном 1 порядке, при квантовании фазы ДОЭ по M уровням

фазовой функцией ДОЭ для формирования распределения интенсивности

$$I_{D_1}(\mathbf{x}) = \begin{cases} I_D(\mathbf{x}), & x \in D_1, \\ 0, & x \notin D_1. \end{cases} \tag{2.342}$$

При этом интенсивности полей, формируемых ДОЭ с ФКП $\exp(i\Phi_2[\varphi_0(\mathbf{u})])$ в основных 1 и -1 порядках ($|c_1|^2 = |c_{-1}|^2 = 0{,}405$), имеют вид

$$I_1(\mathbf{x}) = \left| \frac{c_1}{\lambda f} \int\limits_G \sqrt{I_0(\mathbf{u})} \exp\left(i\varphi_0(\mathbf{u}) - i\frac{k}{f}\mathbf{x}\mathbf{u} \right) \mathrm{d}^2\mathbf{u} \right|^2, \tag{2.343}$$

$$I_{-1}(\mathbf{x}) = \left| \frac{c_{-1}}{\lambda f} \left(\int\limits_G \sqrt{I_0(\mathbf{u})} \exp\left(i\varphi_0(\mathbf{u}) - i\frac{k}{f}(-\mathbf{x})\mathbf{u} \right) \mathrm{d}^2\mathbf{u} \right)^* \right|^2. \tag{2.344}$$

Согласно (2.343), (2.344) распределения интенсивности являются центрально-симметричными:

$$I_1(\mathbf{x}) = I_{-1}(-\mathbf{x}). \tag{2.345}$$

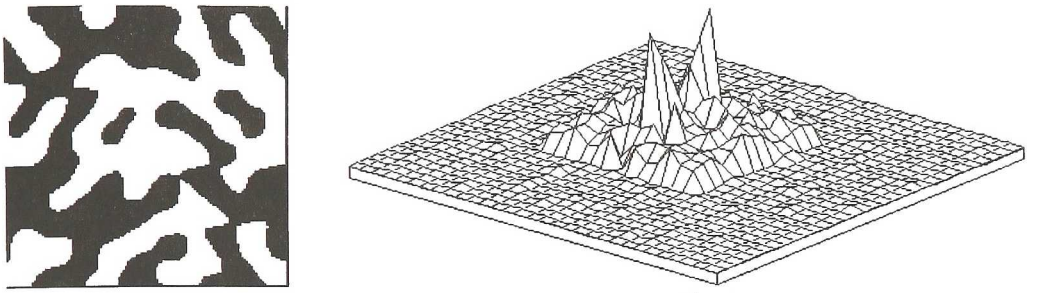
Из (2.345) следует, что если $I_1(\mathbf{x})$ является требуемым распределением интенсивности $I_D(\mathbf{x})$, при $\mathbf{x} \in D_1$, то в -1 порядке будет сформировано распределение интенсивности $I_{-1}(\mathbf{x}) \equiv I_D(\mathbf{x})$, при $\mathbf{x} \in D_2$. Сложение интенсивностей полей, формируемых в 1 и -1 порядках, дает требуемое распределение интенсивности $I_D(\mathbf{x})$ при $\mathbf{x} \in D$. Поскольку $|c_1|^2 + |c_{-1}|^2 = 0{,}81$ то, при учете только главных ± 1 порядков получим, что в требуемой области D фокусируется 81% энергии освещающего пучка. Рассмотренный выбор фазовой функции $\varphi_0(\mathbf{u})$ дает приближенное решение $\Phi_2[\varphi_0(\mathbf{u})]$ задачи фокусировки при $M = 2$. Мы рассмотрели только два главных порядка без учета погрешности расчета функции $\varphi_0(\mathbf{u})$. Итерационные алгоритмы расчета $\varphi_0(\mathbf{u})$, как правило, позволяют сфокусировать в требуемой области 85–90% энергии освещающего пучка при среднеквадратичной ошибке (с.к.о.) формирования требуемого распределения интенсивности в 10–20%. В этом случае, даже при учете только 1-го и -1-го порядков, распределение интенсивности в области D будет отличаться от $I_D(\mathbf{x})$ не только вследствие погрешности расчета фазы $\varphi_0(\mathbf{u})$, но и вследствие интерференции между полями, формируемыми в 1-м и -1-м порядках. Взаимодействием порядков можно пренебречь лишь в случае фокусировки в специфические области, состоящие из двух разнесенных симметричных подобластей. Для практически интересных фокальных областей, таких, как прямоугольник, круг, отрезок, бинарный ДОЭ, как правило, дает энергетическую эффективность 70–75% при с.к.о. формирования заданного распределения интенсивности в 45–55%.

Пример 2.14. Для иллюстрации работы разработанного метода был рассчитан бинарный ДОЭ для фокусировки пучка с гауссовым распределением интенсивности,

$$I_0(\mathbf{u}) = \exp\left(-\mathbf{u}^2/\sigma^2\right), \quad |\mathbf{u}| \leqslant \sigma, \tag{2.346}$$

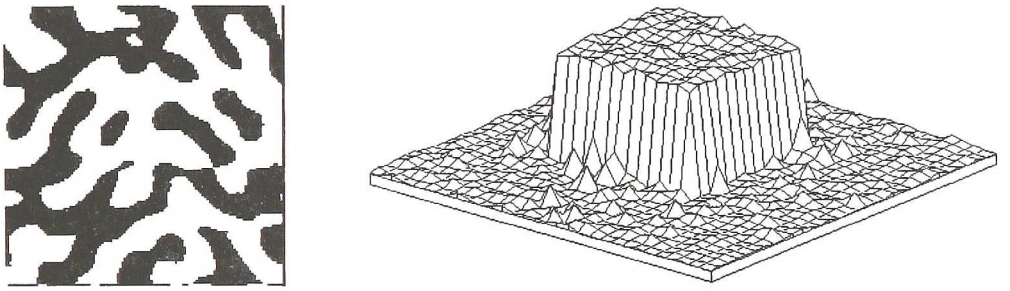
в квадрат $2d \times 2d$ с постоянным распределением интенсивности при $2d = 13\Delta$, где $\Delta = \lambda f/\sigma$ —дифракционный размер. Значения энергетической эффективности E и ошибка δ вычислялись по формулам (2.44) и (2.253).

Как показано выше, приближенное решение для фазовой функции бинарного ДОЭ может быть получено прямой бинаризацией фазовой функции $\varphi_0(\mathbf{u})$, рассчитанной из условия фокусировки пучка (2.346) в «половину квадрата». На рис. 2.52*а*



а *б*

Рис. 2.52. Бинарная фаза ДОЭ для фокусировки в квадрат (приближенное решение) (*а*); распределение интенсивности при фокусировке в квадрат (*б*)



а *б*

Рис. 2.53. Бинарная фаза ДОЭ для фокусировки в квадрат, рассчитанная итерационно при $L_1 = L_2 = 5$ (*а*); распределение интенсивности при фокусировке в квадрат (*б*)

приведена бинарная фаза $\varphi_2(\mathbf{u}) = \Phi_2[\varphi_0(\mathbf{u})]$, полученная при $\varphi_0(\mathbf{u})$, рассчитанной по алгоритму ГС для фокусировки в половину квадрата. Расчетное распределение интенсивности в плоскости $z = f$ для бинарного ДОЭ приведено на рис. 2.52*б*, из которого видно, что хотя энергия сфокусирована в пределах квадрата ($E = 74{,}8\%$), распределение интенсивности является существенно неравномерным ($\delta = 54{,}5\%$). Для получения лучшего решения, ФКП бинарного ДОЭ была аппроксимирована функцией $A_2(\mathbf{u}; L_1, L_2)$ при $L_1 = L_2 = 5$. Расчет $\varphi(\mathbf{u})$ в (2.316) проводился итерационно по формулам (2.319), (2.331) с использованием регуляризации (2.330). В качестве функционала квадрата ошибки (2.318) был использован

$$\varepsilon(\varphi) = \int_{-\infty}^{\infty} \left(\sqrt{\tilde{I}_M(\mathbf{x})} - \sqrt{I_D(\mathbf{x})} \right)^2 \mathrm{d}^2\mathbf{x}. \tag{2.347}$$

В качестве начального приближения для $\varphi(\mathbf{u})$ в (2.316) использовалась фаза $\varphi_0(\mathbf{u})$. Указанный выбор начального приближения позволяет в 2–3 раза сократить число итераций. На рис. 2.53*а* приведена бинарная фаза $\varphi_2(\mathbf{u})$, полученная бинаризацией функции $\varphi(\mathbf{u})$ в (2.316), рассчитанной за 50 итераций. Расчетное распределение интенсивности, формируемое бинарным ДОЭ приведено на рис. 2.53*б*. Для ДОЭ с ФКП $A_2(\mathbf{u}; 5, 5))$ и соответствующего ему бинарного ДОЭ значения E и δ отличались незначительно и составили 73,5%, 1,5% и 72,6%, 2,9% соответственно.

Пример 2.15. Рассчитаны ДОЭ для фокусировки пучка (2.346) в кольцо с радиусами $R_1 = 9\Delta$, $R_2 = 18\Delta$, и в букву "Н", состоящую из трех отрезков с длинами 12Δ, 12Δ и 7Δ. Рассчитанные бинарные фазы и распределения интенсивности, формируемые бинарными ДОЭ, приведены на рисунках 2.54 и 2.55. При фокусировке

в кольцо значения E и δ составили 73,2% и 3,2%, а при фокусировке в букву "Н" – 70,2% и 2,1%, соответственно.

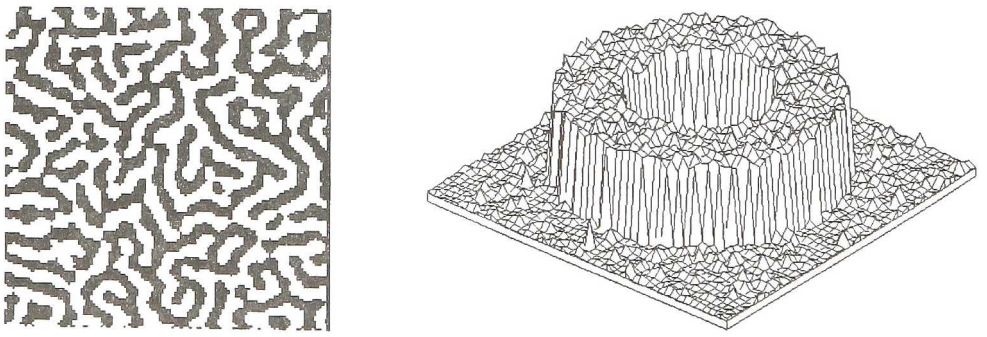
В процессе расчетов получается, что аппроксимация ФКП бинарного ДОЭ функцией $A_2(\mathbf{u}; L_1, L_2)$ является приемлемой уже при $L_1 = L_2 > 4{,}5$. Для приведенных примеров, включающих регуляризованный расчет приращения $\Delta\tilde{A}_2(\mathbf{u}; L_1, L_2)$, обратный переход от функции $A_2(\mathbf{u}; L_1, L_2)$ к бинарной функции при $L_1 = L_2 = 4{,}5$ приводил к снижению энергетической эффективности E всего на 1–2% при незначительном увеличении δ на 1–2%. Отметим, что расчет приращения $\Delta\tilde{A}_2(\mathbf{u}; L_1, L_2)$ без регуляризирующих множителей приводит к большей (в 2–3 раза) ошибке в значениях E и δ при обратном переходе к бинарной функции.

Бинарные ДОЭ позволяют сформировать только центрально-симметричное распределение интенсивности $I_D(\mathbf{x})$. Для фокусировки в несимметричную область D число уровней квантования M должно быть больше 2. В качестве примера фокусировки в несимметричную область, был рассчитан ДОЭ с четырьмя уровнями квантования, для фокусировки в букву "F", состоящую из трех отрезков с длинами 13Δ, 7Δ и 4Δ. Рассчитанная фазовая функция и распределение интенсивности, формируемые четырехуровневым ДОЭ, приведены на рис. 2.56. При фокусировке в букву "F" значения E и δ составили 70,2% и 2,1%, соответственно. Интересно отметить, что полученное значение энергетической эффективности E при $M = 4$ примерно такое же, как и для бинарных ДОЭ на рис. 2.52–2.55. Это объясняется тем, что при расчете бинарных ДОЭ используются два основных порядка: 1 и -1. При этом доля энергии $E = |c_1|^2 + |c_{-1}|^2 = 0{,}81$, фокусируемая в основных порядках 1 и -1 бинарного ДОЭ равна доле энергии E (2.341) (см. рис. 2.51), фокусируемой в основном первом порядке при $M = 4$.

Отметим, что расчет квантованного ДОЭ с квадратной апертурой $G(b \times b)$ для фокусировки пучка с постоянной интенсивностью в прямоугольник $N\Delta \times K\Delta$, где $\Delta = \lambda f / b$ — размер дифракционного пятна, эквивалентен расчету фазовой дифракционной решетки (ФДР), формирующей $N \times K$ порядков. Действительно, для $\mathbf{x} = \mathbf{x}_{j,p} = (j\Delta, p\Delta)$ значения интенсивности в плоскости $z = f$,

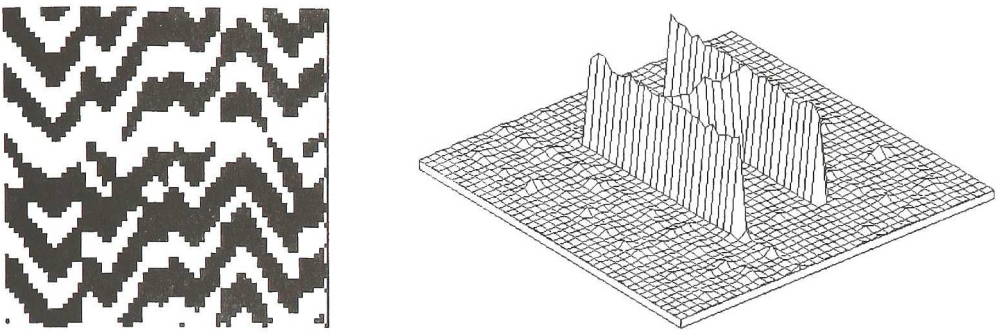
$$I_{j,p} = \left| \frac{1}{\lambda f} \int\limits_G \exp\left(i\varphi_M(\mathbf{u})\right) \exp\left(-i\frac{2\pi}{b}(ju + pv)\right) \mathrm{d}^2\mathbf{u} \right|^2, \qquad (2.348)$$

пропорциональны квадратам модулей коэффициентов Фурье в разложении функции $\exp\left(i\varphi_M(\mathbf{u})\right)$. Поэтому функция $\varphi_M(\mathbf{u})$ соответствует M-уровневой фазовой



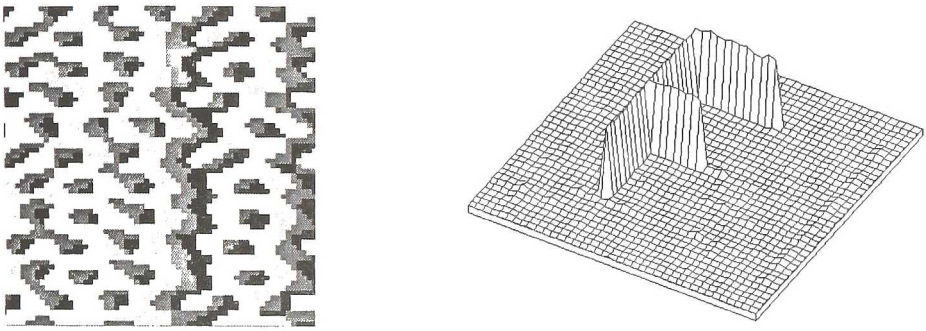
а *б*

Рис. 2.54. Бинарная фаза ДОЭ для фокусировки в кольцо (*а*); распределение интенсивности при фокусировке в кольцо (*б*)



а *б*

Рис. 2.55. Бинарная фаза ДОЭ для фокусировки в букву "H" (*а*); распределение интенсивности при фокусировке в букву "H" (*б*)



а *б*

Рис. 2.56. 4-х уровневая фаза ДОЭ для фокусировки в букву "F" (*а*); распределение интенсивности при фокусировке в букву "F" (*б*)

функции дифракционной решетки с периодом $b \times b$ и интенсивностью порядков

$$I_{j,p} = \begin{cases} I_D\left(j\Delta, p\Delta\right), & |j| \leqslant N/2, |p| \leqslant K/2, \\ 0, & \text{иначе.} \end{cases} \tag{2.349}$$

Пример 2.16. В таблице 2.5 представлены результаты расчета бинарных ФДР для различного числа порядков с равной интенсивностью. Энергетическая эффективность $E = 73$–75% и среднеквадратичная ошибка $\delta = 3$–5%, полученные для ФДР с числом порядков до 51×51, подтверждают высокую работоспособность предложенного метода. При итерационном расчете решеток, в качестве начального приближения $\varphi_0(\mathbf{u})$ использовалась следующая фазовая функция:

$$\varphi_0(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(\frac{N}{4}\left(\left(\frac{u}{b}\right)^2 + 2\left(\frac{v}{b}\right)^2 + 4\left(\frac{u}{b}\right)\right)\right), \quad (u,v) \in [-b/2, b/2] \tag{2.350}$$

Функция (2.350) является фазовой функцией фокусатора [71, 89], рассчитанного в приближении лучевой оптики из условия фокусировки в прямоугольник:

$$\mathbf{x} \in \left[0, \frac{N}{2}\Delta\right] \times \left[-\frac{N}{2}\Delta, \frac{N}{2}\Delta\right].$$

Фазовая функция (2.350) соответствует фазе периода дифракционной решетки с числом порядков $(N/2) \times N$ и обеспечивает фокусировку в половину исходной области из $N \times N$ порядков. Использование аналитического начального приближения

Таблица 2.5. Результаты расчета бинарных ФДР

Число порядков	Энергетическая эффективность $E(\%)$	Среднеквадратичная ошибка $\delta(\%)$
5×5	74,4	3,1
7×7	73,7	3,1
9×9	72,6	2,9
11×11	73,8	3,6
15×15	74,7	3,4
21×21	73,8	3,7
33×33	75,1	4,8
51×51	74,2	4,9

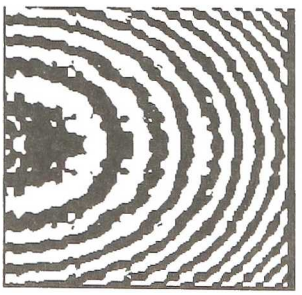
(2.350) упрощает расчетную процедуру и дает более быструю сходимость итерационного процесса (2.319).

На рис. 2.57*а* показана расчетная бинарная фаза ФДР, формирующей 33×33 порядков с равной интенсивностью. Квазирегулярный вид полученной бинарной фазы объясняется использованием в качестве начального приближения гладкой функции (2.350). Распределение интенсивности в плоскости $z = f$ для ФДР показано при числе периодов решетки 2×2.

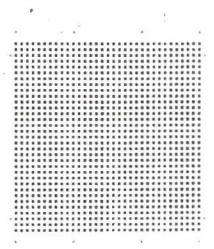
2.8.11. Квантованные ДОЭ для формирования амплитудно-фазовых распределений

Применение рассмотренного в пункте 2.9.10 метода не ограничивается расчетом квантованных ДОЭ для фокусировки в заданную область. Метод может быть также применен к расчету бинарных и квантованных ДОЭ для формирования заданных амплитудно-фазовых распределений.

Пусть требуется рассчитать ДОЭ с квантованной по M уровням фазовой функцией $\varphi_M(\mathbf{u})$ для формирования в области D при $z = f$ заданного распределения комплексной амплитуды $W_D(\mathbf{x})$. С учетом специфики задачи будем предполагать, что ДОЭ расположен в передней плоскости фурье-линзы (при $z = -f$). В этом случае комплексная амплитуда в выходной плоскости (при $z = f$) соответствует



а



б

Рис. 2.57. Бинарная фаза ФДР с числом порядков 33×33 (*а*); распределение интенсивности в Фурье плоскости (*б*)

преобразованию Фурье входного поля:

$$W(\mathbf{x}) = \frac{1}{\lambda f}\int\limits_G \sqrt{I_0(\mathbf{u})}\exp\left(i\varphi_M(\mathbf{u}) - i\frac{k}{f}\mathbf{xu}\right)\mathrm{d}^2\mathbf{u}. \tag{2.351}$$

Расчет квантованных ДОЭ для формирования полей с заданным распределением комплексной амплитуды также основывается на оптимизации непрерывной функции $\tilde{A}_M(\mathbf{u}; L_1, L_2)$ (2.316). Расчет функции $\varphi(\mathbf{u})$ проводится градиентным методом и состоит в итерационной коррекции функции $\varphi(\mathbf{u})$ по правилу (2.319). Функционал $\varepsilon(\varphi)$ в этом случае соответствует некоторому интегральному функционалу

$$\varepsilon(\varphi) = \bar{\varepsilon}\left(\tilde{W}_M(\mathbf{x}), W_D(\mathbf{x})\right), \tag{2.352}$$

характеризующему отличие распределения комплексной амплитуды, формируемой ДОЭ с непрерывной ФКП,

$$\tilde{W}_M(\mathbf{x}) = \frac{1}{\lambda f}\int\limits_G \sqrt{I_0(\mathbf{u})}\tilde{A}_M(\mathbf{u}; L_1, L_2)\exp\left(-i\frac{k}{f}\mathbf{xu}\right)\mathrm{d}^2\mathbf{u}, \tag{2.353}$$

от требуемого амплитудно-фазового распределения $W_D(\mathbf{x}), \mathbf{x} \in D$.

Градиент $\varepsilon'(\varphi)$ функционала также представляется формулами (2.324), (2.331) при функции $\Psi(\mathbf{u})$, определенной в виде

$$\Psi(\mathbf{x}) = \mathcal{F}^{-1}\left\{\frac{d\bar{\varepsilon}\left(\tilde{W}_M, W_D\right)}{d\tilde{W}_M}\right\}. \tag{2.354}$$

Обратный переход от непрерывной функции $\tilde{A}_M(\mathbf{u}; L_1, L_2)$ к квантованной функции (2.310) дает M-уровневую фазовую функцию.

В заключение отметим, что метод может быть также обобщен на случай расчета ДОЭ с ФКП следующего общего вида:

$$A(\mathbf{u}) = \Phi[\varphi(\mathbf{u})], \tag{2.355}$$

где $\varphi(\mathbf{u})$ — непрерывная функция определенная в интервале $[0, 2\pi)$, а $\Phi[\xi]$, $\xi \in \in [0, 2\pi)$ — некоторая разрывная комплексная функция, определяющая поточечное преобразование функции $\varphi(\mathbf{u})$. ФКП квантованного ДОЭ (2.310), является частным случаем представления (2.355) при функции $\Phi[\xi]$, определенной в виде

$$\Phi[\xi] = \exp(i\Phi_M[\xi]) = \exp\left(i\frac{2\pi}{M}\cdot\mathrm{int}\left(\frac{\xi M}{2\pi}\right)\right). \tag{2.356}$$

ФКП (2.355) также может быть представлена в виде ряда по дифракционным порядкам (2.314). В этом случае коэффициенты c_n будут коэффициентами Фурье функции $\Phi[\xi]$. Использование отрезка ряда по дифракционным порядкам для ФКП (2.355) также позволяет заменить задачу оптимизацию разрывной ФКП (2.355) задачей оптимизации непрерывного отрезка ряда (2.314). Конкретный вид функции $\Phi[\xi]$ в (2.355) зависит от расчетной задачи. Например, функция $\Phi[\xi] = (1/2\pi)\Phi_M[\xi]$ может быть использована для расчета квантованных амплитудных фильтров.

Пример 2.17. Работоспособность метода расчета квантованных ДОЭ для формирования полей с заданным распределением комплексной амплитуды проиллюстрируем на задаче расчета моданов — фазовых ДОЭ для формирования полей с комплексной амплитудой, соответствующей модовым функциям оператора распространения света [10, 117]. Метод был применен к расчету бинарных моданов для преобразования гауссова пучка (2.346) в унимодальные распределения Гаусса–Эрмита:

$$M_{(p',l)}(\mathbf{x}) = H_p\left(\sqrt{2}x/\sigma\right) H_l\left(\sqrt{2}y/\sigma\right) \exp\left(-\mathbf{x}^2/\sigma^2\right), \tag{2.357}$$

где (p,l) — индекс моды, а $H_p(x)$ — полином Эрмита степени p. Расчет проводился для модовых распределений $(1,0)$ и $(1,1)$ и был основан на градиентной процедуре (2.319), (2.324), (2.330), (2.331), (2.354) при $L_1 = L_2 = 5$ и следующем функционале ошибки:

$$\varepsilon(\varphi) = \int_{-\infty}^{\infty} \left|\tilde{W}_M(\mathbf{x}) - W_D(\mathbf{x})\right|^2 \mathrm{d}^2\mathbf{x}. \tag{2.358}$$

Расчет производился для следующих физических параметров: длина волны $\lambda =$ $= 0{,}63$ мкм, параметр гауссова пучка $\sigma = 1{,}41$ мм, фундаментальные радиусы мод (1,0) и $(1,1)$ $\sigma_{(0,1)} = \sigma_{(1,1)} = 0{,}123$ мм, фокусное расстояние фурье-линзы $f = 500$ мм, размер области D, в которой формируются модовые распределения, 0,2 мм $\times$ 0,2 мм.

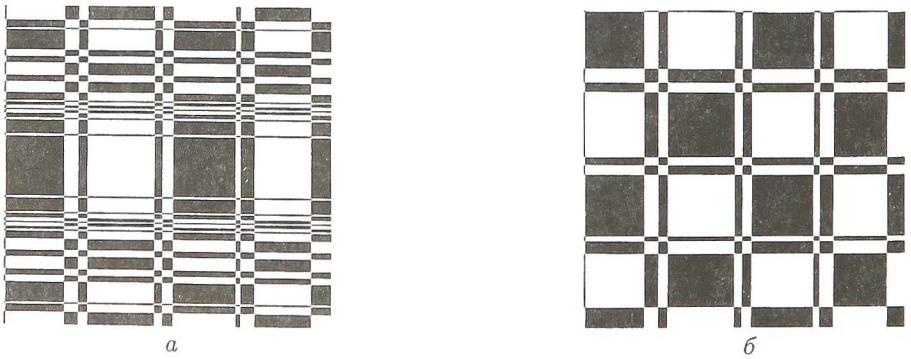
Рассчитанные бинарные фазы ДОЭ для формирования мод Гаусса-Эрмита $(0,1)$ и $(1,1)$ приведены на рис. 2.58а и 2.58б соответственно. Распределения амплитуды и фазы полей, генерируемых ДОЭ в выходной плоскости фурье-линзы, приведены на рис. 2.59. рис. 2.59 ясно показывает структуру полей характерную для модовых распределений $(0,1)$ и $(1,1)$.

Для характеристики работоспособности ДОЭ были рассчитаны значения энергетической эффективности

$$\mathrm{E}_m = \int_D |W_D(\mathbf{x})|^2 \mathrm{d}^2\mathbf{x} \Big/ \int_G I_0(\mathbf{u}) \mathrm{d}^2\mathbf{u}, \tag{2.359}$$

и модового коэффициента

$$\gamma = \left|\int_D W(\mathbf{x}) W_D^*(\mathbf{x}) \mathrm{d}^2\mathbf{x}\right|^2 \Big/ \int_D |W(\mathbf{x})|^2 \mathrm{d}^2\mathbf{x}, \tag{2.360}$$



а *б*

Рис. 2.58. Бинарная фазы моданов для преобразования гауссова пучка в модовые пучки: Гаусса–Эрмита (0,1) (*а*), Гаусса–Эрмита (1,1) (*б*)

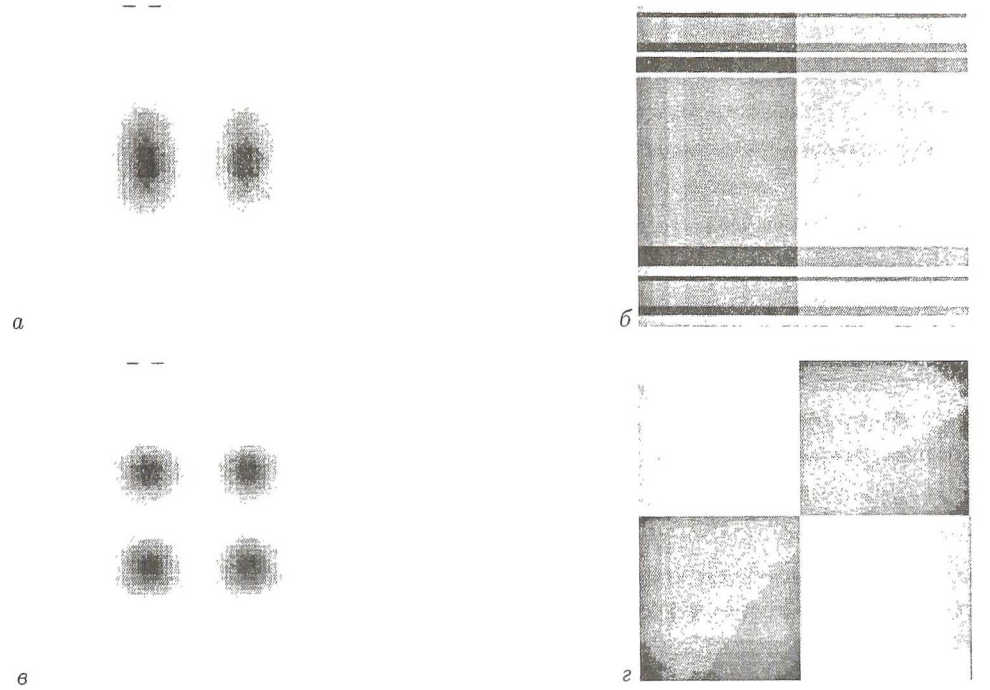


Рис. 2.59. Распределения амплитуды и фазы в Фурье-плоскости: (*а*), (*б*) — для бинарного модана на рис. 2.58*а*; (*в*), (*г*) — для бинарного модана на рис. 2.58*б*

характеризующего содержание моды в полезной области D [117]. Рассчитанные значения E_m и γ составили 47,7%, 0,958 для моды $(0,1)$ и 43,9%, 0,947 для моды $(1,1)$.

Следует отметить, что обобщенный метод кодирования Кирка–Джонса, традиционно используемый для расчета моданов, не позволяет рассчитывать бинарные моданы. Для сравнения были рассчитаны моданы с непрерывной фазой для вышеуказанных параметров по обобщенному методу Кирка–Джонса [117]. Для рассчитанных по методу Кирка–Джонса моданов энергетическая эффективность более чем в 1,5 раза меньше, чем для бинарных моданов на рис. 2.59, что говорит о высокой работоспособности разработанного метода.

2.9. Заключение

Итеративные алгоритмы (ИА) решения обратных задач скалярной теории дифракции для расчета ДОЭ, рассмотренные в этой главе, не охватывают весь спектр известных итеративных алгоритмов. Здесь не были рассмотрены такие популярные в прошлом алгоритмы, как метод прямого стохастического поиска бинарной фазы [52], метод моделирования отжига [54], метод диффузии ошибки [105], метод Ньютона для расчета решеток Дамманна [61] и другие. Эти методы менее эффективны по сравнению с рассмотренными параметрическими и градиентными ИА.

Хотя ИА расчета ДОЭ имеют свои недостатки, например, стагнация ошибки, которая заключается в том, что из-за неудачного выбора начальной оценки искомой фазы ДОЭ, алгоритм может попасть в локальный минимум, когда с увеличением числа итераций ошибка не изменяется, несомненные достоинства ИА объясняют их широкое применение в дифракционной оптике. Укажем некоторые из них.

Универсальность. С помощью ИА можно рассчитывать любые типы ДОЭ: фокусаторы, киноформы, аксиконы, дифракционные решетки, многофокусные линзы, компенсаторы, формирователи мод и бездифракционных пучков, пространственные фильтры.

Модульность. Различные типы ИА можно комбинировать и «стыковать» между собой так, чтобы выходные данные одного алгоритма поступали на вход другого.

Высокая эффективность. По отношению к неитеративным методам кодирования, применяемым для расчета цифровых голограмм, когда с помощью введения несущей пространственной частоты (явной, как в методе Кирка–Джонса, или неявной, как в алгоритме Ломана [74]) рассчитывается голограмма, формирующая в первом порядке дифракции требуемое изображение с эффективностью не более 10–30%, ИА позволяют рассчитывать ДОЭ, работающие в нулевом порядке с эффективностью 70–90%.

Физическая реализуемость. ИА типа «пинг-понг» или «маятника», основанные на последовательном применении пары прямого и обратного интегральных преобразований Фурье, Френеля, Ханкеля могут быть реализованы в оптической схеме с обратной связью [98] или в резонаторе, образованном парой зеркал [118].

Квантование фазы. ИА позволяют удобно выполнять такую нелинейную операцию над фазовой функцией ДОЭ, как квантование (кусочно-постоянная аппроксимация фазы). При этом квантовать фазу можно прямым «жестким» способом на каждом шаге итеративного процесса или «мягким» способом, аппроксимируя ее конечной суммой ряда Гудмена–Сильвестри [112].

Литература к главе 2

1. Василенко Г. И., Тараторин А. М. *Восстановление изображений* (М:Радио и связь, 1986)
2. Гудмен Дж. *Введение в Фурье-оптику* (М.:Мир, 1970)
3. Gerchberg R. W., Saxton W. O. *Optik* **35**(2) 237 (1972)
4. Lin B., Gallagher N. C. *Appl. Opt* **13**(11) 2470 (1974)
5. Fienup J. R. *Appl. Opt* **21**(15) 2758 (1982)
6. Сивоконь В. П. *Формирование световых пучков с заданной структурой для задач лазерной технологии. Кандидатская диссертация* (М.:МГУ, 1986)
7. Fienup J. R. *Opt. Lett.* **3**(1) 27 (1978)
8. Fienup J. R. *Opt. Eng.* **19**(3) 297 (1980)
9. Lu G., Zhang Zh., Yu F T. S., Tanone A. *Opt. Eng.* **33**(2) 548 (1994)
10. Kotlyar V V., Nikolski I. V., Soifer V. A. *Optik* **88**(1) 17 (1991)
11. Kotlyar V. V., Nikolski I. V. *Opt. & Las. in Engin.* **15**(5) 323 (1991)
12. Kotlyar V. V., Seraphimovich P. G., Soifer V. A. *Proceedings of SPIE "Image Processing and Computer Optics"* **2363** 175 (1994)
13. Fienup J. R. *Appl. Opt.* **32**(10) 1737 (1993)
14. Котляр В. В., Серафимович П. Г., Сойфер В. А. *Оптика и спектроскопия* **78**(1) 148 (1995)
15. Vasilyev E. D., Kotlyar V. V., Nikolsky I. V., Soifer V. A. *Proceedings of SPIE* **1983** 1012 (1993)
16. Борн М., Вольф Э. *Основы оптики* (М.:Наука, 1973)
17. Котляр В. В., Серафимович П. Г. *Оптика и спектроскопия* **77**(4) 678 (1994)
18. Тихонов А. Н., Арсенин В. Я. *Методы решения некорректных задач* (М.:Наука, 1979)
19. Gerchberg R. W. *Opt. Acta* **21** 709 (1974)
20. Papoulis A. *IEEE Trans. Circuits Syst.* **22** 735 (1975)
21. Wyrowski F. J. *Opt. Soc. Am.* **7**(6) 961 (1990)

22. Kotlyar V. V., Seraphimovich P. G., Soifer V. A. *Optics and Laser in Engineering* **29**(4-5) 261 (1998)
23. Youla D C *IEEE Trans. Circuits. Syst.* **25** 694 (1979)
24. Aagedal H., Schmid M., Beth T., Teiwes S., Wyrowski F. J. *Mod. Opt.* **43**(7) 1409 (1996)
25. Abbis J. B., Defrise M., DeMol C., Dhadwal H. S. J. *Opt. Soc. Am.* **73** 1470 (1983)
26. Мину М. *Математическое программирование* (М.:Наука, 1990)
27. Siegman A. E. *Opt. Lett.* **1**(1) 13 (1977)
28. Magni V., Gerullo G., Desilvestry S. J. *Opt. Soc. Am. A* **9**(11) 2031 (1992)
29. Khonina S. N., Kotlyar V. V., Soifer V. *Optik* **88**(4) 182 (1991)
30. Doskolovich L. L., Khonina S. N., Kotlyar V. V., Soifer V. A. *Opt. & Quant. Electr.* **25** 801 (1993)
31. Березный А. Е., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Доклады АН СССР* **274**(4) 802 (1984)
32. Khonina S. N., Kotlyar V. V., Shinkaryev M. V., Soifer V. A., Uspleniev G. V. *Opt. Commun.* **91**(3–4) 158 (1992)
33. Котляр В. В., Сойфер В. А. *Письма в ЖТФ* **16** (12) 30 (1990)
34. Khonina S. N., Kotlyar V. V., Shinkaryev M. V., Soifer V. A., Uspleniev G. V *J. Mod. Opt.* **39**(5) 1147 (1992)
35. McLeod J. H. *J. Opt. Soc. Am.* **44**(8) 592 (1954)
36. Belanger P., Rioux M. *Appl. Opt.* **17**(7) 1080 (1978)
37. Голуб М. А., Карпеев С. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Письма в ЖТФ* **7**(10) 618 (1981)
38. Tremblay R., D'Astons Y., Roy G., Blanshard M. *Opt. Commun.* **28**(2) 193 (1979)
39. Brenden B. B., Russel J. T. *Appl. Opt.* **23**(19) 3250 (1984)
40. Michaltsova I. A., Nalivaiko V. I., Soldatenkov I. S. *Optik* **67**(3) 267 (1984)
41. Khonina S. N., Kotlyar V. V., Soifer V A. *J. Mod. Opt.* **40**(5) 761 (1993)
42. Khonina S. N., Kotlyar V. V., Soifer V A. *Proceedings of SPIE* **1780** 263 (1993)
43. Khonina S. N., Kotlyar V. V., Soifer V. A. *Computer Optics* **13** 12 (1993)
44. Kotlyar V. V., Khonina S. N., Soifer V A. *Opt. and Lasers in Technology* **27**(4) 235 (1995)
45. Khonina S. N., Kotlyar V. V., Philippov S. V., Nikolsky I. V., Soifer V. A. *Opt. Design Conf.* **22** 251 (1994)
46. Damman H., Klotz E. *Opt. Acta* **24** 505 (1977)
47. Mait J. N., Brenner K. *Appl. Opt.* **27** 1692 (1988)
48. Veldkamp W. B., Leger J. R., Swanson G. J. *Opt. Lett.* **11** 303 (1986)
49. Simpson M. J. *Appl. Opt.* **26**(9) 1786 (1987)
50. Killat U., Rabe G., Rave W. *Fiber Integ. Opt.* **4** 159 (1982)
51. Arsenault H. H., Szoplik T., Macukow B. *Optical processing and computing* (San Diego:Academic Press, 1989)
52. Досколович Л. Л., Сойфер В. А., Шинкарев М. С. *Автометрия* **3** 104 (1990)
53. Seldowitz M. A., Allebach J. P., Sweeney D. W. *Appl. Opt.* **26** 2788 (1987)
54. Feldman M. R., Gest C. C. *Opt. Lett.* **14**(8) 479 (1989)
55. Turunen J., Vasara A., Westerholm J. *Opt. Eng.* **28**(11) 1162 (1989)
56. Morrison R. L., Walker S. L., Cloonan T. J. *Appl. Opt.* **32** 2512 (1993)
57. Turunen J., Vasara A., Westerholm J. *Opt. Commun.* **74** 245 (1990)
58. Mait J. N. *J. Opt. Soc. Am. A* **7**(8) 1514 (1990)
59. Березный А. Е., Комаров С. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Доклады АН СССР* **287**(3) 623 (1986)
60. Досколович Л. Л., Котляр В. В., Сойфер В. А. *Письма в ЖТФ* **17**(21) 54 (1991)
61. Dammann H., Gortler K. *Opt. Commun.* **3**(5) 312 (1971)
62. Jahns J., Downs M. M., Prise M. E., Streibl N., Walker S. J. *Opt. Eng.* **28** 1267 (1988)
63. Бобров С. Т., Туркевич Ю. Г. *Компьютерная оптика* **4** 38 (1989)
64. Bobrov S., Kotletsov B., Turkevich Y. *Proc. SPIE* **1751** 154 (1992)
65. Pommet A. D., Moharam M. G., Grann E. B. *J. Opt. Soc. Am. A* **11**(6) 1827 (1994)
66. Golub M. A., Sisakyan I. N., Soifer V. A. *Opt. & Las. in Engin.* **15**(5) 297 (1991)

67. Soifer V., Golub M. *Proceedings of SPIE* **1751** 140 (1992)
68. Doskolovich L. L., Soifer V. A., Alessandretti G., Perlo P., Repetto M. *Pure & Appl. Opt.* **3** 921 (1994)
69. Doskolovich L. L., Repetto M., Perlo P., Lavagno L., Sinesi S. *Proc. ISHM «The many aspects of microelectronics»* (1994)
70. Карманов В. Г. *Математическое программирование* (М.:Наука, 1986)
71. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С И *Компьютерная оптика* **10–11** 100 (1992)
72. Leseberg D. *Appl. Opt.* **31**(2) 223 (1992)
73. Воронцов М. А., Корябин А. В., Шмальгаузен В. И. *Управляемые оптические системы* (М.:Наука, 1988)
74. Bryngdahl O., Wyrowski F. *Progress in Optics* **28** 3 (1990)
75. Досколович Л. Л., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **12** 8 (1992)
76. Kotlyar V. V., Khonina S. N., Soifer V. A. *Journal of Modern optics* **43**(7) 1509 (1996)
77. Kotlyar V. V., Nikolsky I. V., Soifer V A. *Optik* **98**(1) 26 (1994)
78. Kazanskiy N. L., Kotlyar V V., Soifer V A. *Opt. Eng.* **33**(10) 3156 (1994)
79. Khonina S. N., Kotlyar V. V., Soifer V. A. *J. Mod. Opt.* **42**(6) 1231 (1995)
80. Golub M. A., Doskolovich L. L., Kazansky N. L., Kharitonov S. I., Soifer V. A. *J. Mod. Opt* **39**(6) 1245 (1992)
81. Khonina S. N., Kotlyar V. V., Soifer V. A. *Proceedings of SPIE* **1780** 263 (1993)
82. Прудников А. П., Брычков Ю. А., Маричев О. И. *Интегралы и ряды. Специальные функции* (М.:Наука, 1983)
83. Veldcamp W. P. *Appl. Opt.* **21**(17) 3209 (1981)
84. Han C. Y., Ishii Y., Murata K. *Appl.Opt.* **22**(22) 3644 (1983)
85. Коронкевич В. П., Ленкова Г. А., Михальцова И. А., Пальчикова И. Г., Полещук А. Г., Седухин А. Г., Чурин Е. Г., Юрлов Ю. И. *Автометрия* **1** 4 (1985)
86. Воронцов М. А., Матвеев А. И., Сивоконь В. П. *Компьютерная оптика* **1** 74 (1987)
87. Roberts N. C. *Appl.Opt.* **28** 31 (1989)
88. Eismann M. T., Tai A. M., Cederquist J. N. *Appl.Opt.* **28**(15) 2641 (1989)
89. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **7** 42 (1990)
90. Aleksoff C. C., Ellis K. K., Neagle S. D. *Opt. Eng.* **30**(5) 537 (1991)
91. Golub M. A., Sisakyan I. N., Soifer V A. *Opt. and Lasers in Engin.* **15**(5) 297 (1991)
92. Golub M. A., Doskolovich L. L., Kotlyar V. V., Nikolsky I. V., Soifer V. A. *Компьютерная оптика* **13** 30 (1993)
93. Cordingley J. *Appl. Opt.* **32**(14) 2538 (1994)
94. Седухин А. Г., Чурин Е. Г. *Автометрия* **6** 75 (1995)
95. Duparre M., Golub M. A., Ludge B., Pavelyev V. S., Soifer V. A., Uspleniev G. V., Volotovskii S. G. *Appl. Opt.* **34**(14) 2488 (1995)
96. Golub M. A., Duparre M., Kley E., Kowarschik R., Ludge B., Rockstroh W., Fuchs H. *Opt. Eng.* **35**(5) 1400 (1996)
97. Farn M. W. **22** 246 (1994)
98. Воронцов М. А., Шмальгаузен В. И. *Принципы адаптивной оптики* (М.:Наука, 1985)
99. Fedotovsky A., Lehovec H. *Appl. Opt.* **13**(12) 2919 (1974)
100. Пальчикова И. Г. *Компьютерная оптика* **6** 9 (1989)
101. Doskolovich L. L., Kasanskiy N. L., Kharitonov S. I., Uspleniev G. V *Opt. and Lasers in Engineering* **15**(5) 311 (1991)
102. Khonina S. N., Kotlyar V. V., Lushpin V. V., Soifer V. A. *Optical Memory and Neural Networks* **6**(3) 213 (1997)
103. Soifer V. A., Kotlyar V. V., Doskolovich L. L. *Iterative methods for diffractive optical elements computation* (Taylor & Francis, 1997)
104. Wyrowski F. *Appl.Opt.* **28** 3864 (1989)
105. Weissbach S., Wirowski F., Bringdahl O. *Opt. Comm.* **72**(1,2) 37 (1982)
106. Wyrowski F., Bryngdahl O. *In Rep. Prog. Phys.* **28** 1481 (1991)

107. Doskolovich L. L., Perlo P., Petrova O. I., Repetto P., Soifer V. A. *J. Mod.Opt.* **44**(4) 685 (1997)
108. Doskolovich L. L., Perlo P., Petrova O. I., Repetto P., Soifer V. A. *Int. Jour. of Optoelectronics* **10**(4) 243 (1995)
109. Досколович Л. Л., Казанский Н. Л., Сойфер В. А. *Автометрия* **5** 42 (1995)
110. Repetto P., Doskolovich L., Bigliati C., Pierro P *Workshop on Diffractive Optics and Microoptics* **7** (1998)
111. Dallas J. M. *Appl. Opt.* **14** 674 (1971)
112. Goodman J. M., Silvestri A. M. *IBM Journ. Res. Develop.* **14** 478 (1969)
113. Ланцош К. *Практические методы прикладного анализа.* (М.:Гос. изд-во физ.-мат. Литературы, 1961)
114. Оппенгейм А. В., Шафер Р. В. *Цифровая обработка сигналов* (М.:Связь, 1979)
115. Старк Г. *Теория и измерение оптических фурье-спектров. В книге: Применение методов фурье-оптики* (М.: Радио и Связь, 1988)
116. Sisakyan I. N., Soifer V. A. *Infrared Physics* **32** 435 (1991)
117. Soifer V. A., Golub M. A. *Laser beam mode selection by computer generated holograms* (*CRC Press, Boca Raton*, 1994)
118. Noel E., Khan R. R., Dhadwal H. S. *Opto. Eng.* **32**(11) 2866 (1993)

ГЛАВА 3

ЭЛЕКТРОМАГНИТНЫЙ ПОДХОД К РАСЧЕТУ ДИФРАКЦИОННЫХ ОПТИЧЕСКИХ ЭЛЕМЕНТОВ

Если характерные параметры среды (средние глубина и период модуляции рельефа), на которой происходит дифракция света, сравнимы с длиной волны, то скалярное приближение, рассмотренное в главе 2, утрачивает свою применимость. Для того чтобы адекватно описать дифракцию света в этом случае требуется использовать общую электромагнитную теорию света, основанную на уравнениях Максвелла. Краткое содержание электромагнитной теории дифракции света можно найти в работах [1–3, 8]. В данной главе приведены решения конкретных прямых и обратных задач дифракции в рамках строгой теории.

3.1. Дифракция на отражающих решетках со ступенчатым профилем

Уравнения Максвелла для электромагнитного поля в гауссовой системе единиц имеют следующий вид [8]:

$$\operatorname{rot}\mathbf{H} = \frac{4\pi\mathbf{j}}{c} + \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t}; \tag{3.1}$$

$$\operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t}; \tag{3.2}$$

$$\operatorname{div}\mathbf{D} = 4\pi\rho; \tag{3.3}$$

$$\operatorname{div}\mathbf{B} = 0, \tag{3.4}$$

где $\mathbf{E}, \mathbf{H}$ — напряженности электрического и магнитного полей, соответственно; $\mathbf{D}$ — индукция (электрическое смещение) электрического поля; $\mathbf{B}$ — индукция магнитного поля; $\mathbf{j}$, ρ — плотности токов и зарядов.

Материальные уравнения для изотропной среды имеют следующий вид: $\mathbf{j} = \sigma\mathbf{E}$, $\mathbf{D} = \varepsilon\mathbf{E}$ и $\mathbf{B} = \mu\mathbf{H}$, где σ — удельная проводимость, ε, μ — диэлектрическая и магнитная проницаемости среды.

Для монохроматического поля, пренебрегая проводимостью среды, получаем:

$$\operatorname{rot}\mathbf{H} = \frac{4\pi j}{c} - \frac{i\omega}{c}\varepsilon\mathbf{E}, \tag{3.5}$$

$$\operatorname{rot}\mathbf{E} = \frac{i\omega}{c}\mathbf{H}, \tag{3.6}$$

где ω — циклическая частота.

Рассмотрим дифракционную решетку, изготовленную из абсолютно проводящего материала и имеющую K прямоугольных штрихов на периоде d (рис. 3.1). Положение l-го штриха определяется координатой x_l начала штриха, шириной штриха c_l и

глубиной штриха h_l. Предположим, что дифракционная решетка освещается плоской монохроматической волной с единичной амплитудой и волновым вектором $\mathbf{k} = k(\sin(\theta)\cos(\gamma_0), -\cos(\theta)\cos(\gamma_0), \sin(\gamma_0))$, где $k = 2\pi/\lambda$, а углы θ и γ_0 определяют направление падающей волны (рис. 3.1).

Зависимость от переменной z напряженностей электрического и магнитного полей, сформированных при отражении плоской волны от решетки, описывается функцией $\exp(ik\gamma z)$ [4, 5]:

$$\begin{aligned} \mathbf{E}(x,y,z) &= \mathbf{E}(x,y)\exp(ik\gamma z), \\ \mathbf{H}(x,y,z) &= \mathbf{H}(x,y)\exp(ik\gamma z), \end{aligned} \tag{3.7}$$

где $\gamma = \sin(\gamma_0)$. Подставив (3.7) в уравнения (3.5) и (3.6), получим уравнения Максвелла для проекций электрического и магнитного векторов монохроматической волны для свободного пространства:

$$\begin{aligned} &\partial_y E_z - ik\gamma E_y = ikH_x; \quad \partial_y H_z - ik\gamma H_y = -ik\varepsilon E_x; \quad ik\gamma E_x - \partial_x E_z = ikH_y; \\ &\partial_x E_y - \partial_y E_x = ikH_z; \quad \partial_x H_y - \partial_y H_x = -ik\varepsilon E_z; \quad ik\gamma H_x - \partial_x H_z = -ik\varepsilon E_y; \end{aligned} \tag{3.8}$$

где $\partial_x H_z = \partial H_z/\partial x$ и т. п. Согласно (3.8), компоненты векторов $\mathbf{E}(x,y)$, $\mathbf{H}(x,y)$ выражаются через компоненты $E_z(x,y), H_z(x,y)$ следующим образом:

$$\begin{aligned} E_x &= \frac{-1}{ik - ik\gamma^2}\left(\partial_y H_z + \gamma\partial_x E_z\right), \quad H_x = \frac{1}{ik - ik\gamma^2}\left(\partial_y E_z - \gamma\partial_x H_z\right), \\ E_y &= \frac{1}{ik - ik\gamma^2}\left(\partial_x H_z - \gamma\partial_y E_z\right), \quad H_y = \frac{-1}{ik - ik\gamma^2}\left(\partial_x E_z + \gamma\partial_y H_z\right), \end{aligned} \tag{3.9}$$

где $\partial_x = \partial/\partial x$, а компоненты $E_z(x,y), H_z(x,y)$ удовлетворяют уравнению Гельмгольца:

$$\begin{aligned} \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + k^2(1-\gamma^2)E_z = 0, \\ \frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + k^2(1-\gamma^2)H_z = 0. \end{aligned} \tag{3.10}$$

Согласно (3.9), (3.10), задача дифракции плоской волны на одномерной дифракционной решетке сводится к рассмотрению двух независимых задач: задачи дифракции плоской волны с ТЕ-поляризацией ($E_z \neq 0, H_z = 0$) и задачи дифракции плоской волны с ТМ-поляризацией ($H_z \neq 0, E_z = 0$) [4, 5]. При этом произвольная

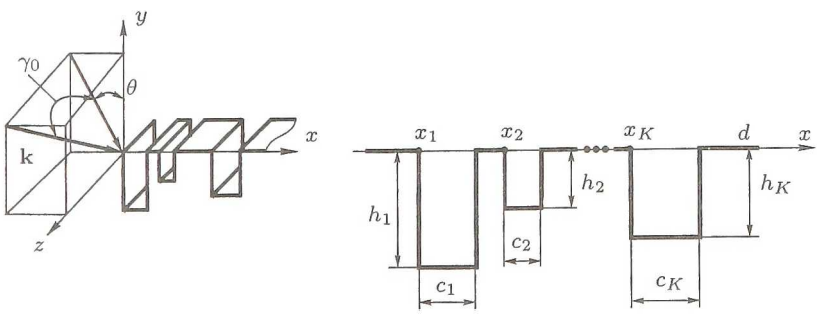


Рис. 3.1. Дифракционная решетка со ступенчатым профилем

плоская волна может быть представлена в виде линейной суперпозиции волн этих двух типов. Уравнения Гельмгольца (3.10) должны быть дополнены граничными условиями. Граничное условие на поверхности абсолютного проводника имеет вид

$$E_t = 0, \tag{3.11}$$

где E_t — тангенциальная составляющая электрического поля. Условие (3.11) с учетом (3.9) принимает вид

$$\begin{aligned} &E_x \sim \partial_y H_z + \gamma \partial_x E_z = 0, \\ &E_z = 0, \end{aligned} \tag{3.12}$$

для горизонтальных участков профиля решетки и

$$\begin{aligned} &E_z = 0, \\ &E_y \sim \partial_x H_z - \gamma \partial_y E_z = 0 \end{aligned} \tag{3.13}$$

для вертикальных участков профиля решетки. Кроме того, компоненты электрического поля должны быть непрерывны при $y = 0$. Граничные условия (3.12), (3.13) должны быть также дополнены условием непрерывности тангенциальных компонентов H_x, H_z магнитного поля при $y = 0$ в области штрихов, то есть при $x_i \leqslant x \leqslant x_i + c_i, i = \overline{1, K}$ [4–7]. Компоненты магнитного поля при $y = 0$ между штрихами, в отличие от компонентов электрического поля, терпят разрыв, равный плотности поверхностного тока.

В дальнейшем изложении ограничимся рассмотрением случая $\gamma = 0$, соответствующего «плоскому падению». При $\gamma = 0$ для ТЕ-поляризации $E_z \neq 0$, $E_x = E_y = H_z = 0$, а для ТМ-поляризации $H_z \neq 0$, $H_x = H_y = E_z = 0$.

Рассмотрим решение задачи дифракции по методу модового разложения [4–7]. Метод основан на сшивке при $y = 0$ компонентов двух полей. Первое решение является решением уравнения Гельмгольца внутри штрихов при $x_i \leqslant x \leqslant x_i + c_i$, $-h_i \leqslant y < 0$, $i = \overline{1, K}$, а второе — решением уравнения Гельмгольца вне штрихов дифракционной решетки при $y > 0$.

Решение вне штрихов представляется в виде разложения Рэлея:

$$u(x, y) = \exp(ik(\alpha_0 x - \beta_0 y)) + \sum_{n=-\infty}^{\infty} R_n \exp(ik(\alpha_n x + \beta_n y)), \tag{3.14}$$

где

$$\alpha_n = \sin(\theta) + n\lambda/d, \tag{3.15}$$

$$\beta_n = \sqrt{1 - \alpha_n^2}. \tag{3.16}$$

Скалярная функция $u(x, y)$ соответствует компоненте $E_z(x, y)$ для ТЕ-поляризации и компоненте $H_z(x, y)$ для ТМ-поляризации. Разложение Рэлея (3.14) является решением уравнения Гельмгольца и содержит однородные плоские волны $(\alpha_n^2 < 1)$ и неоднородные плоские волны $(\alpha_n^2 > 1)$, экспоненциально-затухающие при удалении от поверхности дифракционной решетки. Слагаемое $n\lambda/d$ в (3.15) соответствует теореме Флоке и характеризует наличие постоянного фазового сдвига между соседними периодами решетки.

Представление для компонент поля внутри штрихов дифракционной решетки различно для ТЕ- и ТМ-поляризаций.

ТЕ-поляризация. Для ТЕ-поляризации электрическое поле $E_z^l(x,y)$ внутри l-го штриха удовлетворяет уравнению Гельмгольца

$$\Delta E_z^l + k^2 E_z^l = 0 \tag{3.17}$$

с граничным условием

$$E_z^l(x,y)\big|_{(x,y)\in C_l} = 0, \tag{3.18}$$

где C_l — контур штриха, состоящий из трех отрезков $x = x_l$, $y \in [-h_l, 0]$, $x = x_l + c_l$, $y \in [-h_l, 0]$ и $y = -h_l$, $x \in [x_l, x_l + c_l]$. Условие (3.18) является конкретизацией условий (3.12), (3.13). Используя метод разделения переменных для решения (3.17), получим поле внутри l-го штриха в виде следующего разложения по ортогональным функциям:

$$E_z^l(x,y) = \sum_{n=1}^{\infty} B_n^l \sin\left(\frac{\pi n}{c_l}(x - x_l)\right) \sin\left(\mu_{\ln}(y + h_l)\right), \tag{3.19}$$

где

$$\mu_{\ln}^2 = k^2 - \left(\frac{\pi n}{c_l}\right)^2. \tag{3.20}$$

Дальнейшее решение задачи дифракции сводится к нахождению коэффициентов R_n и B_n^l из граничных условий при $y = 0$. Запишем условие непрерывности $E_z(x,y)$ при $y = 0$:

$$\exp(ik\alpha_0 x) + \sum_{n=-\infty}^{\infty} R_n \exp(ik\alpha_n x) = \\ = \sum_{l=1}^{K} \operatorname{rect}\left(\frac{1}{c_l}(x - x_l - c_l/2)\right) \sum_{n=1}^{\infty} B_n^l \sin\left(\frac{\pi n}{c_l}(x - x_l)\right) \sin(\mu_{\ln} h_l), \tag{3.21}$$

где $\operatorname{rect}(x) = \begin{cases} 1, & |x| \leqslant 0{,}5 \\ 0, & |x| > 0{,}5 \end{cases}$.

Умножая уравнение (3.21) на $\exp(-ik\alpha_p x)$, $(p = -\infty, \infty)$ и интегрируя по периоду, заменим (3.21) условием равенства коэффициентов Фурье в разложении правой и левой частей уравнения. В результате получим следующую систему линейных уравнений:

$$R_p + \delta(p) = \sum_{l=1}^{K} \sum_{n=1}^{\infty} A_{pn}^l B_n^l, \quad (p = -\infty, \infty), \tag{3.22}$$

где $\delta(p) = \begin{cases} 1, & p = 0, \\ 0, & p \neq 0, \end{cases}$

$$A_{pm}^l = \frac{\sin(\mu_{lm} h_l)}{kd} \exp(-ik\alpha_p x_l) \int_0^{kc_l} \sin\left(\frac{\pi m}{kc_l}\xi\right) \exp(-i\alpha_p \xi)\, \mathrm{d}\xi. \tag{3.23}$$

Система линейных уравнений (3.22) содержит две последовательности коэффициентов R_n и B_n^l. Для того, чтобы найти второе уравнение связи между наборами коэффициентов R_n и B_n^l воспользуемся условием непрерывности для тангенциальных составляющих магнитного поля. Тангенциальная составляющая $H_x \sim \partial E_z/\partial y$ должна быть непрерывной при $y = 0$ в области штрихов, то есть при $x_l \leqslant x \leqslant x_l + c_l$, $l = \overline{1, K}$. Применяя при $y = 0$ условие непрерывности $\partial E_z/\partial y$, получаем

$$-ik\beta_0 \exp(ik\alpha_0 x) + \sum_{n=-\infty}^{\infty} R_n ik\beta_n \exp(ik\alpha_n x) = \\ = \sum_{m=1}^{\infty} B_m^l \mu_{lm} \sin\left(\frac{\pi m}{c_l}(x - x_l)\right) \cos(\mu_{lm} h_l), \quad (3.24)$$

при $x_l \leqslant x \leqslant x_l + c_l$, $l = \overline{1, K}$. Умножая (3.24) на $\sin(m(x - x_l)/c_l)$, $m = -\infty, \infty$ и интегрируя по x на отрезке $[x_l, x_l + c_l]$, получаем выражение для коэффициентов разложения внутри l-го штриха в виде

$$B_m^l = \frac{2ik}{\mu_{lm} \cos(\mu_{lm} h_l)} \sum_{n=-\infty}^{\infty} \beta_n (R_n - \delta(n)) D_{mn}^l, \quad (3.25)$$

где

$$D_{mn}^l = \frac{1}{kc_l} \exp(ik\alpha_n x_l) \int_0^{kc_l} \sin\left(\frac{\pi m}{kc_l}\xi\right) \exp(i\alpha_n \xi)\, d\xi. \quad (3.26)$$

Подставляя (3.25) и (3.26) в (3.22), получаем систему линейных уравнений для нахождения коэффициентов Рэлея:

$$\sum_{s=-\infty}^{\infty} A_{ps} \cdot R_s = D_p, \quad p = -\infty, \infty, \quad (3.27)$$

где

$$A_{ps} = \beta_s M_{ps} - \delta(p - s), \quad (3.28)$$

$$M_{ps} = \frac{2ik}{d} \sum_{l=1}^{K} \sum_{n=1}^{\infty} c_l \frac{\operatorname{tg}(\mu_{ln} h_l)}{\mu_{ln}} D_{ns}^l \left(D_{np}^l\right)^*, \quad (3.29)$$

$$D_p = \beta_0 M_{p0} + \delta(p). \quad (3.30)$$

При расчетах система бесконечного числа линейных уравнений (3.27) заменяется конечной системой из $2N + 1$ уравнений в матричном виде:

$$A \cdot \mathbf{R} = \mathbf{D}, \quad (3.31)$$

где

$$A = [A_{ps}]_{-N}^{N}, \quad \mathbf{R} = (R_p)_{-N}^{N}, \quad \mathbf{D} = (D_p)_{-N}^{N}; \quad (3.32)$$

$\mathbf{R}$ и $\mathbf{D}$ — векторы с компонентами (3.32). При этом ряды в (3.29) заменяются суммами из n_l, $l = \overline{1, K}$ членов. Следует отметить, что замена l-го ряда в (3.29) суммой n_l членов предполагает, что поле внутри l-го штриха (3.19) хорошо описывается суммой n_l членов.

ТМ-поляризация. Случай ТМ-поляризации во многом аналогичен рассмотренному — для ТЕ-поляризации. Внутри l-го штриха дифракционной решетки магнитное поле $H_z^l(x,y)$ удовлетворяет уравнению Гельмгольца:

$$\Delta H_z^l + k^2 H_z^l = 0 \tag{3.33}$$

с граничным условием

$$\left.\frac{\partial H_z^l(x,y)}{\partial \mathbf{n}}\right|_{(x,y)\in C_l} = 0, \tag{3.34}$$

где $\mathbf{n}$ — вектор нормали к контуру штриха C_l. Условия (3.34) являются следствием общих граничных условий (3.12), (3.13). Используя метод разделения переменных для уравнения Гельмгольца (3.33), получаем $H_z^l(x,y)$ в виде следующего разложения:

$$H_z^l(x,y) = \sum_{n=1}^{\infty} B_n^l \cos\left(\frac{\pi n}{c_l}(x-x_l)\right)\cos\left(\mu_{ln}(y+h_l)\right), \tag{3.35}$$

где коэффициенты μ_{ln} определены в уравнении (3.20).

Для нахождения коэффициентов R_n и B_n^l разложения компонент поля внутри штрихов, воспользуемся условиями «сшивки» при $y=0$. Условие непрерывности $H_z^l(x,y)$ при $y=0$ в области штрихов решетки дает следующее функциональное уравнение:

$$\exp(ik\alpha_0 x) + \sum_{n=-\infty}^{\infty} R_n \exp(ik\alpha_n x) = \sum_{n=1}^{\infty} B_n^l \cos\left(\frac{\pi n}{c_l}(x-x_l)\right)\cos(\mu_{ln}h_l), \tag{3.36}$$

где $x_l \leqslant x \leqslant x_l + c_l$ и $l = \overline{1,K}$. Граничное условие для тангенциальной составляющей электрического поля (в данном случае для $E_x \sim \partial H_z/\partial y$) требует при $y=0$ равенства нулю производной $\partial H_z/\partial y$ на горизонтальных участках профиля и непрерывности производной $\partial H_z/\partial y$ в области штрихов (при $x_l \leqslant x \leqslant x_l + c_l$, $l = \overline{1,K}$). Данные условия дают второе функциональное уравнение для коэффициентов R_n и B_n^l в виде

$$\begin{aligned} &-ik\beta_0 \exp(ik\alpha_0 x) + \sum_{n=-\infty}^{\infty} R_n ik\beta_n \exp(ik\alpha_n x) = \\ &= -\sum_{l=1}^{K} \operatorname{rect}\left(\frac{1}{c_l}(x - x_l - c_l/2)\right) \sum_{m=1}^{\infty} B_m^l \mu_{lm} \cos\left(\frac{\pi m}{c_l}(x-x_l)\right)\sin(\mu_{lm}h_l). \end{aligned} \tag{3.37}$$

Умножая (3.36) на $\cos(\pi m(x-x_l)/c_l)$, $m = -\infty, \infty$ и интегрируя по x на отрезке $[x_l, x_l + c_l]$, получаем следующее уравнение для коэффициентов разложения внутри l-го штриха:

$$B_m^l = \frac{2}{\cos(\mu_{lm}h_l)} \sum_{n=-\infty}^{\infty} (R_n + \delta(n)) D_{mn}^l, \tag{3.38}$$

где

$$D_{mn}^l = \frac{1}{kc_l} \exp(ik\alpha_n x_l) \int_0^{kc_l} \cos\left(\frac{\pi m}{kc_l}\xi\right) \exp(i\alpha_n \xi)\, \mathrm{d}\xi. \tag{3.39}$$

Далее, заменяя (3.37) условием равенства коэффициентов Фурье для правой и левой частей, получаем следующую систему линейных уравнений:

$$-\beta_0\delta(p)+\beta_p R_p = i\sum_{l=1}^{K}\sum_{m=1}^{\infty} B_m^l A_{pm}^l, \quad p=-\infty,\infty, \tag{3.40}$$

$$A_{pm}^l = \frac{\sin(\mu_{lm}h_l)}{kd}\left(\frac{\mu_{lm}}{k}\right)\exp(-ik\alpha_p x_l)\int_0^{kc_l}\cos\left(\frac{\pi m}{kc_l}\xi\right)\exp(-i\alpha_p\xi)\,\mathrm{d}\xi. \tag{3.41}$$

Подставляя (3.38), (3.39) в (3.40), получаем систему линейных уравнений относительно коэффициентов Рэлея в виде (3.27), (3.31), где

$$A_{ps} = M_{ps} - \beta_p\delta(p-s), \tag{3.42}$$

$$M_{ps} = \frac{2ki}{d}\sum_{l=1}^{K}\sum_{n=1}^{\infty}\frac{c_l}{\mu_{ln}}\operatorname{tg}(\mu_{ln}h_l)\,D_{ns}^l\left(D_{np}^l\right)^*, \tag{3.43}$$

$$D_p = -M_{p0} - \beta_0\delta(p). \tag{3.44}$$

В приближении Кирхгофа под «интенсивностями дифракционных порядков» понимаются квадраты модулей коэффициентов Фурье $|c_n|^2$ в разложении функции $\exp(i\varphi(x))$, где функция $\varphi(x)$ описывает набег фазы при отражении плоской волны от профиля решетки. Определим понятие «интенсивность порядка» для отражающей решетки в рамках электромагнитной теории. Рассмотрим область D ограниченную снизу профилем решетки, сверху отрезком прямой $x=p$, $p>0$, справа и слева — отрезками прямых $x=0$, $x=d$. Используя закон сохранения энергии, приравняем к нулю поток вектора Умова–Пойнтинга $\mathbf{S}=(c/8\pi)re\,[\mathbf{E},\mathbf{H}^*]$ через область D [8]. В результате получим следующее условие нормировки:

$$\beta_0 = \sum_{n\in U}|R_n|^2\beta_n, \tag{3.45}$$

где $U=\{n|\,\alpha_n^2<1\}$ — множество индексов, соответствующих распространяющимся отраженным волнам (порядкам) в (3.14), а β_n определены в (3.16) и равны косинусам углов между осью Oy и направлениями отраженных волн (порядков). Уравнение (3.45) имеет ясную физическую интерпретацию: энергия падающей волны равна сумме энергий отраженных порядков. Согласно (3.45), под интенсивностями порядков следует понимать следующие нормированные значения коэффициентов Рэлея:

$$I_n = |R_n|^2\frac{\beta_n}{\beta_0}, \quad \left(\sum_{n\in U} I_n = 1\right). \tag{3.46}$$

Пример 3.1. Рассмотрим работу простейшей бинарной решетки с одним штрихом длиной $d/2$ на периоде d. На рис. 3.2 приведены зависимости интенсивностей 0-го и -1-го порядков от глубины штриха h. Интенсивности порядков рассчитывались по формулам (3.31), (3.42)–(3.44) при периоде $d=\lambda$ и угле падения $\theta=30°$ для TM-поляризации. Отметим, что при указанных параметрах выполняется условие Брэгга [7] $m\lambda=-2d\sin\theta$ при $m=-1$. В этом случае существуют только 0-й и -1-й распространяющиеся порядки. При этом направление распространения -1-го порядка противоположно направлению падающей волны. Все остальные порядки

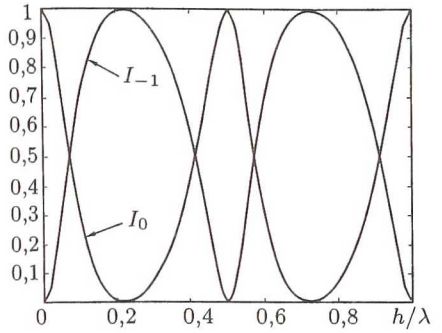


Рис. 3.2. Зависимость интенсивностей -1-го (I_{-1}) и 0-го (I_0) порядков бинарной решетки от высоты штриха для TM-поляризации при $d = \lambda$, $\theta = 30°$

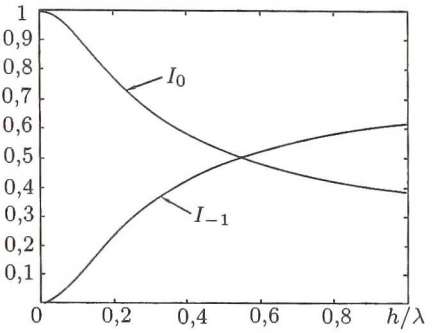


Рис. 3.3. Зависимость интенсивностей -1-го (I_{-1}) и 0-го (I_0) порядков бинарной решетки от высоты штриха для TE-поляризации при $d = \lambda$, $\theta = 30°$

соответствуют неоднородным волнам $\left(\alpha_n^2 > 1\right)$, экспоненциально затухающим при удалении от поверхности решетки. Рис. 3.2 показывает следующие интересные моменты в работе бинарной решетки. Во-первых, при глубинах штриха $h \approx 0{,}1\lambda$, $0{,}42\lambda$, $0{,}57\lambda$ и $0{,}93\lambda$ энергия делится пополам между 0-м и -1-м порядками, т. е. решетка работает как делитель пучка. Во-вторых, при глубинах штриха $0{,}22\lambda$ и $0{,}72\lambda$ энергия падающей волны концентрируется в -1-м порядке, т.е. при данных глубинах решетку можно использовать как дефлектор (отклонитель) пучка. По общему виду графиков можно предположить, что интенсивности 0-го и -1-го порядков меняются периодически. На рис. 3.3 приведена зависимость интенсивностей порядков бинарной решетки от глубины штриха для TE-поляризации. Для TE-поляризации бинарная решетка не обладает свойством концентрировать излучение в -1-м порядке. Как делитель пучка, решетка работает при глубине штриха $h \approx 0{,}55\lambda$.

3.2. Дифракция на отражающих решетках с непрерывным профилем

Прежде чем решать задачу дифракции плоской волны на одномерной идеально отражающей дифракционной решетке, приведем вывод интегрального уравнения, связывающего световое поле и плотность поверхностного тока на идеально проводящей поверхности.

Для получения интегральных уравнений используется векторная форма второй формулы Грина:

$$\int\limits_V (\mathbf{Q}\,\mathrm{rot}\,\mathrm{rot}\,\mathbf{P} - \mathbf{P}\,\mathrm{rot}\,\mathrm{rot}\,\mathbf{Q})\,\mathrm{d}V = \oint\limits_S \{[\mathbf{P}, \mathrm{rot}\,\mathbf{Q}] - [\mathbf{Q}, \mathrm{rot}\,\mathbf{P}]\}\,\mathbf{n}\,\mathrm{d}S. \tag{3.47}$$

Преобразуем подынтегральное выражение в (3.47) к виду

$$\mathbf{Q}\,\mathrm{rot}\,\mathrm{rot}\,\mathbf{P} - \mathbf{P}\,\mathrm{rot}\,\mathrm{rot}\,\mathbf{Q} = \mathbf{P}\Delta Q - \mathbf{Q}\Delta\mathbf{P} + \mathbf{Q}\,\mathrm{grad}\,\mathrm{div}\,\mathbf{P} - \mathbf{P}\,\mathrm{grad}\,\mathrm{div}\,\mathbf{Q} =$$

$$= \mathbf{P}\Delta\mathbf{Q} - \mathbf{Q}\Delta\mathbf{P} + \mathrm{div}\,(\mathbf{Q}\,\mathrm{div}\,\mathbf{P} - \mathbf{P}\,\mathrm{div}\,\mathbf{Q}).$$

Далее используя формулу Остроградского–Гаусса, получаем

$$\int\limits_V (\mathbf{P}\triangle\mathbf{Q} - \mathbf{Q}\triangle\mathbf{P})\,\mathrm{d}V = \oint\limits_S \{\mathbf{n}\,[\mathbf{P}, \operatorname{rot}\mathbf{Q}] - \mathbf{n}\,[\mathbf{Q}, \operatorname{rot}\mathbf{P}] + \mathbf{nP}\operatorname{div}\mathbf{Q} - \mathbf{nQ}\operatorname{div}\mathbf{P}\}\,\mathrm{d}S. \tag{3.48}$$

Из (3.47) и (3.48) нетрудно получить интегральные соотношения для электромагнитного поля в пространстве. Пусть $\mathbf{P} = \mathbf{E}$, $\mathbf{Q} = \mathbf{a}G$, $G(\mathbf{x} - \mathbf{x}_0) = \exp(ikR)/R$, $\mathbf{a}$ — единичный вектор произвольного направления, где $R = |\mathbf{x} - \mathbf{x}_0|$, $\mathbf{x}$ — радиус-вектор точки наблюдения, $\mathbf{x}_0$ — точка на поверхности S. В этом случае функция $\mathbf{Q}$ удовлетворяет уравнению

$$\triangle\mathbf{Q} + k^2\mathbf{Q} = -4\pi\mathbf{a}\delta(\mathbf{x} - \mathbf{x}_0), \tag{3.49}$$

где $\operatorname{rot}\mathbf{Q} = [\operatorname{grad}G, \mathbf{a}]$, $\operatorname{div}\mathbf{Q} = (\mathbf{a}, \operatorname{grad}G)$, $k^2 = \omega^2\bar{\varepsilon}\mu/c^2$.

Уравнения Максвелла (3.5) и (3.6) с учетом наличия магнитной проницаемости (и комплексной диэлектрической проницаемости $\bar{\varepsilon} = \varepsilon + (4\pi i\sigma/w)$, σ — удельная проводимость среды, имеют следующий вид:

$$\operatorname{rot}\mathbf{H} = +\frac{4\pi}{c}\mathrm{j} - \frac{i\omega}{c}\bar{\varepsilon}\mathbf{E}, \tag{3.50}$$

$$\operatorname{rot}\mathbf{E} = \frac{i\omega}{c}\mu\mathbf{H}. \tag{3.51}$$

Будем предполагать, что среда, в которой распространяется электромагнитное поле, однородная и изотропная. Применим операцию rot к правой и левой частям уравнения (3.51) и используя формулу $\operatorname{rot}\operatorname{rot}\mathbf{A} = \operatorname{grad}\operatorname{div}\mathbf{A} - \triangle\mathbf{A}$ и уравнение (3.50), получим

$$\begin{aligned} &\triangle\mathbf{E} + \frac{\omega^2\bar{\varepsilon}\mu}{c^2}\mathbf{E} = -4\pi\mathbf{J}_2, \\ &\mathbf{J}_2 = \frac{i\omega\mu}{c^2}\mathbf{j} - \frac{\operatorname{grad}\operatorname{div}\mathbf{j}}{i\omega\bar{\varepsilon}}, \end{aligned} \tag{3.52}$$

где $\mathbf{j}$ — плотность электрического тока.

Умножим уравнение (3.49) на $E(x)$, а уравнение для E (3.52) — на $\mathbf{Q}(\mathbf{x} - \mathbf{x}_0)$. Используя соотношение (3.48) и формулу $(\mathbf{n}, [\mathbf{E}, [\operatorname{grad}G, \mathbf{a}]]) = (\mathbf{a}, [\operatorname{grad}G, [\mathbf{E}, \mathbf{n}]])$, получаем выражение

$$\begin{aligned} \mathbf{E}(x)\mathbf{a} = \mathbf{a}\Bigg\{ &\int\limits_V \mathbf{J}_2(\mathbf{x}_0)G(\mathbf{x} - \mathbf{x}_0)\,\mathrm{d}V + \frac{1}{4\pi}\oint\limits_S \{\mathbf{n}G(\mathbf{x} - \mathbf{x}_0)\operatorname{div}_0\mathbf{E}(\mathbf{x}_0)\}\,\mathrm{d}S + \\ &+ \frac{1}{4\pi}\oint\limits_S \{[\operatorname{grad}_0 G(\mathbf{x} - \mathbf{x}_0), [\mathbf{n}, \mathbf{E}(\mathbf{x}_0)]] + [\operatorname{rot}_0\mathbf{E}(\mathbf{x}_0), \mathbf{n}]\,G(\mathbf{x} - \mathbf{x}_0) - \\ &- \operatorname{grad}_0 G(\mathbf{x} - \mathbf{x}_0)\,(\mathbf{n}, \mathbf{E}(\mathbf{x}_0))\}\,\mathrm{d}S\Bigg\}. \end{aligned} \tag{3.53}$$

Учитывая, что $\operatorname{div}\mathbf{E} = 0$, и произвольность вектора $\mathbf{a}$, получаем интегральное представление

$$\mathbf{E}(\mathbf{x}) = \mathbf{E}_0(\mathbf{x}) + \\ + \frac{1}{4\pi}\oint\limits_S \{[\operatorname{grad}_0 G(\mathbf{x}-\mathbf{x}_0), [\mathbf{n},\mathbf{E}]] + [\operatorname{rot}_0 \mathbf{E}, \mathbf{n}]\, G(\mathbf{x}-\mathbf{x}_0) - \\ - \operatorname{grad}_0 G(\mathbf{x}-\mathbf{x}_0)\,(\mathbf{n},\mathbf{E})\}\,\mathrm{d}S, \quad (3.54)$$

где

$$\mathbf{E}_0(\mathbf{x}) = \int\limits_V \mathbf{J}_2(\mathbf{x}_0) G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}V.$$

По аналогии с (3.2) можно получить интегральное представление для магнитного поля:

$$\mathbf{H}(\mathbf{x}) = \mathbf{H}_0(\mathbf{x}) + \frac{1}{4\pi} \times \\ \times \oint\limits_S \{[\operatorname{grad}_0 G(\mathbf{x}-\mathbf{x}_0)\,[\mathbf{n},\mathbf{H}(x_0)]] + [\operatorname{rot}_0 \mathbf{H}(\mathbf{x}_0), \mathbf{n}] G(\mathbf{x}-\mathbf{x}_0) - \\ - \operatorname{grad}_0 G(\mathbf{x}-\mathbf{x}_0)\,(\mathbf{n},\mathbf{H}(\mathbf{x}_0))\}\,\mathrm{d}S, \quad (3.55)$$

где $\mathbf{H}_0(\mathbf{x})$ — напряженность магнитного поля в падающей волне.

Уравнения (3.2), (3.55) называются формулами Стреттона–Чу.

Введем далее следующие обозначения для поверхностных плотностей электрического и магнитного токов: $(4\pi/c)\,\mathbf{j}_e(\mathbf{x}_0) = [\mathbf{n},\mathbf{H}(\mathbf{x}_0)]$, $(4\pi/c)\,\mathbf{j}_m(\mathbf{x}_0) = [\mathbf{n},\mathbf{E}(\mathbf{x}_0)]$. Учтем, что $\operatorname{grad} G(\mathbf{x}-\mathbf{x}_0) = -\operatorname{grad}_0 G(\mathbf{x}-\mathbf{x}_0)$, $\operatorname{rot}(\mathbf{F}) = \operatorname{rot}\mathbf{F} + [\operatorname{grad},\mathbf{F}]$, $\operatorname{rot} j_m(\mathbf{x}_0) = 0$, $\operatorname{rot}\mathbf{E} = ik\mathbf{H}$,

$$[\operatorname{grad}_0 G(\mathbf{x}-\mathbf{x}_0), \mathrm{j}_m(\mathbf{x}_0)] = -[\operatorname{grad} G(\mathbf{x}-\mathbf{x}_0), \mathrm{j}_m(\mathbf{x}_0)] = \\ = -\operatorname{rot}(G(\mathbf{x}-\mathbf{x}_0)\mathrm{j}_m(\mathbf{x}_0)) + G(\mathbf{x}-\mathbf{x}_0)\operatorname{rot}\mathrm{j}_m(\mathbf{x}_0).$$

В результате вместо уравнения (3.2) получим следующее представление для электрического поля:

$$\mathbf{E}(\mathbf{x}) = \mathbf{E}_0(\mathbf{x}) - \frac{1}{4\pi}\operatorname{rot}\oint\limits_S \mathrm{j}_m(\mathbf{x}_0) G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S - \\ - \frac{ik}{4\pi}\oint\limits_S \mathrm{j}_e(\mathbf{x}_0) G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S + \frac{1}{4\pi}\operatorname{grad}\oint\limits_S G(\mathbf{x}-\mathbf{x}_0)(\mathbf{n},\mathbf{E}(\mathbf{x}_0))\,\mathrm{d}S. \quad (3.56)$$

Применяя операцию rot к обеим частям равенства (3.56) и применяя формулу $\operatorname{rot}(\operatorname{grad}\varphi) = 0$, получим следующее выражение, содержащее только тангенциальные компоненты полей:

$$\mathbf{H}(\mathbf{x}) = \mathbf{H}_0(\mathbf{x}) - \frac{1}{4\pi ik}\operatorname{rot}\operatorname{rot}\oint\limits_S \mathbf{j}_m(\mathbf{x}_0) G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S - \frac{1}{4\pi}\operatorname{rot}\oint\limits_S \mathbf{j}_e(\mathbf{x}_0) G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S. \quad (3.57)$$

Из уравнения (3.56) по аналогии можно получить интегральное представление для электрического поля:

$$\mathbf{E}(\mathbf{x}) = \mathbf{E}_0(\mathbf{x}) + \frac{1}{4\pi ik}\operatorname{rot}\operatorname{rot}\oint\limits_S \mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S - \frac{1}{4\pi}\operatorname{rot}\oint\limits_S \mathbf{j}_m(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S. \tag{3.58}$$

Учитывая граничные условия на идеально проводящей поверхности $[\mathbf{n},\mathbf{E}] = 0$, $(\mathbf{n},\mathbf{H}) = 0$, уравнения (3.57) и (3.58) можно переписать в виде

$$\mathbf{H}(\mathbf{x}) = \mathbf{H}_0(\mathbf{x}) - \frac{1}{4\pi}\operatorname{rot}\oint\limits_S \mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S, \tag{3.59}$$

$$\mathbf{E}(\mathbf{x}) = \mathbf{E}_0(\mathbf{x}) + \frac{1}{4\pi ik}\operatorname{rot}\operatorname{rot}\oint\limits_S \mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S. \tag{3.60}$$

Для получения интегрального уравнения первого рода для плотности электрического тока на поверхности идеального проводника будем считать, что вектор $\mathbf{x}$ принадлежит поверхности. Умножим соотношение (3.60) на вектор нормали в точке $\mathbf{x}$ и учтем граничное условие для идеального проводника, тогда получим

$$[\mathbf{E}_0,(\mathbf{x})\mathbf{n}(\mathbf{x})] = -\frac{1}{4\pi ik}\operatorname{rot}\operatorname{rot}\oint\limits_S [\mathbf{j}_e(\mathbf{x}_0),\mathbf{n}(\mathbf{x})]\,G(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}S. \tag{3.61}$$

Итак, задача нахождения электромагнитного поля разбивается на два этапа:

1) решение интегрального уравнения (3.61) относительно $\mathbf{j}_e(\mathbf{x}_0)$;
2) вычисление компонентов поля по формулам (3.59) и (3.60).

Далее применим уравнение (3.61) для решения задачи дифракции плоской волны на цилиндрической поверхности, образующая которой параллельна оси z. Пусть вектор электрического поля направлен вдоль образующей цилиндрической поверхности (случай ТЕ-поляризации). Воспользуемся известными формулами векторного анализа:

$$\operatorname{rot}\operatorname{rot}(\mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)) = \operatorname{grad}\operatorname{div}(\mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)) - \Delta(\mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0));$$

$$\operatorname{div}(\mathbf{j}_e(\mathbf{x}_0)G(\mathbf{x}-\mathbf{x}_0)) = (\operatorname{grad}G(\mathbf{x}-\mathbf{x}_0), j_e(\mathbf{x}_0));$$

$$\operatorname{grad}(\operatorname{grad}G(\mathbf{x}-\mathbf{x}_0),\mathbf{j}_e(\mathbf{x}_0)) = (j_e(\mathbf{x}_0),\nabla)\operatorname{grad}G(\mathbf{x}-\mathbf{x}_0) + [j_e(\mathbf{x}_0),\operatorname{rot}\operatorname{grad}G(\mathbf{x}-\mathbf{x}_0)];$$

$$\Delta G(\mathbf{x}-\mathbf{x}_0) = -k^2 G(\mathbf{x}-\mathbf{x}_0).$$

Далее подставляя эти формулы в (3.61) и интегрируя вдоль образующей z цилиндрической поверхности, получаем

$$E_z(\mathbf{x}) = E_{0z}(\mathbf{x}) - ik\oint\limits_l j_{ze}(\mathbf{x}_0)G_2(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}l, \tag{3.62}$$

где $G_2(\mathbf{x}-\mathbf{x}_0) = (i/4)\,H_0^1(k\sqrt{(\mathbf{x}-\mathbf{x}_0)^2})$ — двумерная функция Грина, $H_0^1(x)$ — функция Ханкеля первого рода нулевого порядка, $\mathbf{x} = (x_1, x_2)$ — поперечный вектор. Интегральное уравнение для плотности тока на цилиндрической поверхности приобретает вид

$$E_{0z}(\mathbf{x}) = ik\oint\limits_l j_{ez}(\mathbf{x}_0)G_2(\mathbf{x}-\mathbf{x}_0)\,\mathrm{d}l, \tag{3.63}$$

где $\mathbf{x}$ — поперечный вектор, определяющий точку на поверхности. Интегрирование в (3.63) проводится вдоль образующей поверхности.

Далее для решения задачи дифракции на дифракционной решетке (период d) с непрерывным профилем $y = f(x)$ воспользуемся (3.63) с учетом выражения

$$\mathrm{d}l = \sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}\right)^2}\,\mathrm{d}x_0.$$

Решение задачи дифракции для случая ТЕ-поляризации состоит из следующих двух этапов:

1) решение интегрального уравнения относительно z-компоненты вектора плотности поверхностного тока (3.63):

$$E_{0z}(x, f(x)) = \frac{ik}{c}\int\limits_{-\infty}^{\infty} j_z(x_0) G_2(x - x_0, f(x_0) - f(x))\sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}\right)^2}\,\mathrm{d}x_0, \quad (3.64)$$

где $E_{0z}(x, y) = \exp(ik\alpha_0 x - ik\beta_0 y)$ — z-компонента вектора электрического поля падающей волны;

2) вычисление рассеянного поля:

$$E_z(x, y) = -\frac{ik}{c}\int\limits_{-\infty}^{\infty} j_z(x_0) G_2(x - x_0, y - f(x_0))\sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x}\right)^2}\,\mathrm{d}x_0, \quad (3.65)$$

где $G_2(x_0 - x, y_0 - y) = (i/4)\,H_0^1(k\sqrt{(x - x_0)^2 + (y - y_0)^2})$ — двумерная функция Грина.

В уравнениях (3.64), (3.65) интегрирование производится в бесконечных пределах. Получим интегральные представления, в которых интегрирование проводится только по периоду d. При дифракции на решетке плотность поверхностного тока является квазипериодической функцией:

$$j_z(x + nd) = \exp(ik\alpha_0 nd) j_z(x). \quad (3.66)$$

Используя квазипериодичность тока, представляем (3.65) в виде

$$E_z(x, y) = -\frac{ik}{c}\sum_{n=-\infty}^{\infty} \exp(ik\alpha_0 nd)\times$$
$$\times \int\limits_0^d j_z(x_0)\, G_2(x - x_0 - nd, y - f(x_o))\sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0}\right)^2}\,\mathrm{d}x_0. \quad (3.67)$$

В уравнении (3.67) интегрирование производится только по периоду d дифракционной решетки. Далее, используя известное интегральное представление функции Ханкеля,

$$H_0^1(k\sqrt{x^2 + y^2}) = \frac{1}{\pi}\int\limits_{-\infty}^{\infty} \frac{\exp\left(i\left(|y|\sqrt{k^2 - u^2} + ux\right)\right)}{\sqrt{k^2 - u^2}}\,\mathrm{d}u, \quad (3.68)$$

и формулу для функции «гребенки» из теории обобщенных функций,

$$\sum_{l=-\infty}^{\infty} \exp(i(x-y)la) = \frac{2\pi}{a} \sum_{m=-\infty}^{\infty} \delta\left(y - \left(x + \frac{2\pi m}{a}\right)\right), \tag{3.69}$$

получаем окончательное выражение для рассеянного поля в виде

$$E_z(x,y) = -\frac{ik}{c} \int_0^d j_z(x_0) G(x - x_0, y - f(x_0)) \sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0}\right)}\, \mathrm{d}x_0, \tag{3.70}$$

где функция Грина имеет вид

$$G(x - x_0, y - y_0) = \frac{i}{2kd} \sum_{n=-\infty}^{\infty} \frac{\exp(ik\alpha_n(x - x_0))}{\sqrt{1 - \alpha_n^2}} \exp(ik\sqrt{1 - \alpha_n^2}\,|y - f(x_0)|), \tag{3.71}$$

где α_n определено в уравнении (3.15).

Заменяя точку наблюдения (x, y) в (3.70) точкой $(x, f(x))$ профиля дифракционной решетки и используя граничное условие равенства нулю полного электрического поля $E_z(x,y) + E_{0z}(x,y)$ на поверхности решетки (при $y = f(x)$), получаем интегральное уравнение для плотности поверхностного тока:

$$E_{0z}(x, f(x)) = \frac{ik}{c} \int_0^d j_z(x) G(x - x_0, f(x) - f(x_0)) \sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0}\right)^2}\, \mathrm{d}x_0. \tag{3.72}$$

Согласно (3.70), (3.71), рассеянное отраженное поле (при $y > \max(f(x_0))$, $x_0 \in [0, d]$) представлено разложением Рэлея:

$$E_z(x,y) = \sum_{n=-\infty}^{\infty} R_n \exp(ik(\alpha_n x + \beta_n y)), \tag{3.73}$$

с коэффициентами

$$R_n = \frac{1}{2cd\sqrt{1 - \alpha_n^2}} \int_0^d j_z(x_0) \exp(-ik\alpha_n x_0 - ik\beta_n f(x_0)) \sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0}\right)^2}\, \mathrm{d}x_0, \tag{3.74}$$

представленными через значения плотности тока на поверхности решетки.

Таким образом, решение задачи дифракции сводится к решению интегрального уравнения (3.72) с последующим расчетом коэффициентов Рэлея по формуле (3.74).

Рассмотрим решение (3.72) по методу Фурье [5], основанному на представлении искомой функции плотности тока в виде квазипериодического разложения

$$j_z(x) = \sum_{-\infty}^{\infty} c_n \exp(ik\alpha_n x). \tag{3.75}$$

Подставляя (3.75) в (3.72), получаем следующее функциональное уравнение:

$$\frac{c}{ik}\exp(ik\alpha_0 x - ik\beta_0 f(x)) = \\ = \sum_{-\infty}^{\infty} c_n \int_0^d G(x - x_0, f(x) - f(x_0)) \exp(ik\alpha_n x_0) \sqrt{1 + \left(\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0}\right)^2}\, \mathrm{d}x_0. \quad (3.76)$$

Умножая обе части уравнения (3.76) на $\exp(-ik\alpha_p x)$, $p = -\infty, \infty$, и интегрируя по периоду, получаем систему линейных уравнений для определения искомых коэффициентов разложения функции плотности тока в виде

$$\sum_{n=-\infty}^{\infty} A_{pn} \cdot c_n = d_p, \quad p = -\infty, \infty, \quad (3.77)$$

где

$$d_p = \frac{c}{ik} \int_0^d \exp\left(-ik\beta_0 f(x)\right) \exp(-2\pi i p/d)\, \mathrm{d}x, \quad (3.78)$$

$$A_{pn} = \int_0^d \int_0^d \exp\left(-ik\alpha_p x_0\right) G\left(x_0 - x, f(x_0) - f(x)\right) \exp\left(ik\alpha_n x\right) \sqrt{1 + \left(\frac{df\,(x)}{dx}\right)^2}\, \mathrm{d}x\, \mathrm{d}x_0. \quad (3.79)$$

Согласно (3.71), (3.79), вычисление элементов матрицы A_{pn} в (3.77) сводится к расчету ряда, составленного из двойных интегралов, является сложной вычислительной задачей. Если профиль решетки представляет собой кусочно-линейную функцию

$$f(x) = \xi_m + \eta_m x, \quad x \in [x_m, x_{m+1}], \quad m = \overline{1, K}, \quad (3.80)$$

то элементы матрицы (3.79) могут быть преобразованы к виду

$$A_{pn} = \frac{1}{2ikd} \sum_{j=-\infty}^{\infty} \sum_{m=1}^{K-1} \sum_{k=1}^{K-1} \frac{\sqrt{1 + \eta_k^2}}{\beta_j} I_{pn}^{jmk}, \quad (3.81)$$

где интегралы

$$I_{pn}^{jmk} = \int_{x_k}^{x_{k+1}} \int_{x_m}^{x_{m+1}} \exp\{-ik\left(\alpha_p - \alpha_j\right) x_0 + ik\left(\alpha_n - \alpha_j\right) x + \\ + ik\beta_j \left|\xi_m - \xi_k + \eta_m x_0 - \eta_k x\right|\} \mathrm{d}x\, \mathrm{d}x_0, \quad (3.82)$$

могут быть вычислены аналитически. Формулы (3.81), (3.82) можно использовать для расчета дифракционных решеток с непрерывным профилем общего вида. Для этого достаточно провести аппроксимацию кусочно-линейной функцией.

Приближение Рэлея. Поскольку решение интегрального уравнения (3.72) является сложной математической задачей, то в ряде случаев для расчета поля от

дифракционной решетки используют приближение Рэлея [5]. Согласно (3.73), поле над профилем дифракционной решетки (при $y > a = \max(f(x_0))$, $x_0 \in [0, d]$) представляется в виде разложения Рэлея. Из этого не следует, что поле при $y < a$ также описывается разложением Рэлея. Однако, при неглубоком и гладком профиле эту гипотезу можно считать оправданной. Приближение Рэлея предполагает, что разложение (3.73) или (3.14) описывает поле не только над решеткой при $y > a$, но и внутри решетки при $y < a$. Напомним, что скалярная функция $u(x, y)$ в (3.14) соответствует компоненте E_z электрического поля для ТЕ-поляризации или компоненте H_z магнитного поля для ТМ-поляризации.

Для расчета коэффициентов Рэлея R_n в (3.14) достаточно воспользоваться граничным условием (3.11) при $y = f(x)$. Согласно (3.11), поле (3.14) должно удовлетворять граничному условию

$$u(x, y)|_{y=f(x)} = 0, \tag{3.83}$$

для ТЕ-поляризации, и граничному условию

$$\left.\frac{\mathrm{d}u(x, y)}{\mathrm{d}\mathbf{n}}\right|_{y=f(x)} = 0, \tag{3.84}$$

для ТМ-поляризации, где $\mathbf{n}$ — единичный вектор нормали к профилю решетки.

Подставляя (3.14) в (3.83), получаем в случае ТЕ-поляризации для расчета коэффициентов R_n следующее уравнение:

$$\sum_{n=-\infty}^{\infty} R_n \exp\left(ik\left(\left(\alpha_n - \alpha_0\right)x + \left(\beta_n - \beta_0\right)f(x)\right)\right) = -\exp\left(-2ik\beta_0 f(x)\right). \tag{3.85}$$

Заменяя (3.85) условием равенства коэффициентов Фурье в разложении правой и левой частей, получим для расчета R_n следующую систему линейных уравнений:

$$\sum_{n=-\infty}^{\infty} A_{pn} \cdot R_n = B_p, \quad p = -\infty, \infty, \tag{3.86}$$

$$A_{pn} = \int_0^{2\pi} \exp\left(i\xi(n-p) + i\left(\beta_n - \beta_0\right)H(\xi)\right)\mathrm{d}\xi, \tag{3.87}$$

$$B_p = -\int_0^{2\pi} \exp\left(-2i\beta_0 H(\xi) - ip\xi\right)\mathrm{d}\xi, \tag{3.88}$$

где

$$H(\xi) = kf\left(\frac{\xi d}{2\pi}\right) \tag{3.89}$$

— нормированная высота решетки.

Проводя аналогичные вычисления, несложно получить в случае ТМ-поляризации для расчета коэффициентов R_n систему линейных уравнений вида (3.86), где

$$A_{pn} = \int_0^{2\pi} \left(\beta_n - \alpha_n \frac{\lambda}{d} H'(\xi)\right) \exp\left(i\xi(n-p) + i\left(\beta_n - \beta_0\right)H(\xi)\right)\mathrm{d}\xi, \tag{3.90}$$

$$B_p = \int\limits_0^{2\pi} \left(\beta_0 + \alpha_0 \frac{\lambda}{d} H'(\xi)\right) \exp\left(-2i\beta_0 H(\xi) - ip\xi\right) \mathrm{d}\xi. \tag{3.91}$$

На практике для расчета коэффициентов R_n используется конечная система из $2N+1$ уравнений:

$$A \cdot \mathbf{R} = \mathbf{B}, \tag{3.92}$$

где $A = (A_{pn})_{-N}^{N}$, $\mathbf{R} = (R_n)_{-N}^{N}$, $\mathbf{B} = (B_n)_{-N}^{N}$.

При $d \gg \lambda$ приближении Рэлея переходит в приближение Фраунгофера. Действительно, при $d \gg \lambda$ получим, что $(\beta_n - \beta_0) \approx 0$ и $(\lambda/d) H'(\xi) \approx 0$. В этом случае матрицы AA в (3.86) и (3.92) становятся диагональными, и коэффициенты Рэлея будут пропорциональны коэффициентам Фурье в разложении функции $\exp(i\varphi(\xi))$, где функция $\varphi(\xi) = -2\beta_0 kf(\xi d/2\pi)$ описывает набег фазы при отражении плоской волны от решетки с высотой $f(\xi d/2\pi)$.

Интересно отметить, что приближение Фраунгофера соответствует выбору первого члена в асимптотическом разложении для коэффициентов Рэлея. Действительно, используя матричное разложение $(E+C)^{-1} = \sum\limits_{j=0}^{\infty} (-C)^j$, представим коэффициенты Рэлея в виде

$$\mathbf{R} = A^{-1} \cdot \mathbf{B} = \frac{1}{2\pi} (\mathbf{E} + A_1)^{-1} \mathbf{B} = \frac{1}{2\pi} \sum_{j=0}^{\infty} (-A_1)^j \cdot \mathbf{B} = \pm\mathbf{F} - \sum_{j=1}^{\infty} (-A_1)^k \cdot \mathbf{F}, \tag{3.93}$$

где $\mathbf{F}$ — вектор коэффициентов Фурье, а A_1 — матрица с элементами

$$A_{1pn} = \begin{cases} \dfrac{1}{2\pi} A_{pn}, & p \neq n, \\ \dfrac{1}{2\pi} A_{pn} - 1, & p = n. \end{cases} \tag{3.94}$$

Пример 3.2. Для оценки точности приближения Рэлея рассмотрим работу косинусоидальной решетки с профилем

$$f(x) = \frac{h}{2} \cdot \cos\left(\frac{2\pi x}{d}\right). \tag{3.95}$$

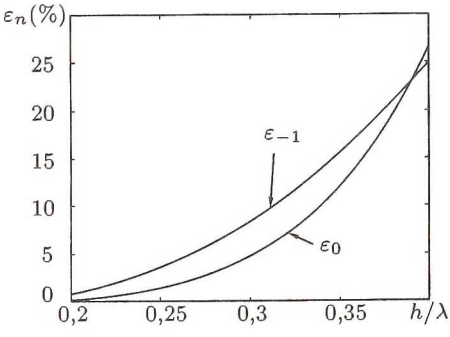


Рис. 3.4. Зависимость относительной ошибки для интенсивностей -1-го ($\varepsilon_{-1}(h)$) и 0-го ($\varepsilon_0(h)$) порядков косинусоидальной решетки от глубины h для ТЕ-поляризации при $d = \lambda$, $\theta = 30^\circ$

Для характеристики точности приближения Рэлея введем относительную ошибку

$$\varepsilon_n(h) = \frac{\left|I_n^{Int}(h) - I_n^{Rel}(h)\right|}{I_n^{Int}(h)},$$

где h — амплитуда профиля, $I_n^{Int}(h), I_n^{Rel}(h)$ — интенсивности n-го порядка, рассчитанные точным интегральным методом по формулам (3.77)–(3.79) и в приближении Рэлея по формулам (3.86)–(3.88). На рис. 3.4 приведены зависимости относительной ошибки $\varepsilon(h/\lambda)$ для интенсивностей 0-го и -1-го порядков от глубины профиля h. Из рис. 3.4 видно, что приближение Рэлея дает ошибку менее 10% при $h < 0{,}325\lambda$. Приведенный пример показывает, что приближение Рэлея применимо при гладких неглубоких профилях ($h < d/3$), даже при периодах близких к длине волны.

3.3. Дифракция на пропускающих диэлектрических решетках

Рассмотрим дифракцию плоской волны с волновым вектором $\mathbf{k} = k(\sin(\theta), -\cos(\theta), 0)$ на диэлектрической решетке с профилем $y = f(x)$ и периодом d. Геометрия задачи дифракции приведена на рис. 3.5, где R_n и T_n — коэффициенты отражения и пропускания дифракционных порядков. Для приведенной геометрии задачи имеются три зоны с различной диэлектрической проницаемостью ε. В зонах 1 и 3, то есть при $y > a$, где a — максимальная высота профиля решетки и при $y < 0$, диэлектрическая проницаемость постоянна. Без ограничения общности будем считать, что в первой зоне $\varepsilon = 1$, а в третьей зоне $\varepsilon > 1$. Во второй зоне $y \in [0, a]$ — зоне модуляции, диэлектрическая проницаемость является функцией $\varepsilon = \varepsilon(x, y)$.

Рассмотрим дифракцию двух независимых TE- и TM-поляризованных плоских волн. В первой ($y > a$) и третьей зонах ($y < 0$) представление поля соответствует разложению Рэлея (3.14). При $y > a$ компоненты поля имеют вид [5]

$$u(x, y) = \exp(i(\alpha_0 x - \beta_0 y)) + \sum_{n=-\infty}^{\infty} R_n \exp(i(\alpha_n x + \beta_n y)), \tag{3.96}$$

где

$$\begin{aligned} &\alpha_n = k_0 \sin(\theta) + n\frac{2\pi}{d}, \quad k_0^2 = \omega^2/c^2 = (2\pi/\lambda_0)^2, \\ &\beta_n = \sqrt{k_0^2 - \alpha_n^2}, \end{aligned} \tag{3.97}$$

а скалярная функция $u(x, y)$ соответствует компоненте $E_z(x, y)$ электрического поля для TE-поляризации и компоненте $H_z(x, y)$ магнитного поля для TM-поляризации.

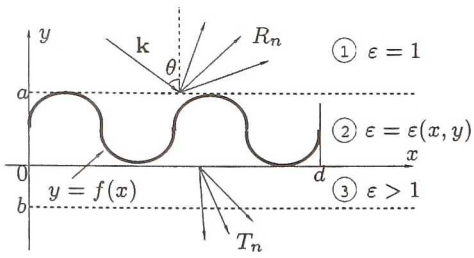


Рис. 3.5. Геометрия задачи дифракции на диэлектрической решетке

Отметим, что в отличие от уравнений (3.14)–(3.16), k_0 для сокращения записей внесено в α_n и β_n, которые более не соответствуют синусам и косинусам углов дифракции. При $y < 0$, $u(x, y)$ имеет вид

$$u(x,y) = \sum_{n=-\infty}^{\infty} T_n \exp\left(i\left(\alpha_n x - \tilde{\beta}_n y\right)\right), \tag{3.98}$$

где

$$\tilde{\beta}_n = \sqrt{k_0^2\varepsilon - \alpha_n^2}. \tag{3.99}$$

Разложения Рэлея (3.96), (3.98) являются решениями уравнения Гельмгольца:

$$\Delta u + k^2 u = 0, \tag{3.100}$$

при $k^2 = k_0^2$ и $k^2 = k_0^2\varepsilon$, соответственно. В зоне модуляции (при $y \in [0, a]$) поле описывается системой дифференциальных уравнений, различных для TE- и TM-поляризаций.

TM-поляризация. Для TM-поляризации при $\mathbf{H}(x, y) = (0, 0, H_z(x, y))$ вместо уравнений (3.8) получим

$$\begin{aligned}
&\frac{\partial H_z(x,y)}{\partial y} = -ik_0\varepsilon E_x(x,y), && \frac{\partial E_y(x,y)}{\partial z} = 0,\\
&-\frac{\partial H_z(x,y)}{\partial x} = -ik_0\varepsilon E_y(x,y), && \frac{\partial E_x(x,y)}{\partial z} = 0,\\
&E_z(x,y) = 0, && \frac{\partial E_y(x,y)}{\partial x} - \frac{\partial E_x(x,y)}{\partial y} = -ik_0 H_z(x,y).
\end{aligned} \tag{3.101}$$

На границе раздела двух диэлектриков с проницаемостями ε_1, ε_2 тангенциальные составляющие векторов $\mathbf{E}$, $\mathbf{H}$ непрерывны. Без ограничения общности будем считать границей раздела ось Ox. При этом тангенциальные составляющие векторов $\mathbf{E}$, $\mathbf{H}$ соответствуют компонентам $H_z(x, y)$ и $E_x(x, y)$. Условие непрерывности имеет вид

$$\frac{-1}{ik_0\varepsilon_1}\frac{\partial H_z(x,+0)}{\partial y} = E_x(x,+0) = E_x(x,-0) = \frac{-1}{ik_0\varepsilon_2}\frac{\partial H_z(x,-0)}{\partial y}. \tag{3.102}$$

Согласно (3.102), при переходе через границу раздела функция $\partial H_z(x, y)/\partial y$ терпит разрыв, что не позволяет использовать в случае TM-поляризации обычное уравнение Гельмгольца. Непрерывной функцией при переходе через границу являются функция $(1/ik_0\varepsilon) \cdot (\partial H_z(x, y)/\partial y)$ и функция

$$\tilde{E}_x(x,y) = \frac{1}{k_0^2\varepsilon}\frac{\partial H_z(x,y)}{\partial y} = \frac{1}{k^2}\frac{\partial H_z(x,y)}{\partial y}. \tag{3.103}$$

Получим из (3.101) аналог уравнения Гельмгольца относительно непрерывных функций $\tilde{E}_x(x, y)$ и $H_z(x, y)$:

$$\frac{\partial}{\partial x}\left(\frac{1}{k^2}\frac{\partial H_z(x,y)}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{k^2}\frac{\partial H_z(x,y)}{\partial y}\right) + H_z(x,y) = 0. \tag{3.104}$$

Представим (3.104), с учетом (3.103), в виде системы

$$\begin{cases} \dfrac{\partial H_z(x,y)}{\partial y} = k^2 \tilde{E}(x,y), \\ \dfrac{\partial \tilde{E}(x,y)}{\partial y} = -\dfrac{\partial}{\partial x}\left(\dfrac{1}{k^2}\dfrac{\partial H_z(x,y)}{\partial x}\right) - H_z(x,y). \end{cases} \tag{3.105}$$

Функция $H_z(x,y)$ является квазипериодической [5]:

$$H_z(x,y) = \nu(x,y)\exp(i\alpha_0 x), \quad \alpha_0 = k_0 \sin(\theta), \tag{3.106}$$

где $\nu(x,y)$ — периодическая по x функция с периодом d. Разлагая функцию $\nu(x,y)$ в ряд Фурье по переменной x, представим $H_z(x,y)$ в виде:

$$H_z(x,y) = \sum_{m=-\infty}^{\infty} H_m(y)\exp(i\alpha_m x), \quad \alpha_m = \alpha_0 + 2\pi m/d. \tag{3.107}$$

Функция $\tilde{E}_x(x,y)$ также является квазипериодической:

$$\tilde{E}_x(x,y) = \sum_{m=-\infty}^{\infty} \tilde{E}_m(y)\exp(i\alpha_m x). \tag{3.108}$$

В дальнейшем будем считать, что функции $\tilde{E}_x(x,y)$, $H_z(x,y)$ в зоне модуляции могут быть аппроксимированы отрезками своих рядов (3.107), (3.108) с $2N+1$ членами.

Функции $k^2(x,y) = k_0^2\varepsilon(x,y)$ и $1/k^2(x,y)$ являются периодическими по x с периодом d:

$$\left.\begin{aligned} k^2(x,y) &= \sum_{n=-\infty}^{\infty} c_n^1(y)\exp\left(i\frac{2\pi}{d}nx\right), \\ \frac{1}{k^2(x,y)} &= \sum_{n=-\infty}^{\infty} c_n^2(y)\exp\left(i\frac{2\pi}{d}nx\right). \end{aligned}\right\} \tag{3.109}$$

Подставляя разложения (3.107)–(3.109) в систему (3.105), получим (при учете $2N+1$ членов разложения в (3.107), (3.108)) следующую систему $4N+2$ дифференциальных уравнений:

$$\begin{cases} \displaystyle\sum_{m=-N}^{N} \frac{dH_m(y)}{dy}\exp(i\alpha_m x) = \sum_{l=-N}^{N} \tilde{E}_l(y)\exp(i\alpha_l x)\cdot \sum_{n=-\infty}^{\infty} c_n^{(1)}(y)\exp\left(i\frac{2\pi}{d}nx\right), \\ \displaystyle\sum_{m=-N}^{N} \frac{d\tilde{E}_m(y)}{dy}\exp(i\alpha_m x) = -\frac{\partial}{\partial x}\left(\sum_{n=-\infty}^{\infty} c_n^{(2)}(y)\exp\left(i\frac{2\pi}{d}nx\right)\times\right. \\ \qquad \left.\times \displaystyle\sum_{l=-N}^{N} i\alpha_l H_l(y)\exp(i\alpha_l x)\right) - \sum_{m=-N}^{N} H_m(y)\exp(i\alpha_m x). \end{cases} \tag{3.110}$$

Умножая каждое из уравнений (3.110) на $\exp(-i\alpha_p x)$, $p = -\overline{N,N}$, и интегрируя по x на периоде, получаем систему уравнений, не зависящих от переменной x:

$$\begin{cases} \dfrac{dH_p(y)}{dy} = \displaystyle\sum_{l=-N}^{N} \tilde{E}_l(y)\, c_{p-l}^{(1)}(y), & p = -\overline{N,N}, \\ \dfrac{d\tilde{E}_p(y)}{dy} = \alpha_p \displaystyle\sum_{l=-N}^{N} \alpha_l c_{p-l}^{(2)}(y)\, H_l(y) - H_p(y), & p = -\overline{N,N}. \end{cases} \tag{3.111}$$

Для бинарной решетки коэффициенты Фурье в (3.111) не зависят от переменной y:

$$c_n^{(1)} = \begin{cases} \dfrac{ik_0^2(\varepsilon-1)}{2\pi n} \displaystyle\sum_{j=1}^{2K} (-1)^j \exp\left(-i\frac{2\pi}{d} n x_j\right), & n \neq 0 \\ k_0^2 + \dfrac{k_0^2(\varepsilon-1)}{d} \displaystyle\sum_{j=1}^{2K} (-1)^j x_j, & n = 0 \end{cases}, \tag{3.112}$$

$$c_n^{(2)} = \begin{cases} \dfrac{i(1/\varepsilon-1)}{2\pi n k_0^2} \displaystyle\sum_{j=1}^{2K} (-1)^j \exp\left(-i\frac{2\pi}{d} n x_j\right), & n \neq 0 \\ 1/k_0^2 + \dfrac{(1/\varepsilon-1)}{k_0^2 d} \displaystyle\sum_{j=1}^{2K} (-1)^j x_j, & n = 0 \end{cases}, \tag{3.113}$$

где K — число штрихов решетки, $x_1, \ldots, x_{2K}$ — координаты границ штрихов (см. рис. 3.6).

Таким образом, в зонах 1 и 3 решения имеют вид (3.96) и (3.98), а в зоне модуляции необходимо решать систему уравнений (3.111). Для поиска общего решения необходимо найти $4N+2$ линейно независимых частных решений. При отсутствии модуляции ($\varepsilon(x,y) = \varepsilon$) базисные решения системы (3.111) имеют вид

$$\left.\begin{aligned} H_p^{\pm}(y) &= \exp\left(\pm i\tilde{\beta}_p y\right), & p = -\overline{N,N}, \\ \tilde{E}_p^{\pm}(y) &= \pm\frac{i\tilde{\beta}_p}{k_0^2\varepsilon} \exp\left(\pm i\tilde{\beta}_p y\right), & p = -\overline{N,N}. \end{aligned}\right\} \tag{3.114}$$

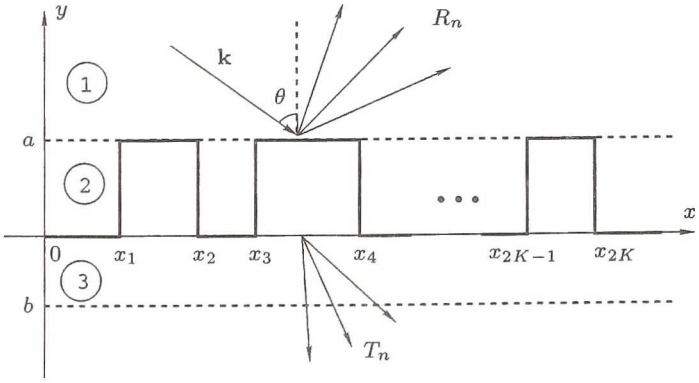


Рис. 3.6. Геометрия задачи дифракции на бинарной диэлектрической решетке

Для согласования решения в зоне модуляции 2 с решением (3.114) в зоне 3, граничные условия для системы (3.111) определим в виде

$$\left.\begin{aligned} H^{\mp}_{mj}(0) &= \delta_{m-j}, && m,j=-\overline{N,N},\\ \tilde{E}^{\mp}_{mj}(0) &= \mp i\tilde{\beta}_m\delta_{m-j}/\left(k_0^2\varepsilon\right), && m,j=-\overline{N,N}, \end{aligned}\right\} \tag{3.115}$$

где $\delta_m = \begin{cases} 1, & m=0,\\ 0, & m\neq 0. \end{cases}$

Для удобства выкладок введем $4N+2$ объединённых векторов граничных условий вида

$$\left\{\begin{pmatrix} \mathbf{H}^{+}_{-N}(0)\\ \tilde{\mathbf{E}}^{+}_{-N}(0)\end{pmatrix},\dots,\begin{pmatrix} \mathbf{H}^{+}_{N}(0)\\ \tilde{\mathbf{E}}^{+}_{N}(0)\end{pmatrix},\begin{pmatrix} \mathbf{H}^{-}_{-N}(0)\\ \tilde{\mathbf{E}}^{-}_{-N}(0)\end{pmatrix},\dots,\begin{pmatrix} \mathbf{H}^{-}_{N}(0)\\ \tilde{\mathbf{E}}^{-}_{N}(0)\end{pmatrix}\right\}. \tag{3.116}$$

Обозначим $\Psi_i(y)$, $i=\overline{1,4N+2}$, вектора базисных функций

$$\left\{\begin{pmatrix} \Psi^{+}_{-N}(y)\\ \tilde{\Psi}^{+}_{-N}(y)\end{pmatrix},\dots,\begin{pmatrix} \Psi^{+}_{N}(y)\\ \tilde{\Psi}^{+}_{N}(y)\end{pmatrix},\begin{pmatrix} \Psi^{-}_{N}(y)\\ \tilde{\Psi}^{-}_{-N}(y)\end{pmatrix},\dots,\begin{pmatrix} \Psi^{-}_{N}(y)\\ \tilde{\Psi}^{-}_{N}(y)\end{pmatrix}\right\}, \tag{3.117}$$

полученные из решения системы (3.111) с граничными условиями (3.116). При этом общее решение системы принимает вид

$$H_m(y) = \sum_{j=-N}^{N} C^{-}{}_j\psi^{-}_{mj}(y) + \sum_{j=-N}^{N} C^{+}{}_j\psi^{+}_{mj}(y), \quad m=-\overline{N,N}, \tag{3.118}$$

$$\tilde{E}_m(y) = \sum_{j=-N}^{N} C^{-}{}_j\tilde{\psi}^{-}_{mj}(y) + \sum_{j=-N}^{N} C^{+}{}_j\tilde{\psi}^{+}_{mj}(y), \quad m=-\overline{N,N}. \tag{3.119}$$

Для определения рэлеевских коэффициентов пропускания T_n и отражения R_n в (3.96), (3.98) воспользуемся условиями непрерывности $H_z(x,y)$ и $\tilde{E}_x(x,y)$ на границах зоны модуляции при $y=0$ и при $y=a$.

Условия непрерывности при $y=0$ дают следующие функциональные уравнения:

$$\begin{aligned} \sum_{n=-N}^{N} T_n\exp(i\alpha_n x) &= \sum_{n=-N}^{N} H_n(0)\exp(i\alpha_n x) = \\ &= \sum_{n=-N}^{N}\left(\sum_{j=-N}^{N} C_j^{-}\psi^{-}_{nj}(0) + \sum_{j=-N}^{N} C_j^{+}\psi^{+}_{nj}(0)\right)\exp(i\alpha_n x) = \\ &= \sum_{n=-N}^{N}\left(C_n^{-}+C_n^{+}\right)\exp(i\alpha_n x), \end{aligned} \tag{3.120}$$

$$-\frac{i}{k_0^2\varepsilon}\sum_{n=-N}^{N}\tilde{\beta}_n T_n \exp(i\alpha_n x) = \sum_{n=-N}^{N}\tilde{E}_n(0)\exp(i\alpha_n x) =$$

$$= \sum_{n=-N}^{N}\left(\sum_{j=-N}^{N} C_j^- \tilde{\psi}_{nj}^-(0) + \sum_{j=-N}^{N} C_j^+ \tilde{\psi}_{nj}^+(0)\right)\exp(i\alpha_n x) =$$

$$= \sum_{n=-N}^{N}\frac{i\tilde{\beta}_n}{k_0^2\varepsilon}\left(C_n^+ - C_n^-\right)\exp(i\alpha_n x). \quad (3.121)$$

Умножая обе части уравнений (3.120), (3.121) на $\exp(-i\alpha_p x)$, $p = -\overline{N,N}$, и интегрируя по периоду, получаем

$$\begin{cases} T_p = C_p^- + C_p^+, \\ -T_p = -C_p^- + C_p^+. \end{cases} \quad (3.122)$$

Решение системы (3.122) имеет вид

$$\begin{cases} T_p = C_p^-, \\ C_p^+ = 0. \end{cases} \quad (3.123)$$

Согласно (3.123) получим, что в зоне модуляции поле имеет вид

$$\begin{pmatrix} H_m(y) \\ \tilde{E}_m(y) \end{pmatrix} = \sum_{j=-N}^{N} C_j^- \begin{pmatrix} \psi_{mj}^-(y) \\ \tilde{\psi}_{mj}^-(y) \end{pmatrix} = \sum_{j=-N}^{N} T_j \begin{pmatrix} \psi_{mj}^-(y) \\ \tilde{\psi}_{mj}^-(y) \end{pmatrix}, \quad m = -\overline{N,N}. \quad (3.124)$$

Условия непрерывности функций $H_z(x,y)$ и $\tilde{E}_x(x,y)$ на верхней границе зоны модуляции (при $y = a$) дают следующие уравнения:

$$\sum_{n=-N}^{N} R_n \exp(i\alpha_n x + i\beta_n a) + \exp(i\alpha_0 x - i\beta_0 a) = \sum_{n=-N}^{N} H_n(a)\exp(i\alpha_n x) =$$

$$= \sum_{n=-N}^{N}\left(\sum_{j=-N}^{N} T_j \psi_{nj}^-(a)\right)\exp(i\alpha_n x), \quad (3.125)$$

$$\frac{i\beta_n}{k_0^2}\sum_{n=-N}^{N} R_n \exp(i\alpha_n x + i\beta_n a) - \frac{i\beta_0}{k_0^2}\exp(i\alpha_0 x - i\beta_0 a) = \sum_{n=-N}^{N}\tilde{E}_n(a)\exp(i\alpha_n x) =$$

$$= \sum_{n=-N}^{N}\left(\sum_{j=-N}^{N} T_j \tilde{\psi}_{nj}^-(a)\right)\exp(i\alpha_n x). \quad (3.126)$$

Умножая обе части (3.125), (3.126) на $\exp(-i\alpha_p x)$ и интегрируя по периоду, получим две системы линейных уравнений вида

$$\sum_{j=-N}^{N} T_j \psi_{pj}^-(a) = R_p \exp(i\beta_p a) + \delta_p \exp(-i\beta_0 a), \quad p = -\overline{N,N}, \quad (3.127)$$

$$\sum_{j=-N}^{N} T_j \tilde{\psi}_{pj}^-(a) = R_p \exp(i\beta_p a)\frac{i\beta_p}{k_0^2} - \delta_p \exp(-i\beta_0 a)\frac{i\beta_0}{k_0^2}, \quad p = -\overline{N,N}. \quad (3.128)$$

Системы линейных уравнений (3.127), (3.128) представим в матричном виде:

$$\begin{aligned} H_{01}\ \ \mathbf{T} &= H_{02}\mathbf{R} + \exp\left(-i\beta_0 a\right)\cdot\boldsymbol{\delta}, \\ H_{01pj} = \psi_{pj}^{-}(a), &\quad H_{02pj} = \delta_{p-j}\cdot\exp(i\beta_p a), \quad p,j = -\overline{N,N}, \end{aligned} \tag{3.129}$$

$$\begin{aligned} H_{11}\cdot\mathbf{T} &= H_{12}\mathbf{R} - \frac{i\beta_0}{k_0^2}\exp\left(-i\beta_0 a\right)\cdot\boldsymbol{\delta}, \\ H_{11pj} = \tilde{\psi}_{pj}^{-}(a), &\quad H_{12pj} = \delta_{p-j}\cdot\frac{i\beta_p}{k_0^2}\exp(i\beta_p a), \quad p,j = -\overline{N,N}. \end{aligned} \tag{3.130}$$

Здесь $\boldsymbol{\delta}$ — вектор столбец, центральный элемент которого равен единице, а остальные элементы равны нулю. Из уравнений (3.129), (3.130) получим вектора коэффициентов Рэлея в виде

$$\mathbf{T} = 2\exp\left(-i\beta_0 a\right)\left(H_{01} - D_\beta\cdot H_{11}\right)^{-1}\cdot\boldsymbol{\delta}, \tag{3.131}$$

$$\mathbf{R} = H_{12}^{-1}H_{11}\ \ \mathbf{T} + \exp\left(-2i\beta_0 a\right)\cdot\boldsymbol{\delta}, \tag{3.132}$$

где D_β — диагональная матрица с элементами $k_0^2/i\beta_p$, $p = -\overline{N,N}$.

Для рассмотренной задачи дифракции несложно учесть влияние подложки на работу решетки. Подложка обозначена пунктирной линией на рис. 3.5. Будем предполагать, что ниже подложки $\varepsilon = 1$. Тогда поле в зоне за подложкой (при $y < b$) имеет вид

$$H_z(x,y) = \sum_{n=-\infty}^{\infty} T_n \exp\left(i\left(\alpha_n x - \beta_n y\right)\right). \tag{3.133}$$

Решение (3.133) отличается от (3.98) в зоне 3 видом коэффициентов β_n. Согласно (3.114), общее решение системы (3.111) при $b < y < 0$ имеет вид

$$\begin{aligned} H_p(y) &= C_p^{+}\exp\left(i\tilde{\beta}_p y\right) + C_p^{-}\exp\left(-i\tilde{\beta}_p y\right), && p = -\overline{N,N}, \\ \tilde{E}_p(y) &= C_p^{+}\frac{i\tilde{\beta}_p}{k_0^2\varepsilon}\exp\left(i\tilde{\beta}_p y\right) - C_p^{-}\frac{i\tilde{\beta}_p}{k_0^2\varepsilon}\exp\left(-i\tilde{\beta}_p y\right), && p = -\overline{N,N}. \end{aligned} \tag{3.134}$$

Воспользуемся условиями непрерывности функций $H_z(x,y)$ и $\tilde{E}_x(x,y)$ на границе подложка — воздух для выражения коэффициентов $C_p^{\pm}$ в (3.134) через коэффициенты T_n:

$$\sum_{n=-N}^{N} T_n\exp\left(i\alpha_n x - i\beta_n b\right) = \sum_{n=-N}^{N} H_n(b)\exp\left(i\alpha_n x\right) =$$

$$= \sum_{n=-N}^{N}\left(C_n^{+}\exp\left(i\tilde{\beta}_n b\right) + C_n^{-}\exp\left(-i\tilde{\beta}_n b\right)\right)\exp\left(i\alpha_n x\right),$$

$$\frac{-i}{k_0^2}\sum_{n=-N}^{N}\beta_n T_n\exp\left(i\alpha_n x - i\beta_n b\right) = \sum_{n=-N}^{N}\tilde{E}_n(b)\exp\left(i\alpha_n x\right) =$$

$$= \frac{i}{k_0^2\varepsilon}\sum_{n=-N}^{N}\tilde{\beta}_n\left(C_n^{+}\exp\left(i\tilde{\beta}_n b\right) - C_n^{-}\exp\left(-i\tilde{\beta}_n b\right)\right)\exp\left(i\alpha_n x\right).$$

Умножая приведенные уравнения на $\exp(-i\alpha_p x)$ и интегрируя по периоду, получаем уравнения связи коэффициентов $C_p^{\pm}$ и T_n в виде

$$\begin{aligned} T_p \exp(-i\beta_p b) &= C_p^+ \exp\left(i\tilde{\beta}_p b\right) + C_p^- \exp\left(-i\tilde{\beta}_p b\right), \\ -T_p \exp(-i\beta_p b) &= \frac{\tilde{\beta}_p}{\beta_p \varepsilon}\left(C_p^+ \exp\left(i\tilde{\beta}_p b\right) - C_p^- \exp\left(-i\tilde{\beta}_p b\right)\right), \quad p = -\overline{N, N}. \end{aligned} \tag{3.135}$$

Из (3.135) получим, что

$$\begin{aligned} C_p^- &= T_p \cdot \gamma_p, \\ C_p^+ &= T_p \cdot \mu_p, \quad p = -\overline{N, N}, \end{aligned} \tag{3.136}$$

где

$$\begin{aligned} \gamma_p &= \frac{\tilde{\beta}_p + \varepsilon\beta_p}{2\tilde{\beta}_p} \cdot \exp\left(-ib\left(\beta_p - \tilde{\beta}_p\right)\right), \\ \mu_p &= \exp\left(-2i\tilde{\beta}_p b\right) \frac{\tilde{\beta}_p - \varepsilon\beta_p}{2\tilde{\beta}_p} \cdot \exp\left(-ib\left(\beta_p + \tilde{\beta}_p\right)\right). \end{aligned} \tag{3.137}$$

Подставляя (3.136) в (3.134), получаем функции $H_p(y), \tilde{E}_p(y)$ в явном виде:

$$\begin{aligned} H_p(y) &= T_p\left(\mu_p \exp\left(i\tilde{\beta}_p y\right) + \gamma_p \exp\left(-i\tilde{\beta}_p y\right)\right), \qquad p = -\overline{N, N}, \\ \tilde{E}_p(y) &= T_p \frac{i\tilde{\beta}_p}{k_0^2 \varepsilon}\left(\mu_p \exp\left(i\tilde{\beta}_p y\right) - \gamma_p \exp\left(-i\tilde{\beta}_p y\right)\right), \quad p = -\overline{N, N}. \end{aligned} \tag{3.138}$$

Таким образом, при учете подложки, в зоне 3 существуют как прошедшие волны, распространяющиеся против направления оси Oy, так и отраженные от границы подложка-воздух волны, распространяющиеся по направлению оси Oy. Для согласования решения в зоне модуляции с решением (3.138) в зоне 3, определим граничные условия для системы (3.111) в виде

$$\begin{aligned} &H^+{}_{mj}(0) = \mu_m \delta_{m-j}, && H^-{}_{mj}(0) = \gamma_m \delta_{m-j}, \\ &\tilde{E}^+{}_{mj}(0) = i\chi_m \mu_m \delta_{mj}, && \tilde{E}^-{}_{mj}(0) = -i\chi_m \gamma_m \delta_{m-j}, \\ &\chi_m = \tilde{\beta}_m / \left(k_0^2 \varepsilon\right). && m, j = -\overline{N, N}. \end{aligned} \tag{3.139}$$

Используя условия непрерывности функций $H_z(x, y)$, $\tilde{E}_x(x, y)$ на нижней границе зоны модуляции, получаем

$$\left.\begin{aligned} T_p(\mu_p + \gamma_p) &= \mu_p C_p^+ + \gamma_p C_p^-, \\ T_p(\mu_p - \gamma_p) &= \mu_p C_p^+ - \gamma_p C_p^-. \end{aligned}\right\} \tag{3.140}$$

Решение системы (3.140) имеет вид

$$T_p = C_p^+ = C_p^-. \tag{3.141}$$

При этом общее решение системы (3.111) принимает вид

$$H_m(y) = \sum_{j=-N}^{N} T_j \left(\psi^-_{mj}(y) + \psi^+_{mj}(y)\right), \quad m = -\overline{N,N}, \tag{3.142}$$

$$\tilde{E}_m(y) = \sum_{j=-N}^{N} T_j \left(\tilde{\psi}^-_{mj}(y) + \tilde{\psi}^+_{mj}(y)\right), \quad m = -\overline{N,N}. \tag{3.143}$$

Далее, для определения коэффициентов пропускания T_n и отражения R_n в уравнениях (3.96), (3.131), воспользуемся условиями непрерывности $H_z(x,y)$ и функции $\tilde{E}_x(x,y)$ на верхней границе зоны модуляции (при $y = a$). Из условий непрерывности несложно получить коэффициенты T_n и R_n в виде (3.131), (3.132) при матрицах H_{01} и H_{11}, имеющих вид

$$\begin{aligned} H_{01pj} &= {\psi^+}_{pj}(a) + {\psi^-}_{pj}(a), \quad p,j = -N,N, \\ H_{11pj} &= {\tilde{\psi}^+}_{pj}(a) + {\tilde{\psi}^-}_{pj}(a), \quad p,j = -N,N. \end{aligned} \tag{3.144}$$

В случае решетки с непрерывным профилем для решения системы (3.111) используются численные методы типа метода Рунге–Кутта. Для бинарной решетки коэффициенты Фурье (3.112), (3.113) не зависят от переменной y. В этом случае поле в зоне модуляции описывается системой дифференциальных уравнений 1-го порядка с постоянными коэффициентами. Решение системы (3.111) с постоянными коэффициентами для граничного условия $\mathbf{X}_0$, определенного в уравнении (3.116) или (3.139), может быть представлено в компактном матричном виде:

$$\begin{pmatrix} \Psi^-_p(y) \\ \tilde{\Psi}^-_p(y) \end{pmatrix} = \exp\left(A^{TM} \cdot y\right) \mathbf{X}_0, \tag{3.145}$$

где A^{TM} — матрица системы (3.111) с постоянными Фурье коэффициентами $c^{(1)}_n$, $c^{(2)}_n$, определенными в уравнениях (3.112), (3.113). Согласно (3.111), матрица A^{TM} имеет вид

$$A^{TM} = \begin{bmatrix} NU & F_1 \\ F_2 & NU \end{bmatrix}, \tag{3.146}$$

где NU — нулевые матрицы размера $(2N+1) \times (2N+1)$, а F_1 и F_2 — матрицы из Фурье коэффициентов:

$$F_{1\,i,j} = c^{(1)}_{i-j}, \quad F_{2\,i,j} = \alpha_{-(N+1)+i}\alpha_{-(N+1)+j}c^{(2)}_{i-j} - \delta_{i-j}, \quad i,j = \overline{1,2N+1}. \tag{3.147}$$

Матричное представление (3.145) позволяет компактно выразить матрицы H_{01} и H_{11} в (3.131), (3.132) через матрицу системы и граничные условия (3.116) или (3.139) в виде

$$\begin{bmatrix} H_{01} \\ H_{11} \end{bmatrix} = \exp\left(A^{TM} \cdot a\right) C, \tag{3.148}$$

где при расчете без учета влияния подложки матрица C составлена из второй половины векторов граничных условий в (3.116):

$$C = \left\{ \begin{pmatrix} \mathbf{H}^-_{-N}(0) \\ \tilde{\mathbf{E}}^-_{-N}(0) \end{pmatrix}, \ldots, \begin{pmatrix} \mathbf{H}^-_{N}(0) \\ \tilde{\mathbf{E}}^-_{N}(0) \end{pmatrix} \right\} = \begin{bmatrix} E \\ D \end{bmatrix}, \tag{3.149}$$

где E — единичная матрица размера $(2N+1)\times(2N+1)$, а D_β — диагональная матрица с элементами

$$D_{\beta jj} = \frac{-i\tilde{\beta}_j}{k_0^2\varepsilon}, \quad j = -\overline{N,N}. \tag{3.150}$$

При учете влияния подложки матрица C составляется из векторов граничных условий (3.139):

$$C = \left[\begin{pmatrix}\mathbf{H}^-{}_{-N}(0)+\mathbf{H}^+{}_{-N}(0)\\ \tilde{\mathbf{E}}^-{}_{-N}(0)+\tilde{\mathbf{E}}^+{}_{-N}(0)\end{pmatrix},\ldots,\begin{pmatrix}\mathbf{H}^-{}_{N}(0)+\mathbf{H}^+{}_{N}(0)\\ \tilde{\mathbf{E}}^-{}_{N}(0)+\tilde{\mathbf{E}}^+{}_{N}(0)\end{pmatrix}\right]. \tag{3.151}$$

Матричное представление (3.148) для бинарной решетки может быть использовано для решения задачи дифракции на решетке с непрерывным профилем. Действительно, введем разбиение y_i, $i=\overline{0,N}$, $y_0=0$, $y_N=a$ зоны модуляции на N слоев. В каждом слое $y_{i-1}\leqslant y\leqslant y_i$ непрерывный профиль может быть аппроксимирован бинарным профилем. Полагая при $y_{i-1}\leqslant y\leqslant y_i$ в (3.111) $c_n^{(1)}(y)\approx c_n^{(1)}(\tilde{y}_i)$, $c_n^{(2)}(y)\approx c_n^{(2)}(\tilde{y}_i)$, где $\tilde{y}_i=(y_{i-1}+y_i)/2$, аппроксимируем при $y_{i-1}\leqslant y\leqslant y_i$ систему (3.111) с переменными коэффициентами системой уравнений с постоянными коэффициентами. Используя аналитическое матричное представление для решения системы дифференциальных уравнений с постоянными коэффициентами, получаем матрицы H_{01} и H_{11} в виде

$$\begin{aligned}\begin{bmatrix}H_{01}\\ H_{11}\end{bmatrix} = \Big(&\exp\left(A^{TM}(\tilde{y}_N)\cdot(a-y_{N-1})\right)\times\ldots\\ &\times\exp\left(A^{TM}(\tilde{y}_2)\cdot(y_2-y_1)\right)\exp\left(A^{TM}(\tilde{y}_1)\cdot y_1\right)\Big)\times\\ &\times C=\left(\prod_{i=N}^{1}\exp\left(\mathbf{A}^{TM}(\tilde{y}_i)\cdot(y_i-y_{i-1})\right)\right)\cdot C,\end{aligned} \tag{3.152}$$

где матрица C — матрица граничных условий (3.149) или (3.151), а $A^{TM}(\tilde{y}_i)$ — матрица системы (3.111) при $c_n^{(1)}(y)=c_n^{(1)}(\tilde{y}_i)$ и $c_n^{(2)}(y)=c_n^{(2)}(\tilde{y}_i)$. При этом коэффициенты Рэлея R_n, T_n также определяются по формулам (3.131) и (3.132). Отметим, что предел произведения матриц

$$\Omega_N(A^{TM})=\prod_{i=N}^{1}\exp\left(A^{TM}(\tilde{y}_i)\cdot(y_i-y_{i-1})\right), \tag{3.153}$$

в правой части (3.152) при $\Delta y_i=(y_i-y_{i-1})\to 0$ и $N\to\infty$ соответствует мультипликативному интегралу $\Omega_0^a\left(A^{TM}\right)$ [9], представляющему решение системы дифференциальных уравнений с переменными коэффициентами. При этом произведение (3.153) соответствует стандартной аппроксимации мультипликативного интеграла при сведении решения системы дифференциальных уравнений с переменными коэффициентами к решению N систем с постоянными коэффициентами.

ТЕ-поляризация. Полагая в базовых уравнениях (3.8) $\mathbf{E}(x,y) = (0,0,E_z(x,y))$, получаем:

$$\begin{aligned} &\frac{\partial E_z(x,y)}{\partial y} = ik_0 H_x(x,y), && \frac{\partial H_y(x,y)}{\partial z} = 0, \\ &\frac{\partial E_z(x,y)}{\partial x} = -ik_0 H_y(x,y), && \frac{\partial H_x(x,y)}{\partial z} = 0, \\ &H_z(x,y) = 0, && \frac{\partial H_y(x,y)}{\partial x} - \frac{\partial H_x(x,y)}{\partial y} = -ik_0 \varepsilon E_z(x,y). \end{aligned} \tag{3.154}$$

На границе раздела двух диэлектриков тангенциальные составляющие $E_z(x,y)$ и $H_x(x,y)$ векторов $\mathbf{E}$, $\mathbf{H}$ непрерывны. В отличие от ТМ-поляризации, производная $(\partial E_z(x,y)/\partial y) = ik_0 H_x(x,y)$ на границе раздела (при $y=0$) является непрерывной:

$$\frac{1}{ik_0}\frac{\partial E_z(x,+0)}{\partial y} = H_x(x,+0) = H_x(x,-0) = \frac{1}{ik_0}\frac{\partial E_z(x,-0)}{\partial y}. \tag{3.155}$$

Поэтому в зоне модуляции $E_z(x,y)$ удовлетворяет уравнению Гельмгольца:

$$\Delta E_z(x,y) + k^2(x,y) E_z(x,y) = 0. \tag{3.156}$$

Функция $E_z(x,y)$ является квазипериодической [5]:

$$E_z(x,y) = \sum_{m=-\infty}^{\infty} E_m(y) \exp(i\alpha_m x). \tag{3.157}$$

Подставляя (3.157) и (3.109) в (3.156), получаем систему $2N+1$ дифференциальных уравнений 2-го порядка:

$$\frac{\mathrm{d}^2 E_n(y)}{\mathrm{d}y^2} - \alpha_n^2 E_n(y) + \sum_{m=-N}^{N} c^{(1)}_{n-m}(y) E_m(y) = 0, \quad n = -\overline{N,N}. \tag{3.158}$$

Для поиска общего решения системы (3.158) необходимо найти $2(2N+1)$ линейно-независимых частных решений. При отсутствии модуляции ($\varepsilon(x,y) = \varepsilon$) базисные решения (3.158) имеют вид

$$E_p^{\pm}(y) = \exp\left(\pm i\tilde{\beta}_p y\right), \quad p = -\overline{N,N}. \tag{3.159}$$

Для согласования решения в зоне модуляции с решением (3.159) в зоне 3, определим $2(2N+1)$ векторов граничных условий для системы (3.158) в виде

$$\begin{aligned} &E^{\mp}_{mj}(0) = \delta_{m-j}, \\ &\frac{\partial E^{\mp}_{mj}(0)}{\partial y} = \mp i\beta_m \delta_{m-j}, \quad m,j = -\overline{N,N}. \end{aligned} \tag{3.160}$$

При этом общее решение системы (3.158) принимает вид

$$E_m(y) = \sum_{j=-N}^{N} C^{-}{}_j E^{-}{}_{mj}(y) + \sum_{j=-N}^{N} C^{+}{}_j E^{+}{}_{mj}(y), \quad m = -\overline{N,N}. \tag{3.161}$$

Подставляя (3.161) в (3.157) получим, что в зоне модуляции $E_z(x,y)$ имеет вид

$$E_z(x,y) = \sum_{m=-N}^{N} \left(\sum_{j=-N}^{N} C_j^- E_{mj^-}(y) + \sum_{j=-N}^{N} C_j^+ E_{mj}^+(y) \right) \exp(ik\alpha_m x). \qquad (3.162)$$

Для определения коэффициентов пропускания T_n и отражения R_n в (3.96), (3.98), из условий непрерывности $E_z(x,y)$ и производной $\partial E_z(x,y)/\partial y$ на границах зоны модуляции при $y=0$ и $y=a$, несложно получить, что при $y=0$, $T_p = C^-{}_p$, $C^+{}_p = 0$, и решение в зоне модуляции принимает вид

$$E(x,y) = \sum_{m=-N}^{N} \sum_{j=-N}^{N} T_j E_{mj}^-(y) \exp(ik\alpha_m x). \qquad (3.163)$$

На верхней границе зоны модуляции условия непрерывности функции $E_z(x,y)$ и ее производной имеют вид

$$\sum_{m=-N}^{N} \left(\sum_{j=-N}^{N} T_j E_{mj}^-(a) \right) \exp(ik\alpha_m x) = \sum_{m=-N}^{N} R_m \exp(i\beta_m a) \exp(i\alpha_m x) + \\ + \exp(-i\beta_0 a) \exp(i\alpha_0 x), \qquad (3.164)$$

$$\sum_{m=-N}^{N} \left(\sum_{j=-N}^{N} T_j \frac{\partial E^-{}_{mj}(a)}{\partial y} \right) \exp(i\alpha_m x) = \sum_{m=-N}^{N} i\beta_m R_m \exp(i\beta_m a) \exp(i\alpha_m x) - \\ - i\beta_0 \exp(-i\beta_0 a) \exp(i\alpha_0 x). \qquad (3.165)$$

Умножая обе части уравнений (3.164), (3.165) на $\exp(-i\alpha_p x)$ и интегрируя по периоду, получим две системы линейных уравнений вида

$$\sum_{j=-N}^{N} T_j E^-{}_{pj}(a) = R_p \exp(i\beta_p a) + \delta_p \exp(-i\beta_0 a), \quad p = -\overline{N,N}, \qquad (3.166)$$

$$\sum_{j=-N}^{N} T_j \frac{\partial E^-{}_{pj}(a)}{\partial y} = i\beta_p R_p \exp(i\beta_p a) - \delta_p i\beta_0 \exp(-i\beta_0 a), \quad p = -\overline{N,N}. \qquad (3.167)$$

Системы (3.166), (3.167) представим в матричном виде

$$E_{01} \cdot \mathbf{T} = E_{02}\mathbf{R} + \exp(-i\beta_0 a)\,\boldsymbol{\delta}, \\ E_{01pj} = E^-{}_{pj}(a), \quad E_{02pj} = \delta_{p-j} \cdot \exp(i\beta_p a), \quad p,j = -\overline{N,N}, \qquad (3.168)$$

$$E_{11} \cdot \mathbf{T} = E_{12}\mathbf{R} - i\beta_0 \exp(-i\beta_0 a)\,\boldsymbol{\delta}, \\ E_{11pj} = \frac{\partial E^-{}_{pj}(a)}{\partial y}, \quad E_{12pj} = \delta_{p-j} \cdot i\beta_p \exp(i\beta_p a), \quad p,j = -\overline{N,N}. \qquad (3.169)$$

Из уравнений (3.168), (3.169) получим вектора коэффициентов Рэлея в виде

$$\mathbf{T} = -2i\beta_0 \exp(-i\beta_0 a) \left(E_{11} - D_\beta \cdot E_{01}\right)^{-1} \boldsymbol{\delta}, \qquad (3.170)$$

$$\mathbf{R} = E_{02}^{-1} E_{01} \cdot \mathbf{T} + \exp(-2i\beta_0 a)\,\boldsymbol{\delta}, \qquad (3.171)$$

где D_β — диагональная матрица с элементами $i\beta_p$, $p = -\overline{N,N}$, а $\boldsymbol{\delta}$ — вектор-столбец, у которого все нули кроме единицы в центре.

Рассмотрим дифракцию на решетке с учетом подложки. $E_z(x,y)$ в зоне за подложкой имеет вид (3.98) с коэффициентами β_n, определенными в уравнении (3.97):

$$E_z(x,y) = \sum_{n=-\infty}^{\infty} T_n \exp\left(i\left(\alpha_n x - \beta_n y\right)\right). \tag{3.172}$$

Общее решение системы (3.158) в зоне 3 имеет вид

$$E_p^{\pm}(y) = C_p^{+} \exp\left(i\tilde{\beta}_p y\right) + C_p^{-} \exp\left(-i\tilde{\beta}_p y\right), \quad p = -\overline{N,N}. \tag{3.173}$$

Используя условия непрерывности функции $E_z(x,y)$ и ее производной на границе подложка — воздух, получим функции $E_p(y)$ в явном виде:

$$E_p(y) = T_p\left(\mu_p \exp\left(i\tilde{\beta}_p y\right) + \gamma_p \exp\left(-i\tilde{\beta}_p y\right)\right), \quad p = -\overline{N,N}, \tag{3.174}$$

где

$$\left.\begin{aligned} \gamma_p &= \frac{\tilde{\beta}_p + \beta_p}{2\tilde{\beta}_p} \cdot \exp\left(-ib\left(\beta_p - \tilde{\beta}_p\right)\right), \\ \mu_p &= \exp\left(-2i\tilde{\beta}_p b\right) \frac{\tilde{\beta}_p - \beta_p}{2\tilde{\beta}_p} \cdot \exp\left(-ib\left(\beta_p + \tilde{\beta}_p\right)\right). \end{aligned}\right\} \tag{3.175}$$

Для согласования решения в зоне модуляции с решением (3.174) в зоне 3, определим граничные условия для системы (3.158) в виде

$$\left.\begin{aligned} &E^{+}{}_{mj}(0) = \mu_m \delta_{m-j}, && E^{-}{}_{mj}(0) = \gamma_m \delta_{m-j}, \\ &\frac{\partial E^{+}{}_{mj}(0)}{\partial y} = i\tilde{\beta}_m \mu_m \delta_{m-j}, && \frac{\partial E^{-}{}_{mj}(0)}{\partial y} = -i\tilde{\beta}_m \gamma_m \delta_{m-j}, \quad m,j = -\overline{N,N}. \end{aligned}\right\} \tag{3.176}$$

Используя граничные условия на нижней границе зоны модуляции, получаем, что как и для ТМ-поляризации

$$T_p = C_p^{+} = C_p^{-}. \tag{3.177}$$

При этом общее решение системы (3.158) принимает вид

$$E_m(y) = \sum_{j=-N}^{N} T_j\left(E_{mj}^{-}(y) + E_{mj}^{+}(y)\right), \quad m = -N, N. \tag{3.178}$$

Далее, для определения коэффициентов пропускания T_n и отражения R_n в (3.96), (3.172), достаточно воспользоваться условиями непрерывности на верхней границе зоны модуляции. Из условий непрерывности несложно получить выражения для векторов коэффициентов Рэлея в виде (3.170), (3.171), где элементы матриц E_{01} и E_{11} имеют вид

$$\left.\begin{aligned} E_{01pj} &= E_{pj}^{-}(a) + E_{pj}^{+}(a), && p,j = -\overline{N,N}, \\ E_{11pj} &= \frac{\partial E^{-}{}_{pj}(a)}{\partial y} + \frac{\partial E^{+}{}_{pj}(a)}{\partial y}, && p,j = -\overline{N,N}. \end{aligned}\right\} \tag{3.179}$$

Для бинарной решетки коэффициенты Фурье (3.112) не зависят от y. В этом случае поле в зоне модуляции описывается системой дифференциальных уравнений 2-го порядка с постоянными коэффициентами. Запишем решение системы с постоянными коэффициентами для граничного условия с номером m в (3.160) или в (3.176) в матричном виде:

$$E_m^-(y) = \cos\left(\sqrt{A^{TE}}\cdot y\right) E_m^-(0) + \frac{\sin\left(\sqrt{A^{TE}}\cdot y\right)}{\sqrt{A^{TE}}} \frac{\partial E_m^-(0)}{\partial y}, \tag{3.180}$$

где

$$A_{i,j}^{TE} = -\alpha^2_{-(N+1)+i}\delta_{i,j} + c_{i-j}^{(1)}, \quad i,j = \overline{1,2N+1} \tag{3.181}$$

— матрица системы (3.158) с постоянными коэффициентами Фурье $c_n^{(1)}$, определенными в (3.112). Приведенное матричное представление позволяет выразить матрицы E_{01} и E_{11} в (3.170), (3.171) через матрицу (3.181) и граничные условия (3.160) или (3.176) в виде

$$\begin{aligned} E_{01} &= \cos\left(\sqrt{A^{TE}}\cdot a\right) E + \frac{\sin\left(\sqrt{A^{TE}}\cdot a\right)}{\sqrt{A^{TE}}} D, \\ E_{11} &= -\sqrt{A^{TE}}\sin\left(\sqrt{A^{TE}}\cdot a\right) E + \cos\left(\sqrt{A^{TE}}\cdot a\right) D. \end{aligned} \tag{3.182}$$

При расчете без учета подложки матрицы E и D в (3.182) являются единичной матрицей и диагональной матрицей с элементами $DE_{jj} = -i\beta_j$, $j = -\overline{N,N}$. При расчете с учетом подложки матрицы E и D являются диагональными матрицами с элементами $E_{jj} = \mu_j + \gamma_j$, $j = -\overline{N,N}$ и $D_{jj} = i\tilde{\beta}_j(\mu_j - \gamma_j)$, $j = -\overline{N,N}$.

Матрицы E_{01} и E_{11} также можно выразить через матричную экспоненту в виде, аналогичном представлению (3.148) для TM-поляризации. Действительно, заменяя систему $2N+1$ дифференциальных уравнений 2-го порядка (3.158) эквивалентной системой $4N+2$ дифференциальных уравнений 1-го порядка, получаем

$$\begin{bmatrix} E_{01} \\ E_{11} \end{bmatrix} = \exp\left(A_1^{TE}\cdot a\right) \begin{bmatrix} E \\ DE \end{bmatrix}, \tag{3.183}$$

где

$$A_1^{TE} = \begin{bmatrix} NU & E \\ -A^{TE} & NU \end{bmatrix},$$

NU — нулевые матрицы размером $(2N+1)\times(2N+1)$, а E — единичная матрица.

Матричные представления (3.182), (3.183) для бинарной решетки могут быть использованы для решения задачи дифракции на решетке с непрерывным профилем. Действительно, аппроксимируя, как для TM-поляризации, непрерывный профиль набором N бинарных слоев и используя для расчета дифракции на слое соотношения (3.182), (3.183), получим коэффициенты R_n, T_n в виде (3.170), (3.171). При расчете дифракции на бинарном слое по формуле (3.183), матрицы $E_{01} = E_{01}(a)$ и $E_{11} = E_{11}(a)$ в (3.170), (3.171) вычисляются через мультипликативный интеграл, как для TM поляризации. При расчете дифракции на слое по формуле (3.182),

матрицы $E_{01}(a)$ и $E_{11}(a)$ вычисляются по следующим рекуррентным формулам:

$$\begin{aligned} E_{01}(y_i) &= \cos\left(\sqrt{A^{TE}(\tilde{y}_i)}\cdot\Delta y_i\right)E_{01}(y_{i-1}) + \\ &\quad + \frac{\sin\left(\sqrt{A^{TE}(\tilde{y}_i)}\cdot\Delta y_i\right)}{\sqrt{A^{TE}(\tilde{y}_i)}}\cdot E_{11}(y_{i-1}), \\ E_{11}(y_i) &= -\sqrt{A^{TE}(\tilde{y}_i)}\sin\left(\sqrt{A^{TE}(\tilde{y}_i)}\cdot\Delta y_i\right)E_{01}(y_{i-1}) + \\ &\quad + \cos\left(\sqrt{A^{TE}(\tilde{y}_i)}\cdot\Delta y_i\right)\cdot E_{11}(y_{i-1}), \\ i &= \overline{1,N}, \quad E_{01}(y_0) = E_{01}(0) = E, \quad E_{11}(y_0) = E_{11}(0) = D, \end{aligned} \tag{3.184}$$

где y_i, $i=\overline{0,N}$, $y_0=0$, $y_N=a$ — разбиение зоны модуляции на N слоев, $\Delta y_i = y_{i+1} - y_i$, $\tilde{y}_i = (y_{i+1}+y_i)/2$, $A^{TE}(\tilde{y}_i)$ — матрица системы (3.158) при $c_n^{(1)}(y) = c_n^{(1)}(\tilde{y}_i)$, E и D — матрицы начальных условий.

Определим из закона сохранения энергии интенсивности порядков, соответствующих отраженным и прошедшим волнам в уравнениях (3.195), (3.197). Обозначим U_1 и U_2 множество индексов, соответствующих распространяющимся отраженным и прошедшим волнам:

$$U_1 = \left\{ n \middle| \left(\frac{\alpha_n}{k_0}\right)^2 < 1 \right\}, \quad U_2 = \left\{ n \middle| \left(\frac{\alpha_n}{k_0\sqrt{\varepsilon}}\right)^2 < 1 \right\}. \tag{3.185}$$

Рассмотрим прямоугольник D ограниченный отрезками прямых $y=0$, $y=a$ и $x=0$, $x=d$. Используя закон сохранения энергии, приравняем к нулю поток вектора Умова–Пойнтинга $\mathbf{S} = (c/8\pi)\,\mathrm{Re}\,[\mathbf{E},\mathbf{H}^*]$ через область D [8]. В результате получим следующие условия нормировки:

$$\cos(\theta) = \sum_{n\in U_1} |R_n|^2 \cos(\theta_n) + \sqrt{\varepsilon}\sum_{n\in U_2} |T_n|^2 \cos\left(\tilde{\theta}_n\right), \tag{3.186}$$

$$\cos(\theta) = \sum_{n\in U_1} |R_n|^2 \cos(\theta_n) + \frac{1}{\sqrt{\varepsilon}}\sum_{n\in U_2} |T_n|^2 \cos\left(\tilde{\theta}_n\right). \tag{3.187}$$

для TE- и TM-поляризации, соответственно. Углы θ_n, $\tilde{\theta}_n$ соответствуют направлениям отраженных и прошедших волн в (3.96),(3.98).

Согласно (3.186), (3.187) под интенсивностями отраженных и прошедших порядков будем понимать следующие нормированные значения коэффициентов Рэлея:

$$I_n^R = |R_n|^2 \frac{\cos(\theta_n)}{\cos(\theta)}, \quad I_n^T = \sqrt{\varepsilon}\,|T_n|^2 \frac{\cos\left(\tilde{\theta}_n\right)}{\cos(\theta)}, \quad \left(\sum_{n\in U_1} I_n^R + \sum_{n\in U_2} I_n^T = 1\right), \tag{3.188}$$

$$I_n^R = |R_n|^2 \frac{\cos(\theta_n)}{\cos(\theta)}, \quad I_n^T = |T_n|^2 \frac{\cos\left(\tilde{\theta}_n\right)}{\sqrt{\varepsilon}\cos(\theta)}, \quad \left(\sum_{n\in U_1} I_n^R + \sum_{n\in U_2} I_n^T = 1\right), \tag{3.189}$$

для TE- и TM-поляризаций, соответственно. Эти формулы аналогичны уравнению (3.46) для случая дифракции на идеально отражающей решетке.

При расчете решеток с учетом подложки положим в уравнениях (3.186) и (3.187) $\varepsilon = 1$ и $\theta_n = \tilde{\theta}_n$. При этом интенсивности порядков определяются по формулам

$$I_n^R = |R_n|^2 \frac{\cos(\theta_n)}{\cos(\theta)}, \quad I_n^T = |T_n|^2 \frac{\cos(\theta_n)}{\cos(\theta)}, \quad \left(\sum_{n \in U_1} I_n^R + \sum_{n \in U_2} I_n^T = 1 \right), \qquad (3.190)$$

и для ТЕ и для ТМ-поляризации.

Пример 3.3. Рассмотрим работу простейшей бинарной решетки с одним штрихом шириной $d/2$ на периоде d. На рис. 3.7 приведены зависимости интенсивностей 0-го отраженного и 0-го и -1-го прошедших порядков от высоты штриха a для ТЕ-поляризации. Расчет интенсивностей порядков проводился по формулам (3.170), (3.171), (3.182) без учета подложки при параметрах $d = \lambda_0$, $\theta = 30°$, удовлетворяющих условию Брэгга $2d\sin(\theta) = m\lambda_0$ при $m = 1$. Рис. 3.7 показывает ряд интересных моментов в работе бинарной решетки. Во-первых, при высоте штриха $a \approx 0{,}75\lambda_0$ энергия равномерно распределена между 0-м и $+1$-м прошедшими порядками, то есть при данной высоте решетку можно использовать в качестве делителя пучка. Во-вторых, при высоте штриха $a \approx 1{,}6\lambda_0$ около 93% энергии падающей волны перетекает из 0-го в -1-й прошедший порядок, то есть при данной высоте решетку можно использовать как дефлектор (отклонитель) пучка. По общему виду графиков можно также предположить, что интенсивности прошедших 0-го и $+1$-го порядков меняются периодически.

На рис. 3.8 приведена зависимость интенсивностей порядков бинарной решетки от высоты штриха для ТМ-поляризации. Для ТМ-поляризации бинарная решетка работает как делитель пучка при высоте штриха $a \approx 1{,}1\lambda_0$. При $a \approx 2{,}15\lambda_0$ более 95% энергии падающей волны содержится в -1-м прошедшем порядке, то есть решетка работает как дефлектор пучка.

На рис. 3.9 показана зависимость интенсивностей порядков бинарной решетки от угла падения θ для ТЕ-поляризации. Из рис. 3.9 видно, что при $\theta \in [25°, 35°]$ в -1-м прошедшем порядке содержится более 80% энергии, а в 0-м порядке — менее 10% энергии. Данный эффект позволяет при изменении угла падения использовать бинарную решетку как световой затвор.

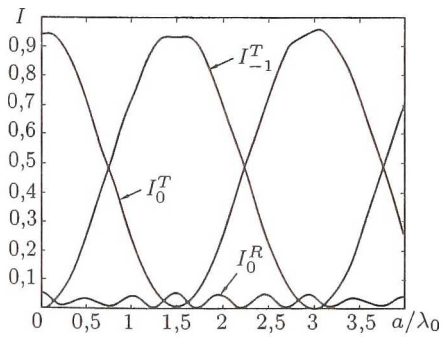


Рис. 3.7. Зависимость интенсивностей -1-го и 0-го прошедших $\left(I_{-1}^T, I_0^T\right)$ и 0-го отраженного $\left(I_0^R\right)$ порядков бинарной решетки от высоты штриха для ТЕ-поляризации при $d = \lambda_0$, $\theta = 30°$, $\varepsilon = 2{,}25$

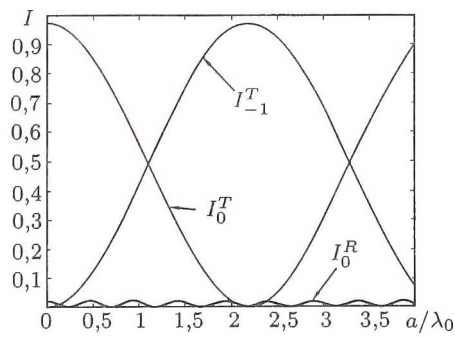


Рис. 3.8. Зависимость интенсивностей -1-го и 0-го прошедших $\left(I_{-1}^T, I_0^T\right)$ и 0-го отраженного $\left(I_0^R\right)$ порядков бинарной решетки от высоты штриха для ТМ-поляризации при $d = \lambda_0$, $\theta = 30°$, $\varepsilon = 2{,}25$

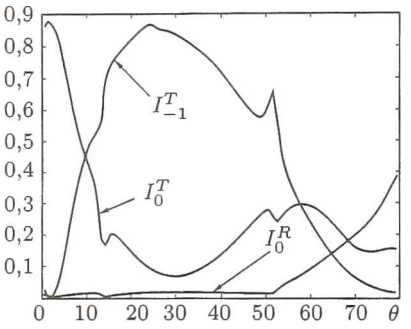


Рис. 3.9. Зависимость интенсивностей -1-го и 0-го прошедших $\left(I_{-1}^T, I_0^T\right)$ и 0-го отраженного $\left(I_0^R\right)$ порядков бинарной решетки от угла падения θ для ТЕ-поляризации при $d = \lambda_0$, $a = 1{,}2\lambda_0$, $\varepsilon = 2{,}5$

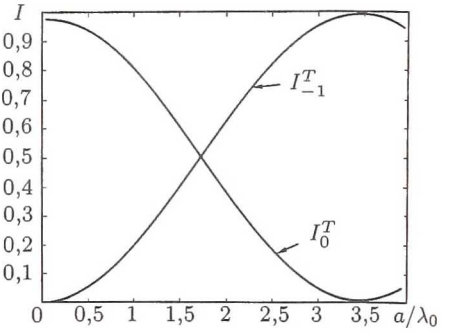


Рис. 3.10. Зависимость интенсивностей -1-го и 0-го прошедших порядков $\left(I_{-1}^T, I_0^T\right)$ симметричной треугольной решетки от высоты профиля для ТМ-поляризации при $d = \lambda_0$, $\theta = 30°$, $\varepsilon = 2{,}25$

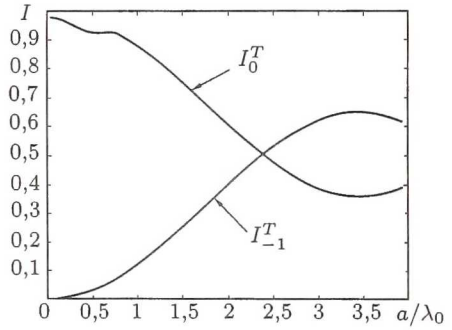


Рис. 3.11. Зависимость интенсивностей -1-го и 0-го прошедших порядков $\left(I_{-1}^T, I_0^T\right)$ правой пилообразной решетки от высоты штриха для ТМ-поляризации при $d = \lambda_0$, $\theta = 30°$, $\varepsilon = 2{,}25$

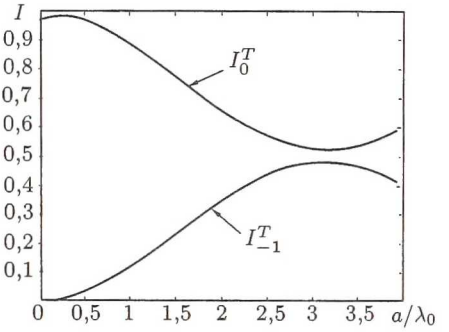


Рис. 3.12. Зависимость интенсивностей -1-го и 0-го прошедших порядков $\left(I_{-1}^T, I_0^T\right)$ левой пилообразной решетки от высоты штриха для ТМ-поляризации при $d = \lambda_0$, $\theta = 30°$, $\varepsilon = 2{,}25$

Пример 3.4. Рассмотрим работу решетки с треугольным профилем:

$$f(x) = \begin{cases} ax/d_1, & x \in [0, d_1], \\ -a(x-d)/(d-d_1) + a, & x \in [d, d_1], \end{cases} \tag{3.191}$$

где d_1 — абсцисса вершины решетки.

На рис. 3.10–3.12 для ТМ-поляризации приведены интенсивности порядков для симметричной треугольной решетки ($d_1 = d/2$), для правой пилообразной решетки ($d_1 = d$) и для левой пилообразной решетки ($d_1 = 0$). Расчет интенсивностей порядков проводился по формулам (3.131), (3.132), (3.152) без учета подложки при периоде $d = \lambda_0$ и угле падения $\theta = 30°$. При расчетах в уравнении (3.152) треугольный профиль был аппроксимирован шестнадцатью бинарными слоями, что соответствует квантованию треугольного профиля решетки по 17 уровням.

Из рис. 3.10 можно видеть, что при высоте профиля $a \approx 1{,}7\lambda$ энергия равномерно распределена между 0-м и -1-м прошедшими порядками и решетка работает как делитель пучка. При $a \approx 3{,}15\lambda$ более 95% энергии содержится в -1-м прошедшем порядке, то есть при данной высоте решетка работает как дефлектор пучка. Таким образом, симметричная треугольная и бинарные решетки обладают свойствами де-

лить пучок и собирать энергию в -1-м порядке. Для симметричной треугольной решетки данные эффекты происходят при примерно в 1,6 раза большей высоте профиля. Графики интенсивностей порядков для правой и левой пилообразных решеток на рис. 3.11 и 3.12 показывают, что варьируя высотой профиля нельзя сконцентрировать энергию в -1-м прошедшем порядке. Таким образом несимметричные пилообразные решетки, в отличие от симметричных бинарной и треугольной решеток, не могут быть использованы как дефлекторы пучка. При этом левая пилообразная решетка (рис. 3.12) в отличие от правой пилообразной решетки (рис. 3.11) не позволяет равномерно распределить энергию между 0-м и -1-м прошедшими порядками.

3.4. Градиентные методы решения обратной задачи расчета дифракционных решеток

3.4.1. Расчет отражающих решеток со ступенчатым профилем

Решение прямой задачи дифракции на решетке со ступенчатым профилем, изготовленной из идеально проводящего материала, приведено в разделе 3.1. Здесь рассматривается градиентный метод решения обратной задачи, состоящий в расчете параметров $(x_1, \ldots, x_K, c_1, \ldots, c_K, h_1, \ldots, h_K)$ профиля решетки (см. рис. 3.1) из условия формирования заданных значений интенсивностей дифракционных порядков [10]. Интенсивности порядков определены в уравнении (3.46) и пропорциональны квадратам модулей коэффициентов Рэлея.

Для построения градиентной процедуры расчета профиля дифракционной решетки введем некоторую функцию ошибки $\varepsilon(\mathbf{p})$, характеризующую отличие рассчитанных значений интенсивности I_n в дифракционных порядках от требуемых значений $\tilde{I}_n$:

$$\varepsilon(\mathbf{p}) = \varepsilon\left(\mathbf{I}(\mathbf{p}), \tilde{\mathbf{I}}\right), \tag{3.192}$$

где $\mathbf{p} = (x_1, \ldots, x_K, c_1, \ldots, c_K, h_1, \ldots, h_K)$ — вектор параметров дифракционной решетки, K — число штрихов на периоде решетки, $\mathbf{I} = \left(I_n\right)_{-M}^{M}$, $\tilde{\mathbf{I}} = \left(\tilde{I}_n\right)_{-M}^{M}$ — вектора рассчитанных и требуемых значений интенсивности порядков.

Градиентная процедура минимизации функции ошибки $\varepsilon(\mathbf{p})$ состоит в итерационной коррекции параметров профиля решетки по правилу

$$\mathbf{p}_{n+1} = \mathbf{p}_n - t \cdot \nabla\varepsilon(\mathbf{p}), \tag{3.193}$$

где n — номер итерации, t — шаг градиентного алгоритма,

$$\nabla\varepsilon(\mathbf{p}) = \left(\frac{\partial\varepsilon(\mathbf{p})}{\partial x_1}, \ldots, \frac{\partial\varepsilon(\mathbf{p})}{\partial x_K}, \frac{\partial\varepsilon(\mathbf{p})}{\partial c_1}, \ldots, \frac{\partial\varepsilon(\mathbf{p})}{\partial c_K}, \frac{\partial\varepsilon(\mathbf{p})}{\partial h_1}, \ldots, \frac{\partial\varepsilon(\mathbf{p})}{\partial h_K}\right)$$

— градиент функции ошибки.

Рассмотрим вычисление градиента функции ошибки $\nabla\varepsilon(\mathbf{p})$. Согласно (3.192) и (3.46), элементы вектора градиента $\partial\varepsilon(\mathbf{p})/\partial p_i$, где p_i — компонент (x_i, c_i или h_i)

вектора $\mathbf{p}$, имеют вид

$$\frac{\partial\varepsilon(\mathbf{p})}{\partial p_i}=\sum_{j=-M}^{M}\frac{\partial\varepsilon\left(\mathbf{I},\tilde{\mathbf{I}}\right)}{\partial I_j}\frac{\partial I_j(\mathbf{p})}{\partial p_i}=\sum_{j=-M}^{M}\frac{\partial\varepsilon\left(\mathbf{I},\tilde{\mathbf{I}}\right)}{\partial I_j}\frac{\partial}{\partial p_i}\left(\frac{\beta_j}{\beta_0}R_j\cdot R_j^*\right)=$$

$$=2\,\mathrm{Re}\left(\sum_{j=-M}^{M}\frac{\partial\varepsilon\left(\mathbf{I},\tilde{\mathbf{I}}\right)}{\partial I_j}\frac{\beta_j}{\beta_0}R_j^*\frac{\partial R_j(\mathbf{p})}{\partial p_i}\right). \quad (3.194)$$

Уравнение (3.194) удобно представить в виде реальной части скалярного произведения векторов:

$$\frac{\partial\varepsilon(\mathbf{p})}{\partial p_i}=2\,\mathrm{Re}\left(\frac{\partial\mathbf{R}(\mathbf{p})}{\partial p_i},\mathbf{L}\right), \quad (3.195)$$

где

$$\frac{\partial\mathbf{R}(\mathbf{p})}{\partial p_j}=\left(\frac{\partial R_j(\mathbf{p})}{\partial p_i}\right)_{-M}^{M},\quad \mathbf{L}=(L_j)_{-M}^{M},\quad L_j=\sum_{j=-M}^{M}\frac{\partial\varepsilon\left(\mathbf{I},\tilde{\mathbf{I}}\right)}{\partial I_j}\frac{\beta_j}{\beta_0}R_j. \quad (3.196)$$

Коэффициенты Рэлея R_j и косинусы углов β_j, определяющих направления распространения порядков, определяются из (3.31) и (3.16), соответственно. Для расчета вектора производных коэффициентов Рэлея в (3.195), продифференцируем (3.31) по переменной p_i:

$$\frac{\partial(A\cdot\mathbf{R})}{\partial p_i}=\frac{\partial A}{\partial p_i}\cdot\mathbf{R}+A\cdot\frac{\partial\mathbf{R}}{\partial p_i}=\frac{\partial\mathbf{D}}{\partial p_i}. \quad (3.197)$$

Согласно (3.197), вектор производных $\partial\mathbf{R}(\mathbf{p})/\partial p_i$ имеет вид

$$\frac{\partial\mathbf{R}}{\partial p_i}=A^{-1}\left(\frac{\partial\mathbf{D}}{\partial p_i}-\frac{\partial A}{\partial p_i}\cdot\mathbf{R}\right). \quad (3.198)$$

Далее, подставляя (3.198) в (3.195), получаем компоненты градиента функции ошибки в виде

$$\frac{\partial\varepsilon(\mathbf{p})}{\partial p_i}=2\,\mathrm{Re}\left(A^{-1}\left(\frac{\partial\mathbf{D}}{\partial p_i}-\frac{\partial A}{\partial p_i}\cdot\mathbf{R}\right),\mathbf{L}\right). \quad (3.199)$$

Матрицы A, $\partial A/\partial p_i$ и вектора $\mathbf{R}$, $\mathbf{D}$, $\partial\mathbf{D}/\partial p_i$ в (3.199) рассчитываются из уравнений (3.27)–(3.30) для TE-поляризации и из уравнений (3.27), (3.42)–(3.44) для TM-поляризации. В частности, для TE-поляризации производные $\partial A/\partial p_i$ и $\partial\mathbf{D}/\partial p_i$ могут быть легко получены из (3.27)–(3.30) в виде

$$\frac{\partial D_p}{\partial p_l}=\frac{\partial A_{p0}}{\partial p_l}, \quad (3.200)$$

$$\frac{\partial A_{ps}}{\partial x_l}=-\beta_s\frac{2k^2}{d}(\alpha_s-\alpha_p)c_l\sum_{n=1}^{\infty}\frac{\mathrm{tg}(\mu_{\mathrm{ln}}h_l)}{\mu_{\mathrm{ln}}}D_{ns}^l\left(D_{np}^l\right)^*, \quad (3.201)$$

$$\frac{\partial A_{ps}}{\partial h_l}=\beta_s\frac{2ik}{d}c_l\sum_{n=1}^{\infty}\frac{1}{\cos^2(\mu_{\mathrm{ln}}h_l)}D_{ns}^l\left(D_{np}^l\right)^*, \quad (3.202)$$

$$\frac{\partial A_{ps}}{\partial c_l} = \beta_s \frac{2ik}{d} \sum_{n=1}^{\infty} \left(\left(\frac{\mathrm{tg}\,(\mu_{\ln} h_l)}{\mu_{\ln}} + \frac{\pi n \left(-\sin\left(2\mu_{\ln} h_l\right) + 2\mu_{\ln} h_l\right)}{2c_l^2 \mu_{\ln}^3 \cos^2\left(\mu_{\ln} h_l\right)} \right) \times \right.$$

$$\left. \times D_{ns}^l \left(D_{np}^l\right)^* + ic_l \frac{\mathrm{tg}\,(\mu_{\ln} h_l)}{\mu_{\ln}} \left(\alpha_s - \alpha_p\right) \cdot \tilde{D}_{ns}^l \cdot \left(D_{np}^l\right)^* \right), \quad (3.203)$$

где

$$\tilde{D}_{ns}^l = \frac{1}{k} \exp\left(ik\alpha_n x_l\right) \int_0^k \sin\left(\frac{\pi m}{k}\xi\right) \exp\left(i\alpha_n c_l \xi\right) \mathrm{d}\xi. \quad (3.204)$$

В вышеприведенных уравнениях α_n и β_n имеют вид (3.15), (3.16). Для ТМ-поляризации расчет производных $\partial A/\partial p_i$, $\partial \mathbf{D}/\partial p_i$ аналогичен и состоит в дифференцировании уравнений (3.42)–(3.44).

Для исследования целесообразности расчета проводящих решеток в рамках электромагнитной теории, оценим работоспособность решеток, рассчитанных в приближении Кирхгофа для формирования $M = 2N + 1$ равных порядков. Для характеристики работы решеток будем использовать значения энергетической эффективности

$$E(M) = \sum_{j=-N}^{N} I_j \quad (3.205)$$

и среднеквадратичной ошибки формирования заданной равной интенсивности порядков

$$\delta(M) = \frac{1}{\bar{I}} \left[\frac{1}{M} \sum_{j=-N}^{N} \left(I_j - \bar{I}\right)^2 \right]^{1/2}, \quad (3.206)$$

где $\bar{I} = E(M)/M$ — среднее значение интенсивности.

Пример 3.5. Для исследования была выбрана 11-порядковая бинарная дифракционная решетка с глубиной штрихов $h = \lambda/4$ и с координатами штрихов $(x_1, c_1) = (0, 0{,}06857)$, $(x_2, c_2) = (0{,}20885, 0{,}23582)$, $(x_3, c_3) = (0{,}5293, 0{,}19171)$, $(x_4, c_4) = (0{,}72854, 0{,}13583)$. Приведенные координаты штрихов нормированы на период решетки. Согласно работе [11] данная решетка в приближении Кирхгофа имеет энергетическую эффективность $E = 76{,}6\%$ при среднеквадратичной ошибке формирования равной интенсивности порядков менее 1%. Для оценки применимости приближения Кирхгофа был проведен расчет интенсивностей порядков решетки при следующих значениях периода d: $7{,}2\lambda$, $15{,}2\lambda$, $25{,}2\lambda$, $35{,}2\lambda$. Расчет проводился для ТЕ-поляризации по формулам (3.27)–(3.32) при нормальном падении плоской волны. Значения среднеквадратичной ошибки δ и энергетической эффективности E составили 30,2% и 83,8% при $d = 7{,}2\lambda$, 17% и 78,9% при $d = 15{,}2\lambda$, 12,1% и 78,5% при $d = 25{,}2\lambda$, 10,6% и 83,8% при $d = 35{,}2\lambda$. Приведенные результаты расчетов показывают, что для 11-порядковой решетки при периоде $d > 15\lambda$ ошибка $\delta \sim 15\%$, то есть скалярное приближение дает приемлемую точность. В то же время при $d = 7{,}2\lambda$ расчет в скалярном приближении приводит к значительной ошибке $\delta > 30\%$. Проведенный пример наглядно демонстрирует актуальность точных процедур синтеза решеток в рамках электромагнитной теории при малых (относительно длины волны) периодах.

Пример 3.6. Точный градиентный метод (3.193), (3.199) был использован для расчета бинарных решеток с равными порядками. В качестве функции ошибки использовалась квадратичная функция ошибки

$$\varepsilon(\mathbf{p}) = \sum_{j=-N}^{N} \left(I_j(\mathbf{p}) - \tilde{I}_j\right)^2. \tag{3.207}$$

В табл. 3.1 приведены результаты расчета решеток при нормальном падении плоской волны для ТЕ-поляризации, которые показывают работоспособность градиентного метода. В частности, рассчитанная для $d = 7{,}2\lambda$ 11-порядковая решетка, в отличие от 11-порядковой решетки рассчитанной в скалярном приближении, имеет в 4 раза меньшую среднеквадратичную ошибку δ. Градиентный метод (3.193), (3.199) позволяет рассчитывать не только бинарные решетки, но и решетки с различной глубиной штрихов. В качестве примера была рассчитана 5-порядковая решетка, имеющая на периоде $d = 3{,}2\lambda$ два штриха с глубиной $0{,}1\lambda$ и $0{,}2584\lambda$ и с координатами: $(x_1, c_1) = (0{,}034224, 0{,}173889)$, $(x_2, c_2) = (0{,}465776, 0{,}308846)$. Значения энергетической эффективности E и среднеквадратичной ошибки δ для этой решетки составили 89,6% и 2,9%. Оптимизация с учетом глубины штрихов h_i дает возможность достичь высокой энергетической эффективности и малой ошибки δ при существенно большей ширине минимального штриха $\Delta = \min_i (c_i, x_{i+1} - (x_i + c_i))$. В частности, 5-порядковая решетка со штрихами разной глубины имеет почти такие же значения E и δ как у бинарной 5-порядковой решетки в табл. 3.1, но при почти в 2 раза большей ширине минимального штриха.

3.4.2. Расчет диэлектрических бинарных решеток

Решение прямой задачи дифракции на диэлектрической решетке рассмотрено в разделе 3.3. В данном пункте рассматривается градиентный метод решения обратной задачи расчета бинарной диэлектрической решетки [12]. Метод состоит в расчете

Таблица 3.1. Результаты градиентного расчета бинарных отражающих решеток в электромагнитном приближении

Число порядков	Число штрихов	Период (d/λ)	Глубина штрихов (h/λ)	Нормированные координаты штрихов $(x_i, c_i)/d$	Энергетическая эффективность E (%)	Среднеквадратичная ошибка d (%)
3	1	3,2	—	(0, 0,5)	92	0
5	2	3,2	0,25	(0, 0,089204), (0,397067, 0,29625)	87,2	3,5
7	2	3,2	0,2241	(0,098974, 0,453849), (0,751026, 0,146244)	99,4	1,9
9	3	5,2	0,2450	(0,117498, 0,095706), (0,401101, 0,066281), (0,643485, 0,289262)	91,2	4,1
11	5	7,2	0,2363	(0,043523, 0,030163), (0,196521, 0,094899), (0,331266, 0,013970), (0,381802, 0,161684), (0,651364, 0,282746)	86,5	7,5

координат штрихов профиля решетки (см. рис. 3.6) из условия формирования заданных интенсивностей дифракционных порядков $\tilde{I}_n$, $n = -\overline{M, M}$, соответствующих прошедшим волнам в уравнении (3.98). Интенсивности дифракционных порядков пропорциональны квадратам модулей коэффициентов пропускания T_n в (3.98):

$$I_n = t_n \cdot |T_n|^2 , \tag{3.208}$$

где коэффициенты t_n можно определить из уравнений (3.188)–(3.189).

Расчет профиля решетки состоит в градиентной минимизации функции ошибки (3.192), которая состоит в итерационной коррекции координат штрихов (вектора $\mathbf{p} = (x_1, \ldots, x_{2K})$) по правилу (3.193). Рассмотрим вычисление градиента функции ошибки $\nabla\varepsilon(\mathbf{p})$ для ТМ-поляризации. Согласно (3.192) частные производные $\partial\varepsilon(\mathbf{p})/\partial x_m$ имеют вид

$$\frac{\partial\varepsilon(\mathbf{p})}{\partial x_m} = \sum_{j=-M}^{M} \frac{\partial\varepsilon\left(\mathbf{I}, \tilde{\mathbf{I}}\right)}{\partial I_j} \frac{\partial I_j(\mathbf{p})}{\partial x_m} =$$

$$= 2\,\mathrm{Re}\left(\sum_{j=-M}^{M} \frac{\partial\varepsilon\left(\mathbf{I}, \tilde{\mathbf{I}}\right)}{\partial I_j} t_j T_j^* \frac{\partial T_j(\mathbf{p})}{\partial x_m} \right) = 2\,\mathrm{Re}\left(\frac{\partial \mathbf{T}(\mathbf{p})}{\partial x_m}, \mathbf{L} \right), \tag{3.209}$$

где

$$\frac{\partial \mathbf{T}(\mathbf{p})}{\partial x_m} = \left(\frac{\partial T_j(\mathbf{p})}{\partial x_m} \right)_{-M}^{M}, \tag{3.210}$$

$$\mathbf{L} = (L_j)_{-M}^{M}, \quad L_j = \sum_{j=-M}^{M} \frac{\partial\varepsilon\left(\mathbf{I}, \tilde{\mathbf{I}}\right)}{\partial I_j} t_j T_j. \tag{3.211}$$

Согласно (3.188) — (3.190), коэффициенты t_j в (3.209), (3.211) имеют вид

$$t_j = \frac{\cos\left(\tilde{\theta}_n\right)}{\sqrt{\varepsilon}\cos(\theta)}, \tag{3.212}$$

при расчете решетки без учета подложки, и

$$t_j = \frac{\cos(\theta_n)}{\cos(\theta)}, \tag{3.213}$$

при расчете решетки с учетом подложки.

Для расчета вектора производных $\partial\mathbf{T}(\mathbf{p})/\partial x_m$ в (3.209) продифференцируем (3.131):

$$\frac{\partial \mathbf{T}}{\partial x_m} = \frac{\partial}{\partial x_m}\left(2 \cdot \exp(-i\beta_0 a) \cdot (H_{01} - D_\beta \cdot H_{11})^{-1} \boldsymbol{\delta} \right) =$$

$$= -(H_{01} - D_\beta \cdot H_{11})^{-1} \left(\frac{\partial H_{01}}{\partial x_m} - D_\beta \cdot \frac{\partial H_{11}}{\partial x_m} \right) \mathbf{T}. \tag{3.214}$$

Для расчета производных $\partial H_{01}/\partial x_m$, $\partial H_{11}/\partial x_m$ воспользуемся аналитическим представлением матриц H_{01} и H_{11} в (3.148). Дифференцируя (3.148), получаем

$$\begin{bmatrix} \dfrac{\partial H_{01}}{\partial x_m} \\ \dfrac{\partial H_{11}}{\partial x_m} \end{bmatrix} = \frac{\partial}{\partial x_m}\left(\exp\left(A^{TM}\cdot a\right)\cdot C\right) = \frac{\partial}{\partial x_m}\left(\sum_{n=1}^{\infty}\frac{a^n}{n!}\left(A^{TM}\right)^n\right)\cdot C =$$

$$= \left(\sum_{n=1}^{\infty}\frac{a^n}{n!}\frac{\partial}{\partial x_m}\left(A^{TM}\right)^n\right)\cdot C =$$

$$= \left(\sum_{n=1}^{\infty}\frac{a^n}{n!}\left(\sum_{j=0}^{n-1}\left(A^{TM}\right)^{n-j-1}\frac{\partial A^{TM}}{\partial x_m}\left(A^{TM}\right)^j\right)\right)\cdot C. \quad (3.215)$$

При суммировании ряда в уравнении (3.215) производные $\partial\left(A^{TM}\right)^n/\partial x_m$ следует вычислять с использованием рекуррентной формулы

$$\frac{\partial}{\partial x_m}(A)^n = A\cdot\frac{\partial}{\partial x_m}(A)^{n-1} + \frac{\partial A}{\partial x_m}\cdot(A)^{n-1}. \quad (3.216)$$

При этом расчет производных матриц H_{01} и H_{11} по порядку сложности эквивалентен вычислению экспоненты в степени матрицы через ряд. Для расчета производной $\partial A^{TM}/\partial x_m$ воспользуемся представлением матрицы A^{TM} в уравнении (3.146). Дифференцируя (3.146), получаем

$$\frac{\partial A^{TM}}{\partial x_m} = \begin{bmatrix} NU & \dfrac{\partial F_1}{\partial x_m} \\ \dfrac{\partial F_2}{\partial x_m} & NU \end{bmatrix}, \quad (3.217)$$

где NU — нулевые матрицы с размером $(2N+1)\times(2N+1)$, а матрицы $\partial F_1/\partial x_m$, $\partial F_2/\partial x_m$ могут быть получены из (3.147) в виде

$$\frac{\partial F_1}{\partial x_m} = \frac{\partial c_{l-j}^{(1)}}{\partial x_m} = (-1)^m\frac{k_0^2(\varepsilon-1)}{d}\exp\left(-i\frac{2\pi}{d}(l-j)x_m\right), \quad l,j=\overline{1,2N+1}, \quad (3.218)$$

$$\frac{\partial F_2}{\partial x_m} = \alpha_{-(N+1)+l}\alpha_{-(N+1)+j}\frac{\partial c_{l-j}^{(2)}}{\partial x_m} = \alpha_{-(N+1)+l}\alpha_{-(N+1)+j}(-1)^m\frac{(1/\varepsilon-1)}{k_0^2 d}\times$$

$$\times\exp\left(-i\frac{2\pi}{d}(l-j)x_m\right), \quad l,j=\overline{1,2N+1}. \quad (3.219)$$

Таким образом, компоненты вектора градиента функции ошибки имеют вид реальной части скалярного произведения:

$$\frac{\partial\varepsilon(\mathbf{p})}{\partial x_m} = 2\,\mathrm{Re}\left(\left(H_{01}-D_\beta\cdot H_{11}\right)^{-1}\left(D_\beta\frac{\partial H_{11}}{\partial x_m}-\frac{\partial H_{01}}{\partial x_m}\right)\mathbf{T},\mathbf{L}\right), \quad (3.220)$$

где вектор $\mathbf{T}$ определен в (3.131), вектор $\mathbf{L}$ определен в (3.211), матрицы H_{01} и H_{11} аналитически выражаются через матрицу системы (3.111) и граничные условия (3.149), (3.151) в виде (3.148), а матрицы производных $\partial H_{01}/\partial x_m$, $\partial H_{11}/\partial x_m$ определены в уравнениях (3.216)–(3.219).

Отметим, что высота профиля решетки a также может рассматриваться как параметр оптимизации. Производная функции ошибки по параметру a также имеет вид (3.220), где матрицы $\partial H_{01}/\partial a$, $\partial H_{11}/\partial a$ могут быть получены из (3.148) в виде

$$\begin{bmatrix} \dfrac{\partial H_{01}}{\partial a} \\ \dfrac{\partial H_{11}}{\partial a} \end{bmatrix} = \frac{\partial}{\partial a}\left(\exp\left(A^{TM}\cdot a\right)\cdot C\right) = A^{TM}\cdot\exp\left(A^{TM}\cdot a\right)\cdot C. \tag{3.221}$$

Рассмотрим вычисление градиента функционала невязки $\nabla\varepsilon\left(\mathbf{p}\right)$ для ТЕ-поляризации. Для ТЕ-поляризации компоненты вектора градиента для функции ошибки также имеют вид (3.209), (3.211), где вектор $\mathbf{T}$ определен в (3.170), а коэффициенты t_j имеют вид

$$t_j = \sqrt{\varepsilon}\frac{\cos\left(\tilde{\theta}_n\right)}{\cos\left(\theta\right)}, \tag{3.222}$$

при расчете решетки без учета подложки, и

$$t_j = \frac{\cos\left(\theta_n\right)}{\cos\left(\theta\right)}, \tag{3.223}$$

при расчете решетки с учетом подложки.

Для вычисления производных $\partial\mathbf{T}\left(\mathbf{p}\right)/\partial x_m$ в (3.209) продифференцируем уравнение (3.170):

$$\frac{\partial\mathbf{T}}{\partial x_m} = -\left(E_{11}-D_\beta\cdot E_{01}\right)^{-1}\left(\frac{\partial E_{11}}{\partial x_m}-D_\beta\cdot\frac{\partial E_{01}}{\partial x_m}\right)\mathbf{T}. \tag{3.224}$$

Для расчета производных $\partial E_{01}/\partial x_m$, $\partial E_{11}/\partial x_m$ от матриц E_{01} и E_{11} воспользуемся аналитическими представлениями (3.182) для матрицы E_{01} и E_{11}. Согласно (3.182) матрицы $\partial E_{01}/\partial x_m$, $\partial E_{11}/\partial x_m$ имеют вид

$$\begin{aligned}
\frac{\partial E_{01}}{\partial x_m} &= \frac{\partial}{\partial x_m}\left(\cos\left(\sqrt{A^{TE}}\cdot a\right)E+\frac{\sin\left(\sqrt{A^{TE}}\cdot a\right)}{\sqrt{A^{TE}}}\cdot D\right) = \\
&= \frac{\partial}{\partial x_m}\left(\left(\sum_{n=0}^{\infty}(-1)^n\frac{a^{2n}}{(2n)!}\left(A^{TE}\right)^n\right)E+\left(\sum_{n=0}^{\infty}(-1)^n\frac{a^{2n+1}}{(2n+1)!}\left(A^{TE}\right)^n\right)D\right) = \\
&= \left(\sum_{n=0}^{\infty}(-1)^n\frac{a^{2n}}{(2n)!}\frac{\partial\left(A^{TE}\right)^n}{\partial x_m}\right)E+\left(\sum_{n=0}^{\infty}(-1)^n\frac{a^{2n+1}}{(2n+1)!}\frac{\partial\left(A^{TE}\right)^n}{\partial x_m}\right)D,
\end{aligned} \tag{3.225}$$

$$\begin{aligned}
\frac{\partial E_{11}}{\partial x_m} &= \frac{\partial}{\partial x_m}\left(-\sqrt{A^{TE}}\sin\left(\sqrt{A^{TE}}\cdot a\right)E+\cos\left(\sqrt{A^{TE}}\cdot a\right)\cdot D\right) = \\
&= \frac{\partial}{\partial x_m}\left(\left(\sum_{n=1}^{\infty}(-1)^n\frac{a^{2n-1}}{(2n-1)!}\left(A^{TE}\right)^n\right)E+\left(\sum_{n=0}^{\infty}(-1)^n\frac{a^{2n}}{(2n)!}\left(A^{TE}\right)^n\right)\cdot D\right) = \\
&= \left(\sum_{n=1}^{\infty}(-1)^n\frac{a^{2n-1}}{(2n-1)!}\frac{\partial\left(A^{TE}\right)^n}{\partial x_m}\right)E+\left(\sum_{n=0}^{\infty}(-1)^n\frac{a^{2n}}{(2n)!}\frac{\partial\left(A^{TE}\right)^n}{\partial x_m}\right)\cdot D,
\end{aligned} \tag{3.226}$$

где

$$\frac{\partial \left(A^{TE}\right)^n}{\partial x_m} = \sum_{j=0}^{n-1} \left(A^{TE}\right)^{n-j-1} \frac{\partial A^{TE}}{\partial x_m} \left(A^{TE}\right)^j. \tag{3.227}$$

При суммировании рядов в уравнениях (3.225), (3.226) производные $\partial \left(A^{TE}\right)^n / \partial x_m$ следует вычислять по рекуррентной формуле (3.216). При этом расчет производных матриц E_{01} и E_{11} по порядку сложности эквивалентен вычислению через ряд косинуса и синуса от матрицы. Матрица $\partial A^{TE}/\partial x_m$ в (3.225), (3.226) может быть получена из (3.180) в виде

$$\frac{\partial A^{TE}}{\partial x_m} = \frac{\partial c_{l-j}^{(1)}}{\partial x_m} = (-1)^m \frac{k_0^2 (\varepsilon - 1)}{d} \exp\left(-i \frac{2\pi}{d} (l-j) x_m\right), \quad l, j = \overline{1, 2N+1}. \tag{3.228}$$

Для ТЕ-поляризации компоненты вектора градиента $\varepsilon(\mathbf{p})$ имеют вид

$$\frac{\partial \varepsilon(\mathbf{p})}{\partial x_m} = 2\,\mathrm{Re}\left(\begin{array}{cc} (E_{11} - D_\beta & E_{01}) \end{array}\right)^{-1} \left(D_\beta \frac{\partial E_{01}}{\partial x_m} - \frac{\partial E_{11}}{\partial x_m}\right) \mathbf{T}, \mathbf{L}\Big), \tag{3.229}$$

где вектор $\mathbf{T}$ определен уравнением (3.170), вектор $\mathbf{L}$ определен уравнениями (3.211), (3.221), (3.222), матрицы E_{01} и E_{11} аналитически выражаются через матрицу системы (3.180) и граничные условия (3.100) или (3.176) в виде (3.182), а матрицы производных $\partial E_{01}/\partial x_m$, $\partial E_{11}/\partial x_m$ определены уравнениями (3.225), (3.226).

При минимизации функции невязки с учетом высоты профиля, производная $\partial \varepsilon(\mathbf{p})/\partial a$ также имеет вид (3.229), где матрицы $\partial E_{01}/\partial a$, $\partial E_{11}/\partial a$ могут быть получены из (3.182) в виде

$$\frac{\partial E_{01}}{\partial a} = \frac{\partial}{\partial a}\left(\cos\left(\sqrt{A^{TE}} \cdot a\right) E + \frac{\sin\left(\sqrt{A^{TE}} \cdot a\right)}{\sqrt{A^{TE}}} \cdot D\right) =$$

$$= -\sqrt{A^{TE}} \cdot \sin\left(\sqrt{A^{TE}} \cdot a\right) E + \cos\left(\sqrt{A^{TE}} \cdot a\right) \cdot D, \tag{3.230}$$

$$\frac{\partial E_{11}}{\partial a} = \frac{\partial}{\partial a}\left(-\sqrt{A^{TE}} \sin\left(\sqrt{A^{TE}} \cdot a\right) E + \cos\left(\sqrt{A^{TE}} \cdot a\right) \quad D\right) =$$

$$= -A^{TE} \cdot \cos\left(\sqrt{A^{TE}} \cdot a\right) E - \sqrt{A^{TE}} \cdot \sin\left(\sqrt{A^{TE}} \cdot a\right) \cdot D. \tag{3.231}$$

Для исследования целесообразности расчета диэлектрических решеток в электромагнитном приближении предварительно был проведен анализ работы решеток, рассчитанных в приближении Кирхгофа для формирования $M = 2N + 1$ равных порядков. Для характеристики работы решеток были использованы значения энергетической эффективности E (см. (3.205)) и среднеквадратичной ошибки δ формирования заданной равной интенсивности порядков (см. (3.206)).

Пример 3.7. Для исследования были выбраны 11 и 7-порядковые дифракционные решетки с глубиной штриха, соответствующей фазе π и с координатами штрихов $(x_1, x_2) = (0, 0{,}06857)$; $(x_3, x_4) = (0{,}20885, 0{,}44467)$; $(x_5, x_6) = (0{,}5293, 0{,}72101)$; $(x_7, x_8) = (0{,}72854, 0{,}86437)$ и $(x_1, x_2) = (0, 0{,}23191)$; $(x_3, x_4) = (0{,}42520, 0{,}52571)$, соответственно. Приведенные координаты штрихов нормированы на период решетки. Согласно [11], в скалярном приближении данные решетки формируют 11 и 7 порядков с энергетическими эффективностями $E(11) = 76{,}6\%$ и $E(7) = 78{,}6\%$ при

неравномерности интенсивности порядков менее 1%. В табл. 3.2 для указанных решеток приведены значения E и δ, рассчитанные в рамках электромагнитной теории при различных длинах периода. Значения E и δ в таблице приведены парами и соответствуют ТМ- и ТЕ- поляризации. Расчет интенсивности прошедших порядков проводился по формулам (3.131), (3.148) и (3.170), (3.182) при $\varepsilon = 2{,}25$ и $\theta = 0°$. Согласно данным табл. 3.2, 11-порядковая решетка фактически не работоспособна при периоде $d < 20\lambda_0$. При этом δ становится менее 10% только при $d > 50\lambda_0$. Для 7-порядковой решетки ошибка становится менее 10% уже при $d > 20\lambda_0$. Лучшая работа 7-порядковой решетки при малых значениях d/λ_0 объясняется большим размером штрихов. В частности для 11-порядковой решетки минимальная ширина штриха $\Delta = 0{,}008d$, а для 7-порядковой решетки $\Delta = 0{,}1d$. Проведенный расчет наглядно демонстрирует актуальность точных процедур синтеза решеток.

Разработанный градиентный метод был использован для расчета бинарных диэлектрических решеток ($\varepsilon = 2{,}25$) с равными порядками без учета подложки. В качестве функции невязки была использована квадратичная функция ошибки (3.207). В табл. 3.3 приведены результаты расчетов решеток с периодом $d = 5{,}5\lambda_0$ при нормальном падении для ТЕ- и ТМ-поляризации, которые показывают, что точный электромагнитный расчет дает решетки, существенно отличающиеся от ранее рассмотренных 11- и 7-порядковых решеток, рассчитанных в скалярном приближении. Использование градиентной процедуры в рамках электромагнитной теории позволяет, как и в скалярном случае, снизить среднеквадратичную ошибку до 1–5%, но при большей на 5–10% энергетической эффективности [11, 13–15]. В частности, рассчитанные 11-порядковые решетки с 3 штрихами имеют энергетическую эффективность более 90%.

Пример 3.8. В скалярном приближении распределение интенсивности в порядках дифракции бинарной решетки может быть только симметричным. Примеры точного электромагнитного расчета поля от простейших решеток, рассмотренные в пунктах 3.1–3.3, показывают, что при параметрах $d = \lambda_0$, $\theta = 30°$, удовлетворяющих условию Брэгга, симметричные решетки позволяют сконцентрировать излучение в -1-м порядке. Такая асимметрия возможна не только при малых периодах. В табл. 3.4 приведены результаты градиентного расчета бинарных дефлекторов с максимумом в -1-м порядке при нормальном падении для ТЕ-поляризации. В качестве функции ошибки была использована квадратичная функция ошибки $\varepsilon(\mathbf{p}) = (1 - I_{-1}(\mathbf{p}))^2 \to \min$.

Таблица 3.2. Характеристики работы решеток, рассчитанных в приближении Кирхгофа ($\varepsilon = 2{,}25, \theta = 0°$)

Период решетки (d/λ_0)	11-порядковая решетка		7-порядковая решетка	
	E(%) (ТМ/ТЕ)	δ(%) (ТМ/ТЕ)	E(%) (ТМ/ТЕ)	δ(%) (ТМ/ТЕ)
5,5	90,3 / 82,9	95,9 / 144,0	81,3 / 79,0	35,8 / 38,2
10	78,9 / 77,6	42,7 / 52,3	76,2 / 75,9	19,6 / 22,2
15	75,9 / 75,9	28,6 / 34,2	75,9 / 75,5	13,4 / 14,1
20	75,4 / 75,4	22,4 / 25,1	75,6 / 75,5	11,4 / 11,7
25	74,9 / 74,8	18,0 / 20,4	75,5 / 75,4	9,1 / 8,4
30	74,6 / 74,6	16,3 / 16,9	75,5 / 75,4	7,8 / 7,1
50	74,0 / 74,0	9,9 / 8,6	75,4 / 75,4	4,8 / 4,1

Таблица 3.3. Результаты градиентного расчета бинарных диэлектрических решеток в электромагнитном приближении ($d = 5{,}5\lambda_0$, $\varepsilon = 2{,}25$, $\theta = 0°$)

Число порядков M	Число штрихов K	Высота штрихов (a/λ_0)	Координаты профиля	$E(\%)$	$\delta(\%)$
ТМ-поляризация					
5	2	0,9	(0,1820, 0,4822), (0,5544, 0,8546)	80,1	3,1
7	2	0,9	(0,2454, 0,4367), (0,5999, 0,7912)	85,1	1,0
9	3	1,56	(0,0925, 0,1963), (0,3728, 0,4786), (0,6155, 0,7185)	96,1	0,6
11	3	1,6	(0,1515, 0,3435), (0,4845, 0,5753), (0,7162, 0,9046)	94,0	5,1
ТЕ-поляризация					
5	2	0,8	(0,1798, 0,4816), (0,5550, 0,8567)	80,7	2,5
7	2	0,875	(0,2579, 0,4297), (0,6070, 0,7787)	83,8	1,1
9	3	1,0	(0,0091, 0,1924), (0,3145, 0,4811), (0,7769, 0,9547)	89,7	3,6
11	3	1,57	(0,0444, 0,3390), (0,5033, 0,5567), (0,8259, 0,8792)	91,5	4,3

Табл. 3.4 показывает, что в -1-м порядке можно собрать более 80% энергии и при «больших» периодах. При этом для достижения высокой ($> 80\%$) эффективности на периоде $d = p\lambda_0$ необходимо порядка p штрихов.

Решетки, позволяющие сконцентрировать излучение в -1-м порядке, имеют большое практическое значение. Рассмотрим возможность использования подобных решеток для расчета различных фокусирующих ДОЭ, например линз. В скалярном приближении для расчета ДОЭ обычно используется приближение тонкого оптического элемента, описывающее ДОЭ через фазовую функцию. Пусть $\varphi(x), x \in [0, d]$ — фазовая функция ДОЭ, рассчитанная в приближении геометрической оптики из условия фокусировки в заданную область. Приближение геометрической оптики основывается на уравнении наклонов, определяющем направление преломленного луча через фазовую функцию ДОЭ в виде

$$\sin(\theta_r) = \frac{1}{k_0\sqrt{\varepsilon}}\left(\frac{\mathrm{d}\varphi(x)}{\mathrm{d}x} + \frac{\mathrm{d}\varphi_0(x)}{\mathrm{d}x}\right) = \frac{\lambda}{2\pi}\left(\frac{\mathrm{d}\varphi(x)}{\mathrm{d}x} + \frac{\mathrm{d}\varphi_0(x)}{\mathrm{d}x}\right), \tag{3.232}$$

где $\lambda = \lambda_0/\sqrt{\varepsilon}$ — длина волны в среде, $\varphi_0(x) = k_0 \sin(\theta) \cdot x$ — фазовая функция освещающего пучка при $y = 0$. Уравнение (3.232) совпадает с уравнением дифракционной решетки:

$$\sqrt{\varepsilon} \cdot \sin(\theta_m) = \sin(\theta) + m\lambda_0/d, \tag{3.233}$$

Таблица 3.4. Результаты градиентного расчета бинарных дефлекторов, концентрирующих излучение в -1-м порядке ($\varepsilon = 2{,}25$, $\theta = 0°$, ТЕ-поляризация)

Период (d/λ_0)	Число штрихов K	Высота штрихов (a/λ_0)	Координаты профиля	$E(\%)$
3,5	2	1,68	(0,2596, 0,4378), (0,6082, 0,6754), (0,8469, 0,8780)	83,5
4,5	4	1,69	(0,2617, 0,4009), (0,5426, 0,6043), (0,7001, 0,7361), (0,8521, 0,8734)	87,7
5,5	5	1,78	(0,2714, 0,3982), (0,4997, 0,5565), (0,6100, 0,6473), (0,7243, 0,7560), (0,8793, 0,8984)	87,6
6,5	5	1,5	(0,1809, 0,4334), (0,4717, 0,5302), (0,6113, 0,6530), (0,7566, 0,7845), (0,8997, 0,9142)	80,0

при периоде $d = 2\pi/|\,\mathrm{d}\varphi(x)/\,\mathrm{d}x|$ и m равном 1 или -1. Знак m определяется знаком производной фазы: $m = \operatorname{sign}(\,\mathrm{d}\varphi\,(x)/\,\mathrm{d}x)$. Данное значение периода совпадает с размером зоны ДОЭ Δx, определяемой из условия

$$\Delta\varphi\,(x) = 2\pi \approx \left|\frac{\mathrm{d}\varphi\,(x)}{\mathrm{d}x}\right| \Delta x \Rightarrow \Delta x = \frac{2\pi}{\left|\dfrac{\mathrm{d}\varphi\,(x)}{\mathrm{d}x}\right|}.$$

Таким образом, приближение геометрической оптики предполагает, что каждая зона ДОЭ работает как «однопорядковая» решетка, концентрирующая излучение в -1-м или $+1$-м порядках. В скалярном приближении Кирхгофа однопорядковой

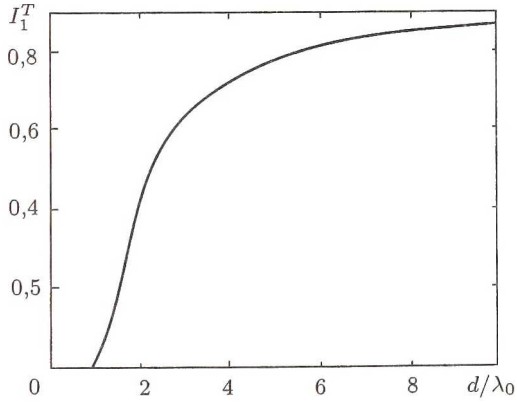


Рис. 3.13. Интенсивность 1-го порядка правой пилообразной решетки в зависимости от длины периода для ТЕ-поляризации при $\varepsilon = 2{,}25$, $\theta = 0°$ и высоте профиля $a = \lambda_0/(\sqrt{\varepsilon} - 1)$

решетке соответствуют левая или правая пилообразная решетка. При малых зонах геометрооптическое и скалярное приближения неработоспособны, поскольку неработоспособной оказывается пилообразная решетка. В частности, для правой пилообразной решетки с высотой $a = \lambda_0/(\sqrt{\varepsilon} - 1)$ интенсивность 1-го порядка стремится к нулю при $d \to \lambda_0$ (рис. 3.13).

Можно предположить, что при замене непрерывных зон ДОЭ бинарными решетками, рассчитанными в электромагнитном приближении из условия максимизации -1-го или 1-го порядка, получится высокоэффективный бинарный ДОЭ. Хотя формулы (3.232), (3.233) являются лишь эвристическим обоснованием метода замены зон ДОЭ «однопорядковой» решеткой, на практике метод оказывается работоспособным. В частности, в работе [19] рассмотрен расчет внеосевого сегмента бинарной отражающей линзы, имеющей для случая ТЕ-поляризации эффективность $\sim 90\%$. Расчет линзы в [19] основывался на замене зон непрерывной дифракционной линзы простейшей отражающей бинарной решеткой с одним штрихом. Для ТЕ-поляризации при $d \sim \lambda_0$, $\theta \sim 30°$ и высоте штриха $a \sim 0{,}25\lambda_0$ такая решетка концентрирует излучение в -1-м порядке (см. рис. 3.2).

Пример 3.9. Проиллюстрируем работоспособность метода на примере расчета преломляющей линзы. Фазовая функция дифракционной линзы имеет вид

$$\varphi(x) = \mathrm{mod}_{2\pi}\left(\frac{2\pi}{\lambda_0}\left(-x\sin(\theta) - \sqrt{\varepsilon}\sqrt{(x-x_0)^2 + f^2} + \mathrm{const}\right)\right), \quad x \in [0, d], \tag{3.234}$$

где (x_0, f) — координаты фокуса. При $d = 20\lambda_0$, $\theta = 30°$ и $(x_0, f) = (-20\lambda_0, -85\lambda_0)$ фазовая функция линзы имеет 20 зон с размером $0{,}89\lambda_0 \leqslant \Delta x \leqslant 1{,}17\lambda_0$. Заменяя зоны линзы (3.234) бинарной решеткой, где

$$h_2(x) = \begin{cases} 0, & x \in [0, \Delta x/2), \\ a\lambda_0, & x \in [\Delta x/2, \Delta x), \end{cases} \tag{3.235}$$

получаем бинарный рельеф. Для ТМ-поляризации при $0{,}89\lambda_0 \leqslant \Delta x \leqslant 1{,}17\lambda_0$, $a \sim 2\lambda_0$, $\theta = 30°$ решетка (3.235) концентрирует излучение в -1-м порядке (см. рис. 3.8). На рис. 3.14 и 3.15*а*) приведены полученный бинарный рельеф линзы и распределение интенсивности

$$I(x,f) = |\mathrm{Re}\,(E_x(x,f)\cdot H_z^*(x,f))| =$$
$$= \frac{1}{k_0\varepsilon}\left|\mathrm{Im}\left\{\left(\sum_{n=-N}^{N} T_n\tilde{\beta}_n \cdot \exp\left(i(\alpha_n x - \tilde{\beta}_n f)\right)\right)\left(\sum_{n=-N}^{N} T_n \exp\left(i(\alpha_n x - \tilde{\beta}_n f)\right)\right)^*\right\}\right|,$$

в фокальной плоскости $y = -85\lambda_0$ при $N = 57$. Указанное выражение для интенсивности $I(x,f)$ пропорционально y-компоненте вектора Умова–Пойнтинга. Для сравнения на рис. 3.15*б*) приведено распределение интенсивности для линзы с непрерывным рельефом,

$$h(x) = \frac{\sqrt{\varepsilon - \sin(\theta)}}{\varepsilon - \cos(\theta)\sqrt{\varepsilon - \sin(\theta)}}\ \mathrm{mod}_{\lambda_0}\left(\mathrm{const} - x\sin(\theta) - \sqrt{\varepsilon}\sqrt{(x-x_0)^2 + f^2}\right), \tag{3.236}$$

$x \in [0, d]$, рассчитанным в приближении тонкого оптического элемента. В распределении интенсивности на рис. 3.15*б* отсутствует четкий фокальный пик интенсивности, как на рис. 3.15*а* с шириной $5\lambda_0$, сформированный бинарной линзой.

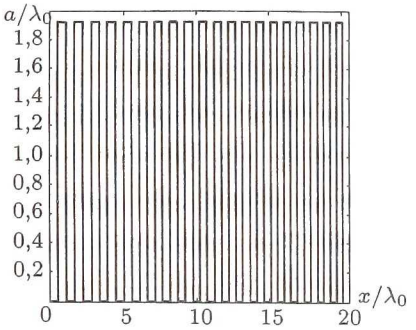


Рис. 3.14. Бинарная линза с апертурой $d = 20\lambda_0$ для фокусировки в точку $(x_0, f) = (-20\lambda_0, -85\lambda_0)$ при $\theta = 30°$, $\varepsilon = 2{,}25$

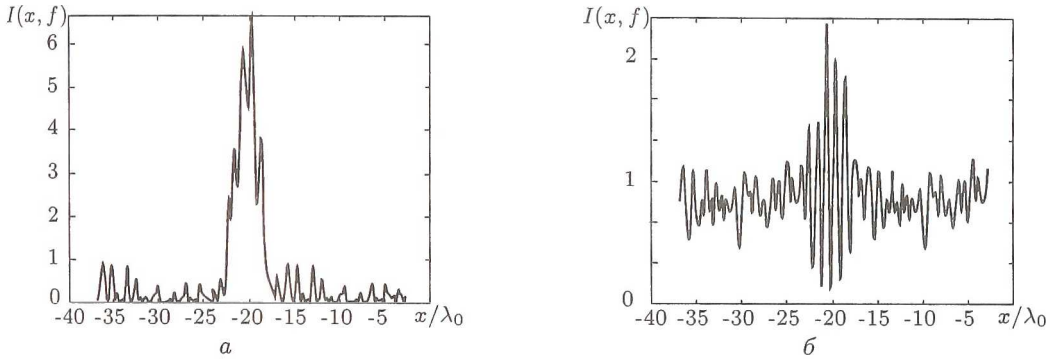


Рис. 3.15. Распределения интенсивности в фокальной плоскости (при $y = -85\lambda_0$) для бинарной линзы на рис. 3.14 (*а*) и для непрерывной линзы (3.236) (*б*) для ТМ- поляризации

Приведенный пример расчета линзы демонстрирует возможности и практическую важность задачи расчета решеток с максимумом энергии в -1-м или 1-м порядках. Большие, по сравнению с длиной волны зоны, требуют более сложных решеток с несколькими штрихами. При этом актуален расчет однопорядковых решеток для некоторого интервала периодов $[d_{\min}, d_{\max}]$, определяющего изменение размера зон ДОЭ. Расчет бинарной решетки с максимальной интенсивностью $I_{-1}(\mathbf{p}, d)$ в -1-м порядке при $d \in [d_{\min}, d_{\max}]$ может быть проведен рассмотренным градиентным методом с функцией ошибки определенной, например, в виде

$$\varepsilon(\mathbf{p}) = \int_{d_{\min}}^{d_{\max}} \left(1 - I_{-1}(\mathbf{p}, \xi)\right)^2 \mathrm{d}\xi \to \min, \tag{3.237}$$

где $\mathbf{p}$ — вектор координат штрихов решетки.

Пример 3.10. Простейшая решетка (3.235) с одной ступенькой может быть использована только при размерах зон близких к длине волны. Приведем пример расчета более сложной бинарной линзы с большими зонами и большим интервалом изменения размеров зон. Линза с непрерывным рельефом (3.236) при следующих параметрах: апертура линзы $d = 30\lambda_0$, координаты точки фокусировки $(x_0, f) = (0, -130\lambda_0)$, $\theta = 10°$, $\varepsilon = 2{,}25$, имеет 10 полных зон с размером от $5\lambda_0$, в начале апертуры, до $2\lambda_0$ на краю апертуры. Для расчета бинарной линзы требуется рассчитать решетку из условия максимума интенсивности в -1-м порядке для интервала периодов $[2\lambda_0, 5\lambda_0]$. Градиентным методом для ТМ-поляризации была рассчитана

решетка с тремя штрихами из условия максимума -1-го порядка для указанного интервала периодов и вышеприведенной функции ошибки (3.237). Нормированные на период координаты штрихов равны $\mathbf{p} = (0{,}0429, 0{,}2981, 0{,}4556, 0{,}5771, 0{,}7745, 0{,}8276)$ при высоте штрихов $a = 2{,}07\lambda_0$. Данная решетка при $d \in [2\lambda_0, 5\lambda_0]$ концентрирует в -1-м порядке не менее 80% энергии. На рис. 3.16*а* приведен профиль бинарной линзы, полученный заменой зон линзы (3.236), рассчитанной бинарной решеткой, а на рис. 3.16*б* — распределение интенсивности, сформированное бинарной линзой при $f = -130\lambda_0$. Для сравнения на рис. 3.16*а,б*, также показаны профиль непрерывной дифракционной линзы (3.236) и распределение интенсивности:

$$I(x,f) = \frac{\sqrt{\varepsilon}}{\lambda_0 f}\left|\int_0^d \exp\left(-2\pi i\frac{\sqrt{\varepsilon}}{\lambda_0 f}xu\right)\mathrm{d}u\right|^2 = \frac{d^2\sqrt{\varepsilon}}{\lambda_0 f}\operatorname{sinc}^2\left(\frac{\pi\sqrt{\varepsilon}\cdot d}{\lambda_0 f}x\right),$$

где $\operatorname{sinc}(x) = \sin(x)/x$, формируемое в скалярном приближении Кирхгофа в фокальной плоскости идеального сферического фронта. Рисунок 3.16 показывает хорошую работоспособность рассчитанной бинарной линзы и демонстрирует возможность замены задачи расчета ДОЭ существенно более простой задачей расчета однопорядковой решетки также и при большом диапазоне изменения размеров зон.

3.4.3. Расчет отражающих решеток с непрерывным профилем в приближении Рэлея

Решение задачи дифракции на решетке с непрерывным профилем в приближении Рэлея, изготовленной из идеально проводящего материала, приведено в пункте 3.2.2. Рассмотрим градиентный метод расчета профиля решетки $f(x)$ из условия формирования заданной интенсивности дифракционных порядков [16]. Интенсивности порядков (см. (3.46)) пропорциональны квадратам модулей коэффициентов Рэлея; $I_n = |R_n|^2 \beta_n/\beta_0$. Для удобства будем работать с нормированной функцией профиля

$$H(x) = k \cdot f\left(\frac{x\cdot d}{2\pi}\right).$$

Для построения градиентной процедуры расчета профиля решетки $H(x)$ введем функционал ошибки $\varepsilon(H)$, характеризующий отличие расчетных интенсивностей порядков I_n от требуемых значений $\tilde{I}_n$:

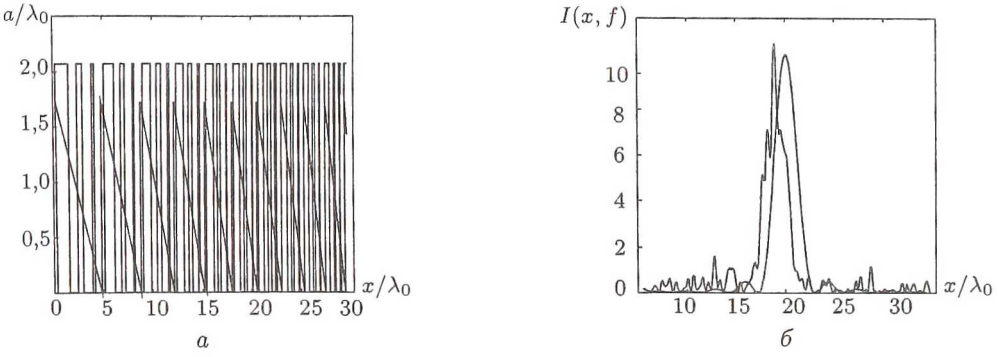
$$\varepsilon(H) = \varepsilon\left(\mathbf{I}(H), \tilde{\mathbf{I}}\right), \tag{3.238}$$

где $\mathbf{I}$ и $\tilde{\mathbf{I}}$ — вектора расчетных и заданных интенсивностей дифракционных порядков. Отметим, что компоненты вектора $\mathbf{I}$ также являются функционалами от профиля решетки $H(x)$.

Градиентная минимизация функционала (3.238) состоит в итерационной коррекции функции $H(x)$ согласно правилу

$$H_{n+1}(x) = H_n(x) - t\cdot\varepsilon'(x), \tag{3.239}$$

где n — номер итерации, t — шаг градиентного метода, $\varepsilon'(x)$ — градиент функционала. Для вычисления градиента $\varepsilon'(x)$ рассмотрим приращение функционала $\varepsilon(H)$,



а *б*

Рис. 3.16. Бинарная линза ($d = 30\lambda_0$) для фокусировки в точку $(x_0, f) = (0, -130\lambda_0)$ при $\theta = 10°$, $\varepsilon = 2{,}25$ и непрерывная линза (3.236) (*а*), распределение интенсивности, сформированное бинарной линзой для ТМ-поляризации и распределение интенсивности при идеальном сферическом фронте (*б*)

вызванное малым приращением высоты профиля $\Delta H(x)$:

$$\Delta\varepsilon(H) = \varepsilon(H + \Delta H) - \varepsilon(H) = \sum_{j=-M}^{M} \frac{\partial\varepsilon\left(\mathbf{I}, \tilde{\mathbf{I}}\right)}{\partial I_j} \Delta I_j(H) =$$

$$= \sum_{j=-M}^{M} \frac{\partial\varepsilon\left(\mathbf{I}, \tilde{\mathbf{I}}\right)}{\partial I_j} \Delta\left(\frac{\beta_j}{\beta_0} R_j(H) \cdot R_j^*(H)\right) = 2\,\mathrm{Re}\left(\Delta\mathbf{R}(H), \mathbf{L}(H)\right), \quad (3.240)$$

где

$$\Delta\mathbf{R}(H) = \left(\Delta R_j(H)\right)_{-M}^{M}, \quad \mathbf{L}(H) = \left(L_j(H)\right)_{-M}^{M}, \quad L_j = \sum_{j=-M}^{M} \frac{\partial\varepsilon\left(\mathbf{I}, \tilde{\mathbf{I}}\right)}{\partial I_j} \frac{\beta_j}{\beta_0} R_j(H). \quad (3.241)$$

Согласно (3.92), вектор приращений $\Delta\mathbf{R}(H)$ имеет вид

$$\Delta\mathbf{R}(H) = A^{-1}\left(\Delta\mathbf{B} - \Delta A \cdot \mathbf{R}\right), \quad (3.242)$$

где матрица A и вектор $\mathbf{B}$ определены в (3.87), (3.88) для ТЕ-поляризации и в (3.90), (3.91) для ТМ-поляризации. Для ТЕ-поляризации из (3.87), (3.88) получим вектор приращений в виде

$$\Delta\mathbf{R}(H) = A^{-1} \cdot \int_0^{2\pi} i\Delta H(\xi)\left(\tilde{\mathbf{B}}(\xi) - \tilde{A}(\xi) \cdot \mathbf{R}\right) \mathrm{d}\xi, \quad (3.243)$$

где

$$\tilde{A}(\xi) = \left[\tilde{A}_{pn}(\xi)\right]_{-N}^{N}, \quad \tilde{A}_{pn}(\xi) = (\beta_n - \beta_0) \cdot \exp\left(i\xi(n-p) + i(\beta_n - \beta_0) H(\xi)\right), \quad (3.244)$$

$$\tilde{\mathbf{B}}(\xi) = \left(\tilde{B}_p(\xi)\right)_{-N}^{N}, \quad \tilde{B}_p(\xi) = -2\beta_0 \exp\left(-2i\beta_0 H(\xi) - ip\xi\right).$$

Подставляя (3.243) в (3.240), получим приращение функционала ошибки $\varepsilon(H)$:

$$\Delta\varepsilon(H) = -2\int_0^{2\pi} \Delta H \cdot \operatorname{Im}\left(A^{-1}\left(\tilde{\mathbf{B}}(\xi) - \tilde{A}(\xi)\cdot\mathbf{R}\right),\mathbf{L}\right)\mathrm{d}\xi. \tag{3.245}$$

Градиент функционала имеет вид мнимой части скалярного произведения:

$$\varepsilon'(x) = -2\cdot\operatorname{Im}\left(A^{-1}\left(\tilde{\mathbf{B}}(x) - \tilde{A}(x)\cdot\mathbf{R}\right),\mathbf{L}\right). \tag{3.246}$$

Проводя аналогичные выкладки для ТМ-поляризации, несложно получить градиент функционала ошибки также в виде (3.246), где вектор $\mathbf{R}$ находим из уравнений (3.90)–(3.92), а матрица $\tilde{A}(\xi)$ и вектор $\tilde{\mathbf{B}}(\xi)$ имеют вид

$$\tilde{A}(\xi) = \left[\tilde{A}_{pn}(\xi)\right]_{-N}^{N}, \quad \tilde{A}_{pn}(\xi) = \left(\beta_n(\beta_n-\beta_0) + \frac{\alpha_n\lambda}{d}(n-p)\right)\times$$
$$\times\exp\left(i\xi(n-p) + i(\beta_n-\beta_0)H(\xi)\right), \tag{3.247}$$

$$\tilde{\mathbf{B}}(\xi) = \left[\tilde{B}_p(\xi)\right]_{-N}^{N}, \quad \tilde{B}_p(\xi) = \left(-2\beta_0^2 + \frac{\alpha_0\lambda}{d}p\right)\exp\left(-2i\beta_0 H(\xi) - ip\xi\right). \tag{3.248}$$

Для расчета t для градиентной процедуры (3.239) рассмотрим функционал $\varepsilon(H)$ вдоль направления, противоположного градиенту, как функцию от t:

$$\varepsilon_1(t) = \varepsilon\left(H - t\cdot\varepsilon'(H)\right). \tag{3.249}$$

Разложим функцию $\varepsilon_1(t)$ в ряд Тейлора в окрестности точки $t=0$:

$$\varepsilon_1(t) = \varepsilon_1(0) + \varepsilon_1'(0)\cdot t. \tag{3.250}$$

Предполагая, что минимальное значение функционала $\varepsilon(H)$ равно нулю, приравняем правую часть уравнения (3.250) к нулю и получим величину шага t в виде

$$t = -\varepsilon_1(0)/\varepsilon_1'(0). \tag{3.251}$$

Для расчета $\varepsilon_1'(0)$ рассмотрим приращение $\Delta\varepsilon_1(t) = \varepsilon_1(t) - \varepsilon_1(0)$ функции $\varepsilon_1(t)$ при $t=0$. Для расчета $\Delta\varepsilon_1(t)$ воспользуемся формулой (3.245) при $\Delta H = -\varepsilon'(H)t$:

$$\Delta\varepsilon_1(t) = t\int_0^{2\pi}\varepsilon'(H)\cdot 2\operatorname{Im}\left(A^{-1}\cdot\left(\tilde{\mathbf{B}}(\xi) - \tilde{A}(\xi)\cdot\mathbf{R}\right),\mathbf{L}\right)\mathrm{d}\xi. \tag{3.252}$$

Согласно (3.252) производная $\varepsilon_1'(0)$ имеет вид

$$\varepsilon_1'(0) = -4\int_0^{2\pi}\left(\operatorname{Im}\left(A^{-1}\cdot\left(\tilde{\mathbf{B}}(\xi) - \tilde{A}(\xi)\cdot\mathbf{R}\right),\mathbf{L}\right)\right)^2\mathrm{d}\xi, \tag{3.253}$$

а значение шага определяется по формуле

$$t = \varepsilon_1(0)\cdot\left(\int_0^{2\pi}\left(2\cdot\operatorname{Im}\left(A^{-1}\left(\tilde{\mathbf{B}}(\xi) - \tilde{A}(\xi)\cdot\mathbf{R}\right),\mathbf{L}\right)\right)^2\mathrm{d}\xi\right)^{-1}. \tag{3.254}$$

Матрица $\tilde{A}(x)$ и вектор $\tilde{\mathbf{B}}(x)$ в (3.254) определены в (3.244) для ТЕ-поляризации и в уравнениях (3.247), (3.248) для ТМ-поляризации.

В **3.2** показано, что при $d \gg \lambda$ приближение Рэлея переходит в приближение Фраунгофера для которого коэффициенты Рэлея соответствуют коэффициентам Фурье в разложении функции $\exp(i\varphi(x))$, где $\varphi(x)$ — функция фазового набега (3.93). Это позволяет рассматривать широко используемые градиентные алгоритмы синтеза решеток в приближении Фраунгофера как частный случай общего градиентного алгоритма (3.239), (3.254), основанного на приближении Рэлея.

Рассмотренный градиентный метод был применен к расчету решеток с $2M+1$ равными порядками. В качестве функционала ошибки был использован квадратичный функционал:

$$\varepsilon(H) = \sum_{j=-M}^{M} \left(I_j(H) - \tilde{I}_j\right)^2. \tag{3.255}$$

Шаг t градиентного метода определялся из условия минимума функции, являющейся линейной аппроксимацией функционала $\varepsilon(H)$ (см. (3.249)).

Пример 3.11. Расчет проводился для нормального падения плоской волны для ТЕ-поляризации. На рис. 3.17–3.20 приведены расчетные профили $H(x)$ и интенсивности дифракционных порядков для 5, 7, 9 и 11-порядковой решеток с периодами $3{,}6\lambda$, $5{,}2\lambda$, $6{,}2\lambda$ и $7{,}2\lambda$, соответственно. В качестве начальной точки для градиентной процедуры в приближении Рэлея использовались профили, рассчитанные градиентным методом в скалярном приближении. Начальные профили показаны на рис. 3.17*а*–3.20*а* пунктирными линиями. Значения энергетической эффективности E (см. (3.205)) и среднеквадратичной ошибки δ (см. (3.206)) для начальных профилей в скалярном приближении составили 95,2% и 1,1% для 5-порядковой решетки, 98,1% и 1,4% для 7-порядковой решетки, 99,6% и 1,9% для 9-порядковой решетки и 98,6% и 0,8% для 11-порядковой решетки.

Интенсивности порядков для начальных профилей, рассчитанные в приближении Рэлея, показаны пунктирной линией на рис. 3.17*б*–3.20*б*. Приближение Рэлея для начальных профилей дает увеличение ошибки δ до 25–30% при снижении эффективности E на 1,5–2,5%. Использование градиентного метода в приближении Рэлея позволяет скорректировать скалярное решение. Профили решеток и интенсивности порядков, рассчитанные в приближении Рэлея, показаны сплошными линиями на рис. 3.17–3.20.

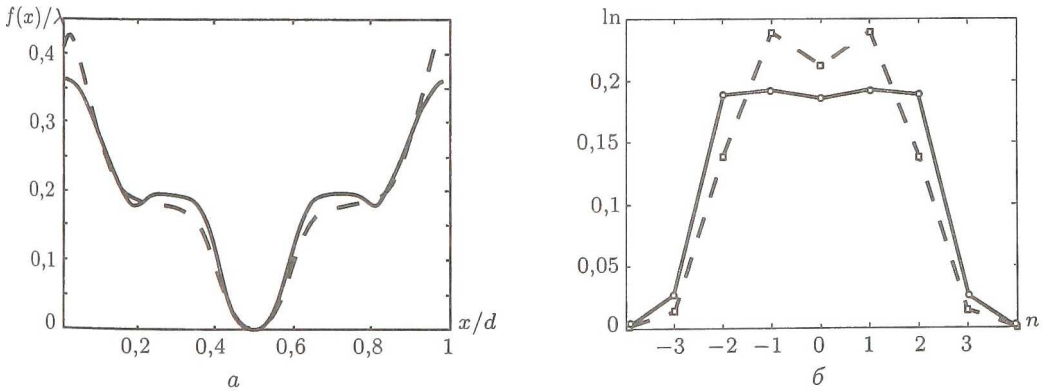


Рис. 3.17. Профили 5-порядковых решеток (*а*) и интенсивности порядков, рассчитанные в скалярном приближении (штриховая) и приближении Рэлея (непрерывная линия) (*б*)

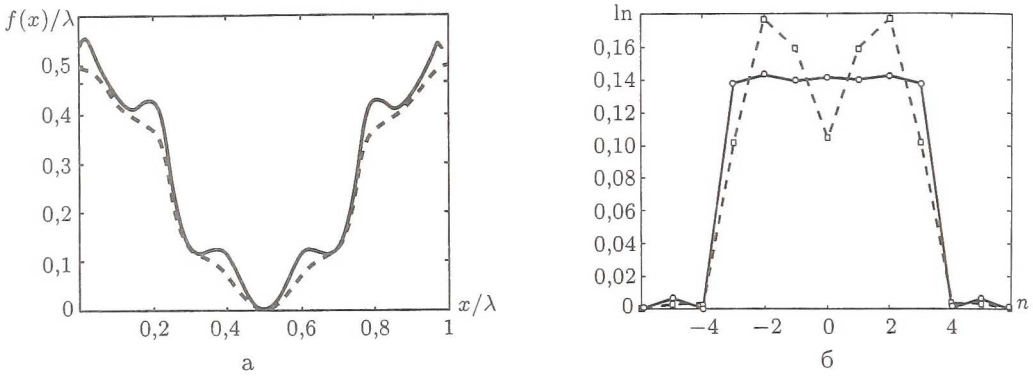


Рис. 3.18. Профили 7-порядковых решеток (а) и интенсивности порядков, рассчитанные в скалярном приближении (штриховая) и приближении Рэлея (непрерывная линия) (б)

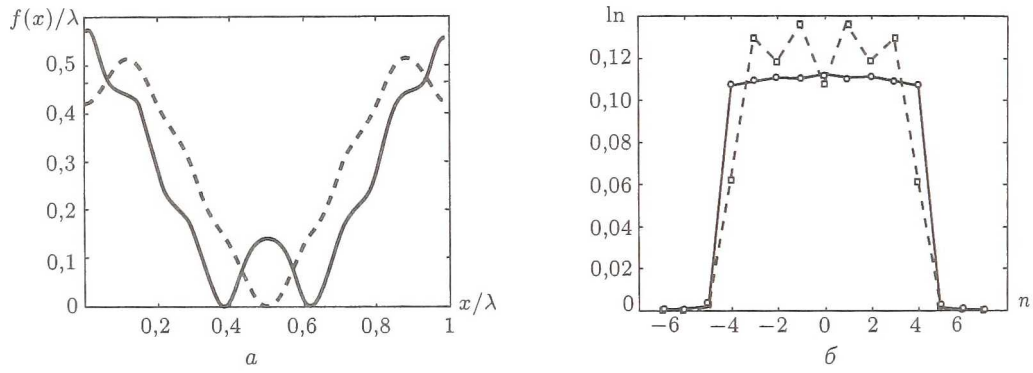


Рис. 3.19. Профили 9-порядковых решеток ($а$) и интенсивности порядков, рассчитанные в скалярном приближении (штриховая) и приближении Рэлея (непрерывная линия) ($б$)

Значения E и δ для рассчитанных в приближении Рэлея решеток составили 92,5% и 1,7% для 5-порядковой решетки, 94,2% и 0,8% для 7-порядковой решетки, 99,3% и 0,5% для 9-порядковой решетки и 96,3% и 0,7% для 11-порядковой решетки. Для оценки точности полученных решений был проведен расчет интенсивности порядков с использованием точного интегрального метода, рассмотренного в 3.2. Результаты расчетов показали, что увеличение ошибки δ при точном расчете интенсивностей порядков составляет всего 2–3%. Приведенные примеры показывают как хорошую сходимость и точность градиентной процедуры в приближении Рэлея, так и необходимость корректировки профилей, рассчитанных в скалярном приближении.

3.5. Дифракция на отражающих двумерных структурах

3.5.1. Дифракция света на криволинейной зоне

Рассмотрим дифракцию монохроматической электромагнитной волны на канавке, вырезанной в идеально проводящем материале в плоскости $z = 0$, где (x, y, z) – декартовы координаты. Стенки канавки параллельны оси z, а дно параллельно плоскости (x, y). Данная задача важна для многих приложений, например, для расчета поля от бинарной линзы Френеля. В дальнейшем криволинейную канавку будем

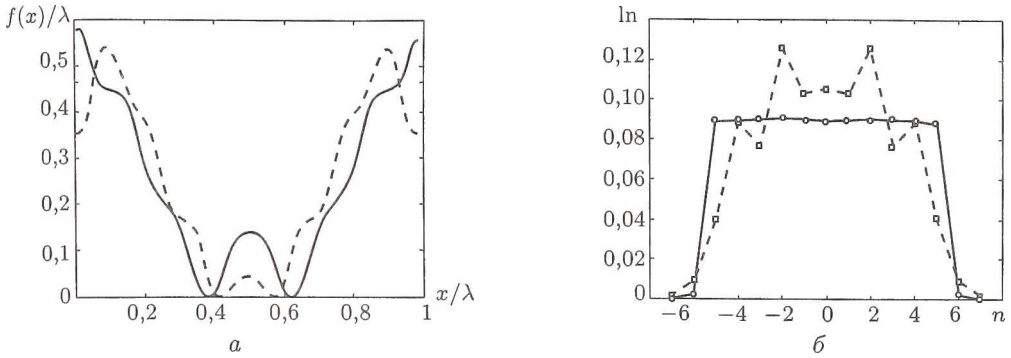


Рис. 3.20. Профили 11-порядковых решеток (*а*) и интенсивности порядков, рассчитанные в скалярном приближении (штриховая) и приближении Рэлея (непрерывная линия) (*б*)

называть зоной. Решать задачу будем с помощью метода модовых разложений [5]. Частный случай метода модовых разложений рассмотрен в разделе 3.1, посвященном решению задачи дифракции на идеально проводящих дифракционных решетках со ступенчатым профилем.

Электрическое и магнитное поля удовлетворяют уравнениям Максвелла для свободного пространства:

$$\operatorname{rot}\mathbf{H} = -ik\mathbf{E}, \tag{3.256}$$

$$\operatorname{rot}\mathbf{E} = ik\mathbf{H}, \tag{3.257}$$

где k — волновое число. Граничные условия на поверхности идеального проводника имеют вид (3.11). Кроме того, тангенциальные компоненты электрического поля должны быть непрерывны при переходе через плоскость $z = 0$, соответствующую верхнему срезу канавки:

$$[\mathbf{n}, \mathbf{E}_1(z=0)] = [\mathbf{n}, \mathbf{E}_2(z=0)], \tag{3.258}$$

где $\mathbf{n}$ — нормаль к плоскости $z = 0$, $\mathbf{E}_1$, $\mathbf{E}_2$ — напряженности поля в области над канавкой (при $z > 0$) и внутри канавки. Тангенциальные компоненты магнитного поля не будут непрерывны при переходе через плоскость $z = 0$:

$$[\mathbf{n}, \mathbf{H}] = 4\pi\mathbf{j}/c, \tag{3.259}$$

где $\mathbf{j}$ — плотность поверхностного тока. Однако, при $z = 0$ в области канавки тангенциальная составляющая магнитного поля также непрерывна:

$$[\mathbf{n}, \mathbf{H}_1(z=0)] = [\mathbf{n}, \mathbf{H}_2(z=0)]. \tag{3.260}$$

Для решения уравнений Максвелла (3.256), (3.257) будем использовать метод разделения переменных, изложенный в 3.1. Представим решение в виде волны, бегущей в направлении оси z:

$$\begin{aligned} \mathbf{E}(\mathbf{x};\chi) &= \tilde{\mathbf{E}}(x,y,\chi)e^{ik\Gamma z}, \\ \mathbf{H}(\mathbf{x};\chi) &= \tilde{\mathbf{H}}(x,y,\chi)e^{ik\Gamma z}, \end{aligned} \tag{3.261}$$

где $\mathbf{x} = (x, y, z)$, $\Gamma = \pm\sqrt{1-\chi^2}$ — постоянная распространения вдоль оси z. Заметим, что если в уравнении (3.7) ось Oz была направлена вдоль штрихов решетки, то в уравнении (3.261) ось Oz направлена перпендикулярно плоскости дна канавки.

Подставляя представление (3.261) в систему уравнений (3.256), (3.257) получаем выражение $\tilde{\mathbf{E}}(x,y,\chi)$ и $\tilde{\mathbf{H}}(x,y,\chi)$ через $\tilde{E}_z(x,y,\chi)$ и $\tilde{H}_z(x,y,\chi)$:

$$\begin{aligned}\tilde{\mathbf{E}}(x,y,\chi) &= -\frac{\Gamma}{ik(1-\Gamma^2)}\nabla_\perp \tilde{E}_z(x,y,\chi) - \frac{1}{ik(1-\Gamma^2)}\operatorname{rot}_\perp \tilde{H}_z(x,y,\chi) + \tilde{E}_z(x,y,\chi)\mathbf{e}_z,\\ \tilde{\mathbf{H}}(x,y,\chi) &= -\frac{\Gamma}{ik(1-\Gamma^2)}\nabla_\perp \tilde{H}_z(x,y,\chi) - \frac{1}{ik(1-\Gamma^2)}\operatorname{rot}_\perp \tilde{E}_z(x,y,\chi) + \tilde{H}_z(x,y,\chi)\mathbf{e}_z,\end{aligned} \tag{3.262}$$

где $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ — орты декартовой системы координат,

$$\nabla_\perp F = \frac{\partial F}{\partial x}\mathbf{e}_x + \frac{\partial F}{\partial y}\mathbf{e}_y, \tag{3.263}$$

$$\nabla_\perp xF = \operatorname{rot}_\perp F = \frac{\partial F}{\partial y}\mathbf{e}_x - \frac{\partial F}{\partial x}\mathbf{e}_y. \tag{3.264}$$

Отметим, что не следует путать $\operatorname{rot}_\perp F$ и $\operatorname{rot}\mathbf{A}$, первая операция применяется к скалярной функции, а вторая — к векторной. Компоненты $\tilde{E}_z$ и $\tilde{H}_z$ в (3.262) удовлетворяют двумерным уравнениям Гельмгольца:

$$\begin{aligned}\nabla_\perp \tilde{E}_z(x,y) + k^2\chi^2\tilde{E}_z(x,y) = 0,\\ \nabla_\perp \tilde{H}_z(x,y) + k^2\chi^2\tilde{H}_z(x,y) = 0.\end{aligned} \tag{3.265}$$

Уравнение Гельмгольца — уравнение на собственные значения. Для свободного пространства собственные значения имеют непрерывный спектр. Для направляющих структур, например волноводов, спектр собственных значений — дискретный. Подставляя (3.262) в (3.261), получаем выражения для бегущих волн:

$$\begin{aligned}\mathbf{E}^s(\mathbf{x},\chi) = \Bigg[&-\frac{s\sqrt{1-\chi^2}}{ik\chi^2}\nabla_\perp \tilde{E}_z(x,y,\chi) - \frac{1}{ik\chi^2}\operatorname{rot}_\perp \tilde{H}_z(x,y,\chi) + \\ &+\tilde{E}_z(x,y,\chi)\mathbf{e}_z\Bigg] \times e^{sik\sqrt{1-\chi^2}z},\\ \mathbf{H}^s(\mathbf{x},\chi) = \Bigg[&-\frac{s\sqrt{1-\chi^2}}{ik\chi^2}\nabla_\perp \tilde{H}_z(x,y,\chi) - \frac{1}{ik\chi^2}\operatorname{rot}_\perp \tilde{E}_z(x,y,\chi) + \\ &+\tilde{H}_z(x,y,\chi)\mathbf{e}_z\Bigg] \times e^{sik\sqrt{1-\chi^2}z}.\end{aligned} \tag{3.266}$$

Здесь $s = \pm 1$, $s = 1$ — для волн, распространяющихся по оси z, а $s = -1$ — для волн, распространяющихся против оси z. В свободном пространстве решения уравнений Гельмгольца (3.265) имеют вид

$$\begin{aligned}\tilde{E}_z(x,y,\chi) = E^s(\alpha,\beta)F_0(\alpha,\beta,x,y),\\ \tilde{H}_z(x,y,\chi) = H^s(\alpha,\beta)F_0(\alpha,\beta,x,y),\end{aligned} \tag{3.267}$$

где $F_0(\alpha,\beta,x,y) = \exp\left(ik(\alpha x + \beta y)\right)$, $\chi^2 = \alpha^2 + \beta^2$, $E^s(\alpha,\beta)$ и $H^s(\alpha,\beta)$ — произвольные функции. Подставляя (3.267) в (3.266), получаем частные решения уравнений

Максвелла (3.256), (3.257), зависящие от параметров (α, β), в виде

$$\begin{aligned}
&\mathbf{E}^s(\mathbf{x}, \chi) = \mathbf{E}^s(\mathbf{x}, \alpha, \beta) = \\
&= \left[-\frac{s\sqrt{1-\chi^2}}{\chi^2} E^s(\alpha,\beta)\mathbf{F}_1(\alpha,\beta,x,y) - \frac{1}{\chi^2} H^s(\alpha,\beta)\mathbf{F}_2(\alpha,\beta,x,y) \right] e^{sik\sqrt{1-\chi^2}z} + \\
&\qquad + \mathbf{e}_z E^s(\alpha,\beta) F_0(\alpha,\beta,x,y) e^{sik\sqrt{1-\chi^2}z}, \\
&\mathbf{H}^s(\mathbf{x}, \chi) = \mathbf{H}^s(\mathbf{x}, \alpha, \beta) = \\
&= \left[-\frac{s\sqrt{1-\chi^2}}{\chi^2} H^s(\alpha,\beta)\mathbf{F}_1(\alpha,\beta,x,y) + \frac{1}{\chi^2} E^s(\alpha,\beta)\mathbf{F}_2(\alpha,\beta,x,y) \right] e^{sik\sqrt{1-\chi^2}z} + \\
&\qquad + \mathbf{e}_z H^s(\alpha,\beta) F_0(\alpha,\beta,x,y) e^{sik\sqrt{1-\chi^2}z},
\end{aligned} \tag{3.268}$$

где $\mathbf{F}_1(\alpha,\beta,x,y) = (1/ik)\,\nabla_\perp e^{ik(\alpha x+\beta y)}$, $\mathbf{F}_2(\alpha,\beta,x,y) = (1/ik)\,\nabla_\perp e^{ik(\alpha x+\beta y)}$.

Решения (3.268) соответствуют линейно поляризованным плоским волнам с направлением распространения $(\alpha, \beta, s\sqrt{1-\alpha^2-\beta^2})$. Согласно принципу суперпозиции, общее решение уравнений (3.256), (3.257) в свободном пространстве представляется в виде суперпозиции плоских волн (3.268) при $s = \pm 1$ и различных (α, β). Поскольку (α, β) пробегают непрерывный спектр значений, то такая суперпозиция имеет вид интегралов:

$$\begin{aligned}
\mathbf{E}(\mathbf{x}) = \sum_s \int\limits_{-\infty}^{\infty} &\left[-\frac{s\sqrt{1-\chi^2}}{\chi^2} E^s(\alpha,\beta)\mathbf{F}_1(\alpha,\beta,x,y) - \frac{1}{\chi^2} H^s(\alpha,\beta)\mathbf{F}_2(\alpha,\beta,x,y) \right] \times \\
&\times e^{sik\sqrt{1-\chi^2}z}\,\mathrm{d}\alpha\,\mathrm{d}\beta + \sum_s \int\limits_{-\infty}^{\infty} \mathbf{e}_z E^s(\alpha,\beta) F_0(\alpha,\beta,x,y) e^{sik\sqrt{1-\chi^2}z}\,\mathrm{d}\alpha\,\mathrm{d}\beta,
\end{aligned}$$

$$\begin{aligned}
\mathbf{H}(\mathbf{x}) = \sum_s \int\limits_{-\infty}^{\infty} &\left[-\frac{s\sqrt{1-\chi^2}}{\chi^2} H^s(\alpha,\beta)\mathbf{F}_1(\alpha,\beta,x,y) + \frac{1}{\chi^2} E^s(\alpha,\beta)\mathbf{F}_2(\alpha,\beta,x,y) \right] \times \\
&\times e^{sik\sqrt{1-\chi^2}z}\,\mathrm{d}\alpha\,\mathrm{d}\beta + \sum_s \int\limits_{-\infty}^{\infty} \mathbf{e}_z H^s(\alpha,\beta) F_0(\alpha,\beta,x,y) e^{sik\sqrt{1-\chi^2}z}\,\mathrm{d}\alpha\,\mathrm{d}\beta.
\end{aligned} \tag{3.269}$$

Особый интерес для дальнейшего представляет случай квазипериодического поля, спектр которого имеет вид

$$\begin{aligned}
E^s(\alpha,\beta) &= \sum_{mn} R^{se}_{mn}\,\delta(\alpha-\alpha_n)\delta(\beta-\beta_n), \\
H^s(\alpha,\beta) &= \sum_{mn} R^{sh}_{mn}\,\delta(\alpha-\alpha_n)\delta(\beta-\beta_n),
\end{aligned} \tag{3.270}$$

где $-\infty < m, n < \infty$, $\alpha_n = \alpha_0 + \lambda n/d_x$, $\beta_m = \beta_0 + \lambda m/d_y$, $\chi^2_{mn} = \alpha^2_{mn} + \beta^2_{mn}$, d_x, d_y — период квазипериодичности. Подставляя (3.270) в (3.269), получаем следующее

разложение для квазипериодических полей:

$$\mathbf{E}(\mathbf{x}) = \sum_s \sum_{mn} \left[-\frac{s\sqrt{1-\chi_{mn}^2}}{\chi_{mn}^2} R_{nm}^{se} \mathbf{F}_1(\alpha_n, \beta_m, x, y) - \right.$$
$$\left. - \frac{1}{\chi_{mn}^2} R_{nm}^{sh} \mathbf{F}_2(\alpha_n, \beta_n, x, y) \right] e^{sik\sqrt{1-\chi_{mn}^2}z} +$$
$$+ \sum_s \sum_{mn} \mathbf{e}_z R_{nm}^{sh} F_0(\alpha_n, \beta_m, x, y) e^{sik\sqrt{1-\chi_{mn}^2}z},$$

$$\mathbf{H}(\mathbf{x}) = \sum_s \sum_{mn} \left[-\frac{s\sqrt{1-\chi_{mn}^2}}{\chi_{mn}^2} R_{nm}^{sh} \mathbf{F}_1(\alpha_n, \beta_m, x, y) - \right.$$
$$\left. - \frac{1}{\chi_{mn}^2} R_{nm}^{se} \mathbf{F}_2(\alpha_n, \beta_n, x, y) \right] e^{sik\sqrt{1-\chi_{mn}^2}z} +$$
$$+ \sum_s \sum_{mn} \mathbf{e}_z R_{nm}^{sh} F_0(\alpha_n, \beta_m, x, y) e^{sik\sqrt{1-\chi_{mn}^2}z}, \quad (3.271)$$

Квазипериодические разложения имеют место при дифракции электромагнитной волны на двумерной дифракционной решетке с периодами d_x, d_y. В случае $s = 1$ получаем волны, отраженные от дифракционной решетки, а в случае $s = 1$ — волны, падающие на дифракционную решетку.

Рассмотрим решение уравнения Максвелла (3.256), (3.257) внутри зоны (канавки). Для этого сначала рассмотрим задачу о распространении электромагнитной волны внутри направляющей структуры (волновода). Предположим, что сечение волновода представляет собой область, расположенную между двумя парами кривых, которые являются координатными линиями одной из криволинейной системы координат: эллиптической, параболической, цилиндрической или декартовой. Название системы координат совпадает с названием кривой второго порядка, которая описывает сечение направляющей структуры. Стенки направляющей структуры будем считать идеально проводящими. Введем ортогональные криволинейные координаты по формулам $x = x(u, v), \quad y = y(u, v)$.

Сечение направляющей структуры представляет собой область ограниченную координатными кривыми, параметрические уравнения которых в декартовой системе координат имеют вид

$$\left.\begin{array}{ll} x = x(t) = x(u_1, t), & y = y(t) = y(u_1, t), \\ x = x(t) = x(u_2, t), & y = y(t) = y(u_2, t), \\ x = x(t) = x(t, v_1), & y = y(t) = y(t, v_1), \\ x = x(t) = x(t, v_2), & y = y(t) = y(t, v_2). \end{array}\right\} \quad (3.272)$$

Выражения для векторных операций в ортогональной криволинейной системе координат принимают вид

$$\nabla_\perp F = \frac{1}{\sqrt{g_{11}}} \frac{\partial F}{\partial u} \mathbf{e}_u + \frac{1}{\sqrt{g_{22}}} \frac{\partial F}{\partial v} \mathbf{e}_v,$$
$$\nabla_\perp \times F = \frac{1}{\sqrt{g_{22}}} \frac{\partial F}{\partial v} \mathbf{e}_u - \frac{1}{\sqrt{g_{11}}} \frac{\partial F}{\partial u} \mathbf{e}_v, \quad (3.273)$$

где g_{ij} — компоненты метрического тензора в криволинейной системе координат.

В этом случае полное поле также выражается через $\tilde{E}_z$ и $\tilde{H}_z$ по формулам (3.262). Сами функции $\tilde{E}_z$ и $\tilde{H}_z$ удовлетворяют уравнениям Гельмгольца, записанным в криволинейных координатах:

$$\begin{aligned}\frac{1}{\sqrt{g}}\left(\frac{\partial}{\partial u}\left(\frac{\sqrt{g}}{g_{11}}\right)\frac{\partial \tilde{E}_z}{\partial u}+\frac{\partial}{\partial v}\left(\frac{\sqrt{g}}{g_{22}}\right)\frac{\partial \tilde{E}_z}{\partial v}\right)+k^2\chi^2\tilde{E}_z=0,\\ \frac{1}{\sqrt{g}}\left(\frac{\partial}{\partial u}\left(\frac{\sqrt{g}}{g_{11}}\right)\frac{\partial \tilde{H}_z}{\partial u}+\frac{\partial}{\partial v}\left(\frac{\sqrt{g}}{g_{22}}\right)\frac{\partial \tilde{H}_z}{\partial v}\right)+k^2\eta^2\tilde{H}_z=0,\end{aligned} \tag{3.274}$$

где χ и η — определяют фазовые скорости распространения электрического и магнитного полей внутри канавки, и на поверхности направляющей структуры удовлетворяют граничным условиям

$$\left.\begin{aligned}(\nabla_\perp \tilde{E}_z, \mathbf{e}_v)=0, \quad \text{если } u=u_1, u=u_2,\\ (\nabla_\perp \tilde{E}_z, \mathbf{e}_u)=0, \quad \text{если } v=v_1, v=v_2,\end{aligned}\right\} \tag{3.275}$$

для электрического поля, и

$$\left.\begin{aligned}(\nabla_\perp \times \tilde{H}_z, \mathbf{e}_v)=0, \quad \text{если } u=u_1, u=u_2,\\ (\nabla_\perp \times \tilde{H}, \mathbf{e}_u)=0, \quad \text{если } v=v_1, v=v_2,\end{aligned}\right\} \tag{3.276}$$

для магнитного поля.

Краевая задача для уравнений (3.274) с граничными условиями (3.275), (3.276) есть задача на собственные значения (χ, η — собственные значения краевой задачи). Спектр собственных значений в данном случае будет дискретным. Обозначим решение для $\tilde{E}_z$, соответствующее собственному значению χ_{nm}, как $E_{nm}(u,v,\chi_{nm})$, а решение для $\tilde{H}_z$, соответствующее собственному значению η_{nm} , через $H_{nm}(u,v,\eta_{nm})$. Общее решение уравнений Максвелла (3.256), (3.257) можно выразить через суперпозицию решений задачи на собственные значения :

$$\begin{aligned}\mathbf{E}^{\pm}(u,v,z)=\sum_m C_{nm}^{\pm^e}\Bigg[-\frac{\pm\sqrt{1-\chi_{nm}^2}}{k\chi_m^2}\nabla_\perp E_{nm}(u,v,\chi_{nm})\exp(\pm ik\sqrt{1-\chi_{nm}^2}z)+\\ +\mathbf{e}_z E_{nm}(u,v,\chi_{nm})\exp(\pm ik\sqrt{1-\chi_{nm}^2}z)\Bigg]-\\ -\sum_m \frac{C_{nm}^{\pm^h}}{k\eta_{nm}^2}\nabla_\perp\times H_{nm}(u,v,\eta_{nm})\exp(\pm ik\sqrt{1-\eta_{nm}^2}z),\\ \mathbf{H}^{\pm}(u,v,z)=\sum_m C_{nm}^{\pm^h}\Bigg[-\frac{\pm\sqrt{1-\eta_m^2}}{ik\eta_m^2}\nabla_\perp H_{nm}(u,v,\eta_{nm})\exp(\pm ik\sqrt{1-\eta_{nm}^2}z))+\\ +\mathbf{e}_z H_{nm}(u,v,\eta_{nm})\exp(\pm ik\sqrt{1-\eta_{nm}^2}z\Bigg]+\\ +\sum_m \frac{C_{nm}^{\pm^e}}{k\chi_{nm}^2}\nabla_\perp\times E_{nm}(u,v,\chi_{nm})\exp(\pm ik\sqrt{1-\chi_{nm}^2}z),\end{aligned} \tag{3.277}$$

где C_{nm} — произвольные коэффициенты. Верхний знак соответствует волнам, распространяющимся вдоль оси z, а нижний знак — волнам, распространяющимся против оси z. Внутри зоны присутствуют волны, распространяющиеся в обоих направлениях (волны, распространяющиеся против оси z, возникают за счет отражения от

дна криволинейной канавки). Граничное условие на дне канавки имеет вид $[E(z = -h), e] = 0$, h — глубина канавки.

Учитывая (3.277) и граничное условия на дне канавки, получаем выражение для векторов электрического и магнитного поля в следующем виде:

$$\mathbf{E}(u,v,z) = \sum_m C^e_{nm} \left[-\frac{\sqrt{1-\chi^2_{nm}}}{k\chi^2_m} \nabla_\perp E_{nm}(u,v,\chi_{nm}) \sin(k\sqrt{1-\chi^2_{nm}}(z+h)) + \right.$$
$$\left. + \mathbf{e}_z E_{nm}(u,v,\chi_{nm}) \cos(k\sqrt{1-\chi^2_{nm}}(z+h)) \right] -$$
$$- \sum_m \frac{C^h_{nm}}{k\eta^2_{nm}} \nabla_\perp \times H_{nm}(u,v,\eta_{nm}) \sin(k\sqrt{1-\eta^2_{nm}}(z+h)),$$

$$\mathbf{H}(u,v,z) = \sum_m C^h_{nm} \left[-\frac{\sqrt{1-\eta^2_m}}{ik\eta^2_m} \nabla_\perp H_{nm}(u,v,\eta_{nm}) \cos(k\sqrt{1-\eta^2_{nm}}(z+h)) + \right.$$
$$\left. + \mathbf{e}_z i H_{nm}(u,v,\eta_{nm}) \sin(k\sqrt{1-\eta^2_{nm}}(z+h)) \right] +$$
$$+ \sum_m \frac{C^e_{nm}}{k\chi^2_{nm}} \nabla_\perp \times E_{nm}(u,v,\chi_{nm}) \cos(k\sqrt{1-\chi^2_{nm}}(z+h)). \quad (3.278)$$

Модовые функции удовлетворяют соотношениям ортогональности:

$$\int E_{ms}(u,v) E^*_{kn}(u,v)\, \mathrm{d}u\, \mathrm{d}v = 0,$$

$$\int (\nabla_\perp E_{ms}(u,v), \nabla_\perp E^*_{nk}(u,v))\, \mathrm{d}u\, \mathrm{d}v = 0,$$

$$\int (\nabla_\perp \times E_{ms}(u,v), \nabla_\perp \times E^*_{nk}(u,v))\, \mathrm{d}u\, \mathrm{d}v = 0,$$

$$\int (\nabla_\perp H_{ms}(u,v), \nabla_\perp H^*_{nk}(u,v)\, \mathrm{d}u\, \mathrm{d}v) = 0, \quad (3.279)$$

$$\int (\nabla_\perp \times H_{ms}(u,v), \nabla_\perp \times H^*_{nk}(u,v))\, \mathrm{d}u\, \mathrm{d}v = 0,$$

$$\int (\nabla_\perp \times E_{ms}(u,v), \nabla_\perp H^*_{nk}(u,v))\, \mathrm{d}u\, \mathrm{d}v = 0.$$

Для дальнейших рассуждений понадобятся выражения для тангенциальных составляющих электрического и магнитного полей при $z = 0$. Тангенциальные компоненты электрического поля должны быть непрерывны при переходе через плоскость раздела сред $z = 0$ (см. (3.258)). С учетом (3.269) получаем

$$\sum_s \int_{-\infty}^{\infty} \left[-\frac{s\sqrt{1-\chi^2}}{\chi^2} E^s(\alpha,\beta) \mathbf{F}_1(\alpha,\beta,x,y) - \frac{1}{\chi^2} H^s(\alpha,\beta) \mathbf{F}_2(\alpha,\beta,x,y) \right] \mathrm{d}\alpha\, \mathrm{d}\beta =$$

$$= \sum_m C^e_{nm} \left[-\frac{\sqrt{1-\chi^2_{nm}}}{k\chi^2_m} \nabla_\perp E_{nm}(u,v,\chi_{nm}) \sin(k\sqrt{1-\chi^2_{nm}}h) - \right.$$

$$\left. - \sum_m \frac{C^h_{nm}}{k\eta^2_{nm}} \nabla_\perp \times H_{nm}(u,v,\eta_{nm}) \sin(k\sqrt{1-\eta^2_{nm}}h) \right). \quad (3.280)$$

Тангенциальная составляющая магнитного поля не является непрерывной. Она остается непрерывной только в точках плоскости $z = 0$, где отсутствует проводник (см. (3.259), (3.260)). В этих точках условие непрерывности имеет вид

$$\sum_s \int_{-\infty}^{\infty} \left[-\frac{s\sqrt{1-\chi^2}}{\chi^2} H^s(\alpha,\beta)\mathbf{F}_1(\alpha,\beta,x,y) + \frac{1}{\chi^2} E^s(\alpha,\beta)\mathbf{F}_2(\alpha,\beta,x,y) \right] \mathrm{d}\alpha\,\mathrm{d}\beta =$$

$$= \sum_m C^h_{nm} \left[-\frac{\sqrt{1-\eta^2_m}}{ik\eta\xi^2_m} \nabla_\perp H_{nm}(u,v,\eta_{nm}) \cos(k\sqrt{1-\eta^2_{nm}}h) + \right.$$

$$\left. + \sum_m \frac{C^e_{nm}}{k\chi^2_{nm}} \nabla_\perp \times E_{nm}(u,v,\chi_{nm}) \cos(k\sqrt{1-\chi^2_{nm}}h) \right). \quad (3.281)$$

Функции $E^s(\alpha,\beta)$ и $H^s(\alpha,\beta)$ при $s = -1$ соответствуют падающей волне.

Выражения (3.280) (3.281) можно рассматривать как систему интегральных уравнений относительно функций $E^s(\alpha,\beta)$ и $H^s(\alpha,\beta)$ и коэффициентов C_{nm}. Учитывая соотношения ортогональности (3.279), выражаем C^e_{nm} и C^e_{nm} из соотношения (3.281) и подставляем в уравнение (3.280). Далее, учитывая соотношения ортогональности, для функций $F_{1,2}$ из уравнений (3.268), получаем систему интегральных уравнений, содержащую только функции $H^s(\alpha,\beta)$, $E^s(\alpha,\beta)$ при $s = 1$. Решая данную систему и подставляя найденные функции в (3.278), получаем выражения для векторов электрического и магнитного полей в полупространстве $z > 0$.

3.5.2. Дифракция на двумерной отражающей бинарной решетке

Алгоритм (п. 3.5.1.) можно использовать для решения задачи дифракции на двумерной решетке. Пусть дифракционная решетка представляет собой периодическую структуру с периодами d_x, d_y по соответствующим координатным осям. Период решетки состоит из набора криволинейных канавок с вертикальными стенками и плоским дном, рассмотренных в предыдущем пункте. Типичным примером такой структуры является растр отражающих бинарных линз. Пусть на такую двумерную дифракционную решетку падает квазипериодическая электромагнитная волна, с векторами электрического и магнитного поля

$$\mathbf{E}(\mathbf{x}) = \sum_{mn} \left[\frac{\sqrt{1-\chi^2}}{\chi^2} R^{se}_{nm}\mathbf{F}_1(\alpha_n,\beta_m,x,y) - \frac{1}{\chi^2} R^{sh}_{nm}\mathbf{F}_2(\alpha_n,\beta_n,x,y) \right] e^{-ik\sqrt{1-\chi^2}z} +$$

$$+ \sum_{mn} \mathbf{e}_z R^{sh}_{nm} F_0(\alpha_n,\beta_m,x,y) e^{-ik\sqrt{1-\chi^2}z},$$

$$\mathbf{H}(\mathrm{x}) = \sum_{mn}\left[\frac{\sqrt{1-\chi^2}}{\chi^2}R^{sh}_{nm}\mathbf{F}_1(\alpha_n,\beta_m,x,y) + \frac{1}{\chi^2}R^{se}_{nm}\mathbf{F}_2(\alpha_n,\beta_m,x,y)\right]e^{-ik\sqrt{1-\chi^2}z} + \\ + \sum_{mn}\mathbf{e}_z R^{sh}_{nm}F_0(\alpha,\beta,x,y)e^{-ik\sqrt{1-\chi^2}z}, \quad (3.282)$$

где $\alpha_n = \alpha_0 + \lambda n/d_x$, $\beta_m = \alpha_0 + \lambda m/d_y$, а коэффициенты R^{se}_{nm} и R^{sh}_{nm} при $s = -1$ считаем известными.

Полное поле представляется (3.271). Задача состоит в отыскании коэффициентов R^{se}_{nm} и R^{sh}_{nm} при s = 1 в разложении для векторов электрического и магнитного полей отраженной волны. Поле внутри канавок имеет вид, аналогичный (3.278):

$$\mathbf{E}(u,v,z) = \sum_L\sum_m C^{eL}_{nm}\left[-\frac{\sqrt{1-{\chi^L_{nm}}^2}}{k{\chi^L_{nm}}^2}\nabla_\perp E^L_{nm}(u,v,\chi_{nm})\sin(k\sqrt{1-{\chi^L_{nm}}^2}(z+h)) + \right. \\ \left. + \mathbf{e}_z E^L_{nm}(u,v,\chi^L_{nm})\cos(k\sqrt{1-{\chi^L_{nm}}^2}(z+h))\right] - \\ - \sum_m \frac{C^{hL}_{nm}}{k{\eta^L_{nm}}^2}\nabla_\perp \times H^L_{nm}(u,v,\eta^L_{nm})\sin(k\sqrt{1-{\eta^L_{nm}}^2}(z+h)),$$

$$\mathbf{H}(u,v,z) = \sum_m C^{hL}_{nm}\left[-\frac{\sqrt{1-\eta^{L\,2}_m}}{ik\eta^2_m}\nabla_\perp H^L_{nm}(u,v,\eta^L_{nm})\cos(k\sqrt{1-{\eta^L_{nm}}^2}(z+h)) + \right. \\ \left. + \mathbf{e}_z i H^L_{nm}(u,v,\eta^L_{nm})\sin(k\sqrt{1-{\eta^L_{nm}}^2}(z+h))\right] + \\ + \sum_m \frac{C^{eL}_{nm}}{k{\chi^L_{nm}}^2}\nabla_\perp \times E^L_{nm}(u,v,\chi^L_{nm})\cos(k\sqrt{1-{\chi^L_{nm}}^2}(z+h)), \quad (3.283)$$

где L — номер канавки.

Равенство тангенциальных компонентов в данном случае имеет вид

$$\sum_s\sum_{mn}\left[-\frac{s\sqrt{1-\chi^2}}{\chi^2_{mn}}R^{se}_{nm}\mathbf{F}_1(\alpha_n,\beta_m,x,y) - \frac{1}{\chi^2_{mn}}R^{sh}_{nm}\mathbf{F}_2(\alpha_n,\beta_n,x,y)\right] = \\ = \sum_L\sum_m C^{eL}_{nm}\left[\left[-\frac{\sqrt{1-{\chi^L_{nm}}^2}}{k{\chi^L_{nm}}^2}\nabla_\perp E^L_{nm}(u,v,\chi_{nm})\sin(k\sqrt{1-{\chi^L_{nm}}^2}h)\right] - \right. \\ \left. - \sum_m \frac{C^{hL}_{nm}}{k{\eta^L_{nm}}^2}\nabla_\perp \times H^L_{nm}(u,v,\eta^L_{nm})\sin(k\sqrt{1-{\eta^L_{nm}}^2}h), \quad (3.284)\right.$$

для электрического поля, и

$$\sum_s\sum_{mn}\left[-\frac{s\sqrt{1-\chi^2}}{\chi^2_{mn}}R^{sh}_{nm}\mathbf{F}_1(\alpha_n,\beta_m,x,y) + \frac{1}{\chi^2_{mn}}R^{se}_{nm}\mathbf{F}_2(\alpha_n,\beta_m,x,y)\right] =$$

$$= \sum_m -C_{nm}^{hL} \frac{\sqrt{1-{\eta_m^L}^2}}{ik\eta_m^2} \nabla_\perp H_{nm}^L(u,v,\eta_{nm}^L) \cos(k\sqrt{1-{\eta_{nm}^L}^2}h) +$$

$$+ \sum_m \frac{C_{nm}^{eL}}{k{\chi_{nm}^L}^2} \nabla_\perp \times E_{nm}^L(u,v,\chi_{nm}^L) \cos(k\sqrt{1-{\chi_{nm}^L}^2}h), \quad (3.285)$$

для магнитного поля.

Систему функциональных уравнений (3.284)–(3.285) можно легко свести к системе линейных алгебраических уравнений. Используя соотношения ортогональности для модовых функций (3.279), выразим коэффициенты C_{nm}^{eL} и C_{nm}^{hL} из уравнения (3.284) и подставим в (3.285). Далее, используя соотношения ортогональности для функций F_1 и F_2, получим систему линейных алгебраических уравнений для определения коэффициентов R_{nm}^{se} и R_{nm}^{sh}.

3.5.3. Дифракция на двумерных пропускающих диэлектрических структурах

Метод анализа дифракции световой волны на двумерной пропускающей диэлектрической структуре основан на численном решении уравнений Максвелла в пространственно-частотных координатах. Систему уравнений Максвелла (3.5), (3.6) запишем в следующем виде, раскрывая векторный оператор:

$$\begin{aligned}
\partial_z E_x &= \frac{i}{k}\partial_x \left(\frac{1}{\varepsilon} \left(\partial_x H_y - \partial_y H_x \right) \right) + ikH_y, \\
\partial_z E_y &= \frac{i}{k}\partial_y \left(\frac{1}{\varepsilon} \left(\partial_x H_y - \partial_y H_x \right) \right) - ikH_x, \\
\partial_z H_x &= \frac{1}{ik}\partial_x \left(\partial_x E_y - \partial_y E_x \right) - ik\varepsilon E_y, \\
\partial_z H_y &= \frac{1}{ik}\partial_y \left(\partial_x E_y - \partial_y E_x \right) + ik\varepsilon E_x.
\end{aligned} \quad (3.286)$$

Представим вектора электрического и магнитного полей и функцию диэлектрической проницаемости в виде разложения в интеграл Фурье по поперечным координатам:

$$\begin{aligned}
\mathbf{E}(x,y,z) &= \int_{-\infty}^{\infty} \mathbf{E}(\alpha,\beta,z) \exp\left(ik\left(\alpha x + \beta y\right)\right) \mathrm{d}\alpha\, \mathrm{d}\beta, \\
\mathbf{H}(x,y,z) &= \int_{-\infty}^{\infty} \mathbf{H}(\alpha,\beta,z) \exp\left(ik\left(\alpha x + \beta y\right)\right) \mathrm{d}\alpha\, \mathrm{d}\beta, \\
\varepsilon(x,y,z) &= \int_{-\infty}^{\infty} \varepsilon(\alpha,\beta,z) \exp\left(ik\left(\alpha x + \beta y\right)\right) \mathrm{d}\alpha\, \mathrm{d}\beta, \\
\varepsilon^{-1}(x,y,z) &= \int_{-\infty}^{\infty} \varepsilon^{-1}(\alpha,\beta,z) \exp\left(ik\left(\alpha x + \beta y\right)\right) \mathrm{d}\alpha\, \mathrm{d}\beta,
\end{aligned} \quad (3.287)$$

где символ $\varepsilon^{-1}(x,y,z) = 1/\varepsilon(x,y,z)$, но $\varepsilon^{-1}(\alpha,\beta,z) \neq 1/\varepsilon(\alpha,\beta,z)$, $\varepsilon^{-1}(\alpha,\beta,z)$ просто удобное обозначение. Функции $\mathbf{E}(\alpha,\beta,z)$ и $\mathbf{H}(\alpha,\beta,z)$ — поперечные пространственно-частотные компоненты.

Подставляя разложения (3.287) в уравнения Максвелла в форме (3.286), получаем уравнения Максвелла (в пространственно-частотных координатах):

$$\partial_z E_x(\alpha,\beta,z) = \\ = -ik\alpha \int\limits_{-\infty}^{\infty} (\omega_x H_y(\omega_x,\omega_y,z) - \omega_y H_x(\omega_x,\omega_y,z))\varepsilon(\alpha-\omega_x,\beta-\omega_y)\,\mathrm{d}\omega_x\,\mathrm{d}\omega_y + \\ + ikH_y(\alpha,\beta,z),$$

$$\partial_z E_y(\alpha,\beta,z) = \\ = -ik\beta \int\limits_{-\infty}^{\infty} (\omega_x H_y(\omega_x,\omega_y,z) - \omega_y H_x(\omega_x,\omega_y,z))\varepsilon(\alpha-\omega_x,\beta-\omega_y)\,\mathrm{d}\omega_x\,\mathrm{d}\omega_y - \\ - ikH_x(\alpha,\beta,z),$$

$$\partial_z H_x(\alpha,\beta,z) = ik\alpha(\alpha E_y(\alpha,\beta,z) - \beta E_x(\alpha,\beta,z)) - \\ - ik\int\limits_{-\infty}^{\infty} \varepsilon(\alpha-\omega_x,\beta-\omega_y)E_y(\omega_x,\omega_y,z)\,\mathrm{d}\omega_x\,\mathrm{d}\omega_y,$$

$$\partial_z H_y(\alpha,\beta,z) = ik\beta(\alpha E_y(\alpha,\beta,z) - \beta E_x(\alpha,\beta,z)) + \\ + ik\int\limits_{-\infty}^{\infty} \varepsilon(\alpha-\omega_x,\beta-\omega_y)E_x(\omega_x,\omega_y,z)\,\mathrm{d}\omega_x\,\mathrm{d}\omega_y. \quad (3.288)$$

Рассмотрим теперь дифракцию на двумерном пропускающем ДОЭ. Решение данной задачи во многом аналогично решению задачи дифракции плоской волны на одномерной пропускающей дифракционной решетке (см. 3.3). Однако, в отличие от рис. 3.5 и рис. 3.6 оси координат расставлены в другом порядке. Ось Oz перпендикулярна плоскости оптического элемента или плоскости дифракционной структуры. Пусть далее первая и третья зоны на рис. 3.5 соответствуют областям $z > a$, где a — максимальная высота рельефа структуры и $z < 0$ с постоянной диэлектрической проницаемостью. Во второй зоне $z \in [0,a]$ — зоне модуляции, диэлектрическая проницаемость является функцией координат $\varepsilon = \varepsilon(x,y,z)$.

Поле в первой зоне состоит из падающей волны, частотные компоненты которой имеют вид (см. обозначения в (3.269))

$$\mathbf{E}(\alpha,\beta,z) = \\ = \left\{ \frac{-\sqrt{1-\alpha^2-\beta^2}}{\alpha^2+\beta^2}(\alpha\mathbf{e}_x + \beta\mathbf{e}_y)I^e(\alpha,\beta) - \frac{\beta\mathbf{e}_x - \alpha\mathbf{e}_y}{\alpha^2+\beta^2}I^h(\alpha,\beta) + I^e(\alpha,\beta)\mathbf{e}_z \right\} \times \\ \times \exp\left(ik\left(-\sqrt{1-\alpha^2-\beta^2}z\right)\right),$$

$$\mathbf{H}(\alpha,\beta,z) =$$
$$= \left\{ \frac{-\sqrt{1-\alpha^2-\beta^2}}{\alpha^2+\beta^2}\left(\alpha\mathbf{e}_x+\beta\mathbf{e}_y\right)I^h(\alpha,\beta) + \frac{\beta\mathbf{e}_x-\alpha\mathbf{e}_y}{\alpha^2+\beta^2}I^e(\alpha,\beta) + I^h(\alpha,\beta)\mathbf{e}_z \right\} \times$$
$$\times \exp\left(ik\left(-\sqrt{1-\alpha^2-\beta^2}z\right)\right), \quad (3.289)$$

и отраженной волны, компоненты которой имеют вид

$$\mathbf{E}(\alpha,\beta,z) =$$
$$= \left\{ \frac{-\sqrt{1-\alpha^2-\beta^2}}{\alpha^2+\beta^2}\left(\alpha\mathbf{e}_x+\beta\mathbf{e}_y\right)R^e(\alpha,\beta) - \frac{\beta\mathbf{e}_x-\alpha\mathbf{e}_y}{\alpha^2+\beta^2}R^h(\alpha,\beta) + R^e(\alpha,\beta)\mathbf{e}_z \right\} \times$$
$$\times \exp\left(ik\left(\sqrt{1-\alpha^2-\beta^2}z\right)\right),$$

$$\mathbf{H}(\alpha,\beta,z) =$$
$$= \left\{ \frac{-\sqrt{1-\alpha^2-\beta^2}}{\alpha^2+\beta^2}\left(\alpha\mathbf{e}_x+\beta\mathbf{e}_y\right)R^h(\alpha,\beta) + \frac{\beta\mathbf{e}_x-\alpha\mathbf{e}_y}{\alpha^2+\beta^2}R^e(\alpha,\beta) + R^h(\alpha,\beta)\mathbf{e}_z \right\} \times$$
$$\times \exp\left(ik\left(-\sqrt{1-\alpha^2-\beta^2}z\right)\right). \quad (3.290)$$

Поле в третьей зоне состоит из прошедшей волны, компоненты которой имеют вид

$$\mathbf{E}(\alpha,\beta,z) =$$
$$= \left\{ \frac{-\sqrt{1-\alpha^2-\beta^2}}{\alpha^2+\beta^2}\left(\alpha\mathbf{e}_x+\beta\mathbf{e}_y\right)T^e(\alpha,\beta) - \frac{\beta\mathbf{e}_x-\alpha\mathbf{e}_y}{\alpha^2+\beta^2}T^h(\alpha,\beta) + T^e(\alpha,\beta)\mathbf{e}_z \right\} \times$$
$$\times \exp\left(ik\left(-\sqrt{1-\alpha^2-\beta^2}z\right)\right),$$

$$\mathbf{H}(\alpha,\beta,z) =$$
$$= \left\{ \frac{-\sqrt{1-\alpha^2-\beta^2}}{\alpha^2+\beta^2}\left(\alpha\mathbf{e}_x+\beta\mathbf{e}_y\right)T^h(\alpha,\beta) + \frac{\beta\mathbf{e}_x-\alpha\mathbf{e}_y}{\alpha^2+\beta^2}T^e(\alpha,\beta) + T^h(\alpha,\beta)\mathbf{e}_z \right\} \times$$
$$\times \exp\left(ik\left(-\sqrt{1-\alpha^2-\beta^2}z\right)\right). \quad (3.291)$$

Пространственно-частотные компоненты в зоне модуляции удовлетворяют системе интегро-дифференциальных уравнений (3.289).

Для удобства дальнейшего анализа перейдем к операторно-матричной форме. В операторной записи (3.286) имеет вид

$$\partial_z W = HW, \quad (3.292)$$

где W — матрица-столбец из 4-х компонентов:

$$W = \begin{pmatrix} E_x \\ E_y \\ H_x \\ H_y \end{pmatrix},$$

H — блочно-матричный дифференциальный оператор:

$$H = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{3.293}$$

$$A = \frac{i}{k} \begin{pmatrix} \partial_x & 0 \\ 0 & \partial_y \end{pmatrix} \begin{pmatrix} \varepsilon^{-1} & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix} \begin{pmatrix} -\partial_y & \partial_x \\ -\partial_y & \partial_x \end{pmatrix} + ik \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$B = \frac{1}{ik} \begin{pmatrix} \partial_x & 0 \\ 0 & \partial_y \end{pmatrix} \begin{pmatrix} -\partial_y & \partial_x \\ -\partial_y & \partial_x \end{pmatrix} + ik \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Представление (3.287) в матричной форме имеет вид

$$W(x, y, z) = \int\limits_{-\infty}^{\infty} W(\alpha, \beta, z) \exp\left(ik\left(\alpha x + \beta y\right)\right) \mathrm{d}\alpha \, \mathrm{d}\beta. \tag{3.294}$$

Система уравнений в пространственно-частотном представлении (3.288) имеет вид

$$\partial_z W(\alpha, \beta, z) = H(\alpha, \beta, z) W(\alpha, \beta, z), \tag{3.295}$$

$$H(\alpha, \beta, z) = \begin{pmatrix} A(\alpha, \beta, z) & 0 \\ 0 & B(\alpha, \beta, z) \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

где двумерные матричные операторы действуют на двумерные матрицы столбцы по формулам

$$A(\alpha, \beta) \begin{pmatrix} \varphi \\ \psi \end{pmatrix} = \frac{i}{k} \begin{pmatrix} ik\alpha & 0 \\ 0 & ik\beta \end{pmatrix} \times$$

$$\times \int\limits_{-\infty}^{\infty} \varepsilon^{-1}(\alpha - \omega_x, \beta - \omega_y, z) \begin{pmatrix} -ik\omega_y & ik\omega_x \\ -ik\omega_y & ik\omega_x \end{pmatrix} \begin{pmatrix} \varphi(\omega_x, \omega_y, z) \\ \psi(\omega_x, \omega_y, z) \end{pmatrix} \mathrm{d}\omega_x \, \mathrm{d}\omega_y +$$

$$+ ik \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \varphi \\ \psi \end{pmatrix},$$

$$B(\alpha, \beta) \begin{pmatrix} \varphi \\ \psi \end{pmatrix} = \frac{1}{ik} \begin{pmatrix} ik\alpha & 0 \\ 0 & ik\beta \end{pmatrix} \begin{pmatrix} -ik\beta & ik\alpha \\ -ik\beta & ik\alpha \end{pmatrix} \begin{pmatrix} \varphi \\ \psi \end{pmatrix} +$$

$$+ ik \int\limits_{-\infty}^{\infty} \varepsilon\left(\alpha - \omega_x, \beta - \omega_y\right) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \varphi(\omega_x, \omega_y) \\ \psi(\omega_x, \omega_y) \end{pmatrix} \mathrm{d}\omega_x \, \mathrm{d}\omega_y. \tag{3.296}$$

Рассмотрим подробнее случай, когда электромагнитное поле распространяется в вакууме. В этом случае система интегро-дифференциальных уравнений (3.288) превращается в систему обыкновенных дифференциальных уравнений:

$$\begin{aligned} \partial_z E_x(\alpha, \beta, z) &= -ik\alpha \left(\alpha H_y(\alpha, \beta, z) - \beta H_x(\alpha, \beta, z)\right) + ikH_y(\alpha, \beta, z), \\ \partial_z E_y(\alpha, \beta, z) &= -ik\beta \left(\alpha H_y(\alpha, \beta, z) - \beta H_x(\alpha, \beta, z)\right) - ikH_x(\alpha, \beta, z), \\ \partial_z H_x\left(\alpha, \beta, z\right) &= ik\alpha \left(\alpha E_y(\alpha, \beta, z) - \beta E_x(\alpha, \beta, z)\right) - ikE_y(\alpha, \beta, z), \\ \partial_z H_y\left(\alpha, \beta, z\right) &= ik\beta \left(\alpha E_y(\alpha, \beta, z) - \beta E_x(\alpha, \beta, z)\right) + ikE_x(\alpha, \beta, z). \end{aligned} \tag{3.297}$$

Общее решение системы уравнений (3.297) имеет вид

$$W(\alpha,\beta,z) = \left(E^+(\alpha,\beta)W^{+e}(\alpha,\beta) + H^+(\alpha,\beta)W^{+h}(\alpha,\beta)\right)\exp\left(ik\sqrt{1-\alpha^2-\beta^2}z\right) + \\ + \left(E^-(\alpha,\beta)W^{-e}(\alpha,\beta) + H^-(\alpha,\beta)W^{-h}(\alpha,\beta)\right)\exp\left(-ik\sqrt{1-\alpha^2-\beta^2}z\right),$$

$$\|W\|\,W^{\pm e}(\alpha,\beta) = \begin{pmatrix} \mp\sqrt{1-\alpha^2-\beta^2}\alpha \\ \mp\sqrt{1-\alpha^2-\beta^2}\beta \\ \beta \\ -\alpha \end{pmatrix}, \tag{3.298}$$

$$\|W\|\,W^{\pm h}(\alpha,\beta) = \begin{pmatrix} -\beta \\ \alpha \\ \mp\sqrt{1-\alpha^2-\beta^2}\alpha \\ \mp\sqrt{1-\alpha^2-\beta^2}\beta \end{pmatrix},$$

$$\|W\| = \sqrt{(\alpha^2+\beta^2)\left(1+\sqrt{1-\alpha^2-\beta^2}\right)^2}$$

где E^+, E^-, H^+, H^- — определяют вклад электрической и магнитной волн в заданном направлении. Верхний знак соответствует волнам, распространяющимся в положительном направлении оси z, а нижний знак — в отрицательном. Отметим, что переход к случаю однородной диэлектрической среды с показателем преломления ε осуществляется с помощью следующего преобразования:

$$\alpha \to \frac{\alpha}{\sqrt{\varepsilon}}, \quad \beta \to \frac{\beta}{\sqrt{\varepsilon}}, \quad z \to \sqrt{\varepsilon}z, \quad H_x \to \sqrt{\varepsilon}H_x, \quad H_y \to \sqrt{\varepsilon}H_y.$$

Перепишем выражение (3.5.3) в блочно-матричной форме:

$$W(\alpha,\beta,z) = W_1 S P, \tag{3.299}$$

$$P = \begin{pmatrix} E^+(\alpha,\beta) \\ H^+(\alpha,\beta) \\ E^-(\alpha,\beta) \\ H^-(\alpha,\beta) \end{pmatrix}, S = \begin{pmatrix} \exp\left(ik\sqrt{1-\alpha^2-\beta^2}z\right)E_2 & 0 \\ 0 & \exp\left(-ik\sqrt{1-\alpha^2-\beta^2}z\right)E_2 \end{pmatrix},$$

где $W_1 = \begin{pmatrix} W^{+e}(\alpha,\beta) & W^{+h}(\alpha,\beta) & W^{-e}(\alpha,\beta) & W^{-h}(\alpha,\beta) \end{pmatrix}$ — квадратная матрица, состоящая из столбцов, а E_2 — единичная матрица:

$$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Связь между пространственно-частотными компонентами в различных плоскостях вдоль направления распространения z имеет вид

$$W(\alpha,\beta,z) = W_1 S M^{-1} W^*{}_1 W(\alpha,\beta,0), \tag{3.300}$$

где $W^*{}_1$ — матрица эрмитово-сопряженная к W_1,

$$M = \begin{pmatrix} E_2 & W^{\pm}E_2 \\ W^{\pm}E_2 & E_2 \end{pmatrix}, \quad W^{\pm} = W^{+e*}W^{-e} = \frac{\left(\alpha^2+\beta^2\right)^2}{\|W\|^2}.$$

Подставляя (3.299) в (3.287), получаем выражение для оператора распространения электромагнитного поля T в свободном пространстве:

$$W(x,y,z) = TW(x,y,0) = \frac{k^2}{4\pi^2} \times$$
$$\times \int\limits_{-\infty}^{\infty} W_1 S M^{-1} W^*{}_1 \left(\int\limits_{-\infty}^{\infty} W(x',y',0) \exp\left(ik\left(\alpha(x-x') + \beta(y-y')\right)\right) \mathrm{d}x'\,\mathrm{d}y' \right) \mathrm{d}\alpha\,\mathrm{d}\beta. \tag{3.301}$$

Преимущество интегрального метода (3.301) состоит в том, что в его основе лежит вычисление прямого и обратного преобразования Фурье. Это позволяет использовать алгоритмы быстрого преобразования Фурье, что в свою очередь позволяет значительно сократить время вычислений по сравнению с решением уравнений Максвелла разностными методами.

Введем в пространстве четырехкомпонентных векторов функций W понятие скалярного произведения. Введем скалярное произведение таким образом, чтобы оно не зависело от координаты z, и в тоже время произведение вектора самого на себя было пропорционально потоку вектора Умова–Пойнтинга. Для этого запишем (3.256), (3.257) в обычной форме для $\mathbf{E}_1$, $\mathbf{H}_2$ и для комплексно сопряженных компонентов:

$$\begin{aligned} \operatorname{rot}\mathbf{E}_1 &= ik\mathbf{H}_1, \qquad \operatorname{rot}\mathbf{E}_2^* = -ik\mathbf{H}_2^*, \\ \operatorname{rot}\mathbf{H}_1 &= -ik\mathrm{E}_1, \quad \operatorname{rot}\mathbf{H}_2^* = ik\mathbf{E}_2^*. \end{aligned} \tag{3.302}$$

Используя известную формулу векторного анализа $\operatorname{div}[\mathbf{a},\mathbf{b}] = \mathbf{b}\operatorname{rot}\mathbf{a} - \mathbf{a}\operatorname{rot}\mathbf{b}$, получаем следующее выражение:

$$\operatorname{div}\left([\mathbf{E_1},\mathbf{H_2^*}] + [\mathbf{E_2^*},\mathbf{H_1}]\right) = 0. \tag{3.303}$$

Далее, используя теорему Остроградского–Гаусса и учитывая условия излучения (с целью зануления интеграла по боковой поверхности), получаем выражение

$$\iint \left([\mathbf{E_1}(x,y,z_1),\mathbf{H_2^*}(x,y,z_1)] + [\mathbf{E_2^*}(x,y,z_1),\mathbf{H_1}(x,y,z_1)]\right) \mathrm{d}x\,\mathrm{d}y =$$
$$= \iint \left([\mathbf{E_1}(x,y,z_2),\mathbf{H_2^*}(x,y,z_2)] + [\mathbf{E_2^*}(x,y,z_2),\mathbf{H_1}(x,y,z_2)]\right) \mathrm{d}x\,\mathrm{d}y. \tag{3.304}$$

Запишем выражение (3.304) в следующем виде:

$$\iint A^T((x,y,z_1))\Omega B^*(x,y,z_1)\,\mathrm{d}x\,\mathrm{d}y = \iint A^T((x,y,z_2))\Omega B^*(x,y,z_2)\,\mathrm{d}x\,\mathrm{d}y, \tag{3.305}$$

где

$$\Omega = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tag{3.306}$$

или

$$\iint \langle A(x,y,z_2), B(x,y,z_2)\rangle \,\mathrm{d}x\,\mathrm{d}y = \iint \langle A(x,y,z_1), B(x,y,z_1)\rangle \,\mathrm{d}x\,\mathrm{d}y, \tag{3.307}$$

где $\langle A,B\rangle = \left(A^T\right)^* \Omega B$, $\left(A^T\right)^*$ — эрмитово-сопряженная матрица. Формула (3.307) означает, что оператор распространения (3.301) сохраняет скалярное произведение

$$\langle A \circ B\rangle = \iint \langle A,B\rangle \,\mathrm{d}x\,\mathrm{d}y, \tag{3.308}$$

т.е. является в пространстве с данным скалярным произведением унитарным оператором. Свойство сохранения скалярного произведения можно использовать для решения обратных задач дифракции. Это в значительной мере упрощает вычисление градиента функционала невязки по аналогии с тем, как это делается в случае скалярного приближения [13–14]. Кроме того, наличие сохраняющейся величины можно использовать для контроля правильности решения прямой задачи дифракции.

Пример 3.12. Приведем результаты расчета интенсивности (рис. 3.21) от геометрооптического фокусатора (см. главу 5) гауссова пучка в тонкое кольцо и в квадрат. В этом случае оператор W в плоскости непосредственно за оптическим элементом имеет вид

$$W(x,y,0) = \frac{1}{2}\sqrt{I_0(x,y)}\exp\left(i\varphi(x,y)\right)\begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \tag{3.309}$$

где $I_0(x,y)$ — интенсивность падающего пучка, $\varphi(x,y)$ — фазовая функция геометрооптического фокусатора. Распространение излучения внутри оптического элемента рассчитывалось в приближении геометрической оптики. В этом случае связь между фазовой функцией и рельефом имеет вид $\varphi(x,y) = 2\,(n-1)\,h(x,y)$, где n — показатель преломления материала ДОЭ, $h(x,y)$ — высота рельефа в данной точке. Были использованы следующие параметры: размер ДОЭ $D_x = 256\lambda$, $D_y = 256\lambda$, параметр гауссова пучка (радиус перетяжки) $\sigma = 64\lambda$, фокусное расстояние $f = 512\lambda$.

Пример 3.13. Приведем сравнение численных результатов расчета дифракции (рис. 3.22) от геометрооптического фокусатора гауссова пучка в квадрат, полученных в приближении Френеля и в рамках электромагнитной теории при следующих

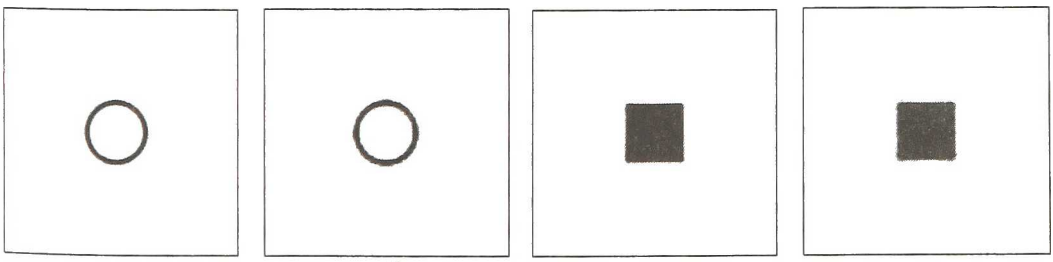
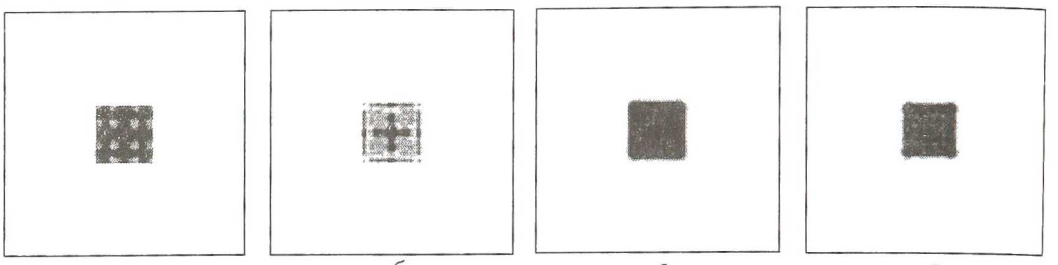


Рис. 3.21. Распределение интенсивности от геометрооптических фокусаторов в кольцо и квадрат: шаг дискретизации $\Delta = \lambda$, радиус кольца $r_0 = 32\lambda$ (*а*); шаг дискретизации $\Delta = 4\lambda$, радиус кольца $r_0 = 32\lambda$(*б*); шаг дискретизации $\Delta = \lambda$, размер квадрата $a \times b = 32\lambda \times 32\lambda$ (*в*); шаг дискретизации $\Delta = 4\lambda$, размер квадрата $a \times b = 32\lambda \times 32\lambda$ (*г*)



а *б* *в* *г*

Рис. 3.22. Распределение интенсивности от фокусатора в квадрат при различных параметрах и различных методах расчета

параметрах: шаг дискретизации на элементе $\Delta = \lambda$, размеры апертуры элемента $D_x = 64\lambda, D_y = 64\lambda$, фокусное расстояние $f = 64\lambda$, размер квадрата в плоскости фокусировки, $a \times b = 16\lambda \times 16\lambda$. На рис. 3.22 показаны результаты расчета: (*а*) $\sigma = 16\lambda$ — параметр гауссова пучка, дифракционная эффективность — 93%, среднеквадратичное отклонение — 20% (электромагнитная теория); (*б*) $\sigma = 16\lambda$ — параметр гауссова пучка, дифракционная эффективность — 94%, среднеквадратичное отклонение — 17% (скалярная теория); (*в*) $\sigma = 32\lambda$ — параметр гауссова пучка, дифракционная эффективность — 96%, среднеквадратичное отклонение — 49% (электромагнитная теория); (*г*) $\sigma = 32\lambda$ — параметр гауссова пучка, дифракционная эффективность — 91%, среднеквадратичное отклонение — 26% (скалярная теория)

Полученные результаты говорят о том, что при данных параметрах применение электромагнитной теории не приводит к существенным изменениям результата фокусировки гауссового пучка с помощью фокусатора (см. главу 5), рассчитанного методом геометрической оптики. Хотя, строго говоря, электромагнитная теория использовалась в этом примере только при распространении света в свободном пространстве, а при распространении в среде ДОЭ выбрано приближение геометрической оптики. Из приведенных примеров можно заметить, что наиболее сильные различия наблюдаются, когда размер освещающего пучка больше апертуры ДОЭ.

Вернемся к решению задачи дифракции электромагнитной волны на пропускающем ДОЭ. Перепишем формулы (3.289)–(3.291) в матричном виде. Поле в первой зоне представляет собой суперпозицию падающей волны, частотные компоненты которой имеют вид

$$W(\alpha,\beta,z) = \\ = \left(I^e(\alpha,\beta)W^{-e}(\alpha,\beta) + I^h(\alpha,\beta)W^{-h}(\alpha,\beta)\right)\exp\left(-ik\sqrt{1-\alpha^2-\beta^2}z\right), \quad (3.310)$$

и отраженной от поверхности оптического элемента,

$$W(\alpha,\beta,z) = \\ = \left(R^e(\alpha,\beta)W^{+e}(\alpha,\beta) + R^h(\alpha,\beta)W^{+h}(\alpha,\beta)\right)\exp\left(ik\sqrt{1-\alpha^2-\beta^2}z\right). \quad (3.311)$$

Поле в третьей зоне (рис 3.6) за оптическим элементом имеет вид

$$W(\alpha,\beta,z) = \\ = \left(T^e(\alpha,\beta)W^{-e}(\alpha,\beta) + T^h(\alpha,\beta)W^{-h}(\alpha,\beta)\right)\exp\left(-ik\sqrt{1-\alpha^2-\beta^2}z\right). \quad (3.312)$$

Поле в зоне модуляции удовлетворяет системе интегро-дифференциальных уравнений (3.295). Рассмотрим метод решения системы (3.295). Вместо непрерывных

функций введем их значения в точках $\alpha_n = \alpha_0 + \lambda n/d_x$, $\beta_m = \beta_0 + \lambda m/d_y$:

$$\left.\begin{aligned} W^{\pm e}_{nm} &= W^{\pm e}(\alpha_n, \beta_m), & W^{\pm h}_{nm} &= W^{\pm h}(\alpha_n, \beta_m), \\ I^e_{nm} &= \frac{I^e(\alpha_n, \beta_m)\lambda^2}{d_x d_y}, & I^h_{nm} &= \frac{I^h(\alpha_n, \beta_m)\lambda^2}{d_x d_y}, \\ R^e_{nm} &= \frac{R^e(\alpha_n, \beta_m)\lambda^2}{d_x d_y}, & R^h_{nm} &= \frac{R^h(\alpha_n, \beta_m)\lambda^2}{d_x d_y}, \\ T^e_{nm} &= \frac{T^e(\alpha_n, \beta_m)\lambda^2}{d_x d_y}, & T^h_{nm} &= \frac{T^h(\alpha_n, \beta_m)\lambda^2}{d_x d_y}. \end{aligned}\right\} \tag{3.313}$$

Система интегро-дифференциальных уравнений (3.313) превращается в систему дифференциальных уравнений первого порядка относительно функций $W_{nm}(z)$:

$$\begin{aligned} \partial_z W_{nm}(z) &= \sum_{sk} H_{nm,sk}(z) W_{sk}(z), \\ H_{nm,sk}(z) &= \begin{pmatrix} A_{nm,sk}(z) & 0 \\ 0 & B_{nm,sk}(z) \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \end{aligned} \tag{3.314}$$

где двумерные матричные операторы действуют на двумерные матрицы-столбцы по формулам, следующим из уравнений (3.296):

$$\begin{aligned} A_{nm,sk} \begin{pmatrix} \varphi_{sk}(z) \\ \psi_{sk}(z) \end{pmatrix} &= \frac{i}{k} \begin{pmatrix} ik\alpha_n & 0 \\ 0 & ik\beta_m \end{pmatrix} \varepsilon^{-1}_{n-s,m-k} \begin{pmatrix} -ik\beta_k & ik\alpha_s \\ -ik\beta_k & ik\alpha_s \end{pmatrix} \begin{pmatrix} \varphi_{sk}(z) \\ \psi_{sk}(z) \end{pmatrix} + \\ &\quad + ik\delta_{ns,mk} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \varphi_{sk}(z) \\ \psi_{sk}(z) \end{pmatrix}, \\ B_{nm,sk} \begin{pmatrix} \varphi_{sk}(z) \\ \psi_{sk}(z) \end{pmatrix} &= \frac{1}{ik} \begin{pmatrix} ik\alpha_n & 0 \\ 0 & ik\beta_m \end{pmatrix} \delta_{nm,sk} \begin{pmatrix} -ik\beta_k & ik\alpha_s \\ -ik\beta_k & ik\alpha_s \end{pmatrix} \begin{pmatrix} \varphi_{sk} \\ \psi_{sk} \end{pmatrix} + \\ &\quad + ik\varepsilon_{n-s,m-k} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \varphi_{sk}(z) \\ \psi_{sk}(z) \end{pmatrix}. \end{aligned} \tag{3.315}$$

Совокупность решений системы уравнений (3.314) с различными индексами удобно объединить в блочную квадратную матрицу $\tilde{W}$. Элементами этой блочной матрицы являются четырехкомпонентные матрицы столбцы. Пусть $\tilde{W}^e_{nm}(z)$, $\tilde{W}^h_{nm}(z)$ — решение системы уравнений (3.314) с начальными условиями $W^{-e}_{nm}(0)$ и $W^{-h}_{nm}(0)$, где $W^{-e}_{nm}(0)$ или $W^{-h}_{nm}(0)$ представляет собой блочную квадратную матрицу, у которой отличен от нуля только элемент, стоящий на пересечении m-й строки и n-го столбца. Этот элемент равен четырехкомпонентному столбцу W^{-e}_{nm} или W^{-h}_{nm} (не следует путать обозначения $W^{-e}_{nm}(0)$, $W^{-h}_{nm}(0)$ и W^{-e}_{nm} или W^{-h}_{nm}). Условие непрерывности пространственно-частотных компонентов при $z = a$ имеет вид

$$\begin{aligned} &\sum_{mn} \left(T^e_{nm} W^{-e}_{nm}(a) + T^h_{nm} W^{-h}_{nm}(a)\right) = \\ &\qquad = \sum_{mn} \left(I^e_{nm} W^{-e}_{nm}(0) + I^h_{nm} W^{-h}_{nm}(0)\right) \exp\left(-ik\sqrt{1 - \alpha_n^2 - \beta_m^2}\, a\right) + \end{aligned}$$

$$+\sum_{mn}\left(R^{e}_{nm}W^{+e}_{nm}+R^{h}_{nm}W^{+h}(0)_{nm}\right)\exp\left(ik\sqrt{1-\alpha_n^2-\beta_m^2}a\right). \quad (3.316)$$

Последнее выражение представляет собой систему линейных уравнений для определения коэффициентов пропускания и отражения. Поле, прошедшее через оптический элемент, имеет вид

$$W(x,y,z)=\sum_{nm}\left(T^{e}_{nm}W^{-e}_{nm}+T^{h}_{nm}W^{-h}_{nm}\right)\times \\ \times\exp\left(-ik\sqrt{1-\alpha_n^2-\beta_m^2}z+ik\left(\alpha_n x+\beta_m y\right)\right). \quad (3.317)$$

3.6. Градиентный метод синтеза ДОЭ

Рассмотрим метод восстановления волновых полей по результатам измерения интенсивности электромагнитного поля в двух различных плоскостях. С помощью этого метода можно рассчитать электромагнитное поле, которое должно быть сразу за ДОЭ и которое при распространении в пространстве сформирует заданное распределение интенсивности на требуемом растоянии от ДОЭ. Далее рассчитанное поле может быть использовано при синтезе рельефа ДОЭ как граничное условие при решении уравнений Максвелла в пропускающей диэлектрической среде оптического элемента.

В качестве интенсивности в данном случае выступает величина, равная проекции вектора Умова–Пойнтинга на ось z декартовой системы координат. Пусть $I_0(x,y)$ – распределение интенсивности в плоскости $z=0$, $I(x,y)$ — распределение интенсивности в плоскости $z=f$. Требуется найти волновое поле $W(x,y,z)$, удовлетворяющее следующим условиям:

$$\begin{cases} W(x,y,f)=TW(x,y,0) \\ I_0(x,y)=\langle W(x,y,0),W(x,y,0)\rangle \\ I(x,y)=\langle W(x,y,f),W(x,y,f)\rangle \end{cases}, \quad (3.318)$$

где T — оператор распространения из (3.301).

Точное решение задачи в общем случае не существует, поэтому заменим точную постановку задачи приближенной:

$$\begin{cases} \varepsilon\left[W(x,y,f\right]\to\min, \\ \varepsilon=\displaystyle\int\left(\sqrt{I\left(x,y\right)}-\sqrt{\langle W(x,y,f),W(x,y,f)\rangle}\right)^2\mathrm{d}x\,\mathrm{d}y, \\ W(x,y,f)=TW(x,y,0), \\ I_0(x,y)=\langle W(x,y,0),W(x,y,0)\rangle. \end{cases} \quad (3.319)$$

Для простоты будем искать минимум в классе функций, имеющих следующий вид:

$$W(x,y,0)=\frac{1}{2}\sqrt{I_0(x,y)}\exp\left(i\varphi(x,y)\right)\begin{pmatrix}\cos\alpha(x,y)\\ -\sin\alpha(x,y)\exp\left(i\delta(x,y)\right)\\ \sin\alpha(x,y)\exp\left(i\delta(x,y)\right)\\ \cos\alpha(x,y)\end{pmatrix}. \quad (3.320)$$

При данной форме записи каждый из параметров в приближении геометрической оптики имеет простой физический смысл. Функция $\varphi(x,y)$ — определяет фазу волны в плоскости $z=0$, $\alpha(x,y), \delta(x,y)$ — определяют поляризацию волны.

Найдем градиент функционала из (3.319):

$$\delta\varepsilon = \mathrm{Re}\left(\langle\psi(\mathbf{x},0), \delta W(\mathbf{x},0)\rangle\right), \tag{3.321}$$

где

$$\psi(x,0) = T^{-1}\left(\sqrt{\frac{I(\mathbf{x})}{\langle W(\mathbf{x},f), W(\mathbf{x},f)\rangle}}W(\mathbf{x},f)\right), \quad \delta W(\mathbf{x},f) = T(\delta W(\mathbf{x},0)).$$

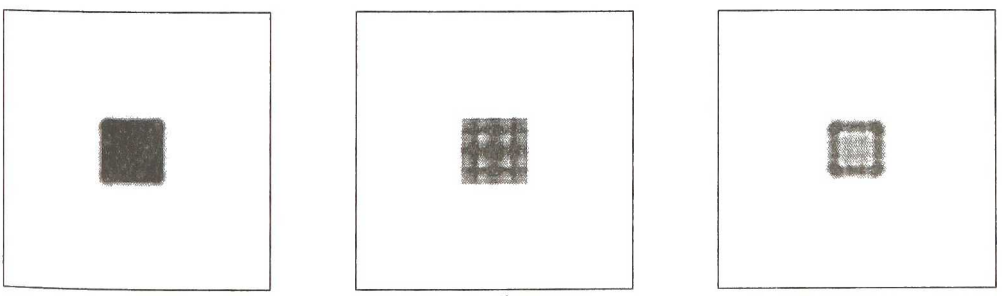
При выводе формулы (3.321) было использовано свойство сохранения псевдоскалярного произведения. Переходя к дискретному представлению и используя (3.318)–(3.321), получим выражение для градиента функционала в каждой точке (u,v):

$$\nabla\varepsilon(\mathbf{x},0) = \langle L(\mathbf{x},0), \psi(\mathbf{x},0)\rangle + \langle\psi(\mathbf{x},0), L(\mathbf{x},0)\rangle = 2\,\mathrm{Re}\,\langle L(\mathbf{x},0), \psi(\mathbf{x},0)\rangle, \tag{3.322}$$

где $L(\mathbf{x},0) = \dfrac{i}{2}\sqrt{I_0(\mathbf{x},0)}\exp(i\varphi(\mathbf{x},0))E(\mathbf{x},0)$, $E(\mathbf{x},0) = \begin{pmatrix}1\\0\\0\\1\end{pmatrix}$.

Далее, с учетом (3.322), можно использовать обычную итеративную процедуру минимизации функционала $\varphi_{n+1}(\mathbf{x},0) = \varphi_n(\mathbf{x},0) - t\nabla\varepsilon$, t — шаг градиентного алгоритма. При этом решение, которое получится в результате итеративного процесса будет зависеть от начального приближения. Это связано с тем, что функционал обычно не имеет глобального минимума, а имеет несколько локальных минимумов. При использовании данного метода для расчета ДОЭ не возникает проблем с множеством решений задачи, так как подходит любое решение, минимизирующее исходный функционал.

Пример 3.14. Приведем расчет интенсивности на выходе ДОЭ, фокусирующего в квадрат, расположенный на расстоянии f от плоскости ДОЭ (рис. 3.23). На рис. 3.23*а* показано распределение интенсивности от фокусатора гауссова пучка в квадрат, синтезированного в приближении Френеля. Расчет интенсивности также приведен в приближении Френеля. На рис. 3.23*б* показано распределение интенсивности от фокусатора гауссова пучка в квадрат синтезированного, в приближении Френеля, но расчет интенсивности проводился в рамках электромагнитной теории. На рис. 3.23*в* изображено распределение интенсивности от фокусатора гауссова пучка в квадрат,



а *б* *в*

Рис. 3.23. Распределение интенсивности от фокусатора гауссова пучка в квадрат, рассчитанного итерационным методом в различных приближениях

синтезированного в рамках электромагнитной теории. Шаг дискретизации на элементе $\Delta = \lambda$, размеры апертуры элемента $D_x = 64\lambda$, $D_y = 64\lambda$, $f = 64\lambda$ — фокусное расстояние, размер квадрата в плоскости фокусировки, $a \times b = 16\lambda \times 16\lambda$, $\sigma = 32\lambda$ – параметр гауссова пучка, дифракционная эффективность E, среднеквадратичное отклонение δ: $E = 94\%$, $\delta = 16\%$ (рис. 3.23а); $E = 96\%$, $\delta = 64\%$ (рис. 3.23*б*); $E = 97\%$, $\delta = 22\%$ (рис. 3.23*в*).

Полученные результаты свидетельствуют о том, что использование электромагнитной теории при описании распространения света в пространстве качественно не меняет результата решения обратной задачи. Количественные отклонения в большей степени зависят от параметров задачи. Результаты, полученные в скалярном приближении и в рамках электромагнитной теории, существенно отличаются в случае, если размер освещающего пучка превышает размер апертуры ДОЭ. В остальных случаях они совпадают с точностью до нескольких процентов. Заметим, что для адекватного описания процесса дифракции света необходимо учитывать прохождение электромагнитной волны через ДОЭ.

3.7. Асимптотический анализ дифракции на зонированных структурах

3.7.1. Решение задачи дифракции на одномерных ДОЭ в скалярном приближении

В настоящее время в электромагнитном приближении решены задачи дифракции света на простейших структурах: дифракционных решетках [4–7], шаре [8] и других. Это в первую очередь связано с тем, что расчет поля от более сложных пространственных структур требует больших вычислительных затрат. В связи с этим большую роль в теории дифракции играют асимптотические методы. В данном разделе рассмотрен асимптотический метод оценки поля при дифракции на бинарной квазипериодической структуре, представляющей совокупность дифракционных решеток с различным периодом. Приведены удобные представления для волновых полей. Данный подход оказывается эффективен при расчете поля от дифракционных оптических элементов.

Рассмотрим одномерную отражающую структуру, представляющую совокупность дифракционных решеток. Введем условные обозначения. Пусть X_M — границы локальных решеток, N_M — число периодов M-й дифракционной решетки. Если $N_M \to \infty$, то квазипериодическая структура превращается в строго периодическую дифракционную решетку. Если $N_M = 1$, все участки квазипериодической структуры имеют различный период. Обозначим x_m — границы периода локальной решетки

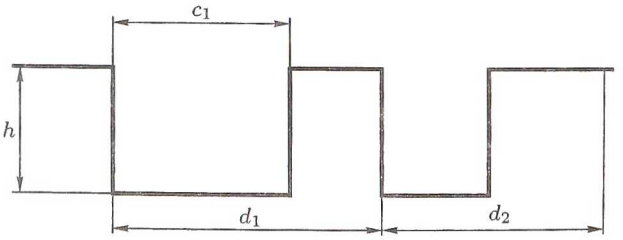


Рис. 3.24. Вид двух зон ДОЭ, которые имеют по одному периоду разных дифракционных решеток

(для границ периодов используется сплошная нумерация, независимо от того, какой решетке принадлежит данный период), $d_m = x_{m+1} - x_m$. Не ограничивая общности подхода, рассмотрим дифракционную решетку с одной канавкой на периоде [4–5] (см. рис. 3.24), c_m — ширина канавки, h — его глубина. Пусть плоская волна $\exp(ik(\alpha_0 x - \beta_0 z))$ падает на описанную выше квазипериодическую структуру.

Рассмотрим две области: 1 — область $z > 0$, 2 — область $z < 0$ (рис. 3.1). Поле в первой области представляет собой совокупность плоских волн:

$$W(x,z) = \exp(ik(\alpha_0 x - \beta_0 z)) + \bar{W}(x,z),$$
$$\bar{W}(x,z) = \int_{-\infty}^{\infty} A(\alpha)\exp(ik(\alpha x + \beta z))\,\mathrm{d}\alpha, \tag{3.323}$$

$$\frac{\partial W(x,z)}{\partial z} = -ik\beta_0 \exp(ik(\alpha_0 x - \beta_0 z)) + \int_{-\infty}^{\infty} ik\beta(\alpha)A(\alpha)\exp(ik(\alpha x + \beta z))\,\mathrm{d}\alpha, \tag{3.324}$$

где $\beta = \sqrt{1-\alpha^2}$, $A(\alpha)$ — спектр пространственных частот волн, отраженных от структуры, $|\alpha| < 1$ соответствует однородным плоским волнам; $|\alpha| > 1$ соответствует неоднородным плоским волнам, амплитуда которых убывает при удалении от поверхности квазипериодической структуры. Спектральная функция выражается через отраженную волну на поверхности следующим образом:

$$A(\alpha) = \frac{k}{2\pi}\int_{-\infty}^{\infty} \bar{W}(x,0)\exp(-ik\alpha x)\,\mathrm{d}x. \tag{3.325}$$

Пусть отраженная волна на поверхности структуры представляется в виде

$$\bar{W}(x) = \exp(ik\alpha_0 x)\sum_m\sum R_n^m \exp\left(ik\frac{\lambda n}{d_m}(x - x_m)\right)\mathrm{rect}(x, x_m, x_{m+1}), \tag{3.326}$$

$$\mathrm{rect}(x,a,b) = \begin{cases} 1, & x \in [a,b], \\ 0, & x \notin [a,b]. \end{cases} \tag{3.327}$$

Данное представление удобно тем, что по форме совпадает с представлением отраженного поля в случае дифракции на отражающей строго периодической структуре. Нахождение поля на поверхности сводится к определению дифракционных коэффициентов R_n^m. Далее будет показано, что в пределе выражения для дифракционных коэффициентов в квазипериодическом случае совпадают с коэффициентами, вычисленными для строго периодической решетки с параметрами d_m, c_m. Этот результат является основным для асимптотической оценки поля при дифракции на квазипериодической структуре.

Функция W и ее производная при $z < 0$ представляется в виде

$$W = \sum_k\sum_q a_q^k \sin\left(\frac{\pi q(x - x_k)}{c_k}\right)\sin(\Gamma_q^k(z+h))S^k(x), \tag{3.328}$$

$$\frac{\partial W}{\partial z} = \sum_k\sum_q a_q^k\Gamma_q^k \sin\left(\frac{\pi q(x - x_k)}{c_k}\right)\cos(\Gamma_q^k(z+h))S^k(x), \tag{3.329}$$

$$\Gamma_q^k = \sqrt{k^2 - \left(\frac{\pi q}{c_k}\right)^2}, \quad q = 1, 2, \dots, \tag{3.330}$$

$$S^k(x) = \begin{cases} 1, & x \in [x_k, x_k + c_k], \\ 0, & x \notin [x_k, x_k + c_k]. \end{cases} \tag{3.331}$$

(3.331) По аналогии с [4–5] запишем условия непрерывности $W(x, z)$ и ее производной на границе раздела между поверхностью дифракционной решетки и свободным полупространством. При $z = 0$ будем иметь

$$\exp(ik\alpha_0 x)\left(1 + \sum_n \sum_m R_n^m \Phi_n^m(x)\right) = \sum_k \sum_q a_q^k \sin\left(\frac{\pi q(x - x_k)}{bd_k}\right) \sin(\Gamma_q^k h) S^k(x),$$

$$\sum_s ik \exp(ik\alpha_0 x)\left(-\beta(\alpha_0) + \sum_n \sum_m R_n^m F_n^m(x)\right) S^s(x) =$$

$$= \sum_k \left(\sum_q a_q^k \sin\left(\frac{\pi q(x - x_k)}{bd_k}\right) \Gamma_q^k \cos(\Gamma_q^k h)\right) S^k(x), \tag{3.332}$$

где

$$\Phi_n^m = \exp\left(\frac{ik\lambda n}{d_m}(x - x_m)\right) \operatorname{rect}(x, x_m, x_{m+1}), \tag{3.333}$$

$$F_n^m(x) = \hat{F}_n^m(x) + \tilde{F}_n^m(x), \qquad \hat{F}_n^m(x) = \beta(\alpha_n^m)\Phi_n^m(x), \tag{3.334}$$

$$\tilde{F}_n^m(x) = \frac{i}{2\pi} \times$$

$$\times \int_{-\infty}^{\infty} (\beta(\alpha) - \beta(\alpha_n^m)) \frac{\exp(ik(\alpha - \alpha_0)(x - x_{m+1})) - \exp(ik(\alpha - \alpha_0)(x - x_m))}{(\alpha - \alpha_0) - \lambda n/d_m} \, \mathrm{d}\alpha, \tag{3.335}$$

$$\alpha_n^m = \alpha_0 + \frac{\lambda n}{d_m}, \quad c_k = bd_k.$$

Систему функциональных уравнений (3.332) можно привести к системе линейных алгебраических уравнений:

$$\sum_{mn} M_{pn}^k R_n^k = D_p^k, \tag{3.336}$$

где

$$M_{pn}^k = iA_{pn}^k + kd_k\delta_{pn}, \quad D_p^k = iA_{p0}^k - kd_k\delta_{p0} - B_p^k, \quad A_{pn}^k = \sum_{qn} \frac{\operatorname{tg}(\Gamma_q^k h)}{N_q^k \Gamma_q^k} \beta(\alpha_n^k) D_{nq}^k \bar{D}_{pq}^k,$$

$$D_{nq}^k = \int_{x_k}^{x_k + c_k} \exp(ik(\alpha_0 x + \lambda n/d_k(x - x_k))) \sin\left(\frac{\pi q(x - x_k)}{c_k}\right) \mathrm{d}x, \tag{3.337}$$

$$B_p^k = \sum \frac{\operatorname{tg}(\Gamma_q^k h)}{N_q^k \Gamma_q^k} \beta(\alpha_n^k) \Delta_q^k \bar{D}_{pq}^k,$$

$$\Delta_q^k = ik \int_{x_k}^{x_k + c_k} \sum_n \exp(ik\alpha_0 x) \tilde{F}_n(x) \sin\left(\frac{\pi q (x - x_k)}{c_k}\right) \mathrm{d}x, \tag{3.338}$$

$$\tilde{F}_n(x) = \sum_m R_n^m \tilde{F}_n^m(x).$$

Полученная система содержит бесконечное число линейных уравнений. Для нахождения дифракционных коэффициентов необходимо оставить конечное число линейных уравнений. Пусть $N = \max(k)$ — максимальное число периодов, $P = \max(n)$ — максимальное число порядков. В этом случае размерность системы уравнений равна $N \times P$.

Найдем коротковолновую асимптотику. Последнее уравнение в (3.337) описывает влияние различных периодов квазипериодической структуры. Анализируя (3.336), отметим, что если $\tilde{F}_n(x) = \sum R_n^m \tilde{F}_n^m(x) \to 0$, то система (3.336) распадается на N независимых систем уравнений, каждая из которых имеет размерность P. Каждая из этих подсистем совпадает с системой уравнений, полученной в случае решения задачи дифракции на идеально отражающей бинарной решетке с параметрами $d_m = x_{m+1} - x_m, c_m, h_m$ (см. раздел 3.1). Это позволяет уменьшить вычислительную ошибку при решении системы уравнений и получить удобные выражения для асимптотической оценки поля.

Для дальнейших рассуждений введем несколько вспомогательных непрерывных функций: $\nu(x)$, $R_n(x)$, $\alpha_n(x)$. Функция $\nu(x)$ определяет границы зон квазипериодической решетки $\nu(x_m) = m\Lambda$, где Λ — параметр квазипериодической структуры (при $\Lambda \to 0$ границы зон стремятся друг к другу), $R_n(x_m) = R_n^m$, $\alpha_n(x_m) = \alpha_n^m$. Перепишем $\tilde{F}_n(x)$ в следующем виде:

$$\tilde{F}_n(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \sum_m \left[\frac{\beta(\alpha) - \beta(\alpha_n^{m+1})}{\alpha - \alpha_n^{m+1}} R_n^{m+1} - \right.$$

$$\left. - \frac{\beta(\alpha) - \beta(\alpha_n^m)}{\alpha - \alpha_n^m} R_n^m \right] \exp(ik(\alpha - \alpha_0)(x - x_{m+1})) \, \mathrm{d}a. \tag{3.339}$$

Рассмотрим предельные случаи, когда влиянием соседних периодов можно пренебречь, т.е. $\tilde{F}_n(x) \to 0$.

1. Периодическая структура $\alpha_n^m = \alpha_n^{m+1}$, $R_n^m = R_n^{m+1}$. В этом случае из (3.337) следует, что $\tilde{F}_n(x)$ тождественно равна нулю.

2. Длина волны освещающего пучка стремится к нулю $\lambda \to 0$ ($k \to \infty$), Λ остается конечной. В этом случае (3.339) есть фурье-преобразование (по переменной α) от низкочастотной функции. При $k(x - x_{m+1}) \to \infty$ значение фурье-образа стремится к нулю.

3. $\Lambda \to 0$, $k \to \infty$, но $k\Lambda$ остается конечной величиной.

Можно показать, что

$$\sum_m \left[\frac{\beta(\alpha) - \beta(\alpha_n^{m+1})}{\alpha - \alpha_n^{m+1}} R_n^{m+1} - \frac{\beta(\alpha) - \beta(\alpha_n^m)}{\alpha - \alpha_n^m} R_n^m \right] \exp(ik(\alpha - \alpha_0)(x - x_{m+1})) \to 0. \tag{3.340}$$

Доказательство утверждения (3.340) сводится к доказательству предельной формулы

$$\sum_m \left[R\left(\frac{x_{m+2}-x_{m+1}}{\Lambda}\right) - R\left(\frac{x_{m+1}-x_m}{\Lambda}\right)\right] \exp\left(\frac{2\pi i p x_{m+1}}{\Lambda}\right) \to 0 \quad \text{при } \Lambda \to 0,$$

где $\nu(x_m) = \Lambda m$, $R(x)$ — произвольная функция. Строгое доказательство можно провести только в случае равноотстоящих зон (в этом случае выражение, стоящее в квадратных скобках в формуле (3.340) тождественно равно нулю), однако численный анализ показывает справедливость данного утверждения для широкого класса последовательностей x_m и функций $R(x)$.

Рассмотрим случай малых Λ и получим в этом случае удобное выражение для амплитуды вблизи поверхности квазипериодической структуры. Если функция $\varphi(x)$ монотонна и дифференцируема, то в пределах отрезка $[x_m x_{m+1}]$ справедлива линейная аппроксимация $\nu(x) = m\Lambda + (\Lambda(x - x_m)/d_m)$, где $d_m = x_{m+1} - x_m$. Тогда, подставляя эту аппроксимацию в (3.327), получим представление

$$\bar{W}(x) = \exp(ik\alpha_0 x) \sum_m \sum R_n^M \exp\left(ik\frac{\lambda n}{\Lambda}\nu(x)\right) \operatorname{rect}(x, x_m, x_{m+1}). \tag{3.341}$$

Нетрудно заметить, что знак суммы можно убрать, и поле на поверхности квазипериодической структуры представляется в виде

$$\bar{W}_0(x) = \exp(ik\alpha_0 x) \sum_n R_n(x) \exp\left(ik\frac{\lambda n}{\Lambda}\nu(x)\right). \tag{3.342}$$

$W(x, z)$ в полупространстве над поверхностью имеет вид

$$W(x,z) = \int_{-\infty}^{\infty} G(x - \bar{x}, z) W_0(\bar{x})\, \mathrm{d}\bar{x}, \tag{3.343}$$

$$G(x,z) = \frac{1}{2i}\frac{\partial H_0^1(k\sqrt{x^2+z^2})}{\partial z}. \tag{3.344}$$

Подставляя (3.342) в (3.343), получим выражение для поля везде при $z > 0$:

$$\bar{W}(x,z) = \sum_n \int_{-\infty}^{\infty} R_n(\bar{x}) G(x - \bar{x}, z) \exp\left(ik\alpha_0\bar{x} + ik\frac{\lambda n}{\Lambda}\nu(\bar{x})\right) \mathrm{d}\bar{x}. \tag{3.345}$$

С идейной точки зрения данный подход аналогичен методу нелинейного предыскажения [15]. Отличие состоит в том, что в работе [15] дифракционные коэффициенты от решетки были найдены, исходя из геометрооптической аппроксимации поля внутри штрихов, которая справедлива, если период решетки составляет несколько десятков длин волн. В электромагнитном подходе значения коэффициентов отражения в (3.345) будут зависеть от типа поляризации и относительной ориентации векторов поляризации и направления штрихов решетки. Существенное отличие выражения (3.345) от формул, полученных с помощью метода нелинейного предыскажения, состоит в наличии зависимости дифракционных коэффициентов от координат.

Пример 3.15. Рассмотрим вычисление поля от бинарной квазипериодической структуры, границы зон которой определяются формулой $\nu(x_m) = m\Lambda$, где $\nu(x)$ имеет вид фазовой функции линзы с фокусным расстоянием f:

$$\nu(x) = x\sin\omega + \sqrt{(x - f\sin q)^2 + f^2\cos^2 q}, \tag{3.346}$$

где ω — угол падения плоской волны на линзу, q — угол, под которым фокусирует линза. Для того, чтобы использовать (3.345) для нахождения поля над линзой необходимо определить функцию $R_n(x)$. Для ее определения необходимо использовать значения дифракционных коэффициентов $R_n^m = Q_n(\lambda/d_m)$, которые находятся из системы (3.332):

$$R_n(x) = Q_n(\lambda/d_m(x)), \tag{3.347}$$

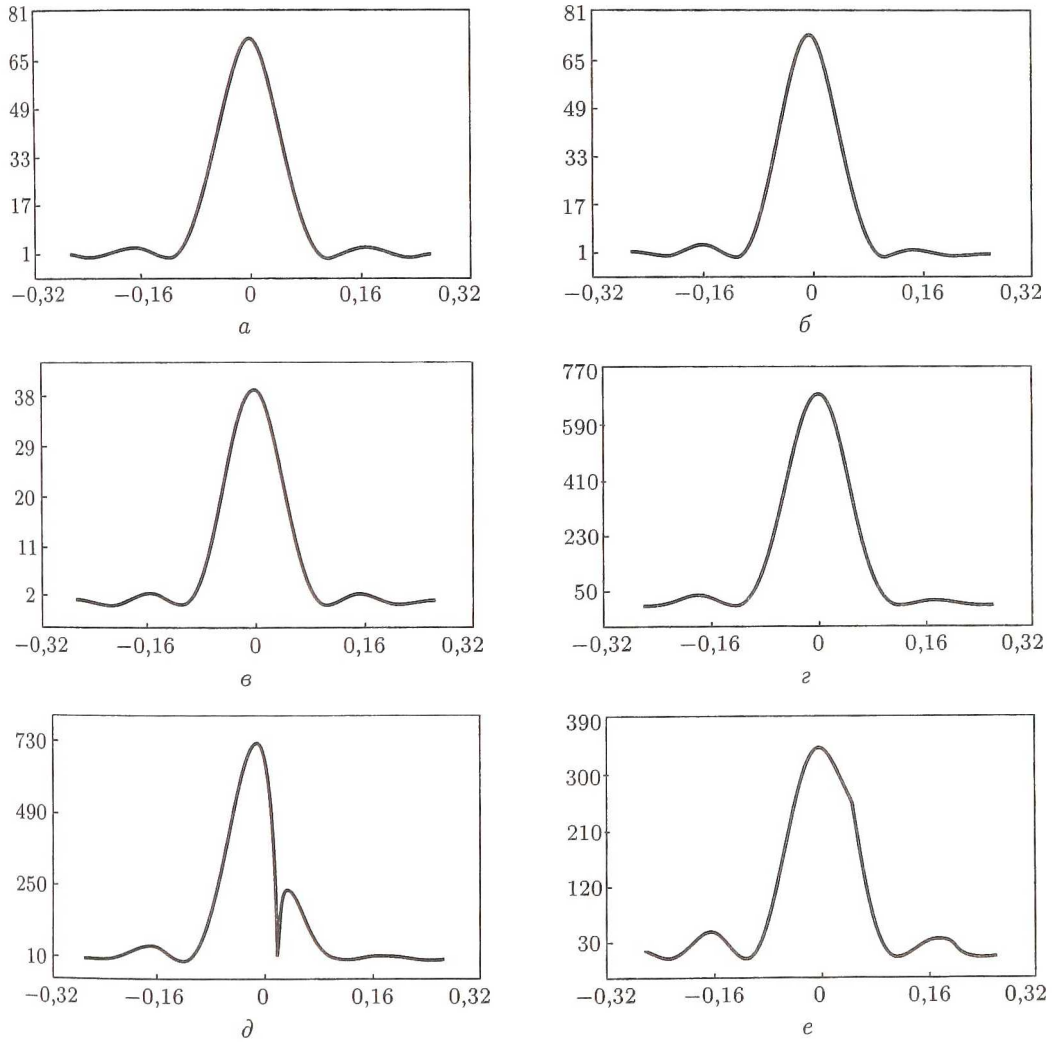


Рис. 3.25. Поле от внеосевого сегмента отражательной зонной пластинки (размер ДОЭ $D = 10$ мм, $\omega = 45°$): (*а*) $q = -25°$, $f = 100$ мм; (*б*) $q = -15°$, $f = 100$ мм; (*в*) $q = 0°$, $f = 100$ мм; (*г*) $q = -25°$, $f = 20$ мм; (*д*) $q = -15°$, $f = 20$ мм; (*е*) $q = 0°$, $f = 20$ мм

где

$$\frac{\Lambda}{d_m(x)} = \frac{(x - f\sin q)}{\sqrt{(x - f\sin q)^2 + f^2\cos^2 q}} + \sin\omega. \tag{3.348}$$

График функции

$$Q(\lambda/d_m) = \frac{\sqrt{1 - (\alpha_0 - \lambda_o/d_m)^2}}{\sqrt{1 - \alpha_o^2}} Q_{-1}(\lambda/d_m),$$
$$0{,}8 < \lambda/d_m < 1{,}8, \quad \alpha_0 = \sin\omega, \quad \omega = 45^\circ,$$

для ТМ-поляризации приведен в работе [19]. Из результатов, приведенных в [19],

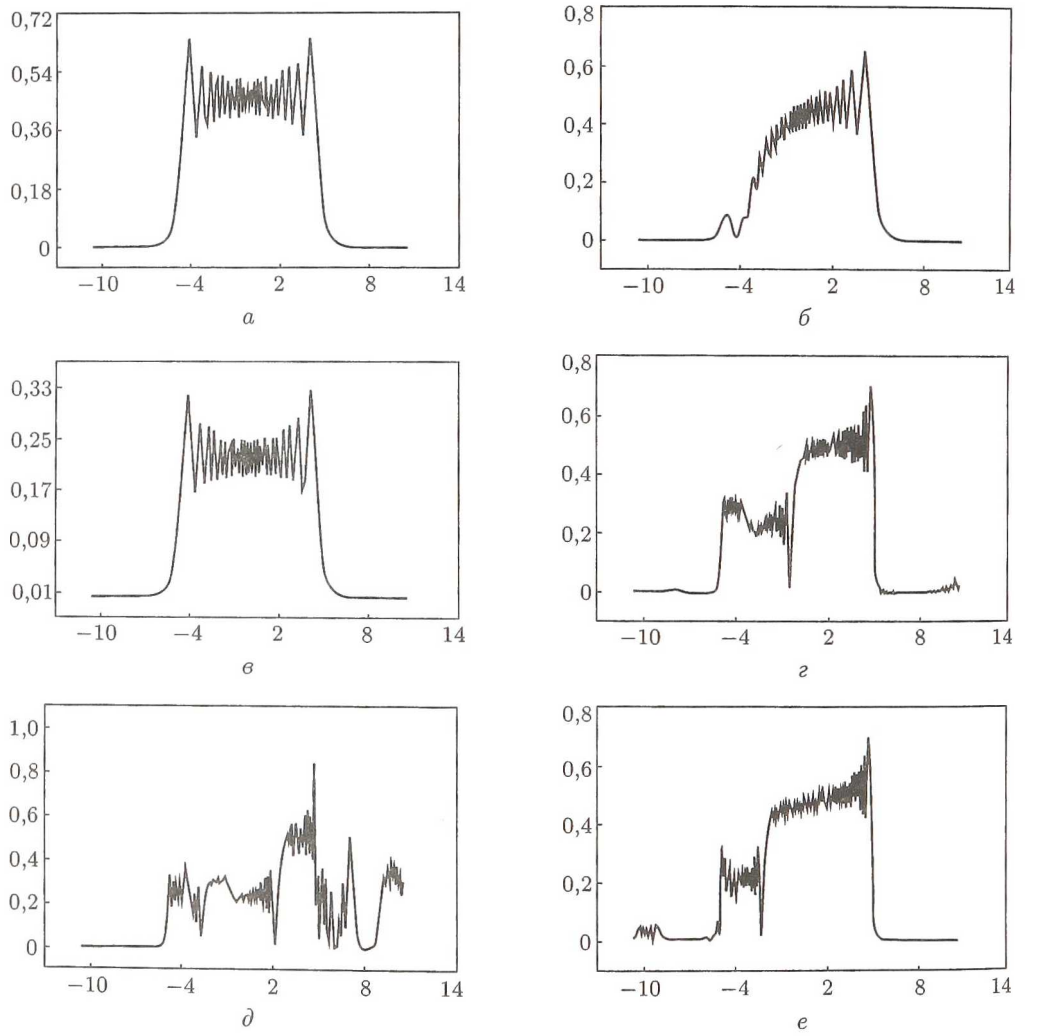


Рис. 3.26. Поле от фокусатора в отрезок для $D = 10$ мм, размер отрезка $L = 10$ мм, $\omega = 45°$: (*а*) $q = -25°$, $f = 100$ мм; (*б*) $q = -15°$, $f = 100$ мм; (*в*) $q = 0°$, $f = 100$ мм; (*г*) $q = -25°$, $f = 10$ мм; (*д*) $q = -15°$, $f = 10$ мм; (*е*) $q = 0°$, $f = 10$ мм

также видно, что $Q_n = 0$, если $n \neq -1, 0$. Функцию Q_0 можно найти из закона сохранения энергии: $Q_{-1}^2 + Q_0^2 = 1$.

Результаты вычисления распределения интенсивности приведены на рис. 3.25. Анализируя полученные результаты, можно заметить, что при коротких фокусных расстояниях (т.е. сравнимых с размером оптического элемента) наблюдается асимметрия формы фокального пятна и смещение относительно точки предполагаемого геометрического фокуса. Это смещение связано с тем, что фаза дифракционных коэффициентов на апертуре слабо, но меняется.

Пример 3.16. Более сложным примером отражающей зонной структуры является фокусатор в поперечный отрезок. На рис. 3.26 приведены примеры фокусировки в отрезок при различных параметрах.

3.7.2. Решение задачи дифракции одномерных ДОЭ в рамках электромагнитной теории

Рассмотрим отражение плоской электромагнитной волны от идеально проводящей квазипериодической структуры. Пусть плоская волна $\exp(ik(\alpha_0 x_\nu - \beta_0 x_N + \gamma x_\tau))$ падает на одномерный оптический элемент, границы зон x_m которого определяются из уравнения $\nu(x_m) = m\lambda$, $m = 1, 2 \ldots$. Чтобы упростить рассуждения рассмотрим одномерный случай. Введем систему координат x, x_N, x_τ и вектора (см. рис. 3.27)

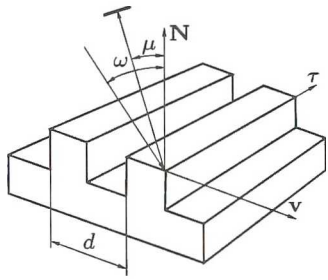


Рис. 3.27. Падение света под углом на бинарную отражательную квазипериодическую структуру

s — определяет направление распространения, $\mathbf{N}$ — вектор нормали к решетке (к плоскости дифракционной структуры), ν — вектор нормали к линиям решетки, τ — вектор, касательный к линиям решетки, $\tau = [\nu \times \mathbf{N}]$, $\nu = \nabla\varphi(\mathbf{u})/|\nabla\varphi(\mathbf{u})|$ — вектор, перпендикулярный к линиям решетки и лежащий в плоскости оптического элемента.

Введем также вектор l, лежащий в плоскости падения волны (т.е. в плоскости, проходящей через вектор нормали к плоскости оптического элемента и вектор падающей волны), и вектор $\mathbf{t}$, перпендикулярный плоскости падения:

$$\mathbf{t} = \frac{[\mathbf{s}, \mathbf{N}]}{|(\mathbf{s}, \mathbf{N})|}, \quad \mathrm{l} = \frac{[\mathbf{s}, \mathbf{t}]}{|(\mathbf{s}, \mathbf{t})|}. \tag{3.349}$$

Электрический вектор в падающей волне представим в виде суммы двух векторов, коллинеарных l и $\mathbf{t}$:

$$\mathbf{E} = \mathbf{E}_t + \mathbf{E}_l. \tag{3.350}$$

Пусть электрический вектор лежит в плоскости оптического элемента:

$$\mathbf{E} = \exp\left[ik\,(\mathbf{s}, \mathbf{x})\right]\mathbf{t}, \tag{3.351}$$

$$\mathbf{H} = \frac{1}{ik}\left[\mathrm{grad}\left(\exp\left(ik\,(\mathbf{s}, \mathbf{x})\right)\right), \mathbf{t}\right], \tag{3.352}$$

$$\mathbf{H} = \exp\left[ik\,(\mathbf{s}, \mathbf{x})\right][\mathbf{s}, \mathbf{t}] = \exp\left[ik\,(\mathbf{s}, \mathbf{x})\right]\mathrm{l}. \tag{3.353}$$

Составляющая электрического поля вдоль штрихов в падающей волне равна

$$E_\tau^0 = (\mathbf{t}, \boldsymbol{\tau}), \tag{3.354}$$

а составляющая магнитного поля вдоль штрихов в падающей волне равна

$$H_\tau^0 = (\mathbf{l}, \boldsymbol{\tau}). \tag{3.355}$$

Если вектор $\mathbf{H}$ магнитного поля параллелен плоскости оптического элемента, то выражения (3.354) и (3.355) примут вид

$$E_\tau^0 = -(\mathbf{l}, \boldsymbol{\tau}), \quad H_\tau^0 = (\mathbf{t}, \boldsymbol{\tau}). \tag{3.356}$$

Компоненты электромагнитного поля можно найти, используя выражения

$$\mathbf{E} = E_\nu \boldsymbol{\nu} + E_N \mathbf{N} + E_\tau \boldsymbol{\tau}, \quad \mathbf{H} = H_\nu \boldsymbol{\nu} + H_N \mathbf{N} + H_\tau \boldsymbol{\tau}, \tag{3.357}$$

$$\begin{aligned} E_\nu &= \frac{i}{k(1-\gamma^2)}(\partial_N H_\tau - \gamma \partial_\nu E_\tau), & H_\nu &= -\frac{i}{k(1-\gamma^2)}(\partial_N E_\tau - \gamma \partial_\nu H_\tau), \\ E_N &= -\frac{i}{k(1-\gamma^2)}(\partial_\nu H_\tau + \gamma \partial_N E_\tau), & H_N &= \frac{i}{k(1-\gamma^2)}(\partial_\nu E_\tau + \gamma \partial_N H_\tau). \end{aligned} \tag{3.358}$$

Рассмотрим зону с номером m, расположенную в интервале $[x_m, x_{m+1}]$. Будем полагать, что поле в этой зоне при $z = 0$ совпадает с полем дифракционной решетки, у которой микрорельеф на периоде совпадает с микрорельефом в данной зоне, и структура микрорельефа во всех зонах подобна с точностью до масштаба. В этом случае все характеристики выходного поля (коэффициенты дифракции) зависят только от соотношения λ/d. Используя формулы (3.358), получим выражения для отраженного от решетки поля:

$$\begin{aligned} E_\tau(\mathbf{x}) = {} & \left|E_\tau^0\right| \exp(ik\alpha_0 x + ik\gamma x_\tau) \times \\ & \times \sum_m \sum_n E_n^m \exp\left(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N\right) \operatorname{rect}(x, x_m, x_{m+1}), \end{aligned} \tag{3.359}$$

$$\begin{aligned} H_\tau(\mathbf{x}) = {} & \left|H_\tau^0\right| \exp(ik\alpha_0 x + ik\gamma x_\tau) \times \\ & \times \sum_m \sum_n H_n^m \exp\left(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N\right) \operatorname{rect}(x, x_m, x_{m+1}), \end{aligned} \tag{3.360}$$

$$\begin{aligned} E_\nu(\mathbf{x}) = {} & -\frac{1}{1-\gamma^2}\left\{ H_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \sum_m \sum_n \beta_n H_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + \right. \\ & + ik\beta_n x_N) rect(x, x_m, x_{m+1}) + \gamma E_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times \\ & \left. \times \sum_m \sum_n \alpha_n E_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N) \operatorname{rect}(x, x_m, x_{m+1}) \right\}, \end{aligned} \tag{3.361}$$

$$\begin{aligned} H_\nu(\mathbf{x}) = {} & \frac{1}{1-\gamma^2}\left\{ E_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times \right. \\ & \times \sum_m \sum_n \beta_n E_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N) \operatorname{rect}(x, x_m, x_{m+1}) - \\ & - \gamma H_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times \\ & \left. \times \sum_m \sum_n \alpha_n H_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N) \operatorname{rect}(x, x_m, x_{m+1}) \right\}, \end{aligned} \tag{3.362}$$

$$E_N(\mathbf{x}) = \frac{1}{1-\gamma^2}\Big\{H_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times$$
$$\times \sum_m \sum_n \alpha_n H_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N)\,\mathrm{rect}(x, x_m, x_{m+1}) -$$
$$- \gamma E_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times$$
$$\times \sum_m \sum_n \beta_n E_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N)\,\mathrm{rect}(x, x_m, x_{m+1})\Big\}, \quad (3.363)$$

$$H_N(\mathbf{x}) = -\frac{1}{1-\gamma^2}\Big\{E_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times$$
$$\times \sum_m \sum_n \alpha_n E_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N)\,\mathrm{rect}(x, x_m, x_{m+1}) +$$
$$+ \gamma H_\tau^0 \exp(ik\alpha_0 x + ik\gamma x_\tau) \times$$
$$\times \sum_m \sum_n \beta_n H_n^m \exp(ik\frac{\lambda n}{d_m}(x - x_m) + ik\beta_n x_N)\,\mathrm{rect}(x_m, x_{m+1})\Big\}, \quad (3.364)$$

где

$$\gamma(\mathbf{u}) = (\mathbf{s}, (\mathbf{u}))\,, \quad \alpha_0(\mathbf{u}) = (\mathbf{s}, \boldsymbol{\nu}(\mathbf{u}))\,, \quad \alpha_n(\mathbf{u}) = \alpha_0(\mathbf{u}) + \lambda n/d,$$
$$\beta_n(\mathbf{u}) = \sqrt{1 - \gamma^2(\mathbf{u}) - \alpha_n^2(\mathbf{u})}, \quad \lambda/d = |\nabla\varphi(\mathbf{u})|\,. \quad (3.365)$$

Учитывая, что

$$\nabla\nu = \lambda/d_m, \quad \nu(x) = m\lambda + \nabla\nu(x - x_m), \quad (3.366)$$

получим выражение для поля, формируемого оптическим элементом в плоскости $x_N = 0$:

$$E_\nu(\mathbf{x}) = -\frac{1}{(1-\gamma^2)}\sum_n \big\{H_\tau^0(\mathbf{x})H_n(\mathbf{x})\beta_n(\mathbf{x}) + \gamma(\mathbf{x})E_\tau^0(\mathbf{x})E_n(\mathbf{x})\alpha_n(\mathbf{x})\big\} \times$$
$$\times \exp(ikn\nu(\mathbf{x}))\exp(ik\,(\mathbf{s}, \mathbf{x})), \quad (3.367)$$

$$H_\nu(\mathbf{x}) = \frac{1}{(1-\gamma^2)}\sum_n \big\{E_\tau^0(\mathbf{x})E_n(\mathbf{x})\beta_n(\mathbf{x}) - \gamma(\mathbf{x})H_\tau^0(\mathbf{x})H_n(\mathbf{x})\alpha_n(\mathbf{x})\big\} \times$$
$$\times \exp(ikn\nu(\mathbf{x}))\exp(ik\,(\mathbf{s}, \mathbf{x})), \quad (3.368)$$

$$E_N(\mathbf{x}) = \frac{1}{(1-\gamma^2)}\sum_n \big\{H_\tau^0(\mathbf{x})H_n(\mathbf{x})\alpha_n(\mathbf{x}) - \gamma(\mathbf{x})E_\tau^0(\mathbf{x})E_n(\mathbf{x})\beta_n(\mathbf{x})\big\} \times$$
$$\times \exp(ikn\nu(\mathbf{x}))\exp(ik\mathbf{s}\mathbf{x}), \quad (3.369)$$

$$H_N(\mathbf{x}) = -\frac{1}{(1-\gamma^2)}\sum_n \big\{E_\tau^0(\mathbf{x})E_n(\mathbf{x})\alpha_n(\mathbf{x}) - \gamma(\mathbf{x})H_\tau^0(\mathbf{x})H_n(\mathbf{x})\beta_n(\mathbf{x})\big\} \times$$
$$\times \exp(ikn\nu(\mathbf{x}))\exp(ik\mathbf{s}\mathbf{x}), \quad (3.370)$$

$$E_\tau(\mathbf{x}) = \sum_n E_\tau^0(\mathbf{x}) E_n(\mathbf{x}) \exp(ikn\nu(\mathbf{x})) \exp(ik\mathbf{sx}), \tag{3.371}$$

$$H_\tau(\mathbf{x}) = \sum_n H_\tau^0(\mathbf{x}) H_n(\mathbf{x}) \exp(ikn\nu(\mathbf{x})) \exp(ik\mathbf{sx}). \tag{3.372}$$

Формулы (3.367)–(3.372) получены для дифракции на одномерном ДОЭ, но их вид остается неизменным и в случае двумерного ДОЭ.

Для нахождения электромагнитного поля во всем пространстве необходимо воспользоваться представлением поля в форме Кирхгофа–Котлера [8]. Вычисление интеграла Кирхгофа–Котлера представляет собой сложную вычислительную задачу, так как в общем случае нельзя воспользоваться алгоритмами быстрого преобразования Фурье. Авторами для вычисления данного интеграла был использован модифицированный метод Гопкинса. Метод Гопкинса рекомендуется использовать, когда функция $\varphi(x, y)$ является дифференцируемой. Модификация метода состояла в том, что при расчете методом Гопкинса отбрасывались члены в интегральной сумме, если аргумент в функции $\mathrm{sinc}(x)$ превышал 10. В этом случае в этих точках отпадала необходимость расчета дифракционных коэффициентов и произведений векторов, вычисление которых занимает большую часть времени.

Пример 3.17. Расчет дифракции от внеосевого сегмента отражательной зонной пластинки, границы зон которого описываются функцией $\nu(\mathbf{x}) = m\lambda$, где

$$\nu(\mathbf{x}) = x \sin\omega + \sqrt{(x - f\sin q)^2 + y^2 + f^2\cos^2 q}, \tag{3.373}$$

ω — угол падения плоской волны на линзу. Функция (3.373) является двумерным обобщением формулы (3.346). Внутри зоны микрорельеф имеет вид бинарной ди-

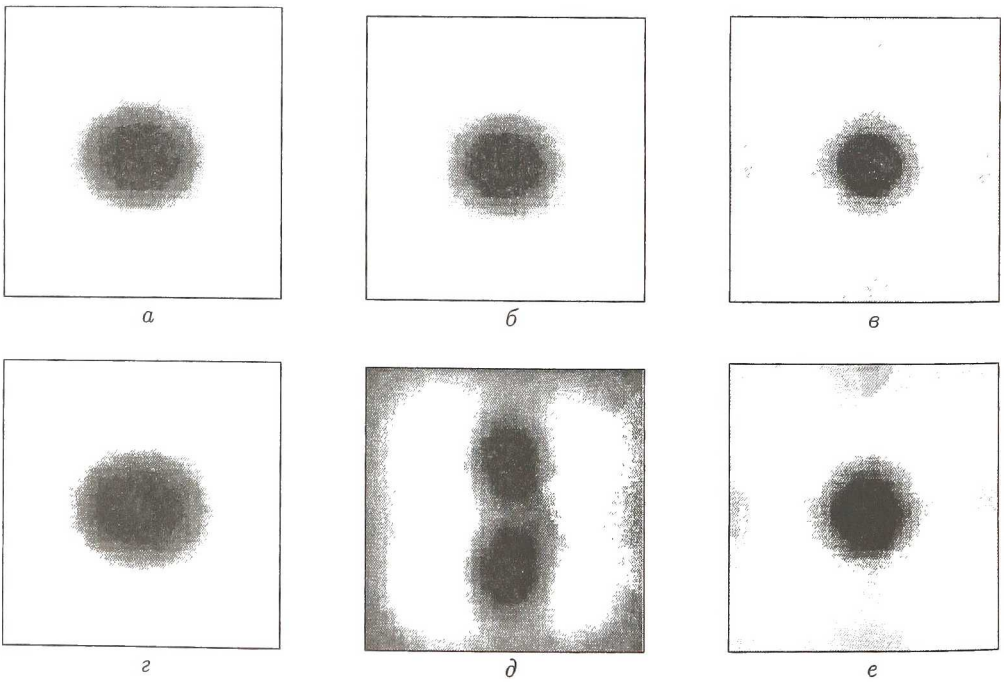


Рис. 3.28. Поле от внеосевого сегмента отражательной зонной пластинки для $D = 10 \times 10$ мм, $\omega = 45°$: (*а*) $q = -25°$, $f = 100$ мм; (*б*) $q = -15°$, $f = 100$ мм; (*в*) $q = -0°$, $f = 100$ мм; (*г*) $q = -25°$, $f = 20$ мм; (*д*) $q = -15°$, $f = 20$ мм; (*е*) $q = -0°$, $f = 20$ мм

фракционной решетки с одним штрихом. Нетрудно заметить, что данный элемент фокусирует излучение в точку с координатами $(x, y, z) = (f \sin q, 0, f \cos q)$.

Результаты вычисления I в фокальной плоскости линзы для случая, когда вектор магнитного поля лежит в плоскости ДОЭ, приведены на рис. 3.28, и показывают, что при коротких фокусных расстояниях форма фокального пятна асимметрична.

На рис. 3.28 введены следующие обозначения ω — угол падения света, q — угол между нормалью к пластине и направлением на фокус, D — размер зонной пластинки, f — фокусное расстояние.

3.8. Моделирование распространения электромагнитного излучения методом конечных разностей

В работе [24] изложен разностный метод решения уравнений Максвелла и получены результаты для оптических интегральных схем. В этом разделе численно проведены исследования градиентных сред и ДОЭ. Для этого применяется разностная схема «push–pull» и схема повышенного порядка точности по времени, которые позволяют исследовать распространение короткого импульса в волокне и ДОЭ, размеры которых сравнимы с длиной волны. Разностный метод реализован для решения уравнений Максвелла в левосторонней декартовой системе координат в случае волны типа H [25]:

$$\begin{gathered}\frac{\partial E_x}{\partial \tau} = \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial z} - \frac{\partial H_z}{\partial y}\right);\\ \frac{\partial H_y}{\partial \tau} = \frac{\partial E_x}{\partial z};\\ \frac{\partial H_z}{\partial \tau} = -\frac{\partial E_x}{\partial y}.\end{gathered} \tag{3.374}$$

где H_y, H_z, E_x — соответствующие проекции векторов напряженности магнитного и электрического полей, $\tau = ct$, c — скорость света в свободном пространстве, t — время, ε — диэлектрическая проницаемость. Среда представляет собой тонкий слой, расположенный в плоскости (y, z), а излучение направлено вдоль оси z.

Будем полагать, что на границе вычислительной области $E_x = 0$ (электрическая стенка) [25]. Это справедливо, например, для планарного волновода, заключенного в металлическую оболочку. Такой волновод будет соответствовать волноводу без оболочки, если:

— изгиб не превышает критического угла α, который соответствует переходу возбужденной моды высшего порядка в первую воздушную радиационную моду;

— неоднородностями показателя преломления, которые вызывают появление воздушных мод, можно пренебречь.

Будем решать систему (3.374) методом конечных разностей «push–pull». Для этого рассмотрим множество функций $E_x(z, y, \tau)$, $H_y(z, y, \tau)$, $H_z(z, y, \tau)$ в области $D = \{0 \leqslant z \leqslant L, -a/2 \leqslant y \leqslant a/2, 0 \leqslant \tau \leqslant T\}$, где L — длина волновода, a — ширина волновода, $T = ct_1$, t_1 — длительность эксперимента. Определим в области D сетку: $\Omega_{\Delta z \Delta y \Delta \tau} = \{(z_i, y_j, \tau_n) \in D\}$, где (z_i, y_j, τ_n) — узлы сетки, причем $0 \leqslant i \leqslant N_1 - 1$; $0 \leqslant j \leqslant N_2 - 1$; $0 \leqslant n \leqslant N_3 - 1$.

Шаги сетки определяются формулами

$$\Delta z = \frac{L}{N_1 - 1}; \quad \Delta y = \frac{a}{N_2 - 1}; \quad \Delta \tau = \frac{T}{N_3 - 1}.$$

В уравнениях Максвелла (3.374) произведем расщепление:

$$\begin{gathered}\frac{\partial \tilde{E}_x}{\partial \tau}=\frac{1}{n^2(y)}\frac{\partial H_y}{\partial z}; \qquad \frac{\partial E_x}{\partial \tau}=\frac{1}{n^2(y)}\frac{\partial H_z}{\partial y};\\ \frac{\partial H_y}{\partial \tau}=\frac{\partial \tilde{E}_x}{\partial z}; \qquad \frac{\partial H_z}{\partial \tau}=-\frac{\partial E_x}{\partial y},\end{gathered} \tag{3.375}$$

где $n^2(y)=\varepsilon$, $\tilde{E}_x$ — расщепленная компонента E_x .

Основываясь на общих принципах, изложенных в [26], построим неявную разностную схему расщепления:

$$\begin{gathered}\frac{\tilde{E}^{n+1}_{x_{i,j}}-E^{n}_{x_{i,j}}}{\Delta\tau}=\frac{1}{n^2(y)}\frac{H^{n+1}_{y_{i,j}}-H^{n+1}_{y_{i-1,j}}}{\Delta z};\\ \frac{E^{n+1}_{x_{i,j}}-\tilde{E}^{n+1}_{x_{i,j}}}{\Delta\tau}=-\frac{1}{n^2(y)}\frac{H^{n+1}_{z_{i,j}}-H^{n+1}_{z_{i,j-1}}}{\Delta y};\\ \frac{H^{n+1}_{y_{i,j}}-H^{n}_{y_{i,j}}}{\Delta\tau}=\frac{\tilde{E}^{n+1}_{x_{i+1,j}}-\tilde{E}^{n+1}_{x_{i,j}}}{\Delta z};\\ \frac{H^{n+1}_{z_{i,j}}-H^{n}_{z_{i,j}}}{\Delta\tau}=-\frac{E^{n+1}_{x_{i,j+1}}-E^{n+1}_{x_{i,j}}}{\Delta y}.\end{gathered} \tag{3.376}$$

Система (3.376) аппроксимирует (3.375) в узлах (z_i, y_j, τ_n) при значении индексов: $0\leqslant i\leqslant N_1-1$; $1\leqslant j\leqslant N_2-1$; $0\leqslant n\leqslant N_3-1$.

Подставив третье уравнение системы (3.376) в первое и четвертое во второе, получим разностные уравнения, которые решаются методом прогонки:

$$\begin{gathered}\frac{\Delta\tau}{n^2(y)\Delta z^2}\tilde{E}^{n+1}_{x_{i-1,j}}+\left(-\frac{1}{\Delta\tau}-2\frac{\Delta\tau}{n^2(y)\Delta z^2}\right)\tilde{E}^{n+1}_{x_{i,j}}+\frac{\Delta\tau}{n^2(y)\Delta z^2}\tilde{E}^{n+1}_{x_{i+1,j}}=\\ =\frac{1}{\Delta\tau}E^{n}_{x_{i,j}}-\frac{1}{n^2(y)\Delta z}H^{n}_{y_{i,j}}+\frac{1}{n^2(y)\Delta z}H^{n}_{y_{i-1,j}};\end{gathered}$$

$$\begin{gathered}\frac{\Delta\tau}{n^2(y)\Delta z^2}E^{n+1}_{x_{i,j-1}}+\left(-\frac{1}{\Delta\tau}-2\frac{\Delta\tau}{n^2(y)\Delta z^2}\right)E^{n+1}_{x_{i,j}}+\frac{\Delta\tau}{n^2(y)\Delta z^2}E^{n+1}_{x_{i,j+1}}=\\ =\frac{1}{\Delta\tau}\tilde{E}^{n+1}_{x_{i,j}}-\frac{1}{n^2(y)\Delta z}H^{n}_{z_{i,j}}+\frac{1}{n^2(y)\Delta z}H^{n}_{z_{i,j-1}}.\end{gathered}$$

В узлах (x_i, y_j, τ_k) при значениях индексов: $0\leqslant i\leqslant N_1-1$; $j=0$; $0\leqslant n\leqslant N_3-1$, построим альтернативную систему аппроксимирующих уравнений:

$$\begin{gathered}\frac{\tilde{E}^{n+1}_{x_{i,j}}-E^{n}_{x_{i,j}}}{\Delta\tau}=\frac{1}{n^2(y)}\frac{H^{n+1}_{y_{i+1,j}}-H^{n+1}_{y_{i,j}}}{\Delta z};\\ \frac{E^{n+1}_{x_{i,j}}-\tilde{E}^{n+1}_{x_{i,j}}}{\Delta\tau}=-\frac{1}{n^2(y)}\frac{H^{n+1}_{z_{i,j+1}}-H^{n+1}_{z_{i,j}}}{\Delta y};\\ \frac{H^{n+1}_{y_{i,j}}-H^{n}_{y_{i,j}}}{\Delta\tau}=\frac{\tilde{E}^{n+1}_{x_{i,j}}-\tilde{E}^{n+1}_{x_{i-1,j}}}{\Delta z};\\ \frac{H^{n+1}_{z_{i,j}}-H^{n}_{z_{i,j}}}{\Delta\tau}=-\frac{E^{n+1}_{x_{i,j}}-E^{n+1}_{x_{i,j-1}}}{\Delta y}.\end{gathered} \tag{3.377}$$

Подставив третье уравнение системы (3.377) в первое и четвертое во второе, получим разностные уравнения:

$$\frac{\Delta\tau}{n^2(y)\,\Delta z^2}\tilde{E}^{n+1}_{x_{i-1,j}} + \left(-\frac{1}{\Delta\tau} - 2\frac{\Delta\tau}{n^2(y)\,\Delta z^2}\right)\tilde{E}^{n+1}_{x_{i,j}} + \frac{\Delta\tau}{n^2(y)\,\Delta z^2}\tilde{E}^{n+1}_{x_{i+1,j}} = -\frac{1}{\Delta\tau}E^{n}_{x_{i,j}} - \frac{1}{n^2(y)\,\Delta z}H^{n}_{y_{i+1,j}} + \frac{1}{n^2(y)\,\Delta z}H^{n}_{y_{i,j}};$$

$$\frac{\Delta\tau}{n^2(y)\,\Delta y^2}E^{n+1}_{x_{i,j-1}} + \left(-\frac{1}{\Delta\tau} - 2\frac{\Delta\tau}{n^2(y)\,\Delta y^2}\right)E^{n+1}_{x_{i,j}} + \frac{\Delta\tau}{n^2(y)\,\Delta y^2}E^{n+1}_{x_{i,j+1}} = -\frac{1}{\Delta\tau}\tilde{E}^{n+1}_{x_{i,j}} - \frac{1}{n^2(y)\,\Delta y}H^{n}_{z_{i,j+1}} + \frac{1}{n^2(y)\,\Delta y}H^{n}_{z_{i,j}},$$

так же решаемые методом прогонки. Теперь, зная отсчеты компонента электрического поля $E^{n+1}_{x_{i,j}}$, определим компоненты магнитного поля,

$$\frac{H^{n+1}_{y_{i,j}} - H^{n}_{y_{i,j}}}{\Delta\tau} = \frac{E^{n+1}_{x_{i+1,j}} - E^{n+1}_{x_{i,j}}}{\Delta z},$$
$$\frac{H^{n+1}_{z_{i,j}} - H^{n}_{z_{i,j}}}{\Delta\tau} = -\frac{E^{n+1}_{x_{i,j+1}} - E^{n+1}_{x_{i,j}}}{\Delta y},$$

при значениях индексов: $0 \leqslant i \leqslant N_1 - 1$; $1 \leqslant j \leqslant N_2 - 1$; $0 \leqslant n \leqslant N_3 - 1$, и

$$\frac{H^{n+1}_{y_{i,j}} - H^{n}_{y_{i,j}}}{\Delta\tau} = \frac{E^{n+1}_{x_{i,j}} - E^{n+1}_{x_{i-1,j}}}{\Delta z},$$
$$\frac{H^{n+1}_{z_{i,j}} - H^{n}_{z_{i,j}}}{\Delta\tau} = -\frac{E^{n+1}_{x_{i,j}} - E^{n+1}_{x_{i,j-1}}}{\Delta y},$$

при значениях индексов: $0 \leqslant i \leqslant N_1 - 1$; $j = 0$; $0 \leqslant n \leqslant N_3 - 1$.

Возьмем в качестве граничного условия $E_x|_\Gamma = 0$ (область Γ — граница) и — торцевого условия (при $z = 0$) $E_x = A(y)\cos\left(\omega t + \frac{\pi}{2}\right)$, где A — действительная амплитуда волны в скалярном приближении.

В системе (3.376) аппроксимации пространственных производных в первых двух уравнениях являются левыми, а во вторых двух уравнениях — правыми. В системе (3.377) — наоборот. Поэтому (3.376) и (3.377) можно условно назвать схемой «push–pull». Она позволяет отказаться от использования краевых условий для магнитного поля, что существенно упрощает задачу. Схема «push–pull», (3.377) аппроксимируют краевую задачу (3.375) с погрешностью аппроксимации $O(\Delta\tau, \Delta z, \Delta y)$.

Построим примеры для тестирования полученной разностной схемы и представим практические приложения.

Пример 3.18. Для планарного вакуумного волновода с металлическими стенками модой будет являться следующая H-волна [30]:

$$\begin{aligned} E_x &= \cos(\gamma y)\cdot\cos\left(\sqrt{k^2-\gamma^2}z - \omega t\right), \\ H_y &= -\frac{1}{k}\sqrt{k^2-\gamma^2}\cdot\cos(\gamma y)\cdot\cos\left(\sqrt{k^2-\gamma^2}z - \omega t\right), \\ H_z &= -\frac{1}{k}\gamma\cdot\sin(\gamma y)\cdot\sin\left(\sqrt{k^2-\gamma^2}z - \omega t\right), \end{aligned} \tag{3.378}$$

где k — волновое число, $\gamma = 2\pi/a$. Зададим начальные условия: $E_x = H_y = H_z =$

$= 0$, приняв $t = 0$. На рис. 3.29 представлен результат, полученный численно для значений параметров: $t_1 = 1 \cdot 10^{-11}$ с, $\Delta t = 4{,}2 \cdot 10^{-15}$ с, $a = 60$ мкм, $L = 3$ мкм, $\Delta x = 2{,}36 \cdot 10^{-2}$ мкм, $\Delta y = 1{,}935$ мкм, $\lambda = 0{,}63$ мкм.

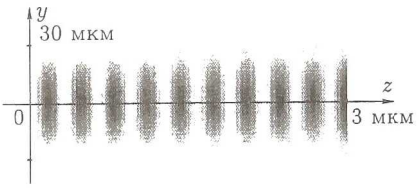


Рис. 3.29. Результат численного моделирования распространения моды (3.391) в вакуумном волноводе; изображен модуль проекции E_x

Визуально результат соответствует ожидаемому представлению электрической составляющей в виде косинусоиды. Численный анализ показывает наличие погрешности, зависящей от величин шагов дискретизации $\Delta\tau$, Δz, Δy.

Пример 3.19. Рассмотрим распространение той же волны [27] с теми же параметрами дискретизации в стеклянном волноводе с показателем преломления $n = 1{,}458$ полагая, что волна падает из вакуума на полубесконечный стеклянный волновод. Из рис. 3.30 можно заметить изменение формы импульса, состоящее в том, что импульсы сжимаются пропорционально n.

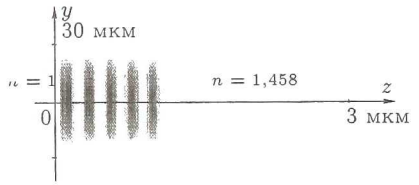


Рис. 3.30. Результат численного моделирования распространения моды (3.385) в стеклянном волноводе с показателем преломления $n = 1{,}458$: изображен модуль проекции E_x

Пример 3.20. Проследим распространение моды Гаусса–Лагерра [28],

$$\Psi_{10} = \frac{2}{\sigma\sqrt{2\pi}} L_1^0\left(\frac{2y^2}{\sigma^2}\right) e^{-y^2/\sigma^2}, \qquad (3.392)$$

в волноводе с градиентным профилем — показатель преломления $n^2(y) = n_1^2\left(1 - 2\Delta\,(y/d)^2\right)$, эффективный радиус моды $\sigma = 5{,}6$ мкм, $n_1 = 1{,}458$, $\Delta = 0{,}009$, $a = 60$ мкм, $L_1^0(x)$ — обобщенный полином Лагерра.

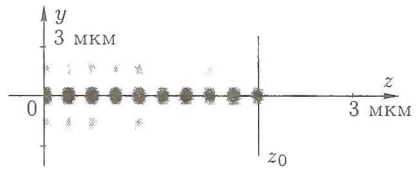


Рис. 3.31. Распространение моды Гаусса–Лагерра в волноводе с градиентным показателем преломления: изображен модуль проекции E_x

Шаги дискретизации были следующими: $\Delta t = 4{,}2 \cdot 10^{-15}$ с, $d = 60$ мкм, $L = 3$ мкм, $\Delta x = 1{,}003 \cdot 10^{-2}$ мкм, $\Delta y = 4{,}72 \cdot 10^{-1}$ мкм, $\lambda = 0{,}63$ мкм. $t = 1 \cdot 10^{-11}$. Результат моделирования представлен на рис. 3.31.

Рассмотрим распределение амплитуды на входном торце волновода ($z = 0$), представленное на рис. 3.32, и в волноводе ($z = z_0$), на рис. 3.33.

Из графиков видно, что мода сохранила характерные максимумы, хотя максимальное значение амплитуды несколько уменьшилось. Это связано с наличием погрешности дискретизации. Необходимо отметить, что с уменьшением шага дискретизации точность эксперимента повышается.

Пример 3.21. Рассмотрим прохождение H-волны ($\lambda = 1$ мкм) по планарному волноводу через цилиндрическую фокусирующую линзу. На рис. 3.34 представлен результат численного моделирования для распределение интенсивности в фокальной плоскости линзы, радиус кривизны которой 5 мкм, апертура 8 мкм, $f(\lambda) = 5$ мкм, $n = 2$, а волна содержит 8 цугов. Заметим, что волновой фронт после линзы является сферическим.

Пример 3.22. Заменим линзу ДОЭ с двумя градациями (бинарным), как показано на рис. 3.35. На рис. 3.37 представлены распределения проекций векторов Умова–Пойнтинга на ось z для четырех линз. Их можно сравнить с аналогичными [29], где линзы с соответствующими уровнями квантования имеют большую эффективность.

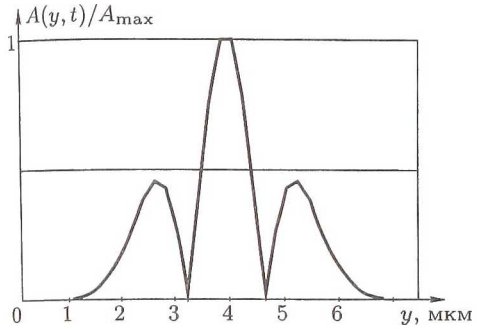


Рис. 3.32. Распределение амплитуды моды Гаусса–Лагерра на входном торце градиентного волновода

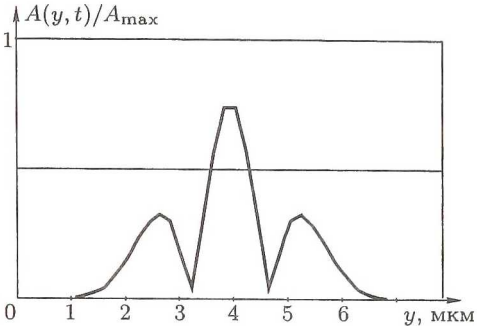


Рис. 3.33. Распределение амплитуды моды Гаусса–Лагерра в градиентном волноводе ($z = z_0$)

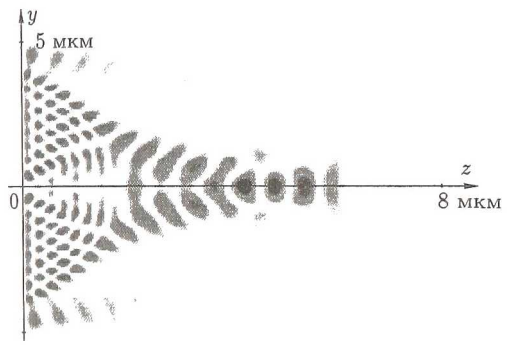


Рис. 3.34. Распределение интенсивности электрического поля в меридиональном сечении цилиндрической линзы

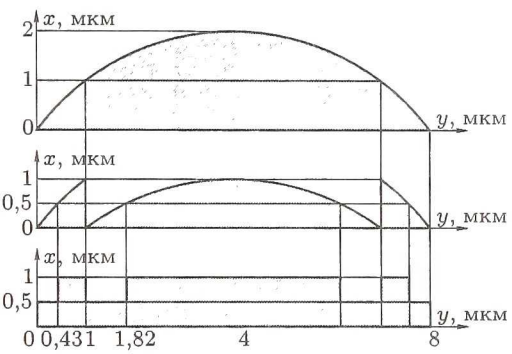


Рис. 3.35. Схема получения бинарной дифракционной цилиндрической линзы. Результаты численного моделирования представлены на рис. 3.36

Отличия вызваны тем, что в работе [29] исследуется тонкая линза в дифракционном приближении, которое не приемлемо в данном случае, когда излучение рассматривается непосредственно за оптическим элементом.

Основываясь на общих принципах построения схем Писмена–Рэкфорда, изложенных в [26], построим неявную разностную схему повышенного порядка точности по времени. Запишем систему разностных уравнений в виде

$$\begin{aligned}
&\frac{E^{n+0,5}_{x_{i,j}} - E^{n}_{x_{i,j}}}{0{,}5\Delta\tau} = \frac{1}{n^2(z,y)}\left(\frac{H^{n+0,5}_{y_{i+1,j}} - H^{n+0,5}_{y_{i,j}}}{\Delta z} - \frac{H^{n}_{z_{i,j}} - H^{n}_{z_{i,j-1}}}{\Delta y}\right);\\
&\frac{E^{n+1}_{x_{i,j}} - E^{n+0,5}_{x_{i,j}}}{0{,}5\Delta\tau} = \frac{1}{n^2(z,y)}\left(\frac{H^{n+0,5}_{y_{i+1,j}} - H^{n+0,5}_{y_{i,j}}}{\Delta z} - \frac{H^{n+1}_{z_{i,j}} - H^{n+1}_{z_{i,j-1}}}{\Delta y}\right);\\
&\frac{H^{n+0,5}_{y_{i,j}} - H^{n-0,5}_{y_{i,j}}}{\Delta\tau} = \frac{E^{n+0,5}_{x_{i,j}} - E^{n+0,5}_{x_{i-1,j}}}{\Delta z};\\
&\frac{H^{n+1}_{z_{i,j}} - H^{n}_{z_{i,j}}}{\Delta\tau} = -\frac{E^{n+1}_{x_{i,j+1}} - E^{n+1}_{x_{i,j}}}{\Delta y}.
\end{aligned} \tag{3.379}$$

Система (3.379) аппроксимирует (3.375) в узлах (z_i, y_j, τ_n) при значениях индексов: $0 \leqslant i \leqslant N_1 - 1$; $2 \leqslant j \leqslant N_2 - 2$; $0 \leqslant n \leqslant N_3 - 1$. Подставив третье уравнение

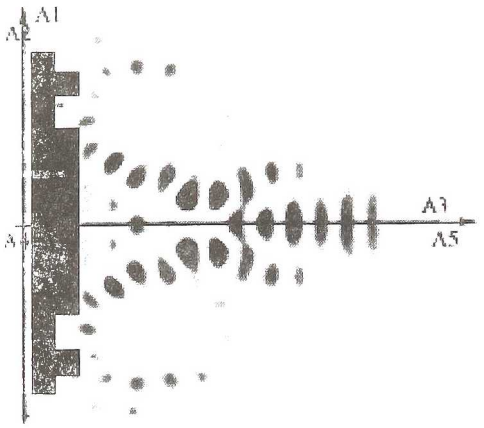


Рис. 3.36. Распределение интенсивности электрического поля в меридиональном сечении дифракционной линзы

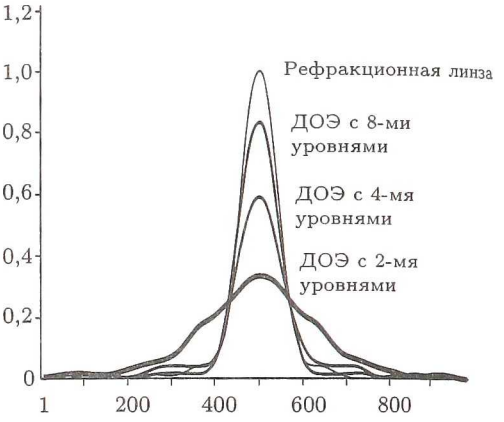


Рис. 3.37. Распределение проекции вектора Умова–Пойнтинга на оси z в фокальной плоскости

системы (3.379) в первое и четвертое во второе, получим разностные уравнения,

$$\frac{\Delta\tau}{n^2(z,y)\,\Delta z^2}E^{n+0,5}_{x_{i-1,j}} + \left(-\frac{2}{\Delta\tau} - 2\frac{\Delta\tau}{n^2(z,y)\,\Delta z^2}\right)E^{n+0,5}_{x_{i,j}} + \frac{\Delta\tau}{n^2(z,y)\,\Delta z^2}E^{n+0,5}_{x_{i+1,j}} =$$
$$= -\frac{2}{\Delta\tau}E^{n}_{x_{i,j}} - \frac{1}{n^2(z,y)\,\Delta z}H^{n-0,5}_{y_{+1i,j}} + \frac{1}{n^2(z,y)\,\Delta z}H^{n-0,5}_{y_{i,j}} +$$
$$+ \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j}} - \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j-1}};$$

$$\frac{\Delta\tau}{n^2(z,y)\,\Delta y^2}E^{n+1}_{x_{i,j-1}} + \left(-\frac{2}{\Delta\tau} - 2\frac{\Delta\tau}{n^2(z,y)\,\Delta y^2}\right)E^{n+1}_{x_{i,j}} + \frac{\Delta\tau}{n^2(z,y)\,\Delta y^2}E^{n+1}_{x_{i,j+1}} =$$
$$= -\frac{2}{\Delta\tau}E^{n+0,5}_{x_{i,j}} + \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j}} - \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j-1}} -$$
$$- \frac{1}{n^2(z,y)\,\Delta z}H^{n+0,5}_{y_{i+1,j}} + \frac{1}{n^2(z,y)\,\Delta z}H^{n+0,5}_{y_{i,j}},$$

которые решаются методом прогонки. В узлах (x_i, y_j, τ_k) при значениях индексов $0 \leqslant i \leqslant N_1 - 1$; $j = 1$; $0 \leqslant n \leqslant N_3 - 1$, построим альтернативную систему аппроксимирующих уравнений:

$$\begin{gathered}
\frac{E^{n+0,5}_{x_{i,j}} - E^{n}_{x_{i,j}}}{0{,}5\Delta\tau} = \frac{1}{n^2(z,y)}\left(\frac{H^{n+0,5}_{y_{i+1,j}} - H^{n+0,5}_{y_{i,j}}}{\Delta z} - \frac{H^{n}_{z_{i,j+1}} - H^{n}_{z_{i,j}}}{\Delta y}\right);\\
\frac{E^{n+1}_{x_{i,j}} - E^{n+0,5}_{x_{i,j}}}{0{,}5\Delta\tau} = \frac{1}{n^2(z,y)}\left(\frac{H^{n+0,5}_{y_{i+1,j}} - H^{n+0,5}_{y_{i,j}}}{\Delta z} - \frac{H^{n+1}_{z_{i,j+1}} - H^{n+1}_{z_{i,j}}}{\Delta y}\right);\\
\frac{H^{n+0,5}_{y_{i,j}} - H^{n-0,5}_{y_{i,j}}}{\Delta\tau} = \frac{E^{n+0,5}_{x_{i,j}} - E^{n+0,5}_{x_{i-1,j}}}{\Delta z};\\
\frac{H^{n+1}_{z_{i,j}} - H^{n}_{z_{i,j}}}{\Delta\tau} = -\frac{E^{n+1}_{x_{i,j}} - E^{n+1}_{x_{i,j-1}}}{\Delta y}.
\end{gathered} \tag{3.380}$$

Подставив третье уравнение системы (3.380) в первое и четвертое во второе, получим разностные уравнения,

$$\frac{\Delta\tau}{n^2(z,y)\,\Delta z^2}E^{n+0,5}_{x_{i-1,j}} + \left(-\frac{2}{\Delta\tau} - 2\frac{\Delta\tau}{n^2(z,y)\,\Delta z^2}\right)E^{n+0,5}_{x_{i,j}} + \frac{\Delta\tau}{n^2(z,y)\,\Delta z^2}E^{n+0,5}_{x_{i+1,j}} = -\frac{2}{\Delta\tau}E^{n}_{x_{i,j}} - \frac{1}{n^2(z,y)\,\Delta z}H^{n-0,5}_{y_{+1i,j}} + \frac{1}{n^2(z,y)\,\Delta z}H^{n-0,5}_{y_{i,j}} + \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j+1}} - \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j}};$$

$$\frac{\Delta\tau}{n^2(z,y)\,\Delta y^2}E^{n+1}_{x_{i,j-1}} + \left(-\frac{2}{\Delta\tau} - 2\frac{\Delta\tau}{n^2(z,y)\,\Delta y^2}\right)E^{n+1}_{x_{i,j}} + \frac{\Delta\tau}{n^2(z,y)\,\Delta y^2}E^{n+1}_{x_{i,j+1}} = -\frac{2}{\Delta\tau}E^{n+0,5}_{x_{i,j}} + \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j+1}} - \frac{1}{n^2(z,y)\,\Delta y}H^{n}_{z_{i,j}} - \frac{1}{n^2(z,y)\,\Delta z}H^{n+0,5}_{y_{i+1,j}} + \frac{1}{n^2(z,y)\,\Delta z}H^{n+0,5}_{y_{i,j}},$$

также решаемые методом прогонки. Теперь, зная отсчеты компонентов электрического поля $E^{n+1}_{x_{i,j}}$, определим компоненты магнитного поля:

$$\frac{H^{n+0,5}_{y_{i,j}} - H^{n}_{y_{i,j}}}{\Delta\tau} = \frac{E^{n+0,5}_{x_{i+1,j}} - E^{n+0,5}_{x_{i,j}}}{\Delta z};$$
$$\frac{H^{n+1}_{z_{i,j}} - H^{n}_{z_{i,j}}}{\Delta\tau} = -\frac{E^{n+1}_{x_{i,j+1}} - E^{n+1}_{x_{i,j}}}{\Delta y},$$

где $0 \leqslant i \leqslant N_1 - 1$; $2 \leqslant j \leqslant N_2 - 2$; $0 \leqslant n \leqslant N_3 - 1$, и

$$\frac{H^{n+0,5}_{y_{i,j}} - H^{n}_{y_{i,j}}}{\Delta\tau} = \frac{E^{n+0,5}_{x_{i,j}} - E^{n+0,5}_{x_{i-1,j}}}{\Delta z};$$
$$\frac{H^{n+1}_{z_{i,j}} - H^{n}_{z_{i,j}}}{\Delta\tau} = -\frac{E^{n+1}_{x_{i,j}} - E^{n+1}_{x_{i,j-1}}}{\Delta y},$$

где $0 \leqslant i \leqslant N_1 - 1$; $j = 1$; $0 \leqslant n \leqslant N_3 - 1$. Схема (3.379), (3.380) аппроксимирует краевую задачу (3.375) с погрешностью аппроксимации $O(\Delta\tau^2, \Delta z, \Delta y)$. Это достигается использованием обеих проекций поля H при определении $E^{n+0,5}$ и E^{n+1}, чего не было в схеме «push-pull» (3.376), (3.377).

Для сравнения эффективности предложенной схемы и схемы «push–pull» проведем оценку относительной погрешности решения задачи расчета прохождения моды по вакуумному волноводу с металлическими стенками. Погрешность равна

$$\delta = \max_{z,y}\left|\frac{E_x^{\text{числ.}}(z,y) - E_x^{\text{аналитич.}}(z,y)}{E_x^{\text{аналитич.}}(z,y)}\right|.$$

На рис. 3.38 представлены зависимости погрешностей от дискретизации по времени для схемы «push-pull» и схемы повышенного порядка точности по времени.

Численные эксперименты проводились при следующих значениях параметров: $t = 10 - 11$ с, $a = 3$ мкм, $L = 3$ мкм, $\Delta z = 0{,}0058$ мкм, $\Delta y = 0{,}023$ мкм, $\lambda = 1$ мкм, $\Delta t_1 = 0{,}333\cdot10^{-12}$, $0{,}166\cdot10^{-12}$, $0{,}111\cdot10^{-12}$, $0{,}0833\cdot10^{-12}$, $0{,}0666\cdot10^{-12}$, $0{,}0556\cdot10^{-12}$, $0{,}0476\cdot10^{-12}$, $0{,}0416\cdot10^{-12}$, $0{,}037\cdot10^{-12}$, $0{,}033\cdot10^{-12}$ с.



Рис. 3.38. Зависимость погрешностей от времени. Тонкая линия — погрешность решения с помощью схемы повышенной точности по времени. Толстая линия — погрешность решения с помощью схемы «push-pull»

Очевидно, при числе отсчетов по времени на заданном интервале более 20, схема повышенного порядка точности является предпочтительной. Отметим, что на рис. 3.38 погрешность решения с помощью схемы повышенного порядка точности по времени имеет квадратичный характер зависимости от дискретизации по времени.

Для уравнений Максвелла в трехмерной правой декартовой системе координат в СИ запишем следующую схему [30]:

$$E_x^{n+1} = \frac{h_t}{\varepsilon}\left(\frac{H_{z_{j+1}}^{n+1} - H_z^{n+1}}{h_y} - \frac{H_{y_{m+1}}^{n+1} - H_y^{n+1}}{h_z}\right) + E_x;$$

$$H_z^{n+1} = -\frac{h_t}{\mu}\left(\frac{E_{y_{i+1}}^{n+1} - E_y^{n+1}}{h_x} - \frac{E_x^{n+1} - E_{x_{j-1}}^{n+1}}{h_y}\right) + H_z;$$

$$E_y^{n+1} = \frac{h_t}{\varepsilon}\left(\frac{H_x^{n+1} - H_{x_{m-1}}^{n+1}}{h_z} - \frac{H_z - H_{z_{i-1}}}{h_x}\right) + E_y;$$

$$H_x^{n+1} = -\frac{h_t}{\mu}\left(\frac{E_{z_{j+1}}^{n+1} - E_z^{n+1}}{h_y} - \frac{E_{y_{m+1}}^{n+1} - E_y^{n+1}}{h_z}\right) + H_x;$$

$$E_z^{n+1} = \frac{h_t}{\varepsilon}\left(\frac{H_{y_{i+1}}^{n+1} - H_y^{n+1}}{h_x} - \frac{H_x - H_{x_{j-1}}}{h_y}\right) + E_z;$$

$$H_y^{n+1} = -\frac{h_t}{\mu}\left(\frac{E_x - E_{x_{m-1}}}{h_z} - \frac{E_z^{n+1} - E_{z_{i-1}}^{n+1}}{h_x}\right) + H_y,$$

где H_x, H_y, H_z, E_x, E_y, E_z — соответствующие проекции векторов напряженности магнитного и электрического полей, h_t, h_x, h_y, h_z — шаги дискретизации на сетке $\Omega_{ht,hx,hy,hz} = \{(x_i, y_j, z_m, t_n) \in D\}$ в области $D = \{0 < x < L_x, 0 < y < L_y, 0 < z < L_z, 0 < t < L_t\}$, где i, j, m, n — узлы сетки, причем $0 < i < N_x - 2$, $0 < j < N_y - 2$, $0 < m < N_z - 2$, $0 < n < N_t - 2$.

Краевые условия выбираются для каждой проекции из следующих соображений: если заключить рассматриваемую область в идеально проводящую оболочку (электрическая стенка), то на ее поверхности тангенциальные проекции вектора электрического поля и нормальные проекции вектора магнитного равны нулю (краевое условие первого рода), а у нормальных проекций вектора электрического поля и тангенциальных — магнитного, производная по нормали к стенке равна нулю (краевое условие второго рода).

Полученная схема аппроксимирует краевую задачу с погрешностью аппроксимации $O(h_t, h_x, h_y, h_z)$.

В качестве теста брались различные электромагнитные моды полого волновода с идеально проводящими стенками. Моды представляют собой гармонические функции, удовлетворяющие граничным условиям. На рис. 3.39 представлено аналитическое и разностное решение в случае E_{11} волны [25].

К недостаткам рассмотренных методов относятся невысокий порядок аппроксимации разностными схемами исходных задач и значительные затраты вычислительных ресурсов, что, в частности, не позволяет использовать методы для расчета полей в волноводах большой протяженности.

3.9. Анализ прохождения электромагнитного импульса через антиотражающую структуру

Актуальной проблемой синтеза ДОЭ являются потери энергии, связанные с френелевским отражением, что особенно существенно при фокусировке излучения мощных технологических лазеров. В главе 4 описана технология синтеза фокусирующих ДОЭ для мощных лазеров ИК диапазона на основе селективной абляции алмазных пластин. Относительно высокий показатель преломления алмаза приводит к тому, что из-за потерь на отражение пропускание алмазной пластины на длине волны 10,6 мкм не превышает 71%. Поэтому особую актуальность приобретает задача просветления алмазных оптических элементов т.е. снижения потерь на френелевское отражение. В работах [31, 32] на результатах натурного эксперимента продемонстрирована эффективность антиотражающих субволновых периодических микроструктур на поверхности алмазной пленки, созданных методом селективного лазерного травления. Численный анализ работы антиотражающих структур в [31, 32] проводился с помощью теории эффективных сред [33]. В рамках этой теории субволновая дифракционная структура рассматривается как градиентная среда с плавно меняющимся поперечным градиентным распределением показателя преломления (рис. 3.40). Однако, приближение теории эффективных сред не учитывает реальных электромагнитных эффектов. В данном разделе излагается метод, позволяющий моделировать прохождение лазерного излучения через антиотражающую структуру в рамках электромагнитной теории.

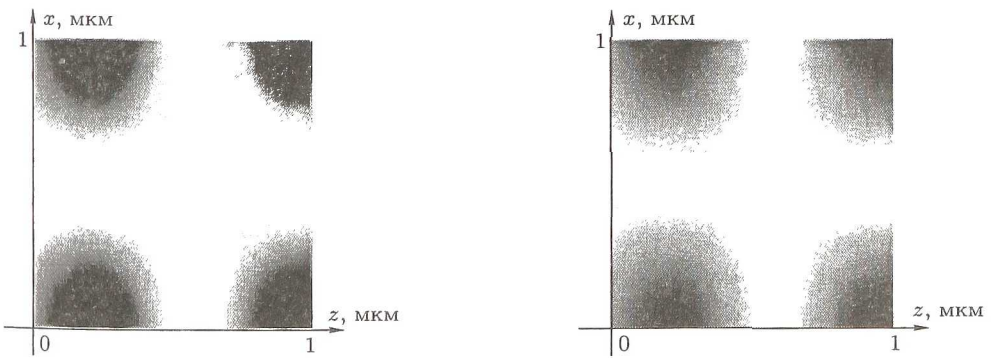


Рис. 3.39. Идеальное и расчетное модовые распределения для волны E_{11} в сечении XZ при $y = Ly/2$

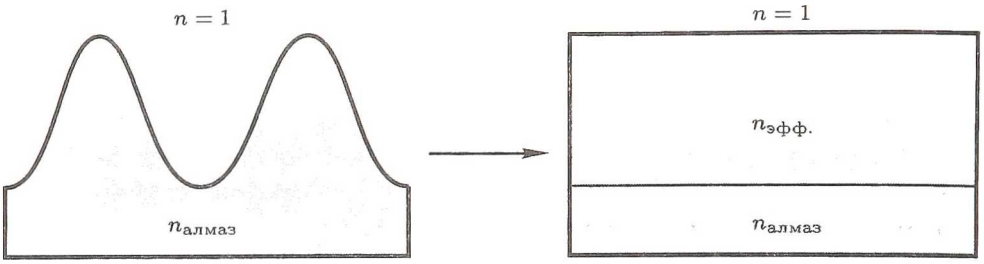


Рис. 3.40. Эквивалентная среда для рельефных субволновых решеток на поверхности алмазной пленки

В работе [30] представлена разностная схема для уравнений Максвелла, записанных в системе СИ в декартовой трехмерной системе координат. Так как дифракционный рельеф наносится в виде полос, то выбор декартовой системы координат с осью X, направленной параллельно полосам, позволяет построить следующие двумерные схемы для ТЕ-поляризации [25]:

$$\begin{aligned}
H_z^{n+1} &= \frac{h_t}{\mu_0\mu}\frac{E_x^{n+1}-E_{x_{j-1}}^{n+1}}{h_y}+H_z;\\
E_x^{n+1} &= \frac{h_t}{\varepsilon_0\varepsilon}\left(\frac{H_{z_{j+1}}-H_z}{h_y}-\frac{H_{y_{m+1}}^{n+1}-H_y^{n+1}}{h_z}\right)+E_x;\\
H_y^{n+1} &= -\frac{h_t}{\mu_0\mu}\frac{E_x^{n+1}-E_{x_{m-1}}^{n+1}}{h_z}+H_y,
\end{aligned}\qquad (3.381)$$

$$\begin{aligned}
H_y^{n+1} &= -\frac{h_t}{\mu_0\mu}\frac{E_x^{n+1}-E_{x_{m-1}}^{n+1}}{h_z}+H_y;\\
E_x^{n+1} &= \frac{h_t}{\varepsilon_0\varepsilon}\left(\frac{H_{z_{j+1}}^{n+1}-H_z^{n+1}}{h_y}-\frac{H_{y_{m+1}}-H_y}{h_z}\right)+E_x;\\
H_z^{n+1} &= \frac{h_t}{\mu_0\mu}\frac{E_x^{n+1}-E_{x_{j-1}}^{n+1}}{h_y}+H_z,
\end{aligned}\qquad (3.382)$$

где H_y, H_z, E_x — соответствующие проекции векторов напряженности электрического и магнитного полей, ε_0, μ_0 — электрическая и магнитная постоянные, ε, μ — диэлектрическая и магнитная проницаемости среды, h_t, h_y, h_z — шаги дискретизации на сетке $\Omega_{ht,hy,hz} = \{(y_j, z_m, t_n) \in D\}$ в области $D = \{0 < y < L_y, 0 < z < L_z, 0 < t < L_t\}$, где j, m, n узлы сетки, причем $0 < j < N_y - 2$, $0 < k < N_z - 2$, $0 < n < N_t - 2$. Краевые условия для E_x и H_y — первого рода, для H_z — второго рода. Считается, что излучение распространяется вдоль оси Z. Полученные схемы аппроксимируют краевую задачу с погрешностью аппроксимации $O(h_t, h_y, h_z)$.

Система (3.381) решается следующим образом: третье уравнение системы подставляется во второе, и получившееся уравнение приводится к трехдиагональному виду,

$$\begin{aligned}
E_{x_{m-1}}^{n+1}\left(-\frac{h_t^2}{\mu_0\mu\varepsilon_0\varepsilon h_z^2}\right)+E_x^{n+1}\left(1+2\frac{h_t^2}{\mu_0\mu\varepsilon_0\varepsilon h_z^2}\right)+E_{x_{m+1}}^{n+1}\left(-\frac{h_t^2}{\mu_0\mu\varepsilon_0\varepsilon h_z^2}\right)=\\
=\frac{h_t}{\varepsilon_0\varepsilon h_y}\left(H_{z_{j+1}}-H_z\right)+\frac{h_t}{\varepsilon_0\varepsilon h_z}\left(H_y-H_{y_{m+1}}\right)+E_x,
\end{aligned}\qquad (3.383)$$

Таблица 3.5. Проявление антиотражающего эффекта при различных формах антиотражающего покрытия

Номер численного эксперимента	Форма антиотражающего покрытия	Доля отраженной энергии (%)
1	без антиотражающего покрытия	19,24
2	Треугольник с базисом 4 мкм и высотой 2,5 мкм	19,24
3	Треугольник с базисом 3 мкм и высотой 2,4 мкм	9,8
4	Треугольник с базисом 3 мкм и высотой 3 мкм	9,8
5	Треугольник с базисом 3 мкм и высотой 2 мкм	11,44
6	Треугольник с базисом 3 мкм и высотой 1,8 мкм	12,39
7	Треугольник с базисом 2 мкм и высотой 2,5 мкм	9,37
8	Треугольник с базисом 1,5 мкм и высотой 2,5 мкм	5,41

и решается стандартной трехдиагональной прогонкой. Найдя проекцию E_x, подставим ее в первое и третье уравнение системы (3.381) и найдем H_y и H_z.

Аналогично решается (3.382), где трехдиагональное уравнение будет иметь вид

$$E^{n+1}_{x_{j-1}}\left(-\frac{h_t^2}{\mu_0\mu\varepsilon_0\varepsilon h_y^2}\right)+E^{n+1}_{x}\left(1+2\frac{h_t^2}{\mu_0\mu\varepsilon_0\varepsilon h_y^2}\right)+E^{n+1}_{x_{j+1}}\left(-\frac{h_t^2}{\mu_0\mu\varepsilon_0\varepsilon h_y^2}\right)=$$
$$=\frac{h_t}{\varepsilon_0\varepsilon h_y}\left(H_{z_{j+1}}-H_z\right)+\frac{h_t}{\varepsilon_0\varepsilon h_z}\left(H_y-H_{y_{m+1}}\right)+E_x. \quad (3.384)$$

Из (3.383) и (3.384) видно, что если число отсчетов вдоль области вычислительного эксперимента значительно превышает число отсчетов поперек области, то предпочтительна схема (3.382), так как можно обойтись меньшей оперативной памятью компьютера. Иначе предпочтительна схема (3.381).

Отметим отличия схем (3.381) (3.382) от схем (3.376), (3.377) и (3.379), (3.380) предыдущего пункта, которые являются неявными. В схеме (3.381) значение продольной составляющей магнитного поля вычисляется явным образом, а в схеме (3.382) явным образом вычисляется значение поперечной составляющей магнитного поля. Но разностные схемы (3.381), (3.382) не являются явными, так как остальные компоненты электромагнитного поля в них вычисляются неявно. Такой подход приносит сокращение объема вычислений на 50 процентов, однако грозит появлением неустойчивости.

Пример 3.23. В табл. 3.5 приведены результаты численных экспериментов, которые состояли в формировании волн типа H_{01} ($\lambda = 10{,}6$ мкм, с цугом в одну длину волны), падающих на границу раздела пластинка–воздух, и моделировании дальнейшего распространения прошедших и отраженных волн. Диэлектрическая проницаемость пластинки $\varepsilon = 5{,}76$. Параметры схемы (3.381) выбирались следующими: $L_y = 500$ мкм, $L_z = 180$ мкм, $L_t = 8{,}1\cdot 10^{-14}$ с, $h_y = 1/3$ мкм, $h_z = 0{,}2$ мкм, $h_t = 3{,}5\cdot 10^{-18}$ с. Распределение энергии электромагнитного поля вдоль оси X определялась как [25] $W = (1/2)\int\limits_D \left(\varepsilon_0\varepsilon\left|E\right|^2+\mu_0\mu\left|H\right|^2\right)dD$, где $D=\{0\leqslant y\leqslant L_y, 0\leqslant z\leqslant L_z\}$.

По результатам экспериментов 3–6 можно предположить наличие оптимальной высоты равнобедренного треугольника, являющегося периодом антиотражающей структуры. Структура с треугольниками, у которых высота меньшей оптимальной,

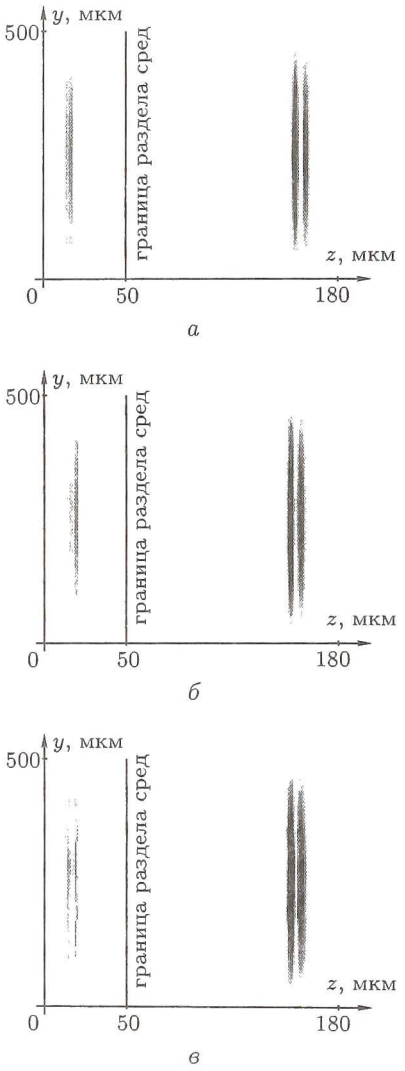


Рис. 3.41. Распределение модуля амплитуды проекции E_z электромагнитного поля в прошедшей и отраженной волне в первом (*а*), в третьем (*б*), восьмом (*в*) экспериментах, соответственно

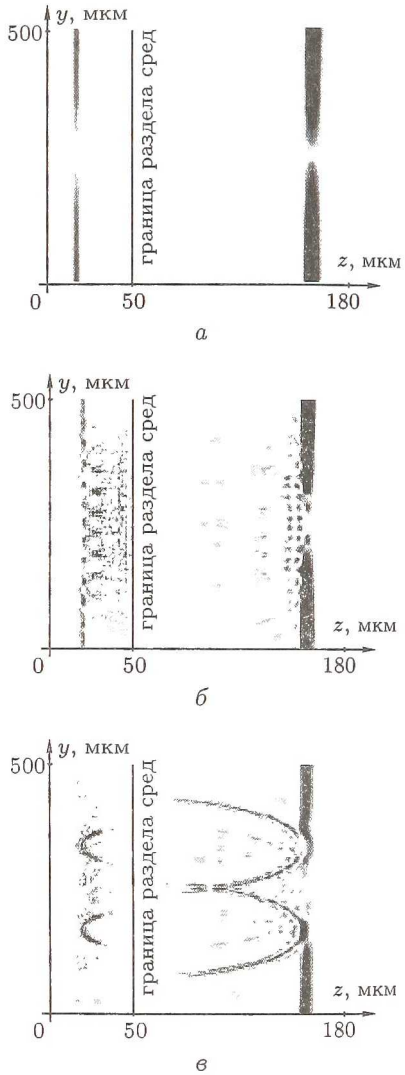


Рис. 3.42. Распределение модуля амплитуды проекции H_z электромагнитного поля в прошедшей и отраженной волне в первом (*а*), третьем (*б*), восьмом (*в*) экспериментах, соответственно

обладает меньшим антиотражающим эффектом (сравним эксперимент 3 с экспериментами 5 и 6). А при высоте треугольника больше оптимальной, антиотражающий эффект не увеличивается (сравним эксперимент 3 с экспериментом 4).

На рис. 3.41*а,б,в* представлено разделение падающей волны (эксперименты 1, 3, 8) на прошедшую через границу раздела алмазная пластинка — воздух и отраженную от этой границы.

На антиотражающий эффект также влияет размер базиса треугольника. С его уменьшением увеличивается доля прошедшей через треугольник энергии (сравним эксперименты 2, 3, 7, 8 и рис. 3.41). Заметно уменьшение доли отраженной энергии на рис. 3.41*а,б* по сравнению с рис. 3.41*в*.

Рис. 3.42*а,б,в* представляют модовый состав электромагнитной волны по разные стороны от границы раздела сред. Модовый состав лучше отслеживается на продольной проекции магнитной составляющей поля. Это происходит из-за того, что моды низких порядков переносят подавляющую часть энергии волны, а продольная проекция магнитной составляющей поля вносит малый вклад в общую энергию и в ней моды низких порядков незначительно превалируют над остальными модами.

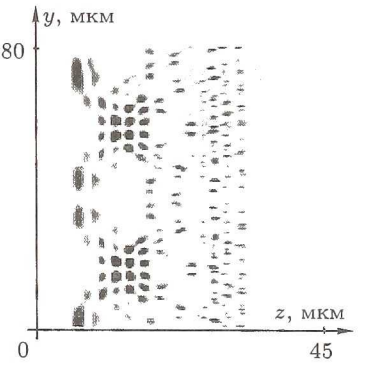


Рис. 3.43. Фрагмент распределения модуля амплитуды проекции H_z электромагнитного поля в отраженной волне в третьем эксперименте

Появление мод высших порядков на рис. 3.42 (в отличии от рис. 3.41, где показана мода низшего порядка) обусловлено непрерывностью электромагнитного поля на антиотражающем рельефе. Каждый фрагмент рельефа, содержащий от одного периода и более, формирует набор мод порядка $L_z a/d$, где d — период антиотражающего рельефа, a — число периодов во фрагменте. Более подробно модовая структура представлена на рис. 3.43, 3.44.

Таким образом, показана возможность использования разностного решения уравнений Максвелла для анализа работы антиотражающих субволновых структур. Результаты численного моделирования находятся в хорошем согласии с результатами натурных исследований, опубликованными в [31, 32], и результатами численного анализа с помощью теории эффективных сред второго порядка [33]. Разностное решение уравнений Максвелла впервые позволило провести моделирование антиотражающих субволновых структур и оценить их эффективность, а также провести анализ тонкой структуры отраженной электромагнитной волны.

3.10. Заключение

Электромагнитный анализ работы простых дифракционных решеток с прямоугольным и треугольным профилем штриха выявляет ряд существенных эффектов, не описываемых в рамках скалярной теории. Это показывает актуальность и необходимость использования точных процедур расчета при анализе ДОЭ с зонами, сравнимыми с длиной волны.

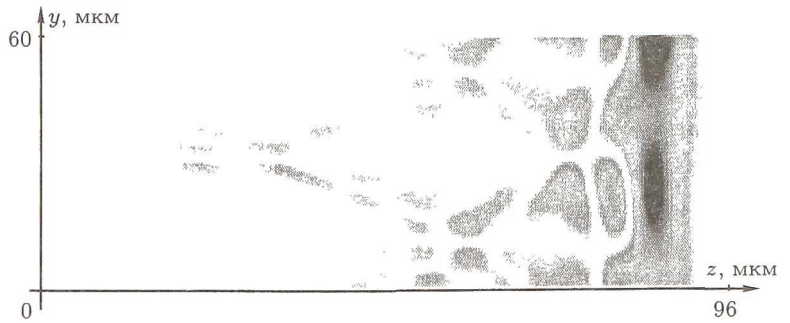


Рис. 3.44. Фрагмент распределения модуля амплитуды проекции H_z электромагнитного поля в прошедшей волне в третьем эксперименте

На основе решений прямой задачи разработаны электромагнитные градиентные методы решения обратной задачи, состоящей в расчете профиля дифракционной решетки из условия формирования заданной интенсивности порядков. Приведенные исследования характеристик работы дифракционных решеток, рассчитанных в скалярном приближении, показывают актуальность точных процедур синтеза, а результаты расчетов профилей решеток подтверждают работоспособность и эффективность разработанных градиентных процедур.

Показана работоспособность градиентной процедуры в задаче расчета однопорядковых бинарных диэлектрических решеток, концентрирующих излучение в 1-м или -1-м порядке. Однопорядковые решетки имеют ряд интересных применений. В частности, преобразование рельефа фокусатора по закону бинарной однопорядковой решетки позволяет достичь энергетической эффективности более 90% при бинарном профиле.

Получен ряд решений задачи дифракции в общем трехмерном случае. Приведены интегральные представления для оператора распространения электромагнитного поля, сводящие решение прямой задачи к четырем преобразованиям Фурье. При этом доказанное свойство унитарности оператора распространения позволяет обобщить «скалярные» итерационные алгоритмы синтеза фазовых волновых полей на случай точного электромагнитного расчета. На основе указанных интегральных представлений, разработан градиентный метод решения обратной задачи восстановления волновых полей.

Рассмотрен асимптотический расчет характеристик поля от ДОЭ, основанный на локальной аппроксимации структуры рельефа элемента двумерной дифракционной решеткой. В рамках данного подхода проведен расчет характеристик поля от отражающей линзы и фокусатора в отрезок. Полученные асимптотики позволяют оценить ряд новых электромагнитных эффектов, не учитываемых в «скалярном случае». Это является актуальным при анализе работы апериодических ДОЭ с большой апертурой, когда точный расчет невозможен по причинам вычислительной сложности. Перспективным является совместное использование полученных асимптотик и градиентных алгоритмов расчета однопорядковых решеток. Это позволяет решать обратную задачу синтеза фокусатора с учетом электромагнитных эффектов.

Разработаны разностные схемы для численного решения системы уравнений Максвелла для всех типов электромагнитных волн в области, ограниченной электрической или магнитной стенкой. Это позволяет моделировать прохождение коротких импульсов света через оптические элементы и волноводы с произвольным профилем показателя преломления. Разностные схемы были использованы для моделирования прохождения H-волны (TE-поляризация) через цилиндрические линзы, в том числе с антиотражающим покрытием.

Литература к главе 3

1. Виноградова М. Б., Руденко О. В., Виноградова А. П., Сухоруков А. П. *Теория волн.* (М.: Наука, 1979)
2. Ильинский А. С., Кравцов В. В., Свешников А. Г. *Математические модели электродинамики* (М.: Высшая школа, 1991)
3. Зверев В. А. *Радиооптика.* (М.: Советское Радио, 1975)
4. Kok Y. L., Gallagher N. C. *J. Opt. Soc. Am. A.* **5** (1) 65 (1988)
5. *Electromagnetic Theory of Gratings: Topics in current physics* (22, Ed. by R.Petit, N.Y.: Springer-Verlag, 1980)
6. Kok Y. L. *Opt. Eng.* **33** (11) 3604 (1994)

7. Moharam M. G., Gaylord T. K. *J. Opt. Soc. Am. A.* **3** (11) 1780 (1986)
8. Борн М., Вольф Э. *Основы Оптики* (Pergamon Press, 1968)
9. Гантмахер Ф. Р. *Теория матриц.* (М.: Наука, 1988)
10. Doskolovich L. L., Kharitonov S. I., Petrova O. I., Soifer V. A. *Opt. And Lasers in Eng.* **29** 249 (1998)
11. Zhou C., Liu L. *Appl.Opt.* **34** 5961 (1995)
12. Досколович Л. Л. *Сб. Компьютерная оптика,* **1** 8 82 (1998)
13. Doskolovich L. L., Soifer V. A., Alessandretti G., Perlo P., Repetto P. *Pure & Appl.Opt.* **3** 921 (1994)
14. Doskolovich L. L., Kazanskiy N. L., Perlo P., Repetto P., Soifer V. A. *Int. Jour. of Optoelectronics* **10** 243 (1995)
15. Soifer V., Kotlyar V., Doskolovich L. *Iterative Methods for Diffractive Optical Elements Computation* (London, Taylor & Francis Ltd, 1997)
16. Doskolovich L. L., Kharitonov S. I., Petrova O. I., Soifer V. A. *Proc. SPIE, Opt. Inf. Science & Tech.* **3348** 13 (1997)
17. *Appl. Opt.* **34** (14) (1995)
18. *J. Opt. Soc. Am. A.* **12** (5) (1995)
19. Veldkamp W. B., Swanson G. C., Shaver D. C. *Optics Communic.* **5** (6) 353 (1984)
20. Федорюк М. В. *Асимптотики: Интегралы и ряды.* (М.: Наука, 1987).
21. Воронцов М. А., Корябин А. В., Шмальгаузен В. И. *Управляемые оптические системы.* (М.:Наука, 1988)
22. Воронцов М. А., Матвеев А. Н., Сивоконь В. П. *Сб. Компьютерная оптика* **1** 74 (1987)
23. Воронцов М. А., Шмальгаузен В. И. *Принципы адаптивной оптики* (М.: Наука, 1985)
24. Chu S. T., Huang W. P., Chaudhuri S. K. *Computer Physics Communications* **68** (1991)
25. Никольский В. В., Никольская Т. И. *Электродинамика и распространение радиоволн* (М.: Наука, 1989)
26. Самарский А. А. *Теория разностных схем* (М.: Наука, 1989)
27. Солимено С., Крозиньяни Б., Ди Порто П. *Дифракция и волноводное распространение оптического излучения* (М.: Мир, 1984)
28. Soifer V. A., Golub M. A. *Laser Beam Mode Selection by Computer Generated Holograms* (CRC Press, Boca Raton, Florida, 1994)
29. Stamnes J. J., Gravelsaeter T., Bentsen O. *Acoustical Imaging* **10** 587 (1982)
30. Головашкин Д. Л., в сб. *Труды девятой межвузовской конференции* (Самара, 1999)
31. Kononenko T. V., Kononenko V. V., Konov V. I., Pimenov S. M., Garnov S. V., Tischenko A. V., Prokhorov A. M., Khomich A. V. *Applied Physics A* **68**(1) 99 (1999)
32. Кононенко В. В., Кононенко Т. В., Конов В. И., Пименов С. М., Гарнов С. В., Тищенко А. В., Прохоров А. М., Хомич А. В. *Квантовая электроника* **26**(2) 158 (1999).
33. Daniel H., Raguin G. *Applied Optics* **32**(7) 1154 (1993).

ГЛАВА 4
ТЕХНОЛОГИЯ СОЗДАНИЯ ДОЭ

4.1. Типы фазовых микрорельефов и способы их получения

ДОЭ представляет собой зонную пластинку с дифракционным микрорельефом, отображающим фазовую функцию. На этапе проектирования ДОЭ рассчитывается его фазовая функция $\varphi(x, y)$ одним из методов, изложенных в предыдущих главах. Полагаем, что зонная пластинка расположена в плоскости (x, y). Высота микрорельефа $h(x, y)$ связана с фазовой функцией пропускающего ДОЭ $\varphi(x, y)$ соотношением [1]

$$h(x,y) = \frac{\lambda}{2\pi\left(\sqrt{n^2 - \sin^2 a} - \cos a\right)} \bmod {}_{2\pi m}(\varphi(x\cos a, y)), \tag{4.1}$$

а в случае отражения световой волны от ДОЭ — соотношением

$$h(x,y) = \frac{\lambda}{4\pi\cos\alpha} \bmod {}_{2\pi m}(\varphi(x\cos\alpha, y)), \tag{4.2}$$

где $\mathrm{mod}_{2\pi m}(t)$ — функция, равная наименьшему положительному остатку от деления t на $2\pi m$; $m = 1, 2, 3, \ldots$ — целое число, λ — длина волны, n — показатель преломления материала подложки, образующей микрорельеф, α — угол между оптической осью светового пучка и нормалью к плоскости оптического элемента (x, y) (рис. 4.1).

Микрорельеф, соответствующий формулам (4.1) или (4.2) с непрерывной фазовой функции $\varphi(x, y)$, называют кусочно-непрерывным или микрорельеф с «блеском» [2]. Согласно формулам (4.1)–(4.2), высота рельефа принимает максимальные значения при значениях фазовой функции $\varphi(x, y)$ кратных $2\pi m$. Максимум

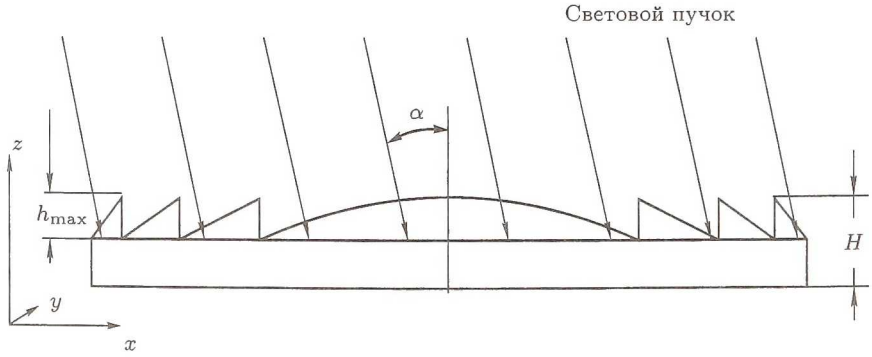


Рис. 4.1. Оптическая схема работы ДОЭ

высоты микрорельефа для пропускающего ДОЭ имеет вид

$$h_{\max} = \frac{m\lambda}{\sqrt{n^2 - \sin^2\alpha} - \cos\alpha}, \tag{4.3}$$

а для отражающего ДОЭ

$$h_{\max} = \frac{m\lambda}{2\cos\alpha}. \tag{4.4}$$

При этом общая толщина H оптического элемента обычно существенно больше $h_{\max}$ с учетом толщины подложки (рис. 4.1).

С точки зрения практической реализации ДОЭ наибольшее распространение получили два типа фазового микрорельефа (рис. 4.2): ступенчатый (бинарный и многоуровневый) и кусочно-непрерывный.

Ступенчатый многоуровневый профиль микрорельефа (см. рис. 1.19 и 1.20) соответствует квантованной по M уровням фазовой функции $\hat{\varphi}(x,y)$ с шагом $\Delta\varphi$, определяемым из соображений технологичности и эффективности ДОЭ. Основной характеристикой ступенчатого рельефа ДОЭ с точки зрения технологии является степень его соответствия профилю фазовой функции $\hat{\varphi}(x,y)$. Наличие технологических ошибок приводит к снижению дифракционной эффективности ДОЭ.

Высота бинарного микрорельефа принимает два значения в пределах одной зоны и определяется соотношением

$$h(x,y) = \begin{cases} 0, & \varphi(x,y) < \pi, \\ h_{\max}, & \varphi(x,y) \geqslant \pi, \end{cases} \tag{4.5}$$

где $\varphi(x,y)$ приведена к интервалу $[0, 2\pi)$.

Получение значений бинарной фазы в формуле (4.5) на основе исходной фазовой функции $\varphi(x,y)$ рассмотрено, например, в разделе 1.2 (см. рис. 1.18).

Задача выбора типа микрорельефа определяется не только набором доступных разработчику технологий формирования дифракционного микрорельефа, но и требованиями к качеству работы и эффективности ДОЭ, предъявляемыми конкретной оптической схемой и функциональным назначением оптического элемента. Программные комплексы, созданные для проектирования ДОЭ, обязательно содержат средства, позволяющие оценить снижение качества и энергетической эффективности работы ДОЭ при переходе от кусочно-непрерывного к многоуровневому или бинарному микрорельефу [3]. Разрабатываются специальные физико-математические методы для повышения эффективности бинарных и многоуровневых ДОЭ [4].

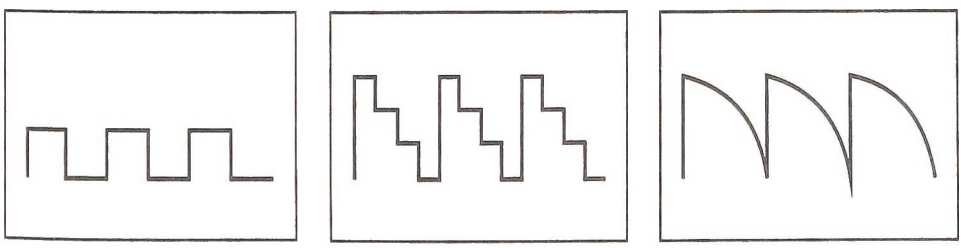


Рис. 4.2. Профили микрорельефов ДОЭ: (*а*) — бинарный; (*б*) — многоуровневый; (*в*) — кусочно-непрерывный

При любых методах расчета и оптимизации фазовой функции ДОЭ, качество работы оптического элемента, в конечном счете, определяется точностью формирования дифракционного микрорельефа. Существуют различные технологии, реализующие запись фазовой функции на оптическую среду. Например, делительные машины позволяют непосредственно получать дифракционные решетки (в т.ч. с блеском) и, поскольку в данном случае границы зон представляют собой прямые линии, управление делительной машиной не требует реализации сложных алгоритмов. Существует также возможность непосредственного перевода на физическую среду радиально-симметричных и даже более сложных фазовых функций с помощью прецизионных механических станков с программным управлением [5, 6]. Такая возможность реально используется для изготовления ДОЭ субмиллиметрового и дальнего ИК диапазонов [7, 8]. Весьма перспективна технология фотолитографии — многоступенчатое химическое травление стеклянной подложки [9]. Для создания оптических элементов видимого диапазона применяют халькогенидные стеклообразные полупроводники (ХСП) [10], материалы с изменяющимся показателем преломления [11]. Перспективными технологиями получения кусочно-непрерывного микрорельефа являются методы, основанные на использовании жидких (и сухих) фотополимеризующихся композиций (ЖФПК) [12, 13].

Особое внимание специалистов в области дифракционной оптики привлекают технологии формирования микрорельефа, основанные на достижениях и оборудовании микроэлектроники [14]. Определяющими в таком выборе являются следующие основные факторы:

— близость геометрических размеров ДОЭ и микросхем;

— близость требуемых характеристик пространственного разрешения;

— высокий уровень автоматизации микроэлектронного оборудования и возможность реализации на нем сложных топологий зон;

— возможность получения многих градаций фазового микрорельефа;

— наличие множества технологических участков микроэлектроники и возможность аренды оборудования;

— высокая степень надежности технологических автоматов и повторяемость результатов;

— возможность использования материалов, имеющихся в микроэлектронике (субстраты, резисты и т.п.);

— наличие автоматизированных средств контроля результата.

Высокая степень автоматизации технологических операций микроэлектроники позволяет говорить об автоматизированном проектировании дифракционного микрорельефа оптических элементов. Такое проектирование включает в себя не только выбор специальных методов, учитывающих технологические ограничения, но и возможность итерационной оптимизации важнейших параметров ДОЭ на основе анализа результатов контроля качества получаемого микрорельефа и эффективности работы ДОЭ в оптической схеме. При этом выбор размера, формы, используемого материала, пространственного разрешения проектируемого микрорельефа (и соответственно выбор технологий) определяется как задачами, стоящими перед ДОЭ, так и оптическим диапазоном работы ДОЭ. Например, возможные границы пространственного разрешения дифракционной решетки могут лежать в пределах от 0,25 мм$^{-1}$ в ИК и субмиллиметровом диапазонах и до 1200 мм$^{-1}$ в УФ.

Модельное представление большинства технологий формирования микрорельефа ДОЭ дано на рис. 4.3. Изменения приведенной к рабочему интервалу $[0, 2\pi m)$ фазовой функции ДОЭ представляются изменениями пропускания маски (шаблона), генерируемой на одном из автоматизированных прецизионных устройств. При

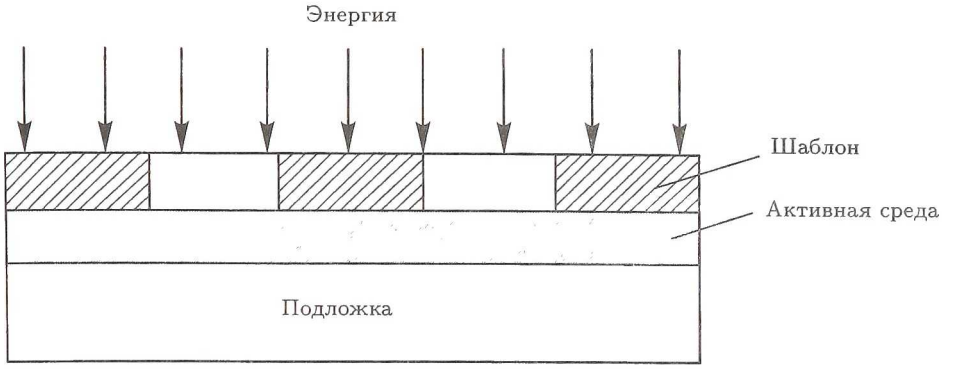


Рис. 4.3. Представление процесса формирования микрорельефа

целенаправленном подводе энергии через шаблон (засветке фотошаблона) активная среда изменяет свои свойства в засвеченных участках и формирует структуру, которая либо сама является дифракционной решеткой (как в слоях ЖФПК), либо способствует созданию ее путем травления подложки.

В настоящее время для формирования дифракционного микрорельефа апробировано и отлажено множество технологий с использованием самых разнообразных (физически, химически, механически) активных сред: фокусируемых электронных и ионных пучков, газов, кислот, резистов, полимерных композиций, алмазных резцов. При этом различные технологии требуют создания различных шаблонов: наборов бинарных масок для фотолитографии, полутоновых фотошаблонов для ЖФПК и отбеливания желатины, тонко-мембранной пленки для рентгеновской литографии или маски-трафарета для использования ионно-лучевой литографии. При использовании полутоновых шаблонов и соответствующих технологий формирование кусочно-непрерывного микрорельефа происходит в один этап (методы формирования кусочно-непрерывного рельефа рассмотрены в п. 4.4). При использовании набора бинарных шаблонов и соответствующих «бинарно-активных» сред для получения многоуровневого рельефа, процесс, показанный на рис. 4.3, приходится повторять несколько раз, перебирая по очереди все шаблоны из набора. Для ряда современных технологий (электронная литография, станки с ЧПУ) изготовления физически существующего шаблона (или набора шаблонов) не происходит: в этих случаях формирование микрорельефа происходит на основе рассчитанного виртуального шаблона.

Важнейшими характеристиками фотошаблона (как реального, так и виртуального) являются параметры имеющихся на нем линий, особенно минимальное значение ширины линии и максимальное значение кривизны линии. Легче всего описать эти параметры для линий, имеющихся на бинарных шаблонах. При изготовлении бинарного ДОЭ фотошаблон является амплитудной зонной пластинкой, разные виды которых подробно описаны в разделе 1.1, там же на простейших примерах охарактеризовано понятие зоны оптического элемента. Линии на фотошаблоне повторяют формы зон, занимая для бинарного ДОЭ около половины площади зоны. В частности, для изготовления фазовой дифракционной решетки в качестве фотошаблона используется амплитудная дифракционная решетка, темные полосы на которой образуют линии фотошаблона (см. рис. 1.1). В случае изготовления более сложных типов ДОЭ линии на фотошаблоне могут быть разной толщины и иметь значительную кривизну (см., например, рис. 1.5 и 1.9).

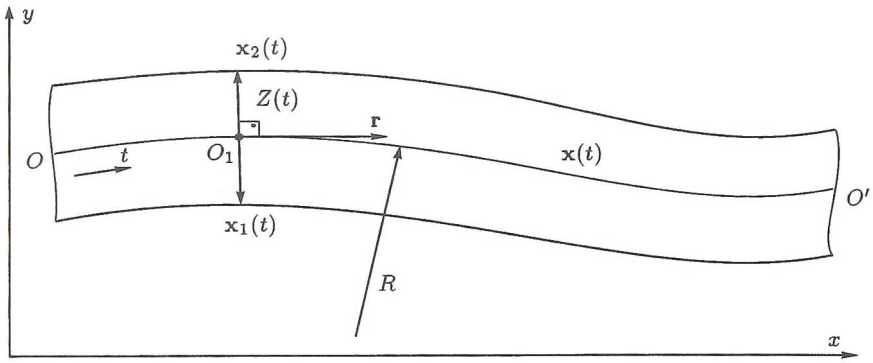


Рис. 4.4. Математическое описание характеристик линии

Математическая модель линии (рис. 4.4) определяется следующими функциями:

$$x = x(t), \quad y = y(t), \quad Z = Z(t), \quad 0 < t < L, \tag{4.6}$$

где $x(t)$, $y(t)$ — дифференцируемые функции, параметрически описывающие кривую центров линии, которую будем называть «трассой»; $Z(t)$ — функция ширины линии (суммарное расстояние от трассы до границ линии, отсчитываемое по перпендикуляру к трассе); t — расстояние от начала линии, измеренное по трассе; L — длина линии (трассы).

Из (4.6) однозначно определяются границы линии, кривизна и радиус кривизны линии в каждой точке трассы.

Кривизна линии в точке t

$$K(t) = \frac{|y'' \cdot x' - x'' \cdot y'|}{\left|(x')^2 + (y')^2\right|}, \tag{4.7}$$

где

$$x' = \frac{dx(t)}{dt}, \quad x'' = \frac{d^2x(t)}{dt^2}.$$

Параметры кривизны и минимальные размеры ширины линий шаблонов определяют выбор вида технологии и набор необходимого технологического оборудования. Очевидно, что чем меньше рабочая длина волны и чем выше оптическая сила ДОЭ, тем меньше ширина линий шаблонов и, соответственно, выше требования к разрешению и точности используемых технологий и технологического оборудования.

Выбор технологии и устройств формирования фотошаблонов предполагает и соответствующую организацию данных расчета оптического элемента. Организация данных будет отличаться для устройств с круговым или построчным сканированием и устройств, имеющих векторный характер прорисовки (фотошаблона или микрорельефа). Полутоновой или бинарный характер используемого автомата также определяет форму передачи данных: либо последовательный набор бинарных массивов (для бинарных устройств), либо один массив с соответствующим объемом байт на отсчет (обычно хватает одного байта), обеспечивающий необходимое число градаций фазы оптического элемента для устройств формирования полутоновых шаблонов.

4.2. Изготовление ДОЭ методами фотолитографии

Литографией называется процесс воспроизведения конфигурации и взаимного расположения элементов изделия на подложке (субстрате, основании) путем формирования на ее поверхности защитного рельефного покрытия и последующего переноса рельефа в подложку. В зависимости от вида облучения литографию делят на фото-, электронно-, рентгено- и ионолитографию [14–16].

Из существующих технологий для изготовления ступенчатого микрорельефа ДОЭ наибольшее распространение получила фотолитография. Ключевым этапом формирования микрорельефа ДОЭ средствами фотолитографии является получение набора прецизионных фотошаблонов.

4.2.1. Изготовление фотошаблонов

Первоначально при изготовлении ДОЭ использовались технологические средства машинной графики: электронные лучевые трубки, графопостроители, бинарные и многоградационные фотопостроители [17].

В частности, для создания синтезированных голограмм и ДОЭ широкое применение в 70-х годах получили электромеханические фотопостроители, управляемые компьютером (рис. 4.5).

Пиксел, поступающий из компьютера, преобразуется в электрический аналоговый сигнал, который модулирует яркость свечения лазерного диода, экспонирующего соответствующее место на фотопленке. Устройство обеспечивает получение растрового изображения с параметрами: размер растра 12,5; 25; 50 и 100 мкм; диапазон оптических плотностей $0–2D$; число уровней квантования до 256; скорость получения пикселов 28 кГц; максимальные размеры изображения 5×10 дюймов. Размер поля пикселов составляет $N_u \times N_v \sim 10^6$.

В 80-е годы появились лазерные фотопостроители. Минимальный размер пятна сфокусированного лазерного излучения He-Ne лазера — около 10 мкм. В зависимости от типа применяемой пленки, обычно это изопанхром, сенсибилизированный к длине волны 0,6328 мкм, а также от выбранных параметров записи и режимов проявления пленки, ширина линии может колебаться в пределах 9–13 мкм.

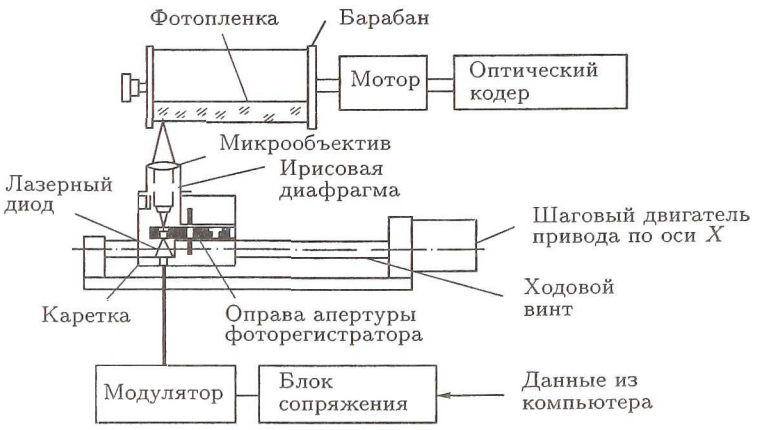


Рис. 4.5. Блок-схема фотопостроителя [17]

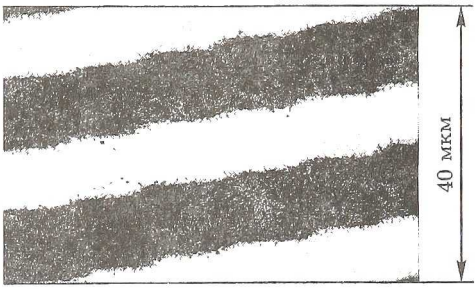


Рис. 4.6. Фрагмент фотошаблона ДОЭ на фотопленке с зернистой структурой

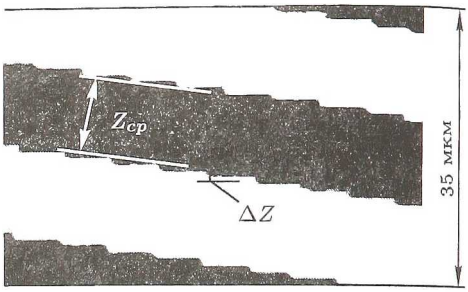


Рис. 4.7. Неровность края линии металлизированного шаблона ДОЭ

Поскольку для получения изображения используется фотохимический процесс, то необходимо стабилизировать и зафиксировать режимы проявления, сильно влияющие на качество и контраст изображения. Отработка параметров, управляющих процессом записи, занимает значительную долю всего времени вывода фотошаблонов ДОЭ. Распределение интенсивности лазерного излучения в пятне — гауссово, поэтому подбираются интенсивность лазерного излучения и параметр дефокусировки пятна. Заполнение линии происходит точками и поскольку важно, чтобы линия была непрерывной, точки должны записываться с перекрытием. Часто на выводимых тестах можно обнаружить либо распадение сплошной линии на чередующиеся точки, либо утолщение линии до недопустимых пределов. Преодолевая эти трудности серией тестов с полным циклом обработки фотопленки, добиваются стабильных параметров записи, после чего, для зафиксированных пленки и проявителя, параметры записи можно считать отработанными. При смене любого из них процесс отработки параметров записи следует повторить. Основным недостатком записи фотошаблонов на фотопленке является ее зернистая структура, сильно влияющая на неровность края и контраст изображения (рис. 4.6).

Для дальнейшей работы фотошаблоны ДОЭ с пленки переводят на стеклянные металлизированные пластинки, покрытые, например, тонким слоем хрома, в результате чего изображение получается более контрастным (рис. 4.7), и на металлизированные шаблоны, пригодные для процесса фотолитографии с многократным экспонированием и совмещением.

Таблица 4.1. Сравнительные характеристики некоторых лазерных генераторов изображений

Система	Размер пикселов (мкм)	Скорость генерации пикселов (МГц)	Максимальная площадь (в дюймах)	Позиционирование
Argis 1	5	25	18 × 32	Стеклянная координатная сетка
Argis 3	5	100	24 × 32	Стеклянная координатная сетка
Argis 4	10	100	24 × 32	Стеклянная координатная сетка
Laserad I	0,5	25	5 × 5	Интерферометр
Laserad II	2,5/1,25/0,5	25	8 × 10	Интерферометр

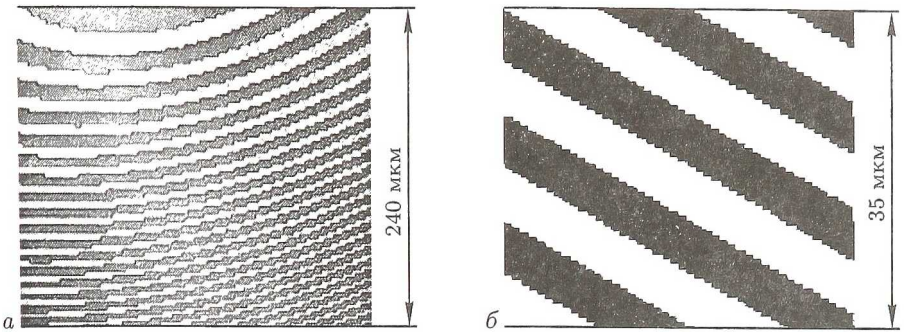


Рис. 4.8. Фрагменты фотошаблонов, выведенных на различных записывающих устройствах: (*а*) лазерный фотопостроитель FEAG, (*б*) электронно-лучевой литограф ZBA-20

Описанные выше электромеханические системы называют устройствами барабанного типа, т.к. пленка в них крепится на барабане за счет вакуумного прижима, а позиционирование вдоль одной из координат происходит за счет вращения микровинта с заданным шагом витков. Для получения фотошаблонов, непосредственно на стеклянной основе, были разработаны сканирующие в X-Y координатной сетке лазерные генераторы изображений, перемещением плоского стола которых управляют шаговые двигатели, а заданные координаты и точность перемещения контролируются с помощью лазерных интерферометров или стеклянной координатной сетки (табл. 4.1). С помощью лазерного линейного генератора изображений были синтезированы многопорядковые дифракционные решетки с пространственной частотой 500 лин.$^{-1}$ [18]. Размер поля пикселов составлял $N_u \times N_v \sim 10^8$.

Поскольку большинство ДОЭ имеет сложную топологию, большое внимание уделяется точности отображения линии. На рис. 4.8 представлены фрагменты фотошаблонов, выведенных на различных записывающих устройствах.

На рис. 4.8*а* и 4.8*б* фотошаблоны ДОЭ состоят из линий, аппроксимируемых набором элементарных изображений — точек или прямоугольников, ориентированных вдоль фиксированных ортогональных осей. При освещении такая структура приводит к рассеянию света на углах прямоугольников, снижению эффективности ДОЭ и дополнительному шуму в формируемом изображении. Существует целый класс оптических элементов, для которых такого рода погрешности недопустимы. Неровность края тем больше, чем ниже разрешение записывающего устройства. Кроме того, для матричных генераторов изображений линии, имеющие значительную кривизну, отображаются набором точек, координаты которых попадают в геометрические пределы линий. Если ширина линии становится сравнимой с размером записывающего пятна, то из-за вычислительных погрешностей на искривленных линиях могут образоваться разрывы. Оценить вероятность их проявления достаточно сложно, поэтому пользуются программой, визуализирующей критические места фотошаблона ДОЭ на экране компьютера еще на этапе синтеза [19], что позволяет заранее скорректировать параметры ДОЭ, не используя дорогостоящее записывающее оборудование (рис. 4.9). В данном случае естественно подобрать другие параметры ДОЭ, например, уменьшить апертуру или увеличить фокусное расстояние.

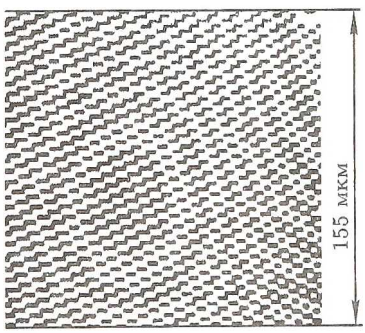


Рис. 4.9. Разрывы линий на фотошаблоне ДОЭ из-за ограниченности разрешающей способности записывающей системы

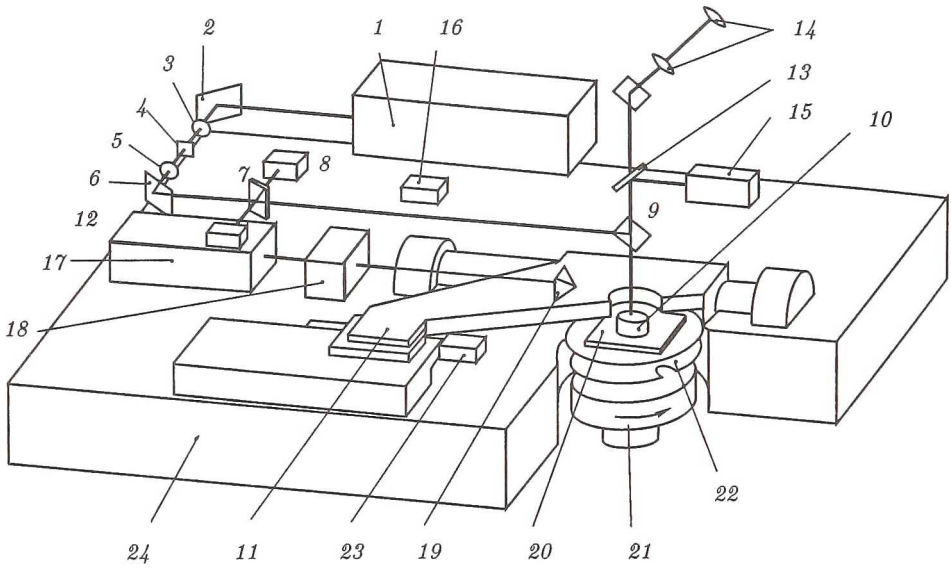


Рис. 4.10. Схема круговой лазерной записывающей системы КЛЗС-300 [20]: *1* — аргоновый лазер; *2, 6* — зеркала; *3, 5* — объективы; *7, 9, 13* — делители пучка; *8, 12* — фотоприемники; *10* — фокусирующий объектив; *11* — платформа; *14* — микроскоп; *15* — датчик фокусировки; *16* — осветитель; *17* — He-Ne-лазер; *18* — интерферометр; *19* — отражатель; *20* — подложка; *21* — шпиндель; *22* — датчик нуля; *23* — линейный движитель; *24* — гранитная плита

Для радиально-симметричных (и близких к ним) ДОЭ с успехом применяются записывающие устройства, которые позволяют существенно снизить эту погрешность, осуществляя вывод шаблонов, например, на круговой лазерной записывающей системе рис. 4.10 [20].

Шаблоны ДОЭ, обладающие осевой симметрей, записываются экспонированием внутренней и внешней границ линии путем кругового сканирования и заполнением внутренней части линии сканированием по спирали. Для аксиальных фотошаблонов, записанных на круговой лазерной записывающей системе, характерны отсутствие биений краев линий и великолепная точность отображения линий как по ширине, так и по месту расположения (рис. 4.11).

Для записи ДОЭ с произвольной структурой используется программное обеспечение, преобразующее прямоугольную систему координат в полярную для модуляции записывающего лазерного пучка в зависимости от углового положения. Пространственное разрешение записываемых структур превышает 1000 мм$^{-1}$ при записи, как кольцевых структур, так и произвольных микроизображений. В качестве светочувствительного материала используются фоторезист или тонкие пленки аморфного хрома и аморфного кремния [21]. Например, на данной установке были записаны бинарные фотошаблоны растра микролинз. Фрагмент микрорельефа на стеклянной подложке, где соединяются четыре линзы растра микролинз, показан на рис. 4.12.

Однако, параметры существующих лазерных генераторов изображений таковы, что для изготовления элементов видимого и даже ближнего ИК диапазонов необходимо использование фотоуменьшительной техники или записывающих устройств с более высоким разрешением.

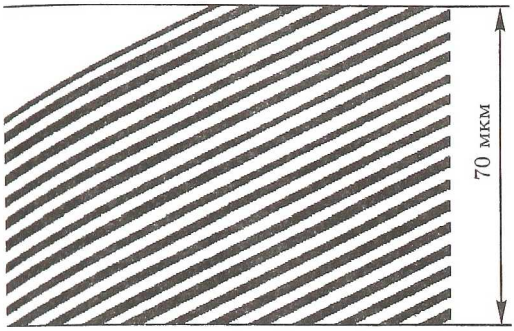


Рис. 4.11. Фрагмент фотошаблона, записанного на КЛЗС-300

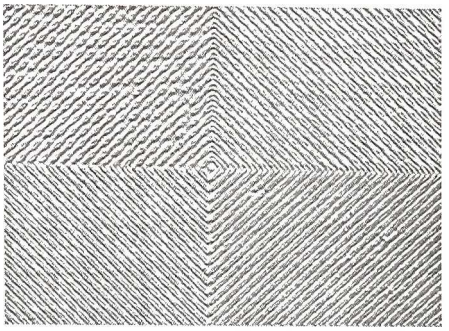


Рис. 4.12. Фрагмент микрорельефа растра микролинз

В 90-е годы основным типом генераторов изображений, используемым при создании ДОЭ, становятся электронно-лучевые генераторы (ЭЛГ). Засветка пластин производится полосами электронного пучка определенной ширины методами векторного или полярного сканирования. Динамическая система отклонения учитывает движение стола во время экспонирования и определяет координаты следующего штампа. Между двумя воздействиями на светочувствительный слой луч бланкируется. Обладая способностью работать при больших плотностях тока в пучке на мишени, а также изменять размеры штампов в широких пределах, ЭЛГ имеют высокую производительность — десятки квадратных сантиметров в минуту или десятки пластин в час. Пространственное разрешение ЭЛГ достигает величины 0,1 мкм, что снимает проблему фотоуменьшения для диапазона видимых волн.

Таблица 4.2. Параметры электронно-лучевых записывающих систем

Параметры	Линейное сканирование	Круговое сканирование
Электронный пучок:		
Напряжение, кВ	15; 30	15; 30
ток, нА	0,1–1,0	0,1–1,0
диаметр пучка, мкм	0,1–1,0	0,1–1,0
Сканируемая область, мм^2	3×3	3×3
Управление	X: аналоговое сканирование прямоугольным сигналом, XY: цифровое сканирование	XY: аналоговое сканирование синусоидальным сигналом
Разрешение (на сканируемую область)	2^{16} точек	2^{14} точек (радиус)
Время сканирования	100 мкс–10 с/лин	10 мс/круг

Как правило, ЭЛГ используются для получения бинарных масок, однако имеется возможность получить и многоградационные изображения. Размеры поля пикселов $N_u \times N_v \sim 10^{10}$. ЭЛГ (в отличие от фотопостроителя) дает амплитудную маску ДОЭ на специальном резисте.

При использовании микролитографии для изготовления ДОЭ не выполняются требования прецизионной микроэлектроники, согласно которым отношение шероховатости края линии к ее ширине $\Delta Z/Z$ (рис. 4.7) должно быть меньшим или

равным 1:100. Для концентрических шаблонов это означает, что ширина линии должна быть как минимум в сто раз больше, чем разрешающая способность записывающего устройства. С учетом количества уровней фазы M допустимая для реализации ширина наименьшей зоны становится чрезвычайно большой, что ограничивает применимость ДОЭ. Поэтому при изготовлении ДОЭ заведомо идут на нарушение этого принципа, чтобы получить возможность создания более короткофокусной и широкоапертурной оптики. На практике в дифракционной оптике можно считать приемлемым отношение $\Delta Z/Z = 1{:}10$, а для периферийных зон ДОЭ даже 1:3. В предельном случае структуру крайней зоны дифракционной линзы можно корректно отобразить линией (линиями), ширина которой (которых) удовлетворяет соотношению

$$Z_{\min} = (\Delta_{\text{spot}} + \Delta_{\text{step}}) \leqslant \frac{f\lambda}{d}\frac{1}{M}, \tag{4.8}$$

где M — число уровней квантования фазы, Δ_{spot} — диаметр сфокусированного записывающего пятна, Δ_{step} — минимальный шаг перемещения каретки записывающего устройства, f — фокусное расстояние, d — диаметр, λ — рабочая длина волны.

Это соотношение можно применять для оценки минимальной ширины линии произвольных ДОЭ, осуществляющих фокусировку.

Например, для $\Delta_{\text{spot}} = 10$ мкм; $\Delta_{\text{step}} = 10$ мкм; $\lambda = 0{,}5$ мкм и 8-ми уровневого ступенчатого микрорельефа минимальное реализуемое отношение фокусного расстояния f к диаметру d становится равным 320, т.е. для линзы с апертурой диаметром 10 мм фокусное расстояние не может быть меньше, чем 3200 мм. Ясно, что при таких условиях, без дополнительных мер, например, фотоуменьшения масок, невозможно создать используемый на практике ДОЭ видимого диапазона. Однако для ДОЭ среднего ИК диапазона ($\lambda = 10{,}6$ мкм) с апертурой диаметром 40 мм фокусное расстояние может быть равным или большим 600 мм, что соответствует практическим требованиям ряда задач.

При использовании фотошаблонов в технологическом процессе осуществляется контроль следующих принципиальных параметров:

1) размеров характеристических элементов фотошаблона (т.е. самых тонких, удаленных от центра или изогнутых линий на шаблоне, а также самых тонких штрихов маркерных знаков, предназначенных для контроля качества совмещения шаблонов);

2) плотности дефектов;

3) точности расположения шаблона.

Минимальный размер характеристических элементов (в дальнейшем просто минимальный элемент) является одной из важнейших характеристик шаблона. Для X-Y координатных генераторов изображений [18] размеры минимальных элементов у шаблонов, имеющих линейный и концентрический рисунок, могут существенно различаться. Для линейной решетки или цилиндрической линзы минимальная ширина линии может быть равна размеру записывающего пятна. Изогнутые линии на таких генераторах получают набором отрезков или точек. Важным параметром, характеризующим набор масок ДОЭ, также является совпадение минимальных элементов (линий), необходимое для совмещения масок из набора. Оба эти параметра (минимальная ширина линии и совпадение минимальных элементов на масках набора) имеют одинаковое влияние на технологический процесс. Правильное определение минимального элемента важно для поддержания необходимой величины экспозиции, при которой отрабатываются все линии структуры ДОЭ. Кроме того, большой разброс минимальных элементов в наборе масок одного ДОЭ заставляет менять диапазон настроек экспозиции при переходе от одной маски к другой.

Относительный уровень плотности дефектов на поверхности фотомаски в первую очередь зависит от степени очистки и подготовки поверхности подложки перед нанесением фоторезиста. Несмотря на то, что по сравнению с производством интегральных микросхем наличие таких незначительных дефектов, как проколы и разрывы линий, не столь критично для ДОЭ, они могут существенно снижать общую эффективность оптического элемента.

Для процесса фотолитографии при экспонировании (засветке) и совмещении необходимо, чтобы используемые шаблоны с изображением рельефа ДОЭ имели достаточно жесткую поверхность рабочего маскирующего слоя. Маскирующий слой должен иметь минимальную толщину для устранения теневого эффекта на границе светлых и темных линий. Плоскостность поверхности шаблона должна обеспечивать наилучший контакт по всей поверхности шаблона и подложки.

Точность позиционирования шаблона, определяющая качество совмещения двух или более уровней из набора шаблонов, обычно характеризуют погрешностью совмещения. Она зависит от точностных параметров и стабильности генератора изображений, обеспечивающего максимальную линейность в изображении прямой линии и перпендикулярность сетки.

При визуальном методе совмещения точность определяется принципом работы и качеством микроманипуляторов, размером и контрастностью знаков совмещения, формой знаков, а также постоянством этих параметров в процессе технологических обработок подложки. В процессе производства оптических элементов определяющим параметром является яркостный контраст, который образуется за счет рассеяния света на «клине травления» (т.е. ухода от вертикальности ступенек микрорельефа), и знаки кажутся темными на световом фоне. Причем размытие тем больше, чем больше глубина травления. Конфигурация знаков на подложке и фотошаблоне взаимосвязана с условием их совмещения, которое может предусматривать контакт элементов знаков, равенство зазоров или площадей между элементами и, наконец, соосность точек пересечения знаков (рис. 4.13).

Оптимальными считаются знаки, образующиеся при совмещении штрихов, вписанных между двумя другими штрихами. Ширина штрихов при изготовлении ДОЭ

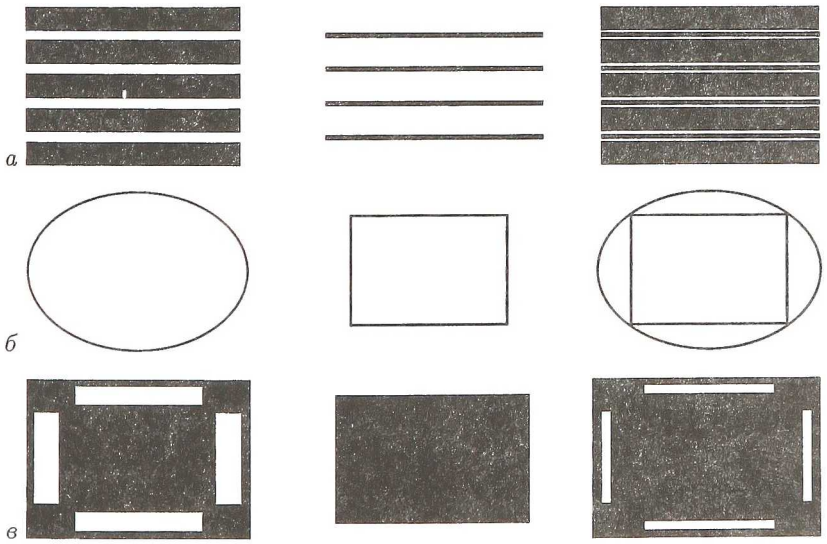


Рис. 4.13. Маркерные знаки: (*а*) равенство зазоров (*б*) касания (*в*) равенство площадей

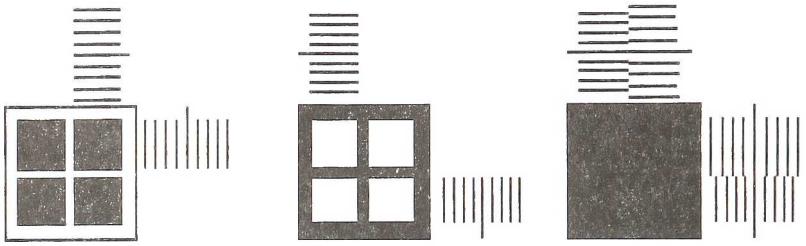


Рис. 4.14. Маркерные знаки со шкалой Нониуса

среднего ИК-диапазона может равняться 3 мкм, длина в 10 раз превышает ширину, контрастность 0,3–0,4 [14]. При выборе знаков совмещения при изготовлении ДОЭ необходимо учитывать структуру рисунка. Например, для радиально-симметричных структур ошибка по углу не является критичной, в то время как небольшой сдвиг по X или Y может привести к полной потере работоспособности элемента. Поэтому наряду с обычными маркерными знаками используется шкала Нониуса, позволяющая оценить погрешность в процессе совмещения и вовремя ее устранить (рис. 4.14).

Точность совмещения с помощью таких маркерных знаков не ниже 1,0 мкм при ширине линий ~ 10 мкм.

4.2.2. Экспонирование и проявление пленок фоторезиста

Экспонирование фоторезиста через шаблоны — одна из самых важных операций в фотолитографическом процессе. Благодаря этой операции происходит физико-химическое превращение фоторезиста, становящегося растворимым (нерастворимым) в проявителе и реализующего структуру ДОЭ.

Формирование изображения на поверхности подложки подчиняется определенным оптическим законам, например, в проекционной литографии главную роль играют характеристики объектива, а при контактной литографии — характеристики освещающего пучка, наличие и величина зазора между фотошаблоном и подложкой. Величину экспозиции, т.е. время засветки и интенсивность освещающего пучка, определяют следующие факторы: характеристика фоточувствительного слоя — тип фоторезиста и режимы его предварительной обработки; параметры шаблона ДОЭ – свойства материала подложки фотомаски и предельные размеры структуры ДОЭ; режимы проявления — концентрация в проявителе реагентов, температура раствора и время проявления. Результатом экспонирования и последующей обработки должен быть полученный в фоторезисте рисунок ДОЭ, определяющий топологию и размеры элементов структуры. Чтобы добиться высокого качества рисунка, применяют различные приемы экспонирования и типы фоторезистов.

Максимальная спектральная чувствительность фоторезистов находится в диапазоне $\lambda = 0{,}3$–$0{,}45$ мкм. Есть фоторезисты, максимальная чувствительность которых адаптирована к области глубокого ультрафиолета ($\lambda = 0{,}193$ мкм или $\lambda = 0{,}248$ мкм, как у KrF-криптоновых и ArF-аргоновых лазеров, соответственно), но при этом возникают трудности с выбором подложек для фотомасок, достаточно прозрачных к свету с такой длиной волны. Чаще экспонирование фоторезиста осуществляется с помощью актиничного излучения, спектральные характеристики которого удовлетворяют условиям технологического процесса и требованию максимальной разрешающей способности в данных условиях. Например, ртутные лампы, име-



Рис. 4.15. Зависимость поглощения фоторезиста от длины волны света (из инструкции фирмы Shipley к фоторезисту марки Microposite S1318)

ющие спектральные линии на длинах волн $\lambda = 0{,}436$ мкм (g-линия) и $\lambda = 0{,}365$ мкм (i-линия), где наблюдается максимальная чувствительность фоторезистов (рис. 4.15), широко используются в установках экспонирования.

Существуют три основных метода экспонирования фоторезистов при изготовлении масок или микрорельефа ДОЭ: а) сканирование сфокусированным пучком, б) печать с зазором или без него, в) проекционная печать (рис. 4.16).

В первом методе (рис. 4.16*а*) имеет место формирование сфокусированного в пятно на подложке, элементарного лучевого пучка частиц, таких как электроны, или светового пучка с гауссовым распределением интенсивности, а для синтеза желаемых шаблонов используется сканирующий аппарат. Сканирование может быть растрового типа, при котором с равномерным шагом сканируется вся апертура ДОЭ, и векторного типа, при котором пучком сканируются только экспонируемые зоны, т.е. сканирование осуществляется в «сплошной» траектории. Второй и третий методы экспонирования (рис. 4.16*б* и 4.16*в*) объединяются в, так называемые, параллельно экспонирующие системы (где могут экспонироваться одновременно большие области с множеством шаблонов) и классифицируются следующим образом:

1) системы, формирующие тень, — контактная печать или печать с зазором;

2) системы, формирующие изображение, — проекционная печать.

При контактной печати маска и образец находятся в близком контакте; при печати с зазором небольшой зазор вводится специально для того, чтобы предотвратить нанесение ущерба фотомаске. При проекционной печати амплитудная маска освещается и пропускает свет, собираемый с помощью объектива, который и формирует изображение на плоскости подложки.

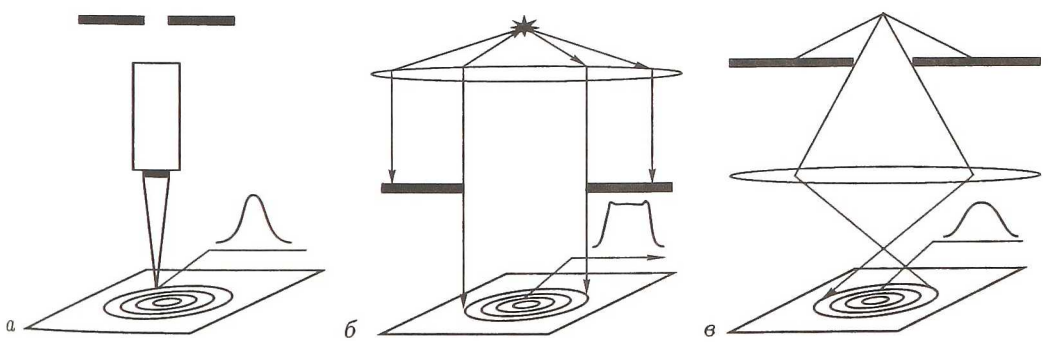


Рис. 4.16. Методы экспонирования фоторезиста: (*а*) сканирование; (*б*) печать с зазором; (*в*) проекционная печать

Разрешающая способность современных установок экспонирования достигает 0,5 мкм на площади в 100×100 мм^2. Точность совмещения не хуже 0,25 мкм. Специальное устройство позволяет выбрать и зафиксировать зазор между шаблоном и образцом от 0 до 50 мкм, сохраняя параллельность плоскостей. Некоторые из устройств могут работать в автоматическом режиме, что целесообразно для поточного производства интегральных микросхем. Однако условия изготовления ДОЭ заставляют использовать ручной режим экспонирования, связанный с нестандартными размерами подложек и некоторыми приемами, улучшающими качество изображения. Для фиксации подложек образца и маски обычно используют вакуум, а прижим друг к другу может быть как вакуумный, так и механический, причем механический обладает некоторыми преимуществами, поскольку позволяет использовать зазор между маской и образцом. Экспозиция определяется фотодатчиком в соответствии с заданными значениями. Контроль совмещения осуществляется с помощью бипольного микроскопа, в разных полях которого в отраженном свете наблюдают метки совмещения (см. рис. 4.13 и 4.14). При совмещении перемещается и вращается образец, в то время как маска остается неподвижной.

Известно, что интенсивность освещающего пучка, прошедшего через шаблон, обратно пропорциональна частоте линий на единицу площади шаблона, и часто бывает так, что на краю шаблона ДОЭ интенсивность в несколько раз меньше, чем в центре. В этом случае применяют переэкспозицию. Еще сильнее разность значений интенсивности прошедшего сквозь маску излучения на краю и в центре ДОЭ, вследствие эффектов дифракции, проявляется при наличии зазора. Редко удается достичь полного контакта по всей площади маска/подложка из-за присутствующих

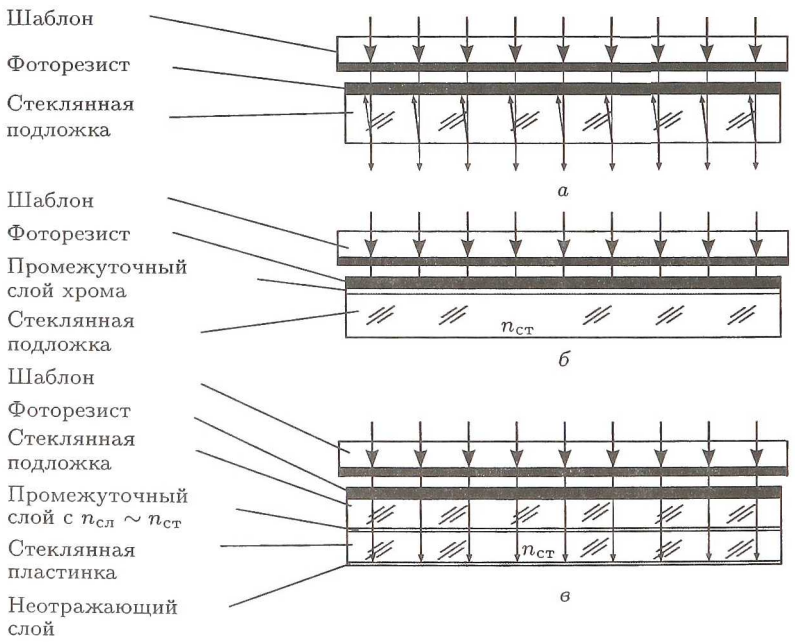


Рис. 4.17. Переотражение (*а*) и способы устранения эффекта переотражения от задней поверхности подложки: (*б*) промежуточным слоем хрома, (*в*) добавлением стеклянной пластинки с неотражающей задней поверхностью ($n_{ст}$ — показатель преломления стеклянной пластинки; $n_{сл}$ — показатель преломления промежуточного слоя)

в воздухе пылинок, неравномерности слоя фоторезиста или температурных деформаций поверхностей. Даже в случае полного контакта роль зазора будет играть слой фоторезиста, что для особо мелких структур может привести к выполнению условий дифракции Френеля [22]. Это приводит к растравливанию узких зон с искажением формы и размеров профиля микрорельефа или даже к полному его исчезновению. Бороться с этим эффектом можно, повышая разрешение метода, например, уменьшая длину волны актиничного излучения. Если по каким-либо причинам не удается обеспечить полного контакта между шаблоном и подложкой, то можно ввести в промежуток между ними слой иммерсионной жидкости, вытесняющий слой воздуха, изменяя тем самым угол дифракции, который меньше в более плотных по сравнению с воздухом средах (иммерсионный метод определения показателя преломления). Промежуточный слой жидкости практически не оказывает химического воздействия на слой фоторезиста, так как поверхность высушенного фоторезистного слоя — гидрофобная. Эту методику следует применять также в случае экспонирования больших площадей, когда есть признаки неравномерности слоя фоторезиста.

Другой причиной непрорабатываемости узких зон ДОЭ может стать отражение световой волны от задней поверхности стеклянной подложки, особенно, если толщина подложки достаточно большая, а дифракционный оптический элемент имеет фокусирующие свойства, как, например, растр микролинз. Отраженная от задней поверхности подложки световая волна интерферирует с падающей, формируя в слое фоторезиста интерференционные максимумы и минимумы в виде колец, увеличивая или подавляя интенсивность излучения. В результате могут не прорабатываться даже не самые узкие зоны. Бороться с этим явлением можно, устраняя отражение лучей от задней поверхности подложки. Например, прикрепив к ней без разделительного слоя толстую стеклянную пластинку с черной или матовой задней поверхностью, работающей подобно черному телу (рис. 4.17), или покрыв поверхность подложки слоем хрома путем напыления в вакуумной установке. Последний способ предпочтительнее, поскольку помогает адгезии фоторезиста, и хром выступает в роли дополнительного защитного слоя при травлении субстрата подложки.

4.2.3. Технология формирования микрорельефа ДОЭ

Подробный анализ факторов, влияющих на значения угла наклона θ боковой стенки фоторезиста и ширины линии W в проявляемом фоторезисте (см. рис. 4.18), приводится в работах [22–27]. Изучению влияния на качество получающейся топологии различных факторов литографического процесса (дозы экспонирования, времени и температуры термообработки, времени и температуры в процессе проявления, а также чувствительности фоторезиста) посвящены работы [25, 27].

С точки зрения дифракционной оптики важным является точность реализации следующих параметров (рис. 4.18):

— высота рельефа h и высота ступеньки h_s;

— как можно более крутой склон ступеньки микрорельефа (то есть угол Θ, характеризующий клин травления, должен стремиться к нулю; а угол θ, характеризующий завал края, — к $\pi/2$);

— минимизация погрешности ширины ступеньки, $\Delta W = W - W_1$.

Анализ работ [22–27] позволяет сделать вывод, что для качественного формирования микрорельефа и соблюдения указанных параметров необходимо использование высококонтрастного резиста.

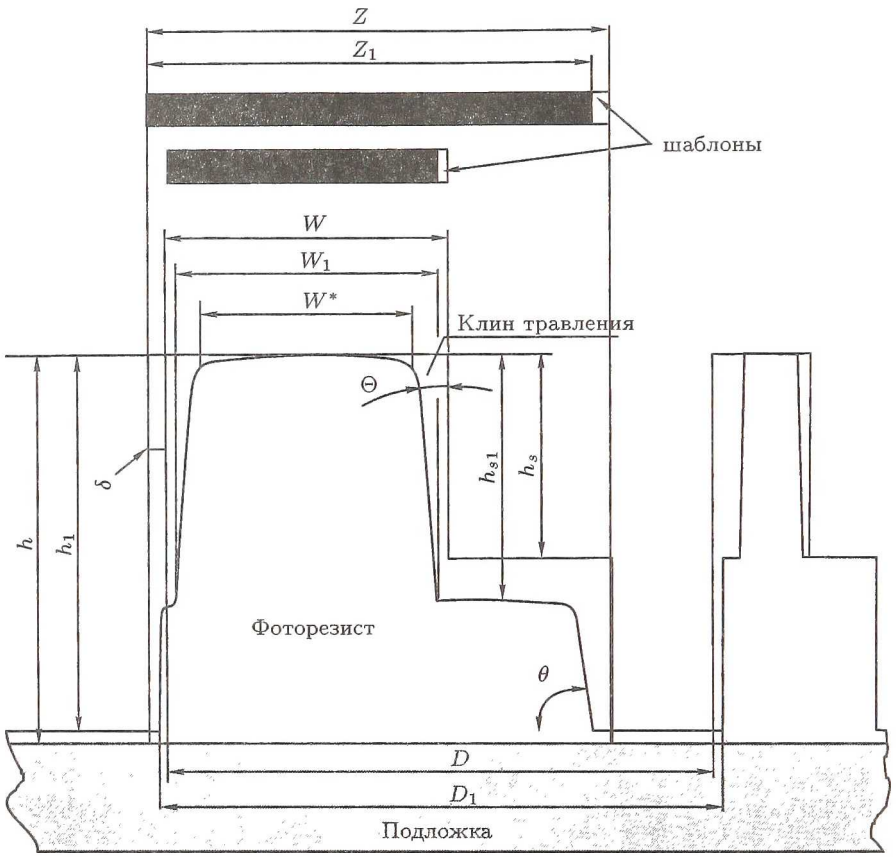


Рис. 4.18. Технологический профиль микрорельефа ДОЭ: Z — расчетная ширина линии на шаблоне; Z_1 — полученная ширина линии; D — ширина зоны идеального профиля; D_1 — ширина зоны полученного профиля; h_1 — полученная высота профиля; h_{s1} — полученная высота ступеньки; W — ширина ступеньки; W_1 — полученная ширина ступеньки; W^* ширина плоской части фоторезистной ступеньки; Θ — клин травления; θ — «завал края»

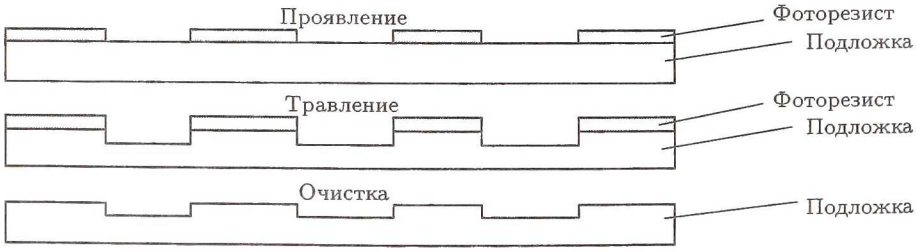


Рис. 4.19. Процесс формирования микрорельефа в подложке

Перенос рисунка ДОЭ в поверхностный слой подложки методом фотолитографии происходит в четыре стадии (рис. 4.19): экспонирование слоя фоторезиста через шаблон и образование скрытого изображения; проявление и задубливание рисунка, т.е. формирование защитной маски; травление поверхностного слоя подложки на незащищенных участках; очистка поверхности подложки от остатков фоторезиста [28, 29].

На каждой из стадий процесса действуют факторы, искажающие исходный рисунок шаблона. При экспонировании имеют место явления дифракции, преломления и отражения света, приводящие к изменению размеров элементов рисунка и размытости их краев. На этапе проявления и задубливания искажения размеров обусловлены набуханием слоя фоторезиста и усадкой фотомаски при тепловой обработке. При травлении, негативным фактором является боковое подтравливание под маску. Условия, в которых происходит обработка на разных стадиях, изменяются как от пластины к пластине, так и в пределах одной пластины. Это приводит к разбросу геометрических параметров элемента и его характеристик, что следует учитывать при выборе технологии изготовления ДОЭ.

Целью литографического процесса является образование фоторезистной маски, которая служит либо для локальной обработки нижележащего слоя подложки, либо сама является требуемым микрорельефом ДОЭ. В схематической форме сечение участка микрорельефа одного периода или зоны ДОЭ показано на рис. 4.18.

На рис. 4.18 показаны основные типы искажений микрорельефа, которые могут появиться в результате ошибок расчета фотошаблонов или неудовлетворительного технологического процесса. Неудовлетворительным назовем технологический процесс, который не удовлетворяет некоторым параметрам, например разрешению, или условиям получения микрорельефа. Ошибки в технологическом процессе могут возникнуть по многим причинам — это и погрешности вывода фотошаблонов, и ошибки в режимах экспонирования и совмещения, при проявлении фоторезистного профиля и травлении рельефа.

Определим основные погрешности микрорельефа:

$\Delta Z = Z - Z_1$ — погрешность ширины линии на шаблоне в результате дискретного характера выводных данных и/или недостаточного разрешения генератора изображения;

$\Delta D = D - D_1$ — погрешность ширины зоны, определяемая искажением размеров шаблона, шероховатостью края линии, недостаточной крутизной склона фоторезистной маски и ошибками совмещения масок ДОЭ;

$\Delta W = W - W_1$ — погрешность ширины ступеньки, причины те же, кроме ошибки совмещения;

$\Delta h = h - h_1$ — погрешность высоты рельефа (результат ошибок травления субстрата подложки);

$\Delta h_s = h_s - h_{s1}$ — погрешность высоты ступеньки (результат ошибки травления одной ступеньки);

Θ — клин травления (подтравливание под фоторезистную маску, изотропный характер травления и недостаточная вымываемость продуктов травления из узких щелей);

θ — «завал края» ступеньки, характерен для фоторезистных профилей микрорельефа ДОЭ;

δ — смещение края ступеньки рельефа из-за погрешностей шаблонов и экспонирующей системы.

Величина

$$\Delta x = \frac{1}{2}(W_1 - W^*), \tag{4.9}$$

где W^* — ширина плоской части фоторезистной ступеньки является шириной края фоторезистной маски, и ее величина существенно зависит от свойств освещающей системы, обеспечивающих передачу изображения шаблона ДОЭ на поверхность подложки. Величина W_1 зависит от параметров процесса обработки резистного слоя. Неконтролируемые вариации параметров резистной пленки и режимов ее обработки будут приводить к разбросу размера W_1 даже при неизменной освещенности; Δx определяет минимальную ширину полоски резистной маски: $W_{\min} = 2\Delta x$. В свою очередь, величина $W_{\min}$ определяет минимальный период решетки $d_{\min}$, воспроизводимой в резистном слое: $d_{\min} = 2W_{\min}$. Величина, обратная минимальному периоду, называется разрешающей способностью R литографического процесса:

$$R = 1/d_{\min}. \tag{4.10}$$

В большинстве случаев предпочтительной является резистная маска, у которой наклон боковых стенок близок к вертикальному. Обычно принимают, что относительная ошибка ширины линии в резистной маске не должна превышать 10%.

Для получения ступенчатого многоуровневого микрорельефа ДОЭ методами литографии используются два способа травления подложек: линейный — равномерное травление субстрата подложки [15]; дихотомический — неравномерное (степенное) травление [30]. Линейный способ создания ступенчатого микрорельефа характеризуется тем, что все уровни рельефа получаются травлением субстрата подложки на одну величину: $h_s = h_{\max}/M$. Количество масок в наборе определяется количеством уровней квантования фазы M, но, поскольку первый уровень соответствует нетронутой плоскости субстрата подложки (рис. 4.20), количество масок получается на одну меньше, чем число уровней квантования фазы, т.е. $M - 1$.

Важным моментом при работе с шаблоном ДОЭ является анализ характеристик линий, получаемых на маске. При изготовлении ступенчатого рельефа ДОЭ

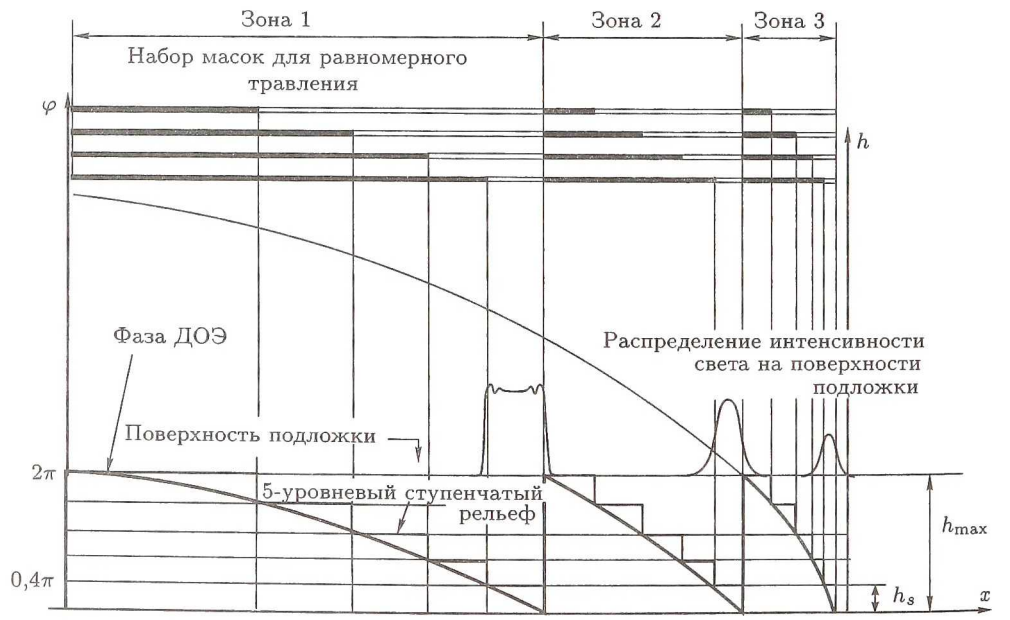


Рис. 4.20. Метод равномерного травления

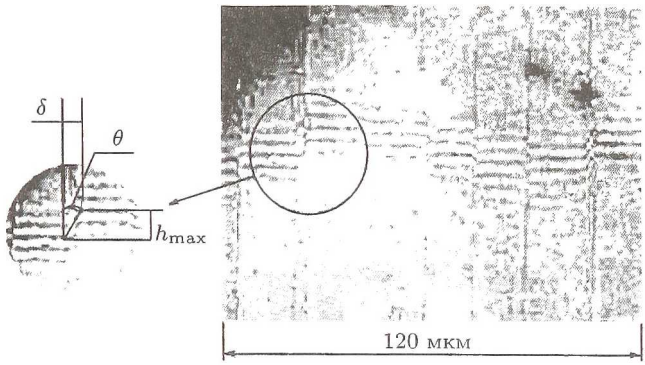


Рис. 4.21. Интерферограмма 5-ти уровневого микрорельефа линзы Френеля: h_{max} — соответствует скачку фазы на $4/5\pi$; θ — «завал края» ступеньки на краю зоны ДОЭ; δ — смещение края ступеньки на 5–8 мкм

равномерным травлением, на каждом шаблоне внутри зоны находится только одна линия. Из рис. 4.20, на котором показаны сечения масок, хорошо видно, что один край линии совпадает с краем зоны (это может не выполняться только для центральных зон ДОЭ), а ширина линии зависит от номера шаблона и определяет форму микрорельефа после травления.

Например, для 8-ми уровневого оптического элемента необходимо 7 масок, имеющих линии разной ширины. Сам процесс фотолитографии для изготовления такого элемента при равномерном травлении повторяется не менее 7 раз, включая 6 этапов совмещения и экспонирования шаблонов, а работоспособным элемент становится только после выполнения всех технологических операций. Этот метод обладает рядом достоинств. Прежде всего, ширина линий на крайних зонах хоть и может достигать относительно малых значений (порядка D_{min}/M, где D_{min} — минимальная ширина зоны), однако указанные линии расположены достаточно далеко друг от друга (рис. 4.20), чтобы оказывать взаимное влияние при экспонировании через шаблон, что несколько повышает разрешение метода. Незначительные ошибки совмещения шаблонов с микрорельефом, полученным на предыдущем этапе, или расхождение размеров зон на маске с расчетными параметрами ДОЭ приводят к незначительным искажениям, а так же к сглаживанию острых кромок границ зон. На рис. 4.21 показана интерферограмма микрорельефа ДОЭ, полученная на микроинтерферометре МИИ-4, с выделением границы перехода от одной зоны к другой, на которой видны сразу несколько уровней, что означает наклон боковой стенки.

Суть формирования микрорельефа ДОЭ неравномерным травлением заключается в том, что на каждом последующем шаге литографии субстрат подложки травится до глубины ровно в два раза меньшей, чем предыдущая. При этом на первом этапе получают бинарный микрорельеф с глубиной травления $h_{max}/2$, где h_{max} определяется по формулам (4.3), (4.4). Дихотомический метод определяет не только глубину травления микрорельефа на каждом шаге, но и принцип синтеза масок ДОЭ, который характеризуется двукратным увеличением числа линий в пределах одной зоны (рис. 4.22) при переходе к каждому следующему шаблону. При выбранном числе уровней квантования M минимальная ширина линии на последнем шаблоне, используемом при дихотомическом методе, совпадает с минимальной шириной линии (порядка D_{min}/M) на последнем шаблоне для равномерного травления. Количество необходимых шаблонов в первом случае значительно меньше (в $(M-1)/\log_2 M$ раз), а количество линий внутри зоны на последнем шаблоне в $M/2$ раз больше, чем при равномерном травлении.



Рис. 4.22. Метод неравномерного травления микрорельефа ДОЭ

При неравномерном травлении количество уровней квантования фазы находится в степенной зависимости от количества масок в наборе:

$$M = 2^N, \tag{4.11}$$

где N — количество масок.

С помощью трех масок можно изготовить 8-ми уровневый микрорельеф ДОЭ, причем на каждом шаге получаем работоспособный элемент, а все последующие этапы травления увеличивают его энергетическую эффективность. Существенным недостатком метода можно считать большую чувствительность к ошибке совмещения и искажениям ширины линий, например, вследствие ошибок режимов экспонирования – проявления и/или генерации масок на записывающем устройстве. В результате на профиле микрорельефа образуются «пики» на стыке совмещаемых линий (рис. 4.23).

Чередующиеся темные и светлые линии на шаблоне в крайних зонах ДОЭ вместе ведут себя как квазипериодическая дифракционная решетка, и чем выше простран-

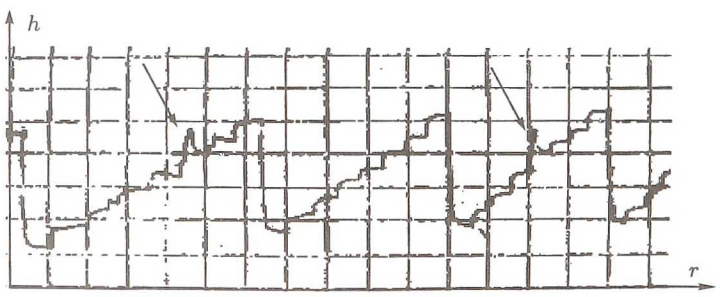


Рис. 4.23. Фрагмент профилограммы микрорельефа 8-ми уровневой линзы Френеля с «пиками»

ственное разрешение шаблона, тем сильнее различие при экспонировании между амплитудой и профилем интенсивности излучения на поверхности подложки в центре и на краю ДОЭ.

С точки зрения оценки реализуемости ДОЭ по технологии фотолитографии существует несколько параметров, характеризующих как сам ДОЭ, так и процесс его синтеза.

Одним из самых важных характеристических параметров ДОЭ является наименьшая ширина зоны (обычно крайней), которую для дифракционной линзы в параксиальном приближении можно оценить выражением [11]

$$D_{\min} = \frac{\lambda f}{a}, \tag{4.12}$$

где f — фокусное расстояние, a — радиус апертуры ДОЭ. Эту формулу можно приближенно использовать и для других ДОЭ, обладающих оптической силой, например, для фокусаторов.

Также необходимо принимать в расчет и то, что для каждой маски с номером $i = 1, 2, 3, \ldots$ минимальная ширина линии в экстремальной зоне дифракционной линзы, определяемая по формуле (4.8), будет существенно меньше, например, при неравномерном травлении $M = 2^i$, благодаря чему ширина линии может уменьшиться до значений, практически не воспроизводимых оборудованием. В этом случае необходимо выбирать параметры ДОЭ в соответствии с возможностями технологии.

Параметры ДОЭ выбирают с таким расчетом, чтобы выполнялся принцип реализуемости и максимальной эффективности ДОЭ. Технология изготовления ДОЭ, базирующаяся на фотолитографии, может быть представлена в виде следующей последовательности операций.

Базовая последовательность технологических операций:

1. Автоматизированное проектирование фотошаблонов на ЭВМ с выдачей кодов управления на генератор изображений.
2. Изготовление фотошаблонов (при использовании оптических генераторов — изготовление промежуточных фотошаблонов путем уменьшения исходных).
3. Подготовка пластин (стекло с нанесенным слоем хрома).
4. Нанесение фоторезиста, проекционное или контактное экспонирование по фотошаблону (из набора фотошаблонов).
5. Проявление, сушка и другие операции с фоторезистом для формирования защитной маски на стеклянных пластинах.
6. Травление хрома, снятие остатков резиста, получение металлизированного шаблона или набора шаблонов (путем повторения этапов 1–6).
7. Контроль параметров металлизированных шаблонов.
8. Подготовка субстрата (стекло, кварц, кремний и т.п.)
9. Повторение операций 4–5 на одном и том же субстрате с совмещением металлизированных шаблонов для получения ступенчатого рельефа на поверхности субстрата, контроль высоты рельефа.
10. Напыление защитных слоев, отражающих покрытий (для элемента, работающего на отражение), снятие реплик, гальванопластика, тиражирование.
11. Заключение в оправу (корпусирование).
12. Аттестация в лазерной системе, паспортизация.

Пример 4.1. Формирование дифракционного микрорельефа послойным наращиванием фоторезиста. Особенность метода состоит в том, что ступенчатый рельеф ДОЭ формируется путем многократного нанесения фоторезистивных слоев (слой на слой), прошедших этапы засветки через соответствующий фотошаблон и все процессы обработки [31].

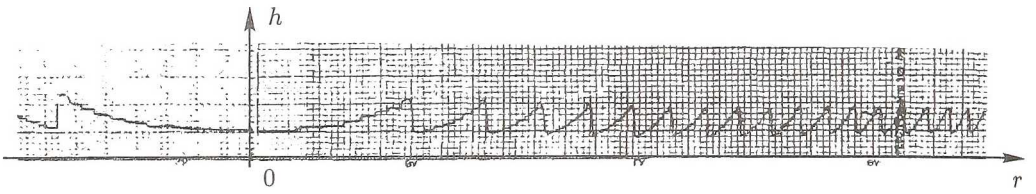


Рис. 4.24. Профилограмма микрорельефа фокусатора

Фоторезист (на основе нафтохинондиазидов) наносится методом центрифугирования на стеклянную подложку с плоскостностью 1–4 кольца Ньютона по всей поверхности. Определяющим фактором в данном процессе является толщина фоторезистивного слоя. Для стандартного фоторезиста отработаны технологические режимы получения слоев определенной толщины [32]. Стабильность толщины наращиваемого слоя обеспечивается фиксированной вязкостью фоторезиста и стабильностью скоростных характеристик центрифуги. Глубокое термическое задубливание предыдущего слоя позволяет без деструкции работать с последующим слоем.

Рассмотрим изготовление фокусатора в отрезок со следующими параметрами:

диаметр апертуры	40 мм,
фокусное расстояние	400 мм,
рабочий угол	45 град.,
длина волны	10,6 мкм,
размеры отрезка	$5 \times 0{,}3$ мм.

При изготовлении многоуровневой структуры фокусатора была использована модернизированная промышленная установка экспонирования и совмещения ЭМ-5006А, обеспечивающая получение минимального размера зоны порядка 3 мкм, по рабочему полю диаметром 100 мм. Она рассчитана на применение фоторезиста (типа ЭЛП-9 или ФП-051) со следующими параметрами: толщина слоя (0,5–1,0) мкм; неравномерность толщины не более 10%; погрешность совмещения элементов шаблона и субстрата по координатам X и Y не более $\pm 0{,}25$ мкм.

Профилограмма изготовленного 8-ми уровневого микрорельефа представлена на рис. 4.24. Профилограмма свидетельствует о высоком качестве микрорельефа, при сравнительно простой технологии его получения.

Пример 4.2. Изготовление ДОЭ с использованием полутоновой маски на основе DLW (Direct Laser Write glass photoplates) стеклянных пластинок в Исследовательском центре ФИАТ (Орбассано, Италия). Как было отмечено ранее, перспективной при изготовлении ДОЭ является технология, позволяющая реализовать фазовую функцию элемента в виде кусочно-непрерывного микрорельефа. В Исследовательском центре ФИАТ за счет применения DLW стеклянных пластин было реализовано преимущество одношагового изготовления полутоновой маски в технологии производства ДОЭ [33–35]. DLW стеклянные пластинки — это монолитные, химически обработанные пластины из силикатного стекла, без какого-либо покрытия, содержащие окрашенные частички серебра в слое толщиной около 1 мкм у поверхности стекла. Сфокусированный лазерный луч любой длины волны в спектральном диапазоне вблизи УФ, видимого ($\lambda = 0{,}524$; $\lambda = 0{,}632$; $\lambda = 0{,}647$ мкм), ближнего ИК ($\lambda = 0{,}82$ и $\lambda = 1{,}06$ мкм) и ИК диапазона ($\lambda = 10{,}6$ мкм) может быть использован для обесцвечивания нагреванием этих окрашенных частичек серебра, заставляя часть (или все окрашенные) частички серебра переходить в бесцветные ионы серебра. Пропускание DLW стеклянных пластинок увеличивается с увеличением плотности энергии записывающего сфокусированного лазерного пучка. Экспозиция этой полутоновой маски на лазерной записывающей установке

позволяет использовать существующее программное обеспечение, уже написанное для изготовления бинарных масок, и достижения прямой записи по фоторезисту при изготовлении ДОЭ. Изготовленные таким образом полутоновые маски могут быть использованы в оптической литографии (проекционная или контактная печать) для массового производства резистных профилей, обеспечивая отсутствие ошибок совмещения. Запись маски осуществляется за счет использования различных плотностей энергии лазерного пучка для генерации полутонов.

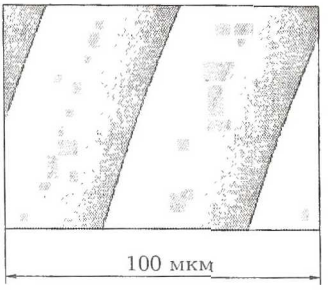


Рис. 4.25. Фрагмент фотошаблона бифокальной микролинзы

На круговой лазерной системе (рис. 4.10) были записаны полутоновые фотошаблоны ДОЭ: бифокальной линзы и растра микролинз. В обоих случаях фазовые функции элементов были заранее рассчитаны и представлены в виде двумерного массива данных. Использовавшийся формат данных позволяет наглядно представить конечный результат синтеза полутонового фотошаблона ДОЭ (рис. 4.25).

Для записи бифокальной линзы использовался одномерный массив, вырезанный вдоль линии, проходящей через центр элемента. Расчетные параметры бифокальной линзы следующие: диаметр линзы — 10 мм, фокусные расстояния, первое и второе — 50 и 100 мм соответственно, количество уровней фазы — 40. Запись производилась по методу симметричных ДОЭ с минимальным радиальным смещением 0,75 мкм при эффективном размере пятна — 0,9 мкм.

На рис. 4.26*а* представлен фрагмент растра микролинз прямоугольной формы. Фазовая функция микролинзы рассчитана с параметрами: размер — 1,4 × 2 мм, фокусное расстояние — 25 мм, число уровней фазы — 8. Размещение растра на подложке — симметричное относительно центра вращения, он содержит 9 микролинз. Запись осуществлялась многопроходным экспонированием с эффективным размером пятна 0,9 мкм. Микроинтерферограмма нескольких зон микролинзы (рис. 4.26*б*) получена с помощью телевизионной приставки к микроинтерферометру МИИ-4 с последующей обработкой данных на компьютере. Симметрично расположенные полосы интерферограммы появились в результате отражения света от поверхности подложки. Микронное разрешение было получено в LDW пластинках с использованием пучка аргонового лазера видимого диапазона, сфокусированного объективом с числовой апертурой 0,25.

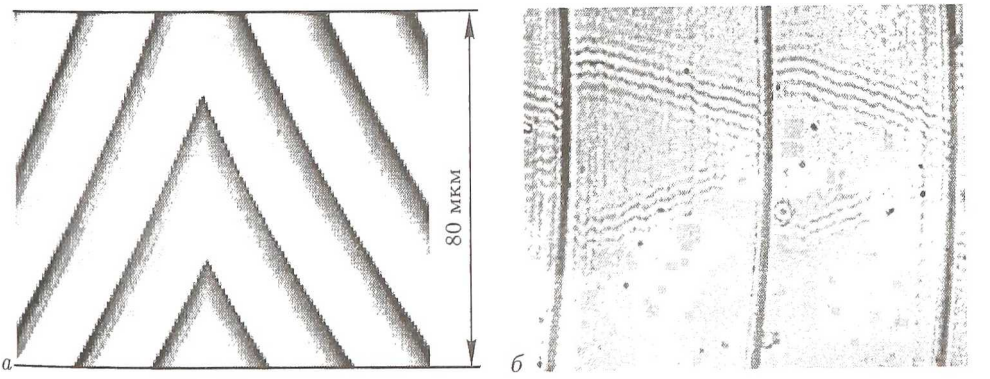


Рис. 4.26. Фрагмент фотошаблона растра микролинз (*а*) и микроинтерферограмма нескольких зон (*б*)

4.3. Изготовление ДОЭ методами электронно-лучевой литографии

Принципиальным отличием электронно-лучевой литографии (ЭЛЛ) от традиционной оптической (фотолитографии) является использование в качестве инициаторов физико-химических процессов в резисте высокоэнергетических электронов [36]. Учитывая дифракционные ограничения фотолитографии в сочетании с известными квантово-механическими представлениями о движении микрочастиц, можно утверждать о существенных преимуществах ЭЛЛ в части разрешающей способности технологии создания топологических рисунков.

В приближении Де-Бройля с движением свободного электрона можно связать волновые представления, количественная характеристика которых приведена в табл. 4.3 (без учета релятивистских эффектов).

Таблица 4.3. Волновое представление движения электронов

Энергия электронов, эВ	1	10^2	10^4
Длина волны, м	10^{-9}	10^{-10}	10^{-11}

Согласно данным таблицы предельная разрешающая способность ЭЛЛ при типичных значениях ускоряющих напряжений ($\sim 10^4$ В) составляет 0,1Å, и не может рассматриваться в качестве ограничения разрешающей способности даже на принципиальном уровне (учитывая, что размер электрона $\sim 2{,}8 \cdot 10^{-5}$Å).

О разрешающей способности и отработанности методов ЭЛЛ можно судить по тому, что в 1998 году состоялся переход ряда производств микросхем к промышленному выпуску чипов по многоуровневой технологии с разрешением 0,18 мкм. Однако существенный отрыв показателей ЭЛЛ в части разрешающей способности процесса «засветки» резиста имеет ряд ограничений на последующих этапах обработки, связанных с особенностями технологии дифракционной оптики.

Принципиальные и технические ограничения ЭЛЛ. Рассматривая возможности использования ЭЛЛ при создании ДОЭ, следует остановиться на некоторых наиболее критических ограничениях ЭЛЛ:

1. Ограничения, обусловленные рассеянием электронов в подложке. Этот тип ограничений является следствием упругих столкновений первичных электронов с атомными ядрами подложки. В результате отражения от подложки электроны возвращаются в фоторезист, обладая энергией, достаточной для инициирования элементарного акта экспонирования. Эффективность отражения в первом приближении является линейной функцией от атомного номера материала подложки.

2. Ограничения, обусловленные статистическим характером потока электронов. Волновая природа электронного пучка сильно проявляет себя при чрезвычайно малых размерах рисунка. Это следует учитывать при проектировании ДОЭ видимого и УФ диапазонов.

3. Ограничения, обусловленные аберрацией фокусирующей системы. Аберрация фокусирующих электронно-оптических систем присуща всем электронно-лучевым методам и нашла свое отражение в «минимальном размере пятна». Диаметр «минимального пятна» современных установок ЭЛЛ составляет порядка 0,005–0,01 мкм. Если при создании ДОЭ используется процедура стыковки соседних рисунков, точность обусловлена применением механических приводов, погрешность которых составляет не менее 0,1 мкм. Поэтому окончательная погрешность рисунка является функцией количества шагов механического перемещения предметного

столика электронного литографа. Проблема уменьшения числа шагов может быть решена путем увеличения геометрических размеров элементарного фрагмента исходного рисунка, в пределах которого может быть достигнута необходимая точность.

4. Ограничения, обусловленные системой управления отклонения пучка. Процесс управления усилителями отклоняющей системы связан с необходимостью перевода управляющих сигналов из цифровой формы в аналоговую, и разрешение системы определяется разрядностью цифро-аналогового преобразователя (ЦАП). Помимо этого, на точность системы управления влияют нелинейные и шумовые эффекты в усилителях отклонения луча. Как показывает опыт эксплуатации электронных литографов, нелинейность и шумы дают вклад в дисперсию отклонения, сравнимый с минимальной единицей информации 16-разрядного преобразователя. Эти факторы влияют на значение минимального шага перемещения луча по фрагменту рисунка ДОЭ.

В табл. 4.4 представлены параметры современных систем ЭЛЛ [23] с различными способами управления электронным пучком.

Вышеизложенные данные характеризуют пределы разрешающей способности ЭЛЛ, характерные для однократного экспонирования без последующего совмещения. При совмещении нескольких масок требования к допускам изготовления шаблонов (и структур) существенно повышаются. Из-за ошибок совмещения линий,

Таблица 4.4. Параметры современных систем ЭЛЛ

Модель	EBES4, Lepton Inc.	JBX-6000FS, JEOL Inc.	EBL Nanowriter, Leica Lithography Systems Ltd.	EL-4, IBM Corp.
Способ управления электронным пучком	Растровое сканирование	Векторное сканирование	SEM система	С изменением формы луча
Минимальный размер пятна	0,125 мкм	0,005 мкм	0,1 мкм	0,15 мкм
Совмещение	Автомат.	Автомат.	Автомат.	Автомат.
Рабочее поле	Полосы 256×32 мкм	Макс. 800 мкм на 50 кВ	До 2 мм при 50 кВ	Макс. 10 мм
Ускоряющее напряжение	20 кВ	25, 50, 100 кВ	10 –100 кВ	75 кВ
Плотность тока	1600 А/см^2	—	—	—
Частота	500 МГц	12 МГц	1 МГц	2, 3 подложки в час
Размер подложки	150 мм	150 мм	50 мм	200 мм
Платформа	Лазерное позиционирование, макс. перемещение 146 мм при минимальном шаге 0,005 мкм	Лазерное позиционирование, макс. перемещение 150 мм при минимальном шаге 0,0006 мкм	Лазерное позиционирование	Лазерное позиционирование на электростатических подшипниках

расположенных в разных уровнях, и возможных отклонений размеров при термообработке имеется неопределенность положения каждого края линии. Для двух линий в этом случае общее число вероятных ошибок равно общему числу степеней свободы. Если принять равенство абсолютных значений возможных ошибок, то при ширине линии 0,5 мкм ошибка совмещения не должна превышать 0,062 мкм. В противном случае обе линии могут оказаться полностью несовмещенными. Этот факт приобретает ключевое значение для таких пространственно-протяженных структур, каковыми являются ДОЭ. В этой связи следует считать, что конструктивные особенности ДОЭ могут сыграть решающую роль в эффективности практической реализации ДОЭ.

Формирование топологии рисунка. Технической основой ЭЛЛ является использование сканирующего остросфокусированного электронного луча, движением и модуляцией интенсивности которого управляет компьютер (рис. 4.27). Подобный подход позволяет реализовать большую гибкость технологического процесса с точки зрения оперативности изменения топологии рисунка.

Под воздействием электронной бомбардировки происходит экспонирование электронного резиста (электроны сшивают либо разрушают атомно-молекулярные связи), и таким образом свойства резиста локально изменяются [37].

Доза электронного облучения регулируется изменением времени сканирования на определенном участке и определяется на основе калиброванной кривой (соотношение между электронной дозой и глубиной стравливания). Соотношение электронной дозы и глубины стравливания определяется с помощью фотографирования ступенчатого профиля решетки на сканирующем электронном микроскопе. После экспонирования в процессе проявления (травления) происходит селективное удаление экспонированной либо неэкспонированной части в зависимости от типа резиста (негативный или позитивный). Сформированный подобным образом рельеф является топологическим рисунком, используемым для дальнейших технологических операций. Представленный процесс создания топологии рисунка является основным в ЭЛЛ. В зависимости от области применения топологический рисунок используется: при создании фотошаблонов, для последующего тиражирования (как правило, с ухудшением разрешающей способности) или непосредственно на обрабатываемой структуре (подложке).

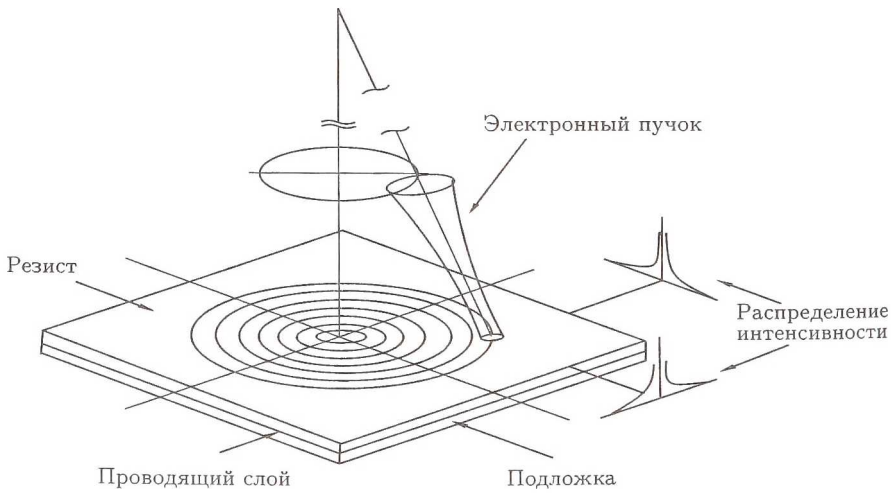


Рис. 4.27. Схема записи на ЭЛЛ

ЭЛЛ на подложке особенно необходима, когда важны гибкость технологического процесса и достижение высокой разрешающей способности [38]. Существующие типы установок ЭЛЛ [23] могут быть с успехом использованы в мелко- и среднесерийном производстве ДОЭ. Учитывая современный опыт использования ЭЛЛ, по прогнозам специалистов можно ожидать увеличения производительности электронно-лучевых установок до уровня рентабельности широкого промышленного использования при изготовлении ДОЭ.

Следует отметить, что при существующих требованиях к разрешающей способности ДОЭ не исключается возможность использования эталонных шаблонов, полученных ЭЛЛ. Последние имеют то преимущество, что при использовании достижений оптической литографии обеспечивают более короткий рабочий цикл мелкосерийного производства ДОЭ с сохранением приемлемого уровня качества.

В качестве примеров можно привести технологические решения, имеющие место при изготовлении ДОЭ с использованием ЭЛЛ.

Изготовление ДОЭ в Институте прикладной физики Университета Фридриха Шиллера (Йена, Германия). В Институте прикладной физики Университета отработаны различные технологии автоматизированного расчета и изготовления ДОЭ [39–43]. На основе процесса пошагового литографического травления резиста применяется технология изготовления ДОЭ для работы в видимом диапазоне с маломощными источниками лазерного излучения [40–42]. Технологическая реализация таких элементов производится равномерным травлением резиста на основе полиметилметакрилата (ПММА), нанесенного на кварцевое стекло по заданному числу уровней квантования фазы (обычно 2–16 уровней). Для осуществления равномерного травления резиста по рассчитанной полутоновой фазовой маске элемента рассчитывается комплект технологических бинарных файлов (виртуальных бинарных масок), определяющих пошаговый процесс литографического травления резиста. Значения фазовой функции, лежащие в промежутке $[0, 2\pi)$, трансформируются в соответствующие значения толщины микрорельефа пропускающего элемента с учетом показателя преломления резиста.

Количество бинарных файлов виртуальных бинарных масок на единицу меньше количества уровней квантования. Схематичное изображение процесса изготовления элемента представлено на рис. 4.28.

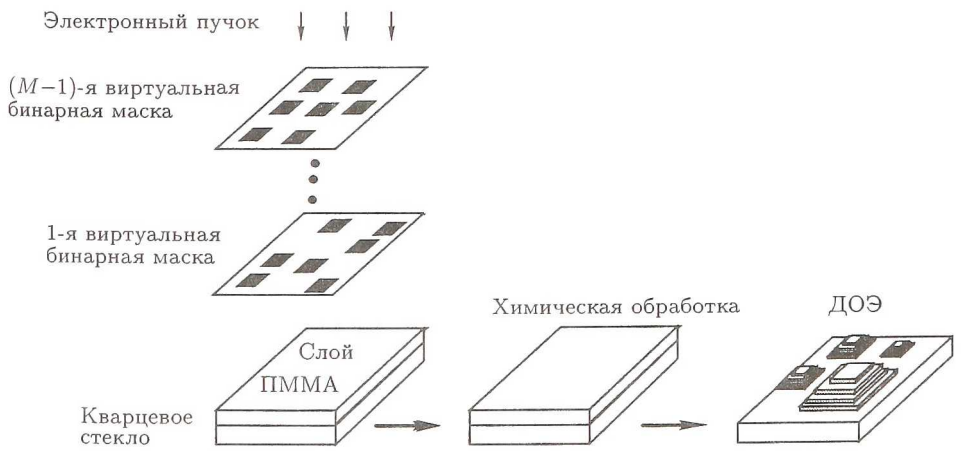


Рис. 4.28. Последовательность технологических операций формирования многоуровневой структуры

В качестве исходной заготовки используется пластина кварцевого стекла, покрытая слоем ПММА (толщина порядка 2–3 мкм). M-уровневый микрорельеф ДОЭ формируется с помощью $(M-1)$-кратного равномерного электронного травления с помощью электронного микролитографа ZBA-23 Jenoptik GmbH по соответствующим виртуальным бинарным маскам. Разогрев резиста, возникающий из-за процедуры травления, приводит к определенным отклонениям от заданного рельефа [40, 41]. Поэтому для того, чтобы достичь реализации расчетного профиля, рекомендуется сначала осуществлять травление лишь до 80% расчетной глубины, замерить полученную глубину травления микропрофилометром и завершить процедуру травления одним или несколькими дополнительными циклами «травление–профилометрия» [41, 42]. Для увеличения эффективности при замере глубины протрава используется регулярная ступенчатая тестовая структура вместо нерегулярного сложного дифракционного профиля, рассчитанного, например, итерационной процедурой. Обычно, такая тестовая структура размещается на подложке вне апертуры ДОЭ для того, чтобы не нарушать функцию комплексного пропускания ДОЭ, или на апертуре тестового ДОЭ. Для контроля микрорельефа используется интерферометрический микропрофилометр MICROMAP-512 с разрешением 3 мкм. Примеры ДОЭ, изготовленных по описанной технологии, приведены в главе 6.

Пример 4.3. В Университете Фридриха Шиллера отработаны технологии производства ДОЭ, работающих в ИК-диапазоне, с использованием метода степенного травления кварцевого стекла и ZnSe-подложек с нанесением покрытия из золота на отражающую поверхность, а также гальванопластического тиражирования изготовленных фокусаторов. Для формирования микрорельефа ДОЭ используется, как правило, плазменное травление подложки по 8 или 16 уровням, для чего на литографе выводится комплект из 3-х или 4-х бинарных фотошаблонов [43].

Натурные исследования изготовленных оптических элементов для работы в видимом и ИК-диапазонах осуществляются в Институте прикладной физики и Институте прикладной оптики Университета. Для этих целей они оснащены необходимым оборудованием: лазерами видимого и ИК-диапазонов, в том числе производства Carl Zeiss Jena, виброустойчивыми платформами, CCD-камерами видимого диапазона (производства SPIRICON Ltd.) и ИК-диапазона системы MAC-2 (производства ACCS GmbH) с соответствующим программным обеспечением.

На рис. 4.29 представлен результат профилометрического исследования микрорельефа фокусатора в прямоугольник, изготовленного шестнадцатиуровневым травлением кварцевого стекла [43]. Представленная профилограмма позволяет говорить о высоком качестве проработки микрорельефа ДОЭ.

Результат измерения камерой MAC-2 (длина волны 10,6 мкм) распределения интенсивности, формируемого фокусатором лазерного излучения ИК-диапазона, представлен на рис. 4.30.

Пример 4.4. Технология изготовления бинарных фазовых ДОЭ в Университете Йоенсуу (Финляндия). Наличие в Университете Йоенсуу низковольтного электронного генератора изображений Leica LION LV1, обеспечивающего предельное разрешение около 0,04 микрона, позволяет изготавливать микрорельеф с использованием одного шаблона [44].

С помощью электронного луча в слое ПММА, толщиной 0,18 мкм, осажденного на подложке из плавленого кварца, рисуется изображение в соответствии с топологией бинарного фазового элемента (рис. 4.31).

После операции проявления осаждением в вакууме на поверхность структуры с бинарным изображением наносится слой хрома толщиной 0,05 мкм. Затем используется процедура вымывания оставшегося резиста на основе ПММА вместе со слоем

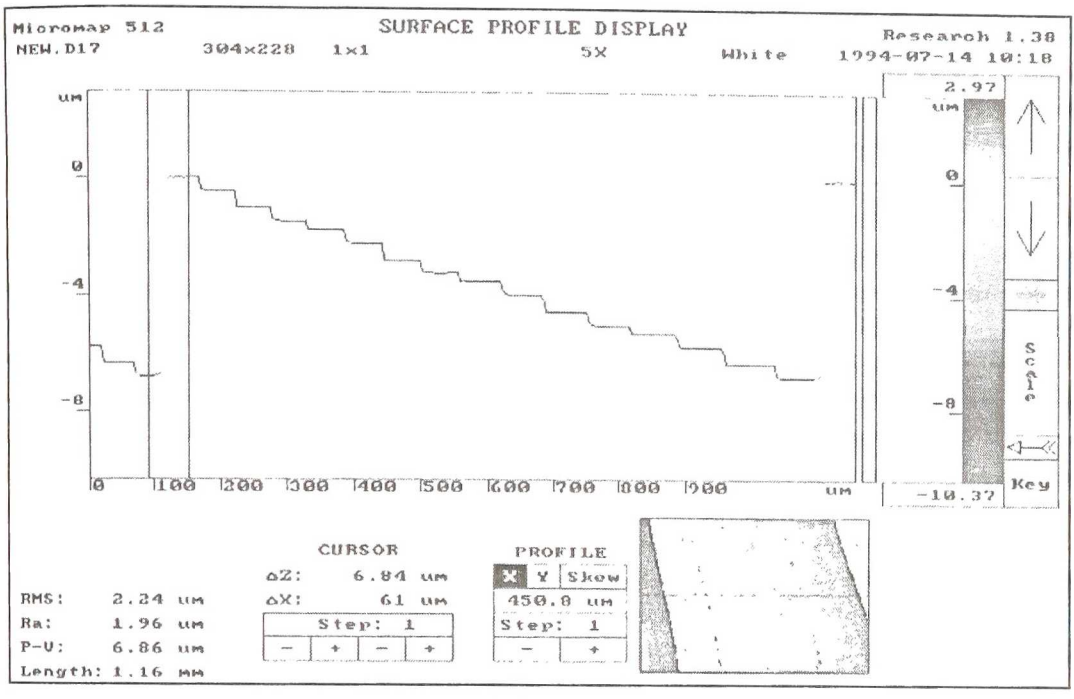


Рис. 4.29. Профилограмма микрорельефа

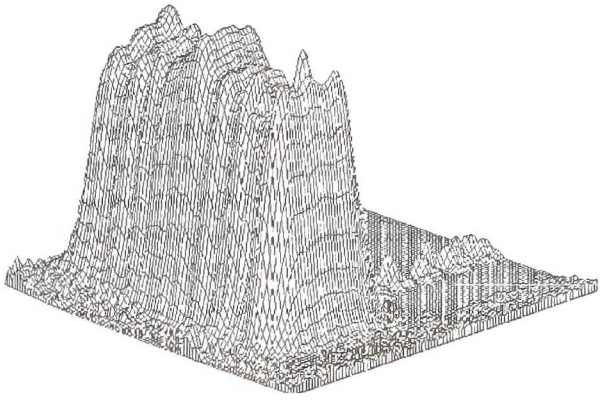


Рис. 4.30. Распределение интенсивности в фокальной плоскости фокусатора в прямоугольник

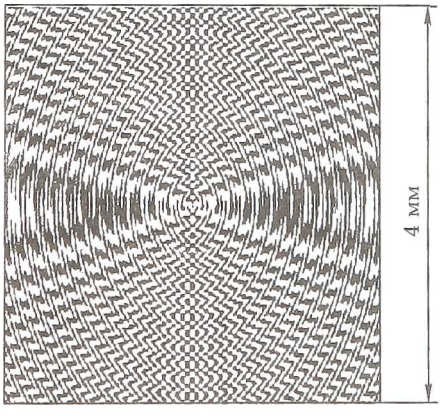


Рис. 4.31. Фрагмент фотошаблона бинарного фазового ДОЭ

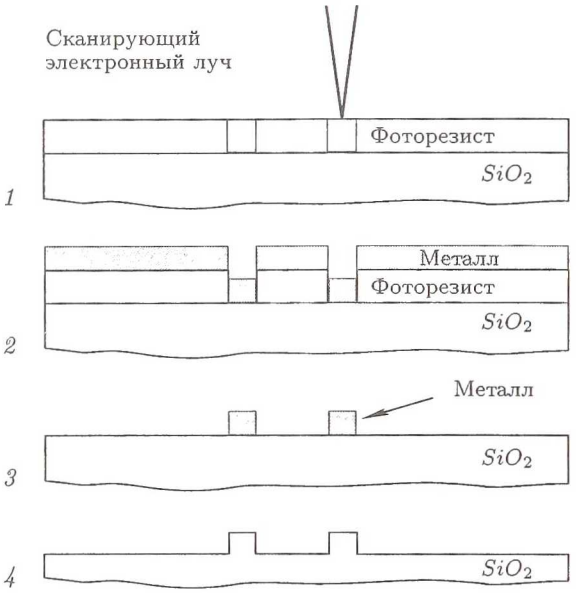


Рис. 4.32. Этапы изготовления бинарного рельефа ДОЭ: *1* — экспонирование резиста электронным пучком; *2* — осаждение хрома и вымывание резиста; *3* — реактивное ионное травление

хрома из областей, не подвергшихся экспонированию. И, наконец, с помощью реактивного ионного травления, при котором слой хрома выступает в качестве защитной маски, в кварце создается требуемый бинарный профиль микрорельефа.

Этапы изготовления бинарного профиля показаны на рис. 4.32.

По описанной технологии был получен ДОЭ с линейными размерами 4×4 мм (1000×1000 отсчетов) для длины волны He-Ne лазера ($\lambda = 0{,}633$ мкм) с глубиной бинарного рельефа равной $\lambda/[2(n-1)] = 0{,}675$ мкм ($n = 1{,}46$ — показатель преломления кварца).

4.4. Формирование кусочно-непрерывного микрорельефа

4.4.1. Методы формирования кусочно-непрерывного микрорельефа

С точки зрения близости параметров физической реализации ДОЭ к расчетным значениям фазовой функции, наиболее естественно создание зон с непрерывным профилем микрорельефа. В настоящее время существует большое количество методов и технологий, позволяющих получать микрорельеф с непрерывным профилем различной высоты и разрешения:

— электронно-лучевые методы травления на полиметилметакрилате;
— лазерно-стимулированное травление;
— травление полимерных материалов жестким УФ-излучением;
— использование массопереноса в полимерных средах;
— отбеливание фотоматериалов;
— проявка полутонового фоторезиста;
— вымывание задубленной желатины или материала «целлофот».

Все они характеризуются своими особенностями и проблемами. Универсального метода, позволяющего создавать непрерывный профиль микрорельефа со сложным рисунком на поверхности и высотой от долей до десятков микрон, для разных областей спектра оптического диапазона, в настоящее время нет.

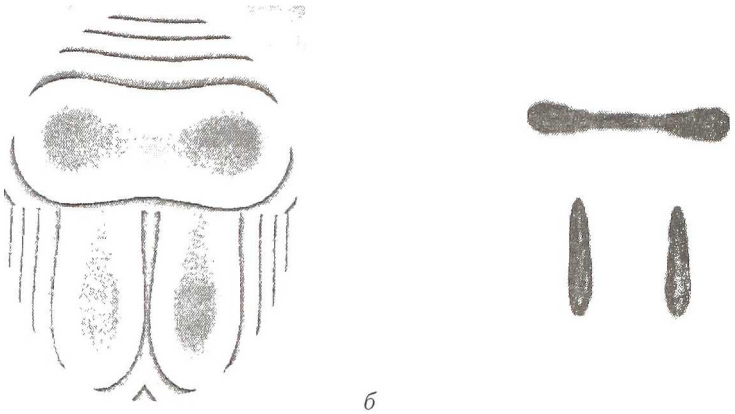
Наиболее известны и доступны способы получения непрерывного рельефа на желатиновых слоях [45–52]. Особенность использования желатина в качестве рельефообразующего материала заключается в необходимости сильного задубливания его в тех местах, где должна быть большая высота рельефа, с последующим вымыванием в горячей воде менее задубленного слоя. Существует три хорошо известных способа селективного задубливания желатина:

1) дубящее проявление;
2) дубящее отбеливание фотографических слоев с галогенидами серебра;
3) световое дубление бихромированного желатина.

Для достижения высоких пространственных частот (свыше 1000 мм$^{-1}$) при высотах рельефа порядка 0,1–0,3 мкм, используются сложные способы проявления, включая термодубление и обработку в растворе серной кислоты. Однако при высотах рельефа порядка нескольких микрон резко падает пространственное разрешение, и вследствие разной задубленности, при изготовлении пластинок, максимально достижимая высота рельефа может отличаться в 2–3 раза.

Еще недавно широко использовалась технология производства ДОЭ с использованием бихромированного желатина (БХЖ). На рис. 4.33*а* показан полутоновой фотошаблон фокусатора в букву «π» и результат фокусировки излучения CO_2-лазера [53]. Качество фокусировки (рис. 4.33*б*) позволяет говорить, в принципе, о широких возможностях фокусаторов, изготовленных на основе БХЖ. Однако, относительно простая технология формирования микрорельефа на основе БХЖ не нашла широкого применения в производстве ДОЭ, главным образом из-за шероховатости поверхности микрорельефа, снижающего эффективность ДОЭ, и ограниченной разрешающей способности, не позволяющей выполнять элементы с числовой апертурой больше 0,1.

Первые фокусирующие ДОЭ ИК-диапазона были получены с помощью негативного пленочного фоторезиста [54, 55]. Высота рельефа непрерывного профиля ограничивается высоким контрастом фоторезиста. С помощью подбора степени контрастности фотошаблонов, режимов засветки и проявления были получены элементы с высотой рельефа до 10 мкм и непрерывно меняющимся профилем для зон



а *б*

Рис. 4.33. Фотошаблон фокусатора в букву «π» (*а*) и результат его работы (*б*)

размером от 5 мм до 150 мкм. К сожалению, у пленочного фоторезиста очень большие вариации по толщине в пределах апертуры ДОЭ, имеющей величину 50–100 мм.

Светорассеяние присуще всем технологиям, основанным на вымывании полимерных материалов с различной степенью полимеризации или задубленности, что обусловлено интенсивным растворением материала в жидком проявляющем растворе. Поэтому изготовление качественных элементов по данной технологии представляет определенную трудность.

Более подробно следует остановиться на рассмотрении метода, основанного на применении жидких фотополимеризующихся композиций (ЖФПК), который используется в ИСОИ РАН, как один из основных технологических процессов изготовления ДОЭ [56–58].

ЖФПК — это композиции на основе олигомеров и мономеров органической природы, имеющие в своем составе фотоинициатор, позволяющий производить отверждение в результате реакции радикальной полимеризации под действием актиничного светового излучения. Наибольшее распространение в голографии нашли композиции на основе олигомеров уретановой природы [59]. Это недорогие доступные материалы, допускающие химический синтез с высокой чистотой состава и воспроизводимостью параметров. При отверждении они образуют сетчатые полимеры в стеклообразном или эластичном состоянии, обладающие высокой лучевой прочностью (500 МВт/см^2) для излучения наносекундной длительности и имеющие окно прозрачности 0,35–2 мкм [60].

4.4.2. Механизм рельефообразования на ЖФПК

Жидкий светочувствительный слой формируется из композиции олигоэфиракрилатов в качестве полимеризующегося соединения и изобутилового эфира бензоина в качестве фотоинициатора. Запись периодической структуры осуществляется на проекционной установке (рис. 4.34), схематично представленной на рис. 4.35. Фотошаблон 1 проецируется в УФ-свете в слой ЖФПК 5.

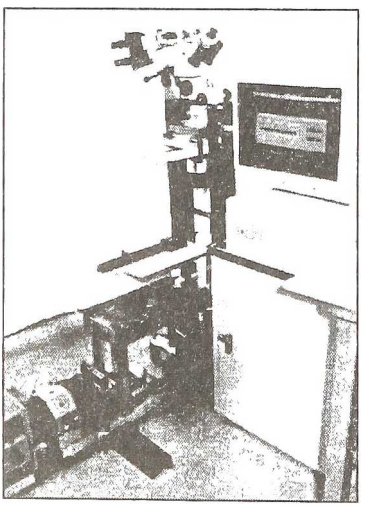


Рис. 4.34. Проекционная установка формирования микрорельефа в ЖФПК

Физику процесса записи рассмотрим с точки зрения термодинамических представлений о трехмерной полимеризации [61]. Ограничимся рассмотрением элементарной ячейки, представленной в виде контактирующих между собой областей узла и пучности светового поля (рис. 4.36). Между указанными областями осуществляется диффузионный обмен частицами.

Полимеризация инициируется в областях максимальной засветки и приводит там к снижению количества молекул олигомера и возрастанию числа молекул полимера, что вызывает появление полимерного геля.

После гелеобразования в местах засветки подвижность молекул полимера оказывается ограниченной, происходит пространственная фиксация образовавшегося полимера на ограничивающих слой поверхностях. При этом начинается процесс диффузии олигомера в полимер или наоборот. Направление диффузии определяется знаком изменения химического потенциала олигомера внутри геля. Если он отрицателен,

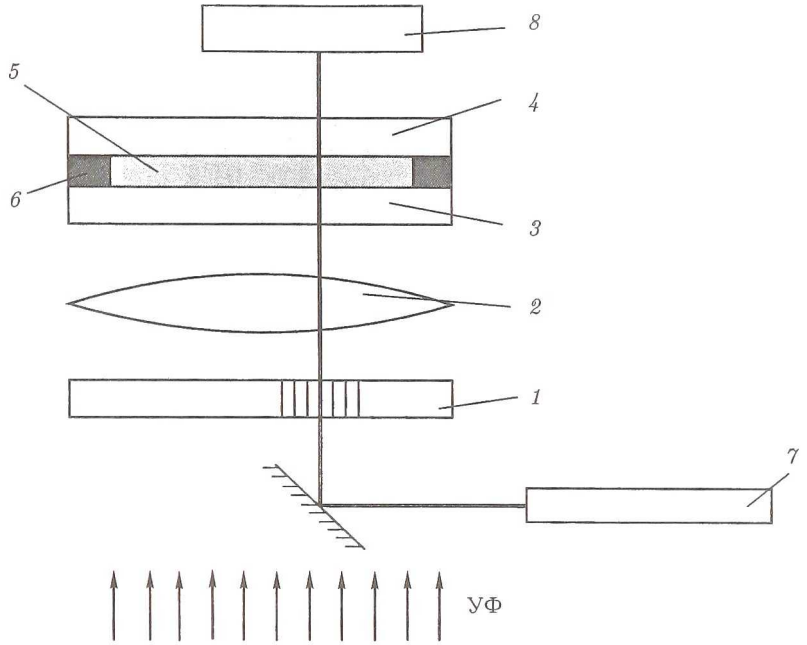


Рис. 4.35. Оптическая схема проекции полутоновых амплитудных масок в слой ЖФПК: *1* — полутоновая маска оптического элемента с дифракционной решеткой; *2* — объектив; *3* — гидрофильная стеклянная пластина; *4* — гидрофобная стеклянная пластина; *5* — слой ЖФПК; *6* — прокладки; *7* — лазер; *8* — измеритель мощности

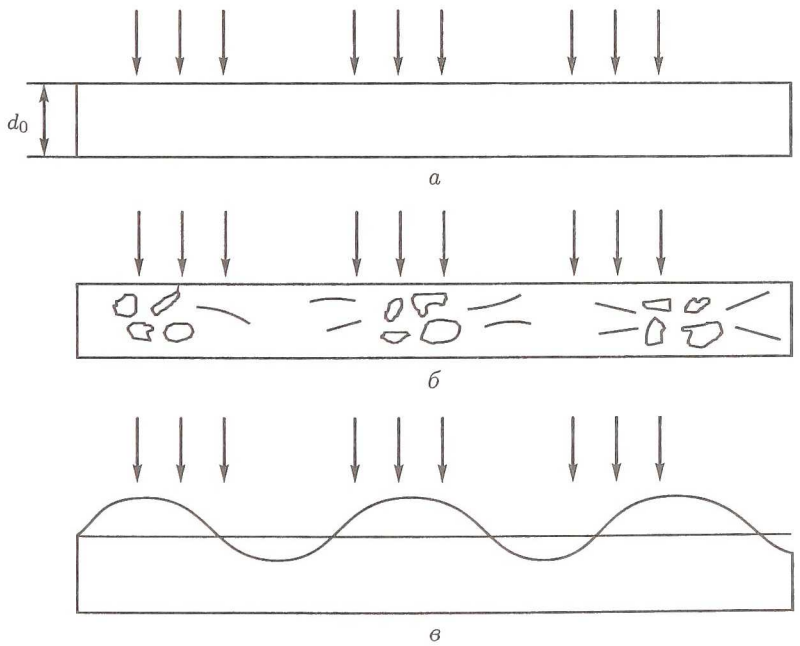


Рис. 4.36. Модель формирования рельефа: (*а*) начальная стадия, (*б*) начало массопереноса, (*в*) окончание процесса формирования

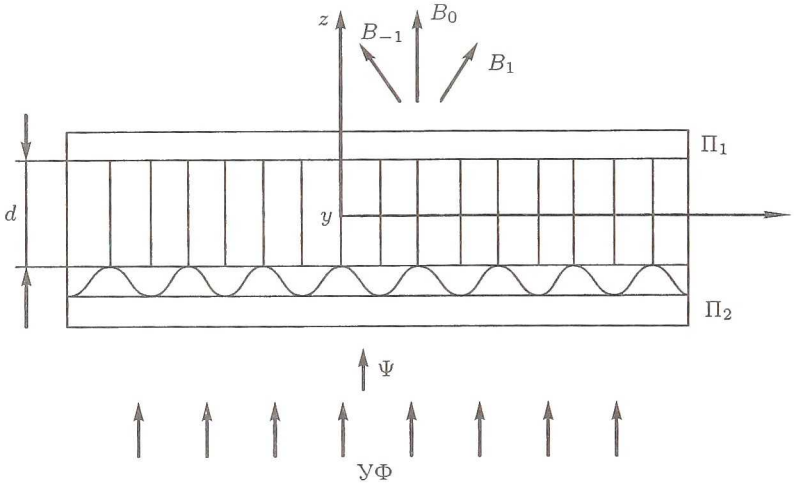


Рис. 4.37. Процесс контроля рельефообразования

олигомер диффундирует в полимер, если положителен — олигомер вытесняется из геля. В случае взаимной компенсации всех изменений свободной энергии процесс массопереноса отсутствует. Т.е. определяющим фактором является характер взаимодействия олигомера с собственным полимером. Процесс рельефообразования усиливается при использовании олигомеров с хорошим термодинамическим сродством [61] к собственным полимерам.

Таким образом, экспонирование проводится до такой степени конверсии первоначально жидкого олигомера, когда он превращается в мягкий резиноподобный слой. Затем снимается верхнее стекло. На поверхности этого слоя можно сразу наблюдать рельефный рисунок экспонируемого фотошаблона. Высота рельефа после снятия верхнего стекла составляет 1/20 часть от максимально возможной высоты и имеет тенденцию к увеличению в темноте за счет эффекта массопереноса [62]. В свою очередь максимально возможная высота составляет 1/3 от толщины первоначально жидкого слоя. Поэтому полученный образец помещается в темное место на время от нескольких часов до 3 суток, в зависимости от необходимой высоты рельефа. Во время темнового роста (т.е. самопроизвольного увеличения высоты рельефа в темноте) периодически проводятся измерения рельефа на интерферометре. Светочувствительность ЖФПК к видимому свету достаточно мала, поэтому нет опасения засветить образец за время контроля. При достижении необходимой высоты производится фиксирование путем общей засветки УФ-излучением в вакууме. При этом проходит полная полимеризация, и образец приобретает свойства жесткого пластика, с которого можно снимать копии, напылять металл и т.д.

Контроль процесса рельефообразования осуществляется с помощью структуры (рис. 4.37), состоящей из двух параллельных пластин $П_1$ и $П_2$ с зазором между ними, заполненным ЖФПК. Используется излучение лазера в виде плоской волны. Прошедший световой поток образует нулевой порядок дифракции B_0 и первые порядки дифракции B_{-1}, B_1. Для создания амплитудной синусоидальной модуляции в тонком приповерхностном слое ЖФПК у внутренней поверхности пластины $П_2$ имеется дифракционная решетка, записанная рядом с полутоновым изображением ДОЭ (рис. 4.38). Максимальная и минимальная оптические плотности D_{max} и D_{min} фотошаблона фокусатора и фотошаблона дифракционной решетки соответственно равны друг другу. Таким образом, степень конверсии ЖФПК в области решетки

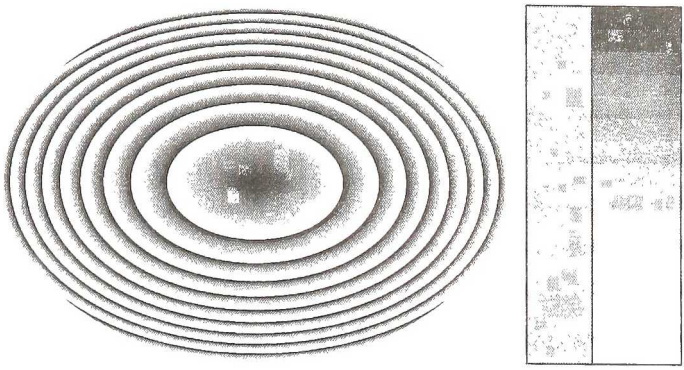


Рис. 4.38. Фотошаблон фокусатора с дифракционной решеткой

будет определять степень конверсии и в области ДОЭ. Достигая минимума нулевого порядка дифракции при экспозиции УФ, необходимо добиваться оптимальной степени конверсии ЖФПК, при которой обеспечиваются наилучшие условия для роста микрорельефа в темновом режиме. Измерение высоты микрорельефа в темновом режиме осуществляется на отражательном микроинтерферометре.

4.4.3. Определение оптимальной оптической плотности фотошаблона при рельефной записи на слоях ЖФПК

Для получения ДОЭ с минимальным рассеянием и максимально высокой передачей резкости края необходимо подобрать оптимальную оптическую плотность фотошаблона. Для выполнения поставленной задачи изготавливался фотошаблон системы дифракционных решеток с оптической плотностью, перекрывающей диапазон $D = (0{,}1\text{–}2)$, и частотами от 1 до 20 мм$^{-1}$ На фотошаблоне (эмульсионная сторона фотопластины) формировался тонкий подслой из олигомера ОКМ-2. Для уменьшения адгезии к фотоэмульсии на подслой наносился тонкий слой фторопласта, растворенного в специальном растворителе. Экспериментально была выбрана решетка с частотой 13 мм$^{-1}$. В дальнейшем перепад C оптической плотности каждой решетки измерялся лазерным сканирующим денситометром. Зависимость максимально достижимой высоты рельефа $h_{\max}$ от перепада оптической плотности решеток приведена на рис. 4.39.

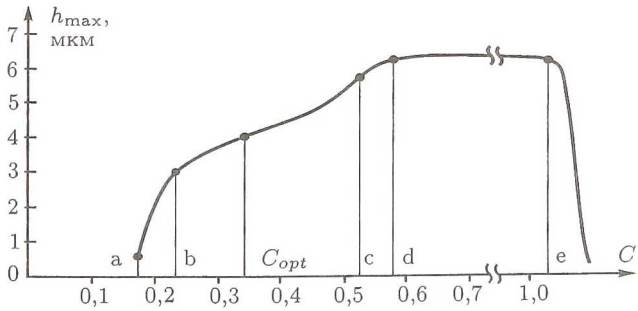


Рис. 4.39. Зависимость максимально достижимой высоты рельефа $h_{\max}$ от перепада оптической плотности изображения C

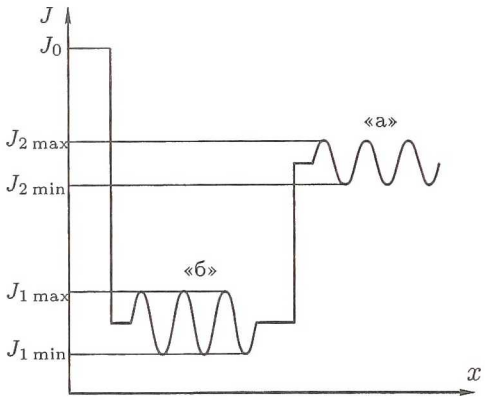


Рис. 4.40. Денситограмма двух тестовых дифракционных решеток «а» и «в» с разным контрастом, соответствующих точкам «а» и «C_{opt}» на рис. 4.39

На графике четко прослеживаются три участка. На первом участке от точки «а» до точки «b» наблюдается резкий рост максимально достижимой высоты рельефа h_{max}. Это объясняется тем, что вначале, до точки «а», когда степень контраста решеток крайне мала, имеет место низкий градиент перехода «свет-тень», поэтому массоперенос практически отсутствует. Затем, когда начинает появляться граница «свет-тень», включается механизм массопереноса (от точки «а» до точки «b»), и наблюдается резкое нарастание h_{max}. Далее от точки «b» до точки «с» идет приблизительно линейный рост h_{max} в соответствии с возрастанием градиента «свет-тень». Начиная с точки «d» появляется зона возрастания шумов и прекращение роста h_{max} за счет того, что образование полимера в темных местах резко снижается. Наконец, при достижении определенного граничного значения контраста, степень конверсии в темных зонах настолько низка, что после разборки системы ЖФПК остается в жидком состоянии. Поэтому после «критической» точки «е» на графике наблюдается резкий спад.

Перепад C вычислялся как разность оптических плотностей в светлых и темных местах, $C_x = D_{max} - D_{min}$, которые определялись из диаграмм, прописанных сканирующим денситометром (рис. 4.40).

Фотоприемник сканирующего денситометра регистрировал сигнал, пропорциональный мощности падающего света. Поэтому оптическая плотность в темных и светлых местах решетки вычислялась как десятичный логарифм коэффициента ослабления K светового потока J, зарегистрированного без фотопластинки: $D_{max} = \lg K_{max}$, где $K_{max} = J_0/J_{min}$; $D_{min} = \lg K_{min}$, где $K_{min} = J_0/J_{max}$, J_{min} — уровень света в темных местах решетки, J_{max} — уровень света в светлых местах решетки.

С точки зрения оптимальности по двум критериям — достижения максимальной высоты рельефа при минимальной толщине слоя, с одной стороны, и достижения максимального отношения сигнал/шум, с другой стороны — на рис. 4.39 экспериментально была выбрана точка $C_{opt} = D_{max} - D_{min} = 0{,}79 - 0{,}45 = 0{,}34$.

Левее этой точки снижаются шумы, но падает h_{max}, правее — h_{max} возрастает, но растут и шумы. Эффект рельефообразования, в основном, реализуется во время темнового роста. Высота рельефа в темновом режиме может вырасти в 10–15 раз по сравнению с первоначально зарегистрированной высотой после экспозиции. Для исследования поведения рельефной составляющей в темновом режиме была выбрана решетка с перепадом $C_{opt} = 0{,}34$.

4.4.4. Изготовление опытных образцов оптических элементов видимого диапазона

Проблема создания слоя фотополимера равной толщины и проецирования изображения маски в фотополимерный слой при изготовлении ДОЭ с апертурой от 40 до 100 мм была решена с помощью контактного метода. Структура многоуровневого сандвича для формирования микрорельефа ДОЭ представлена на рис. 4.41.

Для обеспечения жесткости конструкции и устойчивости ее к деформациям (т.к. на последнем этапе производится разделение сандвича, сопровождающееся механическими нагрузками), к фотопластине с фотоэмульсионным слоем приклеивалось толстое стекло с помощью фоточувствительной композиции толщиной порядка 10 мкм. Отверждение фоточувствительной композиции производилось с помощью парортутной лампы среднего давления. Расстояние до облучающей лампы выбиралось с таким расчетом, чтобы не было значительного нагрева склеиваемых стекол. Время склейки под УФ-облучением длилось не менее 2-х часов. Перед склейкой обратная сторона фотопластинки и склеиваемая сторона силового стекла тщательно обезжиривались и обрабатывались адгезивом.

Следующая очень ответственная операция — формирование, так называемого, жесткого полимерного подслоя на поверхности фоточувствительного слоя фотопластинки. Для обеспечения требуемого оптического качества ДОЭ необходимо обеспечить максимально высокую плоскостность подслоя. Исходя из этого, для формирования подслоя, был выбран олигомер с минимальной усадкой из имеющихся в наличии олигомеров. С точки зрения обеспечения максимальной плоскостности слоя существует некоторый оптимум по толщине. С одной стороны, должны быть устранены все неровности, обусловленные оптической неплоскостностью фотопластинки и неравномерным поливом фоточувствительного слоя. С другой стороны, подслой не должен быть слишком толст, из соображений усадочности слоя и неравномерности его степени конверсии по толщине, вследствие наличия эффекта послойной полимеризации. С учетом этих факторов экспериментально получено, что наилучшая плоскостность подслоя достигается при толщинах порядка 10 мкм. При тонком подслое на этапе разделения сандвича по разделу поверхностей 7–8 (рис. 4.41) происходит отрыв подслоя от поверхности фотоэмульсии из-за недостаточной адгезии к ней.

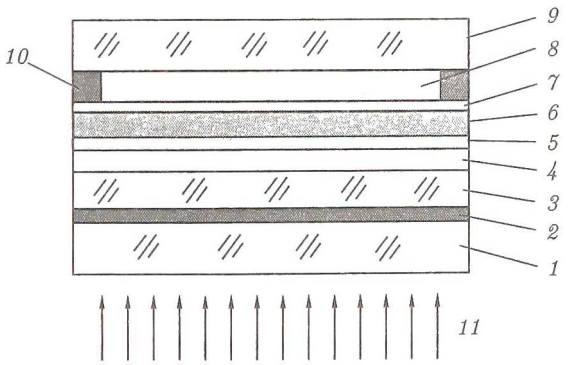


Рис. 4.41. Сандвич для выращивания микрорельефа контактным способом: *1* — толстое стекло, *2* — слой клеевой фоточувствительной композиции, *3* — подложка фотопластины, *4* — фотоэмульсионный слой, *5* — адгезионный слой, *6* — полимерный подслой, *7* — антиадгезионный слой, *8* — ЖФПК, *9* — стекло оптического качества, *10* — калиброванные прокладки, *11* — УФ-свет



а *б*

Рис. 4.42. Интерферограммы сандвичей: (*а*) сандвич, содержащий слои с 1 по 6; (*б*) с 3 по 6 (рис. 4.41)

Для увеличения адгезии подслоя к фоточувствительному желатиновому слою создавался дополнительный адгезионный слой (толщиной 0,1 мкм) на основе раствора олигомеров, имеющих высокую адгезию к олигоэфирам (из которых готовится ЖФПК), на основе изопропилового спирта. Полимеризация этого слоя велась в вакууме путем экспозиции — пароргутной лампой в течение 5 минут. На практике толщина полимерного подслоя делается равной 20–50 мкм в зависимости от апертуры ДОЭ и требований к разрешающей способности, определяемой размером крайней зоны. Контроль сформированного подслоя осуществляется на интерферометре.

Интерферограммы двух образцов сандвичей показаны на рис.4.42. Из сравнения интерферограмм видно, насколько улучшается плоскостность слоя. По кривизне полос (рис. 4.42*б*) можно вычислить, что отклонение от плоскости превышает 10 интерференционных полос на длине волны $\lambda = 0{,}628$ мкм, и составляет порядка 3 мкм. В случае, приведенном на рис. 4.42*а*, плоскость практически идеальна.

После контроля подслоя на плоскостность и получения хороших результатов, на него наносится антиадгезионный слой методом вытяжки из растворов. После высыхания антиадгезионного слоя устанавливаются калиброванные прокладки толщиной порядка 30 мкм и наносится слой ЖФПК, который накрывается пластиной, обработанной адгезивом. Далее производится экспонирование УФ светом при непрерывном контроле с помощью He-Ne лазера.

Пример 4.5. На основе технологии ЖФПК были изготовлены ДОЭ видимого диапазона. На рис. 4.43 и 4.44 показаны фотошаблоны микролинз и их профилограммы.

При толщине подслоя 50 мкм изображение зон ДОЭ «размазывается» на величину 2–3 мкм. Пройдя сквозь слой ЖФПК, изображение края деградирует еще сильнее. В нижней части ЖФПК «нерезкость» края в пределах 4 мкм. Из этого можно сделать вывод, что контактный способ проекции изображения в ЖФПК в нашем случае имеет разрешающую способность порядка 200 мм$^{-1}$.

Были изготовлены линзовые растры, предназначенные для формирования излучения с заданными углами диаграммы направленности по вертикали и горизонтали (так называемые «светорассеиватели» — см. рис. 4.45). На рис. 4.45а и 4.45б представлены светорассеиватели из растров линз с рис. 4.43а и 4.43б, соответственно. Максимальный угол отклонения излучения относительно нормали, который был получен по этой технологии, составил 20 градусов. Для первой зоны с высотой рельефа 135 мкм при длине волны $\lambda = 0{,}6$ мкм и показателе преломления $n = 1{,}5$ набег по фазе составлял порядка $112 \times (2\pi)$.

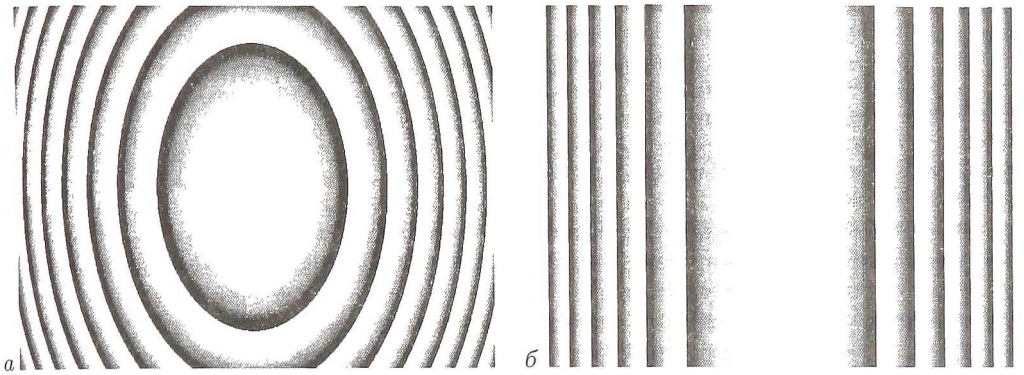
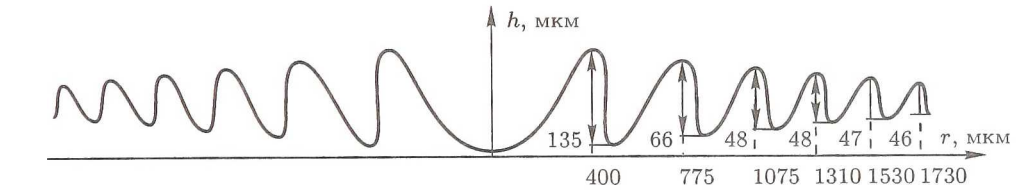
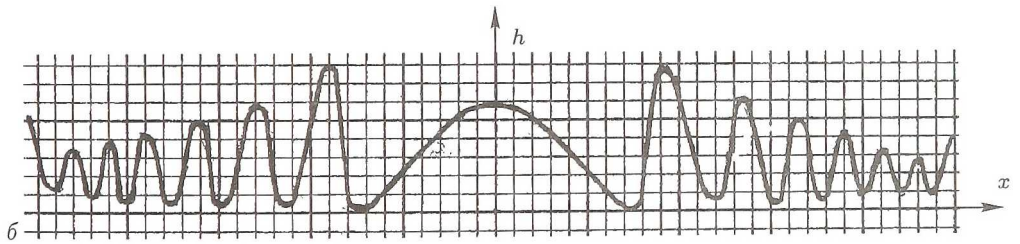


Рис. 4.43. Фотошаблоны микролинз: (*а*) сферическая, (*б*) цилиндрическая

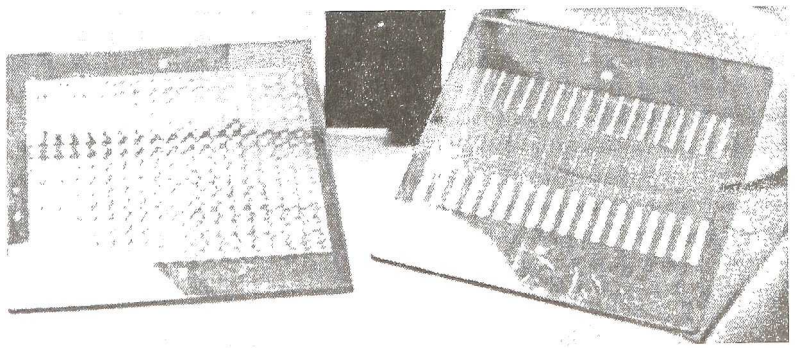


а



б

Рис. 4.44. Профилограммы микролинз с рис. 4.43 — (*а*) и (*б*), соответственно, (масштаб по вертикали — 20 мкм на клетку, по горизонтали — 50 мкм на клетку)



а *б*

Рис. 4.45. Рассеиватели света

4.5. Технология травления

Травление — критическая стадия литографического процесса. Этап переноса изображения травлением диктует следующие требования к технологии, оборудованию и материалам:

1) воспроизводимое и однородное травление структуры на одной пластине или в партии пластин;

2) незначительное удаление материала с боковых стенок формируемой структуры;

3) удаление материала при травлении пластин большой площади (> 200 мм^2) приемлемой производительности.

Независимо от того, где происходит анизотропное травление — в газовой или жидкой фазе — его эффективность зависит от типа ионов и направления электрического поля, в котором находится травитель [63, 64].

Типичный процесс травления включает в себя следующую последовательность реакций:

1) перенос реагента (перенос газа, в среде которого производится травление);

2) адсорбция реагента;

3) реакция на поверхности;

4) десорбция промежуточных и конечных продуктов реакции;

5) перенос продуктов.

На стадии травления жестко испытывается адгезия, непроницаемость, уровень плоскостности и химическая инертность резиста. Стойкость резиста к травлению и его адгезия к подложке, возможно, являются наиболее важными параметрами процесса и в наибольшей степени определяют его успех. Применение резиста с высокой стойкостью к травлению гарантирует минимальное искажение изображения при переносе его в подложку. Практические пределы применимости жидкостного химического травления (ЖХТ) определяются его разрешением (1,5–2,0 мкм) и изменением размеров при травлении $\pm$(0,2–0,5) мкм.

Контроль процесса травления требует знания его статических и динамических параметров. В случае ЖХТ важнейшими параметрами являются толщина пленки, время и температура обработки. Для получения воспроизводимых результатов необходимо четко фиксировать момент окончания травления, а также изменение химического состава реагента.

Простота метода ЖХТ позволила ему выдержать испытание временем и совершенствоваться в направлении уменьшения ширины линий и расстояний между ними. Совместное влияние внутренних эффектов близости при подтравливании и перетравливании ограничивают шаг линий, при ЖХТ металлических пленок, величиной в 6 мкм. Из этого следует, что при изготовлении многоуровневых ДОЭ для технологических лазеров с длиной волны менее 10 мкм использование ЖХТ может оказаться неэффективным.

Исходя из минимальных размеров зон ДОЭ и простоты метода, ЖХТ является приемлемым для изготовления элементов, работающих в ИК диапазоне. Для диапазона меньших длин волн, в которых линейные размеры бинарного рельефа становятся менее одного микрона, возникает необходимость в использовании метода плазменного (сухого) травления.

По физико-химическому механизму взаимодействия с поверхностью обрабатываемого материала различают три основных вида плазменного травления [65]:

— ионное;

— плазмохимическое;

— реактивное ионно-плазменное.

Сухое травление осуществляется в неизотермической низкотемпературной газоразрядной плазме низкого давления, в которой средняя энергия электронов значительно (в 10–100 раз) превышает энергию ионов, атомов, молекул и радикалов.

Важным элементом процесса сухого травления является выбор рабочих газов. Газы должны:

а) быть нетоксичными, невзрывчатыми соединениями, которые легко в газовой или паровой фазе подавать в реактор;

б) давать максимальный выход энергетических и химически активных частиц при разложении в плазме разряда, которые при травлении образуют летучие стабильные соединения;

в) не вызывать коррозии и загрязнения рабочих объемов и насосов установок.

Для плазмохимического травления материалов используются галогеносодержащие соединения в смеси с кислородом, водородом, азотом или инертными газами. При удалении органических материалов применяется кислород.

Использование технологии плазменного травления при изготовлении микрорельефа. Плазменное травление позволяет изготовить профиль в любом материале с точностью и качеством поверхности, недостижимыми для обычного жидкостного травления. Это обусловлено возможностями анизотропного режима травления материала, высокой управляемостью и стабильностью технологических процессов. Исходя из этого, представляется наиболее целесообразным использование плазмохимического травления для формирования микрорельефа ДОЭ видимого и ближнего ИК-диапазонов.

В ИСОИ РАН для плазменной обработки материалов используется установка плазменного травления (УТП ПДЭ-125-009) пластин в ВЧ разряде, которая позволяет травить линии шириной 0,1 мкм. Она обеспечивает: регулирование давления в рабочей камере в диапазоне $0{,}065 \div 1{,}3$ Па; регулировку падающей ВЧ мощности при работе на разрядную камеру в диапазоне от 100 до 700 Вт с точностью $\pm 10\%$; напуск газа по четырем каналам в широких пределах; поддержку расхода газовой смеси с высокой точностью; она снабжена системой подачи части ВЧ мощности на рабочий столик и магнитной системой стабилизации плазмы.

Прокомментируем последовательность технологических операций обработки материалов (полимера, хрома, а также кремниевых, стеклянных и кварцевых подложек).

1. Вывод установки на рабочий режим (т.е. разогрев вакуумных насосов, откачка рабочего объема до 0,001 Па, разогрев высокочастотного генератора).

2. Напуск рабочего газа (или газов), регулировка рабочего давления газа (соотношения концентраций газов); для установки давления требуется время, т.к. существует инерция системы.

3. Включение ВЧ разряда, регулировка согласования с нагрузкой ВЧ генератора. Согласование зависит от давления газа, установившегося в реакторе, а также от вида газа (существует еще очень много факторов). При необходимости регулируется ВЧ смещение на подложку или постоянное отрицательное смещение; регулируется магнитное поле.

4. Выключение подачи ВЧ мощности на индуктор (погашение плазмы).

5. Загрузка обрабатываемого материала (пластины) в шлюзовую камеру, на держатель, откачка шлюзовой камеры, ввод пластины в рабочий объем и закрепление ее на водоохлаждаемом столике.

6. Подача ВЧ мощности с ранее установленными параметрами в рабочий объем (реактор).

Рассмотрим особенности и технологические приемы плазменного травления некоторых материалов, используемых при изготовлении микрорельефа ДОЭ.

Пример 4.6. Травление хрома. Хром — в виде пленки толщиной около (0,1–0,5 мкм), используется в качестве защитной маски при травлении подложек [66]. Слои хрома подвергаются обработке в установке плазменного травления в среде паров четыреххлористого углерода (CCl_4). Хром не образует летучих двухкомпонентных соединений. Например, все галогениды имеют температуру кипения $> 1000°$, но аксихлорид хрома CrO_2Cl_2 кипит при 118°. Поэтому для увеличения скорости травления хрома в пары CCl_4 добавляется кислород O_2. Однако большое содержание кислорода приводит к сильной деградации задубленной фоторезистивной маски, скорость травления фоторезиста может превысить скорость травления хрома.

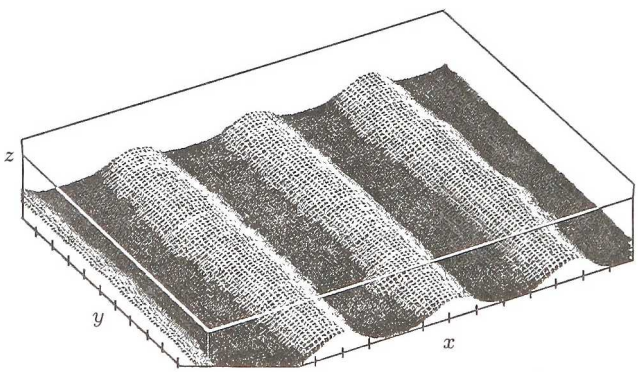
Слои хрома на подложках из стекла при толщине 0,3 мкм обладают достаточной оптической плотностью. Типичная толщина фоторезиста, задубленного при температуре 90°в течение 3 часов, составляет 1,5–2 мкм. Скорость травления хрома в парах $CCl_4 + O_2$ (соотношение 10:1) имеет среднее значение 0,0003 мкм/с, а скорость травления задубленного резиста — 0,0002 мкм/с. Такой режим травления позволяет получать один металлизированный шаблон ДОЭ за 20–30 минут травления в плазме.

Правильно выбранные режимы травления резиста и хрома позволяют получать более качественные металлизированные шаблоны (можно сравнить рис. 4.6 — металлизированный шаблон и рис. 4.5 — исходная маска на фотопленке).

Пример 4.7. Получение решеток с высоким разрешением. Для изготовления высокоэффективной решетки, работающей на отражение и делящей падающее излучение с $\lambda = 0{,}59$ мкм на пучки неодинаковой интенсивности с отклонением на большие углы, необходимо было сформировать микрорельеф с параметрами: период решетки — 0,6 мкм, высота — 0,1 мкм и угол отклонения от основного луча 45°, при отражении в среде с показателем преломления $n = 1{,}5$.

Для изготовления такого элемента приемлемым оказался кремний (полированные пластины диаметром 100 мм), обладающий высокой скоростью травления в плазме. Установка позволяет вести травление пластин кремния с диаметром до 150 мм, обеспечивая равномерность травления слоя по всей поверхности в пределах ±5%. Исходный рисунок решетки получен с помощью электронного литографа в слоях резиста толщиной порядка одного микрона, нанесенного непосредственно на полированные пластины кремния.

Для травления кремния был использован газовый травитель, состоящий из смеси CCl_2F_2 и O_2 в соотношении 5:1 соответственно. Кислород добавлялся для увеличения количества химически активных атомов фтора и хлора и подавления органических загрязнений на поверхности травления. Для увеличения скорости травления крем-



Масштаб по осям: x и y – 1000Å, z – 500Å

Рис. 4.46. Фрагмент микрорельефа решетки на кремнии

ния, покрытого маской фоторезиста, количество кислорода, добавляемого к основному газу, увеличивалось до 25%. Изображение рельефа решетки, полученное с помощью сканирующего зондового микроскопа P4-SPM-MDT, производства фирмы «Нанотехнология-МДТ» (г. Зеленоград), показано на рис. 4.46 [67].

Анализ представленного фрагмента позволяет говорить о высокой воспроизводимости линейных размеров микрорельефа решетки. Отсутствие волнообразований на поверхности дна рельефа соответствует высокому качеству и правильно подобранным режимам травления.

4.6. Формирование дифракционного рельефа с помощью лазерного микроструктурирования алмазных пленок

Рассмотренные выше широко распространенные методы формирования дифракционного микрорельефа, являясь многоэтапными и дорогостоящими, в то же время не всегда позволяют добиться требуемого результата. Сложности начинаются, например, когда необходимо получить микрорельеф на неплоской (сильно выгнутой или вогнутой) поверхности, изготовленной из труднообрабатываемого материала. Метод прямой лазерной записи микрорельефа [68–70] предлагает альтернативный, относительно недорогой, способ получения дифракционного микрорельефа на различных материалах. Использование высокоинтенсивного излучения эксимерного лазера для прямого лазерного травления микрорельефа с помощью селективного удаления (абляции) материала подложки оказывается в современных условиях высокопроизводительной и эффективной технологией изготовления ДОЭ [69, 70]. В настоящее время показана возможность получения этим методом субмикронных структур на ПММА, поликарбонате, алюминии, нержавеющей стали [70]. Экспериментальная схема для прямой записи микрорельефа с помощью эксимерного лазера [70] представлена на рис. 4.47. Устройство основано на высокопрецизионном пьезоэлектрическом перемещении подложки вдоль осей X-Y с точностью 0,05 мкм в

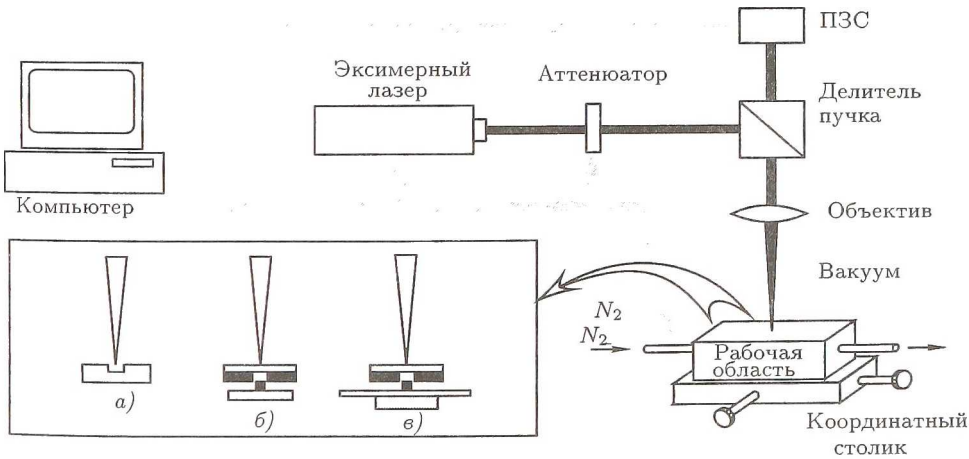


Рис. 4.47. Экспериментальная установка для микроструктурирования с помощью эксимерного лазера: (*а*) микротравление, (*б*) микрокопирование, (*в*) микрокопирование на оптическое волокно

области 25×25 мм. Лазер и рабочий перемещаемый столик управляются компьюте-

ром, рабочая область наблюдается с помощью микроскопа и ПЗС-камеры. Подложка изготавливаемого оптического элемента перемещается в плоскости X-Y с помощью пьезоэлектрических микродвигателей для формирования дифракционного микрорельефа в пошаговом растровом режиме с помощью лазерных импульсов.

На рис. 4.47 изображен режим микротравления подложки (вставка *а*), представлены режимы микрокопирования материала на подложку (вставка *б*) и волокно (вставка *в*). В настоящее время освоены режимы копирования слоев хрома, окиси индия [70], меди [71] и сверхпроводящих пленок [72] в виде структур микронных размеров, перенесенных на плоскую подложку или на искривленную поверхность оптического волокна [70], что демонстрирует уникальные возможности нового метода.

Интересные возможности открывает метод лазерного травления микрорельефа при работе с таким труднообрабатываемым материалом как алмаз — крайне перспективным для создания пропускающей оптики среднего ИК-диапазона ($\lambda =$ $= 10{,}6$ мкм). Достижения в области газофазного синтеза [73] позволяют получать поликристаллические алмазные пленки (АП) с оптическими и теплофизическими свойствами, близкими к монокристаллам алмаза (теплопроводность $\cong$18–20 КВт/см и коэффициент поглощения $\cong 5 \cdot 10^{-2}$ см$^{-1}$, показатель преломления $n = 2{,}38 \div 2{,}42$ для $\lambda = 10{,}6$ мкм). Интерес к использованию подобных алмазных пластин, толщиной до 1–2 мм и площадью до 100 см^2, в качестве выходных окон для CO_2 лазеров мощностью 10–20 кВт [74], обусловлен их более высокими порогами тепловой стабильности и разрушения, чем у традиционных материалов ИК оптики (ZnSe, KCl и др.). Однако задача создания более сложных оптических элементов оставалась нерешенной по двум основным причинам. Во-первых, механическая обработка АП с целью получения требуемого профиля затруднена из-за высокой твердости алмаза. Во-вторых, относительно малая толщина алмазных пластин не позволяет получать традиционным способом оптические элементы с достаточной апертурой. В качестве альтернативы было предложено [75, 76] использовать АП в качестве подложек ДОЭ, фазовый рельеф которых формируется методом селективного лазерного травления.

Другой проблемой алмазной оптики являются также относительно высокие потери энергии, связанные с френелевским отражением, что существенно при фокусировке излучения мощных технологических CO_2-лазеров. Отметим, что относительно высокий показатель преломления алмаза приводит к тому, что из-за потерь на отражение пропускание алмазной пластины на длине волны 10,6 мкм не превышает 71%. Поэтому особую актуальность приобретает задача просветления АП. Обычные пленочные антиотражающие покрытия значительно уступают алмазу по своим свойствам, что не позволяет использовать в полной мере уникальные свойства АП. В работах [77, 78] на результатах натурных и вычислительных экспериментов продемонстрирована эффективность антиотражающих субволновых периодических дифракционных микроструктур на поверхности алмазной пленки, созданных методом селективного лазерного травления, показано, что нанесение антиотражающих субволновых структур на обе стороны алмазной пластины позволяет увеличить пропускание более чем на 10%. На рис. 4.48 приведены результаты исследований антиотражающих структур, реализованных в Институте общей физики РАН.

Численный анализ работы антиотражающих структур в [77, 78] проводился с помощью теории эффективных сред [79]. В то же время приближение теории эффективных сред не учитывает полностью электромагнитных эффектов, возникающих при прохождении ИК-излучения через алмазную субволновую структуру. Для более точной оценки качества работы антиотражающего покрытия необходимо провести моделирование в рамках строгой электромагнитной теории [80, 81, гл. 3]. Численные

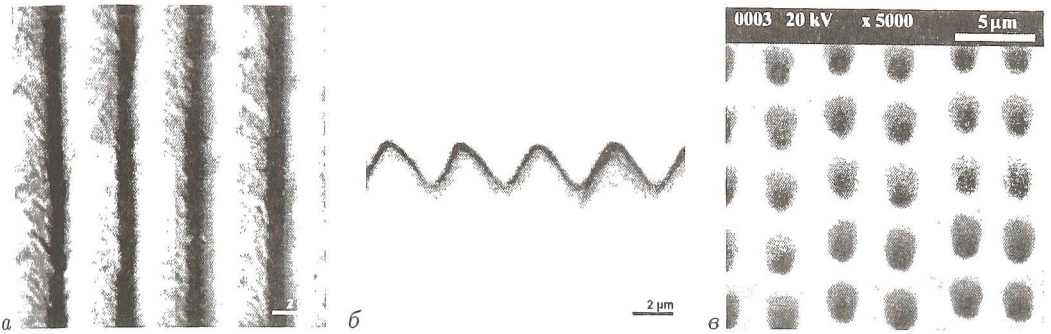


Рис. 4.48. Одномерная структура с периодом 4 мкм — типичный профиль поверхности (условия травления: энергия импульса $E = 12$ Дж/см^2, 15 импульсов на точку): (*а*) вид сверху; (*б*) сечение; (*в*) двумерная структура с периодом 4 мкм, реализованная с помощью Nd:YAG лазера: фотография поверхности ($E = 5$ Дж/см^2, 10 имп/точку)

эксперименты, результаты которых приведены в таблице 4.5, состояли в формировании волн типа H_{01} ($\lambda = 10{,}6$ мкм, с цугом в одну длину волны), падающих на границу раздела пластинка-воздух, и моделировании дальнейшего распространения прошедших и отраженных волн. В работе [78] приведены результаты натурного исследования антиотражающей структуры с периодом 3 мкм и глубиной порядка 1,8–2,0 мкм, реализованной на алмазной пластине. Максимальное увеличение пропускания для алмазной пластины, обработанной с одной стороны, достигало $\approx 6 \div 7\%$, что вполне согласуется с результатами, приведенными в таблице 4.5. Параметры разностной схемы при моделировании выбирались следующими: размеры области исследования — 500×180 мкм по Y и Z, соответственно, временной интервал — $8{,}1 \cdot 10^{-14}$ с, размер ячейки дискретизации — $1/3 \times 0{,}2$ мкм, шаг по времени — $3{,}5 \cdot 10^{-18}$ сек.

На рис. 4.49 представлено разделение падающей волны на прошедшую через границу раздела пластинка-воздух и отраженную от нее.

Результаты численного моделирования находятся в хорошем согласии с результатами натурных исследований [77, 78], и результатами численного анализа с помощью теории эффективных сред второго порядка [78, 79]. Разностное решение уравнений Максвелла впервые позволило провести моделирование алмазных антиотражающих структур (см. гл. 3) и оценить их эффективность в рамках электромагнитной

Таблица 4.5. Сравнение энергии отраженной волны при различных формах дифракционного рельефа

Номер численного эксперимента	Форма дифракционного рельефа	Диэлектрическая проницаемость пластинки (Ф/м)	Доля отраженной энергии (%)
1	без дифракционного рельефа	5,6644	17,2
2	треугольник с базисом 4 мкм и высотой 2,5 мкм	5,6644	12,5
3	треугольник с базисом 3 мкм и высотой 2,5 мкм	5,76	9,8
4	треугольник с базисом 2 мкм и высотой 2,5 мкм	5,76	9,37

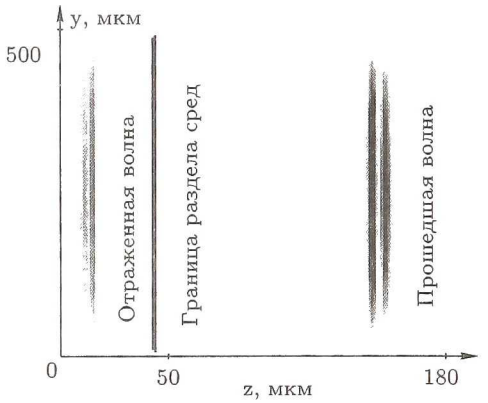


Рис. 4.49. Распределение модуля амплитуды электрической составляющей электромагнитного поля прошедшей и отраженной волны

теории, а также провести анализ тонкой структуры отраженной электромагнитной волны. Таким образом, полученные в [77, 78, гл. 3] результаты позволяют сделать вывод о эффективности реализации субволновых антиотражающих структур на алмазных пластинах. Отметим, что в общем случае реализация антиотражающих алмазных структур предъявляет более жесткие требования к точности формирования микрорельефа, чем реализация рельефа ДОЭ, рассчитанного в дифракционном скалярном приближении.

Пример 4.8. Цилиндрическая дифракционная алмазная линза. В качестве примера возможности лазерного формирования микрорельефа алмазных ДОЭ, работающих в средней ИК области спектра ($\lambda = 10{,}6$ мкм), проанализируем результаты изготовления четырехуровневой цилиндрической линзы [75, 76] с фокусным расстоянием $f = 25$ мм и апертурой $S = 4 \times 4$ мм^2, рассчитанной с использованием программного обеспечения по компьютерной оптике [3]. Уравнение фазовой функции цилиндрической линзы в параксиальном приближении имеет вид (1.15), а максимальная высота микрорельефа, определяемая по формуле (4.3), имеет значение $h_{\max} = 7{,}68$ мкм ($n = 2{,}38$). Значения ширины зон Френеля линзы можно найти из соотношения (1.12). Число полных зон j_0 удовлетворяет соотношению (1.13). Характерным параметром является ширина Z самой узкой, в данном случае – последней периферийной зоны, определяющей требования к технологическому оборудованию. В нашем случае ее значение было порядка 140 мкм. Полученные значения фазовой функции линзы (1.15), приведенные к диапазону $[0, 2\pi)$, были квантованы по 4 уровням, как показано на рис. 1.19. Рассчитанный дифракционный микрорельеф четырехуровневой линзы (рис. 4.50) воспроизводился на поверхности алмазной пластины.

Технология изготовления алмазных пластин [82] содержала следующие стадии: 1) газофазное осаждение АП на кремниевых подложках в плазменном СВЧ реакторе, 2) отделение пленки от подложки путем травления кремния в смеси кислот, 3) лазерную резку, и 4) механическую полировку поверхности АП. Пластины имели размер 0,5–1 см^2 и толщину 300–400 мкм. Для микрообработки поверхности алмаза применялся эксимерный KrF-лазер (модель 1701, длина волны 0,248 мкм, энергия в импульсе порядка 150 мДж). Сформированный на одной из поверхностей пластины рельеф ДОЭ представлял собой набор параллельных каналов шириной 40 мкм с различными глубинами. Отметим, что ширина канала выбиралась как из значения минимальной зоны линзы, так и из технологических соображений.

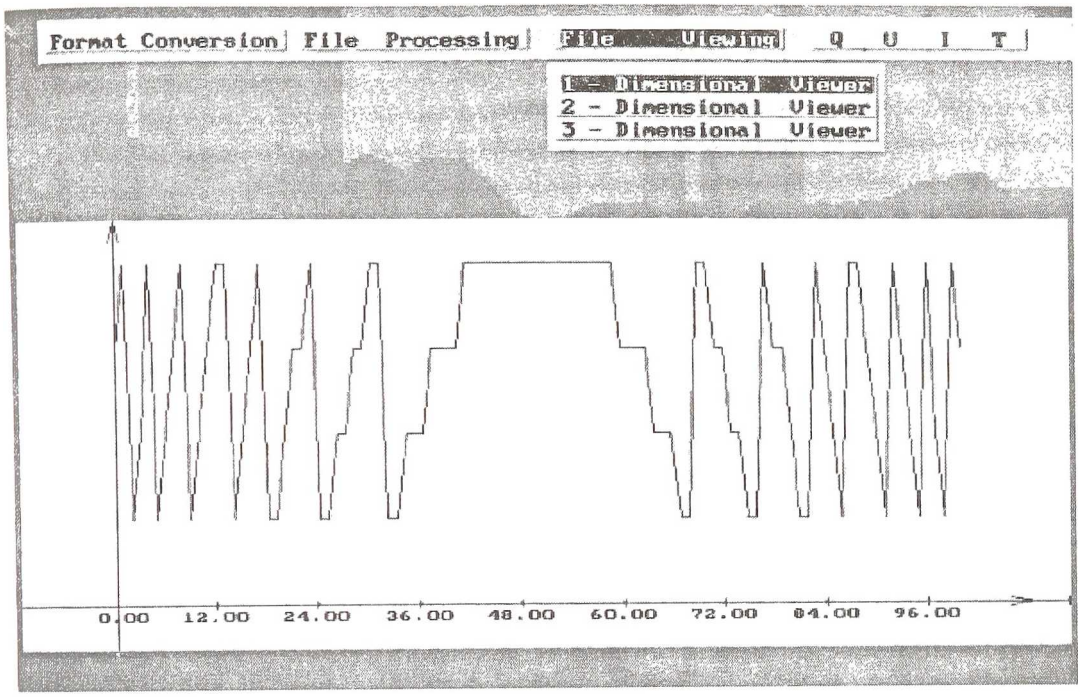


Рис. 4.50. Экран программного обеспечения в момент визуализации рассчитанного четырехуровневого микрорельефа цилиндрической линзы (один отсчет по оси абсцисс соответствует 40 мкм)

Глубина травления контролировалась энергией и количеством импульсов облучения на точку поверхности. Выбор параметров лазерного микроструктурирования поверхности алмазной пленки осуществлялся путем постановки тестовых экспериментов по реализации заданных глубин травления при максимально возможной крутизне края зоны ДОЭ. Некоторые результаты приведены на рис. 4.51.

На рис. 4.52 представлен участок профиля изготовленной алмазной цилиндрической линзы. Величина среднеквадратичного отклонения глубины сформированного рельефа от требуемого δ составила порядка 10%.

Поглощающий слой графита на поверхности алмаза, который образуется в процессе лазерной абляции [83], удалялся путем отжига в плазме дугового разряда

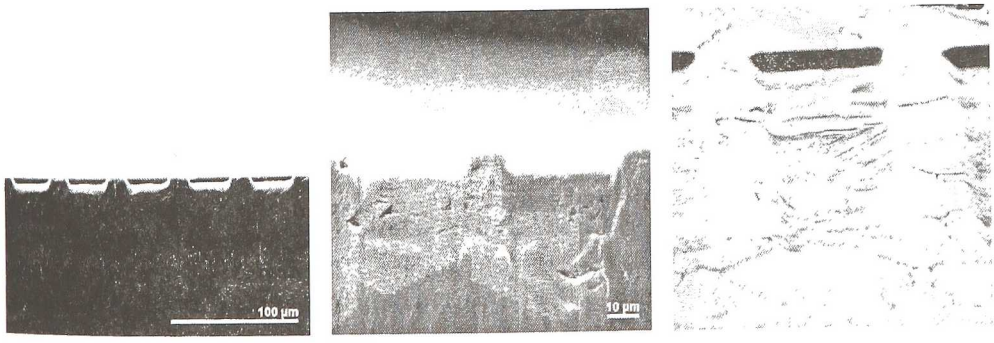


Рис. 4.51. Некоторые результаты экспериментов по выбору параметров микроструктурирования поверхности алмазной пленки

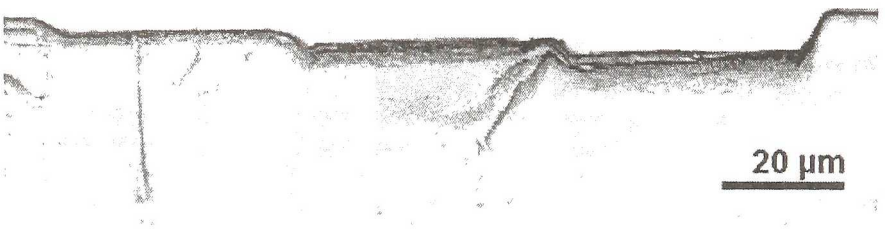


Рис. 4.52. Участок профиля изготовленной алмазной цилиндрической линзы

в атмосфере водорода. Подробный численный анализ работы дифракционной цилиндрической линзы приведен в [84]. В качестве значения дифракционной эффективности ДОЭ принималась доля прошедшей энергии, сосредоточенной в основном максимуме. Компьютерный расчет интенсивности в различных плоскостях фокальной области линзы был выполнен для основной моды TEM_{00} пучка, освещающего ДОЭ. Реализация численного моделирования выполнялась с помощью программного обеспечения по компьютерной оптике [3]. Результаты вычислительного эксперимента по оценке дифракционной эффективности элемента соответствовали данным, приведенным в [84]. При тестировании изготовленного ДОЭ использовался непрерывный CO_2 лазер мощностью 20 Вт. Модулированное обтюратором излучение фиксировалось пироприемником. Распределение плотности энергии в поперечном сечении лазерного пучка было гауссовым с радиусом $\sigma_0 = 1{,}55$ мм. На рис. 4.53 слева представлен результат взаимодействия с бумагой несфокусированного ограниченного квадратной апертурой гауссова пучка CO_2-лазера, справа — в фокусе цилиндрической линзы.

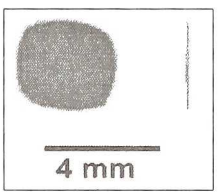


Рис. 4.53. Результат взаимодействия пучка CO_2-лазера с бумагой

В фокальной области ДОЭ распределение интенсивности $I(x)$ сохраняет квазигауссов вид (рис. 4.54*а*). Ширина преобразованного пучка σ_X определялась, как половина расстояния по оси X между двумя точками, в которых интенсивность излучения уменьшается в e^2 раз по сравнению с максимальным значением. Увеличение размера преобразованного линзой пучка $\sigma_X^{\text{эксп}} = 70{,}8$ мкм по сравнению с результатом расчета, $\sigma_X^{\text{теор}} = 65{,}8$ мкм, обусловлено несколькими причинами. Во-первых, отклонениями профиля линзы от расчетного и рассеянием излучения на поликристаллической структуре пленки. Во-вторых, наличием в лазерном пучке мод высших порядков с более высокой угловой расходимостью, чем у основной моды. Дифракционная эффективность ДОЭ, измеренная в ходе натурного эксперимента составила 78%, и близка к оценке 79–80%, полученной в ходе вычислительного эксперимента. На рис. 4.54*б* представлены зависимости $\sigma_X^{\text{эксп}}$ и $\sigma_X^{\text{теор}}$ от расстояния до плоскости элемента. Глубина фокуса определялась как расстояние между точками на оптической оси, в которых интенсивность излучения падает в 2 раза по сравнению с максимальной величиной, и составляла $\Delta f = 5$ мм. Проведенный на примере изготовления цилиндрической дифракционной линзы анализ технологии лазерного травления дифракционного микрорельефа на базе синтетических алмазных пластин продемонстрировал эффективность и перспективность нового метода. Соответствие полученных характеристик с результатами компьютерного моделиро-

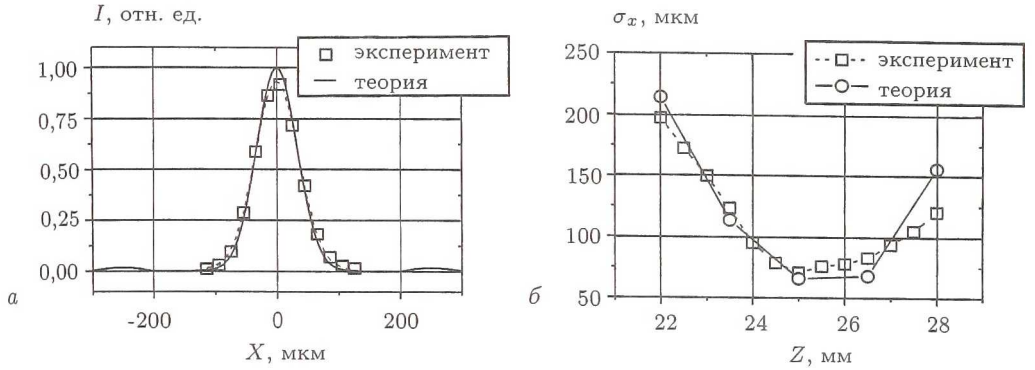


Рис. 4.54. Экспериментальные характеристики изготовленного ДОЭ и соответствующие результаты компьютерного моделирования: (*а*) распределение интенсивности $I(x)$ излучения CO_2-лазера в фокальной плоскости линзы (ось X — перпендикулярна оси симметрии цилиндрической линзы), (*б*) зависимость ширины пучка σ_X от расстояния Z до плоскости элемента. Фокусное расстояние $f = 25$ мм

вания дает основания утверждать, что лазерная обработка обеспечивает необходимую точность формирования фазового микрорельефа для создания многоуровневых ($M \geqslant 4$) алмазных ДОЭ, также как и субволновых антиотражающих структур для среднего ИК-диапазона спектра.

4.7. Копирование микрорельефа ДОЭ

Наиболее сложной и дорогостоящей процедурой является изготовление исходной рельефной матрицы ДОЭ. Далее встает вопрос тиражирования и изготовления промежуточных копий. Из существующих способов переноса фазового микрорельефа наибольшее применение нашли: перенос рельефа горячей штамповкой на полимерный материал из ПММА, копирование с помощью фотополимеризующихся композиций и гальванопластики.

Технология изготовления оптических элементов из ПММА заключается в следующем: между двумя плоскостями, одна из которых представляет собой матрицу с обратной копией ДОЭ, зажимают заготовку из ПММА и нагревают до температуры, при которой ПММА переходит из твердого в вязкое состояние. При приложении достаточного давления ПММА растекается по всей поверхности матрицы и заполняет весь свободный объем. При остывании происходит обратный процесс перехода из вязкого состояния в твердое с сохранением оптических свойств ПММА. Однако следует учесть, что при остывании происходит усадка, и преломляющие поверхности получаются с большим радиусом кривизны для линз и с меньшей глубиной модуляции фазового рельефа для дифракционного элемента. Усадка в среднем составляет 3–5%, и это необходимо учитывать при изготовлении исходных матриц.

Получение реплик с помощью фотополимеризующихся композиций может быть осуществлено при использовании целого ряда органических олигомеров. В зависимости от химического состава и способа инициирования реакций полимеризации используемых смол, существенно изменяется число технологических операций получения реплики и условия их проведения.

При копировании микрорельефа зонных пластинок и других ДОЭ можно использовать смолу олигокарбонатметакрилат — нетоксичный, дешевый, производимый

промышленным способом продукт. Повышенная чувствительность олигокарбонат-метакрилатов к действию УФ-излучения обусловлена, по-видимому, особенностями электронного строения карбонатной группы. В смолу вводится фотосенсибилизатор — метиловый эфир бензоина, образующий свободные радикалы под действием УФ-облучения. Продолжительность фотоинициируемой полимеризации смолы — 20 мин. Коэффициент преломления смолы $n = 1{,}49$.

Технология копирования микрорельефа на смолу заключается в следующем [85]: поверхность стеклянной подложки с вытравленным рельефом обрабатывают специальным раствором для создания антиадгезионного слоя. Поверхность же другой стеклянной подложки, наоборот, обрабатывают так, чтобы повысить адгезию смолы к поверхности стекла. После этого на рабочую поверхность первой подложки наносят каплю смолы и накрывают второй, следя за тем, чтобы смола растекалась равномерным слоем, покрывая весь оптический элемент, и не было пузырьков воздуха. Затем этот сандвич помещают под УФ лампу и выдерживают 15–30 мин. После окончания засветки пластины разбираются, и на одной из них остается слой полимера с копией рельефа, а другая должна быть свободной и готовой к повторению процесса копирования. При таком копировании на жидких олигомерах, сенсибилизированных к УФ, на поверхности копии наблюдается рябь в пределах $0{,}1 \div 1{,}0$ мкм в зависимости от высоты копируемого рельефа (1–10 мкм). Значительно уменьшить величину ряби можно за счет усложнения технологии экспонирования. Достигается это за счет того, что начальное экспонирование проводится через бинарный растр (например, при мощности 1 МВт/см^2 УФ на $\lambda = 0{,}365$ мкм экспонирование через растр длилось 1 минуту [86]). Далее образец экспонируется тем же источником без растра. Отделенный от стеклянной подложки слой смолы представляет собой оптический элемент, работающий на пропускание, поверхность которого представляет негативное изображение рельефа, имеющегося на стеклянной подложке. Затем, если этот рельеф соответствует рельефу оптического элемента, производится технологическая операция по нанесению отражающего покрытия. Если полученный рельеф является негативным изображением требуемого, выполняют повторное копирование и чаще всего методом гальванопластики.

Гальванопластика — это электрохимический способ изготовления разнообразных изделий, в процессе которого выделяющийся при электролизе металл достигает значительных толщин (3–5 мм) и воспроизводит форму поверхности, на которой осаждается.

В технологии создания ДОЭ этот процесс применяется для получения точных, легко отделяемых металлических копий поверхностей сложного и точного рельефа, используемых в качестве отражающей оптики или матриц для тиражирования фазовых ДОЭ.

Основными преимуществами электролитического формирования являются:

— высокая точность воспроизведения микро- и макрогеометрически сложного рельефа поверхности (формы), на которую производится электроосаждение металла;

— низкая стоимость оснастки и оборудования;

— формы для наращивания могут использоваться многократно;

— микрорельефы, снимаемые с одной формы, идентичны;

— при больших сериях производства можно одновременно изготовливать большое количество оптических элементов, которое определяется размерами применяемых гальванических ванн;

— свойства электроосажденных металлов можно изменять в широком диапазоне, изменяя режимы осаждения и используя различные добавки в электролиты;

— достигается значительная экономия металла по сравнению с копированием на смолу.

В связи с этим особое внимание уделяется таким вопросам, как конструирование форм, выбор методов подготовки поверхности перед осаждением, составов и режимов электроосаждения металлов или сплавов.

Технологическая схема процесса электролитического формования состоит из ряда последовательных операций:

1. Изготовление исходного микрорельефа оптического элемента (формы).
2. Подготовка поверхности формы (очистка, обезжиривание).
3. Нанесение разделительного слоя.
4. Нанесение на поверхность формы проводящего слоя.
5. Электроосаждение (наращивание) металла.
6. Упрочнение электроосажденного слоя.
7. Удаление механических припусков.
8. Отделение формы (удаление модели, матрицы, оправки).
9. Отделка изделия.

В качестве модели для гальванического наращивания оптических элементов используется стеклянная подложка или слой полимерного материала, на поверхности которых сформирован микрорельеф ДОЭ. Размер копируемой поверхности определяется размером активной части оптического элемента.

Металлизация поверхности осуществляется в тех случаях, когда форма изготовлена из неметалла. Основное назначение данной операции — создание токопроводящего слоя, позволяющего осаждать металл на поверхности формы. Токопроводящие слои наносят механическим, химическим путем, напылением металлов в высоком вакууме, катодным распылением и восстановлением металлов из газовой фазы.

Удобным и производительным способом получения токопроводящих слоев на формах из неметаллов является металлизация в вакууме путем конденсации паров металла на покрываемой чистой и сухой поверхности [87]. Недостатком метода является необходимость применения дорогостоящей аппаратуры.

При электролитическом формовании микрорельефа оптических элементов возникает необходимость получения толстых слоев, поэтому низкие скорости осаждения резко сокращают производительность процесса. В связи с этим выбирают электролиты, позволяющие вести процесс осаждения при высоких плотностях тока. Наиболее часто используют осаждение меди. Высокие показатели коэффициентов отражения и теплопроводности, дешевизна и доступность используемых реактивов, достаточная изученность процессов электролитического наращивания меди, предопределили выбор меди в качестве основного материала для изготовления фокусирующих оптических фазовых элементов, выполненных в виде отражающих поверхностей. Весь процесс наращивания занимает 72 часа непрерывного осаждения меди, в результате которого достигается толщина ДОЭ до 3–4 мм.

4.8. Автоматизация экспериментальных исследований и технологических испытаний ДОЭ

4.8.1. Операции и оборудование, используемые для контроля процесса создания ДОЭ

Цельное представление о современном состоянии технологий изготовления ДОЭ невозможно без анализа используемых средств автоматизированного контроля па-

раметров ДОЭ. На всех технологических этапах изготовления оптических элементов используются следующие контрольные операции.

1. Входной контроль материалов и полуфабрикатов. Контроль качества материалов подложек, металлов, неорганических продуктов (растворители, щелочи, кислоты), контроль параметров оборудования, стекол, пластин с фотоэмульсией и т.п..
2. Контроль среды, энергоносителей, оборудования. Сюда относится контроль температуры, влажности, давления, запыленности среды; контроль чистоты воды; чистоты, состава, давления и скорости потока газов; температуры рабочих зон установок и оборудования.
3. Контроль технологических операций: проверка физической и химической чистоты поверхности подложек, контроль качества изготовления фотошаблонов (на разных стадиях изготовления), контроль режимов напыления металлов и адгезии, контроль микрорельефа.
4. Экспериментальные исследования изготовленных ДОЭ.

Из рассмотренных операций при создании ДОЭ наибольшего внимания заслуживает контроль:

— исходной подложки, на которой формируется микрорельеф ДОЭ (контроль физической и химической степени чистоты подложки);

— фотошаблонов (плоскостность поверхности пластин и равенство толщин, линейные размеры, совместимость);

— параметров микрорельефа (линейные размеры, физическая чистота протравленных поверхностей, конфигурация профиля);

— параметров изготовленных ДОЭ (энергетическая эффективность; и в зависимости от назначения оптического элемента — распределение энергии в фокальной плоскости или точность создания волнового фронта, или соответствие диаграммы направленности излучения заданной, или соответствие проектируемому модовому составу и т.п.).

Перечислим и кратко охарактеризуем контрольное оборудование и средства автоматизации, используемые при изготовлении ДОЭ.

Профилометр-профилограф. Предназначен для измерения в лабораторных условиях шероховатости и волнистости поверхности изделий, сечение которых в плоскости измерения представляет прямую линию. Действие профилометра-профилографа основано на ощупывании неровностей исследуемой поверхности алмазной иглой датчика и преобразовании возникающих колебаний щупа в электрические сигналы, пропорциональные этим колебаниям. Используется для контроля качества микрорельефа оптических элементов среднего ИК-диапазона (см. рис. 4.24 или рис. 4.29), так как имеет невысокие точностные характеристики (по высоте: 0,2 мкм–500 мкм).

Микроинтерферометр. Предназначен для визуальной оценки, измерения и фотографирования высоты неровностей обработанных поверхностей а также для фотографирования поверхности и ее интерферограммы. Серийно выпускаемые в России на ЛОМО (Санкт-Петербург) микроинтерферометры Линника (МИИ-4, МИИ-10) обеспечивают измерение высот в пределах от 0,1 мкм до 8 мкм с отклонением от предела измерений 37% и 14%. В поле зрения микроинтерферометра одновременно видны интерференционные полосы и измеряемая поверхность. Если на измеряемой поверхности есть выступ или впадина, то произойдет искривление полос интерферограммы (см., например, рис. 4.21). Искривление интерференционной полосы на величину, равную расстоянию между соседними светлыми (или темными) полосами (на один период), означает изменение микрорельефа поверхности на $\lambda/2$.

Микроскоп контроля дефектов (МКД-Р) используется для визуального контроля дефектов фотошаблонов и поверхности подложек, а также микрорельефа ДОЭ.

Рефрактометр предназначен для измерения показателя преломления, линии спектра и средней дисперсии неагрессивных жидкостей и твердых тел. Точное определение показателя преломления n необходимо для правильного расчета высоты микрорельефа ДОЭ, работающего на пропускание.

Фотометр универсальный предназначен для измерения коэффициентов пропускания (или оптической плотности) твердых и жидких прозрачных (нерассеивающих) спектрально-неизбирательных сред, для измерения ряда других параметров и используется для контроля характеристик полутоновых шаблонов при использовании ЖФПК.

Устройства контроля степени чистоты поверхности пластин [88–90].

Автоматизированные устройства контроля плоскостности пластин [91–92].

Устройство для контроля высоты микрорельефа, выращиваемого в слоях ЖФПК.

Электронные и сканирующие зондовые микроскопы.

Средства автоматизации испытаний ДОЭ.

Не останавливаясь на стандартном широко распространенном оборудовании, рассмотрим подробнее новейшие средства высокоточного контроля микрорельефа, основанные на достижениях сканирующей зондовой микроскопии, и средства автоматизированного контроля параметров ДОЭ.

4.8.2. Сканирующий зондовый микроскоп в исследовании ДОЭ

Сканирующий зондовый микроскоп (СЗМ) — это настольный прибор с компьютерным управлением, который производит «ощупывание» объекта микрозондом (рис. 4.55). Причем «ощупывание» возможно как в вакууме, так и на воздухе и даже под водой. В качестве твердого зонда используется микробалка с кантилевером на свободном конце. Кантилевер — это острая игла, например, из осколков монокристалла сапфира. Изменение угла наклона балки регистрируется с помощью чувствительного датчика (лазер–фотодиод).

Изменение положения кантилевера в процессе сканирования можно измерять разными способами: иглой сканирующего туннельного микроскопа (работающего на основе регистрации туннельного тока между проводящим образцом и микроиглой),

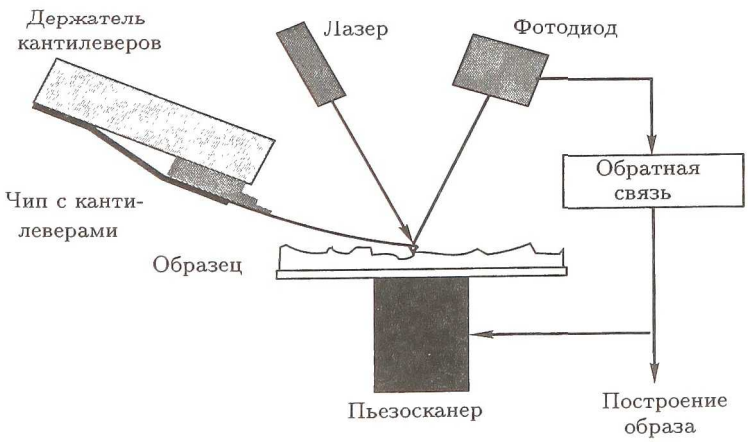
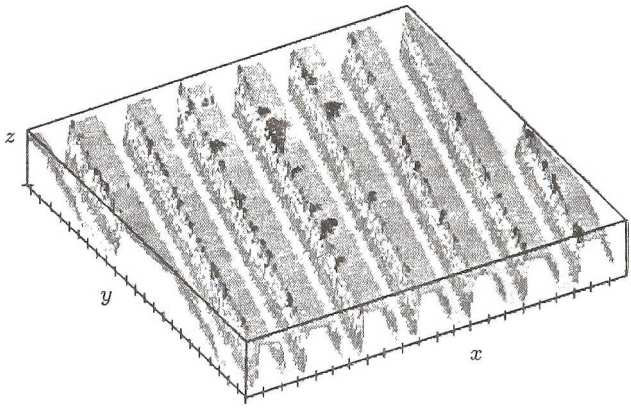
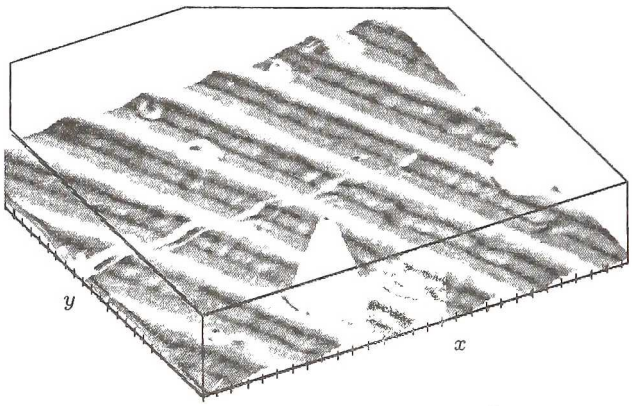


Рис. 4.55. Схема работы сканирующего зондового микроскопа [93]



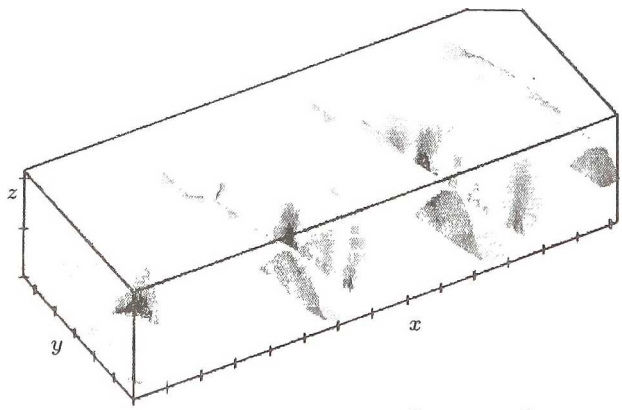
Масштаб по осям: x, y и z — 1 мкм

Рис. 4.56. Фрагмент дифракционного микрорельефа, полученный с помощью СЗМ



Масштаб по осям: x, y и z — 1000Å

Рис. 4.57. Субмикронный микрорельеф в кремнии с конусообразным дефектом



Масштаб по осям: x и y — 1000Å, z — 500Å

Рис. 4.58. Фрагмент микрорельефа с большим разрешением

иглой сканирующего силового микроскопа (работающего на основе регистрации сил межатомного взаимодействия между образцом и иглой), интерферометрически, оптико-позиционной схемой, датчиком давления (тензодатчиком). Наиболее широко распространена оптико-позиционная схема регистрации. В режиме сканирующей силовой микроскопии используется регистрация отклонений кантилевера с радиусом кривизны от 0,01 до 0,05 мкм, которая, являясь относительно несложной, позволяет регистрировать субангстремные отклонения кантилевера. При приближении иглы к образцу между ними возникает межатомное взаимодействие, приводящее к взаимному притяжению. Притяжение начинает чувствоваться кантилевером на расстояниях порядка десятков ангстрем. Сканируя поверхность и поддерживая постоянную силу притяжения, можно получить информацию о рельефе поверхности. Такой режим называется бесконтактной модой. Устойчивость достигается за счет обратной связи.

Источником света является полупроводниковый лазер ($\lambda = 0{,}65$ мкм, $P = 0{,}7$–4 мВт), луч которого фокусируется на зеркальной поверхности кантилевера в районе острия. Отраженный от кантилевера свет попадает на четырехсекционный фотодиод, усиленный разностный сигнал от которого позволяет определить угловое отклонение кантилевера с точностью до 0,1" , что обеспечивает разрешение 0,0001 мкм.

В режиме сканирующей туннельной микроскопии фиксируется ток, протекающий через диэлектрический промежуток между сверхострой иглой (радиус кривизны 0,003–0,005 мкм) и образцом, который возникает при подводе иглы на расстояние около 10 Å к поверхности образца. Для его возникновения необходимо, чтобы образец и игла были проводниками либо полупроводниками. Туннельный ток усиливается в предусилителе, установленном на том же юстировочном столике, что и игла, и попадает в цепь обратной связи, а также после АЦП становится сигналом, который регистрируется программой обработки и построения изображений сканируемой поверхности.

Общее управление и обработка полученной с помощью СЗМ информации ведется с помощью персонального компьютера. Имеющееся программное обеспечение позволяет наглядно и удобно представлять полученные изображения микрорельефа. Пример изображения фрагмента микрорельефа оптического диска, полученного с помощью СЗМ, представлен на рис. 4.56. Полученное изображение позволяет судить о качестве изготовленного микрорельефа, классифицировать и установить причины возникновения дефектов без разрезания оптического элемента. На рис. 4.57 представлен фрагмент микрорельефа субмикронной дифракционной решетки видимого диапазона с характерным конусообразным дефектом, вызванным загрязнением исходного образца пылинкой. Компьютерная обработка полученных экспериментальных данных позволяет детализировать отсканированный фрагмент (рис. 4.58).

Высокая точность и существенно меньшая цена по сравнению с электронными микроскопами (на несколько порядков), неразрушающий способ контроля позволяют классифицировать СЗМ как эффективное средство контроля при создании ДОЭ ближнего ИК, видимого диапазона и более коротких длин волн.

4.8.3. Средства автоматизации испытаний ДОЭ

Конечной технологической операцией при изготовлении ДОЭ является проведение натурных испытаний изготовленных оптических элементов с целью исследования энергетической эффективности и соответствия их реальных и проектируемых параметров. Учитывая разное функциональное назначение ДОЭ, используют различные оптические схемы, устройства ввода и программные продукты для прове-

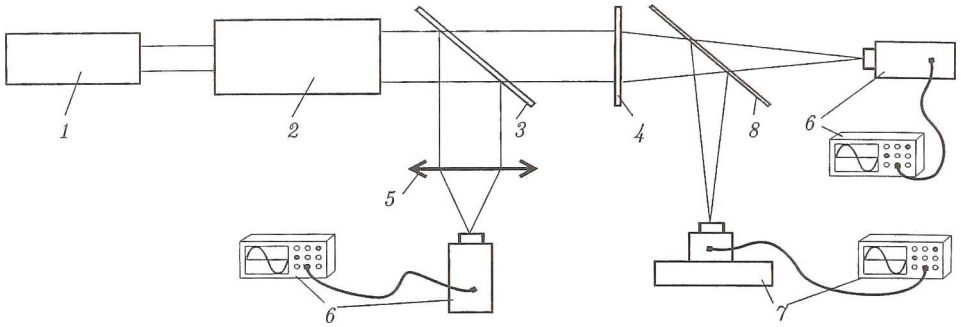


Рис. 4.59. Оптическая схема установки для исследования фокусаторов ближнего ИК-диапазона

дения испытаний. При исследовании компенсаторов можно использовать любую из схем контроля, приведенных в п. 4.8.1, только вместо контролируемой оптической поверхности необходимо использовать эталонную — тогда схема контроля выявит отклонения создаваемого компенсатором волнового фронта от заданного. Автоматизированная обработка полученных таким образом интерферограмм или теневых картин позволит сделать выводы о качестве работы компенсатора и, в случае необходимости, наметить пути устранения выявленных недостатков (совершенствование методов расчета, использование более точной технологии формирования микрорельефа и т.п.). Средства автоматизированного контроля качества формируемого модового состава рассмотрены в главе 6, а средства автоматизации испытаний светотехнических устройств — в разделе 9.6. В настоящем разделе ограничимся рассмотрением средств автоматизации испытаний ДОЭ на примере исследования фокусаторов, теория расчета которых систематически изложена в следующей главе.

На рис. 4.59 представлена оптическая схема для тестирования фокусаторов лазерного излучения ближнего ИК-диапазона. При исследовании фокусаторов другого диапазона используются соответствующие диапазону источники света и средства регистрации сфокусированного излучения, другая оснастка. На рис. 4.60 представлен общий вид установки.

Оптическая схема для тестирования фокусаторов ближнего ИК-диапазона (рис. 4.60) включает в себя: NdGr-лазер непрерывного излучения с длиной волны 1,06 мкм и максимальной мощностью 4 Вт — *1*; кварцевый коллиматор — *2*; делительное стекло — *3*; фокусатор — *4*; линзы — *5*; датчик пироваттметра — *6*; телекамеры с монитором — *7*. Информация с пироваттметров и камеры подается и

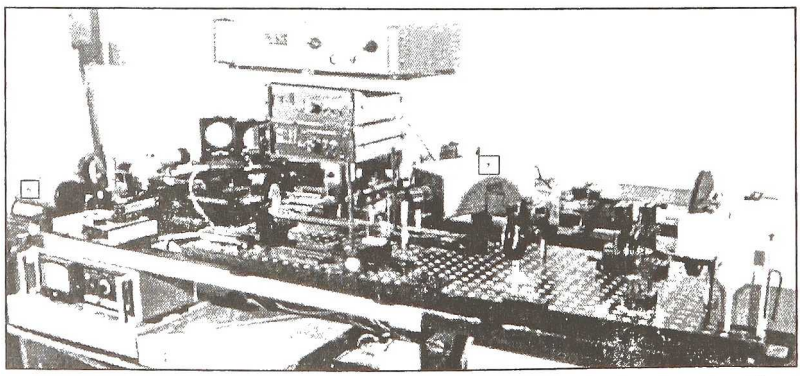


Рис. 4.60. Общий вид лазерного стенда с измерительной оптической системой

обрабатывается на персональном компьютере, управляющем также прецизионным устройством перемещения датчиков в плоскости фокусировки. Излучение лазера *1*, пройдя коллиматор *2*, падает на фокусатор *4*, который фокусирует лазерное излучение в плоскости приемника пироваттметра *6*. При этом часть излучения, отобранная делительной пластиной *3*, фокусируется объективом телекамеры *7*.

Измерения производятся в плоскостях: непосредственно перед фокусатором, для чего служат зеркало *3* и линза *5*, устанавливаемые для измерения интенсивности падающего пучка; непосредственно после фокусатора для определения степени поглощения лазерного излучения материалом подложки фокусатора; и в плоскости фокусировки лазерного излучения на датчике пироваттметра *6*. Отношение интенсивностей сфокусированного излучения к интенсивности излучения сразу же после фокусатора дает представление об эффективности работы фокусатора с точки зрения технологии изготовления микрорельефа; отношение интенсивностей сфокусированного и падающего пучка дает общую энергетическую эффективность фокусатора, учитывая поглощение излучения материалом подложки и отражение части излучения от поверхностей фокусатора. Благодаря средствам автоматизации, можно не только с большой точностью измерять интенсивность лазерного пучка, прошедшего через фокусатор, и сравнивать её с интенсивностью падающего излучения, но также анализировать распределение интенсивности в области фокусировки пучка с помощью телекамеры.

Пример 4.9. Исследование ДОЭ ближнего ИК-диапазона, предназначенных для фокусировки в кольцо и две точки. В качестве примера использования описанного стенда приведем результаты исследования фокусаторов ближнего ИК-диапазона с параметрами: рабочая длина волны $\lambda = 1{,}06$ мкм; диаметр рабочей апертуры $d = 20$ мм, фокусное расстояние $f = 150$ мм; ДОЭ были изготовлены средствами фотолитографии на основе 8-ми уровневого ступенчатого микрорельефа. Первый ДОЭ предназначен для фокусировки в кольцо диаметром 3 мм, второй — в две точки, расположенные на расстоянии 5 мм.

Распределение интенсивности в фокальной плоскости фокусатора в кольцо представлено на рис. 4.61. Дифракционная ширина линии по уровню 0,3 равна 90 мкм. По результатам измерений распределения интенсивности была произведена оценка энергетической эффективности фокусатора. Экспериментальные данные измерений, обработанные на компьютере, показали, что доля энергии, приходящейся на кольцо, составляет 82,5% от всей энергии, попадающей в плоскость фокусировки, а доля энергии, приходящаяся на две точки, — 85,7%.

Результаты фокусировки лазерного излучения, полученные с помощью телевизионной системы и обработанные на компьютере, показаны на рис. 4.62.

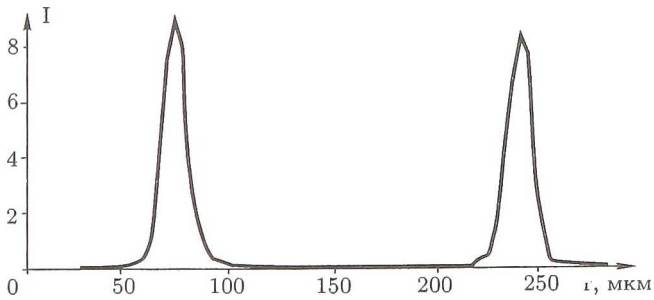


Рис. 4.61. Распределение интенсивности в сечении фокальной плоскости фокусатора в кольцо

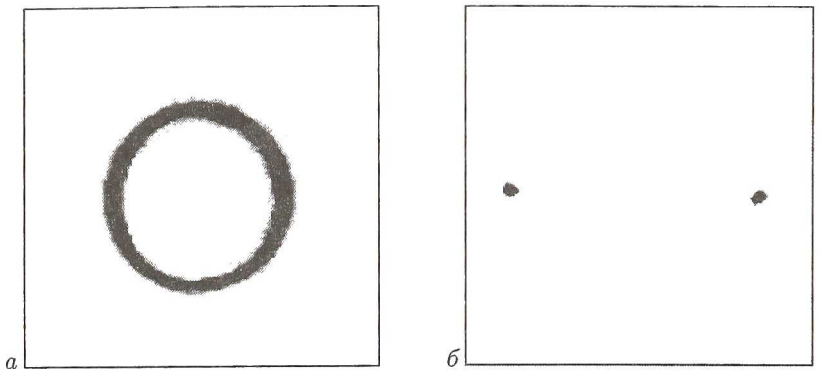


Рис. 4.62. Результаты фокусировки: (*а*) в кольцо и (*б*) в две точки

Используемые программно-аппаратные средства исследования фокусаторов включают в себя оригинальные части, позволяющие проанализировать распределение интенсивности вдоль сложной линии фокусировки (или в заданной плоской области — в случае фокусировки в плоскую область) и оценить на этой основе энергетическую эффективность ДОЭ. Для этого программно (или аппаратно) отслеживают долю энергии, попадающей в заданный (определяемый исследователем на основе данных о конфигурации области фокусировки испытываемого ДОЭ) участок фокальной области.

4.9. Примеры синтеза ДОЭ и использования программных комплексов

4.9.1. Программное обеспечение по компьютерной оптике

В качестве примера возможного оснащения компьютера для специалиста по дифракционной оптике рассмотрим несколько программных комплексов, разработанных в Институте систем обработки изображений РАН.

На рис. 4.63 представлен экран (с поясняющими комментариями) программного обеспечения «QUICK-DOE» [3, 94]. Это программное обеспечение содержит реализацию аналитических методов расчета матриц, описывающих фазовые функции ДОЭ, и методов кодирования функций пропускания оптических элементов. Программное обеспечение также оснащено средствами формирования файлов для вывода на растровые фотопостроители, расчета цифровых голограмм, выполнения ряда вспомогательных функций. В частности, к вспомогательным операциям относятся преобразование форматов файлов, визуализация маски файла, моделирование работы ДОЭ на основе быстрого преобразования Фурье (БПФ), описание и подготовка составного изображения для вывода на фотоплоттер. Программное обеспечение предназначено для использования физиками-оптиками, проектировщиками ДОЭ и программистами, разрабатывающими программы расчета ДОЭ. «QUICK-DOE» содержит программы расчета следующих аналитически описанных ДОЭ:

— различных вариантов дифракционных линз (радиальная, бифокальная, многофокусная, цилиндрическая, скрещенные цилиндрические, с повышенной глубиной фокуса — см. главу 1, а также [95–97]);

— фокусаторов (в кольцо, в крест, в контур квадрата, в отрезок в фокальной плоскости, в прямоугольную область — см. главу 5);

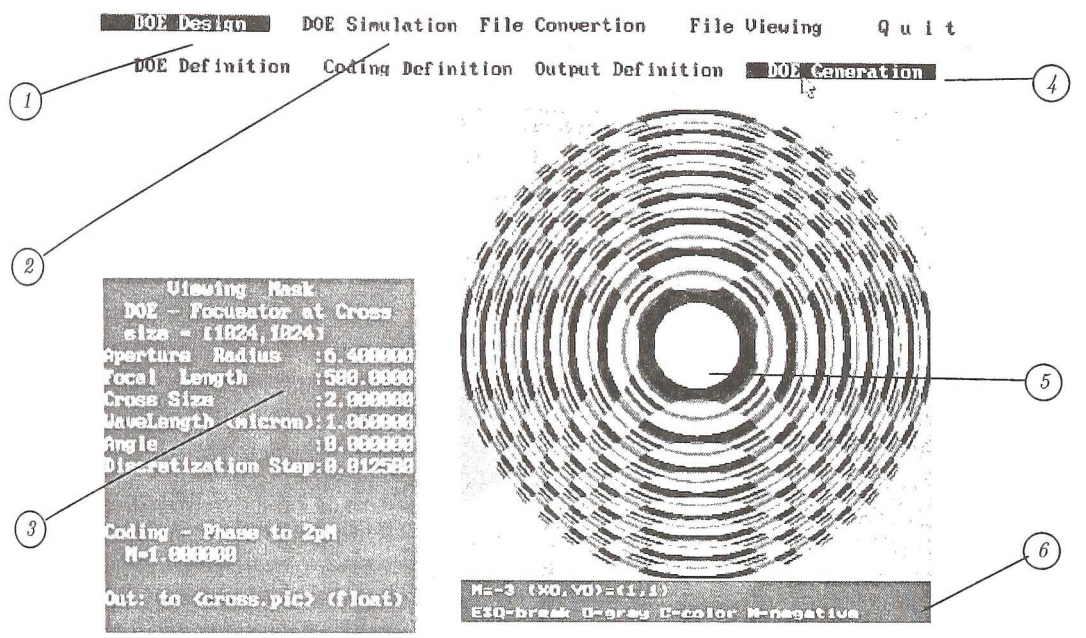


Рис. 4.63. Экран программного обеспечения «QUICK-DOE» с поясняющими комментариями: *1* — меню функций для расчета ДОЭ; *2* — меню функций моделирования ДОЭ; *3* — введенные параметры; *4* — меню функций для расчета ДОЭ; *5* — визуализация рассчитанной фазовой функции ДОЭ; *6* — параметры визуализации

— компенсаторов для преобразования сферического волнового фронта в волновой фронт с осесимметричной поверхностью вращения второго порядка (включая параболоид, эллипсоид и гиперболоид — см. главу 8);

— оптических фильтров (базисные функции Карунена–Лоэва, фильтр Винера, регуляризующий фильтр Тихонова — см. главу 10);

— ДОЭ, согласованных с модами когерентного излучения (моданов), (моды Гаусса–Эрмита и Гаусса–Лагерра — см. главу 6).

Список рассчитываемых ДОЭ может быть дополнен пользовательскими ДОЭ. Разработан и реализован несложный метод включения новых оптических элементов. Для введения нового элемента необходимо отредактировать программу проведения диалога для ввода требуемых параметров и написать программу расчета поля (фазы или амплитуды) в точке.

Программное обеспечение содержит различные варианты кодирования фазы и амплитуды:

— кодирование фазы по модулю 2π,

— кодирование из фазы в амплитуду с наложением несущей частоты,

— кодирование из амплитуды в фазу со значениями $\{0, \pi\}$,

— кодирование из амплитуды в фазу с наложением несущей частоты методом Кирка–Джонса [17, 96],

— кодирование методами Ломана и Ли [17],

— кодирование из амплитуды в амплитуду с наложением несущей частоты,

— преобразование амплитуды по абсолютной величине (амплитуда),

— преобразование амплитуды в интенсивность.

Для моделирования работы ДОЭ в приближении Френеля реализован дифракционный метод расчета, основанный на БПФ [99, 100], а для более сложных оптических схем — менее точный метод расчета трассировки лучей.

Все преобразования могут быть выполнены не только для существующих элементов, но и для элементов, включенных позже пользователем.

При расчете элемента могут быть получены полутоновые или бинарные маски. Маска может быть записана в файл в следующих вариантах:

— каждый отсчет как действительное число,

— байт на отсчет,

— бит на отсчет (только для бинарных масок).

Если нужен набор из нескольких бинарных масок (2–4), то можно рассчитать их последовательно или записать полутоновую маску в файл и затем получить все требуемые бинарные маски, используя преобразование. При расчете маски ДОЭ и изменении формата файла, выполняется синхронная визуализация.

При работе «QUICK-DOE» поддерживается внутренний, растрового, типа формат файлов, записанный построчно с комментариями, содержащимися в заголовке и в окончании файла. Для стыковки с другим программным обеспечением предусмотрены перекодировщики в TIFF-формат байт/отсчет и обратно, и в поле байт/отсчет с построчной организацией. Для облегчения работы пользователя с программным обеспечением имеется удобная оболочка. Если моделирующих средств «QUICK-DOE» не хватает, то для оценки и анализа характеристик ДОЭ возможно использование специализированного программного обеспечения «WAVE-DOE» [101]. В случае необходимости рассчитанные с помощью «QUICK-DOE» фазовые функции оптических элементов могут быть оптимизированы с помощью соответствующего программного обеспечения, реализующего описанные в главе 2 итерационные методы. Такое программное обеспечение используется также, когда нельзя аналитически получить и запрограммировать фазовую функцию ДОЭ. На рис. 4.64 представлен экран программного обеспечения «Iter-DOE», предназначенного для расчета фазовых оптических элементов (киноформов, фокусирующих ДОЭ, многофокусных линз, аксиконов, компенсаторов) на основе быстрых интегральных преобразований. Многооконный интерфейс позволяет наглядно представить все изменения фазовой функции и характеристик качества работы ДОЭ в процессе итерационного расчета.

На рис. 4.65 представлен экран программного обеспечения, основанного на итерационных методах расчета ДОЭ. Это — программное обеспечение «Iter-MODE», предназначенное для расчета ДОЭ, формирующих моды Гаусса–Лагерра, Гаусса–Эрмита и Бесселя на основе алгоритмов аппроксимации комплексной функции конечным числом базисных функций. Многооконный интерфейс позволяет наглядно представить вид фазовой маски рассчитываемого 9-модового пространственного фильтра и результат моделирования воздействия этого фильтра на освещающий пучок с соответствующим модовым составом.

Если в процессе итерационного расчета выясняется, что итерационными методами в приближении Френеля–Кирхгофа невозможно добиться решения поставленной задачи, то приходится использовать более точные методы электромагнитного подхода, изложенные в главе 3. Эти методы приходится применять также в случаях, когда нарушаются условия приближения Френеля. Программное обеспечение «GRATING SOLVER» (рис. 4.66) предназначено для расчета и моделирования дифракционных решеток в рамках электромагнитной теории.

Это программное обеспечение охватывает:

— моделирование отражающих и пропускающих дифракционных решеток с непрерывным и бинарным профилем рельефа;

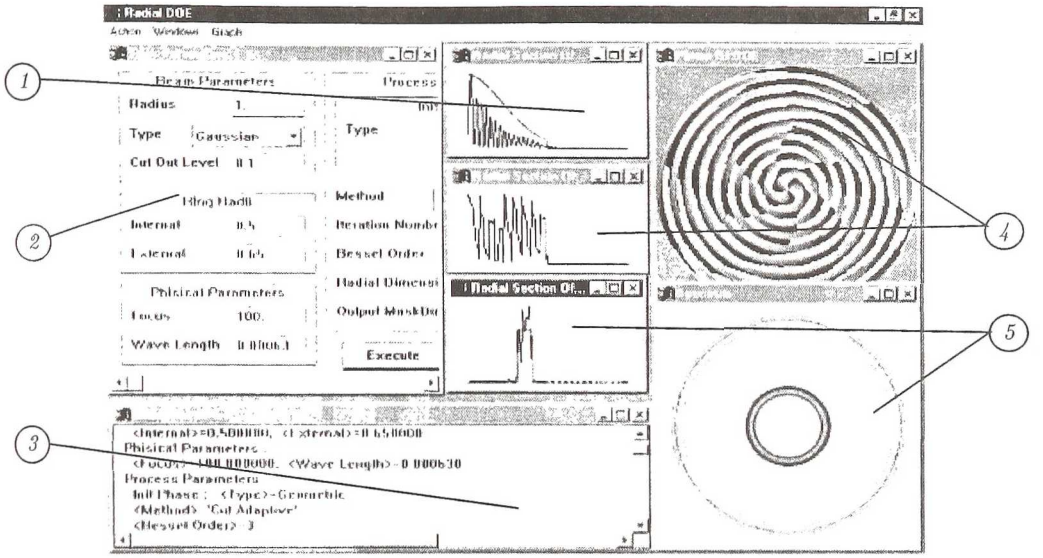


Рис. 4.64. Экран программного обеспечения «Iter-DOE» с поясняющими комментариями (на примере расчета фазового ДОЭ, фокусирующего в кольцо): *1* — радиальное сечение падающего света; *2* — окно параметров (тип падающего света, длина волны, радиусы кольца, тип начальной фазы, тип итеративного метода и т.д.); *3*– - окно протокола (входные параметры, отклонение и эффективность на каждой итерации и т.д.); *4* — рассчитанная фаза (2D-маска и радиальное сечение); *5* — распределение интенсивности в плоскости фокусировки (2D-картина и радиальное сечение)

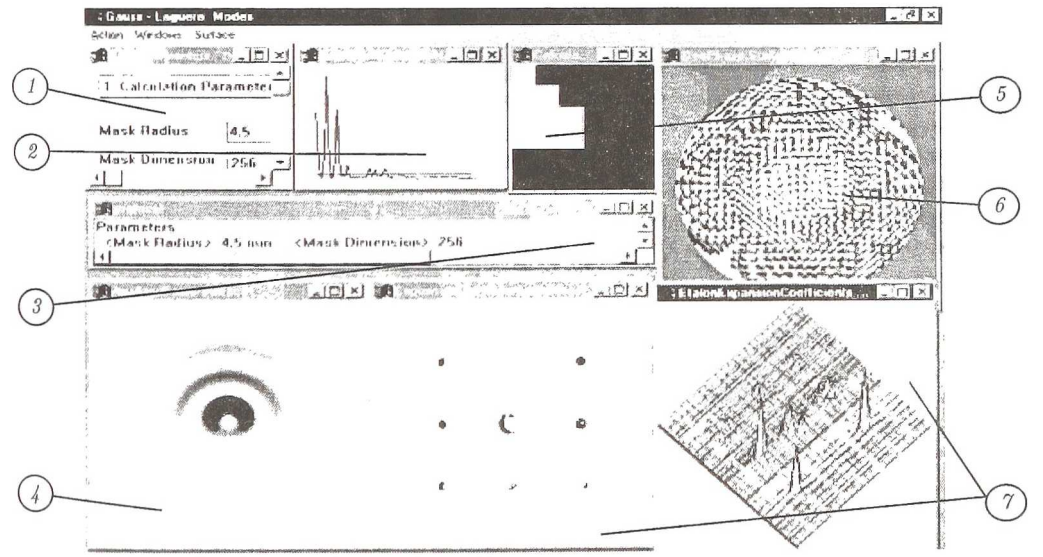


Рис. 4.65. Экран программного обеспечения «Iter-MODE» с поясняющими комментариями (на примере расчета ДОЭ, формирующего многомодовый пучок Гаусса–Лагерра): *1* — окно параметров (модовый состав светового пучка и фильтра, тип начальной фазы, тип итерационного метода, число итераций и т.д.); *2* — радиальное сечение амплитуды; *3* — окно протокола (входные параметры, отклонение и эффективность на итерациях и т.д.); *4* — 3-х модовый световой пучок; *5* — таблица модового состава пространственного фильтра; *6* — фазовая маска пространственного фильтра; — действие фильтра (всплески интенсивности в точках, соответствующих модовому составу освещающего пучка — виды 2D и 3D)

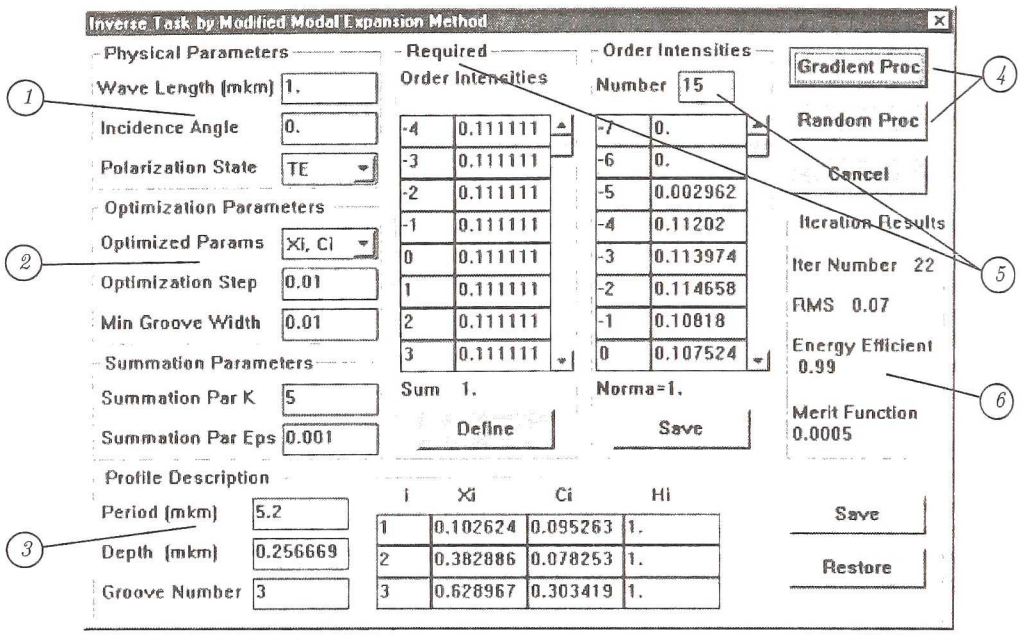


Рис. 4.66. Экран программного обеспечения «GRATING SOLVER» с поясняющими комментариями (на примере расчета 9-порядковой решетки с 3 штрихами): *1* — физические параметры задачи; *2* — параметры оптимизации профиля; *3* — окно параметров рассчитанного бинарного профиля; *4* — тип оптимизации (градиентный или случайный поиск); *5* — расчетные и требуемые интенсивности порядков; *6* — рабочие характеристики решетки (энергетическая эффективность, с.к.о., функция невязки)

— градиентные методы и стохастические алгоритмы решения обратной задачи расчета бинарного профиля из условия формирования заданных порядков дифракции;

— градиентные методы решения обратной задачи расчета непрерывного отражающего профиля в приближении Рэлея.

На рис. 4.66 представлен экран программного обеспечения «GRATING SOLVER» с поясняющими комментариями. Метод расчета основан на оптимизации координат (x_i, c_i, h_i) штрихов профиля из условия минимизации функции невязки расчетных и требуемых интенсивностей порядков.

Для моделирования и оптимизации оптических схем с ДОЭ разработано программное обеспечение «RAY-TRACING-DOE» [102], основанное на специально развитых методах расчета трассировки лучей в оптических системах с ДОЭ (глава 9).

4.9.2. Примеры синтеза ДОЭ

В настоящем разделе приведена таблица ДОЭ, изготовленных с использованием рассмотренных технологий, а также приводится сквозной пример всего процесса изготовления — от расчета до натурных испытаний работоспособности ДОЭ.

Пример 4.10. Рассмотрим синтез однопучкового модана, преобразующего нулевую моду гауссова пучка в моду Гаусса–Эрмита (1,0); рабочая длина волны $\lambda = 0{,}6328$ мкм; размер перетяжки входного пучка $\sigma_{00} = 0{,}835$ мм; диаметр ДОЭ 6,144 мм; число рассчитываемых отсчетов фазовой функции 2048×2048. Экран программного обеспечения «QUICK-DOE», позволяющего рассчитать такой оптический

Таблица 4.6. Примеры ДОЭ, изготовленных с использованием различных технологий

Конфигурация фокальной области	Длина волны, мкм	Апертура, мм	Фокусное расстояние, мм	Размеры фокальной линии	Число уровней квантования фазы	Метод изготовления
«Отрезок»*	10,6	16	250	6 мм	5	Фотолитография
«Отрезок»*	10,6	40	400	10 мм	8	Фотолитография
«Отрезок»*	10,6	50	1200	20 мм	Непрерывный микрорельеф	ЖФПК
«Отрезок»****	10,6	40	500	10 мм	8	ПХТ
«В точку»*	10,6	10	250	—	8	РТМ
«Отрезок»**	1,06	6	80	5 мм	5	Фотолитография
«Кольцо»*	10,6	50	800	10 мм	Непрерывный микрорельеф	БХЖ
«Кольцо»******	10,6	12	400	12 мм	8	РНФ
«Кольцо»**	1,06	20	100	4 мм	8	Фотолитография
«Кольцо»***	1,06	35	150	3×1 мм	2	ПХТ
«Две точки»**	1,06	20	100	3 мм	8	Фотолитография
«Две точки»***	1,06	35	150	$0 \times 1{,}5$ мм	2	ПХТ
«Площадка»***	1,06	35×35	150	1×3 мм	2	ПХТ
Буква «А»**	1,06	5	100	5 мм	2	Фотолитография
Буква «π»*	10,6	18	250	10×10 мм	Непрерывный микрорельеф	БХЖ
Линзовый** растр	0,55	5×5 100×100	3	—	2	ПХТ
Рассеиватель (дифракционная решетка)*****	0,55	1000 лин./мм	—	—	2	ЭЛЛ, ПХТ

Материал подложки:
* — Медь
** — Стекло
*** — Плавленный кварц
**** — Кремний
***** — Полиметилметакрилат
****** — Фоторезист, с алюм./серебр. пленкой

ЖПФК — жидкие фотополимеризующиеся композиции
БХЖ — бихромированный желатин
ПХТ — плазмо-химическое травление
ЭЛЛ — электронно-лучевая литография
РНФ — равномерное наращивание фоторезиста
РТМ — равномерное травление меди

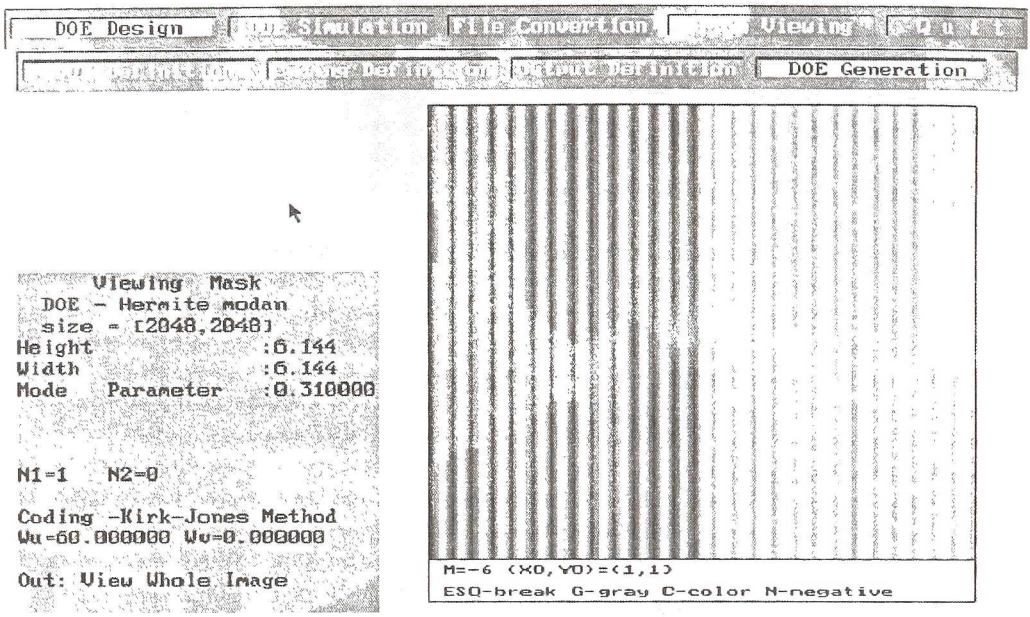


Рис. 4.67. Экран программного обеспечения «QUICK-DOE» с примером расчета модана

элемент с кодированием по методу Кирка–Джонса [98], представлен на рис. 4.67, где центральная (наиболее информативная) часть модана видна на фоне несущей.

Однако вычислительный эксперимент показал, что рассчитанный по методу Кирка–Джонса элемент не обладает достаточной эффективностью, поэтому для расчета высокоэффективного однопучкового модана были разработаны специальные итерационные методы [42]. Для изготовления рассчитанного модана использовалась технология равномерного травления резиста на основе ПММА, нанесенного на кварцевое стекло (сканированием электронным пучком на литографе ZBA-23 — см. раздел 4.3 и рис. 4.28). Для оценки фазового микрорельефа изготовленного однопучкового модана был использован микроинтерферометр MICROMAP-512. В ходе исследования выбирались наиболее характерные участки распределения фазовой функции элемента, квантованной по шестнадцати уровням. Затем прово-

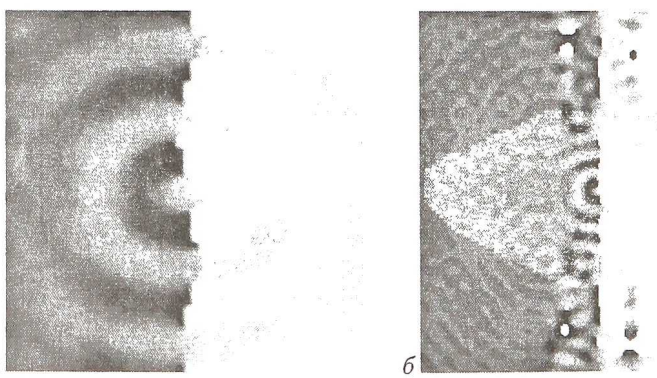
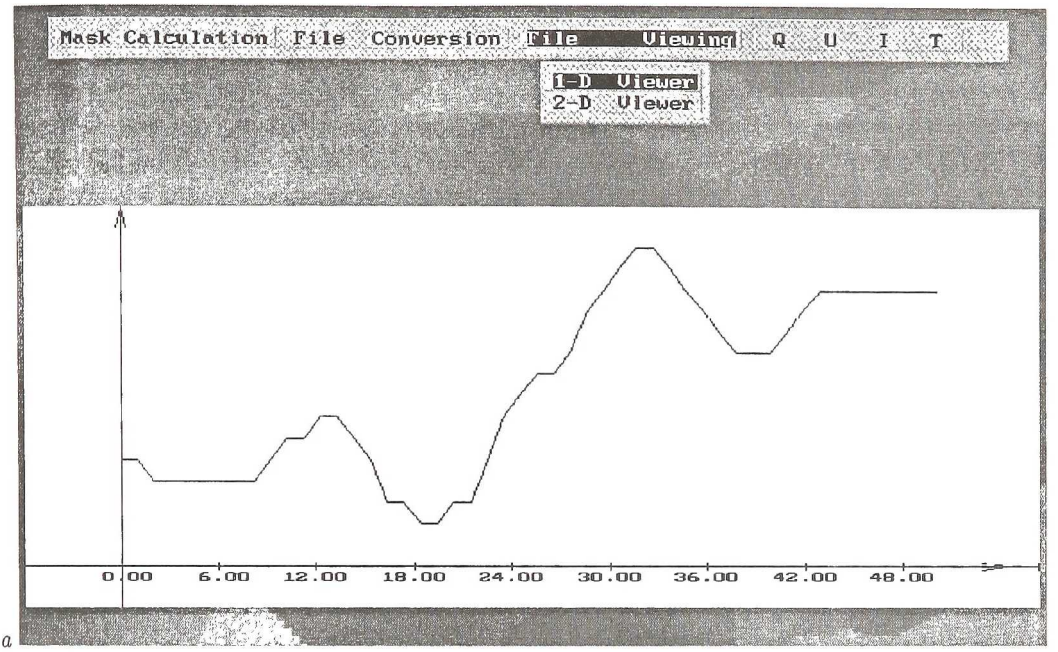
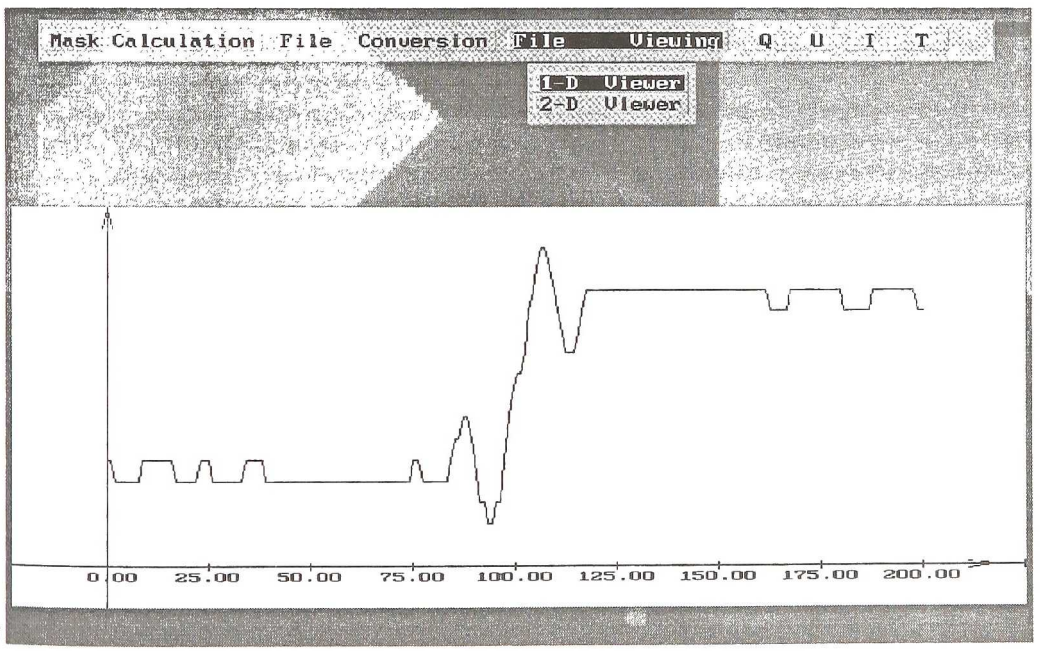


Рис. 4.68. Центральный фрагмент рассчитанной полутоновой фазовой маски модана (*а*) 50×50 пикселов и (*б*) 200×200 пикселов

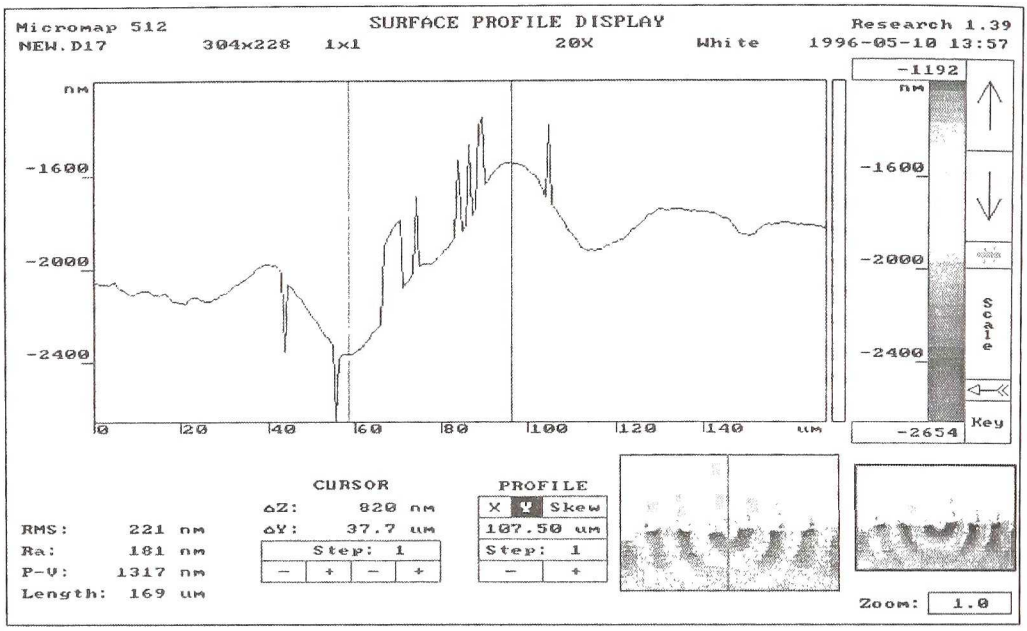


а

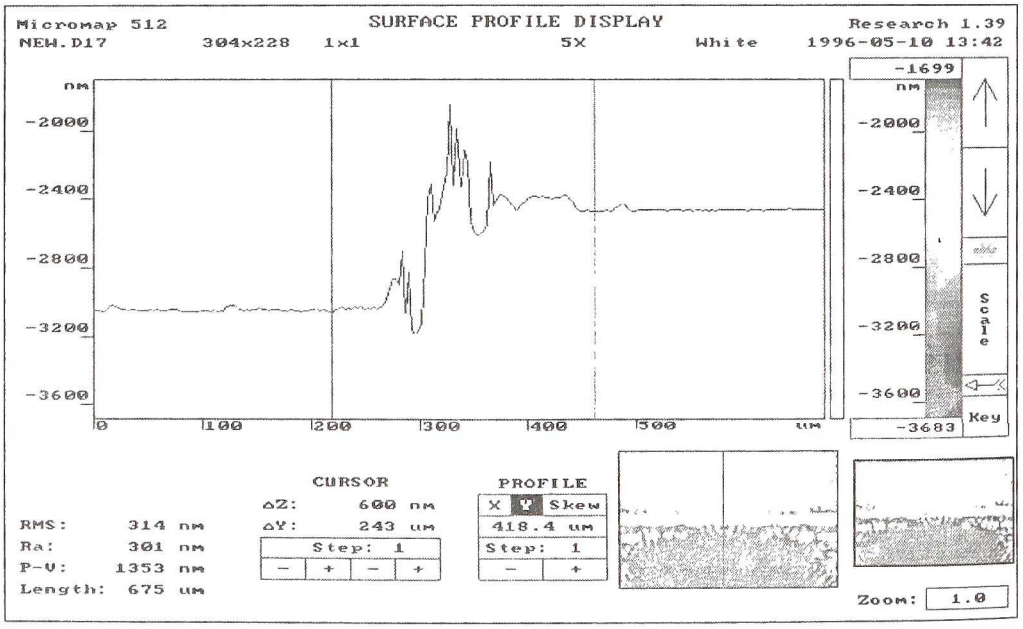


б

Рис. 4.69. Экран программного обеспечения «DOETOOL» в момент визуализации сечений фрагментов маски модана, представленных на рис. 4.68*а*, 4.68*б*



а



б

Рис. 4.70. Результат профилометрии центральных фрагментов изготовленного элемента рис. 4.68*а*, 4.68*б*

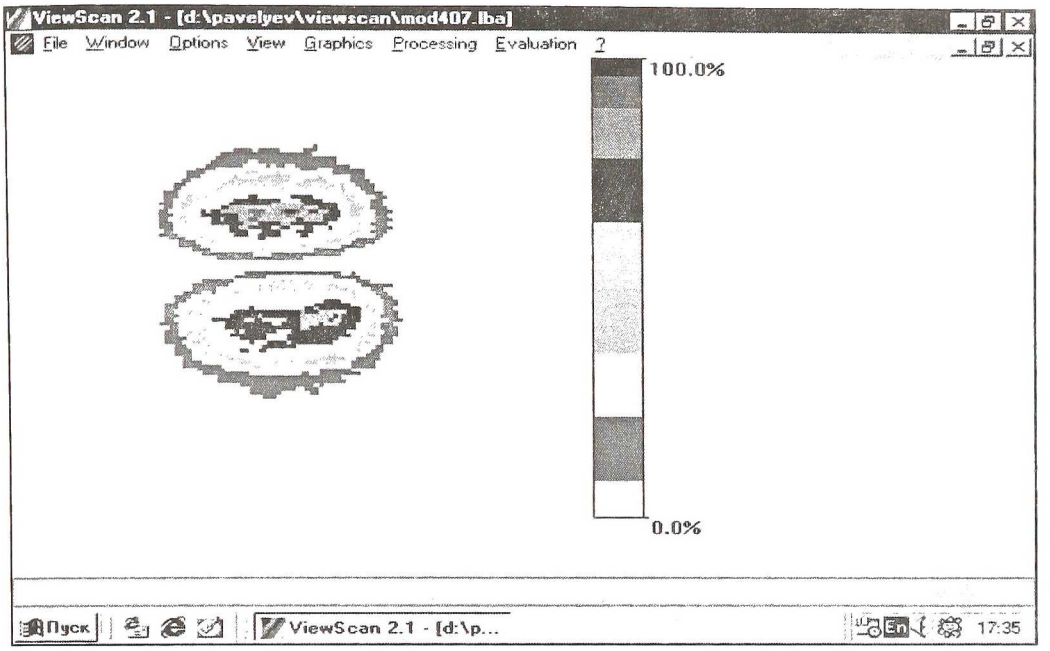


Рис. 4.71. Визуализация результата натурного эксперимента с моданом

дилось сравнение рассчитанного фазового распределения и полученного микрорельефа. На рис. 4.68*а* и 4.68*б* представлены центральные фрагменты рассчитанной полутоновой фазовой маски модана с размерами 50×50 и 200×200 пикселов, а на рис. 4.69 — их центральные сечения.

На рис. 4.70 представлены результаты профилометрии соответствующих фрагментов изготовленного элемента. «Всплески» на перепадах высоты рельефа объясняются наличием интерференционных эффектов при измерении микрорельефа.

На рис. 4.71 представлен результат измерения распределения интенсивности, формируемого изготовленным элементом [40, 42], камерой SPIRICON (длина волны излучения — 0,6328 мкм). Разрешение камеры составляло 13 мкм.

Проведенные исследования показали хорошее качество выполнения микрорельефа итерационно рассчитанных элементов с помощью технологии равномерного травления по шестнадцати уровням резиста ПММА, нанесенного на кварцевую подложку.

Примеры ДОЭ, изготовленных в ИСОИ РАН с использованием различных технологий, приведены в таблице 4.6. Оптические элементы реализованы на различных материалах (медь, стекло, плавленый кварц, кремний, ПММА, фоторезист с алюминиевой или серебряной пленкой) с использованием различных технологий (ЖФПК, бихромированный желатин, плазмо-химическое травление, электронно-лучевая литография, фотолитография, равномерное наращивание фоторезиста и др.) Для расчета и моделирования перечисленных ДОЭ использовалось программное обеспечение «QUICK-DOE».

4.10. Заключение

Проведенный анализ ряда современных технологий формирования дифракционного микрорельефа оптических элементов не может претендовать на полноту, так

как интерес, проявляемый к дифракционной оптике во всем мире, широкий диапазон изменения параметров ДОЭ и требований к микрорельефу породили большое количество оригинальных технологий, не нашедших отражения в данной книге. Проведенный анализ позволяет подчеркнуть главное — подавляющая часть технологий, технологических приемов и технологического оборудования заимствована из других областей техники (в значительной степени из микроэлектроники) и носит соответствующие специфические черты другой отрасли, подчас совершенно не характерные для дифракционной оптики. Такое заимствование можно считать оправданным, когда основное время дорогостоящее прецизионное оборудование используется по своему прямому назначению, а операции по созданию микрорельефа ДОЭ выполняются во время простоя или технологических пауз. Это позволяет существенно удешевить исследования и уменьшить себестоимость изготовления ДОЭ. Однако в последнее время появился ряд технологий (например, прямое лазерное травление микрорельефа), технологических установок (таких, как лазерные записывающие машины нового поколения), которые специально разрабатываются для нужд дифракционной оптики. Авторы в своем изложении старались соблюдать баланс между описанием широко распространенных классических и новейших технологий, подчас еще находящихся в стадии апробации и отработки.

Литература к главе 4

1. Soifer V. A., Golub M. A. *Proceedings SPIE* **1751** 140 (1992)
2. Bryndahl O. *J. Opt. Soc. Am.* **60** 140 (1970)
3. Волотовский С. Г., Голуб М. А., Досколович Л. Л., Казанский Н. Л., Павельев В. С., Серафимович П. Г., Сойфер В. А., Харитонов С. И., Царегородцев А. Е. *Компьютерная оптика* **14–15** 94 (1995)
4. Досколович Л. Л., Казанский Н. Л., Сойфер В. А. *Автометрия* **1** 58 (1993)
5. Curcio M. *Proceedings SPIE* **306** 105 (1981)
6. Clark P. P., Londono C. *Optics News* **15** 39 (1989)
7. Булатов Е. Д., Гридин С. А., Даниленко А. А. *Компьютерная оптика* **1** 167 (1987)
8. Sweeney D. W., Sommargren G. E. *Applied Optics* **34**(14) 2469 (1995)
9. Бобров С. Т., Туркевич Ю. Г. *Применение лазеров в системах преобразования, передачи и обработки информации* (Л., 1978)
10. Гуревич С. Б., Ильяшенко Н. И., Коломиец Б. Т. *Оптические методы обработки информации* (Л., 1974)
11. Грейсух Г. И., Ефименко И. М., Степанов С. А. *Оптика градиентных и дифракционных элементов* (М.:Радио и связь, 1990)
12. Грищенко В. К., Маслюк А. Ф., Гундзера С. С. *Жидкие фотополимеризующиеся композиции* (Киев:Наукова думка, 1985)
13. Boiko Yu. B., Solovjev V. S., Calixto S., Lougnot D. J. *Applied Optics.* **33**(5) 787 (1994)
14. *Введение в фотолитографию* Под ред. Лаврищева В. П. (М.:Энергия, 1997)
15. Swanson G. J. *Technical Report 854, Lincoln Laboratory. — Cambridge: Mass. Inst. of Technology* (1989)
16. Swanson G. J., Veldkamp W. B. *Optical Engineering* **28**(6) 605 (1989)
17. Ярославский Л. П., Мерзляков Н. С. *Цифровая голография* (М.:Наука, 1982)
18. Krackhard U., Schwider J., Schrader M., Stiedl N. *Optical Engineering* **32**(4) 781 (1993)
19. Golub M. A., Rybakov O. E., Uspleniev G. V., Volkov A. V., Volotovsky S. G. *Optics & Laser Technology* **27**(4) 215 (1995)
20. Коронкевич В. П., Корольков В. П., Полещук А. Г. *Автометрия* **6** 5 (1998)
21. Cherkashin V. V., Churin E. G., Burge J. H. *Proceeding SPIE* **3010** 168 (1997)
22. Валиев К. А. *Физика субмикронной литографии* (М.:Наука, 1990)
23. Rai-Choudhury *Proceeding SPIE* (1997)

24. Березин Г. Н., Никитин А. В., Сурис Р. А. *Оптические основы контактной фотолитографии* (М.:Радио и связь, 1982)
25. Watts M. P. C. *Journal Vac. Sci. Tecnology* **1** 434 (1985)
26. Kyser D., Pyle R. *IBM Journal Res. Development.* **24** 426 (1989)
27. Scot Ruska W. *Microelectronic processing* (McGraw-Hill Book Co., 1987)
28. Черняев В. Н. *Технология производства интегральных микросхем* (М.:Энергия, 1977)
29. Ефимов И. Е., Козырь И. Я. *Основы микроэлектроники* (М.: Высшая школа, 1983)
30. Спектор Б. И. *Автометрия* **6** 34 (1985)
31. Волков А. В., Казанский Н. Л., Моисеев О. Ю., Сойфер В. А. *Компьютерная оптика* **16** 12 (1996)
32. Готра З. Ю. *Технология микроэлектронных устройств. Справочник* (М.:Радио и связь, 1991)
33. Перло П., Синези С., Рипетто М., Успленьев Г. В. *Компьютерная оптика* **17** 85 (1997)
34. Корольков В. П., Малышев А. И., Никитин В. Г., Полещук А. Г. *Автометрия* **6** 27 (1998)
35. Корольков В. П., Малышев А. И., Никитин В. Г., Полещук А. Г. *Компьютерная оптика* **18** 121 (1998)
36. *Электронно-лучевая литография в изготовлении микроэлектронных приборов* Перевод с англ. под ред. Дж. Р. Брюэра. (М.:Радио и связь, 1984)
37. Моро У. *Микролитография. В 2-х частях.* Пер. с англ. (М.:Мир, 1980)
38. Вебер Э., Юрке Г. *Электроника* **50**(23) (1977)
39. Aagedal H., Teiwes S., Beth T. *Proceedings SPIE* **2404** 50 (1994)
40. Duparre M., Pavelyev V. S., Luedge B., Kley E. B., Kowarschik R., Soifer V. A. *Proceedings SPIE* **3110** 741 (1997)
41. Duparre M., Pavelyev V. S., Luedge B., Kley E. B., Kowarschik R., Soifer V. A. *Proceedings SPIE* **3134** 357 (1997)
42. Duparre M., Pavelyev V., Luedge B., Kley E., Soifer V., Kowarschik R. *Proceedings SPIE* **3291** 104 (1998)
43. Duparre M., Golub M. A., Ludge B., Pavelyev V S., Soifer V. A., Uspleniev G. V., Volotovskii S. G. *Applied Optics* **34**(14) 2489 (1995)
44. Хонина С. Н., Котляр В. В., Сойфер В. А., Хонканен М., Турунен Я. *Компьютерная оптика* **18** 24 (1998)
45. Бруй Е. Б., Корешев С. Н. *Оптика и спектроскопия* **67**(3) 685 (1989)
46. Корешев С. Н., Гиль С. В. *Оптика и спектроскопия* **68**(2) 422 (1990)
47. Гальперн А. Д., Смаев В. П., Селявко Л. В., Шелехов Н. С. *Сб. научных трудов ЛФТИ* (Л., 1988)
48. Pennington K. S., Harper J. S. *Applied Optics* **9**(7) 1643 (1970)
49. Phillips N. J., Porter D. *J. of Physics Scientific Instruments* **9**(8) 631 (1976)
50. *Справочник по лазерам* Ред. А. М. Прохорова (М.:Советское радио, 1978)
51. Красный-Адмони Л. В., Зайденберг Я. З. *Журнал научной и прикладной фотографии и кинематографии* **10**(1) 8 (1965)
52. Красный-Адмони Л. В. *Малосеребряные фотографические материалы и процессы их обработки* (Л.:Химия, 1986)
53. Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Usplenjev G. V *Optics and Lasers in Engineering* **15**(5) 311 (1991)
54. Голуб М. А., Карпеев С. В., Мурзин С. П., Овчинников К. В., Соловьев В. С., Шинкарев М. В. *Оптическая запись и обработка информации: Сборник научных трудов* 14 (1988)
55. Арефьев Е. Ю., Гилев В. А., Голуб М. А., Казанский Н. Л., Карпеев С. В., Сисакян И. Н., Сойфер В. А., Соловьев В. С., Тихонов Д. Н., Уваров Г. В. *Компьютерная оптика* **5** 49 (1989)
56. *Способ изготовления рельефно-фазовых голограмм фокусирующих элементов* Соловьев В. С., Сойфер В. А., Сисакян И. Н., Бойко Ю. Б., Гранчак В. М., Дилунг И. И. А.С. СССР № 1624864.

57. Карпеев С. В., Соловьев В. С. *Компьютерная оптика* **4** 60 (1989)
58. Соловьев В. С. *Компьютерная оптика* **10–11** 145 (1989)
59. Соловьев В. С. *Диссертация на соискание степени кандидата физико-математических наук* (Самара, 1991)
60. Boiko Y. B., Granchak V. M., Dilung I., Solovjev V. S., Sojfer V. A. *Proceedings SPIE* **1238** 253 (1990)
61. Берлин А. А., Кефели Т. Я., Королев Г. В. *Полиэфирокрилаты* (Москва, 1967)
62. Липатов Ю. С., Грищенко В. К., Гудзера С. С. *Вест. АН УССР* **2** 45 (1985)
63. Никифорова-Денисова С. Н. *Механическая и химическая обработка. Технология полупроводниковых приборов и изделий микроэлектроники* (М.:Высшая школа, 1989)
64. *Плазменная технология в производстве СБИС* Пер. с англ. под ред. Айнспрука Н., Брауна Д. (М.:Мир, 1987)
65. Киреев В. Ю., Данилин Б. С., Кузнецов В. И. *Плазмохимическое и ионно-химическое травление микроструктур* (М.:Радио и связь, 1983)
66. Волков А. В., Казанский Н. Л., Рыбаков О. Е. *Компьютерная оптика* **18** 127 (1998)
67. Волков А. В., Казанский Н. Л., Рыбаков О. Е. *Компьютерная оптика* **18** 130 (1998)
68. Vainos N. A., Mailis S., Pissadakis S., Boutsikaris L., Parmitter P., Dainty P., Hall T. *J. Applied Optics* **35** 6304 (1996)
69. Behrmann G. P., Duignan M. T. *Applied Optics* **36** 4666 (1997)
70. Mailis S., Zergioti I., Koundourakis G., Ikiades A., Patentalaki A., Papakonstantinou P., Vainos, Fotakis N. A. S. *Applied Optics* **38**(11) 2301 (1999)
71. Bohandy J., Kim B. F., Adrian F. J. *J. Appl. Phys.* **60** 1538 (1986)
72. Fogarassy E., Fuchs C., De Unamuno S., Kerherve F., Perriere J. *Mater. Manuf. Processes* **7** 31 (1992)
73. V. Ralchenko, A. Vlasov, I. Vlasov, B. Zubov, A. Nikitin, A. Khomich *Proceedings SPIE* **3484** *Int. Conf.* (Tashkent, Uzbekistan, 1998.)
74. Sussmann R.S., Brandon J.R., Coe S.E., Pickles C.S.J., Sweeney C.G., Wasenczuk A., Wort C.J.H., Dodge C.N. *Finer Points*, 10(2) 6 (1998).
75. Кононенко В. В., Конов В. И., Пименов С. М., Прохоров А. М., Павельев В. С., Сойфер В. А. *Квантовая электроника* **26**(1) 9 (1999)
76. Кононенко В. В., Конов В. И., Пименов С. М., Прохоров А. М., Казанский Н. Л., Павельев В. С., Сойфер В. А. *Компьютерная оптика* **19** (1999)
77. Kononenko T. V., Kononenko V. V., Konov V. I., Pimenov S. M., Garnov S. V., Tischenko A. V., Prokhorov A. M., Khomich A. V. *Applied Physics A* **68**(1) 99 (1999)
78. Кононенко В. В., Кононенко Т. В., Конов В. И., Пименов С. М., Гарнов С. В., Тищенко А. В., Прохоров А. М., Хомич А. В. *Квантовая электроника* **26**(2) 158 (1999)
79. Raguin D. H., Morris G. M. *Applied Optics* **32**(7) 1154 (1993)
80. Головашкин Д. Л., Павельев В. С., Сойфер В. А. *Компьютерная оптика* **19** (1999)
81. *Труды девятой межвузовской конференции* (Самара, 1999)
82. Ralchenko V. G., Smolin A. A., Konov V. I., Sergeichev K. F., Sychov I. A., Vlasov I. I., Migulin V V., Voronina S. V., Khomich A. V *Diamond and Related Materials* **6** 417 (1997)
83. Rothschild M., Arnone C., Ehrlich D. J., *J. Vac. Sci. Technol.* **4** 310 (1986)
84. Голуб М. А., Казанский Н. Л., Сойфер В. А. *Научное приборостроение* **3**(1) 9 (1993)
85. Волков А. В., Волотовский С. Г., Гранчак В. М., Казанский Н. Л., Моисеев О. Ю., Сойфер В. А., Соловьев В. С., Якуненкова Д. М. *Журнал технической физики* **65**(9) 181 (1995)
86. Соловьев В. С., Перло П. *Компьютерная оптика* **18** 130 (1998)
87. Фурман Ш. А. *Тонкие оптические покрытия* (М.:Машиностроение, 1977)
88. Бородин С. А., Волков А. В., Колпаков А. И., Рафельсон Л. Л. *Устройство контроля чистоты поверхности подложек* (М.:ПТЭ, 1990)
89. Рафельсон Л. Л., Волков А. В., Бородин С. А., Иванова В. А. *Устройство контроля чистоты поверхности подложек* А. С. №1741032 от 15.02.1992г.
90. Волков А. В., Колпаков А. И. *Способ определения чистоты поверхности подложки* А. С. №1784868 от 1.09.1992г.

91. Котляр В. В., Сойфер В. А., Храмов А. Г. *Интерферометр для контроля плоскостности отражающих поверхностей* А. С. № 1744452 от 30.06.92, Бюл.24, приоритет от 07.12.90, Класс G01B 11/24.
92. Котляр В. В., Сойфер В. А., Храмов А. Г. *Интерферометр для контроля плоскостности отражающих поверхностей* А. С. № 1760312 от 07.09.92, Бюл. 33, приоритет от 07.12.90, Класс G01B 11/24.
93. Быков В. А., Лазарев М. И., Тавров А. В. *Компьютера* **41** 38 (1997)
94. Голуб М. А., Казанский Н. Л., Шинкарев М. В. *Компьютерная оптика* **5** 43 (1989)
95. Коронкевич В. П., Пальчикова И. Г. *Автометрия* **1** 85 (1992)
96. Doskolovich L. L., Kazanskiy N. L., Soifer V. A., Tzaregorodtzev A. Ye. *Optik* **101**(2) 37 (1995)
97. Golub M. A., Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Soifer V. A. *Journal of Modern Optics* **39**(6) 1245 (1992)
98. Kirk J. P., Jones A. L. *Journal of Optical Society of America* **61**(8) 1023 (1971)
99. Сиклаш Э., Сигмен А. *ТИИЭР* **62**(3) 161 (1974)
100. Воронцов М. А., Шмальгаузен В И *Принципы адаптивной оптики* (М.:Наука, 1985)
101. Базарбаев А. А., Голуб М. А., Казанский Н. Л., Самолинова Е. Б., Сойфер В. А., Тахтаров Я. Е., Шинкарев М. В. *Комплекс программ «Анализ дифракционных характеристик элементов плоской оптики»* (Государственный фонд алгоритмов и программ. — Рег. № 50890001345).
102. Казанский Н. Л., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **14-15** 107 (1995)

ГЛАВА 5
ФОКУСАТОРЫ

5.1. Введение

Один из наиболее интересных классов ДОЭ образуют фокусаторы лазерного излучения. Фокусаторы — это ДОЭ, фокусирующие лазерное излучение в тонкие линии или малые области пространства, рассчитанные в геометрооптическом (лучевом) приближении и имеющие регулярную структуру микрорельефа. Они предложены и впервые исследованы в нашей стране в 1981 г. [1]. Трудами И.Н. Сисакяна, В.А. Сойфера, В.А. Данилова и ряда других отечественных исследователей в первой половине 80-х годов получены основные геометрооптические решения задачи фокусировки и созданы различные фокусаторы [2–8]. В последующие годы в работах [9–13] были развиты теоретические аспекты методов расчета фокусаторов, доказаны теоремы существования и разрешимости задач синтеза фокусаторов.

Фокусаторы по сравнению с фокусирующими ДОЭ, рассчитанными итерационными методами, имеют ряд особенностей. Лучевое приближение, используемое при расчете фокусаторов, дает регулярный зонированный микрорельеф. Итерационные алгоритмы, являются более точными, однако, как правило, дают нерегулярный микрорельеф. Регулярность структуры микрорельефа фокусатора снижает требования к технологии фотолитографии, используемой для производства ДОЭ. Тип микрорельефа фокусаторов также позволяет использовать более широкий круг технологий [14–18]. Например, регулярный рельеф фокусатора, в отличие от стохастического микрорельефа итерационно рассчитанного ДОЭ, может быть произведен прецизионным резанием на станках с ЧПУ [14, 15].

Особенно актуально использование фокусаторов для мощных лазеров, работающих в среднем ИК-диапазоне ($\lambda = 10{,}6$ мкм). Это связано с меньшей стоимостью и отсутствием экологически вредного воздействия производства ДОЭ по сравнению с линзовыми фокусирующими системами лазеров. Особый интерес представляет использование фокусаторов на алмазных пленках, обладающих в диапазоне среднего ИК-излучения высокой термоустойчивостью до 5 Квт/см^2.

Ключевой проблемой при создании фокусаторов является достижение высокой энергетической эффективности ($\sim 90\%$) при формировании требуемого распределения интенсивности в фокальной области. Наличие множества решений, обусловленное некорректностью обратной задачи фокусировки [19], и геометрооптический характер решений приводят к необходимости исследования характеристик фокусаторов. Возникают задачи определения границ применения геометрической оптики для расчета фокусаторов и оценки влияния на работу фокусаторов эффектов дискретизации и квантования, характерных для технологий расчета и изготовления ДОЭ [20–24]. Также актуальной является задача сравнительного анализа различных решений обратной задачи фокусировки.

Среди характеристик процессов создания и работы фокусатора, можно выделить три вида параметров. К первому виду относятся физические параметры, положенные в основу расчета фазовой функции фокусатора, — фокусное расстояние, рабочая длина волны, размеры фокусатора и области фокусировки. Ко второму виду относятся параметры дискретизации и квантования фазовой функции фокусатора, размер и форма элементов (модулей) дискретизации. Эти параметры связаны с выбором устройства регистрации ДОЭ. К третьему виду параметров относятся дифракционные характеристики фокусатора — энергетическая эффективность, ширина фокальной линии, среднеквадратичное отклонение полученного распределения интенсивности от требуемого в фокальной области и т.п. Для проектирования фокусаторов первые два вида параметров являются внутренними, а дифракционные параметры — внешними; последние получаются в результате работы фокусатора с выбранными внутренними параметрами. Для исследования фокусатора важно выявить связь внешних и внутренних параметров. Причем, учитывая трудоемкость и многовариантность процедуры изготовления ДОЭ, исследовать характеристики фокусатора необходимо уже на стадии проектирования.

К настоящему времени разработан ряд дифракционных методов расчета ДОЭ, расширяющих функциональные возможности фокусаторов. Одним из таких методов является метод нелинейного преобразования фазовой функции фокусатора в линию [25–30]. Нелинейное преобразование фазы фокусатора в кривую по закону многопорядковой дифракционной решетки позволяет рассчитывать многофокусные фокусаторы, реализующие фокусировку в набор линий (или точек) переменных (или одинаковых) размеров, расположенных в различных плоскостях вдоль оптической оси. Отметим, что ДОЭ, соответствующий обычной (линейной) суперпозиции фокусатора в кривую и многопорядковой дифракционной решетки, позволяет реализовать фокусировку в набор линий только одинакового размера и только в одной плоскости.

Другим интересным классом ДОЭ являются спектральные фокусаторы, рассчитываемые по методу нелинейного преобразования фазы фокусатора по закону цветоделительной дифракционной решетки [31–33]. Цветоделительные дифракционные решетки впервые предложены в работах H. Dammann [31,32] и предназначены для разделения трех различных спектральных компонент по 0, $+1$ и -1 дифракционным порядкам. Спектральные ДОЭ обобщают цветоделительные решетки и предназначены для фокусировки различных спектральных компонент в различные фокальные области.

Данная глава охватывает материал по фокусаторам от методов расчета и детального исследования их рабочих характеристик до методов конструирования на основе фокусаторов новых классов ДОЭ с заданными многофокусными и спектральными свойствами.

5.2. Геометрооптический расчет фокусаторов в линию

Рассмотрим задачу расчета фазовой функции фокусатора, предназначенного для фокусировки плоского пучка с интенсивностью $I_0(\mathbf{u})$, где $\mathbf{u} = (u, v)$ — декартовы координаты в плоскости фокусатора ($z = 0$), в кривую линию S, заданную параметрическим уравнением

$$\mathbf{X}(\xi) = (X(\xi), Y(\xi), Z(\xi)), \quad \xi \in [0, L]. \tag{5.1}$$

где ξ — натуральный параметр (см. рис. 5.1).

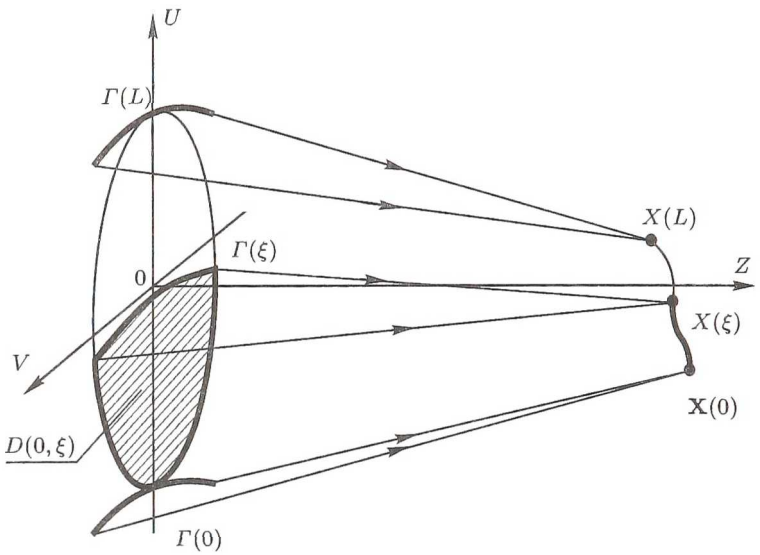


Рис. 5.1. Геометрия задачи фокусировки в пространственную кривую.

Расчет фокусаторов основан на поиске лучевого соответствия $\xi = \xi(\mathbf{u})$ между точками фокальной кривой и точками апертуры фокусатора с последующим восстановлением фазовой функции фокусатора из уравнения наклонов:

$$\begin{aligned}\frac{\partial\varphi(\mathbf{u})}{\partial u} &= \frac{k(X(\xi(\mathbf{u})) - u)}{\sqrt{(u - X(\xi(\mathbf{u})))^2 + (v - Y(\xi(\mathbf{u})))^2 + Z^2(\xi(\mathbf{u}))}},\\ \frac{\partial\varphi(\mathbf{u})}{\partial v} &= \frac{k(Y(\xi(\mathbf{u})) - v)}{\sqrt{(u - X(\xi(\mathbf{u})))^2 + (v - Y(\xi(\mathbf{u})))^2 + Z^2(\xi(\mathbf{u}))}}.\end{aligned} \tag{5.2}$$

Поскольку фокусатор двумерен, а фокальная кривая одномерна, то существует одномерное множество $\Gamma(\xi)$ точек (u, v) на апертуре фокусатора, направляющих излучение в одну и ту же точку $\mathbf{X}(\xi)$ кривой (рис. 5.1). Следуя принятой терминологии, будем называть это множество слоем. Структура лучевого соответствия $\xi(\mathbf{u})$ и вид слоев определяется следующим фундаментальным свойством: все лучи приходящие в данную точку кривой фокусировки находятся на поверхности конуса, ось которого является касательной к кривой фокусировки в данной точке. Поэтому система слоев на апертуре фокусатора соответствует семейству кривых второго порядка, являющихся сечениями конических поверхностей плоскостью $z = 0$. Строгое доказательство данных фактов можно найти в работе [12]. Простая и наглядная интерпретация указанного свойства может быть получена на примере более простой задачи расчета фазовой функции фокусатора в набор из N точек, расположенных на пространственной кривой (5.1). Будем считать, что координаты точек фокусировки (x_i, y_i, z_i) соответствуют возрастающим значениям $\xi_i, i = 0, \ldots, N$ параметра ξ. Для расчета фазовой функции $\varphi(\mathbf{u})$ фокусатора в N точек разобьем апертуру фокусатора D на N областей (сегментов) D_i. Фазовую функцию $\varphi_i(\mathbf{u})$ в пределах сегмента D_i определим из условия фокусировки в точку (x_i, y_i, z_i) на кривой. В этом случае при $\mathbf{u} \in D_i$ фазовая функция фокусатора является фазой линзы с фокусом в точке (x_i, y_i, z_i):

$$\varphi_i(u, v) = -k\sqrt{(u - x_i)^2 + (v - y_i)^2 + z_i^2}. \tag{5.3}$$

В параксиальном приближении, то есть при

$$\frac{(x_i-u)^2+(y_i-v)^2}{z_i^2} \ll 1, \quad (u,v)\in D, \quad i=\overline{1,N},$$

формула (5.3) переходит в известную формулу для фазовой функции сферической линзы:

$$\varphi_i(u,v) = -k\frac{(u-x_i)^2+(v-y_i)^2}{2z_i}. \tag{5.4}$$

Из физических соображений следует наложить требование непрерывности фазовой функции сегментированного фокусатора. Для расчета непрерывной фазовой функции необходимо выбрать границы областей сегментации апертуры фокусатора таким образом, чтобы при переходе через границу в любом месте фазовая функция изменялась на одну и ту же константу. Таким образом, границу раздела между областями D_i и D_{i+1} следует искать из условия $\varphi_i(u,v) - \varphi_{i+1}(u,v) = C(\xi_i)$ или более подробно:

$$k\sqrt{(u-x_{i+1})^2+(v-y_{i+1})^2+z_{i+1}^2} - k\sqrt{(u-x_i)^2+(v-y_i)^2+z_i^2} = C(\xi_i). \tag{5.5}$$

Нетрудно видеть, что выражение (5.5) представляет собой уравнение некоторой кривой, которая является сечением двуполостного гиперболоида вращения,

$$\sqrt{(x-x_{i+1})^2+(y-y_{i+1})^2+(z-z_{i+1})^2} - \\ - \sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2} = C(\xi_i)/k, \tag{5.6}$$

плоскостью оптического элемента. Фокусы данного гиперболоида находятся в точках (x_i,y_i,z_i) и $(x_{i+1},y_{i+1},z_{i+1})$, а ось симметрии гиперболоида совпадает с отрезком, соединяющим точки фокусировки с индексами i и $i+1$. Следует отметить, что константа

$$\Delta\Psi(\xi_i) = C(\xi_i)/k \tag{5.7}$$

в правой части уравнения (5.6) соответствует приращению эйконала при переходе от точки (x_i,y_i,z_i) к точке $(x_{i+1},y_{i+1},z_{i+1})$.

При указанном способе выбора границ сегментов, непрерывная фазовая функция фокусатора в N точек получается добавлением к фазовой функции $\varphi_{i+1}(u)$ константы

$$\tilde{\varphi}_i = \sum_{n=0}^{n=i} C(\xi_n). \tag{5.8}$$

Можно показать, что предложенный способ получения оптического элемента с непрерывной фазовой функцией — единственный, и другие способы сегментации дают разрывную фазовую функцию.

В пределе при стремлении числа точек фокусировки N к бесконечности, так что расстояние между соседними точками стремится к нулю, мы получим фокусатор в непрерывную кривую S. При этом двуполостный гиперболоид вращения (5.6) стремится к конической поверхности. Действительно, заменяя разность в правой части

уравнения (5.6) первым членом ряда Тейлора, получим следующее уравнение конуса :

$$\frac{(x_{i+1}-x_i)(x-x_i)+(y_{i+1}-y_i)(y-y_i)+(z_{i+1}-z_i)(z-z_i)}{\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}}=\Delta\Psi(\xi_i). \qquad (5.9)$$

Ось конуса совпадает с отрезком, соединяющим две соседние точки кривой, и в пределе совпадает с направлением касательной к фокальной кривой в данной точке.

Производя замену

$$(x_{i+1}-x_i)=\frac{\mathrm{d}X(\xi_i)}{\mathrm{d}\xi}\Delta\xi_i,\quad (y_{i+1}-y_i)=\frac{\mathrm{d}Y(\xi_i)}{\mathrm{d}\xi}\Delta\xi_i,\quad (z_{i+1}-z_i)=\frac{\mathrm{d}Z(\xi_i)}{\mathrm{d}\xi}\Delta\xi_i,$$

где $\Delta\xi_i=\xi_{i+1}-\xi_i$, представим (5.9) в виде

$$\frac{\frac{\mathrm{d}X(\xi_i)}{\mathrm{d}\xi}(u-X(\xi_i))+\frac{\mathrm{d}Y(\xi_i)}{\mathrm{d}\xi}(v-Y(\xi_i))+\frac{\mathrm{d}Z(\xi_i)}{\mathrm{d}\xi}(z-Z(\xi_i))}{\sqrt{(u-X(\xi_i))^2+(v-Y(\xi_i))^2+(z-Z(\xi_i))^2}}=c\,(\xi_i)\,, \qquad (5.10)$$

где $c(\xi_i)=\Delta\Psi(\xi_i)/\Delta\xi_i$ — угол при вершине конической поверхности.

В этом случае площадь i-й области на апертуре ДОЭ стремится к нулю, и сама область стягивается в слой, то есть в кривую 2-го порядка, являющуюся сечением конуса (5.10) плоскостью оптического элемента. Полагая $z=0$ в (5.10), уравнение слоя запишем в виде

$$\frac{\frac{\mathrm{d}X(\xi_i)}{\mathrm{d}\xi}(u-X(\xi_i))+\frac{\mathrm{d}Y(\xi_i)}{\mathrm{d}\xi}(v-Y(\xi_i))-\frac{\mathrm{d}Z(\xi_i)}{\mathrm{d}\xi}Z(\xi_i)}{\sqrt{(u-X(\xi_i))^2+(v-Y(\xi_i))^2+Z(\xi_i)^2}}=c\,(\xi_i)\,. \qquad (5.11)$$

Полученные предельным переходом уравнения конуса (5.10) и слоя $\Gamma(\xi)$ (5.11) являются доказательством ранее указанного факта, что все лучи приходящие в данную точку кривой фокусировки находятся на поверхности кругового конуса, ось которого касательна к кривой фокусировки в данной точке. При этом угол при вершине конической поверхности равен производной эйконала вдоль кривой. В дальнейшем изложении мы будем пользоваться непрерывным аналогом уравнения (5.11):

$$\frac{\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi}(u-X(\xi))+\frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi}(v-Y(\xi))-\frac{\mathrm{d}Z(\xi)}{\mathrm{d}\xi}Z(\xi)}{\sqrt{(u-X(\xi))^2+(v-Y(\xi))^2+Z(\xi)^2}}=c\,(\xi)=\frac{\mathrm{d}\Psi(\xi)}{\mathrm{d}\xi}. \qquad (5.12)$$

Угол при вершине конической поверхности $c(\xi)$в (5.12) определяет распределение энергии $I\,(\xi)$, $\xi\in[0,L]$ вдоль кривой фокусировки. Следуя принятой терминологии, будем называть функцию $I\,(\xi)$ линейной плотностью. Данное название связано с определением линейной плотности как интегральной величины, соответствующей потоку энергии на единицу длины фокальной кривой. Заданное распределение линейной плотности $I\,(\xi)$ может быть сформировано за счет выбора распределения эйконала $\Psi\,(\xi)$ на кривой. Действительно, для формирования заданной линейной плотности $I\,(\xi)$, функцию $c(\xi)=\mathrm{d}\Psi(\xi)/\,\mathrm{d}\xi$ в (5.12) определим из закона сохранения энергии в интегральной форме. Для этого достаточно приравнять световой поток падающий на часть апертуры фокусатора $D(0,\xi)$, заключенной между начальным и текущим слоями $\Gamma\,(0)$ и $\Gamma\,(\xi)$ к световому потоку проходящему через часть фокальной кривой, заключенной между точками $\mathbf{X}\,(0)$ и $\mathbf{X}\,(\xi)$ (см.рис. 5.1):

$$\int_0^{\xi} I(t)\,\mathrm{d}t=E(\xi,c(\xi))=\int\limits_{D(0,\xi)} I_0\,(\mathbf{u})\,\mathrm{d}^2\mathbf{u}. \qquad (5.13)$$

Полученные уравнения (5.12), (5.13) позволяют сформулировать следующий метод расчета фазовой функции фокусатора в кривую с заданной линейной плотностью:

1) расчет функции $c(\xi)$ из уравнения (5.13) по заданной линейной плотности;
2) расчет функции лучевого соответствия $\xi(\mathbf{u})$ из уравнения (5.12);
3) восстановление фазовой функции фокусатора из уравнения наклонов (5.2) методами теории потенциала,

$$\varphi(\mathbf{u}) = \varphi(\mathbf{u}_0) + \int\limits_{\mathbf{u}_0}^{\mathbf{u}} \nabla_\perp \varphi(\mathbf{u})\, \mathrm{d}^2\mathbf{u}, \tag{5.14}$$

здесь интегрирование ведется по произвольной кривой, соединяющей точку $\mathbf{u}$ с некоторой начальной точкой $\mathbf{u_0}$, $\nabla_\perp = (\partial/\partial u, \partial/\partial v)$.

Важным практическим случаем является фокусировка в плоскую кривую $\mathbf{X}(\xi) = (X(\xi), Y(\xi), f)$, расположенную в плоскости $z = f$. В этом случае слой (5.11) является гиперболой [9, 10, 12, 13]:

$$\frac{\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi}(u - X(\xi)) + \frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi}(v - Y(\xi))}{\sqrt{(u - X(\xi))^2 + (v - Y(\xi))^2 + f^2}} = c(\xi). \tag{5.15}$$

Ось симметрии гиперболы совпадает с проекцией оси конуса на плоскость фокусатора. В параксиальном приближении, то есть при выполнения условия

$$\sqrt{(u - X(\xi))^2 + (v - Y(\xi))^2 + f^2} \approx f,$$

слои (5.15) вырождаются в прямые, перпендикулярные касательным к фокальной кривой,

$$\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi} u + \frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi} v = p(\xi), \tag{5.16}$$

функция

$$p(\xi) = c(\xi) f + \frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi} X(\xi) + \frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi} X(\xi)$$

имеет смысл расстояния от слоя до начала координат в плоскости фокусатора.

Фазовая функция фокусатора в параксиальном приближении принимает вид

$$\varphi(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \varphi_D(\mathbf{u}), \tag{5.17}$$

где

$$\varphi_D(\mathbf{u}) = \frac{k}{f} \int\limits_{u_0}^{u} X(\xi(u, v_0))\, \mathrm{d}u + \frac{k}{f} \int\limits_{v_0}^{v} Y(\xi(u, v))\, \mathrm{d}v. \tag{5.18}$$

Рассмотрим в качестве важного практического примера расчет фокусатора в отрезок плоскости с уравнением $\mathbf{X}(\xi) = (-d + \xi, 0, f)$, $\xi \in [0, 2d]$, и с распределением линейной плотности $I(\xi)$, $\xi \in [0, 2d]$. При фокусировке в отрезок слои являются отрезками прямых $u = p(\xi)$, перпендикулярных фокальному отрезку. Предположим,

что апертура фокусатора D ограничена кривыми $v = g_1(u)$ и $v = g_2(u)$ и отрезками прямых $u = a$ и $u = b$. В этом случае, полагая в законе сохранения светового потока (5.13) $u = p(\xi)$, получим для расчета функции лучевого соответствия $\xi = \xi(u)$ следующее уравнение:

$$\int_a^u \int_{g_1(u)}^{g_2(u)} I_0(\gamma, \chi)\, \mathrm{d}\chi\, \mathrm{d}\gamma = \int_0^{\xi} I(\chi)\, \mathrm{d}\chi, \tag{5.19}$$

Более удобно работать с функцией $x(u) = \xi(u) - d$, представляющей положение точки на отрезке фокусировки. Закон сохранения светового потока (5.19) более удобно использовать в дифференциальной форме:

$$\frac{\mathrm{d}x(u)}{\mathrm{d}u} = \frac{1}{I(x(u))} \int_{g_1(u)}^{g_2(u)} I_0(u, v)\, \mathrm{d}v, \quad x(a) = -d, \quad x(b) = d, \tag{5.20}$$

При этом фазовая функция фокусатора в отрезок принимает вид

$$\varphi(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \frac{k}{f} \int_0^u x(\gamma)\, \mathrm{d}\gamma. \tag{5.21}$$

Следует отметить, что даже в параксиальном приближении аналитический расчет фокусаторов возможен только для случаев фокусировки в простые линии, такие как отрезок, кольцо, полукольцо и т.п. При фокусировке в более сложные кривые расчет фазы фокусатора является неординарной задачей. Действительно, в общем случае пункты 1), 2), рассмотренного выше метода, являются сложными вычислительными задачами, состоящими в решении нелинейных уравнений. При этом пункт 2 требует решения нелинейного уравнения (5.12) для каждой точки $\mathbf{u}$ апертуры фокусатора. Выполнение пункта 3) также основано на операции двумерного численного интегрирования.

Расчет фокусаторов существенно упрощается при использовании следующей криволинейной системы координат:

$$\left.\begin{aligned} u(\xi, t) &= p(\xi)\, \frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi} - t \cdot \frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi}, \\ v(\xi, t) &= p(\xi)\, \frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi} + t \cdot \frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi}. \end{aligned}\right\} \tag{5.22}$$

Система координат (5.22) связана со слоями на фокусаторе и выражает координаты (u, v) через координату ξ, определяющую слой $\Gamma(\xi)$, содержащий данную точку, и координату t, определяющую положение точки на слое. Отметим, что координата t является расстоянием от точки слоя $(u(0,t), v(0,t))$, являющейся точкой пересечения слоя и нормали к слою, до текущей точки слоя. В последующем изложении мы ограничимся рассмотрением практически важного случая фокусировки в кривую радиальных пучков:

$$I_0(\mathbf{u}) = I_0\left(u^2 + v^2\right), \quad |\mathbf{u}| \leqslant R, \tag{5.23}$$

где R — радиус апертуры фокусатора.

Для радиально-симметричных пучков закон сохранения светового потока (5.13) в координатах (ξ, t) может быть представлен в компактной дифференциальной форме:

$$\mathrm{d}p(\xi) \times \int\limits_{-\sqrt{R^2-p^2(\xi)}}^{\sqrt{R^2-p^2(\xi)}} I_0\left(p^2(\xi)+t^2\right)\,\mathrm{d}t = I(\xi)\,\mathrm{d}\xi. \tag{5.24}$$

Поскольку интеграл в (5.24) обращается в ноль при $p(0) = -R$, $p(L) = +R$, то уравнение (5.24) следует решать относительно функции $\xi(p)$. Расчет $\xi(p)$ состоит в решении дифференциального уравнения 1-го порядка, разрешенного относительно производной. Для решения такого уравнения может быть использован любой стандартный численный метод, например метод Рунге–Кутта. Далее вычисление $p(\xi)$ состоит в обращении гладкой однозначной функции $\xi(p)$.

Основное преимущество в использовании координат (ξ, t) состоит в возможности прямого вычисления фазы фокусатора без расчета функции $\xi(\mathbf{u})$, состоящего в решении нелинейного уравнения для каждой точки $\mathbf{u}$ апертуры фокусатора. Действительно, из уравнений (5.17), (5.22) несложно получить частные производные функции $\varphi_D(\mathbf{u})$ в (5.17) в переменных (ξ, t) в виде:

$$\frac{\partial\varphi_D(\xi,t)}{\partial\xi} = \frac{k}{f}X(\xi)\cdot\left[\frac{\mathrm{d}p(\xi)}{\mathrm{d}\xi}\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi} + p(\xi)\times\frac{\mathrm{d}^2X(\xi)}{\mathrm{d}\xi^2} - t\frac{\mathrm{d}^2Y(\xi)}{\mathrm{d}\xi^2}\right] + \\ + \frac{k}{f}Y(\xi)\times\left[\frac{\mathrm{d}p(\xi)}{\mathrm{d}\xi}\frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi} + p(\xi)\frac{\mathrm{d}^2Y(\xi)}{\mathrm{d}\xi^2} + t\frac{\mathrm{d}^2X(\xi)}{\mathrm{d}\xi^2}\right], \tag{5.25}$$

$$\frac{\partial\varphi_D(\xi,t)}{\partial t} = \frac{k}{f}\left[\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi}Y(\xi) - X(\xi)\frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi}\right]. \tag{5.26}$$

Уравнения (5.25), (5.26) позволяют представить фазовую функцию фокусатора в переменных (ξ, t) в виде

$$\varphi(\xi,t) = -\frac{k}{2f}\left(p^2(\xi)+t^2\right) + \varphi_D(\xi,t), \tag{5.27}$$

где

$$\varphi_D(\xi,t) = +\frac{k}{f}\left[\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi}Y(\xi) - X(\xi)\frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi}\right]\cdot t + \\ + \frac{k}{f}\left[\frac{\mathrm{d}X(\xi)}{\mathrm{d}\xi}X(\xi) + \frac{\mathrm{d}Y(\xi)}{\mathrm{d}\xi}Y(\xi)\right]p(\xi) - \frac{k}{f}\int\limits_0^{\xi} p(\eta)\,\mathrm{d}\eta. \tag{5.28}$$

Согласно уравнениям (5.24), (5.27), расчет фазы фокусатора в переменных (ξ, t) сводится к решению дифференциального уравнения первого порядка, обращению полученного решения и одномерному интегрированию функции $p(\xi)$. Следует отметить, что рассмотренные криволинейные координаты (5.22) несложно обобщить и на непараксиальный случай. Действительно, единственное отличие непараксиального случая от параксиального состоит в том, что слои являются гиперболами (5.15), а не отрезками прямых (5.16). Поэтому в непараксиальном случае следует выразить

координаты (u, v) через координату ξ, определяющую гиперболу $\Gamma(\xi)$, содержащую данную точку и координату t, определяющую положение точки на гиперболе относительно ее вершины. В частности, при фокусировке в отрезок $\mathbf{X}(\xi) = (-d + \xi, 0, f)$, $\xi \in [0, 2d]$ с линейной плотностью $I(\xi)$ в непараксиальном случае, координаты (ξ, t) имеют наиболее простой вид:

$$\begin{cases} u(\xi, t) = \xi - d + \beta(\xi) \sqrt{f^2 + \eta^2}, \\ v(\xi, t) = t, \end{cases} \tag{5.29}$$

где функция

$$\beta(\xi) = c(\xi)/\sqrt{1 - c^2(\xi)} \tag{5.30}$$

соответствует котангенсу угла при вершине конической поверхности, содержащей лучи, приходящие в данную точку отрезка.

Представление параксиальной фазовой функции фокусатора (5.27) в кривую S с уравнением $\mathbf{X}(\xi) = (X(\xi), Y(\xi), f)$ в координатах (ξ, t) (5.22) позволяет установить следующее интересное свойство.

Умножение функции φ_D из (5.27) на константу t дает фазовую функцию ДОЭ, фокусирующего в t раз увеличенную кривую S_t :

$$X_t(\xi_t) = (tX(\xi_t/t), tY(\xi_t/t), f), \quad \xi_t \in [0, tL]. \tag{5.31}$$

При этом линейная плотность энергии

$$I_t(\xi_t) = \frac{1}{t} I(\xi_t/t), \quad \xi_t \in [0, t \cdot L], \tag{5.32}$$

вдоль линии S_t совпадает с функцией $I(\xi)$ для линии S с точностью до масштаба. Данное свойство будет широко использоваться в дальнейшем изложении и легко проверяется непосредственной подстановкой уравнения кривой (5.31) в общее представление фазовой функции (5.27), (5.28).

Операция изготовления фотошаблона фокусатора на литографе или лазерном фотопостроителе предполагает использование декартовых или полярных координат. Согласно (5.27), при фиксированном значении ξ фазовая функция вдоль слоя соответствует суперпозиции фазовых функций линзы и призмы. Это позволяет интерпретировать фокусатор как набор цилиндрических линз и призм, ориентированных вдоль слоев $\Gamma(\xi)$. При этом для точек $\mathbf{u} \in \Gamma(\xi)$ фазовая функция имеет вид

$$\varphi(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \frac{k}{f}(uX(\xi) + vY(\xi)) + c(\xi), \tag{5.33}$$

где

$$c(\xi) = -\frac{k}{f} \int_0^{\xi} p(\eta)\, \mathrm{d}\eta.$$

Согласно (5.33), для расчета фазовой функции (5.10) в декартовых координатах (u, v) достаточно построить систему из N слоев (5.16) по функции $p(\xi)$. При этом система слоев $\Gamma(\xi_n)$, $n = \overline{1, N}$ определит разбиение апертуры фокусатора на N сегментов D_i. Далее, для вычисления фазы фокусатора в точке $\mathbf{u}$ достаточно найти

сегмент D_i, содержащий данную точку и воспользоваться формулой (5.33). Отметим, что поиск сегмента D_i, содержащего данную точку $\mathbf{u}$, сводится к тривиальным операциям определения положения точки относительно прямых.

Проведенный анализ структуры фазы фокусатора вдоль слоев позволяет предложить следующий простой алгоритм расчета фазовой функции фокусатора:

1) расчет функции $p(\xi)$ из уравнения (5.24).

2) переход к декартовым (или полярным) координатам с введением дискретной системы слоев и последующим расчетом фазы в сегментах по формуле (5.33).

В качестве примера рассмотрим расчет фокусатора плоского пучка круглого сечения радиуса R в дугу окружности с постоянной линейной плотностью. Параметрическое уравнение дуги окружности имеет вид

$$\begin{cases} X(\xi) = R_1 \cos\left(\left(\xi - R_1\alpha/2\right)/R_1\right) \\ Y(\xi) = R_1 \sin\left(\left(\xi - R_1\alpha/2\right)/R_1\right) \end{cases}, \quad \xi \in [0, R_1\alpha], \tag{5.34}$$

где R_1 — радиус дуги, α — угловой размер дуги.

Из закона сохранения светового потока (5.24) несложно получить функцию $\xi(p)$ в виде

$$\xi(p) = \frac{R_1\alpha}{\pi}\left(\arcsin\left(\frac{p}{R}\right) + \frac{p}{R}\sqrt{1 - \left(\frac{p}{R}\right)^2}\right). \tag{5.35}$$

Подставляя (5.34) в (5.22) и (5.27), получим фазовую функцию фокусатора в дугу в виде:

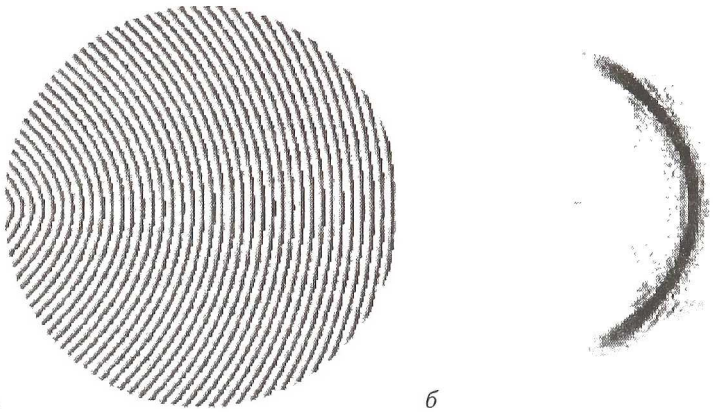
$$\varphi(\xi, t) = -\frac{k}{2f}\left(p^2(\xi) + t^2\right) - \frac{k}{f}\left(t \cdot R_1 + \int_0^{\xi} p(\eta)\, \mathrm{d}\eta\right). \tag{5.36}$$

Простой вид полученной фазы (5.36) демонстрирует достоинство криволинейной системы координат (5.22). Расчет фазы (5.36) сводится к выполнению только двух простых численных операций — обращения функции (5.35) и ее одномерного интегрирования.

Пример 5.1. На рис. 5.2*а* приведено полутоновое изображение фазовой функции фокусатора в дугу окружности, рассчитанной как дополнение к линзе, для следующих параметров $\lambda = 0{,}63$ мкм, $f = 500$ мм, $R = 5$ мм, $R_1 = 1{,}25$ мм, $\alpha = \pi/2$. Для расчета фазы (5.36) в декартовых координатах была построена система из 200 слоев по функции $p(\xi)$, найденной из уравнения (5.35) с последующим использованием формулы (5.33). Полутоновое распределение интенсивности, формируемое фокусатором на рис. 5.2*а*, приведено на рис. 5.2*б* и демонстрирует работоспособность разработанного метода.

5.3. Расчет и исследование геометрооптических фокусаторов

В данном разделе проведено исследование некоторых, наиболее важных, эталонных задач фокусировки. Приведенные здесь результаты расчета и исследования фокусаторов позволяют оценить степень применимости геометрооптического подхода к расчету ДОЭ, установить связь характеристик фокусаторов и оценить влияние технологических погрешностей (дискретизации и квантования фазовой функции)



а *б*

Рис. 5.2. Фазовая функция фокусатора для фокусировки в дугу окружности (*а*), распределение интенсивности в плоскости фокусировки (*б*)

на качество работы фокусаторов. Полученные ниже результаты анализа различных подходов к решению ряда обратных эталонных задачах фокусировки являются основополагающими для оценки использования ДОЭ в конкретных технологических установках.

Все приведенные в данном разделе результаты дифракционного исследования фокусаторов получены численным расчетом интеграла Френеля–Кирхгофа. Для оценки влияния погрешностей дискретизации и квантования фазы на качество работы фокусаторов расчет интеграла Френеля–Кирхгофа проводился с использованием специально разработанных квадратурных формул [20–24], основанных на:

— разбиении апертуры фокусатора на прямоугольные или, в случае радиальной симметрии, на кольцевые модули (ячейки);

— использовании в пределах каждого модуля кусочно-постоянной аппроксимация для функции комплексного пропускания (ФКП) фокусатора;

— использовании базовых аналитических решений задач дифракции на прямоугольных и кольцевых модулях [20–24].

При этом кусочно-постоянная аппроксимация для ФКП фокусатора соответствует процессу отображения фазы фокусатора фотопостроителем с прямоугольным или кольцевым растром, а использование базовых аналитических решений задач дифракции на прямоугольных и кольцевых модулях обеспечивает высокую точность дифракционного расчета [20–24].

5.3.1. Дифракционная линза

Дифракционная линза является наиболее важным и широко используемым дифракционным элементом. Проведенные в данном пункте результаты исследования дифракционной линзы имеют большую значимость и для оценки работы фокусаторов в линию. Данный факт связан с рассмотренной в п. 5.1.2. интерпретацией фокусатора в кривую как набора линз, ориентированных вдоль слоев.

Метод и результаты расчета трехмерного распределения света вблизи фокуса обычной рефракционной линзы имеются в классической работе М. Борна и Э. Вольфа «Основы оптики» [34]. Более точные методы и результаты расчетов приведены в ряде современных исследований [35–41]. Эти исследования позволили выявить асимметрию распределения интенсивности относительно фокальной плоскости

и оценить величину фокального сдвига максимума интенсивности. Аналогичные результаты получены также для случая фокусировки гауссовых пучков [42–47] и для цилиндрических линз [48, 49]. Анализу фокусировки в точку посвящены работы по исследованию зонных пластинок и плоской дифракционной линзы [21, 23, 50–58]. В частности, в работе [55] получены распределения интенсивности в фокальной плоскости дифракционной линзы для различного числа уровней квантования фазы. В работах [21, 57, 58] путем численных расчетов получены распределения интенсивности вдоль оптической оси дифракционной линзы для различного числа уровней квантования фазовой функции.

Также представляет интерес полное исследование структуры сфокусированного излучения во всей фокальной области [23].

Фазовая функция линзы в параксиальном приближении имеет вид

$$\varphi(r) = -\frac{kr^2}{2f}, \quad r \leqslant a, \tag{5.37}$$

где $k = 2\pi/\lambda$, λ — длина волны освещающего пучка, f — фокусное расстояние, r — полярный радиус.

На рис. 5.3 представлены этапы получения плоской линзы. На рис. 5.3*а* показана фазовая функция обычной линзы (5.37). На рис. 5.3*б* — фазовая функция идеальной плоской линзы, т.е. фаза (5.37) взятая по модулю 2π. На рис. 5.3*в* показан реальный ступенчатый профиль плоской линзы, получаемый в соответствии с многоступенчатой технологией изготовления ДОЭ. Аппроксимация непрерывной фазовой функции плоской линзы ступенчатым профилем возможна ступенями равной ширины (равномерная дискретизация) или ступенями равной глубины, как на рис. 5.3*в*. В последнем случае дискретизацию назовем неравномерной. Она определяется выбранным числом M уровней квантования фазы. Для линзы при трех уровнях квантования фазы и трех зонах Френеля получается $N = 9$ кольцевых модулей. Внешние радиусы кольцевых зон для M уровней квантования фазы определяются

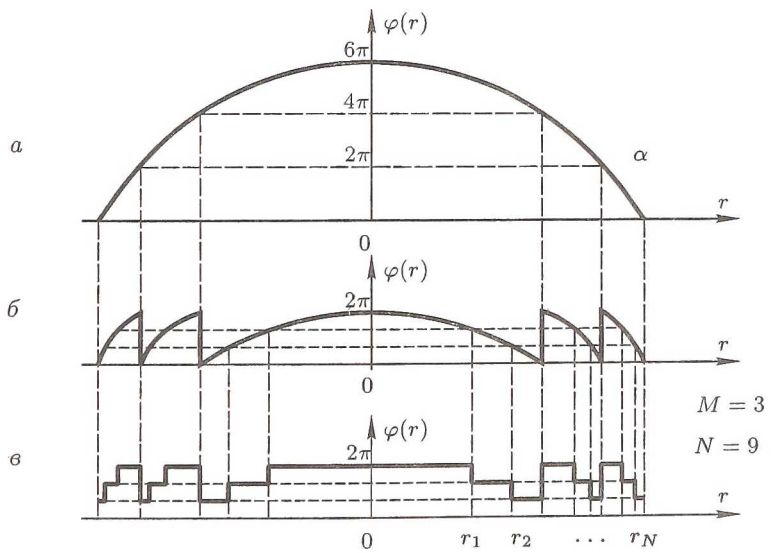


Рис. 5.3. Этапы получения плоской линзы

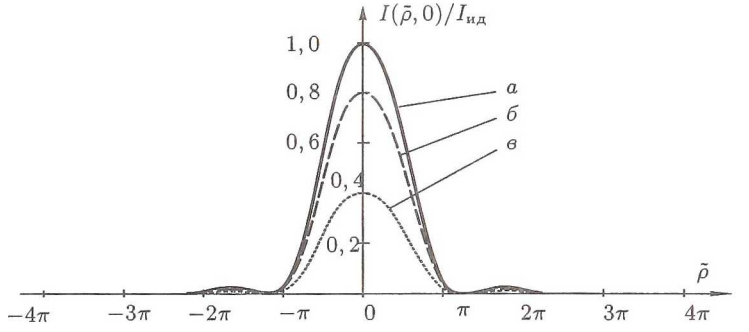


Рис. 5.4. Нормированные распределения интенсивности в фокальной плоскости плоской линзы с неравномерной дискретизацией для различного числа уровней квантования M: $M = 16$ (*а*), $M = 4$ (*б*), $M = 2$ (*в*)

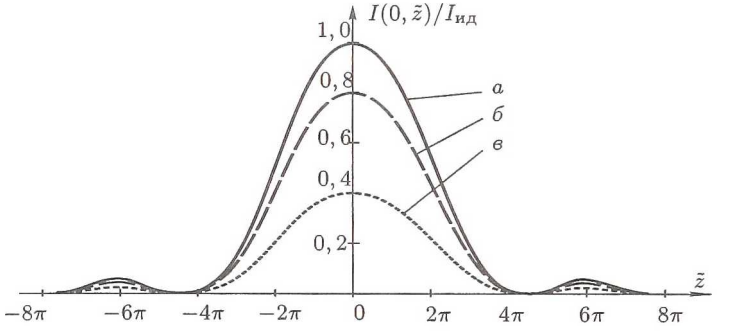


Рис. 5.5. Нормированные распределения интенсивности вдоль оптической оси плоской линзы с неравномерной дискретизацией для различного числа уровней квантования M: $M = 16$ (*а*), $M = 4$ (*б*), $M = 2$ (*в*) для больших чисел Френеля $F = a^2/(\lambda f)$

по формуле

$$r_n = \sqrt{\frac{2nf \cdot \lambda}{M}}, \quad n = \overline{1, N}, \tag{5.38}$$

где $N = \text{int}\left[Ma^2/2\lambda f\right] + 1$. При равномерной дискретизации квантование фазовой функции внутри кольцевого модуля не зависит от числа элементов дискретизации.

На рис. 5.4 представлены распределения интенсивности поля $I\left(\tilde{\rho},0\right)/I_{\text{ид}}$ в фокальной плоскости плоской линзы с дискретизацией (5.38) при 2, 4 и 16 уровнях квантования фазы. На рис. 5.5 приведены аналогичные распределения $I\left(0,\tilde{z}\right)/I_{\text{ид}}$ вдоль оптической оси при больших числах Френеля, когда можно пренебречь асимметрией распределения относительно фокальной плоскости и величиной фокального сдвига. Здесь $I_{\text{ид}}$ — интенсивность в фокусе неквантованной линзы [34]:

$$I_{\text{ид}} = I_0 \left(\frac{ka^2}{2f}\right)^2, \tag{5.39}$$

где I_0 — интенсивность плоского освещающего пучка, а

$$\tilde{\rho} = k\left(\frac{a}{f}\right) \cdot \rho, \quad \tilde{z} = k\left(\frac{a}{f}\right)^2 \cdot z \tag{5.40}$$

— безразмерные координаты в фокальной области плоской линзы с центром в фокусе линзы. На рис. 5.6 представлены изофоты (линии равной интенсивности функции

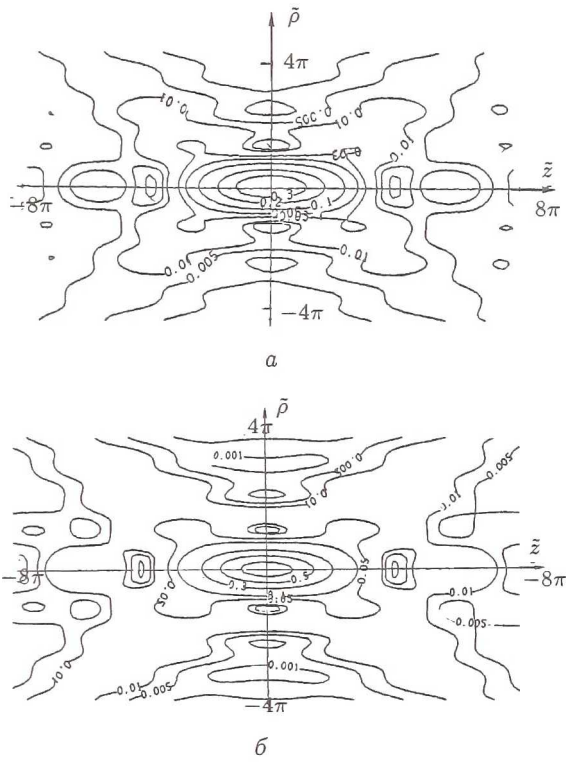


Рис. 5.6. Изофоты нормированных распределений интенсивности $I(\tilde{\rho}, \tilde{z})/I_{\text{ид}}$ в фокальной области плоской линзы с неравномерной дискретизацией: ($а$) $M = 2$, ($б$) $M = 4$.

$I(\tilde{\rho}, \tilde{z})/I_{\text{ид}}$) в меридианальном сечении фокальной области плоской линзы с большим числом Френеля $F = a_2/(\lambda f)$ для 2-х и 4-х уровней квантования M. Изофоты дополняют результаты работ [55, 56], в которых построены лишь графики распределений интенсивности, совпадающие с рис. 5.4 и рис. 5.5. Анализ рисунков 5.4–5.6 подтверждает вывод работы [59] о том, что форма волнового фронта, формируемого в полезном $+1$ порядке дифракции квантованной линзы, фактически не зависит от числа уровней квантования.

Представляет интерес исследование распределения интенсивности вдоль оптической оси плоской линзы с небольшим числом Френеля. На рис. 5.7 и 5.8 приведены распределения нормированной интенсивности $I(0, \tilde{z})/I_{\text{ид}}$ вдоль оптической оси для дифракционной линзы с числом уровней квантования фазы $M = 2, 4, 16$, полученные для равномерного и гауссова $\left(A_0(r) = \exp\left(-r^2/\sigma^2\right)\right)$ пучков при следующих параметрах: радиус $a = 1$ мм; длина волны $\lambda = 10{,}6$ мкм; число Френеля $F = a^2/(\lambda f) = 4$; параметр гауссова пучка $\sigma = 0{,}8$ мм. Графики на рис. 5.7 соответствуют данным работ [57, 58], что свидетельствует о корректности использованного метода расчета интеграла Френеля–Кирхгофа в радиальном случае. Рис. 5.8 показывает «расплывание» фокального пятна вдоль оптической оси для гауссова пучка с радиусом перетяжки, меньшим радиуса линзы. При увеличении числа уровней квантования распределение асимптотически приближается к распределению интенсивности в фокальной области рефракционной линзы, освещаемой гауссовым пучком [42–47].

Эффективность плоской линзы может быть охарактеризована отношением $I(0,0)/I_{\text{ид}}$ интенсивности в фокусе плоской линзы к интенсивности (5.39) в фо-

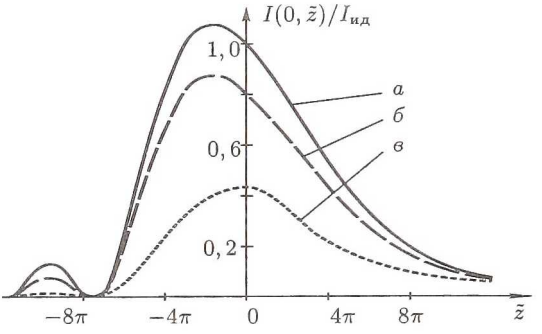


Рис. 5.7. Распределения нормированной интенсивности $I\,(0,\tilde{z})\,/I_{\text{ид}}$ вдоль оптической оси для квантованной дифракционной линзы с числом Френеля $F = 4$ при плоском освещающем пучке: $M = 16$ $(а)$, $M = 4$ $(б)$, $M = 2(в)$

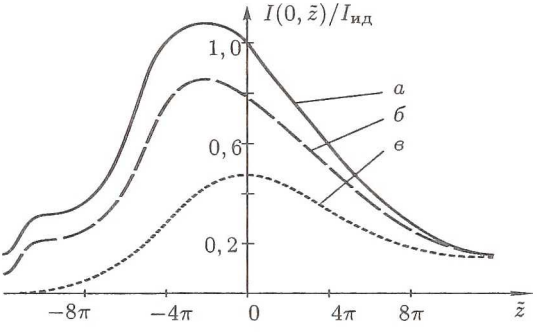


Рис. 5.8. Распределения нормированной интенсивности $I\,(0,\tilde{z})\,/I_{\text{ид}}$ вдоль оптической оси для квантованной дифракционной линзы с числом Френеля $F = 4$ при гауссовом освещающем пучке: $M = 16$ $(а)$, $M = 4$ $(б)$, $M = 2$ $(в)$

кусе идеальной линзы с такими же параметрами. Расчетные значения для эффективности плоской линзы с неравномерной дискретизацией представлены в табл. 5.1 и совпадают с аналогичными данными, полученными методом фурье-анализа [59].

В известных работах [55–59] отсутствуют данные о том, что будет с дифракционной эффективностью и распределением интенсивности в фокальной области плоской линзы в случае отступления от оптимальной неравномерной дискретизации. Анализ в этом случае затруднен тем, что эффективность плоской линзы и структура сфокусированного излучения уже не будут зависеть только от числа уровней квантования, а будут определяться как физическими параметрами, так и параметрами дискретизации. В табл. 5.2 представлена зависимость эффективности плоской линзы с равномерной дискретизацией от числа уровней квантования M и числа кольцевых элементов дискретизации N при следующих параметрах: $f = 100$ мм; диаметр линзы $2a = 6{,}04$ мм; $\lambda = 0{,}633$ мкм; число Френеля $F = a^2/(\lambda f) = 144$. Число элементов дискретизации для плоской линзы с неравномерной дискретизацией равно

Таблица 5.1. Эффективность $I\,(0,0)\,/I_{\text{ид}}$ плоской линзы с неравномерной дискретизацией при различном числе уровней квантования M

M	2	4	5	8	16	32
$I(0,0)/I_{\text{ид}}$	0,405	0,811	0,875	0,9495	0,9871	0,9966

Таблица 5.2. Эффективность $I(0,0)/I_{ug}$ квантованной линзы с равномерной дискретизацией по радиусу в N точках при параметрах: $f = 100$ мм; $a = 3{,}02$ мм; $\lambda = 0{,}633$ мкм

N \ M	∞	16	4	2	$\Gamma = a/N$ (мкм)
128	0,094	0,093	0,076	0,038	23,6
256	0,585	0,577	0,476	0,263	11,8
512	0,877	0,866	0,711	0,356	5,9
1024	0,968	0,955	0,786	0,392	2,95
2048	0,992	0,979	0,804	0,402	1,47
$\Gamma_{\min}$ (мкм) при оптимальной дискретизации	–	1,31	5,25	10,5	–

$N = [M \times F/2] + 1$. При этом ширина кольцевых модулей будет убывать к краям линзы и достигнет у края величины, приведенной в табл. 5.2. Величина Γ_{min} вычисляется, согласно (5.38), по формуле

$$\Gamma_{\min} = r_N - r_{N-1} = \sqrt{\frac{2Nf\lambda}{M}} - \sqrt{\frac{2(N-1)\lambda f}{M}} = \sqrt{\frac{2\lambda f}{M}}\left(\sqrt{N} - \sqrt{N-1}\right). \quad (5.41)$$

Сравнительный анализ данных таблиц 5.1, 5.2 свидетельствует об эффективности неравномерной дискретизации. Для того чтобы добиться при равномерной дискретизации такой же эффективности как при неравномерной с M уровнями квантования фазы, шаг равномерной дискретизации Γ должен быть меньше размера крайнего модуля $\Gamma_{\min}$, определяемого по формулам (5.41), (5.38).

5.3.2. Дифракционная цилиндрическая линза

Исследованию дифракционной цилиндрической линзы в приближении Фраунгофера посвящена работа [60]. В ней по контрастности рассчитанной картины Фраунгофера судят о характеристиках дифракционной цилиндрической линзы. Однако представляет интерес не только исследование контрастности поля в дальней зоне, но и анализ структуры сфокусированного излучения в фокальной области плоской цилиндрической линзы.

Фазовая функция цилиндрической линзы в параксиальном приближении имеет вид

$$\varphi(u) = -\frac{ku^2}{2f}, \quad |u| \leqslant a. \quad (5.42)$$

Координаты границ линейных зон при M уровнях квантования фазы дифракционной цилиндрической линзы имеют вид

$$u_n = \pm\sqrt{2n\lambda f/M}. \quad (5.43)$$

При этом общее количество линейных модулей

$$N = \operatorname{int}\left[Ma^2/\lambda f\right]. \quad (5.44)$$

На рис. 5.9 представлены распределения нормированной интенсивности $I(\tilde{x},0)/I_{\text{ид}}$ в фокальной плоскости дифракционной цилиндрической линзы с неравномерной дискретизацией (5.43) для $M = 2, 4, 16$ уровней градаций фазы. Здесь $\tilde{x} = x/a$ – безразмерная координата в фокальной плоскости, а

$$I_{\text{ид}} = \frac{4a^2}{\lambda f} \cdot I_0 \tag{5.45}$$

— интенсивность в фокусе неквантованной цилиндрической линзы.

Рис. 5.9 показывает, что при уменьшении числа уровней квантования фазы цилиндрической линзы происходит рост «фоновой» интенсивности. Фоновая интенсивность создается паразитными порядками квантования и приводит к подъему уровня боковых лепестков. Отметим, что для радиальной линзы фоновая интенсивность была незаметна (см. рис. 5.4). Несмотря на большее влияние порядков квантование на форму волнового фронта в полезном $+1$ порядке дифракции, расчетная эффективность квантованной цилиндрической линзы фактически совпадает со значениями эффективности для квантованной радиальной линзы, приведенными в табл. 5.1.

На рис. 5.10 представлены распределения нормированной интенсивности вдоль «оптической оси» дифракционной цилиндрической линзы с неравномерной дискретизацией (5.43) для $M = 2, 4, 16$ уровней квантования фазы при параметрах: $a =$ $= 1$ мм; $f = 25$ мм; $\lambda = 10{,}6$ мкм. Анализ рис. 5.10 показывает, что величина фокального сдвига (смещение точки максимума интенсивности из фокальной плоскости) зависит не только от физических параметров, но и от числа уровней квантования фазы и составляет для исследуемой линзы: $\Delta z = -0{,}05f = -1{,}25$ мм при $M = 16$; $\Delta z = 0$ при $M = 4$; и $\Delta z = 0{,}09f = 2{,}25$ мм при $M = 2$. При этом общий характер распределения и величина фокального сдвига с увеличением M асимптотически приближаются к распределению света от неквантованной цилиндрической линзы [48, 49]. Соответствие полученных результатов с данными работ по исследованию рефракционной цилиндрической линзы [48, 49] говорит о корректности использованного метода расчета интеграла Френеля–Кирхгофа.

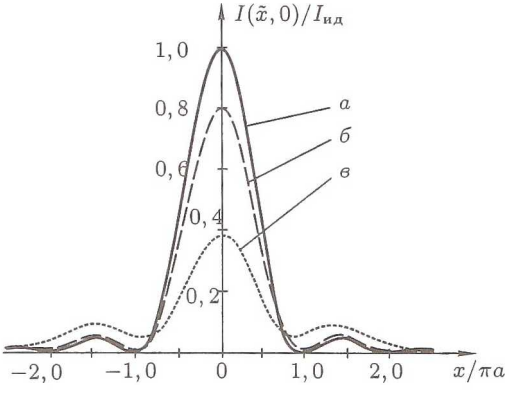


Рис. 5.9. Распределения нормированной интенсивности в фокальной плоскости дифракционной цилиндрической линзы для различного числа уровней квантования M: $M = 16$ (*а*), $M = 4$ (*б*), $M = 2$ (*в*)

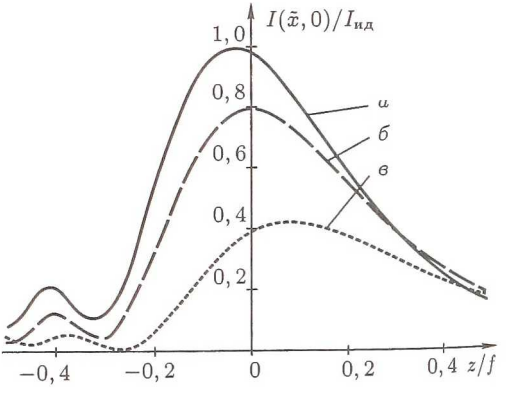


Рис. 5.10. Распределения нормированной интенсивности вдоль «оптической оси» дифракционной цилиндрической линзы для различного числа уровней квантования M: $M = 16$ (*а*), $M = 4$ (*б*), $M = 2$ (*в*)

5.3.3. Фокусатор в кольцо

На основе решения обратной задачи фокусировки лазерного излучения в работах [2, 62–64] предложено несколько типов фазовой функции фокусатора в кольцо. Для фокусировки плоского пучка света в кольцо используется также пара «аксикон — линза» [65, 66]. Фазовую функцию фокусатора в кольцо, действующего аналогично паре «аксикон — линза», несложно получить из общих уравнений (5.12)–(5.14) в виде

$$\varphi(r) = -k\sqrt{f_0^2 + (r - r_0)^2}, \quad 0 \leqslant r \leqslant a, \tag{5.46}$$

где r — полярный радиус, f_0 — фокусное расстояние, a — радиус фокусатора; r_0 — радиус кольца фокусировки. В силу радиальной симметрии задачи фокусировки в кольцо, слои имеют вид прямых, проходящих через центр фокусатора. Поскольку фаза вдоль слоя соответствует фазе линзы, то каждый узкий сектор фокусатора (см. рис. 5.11) представляет собой полоску собирающей линзы, фокусирующей в точку на расстояние r_0 от оптической оси. В параксиальном приближении фазовая функция (5.46) принимает вид

$$\varphi(r) = -kR_0 - \frac{kr^2}{2R_0} + \frac{krr_0}{R_0}, \quad R_0 = \sqrt{f_0^2 + r_0^2}, \tag{5.47}$$

соответствующий комбинации собирающей линзы с фокусом R_0 и аксикона (линейная по r фаза).

В силу широкого использования фокусировки в кольцо, в задачах лазерной технологии представляют интерес оценки ширины кольца, энергетической эффективности фокусировки и глубины резкости кольца, определяющей точность продольной юстировки плоскости фокусировки. Положенные в основу расчета фокусатора (5.46) геометрооптические представления не позволяют оценить указанные дифракционные параметры.

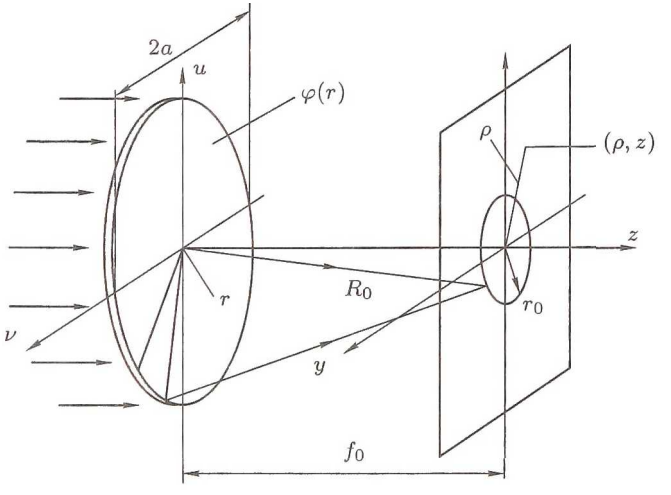


Рис. 5.11. Геометрия фокусировки в кольцо

Таблица 5.3. Зависимость энергетической эффективности E и ширины кольца $\Delta\rho$, определяемой по уровню 0,1 максимальной интенсивности, от радиуса кольца фокусировки r_0 при $f_0 = 750$ мм; $a = 12{,}8$ мм; $\lambda = 0{,}6328$ мкм; $N = \infty$, $M = \infty$

r_0мм	0,1	1,0	5,0
$\Delta\rho$мкм	67	64	65
$\Delta\rho/r_0$	0,67	0,064	0,013
$E(\%)$	89,4	88,5	87,7

Имеется ряд работ, посвященных исследованию фокусировки в кольцо [20, 24, 62–68]. В работах [65, 66] получены распределения интенсивности, формируемые парой «аксикон — линза» вблизи кольца для плоского и гауссова пучков. В работе [63] исследуется фокусировка сходящегося сферического пучка в кольцо с помощью ДОЭ с бинарной функцией комплексного пропускания $\operatorname{sign}(J_0(krr_0/f_0))$. В работах [62, 64] предложено несколько функций комплексного пропускания ДОЭ с кольцевым импульсным откликом и получены интегральные представления для интенсивности поля в плоскости фокусировки вблизи кольца. В то же время остается нерешенной задача исследования объемной структуры светового поля вблизи фокальной плоскости с учетом дискретизации и квантования фокусатора. При решении этой задачи необходимо использование как аналитических методов расчета [64, 68], так и средств вычислительного эксперимента [20, 24]. Это позволит выяснить какие дифракционные эффекты обусловлены несовершенством геометрооптической фазовой функции, а какие — технологией синтеза ДОЭ, т.е. дискретизацией и квантованием фазовой функции.

Проведем исследование фокусатора (5.46), который в параксиальном приближении соответствует паре «аксикон — линза» (5.47). Энергетическую эффективность фокусатора будем оценивать долей энергии освещающего пучка, попадающей в окрестность кольца фокусировки с шириной по уровню 0,1 максимальной интенсивности. Рис. 5.12 и табл. 5.3 позволяют проследить за изменением ширины кольца, энергетической эффективности и распределением интенсивности в плоскости фокусировки в зависимости от радиуса кольца.

Указанные результаты получены на основе аналитических методов расчета [64, 68] без учета эффектов дискретизации и квантования ($N = M = \infty$). Расчетные значения энергетической эффективности соответствуют данным работы [65].

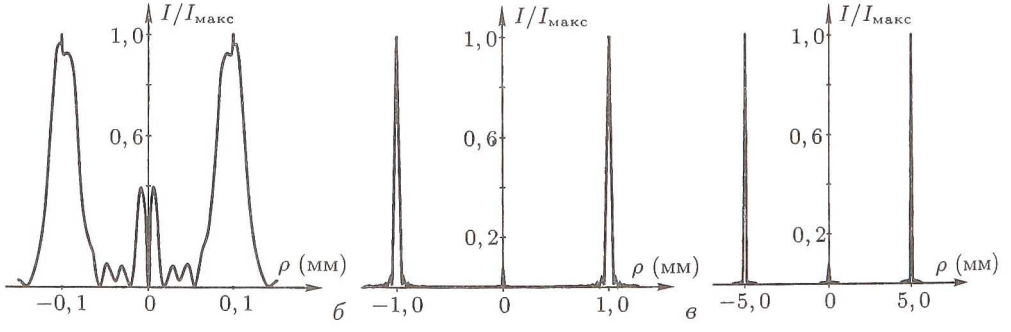


Рис. 5.12. Распределения интенсивности в фокальной плоскости фокусатора в кольцо при $f_0 = 750$ мм; $2a = 25{,}6$ мм; $\lambda = 0{,}6328$ мкм для различных радиусов кольца фокусировки r_0: $r_0 = 0{,}1$ мм ($а$); $r_0 = 1$ мм ($б$); $r_0 = 5$ мм ($в$)

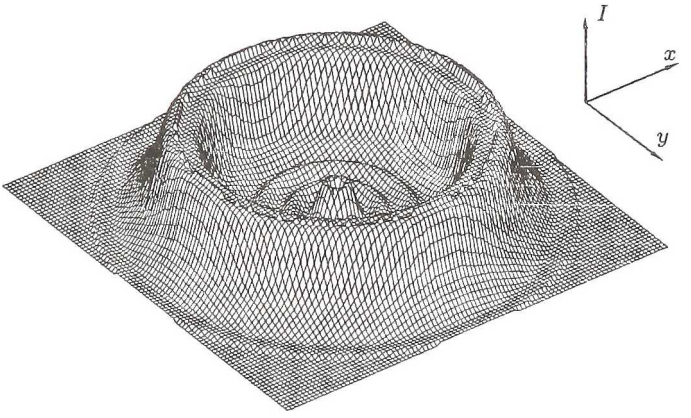


Рис. 5.13. Распределение интенсивности в фокальной плоскости фокусатора в кольцо с параметрами: $f_0 = 750$ мм; $2a = 25{,}6$ мм; $\lambda = 0{,}6328$ мкм; $r_0 = 0{,}1$ мм; $N = \infty$, $M = \infty$

Согласно [65] при указанных параметрах и $r_0 = 1$ мм, в кольцо с дифракционной шириной

$$\Delta = 2{,}28 \cdot (f_0 \lambda / a) = 84 \text{ мкм}, \tag{5.48}$$

определенной по 1-му минимуму, попадает 94% падающей на фокусатор энергии. По нашим данным, в кольцо с дифракционной шириной 64 мкм, определенной по уровню 0,1 максимальной интенсивности, попадает 90% энергии.

Рис. 5.12, 5.13 позволяют выявить интересную дифракционную особенности фокусировки в кольцо. С уменьшением радиуса кольца до $r_0 = 0,1$ мм дифракционная ширина кольца (см. табл. 5.3) становится сравнимой с радиусом кольца, так что кольцо начинает сливаться с центральным пятном.

В работе [62] обоснована целесообразность изготовления аксикона в виде плоского ДОЭ. В соответствии с принятой в [62] терминологией будем называть плоский аксикон фокусатором сходящегося сферического пучка в кольцо. Фазовая функция такого фокусатора имеет вид

$$\varphi(r) = \frac{krr_0}{\sqrt{f_0^2 + r_0^2}}. \tag{5.49}$$

Численное исследование фокусатора (5.49) проведем для $r_0 = 1$ мм. На рис. 5.14 приведены изофоты, характеризующие распределение интенсивности при отклонении от плоскости фокусировки при равномерной дискретизации ($N = 128$) для неквантованного фокусатора. В отличие от изофот линзы (рис. 5.6) видно, что энергия проходит не только через фокус, но и через область кольца. Для оценки влияния эффектов квантования фазы (5.49) при равномерной дискретизации ($N = 128$) были рассчитаны распределения интенсивности в плоскости кольца для $M = 16$; 4; 2 (рис. 5.15). Интенсивность на рис. 5.15 нормирована на значение интенсивности на геометрическом кольце при $M = \infty$. Распределение интенсивности в плоскости кольца при $M = 16$ практически неотличимо от соответствующего распределения на рис. 5.12*б*) при $M = \infty$. При $M = 4$ появляются дополнительные внутренние кольца вблизи усиливающегося центрального пятна. При $M = 2$ даже основное кольцо разбивается на несколько колец, а интенсивность центрального пятна превышает половину максимальной интенсивности. При этом энергетическая интенсивность E фокусировки падает (см. табл. 5.4). Значения энергетической эффективности в табл. 5.4 получены при $N = 128$ и при ширине кольца $\Delta\rho = 0{,}064$ мм,

Таблица 5.4. Зависимость эффективности E фокусатора (5.49) от числа уровней квантования M при равномерной дискретизации по радиусу и параметрах: $f_0 = 750$ мм; $a = 12{,}8$ мм; $\lambda = 0{,}6328$ мкм; $r_0 = 1$ мм; $N = 128$; ширина кольца фокусировки $\Delta\rho = 64$ мкм.

M	∞	16	4	2
$E(\%)$	76,7	75,7	62,4	59,7

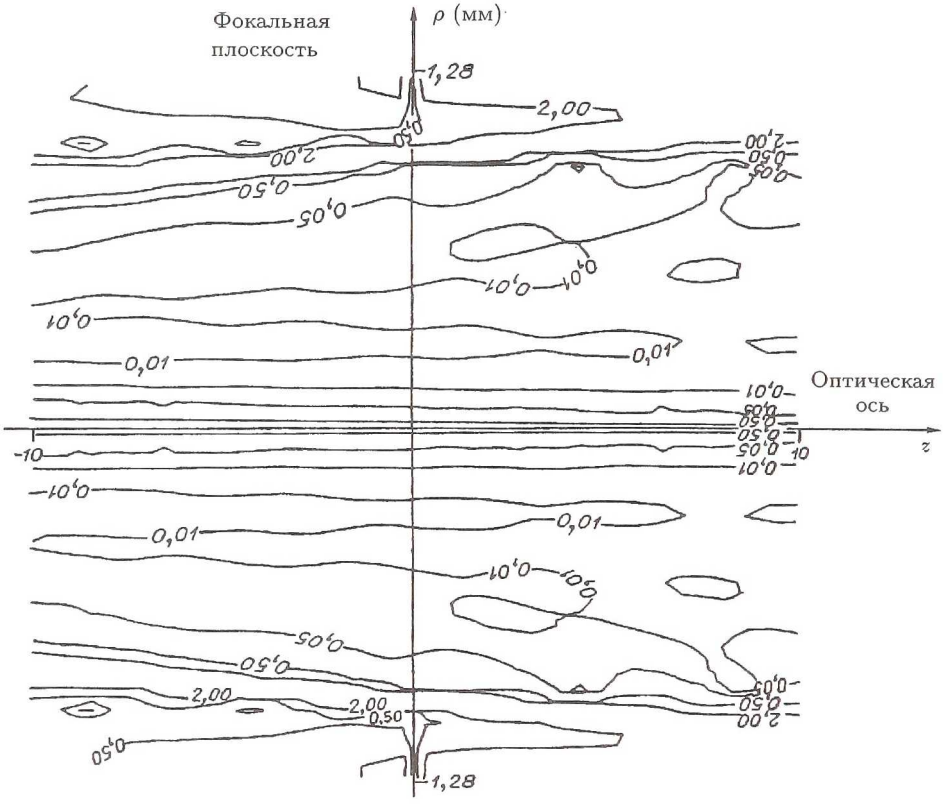


Рис. 5.14. Изофоты нормированного распределения интенсивности $I(\rho, z)/I(0,0)$ в меридиональном сечении фокальной области фокусатора в кольцо с параметрами: $f_0 = 750$ мм; $2a = 25{,}6$ мм; $\lambda = 0{,}6328$ мкм; $r_0 = 1$ мм; $N = 128$; $M = \infty$

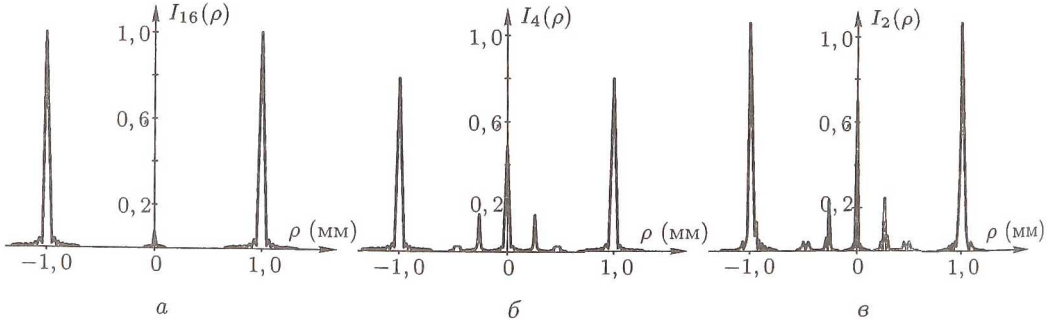


Рис. 5.15. Распределения нормированной интенсивности $I_M(\rho) = I_M(\rho,0)/I_\infty(r_0,0)$ в фокальной плоскости квантованного фокусатора в кольцо при параметрах: $f_0 = 750$ мм; $2a = 25{,}6$ мм; $\lambda = 0{,}6328$ мкм; $r_0 = 1$ мм; $N = 128$ для: $M = 16$ (*а*); $M = 4$ (*б*); $M = 2$ (*в*)

соответствующей уровню 0,1 от максимальной интенсивности при $N = M = \infty$. Сравнивая значение $E = 76,7\%$ при $M = \infty$ из табл. 5.4 с полученным теоретически, для $N = M = \infty$, значением $E = 88,5\%$ из табл. 5.3, можно сделать вывод, что радиальная дискретизация с 128 отсчетами снизила энергетическую эффективность примерно на 10%. Рис. 5.15 и данные табл. 5.4 позволяют заметить интересный эффект совместного влияния ошибок дискретизации и квантования фазы на работу фокусатора. Действительно, при $M = 16$ внутренние паразитные кольца на рис. 5.15*а* отсутствуют. В тоже время появление внутренних паразитных колец на рис. 5.15*б*, в при $M = 2$, 4 не может быть объяснено только квантованием фазы. Согласно результатам п. 2.9, квантование фазы фокусатора (5.49) при $N = \infty$ приводит к появлению дополнительных дифракционных порядков, соответствующих фокусировке в кольца с радиусами

$$R_p = |r_0 \cdot (1 + pM)|, \quad p = 0, \pm 1, \pm 2, \ldots \tag{5.50}$$

Паразитные порядки ($p \neq 0$ в (5.50)) усиливают центральный всплеск, однако, поскольку радиусы (5.50) при $p \neq 0$ превышают радиус r_0 кольца фокусировки, то наличие паразитных порядков квантования не позволяет объяснить появление внутренних колец на рис. 5.15*б,в*. Таким образом, именно совместное влияние ошибок дискретизации и квантования фазы, проявляется в появлении внутренних колец. Интересно также отметить, что фокусатор (5.49) при $M = 2$ дает почти такое же значение энергетической эффективности, как при $M = 4$ (см. табл. 5.4). Это объясняется тем, что при $M = 2$ помимо основного 0-го порядка, также -1-й порядок квантования ($p = -1$ в (5.51)) соответствует фокусировке в кольцо радиуса r_0. Согласно (2.313), в 0 и -1 порядке при $M = 2$ содержится примерно столько же энергии освещающего пучка, что и в основном 0-м порядке при $M = 4$. При этом интерференционное сложение 0-го и 1-го порядков приводит к «изрезанному» виду кольца на рис. 5.15*в*.

Остановимся теперь на исследовании квантованного фокусатора плоского пучка (5.46). Для M уровней квантования фазы из (5.46), (5.47) в параксиальном приближении получается следующая дискретизация:

$$\frac{k}{2R_0}(r_n - r_0)^2 = \frac{2\pi}{M}n, \quad 0 \leqslant r_n \leqslant a, \tag{5.51}$$

откуда

$$|r_n - r_0| = \sqrt{\frac{2\lambda R_0 n}{M}}, \quad 0 \leqslant r_n \leqslant a. \tag{5.52}$$

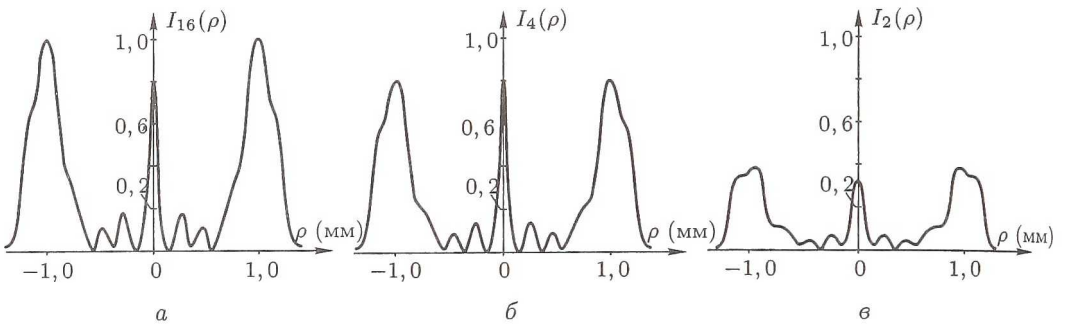
Индекс n в (5.52) определяет радиусы точек дискретизации относительно точки r_0. При этом для каждого значения n возможно одно или два значения радиуса кольца:

$$r_n = r_0 \pm \sqrt{\frac{2\lambda R_0 n}{M}}, \quad 0 \leqslant r_n \leqslant a, \quad n = 1, 2, \ldots, N_1. \tag{5.53}$$

Общее количество кольцевых модулей определяется по формулам

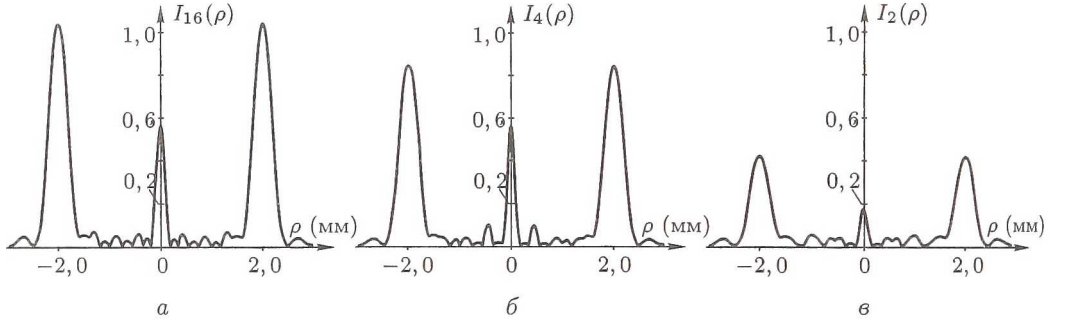
$$N = \operatorname{int}\left[\frac{r_0^2 M}{2\lambda R_0} + \frac{(a - r_0)^2 M}{2\lambda R_0}\right] + 1 \quad \text{при } a > r_0, \tag{5.54}$$

$$N = \operatorname{int}\left[\frac{r_0^2 M}{2\lambda R_0}\right] + 1 \quad \text{при } a < r_0. \tag{5.55}$$



а *б* *в*

Рис. 5.16. Распределения нормированной интенсивности $I_M(\rho) = I_M(\rho, 0)/I_\infty(r_0, 0)$ в фокальной плоскости квантованного фокусатора в кольцо при $f_0 = 500$ мм; $2a = 25{,}6$ мм; $\lambda = 10{,}6$ мкм; $r_0 = 1$ мм для $M = 16$ (*а*); $M = 4$ (*б*); $M = 2$ (*в*)



а *б* *в*

Рис. 5.17. Распределения нормированной интенсивности $I_M(\rho) = I_M(\rho, 0)/I_\infty(r_0, 0)$ в фокальной плоскости квантованного фокусатора в кольцо при $f_0 = 500$ мм; $2a = 25{,}6$ мм; $\lambda = 10{,}6$ мкм; $r_0 = 2$ мм для $M = 16$ (*а*); $M = 4$ (*б*); $M = 2$ (*в*)

Чтобы не увеличивать время расчета, которое прямо пропорционально N и обратно пропорционально λ, для исследования был выбран фокусатор инфракрасного излучения CO_2-лазера с параметрами: $f_0 = 500$ мм; $\lambda = 10{,}6$ мкм; $2a = 25{,}6$ мм. На рис. 5.16, 5.17 и в табл. 5.5 представлены результаты исследования фокусаторов в кольцо с радиусами $r_0 = 1$ мм и $r_0 = 2$ мм для различного числа уровней квантования фазы.

В таблице 5.5 используются следующие обозначения: Θ — уровень спада интенсивности, по которому определяются границы кольца фокусировки (ρ_-, ρ_+); (ρ_-, ρ_+) — границы кольца фокусировки, определяемые по уровню Θ; $\Delta\rho = \rho_+ - \rho_-$ — ширина кольца фокусировки по уровню Θ; $\Delta\rho_\Theta$, E_Θ — теоретические оценки ширины кольца и энергетической эффективности фокусировки для неквантованного фокусатора ($M = N = \infty$) [68].

Анализ приведенных результатов расчета позволяет сделать следующие выводы:

1) расчетный радиус центра кольца фокусировки увеличивается с уменьшением числа уровней квантования;

2) с уменьшением числа уровней квантования происходит увеличение ширины кольца и уменьшение энергетической эффективности;

3) ширина кольца и энергетическая эффективность почти не зависят от радиуса кольца r_0;

4) относительная величина центрального всплеска уменьшается с ростом радиуса кольца фокусировки r_0;

Таблица 5.5. Зависимость энергетической эффективности E и ширины кольца фокусировки $\Delta\rho$ по уровню Θ максимальной интенсивности от числа уровней квантования M для фокусатора в кольцо радиусом r_0 при $f_0 = 500$ мм; $\lambda = 10{,}6$ мкм; $a = 12{,}8$ мм

	$\Theta = 0{,}5$ ($\Delta\rho_\Theta = 0{,}408$ мм)					$\Theta = 0{,}2$ ($\Delta\rho_\Theta = 0{,}613$ мм)					$\Theta = 0{,}1$ ($\Delta\rho_\Theta = 0{,}701$ мм)				
M	E (%)	$\rho-$ (мм)	$\rho+$ (мм)	$\Delta\rho$ (мм)	E/E_Θ	E (%)	$\rho-$ (мм)	$\rho+$ (мм)	$\Delta\rho$ (мм)	E/E_Θ	ε	$\rho-$ (мм)	$\rho+$ (мм)	$\Delta\rho$ (мм)	E/E_Θ
При $r_0 = 1{,}0$ мм															
16	68,2	0,798	1,179	0,381	0,958	84,5	0,652	1,259	0,608	0,992	0,873	0,617	1,306	0,690	0,989
4	57,0	0,811	1,193	0,382	0,801	69,9	0,650	1,266	0,617	0,820	0,721	0,611	1,310	0,700	0,817
2	28,9	0,829	1,218	0,389	0,406	35,3	0,632	1,289	0,657	0,414	0,365	0,574	1,338	0,764	0,413
При $r_0 = 2{,}0$ мм															
16	68,6	1,797	2,189	0,392	0,963	84,4	1,684	2,282	0,588	0,991	0,872	1,643	2,328	0,658	0,988
4	57,8	1,803	2,220	0,417	0,812	69,5	1,687	2,305	0,618	0,816	0,717	1,644	2,348	0,704	0,812
2	28,2	1,863	2,242	0,379	0,396	0,345	1,729	2,330	0,601	0,405	0,364	1,615	2,382	0,767	0,412

5) теоретическая оценка дифракционной ширины кольца фокусировки $\Delta\rho_\Theta$ неквантованного фокусатора неплохо согласуется с численным расчетом.

Работа фокусатора в кольцо (5.46), (5.53)–(5.54) была также исследована для случая гауссова пучка с комплексной амплитудой

$$A_0(r) = \exp\left(-r^2/2\sigma^2\right) \tag{5.56}$$

при следующих параметрах: $f_0 = 500$ мм; $a = 12{,}8$ мм; $\lambda = 10{,}6$ мкм; $\sigma = 0{,}4a$; $r_0 = 1$ мм; $M = 2$; 4; 16.

Результаты расчетов для фокусатора гауссова пучка в кольцо представлены на рис. 5.18. Сравнение рисунков 5.16 и 5.18 позволяет отметить, что при гауссовом пучке увеличивается ширина кольца фокусировки и при $M \geqslant 4$ происходит подавление центрального всплеска, который сливается с окружающим фоном. Увеличение ширины кольца обусловлено фактическим уменьшением величины активной области фокусатора при выбранном σ.

При отсутствии кольцевого фотопостроителя можно изготовить фокусатор в кольцо на растровом генераторе изображений. Результаты исследования фокусатора (5.49) сходящегося сферического пучка в кольцо с дискретизацией, соответствующей

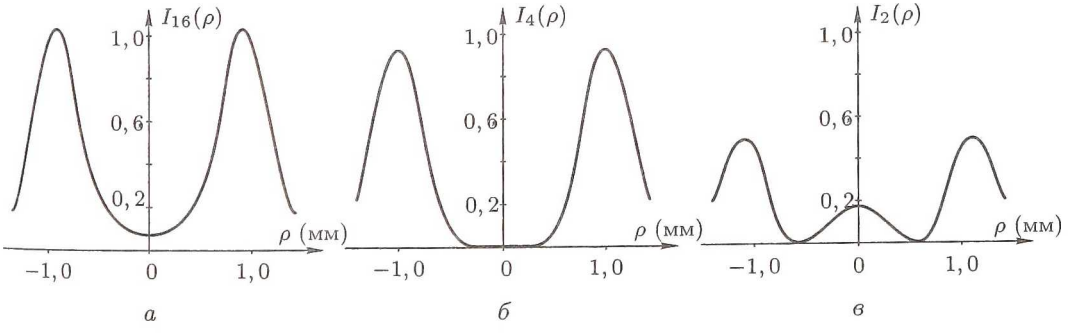
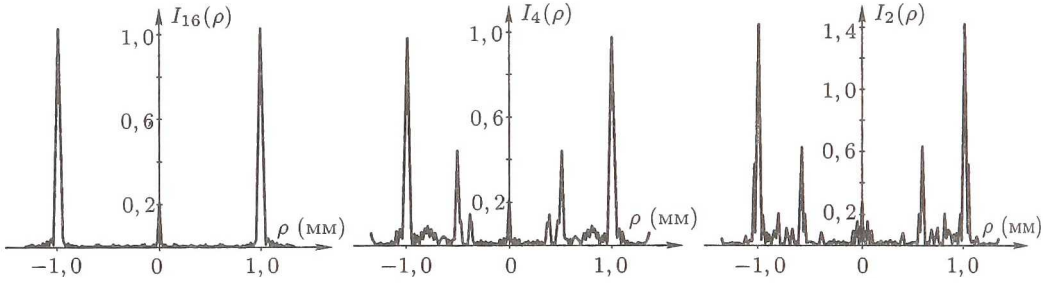


Рис. 5.18. Распределения нормированной интенсивности $I_M(\rho) = I_M(\rho, 0)/I_\infty(r_0, 0)$ в фокальной плоскости квантованного фокусатора в кольцо для гауссова пучка при $f_0 = 500$ мм; $2a = 25{,}6$ мм; $\sigma = 0{,}4a$; $\lambda = 10{,}6$ мкм; $r_0 = 1$ мм и числе уровней квантования $M = 16$ ($а$); $M = 4$ ($б$); $M = 2$ ($в$)

Таблица 5.6. Зависимость энергетической эффективности фокусатора (5.49) с прямоугольным растром дискретизации из $N \times N$ элементов от числа уровней M квантования фазы при ширине кольца $\Delta\rho = 64$ мкм и параметрах: $f_0 = 750$ мм; $\lambda = 0{,}6328$ мкм; $2a = 25{,}6$ мм; $r_0 = 1$ мм

N \ M	∞	16	4	2
128	47,2	47,0	37,8	32,7
256	75,8	74,1	60,2	58,4

разбиению апертуры фокусатора на $N \times N$ квадратных ячеек, представлены на рис. 5.19 и в табл. 5.6. Напомним, что используемые в этом случае квадратурные формулы [19, 20, 21, 45, 47] для расчета интеграла Кирхгофа–Френеля, предпола-



а *б* *в*

Рис. 5.19. Распределения нормированной интенсивности $I_M(\rho) = I_M(\rho, 0)/I_\infty(r_0, 0)$ в фокальной плоскости квантованного фокусатора в кольцо с прямоугольной дискретизацией при $N = 128$, $f_0 = 500$ мм; $2a = 25{,}6$ мм; $\lambda = 10{,}6$ мкм; $r_0 = 1$ мм и числе уровней квантования $M = 16$ (*а*); $M = 4$ (*б*); $M = 2$ (*в*)

гают кусочно-постоянную аппроксимацию функции комплексного пропускания фокусатора на сетке из $N \times N$ квадратных ячеек. Сравнение рис. 5.15, 5.19 и данных табл. 5.4, 5.6 свидетельствует о значительной близости результатов. Однако, для достижения равной энергетической эффективности необходимо, чтобы число элементов дискретизации N вдоль каждой оси прямоугольного растра было в 2 раза большим, чем количество колец при равномерной радиальной дискретизации.

5.3.4. Фокусатор в полукольцо

Фазовая функция геометрооптического фокусатора плоского пучка в полукольцо с радиусом r_0, может быть легко получена из фазовой функции фокусатора в кольцо. Действительно, при фокусировке в кольцо слои имеют вид прямых, а фаза вдоль «слоя-радиуса» соответствует фазе линзы, фокусирующей в точку кольца. Для фокусировки в полукольцо радиусом r_0, фазу линзы, фокусирующей в точку r_0 достаточно определить не вдоль «слоя-радиуса», а вдоль «слоя-диаметра». При этом фазовая функция фокусатора в полукольцо имеет вид [69]

$$\varphi(u, v) = \begin{cases} -k\sqrt{f_0^2 + (r - r_0)^2}, & u \geqslant 0, \\ -k\sqrt{f_0^2 + (r + r_0)^2}, & u < 0, \end{cases} \qquad r = \sqrt{u^2 + v^2} \leqslant a. \tag{5.57}$$

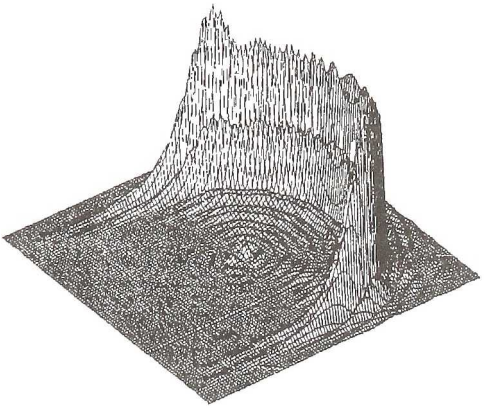


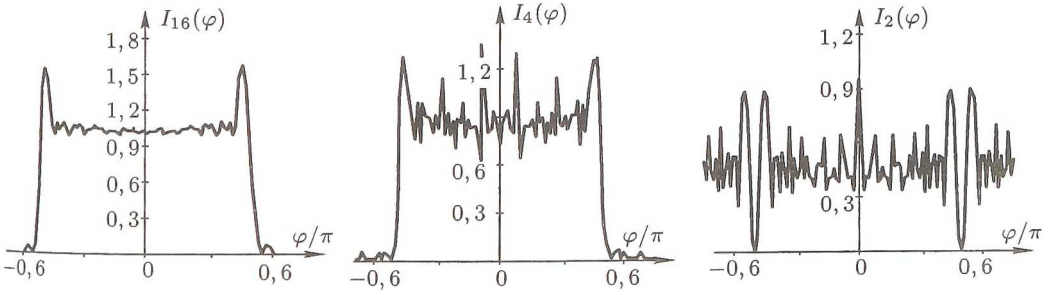
Рис. 5.20. Распределение интенсивности в фокальной плоскости фокусатора в полукольцо

Фазовая функция фокусатора сходящегося сферического пучка в полукольцо при $z = f_0$ может быть получена из (5.57) введением фазы рассеивающей линзы с фокусным расстоянием f_0.

В параксиальном приближении фаза фокусатора сходящегося сферического пучка принимает вид

$$\varphi(u,v) = \frac{krr_0}{\sqrt{f_0^2 + r_0^2}} \operatorname{sgn}(u), \quad r = \sqrt{u^2+v^2} \leqslant a. \tag{5.58}$$

На рис. 5.20, 5.21 представлены результаты исследования фокусатора (5.58) при $f_0 = 750$ мм; $\lambda = 0{,}6328$ мкм; $r_0 = 1$ мм; $2a = 25{,}6$ мм и при растровой дискретизации фазы на сетке из 128×128 отсчетов. Распределение интенсивности в фокальной плоскости фокусатора на рис. 5.20 показывает сохранение всплеска интенсивности на оптической оси, характерного для фокусатора в кольцо. На рис. 5.21 в полярных координатах представлены распределения интенсивности вдоль полукольца (при $\rho = r_0$, $-2\pi/3 \leqslant \varphi \leqslant 2\pi/3$) при различном числе уровней квантования фазы. Всплески интенсивности по краям полукольца могут быть обусловлены разрезом фазовой функции фокусатора (5.58) при $u = 0$. При 4-х уровнях градации фазы (рис. 5.21*б*) происходит уменьшение среднего уровня интенсивности вдоль линии фокусировки



а *б* *в*

Рис. 5.21. Распределения нормированной интенсивности $I_M(\varphi) = I_M(r_0,\varphi)/I_{16}(r_0,0)$ по углу вдоль полукольца фокусировки при (а) M = 16; (б) M = 4; (в) M = 2 и параметрах: $f_0 = 750$ мм; $\lambda = 0{,}6328$ мкм; $r_0 = 1$ мм; $2a = 25{,}6$ мм; $N = 128$

с одновременным увеличением неравномерности распределения интенсивности. Поскольку бинарный ДОЭ позволяет сформировать только центрально-симметричное распределение интенсивности, то при двух уровнях квантования фазы (рис. 5.21в) происходит разрушение области фокусировки, и вместо фокусировки в полукольцо происходит фокусировка в кольцо с перепадами интенсивности вдоль него.

5.3.5. Фокусатор в поперечный отрезок

Рассмотрим квадратный фокусатор плоского пучка в отрезок $|x| \leqslant d$ с постоянной линейной плотностью. Фазовая функция фокусатора может быть получена из общих уравнений (5.20), (5.21) в виде

$$\varphi(u,v) = -k\frac{(1-d/a)\,u^2+v^2}{2f}, \quad |u|<a, \quad |v|<a, \tag{5.59}$$

где $2a$ — размер стороны апертуры фокусатора.

Результаты численного расчета поля от фокусатора (5.59) с параметрами $f = 500$ мм; $\lambda = 10{,}6$ мкм; $2d = 2$ мм; $2a = 12{,}8$ мм при числе элементов дискретизации фазы $N_1 = N_2 = 128$ показаны на рис. 5.22, 5.23. При численном расчете интеграла Френеля–Кирхгофа предполагалось, что фокусатор состоит из $N_1 \times N_2$ прямоугольных модулей, фазовая функция в каждом из которых определяется в соответствии со своим значением (5.59) в центре ячейки.

Распределение интенсивности вдоль фокальной линии для различного числа уровней квантования ($M = 2, 4, 16$) фазы представлено на рис. 5.23, который показывает сохранение формы фокальной области для малых M с соответствующим уменьшением энергетической эффективности: $E = 75{,}3\%$ для $M = 16$, $E = 61{,}8\%$ для $M = 4$ и $E = 31{,}3\%$ для $M = 2$. При этом не представленный на рис. 5.23 график для $M = 256$ ($E = 76{,}3\%$) практически совпадает с графиком для $M = 16$, что доказывает отсутствие практического смысла в использовании более 16 уровней квантования фазы. Ширина фокальной линии ε, определяемая по уровню спада интенсивности $\Theta = 0{,}1$ при $M = 256$ и указанных параметрах составляет 0,62 мм. Приведенные значения энергетической эффективности соответствуют доли энергии, попадающей

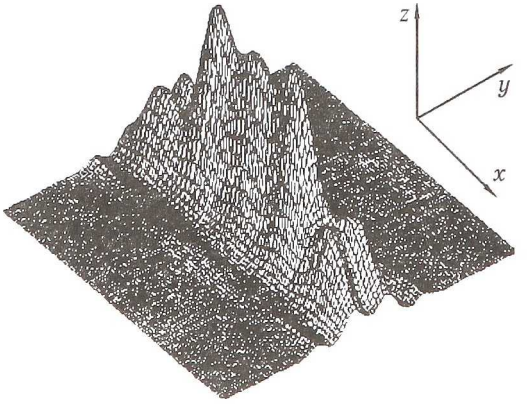


Рис. 5.22. Трехмерное распределение интенсивности в фокальной плоскости фокусатора в отрезок

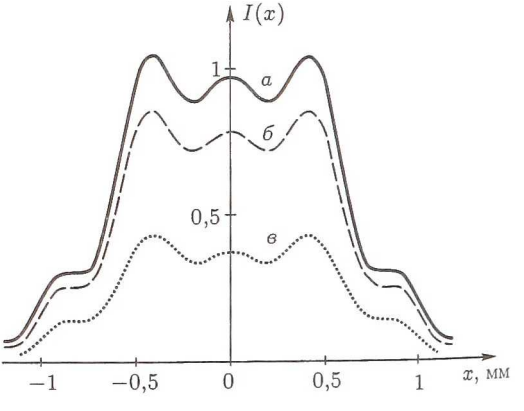


Рис. 5.23. Распределение нормированной интенсивности вдоль отрезка фокусировки в фокальной плоскости фокусатора (5.59)

в прямоугольник $2d\varepsilon$. Заметим, что размер дифракционного пятна $\varepsilon_0 = \lambda f/a$ составляет в этом случае 0,82 мм. Сохранение формы фокальной области при малом числе уровней квантования фазы объясняется структурой дифракционных порядков (2.315) квантованного ДОЭ. Дифракционные порядки, обусловленные квантованием фазы (5.59), соответствуют фокусировке в отрезки в плоскостях

$$z = F_n = \frac{f}{n}, \quad n = 1 + pM, \quad p = 0, \pm 1, \pm 2, \ldots \tag{5.60}$$

При этом изображения, создаваемые квантованным фокусатором в плоскости при $z = f$, соответствуют расфокусированным отрезкам. Ширина расфокусированных отрезков на порядок больше, чем дифракционная ширина отрезка фокусировки, что делает их влияние на полезный отрезок ($p = 0$ в (5.60)) ничтожно малым. Следует также отметить, что приведенное выше изменение энергетической эффективности E с уменьшением числа уровней квантования M, является пропорциональным квадрату модуля коэффициента Фурье c_0 в (2.313), описывающего долю энергии, фокусируемую в полезном +1-м порядке квантованного ДОЭ.

Фазовая функция квадратного ДОЭ, фокусирующего сходящийся сферический пучок в отрезок с равномерным распределением интенсивности имеет вид

$$\varphi(u, v) = \frac{kd}{2af} u^2, \quad |u| < a, \quad |v| < a. \tag{5.61}$$

Результаты исследования фокусатора (5.61) с параметрами $f = 500$ мм; $\lambda = 10{,}6$ мкм; $2d = 10$ мм; $2a = 25{,}6$ мм; $N_1 = 128$, $N_2 = 1$ представлены на рис. 5.24 и в табл. 5.7. Элементы дискретизации ДОЭ в данном случае имеют вид прямоугольников, большая сторона которых равна ширине фокусатора. На рис. 5.24 показано распределение интенсивности $I_M(x) = I_M(x, 0)/I_{256}(x, 0)$ в фокальной плоскости фокусатора при $M = 16$, 4, 2. В табл. 5.7 представлены значения энергетической эффективности $E(\varepsilon)$ для нескольких значений ширины фокальной линии, пропорциональных дифракционному размеру $\varepsilon_0 = \lambda f/a$. Среднеквадратичное отклонение полученного распределения интенсивности вдоль отрезка фокусировки от эталонного

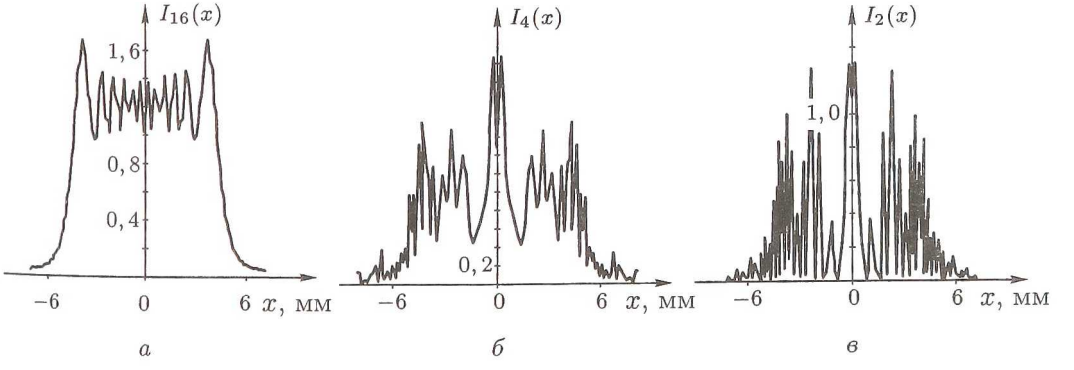


Рис. 5.24. Распределение интенсивности вдоль отрезка фокусировки в фокальной области фокусатора (5.61) для различного числа уровней квантования M: (*а*) $M = 16$, (*б*) $M = 4$, (*в*) $M = 2$

Таблица 5.7. Энергетическая эффективность $E(\varepsilon)$ и среднеквадратичное отклонение δ интенсивности от постоянного значения для различного числа уровней квантования M и элементов дискретизации N_1 фокусатора (5.61). Параметры фокусатора: $f = 500$ мм; $\lambda = 10{,}6$ мкм; $2d = 10$ мм; $2a = 25{,}6$ мм; $N_v = 1$; $\varepsilon_0 = \lambda f/a$

M	$E(\varepsilon)(\%)$					$\delta(\%)$
	$\varepsilon = 0{,}25 \cdot \varepsilon_0$	$\varepsilon = 0{,}5 \cdot \varepsilon_0$	$\varepsilon = 0{,}75 \cdot \varepsilon_0$	$\varepsilon = 1{,}0 \cdot \varepsilon_0$	$\varepsilon = 1{,}5 \cdot \varepsilon_0$	
$N_1 = 128$						
256	43,9	71,5	81,5	82,4	84,0	24,7
4	38,6	63,1	71,9	72,6	74,1	46,1
2	39,2	64,0	72,9	73,8	75,3	97,2
$N_1 = 256$						
256	45,0	73,3	83,5	83,9	84,0	24,3
4	38,8	63,4	72,2	72,6	74,1	46,0
2	39,4	64,3	73,4	73,8	75,3	91,9

определялось по формуле

$$\delta = \sqrt{\int_{-d}^{d} \left[I(x,0) - \tilde{I}(x) \right]^2 \mathrm{d}x \Big/ \int_{-d}^{d} \tilde{I}^2(x)\, \mathrm{d}x}, \tag{5.62}$$

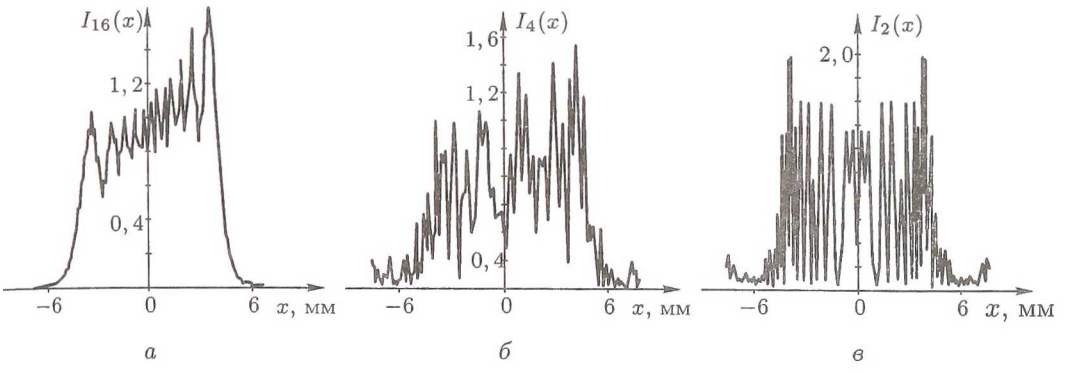
где $\tilde{I}(x)$ — требуемое распределение интенсивности вдоль отрезка фокусировки. Для фокусатора (5.61) $\tilde{I}(x)$ при $|x| \leqslant d$ постоянно и равно среднему значению интенсивности вдоль отрезка фокусировки.

Анализ рис. 5.24 и табл. 5.7 показывает, что при малом числе M градаций фазы (рис. 5.24*в*, *б*) происходит разрушение формы фокальной области, хотя значения энергетической эффективности остаются достаточно высокими. Разрушение отрезка фокусировки при малых M также объясняется структурой дифракционных порядков (2.315), возникающих при квантовании. Дифракционные порядки, обусловленные квантованием фазы (5.61), соответствуют фокусировке в отрезки длиной $d_p = 2d(1 + pM)$, $p = 0, \pm 1, \pm 2, \ldots$ в фокальной плоскости $z = f$.

При этом паразитные изображения отрезков при $p \neq 0$ накладываются на полезный отрезок $|x| \leqslant d$ и, вследствие интерференции, разрушают его. Энергетическая эффективность остается достаточно высокой (см. табл. 5.7) поскольку в данном случае не происходит рассеяние энергии по другим фокальным плоскостям. Разрушение отрезка фокусировки при малых M не удается компенсировать двукратным увеличением дискретизации (табл. 5.7).

Результаты испытаний фокусаторов лазерного излучения дальнего ИК-диапазона в кольцо и отрезок описаны в работе [70]. Энергетическая эффективность фокусировки, измеренная методом калиброванных диафрагм, составила для различных фокусаторов 60–75% [70], что соответствует приведенным данным моделирования фокусаторов в отрезок (5.59) с учетом технологических погрешностей изготовления дифракционного микрорельефа.

Представляет интерес исследование фокусаторов в отрезок с непостоянной линейной плотностью. На рис. 5.25 и в табл. 5.8 приведены результаты исследования фокусатора в отрезок с линейным распределением линейной плотности и с параметрами: $f = 500$ мм; $\lambda = 10{,}6$ мкм; $2d = 10$ мм; $2a = 25{,}6$ мм; $N_1 = 256$,



а *б* *в*

Рис. 5.25. Распределение интенсивности $I_M(x) = I_M(x,0)/I_{256}(x,0)$ вдоль отрезка фокусировки в фокальной области фокусатора (5.63) для различного числа уровней квантования M: (*а*) $M = 16$, (*б*) $M = 4$, (*в*) $M = 2$

$N_2 = 1$. Фазовая функция фокусатора может быть получена из (5.20), (5.21) и, при фокусировке сходящегося сферического пучка в отрезок с двукратным увеличением линейной плотности, имеет вид

$$\varphi(u,v) = -\frac{k}{f}\left[3du - \frac{ad}{9}\sqrt{\left(10+6\frac{u}{a}\right)^3}\right], \quad |u| < a, \quad |v| < a. \tag{5.63}$$

Характер распределения на рис. 5.25*а* в целом отражает двукратное линейное увеличение интенсивности вдоль отрезка фокусировки. Энергетическая эффективность фокусатора (5.63) близка к энергетической эффективности фокусатора в отрезок с равномерным распределением. В тоже время, увеличение среднеквадратичного отклонения интенсивности δ от линейного распределения при уменьшении M является более существенным. Поскольку бинарный ДОЭ позволяет сформировать только центрально-симметричное распределение интенсивности, то при $M = 2$ фокусатор становится не работоспособным (рис. 5.25*в*).

Таблица 5.8. Энергетическая эффективность $E(\varepsilon)$ и среднеквадратичное отклонение δ интенсивности от линейного распределения для различного числа уровней квантования M и элементов дискретизации N_1 фокусатора (5.63). Параметры фокусатора: $f = 500$ мм; $\lambda = 10{,}6$ мкм; $2d = 10$ мм; $2a = 25{,}6$ мм; $N_1 = 1$, $\varepsilon_0 = \lambda f/a$

M	$E(\varepsilon)(\%)$					$\delta(\%)$
	$\varepsilon = 0{,}25\cdot\varepsilon_0$	$\varepsilon = 0{,}5\cdot\varepsilon_0$	$\varepsilon = 0{,}75\cdot\varepsilon_0$	$\varepsilon = 1{,}0\cdot\varepsilon_0$	$\varepsilon = 1{,}5\cdot\varepsilon_0$	
$N_1 = 128$						
256	43,4	70,9	80,8	81,5	82,7	30,1
4	38,2	62,2	70,7	71,3	72,4	49,8
2	38,1	62,2	70,7	71,4	72,6	91,1
$N_1 = 256$						
256	44,5	72,2	82,8	83,6	84,8	29,4
4	38,2	62,2	70,8	71,4	72,4	47,9
2	38,2	62,2	70,7	71,4	72,6	81,5

Выше было проанализировано влияние дискретизации и квантования фазовой функции на работу простейших геометрооптических фокусаторов в отрезок. Для рассмотренных фокусаторов с квадратной апертурой длина слоя была постоянной. Это не позволило оценить эффекты неравномерного дифракционного уширения линии фокусировки при неквадратном освещающем пучке и наметить пути их коррекции.

В связи с этим проведем анализ более общей задачи фокусировки пучка с комплексной амплитудой $F_0(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp(i\varphi_0(\mathbf{u}))$, где $I_0(\mathbf{u})$ — интенсивность освещающего пучка, $\varphi_0(\mathbf{u})$ — фаза пучка, в отрезок $|x| \leqslant d$ в плоскости $z = f$ (см. рис. 5.26). Апертуру фокусатора предполагаем ограниченной кривыми $g_1(u)$, $g_2(u)$ и отрезками прямых $u = a$, $u = b$. Фазовая функция фокусатора в отрезок с распределением линейной плотности $\theta(x), |x| \leqslant d$, имеет вид (5.20), (5.21). Для уяснения сущности геометрооптического решения проведем дифракционный анализ поля в фокальной плоскости фокусатора в отрезок.

В параксиальном приближении интеграла Кирхгофа комплексная амплитуда в плоскости фокусировки имеет вид

$$w(\mathbf{x}) = \frac{k\exp(ikf)}{2\pi if}\int\limits_G w_0(\mathbf{u})\exp(i\varphi(\mathbf{u}))\exp\left(\frac{ik}{2f}(\mathbf{x}-\mathbf{u})^2\right)\mathrm{d}^2\mathbf{u}. \tag{5.64}$$

Подставляя (5.21) в (5.64) получаем

$$w(\mathbf{x}) = \frac{k\exp(ikf)}{2\pi if}\int\limits_G \sqrt{I_0(\mathbf{u})}\exp\left(\frac{ik}{f}\int\limits_{u_0}^{u} x(\xi)d\xi\right)\exp\left(-\frac{ik}{f}\mathbf{xu}\right)\mathrm{d}^2\mathbf{u}, \tag{5.65}$$

где функция $x(\xi)$ описывает лучевое соответствие между точками отрезка фокусировки и слоями на апертуре фокусатора. При дифракционном расчете [71–73] может быть получена аппроксимация интеграла (5.65), основанная на использовании метода стационарной фазы [12] при интегрировании поперек слоя, то есть по переменной u, причем

$$I(\mathbf{x}) = \left|\sqrt{\frac{k}{2\pi f}}\int\limits_{g_1(u_x)}^{g_2(u_x)}\sqrt{I_0(u_x,v)}\left(\left.\frac{\mathrm{d}x}{\mathrm{d}u}\right|_{u=u_x}\right)^{-1/2}\exp\left(-\frac{ikyv}{f}\right)\mathrm{d}v\right|^2, \tag{5.66}$$

где u_x — решение уравнения $x = x(u)$ относительно u.

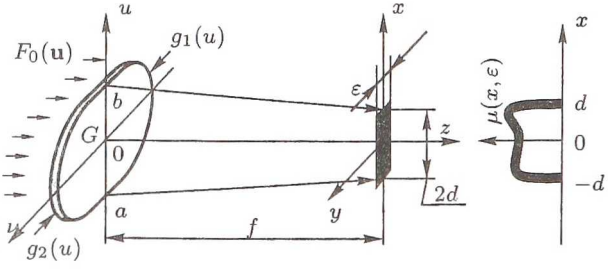


Рис. 5.26. Геометрия фокусировки в отрезок

Аппроксимация (5.66) хорошо описывает дифракционные эффекты в поперечном сечении фокального отрезка, но не вблизи концов. Введем функцию

$$\mu(x,\varepsilon)=\int\limits_{-\varepsilon/2}^{\varepsilon/2} I(x,y)\,\mathrm{d}y, \tag{5.67}$$

характеризующую распределение энергии в ε-окрестности отрезка фокусировки. Подставляя (5.66) в (5.67) получаем

$$\mu(x,\varepsilon)=\int\limits_{-\varepsilon/2}^{\varepsilon/2}\left|\sqrt{\frac{k}{2\pi f}}\int\limits_{g_1(u_x)}^{g_2(u_x)}\sqrt{I_0(u_x,v)}\left(\left.\frac{\mathrm{d}x}{\mathrm{d}u}\right|_{u=u_x}\right)^{-1/2}\exp\left(-\frac{ikyv}{f}\right)\mathrm{d}v\right|^2\mathrm{d}y. \tag{5.68}$$

При $\varepsilon_0\ll\varepsilon$, где ε_0 — дифракционная ширина отрезка, заменим пределы во внешнем интеграле на бесконечные и применим равенство Парсеваля. В результате уравнение (5.68) с учетом (5.20) примет вид

$$\mu(x,\varepsilon)=\theta(x),\quad \varepsilon\gg\varepsilon_0. \tag{5.69}$$

Согласно (5.69), линейной плотностью является интеграл от интенсивности, взятый в направлении поперек отрезка. Таким образом, геометрооптический подход (5.20), (5.21) ограничивается случаем $\varepsilon_0\ll\varepsilon$, что, например, не позволяет создать требуемое распределение интенсивности вдоль фокального отрезка. Используем дифракционное соотношение (5.68) для расчета «фокусатора с дифракционными поправками», формирующего требуемое распределение энергии вдоль фокального отрезка при произвольном ε [72, 73]. Для этого фазовую функцию фокусатора с дифракционными поправками также определим в виде (5.21), а дифференциальное уравнение для расчета $x(u)$ в (5.21) получим из (5.68) в виде

$$\frac{\mathrm{d}x(u)}{\mathrm{d}u}=\frac{1}{\mu\left(x(u),\varepsilon\right)}\int\limits_{-\varepsilon/2}^{\varepsilon/2}\left|\sqrt{\frac{k}{2\pi f}}\int\limits_{g_1(u)}^{g_2(u)}\sqrt{I_0(u,v)}\exp\left(-\frac{ikyv}{f}\right)\mathrm{d}v\right|^2\mathrm{d}y, \tag{5.70}$$
$$x(a)=-d,\quad x(b)=d.$$

При постоянной интенсивности освещающего пучка уравнение (5.70) примет вид

$$\frac{\mathrm{d}x}{\mathrm{d}u}=\frac{c}{\mu\left(x(u),\varepsilon\right)}\cdot\Phi\left(\frac{k\varepsilon}{4f}\left(g_2(u)-g_1(u)\right)\right)\cdot\left(g_2(u)-g_1(u)\right), \tag{5.71}$$
$$x(a)=-d,\quad x(b)=d,$$

где c — константа; $\Phi(\beta)=\mathrm{Si}(2\beta)-\beta\,\mathrm{sinc}^2(\beta)$,

$$\mathrm{Si}(\beta)=\int\limits_0^{\beta}\frac{\sin(x)}{x}\,\mathrm{d}x,\quad \mathrm{sinc}(\beta)=\frac{\sin(x)}{x}.$$

При $\varepsilon\ll\varepsilon_0$ разложим обе части уравнения (5.71) в ряд по степеням ε и учтем только

линейные члены разложения. При этом уравнение для $x(u)$ примет вид

$$\frac{\mathrm{d}x}{\mathrm{d}u} = \frac{1}{I\left(x(u),0\right)} \left| \sqrt{\frac{k}{2\pi f}} \int\limits_{g_1(u)}^{g_2(u)} \sqrt{I_0(u,v)}\,\mathrm{d}v \right|^2 \quad x(a) = -d, \quad x(b) = d, \tag{5.72}$$

где $I(x,0)$ — заданная интенсивность на отрезке при $y = 0$, $|x| \leqslant d$.

Интересно провести сравнительный анализ работоспособности геометрооптического фокусатора (5.20), (5.21) и фокусатора с дифракционными поправками (5.21), (5.70). Сравнение проведем на примерах фокусировки пучков круглого и кольцевого сечений в отрезок с постоянным распределением энергии. Функция $x(u)$ для фокусатора с дифракционными поправками при равномерном освещающем пучке круглого сечения радиуса R ищется из уравнения (5.71) при $g_1(u) = -\sqrt{R^2 - u^2}$ и $g_2(u) = \sqrt{R^2 - u^2}$. При освещающем пучке кольцевого сечения $x(u)$ находим из решения следующего дифференциального уравнения:

$$\frac{\mathrm{d}x(u)}{\mathrm{d}u} = \begin{cases} c \displaystyle\int\limits_{-\varepsilon/2}^{\varepsilon/2} \left[\sqrt{R_2^2 - u^2}\,\mathrm{sinc}\left(\frac{k}{f} y \sqrt{R_2^2 - u^2}\right) - \right. \\ \qquad \left. - \sqrt{R_1^2 - u^2}\,\mathrm{sinc}\left(\frac{k}{f} y \sqrt{R_1^2 - u^2}\right) \right]^2 \mathrm{d}y, & |u| < R_1, \\ c \displaystyle\int\limits_{-\varepsilon/2}^{\varepsilon/2} \left(R_2^2 - u^2\right) \mathrm{sinc}^2\left(\frac{k}{f} y \sqrt{R_2^2 - u^2}\right) \mathrm{d}y, & R_1 \leqslant |u| \leqslant R_2, \end{cases} \tag{5.73}$$

$$x(-R_2) = -d; \quad x(R_2) = d,$$

где c — константа, R_1 и R_2 — внутренний и внешний радиусы кольцевого освещающего пучка.

Следует отметить, что дифференциальные уравнения для функций $\xi(u)$ геометрооптических фокусаторов (п.5.2) имеют существенно другой вид:

$$\frac{\mathrm{d}x\,(u)}{\mathrm{d}u} = c\left(R^2 - u^2\right)^{1/2}, \quad x\,(-R) = -d, \quad x\,(R) = d, \tag{5.74}$$

при круглом освещающем пучке, и

$$\frac{\mathrm{d}x(u)}{\mathrm{d}u} = \begin{cases} \sqrt{R_2^2 - u^2} - \sqrt{R_1^2 - u^2}, & |u| < R_1, \\ \sqrt{R_2^2 - u^2}, & R_1 \leqslant |u| \leqslant R_2, \end{cases} \tag{5.75}$$

при кольцевом освещающем пучке.

Для характеристики качества работы фокусаторов введем значения энергетической эффективности E и среднеквадратичного отклонения δ. Величина

$$\mathrm{E}(\varepsilon) = \frac{\displaystyle\int\limits_{-d}^{d} \mu(x,\varepsilon)\,\mathrm{d}x}{\displaystyle\iint\limits_{G} I_0(\mathbf{u})\,\mathrm{d}^2\mathbf{u}} \tag{5.76}$$

характеризует долю энергии освещающего пучка, фокусируемую в ε-окрестности фокального отрезка. Величина

$$\delta(\varepsilon) = \frac{1}{\overline{I}} \left[\int_{-d}^{d} \left(\mu(x,\varepsilon) - \overline{I} \right)^2 \mathrm{d}x \right]^{1/2} \tag{5.77}$$

характеризует среднеквадратичное отклонение распределения интенсивности $\mu(x,\varepsilon)$ вдоль фокального отрезка от среднего значения $\overline{I} = 1/(2d) \int_{-d}^{d} \mu(x,\varepsilon)\,\mathrm{d}x$. При $\varepsilon \ll \varepsilon_0$, $\delta(\varepsilon)$ соответствует среднеквадратичному отклонению распределения интенсивности вдоль геометрического отрезка от постоянной величины.

Таблица 5.9. Характеристики качества фокусировки для геометрооптического фокусатора и фокусаторов с дифракционными поправками при освещающем пучке круглого сечения

ε	Геометрооптический фокусатор		Фокусаторы с дифракционными поправками	
	$E(\varepsilon),\%$	$\delta(\varepsilon),\%$	$E(\varepsilon),\%$	$\delta(\varepsilon),\%$
$\varepsilon \ll \varepsilon_0$	—	24,3	—	14,6
$\varepsilon_0/2$	65,6	27,1	65,2	13,4
ε_0	85,9	18,1	85,6	14,5
$3\varepsilon_0/2$	88,5	17,6	88,2	16,1
$2\varepsilon_0$	95,1	15,8	93,1	14,9

В левой части табл. 5.9 для различных ε приведены расчетные значения $E(\varepsilon)$ и $\delta(\varepsilon)$ для геометрооптического фокусатора в отрезок при освещающем пучке круглого сечения. В правой части табл. 5.9 приведены аналогичные значения для фокусаторов с дифракционными поправками, рассчитанные при значениях ε, указанных в первой колонке таблицы, где $\varepsilon_0 = \lambda f/R$ — дифракционная ширина в центре отрезка фокусировки. Данные табл. 5.9 получены при следующих параметрах: $\lambda = 1{,}06$ мкм, $f = 100$ мм, $2d = 0{,}68$ мм, радиус освещающего пучка $R = 5$ мм. Анализ данных показывает, что при фактически равной энергетической эффективности $E(\varepsilon)$, фокусатор с дифракционными поправками позволяет сформировать более равномерное распределение энергии в ε-окрестности фокального отрезка. При $\varepsilon \ll \varepsilon_0$, $\varepsilon = \varepsilon_0/2$ и $\varepsilon = \varepsilon_0$, среднеквадратичное отклонение $\delta(\varepsilon)$ для фокусатора с дифракционными поправками меньше, чем для геометрооптического фокусатора в 1,7; 2 и 1,25 раза, соответственно.

Графики функций $\mu(x,\varepsilon)$ для геометрооптического фокусатора приведены на рис. 5.27, а для фокусатора с дифракционными поправками — на рис. 5.28.

При увеличении ε относительно ε_0 фокусатор с дифракционными поправками не позволяет существенно уменьшить среднеквадратичное отклонение $\delta(\varepsilon)$ по сравнению с геометрооптическим фокусатором.

В табл. 5.10 приведены значения $E(\varepsilon)$ и $\delta(\varepsilon)$ для геометрооптического фокусатора и фокусатора с дифракционными поправками при кольцевом освещающем пучке. При кольцевом освещающем пучке дифракционная ширина в центре фокального отрезка $\Delta = 2\lambda f/(R_2 - R_1)$. Расчет проводился при следующих параметрах: $\lambda = 1{,}06$ мкм, $f = 100$ мм, $2d = 1$ мм и радиусах освещающего пучка $R_1 = 3$ мм и $R_2 = 5$ мм.

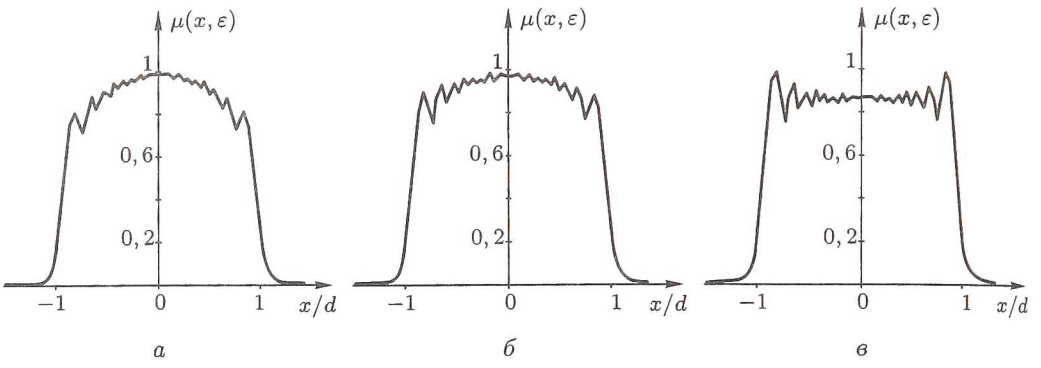


Рис. 5.27. Функция $\mu(x,\varepsilon)$ для геометрооптического фокусатора при равномерном освещающем пучке круглого сечения: (*а*) $\varepsilon \ll \varepsilon_0$, (*б*) $\varepsilon = \varepsilon_0/2$, (*в*) $\varepsilon = \varepsilon_0$

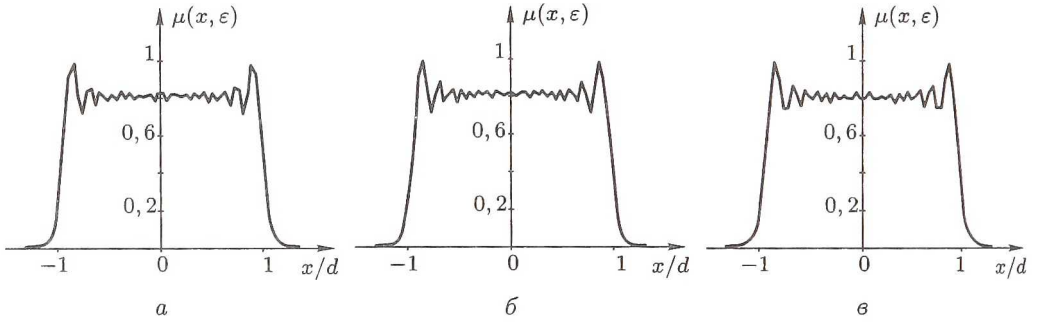


Рис. 5.28. Функция $\mu(x,\varepsilon)$ для фокусатора с дифракционными поправками при равномерном освещающем пучке круглого сечения: (*а*) $\varepsilon \ll \varepsilon_0$, (*б*) $\varepsilon = \varepsilon_0/2$, (*в*) $\varepsilon = \varepsilon_0$

Согласно данным табл. 5.10, для геометрооптического фокусатора распределение энергии $\mu(x,\varepsilon)$ наиболее неравномерно при $\varepsilon \ll \varepsilon_0$ и $\varepsilon = \varepsilon_0/4$: $\delta(\varepsilon \ll \varepsilon_0) = 28,7\%, \delta(\varepsilon_0/4) = 36\%$ (см. рис. 5.29*а* и 5.29*б*). При указанных параметрах фокусатор с дифракционными поправками характеризуется более чем в два раза меньшим среднеквадратичным отклонением (рис. 5.30*а*, 5.30*б*). При $\varepsilon = \varepsilon_0/2$ фокусатор с дифракционными поправками позволяет уменьшить $\delta(\varepsilon)$ уже только в 1,26 раза (см. рис. 5.29*в* и 5.30*в*), а при $\varepsilon = \varepsilon_0$ геометрооптический фокусатор и фокусатор с дифракционными поправками работают практически одинаково.

Таблица 5.10. Характеристики качества фокусировки для геометрооптичекого фокусатора и фокусаторов с дифракционными поправками при кольцевом освещающем пучке

ε	Геометрооптический фокусатор		Фокусаторы с дифракционными поправками	
	$E(\varepsilon),\%$	$\delta(\varepsilon),\%$	$E(\varepsilon),\%$	$\delta(\varepsilon),\%$
$\varepsilon \ll \varepsilon_0$	—	28,3	—	13,3
$\varepsilon_0/4$	62,3	36,3	62,2	16,2
$\varepsilon_0/2$	83,7	18,6	83,6	14,8
$3\varepsilon_0/4$	89,7	18,2	89,5	15,0
$2\varepsilon_0$	90,9	15,6	90,8	15,2

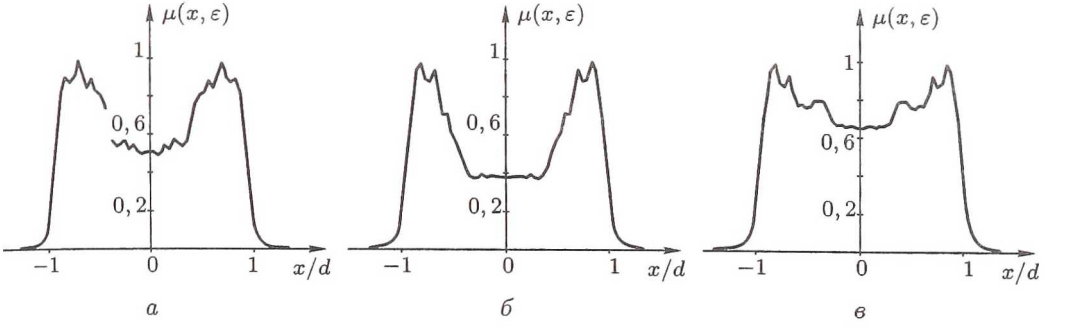


Рис. 5.29. Функция $\mu(x,\varepsilon)$ для геометрооптического фокусатора при равномерном освещающем пучке кольцевого сечения: (*а*) $\varepsilon \ll \varepsilon_0$, (*б*) $\varepsilon = \varepsilon_0/4$, (*в*) $\varepsilon = \varepsilon_0/2$

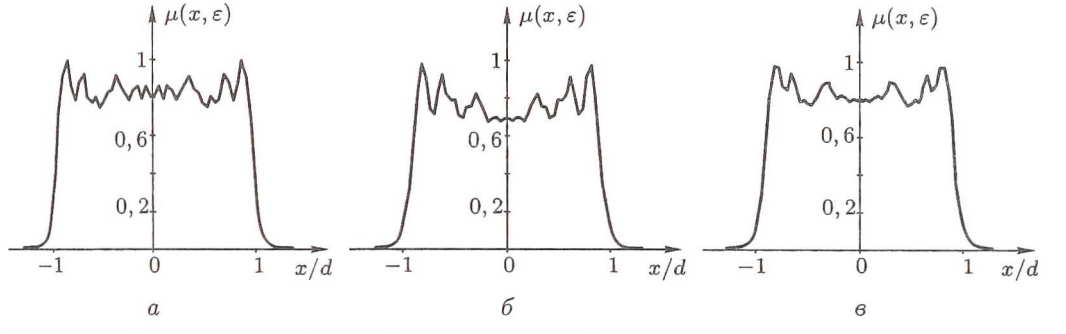


Рис. 5.30. Функция $\mu(x,\varepsilon)$ для фокусатора с дифракционными поправками при равномерном освещающем пучке кольцевого сечения: (*а*) $\varepsilon \ll \varepsilon_0$, (*б*) $\varepsilon = \varepsilon_0/4$, (*в*) $\varepsilon = \varepsilon_0/2$

Проведенное исследование показывает, что фокусаторы с дифракционными поправками, при освещающих пучках круглого и кольцевого сечений, позволяют сформировать требуемое распределение энергии $\mu(x,\varepsilon)$вдоль фокального отрезка при произвольном ε с погрешностью в 13–16%. Геометрооптический фокусатор является частным случаем фокусатора с дифракционными поправками и при значении ε, равном дифракционной ширине отрезка фокусировки, формирует распределение энергии $\mu(x,\varepsilon)$ с ошибкой порядка 15%.

5.3.6. Составной фокусатор в крест

Составным или сегментированным фокусатором называется фокусатор, апертура которого разбита на несколько областей (сегментов), самостоятельно фокусирующих излучение в отдельные области фокусировки. Рассмотренный в пункте 5.3.4 фокусатор в полукольцо занимает промежуточное место между обычным геометрооптическим фокусатором, представляемым в виде непрерывного набора фокусирующих слоев и составным фокусатором, так как, несмотря на наличие характерной линии разрыва фазы при $u = 0$, обе части фокусатора в полукольцо фокусируют в одну и ту же фокальную линию. Особую актуальность составные фокусаторы имеют для лазерной маркировки [74–77], в частности для лазерной маркировки алфавитно-цифровыми символами. Учитывая сложный характер фазовых функций такого рода ДОЭ, представляет интерес дифракционное исследование работоспособности составных фокусаторов.

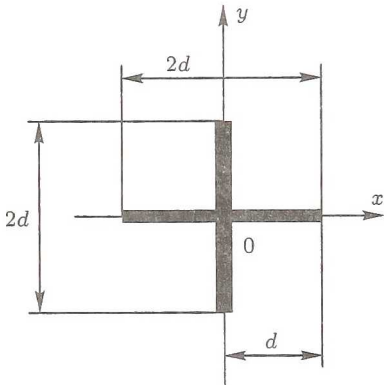


Рис. 5.31. Область фокусировки фокусатора в крест

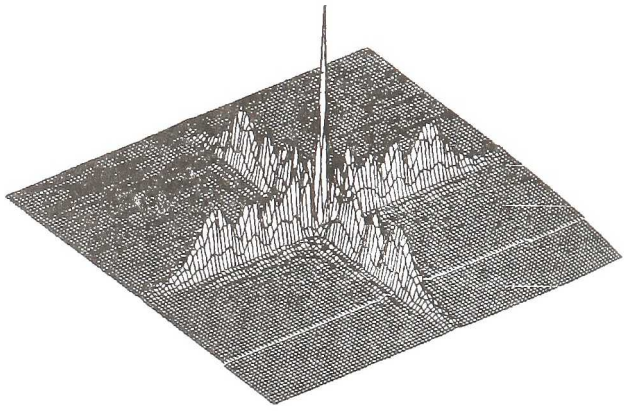


Рис. 5.32. Распределение интенсивности в фокальной плоскости фокусатора в крест

В качестве примера составного фокусатора рассмотрим фокусатор сходящегося сферического пучка в крест (см. рис. 5.31). Апертуру фокусатора $2a \times 2a$ будем предполагать состоящей из четырех сегментов $a \times a$, каждый из которых фокусирует в отрезок креста. В этом случае фазовая функция составного фокусатора имеет вид

$$\varphi(u,v) = \begin{cases} \dfrac{kd}{2af_0}u^2, & u \cdot v \geqslant 0 \\ \dfrac{kd}{2af_0}v^2, & u \cdot v < 0 \end{cases}, \quad |u| \leqslant a, \quad |v| \leqslant a. \tag{5.78}$$

Распределение интенсивности в фокальной области фокусатора (5.78) с параметрами $f_0 = 750$ мм; $\lambda = 0{,}6328$ мкм; $d = 1$ мм; $2a = 25{,}6$ мм; $N_1 = N_2 = 128$ представлено на рис. 5.32, на котором четко видна структура крестообразно сфокусированного излучения. Анализ рис. 5.32 позволяет наметить пути улучшения фазовой функции фокусатора (5.78). Благодаря уникальности точки пересечения фокальных линий креста (оптическая ось) практически отсутствуют интерференционные эффекты, характерные для каустик подобного вида. Однако имеется значительный всплеск интенсивности в области оптической оси, который можно устранить в ходе проектирования фокусатора.

Для этого следует использовать составной фокусатор в крест без центра (см. рис. 5.33), фазовая функция которого имеет вид

$$\varphi(u,v) = \begin{cases} \dfrac{k}{2f_0}\left[\dfrac{2d-\beta}{2a}u^2 + \beta u\right], & u \cdot v \geqslant 0 \\ \dfrac{k}{2f_0}\left[\dfrac{2d-\beta}{2a}u^2 + \beta u\right], & u \cdot v < 0 \end{cases}, \quad |u| \leqslant a, \quad |v| \leqslant a. \tag{5.79}$$

Распределение интенсивности в фокальной плоскости фокусатора (5.79) с параметрами $f_0 = 750$ мм; $\lambda = 0{,}6328$ мкм; $d = 1$ мм; $\beta = 0{,}2$ мм; $2a = 25{,}6$ мм; $N_1 = N_2 = 128$ представлено на рис. 5.34. Сравнение рис. 5.32 и 5.34 позволяет сделать вывод о том, что с помощью оптимизации значения β возможно устранение центрального всплеска.

Значения энергетической эффективности $E(\varepsilon)$ фокусатора (5.79) с различным числом уровней квантования фазовой функции представлены в табл. 5.11.

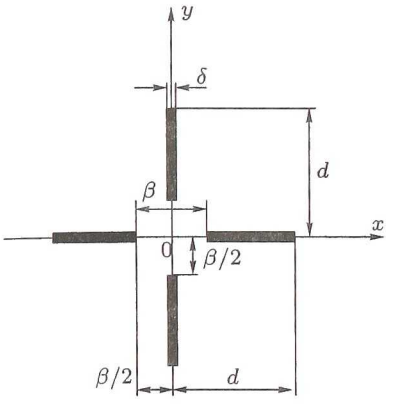


Рис. 5.33. Область фокусировки фокусатора в крест без центра

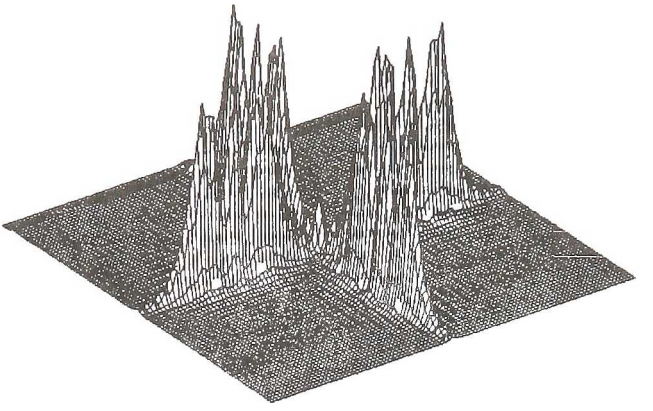


Рис. 5.34. Распределение интенсивности в фокальной плоскости фокусатора в крест без центра

Здесь $E(\varepsilon)$ — доля падающей на фокусатор энергии, попадающая в область фокусировки D (крест без центра с дифракционной шириной линий ε — см. рис. 5.32): $D = \{(x,y) : [(\beta/2 \leqslant |x| \leqslant d) \wedge |y| \leqslant \varepsilon] \vee [(\beta/2 \leqslant |y| \leqslant d) \wedge |x| \leqslant \varepsilon]\}$, где $\wedge, \vee$ — обозначения логических операций над множествами, соответственно конъюнкции и дизъюнкции. В табл. 5.11 приведены значения энергетической эффективности $E(\varepsilon)$ для набора значений дифракционной ширины ε в пределах: 0,02 мм$\leqslant \varepsilon \leqslant$ 0,14 мм. Следует отметить, что ширина фокального пятна линзы с такими же параметрами: по уровню спада интенсивности $\Theta = 0{,}1$ определяется по формуле $\varepsilon_0 = 2{\cdot}2{,}73\,(f_0/ka)$ и для указанных параметров составляет 0,032 мм [34]. Учитывая результаты исследования фокусаторов в отрезок и тот факт, что в каждый отрезок креста фокусируется только четверть апертуры составного фокусатора, следует ожидать увеличения ширины фокального пятна по крайней мере в 4 раза, что согласуется с данными табл. 5.11.

Таблица 5.11. Зависимость энергетической эффективности $E(\varepsilon)$ фокусатора (5.79) с параметрами $f_0 = 750$ мм; $\lambda = 0{,}6328$ мкм; $d = 1$ мм; $\beta = 0{,}2$ мм; $2a = 25{,}6$ мм; $N_1 = N_2 = 128$ от числа уровней M квантования фазы

M \ ε(мм)	0,02	0,04	0,06	0,08	0,1	0,12	0,14
256	0,452	0,712	0,788	0,792	0,805	0,825	0,832
16	0,450	0,709	0,875	0,789	0,802	0,823	0,830
4	0,385	0,608	0,674	0,677	0,688	0,706	0,712
2	0,369	0,593	0,657	0,661	0,671	0,689	0,695

5.3.7. Фокусатор в продольный отрезок

ДОЭ с повышенной глубиной фокуса или фокусатор в отрезок оптической оси впервые был предложен в работе [1]. В последующем было предложено несколько вариантов расчета фазовых функций таких ДОЭ [20, 78–83]. Использование ДОЭ с повышенной глубиной фокуса актуально для лазерных проигрывателей компакт-дисков [84], для получения оптического разряда в газе [85], для создания опорной

световой линии в метрологии [79], для бесконтактных измерений [82], для устройств ввода-вывода изображений в компьютере [79], для лазерных технологических установок [86].

Как известно, глубина фокуса высококачественного объектива оценивается [34] по формуле $L \sim \lambda f_0^2/a^2$ и составляет, например, $L \sim 0,1$ мм для объектива с относительным отверстием 1:10. Для повышения глубины фокуса требуется направленное введение сферической аберрации.

Фазовая функция фокусатора в продольный отрезок может быть получена из общих уравнений (5.13)–(5.15). Согласно (5.15), при фокусировке в отрезок оптической оси слои являются окружностями. Тогда, с учетом радиальной симметрии задачи, для расчета фазовой функции фокусатора получим следующую систему уравнений:

$$\left.\begin{aligned} \varphi(r) &= -k\int_0^r \frac{k\rho}{\sqrt{\rho^2 + (f_0 + z(\rho))^2}}\,\mathrm{d}\rho - \varphi_0(r), \\ \theta(z)\frac{\mathrm{d}z(\rho)}{\mathrm{d}\rho} &= I_0(\rho)\cdot 2\pi\rho, \quad z(0) = 0, \quad z(a) = L. \end{aligned}\right\} \tag{5.80}$$

Здесь r — радиальная координата в плоскости фокусатора; $\theta(z)$ — линейная плотность на отрезке фокусировки $[0, L]$; $I_0(r)$ и $\varphi_0(r)$ — соответственно интенсивность и фаза освещающего пучка; a — радиус фокусатора. Первое уравнение в (5.80) соответствует уравнению (5.14) и получено из уравнения наклонов (5.2) в полярных координатах. Второе уравнение в (5.80) соответствует закону сохранения светового потока (5.13), записанному в дифференциальной форме.

При $f_0 \gg a$ и постоянной линейной плотности $\theta(z) = \mathrm{const}$ получаем из (5.80) формулы

$$\varphi(r) = \varphi_0 - \left(\frac{k}{2c}\right)\ln\left[1 + 2c\sqrt{r^2 + (f_0 - cr^2)^2} - 2cf_0 + 2c^2r^2\right], \qquad c = \frac{L}{a^2}, \quad 0 \leqslant r \leqslant a, \tag{5.81}$$

для плоского освещающего пучка, и

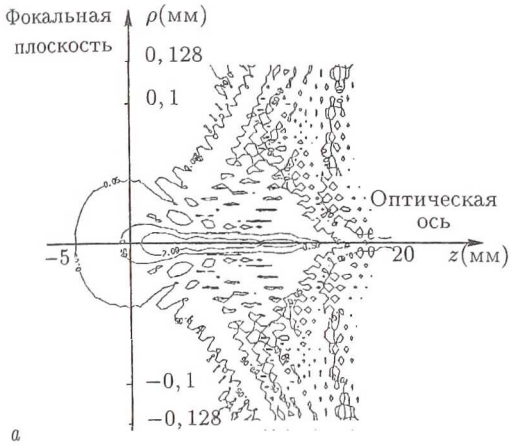
$$\varphi(r) = \varphi_0 - \left(\frac{k}{2c}\right)\ln\left[1 + 2c\sqrt{r^2 + (f_0 - cr^2)^2} - 2cf_0 + 2c^2r^2\right] + k\sqrt{r^2 + f_0^2}, \tag{5.82}$$

для фокусировки в продольный отрезок сходящегося сферического пучка. Остановимся на исследовании фокусатора (5.82) с равномерной дискретизацией.

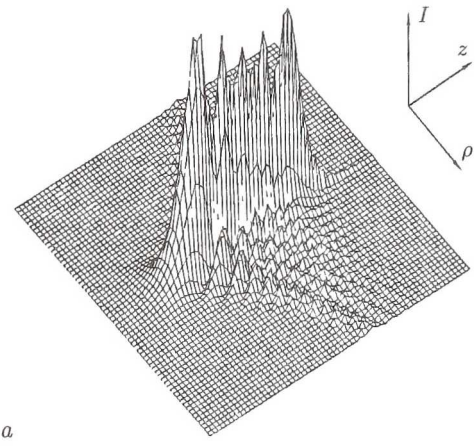
Для определения реальной глубины фокуса и размера кружка рассеяния, обеспечиваемых фокусатором при различных параметрах квантования, был проведен вычислительный эксперимент при значениях $f_0 = 300$ мм, $2a = 25,6$ мм, $L = 15$ мм, $\lambda = 0,6328$ мкм и числе кольцевых модулей дискретизации $N = 128$.

Хорошее представление о работе элемента (5.82) дают изофоты интенсивности в меридиональном сечении фокальной области (рис. 5.35) и соответствующие им трехмерные изображения (рис. 5.36) для различного числа уровней квантования: $M = \infty$; 4; 2.

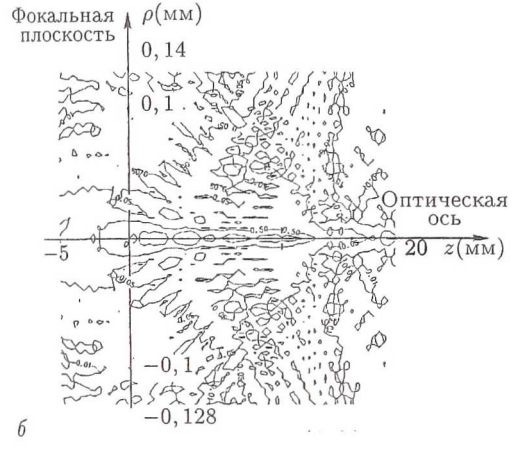
График распределения интенсивности вдоль оптической оси для тех же параметров ДОЭ и $M = \infty$ представлен на рис. 5.37. Нормировка произведена на значение интенсивности в начальной точке отрезка. Среднеквадратичное отклонение интенсивности вдоль отрезка фокусировки от постоянной величины составляет 37,4%. Заметны спад интенсивности к концам отрезка фокусировки и значительные перепады



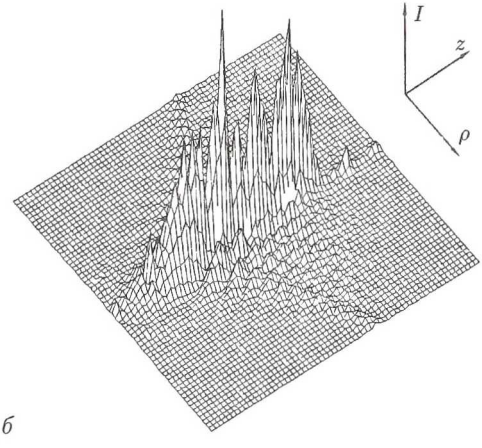
а



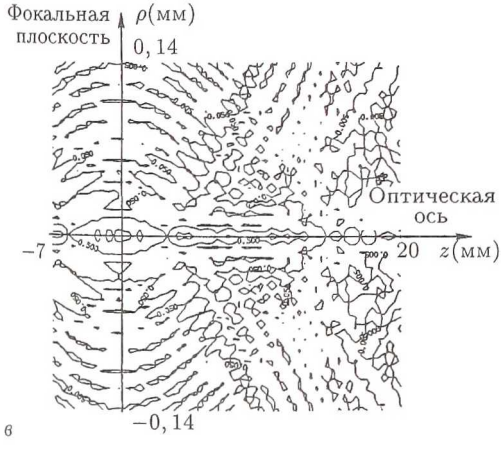
а



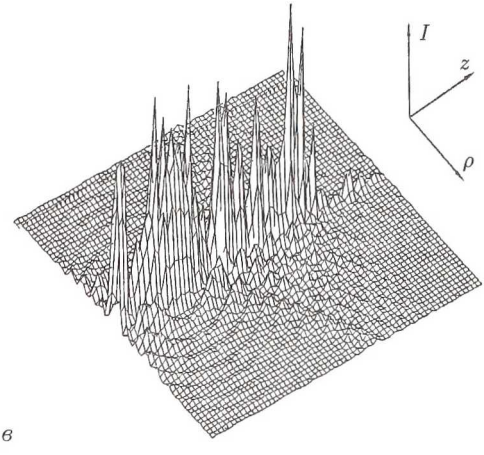
б



б



в



в

Рис. 5.35. Изофоты распределений интенсивности $I_M(\rho, z)/I_\infty(0,0)$ в меридиональном сечении фокальной области ДОЭ с повышенной глубиной фокуса. Параметры ДОЭ: $f_0 = 300$ мм; $\lambda = 0{,}6328$ мкм; $2a = 25{,}6$ мм; $L = 15$ мм; $N = 128$; (*а*) $M = \infty$; (*б*) $M = 4$; (*в*) $M = 2$

Рис. 5.36. Распределения интенсивности в меридиональной плоскости вблизи фокальной линии (-5 мм $< z <$ 20 мм; $0 < \rho <$ 128 мкм) ДОЭ с повышенной глубиной фокуса. Параметры ДОЭ: $f_0 = 300$ мм; $\lambda = 0{,}6328$ мкм; $2a = 25{,}6$ мм; $L = 15$ мм; $N = 128$; (*а*) $M = \infty$; (*б*) $M = 4$; (*в*) $M = 2$

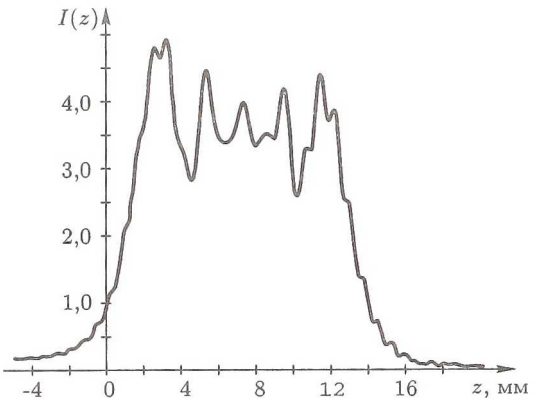


Рис. 5.37. Распределение интенсивности $I(z) = I(0,z)/I(0,0)$ вдоль оптической оси ДОЭ с повышенной глубиной фокуса

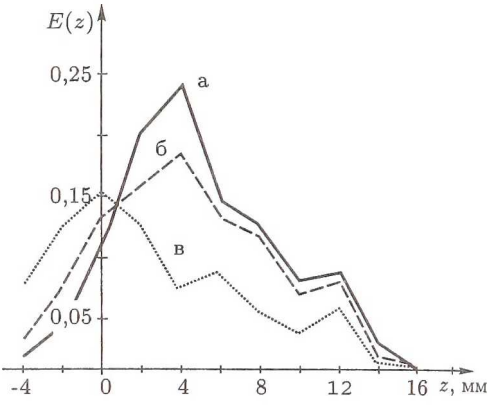


Рис. 5.38. Энергетическая эффективность ДОЭ с повышенной глубиной фокуса: (*а*) $M = \infty$; (*б*) $M = 4$; (*в*) $M = 2$

интенсивности вдоль него. При большом числе уровней квантования фазы ($M = \infty$, рис. 5.35*а* и рис. 5.36*а*) наблюдается вытянутая фокальная область с глубиной $\sim$ $\sim$ 10–15 мм, окруженная поверхностями, которые соответствуют дифракционным кольцам высших порядков. При $M = 4$ и особенно при $M = 2$ фокальная область разбивается на 2–3 подобласти глубиной по 2–5 мм, а количество колец высших порядков возрастает. Анализ рис. 5.35*а* показывает, что размер фокального пятна $\Delta \sim 0{,}06$ мм сохраняется на расстоянии ~ 15 мм, т.е. обеспечивается значительная глубина фокуса.

Энергетическая эффективность может оцениваться долей $E(z)$ энергии, проходящей через круг диаметра $\Delta \sim 0{,}06$ мм, расположенный на глубине z в фокальной области, по отношению к полной энергии, падающей на оптический элемент:

$$E(z) = \frac{\int\limits_0^{\Delta/2} I(\rho, z) \cdot 2\pi\rho \, \mathrm{d}\rho}{\int\limits_0^{\infty} I(\rho, z) \cdot 2\pi\rho \, \mathrm{d}\rho}. \tag{5.83}$$

График функции $E(z)$, приведенный на рис. 5.38, показывает, что $E(z)$ не опускается ниже 10% на протяжении примерно 15 мм, а на протяжении примерно 4 мм имеется всплеск выше 15%. При $M = 2$ среднее значение $E(z)$ вдоль отрезка фокусировки существенно падает, а всплеск смещается к фокусу объектива. Для сравнения заметим, что энергетическая эффективность объектива с обычной глубиной фокуса составляет $\sim 78\%$.

5.4. Фокусаторы в двумерную область. Метод согласованных прямоугольников

Методам расчета фокусаторов в плоскую область посвящено относительно небольшое количество работ. Решение обратной задачи фокусировки в плоскую область в приближении геометрической оптики как правило является очень сложным [12]. Пожалуй, единственным исключением является задача фокусировки прямоугольного пучка в прямоугольную область [12, 87–89]. Данная задача распадается на две од-

номерных задачи и может быть решена аналитически [12, 87–89]. В данном разделе рассмотрен простой численный метод расчета фокусаторов в двумерные области, основанный на базовом аналитическом решении задачи фокусировки в прямоугольник.

Рассмотрим задачу фокусировки в двумерную область произвольной формы. Пусть на фокусатор с апертурой G в плоскости $\mathbf{u} = (u, v)$ при $z = 0$ падает пучок с комплексной амплитудой

$$F_0(\mathbf{u}) = \sqrt{I_0(u)} \exp[i\varphi_0(\mathbf{u})],$$

где $I_0(\mathbf{u})$ — интенсивность пучка, $\varphi_0(\mathbf{u})$ — фаза пучка.

Требуется рассчитать фазовую функцию $\varphi(\mathbf{u})$ фокусатора в область D при $z = f$ с заданным распределением интенсивности $I(\mathbf{x})$, где $\mathbf{x} = (x, y)$ — координаты в плоскости фокусировки.

Геометрооптическая фазовая функция фокусатора $\varphi(\mathbf{u})$ в параксиальном приближении может быть получена из решения следующей системы уравнений:

$$\left.\begin{aligned} &\mathbf{x}(\mathbf{u}) = \mathbf{u} + f\nabla_\perp\varphi(\mathbf{u})/k, \\ &\frac{I_0(\mathbf{u})}{I(\mathbf{x}(\mathbf{u}))} = \frac{\mathrm{d}\mathbf{x}(\mathbf{u})}{\mathrm{d}\mathbf{u}} = |x'_u(\mathbf{u})\,y'_v(\mathbf{u}) - x'_v(\mathbf{u})\,y'_u(\mathbf{u})|, \quad x'_u(\mathbf{u}) = y'_v(\mathbf{u}), \\ &\varphi(\mathbf{u}) = \int\limits_{\mathbf{u}_0}^{\mathbf{u}} \nabla_\perp\varphi(\mathbf{u})\,\mathrm{d}\mathbf{u} - \varphi_0(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \frac{k}{f}\int\limits_{\mathbf{u}_0}^{\mathbf{u}} \mathbf{x}(\mathbf{u})\,\mathrm{d}\mathbf{u} - \varphi_0(\mathbf{u}). \end{aligned}\right\} \tag{5.84}$$

Первое уравнение в (5.84) соответствует уравнению луча и определяет лучевое соответствие между точками $\mathbf{u}$ апертуры фокусатора и точками $\mathbf{x}$ области фокусировки D. Второе уравнение в (5.84) соответствует дифференциальной форме закона сохранения светового потока по лучевым трубкам и позволяет определить функцию $\mathbf{x}(\mathbf{u})$ из условия формирования заданного распределения интенсивности $I(\mathbf{x})$ при $\mathbf{x} \in D$. И наконец третье уравнение в (5.84) соответствует восстановлению фазы по полному дифференциалу, где слагаемое $\varphi_0(\mathbf{u})$ введено для компенсации фазы освещающего пучка.

Решение системы (5.84) в общем случае является очень сложной проблемой. Задача существенно упрощается для прямоугольной апертуры фокусатора G и прямоугольной области фокусировки D при условии факторизуемости функций

$$F_0(\mathbf{u}) = F_1(u) \cdot F_2(v), \quad I(\mathbf{x}) = I_1(x) \cdot I_2(y), \tag{5.85}$$

то есть при

$$I_0(\mathbf{u}) = I_{01}(u)\,I_{02}(v), \quad \varphi_0(\mathbf{u}) = \varphi_{01}(u) + \varphi_{02}(v). \tag{5.86}$$

В этом случае решение двух одномерных задач фокусировки в отрезок позволяет определить двумерную фазовую функцию фокусатора:

$$\varphi(\mathbf{u}) = \varphi_1(u) + \varphi_2(v). \tag{5.87}$$

Фазовая функция $\varphi_1(u)$ одномерного (цилиндрического) фокусатора, осуществляющего заданное преобразование светового пучка, определяется из следующей

системы уравнений:

$$\left.\begin{aligned} & x(u) = u + f\frac{\mathrm{d}\varphi_1(u)}{\mathrm{d}u}\Big/k, \\ & \frac{I_{01}(u)}{I_1(x)} = \frac{\mathrm{d}x(u)}{\mathrm{d}u}, \quad u_0 \leqslant u \leqslant u_1, \quad x_0 \leqslant x \leqslant x_1, \\ & \varphi_1(u) = -\frac{ku^2}{2f} + \frac{k}{f}\int\limits_{u_0}^{u} x(\xi)\,\mathrm{d}\xi - \varphi_{01}(u). \end{aligned}\right\} \tag{5.88}$$

Уравнения системы (5.88) соответствуют уравнению луча, закону сохранения светового потока по лучевым трубкам и уравнению для восстановления фазы. Решение системы (5.88) существенно проще, чем решение (5.84). Например, для случая

$$I_0(x) = \begin{cases} I_1, & x_0 \leqslant x \leqslant x_1, \\ 0, & x < x_0, \quad x > x_1, \end{cases}$$

решение системы уравнений (5.88) имеет вид

$$\varphi_1(u) = -\frac{ku^2}{2f} + \frac{k}{f}\int\limits_{u_0}^{u}\left[\frac{1}{I_1}\int\limits_{u_0}^{\xi} I_{01}(\eta)\,\mathrm{d}\eta - x_0\right]\mathrm{d}\xi - \varphi_{01}(u), \tag{5.89}$$

и, следовательно, двумерная фазовая функция фокусатора с прямоугольной апертурой в прямоугольную область фокусировки с постоянным распределением интенсивности имеет вид

$$\begin{aligned} \varphi(u) = -k\frac{u^2+v^2}{2f} + \frac{k}{f}\int\limits_{u_0}^{u}\left[\frac{1}{I_1}\int\limits_{u_0}^{\xi} I_{01}(\eta)\,\mathrm{d}\eta - x_0\right]\mathrm{d}\xi + \\ + \frac{k}{f}\int\limits_{v_0}^{v}\left[\frac{1}{I_1}\int\limits_{v_0}^{\xi} I_{02}(\eta)\,\mathrm{d}\eta - y_0\right]\mathrm{d}\xi - \varphi_0(u), \end{aligned} \tag{5.90}$$

где (x_0, y_0) — координаты левого нижнего угла прямоугольника фокусировки.

При непрямоугольной форме хотя бы одной из областей G и D, даже при факторизуемых функциях $F_0(\mathbf{u})$, $I(\mathbf{x})$ фазовая функция фокусатора уже не факторизуется и имеет значительно более сложный вид. Приведенные соотношения (5.88)–(5.90) позволяют сформулировать «метод согласованных прямоугольников» для расчета фокусаторов при произвольной форме областей G и D и при факторизующихся функциях (5.85), (5.86). По этому методу апертура фокусатора (область G) и область фокусировки (область D) аппроксимируются наборами апертурных прямоугольников G_i и фокальных прямоугольников D_i, $i = \overline{1, N}$. Далее для каждого апертурного прямоугольника G_i из уравнений (5.87), (5.88) рассчитывается фазовая функция $\varphi_i(\mathbf{u})$ из условия фокусировки в соответствующий фокальный прямоугольник D_i. Рассмотрим описанный метод более детально. Пусть апертура фокусатора G ограничена кривыми $v = g_1(u)$, $v = g_2(u)$ и отрезками прямых $u = u_{\min}$, $u = u_{\max}$ (рис. 5.39), а область фокусировки D — кривыми $y = f_1(x)$, $y = f_2(x)$ и прямыми $x = x_{\min}$, $x = x_{\max}$ (рис. 5.40).

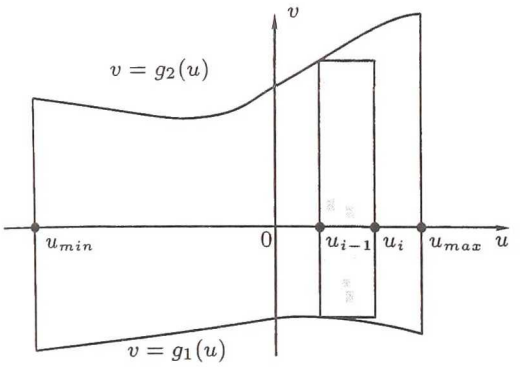


Рис. 5.39. Апертура фокусатора

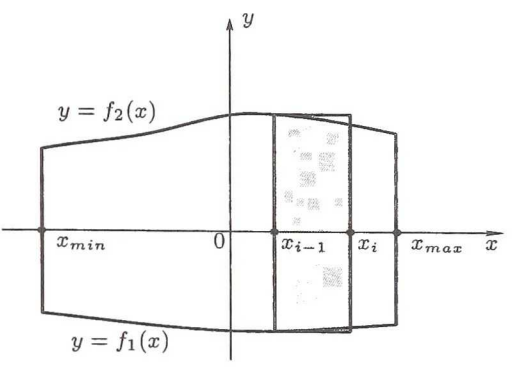


Рис. 5.40. Область фокусировки

Введем разбиение u_i, $i = \overline{0, N}$, $u_0 = u_{\min}$, $u_N = u_{\max}$ отрезка $[u_{\min}, u_{\max}]$ и апертурные прямоугольники $G_i = [u_{i-1}, u_i] \times [g_1(u_{i-1}), g_2(u_{i-1})]$. Разбиение x_i, $i = \overline{0, N}$, $x_0 = x_{\min}$, $x_N = x_{\max}$, отрезка $[x_{\min}, x_{\max}]$ определяется из решения следующего нелинейного рекуррентного уравнения:

$$\int\limits_{u_{i-1}}^{u_i} \int\limits_{g_1(u_{i-1})}^{g_2(u_{i-1})} I_0(\mathbf{u})\, \mathrm{d}^2\mathbf{u} = \int\limits_{x_{i-1}}^{x_i} \int\limits_{f_1(x_{i-1})}^{f_2(x_{i-1})} I(\mathbf{x})\, \mathrm{d}^2\mathbf{x}, \quad i = \overline{1, N}. \tag{5.91}$$

Уравнение (5.91) соответствует закону сохранения светового потока при распространении света из апертурного прямоугольника G_i в фокальный прямоугольник D_i. Решив (5.91), получим приближение области D набором прямоугольников:

$$D_i = [x_{i-1}, x_i] \times [f_1(x_{i-1}), f_2(x_{i-1})], \quad i = \overline{1, N}.$$

Для определения фазовой функции $\varphi_1(u) = \varphi_{11}(u) + \varphi_{12}(v)$ в апертурном прямоугольнике необходимо дважды решить систему уравнений (5.88).

Полученная таким образом фазовая функция фокусатора

$$\varphi(u) = \sum_{i=1}^{N} \varphi_i(u, v)\, \mathrm{rect}\left(\frac{2u - (u_{i-1} + u_i)}{2(u_i - u_{i-1})}\right) \times \\ \times\, \mathrm{rect}\left(\frac{2v - (g_1(u_{i-1}) + g_2(u_{i-1}))}{2(g_2(u_{i-1}) - g_1(u_{i-1}))}\right) - \varphi_0(u, v), \tag{5.92}$$

где

$$\mathrm{rect}(u) = \begin{cases} 1, & |u| < 0{,}5, \\ 0, & |u| \geqslant 0{,}5, \end{cases}$$

терпит разрывы вдоль прямых $u = u_i$, $i = \overline{1, N-1}$, что приводит к интерференции на стыках прямоугольников фокусировки. Для уменьшения величины интерференционных эффектов фазовую функцию $\varphi(\mathbf{u})$ можно сделать непрерывной вдоль некоторой кривой $v = f(u)$. Для этого положим

$$\varphi_i(\mathbf{u}) = \varphi_i(\mathbf{u}) + \tilde{\varphi}_i, \tag{5.93}$$

где

$$\tilde{\varphi}_i = \begin{cases} 0, & i = 0, \\ \sum\limits_{j=1}^{i-1} \left[\varphi_j\left(u_j, f\left(u_j\right)\right) - \varphi_{j+1}\left(u_j, f\left(u_j\right)\right)\right], & 2 \leqslant i \leqslant N. \end{cases} \tag{5.94}$$

В качестве примера рассмотрим синтез фокусатора с круглой апертурой $G = \left\{(u, v) : u^2 + v^2 \leqslant R^2\right\}$, фокусирующего плоский пучок круглого сечения в прямоугольник $D = [-b, b] \times [-a, a]$ с постоянной интенсивностью («круглый фокусатор в прямоугольник»).

В данном случае апертура G аппроксимируется набором прямоугольников

$$G_i = [u_{i-1}, u_i] \times \left[\sqrt{R^2 - u_{i-1}^2}, -\sqrt{R^2 - u_{i-1}^2}\right], \quad i = \overline{1, N}, \quad u_0 = -R, \quad u_N = R, \tag{5.95}$$

а область D — набором прямоугольников

$$D_i = [x_{i-1}, x_i] \times [-a, a], \quad i = \overline{1, N}, \quad x_0 = -b, \quad x_N = b. \tag{5.96}$$

Решение уравнения (5.91), определяющее разбиение фокальной области, имеет вид

$$x_i = x_{i-1} + \frac{4b\left(u_i - u_{i-1}\right)\sqrt{R^2 - u_{i-1}^2}}{\pi R^2}, \quad i = \overline{1, N-1}. \tag{5.97}$$

Фазовая функция $\varphi_i(\mathbf{u})$, являющаяся геометрооптическим решением задачи фокусировки из прямоугольника G_i в прямоугольник D_i, определяется уравнением (5.90). Фазовую добавку $\tilde{\varphi}_1$ в (5.93) выберем из условия непрерывности фазовой функции вдоль оси u:

$$\tilde{\varphi}_i = \begin{cases} 0, & i = 1, \\ \sum\limits_{j=1}^{i-1} \left[\varphi_j\left(u_j, 0\right) - \varphi_{j+1}\left(u_j, 0\right)\right], & 2 \leqslant i \leqslant N. \end{cases} \tag{5.98}$$

Для характеристики качества работы фокусатора введем значения энергетической эффективности E и среднеквадратичного отклонения интенсивности δ. Величина

$$E = \iint\limits_D I\left(\mathbf{x}\right) \mathrm{d}^2\mathbf{x} \Big/ \iint\limits_G I_0\left(\mathbf{u}\right) \mathrm{d}^2\mathbf{u}, \tag{5.99}$$

где E характеризует долю энергии освещающего пучка, попавшую в область фокусировки. Величина

$$\delta = \frac{1}{\bar{I}} \left[\frac{1}{\|D\|} \iint\limits_D \left(I\left(\mathbf{x}\right) - \bar{I}\right)^2 \mathrm{d}^2\mathbf{x}\right]^{1/2} \tag{5.100}$$

характеризует отклонение распределения интенсивности $I(\mathbf{x})$ от среднего значения интенсивности $\bar{I} = (1/\|D\|) \iint\limits_D I\left(\mathbf{x}\right) \mathrm{d}^2\mathbf{x}$ в области фокусировки, $\|D\|$ — площадь области D.

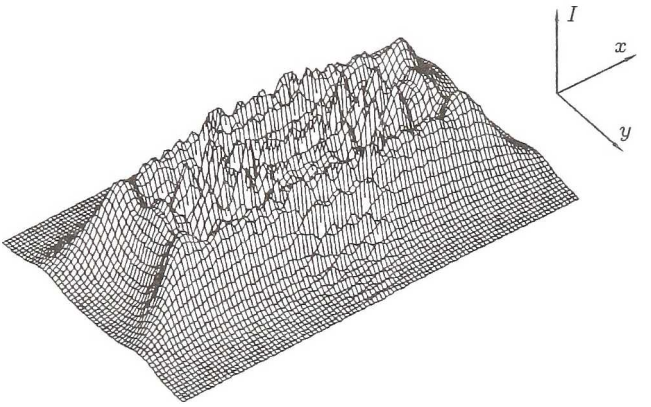


Рис. 5.41. Распределение интенсивности в фокальной области

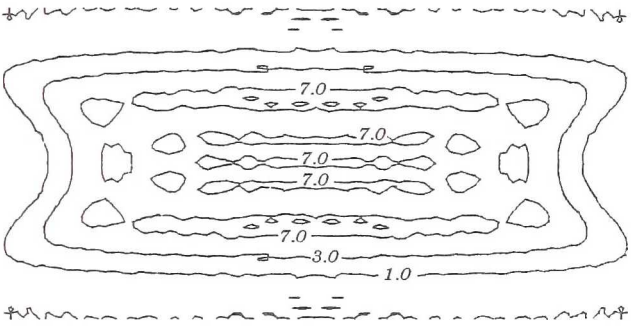


Рис. 5.42. Изофоты распределения интенсивности в фокальной области

Пример 5.2. Расчет поля от «круглого фокусатора в прямоугольник» проводился при следующих физических параметрах: $\lambda = 10{,}6$ мкм, $f = 800$ мм, радиус освещающего пучка $R = 20{,}5$ мм, размер прямоугольника фокусировки — 8×4 мм. Апертура фокусатора аппроксимировалась набором шестидесяти прямоугольников, полученных разбиением отрезка $[-R, R]$ с постоянным шагом. Энергетическая эффективность фокусировки в прямоугольник составила 85,4%, а среднеквадратичное отклонение интенсивности — 34,4%. На рис. 5.41 представлено трехмерное распределение интенсивности в фокальной области фокусатора в прямоугольник, а на рис. 5.42 — изофоты трехмерного распределения. Анализ полученных результатов свидетельствует о высоком качестве фокусировки.

Пример 5.3. По методу согласованных прямоугольников был также синтезирован фокусатор с круглой апертурой, фокусирующий плоский пучок круглого сечения в «усеченный эллипс» (рис. 5.43) с постоянным распределением интенсивности («фокусатор в усеченный эллипс»). Исследование «фокусатора в усеченный эллипс» проводилось при $\lambda = 10{,}6$ мкм; $f = 800$ мм; радиусе освещающего пучка $R = 20{,}5$ мм для следующих характерных размеров усеченного эллипса: $a = 2$ мм, $b = 4$ мм, $b_1 = 3{,}5$ мм. Энергетическая эффективность фокусировки в усеченный эллипс составила 87,6%, а среднеквадратичное отклонение интенсивности —

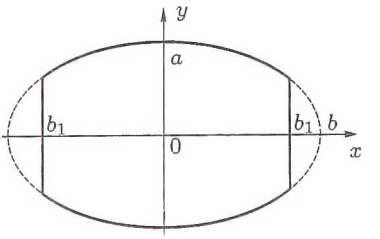


Рис. 5.43. «Усеченный эллипс»

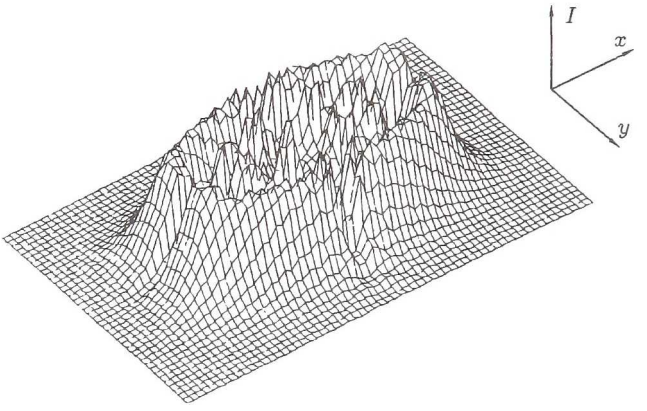


Рис. 5.44. Распределение интенсивности в фокальной плоскости фокусатора в усеченный эллипс

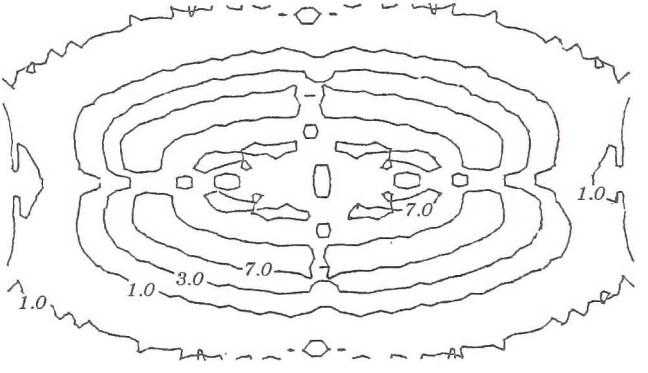


Рис. 5.45. Изофоты распределения интенсивности в фокальной области фокусатора в усеченный эллипс

38,2%. На рис. 5.44 представлено трехмерное распределение интенсивности в фокальной плоскости фокусатора в усеченный эллипс, а на рис. 5.45 представлены изофоты трехмерного распределения интенсивности при фокусировке в усеченный эллипс. Результаты расчетов подтверждают работоспособность метода «согласованных прямоугольников» при расчете фокусаторов в сложные двумерные области.

5.5. Многофокусные фокусаторы

Фокусировка в систему одинаковых, смещенных друг относительно друга фокальных кривых, может быть реализована ДОЭ, фазовая функция которых соответствует суперпозиции фазовых функций фокусатора в кривую и многопорядковой дифракционной решетки. ДОЭ такого типа мультиплицируют (повторяют) заданную фокальную линию в одной плоскости. В данном разделе рассматриваются многофокусные фокусаторы — ДОЭ, предназначенные для фокусировки освещающего пучка в N линий переменных размеров, расположенных в различных фокальных плоскостях вдоль оптической оси. При этом между линиями формируется заданное распределение энергии $I_1, \ldots, I_N$, $\left(\sum_{i=1}^{N} I_i = 1\right)$ (см. рис. 5.46).

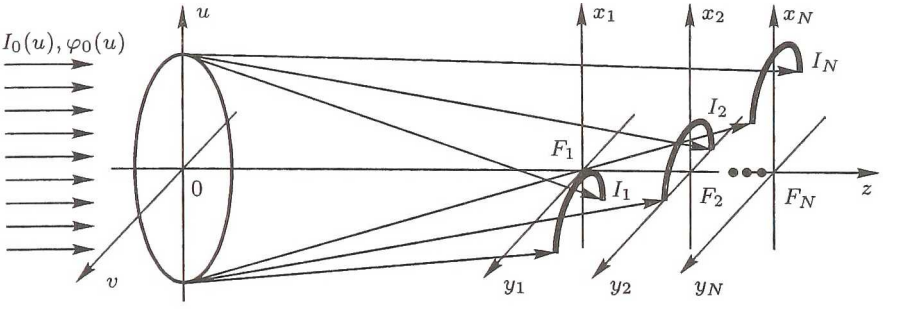


Рис. 5.46. Процесс фокусировки, реализуемый многофокусным фокусатором

Известно, что введение нелинейности в фазовую функцию ДОЭ приводит к появлению дополнительных дифракционных порядков [90–92, 25–30]. При расчете многофокусных ДОЭ основная идея состоит в использовании дифракционных порядков, возникающих при нелинейном преобразовании параксиальной фазовой функции фокусатора в кривую, заданную уравнениями (5.17) и (5.18). При этом изображения, формируемые в дифракционных порядках, соответствуют набору масштабированных линий, расположенных в различных фокальных плоскостях вдоль оптической оси. Тип нелинейного преобразования фазы фокусатора в кривую влияет на распределение энергии между линиями и может быть выбран из условия формирования заданного распределения энергии между линиями фокусировки.

Согласно вышесказанному, фазовую функцию многофокусного фокусатора представим в виде [25–30]

$$\varphi_{mf}(\mathbf{u}) = \varphi_1(\mathbf{u}) + \varphi_{11}(\mathbf{u}) + \Phi[\varphi_d(\mathbf{u})] - \varphi_0(\mathbf{u}), \quad \mathbf{u} \in G, \tag{5.101}$$

где

$$\varphi_d(\mathbf{u}) = \mathrm{mod}_{2\pi}(\varphi_2(\mathbf{u}) + \varphi_{22}(\mathbf{u}) + \varphi_{23}(\mathbf{u})). \tag{5.102}$$

Здесь, $\varphi_0(\mathbf{u})$ — фаза пучка, освещающего ДОЭ с апертурой G, $\varphi_1(\mathbf{u})$ и $\varphi_2(\mathbf{u})$ — параксиальные фазовые функции фокусаторов в линии L с параметрическим уравнением

$$\mathbf{X}(\xi) = (X(\xi), Y(\xi)), \quad \xi \in [0, S], \tag{5.103}$$

расположенные в фокальных плоскостях $z = f_1$ и $z = f_2$. Согласно общим формулам (5.17), (5.18) для фазовой функции фокусатора в кривую, функции $\varphi_1(\mathbf{u})$, $\varphi_2(\mathbf{u})$ имеют вид

$$\varphi_i(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f_i} + \tilde{\varphi}_i(\mathbf{u}), \quad i = 1, 2, \tag{5.104}$$

где

$$\tilde{\varphi}_i(\mathbf{u}) = \frac{k}{f_i}\int\limits_{u_0}^{u} X(\xi(u, v_0))\,\mathrm{d}u + \frac{k}{f_i}\int\limits_{v_0}^{v} Y(\xi(u, v))\,\mathrm{d}v. \tag{5.105}$$

И наконец, функции

$$\varphi_{11}(\mathbf{u}) = \frac{k}{f_1}\mathbf{x}_1\mathbf{u}, \quad \varphi_{22}(\mathbf{u}) = \frac{k}{f_2}\mathbf{x}_2\mathbf{u}, \quad \varphi_{23}(\mathbf{u}) = -\frac{k}{2f_3}\mathbf{u}^2 \tag{5.106}$$

в (5.101) являются параксиальными фазовыми функциями призм и линзы, соответственно. Слагаемое $\varphi_0(\mathbf{u})$ в (5.101) введено для компенсации фазы освещающего пучка. Фаза фокусатора $\varphi_1(\mathbf{u})$, фазы призм $\varphi_{11}(\mathbf{u})$, $\varphi_{22}(\mathbf{u})$ и фаза линзы $\varphi_{23}(\mathbf{u})$ в (5.101) не участвуют явно в формировании дифракционных порядков, а только дают дополнительные степени свободы в выборе координат фокальных плоскостей, расположения линий в этих плоскостях и масштабирующего множителя, определяющего изменение размеров линий фокусировки. Функция $\Phi[\varphi_d(\mathbf{u})] \in [0, 2\pi)$ описывает нелинейное преобразование фазы $\varphi_d(\mathbf{u}) \in [0, 2\pi)$. Механизм нелинейного преобразования показан на рис. 5.47. При этом функция $\Phi[\varphi_0(\mathbf{u})]$ является некоторым аналогом зонной пластинки, которая может быть бинарной, ступенчатой или иметь непрерывно изменяющийся рельеф.

Рассмотрим работу многофокусного фокусатора при освещающем пучке следующего вида:

$$F_0(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp(i\varphi_0(\mathbf{u})).$$

В этом случае поле непосредственно за плоскостью многофокусного фокусатора принимает вид

$$F(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp(i\varphi_1(\mathbf{u}) + i\varphi_{11}(\mathbf{u}) + i\Phi[\varphi_d(\mathbf{u})]). \tag{5.107}$$

Для описания процесса формирования набора линий, разложим функцию $\exp(i\Phi[\xi])$ в ряд Фурье на интервале $[0, 2\pi)$ [25–30, 90–92]:

$$\exp(i\Phi[\xi]) = \sum_{j=-\infty}^{\infty} c_j \exp(ij\xi), \tag{5.108}$$

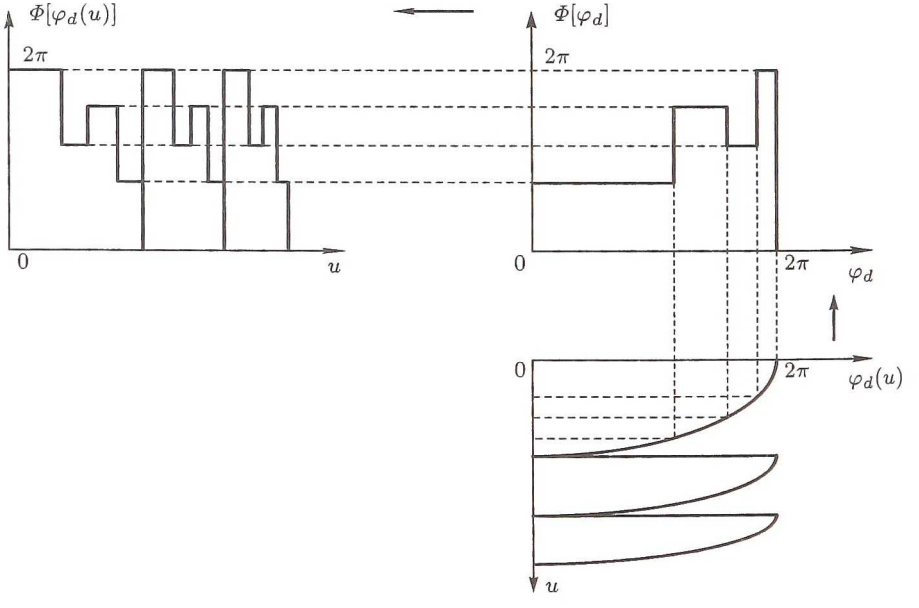


Рис. 5.47. Нелинейное преобразование фазы (1-мерный случай)

где

$$c_j = \frac{1}{2\pi}\int_0^2 \exp\left(i\Phi\left[\xi\right] - ij\xi\right)\,\mathrm{d}\xi \tag{5.109}$$

– коэффициенты Фурье; $\sum\limits_{j=-\infty}^{\infty} |c_j|^2 = 1$. Подставляя в (5.108) $\xi = \varphi_d(\mathbf{u})$, представляем (5.107) с учетом 2π-периодичности в виде

$$F\left(\mathbf{u}\right) = \sqrt{I_0\left(\mathbf{u}\right)}\exp\left(i\varphi_1\left(\mathbf{u}\right) + i\varphi_{11}\left(\mathbf{u}\right)\right)\sum_{j=-\infty}^{\infty} c_j \exp\left(ij\left[\varphi_2\left(\mathbf{u}\right) + \varphi_{22}\left(\mathbf{u}\right) + \varphi_{23}\left(\mathbf{u}\right)\right]\right). \tag{5.110}$$

Уравнение (5.110) соответствует суперпозиции пучков с фазовыми функциями:

$$\varphi_j\left(\mathbf{u}\right) = -\frac{k\mathbf{u}^2}{2F_j} + p_j\tilde{\varphi}_j\left(\mathbf{u}\right) + \frac{k}{F_j}\mathbf{X}_j\mathbf{u}, \tag{5.111}$$

где

$$\tilde{\varphi}_j\left(\mathbf{u}\right) = \frac{k}{F_j}\int_{u_0}^{u} X\left(\xi\left(u, v_0\right)\right)\,\mathrm{d}u + \frac{k}{F_j}\int_{v_0}^{v} Y\left(\xi\left(u, v\right)\right)\,\mathrm{d}v, \tag{5.112}$$

$$F_j = \frac{f_1 f_2}{f_2 + jf_1\left(1 + f_2/f_3\right)}, \tag{5.113}$$

$$p_j = \frac{f_2 + jf_1}{f_2 + jf_1\left(1 + f_2/f_3\right)}, \tag{5.114}$$

$$\mathbf{X}_j = \frac{\mathbf{x}_1 f_2 + j\mathbf{x}_2 f_1}{f_2 + jf_1\left(1 + f_2/f_3\right)}. \tag{5.115}$$

Квадраты модулей коэффициентов Фурье (5.109) представляют распределение энергии между пучками, а функции (5.111) определяют структуру изображений, формируемых в дифракционных порядках многофокусного фокусатора.

В п. 5.2 показано, что при умножении добавки к линзе в фазовой функции фокусатора в кривую на константу p, происходит p-кратное масштабирование линии фокусировки. Согласно указанному свойству и общему виду фазовой функции фокусатора в кривую, фазовая функция (5.111) обеспечивает в плоскости $z = F_j$ фокусировку в линию L_{p_j} с уравнением

$$\mathbf{X}_{p_j}\left(\xi\right) = \left(p_j X\left(\xi/p_j\right), p_j Y\left(\xi/p_j\right)\right), \quad \xi \in \left[0, p_j S\right]. \tag{5.116}$$

Линия L_{p_j} является масштабированной линией L с коэффициентом p_j . Последнее слагаемое в (5.111) соответствует фазовой функции призмы, которая дает сдвиг X_j линии L_{p_j} в плоскости $z = F_j$. Отметим, что процесс формирования линии L_{p_j} не является полностью независимым от других дифракционных порядков. Действительно, изображения создаваемые в плоскости $z = F_j$ пучками с фазовыми функциями $\varphi_k(\mathbf{u})$ при $k \neq j$, соответствуют расфокусированным линиям. В общем случае ширина расфокусированных линий на порядок больше, чем дифракционная ширина

линии L_{p_j}, что делает их влияние ничтожно малым. В некоторых редких случаях, когда взаимодействием порядков пренебречь нельзя, выбором векторов $\mathbf{x}_1$ и $\mathbf{x}_2$ в (5.115), играющих роль параметров несущих, можно разнести паразитные расфокусированные линии относительно полезных линий и тем самым добиться независимого формирования фокальных линий. В связи с изложенным, в последующем мы не будем учитывать взаимодействие между изображениями, формируемыми в различных порядках многофокусного фокусатора.

Обозначим $j_1, \ldots, j_N$ индексы, соответствующие заданным фокальным плоскостям. Для формирования заданного соотношения энергии $I_1, \ldots, I_N$, между кривыми фокусировки, функцию $\Phi[\xi]$, $\xi \in [0, 2\pi)$ достаточно определить из следующего условия:

$$|c_{j_i}|^2 = I_i, \quad i = 1, \ldots, N, \tag{5.117}$$

где c_j — коэффициенты Фурье (5.109) в разложении функции $\exp(i\Phi[\xi])$. Согласно (5.117), будем интерпретировать нелинейное преобразование $\Phi[\xi]$ как фазовую функцию N-порядковой дифракционной решетки с периодом 2π и интенсивностями $I_1, \ldots, I_N$ в порядках $j_1, \ldots, j_N$ [25–30]. Эта интерпретация сводит расчет нелинейного преобразования $\Phi[\xi]$ к хорошо известной задаче синтеза фазовой дифракционной решетки с заданным распределением интенсивности в дифракционных порядках. При этом энергетическая эффективность решетки $E = \sum_{i=1}^{N} |c_{j_i}|^2$ определяет долю энергии освещающего пучка фокусируемую в заданных фокальных линиях, а отклонения значений интенсивности в порядках $j_1, \ldots, j_N$ от заданных значений $I_1, \ldots, I_N$ характеризуют ошибку формирования заданного распределения энергии между линиями фокусировки.

Рассмотрим несколько частных случаев многофокусных фокусаторов. При

$$f_3 = \infty \quad (\varphi_{23}(\mathbf{u}) \equiv 0) \tag{5.118}$$

в (5.101), (5.102) и (5.106), масштабные коэффициенты p_j в (5.114), описывающие изменение размера фокальных линий, равны единице. Соответственно, размер фокальных линий одинаков во всех фокальных плоскостях.

При

$$f_3 = -f_2 \tag{5.119}$$

в (5.101) и (5.102), функции (5.111), (5.112), определяющие структуру изображений формируемых в дифракционных порядках многофокусного фокусатора, принимают вид

$$\varphi_j(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f_1} + p_j \left[\frac{k}{f_1} \int_{u_0}^{u} x(\xi(\mathbf{u}))\, \mathrm{d}u + \frac{k}{f_1} \int_{v_0}^{v} y(\xi(\mathbf{u}))\, \mathrm{d}v \right] + \frac{k}{f_1} \mathbf{X}_j \mathbf{u}, \tag{5.120}$$

где

$$p_j = 1 + j\frac{f_1}{f_2}, \tag{5.121}$$

$$\mathbf{X}_j = \frac{\mathbf{x}_1 f_2 + j\mathbf{x}_2 f_1}{f_2}. \tag{5.122}$$

Таким образом, при выполнении условия (5.119), многофокусный фокусатор обеспечивает фокусировку в N линий в одной фокальной плоскости $z = f_1$. При этом масштаб фокальных линий описывается линейной функцией (5.121).

Приведем ряд примеров, иллюстрирующих рассмотренный метод.

Пример 5.4. Рассмотрим расчет многофокусного фокусатора (5.101)–(5.106), (5.118) для фокусировки в три отрезка равной длины, расположенных в трех разных фокальных плоскостях. Соотношение энергии между отрезками фокусировки будем считать постоянным. Для расчета фазы многофокусного фокусатора определим функции $\varphi_1(\mathbf{u})$ и $\varphi_2(\mathbf{u})$ в (5.101) и (5.102) как фазовые функции фокусаторов в отрезки $|x| \leqslant d$, расположенные в плоскостях $z = f_1$ и $z = f_2$. Для случая квадратной апертуры $G(2a \times 2a)$ и постоянной интенсивности освещающего пучка, функции $\varphi_1(\mathbf{u})$ и $\varphi_2(\mathbf{u})$ несложно получить из уравнений (5.20) и (5.21) в следующем виде:

$$\left.\begin{aligned} \varphi_1(\mathbf{u}) &= -\frac{k\mathbf{u}^2}{2f_1} + \frac{kd}{2f_1 a}u^2, \\ \varphi_2(\mathbf{u}) &= -\frac{k\mathbf{u}^2}{2f_2} + \frac{kd}{2f_2 a}u^2. \end{aligned}\right\} \tag{5.123}$$

Далее определим функцию нелинейного преобразования $\Phi[\xi]$ как фазовую функцию 3-порядковой дифракционной решетки:

$$\Phi[\xi] = \begin{cases} 0, & \xi \in [0, \pi), \\ 2\operatorname{arctg}(\pi/2), & \xi \in [\pi, 2\pi). \end{cases} \tag{5.124}$$

Коэффициенты Фурье функции $\exp(i\Phi[\xi])$ имеют вид

$$c_j = \begin{cases} \dfrac{1-(-1)^j}{2\pi i j}\left[1 - \exp\left(2i \cdot \operatorname{arctg}\dfrac{\pi}{2}\right)\right], & j \neq 0, \quad j = \pm 1, \pm 2, \ldots, \\ \dfrac{1}{2}\left[1 + \exp\left(2i \cdot \operatorname{arctg}\dfrac{\pi}{2}\right)\right], & j = 0. \end{cases} \tag{5.125}$$

Согласно (5.125), $|c_{-1}|^2 = |c_0|^2 = |c_1|^2 = 0{,}2884$, то есть решетка (5.124) концентрирует в основных трех порядках более 86% энергии. Подстановка уравнений (5.123), (5.125) в (5.110) дает, с учетом трех главных дифракционных порядков, фокусировку в три отрезка в плоскостях $z = F_j$, $j = -1, 0, 1$. Центры отрезков в этих плоскостях расположены в точках X_j, $j = -1, 0, 1$, координаты которых определены уравнением (5.115).

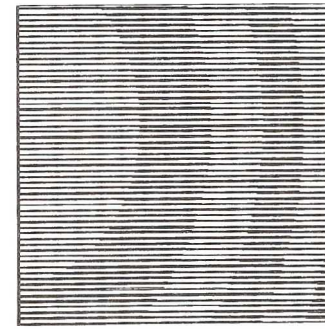
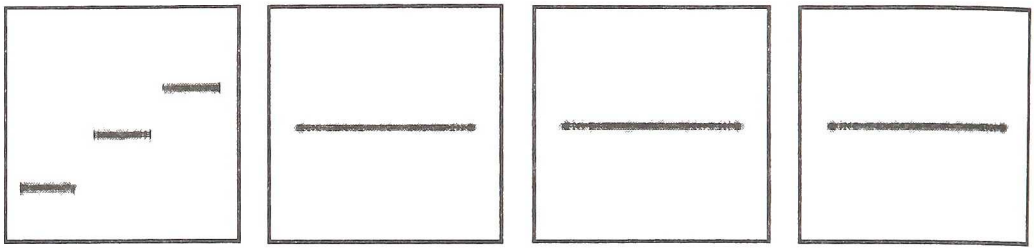


Рис. 5.48. Фрагмент зонной пластинки $\Phi[\varphi_d(\mathbf{u})]$

Для оценки работоспособности многофокусного фокусатора (5.101)–(5.103), (5.118), (5.123), (5.124) проводился расчет интенсивности поля, формируемого многофокусным фокусатором, при следующих параметрах: $\lambda = 1{,}06$ мкм, $2d = 8$ мм, $2a = 9$ мм, $f_1 = 100$ мм, $f_2 = 3000$ мм, $\mathbf{x}_1 = (0, 0)$ мм, $\mathbf{x}_2 = (145, 0)$ мм. Для указанных параметров координаты фокальных плоскостей равны 96,774 мм ($j = 1$), 100 мм ($j = 0$) и 103,448 мм ($j = -1$), а центры отрезков смещены относительно оптической оси в точки $\mathbf{X}_1 = (4{,}677, 0)$ мм, $\mathbf{X}_0 = (0, 0)$ мм, и $\mathbf{X}_{-1} = (-5, 0)$ мм. На рис. 5.48 и 5.49 представлены фрагмент зонной пластинки $\Phi[\varphi_d(\mathbf{u})]$ и распределения интенсивности, формируемые многофокусным фокусатором. Распределения интенсивности на рис. 5.49 получены численным расчетом интеграла Френеля–Кирхгофа по методам [20–24]. Результаты расчетов подтверждают работоспособность представленного метода.



а *б* *в* *г*

Рис. 5.49. Полутоновые распределения интенсивности, формируемые многофокусным фокусатором в три линии, расположенные в трех различных фокальных плоскостях: (*а*) в плоскости XOZ, содержащей фокальные линии; (*б*), (*в*), (*г*) в плоскостях $z = F_j$, $j = -1, 0, 1$

Пример 5.5. Рассмотрим расчет многофокусного фокусатора (5.101)–(5.106), (5.119) с квадратной апертурой $G(2a_{\times}2a)$, фокусирующего пучок с постоянной интенсивностью в пять отрезков переменной длины в одной плоскости $z = f_1$. Пусть 1:2:3:4:5 — заданное соотношение длин фокальных отрезков, а x_0 — заданное расстояние между отрезками (см. рис. 5.50*а*). Указанные соотношение длин и расстояние между отрезками получаются при $f_2 = 3f_1$, $j = \overline{-2,2}$ в (5.121) и при $\mathbf{x}_1 = (0,0)$ мм, $\mathbf{x}_2 = (3x_0, 0)$ мм в (5.122). При этом функции $\varphi_1(\mathbf{u})$ и $\varphi_2(\mathbf{u})$ в (5.101), (5.102) имеют вид

$$\begin{aligned} \varphi_1(\mathbf{u}) &= -\frac{k\mathbf{u}^2}{2f_1} + \frac{kd}{2f_1 a}v^2, \\ \varphi_2(\mathbf{u}) &= -\frac{k\mathbf{u}^2}{6f_1} + \frac{kd}{6f_1 a}v^2. \end{aligned} \qquad (5.126)$$

Далее для окончательного определения фазовой функции многофокусного фокусатора, необходимо определить функцию нелинейного преобразования $\Phi[\xi]$ в (5.101) как фазовую функцию 5-порядковой дифракционной решетки. Для формирования равной интенсивности на отрезках фокусировки, зададим интенсивности дифракционных порядков решетки пропорциональными длине отрезков:

$$I_j/I_0 = 1 + j/3, \quad j = \overline{-2,2}, \qquad (5.127)$$

где, согласно условию нормировки $\left(\sum_{i=1}^{N} I_i = 1\right)$, $I_0 = 1/5$.

Требуемая дифракционная решетка с интенсивностью порядков (5.127) рассчитывалась при помощи АА-алгоритма, описанного в главе 2. Энергетическая эффективность решетки составила 88,8%, при среднеквадратичной ошибке формирования заданного линейного распределения интенсивности (5.127) в 1,4%. Фазовая функция рассчитанной решетки на периоде и интенсивности дифракционных порядков решетки показаны на рис. 5.50*б* и 5.50*в*, соответственно. Для оценки работоспособности многофокусного фокусатора (5.101)–(5.106), (5.119), (5.126) с функцией $\Phi[\xi]$, представленной рис. 5.50*б*, проводился численный расчет интенсивности формируемого фокусатором поля при следующих параметрах: $\lambda = 1{,}06$ мкм, $2d = 0{,}6$ мм, $2a = 10$ мм, $f_1 = 100$ мм, $f_2 = 3f_1 = 300$ мм, $\mathbf{x}_1 = (0,0)$ мм, $\mathbf{x}_2 = (0{,}75, 0)$ мм. Расчетное распределение интенсивности представлено на рис. 5.51 и показывает высокое качество фокусировки.

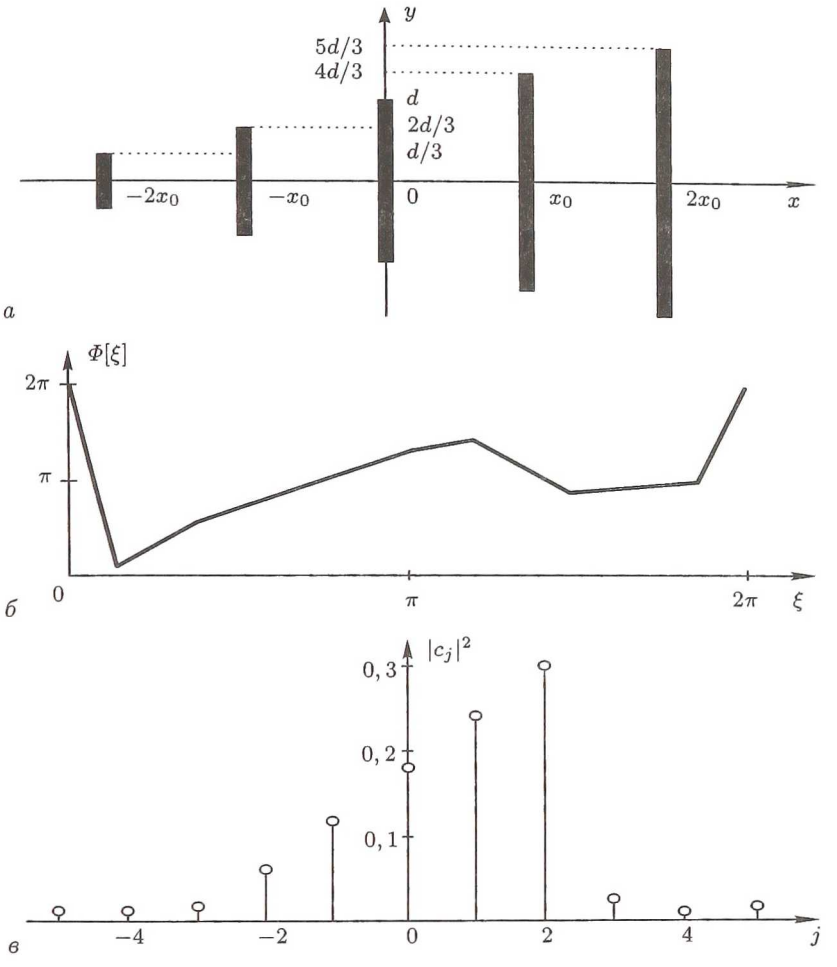


Рис. 5.50. (*а*) Геометрия отрезков фокусировки; (*б*) функция нелинейного преобразования $\Phi[\xi]$; (*в*) квадраты модулей коэффициентов Фурье для функции $\exp(i\Phi[\xi])$

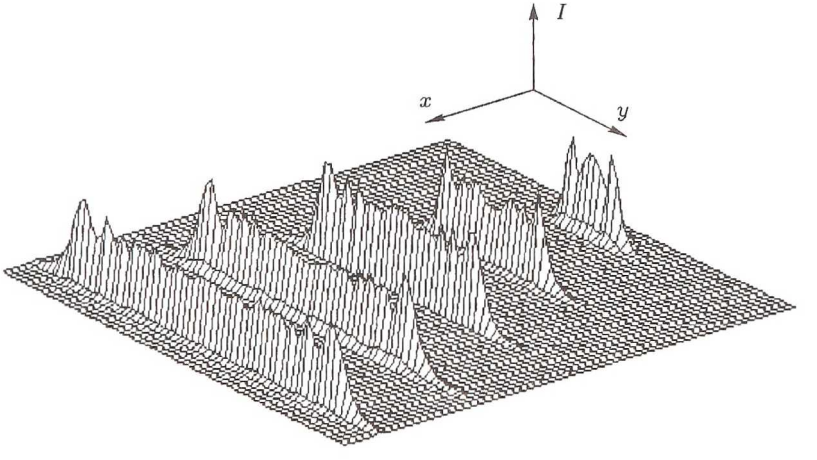


Рис. 5.51. Распределение интенсивности, сформированное многофокусным фокусатором в плоскости $z = f_1$ при фокусировке в набор отрезков на рис. 5.50*а*

5.5.1. Многофокусные бинарные зонные пластинки

Если функция нелинейного преобразования $\Phi[\xi]$ в уравнении (5.101) соответствует многопорядковой бинарной решетке, то фазовая функция

$$\varphi_{zp}(\mathbf{u}) = \Phi[\varphi_d(\mathbf{u})] \tag{5.128}$$

соответствует бинарной зонной пластинке. Зонная пластинка (5.128), (5.102), является частным случаем многофокусного фокусатора (5.101), (5.102). Поскольку бинарные элементы представляют особый интерес в связи с простотой их изготовления, специально рассмотрим многофокусные свойства зонной пластинки. Предположим, что зонная пластинка (5.128), (5.102) освещается сходящимся сферическим пучком с комплексной амплитудой

$$F_0(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp\left(-\frac{ik\mathbf{u}^2}{2f_1}\right). \tag{5.129}$$

Используя уравнения (5.107)–(5.110), представим поле непосредственно за пластинкой в виде

$$F(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp\left(-\frac{ik\mathbf{u}^2}{2f_1}\right)\sum_{j=-\infty}^{\infty} c_j \exp(ij\varphi_d(\mathbf{u})) =$$

$$= \sqrt{I_0(\mathbf{u})}\sum_{j=-\infty}^{\infty} c_j \exp(i\varphi_j(\mathbf{u})), \tag{5.130}$$

где c_j — коэффициенты Фурье (5.109). Функции $\varphi_j(\mathbf{u})$ в (5.130), определяющие структуру изображений, формируемых в дифракционных порядках зонной пластинки, имеют вид (5.111) и соответствуют фазовым функциям фокусатора в масштабированную линию L. Параметры p_j, F_j, X_j определяющие масштаб и расположение фокальных линий, формируемых в j-м дифракционном порядке, задаются соотношениями

$$F_j = \frac{f_1 f_2}{f_2 + jf_1(1 + f_2/f_3)}, \tag{5.131}$$

$$p_j = \frac{jf_1}{f_2 + jf_1(1 + f_2/f_3)}, \tag{5.132}$$

$$\mathbf{X}_j = p_j \mathbf{x}_2. \tag{5.133}$$

При $f_3 = -f_2$, линии фокусировки располагаются в одной фокальной плоскости $z = f_1$, а изменение размера линий задается линейной функцией

$$p_j = j\frac{f_1}{f_2}. \tag{5.134}$$

Таким образом, зонная пластинка (5.128), (5.102) с функцией $\Phi[\xi]$, определенной как фазовая функция N-порядковой решетки с интенсивностями $I_1, \ldots, I_N$ в порядках $j_1, \ldots, j_N$, обеспечивает фокусировку сходящегося сферического пучка (5.129) в N масштабированных линий в плоскостях $z = F_{j_i}$, $i = \overline{1, N}$.

Для функции $\Phi[\xi]$ — фазовой функции N-порядковой бинарной решетки, модули коэффициентов Фурье c_j симметричны:

$$|c_j| = |c_{-j}|. \tag{5.135}$$

Вследствие симметрии расчет многофокусной бинарной пластинки может быть осуществлен только для симметричного набора индексов j_i, и симметричного соотношения энергий $I_1, \ldots, I_N$.

В заключение рассмотрим расчет бинарной зонной пластинки для фокусировки в центрально-симметричный контур C в плоскости $z = f_1$. Контур C можно представить как объединение двух центрально-симметричных линий. Предположим, что $f_3 = -f_2$ и функция $\varphi_d(\mathbf{u})$ в (5.128) является фазовой функцией фокусатора в линию L, равную половине фокального контура C. В этом случае фокальные линии $L_{p_1} \equiv L$ и L_{p-1}, формируемые в $+1$ и -1 дифракционных порядках зонной пластинки являются центрально-симметричными и образуют заданный контур C плоскости $z = f_1$. Для расчета бинарной зонной пластинки, определим функцию $\Phi[\xi]$ как фазовую функцию бинарной 2-порядковой решетки:

$$\Phi[\xi] = \begin{cases} 0, & \xi \in [0, \pi) \\ \pi, & \xi \in [\pi, 2\pi) \end{cases}. \tag{5.136}$$

Коэффициенты Фурье c_j в разложении функции $\exp(i\Phi[\xi])$ имеют вид

$$c_j = \begin{cases} \dfrac{1-(-1)^j}{\pi i j}, & j = \pm 1, \pm 2, \ldots, \\ 0, & j = 0. \end{cases} \tag{5.137}$$

Поскольку $|c_1|^2 = |c_{-1}|^2 = 0{,}405$, то с учетом основных $+1$-го и -1-го порядков в разложении (5.130), бинарная зонная пластинка (5.128), (5.136) фокусирует 81% энергии освещающего пучка в линии L_{p_1} и L_{p-1}, формирующие заданный фокальный контур C.

Рассмотрим ряд примеров расчета многофокусных бинарных пластинок.

Пример 5.6. Рассмотрим расчет бинарной пластинки (5.128), (5.102) для фокусировки сходящегося сферического пучка (5.129) в набор из шести отрезков с соотношением длин 3:2:1:1:2:3 в плоскости $z = f_1$ (см. рис. 5.52a). Согласно (5.131)–(5.133), фокусировка в заданный набор отрезков происходит при следующих параметрах: $f_1 = f_2$, $f_3 = -f_2$, $\mathbf{x}_2 = (x_2, 0)$. При этом функция $\varphi_2(\mathbf{u})$ в (5.102) является фазой фокусатора в отрезок $|x| \leqslant d$ в плоскости $z = f_1$. При освещающем пучке квадратного сечения $(2a \times 2a)$ с постоянной интенсивностью, функция $\varphi_2(\mathbf{u})$ имеет вид

$$\varphi_2(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f_1} + \frac{kd}{2f_1 a} v^2. \tag{5.138}$$

Для формирования равной интенсивности на отрезках фокусировки, функцию $\Phi[\xi]$ зададим как фазу бинарной 6-порядковой решетки с интенсивностями порядков, пропорциональными длине отрезков фокусировки:

$$I_j / I_0 = |j|, \quad j = \overline{-3, 3}, \quad j \neq 0, \tag{5.139}$$

где, согласно условию нормировки $\left(\sum\limits_{i=1}^{N} I_i = 1 \right)$, $I_0 = 1/5$.

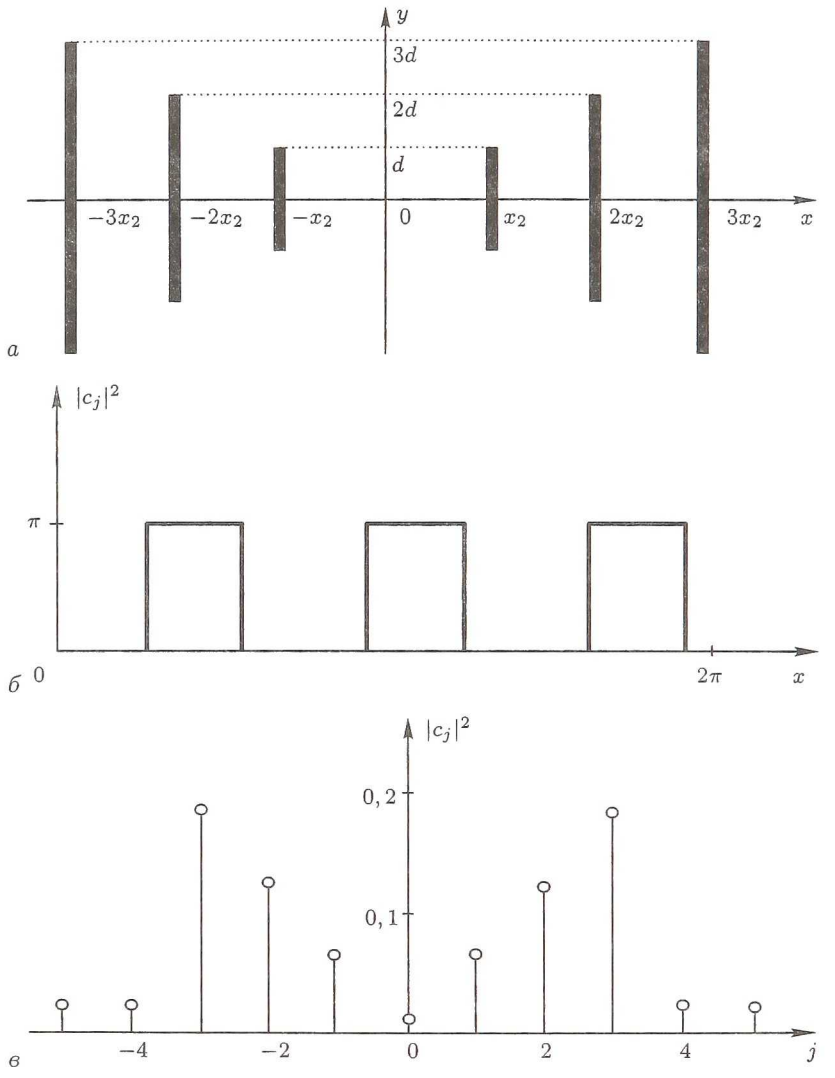


Рис. 5.52. Геометрия отрезков фокусировки (*а*); бинарная функция $\Phi[\xi]$ нелинейного преобразования (*б*); квадраты модулей коэффициентов Фурье для функции $\exp(i\Phi[\xi])$(*в*)

Для расчета бинарной решетки с интенсивностью порядков (5.139) использовался градиентный алгоритм, описанный в главе 2. Рассчитанный бинарный профиль и интенсивности дифракционных порядков решетки (5.139) показаны на рис. 5.52*б* и 5.52*в*, соответственно. Энергетическая эффективность решетки на рис. 5.52б составляет 74,5% при среднеквадратичной ошибке формирования заданной интенсивности порядков (5.139) в 1,6%. На рис. 5.53 приведено расчетное распределение интенсивности, формируемое зонной пластинкой (5.128), (5.102), (5.138) с функцией $\Phi[\xi]$ на рис. 5.52*б*, рассчитанное при следующих параметрах: $\lambda = 1{,}06$ мкм, $2d = 0{,}5$ мм, $2a = 10$ мм, $f_1 = f_2 = -f_3 = 100$ мм, $\mathbf{x}_2 = (0{,}25, 0)$ мм. Рис. 5.53 демонстрирует высокое качество фокусировки; однако в центре имеется острый пик интенсивности. Пик интенсивности объясняется ошибкой в расчете функции $\Phi[\xi]$; коэффициент Фурье c_0 в разложении функции $\exp(i\Phi[\xi])$ не равен нулю, $|c_0|^2 \approx 0{,}01$.

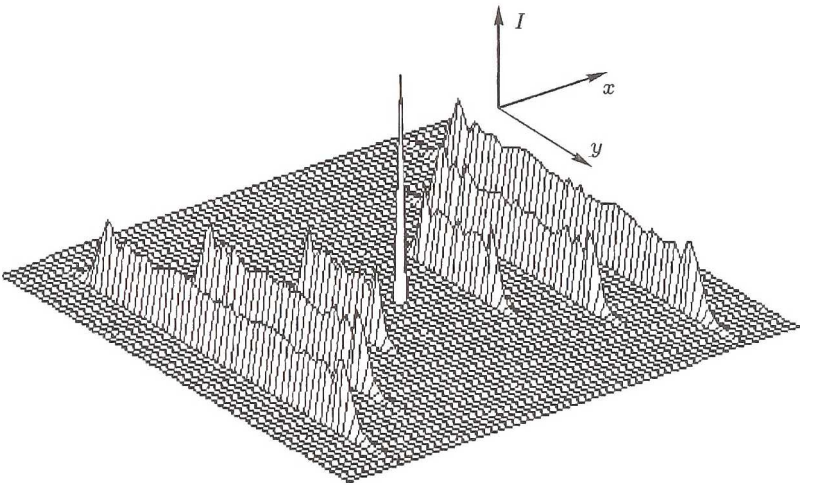
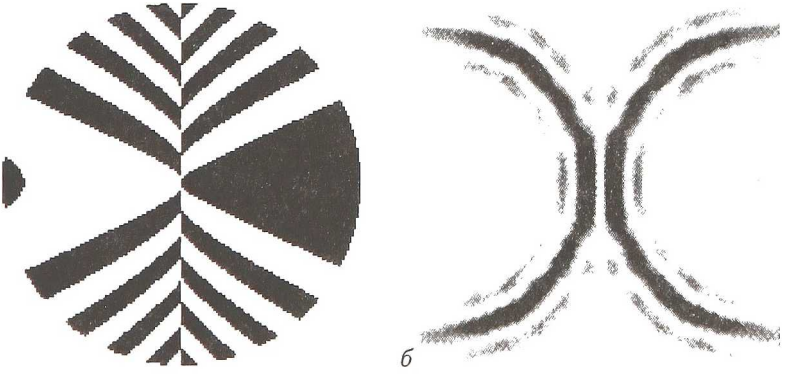


Рис. 5.53. Распределение интенсивности, создаваемое многофокусной бинарной зонной пластинкой в плоскости $z = f_1$ при фокусировке в набор отрезков на рис. 5.52*а*

Согласно (5.132), изображение, формируемое в нулевом порядке, соответствует точке. Так как в заданных порядках $j = -\overline{3,3}$, $j \neq 0$, а энергия распределена равномерно по длине отрезков фокусировки, то центральный пик интенсивности оказывается относительно большим. В то же время доля энергии освещающего пучка, сфокусированная в центральном пике, составляет всего 1%.

Пример 5.7. Рассмотрим расчет бинарной зонной пластинки (5.128), (5.136) для фокусировки в букву «х» в плоскости $z = f_1$. Букву «х» предполагаем состоящей из двух центрально-симметричных полуколец радиуса R_0. В этом случае функция $\varphi_d(\mathbf{u})$ в (5.128) является фазовой функцией фокусатора сходящегося сферического пучка в полукольцо с центром в точке $(-x_0, 0)$ при $z = f_1$. Параметр x_0 определяет расстояние между полукольцами. Для освещающего пучка круглого сечения с постоянной интенсивностью функция $\varphi_d(\mathbf{u})$ может быть получена из (5.17), (5.18) в виде

$$\varphi_d(\mathbf{u}) = \operatorname{mod}_{2\pi}\left[\frac{k}{f_1} R_0 \operatorname{sgn}(u)\,|\mathbf{u}| - \frac{k}{f_1} x_0 u\right], \tag{5.140}$$



а *б*

Рис. 5.54. Бинарная зонная пластинки для фокусировки в букву «х» (*а*), полутоновое распределение интенсивности (*б*), формируемое зонной пластинкой в плоскости $z = f_1$

где

$$\operatorname{sgn}(u) = \begin{cases} 1, & u \geqslant 0, \\ -1, & u < 0. \end{cases}$$

На рис. 5.54*а* показана фотомаска зонной пластинки (5.128), (5.136), (5.140), рассчитанной при следующих параметрах: $\lambda = 10{,}6$ мкм, $R_0 = 5$ мм, $f_1 = 500$ мм, $x_0 = 5{,}5$ мм, радиус освещающего пучка $R = 4$ мм. Смещение $x_0 = 5{,}5$ мм, большее чем радиус полукольца $R_0 = 5$ мм, было использовано для избежания интерференции между полями, формируемыми в $+1$ и -1 порядках. Рассчитанное полутоновое распределение интенсивности, формируемое зонной пластинкой на рис. 5.54*а*, представлено на рис. 5.54*б* и подтверждает работоспособность изложенного метода расчета.

5.5.2. ДОЭ с нелинейно комбинированными фазами

Комбинированный эффект линзы и дифракционной решетки можно создать при помощи одного ДОЭ, фазовая функция которого соответствует сумме фазовых функций линзы и дифракционной решетки. Следующая суперпозиция комплексных функций пропускания любых двух бинарных элементов,

$$T(\mathbf{u}) = \frac{1}{\sqrt{2}} \left(\exp\left(i\varphi_{b1}(\mathbf{u})\right) + i \exp\left(i\varphi_{b2}(\mathbf{u})\right) \right), \tag{5.141}$$

соответствует чисто фазовому элементу с фазовой функцией [121]

$$\varphi\left(\varphi_{b1}(\mathbf{u}), \varphi_{b2}(\mathbf{u})\right) = \begin{cases} \pi/4, & \varphi_{b1}(\mathbf{u}) = 0, \quad \varphi_{b2}(\mathbf{u}) = 0, \\ 3\pi/4, & \varphi_{b1}(\mathbf{u}) = \pi, \quad \varphi_{b2}(\mathbf{u}) = 0, \\ 5\pi/4, & \varphi_{b1}(\mathbf{u}) = \pi, \quad \varphi_{b2}(\mathbf{u}) = \pi, \\ 7\pi/4, & \varphi_{b1}(\mathbf{u}) = 0, \quad \varphi_{b2}(\mathbf{u}) = \pi. \end{cases} \tag{5.142}$$

Используем уравнения (5.141), (5.142) для расчета «ДОЭ с комбинированным эффектом» — ДОЭ, предназначенного для одновременной фокусировки в два набора линий. С этой целью введем функцию $\varphi(\varphi_{zp1}(\mathbf{u}), \varphi_{zp2}(\mathbf{u}))$, соответствующую суперпозиции (5.141) бинарных многофокусных зонных пластинок $\varphi_{zp1}(\mathbf{u})$ и $\varphi_{zp2}(\mathbf{u})$, фокусирующих сходящийся сферический пучок (5.129) в наборы линий S_1 и S_2, соответственно. Далее, определим фазовую функцию ДОЭ с комбинированным эффектом в виде

$$\varphi_{com}(\mathbf{u}) = \varphi_l(\mathbf{u}) + \varphi\left(\varphi_{zp1}(\mathbf{u}), \varphi_{zp2}(\mathbf{u})\right) - \varphi_0(\mathbf{u}), \quad \mathbf{u} \in G, \tag{5.143}$$

где $\varphi_0(\mathbf{u})$ — фаза освещающего пучка, а $\varphi_l(\mathbf{u}) = -\left(k\mathbf{u}^2/2f_1\right)$ — параксиальная фаза линзы.

Согласно (5.143) и (5.141), поле непосредственно за плоскостью ДОЭ с комбинированным эффектом, освещаемым пучком с комплексной амплитудой

$$F_0(\mathbf{u}) = \sqrt{I_0(\mathbf{u})} \exp(i\varphi_0(\mathbf{u})),$$

имеет вид

$$F(\mathbf{u}) = \sqrt{\frac{I_0(\mathbf{u})}{2}} \exp\left(-\frac{ik\mathbf{u}^2}{2f_1} + i\varphi_{zp1}(\mathbf{u})\right) + i\sqrt{\frac{I_0(\mathbf{u})}{2}} \exp\left(-\frac{ik\mathbf{u}^2}{2f_1} + i\varphi_{zp2}(\mathbf{u})\right). \tag{5.144}$$

Члены уравнения (5.144) соответствуют (с точностью до константы) полям, сформированным бинарными зонными пластинками $\varphi_{zp1}(\mathbf{u})$ и $\varphi_{zp2}(\mathbf{u})$ при сходящемся сферическом освещающем пучке (5.129). Согласно линейности оператора распространения света, ДОЭ с комбинированным эффектом (5.143) фокусирует одновременно в наборы линий S_1 и S_2.

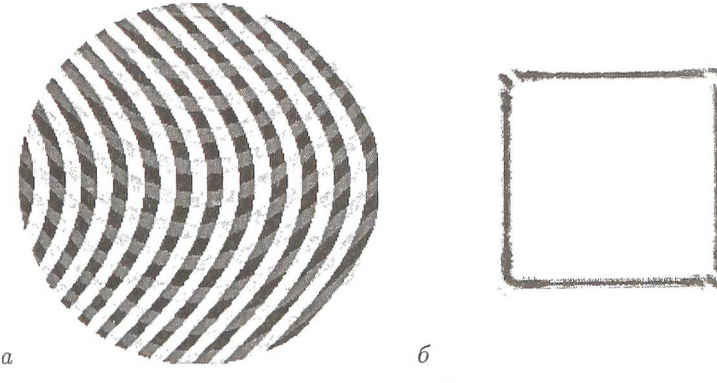
Пример 5.8. Рассмотрим расчет ДОЭ с комбинированным эффектом для фокусировки круглого пучка с постоянной интенсивностью в контур квадрата со стороной $2d$. Контур квадрата представим в виде двух наборов отрезков S_1 и S_2. Набор S_1 состоит из двух центрально-симметричных отрезков $x = d$, $|y| \leqslant d$ и $x = -d$, $|y| \leqslant d$, параллельных оси Oy, а набор S_2 — из двух центрально-симметричных отрезков $y = d$, $|x| \leqslant d$ и $y = -d$, $|x| \leqslant d$, параллельных оси Ox. В этом случае функции $\varphi_{zp1}(\mathbf{u})$ и $\varphi_{zp2}(\mathbf{u})$ в (5.143) соответствуют бинарным зонным пластинкам для фокусировки в наборы отрезков S_1 и S_2 в плоскости $z = f_1$. Причем, в силу симметрии задачи фокусировки, выполняется соотношение

$$\varphi_{zp1}(u, v) = \varphi_{zp2}(v, -u). \tag{5.145}$$

Фазовая функция зонной пластинки, фокусирующей в набор отрезков S_1, имеет вид (5.128), (5.136). Функция $\varphi_d(\mathbf{u})$ в (5.128) имеет вид (5.102), где $f_1 = f_2 = -f_3$, $\mathbf{x}_2 = (0, d)$, а функция $\varphi_2(\mathbf{u})$ вляется фазой фокусатора в отрезок $|x| \leqslant d$, в плоскости $z = f_1$. При круглом освещающем пучке радиуса R с постоянной интенсивностью, фазовую функцию $\varphi_2(\mathbf{u})$ можно получить из (5.20), (5.21) в виде

$$\varphi_2(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \frac{2kd}{\pi f}\left[u \arcsin(u/R) - R\left(1 - (u/R)^2\right)^{1/2} + \frac{R}{3}\left(1 - (u/R)^2\right)^{3/2}\right]. \tag{5.146}$$

Поскольку для бинарной решетки (5.136) $|c_1|^2 = |c_{-1}|^2 = 0{,}405$, то согласно разложению (5.130)–(5.133), зонная пластинка (5.128), (5.102), (5.136), (5.146) фокусирует 81% энергии освещающего пучка в заданный набор отрезков S_1. Для оценки работоспособности ДОЭ с комбинированным эффектом проводился расчет распределения интенсивности в плоскости фокусировки при следующих параметрах: $\lambda = 10{,}6$ мкм, $2d = 8$ мм, $R = 5$ мм, $f_1 = f_2 = -f_3 = 350$ мм. На рис. 5.55*а* показана амплитудная маска рассчитанной зонной пластинки в (5.143). Приведенное на рис. 5.55*б* полутоновое распределение интенсивности, формируемое ДОЭ с комбинированным эффектом, подтверждает работоспособность разработанного подхода.



а *б*

Рис. 5.55. Фотомаска зонной пластинки для фокусировки в контур квадрата (*а*); полутоновое распределение интенсивности при $z = f_1$ (*б*)

5.6. Дифракционные многофокусные линзы

В качестве примера метода расчета многофокусных фокусаторов, рассмотрим расчет многофокусного ДОЭ, обеспечивающего фокусировку в N точек на оптической оси с распределением энергии $I_1, \ldots, I_N$ между точками (фокусами). ДОЭ такого типа будем называть дифракционной многофокусной линзой. Поскольку точка — это частный случай фокальной линии, то для расчета многофокусной линзы достаточно, в общем представлении (5.101) для фазы многофокусного фокусатора, заменить фазовые функции $\varphi_1(\mathbf{u})$, $\varphi_2(\mathbf{u})$ фокусаторов в линию фазовыми функциями линзы. При этом фазовая функция многофокусной линзы принимает вид [28–30]

$$\varphi_{ml}(\mathbf{u}) = \varphi_1(\mathbf{u}) + \Phi[\varphi_2(\mathbf{u})] - \varphi_0(\mathbf{u}), \quad \mathbf{u} \in G, \tag{5.147}$$

где

$$\varphi_1(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f_1}, \quad \varphi_2(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(-\frac{k\mathbf{u}^2}{2f_2}\right) \tag{5.148}$$

— параксиальные фазовые функции линз с фокусами f_1 и f_2. Функция $\Phi[\varphi_2(\mathbf{u})]$ в (5.147) описывает нелинейное преобразование фазы линзы $\varphi_2(\mathbf{u})$. Таким образом, многофокусная линза представлена в виде суперпозиции обычной линзы $\varphi_1(\mathbf{u})$ и зонной пластинки $\Phi[\varphi_2(\mathbf{u})]$.

Согласно общему представлению (5.110), поле непосредственно за плоскостью многофокусной линзы, освещаемой пучком с комплексной амплитудой

$$F_0(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp(i\varphi_0(\mathbf{u})),$$

имеет вид

$$F(\mathbf{u}) = \sqrt{I_0(\mathbf{u})}\exp\left(-\frac{ik\mathbf{u}^2}{2f_1}\right)\sum_{j=-\infty}^{\infty} c_j \exp\left(ij\left(-\frac{k\mathbf{u}^2}{2f_2}\right)\right). \tag{5.149}$$

Выражение (5.149) соответствует, в параксиальном приближении, суперпозиции сходящихся сферических пучков с фокусами

$$F_j = \frac{f_1 f_2}{f_2 + jf_1} \tag{5.150}$$

и распределением энергии $|c_j|^2$ между пучками. Коэффициенты Фурье c_j функции $\exp(i\Phi[\xi])$ определены в уравнении (5.109). Пусть $j_1, \ldots, j_N$ — индексы, соответствующие заданным фокусным расстояниям в (5.150). Тогда для фокусировки в набор точек F_{j_i}, $i = \overline{1,N}$, с заданным распределением энергии $I_1, \ldots, I_N$ между фокусами, достаточно определить функцию нелинейного преобразования $\Phi[\xi]$ как фазовую функцию N-порядковой дифракционной решетки с интенсивностями $I_1, \ldots, I_N$ в порядках $j_1, \ldots, j_N$.

Как правило, расчет функции $\Phi[\xi]$ требует применения численных итерационных алгоритмов. В ряде случаев для расчета $\Phi[\xi]$ можно использовать аналитический метод расчета одномерных фокусаторов (см. (5.88)). Пусть $\tilde{\varphi}(\xi)$, $\xi \in [0,d]$ — фазовая функция одномерного фокусатора сходящегося цилиндрического пучка в отрезок с распределением интенсивности $I(x)$, $x \in [n_1\Delta, n_2\Delta]$, где $\Delta = \lambda f/d$ — размер

дифракционного пятна. Фазовая функция $\tilde{\varphi}(\xi)$ может быть получена из (5.88) в виде

$$\tilde{\varphi}(\xi) = \frac{k}{f}\int_0^{\xi} \chi(\eta)\, \mathrm{d}\eta, \tag{5.151}$$

где функция $\chi(\xi)$ является решением следующего дифференциального уравнения:

$$\frac{\mathrm{d}\chi(\xi)}{\mathrm{d}\xi} = \frac{1}{I(\chi(\xi))}, \tag{5.152}$$

с граничными условиями $\chi(0) = n_1\Delta$, $\chi(d) = n_2\Delta$.

Как показано в главе 2, функцию $\tilde{\varphi}(\xi)$ можно рассматривать как фазовую функцию дифракционной решетки с периодом d и интенсивностью дифракционных порядков

$$I_j = \begin{cases} I(j\Delta), & j = \overline{n_1, n_2}, \\ 0, & \text{иначе.} \end{cases} \tag{5.153}$$

Следовательно, функция нелинейного преобразования $\Phi[\xi] = \tilde{\varphi}(\xi)$ может быть рассчитана из уравнений (5.151), (5.152). Однако, как было установлено в процессе расчетов, геометрооптический подход к расчету функции $\Phi[\xi] = \tilde{\varphi}(\xi)$ оказывается работоспособным только при расчете решеток с большим числом (30–40 и более) подряд идущих порядков.

В заключение рассмотрим расчет «многофокусных линз с комбинированным эффектом», предназначенных для фокусировки одновременно в два набора точек. Первый набор состоит из точек, расположенных в одной плоскости, а второй набор — из точек, расположенных на оптической оси. Такие линзы являются частным случаем ДОЭ с комбинированным эффектом.

Если функция нелинейного преобразования $\Phi[\xi]$ в (5.147) определена для многопорядковой бинарной решетки, то функция $\Phi[\varphi_2(\mathbf{u})]$ является бинарной. Суперпозиция (5.142) функций пропускания бинарной зонной пластинки $\Phi[\varphi_2(\mathbf{u})]$ и многопорядковой бинарной решетки $\varphi_{gr}(\mathbf{u})$ соответствует чисто фазовому элементу с фазовой функцией $\varphi(\Phi[\varphi_2(\mathbf{u})], \varphi_{gr}(\mathbf{u}))$ (5.142). Далее, фазовую функцию многофокусной линзы с комбинированным эффектом предлагается определить в виде

$$\varphi_{cml}(\mathbf{u}) = \varphi_1(\mathbf{u}) + \varphi(\Phi[\varphi_2(\mathbf{u})], \varphi_{gr}(\mathbf{u})) - \varphi_0(\mathbf{u}), \quad \mathbf{u} \in G, \tag{5.154}$$

где $\varphi_0(\mathbf{u})$ — фаза освещающего пучка, $\varphi_1(\mathbf{u})$ — фазовая функция линзы в (5.147).

Согласно (5.141), комплексная функция пропускания линзы (5.154) может быть представлена в виде

$$\exp(i\varphi_{cml}(\mathbf{u})) = \frac{1}{\sqrt{2}}\exp(i\varphi_{ml}(\mathbf{u})) + \frac{i}{\sqrt{2}}\exp(i\varphi_1(\mathbf{u}) + i\varphi_{gr}(\mathbf{u}) - i\varphi_0(\mathbf{u})). \tag{5.155}$$

Первое слагаемое в (5.155) соответствует многофокусной линзе (5.147) с функцией $\Phi[\xi]$. Второе слагаемое содержит фазовые функции линзы и бинарной решетки и, таким образом, обеспечивает фокусировку в набор точек в плоскости $z = f_1$. При этом фаза $\varphi_0(\mathbf{u})$ во втором слагаемом компенсирует фазу освещающего пучка. В силу линейности оператора распространения света, линза (5.154) реализует одновременно фокусировку в наборы точек на оптической оси и в одной фиксированной плоскости.

Рассмотрим ряд примеров расчета многофокусных линз.

Пример 5.9. Рассмотрим расчет бифокальной линзы с фокусами f_a и f_b и равной долей энергии в каждом фокусе. Полагая в общем выражении (5.150) для фокусов многофокусной линзы $F_1 = f_a$, $F_{-1} = f_b$, получаем

$$f_1 = \frac{2f_a f_b}{f_a + f_b}, \quad f_2 = \frac{2f_a f_b}{f_a - f_b}. \tag{5.156}$$

Далее, определим функцию нелинейного преобразования $\Phi[\xi]$ как фазовую функцию 2-порядковой дифракционной решетки (5.136), (5.137). Подставляя (5.156), (5.137) в (5.149) получаем, с учетом главных ненулевых порядков $j = \pm 1$, следующее выражение для поля непосредственно за плоскостью бифокальной линзы:

$$F(\mathbf{u}) = c_1 \exp\left(-\frac{ik\mathbf{u}^2}{2f_a}\right) + c_{-1} \exp\left(-\frac{ik\mathbf{u}^2}{2f_b}\right), \tag{5.157}$$

где $c_{\pm 1}$ — коэффициенты Фурье (5.137). Выражение (5.157) описывает заданный процесс фокусировки в две точки (в параксиальном приближении). Согласно (5.137), $|c_1|^2 = |c_{-1}|^2 = 0{,}405$, то есть 81% энергии освещающего пучка концентрируется в требуемых фокусах f_a, f_b. Представляет интерес сравнение рассмотренной дифракционной линзы (5.147), (5.156), (5.136) с сегментированной бифокальной линзой с апертурой, состоящей из двух равных по площади сегментов. Фазовая функция сегментированной линзы имеет вид

$$\varphi_s(\mathbf{u}) = \begin{cases} -\dfrac{k\mathbf{u}^2}{2f_a}, & |\mathbf{u}| \leqslant \dfrac{R}{\sqrt{2}}, \\ -\dfrac{k\mathbf{u}^2}{2f_b}, & \dfrac{R}{\sqrt{2}} < |\mathbf{u}| \leqslant R. \end{cases} \tag{5.158}$$

На рис. 5.56*а* представлены графики распределения интенсивности I (в условных единицах) вдоль оптической оси z для дифракционной бифокальной линзы (непрерывная кривая) и сегментированной линзы (5.158) (пунктирная кривая). Графики были получены численным расчетом интеграла Френеля–Кирхгофа при следующих параметрах: $\lambda = 0{,}555$ мкм, $f_a = 30$ мм, $f_b = 34$ мм, $R = 1{,}5$ мм. Рис. 5.56*а* показывает, что у дифракционной бифокальной линзы глубина фокусов меньше, чем

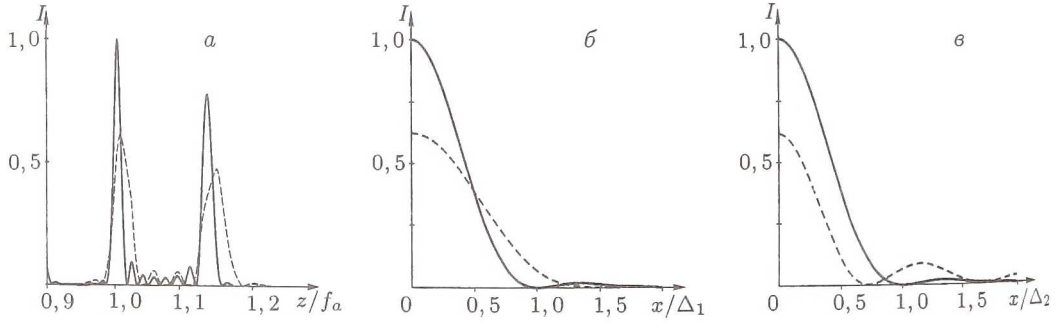


Рис. 5.56. Распределение интенсивности для дифракционной (сплошная линия) и сегментированной (пунктирная линия) бифокальных линз: (*а*) вдоль оптической оси, (*б*),(*в*) в фокальных плоскостях $z = f_a$ и $z = f_b$, где $\Delta_1 = 0{,}61\lambda f_a/R$ и $\Delta_2 = 0{,}61\lambda f_b/R$

у сегментированной линзы. Это объясняется тем, что в каждый фокус дифракционной линзы работает вся апертура. На рис. 5.56*б,в* показана рассчитанная функция рассеяния точки в двух фокальных плоскостях. Введем энергетическую эффективность E как долю энергии падающего пучка, сфокусированную внутри пятна с радиусом $\Delta = 0{,}61\lambda z/R$. В первой фокальной плоскости ($z = 30$ мм) E равно 32,8% и 32,6% для дифракционной и сегментированной линз, соответственно. Во второй фокальной плоскости ($z = 34$ мм) E равно 34,2% и 14,5%, соответственно. Поскольку только центральный сегмент сегментированной линзы фокусирует в первую фокальную плоскость, то ширина дифракционного пятна для сегментированной линзы в 1,5 раза больше, чем для дифракционной бифокальной линзы (см. рис. 5.56*б*). Аналогичный эффект имеет место и во второй фокальной плоскости с учетом двухлепесткового вида функции рассеяния точки для кольцевого зрачка (см. рис. 5.56*в*). Таким образом, проведенное исследование показывает, что дифракционная бифокальная линза работает подобно обычной (несегментированной) линзе в каждом из двух фокусов.

Пример 5.10. Расчет дифракционной линзы (5.147) с числом фокусом больше двух не представляет сложности. Рассмотрим расчет 7-ми фокусной линзы. Фаза 7-ми фокусной линзы равна сумме фазы обычной линзы $\varphi_1(\mathbf{u})$ и зонной пластинки $\Phi[\varphi_2(\mathbf{u})]$, соответствующей преобразованию линзы $\varphi_2(\mathbf{u})$ по закону 7-порядковой решетки. На рис. 5.57*а* приведен профиль бинарной зонной пластинки $\Phi[\varphi_2(\mathbf{u})]$, полученной преобразованием фазы $\varphi_2(\mathbf{u})$ по закону 7-порядковой бинарной решетки, концентрирующей излучение в порядках $-3, -2, -1, 0, 1, 2, 3$, при параметрах: $\lambda = 1{,}06$ мкм, $f_1 = 34$ мм, $f_2 = 1000$ мм и $R = 4$ мм. При этом в требуемые фокусы направляется более 80% энергии освещающего пучка. Доли энергии освещающего пучка, фокусируемые в дифракционной окрестности с радиусом $\Delta = 0{,}61\lambda F_j/R$, $j = -\overline{3,3}$ фактически постоянны для требуемых плоскостей $z = F_j$, $j = -\overline{3,3}$, а небольшое уменьшение величины интенсивности фокальных пиков с ростом

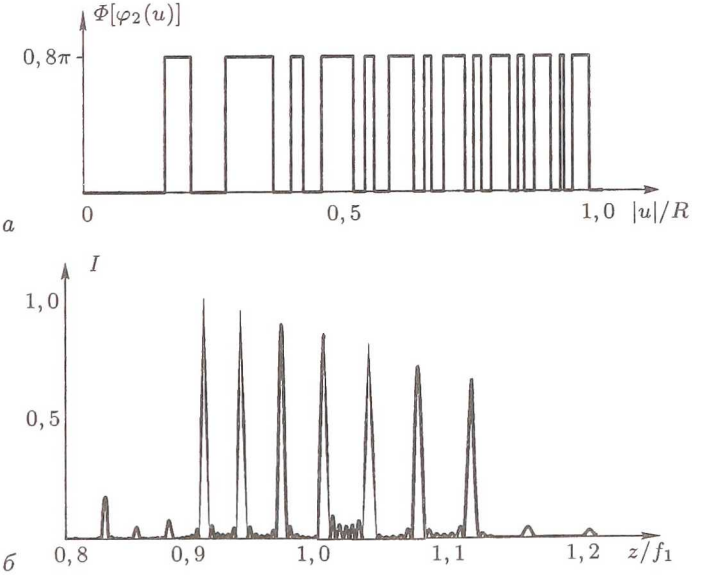


Рис. 5.57. Функция $\Phi[\varphi_2(\mathbf{u})]$ для семифокусной линзы (*а*); Распределение интенсивности вдоль оптической оси для семифокусной линзы (*б*)

фокусного расстояния связано с увеличением дифракционного размытия (интенсивность в геометрическом фокусе обратно пропорциональна фокусному расстоянию).

Пример 5.11. Рассмотрим аналитический расчет многофокусной линзы, основанный на использовании фазовой функции (5.151), (5.152) геометрооптического фокусатора в отрезок в качестве функции нелинейного преобразования $\Phi[\xi]$. Напомним, что использование геометрооптического подхода позволяет рассчитать многофокусную линзу для формирования только большого числа (30–40 и более) подряд идущих фокусов. Из (5.147), (5.148), (5.151) и (5.152) несложно получить фазовую функцию линзы с числом фокусов $N = n_2 - n_1$ равной интенсивности в виде

$$\varphi_{mf}(\mathbf{u}) = \varphi_1(\mathbf{u}) + \varphi_2(\mathbf{u})\left[n_1 + \frac{n_2 - n_1}{4\pi}\varphi_2(\mathbf{u})\right] - \varphi_0(\mathbf{u}). \tag{5.159}$$

Координаты фокусов линзы (5.159) описываются общим выражением (5.150) при $j = \overline{n_1, n_2}$. На рис. 5.58 приведено распределение интенсивности вдоль оптической оси линзы (5.159) при следующих параметрах: число фокусов $N = 40$ ($n1 = 20$, $n2 = 60$), $\lambda = 1{,}06$ мкм, $f_1 = 30$ мм, $f_2 = 4000$ мм, $R = 4$ мм. Неравномерность фокальных пиков объясняется как погрешностью геометрооптического расчета функции $\Phi[\xi]$, так и усилением дифракционного размытия с ростом фокусного расстояния. Среднеквадратичное отклонение δ распределения интенсивности по фокальным плоскостям линзы от постоянной величины, определяется неравномерностью распределения интенсивности, формируемого геометрооптическим фокусатором в отрезок $[n_1\Delta, n_2\Delta]$. Для приведенного примера значение δ составляет 25,2%.

Согласно (5.149), многофокусная линза (5.147) работает в каждом фокусе подобно обычной линзе. Если

$$f_2 = \frac{R^2}{2\lambda}, \tag{5.160}$$

то линза $\varphi_2(\mathbf{u})$ в (5.147) имеет только одну центральную зону. Для многофокусной линзы (5.147), (5.160) расстояние между соседними фокусами

$$l_j = F_{j-1} - F_j \approx 2\lambda\left(\frac{F_j}{R}\right)^2 \tag{5.161}$$

равно половине длины Френеля $l_f = 4\lambda f^2/R2$, определяющей размер дифракционного пятна вдоль оптической оси для линзы с фокусом f. В этом случае многофокусная

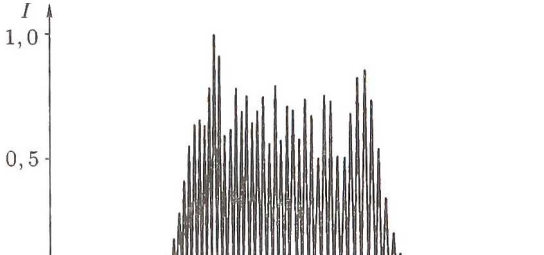


Рис. 5.58. Осевое распределение интенсивности для аналитически полученной 40-фокусной линзы (5.159)

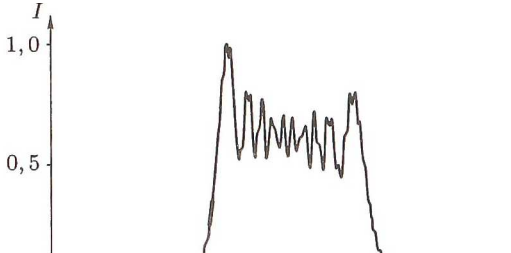


Рис. 5.59. Распределение интенсивности вдоль оптической оси для многофокусной линзы (5.159) при $f_2 = R_2/2\lambda$

линза (5.147), (5.148), (5.160) не создает отдельных фокусов, а формирует непрерывное распределение интенсивности вдоль оптической оси. На рис. 5.59 приведено расчетное распределение интенсивности вдоль оптической оси линзы (5.159), (5.160) при параметрах: $N = 40$ ($n_1 = 20$, $n_2 = 60$), $\lambda = 1{,}06$ мкм, $R = 4$ мм, $f_1 = 30$ мм, $f_2 = R2/2\lambda = 7547{,}17$ мм. Распределение интенсивности на рис. 5.59, как и ожидалось, имеет непрерывный характер.

5.7. Двухпорядковые фокусаторы

Многофокусные фокусаторы позволяют сфокусировать излучение в набор фокальных линий разного масштаба, но одинаковой формы. В данном разделе приводится расчет двухпорядковых фокусаторов, предназначенных для фокусировки в две различные фокальные линии L_1 и L_2. Исходными данными для такого расчета являются фазовые функции $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$ фокусаторов в линии L_1 и L_2, соответственно. Фазовая функция двухпорядкового фокусатора имеет вид [26]

$$\varphi(\mathbf{u}) = \frac{1}{2}\left[\varphi_a(\mathbf{u}) + \varphi_b(\mathbf{u})\right] + \Phi\left[h(\mathbf{u})\right], \tag{5.162}$$

где

$$h(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(\frac{1}{2}\left[\varphi_a(\mathbf{u}) - \varphi_b(\mathbf{u})\right]\right).$$

Как и ранее, $\Phi[h(\mathbf{u})]$ в (5.162) описывает специально подобранное нелинейное преобразование непрерывной функции $h(\mathbf{u}) \in [0, 2\pi]$.

Для определения вида нелинейного преобразования $\Phi[\xi]$, проведем анализ работы ДОЭ (5.162). Разлагая функцию $\exp(i\Phi[\xi])$ в ряд Фурье на интервале $[0, 2\pi)$ и полагая $\xi = h(\mathbf{u})$, запишем функцию пропускания ДОЭ (5.162) в виде

$$\exp\left[i\varphi(\mathbf{u})\right] = \sum_{j=-\infty}^{\infty} c_j \exp\left[\frac{i}{2}\left(\varphi_a(\mathbf{u})(1+j) + \varphi_b(\mathbf{u})(1-j)\right)\right], \tag{5.163}$$

где c_j — коэффициенты Фурье (5.109) функции $\exp(i\Phi[\xi])$.

Согласно (5.163), при дифракции освещающего пучка на двухпорядковом фокусаторе (5.162) формируется много порядков. Изображение, формируемое в j-м дифракционном порядке, соответствует преобразованию освещающего пучка ДОЭ с фазовой функцией

$$\varphi_j(\mathbf{u}) = \frac{1}{2}\left[\varphi_a(\mathbf{u})(1+j) + \varphi_b(\mathbf{u})(1-j)\right]. \tag{5.164}$$

Фокусировка в заданные фокальные линии L_1 и L_2 происходит в 1-м и -1-м дифракционных порядках: $\varphi_1(\mathbf{u}) = \varphi_a(\mathbf{u})$, $\varphi_{-1}(\mathbf{u}) = \varphi_b(\mathbf{u})$. Доля энергии освещающего пучка, фокусируемая в j-м порядке, пропорциональна квадратам модулей коэффициентов Фурье c_j. Поэтому функция нелинейного преобразования $\Phi[\xi]$ должна выбираться из условия равенства нулю коэффициентов Фурье c_j при $j \neq \pm 1$. Такой нелинейности соответствует фазовая функция двухпорядковой дифракционной решетки, концентрирующей излучение в 1-м и -1-м порядках. Для формирования равной интенсивности вдоль фокальных линий L_1, L_2, значения $|c_1|^2$ и $|c_{-1}|^2$ при синтезе решетки должны быть выбраны пропорциональными геометрическим

размерам линий L_1 и L_2. При $|c_1|^2 \neq |c_{-1}|^2$, расчет двухпорядковой решетки $\Phi[\xi]$ осуществляется с использованием итерационных процедур. При $|c_1|^2 = |c_{-1}|^2$, функция $\Phi[\xi]$ соответствует фазовой функции двухпорядковой бинарной решетки (5.136), (5.137). Для решетки (5.136) $|c_1|^2 = |c_{-1}|^2 = 0{,}405$. В этом случае двухпорядковый фокусатор (5.162), (5.136) фокусирует 81% энергии освещающего пучка в 1-м и -1-м дифракционных порядках.

Не существует фазовых двухпорядковых решеток с энергетической эффективностью в 100%. Поэтому, наряду с заданными фокальными линиями L_1 и L_2, двухпорядковый фокусатор (5.162) также создаст паразитные изображения, соответствующие ненулевым коэффициентам Фурье c_j при $j \neq \pm 1$. Степень влияния паразитных изображений оценивается отдельно в каждой конкретной задаче и при необходимости может быть уменьшена путем «внеосевой» фокусировки. Внеосевая фокусировка состоит во введении фазовых функций призм в функции $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$. Призмы играют роль несущих и позволяют сдвинуть паразитные изображения относительно полезных.

Рассмотренный метод расчета двухпорядковых фокусаторов не накладывает никаких ограничений на структуру функций $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$. Например, функции $\varphi_a(\mathbf{u})$, $\varphi_b(\mathbf{u})$ в (5.162) могут быть рассчитаны с использованием итерационных процедур. Если функции $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$ вляются фазовыми функциями многофокусных фокусаторов, формирующих наборы линий S_1 и S_2, то двухпорядковый ДОЭ формирует в 1-м и -1-м порядках наборы фокальных линий S_1 и S_2.

В последующих примерах мы рассмотрим наиболее практически важные случаи фокусировки в две точки и в два отрезка с различной ориентацией.

Пример 5.12. Рассмотрим расчет двухпорядковой линзы, фокусирующей плоский пучок в две точки $z = f_a$ и $z = f_b$ на оптической оси. Фазовая функция двухпорядковой линзы имеет вид (5.162), где $\Phi[\xi]$ — фазовая функция двухпорядковой решетки (5.136), а функции $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$ соответствуют фазовым функциям линз с фокусами f_a и f_b. В непараксиальном приближении, $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$ имеют вид

$$\left.\begin{aligned} \varphi_a(\mathbf{u}) &= -k\sqrt{f_a^2 + \mathbf{u}^2}, \\ \varphi_b(\mathbf{u}) &= -k\sqrt{f_b^2 + \mathbf{u}^2}. \end{aligned}\right\} \tag{5.165}$$

В отличие от бифокальной линзы (5.147), (5.148), (5.136), (5.156) двухпорядковая линза (5.162), (5.136), (5.165) является обобщением на случай непараксиального приближения.

Пример 5.13. Рассмотрим расчет двухпорядкового фокусатора в крест, состоящий из двух перпендикулярных отрезков длины $2d$. В этом случае, функция $\Phi[\xi]$ также имеет вид (5.136), а функции $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$ в (5.162) являются фазовыми функциями фокусаторов в отрезки, составляющие крест. При квадратной апертуре G $(2a \times 2a)$ фокусатора и плоском освещающем пучке функции $\varphi_a(\mathbf{u})$ и $\varphi_b(\mathbf{u})$ могут быть получены из (5.20), (5.21) в виде

$$\left.\begin{aligned} \varphi_a(\mathbf{u}) &= -\frac{k\mathbf{u}^2}{2f} + \frac{kd}{2fa}u^2, \\ \varphi_b(\mathbf{u}) &= -\frac{k\mathbf{u}^2}{2f} + \frac{kd}{2fa}v^2, \end{aligned}\right\} \tag{5.166}$$

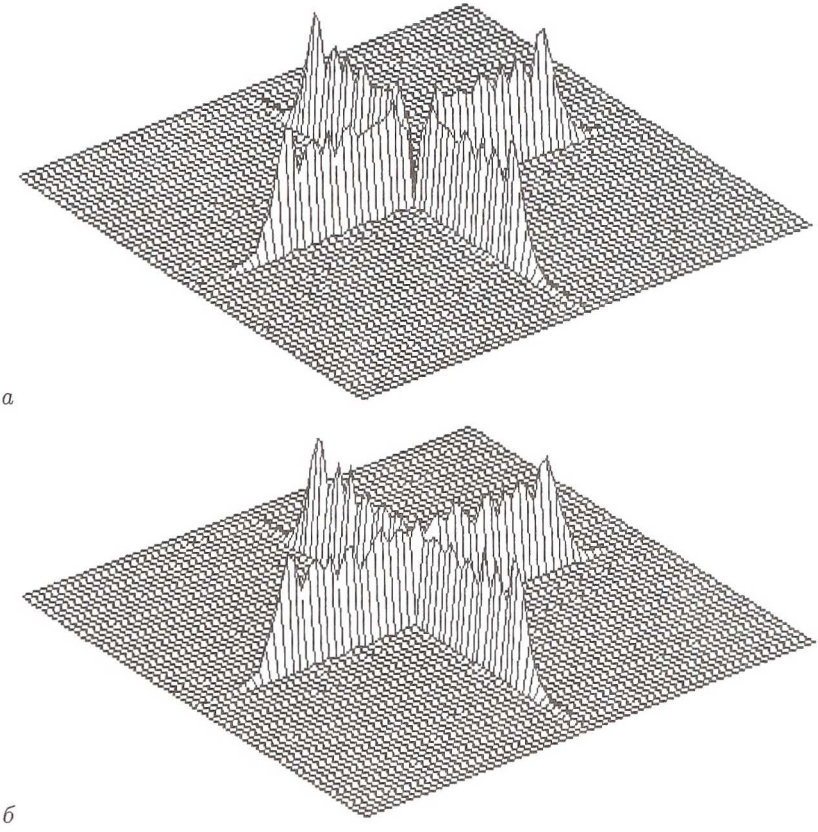
где f — расстояние от двухпорядкового фокусатора до фокальной плоскости.

Проведем анализ паразитных изображений, возникающих при фокусировке в крест. Согласно уравнению (5.164), структура паразитных изображений определяется

фазовыми функциями

$$\varphi_j(\mathbf{u}) = -\frac{k\mathbf{u}^2}{2f} + \frac{kd}{4fa}\left((1+j)\,u^2 + (1-j)\,v^2\right). \tag{5.167}$$

Функция (5.167) является фазовой функцией фокусатора плоского пучка в прямоугольник [87–89] со сторонами $|d(1+j)|$ и $|d(j-1)|$. Благодаря равномерному распределению энергии по площади прямоугольников интенсивность паразитных порядков будет значительно меньше, чем интенсивность на фокальных линиях, составляющих крест. Это позволяет предположить, что паразитные порядки не должны существенно влиять на качество создаваемого полезного изображения. Для оценки разработанного подхода был проведен дифракционный расчет интенсивности в фокальной плоскости двухпорядкового фокусатора в крест (5.162), (5.136), (5.166) при следующих параметрах: $\lambda = 1{,}06$ мкм, $f = 100$ мм, $2d = 1{,}2$ мм, $2a = 5$ мм (см. рис. 5.65*а*). На рис. 5.60*а* четко видна структура излучения, сфокусированного в форме креста. Паразитные изображения незаметны, однако в центре креста наблюдается спад интенсивности. Спад интенсивности возникает из-за разных знаков коэффициентов Фурье c_1 и c_{-1} в (5.137). Согласно (5.163), это приводит к вычитанию полей, соответствующих фокусировке в отрезки креста. Спад интенсивности в центре креста можно компенсировать при помощи циклического сдвига профиля



а

б

Рис. 5.60. Распределение интенсивности в фокальной плоскости двухпорядковых ДОЭ фокусирующих в крест: (*а*) для функции $\Phi[\xi]$ в уравнении (5.136); (*б*) при циклическом сдвиге профиля $\Phi[\xi]$ на $\pi/6$

решетки (5.136). При циклическом сдвиге профиля решетки (5.136) на величину φ_0, коэффициенты Фурье (5.137) принимают вид

$$c_j^{\varphi_0} = c_j \exp(ij\varphi_0). \tag{5.168}$$

Амплитуда поля в центре креста пропорциональна величине $\eta = c_1 \exp(i\varphi_0) + c_{-1} \exp(-i\varphi_0) = 2ic_1 \sin(\varphi_0)$. Поэтому при $\varphi_0 = \pi/6$, $\eta = ic_1$, и спад интенсивности в центре должен исчезнуть. Распределение интенсивности поля в фокальной плоскости двухпорядкового фокусатора на рис. 5.60*б*, полученное при циклическом сдвиге профиля (5.136) на $\pi/6$, подтверждает отсутствие спада интенсивности в центре креста.

Интересно сравнить двухпорядковый фокусатор в крест с сегментированным фокусатором в крест, рассмотренным в пункте 5.3.6. Апертура $2a \times 2a$ сегментированного фокусатора в пункте 5.3.6 состоит из четырех квадратных сегментов $a \times a$, фокусирующих в четыре отрезка длиной d, составляющих крест. Сегментация апертуры при фокусировке в крест приводит к двукратному уменьшению длины слоя и, как следствие, к двукратному увеличению дифракционной ширины отрезков креста. Поэтому, несмотря на то, что двухпорядковый фокусатор теряет 19% энергии освещающего пучка в паразитных порядках, интенсивность на отрезках креста, формируемого двухпорядковым фокусатором, должна быть выше за счет меньшей (в 2 раза) дифракционной ширины линий. Проведенный расчет средней интенсивности поля на геометрическом кресте при вышеуказанных параметрах показал, что для двухпорядкового фокусатора интенсивность на кресте в 1,8 раза выше, чем для сегментированного фокусатора. Для двухпорядкового фокусатора также удалось аналитически скомпенсировать интерференционные эффекты в центре креста; для сегментированного фокусатора подобная коррекция осуществлялась подбором параметров сдвига отрезков креста.

5.8. Расчет спектральных фокусаторов

Для работы с различными спектральными компонентами можно использовать цветоделительные фазовые дифракционные решетки, позволяющие разделить три различных спектральных компонента освещающего пучка по -1, 0 и $+1$ дифракционным порядкам [31–33].

В данном разделе рассмотрен расчет спектральных фокусаторов — фазовых ДОЭ, которые позволяют:

— сфокусировать три различных спектральных компонента освещающего пучка в три одинаковые фокальные области;

— при изменении длины волны изменять конфигурацию области фокусировки для трех различных длин волн. Основная идея расчетного метода состоит в нелинейном преобразовании фазовой функции фокусатора по закону цветоделительной дифракционной решетки.

5.8.1. Цветоделительные решетки

Для описания работы спектральных фокусаторов предварительно рассмотрим принцип работы цветоделительных фазовых дифракционных решеток [31–33]. Решетка, позволяющая разделить три спектральных компонента, соответствующих

длинам волн

$$\lambda_0, \quad \lambda_{+1} = \frac{\lambda_0 N}{(N+1)}, \quad \lambda_{-1} = \frac{\lambda_0 N}{(N-1)}, \tag{5.169}$$

по 0, +1 и −1 дифракционным порядкам, имеет на периоде N ступенек равной ширины. Высота ступенек определяется по формуле

$$d_i = \frac{\lambda_0 i}{(n_0 - 1)}, \quad i = \overline{0, N-1}, \tag{5.170}$$

где n_0 — показатель преломления материала решетки для длины волны λ_0. Пренебрегая дисперсией материала решетки, получается, что решетка (5.170) для плоских пучков с длинами волн (5.169) формирует следующие фазовые набеги:

$$\varphi_0 = 2\pi i, \quad \varphi_{-1} = 2\pi i \frac{\lambda_0}{\lambda_{-1}} = 2\pi i - \frac{2\pi i}{N}, \quad \varphi_{+1} = 2\pi i \frac{\lambda_0}{\lambda_{+1}} = 2\pi i + \frac{2\pi i}{N}. \tag{5.171}$$

С учетом 2π-периодичности фазы, представим (5.171) в виде

$$\varphi_0 = 0, \quad \varphi_{-1} = -\frac{2\pi i}{N}, \quad \varphi_{+1} = \frac{2\pi i}{N}, \quad i = \overline{0, N-1}. \tag{5.172}$$

Согласно (5.172), для длины волны λ_0 фазовый набег равен нулю и, следовательно, данный спектральный компонент направляется в нулевой порядок. Для длин волн $\lambda_{\pm 1}$ фазовые набеги (5.172) соответствуют фазовым функциям призм, квантованных по N уровням:

$$\varphi_{\pm 1} = F_N\left(\pm 2\pi u/d\right), \tag{5.173}$$

где d — период решетки, а функция

$$F_N(\xi) = \frac{2\pi}{N} \operatorname{int}\left(\frac{\xi N}{2\pi}\right) \tag{5.174}$$

описывает операцию квантования по N уровням. Это обеспечивает фокусировку длин волн $\lambda_{\pm 1}$ в порядках ± 1. Вследствие квантования, доля энергии спектральных компонентов $\lambda_{\pm 1}$, фокусируемая в порядках ± 1, имеет вид

$$E(N) = \operatorname{sinc}^2(\pi/N). \tag{5.175}$$

График функции $E(N)$ приведен на рис. 5.61. Например, при $N = 4$ решетка (5.170) направляет 100% энергии спектрального компонента λ_0 в нулевой порядок и 81% энергии компонентов $\lambda_{+1} = 4\lambda_0/5$ и $\lambda_{-1} = 4\lambda_0/3$ в порядки +1 и −1.

Длины волн (5.169) и номера дифракционных порядков для решетки (5.170) не являются независимыми между собой. При работе только с двумя спектральными компонентами, связь длин волн $\lambda_0/\lambda_{\pm 1} = (N \pm 1)/N$ и номеров порядков можно обобщить. Рассмотрим расчет двухволновой решетки для разделения по двум порядкам с номерами m_0, m_1 длин волн λ_0, λ_1, связанных существенно более общим соотношением

$$\lambda_0/\lambda_1 = M/N, \tag{5.176}$$

где M и N — взаимно-простые числа. Двухволновую решетку предлагается строить

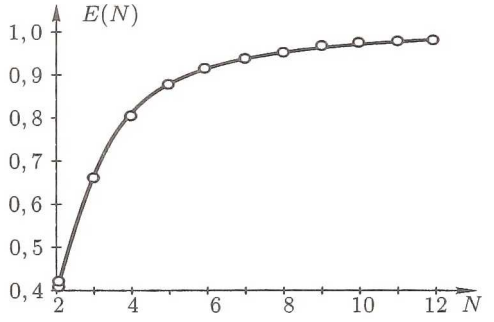


Рис. 5.61. Энергетическая эффективность квантованной линейной решетки

из фазовых функций призм:

$$\begin{aligned} F_M(\varphi(u)) &= F_M\left(\mathrm{mod}_{2\pi}\left(\frac{2\pi}{d}n_0u\right)\right) = \frac{2\pi}{M}\,\mathrm{int}\left[\frac{M}{2\pi}\,\mathrm{mod}_{\,2\pi}\left(\frac{2\pi}{d}m_0u\right)\right], \\ F_N(\varphi_1(u)) &= F_N\left(\mathrm{mod}_{2\pi}\left(\frac{2\pi}{d}n_1u\right)\right) = \frac{2\pi}{N}\,\mathrm{int}\left[\frac{N}{2\pi}\,\mathrm{mod}_{\,2\pi}\left(\frac{2\pi}{d}m_1u\right)\right], \end{aligned} \tag{5.177}$$

квантованных по M и N уровням. Определим функцию высоты рельефа двухволновой решетки в виде

$$h(u) = \frac{\lambda_0}{2\pi(n_0-1)}\left[F_M(\varphi(u)) + 2\pi M_0\left(F_M(\varphi(u)), F_N(\varphi_1(u))\right)\right], \tag{5.178}$$

где $M_0\left(F_M(\varphi(u)), F_N(\varphi_1(u))\right)$ — некоторая неизвестная функция, принимающая целочисленные положительные значения. Для удобства выкладок удобно представить высоту рельефа (5.178) через индексы значений $2\pi i/M$, $i = \overline{0, M-1}$, $2\pi j/N$, $j = \overline{0, N-1}$, принимаемых квантованными функциями $F_M(\varphi(\mathbf{u}))$, $F_N(\varphi_1(\mathbf{u}))$:

$$h(i,j) = \frac{\lambda_0}{2\pi(n_0-1)}\left[\frac{2\pi}{M}i + 2\pi M_0(i,j)\right], \tag{5.179}$$

где

$$i = \frac{M}{2\pi}F_M(\varphi(u)), \quad i = \overline{0, M-1}, \quad j = \frac{N}{2\pi}F_N(\varphi_1(u)), \quad j = \overline{0, N-1}. \tag{5.180}$$

Решетка (5.179) при $\lambda = \lambda_0$ формирует фазовый набег

$$\Phi(i,j;\lambda_0) = \frac{2\pi i}{M} + 2\pi M_0(i,j)), \quad i = \overline{0, M-1}, \quad j = \overline{0, N-1}, \tag{5.181}$$

равный по модулю 2π фазе квантованной призмы $F_M(\varphi(u))$. Это обеспечивает фокусировку компонента λ_0 в порядке m_0. При $\lambda = \lambda_1$ фазовый набег, формируемый решеткой, имеет вид

$$\Phi(i,j;\lambda_{+1}) = 2\pi\frac{M}{N}\left[\frac{i}{M} + M_0(i,j)\right], \quad i = \overline{0, M-1}, \quad j = \overline{0, N-1}. \tag{5.182}$$

Для фокусировки компонента λ_1 в порядке m_1, функция $M_0(i,j)$ должна выбираться из условия равенства по модулю 2π выражения (5.182) квантованной фазе

$F_N(\varphi_1(u))$. Данное условие дает для расчета функции $M_0(i,j)$ следующее уравнение:

$$2\pi\frac{M}{N}\left[\frac{i}{M}+M_0(i,j)\right]=2\pi\left[\frac{j}{N}+M_1(i,j)\right],\quad i=\overline{0,M-1},\quad j=\overline{0,N-1},\tag{5.183}$$

где $M_1(i,j)$ — произвольная функция, принимающая целочисленные значения. Непосредственной подстановкой можно убедиться, что решение уравнения (5.183) имеет вид

$$\left.\begin{aligned}&M_0(i,j)=\alpha_1(j-i)+Nz=\mathrm{mod}_N(\alpha_1(j-i)),\\&M_1(i,j)=\alpha_2(j-i)+Mz,\quad i=\overline{0,M-1},\ j=\overline{0,N-1},\end{aligned}\right\}\tag{5.184}$$

где α_1, α_2 — целые числа, определяемые из уравнения

$$M\alpha_1-N\alpha_2=1,\tag{5.185}$$

а z — наименьшее целое число, выбираемое из условия неотрицательности функции $M_0(i,j)$. Поскольку числа M и N взаимно-простые, то уравнение (5.185) всегда имеет решение в целых числах по теореме о наибольшем общем делителе. Согласно (5.184), рельеф двухволновой решетки имеет вид

$$h(i,j)=\frac{\lambda_0}{2\pi(n_0-1)}\left[\frac{2\pi}{M}i+2\pi\ \mathrm{mod}_N(\alpha_1(j-i))\right],\tag{5.186}$$

где индексы (i,j) определены в (5.180). Частным случаем решетки (5.186) является решетка

$$h(j)=\frac{\lambda_0}{(n_0-1)}\ \mathrm{mod}_N(\alpha_1 j),\quad j=\frac{N}{2\pi}F_N\left(\mathrm{mod}_{2\pi}\left(\frac{2\pi}{d}m_1u\right)\right),\tag{5.187}$$

для разделения спектральных компонент

$$\lambda_0,\quad \lambda_{+1}=\lambda_0N/M=\lambda_0N/(N+p),\quad \lambda_{-1}=\lambda_0N/(N-p),\quad p=M-N\tag{5.188}$$

по порядкам 0 и $\pm m_1$. Решетка (5.170) является частным случаем решетки (5.188) при $M=N+1$ и $m_1=1$. В дальнейшем будет использована двухволновая решетка для разделения спектральных компонент (5.188) по порядкам 0 и ±1. Согласно (5.187), такая решетка имеет на периоде N ступенек равной ширины с высотой

$$d_i=\frac{\lambda_0}{(n_0-1)}\ \mathrm{mod}_N(\alpha_1 i),\quad i=\overline{0,N-1},\tag{5.189}$$

где α_1 определяется из (5.185).

5.8.2. Спектральные фокусаторы в набор одинаковых фокальных областей

Использование комбинации цветоделительная решетка — линза позволяет сфокусировать три спектральных компонента в три точки в фокальной плоскости линзы. Рассмотрим расчет спектрального фокусатора — фазового ДОЭ, выполняющего разделение и фокусировку трех длин волн (5.188) в три одинаковых фокальных области

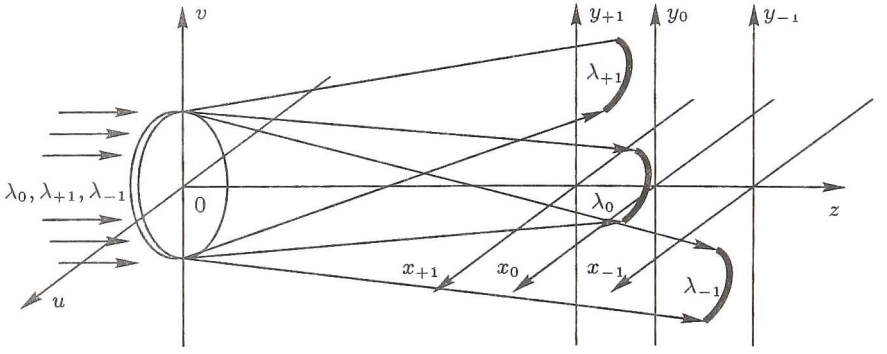


Рис. 5.62. Геометрия фокусировки спектральных компонентов в набор областей

заданной формы, расположенных в одной или различных плоскостях вдоль оптической оси (рис. 5.62). Микрорельеф спектрального фокусатора определим в виде

$$h(\mathbf{u}) = \frac{\lambda_0}{2\pi(n_0 - 1)}\Phi(\mathbf{u}; \lambda_0) = \frac{\lambda_0}{2\pi(n_0 - 1)}\left[\varphi(\mathbf{u}) + G(\tilde{\varphi}(\mathbf{u}))\right], \tag{5.190}$$

где $\varphi(\mathbf{u}) \in [0, 2\pi)$ — фазовая функция фокусатора, рассчитанная из условия фокусировки плоского монохроматического пучка с длиной волны λ_0 в область D в плоскости $z = f$. Для удобства дальнейших выкладок выделим из функции $\varphi(\mathbf{u})$ фазовую функцию линзы в явном виде

$$\varphi(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(-\frac{\pi\mathbf{u}^2}{\lambda_0 f} + \varphi_D(\mathbf{u})\right), \tag{5.191}$$

где добавка к линзе $\varphi_D(\mathbf{u})$ обеспечивает фокусировку сходящегося сферического пучка с фокусом f в область D при $z = f$.

Функция G в (5.190) описывает нелинейное преобразование фазы

$$\tilde{\varphi}(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(-\frac{p}{N}\cdot\varphi(\mathbf{u}) + \varphi_{pr}(\mathbf{u}) + \varphi_l(\mathbf{u})\right) \tag{5.192}$$

по закону фазового набега цветоделительной решетки для длины волны λ_0:

$$G(\xi) = 2\pi \bmod_N (\alpha_1 \,\mathrm{int}\,[N\xi/2]), \quad \xi \in [0, 2\pi). \tag{5.193}$$

Функции

$$\varphi_{pr}(\mathbf{u}) = \frac{2\pi}{\lambda_0 f}\mathbf{x}_0\mathbf{u}, \quad \varphi_l(\mathbf{u}) = -\frac{\pi}{\lambda_0 f_l}\mathbf{u}^2 \tag{5.194}$$

в (5.192) являются фазовыми функциями призмы и линзы и предназначены для разделения областей фокусировки для различных длин волн.

Отметим, что функция $\Phi(\mathbf{u}; \lambda_0)$ в (5.190) соответствует фазовому набегу, формируемому спектральным фокусатором для длины волны λ_0. При этом для длины волны $\lambda \neq \lambda_0$, фазовый набег имеет вид

$$\Phi(\mathbf{u}; \lambda) = \frac{\lambda_0}{\lambda}\Phi(\mathbf{u}; \lambda_0). \tag{5.195}$$

Проведем анализ работы спектрального фокусатора для освещающего пучка, состоящего из трех некогерентных плоских пучков с длинами волн (5.188). Для описания связи распределения интенсивности поля с фазовым набегом $\Phi(\mathbf{u};\lambda)$ будем использовать интеграл Кирхгофа в приближении Френеля:

$$I(\mathbf{x};z,\lambda)=\left|\frac{1}{\lambda z}\int \exp(i\frac{\lambda_0}{\lambda}\Phi(\mathbf{u};\lambda_0))\exp\left(\frac{i\pi}{\lambda z}(\mathbf{x}-\mathbf{u})^2\right)\mathrm{d}^2\mathbf{u}\right|^2. \tag{5.196}$$

При $\lambda=\lambda_0$ фазовый набег, вносимый спектральным ДОЭ, соответствует фазовой функции $\varphi(\mathbf{u})$, что обеспечивает фокусировку в область D в плоскости $z=f$. Для пучков с длинами волн λ_{+1} и λ_{-1} фазовые набеги, с учетом 2π-периодичности фазы, описываются следующими функциями:

$$\begin{aligned}\Phi(\mathbf{u};\lambda_{+1})&=\frac{N+p}{N}\left(\varphi(\mathbf{u})+G\left(\tilde{\varphi}(\mathbf{u})\right)\right)=\frac{N+p}{N}\varphi(\mathbf{u})+F_N\left(\tilde{\varphi}(\mathbf{u})\right),\\ \Phi(\mathbf{u};\lambda_{-1})&=\frac{N-p}{N}\left(\varphi(\mathbf{u})+G\left(\tilde{\varphi}(\mathbf{u})\right)\right)=\frac{N-p}{N}\varphi(\mathbf{u})-F_N\left(\tilde{\varphi}(\mathbf{u})\right).\end{aligned} \tag{5.197}$$

Проанализируем работу спектрального фокусатора при длине волны λ_{+1}. Нелинейная операция квантования $F_N(\tilde{\varphi}(\mathbf{u}))$ в (5.197) приводит к появлению дополнительных дифракционных порядков [90–92, 25–30]. Для анализа эффектов квантования разложим комплексную амплитуду поля $F(\mathbf{u};\lambda_{+1})=\exp(i\Phi(\mathbf{u};\lambda_{+1}))$ непосредственно за плоскостью ДОЭ в ряд по дифракционным порядкам [25–30]:

$$\begin{aligned}&\exp(i\Phi(\mathbf{u};\lambda_{+1}))=\\ &=\sum_{m=-\infty}^{m=\infty}c_m\exp\left(i\frac{N+p}{N}\varphi(\mathbf{u})+im\left(\left(1-\frac{N+p}{N}\right)\varphi(\mathbf{u})+\varphi_{pr}(\mathbf{u})+\varphi_l(\mathbf{u})\right)\right),\end{aligned} \tag{5.198}$$

где

$$c_m=\begin{cases}\operatorname{sinc}(\pi m/N)\exp(-i\pi/N), & m=1+pN,\quad p=0,\pm1,\pm2,\dots,\\ 0, & m\neq 1+pN\end{cases} \tag{5.199}$$

– коэффициенты Фурье функции $\exp(iF_N(\xi))$. Согласно (5.198), в 1-м порядке дифракции (при $m=1$ в (5.198)) формируется пучок с фазовой функцией

$$\Phi_{+1}(\mathbf{u};\lambda_{+1})=\varphi(\mathbf{u})+\varphi_{pr}(\mathbf{u})+\varphi_l(\mathbf{u}). \tag{5.200}$$

Для оценки структуры поля, формируемого пучком с фазовым фронтом $\Phi_{+1}(\mathbf{u};\lambda_{+1})$, подставим фазовую функцию (5.200) в интеграл Кирхгофа–Френеля (5.196). Проведя несложные преобразования, представим распределение интенсивности, формируемое в 1-м порядке при

$$z=z_{+1}=\frac{N+p}{N}\cdot\frac{ff_l}{f+f_l} \tag{5.201}$$

в виде

$$\begin{aligned}I(\mathbf{x};z,\lambda_{+1})&=\left|\frac{1}{\lambda_{+1}z_{+1}}\int\limits_G\exp(i\varphi_D(\mathbf{u})-\varphi_{pr}(\mathbf{u}))\exp\left(\frac{i\pi}{\lambda_{+1}z_{+1}}\mathbf{x}\mathbf{u}\right)d^2\mathbf{u}\right|^2=\\ &=p_{+1}^2\cdot I\left((\mathbf{x}\cdot p_{+1}-\mathbf{x}_0);f,\lambda_0\right),\end{aligned} \tag{5.202}$$

где

$$I(\mathbf{x}; f, \lambda_0) = \left| \frac{1}{\lambda_0 f} \int\limits_G \exp(i\varphi_D(\mathbf{u})) \exp\left(\frac{i\pi}{\lambda_0 f}\mathbf{x}\mathbf{u}\right) \mathrm{d}^2\mathbf{u} \right|^2 \tag{5.203}$$

— распределение интенсивности при $z = f$, формируемое при фазовом набеге $\varphi_D(\mathbf{u})$ для пучка с длиной волны λ_0. По определению функции $\varphi_D(\mathbf{u})$, распределение интенсивности (5.203) соответствует фокусировке в область D. Согласно (5.202), (5.203), распределение интенсивности, формируемое спектральным фокусатором в 1-м порядке при $\lambda = \lambda_{+1}$, соответствует фокусировке в смещенную на вектор $\mathbf{x}_0$ область D в плоскости $z = z_{+1}$. Размер области D в дифракционных пятнах не изменяется. При этом масштабный множитель

$$p_{+1} = 1 + f/f_l \tag{5.204}$$

в (5.202) описывает изменение размера области D, вызванное уменьшением размера дифракционного пятна:

$$\Delta(z_{+1}, \lambda_{+1}) = \lambda_{+1} z_{+1}/a = \frac{1}{p_{+1}}(\lambda_0 f/a), \tag{5.205}$$

где a — характерный размер апертуры ДОЭ. Доля энергии E спектрального компонента λ_{+1} соответствует квадрату модуля коэффициента Фурье c_1 (см. (5.198)) и при $N > 3$ составляет более 80%.

Повторяя приведенные рассуждения для пучка с длиной волны λ_{-1} получим, что фазовый фронт, формируемый ДОЭ (5.190)–(5.194) в основном 1-м порядке дифракции, соответствует фокусировке при

$$z = z_{-1} = \frac{N-p}{N} \frac{f f_l}{f_l - f} \tag{5.206}$$

в смещенную на вектор $\mathbf{x}_0$ область D. При этом изменение физического размера области D описывается множителем

$$p_{-1} = 1 - f/f_l. \tag{5.207}$$

Параметр f_l в (5.192), (5.194) позволяет варьировать положением плоскостей фокусировки. В частности, при $f_l = (1/p)Nf$, $z_{+1} = z_{-1} = f$. В этом случае длины волн (5.188) фокусируются в три разнесенные на вектор $\mathbf{x}_0$ области D в одной плоскости $z = f$.

В заключение отметим, что для снижения требований к разрешающей способности фотолитографической технологии, целесообразно рассчитывать спектральный фокусатор как дополнение к тонкой линзе с фокусом f. В этом случае вместо функции $\varphi(\mathbf{u})$ в (5.190) следует использовать функцию $\varphi_D(\mathbf{u})$, рассчитанную на фокусировку сходящегося сферического пучка с фокусом f в область D при $z = f$.

5.8.3. Спектральные фокусаторы в различные фокальные области

Рассмотренный в п. 5.8.2. метод расчета позволяет разделить и сфокусировать волны (5.188) в три области одинаковой формы, расположенные в одной заданной плоскости или в трех различных плоскостях вдоль оптической оси. Рассмотрим

расчет спектральных фокусаторов, позволяющих при изменении длины волны по закону (5.188) изменять конфигурацию области фокусировки.

Длины волн в (5.188) не являются независимыми между собой. Поэтому в общем случае метод расчета (5.190)–(5.194) может быть модифицирован для случая фокусировки двух различных длин волн из (5.188) в две различные области D_0 и D_1. Обозначим $\varphi(\mathbf{u})$, $\varphi_1(\mathbf{u})$ — фазовые функции фокусаторов для фокусировки пучков с длинами волн λ_0, λ_{+1} в области D_0 и D_1. Переопределим функцию $\tilde{\varphi}(\mathbf{u})$ в общем выражении для микрорельефа спектрального ДОЭ (5.190) в виде

$$\tilde{\varphi}(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(-\frac{N+p}{N}\varphi(\mathbf{u}) + \varphi_1(\mathbf{u})\right). \tag{5.208}$$

При $\tilde{\varphi}(\mathbf{u})$ (5.208), спектральный ДОЭ для пучков с длинами волн λ_0 и λ_{+1} формирует в 0-м и +1-м порядках дифракции пучки с фазовыми функциями

$$\Phi_0(\mathbf{u};\lambda_0) = \varphi(\mathbf{u}), \quad \Phi_{+1}(\mathbf{u};\lambda_{+1}) = \varphi_1(\mathbf{u}), \tag{5.209}$$

что обеспечивает фокусировку в области D_0 и D_1 соответственно. При длине волны λ_{-1} ДОЭ (5.190), (5.193), (5.208) формирует в основном 1-м порядке пучок с фазовой функцией

$$\Phi_{-1}(\mathbf{u};\lambda_{-1}) = 2\varphi(\mathbf{u}) - \varphi_1(\mathbf{u}). \tag{5.210}$$

В случае произвольных функций $\varphi(\mathbf{u})$ и $\varphi_1(\mathbf{u})$ структура области, формируемой при $\lambda = \lambda_{-1}$, не определена. Однако, при некотором выборе функций $\varphi(\mathbf{u})$ и $\varphi_1(\mathbf{u})$, возможна фокусировка длины волны λ_{-1} в область, являющуюся суперпозицией областей D_0 и D_1. Предположим, что спектральный фокусатор рассчитывается как добавка к тонкой линзе с фокусом f. При этом функции $\varphi_0(\mathbf{u})$, $\varphi_1(\mathbf{u})$ являются фазовыми функциями фокусаторов сходящихся сферических пучков с длинами волн λ_0, λ_{+1} в области D_0 и D_1 при $z = f$. Напомним, что при умножении фазы фокусатора сходящегося сферического пучка на константу p происходит p-кратное масштабирование области фокусировки.

Приведем три примера выбора функции $\varphi(\mathbf{u})$, $\varphi_1(\mathbf{u})$, позволяющих работать с тремя длинами волн λ_0, λ_{+1}, λ_{-1}. Пусть $\varphi(\mathbf{u})$ — фазовая функция фокусатора, а $\varphi_1(\mathbf{u})$ — фазовая функция N-порядковой дифракционной решетки. Тогда при $\lambda = \lambda_{-1}$ фазовая функция $2\varphi(\mathbf{u})$ в (5.210) реализует фокусировку в область $2 \times D$, а функция $\varphi_1(\mathbf{u})$ выполняет мультипликацию области $2 \times D$ по N порядкам. В этом случае спектральный фокусатор, как дополнение к тонкой линзе, при $\lambda = \lambda_0$ фокусирует в область D, при $\lambda = \lambda_{+1}$ — в набор из N точек, а при $\lambda = \lambda_{-1}$ — в набор из N областей с размером $2D$.

Пусть $\varphi(\mathbf{u})$ — фазовая функция фокусатора, а $\varphi_1(\mathbf{u})$ — фазовая функция дифракционной N-фокусной линзы, рассмотренной в разделе 5.6. Тогда при $\lambda = \lambda_{-1}$ фазовая функция $2\varphi(\mathbf{u})$ в (5.210) реализует фокусировку в область $2 \times D$, а функция $\varphi_1(\mathbf{u})$ выполняет мультипликацию области $2 \times D$ по N фокальным плоскостям. В этом случае спектральный фокусатор при $\lambda = \lambda_0$ фокусирует в область D, при $\lambda = \lambda_{+1}$ — в набор из N точек на оптической оси, а при $\lambda = \lambda_{-1}$ — в набор из N масштабированных областей D в различных фокальных плоскостях.

Пусть $\varphi(\mathbf{u})$, $\varphi_1(\mathbf{u})$ — фазовые функций фокусаторов пучков прямоугольного сечения в отрезки $|x| \leqslant d$, $y = 0$ и $|y| \leqslant d_1$, $x = 0$. Тогда при $\lambda = \lambda_{-1}$ фазовая функция (5.210) соответствует фокусировке в прямоугольник с размерами $(4d \times 2d_1)$ [87–89]. При указанном выборе функций $\varphi(\mathbf{u})$, $\varphi_1(\mathbf{u})$ спектральный ДОЭ при $\lambda = \lambda_0$ реализует фокусировку в отрезок оси Ox, при $\lambda = \lambda_{+1}$ — в отрезок оси Oy, а при $\lambda = \lambda_{-1}$ — в прямоугольник.

5.8.4. Расчет квантованных спектральных ДОЭ

Технология фотолитографии, традиционно используемая для изготовления ДОЭ, предусматривает квантование дифракционного микрорельефа по M уровням. Микрорельеф спектрального фокусатора (5.190) содержит компонент, соответствующий функции $\varphi(\mathbf{u})$ и N-уровневую ступенчатую функцию $G(\tilde{\varphi}(\mathbf{u}))$. При квантованной по L уровням функции $\varphi(\mathbf{u})$ микрорельеф (5.174) становится квантованным по $N_q = L \cdot N$ уровням.

Операция квантования приводит к снижению эффективности фокусировки и появлению дополнительных дифракционных порядков, снижающих качество формируемых изображений. Для компенсации погрешностей квантования на этапе расчета спектрального фокусатора, функцию $\varphi(\mathbf{u})$ в (5.190) целесообразно рассчитывать на основе итерационных алгоритмов расчета квантованных фазовых функций, рассмотренных в разделе 2.8.10. Напомним, что рассмотренные в 2.8.10 алгоритмы основаны на градиентной оптимизации непрерывной функции (2.316), являющейся отрезком ряда разложения по дифракционным порядкам квантованной комплексной функции пропускания.

Спектральные фокусаторы (5.190), (5.208), предназначенные для фокусировки двух длин волн λ_0, λ_{+1} в две различные области D_0 и D_1, будем называть двухволновыми ДОЭ. Отметим, что для двухволновых ДОЭ тип функций $\varphi(\mathbf{u})$, $\varphi_1(\mathbf{u})$ не играет роли. Действительно, предположение о геометрооптическом типе функций $\varphi(\mathbf{u})$, $\varphi_1(\mathbf{u})$ в (5.190), (5.208) используется только при описании структуры области, формируемой при $\lambda = \lambda_{-1}$.

Расчет двухволновых ДОЭ можно провести с полным учетом влияния паразитных порядков квантования. Согласно (5.179), расчет квантованных двухволновых ДОЭ с числом уровней квантования $N_q = M \cdot N$ может быть проведен через квантованные по M и N уровням фазовые функции $F_M(\varphi(\mathbf{u}))$ и $F_N(\varphi_1(\mathbf{u}))$. При этом высота рельефа в каждой точке u апертуры двухволнового ДОЭ определяется по формуле (5.179), где числа (i,j) соответствуют индексам значений квантованных функций $F_M(\varphi(\mathbf{u}))$, $F_N(\varphi_1(\mathbf{u}))$ в данной точке.

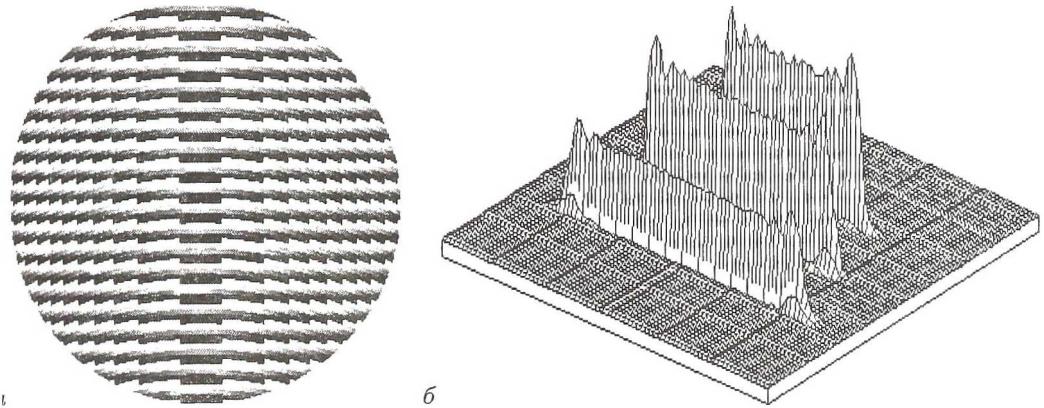
Число уровней квантования $N_q = M \cdot N$ микрорельефа можно уменьшить. Пусть d_1, d_2 наибольшие делители чисел N, M; такие, что $N = p_1 d_1$, $p_1 > 1$, $M = p_2 d_2$, $p_2 > 1$. В этом случае, из квантованных функций $F_{p_2}(\varphi(\mathbf{u}))$ и $F_{p_1}(\varphi_1(\mathbf{u}))$, рассчитанных из условия фокусировки пучков с длинами волн λ_0, λ_{+1} в области D_0 и D_1, можно построить двухволновый ДОЭ с числом уровней рельефа $N_q = p_1 \cdot p_2$ по следующему правилу:

$$h(i,j) = \frac{\lambda_0}{2\pi(n_0 - 1)} \left[\frac{2\pi}{p_2} i + 2\pi \bmod_N (\alpha_1 (d_1 j - d_2 i)) \right], \qquad (5.211)$$
$$i = \overline{0, p_1 - 1}, \quad j = \overline{0, p_2 - 1}.$$

Таким образом, расчет квантованных двухволновых ДОЭ сводится к двум независимым задачам, для решения которых разработаны специальные итерационные алгоритмы, рассмотренные в разделе 2.8.10.

Для оценки работоспособности предложенного метода расчета приведем ряд примеров расчета спектральных фокусаторов.

Пример 5.14. Рассчитывается спектральный фокусатор (5.190)–(5.194) как дополнение к тонкой линзе с фокусом f для разделения и фокусировки спектральных компонентов (5.188) в три отрезка в плоскости $z = f$. В качестве функции $\varphi(\mathbf{u})$ в (5.174)–(5.178) использовалась фазовая функция фокусатора с дифракционными поправками (5.21), (5.70) для фокусировки сходящегося сферического пучка



а *б*

Рис. 5.63. Полутоновой рельеф спектрального фокусатора для разделения и фокусировки длин волн $\lambda_0 = 0{,}525$ мкм, $\lambda_{+1} = 0{,}42$ мкм, $\lambda_{-1} = 0{,}7$ мкм в три отрезка (*а*); распределение интенсивности в плоскости фокусировки (*б*)

с фокусом f в отрезок с постоянной интенсивностью:

$$\varphi(u) = \mathrm{mod}_{2\pi}\left(\frac{3\pi d}{\lambda_0 f R^3}\left(R^2\frac{u^2}{2} - \frac{u^4}{12}\right)\right), \tag{5.212}$$

где R — радиус апертуры ДОЭ, d — длина отрезка фокусировки.

Расчет фокусатора проводился для следующих параметров: $\lambda_0 = 0{,}525$ мкм, $\lambda_{+1} = 4\lambda_0/5 = 0{,}42$ мкм, $\lambda_{-1} = 4\lambda_0/3 = 0{,}7$ мкм ($N = 4$, $p = 1$ в (5.188)), длина отрезков фокусировки $d = 30\Delta(\lambda)$, ($\Delta(\lambda) = \lambda f/R$), радиус апертуры ДОЭ $R = 2{,}5$ мм, фокус линзы $f = 500$ мм, параметры призмы и линзы в (5.194), $\mathbf{x}_0 = (0{,}10\Delta(\lambda_0))$, $f_l = Nf = 2000$ мм. Полутоновое изображение рассчитанного рельефа спектрального фокусатора приведено на рис. 5.63*а*. Распределение интенсивности, формируемое спектральным ДОЭ (5.190)–(5.194), (5.212) для освещающего пучка, состоящего из трех взаимно некогерентных плоских пучков с указанными выше длинами волн приведено на рис. 5.63*б*. Рисунок 5.63*б* показывает высокое качество фокусировки в три отрезка. Различные длины и интенсивности отрезков объясняются различным размером дифракционного пятна $\Delta(\lambda)$ для различных длин волн. Левый, центральный и правый отрезки на рис. 5.63*б* соответствуют компонентам $\lambda_{-1} = 0{,}7$ мкм, $\lambda_0 = 0{,}525$ мкм и $\lambda_{+1} = 0{,}42$ мкм, соответственно.

Пример 5.15. Рассчитан спектральный фокусатор (5.190), (5.208), позволяющий при изменении длины волны по закону (5.188), изменять конфигурацию фокальной области. Фокусатор был рассчитан как дополнение к линзе для фокусировки компонентов $\lambda_0 = 0{,}525$ мкм в отрезок длиной $20\Delta(\lambda_0)$, $\lambda_{+1} = 0{,}42$ мкм — в четыре точки и $\lambda_{-1} = 0{,}7$ мкм — в четыре отрезка с длиной $40\Delta(\lambda_{-1})$. В качестве функции $\varphi(\mathbf{u})$ в (5.190), (5.208) использовалась фазовая функция (5.212) фокусатора в отрезок. Для компенсации влияния паразитных дифракционных порядков при фокусировке компонентов λ_{+1}, λ_{-1}, в фазу (5.212) была дополнительно введена фазовая функция призмы $\varphi_{pr}(v) = (2\pi/\lambda_0 f)\, y_0 v$ при $y_0 = 2\Delta(\lambda_0)$. В качестве функции $\varphi_1(\mathbf{u})$ в (5.208) использовалась фазовая функция 4-порядковой дифракционной решетки, принимающая в интервалах периода $[(l-1)\cdot d/4, l\cdot d/4]$, $l = \overline{1,4}$, значения $(0, \pi, \pi/2, 3\pi/2)$. Данная решетка фокусирует 81% энергии освещающего пучка в порядках $-2, -1, +1, +2$.

Полутоновое изображение рассчитанного рельефа спектрального фокусатора приведено на рис. 5.64*а*. Распределения интенсивности, формируемое спектральным

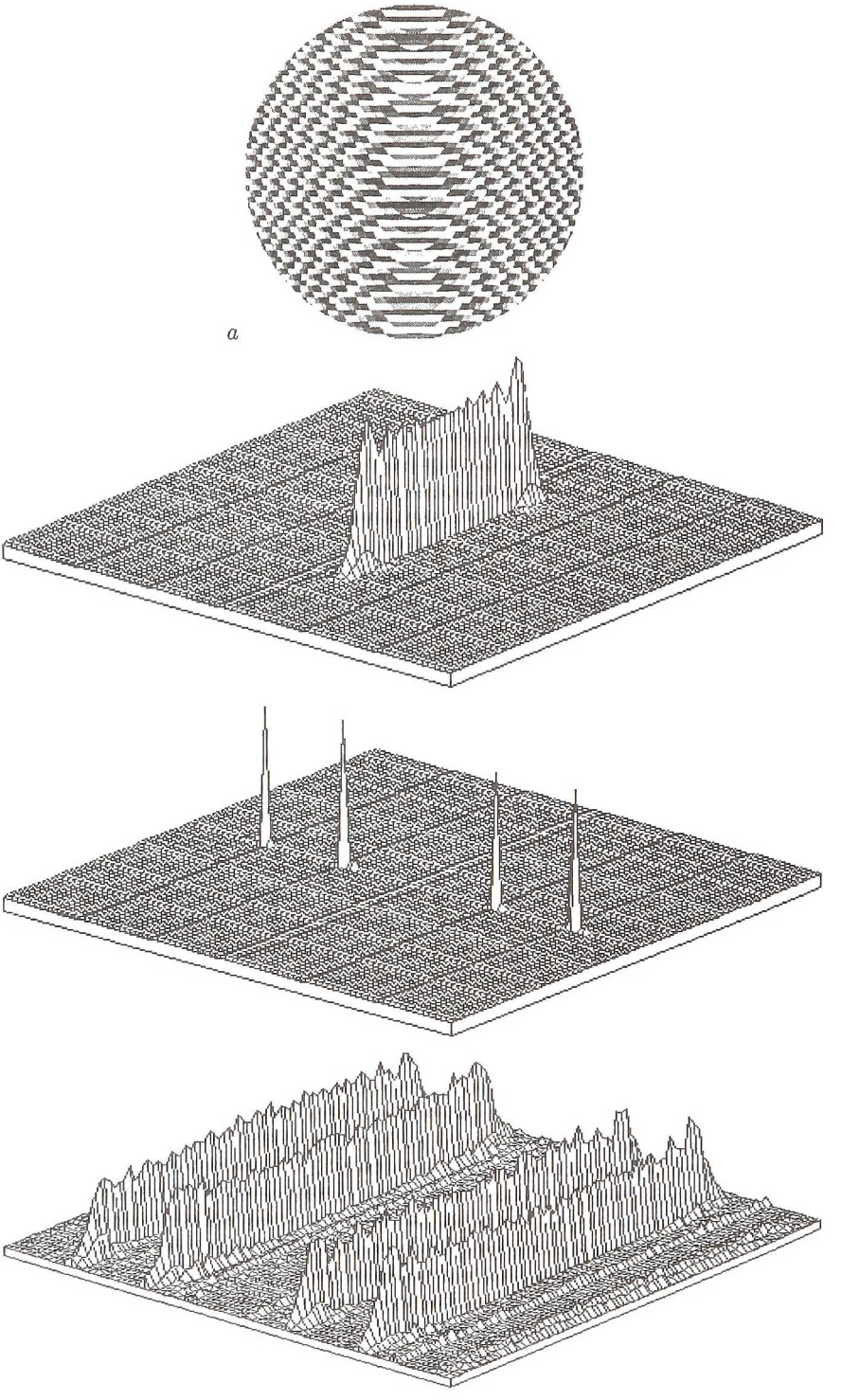


Рис. 5.64. Полутоновой рельеф спектрального фокусатора для фокусировки компонентов с длинами волн (5.188) в отрезок, 4 точки и 4 отрезка (*а*); распределения интенсивности в плоскости фокусировки для $\lambda_0 = 0{,}525$ мкм (*б*), $\lambda_{+1} = 0{,}42$ мкм (*в*), $\lambda_{-1} = 0{,}7$ мкм (*г*)

фокусатором для плоских пучков с длинами волн $\lambda_0 = 0{,}525$ мкм, $\lambda_{+1} = 0{,}42$ мкм и $\lambda_{-1} = 0{,}7$ мкм приведены на рис. 5.64*б,в,г* и ясно показывают структуру излучения, сфокусированного в отрезок, четыре точки и четыре отрезка.

Пример 5.16. Рассчитан спектральный фокусатор (5.190), (5.208) с квадратной апертурой 2,5 × 2,5 мм для фокусировки компонента $\lambda_0 = 0{,}525$ мкм в квадрат $20\Delta(\lambda_0) \times 20\Delta(\lambda 0)$ и компонент $\lambda_{+1} = 0{,}42$ мкм, $\lambda_{-1} = 0{,}7$ мкм — в отрезки осей Ox и Oy с длинами $40\Delta(\lambda_{+1})$ и $40\Delta(\lambda_{-1})$. В качестве функций $\varphi(\mathbf{u})$ и $\varphi_1(\mathbf{u})$ в (5.190), (5.208) использовались фазовые функции фокусаторов сходящегося сферического пучка в квадрат со стороной $K\Delta(\lambda_0)$ и в отрезок оси Ox длиной $2K\Delta(\lambda_{+1})$:

$$\varphi(\mathbf{u}) = \mathrm{mod}_{2\pi}\left(\frac{\pi K}{a^2}\left(u^2 + v^2\right)\right), \tag{5.213}$$

$$\varphi_1(\mathbf{u}) = \frac{2\pi K}{a^2}u^2, \tag{5.214}$$

где a — размер стороны апертуры фокусатора.

ДОЭ (5.190), (5.208), (5.213), (5.214) при длине волны λ_{-1} формирует в основном -1-м порядке пучок с фазовой функцией

$$\Phi_{-1}(\mathbf{u}; \lambda_{-1}) = 2\varphi(\mathbf{u}) - \varphi_1(\mathbf{u}) = \frac{2\pi K}{a^2}v^2. \tag{5.215}$$

Фаза (5.215) соответствует фазовой функции фокусатора сходящегося сферического пучка в отрезок оси Oy длиной $2K\Delta(\lambda_{-1})$.

Полутоновое изображение рельефа спектрального фокусатора приведено на рис. 5.65*а*. Распределения интенсивности приведены на рис. 5.65*б,в,г* и ясно показывают структуру излучения, сфокусированного в квадрат и отрезки.

Пример 5.17. Рассчитаны две квантованные спектральные решетки для формирования четырех порядков $-2, -1, +1, +2$ при длине волны λ_0 и трех порядков $-1, 0, +1$ при длинах волн $\lambda_{+1} = 3\lambda_0/4$ и $\lambda_{+1} = 9\lambda_0/4$, соответственно. Расчет квантованных спектральных решеток (период d) проводился по формулам (5.179), (5.184), (5.185) на основе квантованных фазовых функций 4-х и 3-порядковой решеток, принимающих в интервалах периода $[(l-1)\cdot d/4, l\cdot d/4]$, $l = \overline{1,4}$ значения $(0, \pi, \pi/2, 3\pi/2)$ и $(0, 0, 2\pi/3, 2\pi/3)$, соответственно. Напомним, что значения интенсивностей порядков I_j для решетки с фазой $\varphi(u)$ определяются как квадраты модулей коэффициентов Фурье функции $\exp(i\varphi(u))$. Для 4-порядковой решетки $I_{-2} = I_{-1} = I_1 = I_2 = 0{,}205$, а для 3-порядковой решетки $I_0 = 0{,}304$, $I_{-1} = I_1 = 0{,}25$. Следовательно, приведенные решетки концентрируют более 80% энергии в требуемых порядках $-2, -1, +1, +2$ и $-1, 0, +1$.

При $\lambda_{+1} = 3\lambda_0/4$ из (5.184), (5.185) получим $\alpha_1 = 1$ и $M_0(i,j) = \mathrm{mod}_3(j-i)$. При этом рельеф спектральной решетки для длин волн λ_0 и $\lambda_{+1} = 3\lambda_0/4$ принимает вид

$$h(u) = \frac{\lambda_0}{2\pi(n_0 - 1)}\begin{cases} 0, & u \in [0, p/4), \\ 3\pi, & u \in [p/4, p/2), \\ 0{,}5\pi, & u \in [p/2, 3p/4), \\ 3{,}5\pi, & u \in [3p/4, p). \end{cases} \tag{5.216}$$

При $\lambda_{+1} = 9\lambda_0/4$ из (5.184), (5.185) получим $\alpha_1 = -2$ и $M_0(i,j) = \mathrm{mod}_9(-2(j-i))$.

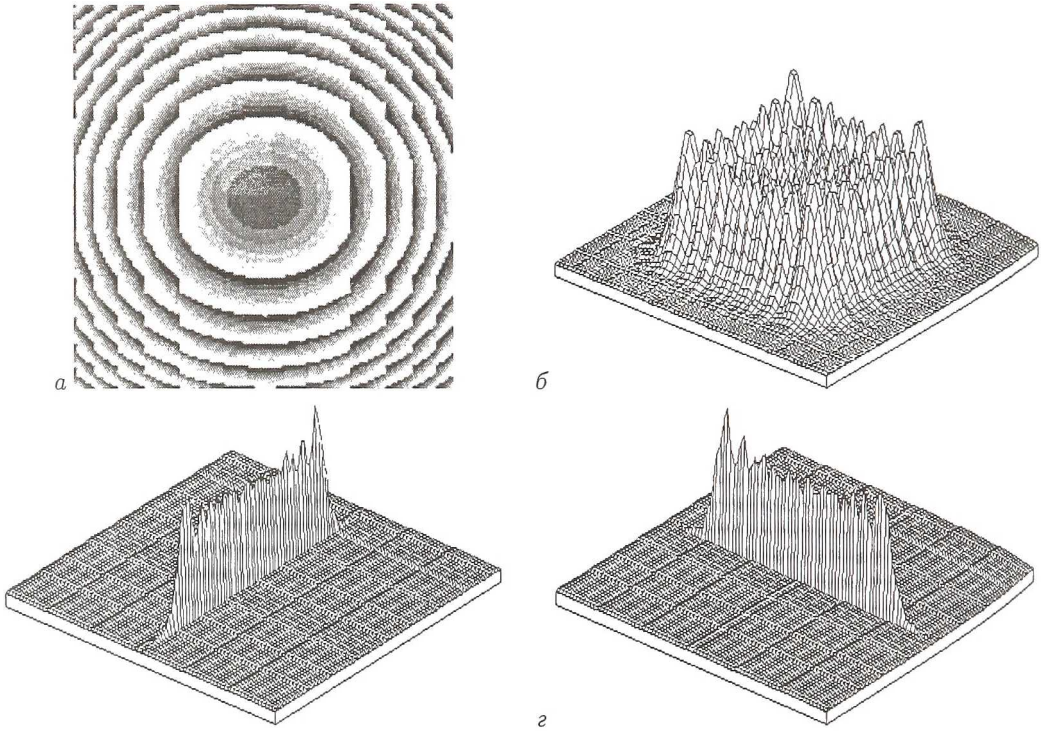


Рис. 5.65. Полутоновой рельеф спектрального фокусатора для фокусировки компонентов с длинами волн (5.188) в квадрат, отрезок оси Ox и отрезок оси Oy (*а*); распределения интенсивности в плоскости фокусировки для $\lambda_0 = 0{,}525$ мкм (*б*), $\lambda_{+1} = 0{,}42$ мкм (*в*), $\lambda_{-1} = 0{,}7$ мкм (*г*)

В результате рельеф решетки для длин волн λ_0 и $\lambda_{+1} = 9\lambda_0/4$ принимает вид

$$h(u) = \frac{\lambda_0}{2\pi(n_0 - 1)} \begin{cases} 0, & u \in [0, p/4), \\ 9\pi, & u \in [p/4, p/2), \\ 10{,}5\pi, & u \in [p/2, 3p/4), \\ 1{,}5\pi, & u \in [3p/4, p). \end{cases} \tag{5.217}$$

Интенсивности дифракционных порядков, формируемых решетками (5.216) и (5.217), при освещении решеток плоскими пучками с длинами волн λ_0, $\lambda_{+1} = 3\lambda_0/4$ и λ_0, $\lambda_{+1} = 9\lambda_0/4$, совпадают и показаны на рис. 5.66.

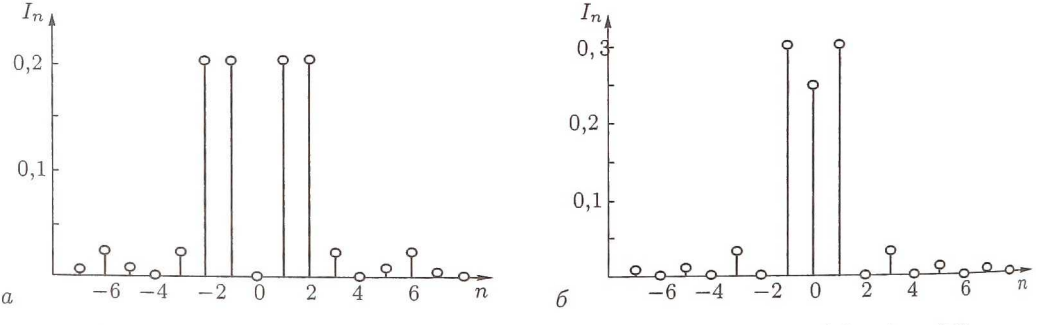


Рис. 5.66. Интенсивности дифракционных порядков для λ_0 (*а*) и λ_{+1} (*б*)

5.9. Заключение

В главе приведены расчет фокусаторов и детальное исследование их работы при различных параметрах с учетом погрешностей квантования и дискретизации фазы, присущих технологии фотолитографии. Методы расчеты фокусаторов в кривые, основанные на построении гладких лучевых соответствий, дополнены дифракционными методами на основе нелинейного преобразования фазы. Метод нелинейного преобразования фазы по закону многопорядковой дифракционной решетки позволил придать фокусатору многофокусные свойства без негативных эффектов сегментации апертуры. Преобразование фазы фокусатора по закону цветоделительной решетки позволяет получить новые элементы, выполняющие одновременно разделение и фокусировку пучков с различными длинами волн и позволяющие изменять конфигурацию области фокусировки для различных длин волн.

Несмотря на двадцатилетнюю историю фокусаторов в кривую, они не получили широкого распространения. Совершенствование технологий производства ДОЭ и дальнейшее развитие методов расчета оставляют фокусаторам множество потенциальных применений. Например, изготовление дифракционного рельефа на алмазных пленках позволит использовать фокусаторы для лазеров с мощностью 5–10 Квт, а методы преобразования рельефа зон фокусатора по закону однопорядковых дифракционных решеток, рассчитываемых в рамках строгой электромагнитной теории, позволят достичь энергетической эффективности фокусировки близкой к 100%.

Литература к главе 5

1. Голуб М. А., Карпеев С. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Письма в ЖТФ.* **7**(10) 618 (1981)
2. Голуб М. А., Дегтярева В. П., Климов А. Н., Попов В. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Письма в ЖТФ.* **8**(13) 449 (1982).
3. Данилов В. А., Попов В. В., Прохоров А. М., Сагателян Д. М., Сисакян Е. В., Сисакян И. Н., Сойфер В. А. *Оптические элементы, фокусирующие когерентное излучение в произвольную фокальную линию* (М.: ФИАН СССР., 1983).
4. Гончарский А. В., Данилов В. А., Попов В. В., Сисакян И. Н., Сойфер В. А., Степанов В. В. *Квантовая электроника* (Москва). **13**(3) 660 (1986)
5. Васин А. Г., Голуб М. А., Данилов В. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Уваров Г. В. *Расчет и исследование когерентного волнового поля в фокальной области радиально-симметричных оптических элементов* (М.: ФИАН, 1983)
6. Гончарский А. В., Данилов В. А., Попов В. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А., Степанов В. В. *Доклады АН СССР* **273**(3) 605 (1983)
7. Данилов В. А., Попов В. В., Прохоров А. М., Сагателян Д. М., Сисакян И. Н., Сойфер В. А. *Письма в ЖТФ.* **8**(13) 810 (1982)
8. Гончарский А. В., Данилов В. А., Попов В. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А., Степанов В. В. *Квантовая электроника* **11**(1) 166 (1984)
9. Гончарский А. В., Степанов В. В. *Доклады АН СССР* **279**(4) 788 (1984)
10. Гончарский А. В., Степанов В. В. *Журнал вычислительной математики и математической физики.* **26**(1) 80 (1986)
11. Гончарский А. В. *Компьютерная оптика* **1** 19 (1987)
12. Гончарский А. В., Попов В. В., Степанов В. В. *Введение в компьютерную оптику* (М.: Изд-во МГУ, 1991)
13. Данилов В. А., Кинбер Б. Е., Шилов А. Е. *Компьютерная оптика* **1** 40 (1987)
14. Curcio M. *Proceedings SPIE* **306** 105 (1981)
15. Булатов Е. Д., Гридин С. А., Даниленко А. А. *Компьютерная оптика* **1** 167 (1987)

16. Boiko Yu. B., Solovjev V. S., Calixto S., Lougnot D. J. *Applied Optics* **33** 787 (1994)
17. Соловьев В. С. *Компьютерная оптика* **10-11** 145 (1992)
18. Волков А. В., Волотовский С. Г., Гранчак В. М., Казанский Н. Л., Моисеев О. Ю., Сойфер В. А., Соловьев В. С., Якуненкова Д. М. *Журнал технической физики* **65**(9) 181 (1995)
19. Сисакян И. Н., Сойфер В. А. *Компьютерная оптика* **1** 5 (1987)
20. Голуб М. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А. *Автометрия* **1** 70 (1988)
21. Голуб М. А., Казанский Н. Л., Сойфер В. А. *Научное приборостроение* **1** 9 (1993)
22. Голуб М. А., Казанский Н. Л., Шинкарев М. В. *Компьютерная оптика* **5** 43 (1989)
23. Казанский Н. Л. *Компьютерная оптика* **3** 22 (1988)
24. Казанский Н. Л. *Компьютерная оптика* **10-11** 128 (1992)
25. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Сойфер В. А., Харитонов С. И. *Оптика и спектроскопия* **73**(1) 191 (1992)
26. Досколович Л. Л., Казанский Н. Л., Сойфер В. А. *Автометрия* **1** 58 (1993)
27. Досколович Л. Л., Казанский Н. Л., Сойфер В. А., Харитонов С. И. *Научное приборостроение* **3**(1) 24 (1993)
28. Golub M. A., Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Soifer V. A. *Journal of Modern Optics* **39**(6) 1245 (1992)
29. Soifer V. A., Doskolovich L. L., Golub M. A., Kazanskiy N. L., Kharitonov S. I., Perlo P. *Proceedings SPIE* **1992** 226 (1993)
30. Soifer V. A., Doskolovich L. L., Kazanskiy N. L. *Optical Engineering* **33**(11) 3610 (1994)
31. Dammann H. *Appl. Opt.* **17**(15) 2273 (1978)
32. Dammann H. *Gratings. Optic* **53** 409 (1979)
33. Farn M. W., Stern M. B. Opt.Lett. 18 1214 (1993)
34. Борн М., Вольф Э. *Основы оптики* (М.: Наука, 1973)
35. Li Y., Wolf E. *Optics Communications* **39**(4) 211 (1981)
36. Erkkila J. H. *Optics Communications* **43**(5) 313 (1982)
37. Givens M. P. *Optics Communications* **41**(3) 145 (1982)
38. Li Y., Wolf E. *Journal of Optical Society of America A* **1**(8) 801 (1984)
39. Дементьев А. С., Домаркене Д. П. *Оптика и спектроскопия* **56**(5) 869 (1984)
40. Кинбер Б. Е., Новоселов С. В. *Радиотехника и электроника* **30**(8) 1469 (1985)
41. Wenzel R. G. *Journal Optical Society of America A* **4**(2) 340 (1987)
42. Carter W. H. *Applied Optics* **21**(11) 1989 (1982)
43. Li Y., Wolf E. *Optics Communications* **42**(3) 151 (1982)
44. Федоров В. Б., Митяков В. Г. *Оптика и спектроскопия* **56**(5) 878 (1984)
45. Tanaka K., Saga N., Hauchi K. *Applied Optics* **24**(8) 1098 (1985)
46. Herman R. M., Pardo J., Wiggins T. A. *Applied Optics* **24**(9) 1346 (1985)
47. Campbell Ch. *Optical Engineering* **26**(3) 270 (1987)
48. Stamnes J. J. *Journal Optical Society of America* **71**(1) 15 (1981)
49. Дементьев А. С., Домаркене Д. П. *Литовский физический сборник* **24**(3) 59 (1984)
50. Bottema M. *Journal of Optical Society of America* **59**(12) 1632 (1969)
51. Mozer J., Janicijevic Lj., Bejtulahu R. J. *Optics (Paris)* **12**(5) 309 (1981)
52. Janicijevic Lj. J. *Optics (Paris)* **13**(4) 199 (1982)
53. Janicijevic Lj. J. *Optics (Paris)* **18**(1) 23 (1987)
54. Байбулатов Ф. Х., Минин И. В., Минин О. В. *Радиотехника и электроника* **30** 1681 (1985)
55. Stamnes J. J., Gravelsaeter T., Bentsen O. *Acoustical Imaging* **10** 587 (1980)
56. Stamnes J. J., Gravelsaeter T. *Acoustical Imaging* **10** 575 (1980)
57. Пальчикова И. Г. *Математическое моделирование процесса дифракции световых волн на фазовых зонных пластинках* (Новосибирск, 1989)
58. Коронкевич В. П., Пальчикова И. Г. *Автометрия* **1** 85 (1992)
59. Бобров С. Т., Грейсух Г. И., Туркевич Ю. Г. *Оптика дифракционных элементов и систем* (Л.:Машиностроение, 1986)
60. Мещеряков Н. А., Томилина Е. А. *Рук. деп. в ВИНИТИ* 4784 (1987)
61. Dammann H. *Optik* **31**(1) 95 (1970)

62. Коронкевич В. П., Ленкова Г. А., Михальцова И. А., Пальчикова И. Г., Полещук А. Г., Седухин А. Г., Чурсин Е. Г., Юрлов Ю. И. *Автометрия* **1** 4 (1985)
63. Fedotowsky A., Lehovec K. *Applied Optics* **13**(12) 2919 (1974)
64. Doskolovich L. L., Khonina S. N., Kotlyar V. V., Nikolsky I. V., Soifer V. A., Uspleniev G. V. *Optical and Quantum Electronics* **25** 801 (1993)
65. Belanger P. A., Rioux M. *Journ. Canadien de Physique* **54** 1774 (1976)
66. Belanger P. A., Rioux M. *Applied Optics* **17**(7) 1080 (1978)
67. Farn M. W., Goodman J. W. *Proceedings SPIE* **1211** 1256 (1990)
68. Голуб М. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Автометрия* **6** 8 (1987)
69. Golub M. A., Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Orlova N. G., Sisakian I. N., Soifer V. A. *Proceedings SPIE* **1500** 194 (1991)
70. Данилов В. А., Дубовский П. Е., Ковш И. Б., Политов М. В., Сисакян И. Н. *Научное приборостроение* **3**(1) 64 (1993)
71. Голуб М. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Оптика и спектроскопия* **67**(6) 1387 (1989)
72. Голуб М. А., Досколович Л. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Оптика и спектроскопия* **6** 1069 (1991)
73. Soifer V. A., Doskolovich L. L., Golub M. A., Kazanskiy N. L. *Proceedings SPIE* **1718** 33 (1992)
74. Golub M. A., Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Sisakian I. N., Soifer V. A. *Proceedings SPIE* **1500** 211 (1991)
75. Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Usplenjev G. V. *Optics and Lasers in Engineering* **15**(5) 311 (1991)
76. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Харитонов С. И. *Компьютерная оптика* **12** 3 (1992)
77. Doskolovich L. L., Golub M. A., Kazanskiy N. L., Soifer V. A., Usplenjev G. V. *Proceedings SPIE* **1983** 647 (1993)
78. Казанский Н. Л. *Компьютерная оптика* **1** 90 (1987)
79. Пальчикова И. Г. *Компьютерная оптика* **6** 9 (1989)
80. Doskolovich L. L, Kazanskiy N. L., Soifer V. A., Tzaregorodtzev A. Ye. *Optik* **101**(2) 37 (1995)
81. Khonina S. N., Kotlyar V. V., Soifer V. A. *Journal of Modern Optics* **40** 761 (1993)
82. Michaltsova I. A., Nalivaiko V. I., Soldatenkov I. S. *Optik* **67**(3) 267 (1984)
83. Kolodziejczyk A., Bara S., Jaroszewicz Z., Sypek M. *Journal of Modern Optics* **37**(8) 1283 (1990)
84. Brenden B. B., Russel J. T. *Applied Optics* **23**(19) 3250 (1984)
85. Tremblay R., D'Astons Y., Roy G., Blanshard M. *Optics Communications* **28**(2) 193 (1979)
86. Rioux M., Tremblay R., Belanger P. A. *Applied Optics* **17**(10) 1532 (1978)
87. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **7** 42 (1990)
88. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **10-11** 110 (1992)
89. Голуб М. А., Досколович Л. Л., Казанский Н. Л., Сисакян И. Н., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **10-11** 100 (1992)
90. Goodman J. W., Silvestri A. M. *IBM Journ. Res. Develop* **14**(9) 478 (1969)
91. Dallas W. J., Lohmann A. W. *Optics Communications* **15**(2) 78 (1972)
92. Dallas J. M. *Applied Optics* **10** 674 (1971)

ГЛАВА 6
СЕЛЕКЦИЯ МОД ЛАЗЕРНОГО ИЗЛУЧЕНИЯ

6.1. Моды лазерного излучения

Концепция волноводных мод, или просто мод, имеет большое значение в лазерной оптике, в частности, для описания когерентных световых пучков в волноводах с квантованной постоянной распространения [1]. Назовем модой пучок света, обладающий свойством *самовоспроизводимости* при распространении в соответствующей волноводной среде. Под *самовоспроизводимостью* будем понимать свойство сохранения поперечной структуры амплитудно-фазового распределения с точностью до масштаба декартовых или цилиндрических координат. Это позволит нам не ограничиваться рассмотрением только функций, являющихся точными решениями уравнения Гельмгольца для различных сред, но и описать в тех же терминах, например, поведение гауссовых пучков в свободном пространстве (гауссовых мод). Каждая мода характеризуется собственным коэффициентом затухания и фазовой задержкой, пропорциональной оптическому пути и константе распространения. Фазовая задержка непрерывно накапливается в ходе распространения моды в среде. Отметим, что в случае рассмотрения волноводов, для нас представляют интерес лишь моды, сохраняющие свою энергию при прохождении любого расстояния в волноводе с нулевым поглощением, так называемые *каналируемые* или *направляемые* моды [1]. Каналируемая мода воспроизводит свою структуру после прохождения оптического пути произвольной длины в волокне. Мода, распространяющаяся в резонаторе, воспроизводит себя каждый раз только после прохождения полного расстояния от первого зеркала до второго и назад [2, 3].

Данный раздел посвящен детальному рассмотрению концепции мод когерентного излучения, а также модовых свойств, существенных при компьютерном синтезе ДОЭ, согласованных с комплексными амплитудами модовых функций. Будет показано, что использование свойств мод лазерного излучения и методов компьютерной оптики позволяет решать как важнейшие фундаментальные физические задачи, так и прикладные, например задачи повышения пропускной способности линий оптической связи.

6.1.1. Модовые пучки в скалярном приближении

Начнем с математических формул, описывающих моды как собственные функции оператора распространения. Введем декартовы координаты $(x, y, z) = (\mathbf{x}, z)$ в среде распространения пучка. Двумерный вектор $\mathbf{x} = (x, y)$ представляет поперечные координаты; z — продольная координата вдоль оптической оси. Введем различные обозначения для поперечных координат в параллельных плоскостях $\mathbf{u} = (u, v)$, $\mathbf{x}' = (x', y')$, и т.д. Пусть рассматриваемые моды локализованы внутри области

$\mathbf{x} \in G$ в сечении пучка. Применяя скалярное представление светового поля и скалярную дифракционную теорию без учета поляризационных эффектов, опишем монохроматическое или квазимонохроматическое поле комплексной амплитудой $F(\mathbf{x}, z)$, соответствующей длине волны λ. Ниже мы рассмотрим вопрос о границах применимости скалярной дифракционной теории к описанию волноводных сред. Комплексная амплитуда моды $\psi_{\mathbf{p}}(\mathbf{x})$ с номером $\mathbf{p} = (p, l)$ рассматривается на поверхности волнового фронта, и $\mathbf{x}$ — координаты проекции точки волнового фронта на ближайшую плоскость, перпендикулярную направлению распространения моды. В случае линейной среды можно ввести линейный оператор распространения $\hat{\mathrm{P}}$, связывающий комплексные распределения F и F_1 на двух волновых фронтах, разделенных некоторым расстоянием:

$$\hat{\mathrm{P}} F(\mathbf{x}) = F_1(\mathbf{x}').$$

Моды конкретной линейной среды $\psi_{\mathbf{p}}(\mathbf{x})$ не меняют своей поперечной конфигурации при распространении в ней и, таким образом, удовлетворяют уравнению на собственные значения

$$\hat{\mathrm{P}}_{\mathbf{p}} \psi_{\mathbf{p}} = \gamma_{\mathbf{p}} \psi_{\mathbf{p}}, \tag{6.1}$$

где $\gamma_{\mathbf{p}}$ — комплексные собственные значения. Отметим, что оператор $\hat{\mathrm{P}}$ и его собственные значения $\gamma_{\mathbf{p}}$ зависят от расстояния между рассматриваемыми волновыми фронтами. Таким образом, моды $\psi_{\mathbf{p}}$ являются собственными функциями оператора распространения $\hat{\mathrm{P}}$. Как известно [4], если выполняются математические свойства нормировки и компактности

$$\hat{\mathrm{P}} \hat{\mathrm{P}}^* = \hat{\mathrm{P}}^* \hat{\mathrm{P}} \equiv \hat{\mathrm{Q}}, \tag{6.2}$$

где * — символ эрмитово-сопряженного оператора, оператор $\hat{\mathrm{P}}$ имеет полное и счетное множество ортогональных собственных функций. В дальнейшем будем предполагать, что свойство ортогональности выполняется с учетом некоторой весовой функции $\rho_0(\mathbf{x}) \geqslant 0$, $\mathbf{x} \in G$, зависящей от различного наклона плоскости волнового фронта мод. Естественно, можно выбрать ортонормированные моды $\psi_{\mathbf{p}}$, используя соответствующую нормировку

$$\iint\limits_G \psi_{\mathbf{p}}(\mathbf{x}) \psi^*_{\mathbf{p}'}(\mathbf{x}) \rho_0(\mathbf{x}) \, \mathrm{d}^2 x = \delta_{\mathbf{p}\mathbf{p}'}, \tag{6.3}$$

где * — символ комплексного сопряжения и

$$\delta_{\mathbf{p}\mathbf{p}'} = \delta_{pp'} \delta_{ll'} = \begin{cases} 1, & \mathbf{p} = \mathbf{p}', \\ 0, & \mathbf{p} \neq \mathbf{p}' \end{cases}$$

— символ Кронекера. Заметим, что ρ_0 меняет свою форму по координате $\mathbf{x}$. Мощности $\hat{\mathrm{P}}^{m_0}$ оператора $\hat{\mathrm{P}}$ соответствуют числу m_0 полных проходов пучка. Операторы $\hat{\mathrm{P}}^*$ и $\hat{\mathrm{P}}^{m_0}$ являются коммутативными с оператором $\hat{\mathrm{P}}$ и, таким образом, имеют те же собственные функции $\psi_{\mathbf{p}}$ и собственные значения, соответственно, связанные уравнениями

$$\hat{\mathrm{P}}^* \psi_{\mathbf{p}} = \gamma^*_{\mathbf{p}} \psi_{\mathbf{p}}; \quad \hat{\mathrm{P}}^{m_0} \psi_{\mathbf{p}} = (\gamma_{\mathbf{p}})^{m_0} \psi_{\mathbf{p}}. \tag{6.4}$$

Оператор $\hat{\mathrm{Q}}$ (6.2) имеет неотрицательные собственные значения

$$\lambda_{\mathbf{p}} = |\gamma_{\mathbf{p}}|^2, \quad \hat{\mathrm{Q}}\psi_{\mathbf{p}} = \lambda_{\mathbf{p}}\psi_{\mathbf{p}}, \tag{6.5}$$

и может быть назван энергетическим оператором. Значения $\lambda_{\mathbf{p}}$, (6.5), характеризуют затухания модовых мощностей за один полный проход, и значения $\arg\gamma_{\mathbf{p}}$ дают фазовую задержку за один проход. В пассивном волноводе $0 \leqslant \gamma_{\mathbf{p}} \leqslant 1$, а величина

$$\alpha_{\mathbf{p}} = 1 - \lambda_{\mathbf{p}} \tag{6.6}$$

характеризует потери мощности для $\psi_{\mathbf{p}}$ – моды (включая дифракционные потери) на один проход волновода. Моду, имеющую максимальное значение $\lambda_{\mathbf{p}}$ (и, таким образом, минимальные потери) называют *основной* или *принципиальной*. Обычно она обозначается нулевым индексом $\mathbf{0} = (0,0)$:

$$\lambda_{\mathbf{0}} = \max_{\mathbf{p}} \lambda_{\mathbf{p}}. \tag{6.7}$$

В соответствие с (6.4), затухание модовой мощности $\psi_{\mathbf{p}}$ для m_0 проходов равно $|\gamma_{\mathbf{p}}|^{2m_0}$, и фазовая задержка равна $m_0 \arg\gamma_{\mathbf{p}}$. Потери $\mathbf{p}$-й моды за m_0 проходов определяются по формуле

$$\alpha_{\mathbf{p}}^{m_0} = 1 - \lambda_{\mathbf{p}}^{m_0}$$

и упорядочены также, как потери за проход. Рассмотрим моду $\psi_{\mathbf{p}}$, имеющую более высокий порядок по сравнению с $\psi_{\mathbf{p}'}$. Мода $\psi_{\mathbf{p}}$ характеризуется более высокими дифракционными потерями, т.е. $\alpha_{\mathbf{p}} > \alpha_{\mathbf{p}'}$ или $\lambda_{\mathbf{p}} < \lambda_{\mathbf{p}'}$. Моды $\psi_{\mathbf{p}}$ и $\psi_{\mathbf{p}'}$ с $\lambda_{\mathbf{p}} = \lambda_{\mathbf{p}'}$ для $\mathbf{p} \neq \mathbf{p}'$ соответствуют вырожденным собственным значениям оператора $\hat{\mathrm{Q}}$ и могут быть упорядочены, например, по возрастанию постоянной распространения, пропорциональной $|\arg\gamma_{\mathbf{p}}|$. Будучи собственными колебаниями волновода, моды могут быть охарактеризованы следующими инвариантными и экстремальными свойствами, объясняющими широкое использование мод, а не других базисов для описания распространения когерентного света.

1. Моды — это *единственный* двумерный базис функций, сохраняющих взаимную ортогональность при распространении в собственной волноводной среде.
2. Одномодовый пучок характеризуется наименьшими потерями мощности по сравнению с любыми другими пучками, содержащими моды более высоких порядков.

Свойство мод воспроизводить свою поперечную структуру в процессе их распространения было сформулировано выше и может быть добавлено к этим свойствам. Определив понятие моды уравнением (6.1), можем представить полное поле $F(\mathrm{x}, z)$ в волноводной среде как суперпозицию мод:

$$F(\mathbf{x}, z) = \sum_{\mathbf{p}} F_{\mathbf{p}} \psi_{\mathbf{p}}(z)\, \psi_{\mathbf{p}}(\mathbf{x}), \quad \gamma_{\mathbf{p}}(0) = 1. \tag{6.8}$$

Будем считать, что коэффициенты $F_{\mathbf{p}}$ ортогонального разложения определяют *модовый состав* поля:

$$F_{\mathbf{p}} = \iint_G F(\mathbf{x}, 0)\psi_{\mathbf{p}}(\mathbf{x})\, \rho_0(\mathbf{x})\, \mathrm{d}^2\mathbf{x}. \tag{6.9}$$

Так как разложение Фурье удовлетворяет условию теоремы Парсеваля, можно записать полную энергию поля F, формируемую парциальными мощностями $|F_{\mathbf{p}}|^2$

отдельных мод:

$$\iint\limits_G |F(\mathbf{x},0)|^2 \rho_0(\mathbf{x})\, \mathrm{d}^2\mathbf{x} = \sum_{\mathbf{p}} |F_{\mathbf{p}}|^2. \tag{6.10}$$

Определим распределение мощности поля по модам как нормированную величину следующего вида:

$$\frac{|F_{\mathbf{p}}|^2}{\iint\limits_G |F(\mathbf{x},0)|^2 \rho_0(\mathbf{x})\, \mathrm{d}^2\mathbf{x}}.$$

Изменения коэффициентов $F_{\mathbf{p}}$ в зависимости от значения $\mathbf{p}$ определяют модовый состав поля и определяются состоянием возбужденной волноводной среды.

6.1.2. Возбуждение мод в оптических волокнах и резонаторах

Тип модовых функций определяется типом распределения показателя преломления в сечении волновода или оптоволокна (профилем оптоволокна) или конфигурацией и формой зеркал резонатора лазера. Возбуждение мод с определенными порядковыми номерами сильно зависит от различных факторов, включающих условия возбуждения, геометрию сердечника оптоволокна, длину волны излучения, тип активной среды и дифракционные потери в резонаторе. Классические уравнения

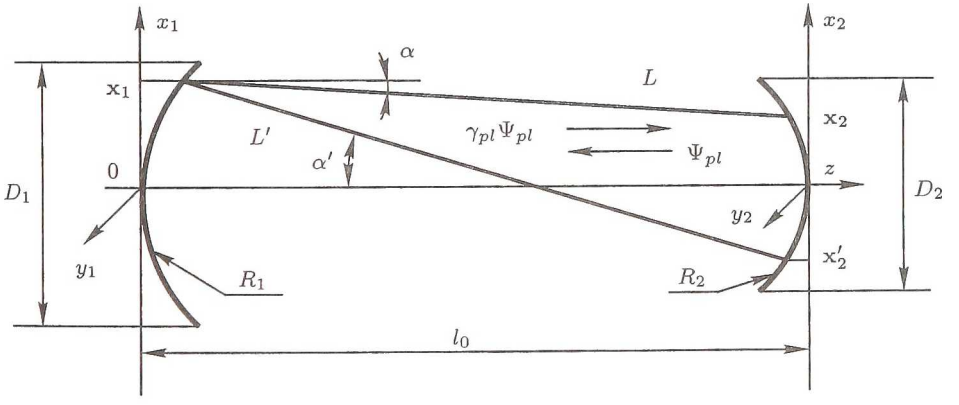


Рис. 6.1. Геометрия расчета мод открытого пассивного резонатора

для мод и условий их возбуждения будут приведены ниже. Если рассматривать моды пассивного резонатора, то оператор $\hat{\mathrm{P}}$ — это некоторый интегральный оператор, описывающий распространение света в резонаторе от первого зеркала до второго и назад, то есть на расстояние $2l_0$, где l_0 — длина резонатора (рис. 6.1). Комплексные амплитуды мод $\psi_{\mathbf{p}}(\mathbf{x}_2)$, взятые на выходном зеркале открытого пассивного резонатора, удовлетворяют интегральному уравнению вида Фокса и Ли [3]:

$$\iint\limits_{D_2} K(\mathbf{x}_2, \mathbf{x}_2')\psi_{\mathbf{p}}(\mathbf{x}_2')\, \mathrm{d}^2\mathbf{x}_2' = \gamma_{\mathbf{p}}\psi_{\mathbf{p}}(\mathbf{x}_2), \tag{6.11}$$

где

$$K(\mathbf{x}_2, \mathbf{x}_2') = \left(\frac{k}{2\pi i}\right)^2 \iint\limits_{D_1} \frac{\exp\left[ik\left(L + L'\right)\right]}{LL'} \cdot \frac{(1+\cos\alpha)}{2} \cdot \frac{(1+\cos\alpha')}{2} \cos\alpha \cdot \cos\alpha' \, \mathrm{d}^2\mathbf{x}_1, \tag{6.12}$$

$\mathbf{x}_2$, $\mathbf{x}_2'$ — двумерные декартовы координаты проекции точки первого или второго зеркала, соответственно, на плоскость, перпендикулярную оптической оси, D_1, D_2 — области, соответствующие световым отверстиям зеркал; L, α — длина и угол наклона к оптической оси отрезка, соединяющего точку $\mathbf{x}_1$ первого зеркала с точкой $\mathbf{x}_2$ второго зеркала (рис. 6.1); L', α' — аналогичные величины для точек с координатами $\mathbf{x}_1'$ и $\mathbf{x}_2'$, k — волновое число.

Например, в параксиальном приближении можем положить, что

$$L \cong l_0 + \frac{g_1}{2l_0}\mathbf{x}_1^2 + \frac{g_2}{2l_0}\mathbf{x}_2^2 - \frac{1}{l_0}\mathbf{x}_1\mathbf{x}_2, \quad \cos\alpha \cong 1,$$

где $g_1 = 1 - l_0/R_1$, $g_2 = 1 - l_0/R_2$, R_1 и R_2 — радиусы кривизны в вершинах зеркал.

Для градиентного волоконного световода с поперечно-неоднородным показателем преломления $n(\mathbf{x})$ волновые фронты направляемых мод являются плоскими [1]. В этом случае оператор $\hat{\mathrm{P}}$ связывает решение $F(\mathbf{x}, z)$ уравнения Гельмгольца

$$\nabla_\perp^2 F(\mathbf{x}, z) + \frac{\partial^2 F(\mathbf{x}, z)}{\partial^2 z} + n^2(\mathbf{x})\, k^2 F(\mathbf{x}, z) = 0 \tag{6.13}$$

с граничным значением комплексной амплитуды $F|_{z=0} = F(\mathbf{x}, 0)$. Здесь $\nabla_\perp = (\partial/\partial x, \partial/\partial y)$ — поперечный дифференциальный оператор Гамильтона (рис. 6.2). При учете конечного диаметра волновода появятся дополнительные граничные условия на границе сердечника. Моды градиентного волокна имеют плоский волновой фронт и удовлетворяют уравнению [5]

$$\nabla_\perp^2 \psi_{\mathbf{p}}(\mathbf{x}) + \left[k^2 n^2(\mathbf{x}) - \beta_{\mathbf{p}}^2\right] \psi_{\mathbf{p}}(\mathbf{x}) = 0. \tag{6.14}$$

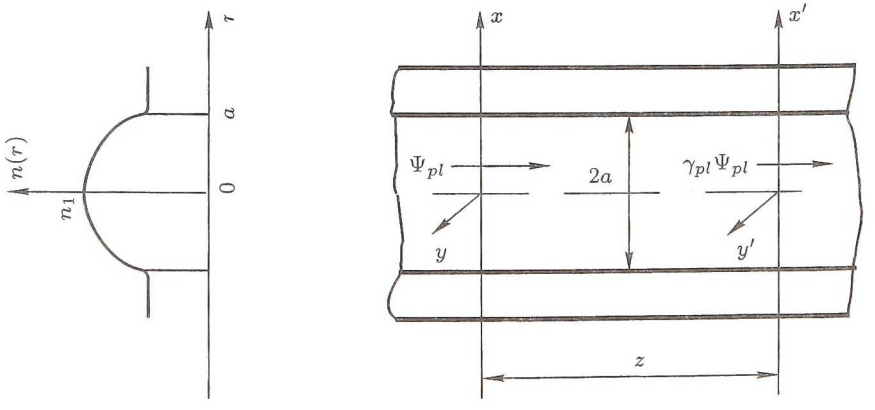


Рис. 6.2. Моды в оптоволокне с градиентным профилем $r = \sqrt{x^2 + y^2} \leqslant a$

Для произвольного значения z имеем

$$F(\mathbf{x}, z) = \gamma_{\mathbf{p}} \psi_{\mathbf{p}}(\mathbf{x}), \tag{6.15}$$

$$\gamma_{\mathbf{p}} = \exp(iK_{\mathbf{p}} z), \tag{6.16}$$

$$K_{\mathbf{p}} = \beta_{\mathbf{p}} + iq_{\mathbf{p}}, \tag{6.17}$$

где $\beta_{\mathbf{p}}$ — константа распространения, $q_{\mathbf{p}}$ — коэффициент затухания моды $\psi_{\mathbf{p}}$ (рис. 6.2). Моды градиентного оптоволокна удовлетворяют уравнению на собственные значения (6.1) для любого расстояния z; собственные значения определяются уравнением (6.16), а самовоспроизводимость мод в данном случае происходит с постоянным масштабом по декартовым координатам $\mathbf{x} = (x, y)$.

6.1.3. Метод комплексного эйконала

Механизм распространения модового пучка в свободном пространстве является чисто дифракционным эффектом и следовательно, не может быть описан с помощью эйконала в его традиционной интерпретации. В то же время, для описания дифракционных световых полей, возникающих внутри каустик, и при полном внутреннем отражении применяется комплексный эйконал, описывающий поля с затухающей амплитудой. В [6] показано, что принципиальная мода гауссова пучка может быть описана с применением метода комплексного эйконала в параксиальном приближении. Были найдены более общие решения уравнения Гельмгольца [8, 9].

В этом пункте мы исследуем такие модовые решения уравнения Гельмгольца, которые имеют волновые фронты заданной формы, являющиеся общими для всей системы мод. Комплексный эйконал, так же как уравнение комплексного переноса, использован как основа для построения мод высших порядков в криволинейных координатах. Такой подход дает, в частности, непараксиальные пучки, параксиальное приближение которых совпадает с модами Гаусса–Эрмита и Гаусса–Лагерра. Метод комплексного эйконала непосредственно обобщает геометрооптическое приближение [10] для среды с произвольным распределением показателя преломления $n = n(\mathbf{r})$, в частности, для однородной среды $n = n_0 \equiv \text{const}$. Требуемые волновые фронты будут выбраны из некоторого множества, порожденного вещественной функцией $S(\mathbf{r})$ трехмерной координаты $\mathbf{r}$, в соответствии с уравнением $S(r) = \text{const}$. Из [7] следует, что добавочная функция $B(\mathbf{r})$ должна быть введена таким образом, чтобы уравнение $B(\mathbf{r}) = \text{const}$ определяло поверхность заданного спада интенсивности. Эти поверхности похожи на гиперболоид гауссова пучка (рис. 6.3). Комплексный эйконал определяется уравнением

$$V(\mathbf{r}) = S(\mathbf{r}) + iB(\mathbf{r}). \tag{6.18}$$

Каждая из функций $V(\mathbf{r})$, $S(\mathbf{r})$, $B(\mathbf{r})$ не зависит ни от длины волны, ни от номера моды. Комплексная амплитуда при использовании метода комплексного эйконала представима в той же форме, что и в геометрической оптике:

$$F(\mathbf{r}) = A_0(\mathbf{r}) \tilde{F}(\mathbf{r}) \exp[ikV(\mathbf{r})], \quad k = 2\pi/\lambda, \tag{6.19}$$

где $A_0(\mathbf{r})$ — не зависящий от λ основной член асимптотического ряда по степеням $1/k$ [10], и $\tilde{F}(\mathbf{r})$ — некоторый мультипликативный компонент, учитывающий наличие остальных членов асимптотического ряда и представимый в виде

$$\tilde{F}(\mathbf{r}) = 1 + O(1/k), \quad \nabla \tilde{F}(\mathbf{r}) = O(1/k), \tag{6.20}$$

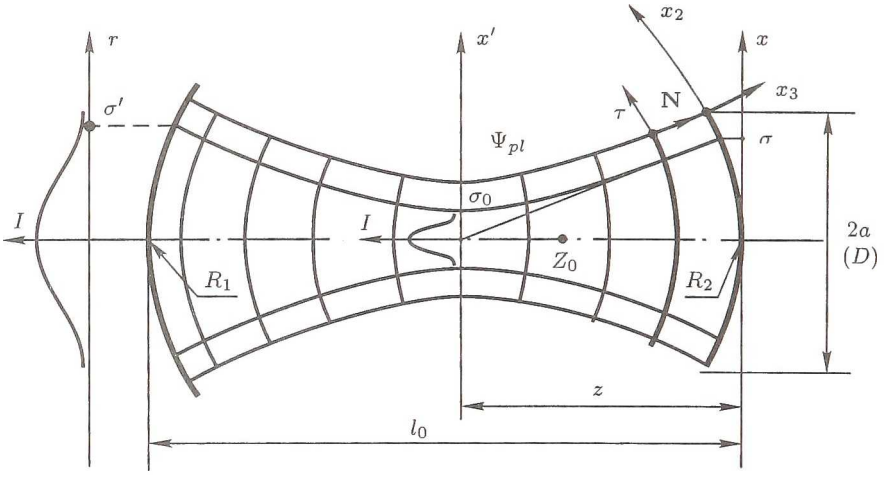


Рис. 6.3. Гауссовы пучки в свободном пространстве. σ_0 — радиус перетяжки основной моды, σ, σ' — радиусы основной моды на поверхностях зеркал, $2z_0$ — конфокальный параметр, I — интенсивность гауссова пучка, l_0 — длина резонатора, R_1, R_2 — радиусы кривизны зеркал, $2a$ — размер апертуры выходного зеркала

где величина $O(1/k)$ является величиной такого же порядка бесконечности, как и $1/k$ при $\lambda \to 0$, ∇ — оператор трехмерного градиента по $\mathbf{r}$. Мнимая часть B эйконала V соответствует затухающему изменению амплитуды поля по закону $\exp(-kB)$ вследствие дифракционного расплывания пучка. Существенно, что функции A_0, F и все входящие в них компоненты предполагаются медленно меняющимися по сравнению с быстро осциллирующим множителем $\exp(ikS)$. Подстановка (6.19) в уравнение Гельмгольца $\nabla^2 F + k^2 n^2 F = 0$ дает следующее уравнение:

$$\nabla^2 \left(A_0 \tilde{F}\right) + 2ikA_0 \nabla \tilde{F} \nabla V + ik\tilde{F} \left[2\nabla A_0 \nabla V + A_0 \nabla^2 V\right] + k^2 A_0 \tilde{F} \left[n^2 - (\nabla V)^2\right] = 0. \tag{6.21}$$

В силу (6.20), первые два члена (6.21) имеют порядок $O(k^0)$, вторые два — порядок $O(k^1)$, и два последних члена имеют порядок $O(k^2)$. Потребуем, чтобы функция V удовлетворяла уравнению комплексного эйконала

$$(\nabla V)^2 = n^2, \tag{6.22}$$

а основной член A_0 асимптотического разложения удовлетворял уравнению переноса

$$2\nabla A_0 \nabla V + A_0 \nabla^2 V = 0. \tag{6.23}$$

При выполнении (6.22) и (6.23), уравнение (6.21) для $\tilde{F}$ принимает форму

$$\nabla^2 \left(A_0 \tilde{F}\right) + 2ikA_0 \nabla \tilde{F} \nabla V = 0. \tag{6.24}$$

Уравнение (6.24) характеризует дифракционное расплывание светового пучка в процессе распространения. Уравнения (6.22) и (6.23) остаются в той же форме, что и в геометрической оптике для вещественного эйконала V Учитывая представление

(6.18) комплексного эйконала, получаем специфическое уравнение, учитывающее дифракционное расплывание,

$$(\nabla S)^2 - (\nabla B)^2 = n^2, \tag{6.25}$$

$$\nabla S \cdot \nabla B = 0, \tag{6.26}$$

$$2\nabla \ln A_0 \nabla B + \nabla^2 B - 2i\nabla \ln A_0 \cdot \nabla S - i\nabla^2 S = 0, \tag{6.27}$$

$$\nabla^2 \left(A_0 \tilde{F}\right) - 2ikA_0 \nabla \tilde{F} \nabla B + 2ikA_0 \nabla \tilde{F} \nabla S = 0. \tag{6.28}$$

В отличие от соответствующих уравнений, приведенных в [7], в данном случае мы не разделяем вещественную и мнимую части величины $\ln A_0$ для удобства дальнейших преобразований. Уравнения (6.25)–(6.28) определяют амплитудно-фазовые соотношения пучка. Будем называть поверхности вида $S(\mathbf{r}) = \text{const}$ волновыми фронтами S. Они отличаются от волновых фронтов, рассматриваемых в геометрической оптике. Множество кривых, ортогональных к волновым фронтам, отличается от прямых отрезков геометрооптических лучей (даже в свободном пространстве) и могут быть названы обобщенными лучами. Обобщенные лучи имеют в каждой точке единственный вектор касательной $\mathbf{N} = \nabla S/|\nabla S|$, направленный по градиенту S. Направление затухания амплитуды в каждой точке имеет единичный вектор $\boldsymbol{\tau} = \nabla B/|\nabla B|$, направленный по градиенту функции B и ортогональный к вектору $\mathbf{N}$ (в силу (6.26)) и, таким образом, лежащий в плоскости, касательной к волновому фронту S. Поле (6.19) с медленно меняющейся функцией $A_0\tilde{F}$, имеет поверхность с постоянной фазой $kS + \arg\left(A_0\tilde{F}\right) = \text{const}$, мало отличающуюся от волнового фронта $S(\mathbf{r}) = \text{const}$ при $k \to \infty$. Существенно, что метод комплексного эйконала, также как традиционный метод эйконала, дает нам возможность заменить световой пучок вблизи некоторой точки локальной плоской волной. Действительно, разложим медленно меняющиеся функции $S(\mathbf{r}')$ и $B(\mathbf{r}')$ в трехмерные ряды, по степеням $(\mathbf{r}' - \mathbf{r})$ до линейных членов. Учитывая (6.19) и условие $n = \text{const}$, получаем

$$F(\mathbf{r}') = A_0\tilde{F} \exp(ikS - kB) \exp\left[-kn\alpha\boldsymbol{\tau}(\mathbf{r}' - \mathbf{r})\right] \exp\left[ikn\mathbf{N}(\mathbf{r}' - \mathbf{r})\right], \tag{6.29}$$

где

$$\alpha = \frac{1}{n}|\nabla B|, \quad \beta = \frac{1}{n}|\nabla S|.$$

Уравнение (6.29) характеризует затухающую плоскую волну около точки $\mathbf{r}(\mathbf{r}' \cong \mathbf{r})$, распространяющуюся в направлении $\mathbf{N}$ и имеющую поперечное затухание τ вдоль волнового фронта. Отметим, что такая плоская волна не является геометрооптической, так как модуль $kn\beta$ ее волнового вектора $\mathbf{k} = kn\beta\mathbf{N}$ не равен волновому числу kn в данной среде. В данном случае вместо геометрооптического равенства имеем $\beta \neq 1$, и в силу (6.25) величина β вместе с функцией α удовлетворяет уравнению

$$\beta^2(\mathbf{r}) - \alpha^2(\mathbf{r}) = 1. \tag{6.30}$$

Важно подчеркнуть, что фаза поперечно затухающей волны (6.29) линейно зависит от координат:

$$kS = kn\beta\mathbf{N}\mathbf{r}'. \tag{6.31}$$

Переходя к традиционной геометрической оптике, имеем $\beta \to 1$ и $kS = kn\beta\mathbf{N}\mathbf{r}'$.

6.1.4. Амплитудно-фазовые соотношения для модовых пучков в свободном пространстве

Моды свободного пространства в обобщенных координатах. В [11, 12] были описаны два хорошо известных типа гауссовых пучков — пучки Гаусса–Лагерра и Гаусса–Эрмита, с использованием лишь параксиального приближения, что соответствует переходу от уравнения Гельмгольца (6.13) к параболическому уравнению. В данном разделе мы рассмотрим более общий класс пучков, имеющих более общие асферические волновые фронты, но тем не менее обладающих свойством самовоспроизведения при распространения. Поставив задачу селекции мод лазерного излучения с помощью оптических элементов (моданов), необходимо учитывать изменение размера модового пучка во время распространения в свободном пространстве между плоскостью оптического элемента и оптоволокном или лазером. Например, если луч расходится от выходного торца оптоволокна с высокой числовой апертурой, имея плоский волновой фронт, то в процессе распространения в свободном пространстве он приобретет сложный асферический волновой фронт. Похожие пучки распространяются также как внутри, так и вне открытых резонаторов лазеров. Особый интерес представляют пучки, изменяющие при распространении только масштаб и форму волнового фронта, но не меняющие поперечного распределения амплитуды в сечении пучка. Такие пучки как бы воспроизводят себя при распространении. Пучок, обладающий способностью *самовоспроизводства*, позволяет работать в плоскости модана с теми же модами, что и в оптоволокне с идеальным параболическим профилем. В то же время способность моды воспроизводить свое сечение при распространении в волноводной среде было нами охарактеризовано как фундаментальное свойств мод. Например, в оптических волокнах, это проявляется в непрерывном воспроизведении пространственной структуры вдоль оси z (6.15), (6.16). В лазерных резонаторах воспроизведение поперечной структуры поля происходит с точностью до масштаба и носит периодический, дискретный характер. Хорошо известны моды Гаусса–Эрмита и Гаусса–Лагерра, воспроизводящие свою поперечную структуру в свободном пространстве с точностью до масштабирующего множителя. Однако введя обобщенные координаты, мы сможем сказать, что они точно воспроизводят свою структуру. Гауссовы моды являются решениями уравнений (6.1) и (6.14) в параксиальном приближении с параметрами σ_0 и R, радиусом принципиальной моды по уровню спада интенсивности $\exp(-2) \cong 0{,}14$ и радиусом кривизны волнового фронта, соответственно. Гауссова мода в волноводе с градиентным профилем имеет постоянный размер $\sigma \equiv \sigma_0$, плоский волновой фронт ($R = \infty$) и обладает свойством непрерывного самовоспроизведения в соответствии с (6.15) при распространении (рис. 6.4). Гауссовы моды, рассматриваемые в пассивном резонаторе или в свободном пространстве, изменяют свой размер $\sigma = \sigma(z)$ и радиус волнового фронта $R = R(z)$ (рис. 6.3).

Комплексная амплитуда гауссовой моды на поверхности фронта может рассматриваться как самовоспроизводящаяся с точностью до меняющегося масштаба координат $\sigma(z)$, то есть, для мод, определенных в новых нормализованных координатах

$$x_1 = \frac{x}{\sigma(z)}, \quad x_2 = \frac{y}{\sigma(z)}, \quad x_3 = z, \tag{6.32}$$

с масштабом, меняющимся вдоль оси z. В свободном пространстве гауссовы пучки не продуцируют сами себя в форме простого повторения их поперечной структуры, как в случае с градиентными волокнами, но демонстрируют это свойство в системе специально отмасштабированных координат. Проблема формирования волновых

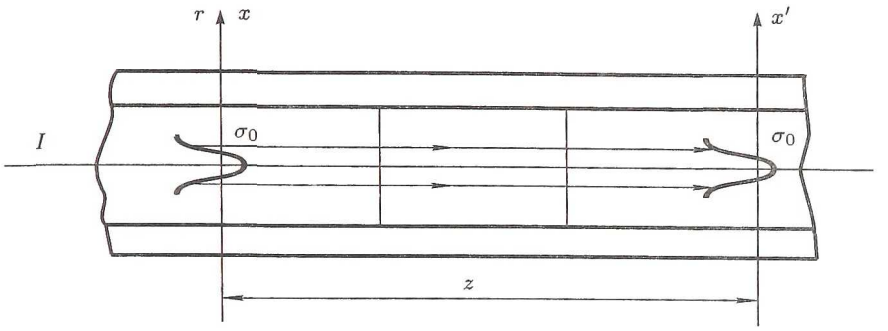


Рис. 6.4. Гауссовы моды в волокне. σ_0 — радиус принципиальной моды; z — расстояние, $\mathrm{p} = (p, l)$ — номера моды, ψ_{p} — комплексная амплитуда моды на поверхности волнового фронта, соответствующая значению координаты $x_3 = 0$, и F_{p} — комплексная амплитуда моды на некотором выходном волновом фронте с координатой $x_3 = \mathrm{const}$

фронтов с заданным распределением интенсивности обсуждалась в [13]. Формирование самовоспроизводящихся модовых пучков света в свободном пространстве также может рассматриваться в контексте решения этой проблемы. В данном случае существенным является то, что все эти пучки из полного множества должны иметь общий волновой фронт в определенной плоскости в свободном пространстве. Моды свободного пространства будут получены ниже на основе требуемого множества волновых фронтов, соответствующих произвольным ортогональным координатам (x_1, x_2, x_3), которые обобщают нормализованные координаты (6.32). «Продольная» координата x_3 соответствует направлению распространению пучка, а «поперечные» координаты (x_1, x_2) характеризуют расположение точки на поверхности $x_3 = \mathrm{const}$. Уравнение, предназначенное для нахождения мод, самовоспроизводящихся непрерывно, может быть записано в криволинейных координатах (x_1, x_2, x_3) в собственных значениях уравнения (6.2). В нашем случае (6.2) на волновом фронте $x_3 = \mathrm{const}$ принимает вид

$$F_{\mathrm{p}}(x_1, x_2, x_3) = \psi_{\mathrm{p}}(x_1, x_2) \cdot \gamma_{\mathrm{p}}(x_3); \quad \gamma_{\mathrm{p}}(0) = 1. \tag{6.33}$$

Комплексная амплитуда мод, взятая в произвольной точке, зависит как от (6.33), так и от формы волнового фронта, определяющего фазу. Таким образом, моды с заданным волновым фронтом имеют специфическую связь амплитуды и фазы; эта связь стабильна в силу свойства самовоспроизводимости.

Уравнения для мод с заданным общим волновым фронтом. Найдем моды с заданным семейством волновых фронтов $S(\mathbf{r}) = \mathrm{const}$ в однородной среде методом комплексного эйконала. Для этого перейдем к ортогональным криволинейным координатам (x_1, x_2, x_3), связанными с декартовыми координатами (x, y, z) соотношением $\mathbf{r} = \mathbf{r}(x_1, x_2, x_3)$, т.е. $x = x(x_1, x_2, x_3)$, $y = y(x_1, x_2, x_3)$, $z = z(x_1, x_2, x_3)$, так, что координатные плоскости $x_3 = \mathrm{const}$ соответствуют волновым фронтам $S(\mathbf{r}) = \mathrm{const}$, а координатные плоскости $x_1 = \mathrm{const}$ — определенному спаду интенсивности. Таким образом, можем записать,что

$$S = S(x_3), \quad B = B(x_1). \tag{6.34}$$

При этом криволинейные координатные линии продольной координаты x_3 ортогональны волновым фронтам $S = \mathrm{const}$ и характеризуют распространение модового пучка. В то же время, координатные линии по x_1 характеризуют направление спада интенсивности в поперечном сечении пучка.

Сопоставление (6.33), (6.34) и (6.19), (6.18) позволяет нам записать

$$F(x_1, x_2, x_3) = \psi_{\mathbf{p}}(x_1, x_2)\,\gamma_{\mathbf{p}}(x_3)\exp[ikS(x_3)], \tag{6.35}$$

$$\psi_{\mathbf{p}}(x_1, x_2)\,\gamma_{\mathbf{p}}(x_3) = A_0(x_1, x_2, x_3)\,\tilde{F}(x_1, x_2, x_3)\exp[-ikB(x_1)], \tag{6.36}$$

где $\psi_{\mathbf{p}}$ — собственные функции с номером $\mathbf{p}$, $\gamma_{\mathbf{p}}$ — соответствующие собственные значения в (6.2), и S — функция, характеризующая форму волнового фронта, общую для всех мод. В силу разделения поперечных переменных (x_1, x_2) с продольной переменной x_3 в уравнениях (6.35) и (6.36), функции $\tilde{F}$ и A_0 также должны быть разделяемы по переменным, то есть могут быть записаны в виде

$$A_0(x_1, x_2, x_3) = A_0(x_1, x_2)\gamma_0(x_3), \tag{6.37}$$

$$\tilde{F}(x_1, x_2, x_3) = \tilde{F}_{\mathbf{p}}(x_1, x_2)\,f_{\mathbf{p}}(x_3), \tag{6.38}$$

где $A_0(x_1, x_2)$, $\tilde{F}_{\mathbf{p}}(x_1, x_2)$, $\gamma_0(x_3)$, $f_{\mathbf{p}}(x_3)$ — некоторые введенные функции.

Будучи основным при $k \to \infty$, член $A_0(x_1, x_2)$ может быть интерпретирован как соответствующий принципиальной моде.

Далее, из (6.36) и (6.20) можем получить

$$\psi_0(x_1, x_2) = A_0(x_1, x_2)\exp[-kB(x_1)]. \tag{6.39}$$

Множитель $\tilde{F}$, определенный в (6.20) и примененный к (6.36), определяет разницу между модами высших порядков и основной модой. Из (6.36), (6.39) имеем

$$\psi_{\mathbf{p}}(x_1, x_2) = \psi_0(x_1, x_2)\,\tilde{F}_{\mathbf{p}}(x_1, x_2), \tag{6.40}$$

$$\gamma_{\mathbf{p}}(x_3) = \gamma_0(x_3)\,f_{\mathbf{p}}(x_3). \tag{6.41}$$

Перепишем (6.25), используя (6.28), в криволинейных ортогональных координатах, используя соответствующую формулу из [14] для операторов ∇, ∇^2. Благодаря разделению координат в представлении (6.37) и (6.38), получим

$$\left(\frac{1}{H_3}S'\right)^2 - \left(\frac{1}{H_1}B'\right)^2 = n^2, \tag{6.42}$$

$$\frac{1}{A_0\tilde{F}_{\mathbf{p}}}\frac{\partial}{\partial x_1}\left(\rho_1\frac{\partial\left(A_0\tilde{F}_{\mathbf{p}}\right)}{\partial x_1}\right) - 2k\rho_1 B'\frac{1}{\tilde{F}_{\mathbf{p}}}\frac{\partial\tilde{F}_{\mathbf{p}}}{\partial x_1} + \frac{1}{A_0\tilde{F}_{\mathbf{p}}}\frac{\partial}{\partial x_2}\left(\rho_2\frac{\partial\left(A_0\tilde{F}_{\mathbf{p}}\right)}{\partial x_2}\right) + \tag{6.43}$$

$$+\frac{1}{\gamma_0 f_{\mathbf{p}}}\frac{\partial}{\partial x_3}\left(\rho_3\frac{\partial(\gamma_0 f_{\mathbf{p}})}{\partial x_3}\right) + 2ik\rho_3\frac{1}{f_{\mathbf{p}}}\frac{\partial f_{\mathbf{p}}}{\partial x_3} = 0,$$

$$2\rho_1 B'\frac{\partial lnA_0}{\partial x_1} + \frac{\partial}{\partial x_1}(\rho_1 B') - 2i\rho_3 S'\frac{\partial\ln\gamma_0}{\partial x_3} - i\frac{\partial}{\partial x_3}(\rho_3 S') = 0, \tag{6.44}$$

$$\rho_1 = \frac{H_2H_3}{H_1}, \quad \rho_2 = \frac{H_1H_3}{H_2}, \quad \rho_3 = \frac{H_2H_1}{H3}, \tag{6.45}$$

где H_1, H_2, H_3 — коэффициенты Ламэ в декартовых координатах, а B', S' соответствуют производным функций B и S (6.34) по переменным x_1 и x_3 соответственно. Уравнение (6.43) показывает, что функция A_0 может быть выбрана независящей от переменной x_2. После того как функции S, B, A_0, γ_0 ,$F_{\mathbf{p}}$, $f_{\mathbf{p}}$ найдены из (6.42), используя (6.44), можем записать моды (6.39), (6.40) и собственные значения (6.41).

Особый интерес представляют параксиальные пучки, распространяющиеся вблизи оси z. Формальными условиями являются соотношения

$$\frac{|x|}{z} \ll 1, \quad \frac{|y|}{z} \ll 1, \tag{6.46}$$

Для этих пучков можем записать

$$A_0 \cong \text{const}, \tag{6.47}$$

$$\left| \frac{\partial}{\partial x_3} \left(\rho_3 \frac{\partial (\gamma_0 f_{\mathrm{p}})}{\partial x_3} \right) \right| \ll k \rho_3 \left| S' \gamma_0 \frac{\partial f_{\mathrm{p}}}{\partial x_3} \right|, \tag{6.48}$$

а вместо соотношений (6.43) и (6.44) получить параксиальные уравнения в форме

$$\frac{\partial}{\partial x_1} (\rho_1 B') - 2i\rho_3 S' \frac{\partial \ln \gamma_0}{\partial x_3} - i \frac{\partial}{\partial x_3} (\rho_3 S') = 0, \tag{6.49}$$

$$\frac{1}{\tilde{F}_{\mathrm{p}}} \frac{\partial}{\partial x_1} \left(\rho_1 \frac{\partial \tilde{F}_{\mathrm{p}}}{\partial x_1} \right) - 2k\rho_1 B' \frac{1}{\tilde{F}_{\mathrm{p}}} \frac{\partial \tilde{F}_{\mathrm{p}}}{\partial x_1} + \frac{1}{\tilde{F}_{\mathrm{p}}} \frac{\partial}{\partial x_2} \left(\rho_2 \frac{\partial \tilde{F}_{\mathrm{p}}}{\partial x_2} \right) + 2ik\rho_3 S' \frac{1}{f_{\mathrm{p}}} \frac{\partial f_{\mathrm{p}}}{\partial x_3} = 0. \tag{6.50}$$

Одномерные модовые пучки. Рассмотрим одномерные пучки, имеющие независимые от x_2 распределения комплексной амплитуды, порождаемые криволинейными цилиндрическими координатами, удовлетворяющими условиям типа Коши–Римана [14]:

$$\begin{cases} x = x(x_1, x_3), \\ y = x_2, \\ z = z(x_1, x_3), \end{cases} \tag{6.51}$$

$$\begin{cases} \dfrac{\partial x}{\partial x_1} = \dfrac{\partial z}{\partial x_3}, \\ \dfrac{\partial x}{\partial x_3} = -\dfrac{\partial z}{\partial x_1}. \end{cases} \tag{6.52}$$

Для пучков вида (6.51) имеем

$$H_1^2 = H_3^2 = \left(\frac{\partial x}{\partial x_1} \right)^2 + \left(\frac{\partial z}{\partial x_1} \right)^2; \quad H_2^2 = 1, \tag{6.53}$$

$\rho_1 = \rho_3 = 1$; $\rho_2 = H_1^2$, где ρ_2 является независимым от x_2. Уравнение эйконала для одномерных мод имеет вид

$$[S'(x_3)]^2 - [B'(x_1)]^2 = H_1^2 \tag{6.54}$$

и накладывает некоторые ограничения на возможные преобразования координат (6.51). Так как переменная H_1^2, введенная в (6.54), должна быть суммой функций x_1 и x_3, она должна удовлетворять уравнению

$$\frac{\partial^2 H_1^2}{\partial x_1 \partial x_3} = 0. \tag{6.55}$$

Удобно записать преобразование (6.51) в комплексном виде $X = x+iz$, $U = x_1 + ix_3$. В силу (6.52) комплексная функция $X = X(U)$ должна удовлетворять условиям Коши–Римана, приведенным в [14]. Таким образом, $X(U)$ — аналитическая функция, а преобразование $X(U)$ является конформным ($H_1 \neq 0$). При этом $H_1^2 = |X'(U)|^2 = X' \cdot X'^*$, где X' является производной X по комплексной переменной U. Уравнение (6.55) может быть переписано в форме

$$\frac{\partial^2}{\partial x_1 \partial x_3}\left(X' \cdot X'^*\right) = i\left(X'^* \cdot X''' - X' \cdot X'''^*\right) = 0$$

или

$$\frac{X'}{X'''} = \frac{X'^*}{X'''^*}. \tag{6.56}$$

Условие (6.56) означает, что функция X'/X''' является вещественной. Таким образом, аналитическая функция $X'(U)$ удовлетворяет уравнению

$$(X')'' + \omega_0^2(U)X' = 0 \tag{6.57}$$

с действительным коэффициентом ω_0^2. Условие (6.57) определяет класс допустимых преобразований координат. Весь класс допустимых преобразований координат для цилиндрических волновых фронтов может быть получен путем варьирования гладкой вещественной функции $\omega_0(U)$ в ходе решения (6.57). Уравнение (6.57) дает такие коэффициенты Ламэ H_1^2, которые могут быть представлены как сумма функций переменных x_1 и x_3. Для того чтобы найти эйконалы B и S, зная H_1^2, используем произвольный характер выбора функций $\theta(x_1, x_3) > 1$ и $\chi(x_1, x_3)$ в равенстве

$$H_1^2 = \left(\frac{\theta^2}{\theta^2 - 1}H_1^2 + \chi\right) - \left(\frac{1}{\theta^2 - 1}H_1^2 + \chi\right).$$

Выберем их таким образом, чтобы первая скобка зависела от переменной x_1, а вторая — от x_3. Далее, полагая производные выражений в скобках равными нулю, получаем уравнения

$$\begin{cases} \dfrac{\partial \chi}{\partial x_1} = -\dfrac{\partial}{\partial x_1}\left(\dfrac{\theta^2}{\theta^2 - 1}H_1^2\right), \\ \dfrac{\partial \chi}{\partial x_3} = -\dfrac{\partial}{\partial x_3}\left(\dfrac{1}{\theta^2 - 1}H_1^2\right). \end{cases}$$

Так как для произвольного θ условие $\frac{\partial}{\partial x_3}\frac{\partial \chi}{\partial x_1} - \frac{\partial}{\partial x_1}\frac{\partial \chi}{\partial x_3} = \frac{\partial^2}{\partial x_1 \partial x_3}H_1^2 = 0$ выполняется в силу (6.54), можем восстановить функцию $\chi(x_1, x_3)$ с точностью до некоторой константы χ_0, зная ее производную для любого θ, в частности $\theta = \text{const}\,(\theta > 1)$. Далее можем восстановить компоненты эйконала B и S из уравнений

$$\frac{\partial S}{\partial x_3} = \left(\frac{\theta^2}{\theta^2 - 1}H_1^2 + \chi\right)^{1/2}, \quad \frac{\partial B}{\partial x_1} = \left(\frac{H_1^2}{\theta^2 - 1} + \chi\right)^{1/2},$$

записанных таким образом, чтобы получить уравнение комплексного эйконала. Здесь произвольная константа χ_0 функции χ выбирается так, чтобы обеспечить положительное значение выражение в скобках в предыдущем уравнении. Ввиду (6.53),

уравнение переноса (6.43) и уравнение (6.44) допускают разделение переменных, что дает нам уравнения

$$2B'\frac{\partial \ln A_0(x_1)}{\partial x_1} + B'' - C = 0, \tag{6.58}$$

$$2S'\frac{\partial \ln \gamma_0}{\partial x_3} + S'' + iC = 0, \tag{6.59}$$

$$\frac{1}{A_0}\frac{\mathrm{d}^2\left(A_0\tilde{F}_p\right)}{\mathrm{d}x_1^2} - 2kB'\frac{\mathrm{d}\tilde{F}_p}{\mathrm{d}x_1} + 2kn_0\beta_p\tilde{F}_p = 0, \tag{6.60}$$

$$\frac{1}{\gamma_0}\frac{\mathrm{d}^2}{\mathrm{d}x_3^2}(\gamma_0 f_p) + 2ikS'\frac{\mathrm{d}f_p}{\mathrm{d}x_3} - 2kn_0\beta_p f_p = 0, \tag{6.61}$$

где константа разделения для (6.44) обозначена как $2kn_0\beta_0$. В этом случае уравнения переноса (6.58), (6.59) могут быть проинтегрированы в общей форме:

$$A_0(x_1) = \exp\left[\frac{1}{2}\int_0^{x_1}\frac{C - B''}{B'}\,\mathrm{d}x_1\right], \tag{6.62}$$

$$\gamma_0(x_3) = \exp\left[-\frac{1}{2}\int_0^{x_3}\frac{iC - S''}{S'}\,\mathrm{d}x_3\right]. \tag{6.63}$$

Уравнение (6.60) является интегрируемым только для некоторых специфических типов эйконала. В частности, уравнения (6.62), (6.63), (6.61), и (6.60) принимают следующий вид в параксиальном приближении:

$$A_0(x_1) \cong \text{const}, \tag{6.64}$$

$$B'' - C = 0, \quad B(x_1) = C(x_1 - x_{10})^2, \quad C = \frac{1}{2}B''(0), \tag{6.65}$$

$$f_p(x_3) = \exp\left[-i\beta_p n_0\int_0^{x_3}\frac{\mathrm{d}x_3}{S'(x_3)}\right], \tag{6.66}$$

$$\frac{\mathrm{d}^2\tilde{F}_p}{\mathrm{d}x_1^2} - 2kB'\cdot\frac{\mathrm{d}\tilde{F}_p}{\mathrm{d}x_1} + 2kn_0\beta_p\tilde{F}_p = 0. \tag{6.67}$$

Зная функции $B, S, A_0, f_p, \tilde{F}_p$, можем получить формулы для одномерных мод и собственных функций, используя формулы (6.39)–(6.41):

$$\psi_0(x_1) = A_0(x_1)\exp[-kB(x_1)], \tag{6.68}$$

$$\psi_p(x_1) = \psi_0(x_1)\tilde{F}_p(x_1), \tag{6.69}$$

$$\gamma_p(x_3) = \gamma_0(x_3) f_p(x_3). \tag{6.70}$$

Выше было показано, что одномерные пучки способны порождать двумерные пучки с разделенными поперечными координатами. Рассмотрим разделение переменных в декартовых координатах в рамках параксиального приближения. Предположим, что мы уже нашли одномерную модовую комплексную амплитуду

$$F_p(x,z) = \exp[ikS(x,z)]\,\psi_p(x,z)\,\gamma_p(z), \tag{6.71}$$

удовлетворяющую уравнению Гельмгольца и выраженную в декартовой системе координат вместо криволинейных координат x_1, x_3. Аналогичным образом могут быть найдены функции $F_l(y,z)$ для оси y. Уравнение Гельмгольца в параксиальном приближении переходит в параболическое уравнение, с разделением переменных в виде

$$F_{\mathbf{p}}(\mathbf{x},z) = F_p(\mathbf{x},z)\, F_l(\mathbf{x},z) \exp(-ikn_0 z). \tag{6.72}$$

Все методы, рассмотренные в этом пункте, применимы к двумерным модам. Однако, необходимо отметить, что эти моды должны иметь более сложный волновой фронт, определяемый из уравнения

$$S = S(x,z) + \tilde{S}(y,z) - n_0 z = \text{const}, \tag{6.73}$$

где $S(x,z)$, $\tilde{S}(y,z)$ — эйконалы S для соответствующих одномерных случаев. Двумерные моды и их собственные значения $\gamma_{\mathbf{p}}(z)$ могут быть построены на основе одномерных по формулам

$$\psi_{\mathbf{p}}(\mathbf{x},z) = \psi_p(x,z)\,\psi_l(y,z), \tag{6.74}$$

$$\gamma_{\mathbf{p}}(z) = \gamma_p(z)\,\gamma_l(z), \tag{6.75}$$

что следует из (6.72), (6.73) и (6.35). Таким образом, стратегия разделения поперечных координат дает нам удобный метод построения модовых пучков различных порядков, обладающих способностью к самовоспроизводству в ходе распространения. Следует отметить, что поперечная структура моды, факторизуемой по x, y, может воспроизводить себя не только в двумерном случае (6.33), но также для любого одномерного пучка, порожденного в плоскости, параллельной к плоскости (x,z) или (y,z).

6.1.5. Гауссовы и бесселевы моды

Обобщенные моды Гаусса–Лагерра. В [7] было показано, что в однородной среде с показателем преломления $n_0 = \text{const}$ существует важный класс световых пучков с волновыми фронтами в форме эллипсов вращения. Там же введена система вырожденных эллипсоидальных координат сжатого эллипсоида вращения:

$$\begin{cases} x = b \sin x_1 \operatorname{ch} x_3 \cos x_2, \\ y = b \sin x_1 \operatorname{ch} x_3 \sin x_2, \\ z = b \cos x_1 \operatorname{sh} x_3, \end{cases} \tag{6.76}$$

где $0 \leqslant x_1 \leqslant \pi/2$, $0 \leqslant x_2 \leqslant 2\pi$, $0 \leqslant x_3 < \infty$.

Поверхности постоянного спада интенсивности $x_1 = \text{const}$ в данном случае представляют собой однополостные гиперболоиды вращения. Согласно [14], коэффициенты Ламэ могут быть найдены из уравнений

$$\begin{cases} H_1 = b\left(\operatorname{ch}^2 x_1 - \sin^2 x_3\right)^{1/2}, \\ H_2 = b \sin x_1 \operatorname{ch} x_3, \\ H_3 = H_1. \end{cases} \tag{6.77}$$

В [7] описаны компоненты эйконала S и B, соответствующие (6.76):

$$\begin{aligned} S &= n_0 b \operatorname{sh} x_3, \\ B &= n_0 b (1 - \cos x_1). \end{aligned} \tag{6.78}$$

В рамках параксиального приближения, решения уравнений (6.47)–(6.50), (6.35) и (6.39)–(6.41) могут быть найдены в аналитической форме и соответствуют хорошо известным модам Гаусса–Лагерра. Моды могут быть получены из решения общих уравнений (6.49) и (6.50), взятых в системе координат (6.76). Уравнения (6.76), записанные в форме $r = b \sin x_1 \operatorname{ch} x_3$, $z = b \cos x_1 \operatorname{sh} x_3$, где $r = \sqrt{x^2 + y^2}$, позволяют нам заключить, что преобразование координат имеет место в плоскости r, z вокруг оси z. Учитывая (6.45) и (6.77), имеем

$$\rho_1 = \rho_3 = b \sin x_1 \operatorname{ch} x_3, \quad \rho_2 = \frac{b\left(\operatorname{ch}^2 x_3 - \sin^2 x_1\right)}{\sin x_1 \operatorname{ch} x_3}.$$

Функции ρ_1 и ρ_3 имеют разделяемые переменные x_1, x_3, и функция ρ_2 является независимой от переменной x_2. Эйконал может быть определен в точной форме (6.78). Опишем множитель $\tilde{F}_{\mathbf{p}}$, нормирующую константу для функции $\psi_{\mathbf{p}}$. Уравнение переноса рассматривается в виде

$$\frac{\partial \ln A_0}{\partial x_1} = \frac{C - \cos x_1}{\sin x_1} \equiv \frac{\sin(0{,}5x_1)}{\cos(0{,}5x_1)}; \quad \frac{\partial \ln \gamma_0}{\partial x_3} = -\frac{iC + \operatorname{sh} x_3}{\operatorname{ch} x_3} = -\frac{i \operatorname{ch} x_3}{1 + i \operatorname{sh} x_3}$$

и легко интегрируется в элементарных функциях:

$$A_0(x_1) = \frac{1}{\cos^2(0{,}5x_1)}, \tag{6.79}$$

$$\gamma_0(x_3) = \frac{1}{1 + i \operatorname{sh} x_3} = \frac{1}{\operatorname{ch} x_3} \exp\left(i \cdot \operatorname{arctg}(\operatorname{sh}(x_3))\right). \tag{6.80}$$

Координата x_2 может быть легко отделена от координат (x_1, x_3) в (6.44) делением обеих сторон уравнения на ρ_2. Координаты x_1 и x_2 также разделяются,

$$\tilde{F}_{\mathbf{p}}(x_1, x_2) = \tilde{F}_p(x_1)\, \tilde{F}_l(x_2), \tag{6.81}$$

и для $\tilde{F}_l$ по x_2 получается уравнение осциллятора :

$$\frac{\mathrm{d}^2 \tilde{F}_l}{\mathrm{d}x_2^2} + l^2 \tilde{F}_l = 0, \quad \tilde{F}_l(x_2) = \exp(\pm l x_2), \tag{6.82}$$

где l^2 обозначает константу разделения. Требование к функции $\tilde{F}_l(x_2)$ — иметь единственное значение в каждой точке x_2, приводит к требованию целых значений номера l, $l = 1, 2, 3, \ldots$.

Отделение переменных x_1 и x_3 от константы, обозначенной как $-2kbn_0(2p + l)$, где p — некоторый порядковый номер, далее дает нам следующее:

$$\frac{1}{A_0 \sin x_1} \cdot \frac{\partial}{\partial x_1}\left(\sin x_1 \frac{\partial\left(A_0 \tilde{F}_{\mathbf{p}}\right)}{\partial x_1}\right) - 2kbn_0 \sin x_1 \frac{\partial \tilde{F}_{\mathbf{p}}}{\partial x_1} + \\ + \left[2kbn_0(2p + l) - \frac{l^2}{sin^2 x_1}\right] \tilde{F}_{\mathbf{p}} = 0, \tag{6.83}$$

$$\frac{1}{\gamma_0 \operatorname{ch} x_3} \cdot \frac{\partial}{\partial x_3}\left(\operatorname{ch} x_3 \frac{\partial(\gamma_0 f_{\mathbf{p}})}{\partial x_3}\right) + 2kbn_0 \operatorname{ch} x_3 \frac{\partial f_{\mathbf{p}}}{\partial x_3} - \\ - \left[2kbn_0(2p + l) - \frac{l^2}{\operatorname{ch} x_3}\right] f_{\mathbf{p}} = 0, \tag{6.84}$$

где A_0, γ_0 определяются уравнениями (6.79) и (6.80). Уравнения (6.79) и (6.80), (6.82)–(6.84) определяют непараксиальные пучки Гаусса–Лагерра, распространяющиеся в однородной среде подобно решениям [8]. Уравнения (6.47)–(6.50), взятые в параксиальном приближении, приводят к обычным модам Гаусса–Лагерра. Действительно, для рассмотренного параксиального приближения, функции B и S в уравнении (6.78) и соотношения для преобразований координат могут быть разложены по степеням x_1. Таким образом, мы можем записать следующие соотношения:

$$r \cong bx_1 \operatorname{ch} x_3 \cong bx_1\sqrt{1+\operatorname{sh}^2 x_3}, \quad z \cong b\operatorname{sh} x_3\left(1-\frac{x_1^2}{2}\right) \cong b\operatorname{sh} x_3,$$

$$\operatorname{sh} x_3 \cong \frac{z}{b}\left(1+\frac{x_1^2}{2}\right) \cong \frac{z}{b}, \quad x_1 = \frac{r}{b\sqrt{1+z^2/b^2}}, \tag{6.85}$$

$$B \cong n_0 b\frac{x_1^2}{2} = \frac{n_0 r^2}{2b\left(1+z^2/b^2\right)} = \frac{r^2}{k\sigma^2}, \tag{6.86}$$

$$S \cong n_0 z\left(1+\frac{r^2}{2b^2\left(1+z^2/b^2\right)}\right) = n_0 z + \frac{n_0 r^2}{2R}, \tag{6.87}$$

где

$$R = z\left(1+\frac{b^2}{z^2}\right), \quad \sigma^2 = \sigma_0^2\left(1+\frac{z^2}{b^2}\right), \quad \sigma_0^2 = \frac{2b}{kn_0}. \tag{6.88}$$

Далее, уравнения (6.79), (6.80), и (6.85) дают нам

$$A_0(x_1) \cong 1, \quad \gamma_0(x_3) = \frac{\sigma_0}{\sigma}\exp\left[-i\eta_0(z)\right]; \tag{6.89}$$

из уравнения (6.48), $l \ll b/\lambda$, получаем

$$f_{\mathbf{p}}(x_3) = \exp\left[-i(2p+l)\eta_0(z)\right], \tag{6.90}$$

где

$$\eta_0(z) = \operatorname{arctg}\frac{z}{b}. \tag{6.91}$$

В параксиальном приближении множители γ_0, $f_{\mathbf{p}}$, зависящие от x_3, зависят лишь от z и не зависят от x. Уравнение (6.83) принимает вид

$$\frac{1}{x_1}\cdot\frac{\partial}{\partial x_1}\left(x_1\frac{\partial\tilde{F}_{\mathbf{p}}}{\partial x_1}\right) - 2kbn_0x_1\frac{\partial\tilde{F}_{\mathbf{p}}}{\partial x_1} + \left[2kbn_0(2p+l) - \frac{l^2}{x_1^2}\right] = 0. \tag{6.92}$$

Используя традиционную подстановку

$$\tilde{F}_{\mathbf{p}} = t^{1/2}\cdot L_p^l(t), \quad t = kn_0x_1^2, \tag{6.93}$$

можно свести (6.92) к каноническому уравнению для лагерровских многочленов,

$$t\frac{\mathrm{d}^2L_p^l}{\mathrm{d}t^2} + (l+1-t)\frac{\mathrm{d}L_p^l}{\mathrm{d}t} + pL_p^l = 0,$$

которое имеет конечные решения в нуле и бесконечные точки только для целых значений $p = 0, 1, 2, 3 \ldots$. Подставляя полученные уравнения (6.86), (6.87), и (6.89)–(6.93) в (6.35) и (6.39)–(6.41), и учитывая (6.85), получаем хорошо известную формулу для комплексных амплитуд мод Гаусса–Лагерра, описанных в [15]:

$$F_{\mathbf{p}}(\mathbf{x}, z) = \psi_{\mathbf{p}}(\mathbf{x})\,\gamma_{\mathbf{p}}(z)\exp\left[ikn_0\left(z + \frac{\mathbf{x}^2}{2R}\right)\right], \tag{6.94}$$

$$\gamma_{\mathbf{p}}(z) = \exp\left[-i\left(r_{\mathbf{p}} + 1\right)\eta_0(z)\right], \tag{6.95}$$

$$\psi_{\mathbf{p}}(x) = E_{pl}\left(\frac{\sqrt{2}\,|\mathbf{x}|}{\sigma}\right)^l L_p^l\left(\frac{2\mathbf{x}^2}{\sigma^2}\right)\exp\left(-\frac{\mathbf{x}^2}{\sigma^2}\right)\exp\left(\pm il\alpha\right). \tag{6.96}$$

где α — полярный угол вектора x, L_p^l — обобщенный полином Лагерра [14] с номерами $p, l = 0, 1, 2 \ldots$, $r_{\mathbf{p}} = 2p + l$,

$$E_{pl} = \frac{2}{\sigma\sqrt{2\pi C_{p+l}^l l!}} \tag{6.97}$$

– константа нормировки. В некоторых случаях моды (6.96) записывают в вещественной форме, используя $\sin(l\alpha)$ и $\cos(l\alpha)$ вместо $\exp(\pm il\alpha)$. Значение $z = 0$ соответствует точке перетяжки пучка, $2b$ является конфокальным параметром гауссова пучка, а параметры σ, R и η определяются (6.88) и (6.91). Таким образом, для введенной системы эллиптических координат метод комплексного эйконала в параксиальном приближении позволяет получить классические моды Гаусса–Лагерра. На основе квазиклассического приближения [16] можно легко оценить диаметр $2a = 2\sigma\sqrt{(2p + l)_{\max} + 1}$ для мод Гаусса–Лагерра, пронумерованных таким образом, что $2p + l \leqslant (2p + l)_{\max}$, для больших значений номеров p и l.

Обобщенные моды Гаусса–Эрмита. При $\omega_0 = \text{const}$ одним из решений уравнения (6.57) является функция

$$X(U) = b\sin U. \tag{6.98}$$

Переходя к вещественной и мнимой частям $X = x + iz$, получаем $x = b\sin x_1 \operatorname{ch} x_3$, $z = b\cos x_1 \operatorname{sh} x_3$, $H_1^2 = b^2\left(\operatorname{ch}^2 x_3 - \sin^2 x_1\right) = H_3^2$.

Эйконал имеет здесь тот же вид (6.78), что и для мод Гаусса–Лагерра. Как показано ниже, параксиальные уравнения с учетом (6.72) дают решение в виде [15] для мод Гаусса–Эрмита. Для непараксиального варианта, (6.98) описывает обобщение одномерной моды Гаусса–Эрмита. Другие обобщения одномерных мод могут основываться на конформных преобразованиях координат, отличных от (6.98). В силу (6.57), такие преобразования порождаются вещественнозначными функциями ω_0^2, изменяющимися по x_1, x_2. Одномерные моды Гаусса–Эрмита могут быть получены из уравнений (6.68)–(6.72), (6.74), содержащих функции S, B, $A_0(x_1)$, $\gamma_0(x_3)$, $\tilde{F}_p(x_1)$.

Из уравнений (6.62), (6.63), и (6.60), (6.61) следует, что

$$A_0(x_1) = \frac{1}{\cos(x_1/2)}, \tag{6.99}$$

$$\gamma_0(x_3) = \frac{1}{\sqrt{1 + i\operatorname{sh} x_3}} = \frac{1}{\sqrt{\operatorname{ch} x_3}}\exp\left(-\frac{1}{2}i\operatorname{arctg}\left[\operatorname{sh} x_3\right]\right), \tag{6.100}$$

$$f_p(x_3) = \exp\left(-ip\operatorname{arctg}\left[\operatorname{sh} x_3\right]\right), \tag{6.101}$$

где

$$p = \beta l / b. \tag{6.102}$$

Функция $\tilde{F}_p$ найдена путем решения обычного дифференциального уравнения второго порядка с изменяющимися коэффициентами (6.60):

$$\cos\frac{x_1}{2}\cdot\frac{\mathrm{d}^2}{\mathrm{d}x_1^2}\left(\frac{\tilde{F}_p}{\cos(x_1/2)}\right) - 2kbn_0\sin x_1\frac{\mathrm{d}\tilde{F}_p}{\mathrm{d}x_1} + 2kn_0b_p\tilde{F}_p = 0. \tag{6.103}$$

В случае параксиального приближения запишем

$$\frac{|x|}{z} \ll 1, \quad \cos\frac{x_1}{2} \cong 1; \quad \sin x_1 \cong x_1, \quad A_0 \cong 1. \tag{6.104}$$

При этом выражение (6.103) для $\tilde{F}_p$ (заменой переменных вида $t = \sqrt{kbn_0}x_1$) сводится к стандартным полиномам Эрмита [14]:

$$\frac{\mathrm{d}^2\tilde{F}_p}{\mathrm{d}t^2} - 2t\frac{d\tilde{F}_p}{\mathrm{d}t} = -2p\tilde{F}_p.$$

Последнее уравнение имеет решения для $t = 0$ и $t \to \infty$ только для целых значений $p = 0, 1, 2, \ldots$. Решения для $\tilde{F}_p$, таким образом, могут быть записаны в форме

$$\tilde{F}_p = \mathrm{const}H_p\left(\sqrt{bkn_0x_1}\right). \tag{6.105}$$

Далее, используя (6.99)–(6.103) и (6.104), получаем (6.85)–(6.88), где σ^2 и R определяются (6.88) и (6.91), как в случае с модами Гаусса–Лагерра. Для γ_0 и f_p, имеем

$$\gamma_0(x_3) = \sqrt{\frac{\sigma_0}{\sigma}}\exp\left(-i\frac{1}{2}\operatorname{arctg}\frac{z}{b}\right), \tag{6.106}$$

$$f_p(x_3) = \exp\left(-ip\operatorname{arctg}\frac{z}{b}\right). \tag{6.107}$$

В параксиальном приближении значения γ_0 и f_p зависят только от z и не зависят от x. В заключение получаем, что двумерные моды и соответствующие собственные значения могут быть получены из знания соответствующих одномерных мод и, используя (6.105)–(6.106), могут быть записаны в следующем виде [15]:

$$\psi_{\mathbf{p}}(\mathbf{x}) = E_{pl}H_p\left(\frac{\sqrt{2}x}{\sigma}\right)H_l\left(\frac{\sqrt{2}y}{\sigma}\right)\exp\left(-\frac{\mathbf{x}^2}{\sigma^2}\right), \tag{6.108}$$

$$r_{\mathbf{p}} = p + l, \tag{6.109}$$

и

$$E_{pl} = \frac{1}{\sigma}\sqrt{\frac{2}{2^{p+l}p!l!}}, \tag{6.110}$$

где $H_p(\cdot)$ — полином Эрмита p-го порядка. Используя квазиклассическое приближение [16] для больших значений p и l, легко определить размер $2a \times 2b$ области, содержащей моды до номеров (p, l):

$$2a = 2\sigma\sqrt{p_{\max} + 1/2}, \quad 2b = 2\sigma\sqrt{l_{\max} + 1/2}, \tag{6.111}$$

где $p_{\max}$ и $l_{\max}$ — максимальные значения модовых номеров p или l.

6.1.6. Гауссовы моды в пассивных резонаторах и градиентных волокнах

Гауссовыми модами называются модовые функции из базисов Гаусса–Эрмита и Гаусса–Лагерра, обладающие несколькими общими свойствами. Строго говоря, гауссовы моды ортогональны на бесконечной области. Однако, из-за быстрого убывания амплитуды, их можно рассматривать как ортогональный базис на некоторой конечной области G. Эти модовые функции описывают моды не очень высоких порядков в открытых пассивных резонаторах с зеркалами сферического профиля. Кроме того, гауссовы моды характеризуют моды градиентных волокон с параболическим профилем и моды других линзоподобных сред. Моды Гаусса–Лагерра соответствуют круглым зеркалам или градиентным волокнам с цилиндрическим сердечником. Моды Гаусса–Эрмита соответствуют квадратным зеркалам или градиентным волокнам с сердечником квадратного профиля. Для получения гауссовых мод резонатора необходимо соблюдение условия параксиального приближения:

$$\frac{2a2b}{R_1R_2} \ll 1 \quad \text{или} \quad \frac{(2a)^2}{R_1R_2} \ll 1, \tag{6.112}$$

где $2a \times 2b$ — размеры апертуры наибольшего сферического зеркала, R_1, R_2 — радиусы кривизны сферических зеркал. Для резонаторов параметр σ_0 может быть найден из уравнения [17]

$$\sigma_0 = \left(\frac{\lambda}{\pi}\right) \frac{\left[l\left(R_1 - l\right)\left(R_2 - l\right)\left(R_1 + R_2 - l\right)\right]^{1/2}}{\left(R_1 + R_2 - 2l\right)}. \tag{6.113}$$

Перетяжка располагается в резонаторе на расстоянии

$$z_2 = \frac{l\left(R_1 - l\right)}{R_1 + R_2 - 2l}. \tag{6.114}$$

от выходного зеркала. Для случая распространения моды в волокне с квадратическим профилем преломления

$$n^2\left(r\right) = n_1^2\left(1 - 2\Delta\frac{r^2}{a^2}\right), \tag{6.115}$$

имеем

$$\gamma_{\mathbf{p}}\left(z\right) = \exp\left(i\beta_{\mathbf{p}}z\right), \tag{6.116}$$

где β_p найдем по формуле [15]:

$$\beta_{\mathbf{p}} = \left[k^2n_0^2 - \frac{4}{\sigma_0^2}\left(r_{\mathbf{p}} + 1\right)\right]^{1/2}, \tag{6.117}$$

$$\sigma_0 = \left(\frac{\lambda a}{\pi n_1}\right)^{1/2}\left(\frac{1}{2\Delta}\right)^{1/4} = \text{const}, \quad R = \infty. \tag{6.118}$$

Гауссовы моды сохраняют свою структуру не только при распространении в свободном пространстве, но и при прохождении через линзы и линзоподобные среды. При прохождении гауссовой моды через фурье-каскад [15] с фокусным расстоянием линзы f, параметры пучка σ и $R = 0$, изменяются на σ_F и R_F, соответственно, где

$$\sigma_F = \frac{f}{\pi\sigma}, \quad R_F = \infty. \tag{6.119}$$

Фаза пучка меняется в этом случае на следующую величину:

$$kf + \arg\left(i^{2p+l}\right) = kf + r_{\mathbf{p}}\frac{\pi}{2}. \tag{6.120}$$

Благодаря инвариантности гауссовых мод к прохождению через фурье-каскад, последний может быть использован в качестве оптической проекционной системы для проекции пучка с заданным модовым составом с выходного торца волокна в плоскость ДОЭ или наоборот.

Бесселевы моды. Амплитудно-фазовое распределение в сечении бесселевой моды описывается следующим уравнением [5]:

$$\psi_{\mathbf{p}}(\mathbf{x}) = E_{pl}J_l\left(\rho_{pl}\frac{|\mathbf{x}|}{a}\right)\exp(\pm il\alpha), \quad |\mathbf{x}| \leqslant a; \quad p, l = 0, 1, 2\ldots, \tag{6.121}$$

$$E_{pl} = \frac{1}{a\sqrt{\pi}J_l'(\rho_{pl})}, \tag{6.122}$$

где J_l — бесселева функция целого порядка, ρ_{pl} — p-й нуль функции J_l. Бесселевы моды являются собственными функциями оператора распространения света в волокне со ступенчатым профилем в свободном пространстве.

Косинусоидальные моды. Косинусоидальные моды составляют тригонометрический базис, ортогональный на прямоугольнике G с характерными размерами $[-R_x, R_x] \times [-R_y, R_y]$:

$$\psi_{\mathbf{p}}(\mathbf{x}) = \cos\left(\frac{\pi p}{2R_x}x\right)\cdot\cos\left(\frac{\pi l}{2R_y}y\right)\cdot\exp\left[-\frac{\pi\lambda}{16}\left[\left(\frac{m}{R_x}\right)^2 + \left(\frac{l}{R_y}\right)^2\right]z\right]\cdot\exp\left[i\frac{2\pi}{\lambda}z\right]. \tag{6.123}$$

Такими функциями описываются, например, моды для прямоугольного полого волновода с размерами $[-R_x, R_x] \times [-R_y, R_y]$ и идеально проводящими стенками.

6.2. Формирование и селекция мод лазерного излучения с помощью ДОЭ

6.2.1. Постановка задачи синтеза ДОЭ, согласованных с модами лазерного излучения

Как отмечалось в предыдущем разделе, в резонаторах лазеров, линзоподобных средах, волоконных световодах наблюдаются и требуются пучки с различным распределением мощности по модам [7, 15, 18]. В то же время имеются задачи, где требуется селективно работать с одной или определенной группой мод, например, с группой мод с заданным распределением постоянной распространения по модам [19, 20]. При построении волоконно-оптических систем связи возникает актуальная проблема измерения и/или коррекции дифференциального затухания мод, их дифференциальных модовых задержек, вызывающих уширение импульса [18, 19]. В каждом случае, с формальной точки зрения речь идет об измерении или коррекции амплитуды и фазы коэффициентов разложения светового пучка по модам, т.е. об анализе или фильтрации мод. Близкие задачи возникают при работе с переменным во времени световым пучком, используемым для построения волоконно-оптической линии связи с модовым уплотнением каналов [19]. В последнем случае

нужно уметь независимо менять во времени мощности мод, складывать моды в темпе передачи информации и анализировать модовый состав на выходе линии связи (задача анализа светового поля в режиме реального времени).

Сформулируем три основных типа задач селекции мод:

задача модового анализа — разделение многомодового светового пучка на отдельные модовые составляющие с измерением распределения мощности по модам и межмодовых фазовых сдвигов (рис. 6.5);

задача модового синтеза — возбуждение эталонов требуемых мод или многомодовых пучков с требуемым распределением мощности и фаз по модам (рис. 6.6);

модовая фильтрация на основе модовых преобразований, изменение модового состава пучка как по мощности, так и по межмодовым фазовым сдвигам (рис. 6.7).

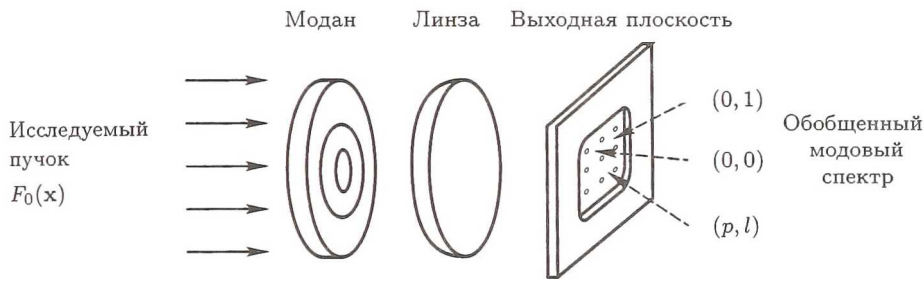


Рис. 6.5. Постановка задачи расчета модана — ДОЭ для измерения обобщенного модового спектра

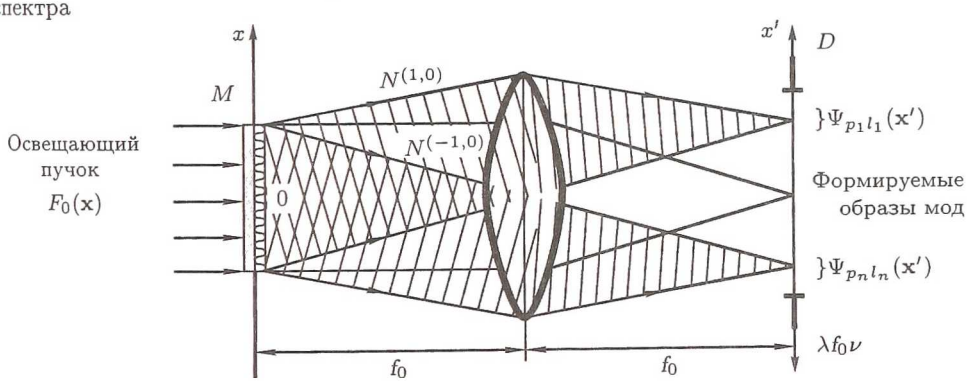


Рис. 6.6. Постановка задачи расчета ДОЭ, формирующего образы мод из освещающего пучка $F_0(\mathbf{x})$ (M — фазовый модан, f_0 — фокусное расстояние тонкой линзы)

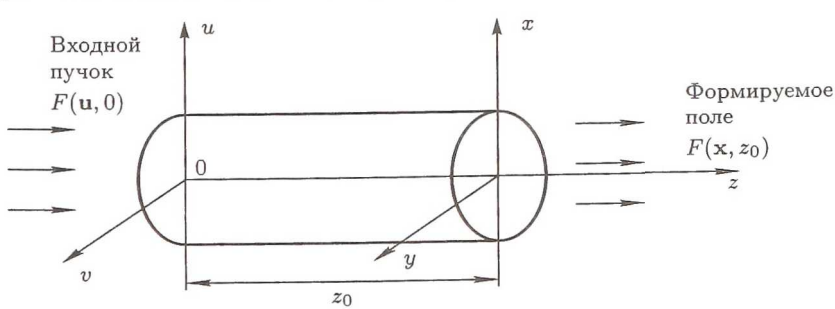


Рис. 6.7. Постановка задачи изменения модового состава пучка

В работе [19] обсуждается важная для оптики проблема: существуют ли моды физически по отдельности в виде некоторых эталонов, или они являются лишь плодом абстракции, т. е. одним из множества ортогональных математических базисов для удобного представления пучков когерентного излучения? В свое время та же проблема обсуждалась по отношению к «продольному» (хроматическому) спектру света: существуют ли монохроматические составляющие света физически или они являются лишь удобной математической формой представления световых колебаний в виде разложения поля по продольным синусоидальным гармоникам. С современной точки зрения на данный вопрос правомерен такой ответ: монохроматические гармоники существуют, т.к. а) они распространяются в свободном пространстве, не изменяя своей продольной структуры и длины волны; б) имеются спектральные приборы, позволяющие селектировать, возбуждать, наблюдать и измерять гармоники в виде спектра.

При распространении в свободном пространстве монохроматические плоские и сферические волны не размываются по спектру цветов, не теряют своей индивидуальной пространственной формы, а лишь приобретают фазовый набег, что и позволило ввести понятие пространственного («продольного») спектра в виде разложения по плоским волнам [21]. Свойство (б) реализуется спектрографами, выполняющими спектральный анализ света, а также монохроматорами, светофильтрами и лазерными источниками, позволяющими выполнить спектральный синтез, т. е. селективно возбуждать отдельные монохроматические составляющие или их группы. Таким образом, спектральные гармоники имеют свойства (а), (б) и, следовательно, существуют физически в виде эталонов продольных мод. Подобно монохроматическим гармоникам, собственные функции оператора распространения — поперечные моды когерентного излучения можно рассматривать как «обобщенные спектральные» компоненты, поскольку аналогично свойству (а) продольного спектра они распространяются в соответствующей среде, не теряя своей индивидуальной формы, а лишь приобретая фазовый набег. Решение проблемы физического существования мод таким образом оказывается полностью связано с возможностью создания приборов, формирующих эталоны модовых пучков, и приборов, измеряющих модовый состав пучков (свойство б). Такие приборы смогут осуществить индивидуальное возбуждение, измерение мощности и обнаружение определенных мод когерентного излучения. Классические оптические элементы не позволяют сформировать эталоны одномодовых пучков и требуемых групп мод когерентного излучения (модовых пакетов). Для формирования световых пучков заданной поперечно-модовой структуры и для анализа поперечно-модовой структуры пучка в [19, 22] был предложен новый тип фазового ДОЭ — модан. В работах [19, 20, 22– 36] описано применение различных методов компьютерной оптики и цифровой голографии для синтеза фазовых моданов, результаты численного моделирования моданов и результаты физических экспериментов по селекции мод и анализу модового состава излучения. Необходимо отметить, что значительная часть работ по этой тематике посвящено гауссовым модам — модам Гаусса–Эрмита и Гаусса–Лагерра. Это объясняется отчасти «универсальностью» мод данного типа, — являясь модами градиентных сред с параболическим профилем, гауссовы пучки с точностью до масштаба сохраняют свою структуру при распространении в свободном пространстве и при прохождении через фурье-каскад [15]. С другой стороны, большое внимание к данному типу мод объясняется возрастающим интересом к так называемым «линзоподобным» средам и их применению в системах обработки изображений, интегральной и волноводной оптике [5, 15]. Последние достижения в области технологии изготовления градиентных оптоволокон позволяют изготовить многомодовые волокна с профилем

весьма близким к идеальному параболическому. Тем не менее, стоит отметить, что значительная часть подходов, развитых для расчета и исследования моданов, согласованных с гауссовыми модами, может быть без труда обобщена на расчет моданов, предназначенных для работы с модами градиентных волноводов произвольного непараболического профиля в тех случаях, когда поляризационными эффектами можно пренебречь. Подробный анализ применимости скалярной теории дифракции для описания градиентных волноводов можно найти, например, в [18]. Однако, следует помнить, что моды градиентных волноводов с произвольным непараболическим профилем не сохраняют своей структуры в свободном пространстве, поэтому для расчета комплексной функции пропускания ДОЭ необходимо в общем случае использовать точное вычисление интеграла Френеля–Кирхгофа. Результаты расчета моданов, согласованных с бесселевыми модами, можно найти в работах [23, 36], а также в главе 7 данной книги.

6.2.2. Методы расчета фазовых моданов

Соображения эффективного использования мощности источника излучения приводит к идее реализации модана в виде прозрачного или хорошо отражающего, т.е. чисто фазового оптического элемента. В [19] было показано, что в случае комплексной функции пропускания энергетическая эффективность амплитудного ДОЭ будет более чем на порядок ниже эффективности фазового ДОЭ с той же функцией пропускания. Как известно [22], тонкий чисто фазовый оптический элемент характеризуется фазовой функцией $\varphi(\mathbf{u})$, где $\mathbf{u} = (u, v)$ — поперечные координаты точки, лежащей на поверхности подложки элемента. Будем считать, что при освещении тонкого ДОЭ пучком с комплексной амплитудой $F_0(\mathbf{u})$, непосредственно за элементом формируется поле с комплексной амплитудой

$$F(\mathbf{u}) = F_0(\mathbf{u}) \exp[i\varphi(\mathbf{u})]. \tag{6.124}$$

6.2.3. Расчет одномодовых моданов

Задача формирования заданного одномодового комплексного распределения с помощью фазового ДОЭ, т.е. путем воздействия только на фазу освещающего пучка в общем случае является некорректной. Некорректность обусловлена свойствами оператора распространения света [10]: сигнал не может быть однозначно определен лишь фазой своего спектра. Поэтому для расчета фазового тонкого ДОЭ, формирующего заданное одномодовое распределение, используется два подхода. Первый подход заключается в расчете фазовой функции элемента, формирующего из освещающего пучка некоторое световое распределение с максимально возможным содержанием полезной моды [37].

Впоследствии этот же подход был использован для расчета моданов, формирующих одномодовое распределение Гаусса–Лагерра и Гаусса–Эрмита [25, 26]. В работе [25] показано, что выбор в качестве фазовой функции модана знаковой функции заданной моды Гаусса–Лагерра позволяет сформировать моду с эффективностью порядка 77–81% в зависимости от конкретного номера моды. В работе [26] продемонстрирована возможность расчета модана Гаусса–Эрмита также путем выбора в качестве фазовой функции знаковой функции формируемой моды. Подробно об этом методе можно прочитать в главе 7 данной книги. Однако подход, заключающийся в подборе фазы элемента, максимизирующей содержание заданной

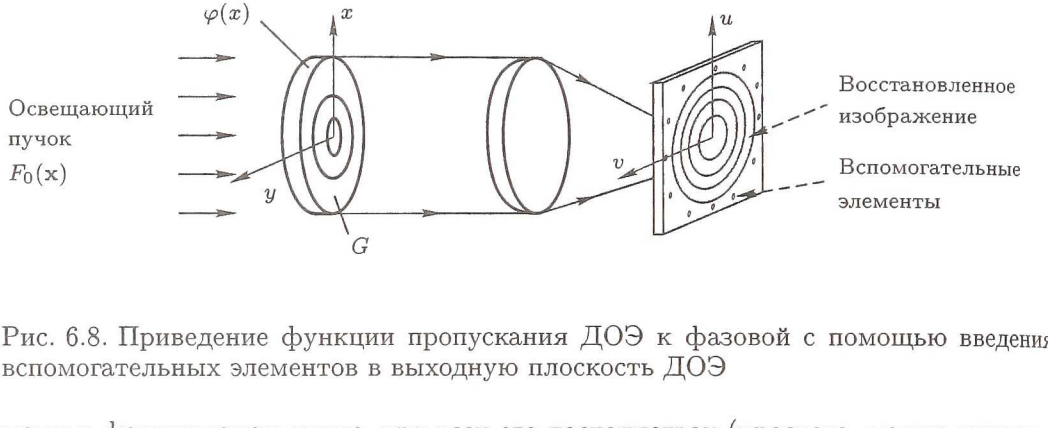


Рис. 6.8. Приведение функции пропускания ДОЭ к фазовой с помощью введения вспомогательных элементов в выходную плоскость ДОЭ

моды в формируемом пучке, при всех его достоинствах (простота, малые вычислительные затраты, возможность расчета бинарного элемента) неизбежно приводит к уширению импульса, если сформированный квазиодномодовый пучок предназначен для возбуждения волновода. Это объясняется тем, что паразитные моды, присутствующие в пучке, в общем случае имеют значения постоянной распространения отличные от постоянной распространения полезной моды. Кроме того, наличие паразитных мод недопустимо по определению в задачах формирования эталонных модовых пучков. Поэтому основу другого подхода к решению задачи формирования одномодового распределения составляет идея введения вспомогательных элементов для приведения комплексной функции пропускания элемента к фазовой [19]. Будучи пространственно разделены с полезным изображением, вспомогательные элементы не влияют на качество его восстановления, а лишь снижают долю энергии, идущую на формирование полезного изображения (рис. 6.8).

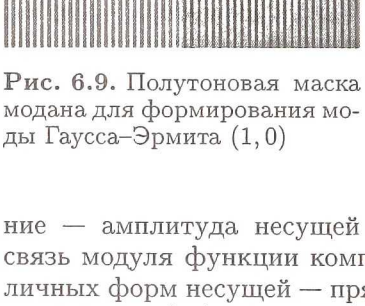


Рис. 6.9. Полутоновая маска модана для формирования моды Гаусса–Эрмита $(1,0)$

Известна идея формирования вспомогательных элементов в виде паразитных дифракционных порядков с помощью введения модулированной несущей в фазу элемента [19, 38]. Форма несущей определяется выбором конкретного метода кодирования — прямоугольно-импульсная несущая соответствует методу синтетических коэффициентов [39], косинусидальная несущая соответствует методу Кирка–Джонса [38]. Амплитуда несущей выбирается локально зависящей от функции пропускания–в точках, где амплитуда функции пропускания максимальна, она равна нулю, в точках, где амплитуда функции комплексного пропускания имеет минимальное значение — амплитуда несущей максимальна. Уравнения, описывающие нелинейную связь модуля функции комплексного пропускания с амплитудой несущей для различных форм несущей — прямоугольно-импульсной, синусоидальной, пилообразной приведены в [19]. На рис. 6.9 представлена полутоновая маска модана, предназначенного для формирования моды Гаусса–Эрмита (1,0) в нулевом порядке дифракции. Для кодирования был использован обобщенный метод Кирка–Джонса [19]. Анализ рис. 6.9 дает общее представление об элементах, рассчитанных методом [19], — дифракционных решетках специального вида. Несколько забегая вперед, отметим, что

моданы, рассчитанные с помощью введения несущей, могут быть использованы как для формирования заданных модовых пучков, так и для анализа модовой структуры пучка [19].

Достоинством такого метода расчета моданов является возможность аналитического расчета функции пропускания элемента, однако далеко не всегда выбор вспомогательных элементов в виде паразитных дифракционных порядков является оптимальным с точки зрения соотношения качества формирования моды (содержания эталонной моды (6.9) в формируемом пучке) и энергетической эффективности ДОЭ, т.е. соотношения долей энергии, приходящихся на изображение моды и паразитные дифракционные порядки. Применение метода [19] позволяет создавать заданное одномодовое распределение с высокой точностью. Практически, ошибка формирования заданного комплексного распределения в данном случае определяется лишь погрешностями квантования фазы и технологическими погрешностями при изготовлении ДОЭ. Однако, появление паразитных дифракционных порядков, вызванное введением опорной волны в фазу, приводит к низкой дифракционной эффективности элемента. Это объясняется сильным отличием структуры амплитудного распределения формируемой моды от типичного амплитудного распределения освещающего пучка, гауссова или равномерного. На рис. 6.10 приведен результат моделирования работы модана, фазовая маска которого приведена на рис. 6.9. Отчетливо виден рабочий дифракционный порядок (в центре) и паразитные дифракционные порядки. Из анализа рис. 6.10 можно сделать вывод, что доля энергии, идущей в паразитные порядки, превышает долю энергии, идущей на формирование полезного изображения.



Рис. 6.10. Амплитуда, сформированная в выходной плоскости моданом, фазовая маска которого приведена на рис. 6.9

Для фазового модана, предназначенного для формирования моды Гаусса–Эрмита (1,0) из гауссова пучка, верхняя оценка энергетической эффективности в случае использования обобщенного метода Кирка–Джонса [19] составляет всего около 33%. Низкая дифракционная эффективность элементов, рассчитанных с помощью введения несущей в фазу, является одним из основных факторов, сдерживающих их применение на практике.

Понятно, что вопрос о приоритетности точности формирования моды или энергетической эффективности модана решается каждый раз исходя из специфики конкретной задачи. Напрашивается мысль о необходимости наличия «инструментария» численных методов, позволяющих находить необходимый компромисс в каждом конкретном случае.

6.2.4. Построение итеративной процедуры расчета одномодового модана

В работах [27–31] описано применение итеративных процедур к расчету одномодовых фазовых ДОЭ, формирующих заданное одномодовое распределение. Итеративная процедура в этом случае используется для поиска вспомогательных элементов, обеспечивающих минимальную разницу между амплитудой спектра формируемого изображения и амплитудой освещающего пучка. Представлены различные подходы к построению итеративной процедуры: проекционный [27, 28], градиентный [31]. Понятно, что для решения этой задачи также могут быть использованы и другие методы теории оптимизации.

Итеративная процедура расчета одномодового модана, описанная в [27, 28], была построена на базе математического метода обобщенных проекций. Метод обобщенных проекций был разработан и строго обоснован для восстановления функции по имеющейся априорной информации [40]. Суть метода — в последовательном построении математических проекций на замкнутые множества, соответствующие ограничениям, наложенным на функцию и ее спектр. Введем несколько основных понятий. Для любого замкнутого множества G_j некоторая функция $p_j = \hat{\mathrm{P}}_j q$ может быть названа проекцией q на G_j если $p_j \in G_j$ и

$$\|p_j - q\| = \min_{f \in G_j} \|f - q\|. \tag{6.125}$$

В качестве нормы $\|\,.\,\|$ обычно используется норма в пространстве $L_{2,2}$ [40]:

$$\|F(\mathrm{x})\|^2_{L_{2,2}} = \iint\limits_{\infty} F(\mathrm{x}) F^*(\mathrm{x})\, \mathrm{d}^2\mathrm{x}. \tag{6.126}$$

Оператор $\hat{\mathrm{P}}_j$ будем называть проекционным оператором (или проектором) на множество G_j. Будем считать, что известны m априорных свойств функции F, каждое из которых относит ее к одному из m замкнутых множеств $G_1 \ldots G_m$. Тогда решением задачи восстановления будет являться функция $F \in G_0$, где

$$G_0 = \bigcap_{j=0}^{m} G_j. \tag{6.127}$$

Назовем G_0 множеством допустимых решений. Отметим, что в случае, когда G_0 является пустым множеством, решения существовать не будет; в случае, если G_0 содержит только один элемент, решение единственно, и если G_0 содержит более одного элемента, любой из них может быть выбран в качестве решения поставленной задачи восстановления. В работе [40] для нахождения функции F было предложено использовать рекурсивную процедуру следующего вида:

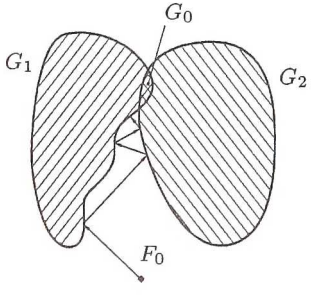


Рис. 6.11. Проекционное восстановление функции

$$F_{n+1} = \hat{\mathrm{T}}_{1,n} \hat{\mathrm{T}}_{2,n} \ldots \hat{\mathrm{T}}_{m,n} F_n, \tag{6.128}$$

где F_0 — выбранное начальное приближение,

$$\hat{\mathrm{T}}_{j,n} = \hat{\mathrm{E}} + \alpha_{j,n} \left(\hat{\mathrm{P}}_j - \hat{\mathrm{E}} \right), \quad j = 1, 2, \ldots m, \tag{6.129}$$

$\hat{\mathrm{E}}$ — единичный оператор и $\alpha_{j,n}$, $0 \leqslant \alpha_{j,n} \leqslant 1$, — некоторые параметры, вводимые в общем случае для повышения скорости поиска точки F, называемые параметрами релаксации. В общем случае при $m = 2$ алгоритм (6.128) обладает лишь свойством уменьшения ошибки суммарного расстояния (ОСР) [40], то есть функционал вида

$$J(F_n) = \left\| \hat{\mathrm{P}}_1 F_n - F_n \right\| + \left\| \hat{\mathrm{P}}_2 F_n - F_n \right\| \tag{6.130}$$

обладает следующим свойством:

$$J(F_{n+1}) \leqslant J(F_n). \tag{6.131}$$

Геометрическая интерпретация проекционного подхода к восстановлению функции по двум наложенным на нее ограничениям, соответствующим принадлежности замкнутым множествам, приведена на рис. 6.11.

Для того чтобы построить итеративную процедуру расчета фазовых ДОЭ на основе метода обобщенных проекций, необходимо определить множества G_1 и G_2 и, соответственно, построить проекционные операторы $\hat{\mathrm{P}}_1$ и $\hat{\mathrm{P}}_2$. Обычно в качестве множеств G_1 и G_2 выбираются замкнутые множества, соответствующие ограничениям, налагаемым на распределение $F(\mathbf{x})$, формируемое ДОЭ в плоскости наблюдения, и его спектр $F_1(\mathbf{u})$. Построение проекционных операторов не вызывает в большинстве случаев особых затруднений [27, 28].

Рассмотрим построение итеративной процедуры расчета однопучкового модана. В случае выполнения условия параксиального приближения и условия дальней зоны, преобразование Френеля–Кирхгофа сводится к преобразованию Фурье [10]:

$$\begin{cases} F(\mathbf{x}) = \Im\left(F_1(\mathbf{u})\right), \\ F_1(\mathbf{u}) = \Im^{-1}\left(F(\mathbf{x})\right). \end{cases} \tag{6.132}$$

При соблюдении упомянутых условий можно сформулировать задачу синтеза фазового ДОЭ, формирующего заданное комплексное одномодовое распределение $\psi_p(x)$ в области M плоскости наблюдения, как задачу поиска такой функции $F(\mathbf{x})$, которая удовлетворяя условию $\iint\limits_M |F(\mathbf{x}) - \psi_{\mathbf{p}}(\mathbf{x})|^2 \, \mathrm{d}^2\mathbf{x} = 0$, имела бы амплитуду фурье-спектра $C\sqrt{I_0(\mathbf{u})}$, где $I_0(\mathbf{u})$ — распределение интенсивности освещающего пучка, C — константа, выбираемая из условия сохранения энергии. Ограничению на амплитуду фурье-спектра искомой функции соответствует множество G_1 функций $F(\mathbf{x})$ с амплитудой фурье-спектра

$$|F_1(\mathbf{u})| = \begin{cases} C\sqrt{I_0(\mathbf{u})}, & (\mathbf{u}) \in D \\ 0 \quad , & (\mathbf{u}) \notin D \end{cases} \tag{6.133}$$

с соответствующим проектором [41]

$$\hat{\mathrm{P}}_1 F = \Im\left(C\sqrt{I_0(\mathbf{u})}\exp\left(i \cdot \arg\left[F_1(\mathbf{u})\right]\right)\right). \tag{6.134}$$

В качестве множества, описывающего ограничение, наложенное на распределение в фокальной плоскости, естественно определить замкнутое множество

$$G_2 = \left\{ F : \iint\limits_{\mathbf{x}\in M} |F(\mathbf{x}) - \psi_{\mathbf{p}}(\mathbf{x})|^2 \, \mathrm{d}^2\mathbf{x} = 0; \quad \iint\limits_{\mathbf{x}\notin M} |F(\mathbf{x})|^2 \, \mathrm{d}^2\mathbf{x} = 0 \right\}. \tag{6.135}$$

Оператор

$$\hat{\mathrm{P}}_2 F = \begin{cases} \psi_{\mathbf{p}}(\mathbf{x}), & (\mathbf{x}) \in M, \\ 0 \quad , & (\mathbf{x}) \notin M \end{cases} \tag{6.136}$$

является проекционным на множество G_2. Однако в случае, если амплитуда освещающего пучка отлична от амплитуды фурье-образа восстанавливаемой моды, множество допустимых решений будет пусто. Стагнация итеративной процедуры, таким образом, начнется в этом случае с первых же итераций. Поэтому предлагается [27, 28] использовать построение проекции не на множество G_2, а на множество G_2':

$$G_2' = \left\{ F : \iint\limits_{\mathbf{x}\in M} |\psi_{\mathbf{p}}(\mathbf{x}) - F(\mathbf{x})|^2 \, \mathrm{d}^2\mathbf{x} = 0; \quad \iint\limits_{r\geqslant R_2} |F(\mathbf{x})|^2 \, \mathrm{d}^2\mathbf{x} = 0 \right\}, \tag{6.137}$$

где $r = \sqrt{x^2 + y^2}$, $\mathbf{x} = (x, y)$, R_2 — радиус некоторой области M' такой, что для любого $\mathbf{x} \in M$ верно $\mathbf{x} \in M'$. Очевидно, что $G_2 \subset G_2'$. Оператор

$$\hat{\mathrm{P}}_2' F(x) = \begin{cases} \psi_{\mathbf{p}}(\mathbf{x}), & (\mathbf{x}) \in M, \\ F(\mathbf{x}), & (\mathbf{x}) \in M'/M, \\ 0, & r \geqslant R_2, \end{cases} \tag{6.138}$$

является проекционным на множество G_2'. Значения комплексной амплитуды в точках внутри вспомогательной области M'/M являются свободными параметрами, необходимыми для работы итеративной процедуры. Доля энергии освещающего пучка, дифрагируемая в точки вспомогательной области, является «платой» за приведение амплитуды фурье-образа формируемого изображения к амплитуде освещающего пучка. Интересно отметить, что размер области M' определяет степень «размытия» множества G_2. Итеративная процедура, использующая операторы (6.134), (6.138), обладает свойством уменьшения ошибки суммарного расстояния до множеств G_1, G_2'. Второе слагаемое функционала (6.130) в этом случае описывает отклонение формируемого комплексного распределения $F(\mathbf{x})$ от эталонного одномодового распределения $\psi_{\mathbf{p}}(\mathbf{x})$ при $(x, y) \in M$, первое — отклонение амплитуды $|F_1(\mathbf{u})|$ фурье-спектра от распределения амплитуды освещающего пучка $C\sqrt{I_0(\mathbf{u})}$ на апертуре D рассчитываемого элемента на данной итерации. Очевидным образом построенный алгоритм может быть применен к расчету ДОЭ, формирующего в заданной области M требуемое распределение комплексной амплитуды произвольного вида. Отметим, что выбор размера области M' соответствует поиску необходимого компромисса между энергетической эффективностью и точностью формирования заданного одномодового распределения. В результате, рассчитанный ДОЭ с заданной эффективностью формирует пучок, состоящий из полезной моды и допустимым вкладом «паразитных» мод. Если множество точек области M'/M пусто, то результат поиска фазовой функции ДОЭ будет совпадать с найденным в [25, 26] решением в виде знаковой функции моды. Достоинством предлагаемого подхода к расчету однопучковых моданов является возможность построения высокоэффективных итеративных процедур, если формируемое одномодовое распределение обладает радиальной симметрией [42], что позволяет свести двумерный пересчет поля из плоскости ДОЭ в плоскость формирования изображения к одномерному. Таким случаем является, например, задача формирования мод Гаусса–Лагерра, рассмотренная в [42], или задача формирования мод Бесселя из главы 7 данной книги.

6.2.5. Быстрый расчет ДОЭ, формирующего заданное одномодовое распределение радиальных мод

В [43] показано, что если формируемое изображение $F(x) = F(r)$ и освещающий пучок $I_0(\mathbf{u}) = I_0(\rho)$, $r = \sqrt{x^2 + y^2}$, $\rho = \sqrt{u^2 + v^2}$, $\mathbf{u} = (u, v)$, $\mathbf{x} = (x, y)$ обладают радиальной симметрией, то преобразование (6.132) может быть сведено к преобразованию Ханкеля 0-го порядка:

$$\begin{cases} F(r) = \mathcal{H}_0(F_1(\rho)), \\ F_1(\rho) = \mathcal{H}_0(F(r)). \end{cases} \tag{6.139}$$

В этом случае задача расчета ДОЭ ставится следующим образом (рис. 6.12): необходимо найти функцию $F(r)$, которая при условии $\iint\limits_{r<R_1} |F(r) - \psi_{\mathbf{p}}(r)|^2 \, r \, \mathrm{d}r = 0$,

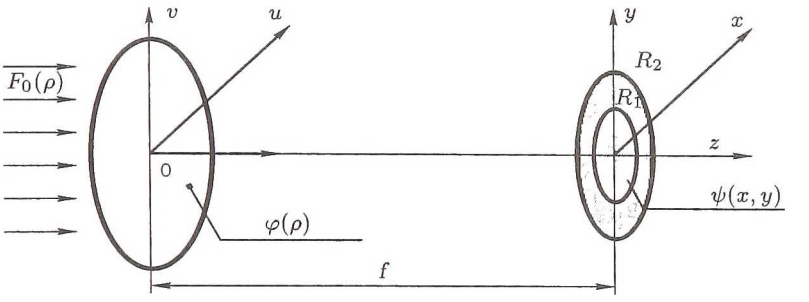


Рис. 6.12. Постановка задачи расчета модана, формирующего моду Гаусса–Лагерра

имела бы амплитуду фурье-спектра $C\sqrt{I_0(\rho)}$, R_1 — радиус полезной области (например, радиус сердцевины оптического градиентного волновода). В работе [44] показано, что вычисление преобразования Ханкеля (6.139) легко сводится к вычислению трех одномерных преобразований Фурье.

В предположении, что «вспомогательная» область также обладает радиальной симметрией, итеративная процедура [27] существенно упрощается и требует вычисления лишь трех одномерных преобразований Фурье на итерацию. Интересно отметить, что обладая эффективной процедурой расчета радиально-симметричных моданов, можно получить распределение, описываемое любой обобщенной модой Гаусса–Лагерра, формируя винтовую «добавку» сответствующим фазовым фильтром, выбираемым из следующих соображений. Для формирования в дальней зоне Фраунгофера обобщенной моды Гаусса–Лагерра [19]

$$\psi_{\mathbf{p}}^{\pm}(r,\theta)=\frac{2\sqrt{\pi(p-l)!}}{\sigma\sqrt{(p!)^3}}\left(\frac{r}{\sigma}\right)^l\exp\left[-\frac{r^2}{2\sigma^2}\right]L_p^l\left(\frac{r^2}{\sigma^2}\right)\exp\left(\pm il\theta\right)=\tilde{\psi}_{\mathbf{p}}(r)\exp\left(\pm il\theta\right),\tag{6.140}$$

где σ — фундаментальный радиус моды, θ — угол в полярных координатах, L_p^l — полином Лагерра, $\mathbf{p}=(p,l)$ — порядковые номера мод, необходимо на входе иметь следующее распределение комплексной амплитуды:

$$\Psi(\rho,\nu)=\int_0^{2\pi}\int_0^{\infty}\psi_{\mathbf{p}}^{\pm}(r,\theta)\exp\left[-i\cdot 2pr\cos(\theta-\nu)\right]r\,\mathrm{d}r\,\mathrm{d}\theta=$$

$$=\exp\left(\pm il\nu\right)\int_0^{\infty}\tilde{\psi}_{\mathbf{p}}(r)\left[\int_0^{2\pi}\exp\left(\pm ilt\right)\exp\left[-i2\pi r\rho\cos(t)\right]\,\mathrm{d}t\right]r\,\mathrm{d}r.\tag{6.141}$$

Используя справочный интеграл [43]

$$J_l(x)=\frac{i^l}{2\pi}\int_0^{2\pi}\exp\left(\pm ilt\right)\exp\left[-ix\cos(t)\right]\,\mathrm{d}t,\tag{6.142}$$

приведем (6.141) к следующему виду:

$$\Psi(\rho,\nu)=\exp\left(\pm il\nu\right)\frac{2\pi}{i^l}\int_0^{\infty}\tilde{\psi}_{\mathbf{p}}(r)J_l(r\rho)r\,\mathrm{d}r=\tilde{\Psi}(\rho)\exp\left(\pm il\nu\right).\tag{6.143}$$

Для формирования функции (6.141) нам нужно рассчитать только радиальный множитель искомой функции:

$$\tilde{\Psi}(\rho) = \frac{2\pi}{i^l} \int_0^\infty \tilde{\psi}_{\mathbf{p}}(r) J_l(r\rho) r \, \mathrm{d}r, \tag{6.144}$$

который представляет собой преобразование Ханкеля l-го порядка. Отметим также, что в случае, если формируемое изображение содержит только нулевые значения в точках вблизи оптической оси, пересчет поля с апертуры в фокальную плоскость и обратно может осуществляться, используя вычисление всего лишь двух одномерных преобразований Фурье на итерацию [45]. Примером такого распределения являются обобщенные моды Гаусса–Лагерра с равными значениями первого и второго модовых номеров.

Пример 6.1. Итеративный расчет одномодового модана. В работе [46] исследовалось применение итеративной процедуры с проекторами (6.138), (6.134) к синтезу ДОЭ, формирующих из гауссова пучка с распределением интенсивности

$$I_0(\mathbf{u}) = \exp\left[-\frac{2\mathbf{u}^2}{\sigma_{00}^2}\right], \tag{6.145}$$

эталонное распределение, описываемое модой лазерного излучения $\psi_{\mathbf{p}}(\mathbf{x})$, внутри области M с радиусом R_1. В качестве эталонного распределения выбиралось распределение, описываемое модовыми функциями Гаусса–Эрмита $(1,0)$ и Гаусса–Лагерра $(1,0)$. В качестве начального приближения выбиралась функция

$$F_0(\mathbf{x}) = \begin{cases} \psi_{\mathbf{p}}(\mathbf{x}), & \mathbf{x} \in M, \\ \tilde{\mu}\chi(\mathbf{x}), & \mathbf{x} \in M'/M, \\ 0, & r \geqslant R_2, \end{cases} \tag{6.146}$$

где $\chi(\mathbf{x})$ — случайная вещественная функция, равномерно распределенная на отрезке $\left[0, \underset{\mathbf{x}\in M}{max}|\psi_{\mathbf{p}}(\mathbf{x})|\right]$, параметр $\tilde{\mu}$ определяет соотношение долей энергии, приходящейся на полезную и вспомогательную области начального приближения. Введение случайного предыскажения во вспомогательную область приводит к появлению свободных параметров, необходимых для работы итеративной процедуры. На рис. 6.13 приведена амплитуда начального распределения $F_0(\mathbf{x})$, содержащего эталонное распределение в полезной области и случайное предыскажение во вспомогательной. На рис. 6.14 приведена фазовая маска итеративно рассчитанного модана Гаусса–Эрмита $(1,0)$. На рис. 6.15 приведена амплитуда изображения, сформированного рассчитанным элементом (результат численного моделирования).

Были выбраны следующие критерии качества формирования моды [27]: энергетическая эффективность

$$E_m = \frac{\iint\limits_M |F_m(\mathbf{x})|^2 \, \mathrm{d}^2\mathbf{x}}{\iint\limits_D I_0(\mathbf{u}) \, \mathrm{d}^2\mathbf{u}}, \tag{6.147}$$

где $F_m(\mathbf{x})$ комплексное распределение, формируемое рассчитанным элементом на

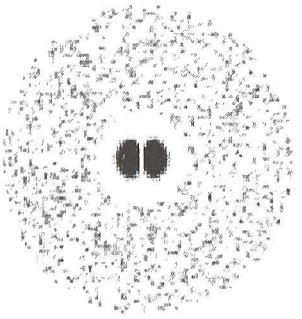


Рис. 6.13. Амплитуда начального распределения

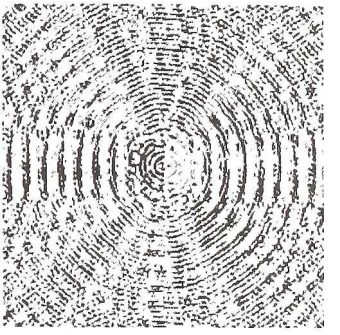


Рис. 6.14. Фазовая маска рассчитанного модана

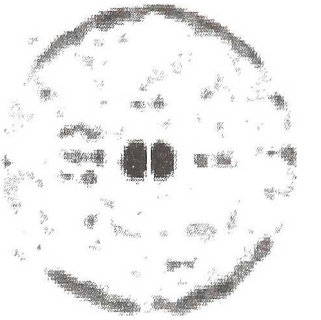


Рис. 6.15. Результат моделирования рассчитанного модана

m-ом шаге итеративной процедуры, и удельная мощность заданной моды в области M

$$\zeta_m = \left| C_0 \iint\limits_M F_m(\mathbf{x})_p^* \psi(\mathbf{x})\, \mathrm{d}^2\mathbf{x} \right|^2, \tag{6.148}$$

характеризующая долю энергии формируемой моды в данной области. C_0 выбиралась из условия ортонормированности базиса модовых функций $\iint\limits_\infty \psi_{\mathbf{p}}(\mathbf{x})\, \psi_{\mathbf{p}}^*(\mathbf{x})\, \mathrm{d}^2\mathbf{x} = 1$

$$C_0 = \frac{1}{\sqrt{\iint\limits_M |F_m(\mathbf{x})|^2\, \mathrm{d}^2\mathbf{x}}}. \tag{6.149}$$

Так как амплитуда модовых функций достаточно быстро спадает при удалении от точки $(0,0)$, при соответствующем выборе размера полезной области M оправдан переход к интегрированию по M в (6.148). В силу ортогональности взаимодействие мод происходит так, что при интегрировании (6.148) паразитные моды уничтожаются и величина ζ определяет долю энергии в области M, приходящуюся на формируемую моду $\psi_{\mathbf{p}}(\mathbf{x})$. Для вычислительного эксперимента были использованы следующие параметры: радиус апертуры элемента $a = 10$ мм, гауссов параметр освещающего пучка $\sigma_{00} = 3{,}5$ мм, длина волны освещающего пучка $\lambda = 0{,}6328$ мкм, фокусное расстояние фурье-линзы $f = 50$ мм, внешний радиус полезной области $R_1 = 30$ мкм, внешний радиус вспомогательной области $R_2 = 95$ мкм, фундаментальный радиус формируемой моды в выходной фокальной плоскости линзы $\sigma = 5{,}6$ мкм, число отсчетов матрицы $N_2 = 128 \times 128$. В работе [46] проводилось сравнение характеристик моданов, рассчитанных с помощью предложенной итеративной процедуры с характеристиками моданов, рассчитанных с помощью метода Кирка–Джонса. Согласно [19], энергетическая эффективность однопучкового модана, рассчитанного обобщенным методом Кирка–Джонса, может быть найдена по формуле

$$E_k = g\tilde{C}^2, \tag{6.150}$$

где $g \leqslant 1$ определяется типом базовой дифракционной решетки; в случае выбора синусоидальной или прямоугольной несущей и выбора в качестве рабочего 0-го

Таблица 6.1. Результаты вычислительного эксперимента по итеративному расчету однопучковых моданов (в скобках даны значения энергетической эффективности для метода Кирка–Джонса)

Критерий	Мода Гаусса–Эрмита $(1,0)$	Мода Гаусса–Лагерра $(1,0)$
Эффективность $E, \%$	53,3(26,3)	59,4(26,4)
Удельная мощность заданной моды $\zeta, \%$	95,2	98,6

дифракционного порядка, $g = 1$,

$$\tilde{C}^{-1} = \max_{\mathbf{u}\in D} \left| \frac{\psi_{\mathbf{p}}(\mathbf{u})/\sqrt{E_1}}{\sqrt{I_0(\mathbf{u})}/\sqrt{E_0}} \right|, \tag{6.151}$$

$$E_0 = \iint\limits_G I_0(\mathbf{u})\, \mathrm{d}^2\mathbf{u}, \tag{6.152}$$

$$E_1 = \iint\limits_G |\psi_{\mathbf{p}}(\mathbf{u})|^2\, \mathrm{d}^2\mathbf{u}, \tag{6.153}$$

$\psi_{\mathbf{p}}(\mathbf{u}) = F^{-1}(\psi_{\mathbf{p}}(\mathbf{x}))$ — формируемое одномодовое распределение в плоскости элемента. Результаты вычислительного эксперимента приведены в табл. 6.1.

Как видно из табл. 6.1, предлагаемый метод позволяет рассчитывать однопучковые моданы с энергетической эффективностью более чем в два раза большей, чем обобщенный метод Кирка–Джонса. При этом формируемое квазиодномодовое распределение на 95–98% состоит из заданной моды.

Пример 6.2. Расчет одномодового модана, формирующего моду Гаусса–Лагерра. В работе [42] производился итеративный расчет ДОЭ, формирующего из гауссова пучка с распределением интенсивности $I_0(\rho) = \exp\left[-2\rho^2/\sigma_{00}^2\right]$ эталонное распределение, описываемое модой лазерного излучения $\psi_{\mathbf{p}}(r)$, внутри области M с радиусом R_1. В качестве эталонных распределений выбирались обобщенные моды Гаусса–Лагерра $(1,0)$, $(4,0)$, $(2,2)$. В начальном приближении выбиралась функция

$$F_0(r) = \begin{cases} \psi_{\mathbf{p}}(r), & r \leqslant R_1, \\ \tilde{\mu}\chi(r), & R_1 < r < R_2, \\ 0, & r \geqslant R_2. \end{cases}$$

В качестве $\chi(r)$ выбиралась случайная вещественнозначная функция, равномерно распределенная на отрезке $[0, \max|\psi_{\mathbf{p}}(r)|]$, параметр $\tilde{\mu}$ определяет соотношение долей энергии, приходящихся на «полезную» и «вспомогательную» области начального приближения. Для пересчета поля из плоскости элемента в фокальную

Таблица 6.2.

Номера моды	σ, mm	$E, \%$	$\zeta, \%$
$(p,l) = (1,0)$	3	40,1	97,5
$(p,l) = (2,0)$	3,3	35,4	96,4
$(p,l) = (2,2)$	3,5	62,8	96,7
$(p,l) = (4,0)$	3,5	33,7	96,4

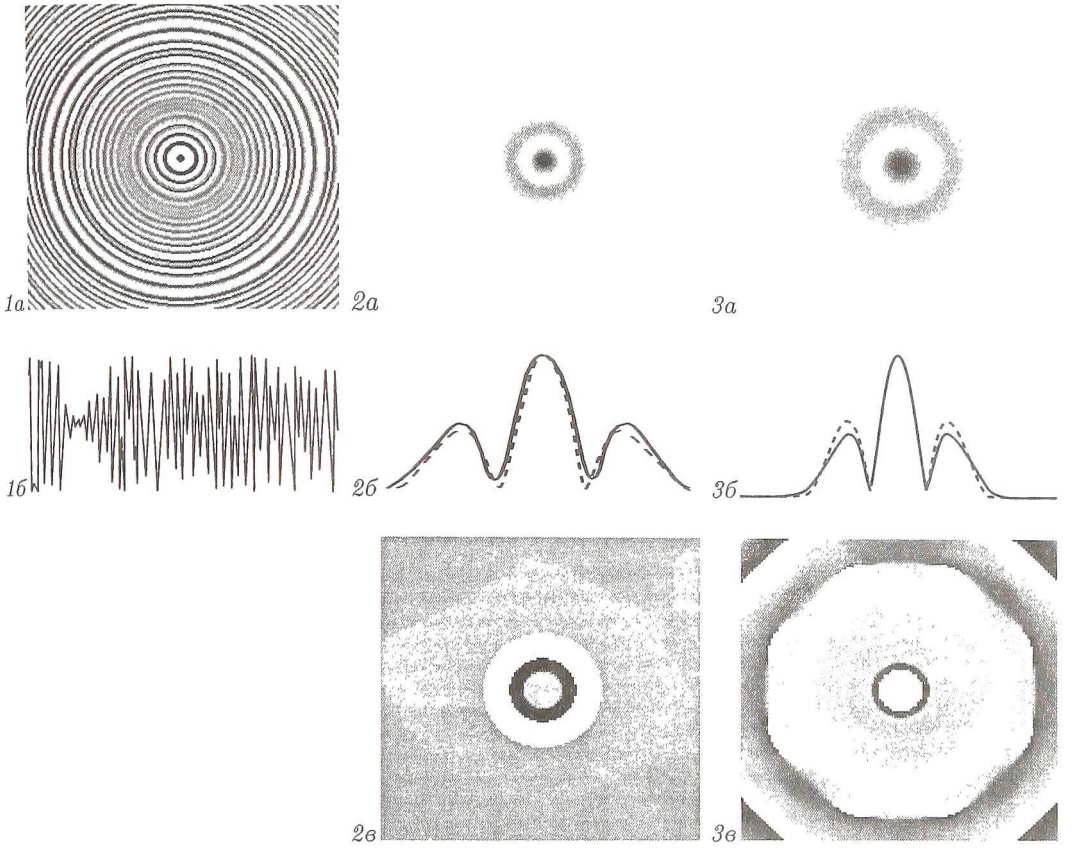


Рис. 6.16. Формирование моды Гаусса–Лагерра $(1,0)$: распределение фазы ДОЭ (*1а*), ее радиальное сечение (*1б*), сформированное амплитудное распределение в полезной области (*2а*), его сечение (штриховой линией показано сечение эталонной моды) (*2б*), сформированное фазовое распределение в полезной области (*2в*) (диапазон значений $[0, 2\pi]$), в третьем столбце приведены результаты моделирования прохождения сформированной моды через фурье-каскад

плоскость и обратно использовалось процедура быстрого вычисления преобразования Ханкеля. Элементы рассчитывались как добавка к линзе со следующими значениями физических параметров: фокусное расстояние $f = 50$ мм, длина волны $\lambda = 0{,}6328$ мкм, радиус апертуры ДОЭ $R_{ap} = 10$ мм, фундаментальный радиус формируемой моды $\omega = 5{,}6$ мкм, радиус освещающего гауссова пучка $\sigma = 3$–$3{,}5$ мм, количество итераций $N = 10$. Результаты расчета и моделирования ДОЭ приведены в табл. 6.2 и на рис. 6.16 и рис. 6.17. Результаты вычислительного эксперимента показали, что предложенная итеративная процедура позволяет рассчитывать моданы, формирующие обобщенные моды Гаусса–Лагерра с энергетической эффективностью в 2–3 раза большей, чем моданы, рассчитанные с помощью метода Кирка–Джонса. При этом содержание заданной моды в формируемом пучке составило свыше 95% (табл. 6.2).

Интересно отметить «вращение» моды $(2,2)$ (рис. 6.17) при прохождении Фурье-каскада, что объясняется наличием фазовой винтовой добавки. Подробнее о замечательных свойствах вращающихся пучков или пучков с периодическими свойствами можно прочитать в работе [47] или главе 7 данной книги.

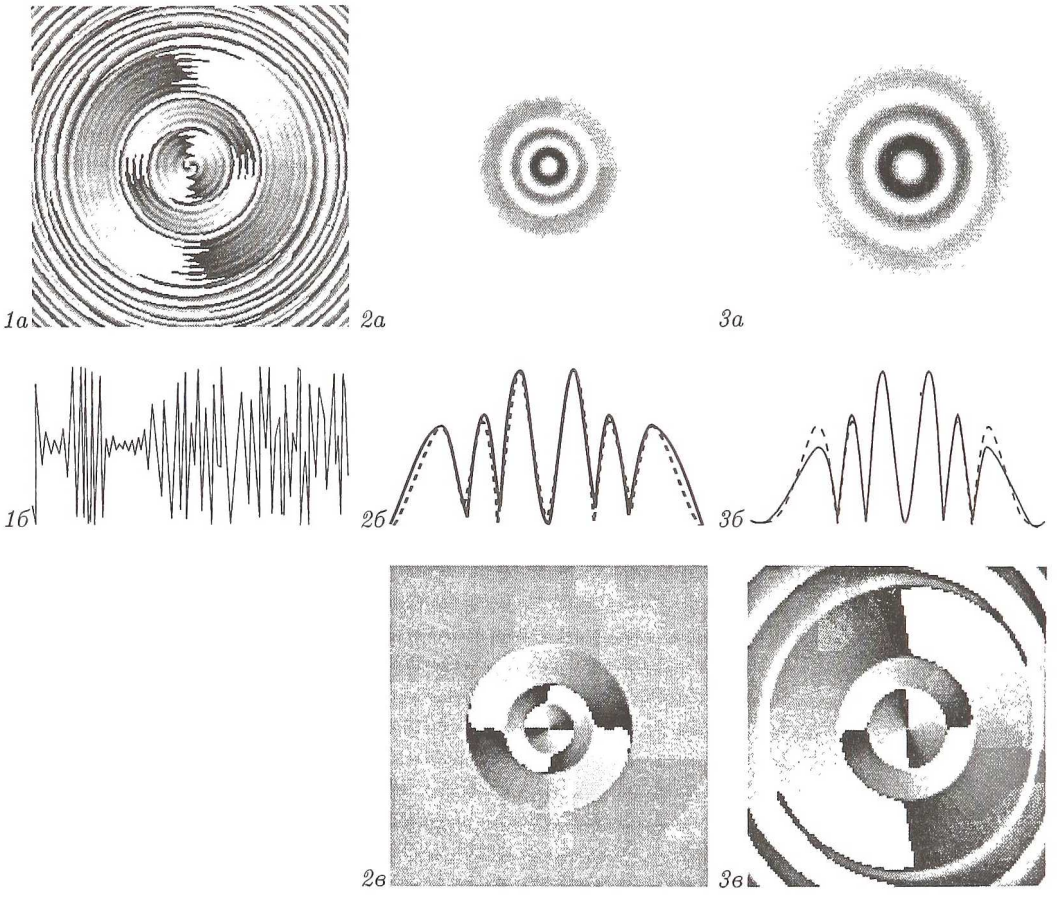


Рис. 6.17. Формирование моды Гаусса–Лагерра $(2, 2)$: распределение фазы ДОЭ (*1а*), ее радиальное сечение (*1б*), сформированное амплитудное распределение в полезной области (*2а*), его сечение (штриховой линией показано сечение эталонной моды) (*2б*), сформированное фазовое распределение в полезной области (*2в*) (диапазон значений $[0, 2\pi]$); в третьем столбце приведены результаты моделирования прохождения сформированной моды через фурье-каскад

6.2.6. Расчет ДОЭ, формирующих группу мод лазерного излучения

Сечение пучка, содержащего группу мод лазерного излучения с заданными весами, также описывается комплексной функцией соответствующего вида и, следовательно, для расчета фазового ДОЭ, формирующего группу мод, могут быть применены методы кодирования, основанные на введении несущей в фазу элемента [19]. Также как при расчете ДОЭ, предназначенного для формирования заданного одномодового распределения (6.2.3), применение аналитических методов кодирования, основанных на внесении несущей в фазу элемента, к расчету модана, формирующего заданную группу мод, приводит в общем случае к низкой дифракционной эффективности элемента.

Однако, если необходимо сформировать одновременно несколько модовых пучков одним элементом, и соотношение начальных фазовых межмодовых задержек при этом не является значимым, то произвол выбора этих задержек дает дополнительную

свободу для расчета фазовой функции элемента без введения вспомогательных элементов и, следовательно, без значительных энергетических потерь. Возникает заманчивая мысль о построении итеративной процедуры расчета многомодового модана, использующей в качестве свободных параметров в плоскости элемента отсчеты фазовой функции модана, а в выходной плоскости — взаимные фазовые сдвиги формируемых мод [24]. Напрашивается аналогия с итеративным расчетом ДОЭ, предназначенным для формирования заданного распределения интенсивности, с помощью процедуры типа Фьенапа [48]. Отличие задачи расчета модана от задачи расчета ДОЭ, предназначенного для формирования заданного распределения интенсивности, состоит лишь в том, что в последнем случае в качестве свободных параметров в выходной плоскости используются отсчеты фазы формируемого светового распределения.

В работах [23, 24] рассмотрен расчет фазовых моданов, согласованных с гауссовыми лазерными модами и формирующих световые пучки в виде суперпозиции мод Гаусса–Эрмита или мод Гаусса–Лагерра с заданными амплитудами модовых коэффициентов.

Рассмотрим построение итеративной процедуры расчета многопучковых моданов Гаусса–Эрмита. Легко показать [19, 23], что существует частное решение параксиального уравнения в одномерном случае

$$2ik\frac{\partial F(u,z)}{\partial z}+\frac{\partial^2 F(u,z)}{\partial^2 u}=0, \tag{6.154}$$

где z — направление распространения пучка, которое может быть представлено в виде

$$F_n(u,z)=a^{-n-1}\exp\left(-(u/a)^2\right)H_n(u/a), \tag{6.155}$$

где $a^2(z)=a_0^2+2iz/k$, a_0^2 — радиус пучка при $z=0$, $H_n(u)$ — полином Эрмита. Следовательно, общее решение уравнения (6.154) можно записать в виде линейной комбинации его частных решений, представленных в уравнении (6.155):

$$F(u,z)=\exp\left(-(u/a)^2\right)\sum_{n=0}^{\infty}C_n a^{-n-1}H_n(u/a). \tag{6.156}$$

Чтобы световое поле с амплитудой (6.156) распространялось вдоль оси z, в плоскости $z=0$ следует сформировать следующую комплексную амплитуду:

$$F_0(u)=\exp\left(-(u/a_0)^2\right)\sum_{n=0}^{\infty}C_n a_0^{-n-1}H_n(u/a_0). \tag{6.157}$$

Задача расчета фазовых оптических элементов, генерирующих пучки Эрмита с произвольным модовым составом, может быть сформулирована следующим образом [23]. Необходимо найти фазу $\varphi(u)$, удовлетворяющую следующей системе $N+1$ алгебраических уравнений:

$$|C_n|=\left|\int_{-\infty}^{\infty}\sqrt{I_0(u)}\exp\left(i\varphi(u)\right)P_n(u)\,\mathrm{d}u\right|,\quad n=\overline{0,N}, \tag{6.158}$$

где

$$\sqrt{I_0(u)}\exp(i\varphi(u)) = \sum_{n=0}^{N} C_n P_n(u), \tag{6.159}$$

$$P_n(u) = \left(2^n n!\sqrt{n}\right)^{-1/2} \exp\left(-\frac{u^2}{2}\right) H_n(u), \tag{6.160}$$

$P_n(u)$ — функция Эрмита, $\sqrt{I_0(u)}$ — амплитуда освещающего пучка. Модули коэффициентов $|C_n|$ заданы и определяют энергетический вклад каждой моды Эрмита в общий световой пучок. Функция Эрмита в уравнении (6.160) удовлетворяет соотношениям ортогональности

$$\int_{-\infty}^{\infty} P_n(u) P_m(u)\, \mathrm{d}u = \delta_{mn}, \tag{6.161}$$

где δ_{mn} — символ Кронекера. Начальная оценка $\varphi_0(u)$ фазы выбирается случайной [23]. Тогда $N+1$ коэффициентов C_n находятся при помощи уравнений (6.158). Коэффициенты $C_n^{(k)}$, найденные на k-м шаге итераций, подвергаются замене

$$\bar{C}_n^{(k)} = B_n \frac{C_n^{(k)}}{\left|C_n^{(k)}\right|}, \tag{6.162}$$

где $B_n \geqslant 0$ — заданные вещественные числа, характеризующие распределение световой энергии между модами. Затем рассчитывается сумма (6.159), чей аргумент берется в качестве новой оценки $\varphi_{k+1}(u)$ искомой фазы и т.д. Следует отметить, что все приведенные выше рассуждения могут быть полностью распространены на двумерный случай, если требуется получить фазу $\varphi(\mathbf{u})$ оптического элемента, формирующего световой пучок, состоящий из заданного набора пучков Эрмита с заданными весами энергий:

$$\sqrt{I_0(\mathbf{u})}\exp(i\varphi(\mathbf{u})) = \sum_{n=0}^{N}\sum_{m=0}^{M} C_{mn} P_n(u) P_m(v). \tag{6.163}$$

Процедура, аналогичная (6.158)–(6.162), может быть построена на основе метода обобщенных проекций (6.128). В самом деле, нам необходимо найти такую комплексную функцию F, чтобы ее модуль удовлетворял условию $|F(\mathbf{u})| = \sqrt{I_0(\mathbf{u})}$, накладываемому формой освещающего пучка, а модули комплексных коэффициентов разложения функции по модовому базису определялись следующим образом:

$$\begin{cases} B_1 = \left| \iint\limits_G F(\mathbf{u})\, \psi^*_{\mathbf{p_1}}(\mathbf{u})\, \mathrm{d}^2\mathbf{u} \right|, \\ B_2 = \left| \iint\limits_G F(\mathbf{u})\, \psi^*_{\mathbf{p_2}}(\mathbf{u})\, \mathrm{d}^2\mathbf{u} \right|, \\ \ldots, \end{cases}$$

где $B_n \geqslant 0$ — заданные числа, характеризующие распределение световой энергии между модами. Аргумент $\varphi(\mathbf{u}) = \arg[F(\mathbf{u})]$ искомой функции $F(\mathbf{u})$ является фазовой функцией рассчитываемого модана. Таким образом, необходимо решить задачу восстановления комплекснозначной функции по ее амплитуде и амплитуде поперечно-модового спектра. Эта задача отличается от задачи расчета ДОЭ, фокусирующего лазерное излучение в заданную двумерную область, лишь тем, что вместо амплитуды фурье-спектра в нашем случае фиксирована амплитуда поперечно-модового спектра. Построение итеративной процедуры восстановления комплекснозначной функции по ее амплитуде и амплитуде ее фурье-спектра на основе метода обобщенных проекций рассмотрено в работах [40, 41]. Рассмотрим построение итеративной процедуры восстановления фазы по амплитуде и амплитуде поперечно-модового спектра (6.9).

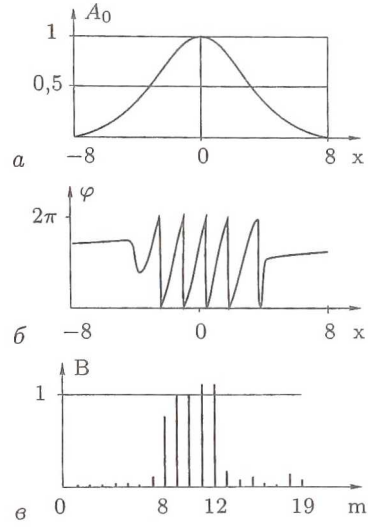


Рис. 6.18. Расчет пучка Эрмита: (*а*) амплитуда освещающего пучка, (*б*) фаза ДОЭ, (*в*) модули коэффициентов C_k

Пример 6.3. Расчет ДОЭ, формирующего несколько мод Гаусса–Эрмита одновременно. При численном моделировании отсчеты вдоль переменной u брались с интервалом в 0,02 мм. Взятие неопределенного интеграла в бесконечных пределах (6.158) заменялось интегрированием на интервале $[-a, a]$, $a = 6$ мм. С ростом a коэффициенты C_n изменялись на десятую долю процента. Амплитуда $\sqrt{I_0(u)}$ освещающего пучка выбиралась гауссовой. Число слагаемых N в сумме (6.159) равнялось 20. На рис. 6.18*б* представлена фаза оптического элемента, освещаемого плоским пучком с гауссовым распределением интенсивности, который формирует пучок Эрмита, эффективно представляющего собой суперпозицию пяти мод Гаусса–Эрмита с постоянными модулями коэффициентов C_k (рис. 6.18*в*).

Модули коэффициентов представлены на рис. 6.18*в*. Отметим, что общий наклон фазы оптического элемента (см. рис. 6.18*б*), приводящий к осевому смещению пучка Эрмита при его распространении, является результатом существования мод с нечетными номерами в сумме (6.159). На рис. 6.19 показаны среднеквадратичные отклонения для амплитуды в плоскости оптического элемента (кривая 1) и для коэффициентов ряда (6.159) (кривая 2) в зависимости от числа итераций;

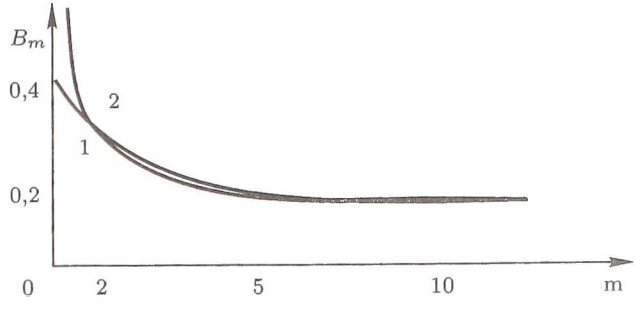


Рис. 6.19. Зависимость среднеквадратичных отклонений от числа итераций

среднеквадратичные отклонения рассчитывались по формулам

$$\delta_A = \left[\int_{-a}^{a} \left||F_{k+1}(u)| - \sqrt{I_0(u)}\right|^2 \mathrm{d}u\right]^{1/2} \left[\int_{-a}^{a} I_0(u)\,\mathrm{d}u\right]^{-1/2} \tag{6.164}$$

$$\delta_C = \left[\sum_{n=0}^{N} \left\{\left|C_n^{(k)}\right| - B_n\right\}^2\right]^{1/2} \left[\sum_{n=0}^{N} B_n^2\right]^{-1/2} \tag{6.165}$$

Из рис. 6.19 видно, что отклонения монотонно уменьшаются с ростом числа итераций. Отметим, что эффективность представлений пучка Эрмита через пять мод составляет в данном случае 90% (рис. 6.18*в*). Эффективность можно найти из следующего соотношения:

$$E = \left[\sum_{\Omega} \left|C_n^{(k)}\right|^2\right] \left[\sum_{n=0}^{N} \left|C_n^{(k)}\right|^2\right]^{-1}, \tag{6.166}$$

где k — номер итерации, Ω — множество тех номеров мод, которые были заданы ненулевыми коэффициентами.

Следующая задача имеет некоторое отличие от предыдущей, так как включает расчет фазового оптического элемента, формирующего пучки Гаусса–Эрмита в заданных дифракционных порядках и с заданным распределением энергии света между этими порядками. В данном случае каждая световая мода распространяется под своим собственным углом к оптической оси. Следовательно, чтобы найти фазу ДОЭ $\varphi(u)$, следует вместо уравнения (6.159) использовать следующее соотношение:

$$\sqrt{I_0(u)} \exp(i\varphi(u)) = \sum_{n=0}^{N} C_n P_n(u) \exp(-i\nu_n u), \tag{6.167}$$

где $P_n(u)$ — функция Эрмита, заданная в уравнении (6.160), ν_n — несущая пространственная частота: $\nu_1 < \nu_2 < \ldots < \nu_N$. Модули коэффициентов C_n, как и ранее, являются фиксированными значениями, а их аргументы являются свободными параметрами задачи. Функции Эрмита являются собственными функциями преобразования Фурье:

$$\int_{-\infty}^{\infty} P_n(u) \exp(-i\nu u)\,\mathrm{d}u = (-i)^n P_n(\nu). \tag{6.168}$$

Воспользовавшись уравнением (6.168), мы получим следующее выражение:

$$\int_{-\infty}^{\infty} P_n(u) P_m(u) \exp[-i(\nu_n - \nu_m)u]\,\mathrm{d}u = i^{n+m} \int_{-\infty}^{\infty} P_n(\zeta) P_m(\zeta + \nu_n - \nu_m)\,\mathrm{d}\zeta. \tag{6.169}$$

Функция в правой части уравнения (6.169) является корреляцией двух функций Эрмита. Поэтому, выбирая разность между двумя последовательными пространственными несущими частотами $|\nu_{n+1} - \nu_n|$ достаточно большой (на практике это

означает, что формируемые пучки Эрмита должны быть разделены некоторым порогом), мы можем рассматривать функции $P_n(u)\exp[-i\nu_n u]$ как приближенно ортогональные и можем получить коэффициенты суммы (6.167) из соотношения

$$C_n = \int_{-\infty}^{\infty} \sqrt{I_0(u)}\exp\left(i\varphi\left(u\right) + i\nu_n u\right) P_n\left(u\right)\,\mathrm{d}u, \quad n = \overline{0,N}. \tag{6.170}$$

Используя уравнения (6.167) и (6.170), мы можем организовать итеративную процедуру поиска фазы $\varphi(u)$ оптического элемента, аналогичную алгоритму, описанному в уравнениях (6.158)–(6.160). Коэффициенты C_n, полученные из уравнения (6.170) на k-м шагу итераций, подвергаются следующей замене: их модули заменяются заданными числами B_n, при этом фазы (аргументы) остаются без изменений. Затем находится сумма (6.167) и полученная функция подвергается схожей замене: ее модуль заменяется амплитудой $\sqrt{I_0(u)}$ освещающего пучка, при этом фаза остается без изменений и представляет собой новую оценку искомой фазы. Очевидно, что вышеприведенные соображения можно применить к 2-мерному случаю, если вместо уравнения (6.167) использовать следующее соотношение:

$$\sqrt{I_0\left(\mathbf{u}\right)}\exp\left(i\varphi\left(\mathbf{u}\right)\right) = \sum_{n=0}^{N}\sum_{m=0}^{M} C_{nm}P_n\left(u\right)P_m\left(v\right)\exp\left(-i\left(\nu_n^1 u + \nu_m^2 v\right)\right). \tag{6.171}$$

Пример 6.4. Расчет четырехканального модана Гаусса–Эрмита (двумерный случай). Для конкретного расчета мы выбрали следующие параметры: переменные x и y менялись на интервале $[-10, 10]$ с шагом 0,02. Общее число отсчетов составляло 128×128. Число слагаемых в сумме (6.171) выбиралось $M = N = 9$. Амплитуда освещающего пучка выбиралась постоянной $\sqrt{I_0\left(\mathbf{u}\right)} = \mathrm{const}$. На рис. 6.20 представлены: (а) фаза четырехканального оптического элемента, рассчитанного за 8 итераций, который может формировать в фокальной плоскости линзы распределение интенсивности (б) и (в) — распределение фазы. На рис. 6.20*б* показаны: мода $(0,1)$ в верхнем левом углу, мода $(1,1)$ в верхнем правом углу, мода $(1,2)$ в нижнем левом углу, и мода $(2,0)$ в нижнем правом углу. На формирование этих четырех мод идет около 95% энергии освещающего пучка, при этом энергия распределяется в равных пропорциях между этими модами со средней ошибкой в 2%. Среднее отклонение интенсивности сформированных мод от идеальных значений достигает 21%. На рис. 6.21 показано, как меняется ошибка с ходом итераций. Видно, что ошибка монотонно уменьшается. Увеличение скорости сходимости после шестой итерации

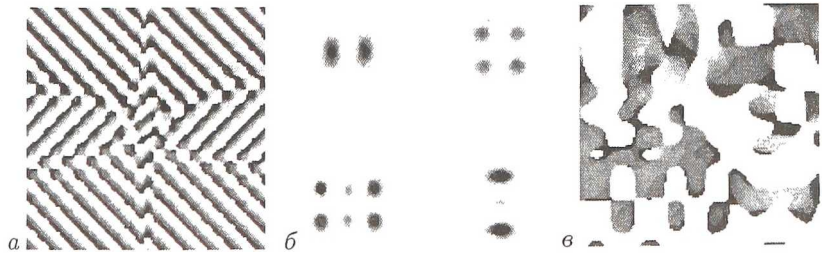


Рис. 6.20. Расчет четырехканального формирователя мод Эрмита: (*а*) фаза ДОЭ, (*б*) интенсивность, (*в*) распределение фазы в фокальной плоскости

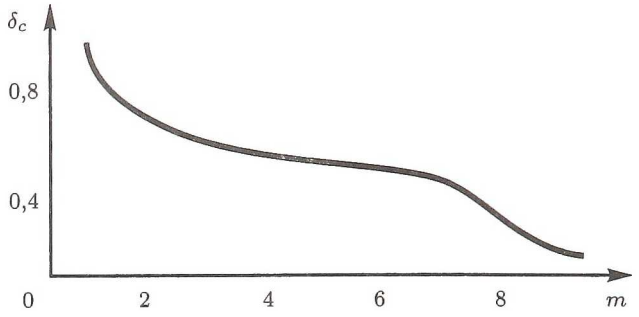


Рис. 6.21. Зависимость ошибки (6.165) от числа итераций

(см. рис. 6.21) происходит благодаря применению адаптивно-аддитивного алгоритма [49]. На рис. 6.20*в* представлена фаза в частотной плоскости. Видно, что в районах локализации четырех мод имеются характеристические скачки фазы на π при переходе через линии нулевой интенсивности.

Оптический элемент с фазой, представленной на рис. 6.20*а*, можно использовать не только для формирования пространственно-разделенных мод Гаусса–Эрмита, но и в качестве пространственного фильтра, согласованного с четырьмя вышеназванными фильтрами (6.2.8).

Если такой оптический элемент освещается световым пучком с амплитудой, пропорциональной амплитуде некоторой моды $T_{mn} = P_n(u)P_m(v)$, появляется корреляционный пик интенсивности в частотной плоскости в точке $\left(\nu_n^1, \nu_m^2\right)$. На рис. 6.22 показаны распределения интенсивности в частотной плоскости, если оптический элемент (рис. 6.20*а* освещается: (*а*) модой T_{01}, (*б*) модой T_{20}, (*в*) модой $T_{11} + T_{12} + T_{20}$, (*г*) модой T_{02}. Отметим, что последняя мода отсутствовала при расчете ДОЭ. Описание итеративной процедуры и соответствующие результаты расчета элементов, согласованных с группой мод Гаусса–Лагерра можно найти в главе 7 данной книги. Отметим, что попытка расчета модана, формирующего одномодовое распределение, с помощью процедуры (6.158)–(6.162) приведет после первой же итерации к стагнации, а в качестве полученного решения будет знаковая функция формируемой моды, как в [25, 26].

6.2.7. Организация многоканальной связи в идеальной линзоподобной среде с минимальными потерями энергии

Если методы расчета моданов, основанные на внесении несущей в фазу элемента [19], приводят к низкой энергетической эффективности элементов, то главным недостатком процедур расчета фазовых ДОЭ, основанных на последовательном построении проекций на замкнутые множества [27–30], является принципиальное отсутствие сходимости [40]. После некоторого числа итераций приходится столкнуться со стагнацией алгоритма, т.е. наступает момент, когда дальнейшее увеличение числа итераций не приводит к заметному улучшению характеристик рассчитываемого элемента [23, 46]. При этом стагнация итеративной процедуры вовсе не обязательно означает, что найдено наилучшее решение. Подробнее о различных причинах возникновения стагнации и подходах к борьбе с ней можно прочитать в главе 2 данной книги и в работах [23, 40, 46]. Определенной альтернативой градиентным и проекционным процедурам в случае наличия у исследователя высокопроизводительной

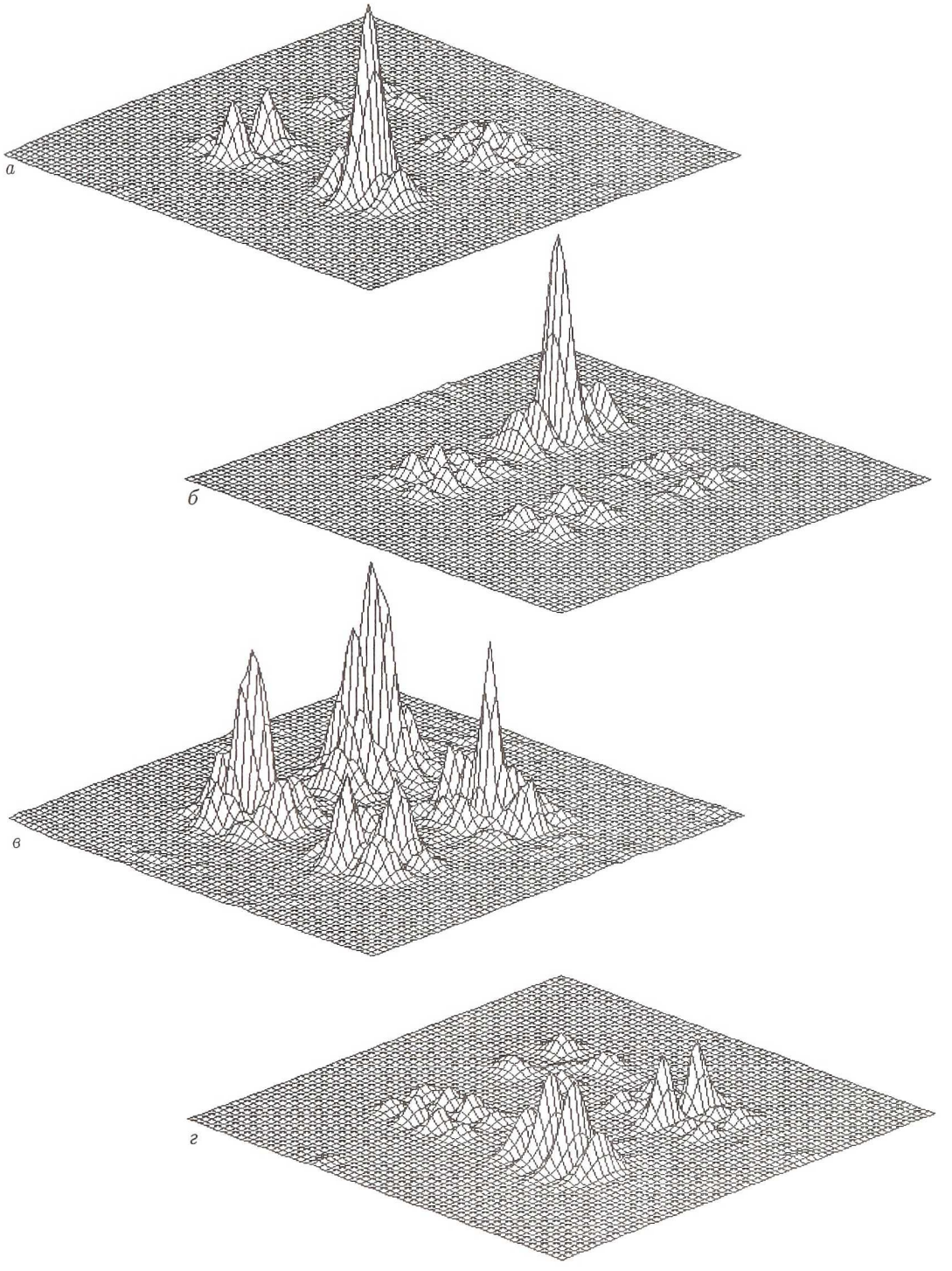


Рис. 6.22. Распределение интенсивности в частотной плоскости, если ДОЭ (рис. 6.20*а*) освещается: (*а*) модой T_{01}, (*б*) модой T_{20}, (*в*) модой $T_{11} + T_{12} + T_{20}$ и (*г*) модой T_{02}

вычислительной техники могут выступить стохастические оптимизационные процедуры, все чаще применяемые в последнее время в задачах расчета ДОЭ [50]. В любом случае, результат работы итеративной процедуры напрямую связан с наличием свободных параметров, которыми можно варьировать для приведения функции комплексного пропускания к чисто фазовой. Для качественного формирования модовых пучков с помощью алгоритма (6.158)–(6.162) требуется, как правило, чтобы число мод было более 4–5 [23]. В противном случае, доля паразитных мод в формируемых квазиодномодовых пучках будет чересчур высока. Это может привести к заметному уширению импульса в канале связи при использовании такого элемента в оптической телекоммуникационной системе. Количество формируемых мод определяется конкретной спецификой задачи, например, количество мод может соответствовать числу независимых каналов в оптико-волоконной системе связи. Передача одного канала связи с помощи произвольной группы из нескольких мод приведет к появлению эффекта уширения импульса в канале [18]. Для увеличения числа свободных параметров, необходимых для работы итеративной процедуры, в [20] предложено для передачи каждого канала связи в параболическом волокне использовать пакет двумерных гауссовых мод с одинаковыми значениями постоянной распространения β. Напомним, что из (6.94), (6.95), (6.97) следует, что двумерные моды Гаусса–Лагерра с номерами, удовлетворяющими условию $r_{\text{р}} = 2p + l = \text{const}$ обладают одинаковой постоянной распространения $\beta_{r_{\text{р}}}$ в линзоподобной среде. Из (6.94), (6.95), (6.109) следует аналогичное условие для мод Гаусса–Эрмита – $r_p = p + l = \text{const}$. Таким образом, возможно увеличить число свободных параметров, не сталкиваясь с проблемой уширения импульсов, если использовать для передачи одного канала информации в идеальной линзоподобной среде суперпозицию гауссовых мод с одинаковым значением постоянной распространения. На рис. 6.23 представлено трехмерное распределение амплитуды четырех пакетов мод Гаусса–Эрмита с одинаковым значением постоянной распространения внутри каждого пакета, формируемых ДОЭ в разных дифракционных порядках [20]. На рис. 6.24 представлен аналогичный результат для четырех пакетов мод Гаусса–Лагерра [20].

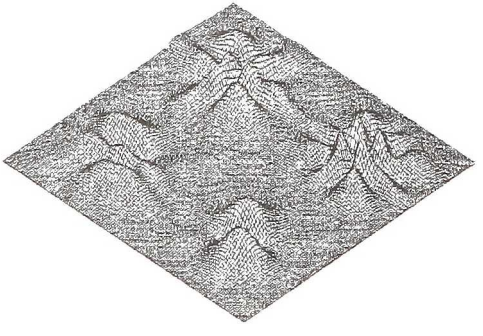


Рис. 6.23. Распределение амплитуды четырех модовых пакетов, состоящих из мод Гаусса–Эрмита с одинаковым значением постоянной распространения

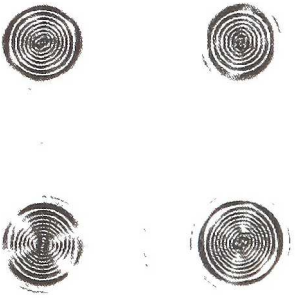


Рис. 6.24. Распределение амплитуды четырех модовых пакетов, состоящих из мод Гаусса–Лагерра с одинаковым значением постоянной распространения

Отметим особенность пучков, состоящих из гауссовых мод с одинаковым значением постоянной распространения. Комплексная амплитуда в сечении таких пучков имеет следующий вид:

$$\chi_{r_{\mathbf{p}}}(\mathbf{x}) = \sum_{r_{\mathbf{p}}=\text{const}} \tilde{C}_{\mathbf{p}} \psi_{\mathbf{p}}(\mathbf{x}). \tag{6.172}$$

Распространение пучков вида (6.172) напоминает распространение отдельных мод $\psi_{\mathbf{p}}(\mathbf{x})$. В самом деле, пучок когерентного света с поперечным сечением вида (6.172) обладает свойством самовоспроизведения амплитудно-фазовой структуры в ходе распространения в соответствующей среде. Такие пучки называют инвариантными модовыми пакетами [20] (или многомодовыми бездисперсионными импульсами).

Выпишем основные свойства инвариантных модовых пакетов.

1. Дискретный набор мод волновода $\{\psi_{\mathbf{p}}(\mathbf{x})\}$ порождает континуальное множество инвариантных модовых пакетов в силу произвольности выбора коэффициентов $\tilde{C}_{\mathbf{p}}$ в (6.172).

2. Самовоспроизводство: инвариантные модовые пакеты (6.172) не изменяют своей амплитудно-фазовой структуры в ходе распространения в соответствующей волноводной среде в силу (6.94), (6.95), (6.97).

3. Гауссовы инвариантные модовые пакеты не изменяют своей амплитудно-фазовой структуры в ходе распространении в линзоподобных средах, свободном пространстве, при прохождении через фурье-каскад, изменяя лишь фундаментальный модовый радиус в соответствии с (6.118).

4. Инвариантные модовые пакеты распространяются в волноводе без уширения импульса, вызываемого межмодовой дисперсией, описанной, например, в [5].

На рис. 6.25–6.26 приведены результаты моделирования прохождения суммы двух мод Гаусса–Эрмита с номерами $(2,2)$ и $(4,0)$, взятых с единичными весами $\tilde{C}_{22} = 1$, $\tilde{C}_{40} = 1$, через фурье-каскад. На рис. 6.25 представлена амплитуда двухмодового пучка до прохождения фурье-каскада, на рис. 6.26 — после. Результат моделирования подтверждает, что пучок, состоящий из двух мод с одинаковой постоянной распространения, сохраняет свою структуру подобно одной моде.

На рис. 6.27–6.30 представлены результаты натурного исследования элемента, формирующего амплитудно-фазовое распределение, описываемое суммой двух мод Гаусса–Эрмита с номерами $(2,2)$ и $(4,0)$, взятых с единичными весами $\tilde{C}_{22} = 1$, $\tilde{C}_{40} = 1$. Соответствующий ДОЭ был изготовлен по технологии, описанной в [29]. Для

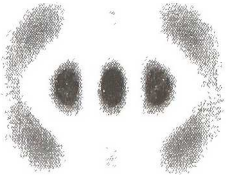


Рис. 6.25. Амплитуда пучка, состоящего из двух мод Гаусса–Эрмита с номерами $(2,2)$ и $(4,0)$, сформированного ДОЭ

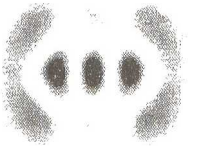


Рис. 6.26. Результат моделирования прохождения пучка, представленного на рис. 6.25, через фурье-каскад

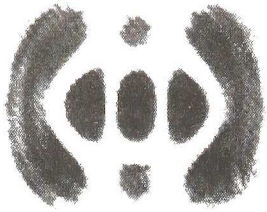


Рис. 6.27. Амплитуда в фокальной плоскости элемента, формирующего распределение суммы двух мод Гаусса–Эрмита с номерами $(2,2)$ и $(4,0)$ (восстановлена по результатам измерения интенсивности)



Рис. 6.28. Результат натурного исследования прохождения суммы двух мод Гаусса–Эрмита с номерами $(2,2)$ и $(4,0)$ через фурье-каскад

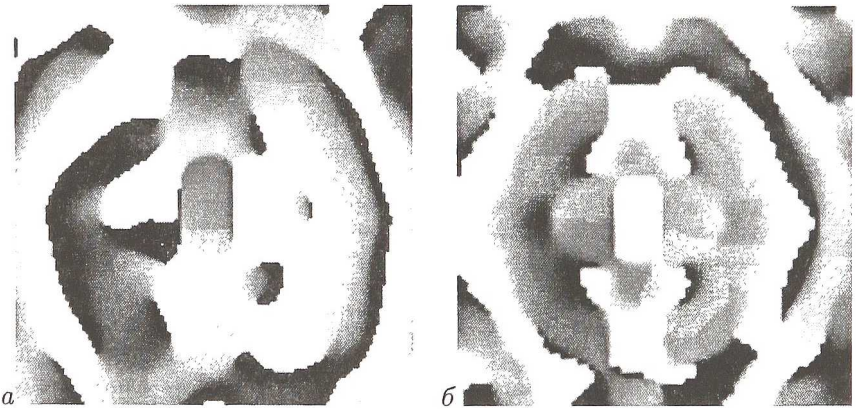


Рис. 6.29. Результаты численного восстановления фазы пучка по распределениям амплитуды во входной и выходной плоскостях фурье-каскада, представленным на рис. 6.27–6.28



Рис. 6.30. Результат интерферометрического исследования фазы пучка, состоящего из суммы двух мод Гаусса–Эрмита с номерами $(2,2)$ и $(4,0)$

кодирования его комплексной функции пропускания применялся обощенный метод Кирка–Джонса, описанный в [19]. Приведенные физические результаты наглядно демонстрируют принципиальную возможность формирования инвариантных модовых пакетов с помощью ДОЭ.

Использование пакетов мод с одинаковым значением постоянной распространения для передачи отдельных информационных каналов позволяет выбирать в качестве свободных параметров для работы итеративной процедуры расчета фазового модана не только межмодовые сдвиги $\arg\left(\tilde{C}_{\mathbf{p}}\right)$, но и распределение энергии между модами внутри одного пакета $\left|\tilde{C}_{\mathbf{p}}\right|^2$ при фиксированной энергии каждого пакета $\sum\limits_{r_{\mathbf{p}}}\left|\tilde{C}_{\mathbf{p}}\right|^2 = B_m$ ($r_{\mathbf{p}} = p + l$ в случае мод Гаусса–Эрмита и $r_{\mathbf{p}} = 2p + l$ в случае мод Гаусса–Лагерра,) где m — номер информационного канала, B — некоторая наперед заданная неотрицательная константа.

Пример 6.5. Организация эффективной многоканальной системы связи в идеальной линзоподобной среде. Предположим, что требуется создать N_k информационных каналов передачи цифровой информации в идеальной линзоподобной среде, используя моды Гаусса–Эрмита, с равным распределением энергии между каналами $B_1 = B_2 = \ldots = B_{N_k}$. Общее число используемых мод определяется, как известно, числом отсечки $r_{\max} = N_{\text{cut}}$ [15]:

$$B_m = \sum_{p+l=m}\left|\tilde{C}_{\mathbf{p}}\right|^2, \quad \sum_{m=1}^{N_k} B_m = E_0, \tag{6.173}$$

где E_0 — полная энергия коллимированного источника лазерного излучения L в плоскости установки ДОЭ.

В дальнейшем не будем учитывать потери, связанные с поглощением и френелевским отражением. Для пространственного разделения формируемых инвариантных модовых пакетов используем модификацию разложения, описанного ранее в [23]:

$$A_0(\mathbf{x})\exp(i\varphi(\mathbf{x})) = \sum_{j=1}^{N_k}\exp(i2\pi\nu_j\mathbf{x})\sum_{p+l=j}\tilde{C}^j_{\mathbf{p}}\psi^j_{\mathbf{p}}(\mathbf{x}), \quad N_k \leqslant N_{\text{cut}}, \tag{6.174}$$

где $A_0(\mathbf{x})$ — амплитудное распределение в поперечном сечении освещающего коллимированного пучка, $\varphi(\mathbf{x})$ — фазовая функция модана M и ν_j — частоты несущих,

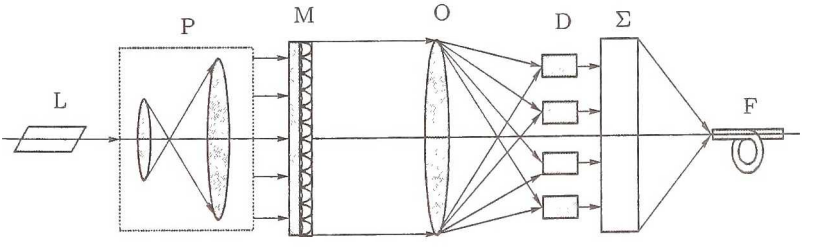


Рис. 6.31. Общая схема многоканальной телекоммуникационной системы. L — источник лазерного излучения, P — коллиматор лазерного излучения, M — модан, O — фурье-каскад, D — матрица модуляторов, F — идеальный линзоподобный волновод, Σ — оптический сумматор пучков

вводимых для пространственного разделения формируемых пучков [23]. Теперь, для нахождения коэффициентов $\tilde{C}_{\mathbf{p}}$ для каждой моды в (6.174) можно использовать какую-либо оптимизационную процедуру, скажем стохастическую [50], минимизирующую функционал невязки следующего вида $\delta_m = \sum_j^{N_k} \left| \frac{1}{N_k} - \sum_{p+l=j} \left| \tilde{C}_{\mathbf{p}}^{jm} \right|^2 \right|$, где коэффициенты находятся следующим образом:

$$\tilde{C}_{\mathbf{p}}^{jm} = \iint_G A_0(\mathbf{x}) \exp\left(i\varphi_m(\mathbf{x})\right) \psi_{\mathbf{p}}^j(\mathbf{x}) \exp\left(i2\pi\nu_j \mathbf{x}\right) \mathrm{d}^2\mathbf{x}, \tag{6.175}$$

а $\varphi_m(\mathbf{x})$ — оценка фазы ДОЭ на m-ой итерации оптимизационной процедуры.

6.2.8. Расчет ДОЭ для анализа поперечно-модового состава пучков когерентного излучения

Выше отмечалось (п. 6.2.1), что под анализом поперечно-модового состава пучка целесообразно понимать измерение распределения мощности по модам (т.е. значений квадратов модулей коэффициентов (6.9)) и межмодовых фазовых сдвигов (аргументов коэффициентов (6.9)). Опираясь на вышеизложенный материал, рассмотрим расчет функции комплексного пропускания ДОЭ, предназначенного для анализа поперечно-модового состава пучков когерентного излучения.

Предположим, что $\{\psi_{\mathbf{p}}(x)\}$ — комплексные амплитуды мод в сечении пучка, освещающего модан, которые подчиняются условию ортогональности (6.1)

$$\int_G \psi_{\mathbf{p}}(\mathbf{x}) \psi_{\mathbf{p}'}^*(\mathbf{x}) \, \mathrm{d}^2\mathbf{x} = \delta_{\mathbf{p}'\mathbf{p}}, \tag{6.176}$$

где $\mathbf{x} = (x, y)$ — двумерные декартовы координаты в сечении пучка, G — световая апертура, * — символ комплексного сопряжения и

$$\delta_{\mathbf{p}\mathbf{p}'} = \delta_{pp'}\delta_{ll'} = \begin{cases} 1, & \mathbf{p} = \mathbf{p}' \\ 0, & \mathbf{p} \neq \mathbf{p}' \end{cases} \tag{6.177}$$

— символ Кронекера. Если мы разложим комплексную амплитуду пучка в ряд по модовым функциям, наша задача сведется к нахождению значений (6.9) комплекснозначных коэффициентов $F_{\mathbf{p}}$. Значения квадратов модулей коэффициентов $|F_{\mathbf{p}}|^2$ описывают распределение мощностей мод в пучке, а распределение аргументов коэффициентов $\arg(F_{\mathbf{p}})$ — межмодовые сдвиги в пучке. Задачу измерения модового состава $\{\psi_{\mathbf{p}}(\mathbf{x})\}$ поля с комплексной амплитудой

$$F(\mathbf{x}) = \sum_{\mathbf{p}} F_{\mathbf{p}} \psi_{\mathbf{p}}(\mathbf{x}), \quad \mathbf{x} \in G, \tag{6.178}$$

в [19] предлагается решать в два этапа.

Первый этап представляет собой оптическую селекцию мод на основе их свойств ортогональности. При этом распределение мощности по модам преобразуется в распределение интенсивности света в заданных точках фокальной плоскости (набор соответствующих корреляционных пиков).

Второй этап состоит в измерении мощности каждой моды $|F_{\mathbf{p}}|^2$ обычными фотоприемниками, установленными в центральных точках корреляционных пиков, соответствующих заданным модам.

Покажем, что при установке на вход фурье-каскада пространственного фильтра с математической функцией комплексного пропускания $T(\mathbf{x}) = \psi^*_{\mathbf{p}}(\mathbf{x})$, освещаемого пучком с поперечным распределением комплексной амплитуды $F(\mathbf{x})$, значение комплексной амплитуды в центре выходной плоскости будет однозначно определяться значением коэффициента $F_{\mathbf{p}}$ моды в (6.178). В самом деле, умножая комплексную амплитуду пучка $F(\mathbf{x})$ на функцию комплексного пропускания элемента $T(\mathbf{x})$, где $T(\mathbf{x}) = \psi^*_{\mathbf{p}}(\mathbf{x})$ — комплексно-сопряженная к $\psi_{\mathbf{p}}(\mathbf{x})$, получаем

$$F(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x}) = \sum_{\mathbf{p}'} F_{\mathbf{p}'}\psi_{\mathbf{p}'}(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x}). \tag{6.179}$$

Преобразование Фурье, взятое в центре $\mathbf{u} = 0$ выходной плоскости пространственных частот $\mathbf{u} = (u, v)$, дает операцию интегрирования комплексной амплитуды.

Двумерное интегрирование по апертуре G дает нам

$$F_1(0) = \int\limits_G F(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x})\,\mathrm{d}^2\mathbf{x}. \tag{6.180}$$

В силу ортогональности взаимодействие мод происходит так, что при интегрировании комплексной амплитуды (6.180) в точке $\mathbf{u} = 0$ «фон» взаимоуничтожается и получается значение поля, равное

$$F_{\mathbf{p}} = \int\limits_G F(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x})\,\mathrm{d}^2\mathbf{x}. \tag{6.181}$$

Мощность $|F_{\mathbf{p}}|^2$ моды $\psi_{\mathbf{p}}(\mathbf{x})$ будет измерена уже обычным фотоприемником с точечной апертурой. Последовательно меняя фильтры $\psi_{\mathbf{p}}(\mathbf{x})$, можно измерить распределение мощности по всему поперечно-модовому спектру. Однако, в практических задачах, требующих измерения поперечно-модового состава излучения [19], требуется измерение значений мощностей нескольких мод в режиме реального времени. Соответствующий «многоканальный» вариант должен обеспечить параллельное формирование коэффициентов $F_{\mathbf{p}}$ в виде соответствующих значений комплексной амплитуды в наборе точек плоскости $\mathbf{u} = (u, v)$. В рамках этой схемы, измерение интенсивности $|F_1(\mathbf{u})|^2$ в соответствующей точке в фурье-плоскости с помощью малоапертурного детектора или элементов матрицы CCD-камеры позволяет измерять требуемые значения $|F_{\mathbf{p}}|^2$. Для реализации этой схемы апертура G может быть интерпретирована как сумма двух (или более) связанных «виртуальных» субапертур, каждая из которых описывается своей функцией комплексного пропускания вида $T_{\mathbf{p}}(\mathbf{x}) = \xi_{\mathbf{p}}\psi^*_{\mathbf{p}}(\mathbf{x})\exp(i2\pi\nu_{\mathbf{p}}\mathbf{x})$, согласованной с соответствующей модой $\psi_{\mathbf{p}}(\mathbf{x})$ и связанной с несущей ν_p, введенной для углового разделения парциальных пучков, формирующих комплексные амплитуды $F_{\mathbf{p}}$ в соответствующих точках, $\xi_{\mathbf{p}}$ — введенные весовые неотрицательные коэффициенты, в простейшем случае $\xi_{\mathbf{p}} = 1$.

Функция комплексного пропускания ДОЭ в этом случае примет вид

$$T(\mathbf{x}) = \sum_{\mathbf{p}} \xi_{\mathbf{p}}\psi^*_{\mathbf{p}}(\mathbf{x})\exp(i2\pi\nu_{\mathbf{p}}x).$$

При помещении в оптическую схему поперечно-модового анализатора модан выполняет параллельное формирование набора модовых пучков, разделяющихся по углам в плоскости модана и по координате в фурье-сопряженной плоскости. В точках фурье-плоскости с координатами

$$\mathbf{u}_\mathbf{p} = f\lambda\nu_\mathbf{p}, \tag{6.182}$$

соответствующими пространственным частотам [19], мы получим значения комплексной амплитуды, пропорциональные

$$F_\mathbf{p} = \int\limits_G F(\mathbf{x})\xi_\mathbf{p}\psi^*_\mathbf{p}(\mathbf{x})\exp(-i2\pi\nu_\mathbf{p}\mathbf{x})\,\mathrm{d}^2\mathbf{x}. \tag{6.183}$$

Полагая весовой коэффициент k-го канала многоканального модана $\xi^{(\mathbf{k})}_\mathbf{p} = \delta_{\mathbf{pk}}$, мы пришли (6.179)–(6.183) к возможности формирования в каждом канале комплексной амплитуды, пропорциональной модовому коэффициенту $F_\mathbf{p}$, квадрат модуля которого можно измерить $\left(\sim |F_\mathbf{p}|^2\right)$, используя фотоприемник. Измерение фазовых межмодовых сдвигов $\arg F_\mathbf{p}$ требует использования весового коэффициента более сложного вида.

Рассмотрим многоканальный анализирующий модан с коэффициентами $\zeta^{(\mathbf{k_c})}_\mathbf{p}$ и $\zeta^{(\mathbf{k_s})}_\mathbf{p}$, соответствующими каналам $\mathbf{k_c}$ и $\mathbf{k_s}$:

$$\begin{aligned}\zeta^{(\mathbf{k_c})}_\mathbf{p} &= \delta_{\mathbf{p_1 k_c}} + \delta_{\mathbf{p_2 k_c}}\\ \zeta^{(\mathbf{k_s})}_\mathbf{p} &= \delta_{\mathbf{p_1 k_s}} + \exp(i\Delta\varphi)\,\delta_{\mathbf{p_2 k_s}}, \quad \Delta\varphi = \mathrm{const}.\end{aligned} \tag{6.184}$$

Освещение такого модана пучком с поперечным распределением комплексной амплитуды $F(\mathbf{x})$ приводит к формированию комплексных коэффициентов,

$$\begin{aligned}F^{(\mathbf{k_c})} &= F_{\mathbf{p_1}} + F_{\mathbf{p_2}}\\ F^{(\mathbf{k_s})} &= F_{\mathbf{p_1}} + \exp(i\Delta\varphi)\,F_{\mathbf{p_2}},\end{aligned} \tag{6.185}$$

в каналах с номерами $\mathbf{k_c}$ и $\mathbf{k_s}$, соответственно. Фотоприемник с точечной апертурой, измеряющий интерференцию коэффициентов $F_{\mathbf{p_1}}$ и $F_{\mathbf{p_2}}$ в точках $\lambda f\nu^{(k_s)}$ и $\lambda f\nu^{(k_c)}$, не позволяет нам непосредственно восстановить фазы $\varphi_\mathbf{p} = \arg F_\mathbf{p}$ коэффициентов $F_\mathbf{p}$. Восстановление фаз $\arg F_\mathbf{p}$ требует решения системы уравнений подобно тому, как это делается при обработке квадратурных интерферограмм. Для того, чтобы получить требуемые уравнения, запишем уравнения (6.184) в следующей форме:

$$\begin{aligned}\left|F^{(\mathbf{k_c})}\right|^2 &\cong |F_{\mathbf{p_1}}|^2 + |F_{\mathbf{p_2}}|^2 + 2\cdot|F_{\mathbf{p_1}}|\cdot|F_{\mathbf{p_2}}|\cos(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}})\\ \left|F^{(\mathbf{k_s})}\right|^2 &\cong |F_{\mathbf{p_1}}|^2 + |F_{\mathbf{p_2}}|^2 + 2\cdot|F_{\mathbf{p_1}}|\cdot|F_{\mathbf{p_2}}|\times\\ &\quad\times\left[\cos(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}})\cos\Delta\varphi - \sin(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}})\sin\Delta\varphi\right].\end{aligned} \tag{6.186}$$

Предположим, что мощность $|F_{\mathbf{p_1}}|^2$ замерена в канале с номером $\mathbf{k_1}$ и мощность $|F_{\mathbf{p_2}}|^2$ — в канале с номером $\mathbf{k_2}$, так что

$$\left|F^{(\mathbf{k_1})}\right|^2 \approx |F_{\mathbf{p_1}}|^2, \quad \left|F^{(\mathbf{k_2})}\right|^2 \approx |F_{\mathbf{p_2}}|^2. \tag{6.187}$$

В этом случае мощности мод с номерами $\mathbf{p_1}$ и $\mathbf{p_2}$ могут быть найдены (6.181). Далее, для того чтобы восстановить фазовый сдвиг $\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}}$, необходимо найти косинус и синус сдвига:

$$\cos(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}}) = \frac{\left|F^{(\mathbf{k_c})}\right|^2 - \left|F^{(\mathbf{k_1})}\right|^2 - \left|F^{(\mathbf{k_2})}\right|^2}{2\left[\left|F^{(\mathbf{k_1})}\right|^2 \cdot \left|F^{(\mathbf{k_2})}\right|^2\right]^{1/2}}, \tag{6.188}$$

$$\sin(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}}) = \frac{\cos(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}})}{\operatorname{tg}\Delta\varphi} - \frac{\left|F^{(\mathbf{k_s})}\right|^2 - \left|F^{(\mathbf{k_1})}\right|^2 - \left|F^{(\mathbf{k_2})}\right|^2}{2\sin\Delta\varphi\left[\left|F^{(\mathbf{k_1})}\right|^2 \cdot \left|F^{(\mathbf{k_2})}\right|^2\right]^{1/2}}. \tag{6.189}$$

Затем, учитывая знак функции $\cos(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}})$, фазовый сдвиг $\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}}$ может быть определен из уравнения

$$\operatorname{tg}(\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}}) = \frac{1}{\operatorname{tg}\Delta\varphi} - \frac{1}{\sin\Delta\varphi} \cdot \frac{\left|F^{(\mathbf{k_s})}\right|^2 - \left|F^{(\mathbf{k_1})}\right|^2 - \left|F^{(\mathbf{k_2})}\right|^2}{\left|F^{(\mathbf{k_c})}\right|^2 - \left|F^{(\mathbf{k_1})}\right|^2 - \left|F^{(\mathbf{k_2})}\right|^2}. \tag{6.190}$$

Можно определить соответствующие значения фаз $\varphi_{\mathbf{p}}$ либо из рекуррентных соотношений $\varphi_{\mathbf{p_1}} - \varphi_{\mathbf{p_2}}$, либо с использованием опорного пучка. Для кодирования рассчитанной функции комплексного пропускания модана, предназначенного для анализа модового состава пучка, в чисто фазовую в [19] было предложено использовать обобщенный метод Кирка–Джонса, используемого также для кодирования моданов, предназначенных для формирования заданного модового распределения. Нетрудно видеть, что фазовая функция пропускания модана, предназначенного для измерения только мощностей гауссовых мод в исследуемом пучке, может быть рассчитана с помощью итеративной процедуры (6.167)–(6.170), если в качестве освещающего пучка $\sqrt{I_0(\mathbf{x})}$ в данной процедуре будет использован плоский пучок с равномерным распределеним интенсивности $\sqrt{I_0(\mathbf{x})} = \text{const}$. Соответствующие результаты моделирования измерения мощностей мод Гаусса–Эрмита в освещающем пучке приведены на рис. 6.22.

В силу замечательных свойств гауссовых мод, для расчета моданов, предназначенных для анализа поперечно-модового состава излучения могут быть использованы подходы, рассмотренные выше для расчета моданов, формирующих заданное модовое распределение. Результаты натурного эксперимента по селекции мод Гаусса–Эрмита с помощью ДОЭ, рассчитанного по описанной выше методике, приведены в 6.2.10.

6.2.9. Результаты экспериментального исследования фундаментальных свойств гауссовых мод с помощью ДОЭ

Первая работа по синтезированным на компьютере оптическим элементам для анализа и формирования поперечно-модового состава излучения была опубликована в 1982 г. [51]. В то время ДОЭ, согласованные с поперечными модами, представляли собой бинарные амплитудные (а впоследствии — фазовые) нерегулярные решетки, рассчитанные с помощью компьютера. Близкая к [51] по постановке задача содержалась в работе [52], опубликованной в 1983 г. До появления ДОЭ предлагались сложные диафрагмы с газовыми ячейками, повторяющие форму нулей мод [53], однако формируемые ими изображения содержали много паразитных мод. Впослед-

ствии были изготовлены фазовые моданы, рассчитанные с помощью введения несущей в фазу элемента [19]. В работе [54] приведены результаты возбуждения мод в градиентном оптоволокне с помощью ДОЭ. Результаты по экспериментальному исследованию фазовых моданов, формирующих гауссовы моды, и фундаментальных свойств сформированных мод приведены в работах [19, 27–30, 33–35].

Экспериментальное исследование фазовых моданов Гаусса–Эрмита. Для экспериментального исследования свойств элементов, рассчитанных с помощью итеративной процедуры (6.128)–(6.129) с операторами (6.134) и (6.138), был изготовлен и исследован модан, предназначенный для формирования моды Гаусса–Эрмита $(1,0)$ [27–30]. Интересно отметить особенность мод Гаусса–Эрмита, заключающуюся в том, что поворот модана на 900 приводит к взаимной смене значений первого и второго номеров формируемой моды. Таким образом, синтезировав два модана Гаусса–Эрмита $(1,0)$ с помощью одного комплекта технологических бинарных масок, можно получить возможность формирования двух мод — Гаусса–Эрмита $(1,0)$ и Гаусса–Эрмита $(1,0)$ с одинаковыми значениями фундаментального модового радиуса, что и было продемонстрировано в [29]. Элемент был рассчитан как добавка к линзе. Исходя из конкретных условий постановки натурного эксперимента, были выбраны следующие физические параметры элемента: длина волны освещающего пучка $\lambda = 0{,}6328$ мкм, радиус апертуры элемента $a = 3{,}072$ мм, радиус освещающего гауссова пучка $\sigma_{00} = 0{,}835$ мм, внешний радиус полезной области $R_1 = 2$ мм, внешний радиус вспомогательной области $R_2 = 10$ мм, фокусное расстояние линзы $f = 452$ мм, число отсчетов фазовой функции рассчитываемого элемента $N = 2048 \times 2048$, количество итераций итеративной процедуры [27] $n = 10$. На рис. 6.32 приведено рассчитанное распределение фазовой функции на апертуре модана, приведенное к диапазону $[0, 2\pi]$. С помощью процедуры быстрого преобразования Фурье было проведено компьютерное моделирование работы рассчитанного элемента. В результате моделирования получены следующие значения критериев качества: энергетическая эффективность $E = 0{,}455$; удельная мощность $\zeta = 0{,}88$.

Значение оценки (6.150) энергетической эффективности модана, рассчитанного по методу Кирка–Джонса, для данного набора параметров составило $E_K = 0{,}16$. Технологическая реализация рассчитанного модана производилась равномерным шестнадцатиуровневым травлением резиста ПММА, нанесенного на кварцевое стекло, в Институте прикладной физики Университета Фридриха Шиллера (г. Йена, Германия). Способность изготовленного элемента формировать заданное квазиодномодовое распределение была исследована с помощью установки, оптическая схема

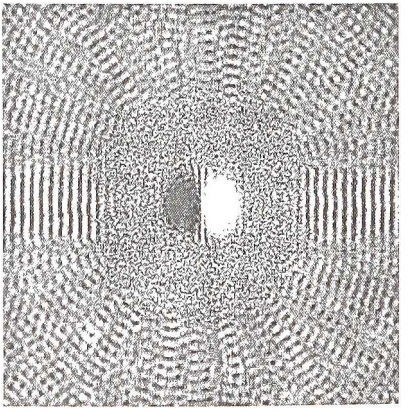


Рис. 6.32. Фазовая функция изготовленного модана

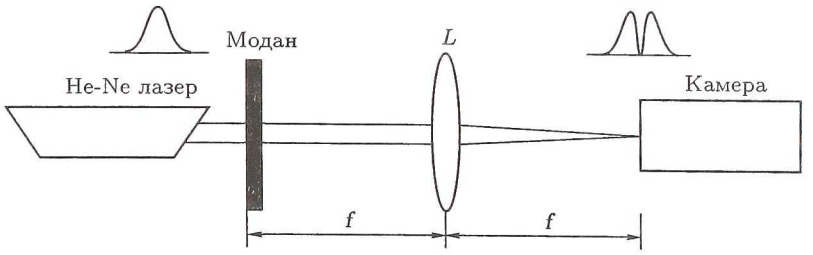


Рис. 6.33. Схема оптического эксперимента

которой приведена на рис. 6.33. Изготовленный элемент освещался пучком лазера HNA-180 (производство фирмы «Carl-Zeiss Jena», Германия). За фокальной плоскостью собирающей линзы с помощью камеры SPIRICON исследовалось формируемое распределение интенсивности и замерялась энергетическая эффективность элемента. Для измерения энергетической эффективности вспомогательная область в фокальной плоскости собирающей линзы L маскировалась с помощью непрозрачного экрана с отверстием в центре. Энергетическая эффективность в данном натурном эксперименте составила $E = 0{,}377$, что несколько меньше, чем значение, полученное в вычислительном эксперименте — $E = 0{,}455$. Данное расхождение может быть объяснено наличием погрешностей изготовления ДОЭ, а также погрешностей квантования и дискретизации.

На рис. 6.34 представлен результат натурного эксперимента — замеренная интенсивность сформированного квазиодномодового пучка.

Способность сформированного квазиодномодового пучка сохранять структуру, близкую к структуре моды Гаусса–Эрмита $(1,0)$, при прохождении через фурье-каскад исследовалась в ходе натурного эксперимента, схема которого представлена на рис. 6.35.

Вспомогательная область в фокальной плоскости собирающей линзы L_1 маскировалась с помощью непрозрачного экрана с отверстием в центре. Радиус отверстия выбирался несколько меньше внешнего радиуса полезной области — 1,7 мм. Полученный пучок формировался на входе фурье-каскада с линзой L_2, что позволило оценить модовые свойства путем оценки изменения структуры пучка после прохо-

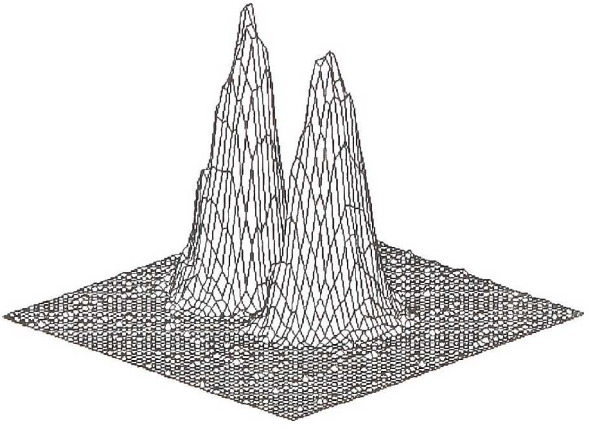


Рис. 6.34. Замеренная интенсивность сформированного пучка

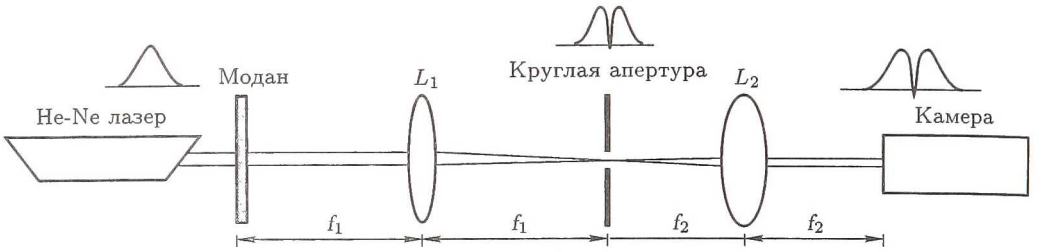


Рис. 6.35. Схема оптического эксперимента

ждения Фурье-каскада. Значение энергетической эффективности в данном случае составило $E = 0{,}362$, что вполне согласуется с полученными ранее результатами. На рис. 6.36 представлено двухмерное распределение интенсивности, замеренное в фокальной плоскости линзы L_2.

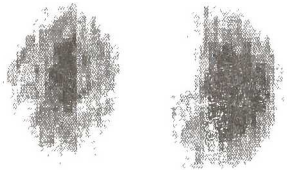


Рис. 6.36. Интенсивность пучка после прохождения Фурье-каскада

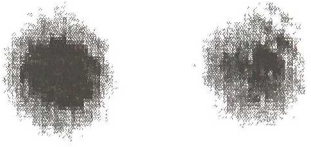


Рис. 6.37. Интенсивность пучка в плоскости, отстоящей на расстоянии $z = 240$ мм

Можно отметить хорошую устойчивость сформированного квазиодномодового пучка к преобразованию Фурье. Модовая структура исследовалась также в плоскости, отстоящей от фокальной плоскости линзы L_2 на расстоянии $z = 240$ мм. На рис. 6.37 представлено распределение интенсивности, замеренное в этой плоскости. Сравнение рис. 6.36 и 6.37 наглядно подтверждает устойчивость сформированного квазиодномодового пучка к распространению в свободном пространстве. Для исследования фазовой структуры сформированного распределения была использована установка, оптическая схема которой приведена на рис. 6.38. На рис. 6.39 приведено полученное распределение интенсивности.

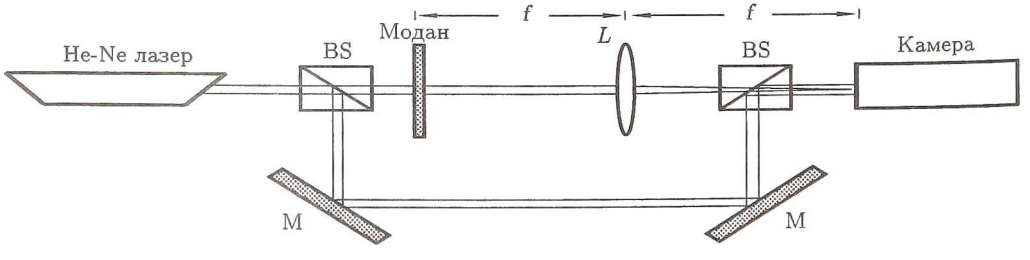


Рис. 6.38. Схема оптического эксперимента

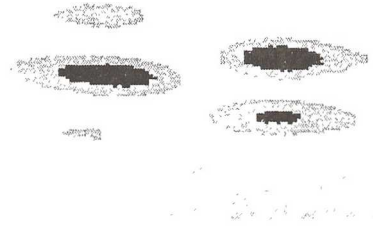


Рис. 6.39. Распределение интенсивности, полученное в ходе эксперимента, схема которого приведена на рис. 6.38

По сдвигу между интерференционными полосами был сделан вывод, что разность фаз на «долях» квазиодномодового распределения составляет величину, близкую к π рад. Это подтверждает способность изготовленного модана формировать распределение заданной фазовой структуры. Фазовая структура квазиодномодово-

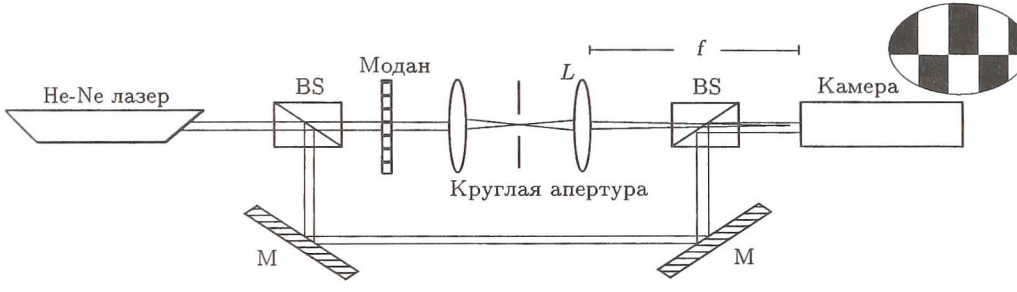


Рис. 6.40. Схема оптического эксперимента

го распределения после прохождения фурье-каскада исследовалась с помощью установки, оптическая схема которой приведена на рис. 6.40. На рис. 6.41 приведено распределение интенсивности, полученное в ходе натурного эксперимента, схема которого представлена на рис. 6.40. Результаты проведенного исследования показывают устойчивость фазовой структуры сформированного пучка к оптическому аналогу фурье-преобразования.

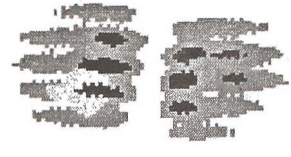


Рис. 6.41. Распределение интенсивности, полученное в ходе эксперимента, схема которого приведена на рис. 6.40

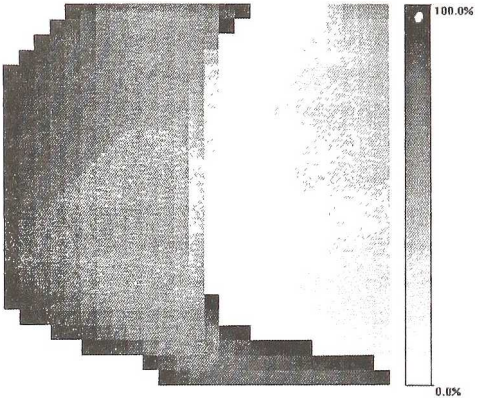


Рис. 6.42. Итеративно восстановленная фаза моды Гаусса–Эрмита $(0,1)$ (белый цвет соответствует значению фазы π, черный — 0)

В дополнение к результатам интерферометрического исследования фазовой структуры применялся численный метод восстановления фазы [48] пучка в фокальной плоскости первой Фурье-линзы: распределение фазы восстанавливалось по результатам измерения распределения интенсивности во входной и выходной плоскостях, соответственно, фурье-линзы L_2, в ходе 30 итераций процедуры [48]. Схема экспериментальной установки для получения двух распределений приведена на рис. 6.35. После 30 итераций, среднеквадратичное отклонение экспериментально полученного амплитудного распределения от его оценки на последней итерации составляло менее 17%. Восстановленное фазовое распределение во входной плоскости фурье-линзы представлено на рис. 6.42. Фазовый сдвиг между половинками моды составляет около $0{,}85\pi$, что согласуется с результатами интерферометрии и теоретической оценкой — π. Таким образом, устойчивость амплитудно-фазовой структуры гауссовых мод к фурье-преобразованию позволяет использовать итеративную процедуру [48], основанную на вычислении прямого и обратного преобразований Фурье, для верификации результатов интерферометрического исследования фазовой структуры сформированного модового пучка (см. рис. 6.39, 6.41, 6.42).

Результаты проведенного натурного эксперимента полностью согласуются с результатами вычислительного эксперимента, что подтверждает перспективность подхода [27, 28] к решению задачи формирования одномодовых пучков с высокой эффективностью.

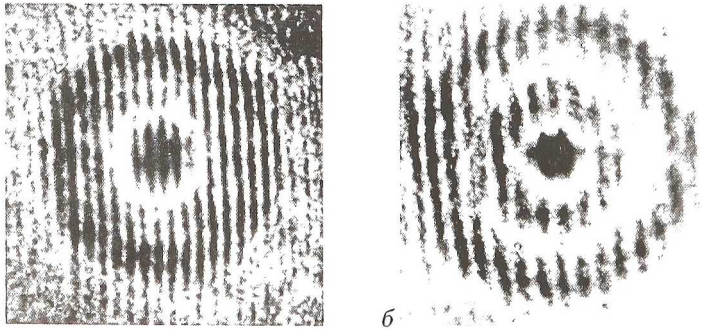


Рис. 6.43. ($а$) Интерферограмма моды Гаусса–Лагерра $(1,0)$, ($б$) интерферограмма моды Гаусса–Лагерра $(2,0)$

Экспериментальное исследование фазовых моданов Гаусса–Лагерра. Результаты экспериментального исследования моданов Гаусса–Лагерра, рассчитанных с помощью обобщенного метода Кирка–Джонса [19], приведены в работе [33]. На рис. 6.43 представлены интерферограммы сформированных модовых пучков Гаусса–Лагерра $(1,0)$ и $(2,0)$ [33]. Отметим изменение фазы на π на участках с отрицательными значениями модовой функции. На рис. 6.44 представлена осциллограмма

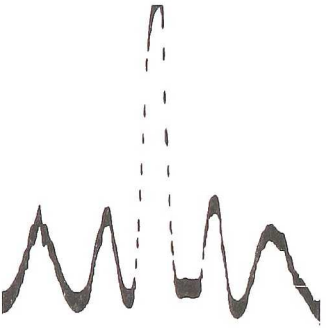


Рис. 6.44. Осциллограмма сигнала с ПЗС-линейки, полученная при измерении моды Гаусса–Лагерра $(2,0)$

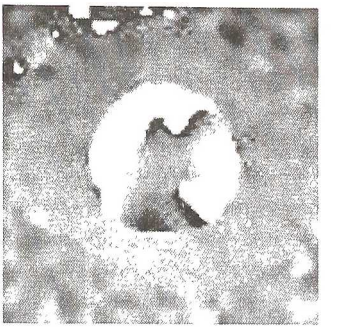


Рис. 6.45. Фаза моды Гаусса–Лагерра $(2,0)$, итеративно восстановленная по двум измерениям интенсивности (белый цвет соответствует значению фазы 2π, черный — 0)

сигнала с ПЗС-линейки при исследовании распределения интенсивности моды Гаусса–Лагерра $(2,0)$ [33]. Вследствие радиальной симметрии при $l = 0$ для исследования достаточно одного сечения, проходящего через центр симметрии. Симметрия кривой на рис. 6.44 несколько нарушена из-за наличия шумов. В целом этот результат демонстрирует хорошее соответствие эксперимента с теорией. Среднеквадратичное отклонение составило порядка 12%. В дополнение к результатам интерферометрического исследования фазовой структуры [33] в [34] также применялся численный метод восстановления фазы [48] пучка в фокальной плоскости первой фурье-линзы: распределение фазы восстанавливалось по результатам измерения распределения интенсивности во входной и выходной плоскостях фурье-линзы. Восстановленное фазовое распределение во входной плоскости фурье-линзы представлено на рис. 6.45.

6.2.10. Натурное исследование возможности уплотнения каналов оптической связи с помощью селективного возбуждения мод Гаусса–Эрмита

В [29, 34] приведены результаты натурного эксперимента по построению и исследованию модели двухканальной связи, использующей для передачи двух независимых каналов информации моды Гаусса–Эрмита $(1,0)$ и $(0,1)$, сформированные с помощью одномодовых моданов, рассчитанных итеративной процедурой (6.128), (6.129) с операторами (6.134) и (6.138). Для постановки этого эксперимента была собрана оптическая установка, схема которой представлена на рис. 6.46, имевшая следующие физические параметры: $f_1 = 452$ мм, $f_2 = 840$ мм, $f_3 = 300$ мм, диаметр отверстия круговой апертуры составлял 2,7 мм, модовый параметр пучка $\sigma_{10} =$

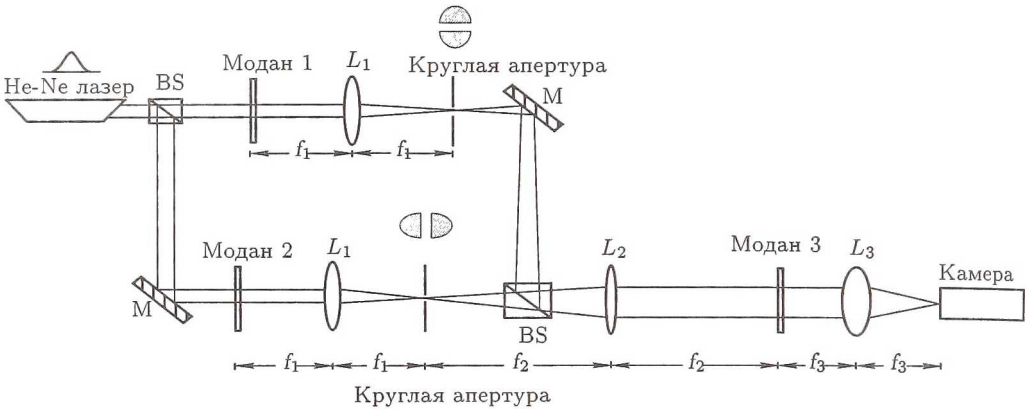


Рис. 6.46. Схема оптической установки для экспериментального исследования селекции мод с помощью ДОЭ: BS — делитель пучка, M — зеркало, L_1, L_2, L_3 — линзы

$= \sigma_{01} = 0{,}58$ мм в плоскости анализирующего модана 3, выбранный в соответствии с размером эффективной области дифракции 4,096 мм $\times$ 4,096 мм, размер пиксела использованной CCD камеры составлял 13 мкм. Для пространственного разделения корреляционных пиков, соответствующих модам Гаусса–Эрмита $(1,0)$ и $(0,1)$, в фазовую функцию модана 3 вводилась пространственная несущая, соответствующая расстоянию 30 мм между пиками в фокальной плоскости линзы L_3.

Линза L_2 в данной схеме выполняет роль «модели» линзоподобной среды, в которой распространяются сгенерированные моды: с другой стороны поведение модового пучка в ее выходной плоскости соответствует дифракции в дальней зоне, например при организации оптической связи в свободном пространстве. Подробно о линзоподобных средах и линзовых волноводах можно прочитать в [15]. Анализирующий модан 3, согласованный с модами $(1,0)$ и $(0,1)$, был рассчитан с помощью метода Кирка–Джонса [19] и изготовлен с помощью той же технологии, что и формирующие моданы [27–29]. Приведение функции комплексного пропускания модана 3 к чисто фазовому виду осуществлялось с помощью введения в фазу ДОЭ прямоугольно-импульсной несущей с частотой 41,7 линий/мм.

На первом этапе с помощью вычислительного эксперимента была исследована работа анализирующего модана 3, освещенного одномодовым пучком Гаусса–Эрмита $(1,0)$ или Гаусса–Эрмита $(0,1)$ радиусом $\sigma_{10} = 0{,}58$ мм. После применения быстрого преобразования Фурье было получено расчетное амплитудно-фазовое распределение в фокальной плоскости фурье-линзы L_3. Рассчитанные распределения амплитуды, фазы и интенсивности в фокальной плоскости фурье-линзы L_3 представлены на рис. 6.47. Расчет велся без учета квантованного характера фазы ДОЭ.

Отметим наличие на рисунках *б*), *г*) корреляционного пика, соответствующего наличию освещающей моды, и его отсутствие на рисунках *а*), *в*), что соответствует отсутствию детектируемой моды $(0,1)$. На рисунках можно наблюдать лишь четыре паразитных пятна вокруг точки, соответствующей положению корреляционного пика. Расстояние между точкой, соответствующей положению корреляционного пика, и центрами паразитных пятен составляло 150 мкм.

Соотношение максимумов интенсивности корреляционных пиков, наблюдаемых в случае наличия детектируемой моды, и максимумов интенсивности паразитных пятен в случае отсутствия детектируемой моды составляло в центре примерно 3:1. В силу специфики микролитографической технологии, изготовленный элемент имел 15 уровней квантования фазы, теоретически соответствующих [29, 30] 1200 нм.

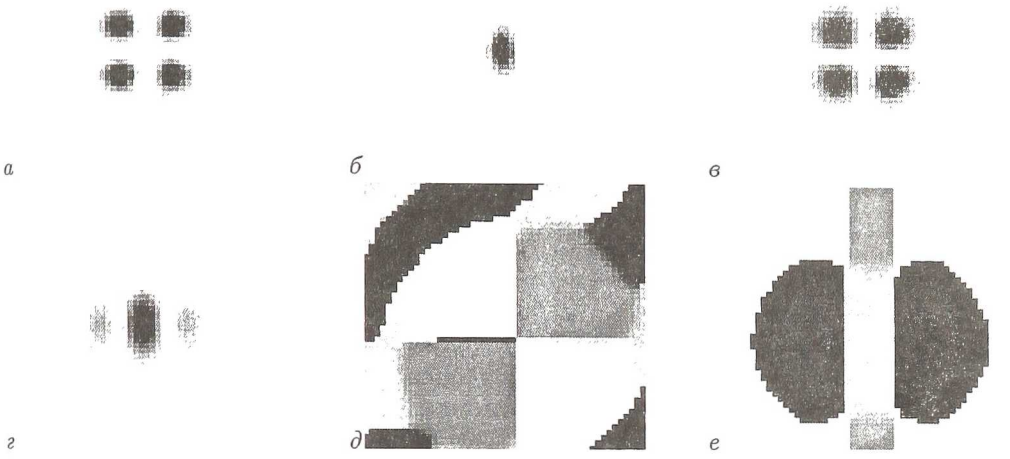


Рис. 6.47. Результаты вычислительного эксперимента по анализу входного пучка с помощью рассчитанного ДОЭ (без учета квантования): расчетные распределения интенсивности (*а*), (*б*), амплитуды (*в*), (*г*) и фазы (*д*), (*е*) в окрестности точек формирования корреляционных пиков, соответствующих модам Гаусса–Эрмита $(0,1)$ и $(1,0)$

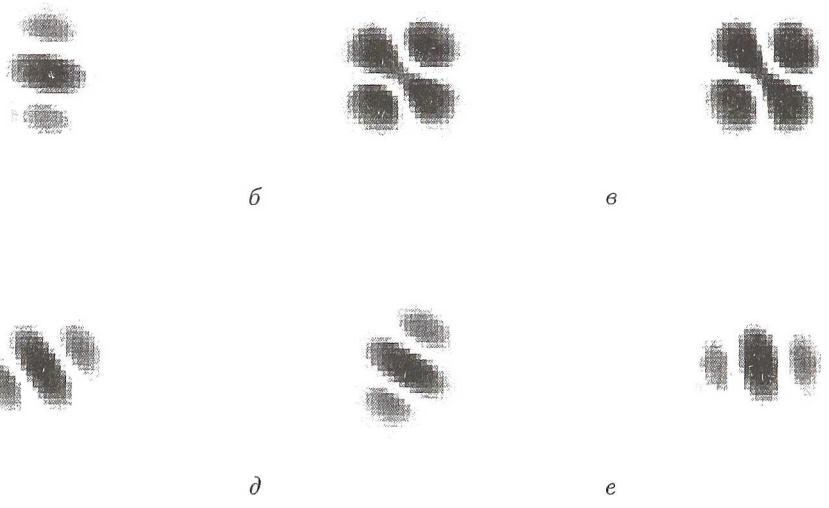
Однако, из-за трудностей контроля глубины травления максимальная глубина травления рельефа составляла лишь 70% от рассчитанного значения 1200 нм. Исходя из этого, осуществлялось моделирование с помощью быстрого преобразования Фурье работы элемента с максимальной глубиной травления 70% от требуемого для случаев, когда элемент освещен одномодовым пучком $(1,0)$ или $(0,1)$ и пучком, содержащим обе моды с равными мощностями. Также учитывалось квантование фазы элемента по 16 уровням. Полученные в ходе этого вычислительного эксперимента распределения интенсивности представлены на рис. 6.48.

На рис. 6.48*а,в,д* представлено распределение интенсивности в окрестности левого корреляционного пика, соответствующего моде Гаусса–Эрмита $(0,1)$, на рис. 6.48*б,г,е* — распределение интенсивности в окрестности правого корреляционного пика, соответствующего моде Гаусса–Эрмита $(1,0)$.

Рис. 6.48*а,б* соответствуют случаю, когда элемент освещен одномодовым пучком Гаусса–Эрмита $(0,1)$, рис. 6.48*в,г* соответствуют моде Гаусса–Эрмита $(1,0)$, рис. 6.48*д,е* соответствуют случаю освещения элемента суперпозицией этих двух мод с равными мощностями.

В заключение, на рис. 6.49 представлены результаты экспериментального исследования, выполненного с использованием экспериментальной установки, представленной на рис. 6.46, при том же наборе физических параметров, что использовались в компьютерном моделировании, результаты которого представлены на рис. 6.48.

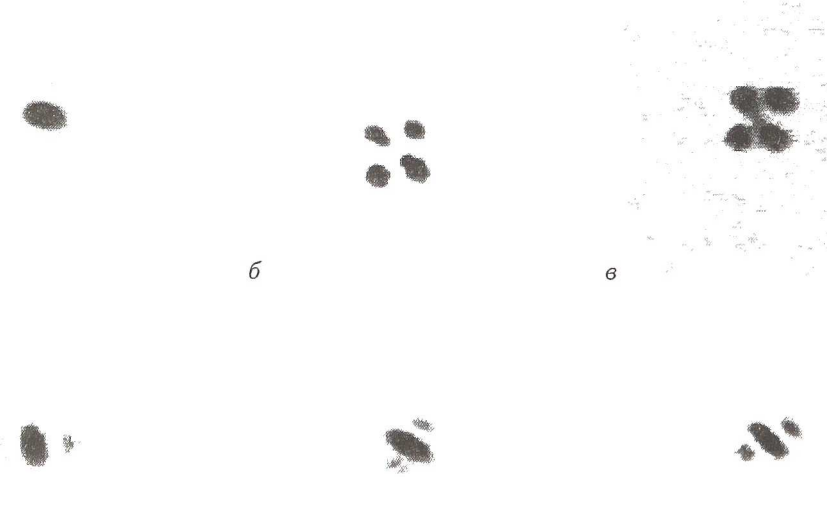
Отметим хорошее соответствие результатов физического эксперимента результатам компьютерного моделирования, полученных для случая 70% реализации расчетной глубины профиля. Учитывая физический размер чувствительного элемента CCD-камеры и результаты моделирования, можно сделать вывод о качественном соответствии экспериментально полученных распределений в фокальной плоскости линзы L_3 расчетным результатам, даже несмотря на небольшие различия в форме корреляционных пиков.



а *б* *в*

г *д* *е*

Рис. 6.48. Результаты вычислительного эксперимента по анализу входного пучка с помощью рассчитанного ДОЭ: расчетные распределения интенсивности в окрестности точек формирования корреляционных пиков, соответствующих модам Гаусса–Эрмита $(0,1)$ и $(1,0)$ при освещении модана модовым пучком Гаусса–Эрмита $(0,1)$ (*а*), (*б*); Гаусса–Эрмита $(1,0)$ (*в*), (*г*) и пучком, содержащим обе моды (*д*), (*е*).



а *б* *в*

г *д* *е*

Рис. 6.49. Результаты измерений сформированной интенсивности в окрестности центров корреляционных пиков, соответствующих модам Гаусса–Эрмита $(0,1)$ и $(1,0)$ при освещении модана модовым пучком Гаусса–Эрмита $(0,1)$ (*а*), (*б*), Гаусса–Эрмита $(1,0)$ (*в*), (*г*) и пучком, содержащим обе моды (*д*), (*е*)

В работе [29] на основе результатов поставленных вычислительных и натурных экспериментов была продемонстрирована принципиальная осуществимость концепции применения поперечных мод лазерного излучения в качестве носителей независимых каналов телекоммуникационной системы, рассмотренной в [19]. Показана целесообразность применения моданов, рассчитанных с помощью итеративной процедуры [27] и метода синтетических коэффициентов [39], и изготовленных с помощью электронной микролитографии.

Аналогичные результаты для мод Гаусса–Лагерра приведены в [33]. Для измерения поперечно-модового состава эталонного многомодового пучка в качестве фильтра ставился пространственный фильтр, рассчитанный с помощью метода Кирка–Джонса [19], функция комплексного пропускания которого представляла собой суперпозицию заданных мод Гаусса–Лагерра. В точках выходной плоскости, соответствующих корреляционным пикам измеряемых мод, помещалась ПЗС–линейка. В соответствии с (6.177) мощность измеряемой моды с точностью до постоянного множителя равна интенсивности света в центре соответствующего порядка дифракции.

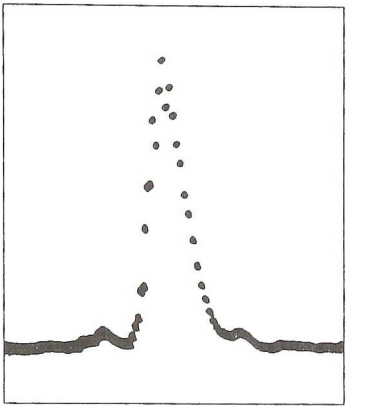


Рис. 6.50. Осциллограмма сигнала с ПЗС–линейки, при детектировании моды Гаусса–Лагерра $(1,0)$ возбужденной с помощью модана

На рис. 6.50 представлена осциллограмма сигнала с ПЗС-линейки, полученная при измерении распределения интенсивности в 1-м порядке дифракции модана, согласованного с модой Гаусса–Лагерра $(1,0)$ в момент детектирования моды. Отклонение измеренных коэффициентов от теоретических составляло не более 5% от максимального значения [33].

6.2.11. Расчет ДОЭ, согласованных с модами градиентных волокон с непараболическим профилем

Современные технологии изготовления градиентных оптических волокон далеко не всегда обеспечивают формирование заданного профиля с высокой точностью [18]. На рис. 6.51 представлено характерное распределение показателя преломления реального оптоволокна, производимого промышленностью. Схематично показан центральный провал показателя преломления, типично возникающий из-за специфики технологии парафазного осаждения изготовления оптоволокон [18].

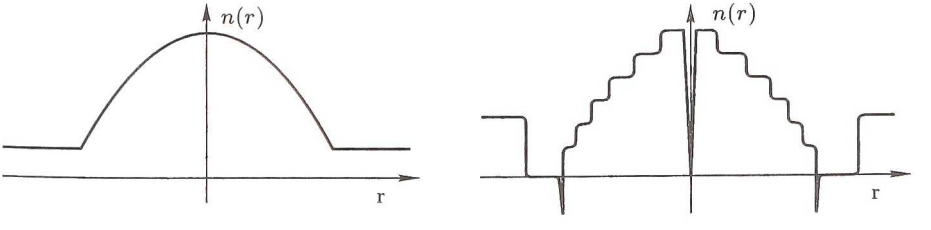


Рис. 6.51. Слева — идеальный параболический профиль, справа — типичный профиль кварцевого волокна, производимого промышленностью

Кроме технологических отклонений от заданного параболического профиля, имеются волноводы с расчетной формой профиля, отличной от параболической [18]. Поэтому представляется актуальной задача расчета ДОЭ, согласованных с модами градиентных волокон с непараболическим профилем. В [18] показано, для описания распространения света в слабонаправляющих градиентных волокнах вполне применима скалярная дифракционная теория без учета поляризационных эффектов. Там же рассмотрен поиск собственных функций оператора распространения света в градиентных волокнах с различными заданными формами профиля. В случае небольших отклонений реального профиля от заданного целесообразно применить теорию возмущений. В [18] теория возмущений первого порядка была использована для определения постоянных распространения и решений скалярного волнового уравнения. Было показано [18], что для градиентных волокон с круговой симметрией возмущенного показателя преломления в приближении первого порядка можно получить следующие оценки постоянной распространения и модовых распределений:

$$\beta_{\mathbf{p}}'^2 = \beta_{\mathbf{p}}^2 + k^2 \int_0^\infty \psi_{\mathbf{p}}^2(r) n^2(r) r \, \mathrm{d}r, \tag{6.191}$$

$$\psi_{\mathbf{p}}'(r) = \psi_{\mathbf{p}}(r) + \sum_{\substack{k \\ (p \neq k)}} \frac{\psi_{\mathbf{p}}(r) \int_0^\infty \psi_{\mathbf{p}}(r) \psi_{kl}(r) n^2(r) r \, \mathrm{d}r}{\left(\beta_{\mathbf{p}}^2 - \beta_{kl}^2\right)}, \tag{6.192}$$

где $\beta_{\mathbf{p}}'$ и $\psi_{\mathbf{p}}'(r)$ — постоянная распространения и поле моды $\mathbf{p} = (p, l)$ для возмущенного профиля, а β_{pl} и ψ_{pl} — соответствующие величины для невозмущенного профиля $n(r)$, $\delta n^2(r)$ — возмущение профиля. Таким образом, появляется возможность расчета комплексных полей мод градиентных волокон с профилем, отклоняющимся от идеального и, соответственно, расчета ДОЭ, согласованных с модами таких волокон. Однако, необходимо помнить, что моды градиентных волокон с профилем, отличным от параболического, не обладают свойством самовоспроизводимости в пространстве в отличие от гауссовых мод. Поэтому пересчет поля с торца волокна с таким профилем в плоскость модана, согласованного с заданным модовым распределением, в этом случае должен производиться с применением строгого вычисления интеграла Кирхгофа–Френеля [10]. За исключением этого момента все развитые выше подходы к расчету ДОЭ, согласованных с гауссовыми модами, могут быть без труда перенесены на ДОЭ, согласованные с модами реальных градиентных непараболических волноводов. Расчет ДОЭ, формирующего заданную моду непараболического волокна, может осуществляться как с помощью внесения несущей в фазу элемента подобно [19], так и с помощью итеративных процедур типа [27]. Если соотношение фазовых межмодовых задержек в формируемом многомодовом пучке не является значимым, итеративный расчет ДОЭ формирующего группу заданных мод непараболического волокна может быть осуществлен итеративно, используя ортогональность множества модовых функций (6.3), аналогично процедуре (6.158)–(6.160), построенной для формирования группы гауссовых мод. Однако, как и было отмечено выше, в случае негауссовых мод разложение, аналогичное (6.167), следует производить, используя вычисление интеграла Кирхгофа–Френеля [10] для пересчета комплексной амплитуды.

Дифракционный анализ основан на решении (разными путями) волнового уравнения, которое верно описывает распространение света (и в частности дифракцию) только при одном из поляризационных состояний TE или TM. Это связано с тем,

что электромагнитную волну можно полностью описать, зная как минимум две ее продольные составляющие — электрическую (TE волна) и магнитную (TM волна). Составление и решение волнового уравнения относительно одной продольной составляющей обосновано в случаях однородной среды или среды с цилиндрическими неоднородностями. Иначе неизбежно появление мод с иной поляризацией. В таком случае решение волнового уравнения даст искаженный модовый состав, в котором будут присутствовать моды слишком высоких порядков, не наблюдающихся в реальности, и отсутствовать реальные моды, иначе поляризованные. Следовательно, дифракционная теория перестанет быть удовлетворительной при расчете прохождения света по волноводу с нецилиндрическими неоднородностями, геометрические размеры которых превышают четверть волны.

Таким образом, в тех случаях, когда для описания градиентного многомодового оптического волокна можно использовать дифракционную скалярную теорию, для возбуждения и селекции заданных мод волновода целесообразно использовать дифракционные оптические элементы, методы расчета которых описаны в 6.2.3–6.2.6.

6.3. Применение ДОЭ в системах сбора, передачи и хранения информации

Обладая уникальными по сравнению с классическими оптическими элементами качественными характеристиками [23], элементы компьютерной оптики находят широкое применение для решения различных задач, возникающих при построении систем сбора, передачи и хранения информации, в частности, оптических телекоммуникационных систем [19], систем параллельной обработки сигналов и нейрооптических сетей [55, 56, 57], построении когерентных оптических вычислительных машин [58, 59], построении высокочувствительных волоконно-оптических датчиков [60, 61], считывающих головок для CD-проигрывателей [62] и др. Рассмотрим применение ДОЭ для решения некоторых из этих задач более подробно в данном пункте. Расчет и реализация с помощью технологий дифракционной оптики оптических пространственных фильтров, широко применяемых в оптических информационных системах, подробно рассмотрены в главе 10 данной книги, а также в работах [19, 58, 59].

6.3.1. Повышение пропускной способности систем оптической связи

Повышение пропускной способности современных телекоммуникационных систем является важнейшей научно-технической задачей, требующей дальнейшего исследования физических (в том числе оптических) эффектов. Одним из наиболее привлекательных подходов к решению этой задачи является поиск возможностей увеличения числа каналов без построения дополнительных физических линий связи. В случае, если в качестве носителя информации рассматривается когерентный световой пучок, а в качестве линии связи используется линейная волноводная среда (например, оптический световод), весьма перспективным представляется параллельное использование различных мод лазерного излучения (как продольных так и поперечных) для передачи отдельных каналов. В настоящее время уже получили достаточно широкое распространение телекоммуникационные системы, использующие для передачи различных каналов различные продольные моды, то есть различные длины волн излучения, распространяющегося в волноводе [63–65]. Такие системы, получившие название WDM-систем (Wavelength Division Multiplexing systems), ис-

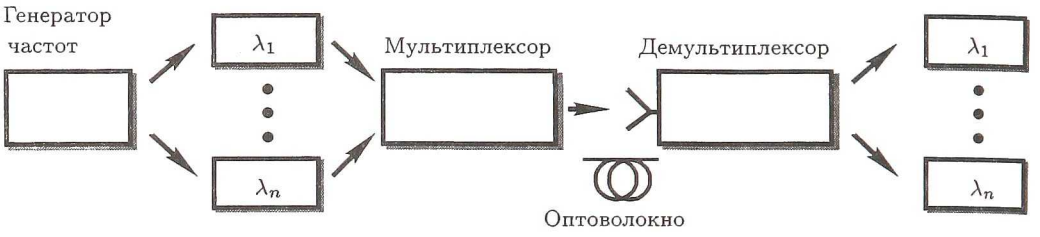


Рис. 6.52. Блок-схема телекоммуникационной WDM-системы

пользуют в качестве демультиплексоров оптические фильтры, разделяющие различные длины волн (рис. 6.52). Последние достижения в разработке диэлектрических полосовых фильтров с шириной полосы пропускания $\Delta\lambda \leqslant 1$ нм, пригодных к массовому производству, позволили уже сегодня достичь значительного уплотнения каналов связи в подобных системах.

В работе [64] рассмотрено построение двухканальной волоконно-оптической системы связи, основанной на передаче по одномодовому волокну двух независимых каналов с длинами волн 1,3 и 1,55 мкм. Для селекции каналов на выходе волокна использовалась голографическая дифракционная решетка. Для этих же целей помимо дифракционных решеток могут применяться спектральные дифракционные элементы, согласованные с несколькими длинами волн [66, 67]. В работе [65] предложена система прямой передачи изображений по оптическому волокну с использованием разложения белого света по спектральным компонентам. Селекция компонент в [65] осуществляется с помощью сегментированного голографического оптического элемента, каждый сегмент которого согласован с определенным диапазоном спектра.

Интересно рассмотреть также поперечные моды в качестве независимых носителей информационных каналов вместо используемых продольных мод (а может быть, и в дополнение к ним). Как было сказано выше, поперечные моды лазерного излучения представляют собой пучки света, распределение комплексной амплитуды в сечении которых описывается собственными функциями оператора распространения света в соответствующей среде. Фундаментальным свойством мод является сохранение структуры и взаимной ортогональности при распространении в среде. Именно это свойство поперечных мод является основой для построения систем связи с модовым уплотнением каналов. Интерес к поперечным модам как носителям независимых каналов передачи информации связан, во-первых, с постоянным повышением качества производимых многомодовых волокон [см., например, 68], во-вторых, с разработкой методов качественного синтеза дифракционных оптических элементов — моданов [19, 27–30], способных эффективно формировать и селектировать поперечные моды лазерного излучения (см. также 6.2 данной книги). Общая теория построения телекоммуникационных систем с уплотнением каналов, основанном на использовании поперечных мод, детально изложена в [19]. Отметим, что селективное возбуждение поперечных мод оптоволокна позволит увеличить пропускную способность линии связи не только за счет параллельной передачи нескольких каналов по одному волокну, но и за счет решения проблемы уширения импульса, вызываемого наличием межмодовой дисперсии [18–20, 6.2.7]. Одна из предполагаемых инженерных реализаций волоконно-оптической связи с использованием селективного возбуждения поперечных мод [19] представлена на рис. 6.53. Пространственный фильтр МА является матрицей электрооптических модуляторов, освещаемых плоской волной когерентного света $F_0(\mathbf{x})$. На матрицу электрооптических модуляторов непосредственно подается вектор промодулированных по времени сигналов $\{S_i\}$.

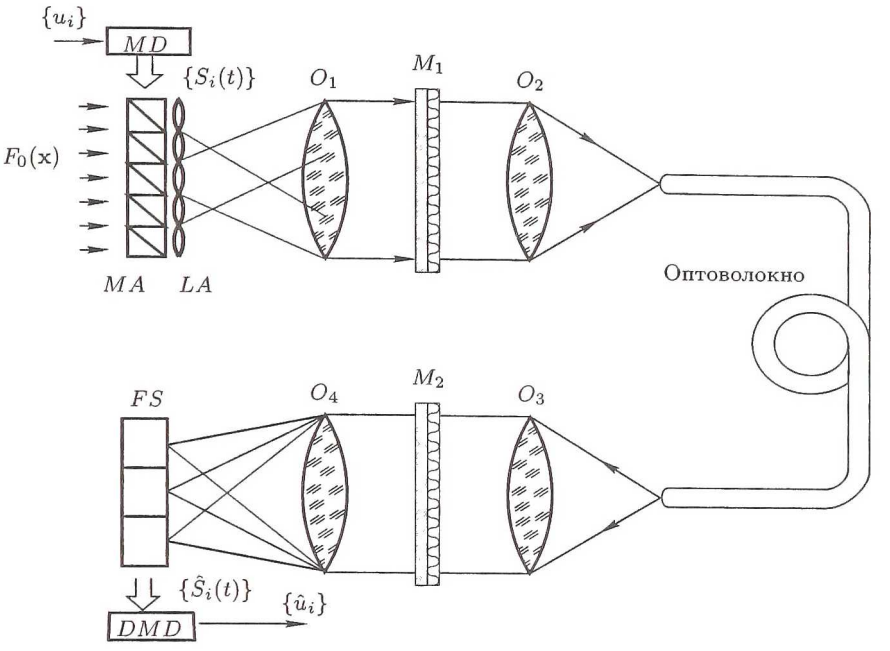


Рис. 6.53. Схема волоконно-оптической линии связи с использованием модовой селекции; $F_0(\mathrm{x})$ — освещающий пучок; M_1 — вектор входного сигнала; $\{S_i(t)\}$ — вектор-сигнал, промодулированный по времени; MD — модулятор электрического сигнала, MA — матрица акусто- или электро-оптических модуляторов; LA — матрица микролинз; O_1, O_2, O_3, O_4 — объективы; M_1, M_2 — многомодовые моданы; FS — матрица фотоприемника; DMD — демодулятор электрического сигнала; $\{\hat{S}_i(t)\}$ — электрический вектор-сигнал, промодулированный по времени.

Фурье-объектив O_1 преобразует набор точечных источников в набор плоских волн, освещающих модан M_1.

Модан M_1 является пространственным фильтром, согласованным с множеством мод, кодированных с внесением различных пространственных частот [19, 6.2.]. Значения частот пространственных несущих модана M_1 согласованы с координатами ячеек модулятора и их расстояний от объектива O_1. На выходе модана M_1 формируется суперпозиция мод, с мощностями, пропорциональными интенсивности света, прошедшего через соответствующие ячейки модулятора MA. Микрообъектив O_2 предназначен для приведения радиуса сформированных мод к радиусу принципиальной моды «идеального» оптоволокна F.

Многомодовый пучок, введенный в волокно анализируется на выходе оптоволоконной линии. Микрообъектив O_3 согласует радиус принципиальной моды оптоволокна и радиус мод в плоскости модана M_2. Модан M_2 согласован с сопряженными модами, кодированными с использованием различных несущих. Значения интенсивности света на различных ячейках фотоприемника FS пропорциональны мощности соответствующих мод. Заметим, что из-за разных значений фазовых скоростей различных мод значения их мощностей должны замеряться с временными задержками, пропорциональными константам распространения. Для компенсации различных значений модовой задержки и затухания целесообразно ввести компенсирующие фазовые добавки в каналах модана M_2, определяемые уравнением

$$\alpha^{(\mathrm{p})} = (\beta_{\max} - \beta_{\mathrm{p}})\, z,$$

где $\beta_{\mathbf{p}}$ — константа распространения для моды с индексом $\mathbf{p}$, $\beta_{\max} = \max\limits_{\mathbf{p}} \beta_{\mathbf{p}}$, и z — длина оптоволокна. Значения мощностных коэффициентов в каналах выбираются следующим образом:

$$P^{(\mathbf{p})} = \frac{1}{\rho_{\mathbf{p}}}, \tag{6.193}$$

где $\rho_{\mathbf{p}}$ — различные значения затухания мод. В случае необходимости осуществления фильтрации целесообразно выбирать коэффициенты фильтрации в виде

$$P^{(\mathbf{p})} = \frac{1}{\rho_{\mathbf{p}}} \cdot \frac{1}{1 + S_{\mathbf{p}}/N_{\mathbf{p}}}, \tag{6.194}$$

где $S_{\mathbf{p}}$ — математическое ожидание мощности сигнала, соответствующего моде с номером $\mathbf{p}$ на входе анализатора, и $N_{\mathbf{p}}$ — математическое ожидание шума в канале, соответствующего моде с номером $\mathbf{p}$.

Выше речь шла о случае линейного распространения светового пучка в волноводе (волокне). В настоящее время большой интерес также вызывает возможность применения нелинейных оптических эффектов для повышения пропускной способности волоконных линий связи [69, 70]. В работе [71] рассматривается возможность использования селективного возбуждения поперечных мод градиентного многомодового волокна в условиях нелинейности для повышения пропускной способности. Там же приведены результаты численного эксперимента по моделированию работы такого канала связи. Формирование заданного модового состава может быть полезно для повышения качества не только оптической волноводной связи, но и оптической связи в свободном пространстве. В работе [72] приводятся результаты сравнительного исследования применения гауссова пучка (гауссовой моды $(0,0)$) и нулевой моды Бесселя, являющейся модой свободного пространства [23], для оптической передачи информации в свободном пространстве.

Описанный в 6.2.8. способ измерения модового состава пучка с помощью модана может использоваться не только для построения многоканальной оптической связи (рис. 6.53), но и для контроля модового состава выходного пучка лазера. В работах [73, 74] описано численное восстановление комплексной поперечной структуры выходного пучка исследуемого лазера по результатам измерений интенсивности в ближней и дальней зоне. Такой подход, однако, не дает возможности анализировать модовый состав в режиме реального времени, в отличие от метода, представленного в 6.2.8.

6.3.2. Волоконно-оптические датчики

Широкое распространение получили оптоволоконные датчики микроперемещений, основанные на измерении полной световой энергии пучка, прошедшего через деформированное оптоволокно [75, 76]. Однако, это означает, что на выходе оптоволокна происходит измерение энергии всех выходящих мод. Так как в этом случае происходит усреднение мощности суммируемых мод, изменение полной энергии происходит достаточно медленно по сравнению с изменением мощности отдельных мод.

В работах [19, 60] был предложен альтернативный подход, заключающийся в селективном возбуждении мод в оптоволокне и измерении мощности этих мод на выходе деформированного волокна. В работах [77, 78] было показано, что чувствительность изменения мощности отдельной моды к деформации волокна гораздо

выше чувствительности изменения полной мощности светового пучка. Эта идея и послужила основой создания высокочувствительного волоконно-оптического датчика перемещений. Основная проблема в решении такой задачи — реализация селективного возбуждения заданных мод в оптоволокне и селективное измерение их мощности на выходе. Эта проблема может быть эффективно решена путем использования моданов, расчет и исследование которых рассмотрены в 6.2.

Рассмотрим датчик, основанный на использовании оптоволокна с направляемыми модами, описываемыми комплексными амплитудами $\{\psi_{\mathbf{p}}(\mathbf{x})\}$ в скалярном приближении с номерами мод $\mathbf{p} = (p, l)$. Декартовы координаты $\mathbf{x} = (x, y)$ введены в сечении оптоволокна. Условие ортогональности мод (6.3) выполняется на области $G(\rho_0 \equiv 1)$. Возбуждение оптоволокна означает освещение пучком, характеризуемым комплексной амплитудой

$$F(\mathbf{x}) = \sum_{\mathbf{p}} F_{\mathbf{p}} \psi_{\mathbf{p}}(\mathbf{x}) \tag{6.195}$$

с коэффициентами $\{F_{\mathbf{p}}\}$. Для описания распространения пучка света в оптоволокне, используем линейный оператор $\hat{\mathrm{T}}$ действующий на функцию $F(\mathbf{x})$. Комплексная амплитуда $F^1(\mathbf{x})$ пучка на выходе оптоволокна описывается уравнением $F^1 = \hat{\mathrm{T}}F$.

Модовый спектр $\{F_1\}$ выходного пучка

$$F_1(\mathbf{x}) = \sum_{\mathbf{p}'} F^1_{\mathbf{p}'} \psi_{\mathbf{p}'}(\mathbf{x}) = \sum_{\mathbf{p}} F_{\mathbf{p}} \hat{\mathrm{T}}_{\mathbf{p}}(\mathbf{x}) \tag{6.196}$$

определяется модовой композицией $\{F_{\mathbf{p}}\}$ входного пучка и свойствами волокна, определяемыми оператором $\hat{\mathrm{T}}$. Вид оператора $\hat{\mathrm{T}}$ зависит от состояния волокна, т.е. от возмущений его физических параметров, характеризуемых измеряемыми параметрами $\mathbf{V} = (V_1, V_2, \ldots, V_n)$. Таким образом, запишем $\hat{\mathrm{T}} = \hat{\mathrm{T}}(\mathbf{V})$, $F^1_{\mathbf{p}} = F^1_{\mathbf{p}}(\mathbf{V})$.

В отсутствие возмущений волокна имеем $\mathbf{V} = 0$, $\hat{\mathrm{T}} = \hat{\mathrm{T}}(0)$, где $\hat{\mathrm{T}}(0)$ — оператор, описывающий распространение моды в собственной среде. Используя определение константы распространения $\beta_{\mathbf{p}}$, запишем

$$\hat{\mathrm{T}}(0)\,\psi_{\mathbf{p}}(\mathbf{x}) = \exp(i\beta_{\mathbf{p}} z)\,\psi_{\mathbf{p}}(\mathbf{x}), \tag{6.197}$$

где z — пройденное модой расстояние в волокне. В случае отсутствия возмущений межмодовых взаимодействий нет:

$$F^1_{\mathbf{p}'} = \exp(i\beta_{\mathbf{p}'} z)\, F_{\mathbf{p}'}, \quad \left|F^1_{\mathbf{p}'}\right|^2 = \left|F_{\mathbf{p}'}\right|^2, \tag{6.198}$$

Это объясняется тем, что моды являются собственными функциями оператора распространения света в невозмущенном оптоволокне. Значения $\mathbf{p}$ соответствуют номерам направляемых мод волокна.

В присутствии ненулевого возмущения $\mathbf{V}$ оптоволокна, вместо уравнения (6.196) введем коэффициенты $T_{\mathbf{p}\mathbf{p}'}$, так что

$$\hat{\mathrm{T}}(\mathbf{V})\,\psi_{\mathbf{p}}(\mathbf{x}) = \sum_{\mathbf{p}'} T_{\mathbf{p}\mathbf{p}'}(\mathbf{V})\,\psi_{\mathbf{p}'}(\mathbf{x}). \tag{6.199}$$

Коэффициенты $T_{\mathbf{p}\mathbf{p}'}(\mathbf{V})$ представляют собой матрицу элементов оператора $\hat{\mathrm{T}}$ и соответствуют значениям интегралов перекрытия мод. Свойство ортогональности

(6.3) модового базиса $\psi_{\mathbf{p}}$ дает нам следующее уравнение:

$$T_{\mathbf{pp}'}(\mathbf{V}) = \int_G \psi^*_{\mathbf{p}'}(\mathbf{x})\hat{\mathrm{T}}\psi_{\mathbf{p}}(\mathbf{x})\,\mathrm{d}^2\mathbf{x}.$$

Коэффициенты $F^1_{\mathbf{p}'}$ могут быть определены уравнениями

$$F^1_{\mathbf{p}'}(\mathbf{V}) = \int_G F^1(\mathbf{x})\,\psi_{\mathbf{p}'}(\mathbf{x})\,\mathrm{d}^2\mathbf{x}, \tag{6.200}$$

$$F^1_{\mathbf{p}'}(\mathbf{V}) = \sum_{\mathbf{p}} T_{\mathbf{pp}'}(\mathbf{V})\,F_{\mathbf{p}}. \tag{6.201}$$

В силу выполнения равенства Парсеваля для ортогональных базисов мощности пучков на входе и выходе оптоволокна определяются рядами

$$E = \sum_{\mathbf{p}} |F_{\mathbf{p}}|^2, \quad E'(\mathbf{V}) = \sum_{\mathbf{p}'} \left|F^1_{\mathbf{p}'}(\mathbf{V})\right|^2 \tag{6.202}$$

Большинство известных волоконно-оптических датчиков основано на измерении полной энергии выходного пучка $E'(\mathbf{V})$ в зависимости от возмущений $\mathbf{V}$. Как видно из (6.202), значение $E'(\mathbf{V})$ вляется усредненным по мощностям отдельных мод $|F_{\mathbf{p}'}(\mathbf{V})|^2$ по всему ансамблю возбужденных мод с номерами $\mathbf{p}'$. Предложенный в [60] сенсор выполняет селективное измерение мощностей отдельных мод $\left|F^1_{\mathbf{p}'}(\mathbf{V})\right|^2$. Заметим, что в специальном случае возбуждения одной моды ненулевой коэффициент $F_{\mathbf{p}}$ будет только для одного значения $\mathbf{p}$. Значение $|T_{\mathbf{p}'\mathbf{p}}(\mathbf{V})|^2$ назовем коэффициентом возбуждения моды, так как оно определяет мощность моды с номером $\mathbf{p}'$ для случая, когда волокно возбуждается модой с номером $\mathbf{p}$. В работах [79–80] исследовалась зависимость значений коэффициентов возбуждения от различных значений возбуждаемой моды и условий возбуждения.

6.3.3. Экспериментальное исследование датчика микроперемещений, основанного на использовании модана

Создание модана позволило разработать семейство датчиков микроперемещений с высокой и управляемой чувствительностью [60]. Датчик микроперемещений, схема которого приведена на рис. 6.54, имеет повышенную чувствительность в силу того, что нтенсивность отдельно взятой моды более чувствительна к микросдвигам, чем полная интенсивность пучка. В этом состоит отличие приведенной схемы от традиционно используемой схемы датчика перемещений. В ходе экспериментального исследования высокочувствительного датчика было использовано градиентное волокно с параболическим профилем, имеющее длину 1200 мкм. Волокно испытывало единичный квазисинусоидный прогиб с варьируемой амплитудой h (рис. 6.54). Прогиб осуществлялся с помощью рабочей мембраны 4, имеющей одну точку касания с волокном, и контрольной мембраны с двумя точками касания (рис. 6.55).

Расстояние 1,0 мм между точками касания было выбрано для обеспечения создания резонанса мод [5]. Обе мембраны фиксировались между корпусом *3* и крышками *5* и *8*. Винт *9* производит начальную деформацию волокна. Измерение сдвига рабочей мембраны выполнялось с помощью индикатора *1*. Волокно F было пропущено

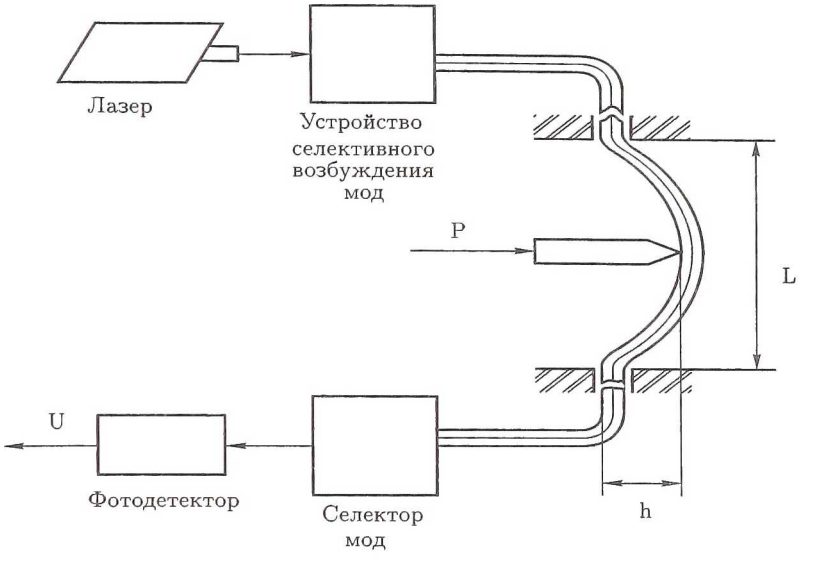


Рис. 6.54. Блок-диаграмма волоконно-оптического датчика микроперемещений. P — механический микроштифт, h — амплитуда микроизгиба волокна, U — выходное электрическое напряжение

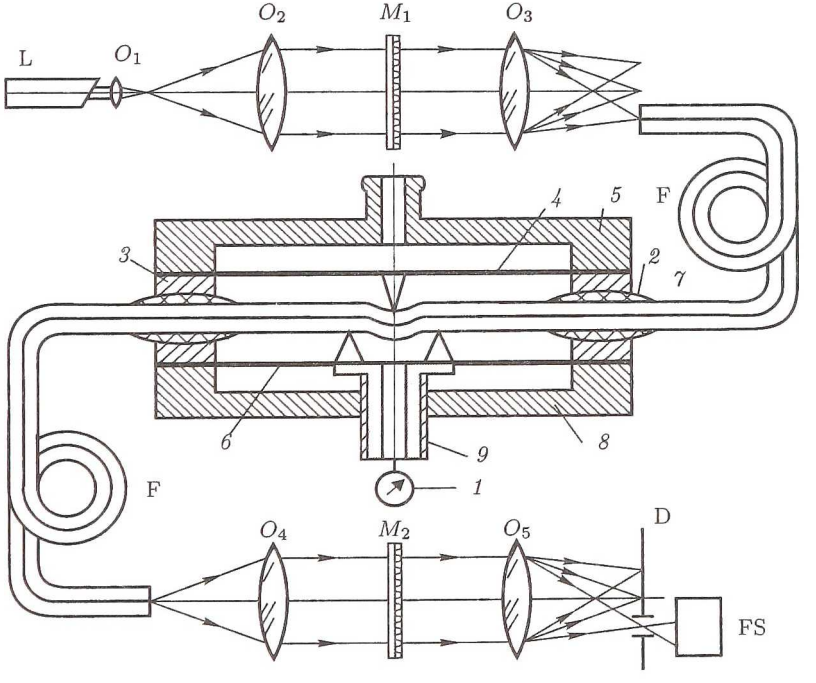


Рис. 6.55. Оптическая схема волоконно-оптического датчика микроперемещений. L — лазер; O_1, O_2, O_5 — объективы; O_3, O_4 — микрообъективы; M_1, M_2 — многопучковый модан; F — оптоволокно; D — диафрагма; FS — фотоприемник; 1 — индикатор перемещения мембраны; 2 — металлические микротрубки; 3 — корпус датчика; 4 — рабочая мембрана; 5 — верхняя крышка; 8 — нижняя крышка; 6 — контрольная мембрана; 7 — герметичный силиконовый фильтр; 9 — микровинт для получения деформации оптоволокна

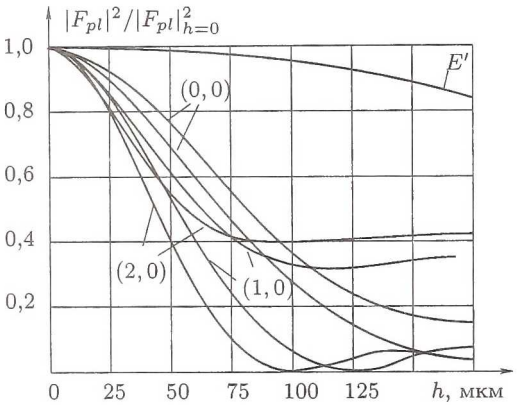


Рис. 6.56. Выходной сигнал датчика в зависимости от сдвига h рабочей диафрагмы E' — общий световой поток, прошедший через оптоволокно; парами чисел (0,0), (1,0), (2,0) помечены кривые, соответствующие селектируемой моде Гаусса–Лагерра

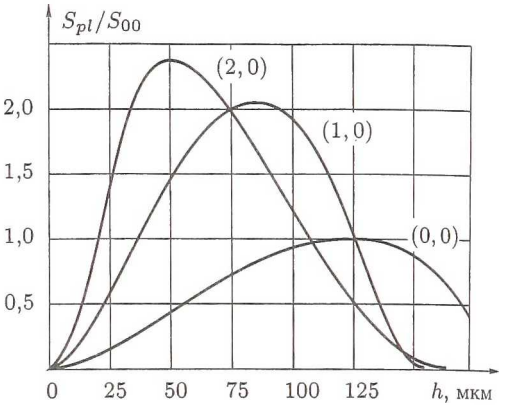


Рис. 6.57. Чувствительность датчика в зависимости от сдвига h рабочей диафрагмы. Парами чисел (0,0), (1,0), (2,0) помечены кривые, соответствующие селектируемым модам

через две металлических микротрубки *2* вмонтированных в корпус *3* и заполненных силиконовым герметическим фильтром *7*. Селективное возбуждение отдельных поперечных мод в оптоволокне происходит в формирователе мод, включающем в себя лазер L, коллиматор $O_1 - O_2$, модан M_1 и объектив O_3. Комплексная функция пропускания модана M_1 согласована с одной из поперечных мод волокна. Селективное выделение заданной моды на выходе волокна и измерение ее интенсивности происходит в анализаторе, состоящем из объектива O_4, второго модана M_2, объектива O_5 и фотоприемника FS с диафрагмой D. Комплексная функция пропускания модана M_2 также согласована с заданной модой волокна. Селективное возбуждение и детектирование моды обеспечивается выполнением одномодовой операции многомодовым фильтром. Число мод определяется количеством каналов многоканального модана. Таким образом, с одним фильтром можно возбуждать различные моды.

Путем юстировки сформированного модового пучка относительно оси оптоволокна по масштабу, углу и сдвигу можно добиться возбуждения заданной моды с максимальной точностью. На рис. 6.56 показана зависимость выходного сигнала (экспериментальные результаты и теоретическая оценка), пропорциональная интенсивности заданной моды, от величины изгиба волокна h. Выходной сигнал представляет собой мощность одной моды, нормализованную по уровню максимального значения, соответствующего нулевому изгибу. На рис. 6.56 также показаны изменения полного светового потока E' в случае использования гауссова входного пучка. Как видно из экспериментально полученных кривых на рис. 6.56, изменения нормализованных модовых мощностей, измеряемых с помощью моданов, лежат в пределах интервала 0,4–1, в то время как изменения полного светового потока E находятся в пределах существенно меньшего интервала 0,85–1, причем в обоих случаях значения велечины изгиба лежат в пределах одного и то же интервала 0–80 мкм.

Часть экспериментальных результатов, приведенных на рис. 6.56 отличается от теоретических оценок [15]. Это соответствует проявлениям эффектов, связанных с неточной юстировкой установки, технологическими искажениями расчетной функции пропускания моданов. На рис. 6.57 представлена чувствительность датчика $S_{\text{р}} = \partial\,|F_{\text{р}}|^2/\partial h$ в зависимости от величины микроизгиба h, нормализованного к

максимальной чувствительности основной моды $(0,0)$. Анализ рис. 6.56 и 6.57 показывает, что максимальная чувствительность мод в зависимости от изгиба оптоволокна приходится на диапазон уровня выходного сигнала 0,4–0,6. Оценим выигрыш по чувствительности датчика, основанного на использовании селективно возбуждаемых мод, по сравнению с традиционным волоконно-оптическим датчиком, действие которого основано на измерении полного светового потока. Для сравнение было выбрано значение изгиба, равное 175 мкм. В случае использования традиционно используемого датчика, такой изгиб приводит к выходному сигналу в 0,85. Анализ графиков, приведенных на рис. 6.57 и табл. 6.3 показывает, что чувствительность основной моды $(0,0)$ в 3,9 раз выше чувствительности полного светового пучка. Чувствительность к такому изгибу моды $(1,0)$ в 2,1 раза выше, а моды $(2,0)$ — в 2,4 раз выше чувствительности основной моды (табл. 6.3).

Таблица 6.3. Результаты экспериментального исследования датчика

Номера мод (p,l)	(0,0)	(1,0)	(2,0)
S_{pl}, Теоретическая оценка	1,0	1,9	2,3
S_{pl}, Экспериментальный результат	1,0	1,8	2,1

Таким образом, применение волоконно-оптического датчика перемещений, основанного на использовании селективного возбуждения поперечных мод лазерного излучения, позволяет в 4–8 раз увеличить чувствительность по сравнению с традиционным волоконно-оптическим датчиком.

6.3.4. Применение ДОЭ для коллимации излучения полупроводникового лазера

Широкое использование полупроводниковых лазеров в оптической связи, аппаратуре для записи-считывания с компакт-дисков, лазерных целеуказателях делает весьма актуальной задачу совершенствования соответствующей оптики. Специфика оптических свойств полупроводниковых лазеров состоит, как известно, в существенной асимметрии диаграммы направленности излучения, а также в весьма высокой числовой апертуре по одному из сечений пучка. Особую актуальность эта проблема приобретает при построении линий связи с использованием одномодовых волокон. В качестве источников излучения в волоконно-оптических системах связи используются, как правило, лазерные диоды или линейки лазерных диодов видимого или ИК-диапазона. На рис. 6.58 представлен результат измерения распределения интенсивности, формируемого лазерным диодом мощностью 5 мВт и длиной волны 0,67 мкм.

Указанная специфика излучения лазерного диода приводит, как правило, к большим или меньшим (в зависимости от задачи) потерям световой энергии, достигающим в ряде случаев 80%. Таким образом, узкий ($\Delta\lambda < 1$ нм) спектральный диапазон с одной стороны, сложный асимметричный характер амплитудно-фазового распределения и, как следствие, высокие потери в традиционных оптических элементах с другой, делают дифракционную оптику в данном случае вполне конкурентоспособной. Известен дифракционный микрообъектив [81], предназначенный для лазерного проигрывателя, представляющий собой бинарную микролинзу, однако такая линза не устраняет асимметрию пучка, имеет низкую эффективность и весьма ограниченное применение. Более совершенная линза Френеля для коллимации излучения полупроводникового лазера [82] имеет непрерывный профиль и учитывает изменения

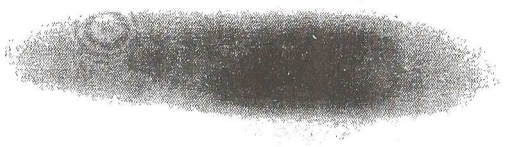


Рис. 6.58. Распределение интенсивности излучения лазерного диода видимого диапазона

температуры ДОЭ. Асимметрия выходного пучка здесь существенно уменьшена путем подбора параметров при расчете в геометрооптическом приближении, однако все же сохраняется. Другой вариант использования ДОЭ [83] — линейки анаморфотных микролинз с эллиптическими зонами для изменения соотношения расходимостей по осям X и Y. Такие микролинзы могут давать на определенном расстоянии пучок круглого сечения, что основано на известном факте, что при малых (дестки микрон) значениях радиусов перетяжек расходимости пучков велики и соответствующим их подбором можно на небольшом (1,5–2 мм) расстоянии от линзы получить круглое пятно. Однако такая плоскость единственная и при дальнейшем распространении пучок вновь станет асимметричным. Указанная особенность предопределяет использование таких элементов преимущественно для ввода пучка в волоконные световоды.

6.3.5. Дифракционные делители пучка

Выше были описаны подходы к параллельной передаче и обработке информации, основанные на использовании продольных и поперечных мод лазерного излучения. Одновременную передачу оптических каналов связи можно вести также путем «простого» пространственного разнесения пучков, соответствующих различным каналам. Такой подход получил широкое распространение в решении задач параллельной обработки сигналов [55–57, 84].

Для этой цели используются дифракционные делители пучка, представляющие собой многопорядковые дифракционные решетки [84]. В монографии [23] описаны различные подходы к расчету делителей пучка, в том числе и с квантованной фазой. Являясь фазовыми элементами, предназначенными для формирования заданного распределения интенсивности, дифракционные делители пучка могут быть рассчитаны с помощью построения соответствующей итеративной градиентной процедуры [84] или процедуры типа Герчберга–Секстона [56, 57]. В пунктах 2.8.4–2.8.5 данной книги представлены результаты расчетов многопорядковых дифракционных делителей пучка с помощью различных итеративных процедур.

Следует отметить, что нередко возникает задача не просто сформировать из освещающего пучка набор дифракционных порядков с заданным распределением энергии между ними, но и контролировать распределение комплексной амплитуды в каждом порядке. Такая задача возникает например в случае, если требуется одновременно возбудить несколько оптических волокон (рис. 6.59).

Если в качестве элемента M использовать простую дифракционную решетку, формирующую на торцах волокон комплекснозначные распределения с неконтролируемым модовым составом, то неизбежны потери энергии, связанные с «утечкой» неканалируемых мод [18]. В [19] предлагалось использовать для решения этой

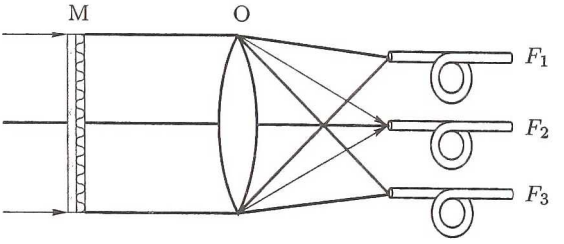


Рис. 6.59. Параллельное возбуждение нескольких волокон одним пучком: F_1, F_2, F_3 — оптоволокна, M-ДОЭ, O — линза

проблемы модан, рассчитанная функция комплексного пропускания которого кодировалась в чисто фазовую с помощью внесения в фазу опорной несущей. Нетрудно видеть, что таким образом можно рассчитать, например, элемент, формирующий из освещающего гауссова пучка лазера фундаментальные $(0,0)$ моды волокон F_1, F_2, F_3, F_4 на их торцах.

6.4. Заключение

В данной главе рассмотрена возможность формирования, селекции мод лазерного излучения с помощью специальных ДОЭ — моданов, а также вопросы, связанные с использованием ДОЭ в системах сбора, хранения и передачи информации. Показано, что разработка и создание дифракционных оптических элементов нового типа — моданов позволили решить фундаментальные задачи, неразрешимые с помощью традиционных оптических элементов — задачи формирования пучков лазерного излучения с заданным поперечно-модовым составом и определения поперечно-модового состава пучка в режиме реального времени. Были описаны возможные приложения ДОЭ в системах сбора, хранения и передачи информации: для повышения пропускной способности волоконно-оптических систем связи, создания высокочувствительных датчиков перемещения, анализа амплитудно-фазовых характеристик лазерного излучения в режиме реального времени.

Необходимо отметить, что до настоящего времени увеличение пропускной способности волоконно-оптических линий связи с помощью селективного возбуждения поперечных мод лазерного излучения не нашло практического воплощения. Это объясняется несколькими причинами:

— реализация качественного профиля градиентных волноводов, позволяющего передавать пучки без существенного изменения модового состава, требует значительно больших затрат, чем хорошо отлаженное производство традиционных кварцевых многомодовых волокон;

— наличие альтернативных подходов к уплотнению волоконно-оптических линий связи — WDM-уплотнения, не налагающего особых требований на качество исполнения волокна, и нелинейных методов уплотнения каналов передачи данных в одномодовых волокнах, позволяющих достичь больших скоростей передачи данных.

— недостаточная отработанность технологии синтеза моданов на этапе развития описываемой концепции.

Авторы полагают, что с развитием волоконной и интегральной оптики, а также с повышением требований к разрабатываемыми оптическим системам связи описанный подход найдет свою «нишу» в перспективных системах телекоммуникаций. Отметим также, что развитие технологий, в частности технологий производства пропускающих ДОЭ для мощных лазеров (4.6), позволяет рассчитывать, что синтез

ДОЭ, согласованных с поперечно-модовым составом лазерного излучения, может быть с успехом использован для измерения поперечно-модового состава излучения и восстановления амплитудно-фазового распределения в его поперечном сечении. Кроме того, разработка методов синтеза моданов дает значительную информацию о возможных подходах к решению задачи синтеза дифракционных оптических элементов, формирующих произвольные амплитудно-фазовые распределения. В этой связи интересно отметить, что вопрос о приоритете точности формирования моды или энергетической эффективности модана решался каждый раз исходя из специфики конкретной задачи, будь то построение волоконно-оптической линии связи или разработка волоконно-оптического датчика давления. Поэтому был разработан определенный «инструментарий» численных методов, позволяющий находить необходимый компромисс в каждом конкретном случае. Этот подход вполне может быть обобщен на расчет ДОЭ, формирующего произвольное амплитудно-фазовое распределение. Обобщая вышесказанное, можно сказать, что дифракционные оптические элементы, благодаря своим уникальным характеристикам, вместе с элементами волноводной и интегральной оптики формируют элементную базу высокоэффективных оптических и оптико-электронных систем сбора, обработки и передачи информации.

Литература к главе 6

1. Снайдер А., Лав Дж. *Теория оптических волноводов* (М.: Радио и связь, 1987)
2. Ананьев Ю. А. *Оптические резонаторы и лазерные пучки* (М. : Наука, 1990)
3. Звелто О. *Принципы лазеров* (М.: Мир, 1990)
4. Rudin W. *Functional Analysis* (McGraw–Hill, New York, 1973)
5. Унгер Х. Г., *Планарные и волоконные оптические волноводы* (М.: Мир, 1980)
6. Felsen L. B., Marawitz M. *Radiation and scattering of waves*(Prentice–Hall, Englewood Cliffs, New Jersey, 1973)
7. Солимено С., Крозиньяни Б., Порто П. Ди *Дифракция и волноводное распространение оптического излучения* (М.: Мир, 1989)
8. Landesman B. T., Barrett H. H. *J. Opt. Soc. Am. A* **5** 1610 (1988)
9. Nemoto S. *Appl. Opt.* **29** 1940 (1990)
10. Борн М., Вольф Э. *Основы оптики* (М. : Наука, 1973)
11. Haus H. A. *Waves and Fields in Optoelectronics* (Prentice Hall, Inc. Englewood Cliffs, 1984)
12. Marcuse D. *Light transmission optics* (Van Nostrand Reinhold, New York, 1982)
13. Golub M. A., Sisakyan I. N., Soifer V. A. *Modern Optics* **38** 1067 (1991)
14. Корн Г., Корн Т. *Справочник по математике* (М.: Наука, 1974)
15. Ярив А. *Квантовая электроника* (М.: Сов. Радио, 1980)
16. Ландау Л. Д., Лифшиц Е. М. *Квантовая механика* (М.: Наука, 1974)
17. Pressley R. I. *Handbook of lasers with selected data on optical technology* (Chemical Rubber Co., Cleveland, 1971)
18. Адамс М. *Введение в теорию оптических волноводов* (М.: Мир, 1984)
19. Soifer V. A., Golub M. A. *Laser Beam Mode Selection by Computer Generated Holograms* (CRC Press 1994)
20. Бахарев М. А., Котляр В. В., Павельев В. С., Сойфер В. А., Хонина С. Н. *Компьютерная оптика* **17** 21 (1997)
21. Goodman J. W. *Introduction to Fourier Optics* (McGraw–Hill, New-York, 1968)
22. Голуб М. А., Сисакян И. Н., Сойфер В. А. *Компьютерная оптика* **8** 3 (1990)
23. Soifer V. A., Kotlyar V. V., Doskolovich L. L. *Iterative Methods for Diffractive Optical Elements Computation*, "Taylor & Francis Ltd. 1997
24. Kotlyar V. V., Nikolsky I. V., Soifer V. A. *Optik* **98** 1 26 (1994)
25. Хонина С. Н., Котляр В. В., Сойфер В. А. *Компьютерная оптика* **17** 25 (1997)
26. Котляр В. В., Сойфер В. А., Хонина С. Н. *Компьютерная оптика* **17** 31 (1997)

27. Duparre M., Pavelyev V. S., Luedge B., Kley E. B., Kowarschik R., Soifer V. A. *Proc. SPIE* **3110** 741 (1997)
28. Duparre M., Pavelyev V. S., Luedge B., Kley E. B., Kowarschik R., Soifer V. A. *Proc. SPIE* **3134** 357 (1997)
29. Duparre M., Pavelyev V., Luedge B., Kley B., Soifer V., Kowarschik R. *Proc. SPIE* **3291** 104 (1998)
30. Pavelyev V. S., Soifer V. A., Duparre M., Kowarschik R., Ludge B., Kley B. *Optics and Lasers in Engineering* **29** 269 (1998)
31. Doskolovich L. L., Perlo P., Petrova O. I., Repetto P., Soifer V. A. *Proc. SPIE* **3348** 37
32. Soifer V. A., Golub M. A., Pavelyev V. S. *OSA: Diffractive Optics: Design, Fabrication and Applications Technical Digest Series* **11** 209 (1994)
33. Карпеев С. В. *Пространственные фильтры для анализа и формирования поперечно-модового состава когерентного излучения в волоконных световодах* (Диссертация на соискание ученой степени кандидата физико-математических наук, Куйбышев,1990)
34. Павельев В. С., Сойфер В. А., Дюпарре М., Коваршик Р., Людге Б., Клей Б., Карпеев С. В. *Компьютерная оптика* **18** 115 (1998)
35. Berezny A. E., Karpeev S. V., Uspleniev G. V. *Optics and Lasers in Engineering* **15** 331 (1991)
36. Khonina S. N., Kotlyar V. V. *Proc. SPIE* **2363** 184 (1994)
37. Thornburg W. Q., Corrado B. J., Zhu X. D. *Optics letters* **19** 7 454
38. Kirk J. P., Jones A. L. *J. Opt. Soc. Am.* **61**(8) 1023 (1971)
39. Chu D. C., Fienup J. R. *Optical Engineering* **13**(3) 189 (1974)
40. Старк Г *Реконструкция изображений* (М.:Мир, 1992)
41. Levi A., Stark H. *J. Opt. Soc. Am.* **1**(2) 932 (1984)
42. Павельев В. С., Хонина С. Н. *Компьютерная оптика* **17** 15 (1997)
43. Корн Г., Корн Т. *Справочник по математике для научных работников и инженеров* (М.: Наука, 1984)
44. Khonina S. N., Kotlyar V. V., Soifer V. A. *Optik* **88**(4) 182 (1991)
45. Soifer V. A., Doskolovich L. L., Kazanskiy N. L., Pavelyev V. S. *Proc. SPIE* **2426** 358
46. Павельев В. С. *Расчет дифракционных оптических элементов методом обобщенных проекций* (Диссертация на соискание ученой степени кандидата физико-математических наук, Самара, 1996)
47. Bagini V., Gori F., Santasiero M., Frezza F., Schettini G., Richetta M., Spagnolo G. *Laser Beam Characterization (SEDO, Madrid, 1993)*, 31
48. Fienup J. R. *Applied optics* **21**(15) 2758 (1982)
49. Kotlyar V. V., Nikolski I. V., Soifer V. A. *Optik* **88**(1) 17 (1991)
50. Lu C. Y., Liao H. Z., Lee C. K., Wang J. S. *Applied Optics* **36**(20) (1997)
51. Голуб М. А., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Квантовая электроника* **9**(9) 1866 (1982)
52. Bartelt H. O., Lohmann A. W., Freude W., Grau G. K. *Electronic letters* **19**(7) 247 (1983)
53. Berdague S., Facq P. *Applied Optics* **21**(11) 1950 (1982)
54. Dubois F., Emplit Ph., Hugon O. *Optics Letters* **19** 7 433 (1994)
55. Lalanne Ph., Chavel P. *Perspectives for parallel optical interconnects* (DG 3 Comiss Of the Europ Communities; - Berlin etc.: Springer, Cop. 1993)
56. Prokopenko S. A., Palagushkin A. N., Mikaelian A. L., Kiseliov B. S., Shkitin V. A. *Proc. SPIE* **3348** 22
57. Prokopenko S. A., Palagushkin A. N., Mikaelian A. L., Kiseliov B. S., Shkitin V. A. *Proc. SPIE* **3348** 83
58. Акаев А. А., Майоров С. А. *Когерентные оптические вычислительные машины* (Л.: Машиностроение, 1977)
59. Акаев А. А., Майоров С. А. *Оптические методы обработки информации* (М.:Высшая школа, 1988)
60. Уваров Г. В. *Волоконно-оптические преобразователи на основе разделения поперечных мод* (Диссертация на соискание ученой степени кандидата технических наук, Куйбышев, 1988)
61. Golub M. A., Sisakyan I. N., Soifer V. A., Uvarov G. V. *Proc. SPIE* **1572** 101 (1991)
62. Bazhanov Yu., Shaposhnikov S. N., Adrianov S. N. *Proc. SPIE* **3348** 130 (1998)

63. Chang J. T., Su D. C., Huang Y. T. *Appl. Opt.* **33**(35) 8143 (1994)
64. Fainman Y., Xu F., Tyan R., Sun P. C., Ford J. E., Krishnamoorthy A., Scherer A. *Proc. SPIE* **3348** 152
65. Бобринев В. И., Сон Дж. Я., Чон Х. В. *Квантовая электроника* **22** (1995)
66. Досколович Л. Л. *Компьютерная оптика* **18** 16 (1998)
67. Сойфер В. А. *Введение в дифракционную микрооптику* (Самара: СГАУ, 1996)
68. Коеррен С., Shi R. F., Chen W. D., Garito F. *J. Opt. Soc. Am. B* **15**(2) 727 (1998)
69. Хасэгава А., Кодама Ю. *ТИИЭР* **69**(9) 57 (1981)
70. Ахманов С. А., Выслоух В. А., Чиркин А. С. *Успехи физических наук* **149**(3) 450
71. Шерман А. Ю. *Компьютерная оптика* **6** 32
72. Gagliardi R. M. *Optical Communications* (Wiley Series in Telecommunications and Signal Processing; John Wiley & Sons, Inc. ; USA, 1995)
73. Cutolo, T. Pierri, L. Zeni *Laser Beam Characterization* (SEDO, Madrid, 1993), 263
74. Du K.-M., Herziger G., Loosen P., Ruhl F. *Optical and Quantum Electronics* **24** 1119 (1992)
75. Fields J. N. *Appl. Opt. Phys. Lett.* **36** 799 (1980)
76. Fields J. N., Asawa C. K., Ramer O. G., Barnoski M. K. *J. Acoust. Soc. Am.* **67** 816 (1980)
77. Grau G. K., Leminger O. G., Sauter E. G. *Archiv fur Electronik und Ubertragungstechnik, Electronics and Communication* **34** 259 (1980)
78. Krivoshlykov S. G., Petrov N. I., Sisakyan I. N. *Opt. and Quant. Electr.* **15** 193 (1983)
79. Krivoshlykov S. G., Sauter E. *Journal of Physics A* **20** 3805 (1987)
80. Гаричев В. П., Голуб М. А., Карпеев С. В., Кривошлыков С. Г., Сисакян И. Н., Сойфер В. А., Уваров Г. В. *Компьютерная оптика* **3** 103 (1988)
81. Ogata S., Ito Y. *Optical Engineering* **33**(11) 3656 (1994)
82. Chen W., Roychoudhuri C. S., Banas C. M. *Optical Engineering* **33**(11) 3662 (1994)
83. Rossi M., Bona G. L., Kunz R. E. *Applied optics* **34**(14) 2483 (1995)
84. Doscolovich L. L., Soifer V. A., Alexandretti G., Perlo P., Repetto P. *Pure and Appl. Optics (JEOS)* **3** 921 (1994)
85. Wiitrock U., Kumkar M., Weber H. *Laser Beam Characterization*, (SEDO, Madrid, 1993) 41
86. Виноградова М. Б., Руденко Д. В., Сухоруков А. П. *Теория волн* (М.: Наука, 1979)
87. Прудников А. П., Брычков Я. А., Маричев О. И. *Интегралы и Ряды Специальные функции* (М.: Наука, 1979)

ГЛАВА 7

СВЕТОВЫЕ ПУЧКИ С ПЕРИОДИЧЕСКИМИ СВОЙСТВАМИ

7.1. Некоторые типы световых пучков с периодическими свойствами

Периодическое самовоспроизведение изображений известно в оптике с 1836 года после работы Х. Ф. Тальбота, в которой он рассматривал световые пучки, формируемые планарными масками, имеющими поперечно-периодическую функцию пропускания (линейные и скрещенные решетки) [1]. С тех пор явление самовоспроизведения когерентного светового пучка, имеющего поперечную периодичность, на определенных расстояниях при распространении (продольная периодичность) называется *эффектом Тальбота*. Для некогерентного пучка это явление называется *эффектом Лау* [2].

Затем в 1881 году Рэлей показал, что для наблюдения хорошего самовоспроизведения решеток, они должны освещаться монохроматическим коллимированным пучком и вывел формулу для расстояния (получившего впоследствии название период Тальбота), на котором решетка повторится первый раз [3]:

$$d = 2a^2/\lambda, \tag{7.1}$$

где a — период решетки, λ — длина волны освещающего пучка. Последующие работы касались в основном вариаций в свойствах решеток (фазовые, прямоугольные).

Таким образом, исторически первыми были отмечены периодические свойства световых полей, состоящих из конечного числа плоских волн.

В [4] было показано, что линейные решетки могут формировать периодически повторяющиеся изображения, отличающиеся от самого объекта. Затем в [5–7] были получены выражения для распределения светового поля на долях периода Тальбота (7.1). Так, прямоугольная решетка с периодом $a \times a$ и с функцией пропускания

$$f(x, y, 0) = f(x + pa, y + qa, 0) = \sum_{n,m} F_{n,m} \exp\left[i2\pi \frac{nx + my}{a}\right], \tag{7.2}$$

где $F_{n,m}$ — пространственный спектр $f(x, y)$, в приближении Френеля формирует на расстоянии z световое поле с амплитудой

$$f(x, y, z) = \exp(ikz) \sum_{n,m} F_{n,m} \exp\left[i2\pi \frac{nx + my}{a}\right] \exp\left[-i2\pi \frac{(n^2 + m^2)\, z}{d}\right], \tag{7.3}$$

где d — период Тальбота (7.1). На расстояниях кратных периоду $z = ld$ (так как $\exp[-i2\pi(n^2 + m^2)l] = 1$) происходит повторение распределения интенсивности светового поля:

$$|f(x,y,z)|_{z=ld}|^2 = \left|\exp(ikld)\sum_{n,m} F_{n,m}\exp\left[i2\pi\frac{nx+my}{a}\right]\right|^2 = |f(x,y,0)|^2, \qquad (7.4)$$

то есть решетка самовоспроизводится. На расстоянии, равном половине периода d (при $l = 1/2$), получится копия решетки, сдвинутая на половину поперечного периода a:

$$f(x,y,z)|_{z=d/2} = \exp(ikd/2)\sum_{n,m} F_{n,m}\exp\left[i2\pi\frac{nx+my}{a}\right]\times$$
$$\times\exp\left[-i\pi\left(n^2+m^2\right)\right] = \exp(ikd/2)\, f(x+a/2, y+a/2, 0), \qquad (7.5)$$

а на расстоянии $4q$ долей периода d получим

$$f(x,y,z)|_{z=d/4q} = \exp(ikd/4q)\,\frac{\exp(-i\pi/2)}{2q}\sum_{s,r}^{2q-1}\exp\left[i\pi\frac{\left(s^2+r^2\right)}{2q}\right]\times$$
$$\times f(x+sa/2q, y+ra/2q, 0). \qquad (7.6)$$

То есть, на расстояниях, составляющих доли периода Тальбота происходит сложение сдвинутых копий первоначального изображения.

В 1967 году У. Д. Монтгомери показал, что объект с бесконечной апертурой будет самовоспроизводиться при распространении, если его спектр лежит на кольцах зонной пластинки [8]. Исходя из интегрального представления решения уравнения Гельмгольца,

$$f(\mathbf{x},z) = \int_{R^2} F(\boldsymbol{\xi})\exp\left[iK(\boldsymbol{\xi})z\right]\exp(i2\pi\mathbf{x}\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}, \qquad (7.7)$$

где $F(\boldsymbol{\xi})$ — Фурье-образ функции $f(\mathbf{x})$, $K(\boldsymbol{\xi}) = k\sqrt{1-(\lambda\boldsymbol{\xi})^2}$, $\boldsymbol{\xi} = (\xi,\eta)$ — вектор пространственных частот, k — волновое число, и, ставя условие самоповторения на расстоянии d с точностью до комплексной константы,

$$f(\mathbf{x},d) = wf(\mathbf{x},0) \quad \forall\mathbf{x}, \quad w = \mathrm{const}, \quad \mathbf{x} = (x,y), \qquad (7.8)$$

Монтгомери получил следующее выражение для пространственных частот:

$$\frac{d}{\lambda}\left[1-\sqrt{1-(\lambda\boldsymbol{\xi})^2}\right] = m(\boldsymbol{\xi}), \qquad (7.9)$$

где $m(\boldsymbol{\xi})$ — целое число, зависящее от $\boldsymbol{\xi}$.

Разделив спектр $F(\boldsymbol{\xi})$ на два типа, можно выделить два типа самовоспроизведения.

а) *«Слабое» самовоспроизведение*: пространственные частоты $\boldsymbol{\xi}$ объекта $f(\mathbf{x})$, для которых $F(\boldsymbol{\xi}) \neq 0$, лежат внутри круга радиуса $R \ll 1/\lambda$. Тогда самоповторение на расстоянии d происходит, если

$$|\boldsymbol{\xi}| = \sqrt{2/\lambda d}\sqrt{m(\boldsymbol{\xi})}, \qquad (7.10)$$

т. е., пространственные частоты лежат на кольцах зонной пластинки с радиусами $R_m = \sqrt{2\lambda dm}$.

Для поперечно-периодических объектов $f(\mathbf{x}+\mathbf{a_n}) = f(\mathbf{x})$ пространственные частоты будут дискретными:

$$|\mathbf{b_n}| = \sqrt{2/\lambda d}\sqrt{m(\mathbf{n})}, \tag{7.11}$$

откуда, в частности, для периодического по одной координате объекта $f(x+na, y) = f(x,y)$, $m(n) = n^2$, получается формула Рэлея (7.1).

б) *«Сильное» самовоспроизведение*: пространственные частоты $\boldsymbol{\xi}$ объекта $f(\mathbf{x})$, для которых $F(\boldsymbol{\xi}) \neq 0$, лежат внутри круга радиуса вплоть до $1/\lambda$. Тогда самоповторение на расстоянии d происходит, если

$$|\boldsymbol{\xi}|^2 + (m(\boldsymbol{\xi})/d - 1/\lambda)^2 = (1/\lambda)^2, \tag{7.12}$$

что представляет собой уравнение окружности радиуса $1/\lambda$ с центром в точке $(0, 1/\lambda)$. Для $d \gg \lambda$ получается набор окружностей в частотной плоскости:

$$|\boldsymbol{\xi}|^2 = 2m(\boldsymbol{\xi})/\lambda d. \tag{7.13}$$

Учитывая, что $\mathbf{x} = (x,y)$, $\boldsymbol{\xi} = (\xi, \eta)$, $|\boldsymbol{\xi}| = \sqrt{\xi^2+\eta^2}$, перепишем (7.7) в скалярной форме:

$$f(x,y,z) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} F(\xi,\eta)\exp\left[ikz\sqrt{1-\lambda^2(\xi^2+\eta^2)}\right] \times$$

$$\times \exp\left[i2\pi(x\xi + y\eta)\right] \mathrm{d}\xi\, \mathrm{d}\eta. \tag{7.14}$$

На расстоянии d произойдет самовоспроизведение, если выполнено (7.10) или (7.13):

$$\xi^2 + \eta^2 = \frac{2m}{\lambda d}. \tag{7.15}$$

Подставляя (7.15) в (7.14), получаем

$$f(x,y,z) = \exp\left[ikz\sqrt{1-\frac{2m\lambda}{d}}\right] \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} F(\xi_m,\eta_m) \times$$

$$\times \exp\left[i2\pi(x\xi_m + y\eta_m)\right] \mathrm{d}\xi_m\, \mathrm{d}\eta_m = \exp(i\beta_m z)\, f(x,y), \tag{7.16}$$

где $\beta_m = k\sqrt{1-2m\lambda/d}$, а $f(x,y)$ — поперечно-периодическая функция.

Пусть в полярных координатах $\mathbf{x} = (r,\varphi)$, $\boldsymbol{\xi} = (\rho,\theta)$, спектр имеет частоты только на одном кольце $F(\rho,\theta) = A_m(\theta)\delta(\rho - \sqrt{2m\,/\,\lambda d})$, тогда вместо (7.16) получим формулу для комплексной амплитуды пучков с продольной периодичностью:

$$f(r,\varphi,z) = \exp(i\beta_m z) \int\limits_0^{2\pi} A_m(\theta)\exp\left[i2\pi r\sqrt{2m/\lambda d}\cos(\varphi-\theta)\right] \mathrm{d}\theta =$$

$$= \exp(i\beta_m z) \int\limits_0^{2\pi} A_m(\theta)\exp\left[ir\alpha_m\cos(\varphi-\theta)\right] \mathrm{d}\theta, \tag{7.17}$$

где $A_m(\theta)$ — некоторая комплексная функция, $\alpha_m = 2\pi\sqrt{2m/\lambda d}$. Причем, легко видеть, что $\alpha_m^2 + \beta_m^2 = k^2$. При $A_m(\theta) = \exp(im\theta)$ из (7.17) следует, что $f(r,\varphi) = 2\pi(i)^m J_m(\alpha_m r)\exp(im\varphi)$, которая уже не является поперечно-периодической функцией ($J_m(x)$ — функция Бесселя m-го порядка).

Таким образом, был определен еще один тип самовоспроизводящихся функций – моды Бесселя. Методы формирования этих мод рассматриваются в разделах 7.2–7.4.

В [9] Монтгомери обобщил результаты, полученные в [8]. Рассматривая собственные функции оператора распространения

$$\mathrm{D}_z f = wf, \tag{7.18}$$

где

$$\mathrm{D}_z f = \int f(\mathbf{u})\,\mathrm{D}_z(\mathbf{x}-\mathbf{u})\,\mathrm{d}\mathbf{u}, \tag{7.19}$$

$$\mathrm{D}_z(\mathbf{x}) = \int \exp\left[iK(\boldsymbol{\xi})z\right]\exp\left(i2\pi\mathbf{x}\boldsymbol{\xi}\right)\,\mathrm{d}\boldsymbol{\xi}, \tag{7.20}$$

он показал, что в случае *бесконечной апертуры* дискретный спектр этих функций пуст. Могут существовать поля только с непрерывным спектром, и для самовоспроизведения таких полей энергия спектра должна концентрироваться на кольце или наборе колец:

$$\exp\left[iK(\boldsymbol{\xi})\right] = w. \tag{7.21}$$

В случае же *конечной апертуры* (что соответствует введению черного экрана с отверстием Σ) дискретный спектр не пуст, и существуют собственные функции оператора:

$$\mathrm{B}_z Af = wf, \quad |w| < 1, \tag{7.22}$$

где

$$\mathrm{B}_z Af = \int\limits_{\Sigma} f(\mathbf{u})\,\mathrm{B}_z(\mathbf{x}-\mathbf{u})\,\mathrm{d}\mathbf{u}, \tag{7.23}$$

$$\mathrm{B}_z(\mathbf{x}) = \int\limits_{|\xi|\leqslant 1/\lambda} \exp\left[iK(\boldsymbol{\xi})z\right]\exp\left(i2\pi\mathbf{x}\boldsymbol{\xi}\right)\,\mathrm{d}\boldsymbol{\xi}. \tag{7.24}$$

Искать самовоспроизводящуюся функцию f предлагается через собственные функции ϕ_n:

$$f = \sum_n C_n\phi_n, \tag{7.25}$$

$$B_0 A\phi_n = \lambda_n\phi_n, \quad (\phi_n,\phi_k) = \delta_{n,k}, \tag{7.26}$$

решая матричное уравнение

$$M_z\Lambda\mathbf{C} = w\mathbf{C}, \tag{7.27}$$

где M_z — матрица элементов $m_{n,k} = (B_z\phi_n,\phi_k)$, Λ — диагональная матрица собственных значений ϕ_n, $\mathbf{C}$ — вектор коэффициентов C_n.

Известно [10], что в случае квадратной апертуры свойствами ϕ_n обладают вытянутые сфероидальные функции.

В [11] была показана взаимосвязь поперечной и продольной периодичности световых полей. Так, световое поле, созданное эталоном Фабри–Перо с расстоянием между зеркалами Δ, самовоспроизводится с периодом $d = 2\Delta$ и имеет спектр, состоящий из колец. При Δ равном целому числу полуволн:

$$\Delta = N\lambda/2, \quad N\text{— целое}, \tag{7.28}$$

получается спектр с кольцами, имеющими радиусы в пространственно-частотной плоскости:

$$\rho_m = \sqrt{m/\lambda\Delta} = \sqrt{2m/\lambda d}, \quad m = 0, 1, 2, \ldots \tag{7.29}$$

Если выбрать только первые два кольца, получится спектр $\delta(\rho) + \delta(\rho - \rho_1)$, что соответствует при $z = 0$ амплитуде $1 + J_0(2\pi\rho_1 r)$, где $J_0(x)$ — функция Бесселя нулевого порядка.

Полученные Монтгомери результаты были обобщены для случая частично-когерентных полей в [8]. Исходя из того, что взаимная корреляционная функция частично-когерентного светового поля удовлетворяет паре уравнений Гельмгольца:

$$\begin{aligned} \left(\nabla_1^2 + k^2\right) W(x_1, y_1, x_2, y_2, z) = 0, \\ \left(\nabla_2^2 + k^2\right) W(x_1, y_1, x_2, y_2, z) = 0, \end{aligned} \tag{7.30}$$

и накладывая условие инвариантности распространения указанного поля

$$W(x_1, y_1, x_2, y_2, z) = W(x_1, y_1, x_2, y_2, 0), \quad z \geqslant 0, \tag{7.31}$$

было получено, что спектр взаимной корреляционной функции при $z = 0$ (в полярных координатах)

$$\mathcal{F}\left[W(x_1, y_1, x_2, y_2, 0)\right] = A(r_1, \varphi_1, r_2, \varphi_2) = S(r_1, \varphi_1, \varphi_2)\delta(r_1 - r_2), \tag{7.32}$$

не коррелирован по радиальной составляющей.

Вращающиеся при распространении световые поля можно считать инвариантными с точностью до поворота или инвариантными во вращающейся системе координат. В [12] *вращающиеся волны* определяются как движущиеся волны, распространяющиеся кругами, но имеющие модовую структуру (или дискретный спектр) стоячих волн. В [12, 13] выделены два типа вращающихся волн: *цилиндрические*

$$f(r, \varphi, z, t) = H_m^{(\pm)}(\alpha r) \exp\left(i\beta z + im\varphi - i\omega t\right), \tag{7.33}$$

где $H_m^{(\pm)}(x) = J_m(x) \pm iN_m(x)$, $H_m^{(\pm)}(x)$ — функция Ханкеля, $J_m(x)$ — функция Бесселя, $N_m(x)$ — функция Неймана; и *сферические*

$$f(R, \theta, \varphi, t) = P_l^m(\cos\theta) h_l^{(\pm)}(\alpha R) \exp\left(im\varphi - i\omega t\right), \tag{7.34}$$

где $h_l^{(\pm)}(x) = j_l(x) \pm in_l(x)$, $h_l^{(\pm)}(x)$ — сферическая функция Ханкеля, $j_l(x)$ — сферическая функция Бесселя, $n_l(x)$ — сферическая функция Неймана, $P_l^m(x)$ — функция Лежандра.

Хотя описанные функции несут в себе вращательный момент $\exp(im\varphi)$, их вращение не наблюдаемо, так как интенсивность таких функций радиально-симметричная. Чтобы сформировать вращающийся световой пучок, наблюдаемый визуально,

необходима композиция хотя бы из двух таких функций. В [13, 14] рассматривается композиция двух функций $J_n(\alpha r)\exp(in\varphi)$ и $J_m(\beta r)\exp(im\varphi)$, которая при $n \neq m$ вращается как единое целое, делая полный оборот на расстоянии

$$\Delta z_{2\pi} \approx 4\pi k(n-m)/(\alpha^2-\beta^2), \tag{7.35}$$

и имеет период повторения картины поперечной интенсивности, равный

$$\Delta z \approx 4\pi k/(\alpha^2-\beta^2), \tag{7.36}$$

что связано с симметрией интенсивности порядка $(n-m)$.

Общее условие периодичности бесселевых пучков и их вращения как единого целого для композиции из более чем двух составляющих рассматривается в разделах 7.2 и 7.3.

В [15, 16] рассматривается еще один тип функций, несущих в себе вращательный момент, и представляющих собой гауссовы пучки:

$$f(x,y) = \exp\left[-\frac{x^2+y^2}{\sigma^2}\right] \prod_{m=0}^{M} \prod_{p_m=1}^{P_m} \left[(x-x_{p_m}) + i\,(y-y_{p_m})\right]^m \times$$

$$\times \prod_{q_m=1}^{Q_m} \left[(x-x_{q_m}) - i\,(y-y_{q_m})\right]^m, \tag{7.37}$$

где P_m — число положительных вращательных моментов:

$$\begin{aligned} &(x-x_{p_m}) + i\,(y-y_{p_m}) = r\exp(i\varphi) - r_{p_m}\exp(i\varphi_{p_m}),\\ &r = \sqrt{x^2+y^2}, \quad \varphi = \mathrm{tg}^{-1}(y/x), \quad r_{p_m} = \sqrt{x_{p_m}^2+y_{p_m}^2},\\ &\varphi_{p_m} = \mathrm{tg}^{-1}(y_{p_m}/x_{p_m}), \end{aligned} \tag{7.38}$$

а Q_m — число отрицательных вращательных моментов:

$$(x-x_{q_m}) - i\,(y-y_{q_m}) = r\exp(-i\varphi) - r_{q_m}\exp(-i\varphi_{q_m}), \tag{7.39}$$

(x_{p_m}, y_{p_m}) и (x_{q_m}, y_{q_m}) — центры соответствующих воронок.

В [17] получено условие сохранения структуры многомодового пучка Гаусса–Лагерра с точностью до масштаба и поворота при распространении в свободном пространстве. Композиция

$$f(x,y) = \sum_{n,m} C_{n,m}\Psi_{n,m}(x,y), \tag{7.40}$$

где

$$\Psi_{n,m}(x,y) = \exp\left(-x^2-y^2\right)\left(x + i\,\mathrm{sgn}(m)y\right)^{|m|} L_n^{|m|}\left(2x^2+2y^2\right), \tag{7.41}$$

будет вращаться, если

$$2n + |m| + \mu m = \gamma - 1, \tag{7.42}$$

$L_n^m(x)$ — полином Лагерра, μ, γ — константы.

Интересно, что все упомянутые выше световые моды свободного пространства можно получить как решение изопериметрической задачи [18], то есть из условия минимума некоторого функционала, который пропорционален параметру расходимости светового пучка. В этом случае моды оказываются такими состояниями, которые при прочих равных условиях (заданной энергии и радиуса пучка) обеспечивают минимальную расходимость и максимальную яркость в фокальной плоскости линзы.

Вариационная задача поиска пучков с минимальной расходимостью при заданных значениях энергии и радиуса пучка или диафрагмы приводит к явным выражениям для известных четырех типов световых мод свободного пространства: Гаусса—Эрмита, Гаусса—Лагерра, Бесселя и косинусоидальных [19]. Эти моды описываются амплитудно-фазовыми функциями. Методы получения чисто фазовой функции пропускания ДОЭ, согласованной с модами свободного пространства, будут рассмотрены в следующих разделах.

7.2. Фазовые формирователи световых полей с продольной периодичностью

Из упомянутых в 7.1 световых мод пучки Бесселя вызывают особый интерес благодаря свойству распространяться в свободном пространстве практически без дифракции. В работах [20, 21] изучался световой пучок, описываемый функцией Бесселя первого рода нулевого порядка, а в [22] — пучок, амплитуда которого пропорциональна произведению функции Бесселя на функцию Гаусса. В [23, 24] рассмотрены бездифракционные пучки высших порядков, описываемые функциями Бесселя произвольного порядка, т. н. бесселевыми модами [25]. Они распространяются, например, внутри сердцевины круглого оптического волокна со ступенчатым показателем преломления, а также появляются на выходе резонатора с круглыми плоскими зеркалами одинакового радиуса.

В [26] получено общее выражение для комплексной амплитуды бездифракционных пучков, которые при распространении не изменяются. Разработан итеративный алгоритм [27], с помощью которого можно рассчитывать фазовые ДОЭ, формирующие световые пучки с произвольной композицией бесселевых мод. Такие ДОЭ работают по типу винтового аксикона [14], обеспечивая инвариантные свойства сформированного пучка на расстоянии, пропорциональном радиусу ДОЭ и обратно пропорциональном углу наклона плоских волн пространственного спектра поля. Дифракционное расширение диаметра пучка компенсируется за счет притока энергии из периферийных областей ДОЭ. То есть, с увеличением расстояния z от ДОЭ до рассматриваемой плоскости растет и радиус зоны (кольца) ДОЭ, которая отвечает за формирование светового поля на этой плоскости.

Пучки с продольной периодичностью возникают в случае, когда несколько конических волн с различными углами при вершине конуса распространяются вдоль оптической оси. В результате их интерференции вдоль оптической оси образуется интерференционная картина, амплитуда модуляции которой примерно постоянная на отрезке от ДОЭ до расстояния инвариантности, о котором говорилось выше. При этом волна, распространяясь вдоль оптической оси, в своем поперечном сечении испытывает периодические изменения. Минимальный период интерференционной картины будет определяться максимальной разностью углов наклона различных конических волн.

Вопросы, связанные с синтезом фазовых ДОЭ генерирующих волны с заданным продольным, в частности, продольно-периодическим, распределением интенсивности, рассматривались в [28–30]. Предложенные в этих работах алгоритмы позволяют рассчитывать фазовые ДОЭ типа линзы со сферической аберрацией, которая увеличивает размер фокального пятна. Такие ДОЭ формируют заданное продольное распределение интенсивности на малом участке оптической оси.

В данном разделе рассматривается итеративный алгоритм расчета ДОЭ типа обобщенных аксиконов, формирующих волны с продольной периодичностью, обладающие модовым характером. Пусть комплексная амплитуда монохроматического светового поля $U(x, y, z)$, удовлетворяет дифференциальному уравнению Гельмгольца (см. (6.13) при $n = 1$):

$$(\nabla^2 + k^2)U(x, y, z) = 0. \tag{7.43}$$

Если комплексная амплитуда приводима в цилиндрических координатах

$$U(x, y, z) = U_1(r)U_2(\varphi)U_3(z),$$

то частное решение уравнения (7.43) можно записать в виде [31]

$$U_{qm}(r, \varphi, z) = e^{qz} J_m\left(r\sqrt{k^2 + q^2}\right) \exp(im\varphi), \tag{7.44}$$

где (r, φ) — поперечные полярные координаты светового поля, q — постоянная разделения, $J_m(x)$ — функция Бесселя m-го порядка. Постоянную разделения q можно найти из требования периодичности амплитуды волны вдоль оси z:

$$U(r, \varphi, z + z_0) = U(r, \varphi, z), \tag{7.45}$$

где z_0 — период. Из условия (7.45) следует возможность представления амплитуды светового поля функциональным рядом Фурье:

$$U(r, \varphi, z) = \sum_{n=-\infty}^{\infty} U_n(r, \varphi) \exp\left(\frac{i2\pi nz}{z_0}\right), \tag{7.46}$$

где

$$U_n(r, \varphi) = \frac{1}{z_0} \int_0^{z_0} U(r, \varphi, z) \exp\left(-\frac{i2\pi nz}{z_0}\right) \mathrm{d}z. \tag{7.47}$$

Сравнивая выражения (7.44) и (7.46), получим явный вид функции $U_n(r, \varphi)$:

$$U_n(r, \varphi) = \sum_{m=-\infty}^{\infty} C_m J_m\left(r\sqrt{k^2 + q_n^2}\right) e^{im\varphi}, \tag{7.48}$$

где $q_n = i2\pi n/z_0$.

Таким образом, общий вид функции $U(r, \varphi, z)$, обладающей продольной периодичностью в соответствии с условием (7.45), должен быть следующим:

$$U(r, \varphi, z) = \sum_{n,m=-\infty}^{\infty} C_m J_m(kr\bar{\rho}_n) \exp\left\{\frac{i2\pi nz}{z_0} + im\varphi\right\}, \tag{7.49}$$

где $\bar{\rho}_n = \sqrt{1 - (n\lambda/z_0)^2}$. Требование положительности подкоренного выражения для $\bar{\rho}_n$ ограничивает число слагаемых в сумме (7.49) по индексу n:

$$|n| \leqslant z_0/\lambda. \tag{7.50}$$

Слагаемые в уравнении (7.49) при $|n| > z_0/\lambda$ будут пропорциональны модифицированным функциям Бесселя $I_m(kr\bar{\rho}_n)$, которые расходятся при стремлении r к бесконечности. Эти нефизические слагаемые исключаются из рассмотрения.

Следуя работе [26], из условия продольной периодичности $U(r, \varphi, z)$ по модулю

$$|U(r, \varphi, z + z_0)| = |U(r, \varphi, z)|, \tag{7.51}$$

можно получить отличное от (7.49) выражение:

$$U(r, \varphi, z) = \sum_{n=n_1}^{n_2} U_n(r, \varphi) \exp\left\{i2\pi z \left(\frac{\sigma_0}{\lambda} + \frac{n}{z_0}\right)\right\}, \tag{7.52}$$

где

$$U_n(r, \varphi) = \frac{1}{2\pi} \int_0^{2\pi} F_n(\theta) \exp\left[-ikr\rho_n \cos(\varphi - \theta)\right] \mathrm{d}\theta, \tag{7.53}$$

$$\rho_n = \sqrt{1 - (\sigma_0 + n\lambda/z_0)^2}, \tag{7.54}$$

$\sigma_0 = \cos\alpha_0$, α_0 — минимальный угол наклона плоских волн пространственного спектра для данного поля, n_1 и n_2 находятся из условия положительности подкоренного выражения в уравнении (7.54):

$$n_1 = -\left[\sigma_0 z_0/\lambda\right], \quad n_2 = \left[(1 - \sigma_0) z_0/\lambda\right], \tag{7.55}$$

где $[\ldots]$ — знак целой части числа.

В уравнении (7.53) произвольная функция $F_n(\theta)$ описывает угловую зависимость спектра плоских волн саморепродуцирующегося пучка. Подставив в (7.53) выражение для разложения этой периодической функции в ряд Фурье

$$F_n(\theta) = \sum_{m=-\infty}^{\infty} C_{m,n} \exp(im\theta) \tag{7.56}$$

и используя интегральное представление для функции Бесселя, получаем:

$$U(r, \varphi, z) = \exp(ik\sigma_0 z) \sum_{n=n_1}^{n_2} \sum_{m=-\infty}^{\infty} C_{m,n} J_m(kr\rho_n) \exp\left\{\frac{i2\pi nz}{z_0} + im\varphi\right\}. \tag{7.57}$$

Из сравнения (7.49) и (7.57) видно, что при $\sigma_0 = 0$ они функционально тождественны. Световые поля, представленные в (7.57), являются более широким семейством полей, чем те, которые представлены в (7.49), т. к. условие (7.51) более слабое, чем (7.45).

Из (7.57) также видно, что в световое поле с продольной периодичностью вносят отличный от нуля вклад плоские волны, волновые векторы которых лежат на конических поверхностях с образующими, наклоненными к оси z под углами

$$\cos\alpha_n = \cos\alpha_0 + n\lambda/z_0, \quad n \in [n_1, n_2]. \tag{7.58}$$

Далее на основе (7.57) рассматриваются алгоритмы расчета фазовых ДОЭ.

Рассмотрим полную систему ортогональных функций в круге радиуса R [25]

$$\Omega_{m,n}(r,\varphi) = A_{m,n} J_m\left(\frac{r}{R}\gamma_{m,n}\right)\exp\left(im\varphi\right), \tag{7.59}$$

$$A_{m,n} = \left\{\sqrt{\pi}RJ'_m(\gamma_{m,n})\right\}^{-1}, \tag{7.60}$$

где $J'_m(x)$ — производная функции Бесселя, $\gamma_{m,n}$ — корни функции Бесселя: $J_m(\gamma_{m,n}) = 0$, $\gamma_{-m,n} = \gamma_{m,-n} = \gamma_{-m,-n} = \gamma_{m,n}$.

Условие ортогональности функций (7.59) имеет вид:

$$\int_0^R\int_0^{2\pi} \Omega_{m,n}(r,\varphi)\Omega^*_{p,q}(r,\varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi = \delta_{m,p}\,\delta_{n,q}. \tag{7.61}$$

Любую непрерывную и ограниченную в круге функцию $U(r,\varphi)$ можно разложить по базису (7.59)

$$U(r,\varphi) = \sum_{m,n=-\infty}^{\infty} C_{m,n}\Omega_{m,n}(r,\varphi). \tag{7.62}$$

Для того, чтобы сформировать световое поле $U(r,\varphi,z)$ из уравнения (7.57), следует при $z=0$ сформировать световое поле с комплексной амплитудой вида

$$U_0(r,\varphi) = \sum_{n=n_1}^{n_2}\sum_{m=-\infty}^{\infty} \bar{C}_{m,n} J_m(kr\rho_n)\exp(im\varphi). \tag{7.63}$$

Сравнивая уравнения (7.62) и (7.63), находим, что в (7.63) только те коэффициенты $\bar{C}_{m,n}$ отличны от нуля, для которых выполняется условие:

$$kR\rho_n = \gamma_{m,n}. \tag{7.64}$$

Заметим, что корни функции Бесселя перемежаются согласно неравенствам $0 < \gamma_{m,1} < \gamma_{m+1,1} < \gamma_{m,2} < \gamma_{m+1,2} < \gamma_{m,3} \dots$.

Разложение (7.62) с условием (7.64) позволяет организовать итеративную процедуру расчета фазового ДОЭ, генерирующего световое поле с продольной периодичностью. Выбор величины z_0 и параметров σ_0 , λ, R определяет значения ρ_n, n_1, n_2 с помощью уравнений (7.54) и (7.55). Введя обозначение для p-ой оценки искомой фазы комплексной функции пропускания ДОЭ

$$S_p(r,\varphi) = \arg U_0^{(p)}(r,\varphi), \quad p = 0,1,2,\dots,$$

получим итеративный алгоритм ее расчета:

$$S_{p+1}(r,\varphi) = \arg\left\{\sum_{m,n=-\infty}^{\infty} B_{m,n}\exp\left[i\arg C^{(p)}_{m,n}\right]\Omega_{m,n}(r,\varphi)\right\}, \tag{7.65}$$

$$C^{(p)}_{m,n} = \int_0^R\int_0^{2\pi} \exp\left[iS_p(r,\varphi)\right]\Omega^*_{m,n}(r,\varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi, \tag{7.66}$$

$$B_{m,n} \neq 0 \quad \text{при} \quad Rk\rho_n = \gamma_{m,n}, \tag{7.67}$$

где $B_{m,n}$ — неотрицательные числа, которые задаются произвольно и определяют вид поперечного распределения интенсивности светового поля. Алгоритм (7.65)–(7.67) обладает свойством релаксации ошибки, т. е., можно показать, что для любого p имеет место неравенство:

$$\sum_{m,n=-\infty}^{\infty} \left(\left| C_{m,n}^{(p+1)} \right| - B_{m,n} \right)^2 \leqslant \sum_{m,n=-\infty}^{\infty} \left(\left| C_{m,n}^{(p)} \right| - B_{m,n} \right)^2. \tag{7.68}$$

Заметим, однако, что алгоритм (7.65)–(7.67) не позволяет рассчитывать фазовые ДОЭ, формирующие продольно-периодические световые поля с заданным поперечным распределением интенсивности. Вид поперечного распределения интенсивности является не входной, а выходной функцией для данного алгоритма.

Рассмотрим алгоритм для расчета ДОЭ, формирующих радиальные волны с продольной периодичностью и с заданным поперечным распределением интенсивности вдоль оси. Алгоритм основан на разложении амплитуды радиального поля в ряд по ортогональным базисным функциям

$$Q_{m,n}(r,z) = D_{m,n} J_m\left(\frac{r}{R}\gamma_{m,n}\right) \exp\left[i2\pi\left(\frac{n}{z_0} + \frac{\sigma}{\lambda}\right)z\right], \tag{7.69}$$

где

$$D_{m,n} = \left\{\sqrt{z_0} R J'_m(\gamma_{m,n})\right\}^{-1}, \quad U(r,z) = \sum_{m,n=-\infty}^{\infty} \bar{C}_{m,n} Q_{m,n}(r,z). \tag{7.70}$$

Вместо (7.65), (7.66) (уравнение (7.67) сохраняется) запишем аналогичные уравнения для итеративного поиска фазы $T_p(r,z) = \arg U^{(p)}(r,z)$ светового поля:

$$T_{p+1}(r,z) = \arg\left\{ \sum_{m,n=-\infty}^{\infty} B_{m,n} \exp\left[i \arg \bar{C}_{m,n}^{(p)}\right] Q_{m,n}(r,z) \right\}, \tag{7.71}$$

$$\bar{C}_{m,n}^{(p)} = \int_0^R \int_0^{z_0} \sqrt{I_0(r,z)} \exp\left[iT_p(r,z)\right] Q_{m,n}^*(r,z) r \, \mathrm{d}r \, \mathrm{d}z, \tag{7.72}$$

где $I_0(r,z)$ — заданное периодическое (с периодом z_0) радиальное распределение интенсивности вдоль оси z, причем при $z = 0$, функция $I_0(r)$ описывает интенсивность освещающего пучка, $B_{m,n}$ — произвольные неотрицательные числа, от которых зависит вид фазы и скорость сходимости алгоритма. Фаза радиального ДОЭ, полученная за p шагов алгоритма (7.67), (7.71)–(7.72), равна функции $T_p(r, z = 0)$.

Пример 7.1. Расчет ДОЭ [32], формирующих продольно-периодические поля, производился с помощью алгоритма (7.65)–(7.67). Параметры расчета выбирались следующие: $R = 6$ мм, $\lambda = 1{,}06$ мкм, $z_0 = 20$ мм, $\rho_0 = 0{,}01$ мм. При этих параметрах модовый характер светового поля, прошедшего ДОЭ, сохраняется приблизительно на расстоянии [14, 20]

$$z_1 = R\rho_0^{-1}, \tag{7.73}$$

где ρ_0 — максимальное значение параметра ρ_n в уравнении (7.54). Это расстояние тем меньше, чем больше отличных от нуля слагаемых в сумме (7.62).

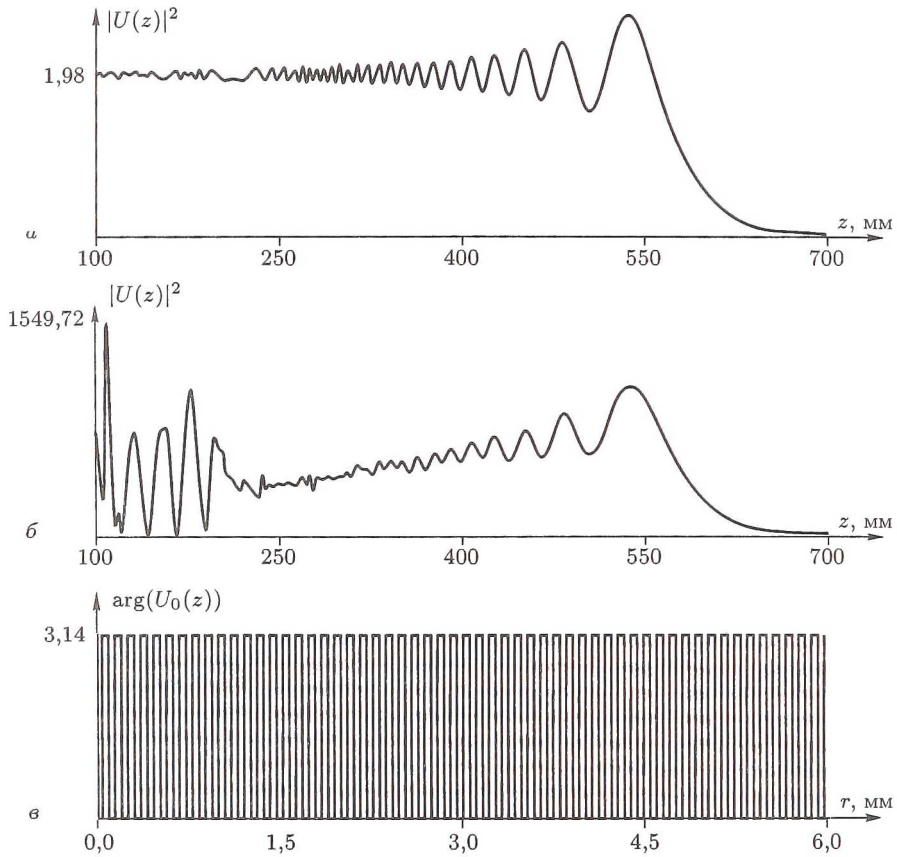


Рис. 7.1. Распределения интенсивности вдоль оси z для амплитудно-фазового ($а$) и фазового ($б$) ДОЭ, рассчитанных итеративным алгоритмом с одним слагаемым суммы (7.62), а также радиальное сечение фазы этого ДОЭ ($в$)

На рис. 7.1 показаны распределения нормированной интенсивности на оптической оси для ДОЭ, который получен при выборе отличным от нуля одного первого коэффициента суммы (7.62). Это единственное слагаемое пропорционально функции Бесселя нулевого порядка $J_0(kr\rho_0)$. Распространение такого поля вдоль оси z рассчитывалось с помощью преобразования Френеля, которое, в свою очередь, вычислялось с помощью алгоритма быстрого преобразования Фурье.

Из рисунков 7.1$а$ и 7.2$а$ видно, что бесселевый пучок сохраняется по форме, испытывая только колебания интенсивности с постоянной амплитудой на оптической оси до расстояния оцениваемого по формуле (7.73) $z = 300$ мм. Далее по мере распространения пучка, из-за конечности апертуры ДОЭ, амплитуда осцилляций нарастает, и при $z > 550$ мм осевая интенсивность пучка монотонно спадает, что означает начало дифракционной расходимости пучка. Радиус бесселевого пучка $\approx 0{,}02$ мм.

На рисунках 7.1$б$ и 7.2$б$ показаны аналогичные кривые распределения интенсивности для оптического элемента, полученного из предыдущего ДОЭ с заменой амплитуды на единичную и без изменения фазы. Фактически такой ДОЭ представляет собой фазовый бинарный аксикон. Из рисунков 7.1$б$ и 7.2$б$ видно, что хотя амплитуда осцилляций интенсивности вдоль оси увеличивается, модовый характер пучка, то есть распространение без изменения вида, сохраняется. Понятно, что вы-

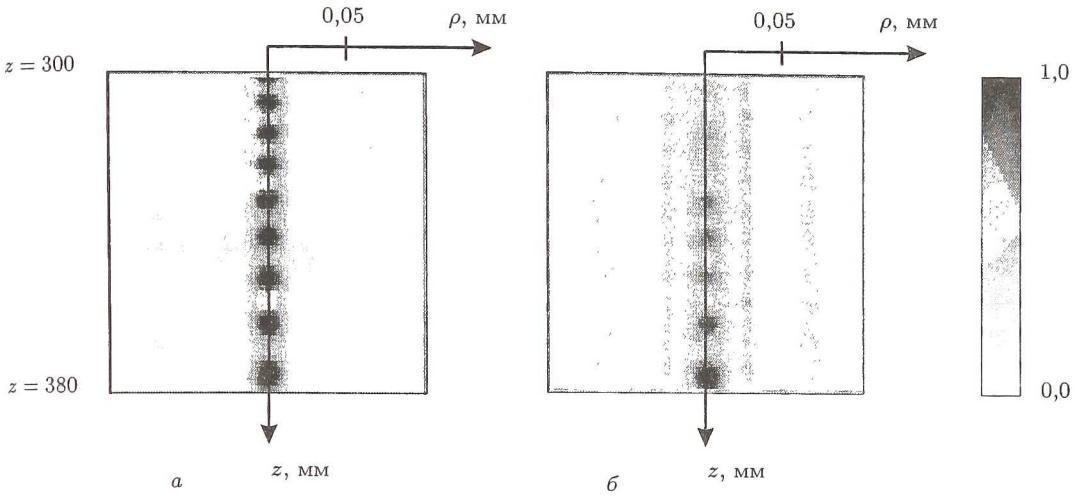


Рис. 7.2. Полутоновые нормированные распределения интенсивности в координатах (ρ, z) для амплитудно-фазового $(а)$ и фазового $(б)$ ДОЭ, рассчитанных итеративным алгоритмом с одним слагаемым суммы (7.62). Черный цвет соответствует максимальному значению интенсивности

бором только одного слагаемого в сумме (7.62) получить продольно периодическое поле нельзя. Для этого требуется хотя бы два слагаемых отличных от нуля.

Пример 7.2. На рис. 7.3 показаны осевые распределения нормированной интенсивности для амплитудно-фазового $(а)$ и только фазового $(б)$ ДОЭ, рассчитанных по методу (7.65)–(7.67), если отличны от нуля коэффициенты с номерами $m = 0$, $n = 0, 1$, и радиальное сечение фазы этого ДОЭ $(в)$. Видно, что световое поле обладает продольной периодичностью с периодом $z_0 = 20$ мм на расстоянии около 400 мм ($R = 6$ мм, $\lambda = 1{,}06$ мкм). Из рис. 7.3$б$ видно, что фазовый ДОЭ также имеет вдоль оптической оси чередование максимумов и минимумов интенсивности через каждые 20 мм, хотя амплитуда модуляции света в каждом периоде меняется (растет). На рис. 7.4 показаны полутоновые нормированные распределения интенсивности в координатах (ρ, z) в диапазоне $|\rho| < 0{,}1$ мм, 295 мм $< z <$ 375 мм, сформированные амплитудно-фазовым $(а)$ и фазовым $(б)$ ДОЭ. Заметим, что в этом и предыдущем случаях фаза ДОЭ была радиально симметричной: $m = 0$ (рис. 7.1$в$, 7.3$в$).

Пример 7.3. В этом примере фаза ДОЭ уже не будет радиальной (рис. 7.5$в$). В этом случае отличными от нуля выбирались коэффициенты с номерами $m = 1$, $n = 0, 1, 2, 3$. Из рис. 7.5 и 7.6 видно, что характер периодичности поля усложняется, и расстояние, на котором проявляется модовый характер поля, сокращается с 400 мм до 300 мм. Поперечное распределение интенсивности носит кольцевой характер: на оптической оси — нулевая интенсивность. На рис. 7.5 показаны кривые интенсивности при $\rho = 0{,}02$ мм.

Число итераций, при котором наступает стагнация сходимости алгоритма, то есть не происходит заметного изменения ошибки, пропорционально числу ненулевых слагаемых (7.62). Так, при выборе одного или двух коэффициентов (7.62), ошибка (7.68) устанавливается за 1–2 итерации, при выборе 4-х коэффициентов — за 5–6.

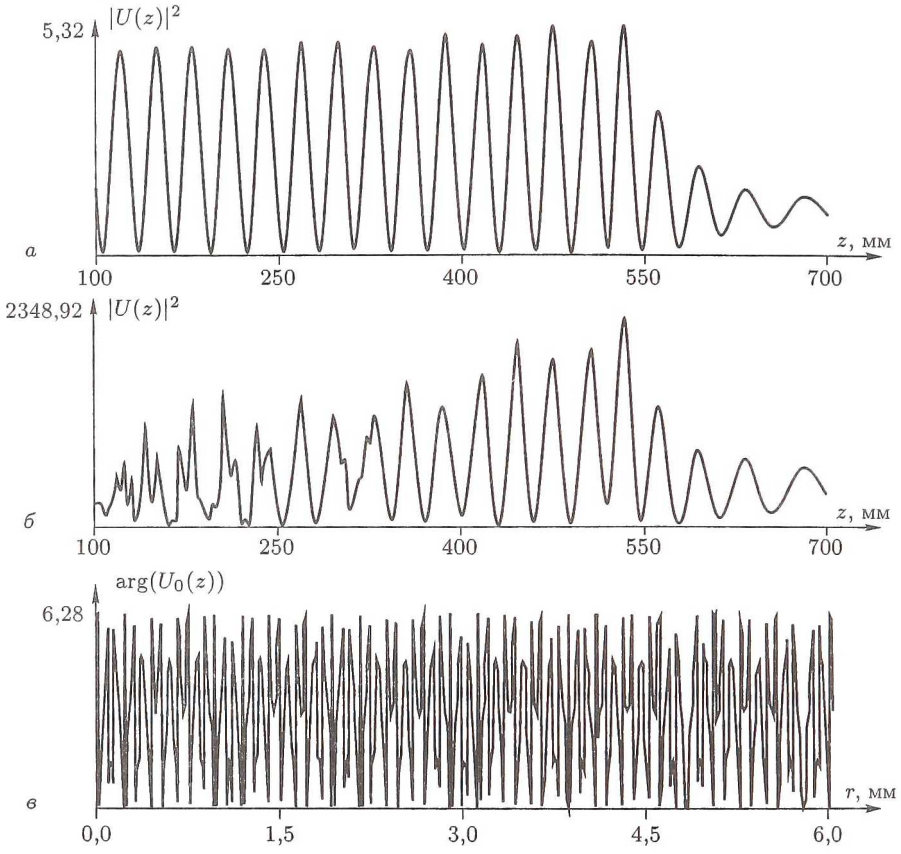


Рис. 7.3. Распределения интенсивности вдоль оси для амплитудно-фазового (*а*) и фазового (*б*) ДОЭ, рассчитанных итеративным алгоритмом с двумя первыми слагаемыми суммы (7.62), а также радиальное сечение фазы этого ДОЭ (*в*)

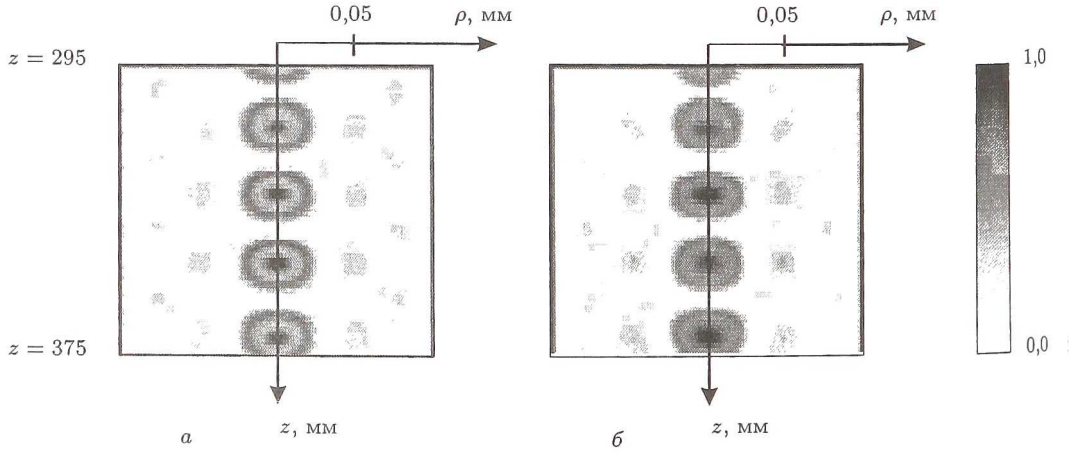


Рис. 7.4. Полутоновые нормированные распределения интенсивности в координатах (ρ, z) для амплитудно-фазового (*а*) и фазового (*б*) ДОЭ, рассчитанных итеративным алгоритмом с двумя первыми слагаемыми суммы (7.62)

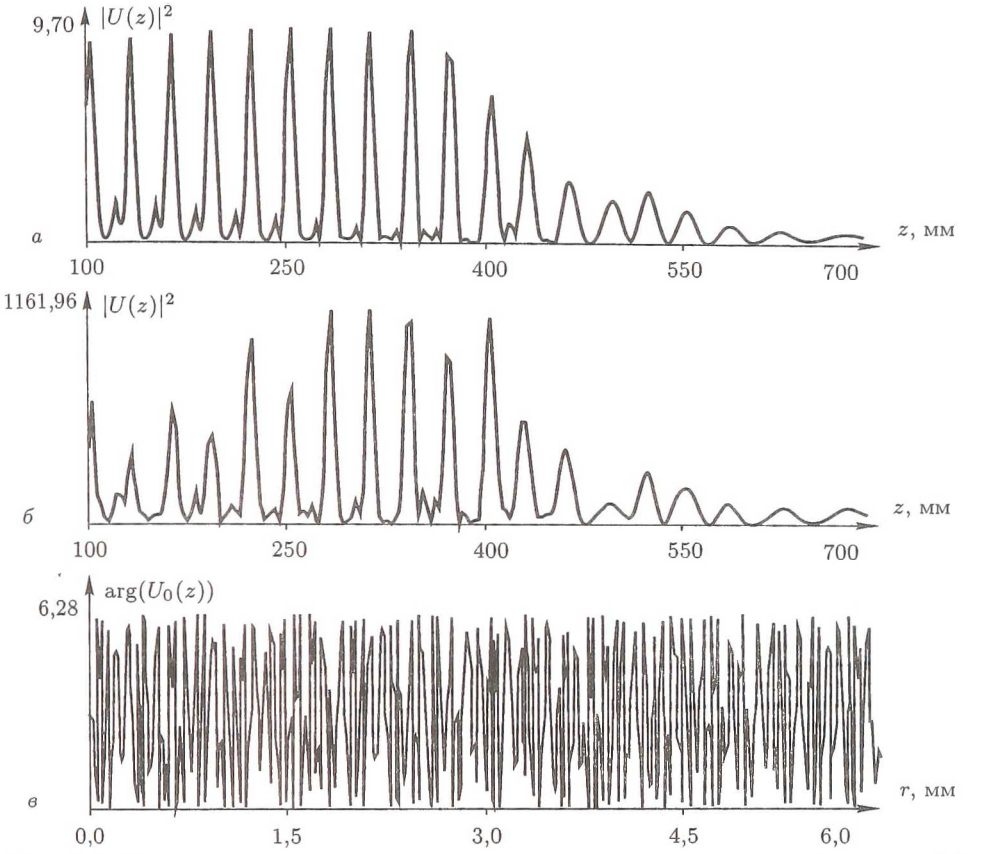


Рис. 7.5. Распределения интенсивности вдоль оси для амплитудно-фазового (*а*) и фазового (*б*) ДОЭ, рассчитанных итеративным алгоритмом с четырьмя первыми слагаемыми суммы (7.62), а также радиальное сечение фазы этого ДОЭ (*в*)

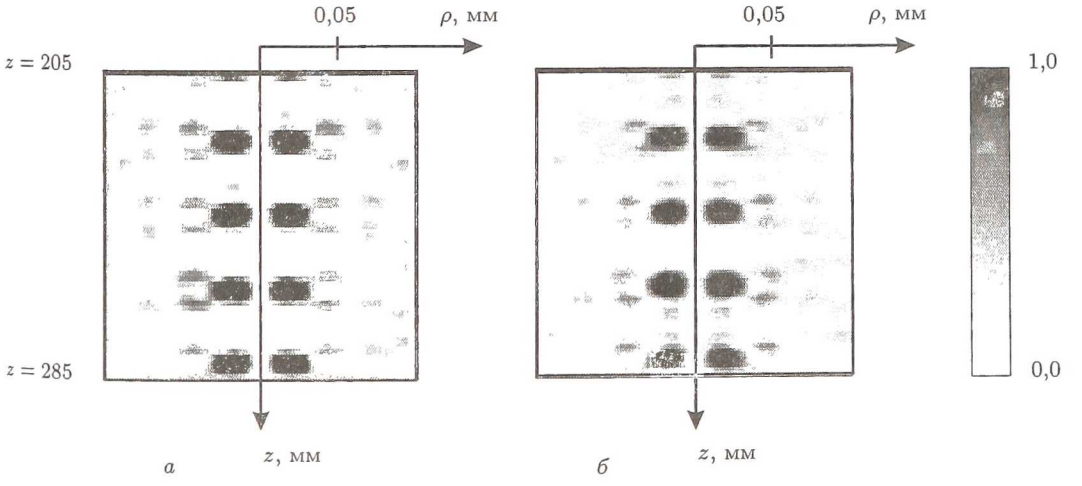


Рис. 7.6. Полутоновые нормированные распределения интенсивности в координатах (ρ, z) для амплитудно-фазового (*а*) и фазового (*б*) ДОЭ, рассчитанных итеративным алгоритмом с четырьмя первыми слагаемыми суммы (7.62)

7.3. Алгоритм расчета ДОЭ для генерации вращающихся многомодовых пучков Бесселя

Эффект вращения поперечного сечения бесселевого пучка объясняется тем, что в результате распространения двух конических волн с различными углами при вершинах конусов и с различными «винтовыми» составляющими, возникает интерференционная картина, которая вращается вдоль оптической оси. При распространении более двух таких волн характер интерференционной картины усложняется, и для получения пучка, поперечное сечение которого вращается как единое целое, необходимо наложить определенные условия на параметры конических волн.

Из уравнения (7.57) следует, что если отличные от нуля коэффициенты $C_{m,n}$ имеют одинаковый номер $n = n_0$, то волна распространяется без дифракции и без изменения вида, как бесселева мода:

$$U(r,\varphi,z) = \exp\left(ik\sigma_0 z + i\frac{2\pi n_0 z}{z_0}\right)\sum_{m=-\infty}^{\infty} C_m J_m(kr\rho_0)\exp(im\varphi), \tag{7.74}$$

где $C_m = C_{m,n_0}$, $\rho_0 = \rho_{n_0}$.

Продольная периодичность светового поля (7.57) имеет место, когда хотя бы два слагаемых с различными номерами n_1 и n_2 отличны от нуля. Вращение поперечного сечения светового пучка, представленного в (7.57), является частным случаем продольной периодичности и обеспечивается согласованным выбором отличных от нуля слагаемых с разными номерами n и m. Для получения условия вращения пучка вместо (7.57) запишем выражение для интенсивности:

$$I(r,\varphi,z) = \sum_{n,m=-\infty}^{\infty} |C_{m,n}J_m(kr\rho_n)|^2 + \\ + 2\sum_{m,n=-\infty}^{\infty}\sum_{m',n'=-\infty}^{\infty} |C_{m,n}C_{m',n'}|\, J_m(kr\rho_n)J_{m'}(kr\rho_{n'})\cos\Phi_{m,n}^{m',n'}(r,\varphi), \tag{7.75}$$

$$\Phi_{m,n}^{m',n'}(r,\varphi) = \arg C_{m,n} - \arg C_{m',n'} + \frac{2\pi(n-n')z}{z_0} + (m-m')\varphi. \tag{7.76}$$

Из (7.76) следует, что вид функции $I(r,\varphi+\Delta\varphi,z+\Delta z)$ не изменится при условии

$$(m-m')/(n-n') = A, \quad n \neq n', \quad m \neq m', \tag{7.77}$$

где A — постоянная величина, $\Delta\varphi$, Δz — малые произвольные приращения величин φ и z. Из (7.77) следует, что если в (7.57) отличны от нуля только два слагаемых, то при любых значениях их номеров $n \neq n'$, $m \neq m'$ пучок будет вращаться вокруг оси распространения. Из (7.77) также следуют менее общие, но более удобные для использования условия вращения пучка:

$$m/n = A, \tag{7.78}$$

$$m - n = A. \tag{7.79}$$

При выполнении условия (7.77), (7.78) или (7.79) с изменением расстояния на величину Δz значение интенсивности $I(r,\varphi,z)$ в уравнении (7.75) не изменится при выборе величины угла $\varphi + \Delta\varphi$, где

$$\Delta\varphi = -2\pi\Delta z/Az_0. \tag{7.80}$$

Т. к. Δz и $\Delta\varphi$ произвольные, это означает наличие вращения пучка при распространении. Из уравнения (7.80) видно, что если A^{-1} — целое число, то на длине периода z_0 будет совершаться целое число оборотов при вращении пучка. Направление вращения (левое или правое) определяется знаком постоянной A.

Рассмотрим ортогональный базис собственных функций колебаний круглой тонкой мембраны с радиусом R [33], который описывается уравнениями (7.59), (7.60).

Функцию пропускания ДОЭ $U_0(r,\varphi)$ можно искать как линейную комбинацию функций (7.59). Модули коэффициентов при этом считаются заданными, а их аргументы являются свободными параметрами. Модуль самой функции пропускания $|U_0(r,\varphi)|$ также задан и имеет смысл амплитуды освещающего ДОЭ пучка. Данная линейная комбинация имеет вид

$$U_0(r,\varphi) = \sum_{m,n=-\infty}^{\infty} C_{m,n}\Omega_{m,n}(r,\varphi), \tag{7.81}$$

при условии, что

$$kR\rho_n = \gamma_{m,n}. \tag{7.82}$$

Задачу поиска фазы функции $U_0(r,\varphi)$ и, соответственно, аргументов $C_{m,n}$ в сумме (7.81) можно сформулировать как задачу минимизации квадратичного критерия:

$$K = \int_0^{\infty}\int_0^{2\pi}\left[|U_0(r,\varphi)| - \left|\sum_{m,n=-\infty}^{\infty} C_{m,n}\Omega_{m,n}(r,\varphi)\right|\right]^2 r\,\mathrm{d}r\,\mathrm{d}\varphi. \tag{7.83}$$

Градиентный метод минимизации критерия (7.83) нетрудно построить, используя результаты работы [34]. При этом итеративная процедура поиска аргументов коэффициентов $\nu_{m,n} = \arg C_{m,n}$, будет выглядеть следующим образом:

$$\nu_{m,n}^{(p+1)} = \nu_{m,n}^{(p)} + \tau_p \frac{\partial K}{\partial \nu_{m,n}^{(p)}}, \tag{7.84}$$

$$\frac{\partial K}{\partial \nu_{m,n}^{(p)}} = 2\left|C_{m,n}\overline{C}_{m,n}^{(p)}\right| \sin\left(\nu_{m,n}^{(p)} - \overline{\nu}_{m,n}^{(p)}\right), \tag{7.85}$$

$$\overline{C}_{m,n} = \int_0^{\infty}\int_0^{2\rho} F(r,\varphi)\Omega_{m,n}^{*}(r,\varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi, \tag{7.86}$$

$$F(r,\varphi) = |U_0(r,\varphi)|\,\frac{U(r,\varphi)}{|U(r,\varphi)|} - U(r,\varphi), \tag{7.87}$$

$$U(r,\varphi) = \sum_{m,n=-\infty}^{\infty} C_{m,n}\Omega_{m,n}(r,\varphi), \tag{7.88}$$

где $\overline{\nu}_{m,n} = \arg \overline{C}_{m,n}$, p — номер итерации. Модули $|C_{m,n}|$ задаются произвольно и являются параметрами задачи, от которых зависит вид поперечного сечения формируемого пучка.

Шаг τ_p в уравнении (7.84), следуя работе [34], можно представить в виде

$$\tau_p = \frac{K}{\overline{K}_p}, \tag{7.89}$$

$$\overline{K}_p = -\sum_{m,n=-\infty}^{\infty} \left| \frac{\partial K}{\partial \nu_{m,n}^{(p)}} \right|^2, \tag{7.90}$$

Этот итеративный алгоритм отличается от (7.65)–(7.66) и (7.71)–(7.72).

Итеративным алгоритмом рассчитывались фазовые оптические элементы, формирующие световые поля, состоящие эффективно (на 80–90%) из 2–3 мод Бесселя.

Пример 7.4. Пример бездифракционного пучка или стабильной моды Бесселя показан на рис. 7.7 в левом столбце. Параметры расчета следующие: $\lambda = 1{,}06$ мкм, $z_0 = 20$ мм, $R = 1$ мм, $\sigma_0 = 0{,}01$. На рис. 7.7 в левом столбце показаны: бинарная фаза ДОЭ (а1), формирующего световой пучок, эффективно состоящий из двух слагаемых в уравнении (7.57) с номерами (m, n): $(-3, -3)$, $(3, -3)$. Так как оба слагаемых имеют одинаковый номер $n = -3$, то согласно (7.74) пучок при распространении не изменяет свою структуру. В левом столбце показаны распределения поперечной интенсивности данного пучка (негатив) на расстояниях z: 25 мм (а1), 30 мм (в1), 35 мм (г1), 40 мм (д1), 45 мм (е1).

Пример светового поля с продольной периодичностью показан в среднем столбце рис. 7.7. Параметры расчета: $\lambda = 1{,}06$ мкм, $z_0 = 20$ мм, $R = 3$ мм, $\sigma_0 = 0{,}02$. В (7.57) сохранялись отличными от нуля только три слагаемых с номерами (m, n): $(-2, -3)$, $(0, -1)$, $(2, -2)$. При этом условия вращения (7.77) и (7.78) не соблюдаются. На рис. 7.7 показаны: полутоновая фаза такого ДОЭ (а2) (черный цвет — 2π, белый — 0) и поперечные распределения интенсивности (негатив), сформированные на расстояниях z: 90 мм (а2), 95 мм (в2), 100 мм (г2), 105 мм (д2), 110 мм (е2).

В правом столбце рис. 7.7 показан пример вращающегося многомодового пучка Бесселя. Параметры расчета: $\lambda = 1{,}06$ мкм, $z_0 = 20$ мм, $R = 2$ мм, $\sigma_0 = 0{,}015$. В (7.57) отличными от нуля выбирались только три слагаемых с номерами (m, n): $(-3, -3)$, $(-1, -1)$, $(1, 1)$. Условия вращения (7.77) и (7.78) для данных номеров соблюдаются. В правом столбце показаны полутоновая фаза ДОЭ (а3), формирующего световой пучок, эффективно содержащий три моды Бесселя, поперечные распределения интенсивности которого (негатив) показаны на расстояниях z: 70 мм (а3), 75 мм (в3), 80 мм (г3), 85 мм (д3), 90 мм (е3). Из рисунка видно, что имеет место вращение пучка при его распространении, но также видно, что происходят некоторые изменения вида картины поперечного сечения пучка, обусловленные интерференцией трех главных мод (около 90% энергии) с модами высших порядков.

Поперечные картины дифракции на рис. 7.7 вычислялись с помощью преобразования Френеля, которое, в свою очередь, вычислялось с помощью алгоритма быстрого преобразования Фурье.

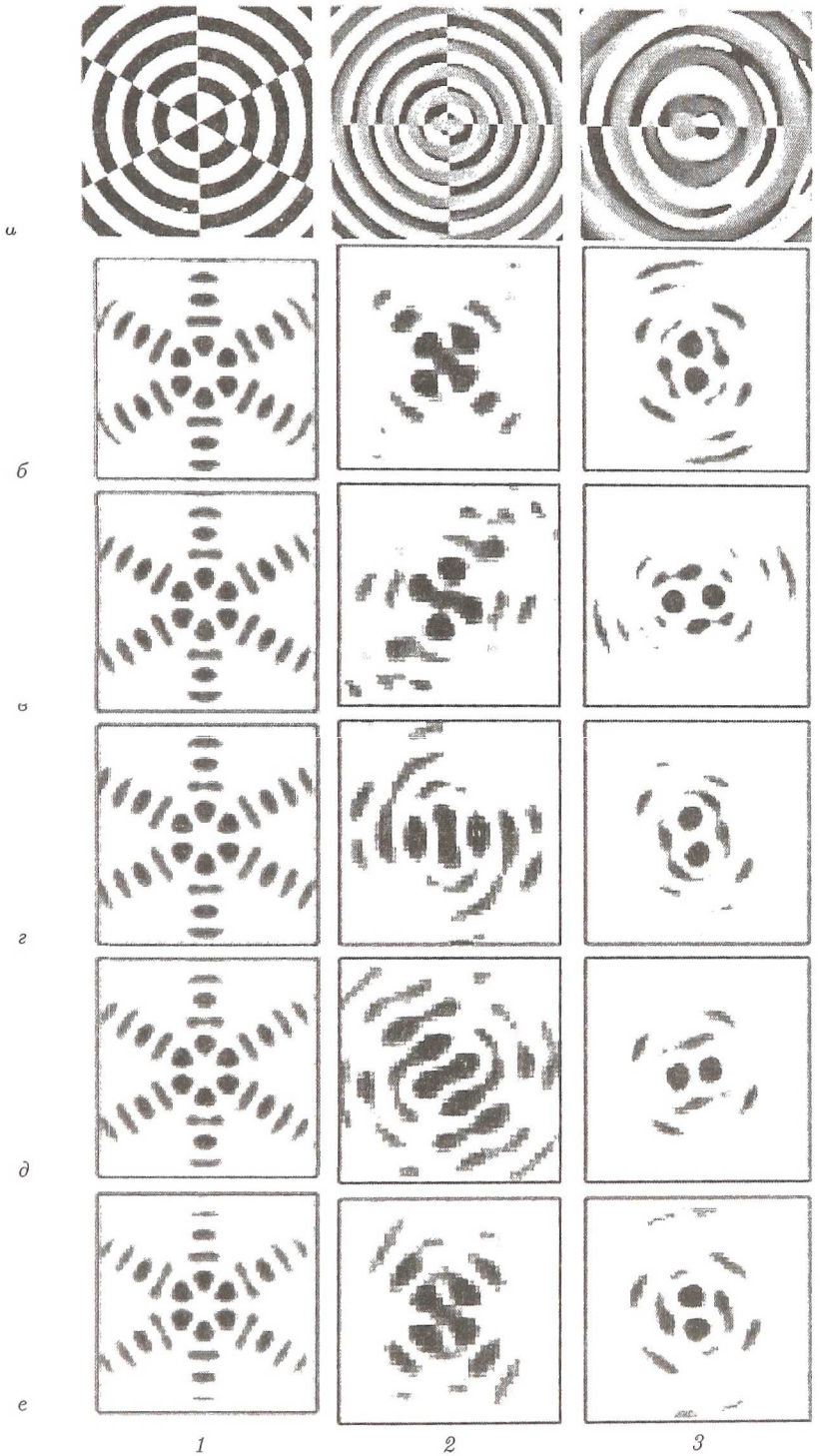


Рис. 7.7. Итеративно рассчитанные фазы ДОЭ (*а1—а3*) и поперечные сечения многомодовых бесселевых пучков (*б1—е3*), генерируемых такими ДОЭ

7.4. Формирование пары вращающихся бездифракционных пучков с помощью бинарного фазового ДОЭ

Из предыдущих разделов ясно, что если пространственный спектр когерентного светового поля состоит из конечного набора узких концентрических колец, то такое поле будет самовоспроизводиться через равные расстояния вдоль оси распространения [8]. Частным случаем таких световых полей являются бездифракционные бесселевы пучки, комплексная амплитуда которых является суперпозицией функций Бесселя первого рода с угловыми гармониками [20, 35, 36]. Заметим, что такие бездифракционные световые поля можно сформировать и для частично-когерентного света [37, 38]. Так как идеальные самовоспроизводящиеся световые поля имеют бесконечные размеры и энергию, то для их практического формирования требуется использовать аподизацию, например, гауссовой апертурой [22, 39]. Для формирования стабильных (период саморепродукции равен нулю) бездифракционных бесселевых пучков можно использовать: интерферометр Фабри–Перо [11, 40]; узкую кольцевую щель в экране, расположенном на фокусном расстоянии перед сферической линзой [21]; бинарные амплитудные голограммы с несущей пространственной частотой [24, 41]; амплитудные [42] и фазовые [43, 44] дифракционные аксиконы.

Чтобы сформировать стабильный многомодовый бесселев пучок с заданным вкладом каждой отдельной моды, можно итеративным способом спроектировать нужный для этого фазовый ДОЭ [27], как это было показано в предыдущих разделах. Свойствами, близкими к бездифракционным пучкам, обладают и другие типы световых волн, амплитуда которых описывается функциями Бесселя второго рода [45], или функциями Неймана [13]. Близкие к бездифракционным пучки можно сформировать амплитудными дифракционными элементами, которые рассчитываются по некоторым специальным итеративным процедурам [46].

Частным случаем самовоспроизводящихся пучков являются пучки, обладающие вращением вдоль оси распространения. Простейшие из них — суперпозиция двух бесселевых функций [12] с разными показателями и масштабами [13, 14]. В [47, 48] получено общее условие для формирования вращающихся бездифракционных пучков. Сформировать такие пучки можно с помощью фазовых трохосонов (trochosons) [43], бинарных геликонов (helicons) [14] или бинарных фазовых винтовых аксиконов [14, 44].

В данном разделе рассмотрены вращающиеся бесселевы пучки, сформированные с помощью бинарных фазовых ДОЭ с пространственной несущей частотой, которые так же как и в работе [14] имеют около 40% энергетической эффективности, но, в отличие от винтовых аксиконов, эти ДОЭ формируют сразу два одинаковых пучка (по 40% энергии в каждом), вращающиеся в разные стороны. Причем, в отличие от [14], каждый из этих пучков можно описать линейной комбинацией функций Бесселя с угловыми гармониками и с произвольно заданным весовым вкладом каждой из них.

В [14] предложено формировать многомодовые пучки Бесселя с помощью ДОЭ, функция пропускания которого $U(r,\theta)$ является линейной комбинацией аксиконов с угловыми гармониками. Тогда вместо (7.57), при $z=0$, следует записать, что

$$U(r,\varphi)=\sum_{m=-\infty}^{\infty}\sum_{n=0}^{N}C_{mn}\exp\left(-ik\rho_n r+im\varphi\right). \tag{7.91}$$

Аксикон с пропусканием $\exp\left(-ik\rho_n r+im\varphi\right)$ эффективно сформирует световое поле вблизи оптической оси на отрезке $0<z<R/\rho_n$, амплитуда которого пропорциональна функции Бесселя [42]: $J_m(k\rho_n r)\exp\left(im\varphi\right)$. Для формирования

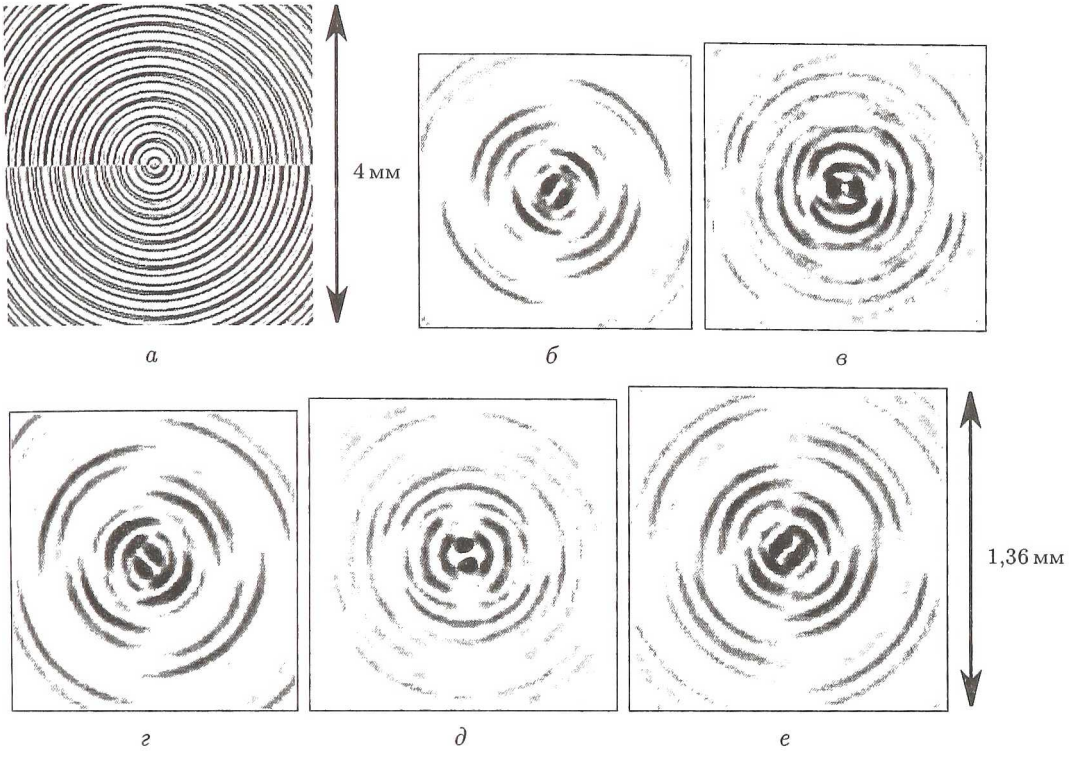


Рис. 7.8. Результат моделирования: фаза ДОЭ ($а$) и распределение поперечной интенсивности для бесселевого пучка с номерами $(-1,0)$ и $(1,1)$ на расстояниях z: 115 мм ($б$), 120 мм ($в$), 125 мм ($г$), 130 мм ($д$), 135 мм ($е$)

вращающихся пучков Бесселя в [14] предложено использовать бинарные аксиконы с угловыми гармониками. В этом случае, разлагая функцию пропускания бинарного аксикона в функциональный ряд Гудмена–Сильвестри [49], получим

$$\exp\left[i\left\{-k\rho_0 r+m\varphi\right\}_2\right]=\sum_{p=-\infty}^{\infty}\operatorname{sinc}\left[\pi\left(p+\frac{1}{2}\right)\right]\exp\left[i(2p+1)\left(-k\rho_0 r+m\varphi\right)\right], \quad (7.92)$$

где $\{\ldots\}_2$ — обозначает операцию квантования по двум уровням: 0 и π. В (7.92) первые два слагаемых при $p=0,1$ имеют вид

$$\frac{2}{\pi}\exp\left[-ik\rho_0 r+im\varphi\right]+\frac{2}{3\pi}\exp\left[-3ik\rho_0 r+3im\varphi\right].$$

Вблизи оптической оси на отрезке $0<z<R/3\rho_0$ эти два слагаемых в результате интерференции эффективно сформируют функцию вида

$$\frac{2}{\pi}J_1(k\rho_0 r)\exp\left(im\varphi+i\frac{kz\rho_0^2}{2}\right)+\frac{2}{3\pi}J_1(3k\rho_0 r)\exp\left(3im\varphi+i\frac{9kz\rho_0^2}{2}\right).$$

Эффективность такого поля составляет немногим более $(2/\pi)^2=40,5\%$. Из (7.92) следует, что с помощью бинарного аксикона можно формировать вращающиеся пучки Бесселя только специального вида. Фазовый ДОЭ, рассчитанный согласно (7.63), позволяет формировать вращающиеся бездифракционные пучки более общего вида.

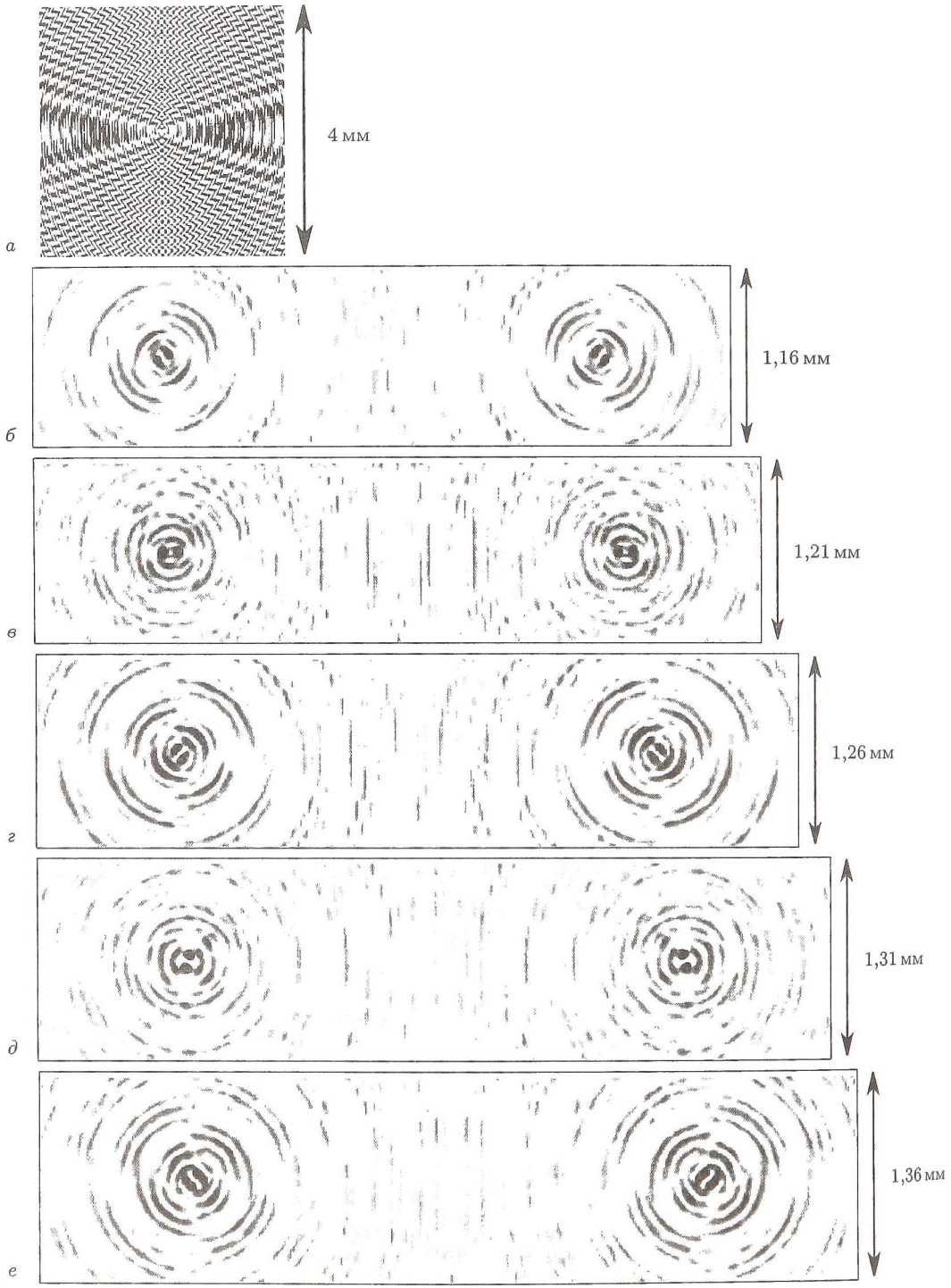


Рис. 7.9. Результат моделирования: бинарная фаза ДОЭ (*а*) и поперечные распределения интенсивности бесселевых пучков с номерами $(-1, 0)$, $(1, 1)$ и $(1, 0)$, $(-1, 1)$ на расстояниях z: 115 мм (*б*), 120 мм (*в*), 125 мм (*г*), 130 мм (*д*), 135 мм (*е*)

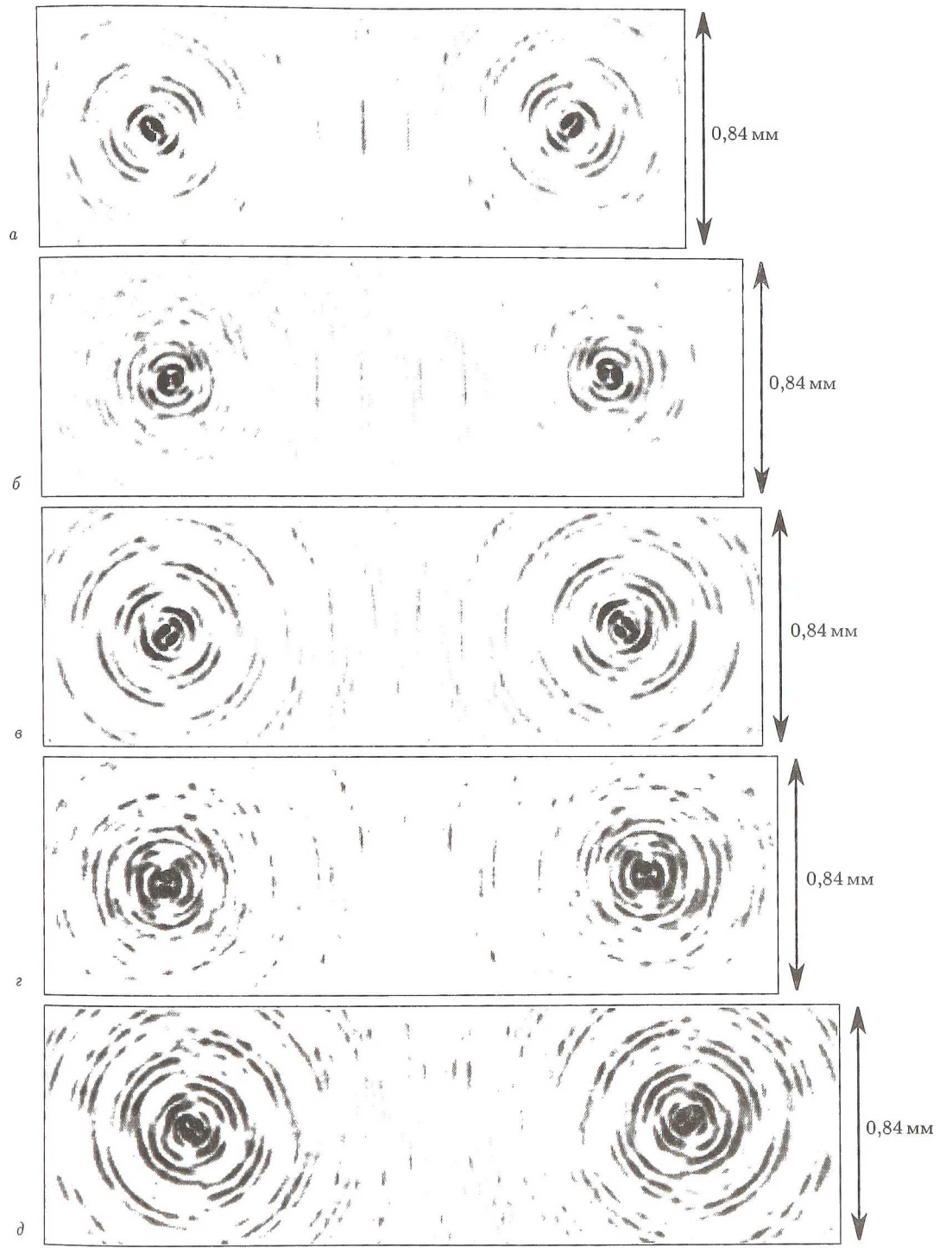


Рис. 7.10. Экспериментально зарегистрированные поперечные распределения интенсивности (негатив) после ДОЭ с фазой на рис. 7.9*а* на расстояниях z: 115 мм (*а*), 120 мм (*б*), 125 мм (*в*), 130 мм (*г*), 135 мм (*д*)

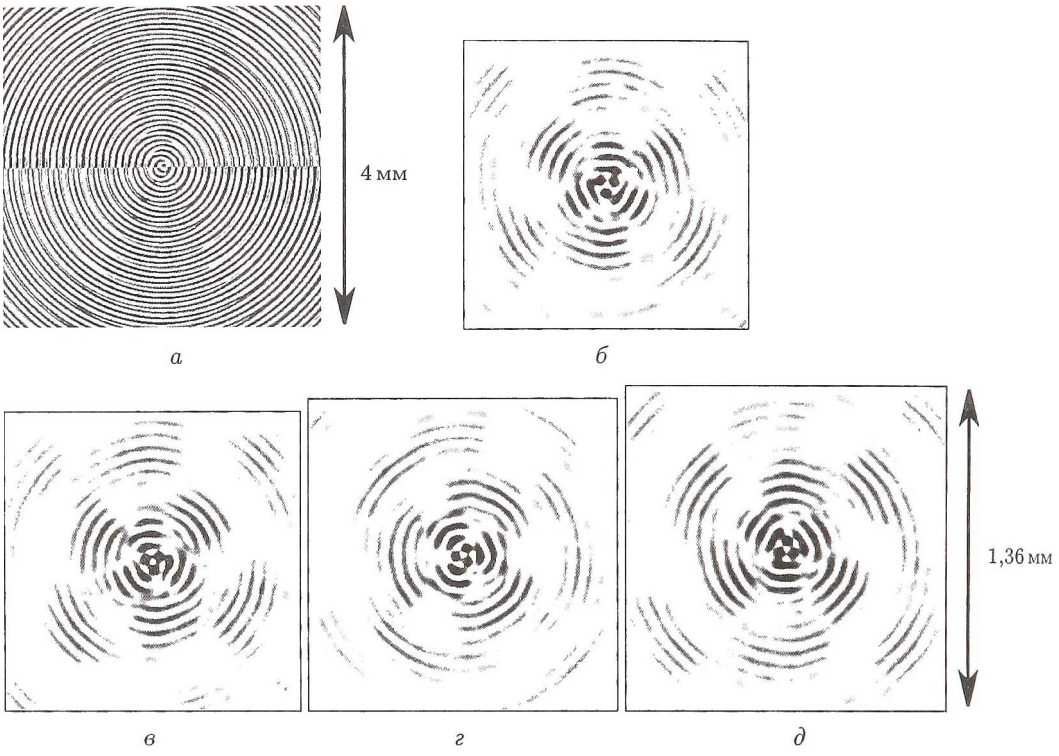


Рис. 7.11. Результат моделирования: фаза ДОЭ ($а$) и поперечное распределение интенсивности для бесселевого пучка с номерами мод $(-1, 0)$ и $(2, 1)$ на расстояниях z: 170 мм ($б$), 175 мм ($в$), 180 мм ($г$), 185 мм ($д$)

Бинарный вариант ДОЭ, основанный на фазе функции (7.63), может быть получен с помощью добавления линейной фазы по координате $x = r\cos\varphi$ и квантования по двум уровням 0 и π:

$$S_2(r,\varphi) = \{S(r,\varphi) + 2\pi\nu r\cos\varphi\}_2\,, \tag{7.93}$$

где ν — пространственная несущая частота, символ $\{\ldots\}_2$ имеет тот же смысл, что и в (7.92), $S(r,\varphi)$ — аргумент функции $U_0(r,\varphi)$ из уравнения (7.63) после достаточного числа итераций.

Используя снова разложение в функциональный ряд [49], получим для бинарной фазы из (7.93) следующее представление:

$$\exp\left[iS_2(r,\varphi)\right] = \sum_{p=-\infty}^{\infty} \operatorname{sinc}\left[\pi\left(p+\frac{1}{2}\right)\right]\exp\left\{i(2p+1)\left[S(r,\varphi)+2\pi\nu r\cos\varphi\right]\right\}. \tag{7.94}$$

В сумме (7.94) только два первых слагаемых при $p = 0, -1$ будут описывать два одинаковых вращающихся в разные стороны и распространяющиеся под углом друг к другу бесселевых пучка, каждый с эффективностью около 40%:

$$\frac{2}{\pi}\exp\left[iS(r,\varphi) + i2\pi\nu r\cos\varphi\right] + \frac{2}{\pi}\exp\left[-iS(r,\varphi) - i2\pi\nu r\cos\varphi\right]. \tag{7.95}$$

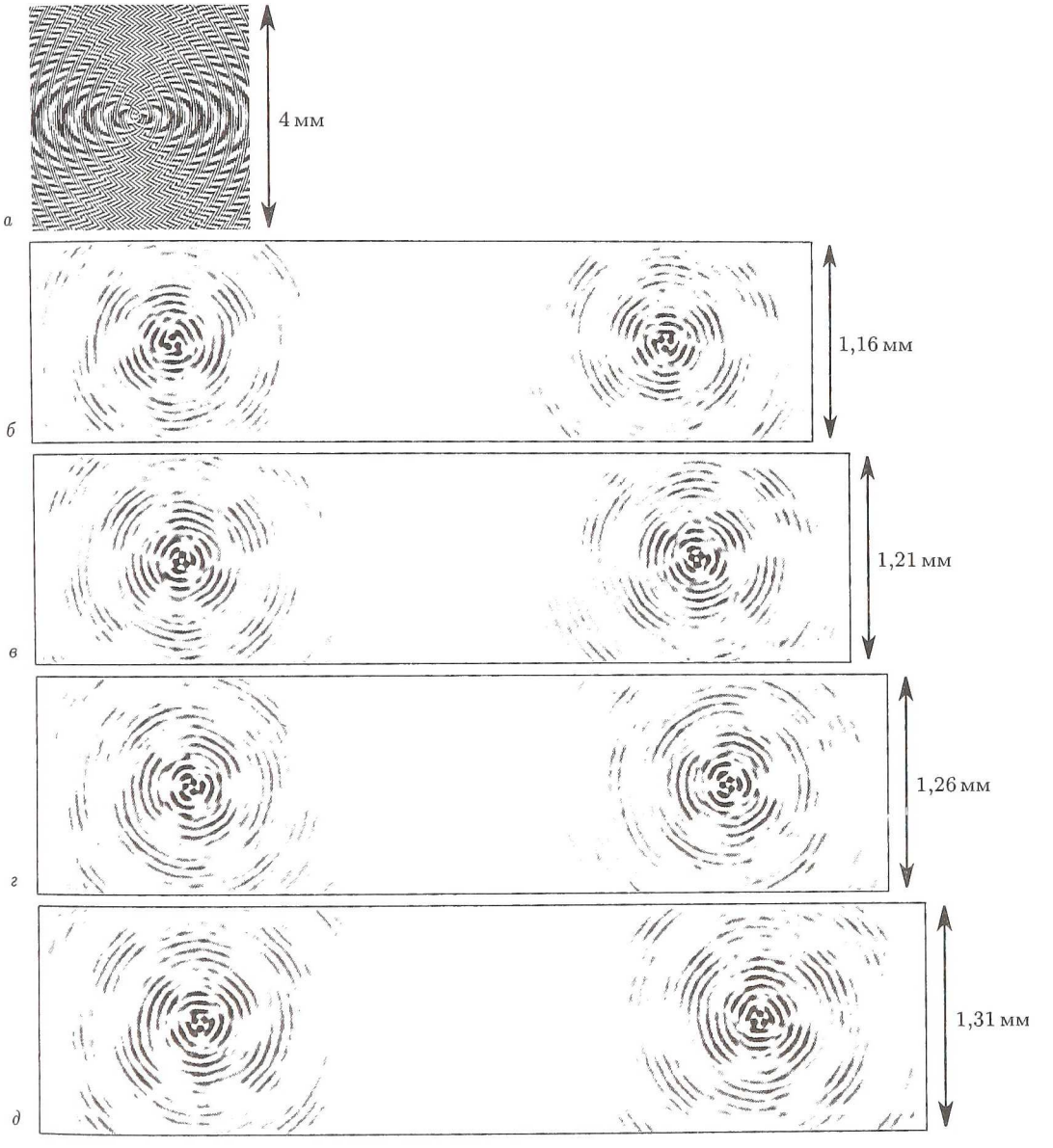


Рис. 7.12. Результат моделирования: бинарная фаза ДОЭ с несущей частотой ($а$) и поперечные распределения интенсивности для двухмодового пучка Бесселя на расстояниях z: 170 мм ($б$), 175 мм ($в$), 180 мм ($г$), 185 мм ($д$)

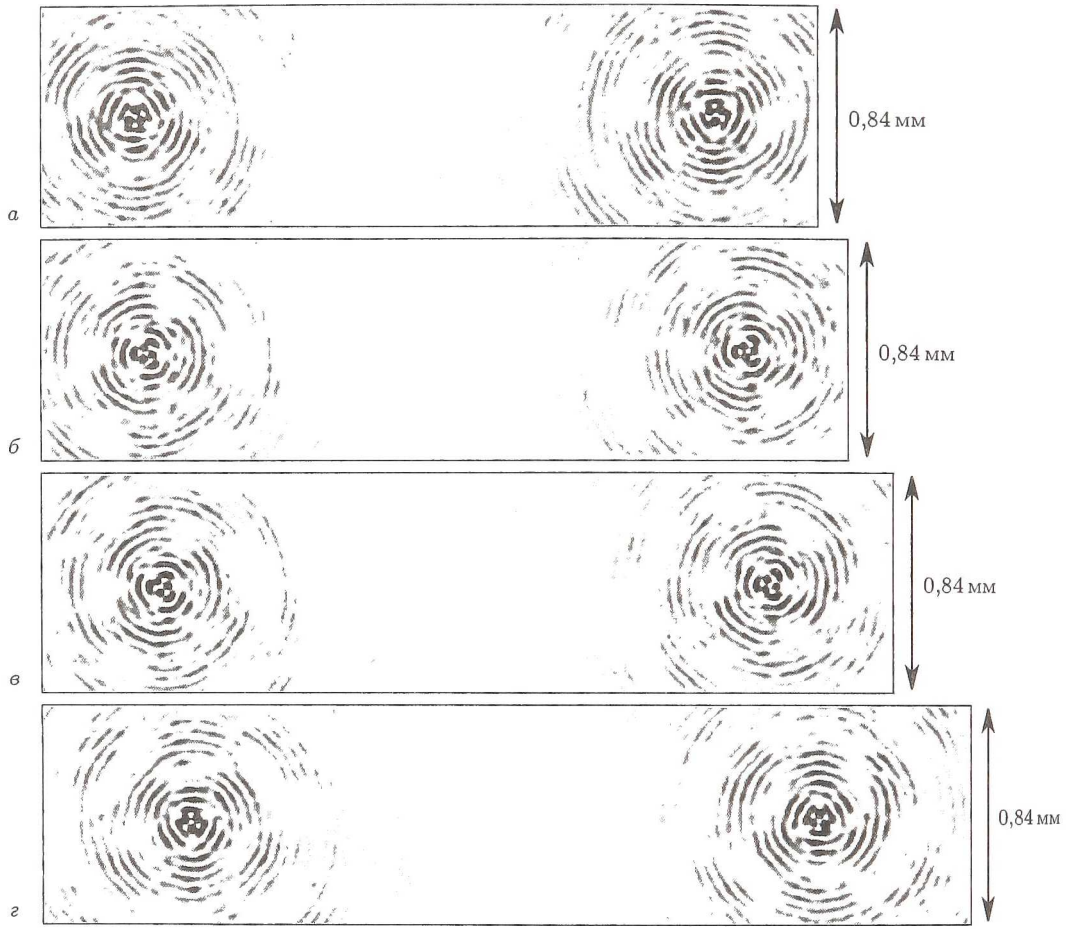


Рис. 7.13. Экспериментальные распределения интенсивности пучков, сформированных ДОЭ с фазой на рис. 7.12*а* на расстояниях z: 170 мм (*а*), 175 мм (*б*), 180 мм (*в*), 185 мм (*г*)

Пример 7.5. На рис. 7.8*а* показана полутоновая фаза ДОЭ (255 уровней квантования в диапазоне $[-\pi, \pi]$), рассчитанная как аргумент (7.57). При этом в сумме (7.57) отличными от нуля выбирались два слагаемых с номерами (m, n): $(-1, 0)$ и $(1, 1)$. Условие вращения (7.77) выполняется. Период вращения выбирался равным $z_0 = 20$ мм.

Размер фазовой функции на рис. 7.8*а*: 4 мм×4 мм (512×512 пикселов). На рис. 7.8*б–е* показаны полутоновые распределения поперечной интенсивности поля (негатив), сформированного с помощью фазового ДОЭ, рис. 7.8*а*. Распределения интенсивности в поперечном сечении пучка вычислены с помощью преобразования Френеля на расстояниях: $z = 115$ мм (7.8*б*), $z = 120$ мм (7.8*в*), $z = 125$ мм (7.8*г*), $z = 130$ мм (7.8*д*), $z = 135$ мм (7.8*е*). Таким образом охвачен полный период. Видно, что фазовый ДОЭ, полученный с помощью устранения амплитудной информации, успешно формирует бездифракционный вращающийся пучок.

На рис. 7.9*а* показана бинарная фаза, полученная из полутоновой фазы (рис. 7.8*а*) с помощью добавления пространственной частоты $\nu = 10\,\text{мм}^{-1}$ по формуле (7.93). На рис. 7.9*б–е* показаны полутоновые распределения поперечной интенсивности (негатив), сформированные бинарным фазовым ДОЭ (рис. 7.9*а*),

рассчитанные с помощью преобразования Френеля на тех же расстояниях. Видно, что бинарный ДОЭ в соответствии с формулой (7.95) формирует два одинаковых бездифракционных пучка, вращающихся в разные стороны. Из сравнения результатов моделирования на рис. 7.8*б–е* и рис. 7.9*б–е* видно, что оба элемента формируют очень похожие световые поля.

Бинарный фазовый ДОЭ (рис. 7.9*а*) был изготовлены с помощью низковольтового электронного генератора изображений Leica LION LV1 в Университете Йоенсуу, Финляндия (см. главу 4). Предельное разрешение этого литографа около 0,04 мкм. Размер ДОЭ 4 мм×4 мм (1000×1000 пикселов) с глубиной бинарного рельефа для $\lambda = 0{,}633$ микрона. Изготовленный по шаблону (рис. 7.9*а*) бинарный фазовый ДОЭ освещался коллимированным пучком He-Ne лазера. Распределение интенсивности на разных расстояниях после ДОЭ регистрировалось с помощью ПЗС-камеры с окном 6,6 мм×4,78 мм с числом пикселов 768×567.

На рис. 7.10 показана экспериментально зарегистрированная интенсивность света после ДОЭ на разных расстояниях в пределах одного периода. Видны отсутствие нулевого порядка, эффект вращения пучка и хорошее согласие с теорией (рис. 7.9*б–е*).

Пример 7.6. На рис. 7.11, 7.12 и 7.13 показаны результаты численного (рис. 7.11, 7.12) и натурного (рис. 7.13) экспериментов для случая распространения другого двухмодового пучка Бесселя с номерами мод (m, n): $(-1, 0)$ и $(2, 1)$. Период вращения равен $z_0 = 30$ мм. На рис. 7.11*а* показана полутоновая фаза ДОЭ, полученная как аргумент (7.57). На рис. 7.11*б–д* показаны вычисленные распределения интенсивности сформированного пучка на половине периода вращения.

На рис. 7.12*а* показана бинарная фаза, полученная по формуле (7.93) из фазы, представленной на рис. 7.11*а*, с помощью добавления пространственной частоты $\nu = 20\,\text{мм}^{-1}$.

На рис. 7.12*б–д* показаны вычисленные распределения интенсивности светового поля, сформированного бинарным фазовым ДОЭ (рис. 7.12*а*) на тех же расстояниях.

На рис. 7.13 показаны экспериментальные распределения интенсивности.

Из сравнения теории и эксперимента видно их хорошее согласие. Неполное сохранение вида поперечного сечения вращающегося пучка объясняется пренебрежением амплитудной информацией при расчете фазы как аргумента (7.57).

7.5. ДОЭ для формирования многомодовых пучков Гаусса–Лагерра

В главе 6 рассматриваются моды лазерного излучения. Они не расплываются и не изменяют пространственной структуры в процессе распространения в своей среде, приобретая лишь фазовый набег. Гауссовы моды сохраняют свою структуру и в свободном пространстве, изменяясь масштабно.

В работах [50–52] рассматриваются ДОЭ для генерации многомодовых пучков Гаусса–Лагерра (ГЛ), рассчитанные с помощью методов компьютерной оптики. Особый интерес представляют фазовые ДОЭ, имеющие повышенную энергетическую эффективность и многоканальный характер работы, позволяющий сформировать несколько модовых пучков.

В отличие от метода кодирования с пространственной несущей частотой, рассмотренного в [50], и итеративного метода с использованием вспомогательной области [51], в данном разделе используется метод, который состоит в итеративной аппроксимации функции пропускания ДОЭ конечной суммой гауссовых мод [52]. В [52, 53] были рассмотрены только радиально-симметричные моды ГЛ. В данном

разделе этот метод применен к обобщенным модам ГЛ. Такое расширение метода позволяет с высокой эффективностью формировать модовые пучки, поперечное сечение которых представляет собой заданное изображение. Кроме того, хотя рассматриваемый метод работает тем лучше, чем больше задействовано мод в пучке, он работоспособен и при формировании одномодового пучка. В этом случае за одну итерацию получается ДОЭ с фазовой функцией, пропорциональной знаковой функции моды. В данном разделе показано, что для эффективного формирования отдельных мод ГЛ можно использовать фазовые элементы, пропускание которых равно знаковой функции соответствующего многочлена Лагерра.

Алгоритм, рассмотренный в [52, 53], в случае обобщенных мод ГЛ выглядит следующим образом. Рассчитывается фаза $S(r,\varphi)$ ДОЭ в полярных координатах (r,φ), удовлетворяющая соотношению

$$f(r,\varphi) = A_0(r,\varphi)\exp\left[iS(r,\varphi)\right] = \sum_{n=0}^{N}\sum_{|m|\leqslant n} C_{n,m}\Psi_{n,m}(r,\varphi), \tag{7.96}$$

в котором $\Psi_{n,m}(r,\varphi)$ —обобщенная функция ГЛ в полярных координатах (вид в декартовых координатах см. (6.96), (6.97)):

$$\Psi_{n,m}(r,\varphi) = \frac{\sqrt{n!}}{a\sqrt{\pi(n+|m|)!}}\left(\frac{r}{a}\right)^{|m|}\exp\left[-\left(\frac{r}{\sqrt{2}a}\right)^2\right]L_n^{|m|}\left(\frac{r^2}{a^2}\right)\exp\left[im\varphi\right], \tag{7.97}$$

$$r^2 = x^2 + y^2, \quad \varphi = \operatorname{arctg}\frac{y}{x},$$

где $L_n^m(x) = (-1)^m\,\mathrm{d}^m\left[L_{m+n}(x)\right]/\,\mathrm{d}x^m$ — обобщенный многочлен Лагерра, $a = \sigma_0/\sqrt{2}$, σ_0 — радиус перетяжки, $A_0(r,\varphi)$ — известная амплитуда освещающего пучка, модули коэффициентов $|C_{n,m}|$ задаются произвольно, а аргументы $C_{n,m}$ являются свободными параметрами задачи. Квадраты модулей коэффициентов $|C_{n,m}|^2$ характеризуют энергетический вклад каждой моды в пучок.

Функции ГЛ ортогональны:

$$\int_0^{2\pi}\int_0^{\infty}\Psi_{n,m}(r,\varphi)\Psi_{k,l}(r,\varphi)r\,\mathrm{d}r\,\mathrm{d}\varphi = \delta_{n,k}\delta_{m,l}, \tag{7.98}$$

где $\delta_{n,m} = \begin{cases} 1, & n = m, \\ 0, & n \neq m. \end{cases}$ — символ Кронекера.

Тогда коэффициенты в уравнении (7.96) вычисляются по формуле

$$C_{n,m} = \int_0^{2\pi}\int_0^{\infty} A_0(r,\varphi)\exp\left[iS(r,\varphi)\right]\Psi^*_{n,m}(r,\varphi)\,r\,\mathrm{d}r\,\mathrm{d}\varphi. \tag{7.99}$$

Предлагаемый итеративный алгоритм основан на последовательном вычислении сумм (7.96) и интегралов (7.99) с помощью алгоритма быстрого преобразования Фурье с наложением соответствующих ограничений. Так, на k-й итерации рассчитанные коэффициенты $C_{nm}^{(k)}$ заменяются на $\bar{C}_{n,m}^{(k)}$ следующим образом:

$$\bar{C}_{n,m}^{(k)} = D_{n,m} C_{n,m}^{(k)} \left| C_{n,m}^{(k)} \right|^{-1}, \tag{7.100}$$

где $D_{n,m}$ — неотрицательные числа, характеризующие распределение энергии между модами. Ограничения типа (7.100) характерны для итеративного алгоритма Герчберга–Секстона [34]. Однако для улучшения сходимости можно применять модификации замены (7.100), вводя некоторый коэффициент адаптивности $0 < \alpha \leqslant 1$, аналогичный рассмотренному (2.40):

$$\bar{C}_{n,m}^{(k)} = \begin{cases} \left[D_{n,m} - \alpha \left(\left| C_{n,m}^{(k)} \right| - D_{n,m} \right) \right] C_{n,m}^{(k)} \left| C_{n,m}^{(k)} \right|^{-1}, & (n,m) \in \Omega, \\ 0, & (n,m) \notin \Omega, \end{cases} \tag{7.101}$$

или

$$\bar{C}_{n,m}^{(k)} = \begin{cases} \left[D_{n,m} - \alpha \left(\left| C_{n,m}^{(k)} \right| - D_{n,m} \right) \right] C_{n,m}^{(k)} \left| C_{n,m}^{(k)} \right|^{-1}, & (n,m) \in \Omega, \\ C_{n,m}^{(k)}, & (n,m) \notin \Omega, \end{cases} \tag{7.102}$$

где Ω — множество пар индексов, для которых числа $D_{n,m}$ отличны от нуля.

Таким образом, можно предложить следующий алгоритм для нахождения фазы оптического элемента $S(r,\varphi)$:

1. Начальная фаза выбирается как случайная величина $S_0(r,\varphi)$.

2. Пусть на k-м шаге по формуле (7.96) получается значение k-й оценки фазы $S_k(r,\varphi)$. Используя $S_k(r,\varphi)$, из уравнения (7.99) рассчитываются коэффициенты $C_{n,m}^{(k)}$.

3. Затем они заменяются коэффициентами $\bar{C}_{n,m}^{(k)}$, исходя из сформулированных выше правил (7.100), (7.101) или (7.102).

4. Коэффициенты $\bar{C}_{n,m}^{(k)}$ подставляются в (7.96). В результате получается функция $f_k(r,\varphi)$, аргумент которой служит последующей оценкой фазы:

$$f_k(r,\varphi) = \sum_{n=0}^{N} \sum_{|m| \leqslant n} \bar{C}_{n,m}^{(k)} \Psi_{n,m}(r,\varphi), \quad S_{k+1}(r,\varphi) = \arg\left\{ f_k(r,\varphi) \right\}. \tag{7.103}$$

Переходим к 2. И так далее.

Сходимость алгоритма контролируется по средним отклонениям:

$$\delta_A = \left\{ \int_0^{2\pi} \int_0^{\infty} \left[A_0(r,\theta) - |f_k(r,\varphi)| \right]^2 r \, \mathrm{d}r \, \mathrm{d}\varphi \right\}^{1/2} \left\{ \int_0^{2\pi} \int_0^{\infty} A_0^2(r,\varphi) \, r \, \mathrm{d}r \, \mathrm{d}\varphi \right\}^{-1/2}, \tag{7.104}$$

$$\delta_C = \left\{ \sum_{n=0}^{N} \sum_{|m| \leqslant n} \left[D_{n,m} - \left| C_{n,m}^{(k)} \right| \right]^2 \right\}^{1/2} \left\{ \sum_{n=0}^{N} \sum_{|m| \leqslant n} D_{n,m}^2 \right\}^{-1/2}. \tag{7.105}$$

Можно показать, как это сделано в [53], что ошибки (7.104) и (7.105) не возрастают:

$$\delta_A^{(k+1)} \leqslant \delta_A^{(k)}, \quad \delta_C^{(k+1)} \leqslant \delta_C^{(k)}. \tag{7.106}$$

Функции ГЛ являются собственными функциями преобразования Фурье,

$$\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}\Psi_{n,m}(r,\varphi)\exp\left[-ir\rho\cos\left(\varphi-\theta\right)\right]r\,\mathrm{d}r\,\mathrm{d}\varphi=(-1)^n\,(-i)^m\,\Psi_{n,m}(\rho,\theta), \qquad (7.107)$$

что позволяет использовать их для эффективного ввода излучения в световые волокна [50]. Так, если сформирован световой пучок, сечение которого имеет вид заданного изображения, и который является суперпозицией мод ГЛ,

$$g(r,\varphi)=\sum_{n=0}^{N}\sum_{|m|\leqslant n}C^g_{n,m}\Psi_{n,m}(r,\varphi), \qquad (7.108)$$

то в зоне дифракции Фраунгофера или в фокальной плоскости линзы получится следующее изображение:

$$\mathcal{F}\left\{g(r,\varphi)\right\}=G(\rho,\theta)=\sum_{n=0}^{N}\sum_{|m|\leqslant n}\mu_{n,m}C^g_{n,m}\Psi_{n,m}(\rho,\theta), \qquad (7.109)$$

где $\mathcal{F}$ — обозначение фурье-преобразования, (ρ,θ) — полярные координаты в фурье-плоскости.

Следовательно, если в суперпозиции участвовали только моды с одинаковыми значениями собственных чисел

$$\mu_{n,m}=(-1)^n(-i)^{|m|}=\mu_0, \qquad (7.110)$$

то вид изображения в сечении пучка не изменится.

Из формулы (7.110) видно, что собственные числа функций ГЛ принимают всего четыре значения:

$$\begin{aligned}
&\mu_{n,m}=1:\ (n=2k,m=4l),\ (n=2k+1,m=4l+2),\\
&\mu_{n,m}=-1:\ (n=2k,m=4l+2),\ (n=2k+1,m=4l),\\
&\mu_{n,m}=i:\ (n=2k,m=4l+3),\ (n=2k+1,m=4l+1),\\
&\mu_{n,m}=-i:\ (n=2k,m=4l+1),\ (n=2k+1,m=4l+3),\quad k,l=0,1,2,\ldots
\end{aligned} \qquad (7.111)$$

Более наглядно распределение одинаковых значений собственных чисел в зависимости от номера функции ГЛ приведено в таблице 7.1.

Таким образом, вполне реально подобрать такую суперпозицию мод ГЛ с одинаковыми собственными значениями, чтобы сформировать некоторое изображение, не меняющее своей структуры при прохождении фурье-каскада. Заметим, что функции ГЛ с взаимно противоположной «закруткой» $\exp\left(\pm i|m|\varphi\right)$ имеют одинаковые собственные значения. Совокупность мод ГЛ с одинаковыми собственными значениями можно считать модовой группой, так как она также представляет собой собственную функцию преобразования Фурье.

Понятно, что рассматриваемый метод работает тем лучше, чем больше задействовано членов в сумме (7.96), а следовательно мод в пучке. Чем больше степеней свободы (фазы коэффициентов $C_{n,m}$), тем точнее аппроксимируется желаемая функция $f(r,\varphi)$, в частности, амплитуда $A_0(x,y)$ (7.96).

Данный метод работоспособен и при формировании одномодового пучка. В этом случае за одну итерацию (с дальнейшей стагнацией) рассчитываются ДОЭ с фазой, пропорциональной знаковой функции от многочлена Лагерра

$$\tau(r,\varphi) = \arg\{\Psi_{n,m}(r,\varphi)\} = \frac{\pi}{2}\left\{1 - \operatorname{sgn}\left[L_n^{|m|}\left(r^2/a^2\right)\right]\right\} + im\varphi. \tag{7.112}$$

Покажем, что такая аппроксимация функции ГЛ является эффективной.

Выражение (7.112) для фазовой функции эквивалентно следующему выражению для амплитуды на участке $[-R, R]$, где R — размер апертуры, $t = (r/a)^2$:

$$\tau_{n,m}(t,\varphi) = \operatorname{sgn}\left[L_n^{|m|}(t)\right]\exp(im\varphi). \tag{7.113}$$

Разложим (7.113) по ортогональным многочленам Лагерра с «закруткой»,

$$\tau_{n,m}(t,\varphi) = \sum_{p=0}^{\infty}\sum_{|q|\leqslant p} C_{p,q}^{(n,m)} L_p^{|q|}(t)\exp(iq\varphi), \tag{7.114}$$

и найдем коэффициенты разложения

$$C_{p,q}^{(n,m)} = \sqrt{\frac{p!}{(p+|q|)!}}\int_0^{2\pi}\int_0^{\infty} \operatorname{sgn}\left[L_n^{|m|}(t)\right] e^{im\theta} L_p^{|q|}(t)\, e^{iq\theta} t^{|q|} e^{-t}\,\mathrm{d}t\,\mathrm{d}\varphi =$$

$$= \sqrt{\frac{p!}{(p+|q|)!}}\int_0^{\infty} \operatorname{sgn}\left[L_n^{|m|}(t)\right] L_p^{|q|}(t)\cdot t^{|q|} e^{-t}\,\mathrm{d}t \underbrace{\int_0^{2\pi} e^{im\theta} e^{iq\theta}\,\mathrm{d}\varphi}_{\delta_{m,q}} =$$

$$= \sqrt{\frac{p!}{(p+|m|)!}}\int_0^{\infty} \operatorname{sgn}\left[L_n^{|m|}(t)\right] L_p^{|m|}(t)\cdot t^{|m|} e^{-t}\,\mathrm{d}t = C_{p,m}^{(n,m)}. \tag{7.115}$$

Таким образом, ненулевыми являются только коэффициенты с таким же вторым индексом, что и раскладываемая функция $\tau_{n,m}(t,\varphi)$.

Таблица 7.1. Собственные значения для собственных функций ГЛ

n	m									
	0	1	2	3	4	5	6	7	8	9
0	1									
1	−1	i								
2	1	$-i$	−1							
3	−1	i	1	$-i$						
4	1	$-i$	−1	i	1					
5	−1	i	1	$-i$	−1	i				
6	1	$-i$	−1	i	1	$-i$	−1			
7	−1	i	1	$-i$	−1	i	1	$-i$		
8	1	$-i$	−1	i	1	$-i$	−1	i	1	
9	−1	i	1	$-i$	−1	i	1	$-i$	−1	i

Обозначив через $t_{k,n}$ нули многочлена $L_n^m(t_{k,n}) = 0$ и считая $t_{0,n} = 0$, представим выражение (7.115) в виде суммы интегралов

$$C_{p,m}^{(n,m)} = \sqrt{\frac{p!}{(p+|m|)!}} \sum_{k=0}^{n-1} \left[(-1)^k \int_{t_{k,n}}^{t_{k+1,n}} t^{|m|} e^{-t} L_p^{|m|}(t)\, \mathrm{d}t \right]. \tag{7.116}$$

С учетом справочного интеграла [54]:

$$\int x^\alpha e^{-x} L_n^\alpha(x)\, \mathrm{d}x = \frac{1}{n} x^{\alpha+1} e^{-x} L_{n-1}^{\alpha+1}(x), \tag{7.117}$$

вместо (7.116) получим следующее выражение для коэффициентов разложения (7.113):

$$C_{p,m}^{(n,m)} = \frac{2}{p} \sqrt{\frac{p!}{(p+|m|)!}} \sum_{k=1}^{n} \left[(-1)^{k-1} (t_{k,n})^{|m|+1} e^{-t_{k,n}} L_{p-1}^{|m|+1}(t_{k,n}) \right]. \tag{7.118}$$

При этом коэффициенты $C_{p,q}^{(n,m)}$ с $q \neq m$ равны нулю.

Пример 7.7. Пусть требуется сформировать моду ГЛ $\Psi_{1,0}(t, \varphi)$. Для этого разместим в плоскости $z = 0$ фазовый элемент с пропусканием $\tau_{1,0}(t) = \operatorname{sgn}\left[L_1^0(t)\right]$. Тогда, согласно уравнению (7.118), в пространстве за фильтром сформируется суперпозиция световых мод

$$\tau_{1,0}(t) = \sum_{p=1}^{\infty} C_{p,0}^{(1,0)} L_p^0(t) = \\ = 0{,}736 L_1^0(t) - 0{,}184 L_2^0(t) + 0{,}020 L_3^0(t) - 0{,}001 L_4^0(t) + \ldots . \tag{7.119}$$

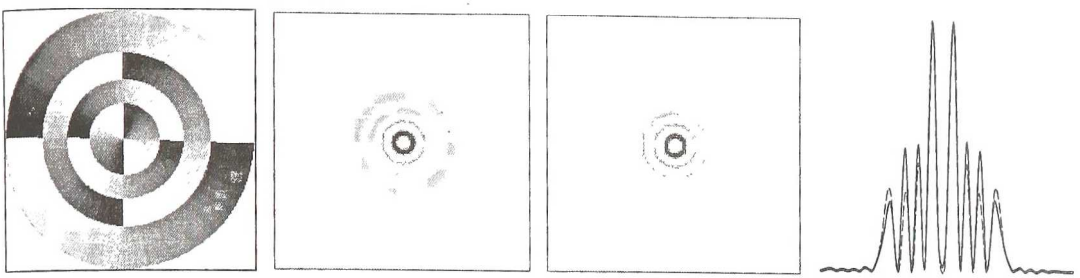
Освещая такой фазовый элемент коллимированным гауссовым пучком, получаем, согласно уравнению (7.107), в частотной плоскости световое поле с амплитудой, пропорциональной выражению ($a = 1$ мм)

$$\int_0^\infty e^{-r^2/2} \operatorname{sgn}\left[L_1^0(r^2)\right] J_0(r\rho) \cdot r\, \mathrm{d}r = \sum_{p=1}^{\infty} (-1)^p C_{p,0}^{(1,0)} e^{-\rho^2/2} L_p^0(\rho^2) = \\ = -e^{-\rho^2/2} \left[0{,}736 L_1^0(\rho^2) + 0{,}184 L_2^0(\rho^2) + 0{,}020 L_3^0(\rho^2) + 0{,}001 L_4^0(\rho^2) + \ldots\right], \tag{7.120}$$

где $J_0(x)$ — функция Бесселя.

Отношение энергий световых пучков, описываемых вторым и первым слагаемыми в уравнении (7.120), равно

$$\frac{(0,184)^2 \int_0^\infty e^{-\rho^2} \left[L_2^0(\rho^2)\right]^2 \rho\, \mathrm{d}\rho}{(0,736)^2 \int_0^\infty e^{-\rho^2} \left[L_1^0(\rho^2)\right]^2 \rho\, \mathrm{d}\rho} = \frac{(0,184)^2 (2!)^2}{(0,736)^2 (1!)^2} \approx 0{,}25.$$



а *б* *в* *г*

Рис. 7.14. Одномодовый пучок: фаза ДОЭ (*а*); распределения интенсивности (негатив) в Фурье-спектре для рассчитанной (*б*) и эталонной (*в*) мод ГЛ $\Psi_{3,2}(r,\varphi)$; интенсивность в радиальном сечении (*г*) (сплошная линия — расчет, пунктирная линия — эталон)

Таким образом, в первом слагаемом суммы (7.120), описывающем моду ГЛ $\Psi_{1,0}(t,\varphi)$, содержится более 70% всей энергии светового пучка.

Пример 7.8. Еще более эффективным способом формирования моды ГЛ $\Psi_{n,m}(r,\varphi)$ является освещение не гауссовым, а плоским световым пучком фазового ДОЭ с пропусканием ($a = 1$ мм):

$$\tilde{\tau}_{n,m}(r,\varphi) = \operatorname{circ}\left(\frac{r}{R}\right) \operatorname{sgn}\left[L_n^{|m|}\left(r^2\right)\right] \exp(im\varphi), \tag{7.121}$$

где $\operatorname{circ}\left(\dfrac{t}{t_0}\right) = \begin{cases} 1, & t \leqslant t_0, \\ 0, & t > t_0. \end{cases}$

В этом случае значение R следует выбирать так, чтобы выражение, определяющее эффективность возбуждения отдельной моды,

$$\eta = \frac{\left|\int\limits_0^{2\pi}\int\limits_0^{R} \tilde{\tau}_{n,m}(r,\varphi)\cdot\Psi^*_{n,m}(r,\varphi) r\,\mathrm{d}r\,\mathrm{d}\varphi\right|^2}{\int\limits_0^{2\pi}\int\limits_0^{R} |\tilde{\tau}_{n,m}(r,\varphi)|^2 r\,\mathrm{d}r\,\mathrm{d}\varphi \cdot \int\limits_0^{2\pi}\int\limits_0^{\infty} |\Psi_{n,m}(r,\varphi)|^2 r\,\mathrm{d}r\,\mathrm{d}\varphi}, \tag{7.122}$$

было максимальным.

Рассмотрим пример с модой ГЛ $\Psi_{1,0}(r,\varphi)$. В этом случае формула (7.122) записывается следующим образом:

$$\eta = \frac{\left|2\left[\int\limits_0^1 e^{-r^2/2}\left(1-r^2\right) r\,\mathrm{d}r - \int\limits_1^R e^{-r^2/2}\left(1-r^2\right) r\,\mathrm{d}r\right]\right|^2}{R^2} = \frac{4\left[1{,}426 - e^{-R^2/2}\left(R^2+1\right)\right]^2}{R^2} \tag{7.123}$$

и достигает максимального значения $\eta = 0{,}786$ при $R = 2{,}5$ мм. Значит, световое поле (7.121) формирует моду ГЛ $\Psi_{1,0}(r,\varphi)$ с эффективностью около 79%.

Пример 7.9. При численном моделировании использовались следующие параметры: 128 пикселов по радиусу r и 128 пикселов по угловой составляющей φ, диапазон изменения аргументов $r \in [0, 7\,\text{мм}]$, $\varphi \in [0, 2\pi]$, длина волны $\lambda = 0{,}63\,\text{мкм}$, фокусное расстояние $f = 100\,\text{мм}$, радиус гауссового пучка в перетяжке $a = 1\,\text{мм}$. В формуле (7.96) рассматривались члены ряда с номерами $n, m \leqslant N = 7$. Действие рассчитанных ДОЭ [55] моделировалось с помощью численного преобразования Фурье (7.107).

На рис. 7.14 представлены результаты формирования моды ГЛ с номером $(3,2)$ $\Psi_{3,2}(r,\varphi)$. Фазовая функция (рис. 7.14*а*) рассчитывалась по формуле (7.121). На рис. 7.14*б* показан сформированный таким ДОЭ пучок и для сравнения на рис. 7.14*в* приведена эталонная мода ГЛ с номером $(3,2)$. Их радиальные сечения показаны на рис. 7.14*г* (сплошная линия — рассчитанная мода, пунктирная линия — эталонная мода).

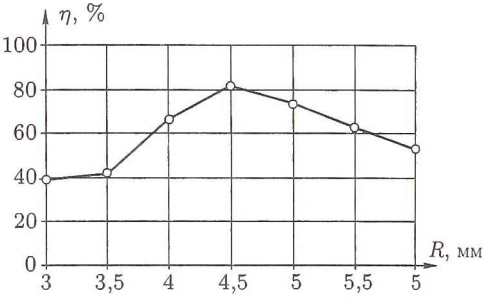


Рис. 7.15. Зависимость эффективности возбуждения моды η от радиуса апертуры R для ДОЭ, формирующего моду ГЛ $\Psi_{3,2}(r,\varphi)$

Оптимальный радиус апертуры R определялся из условия максимума эффективности η (7.122) сформированного пучка. График зависимости η от R для моды ГЛ $\Psi_{3,2}(r,\varphi)$ представлен на рис. 7.15. Из графика видно, что оптимальный размер апертуры ДОЭ для формирования моды ГЛ $\Psi_{3,2}(r,\varphi)$ равен $R = 4{,}5\,\text{мм}$. При этом эффективность возбуждения этой моды достигает значения 81%.

Проведя аналогичные исследования, можно подобрать оптимальный радиус ДОЭ для формирования любой моды. С помощью численного моделирования было показано, что фазовые ДОЭ, рассчитанные по формуле (7.121), позволяют формировать одномодовые пучки с эффективностью 77–81% в зависимости от номера моды. В таблице 7.2 приведены значения R и η для нескольких мод. Можно предположить, что для суперпозиции мод, оптимальным радиусом апертуры ДОЭ является средний из оптимальных радиусов, входящих в суперпозицию мод.

Пример 7.10. На рис. 7.16 приведены примеры фазовых ДОЭ (рис. 7.16*а*), формирующих многомодовые пучки ГЛ в плоскости пространственного спектра (рис. 7.16*в*). Верхняя строка рисунка относится к группе из 5 мод: $(1,1)$, $(3,1)$, $(4,-3)$, $(5,1)$, $(5,-5)$, имеющих собственные числа, равные i. Квадраты модулей коэффициентов разложения пучков представлены на рис. 7.16*б* в полутонах. Нижняя строка рисунка относится к группе из 6 мод: $(2,0)$, $(3,-2)$, $(3,2)$, $(4,-4)$, $(4,0)$, $(4,4)$, имеющих собственные числа, равные 1. Оба пучка являются собственными функциями преобразования Фурье. Для сравнения на рис. 7.16*г* показаны эталонные распределения интенсивности указанных композиций мод.

Таблица 7.2. Значения оптимального радиуса апертуры ДОЭ R и эффективности η ($a = 1\,\text{мм}$)

Номер моды ГЛ	$(1,0)$	$(2,1)$	$(3,2)$	$(4,2)$	$(5,1)$
Оптимальный радиус R, мм	2,5	4	4,5	4,7	5
Эффективность η, %	79	77	81	78	77

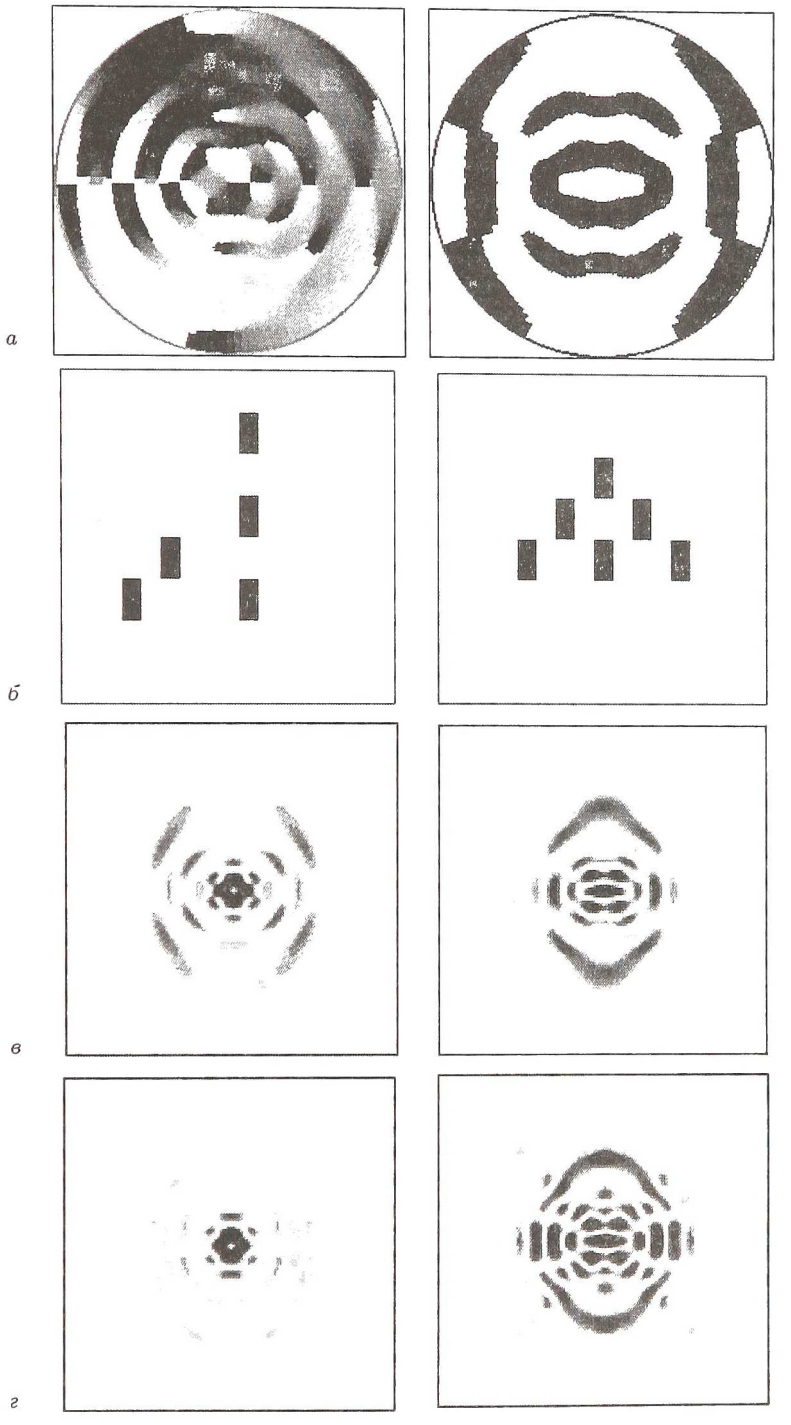


Рис. 7.16. Многомодовые пучки: фазы ДОЭ (*а*); квадраты модулей коэффициентов в разложении (7.96) (*б*); распределения интенсивности в фокальной плоскости для рассчитанной (*в*) и эталонной (*г*) композиций мод ГЛ (левый столбец для 5-ти модового пучка, правый столбец для 6-ти модового пучка)

7.6. Вращение световых многомодовых пучков Гаусса–Лагерра в свободном пространстве и волокне

Моды в свободном пространстве. В данном разделе рассмотрены условия, при которых наблюдается эффект вращения многомодового пучка ГЛ вокруг оси распространения и получено выражение для полного числа оборотов.

Известно [56], что уравнению Гельмгольца удовлетворяют световые поля, являющиеся суперпозицией мод ГЛ. Комплексную амплитуду таких полей в свободном пространстве можно представить в цилиндрических координатах (r, φ, z) в следующем виде:

$$U(r,\varphi,z) = \exp\left[ikz + ikr^2/2R - r^2/\sigma^2\right] \times \\ \times \sum_{n=0}^{\infty} \sum_{|m|\leqslant n} C_{n,m} \left(\frac{\sqrt{2}r}{\sigma}\right)^{|m|} L_n^{|m|}\left(2r^2/\sigma^2\right) \exp\left[-i\beta_{n,m}(z) + im\varphi\right], \quad (7.124)$$

где

$$\beta_{n,m}(z) = (2n + |m| + 1)\operatorname{arctg}\left(z/z_0\right), \quad (7.125)$$

$R = z\left(1 + z_0^2/z^2\right)$ — радиус кривизны параболического фронта светового поля, $\sigma^2 = \sigma_0^2\left(1 + z^2/z_0^2\right)$ — эффективный радиус пучка, $2z_0 = 2\pi\sigma_0^2/\lambda$ — конфокальный параметр, σ_0 — радиус перетяжки, $C_{n,m}$ — постоянные коэффициенты.

Получим условие для наблюдения эффекта вращения поперечного сечения многомодового пучка. На основании (7.124) запишем выражение для интенсивности $I(r,\varphi,z) = |U(r,\varphi,z)|^2$

$$I(r,\varphi,z) = \exp\left[-\frac{2r^2}{\sigma^2}\right]\left\{\sum_{n=0}^{\infty}\sum_{|m|\leqslant n} |C_{n,m}|^2 \left(\frac{\sqrt{2}r}{\sigma}\right)^{2|m|} \left|L_n^{|m|}\left(\frac{2r^2}{\sigma^2}\right)\right|^2 + \right.$$

$$+ 2\sum_{n\neq n'}\sum_{m\neq m'} |C_{n,m}C_{n',m'}| \left(\frac{\sqrt{2}r}{\sigma}\right)^{|m|+|m'|} \left|L_n^{|m|}\left(\frac{2r^2}{\sigma^2}\right) L_{n'}^{|m'|}\left(\frac{2r^2}{\sigma^2}\right)\right| \times$$

$$\left. \times \cos\Phi_{n',m'}^{n,m}(r,\varphi)\right\}, \quad (7.126)$$

где

$$\Phi_{n',m'}^{n,m}(r,\varphi) = \arg C_{n,m} - \arg C_{n',m'} + \\ + \left[2(n-n') + (|m| - |m'|)\right]\operatorname{arctg}(z/z_0) + (m-m')\,\varphi. \quad (7.127)$$

Из уравнения (7.127) видно, что вид функции $I(r, \varphi+\Delta\varphi, z+\Delta z)$ не изменится, если в соотношении

$$\varphi + \Delta\varphi = B\operatorname{arctg}[(z + \Delta z)/z_0] \quad (7.128)$$

величина B будет постоянной:

$$B = \frac{2(n-n') + (|m| - |m'|)}{m - m'} = \text{const}, \quad n, n', m, m' = 0, 1, 2, \ldots \quad (7.129)$$

Строго говоря, если в (7.126) и (7.127) сделать замену переменных

$$r' = r/\sigma, \quad \varphi' = \varphi - B\,\mathrm{arctg}(z/z_0), \quad z' = z,$$

то интенсивность $I(r', \varphi')$ в такой растягивающейся и одновременно вращающейся системе координат не будет зависеть от z'. Условие вращения (7.129) многомодового пучка ГЛ как единого целого в свободном пространстве аналогично условию (7.42), полученному в [17].

Из (7.128) можно найти расстояния z_p, на которых пучок сделает p оборотов

$$z_p = z_0 \,\mathrm{tg}\,(2\pi p/B), \quad p = 1, 2, \ldots, N, \tag{7.130}$$

где $N = B/4$ — максимальное число оборотов, которые пучок способен сделать от $z = 0$ до $z = \infty$ ($z \gg z_0$). Скорость вращения v поперечного сечения многомодового пучка зависит от расстояния z следующим образом:

$$v = \frac{\mathrm{d}\varphi}{\mathrm{d}z} = B\left[1 + \left(\frac{z}{z_0}\right)^2\right]^{-1}. \tag{7.131}$$

Из вышеизложенного следует, что подбором отличных от нуля слагаемых в уравнении (7.124), с номерами, удовлетворяющими условию (7.129), можно по формулам (7.99)–(7.103) рассчитывать фазовые ДОЭ, которые при освещении их лазерным светом формируют с высокой эффективностью многомодовые нерадиально-симметричные пучки ГЛ, поперечное сечение которых испытывает вращение вокруг оси распространения.

Пример 7.11. Параметры численного моделирования: число пикселов 256×256, радиус перетяжки $\sigma_0 = 0{,}1$ мм, длина волны $\lambda = 0{,}63$ мкм, радиус ДОЭ $R_0 = 0{,}5$ мм, конфокальный параметр $z_0 = 49{,}86$ мм. На рисунке 7.17 представлены численные примеры. В первом столбце показана фаза ДОЭ (*а1*), рассчитанная по формулам (7.99)–(7.103) с отличными от нуля двумя слагаемыми суммы (7.124) с коэффициентами $C_{1,-1}$ и $C_{11,2}$. Условие (7.129) для двух слагаемых очевидно выполняется: $B = 7$. Максимальное число оборотов равно $N = 1{,}75$. На расстоянии $z = 62{,}53$ мм, как следует из уравнения (7.130), пучок сделает полный оборот, а на расстоянии $z = 217{,}76$ мм — еще пол-оборота. На рисунке показаны полутоновые поперечные нормированные распределения интенсивности светового пучка (негатив), сформированного фазовым ДОЭ (*а1*), и вычисленные с помощью преобразования Френеля на плоскостях: $z = 62{,}53$ мм (*б1*), $z = 101{,}34$ мм (*в1*), $z = 140{,}15$ мм (*г1*), $z = 178{,}95$ мм (*д1*), $z = 217{,}76$ мм (*е1*). Вращение картины происходит против часовой стрелки.

Во втором столбце показаны аналогичные результаты для вращающегося пучка, состоящего из трех мод с коэффициентами: $C_{1,-1}$, $C_{5,0}$ и $C_{11,2}$. Условие (7.129) выполняется: $B = 7, 7, 7$. Фазовая функция такого элемента (*а2*) показана в полутонах (черный цвет — 2π, белый цвет — 0). Поперечные распределения интенсивности для такого пучка (*б2–е2*) показаны на тех же расстояниях, что и в первом столбце (негатив).

В третьем столбце приведены фаза (*а3*) и сечения интенсивности (*б3–е3*) для трехмодового пучка с коэффициентами: $C_{2,-2}$, $C_{5,0}$ и $C_{15,2}$. Условие (7.129) при этом не выполняется: $B = 2, 6, 11$, и разные элементы пучка вращаются с различной скоростью, что приводит к искажению картины.

Моды в градиентном волокне. В многомодовом световом волокне круглого сечения и с квадратичной зависимостью показателя преломления

$$n^2(r) = n_0^2\left(1 - 2\Delta r^2/r_0^2\right), \tag{7.132}$$



Рис. 7.17. Фазы ДОЭ, формирующих многомодовые пучки ГЛ (*а*) и нормированные поперечные распределения интенсивности на разных расстояниях от ДОЭ (*б–е*)

где n_0 — максимальное значение показателя преломления на оси волокна, r_0 — радиус волокна, Δ — параметр дисперсии показателя преломления, распространяется световое поле, являющееся суперпозицией мод ГЛ [57]

$$U(r,\varphi,z)=\sum_{n=0}^{\infty}\sum_{|m|\leqslant n}C_{n,m}\left(r\sqrt{2}\big/\sigma\right)^{|m|}L_n^{|m|}\left(2r^2/\sigma^2\right)\times$$
$$\times\exp\left[i\alpha_{n,m}z-r^2/\sigma^2+im\varphi\right], \quad (7.133)$$

где $\sigma=\left(\lambda r_0/\pi n_0\right)^{1/2}(2\Delta)^{-1/4}$, $\alpha_{n,m}=\left[k^2n_0^2-4\sigma^{-2}(2n+|m|+1)\right]^{1/2}$ σ — эффективный радиус первой моды, $\alpha_{n,m}$ — коэффициент, пропорциональный фазовой скорости соответствующей моды.

Из-за дисперсии фазовых скоростей мод ГЛ, сформированное на входе волокна изображение $U(r,\varphi,z=0)$ будет деградировать по мере распространения. Чтобы частично компенсировать дисперсию фазовых скоростей мод нужно из суммы (7.133) выбрать такие слагаемые, для которых фазовые скорости некоторым образом согласованы. Например, можно заставить световой многомодовый пучок, сформированный фазовым ДОЭ, вращаться вокруг оси по спирали. Условия вращения поперечного сечения пучка (7.133) аналогичны условиям (7.77) и (7.78) для многомодовых пучков Бесселя и имеют соответственно вид:

$$(\alpha_{n,m}-\alpha_{n',m'})/(m-m')=B, \quad m\neq m', \quad n\neq n', \quad (7.134)$$

$$\alpha_{n,m}/m=B, \quad (7.135)$$

где B — постоянная величина. В уравнении (7.134) величины $\alpha_{n,m}$ можно приближенно записать:

$$\alpha_{n,m}\cong kn_0-\alpha_0(2n+|m|+1), \quad (7.136)$$

так как $kn_0\gg\alpha_0$, $\alpha_0=2(kn_0\sigma^2)^{-1}$. Например, при характерных параметрах волокна: $r_0=10$ мкм, $\Delta=0{,}01$, $n_0=1{,}5$, и для видимого света: $k=104\,\text{мм}^{-1}$, получим следующие значения: $kn_0=1{,}5\times10^4\,\text{мм}^{-1}$ и $\alpha_0=15\,\text{мм}^{-1}$.

С учетом уравнения (7.136) условие (7.134) будет иметь более простой вид:

$$\alpha_0\left[\frac{2(n-n')+(|m|-|m'|)}{m-m'}\right]=B, \quad n,n',m,m'=0,1,2,\ldots \quad (7.137)$$

В данном случае условие (7.137) похоже на (7.77) для бесселевых мод и на (7.129) для мод ГЛ в свободном пространстве. При небольших значениях номеров мод ($n<$ <10), приближенное равенство в (7.136) соблюдается при выбранных параметрах волокна с точностью около 2%. Период z_0, на котором будет происходить полный оборот, рассчитывается по формуле

$$z_0=2\pi/|B| \quad (7.138)$$

и для выбранных параметров равен десяткам и сотням микрон. Аналогично модам Бесселя итеративно рассчитывались фазовые ДОЭ, формирующие световые поля, состоящие эффективно (на 80–90%) из двух-четырех мод ГЛ и распространяющиеся в многомодовом волокне с квадратичной зависимостью показателя преломления (7.132).

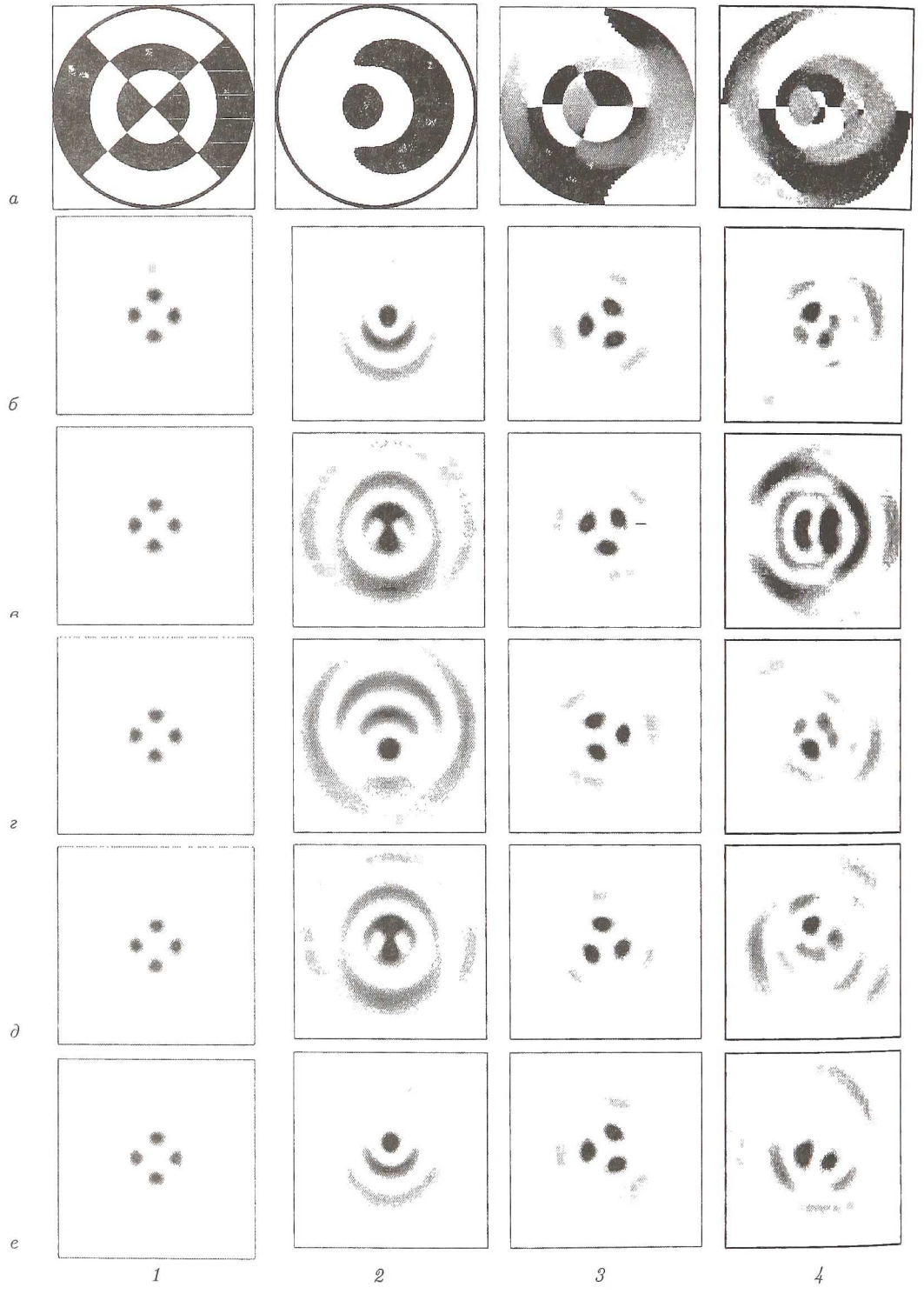


Рис. 7.18. Итеративно рассчитанные фазы ДОЭ (*а1–а4*) и поперечные сечения многомодовых пучков ГЛ (*б1–е4*), генерируемых такими ДОЭ

Пример 7.12. На рис. 7.18 показаны некоторые примеры. Параметры расчета: $r_0 = 10$ мкм, $\Delta = 0{,}01$, $n_0 = 1{,}5$, $\lambda = 0{,}63$ мкм, $R = 0{,}01$ мм, $\alpha_0 = 15$ мм$^{-1}$.

На рис. 7.18 в левом столбце приведен пример моды ГЛ, распространяющейся в волокне почти без дифракции. В уравнении (7.133) отличными от нуля выбирались только два слагаемых с номерами (n, m): $(2, -2)$ и $(2, 2)$. При этом уравнения (7.136)–(7.138) приводят к бесконечному периоду вращения $z_0 = \infty$, $B = 0$. На рис. 7.18 показаны бинарная фаза такого ДОЭ (*а1*) и поперечные сечения сформированного пучка (негатив), рассчитанного на разных расстояниях z: 1 мм (*б1*), 2 мм (*в1*), 3 мм (*г1*), 4 мм (*д1*), 5 мм (*е1*). Видно, что при распространении такого пучка происходит незначительная модуляция интенсивности, обусловленная периодически изменяющимся влиянием высших мод ГЛ. В расчете учитывались всего 25 мод с соседними номерами.

Пример периодически изменяющегося многомодового пучка ГЛ показан во втором слева столбце на рис. 7.18. В этом случае в уравнении (7.133) отличными от нуля считались слагаемые с номерами (n, m): $(1, -1)$, $(3, 0)$, $(1, 1)$. Условия вращения (7.134) и (7.135) не соблюдаются, однако полученные значения B оказались кратными, что позволило выделить общий период $z_0 = 0{,}2087$ мм. На рис. 7.18 во втором слева столбце показаны бинарная фаза ДОЭ (*а2*) и поперечные сечения светового пучка (негатив) на разных расстояниях z: 2,087 мм (*б2*), 2,139 мм (*в2*), 2,191 мм (*г2*), 2,243 мм (*д2*), 2,295 мм (*е2*). Видно, что происходит периодическая смена картин дифракции при распространении пучка в волокне.

В третьем слева столбце на рис. 7.18 приведен пример вращающегося многомодового пучка ГЛ. В сумме (7.133) сохранялись два слагаемых с номерами (n, m): $(1, -1)$, $(2, 2)$. Условие вращения (7.134) выполняется, период вращения $z_0 = 0{,}626$ мм. На рис. 7.18 в третьем слева столбце показаны полутоновая фаза такого ДОЭ (*а3*) и поперечные сечения пучка (негатив) на разных расстояниях z: 1 мм (*б3*), 1,156 мм (*в3*), 1,313 мм (*г3*), 1,469 мм (*д3*), 1,626 мм (*е3*). Заметим, что условию вращения (7.134) удовлетворяют, например, трехмодовые пучки ГЛ со следующими номерами (n, m): $(1, -1)$, $(5, 0)$, $(11, 2)$, а также $(2, -2)$, $(8, 0)$, $(14, 3)$.

На рис. 7.18 в первом справа столбце приведен пример четырехмодового пучка ГЛ, для которого условие вращения (7.134) не выполняется, и поэтому происходит вращение с искажением вида поперечной картины пучка. На рис. 7.18 в первом справа столбце показана полутоновая фаза ДОЭ (*а4*), рассчитанная итеративно с учетом четырех отличных от нуля слагаемых в уравнении (7.133) с номерами (n, m): $(2, -2)$, $(1, -1)$, $(2, 1)$, $(3, 2)$. Поперечные сечения пучка, сформированного таким ДОЭ, показаны в правом столбце на разных расстояниях z: 1 мм (*б4*), 1,094 мм (*в4*), 1,187 мм (*г4*), 1,281 мм (*д4*), 1,375 мм (*е4*). Искажение картины поперечного сечения пучка происходит из-за того, что отдельные моды вращаются с разными угловыми скоростями и распространяются с разными фазовыми скоростями, и скорости эти не согласованы друг с другом, как это было в случаях, относящихся к рис. 7.18.

7.7. Формирование вращающихся пучков Гаусса–Лагерра с помощью фазовой бинарной дифракционной оптики

Экспериментальные результаты по вращению лазерного пучка, состоящего из двух мод ГЛ, представлены, например, в [58]. Пучок был сформирован с помощью амплитудной голограммы, фазовая информация в которой была закодирована по методу Ли, поэтому энергетическая эффективность такого пучка была небольшая. В [59] приведены численные результаты по вращению пучков ГЛ.

В данном разделе приведены экспериментальные результаты (и их сравнение с численными) по формированию в свободном пространстве двух одинаковых лазерных пучков, вращающихся в противоположные стороны и состоящих из двух мод ГЛ. Эти вращающиеся пучки формируются с помощью бинарного фазового ДОЭ, и энергетическая эффективность каждого из них около 40%. Показано хорошее соответствие теории и эксперимента.

Пример 7.13. На рис. 7.19*а* показана полутоновая фаза ДОЭ (255 уровней градаций в диапазоне $[-\pi, \pi]$), рассчитанная за две итерации методом (7.99)–(7.103). При этом в сумме (7.124) отличны от нуля только два слагаемых с номерами (n, m): $(1, -1)$ и $(15, 1)$. Условие вращения (7.129) для двух слагаемых очевидно выполняется: $B = 14$. На расстоянии $z_1 = 78$ мм пучок делает первый оборот. Всего пучок сделает $N = B/4 = 3{,}5$ оборотов.

На рис. 7.19*б–е* показаны полутоновые распределения поперечной интенсивности (негатив), сформированные ДОЭ (рис. 7.19*а*) и вычисленные с помощью преобразования Френеля на расстояниях: $z = 215$ мм (7.19*б*), $z = 245$ мм (7.19*в*), $z = 310$ мм (7.19*г*), $z = 355$ мм (7.19*д*) и $z = 440$ мм (7.19*е*). Радиус ДОЭ $R = 1$ мм, число пикселов вдоль радиуса — 256. Перетяжка гауссова пучка при $z = 0$ равна $\sigma_0 = 0{,}18$ мм.

Однако, изготовление полутонового фазового ДОЭ с помощью литографии — трудная задача. Поэтому можно выбрать более простой путь: изготовление бинарного фазового ДОЭ. Если рассчитанную по алгоритму (7.99)–(7.103) фазу $S(r, \varphi)$ проквантовать по двум уровням 0 и π, то получившийся бинарный фазовый ДОЭ не будет генерировать вращающийся пучок. Чтобы сохранить эффект вращения, перед процедурой квантования нужно добавить к фазе $S(r, \varphi)$ «слагаемое призмы»:

$$\hat{S}(r, \varphi) = S(r, \varphi) + 2\pi\nu r \cos\varphi. \tag{7.139}$$

На рис. 7.20*а* показана бинарная фаза, которая получена из полутоновой фазы (рис. 7.19*а*) с помощью (7.139) и последующего квантования по двум уровням 0 и π. Несущая частота $\nu = 10\,\text{мм}^{-1}$.

На рис. 7.20*б–е* показаны полутоновые распределения поперечной интенсивности (негатив), сформированные бинарным фазовым ДОЭ (рис. 7.20*а*) и вычисленные с помощью преобразования Френеля на плоскостях: $z = 215$ мм (7.20*б*), $z = 245$ мм (7.20*в*), $z = 310$ мм (7.20*г*), $z = 355$ мм (7.20*д*) и $z = 440$ мм (7.20*е*).

Из рис. 7.20 видно, что правый пучок вращается в ту же сторону, что и пучок на рис. 7.19, а левый — в противоположную. Если бы в (7.139) мы добавили «слагаемое призмы» с отрицательным знаком, то вращение этих пучков (рис.7.20) сменилось бы на противоположное. Более 80% всей световой энергии идет на формирование этих двух пучков. Видно также, что в таком способе кодирования фазы не появляется нулевой порядок, а появляются ± 1 порядки дифракции. Это следует из функционального ряда Гудмена–Сильвестри [49]:

$$\exp\left[i\hat{S}_M(r, \varphi)\right] = \sum_{m=-\infty}^{\infty} \operatorname{sinc}\left[\pi\left(m + \frac{1}{M}\right)\right] \exp\left[i(Mm + 1) \cdot \hat{S}(r, \varphi)\right], \tag{7.140}$$

где $\hat{S}(r, \varphi)$ — полутоновая фаза из (7.139), $\hat{S}_M(r, \varphi)$ — квантованная по M уровням фаза ДОЭ.

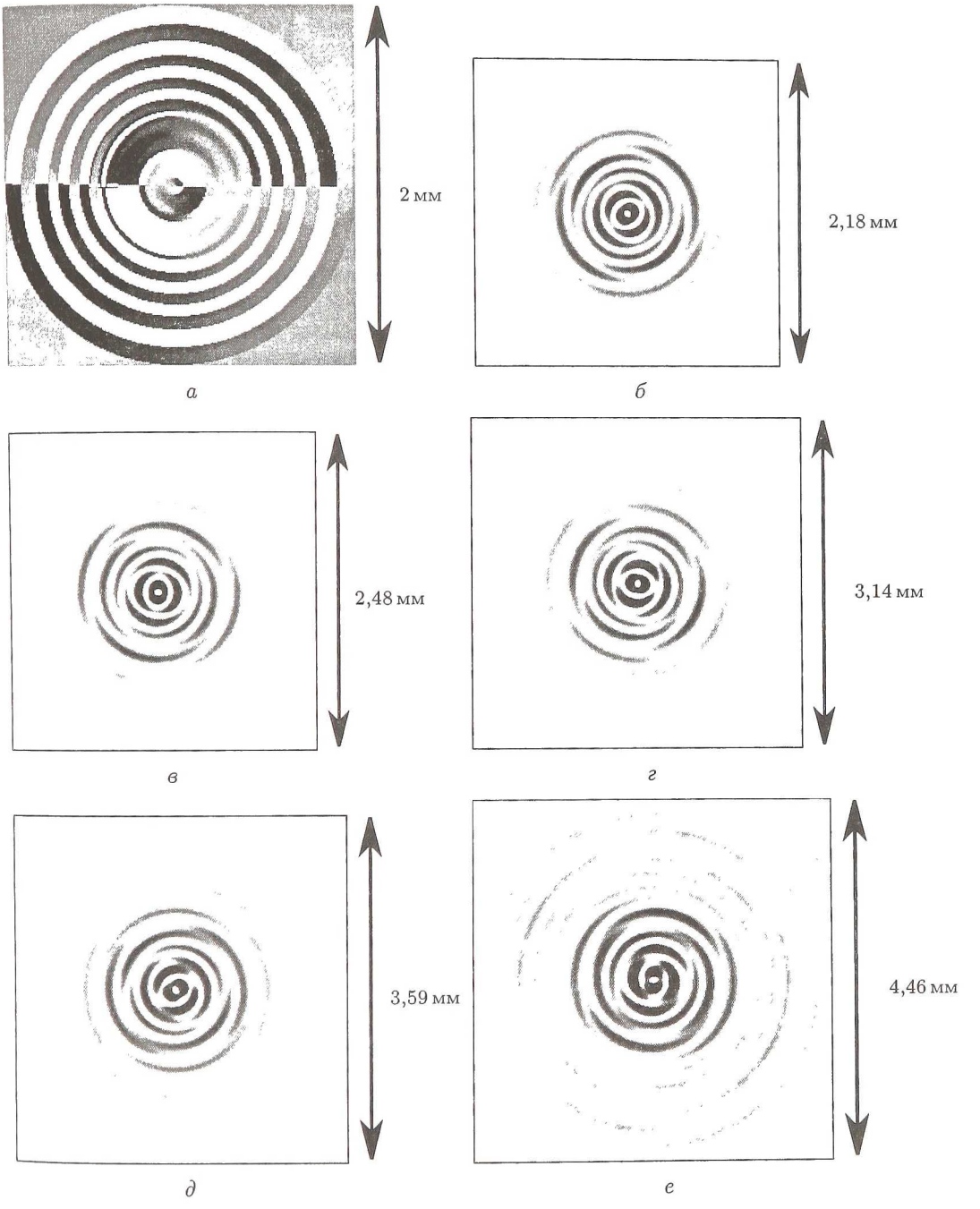


Рис. 7.19. Результат моделирования: фаза ДОЭ (*а*), формирующего двухмодовый пучок ГЛ, и поперечные распределения интенсивности пучка на разных расстояниях z: 215 мм (*б*), 245 мм (*в*), 310 мм (*г*), 355 мм (*д*), 440 мм (*е*)

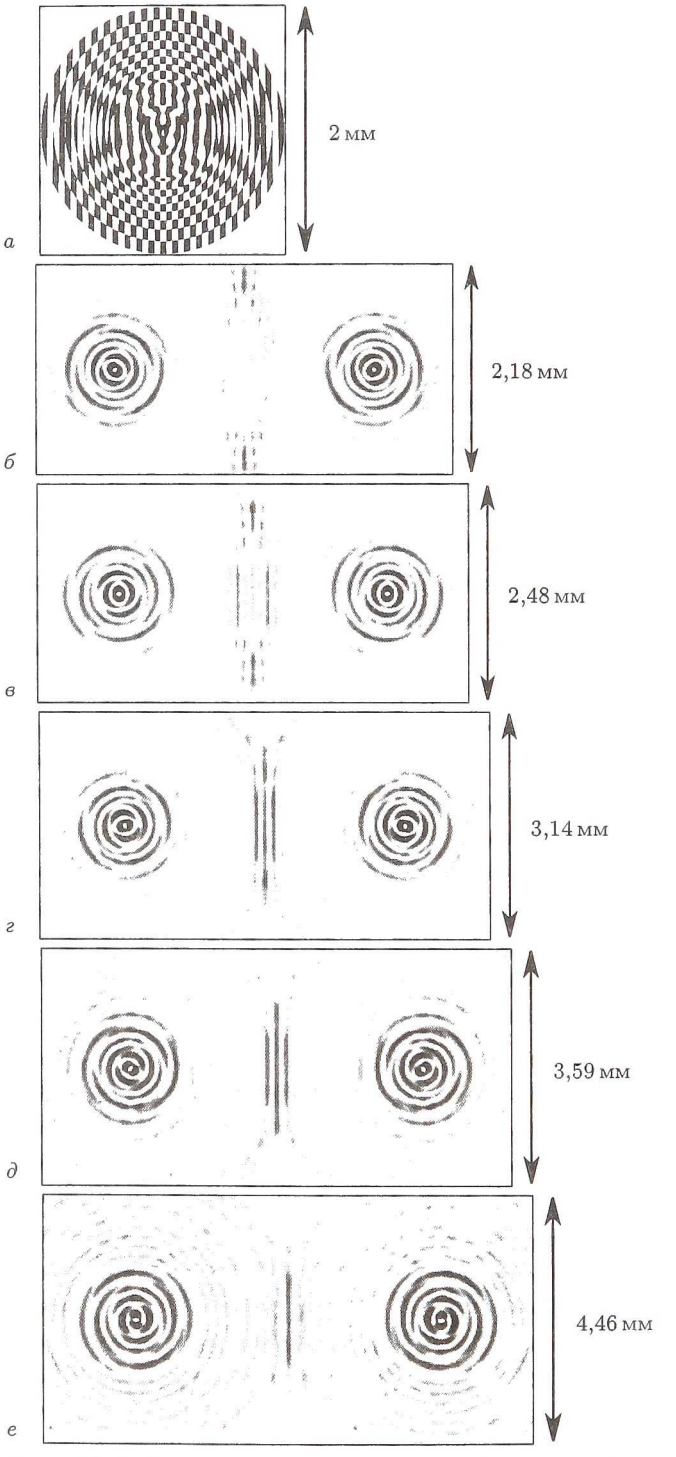


Рис. 7.20. Результат моделирования: бинарная фаза ДОЭ (*а*) и распределения интенсивности в поперечном сечении пучков на разных расстояниях z: 215 мм (*б*), 245 мм (*в*), 310 мм (*г*), 355 мм (*д*), 440 мм (*е*)

В нашем случае $M = 2$. Тогда вместо (7.140) получим:

$$\exp\left[i\hat{S}_2(r,\varphi)\right] = \sum_{m=0}^{\infty} \operatorname{sinc}\left[\pi\left(m+\frac{1}{2}\right)\right] \exp\left[i(2m+1)\cdot\hat{S}(r,\varphi)\right] =$$
$$= (2/\pi)\exp\left[iS(r,\varphi) + i2\pi\nu r\cos\varphi\right] + (2/\pi)\exp\left[-iS(r,\varphi) - i2\pi\nu r\cos\varphi\right] -$$
$$- (2/3\pi)\exp\left[3iS(r,\varphi) + i6\pi\nu r\cos\varphi\right] - (2/3\pi)\exp\left[-3iS(r,\varphi) - i6\pi\nu r\cos\varphi\right] +$$
$$+ \ldots \quad (7.141)$$

Дополнительные порядки дифракции, которые появляются в соответствии с (7.141) при дифракции плоской волны на бинарном фазовом ДОЭ, не показаны на рис.7.20*б–е*. Показаны только ± 1 порядки, которые соответствуют двум первым слагаемым в (7.141), эффективность каждого из которых равна $4/\pi^2 = 40{,}5\%$.

С помощью процедуры, описанной в разделе 4, был изготовлен ДОЭ с бинарной фазой, аналогичной фазе на рис. 7.20*а*. Этот ДОЭ освещался коллимированным пучком He-Ne лазера с круговой апертурой. Поперечное распределение интенсивности регистрировалось на различных расстояниях с помощью ПЗС-камеры с окном 6,6 мм×4,78 мм и числом пикселов 768×567. На рис. 7.21*а* показано полученное экспериментально распределение интенсивности на расстоянии $z = 110$ мм от ДОЭ. Видно отсутствие нулевого порядка и наличие ± 1 порядков дифракции. Далее на камеру регистрировался один левый порядок дифракции. На рис. 7.21*б–е* показаны распределения интенсивности в -1 порядке на расстояниях: $z = 215$ мм (7.21*б*), $z = 245$ мм (7.21*в*), $z = 310$ мм (7.21*г*), $z = 355$ мм (7.21*д*) и $z = 440$ мм (7.21*е*). Из сравнения рис. 7.19*б–е*, рис. 7.20*б–е* и 7.21*б–е* видно вращение поперечного сечения лазерного пучка и хорошее согласие теории и эксперимента.

Пример 7.14. На рис. 7.22 и 7.23 показаны результаты численного и натурного эксперимента для случая распространения другого двухмодового пучка ГЛ с номерами (n, m): $(1, -1)$ и $(16, 2)$. На рис. 7.22*а* показана полутоновая фаза ДОЭ, полученная за две итерации по алгоритму (7.159), (7.160). На рис. 7.22*б* показана бинарная фаза, полученная из полутоновой фазы (рис. 7.22*а*) с помощью (7.139) и квантования по двум уровням 0 и π. На рис. 7.22*в–ж* показаны вычисленные с помощью преобразования Френеля поперечные распределения интенсивности светового поля (негатив), сформированного бинарным фазовым ДОЭ (рис. 7.22*б*) на расстояниях: $z = 190$ мм (7.22*в*), $z = 225$ мм (7.22*г*), $z = 295$ мм (7.22*д*), $z = 335$ мм (7.22*е*) и $z = 465$ мм (7.22*ж*). Радиус ДОЭ $R = 1$ мм, число пикселов вдоль радиуса — 256. Перетяжка гауссова пучка при $z = 0$ равна $\sigma_0 = 0{,}17$ мм.

На рис. 7.23*а* показано зарегистрированное ПЗС-камерой распределение интенсивности светового поля, сформированного бинарным фазовым ДОЭ, изготовленным по описанной выше технологии, на расстоянии $z = 167$ мм. На рис. 7.23*б–е* показаны поперечные распределения интенсивности только в правом пучке (рис. 7.23*а*) на расстояниях: $z = 190$ мм (7.23*б*), $z = 225$ мм (7.23*в*), $z = 295$ мм (7.23*г*), $z = 335$ мм (7.23*д*) и $z = 465$ мм (7.23*е*).

Из сравнения рисунков 7.22 и 7.23 также хорошо видно согласие теории и эксперимента и наличие эффекта вращения пучка. Световое кольцо на рис. 7.21*а* и 7.23*а* появляется из-за того, что диаметр апертуры освещающего лазерного пучка был больше, чем диаметр ДОЭ: диаметр ДОЭ — 2 мм, диаметр апертуры — 3 мм. Больший размер освещающего пучка был выбран специально, чтобы создать нулевой порядок, интерферирующий на близких расстояниях с расходящимися пучками.

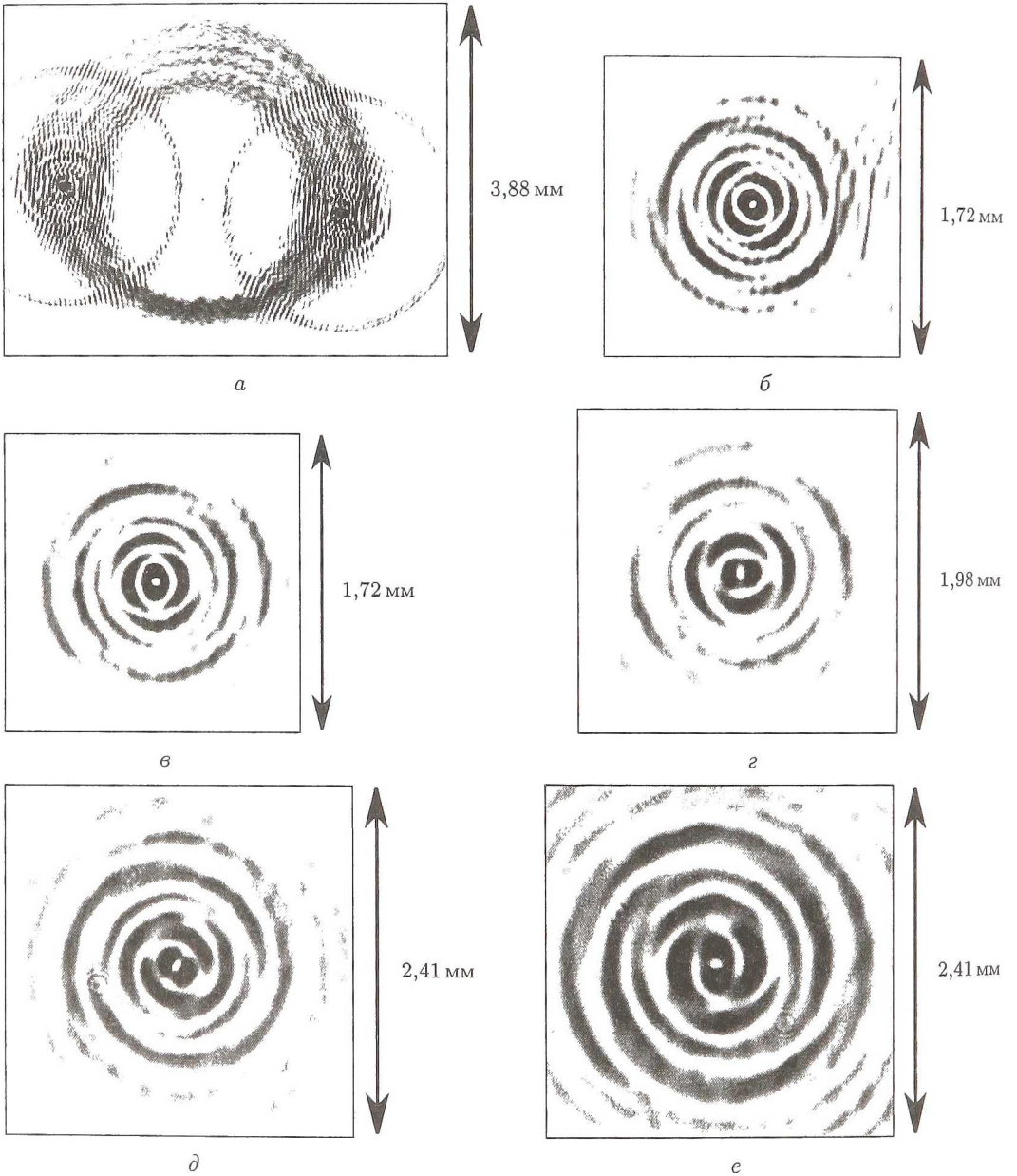


Рис. 7.21. Эксперимент: распределения интенсивности в поперечном сечении пучка на разных расстояниях z: 110 мм (*а*), 215 мм (*б*), 245 мм (*в*), 310 мм (*г*), 355 мм (*д*), 440 мм (*е*)

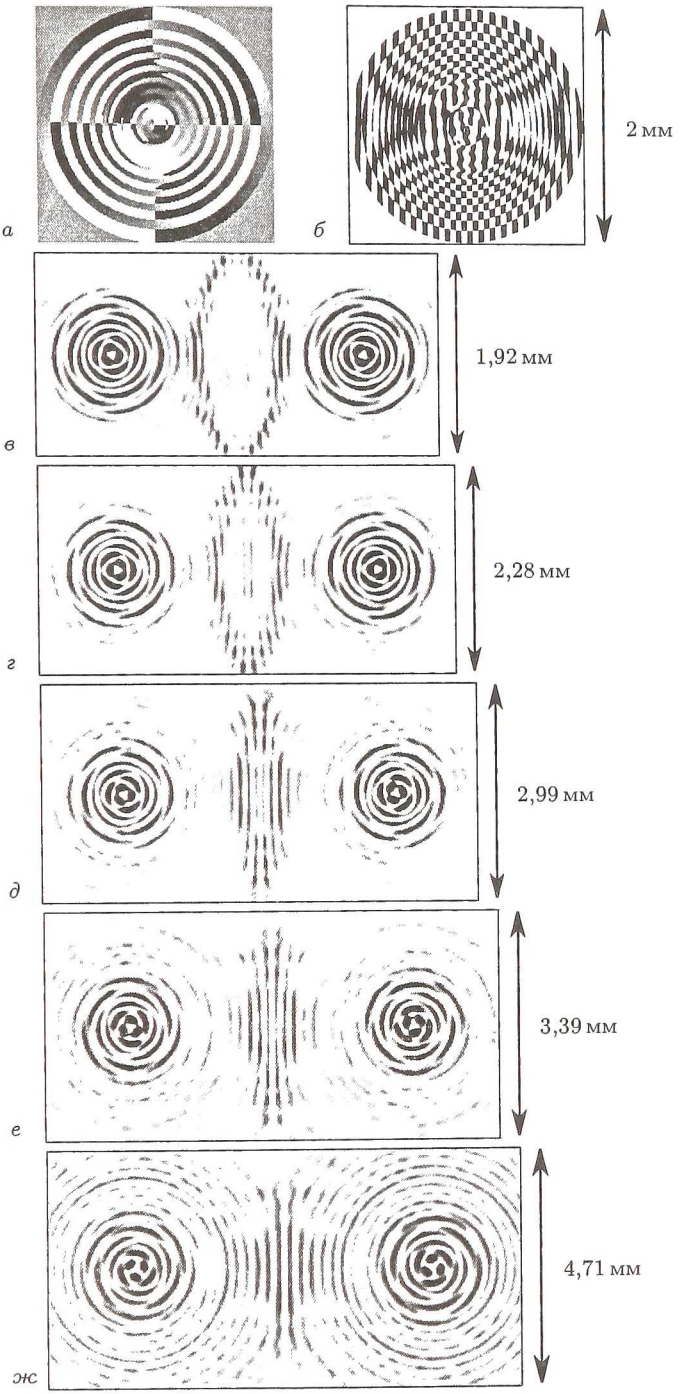


Рис. 7.22. Результат моделирования: полутоновая (*а*) и бинарная с несущей частотой (*б*) фазы ДОЭ, формирующего двухмодовые пучки ГЛ; распределения интенсивности поперечного сечения пучков на разных расстояниях z: 190 мм (*в*), 225 мм (*г*), 295 мм (*д*), 335 мм (*е*), 465 мм (*ж*)

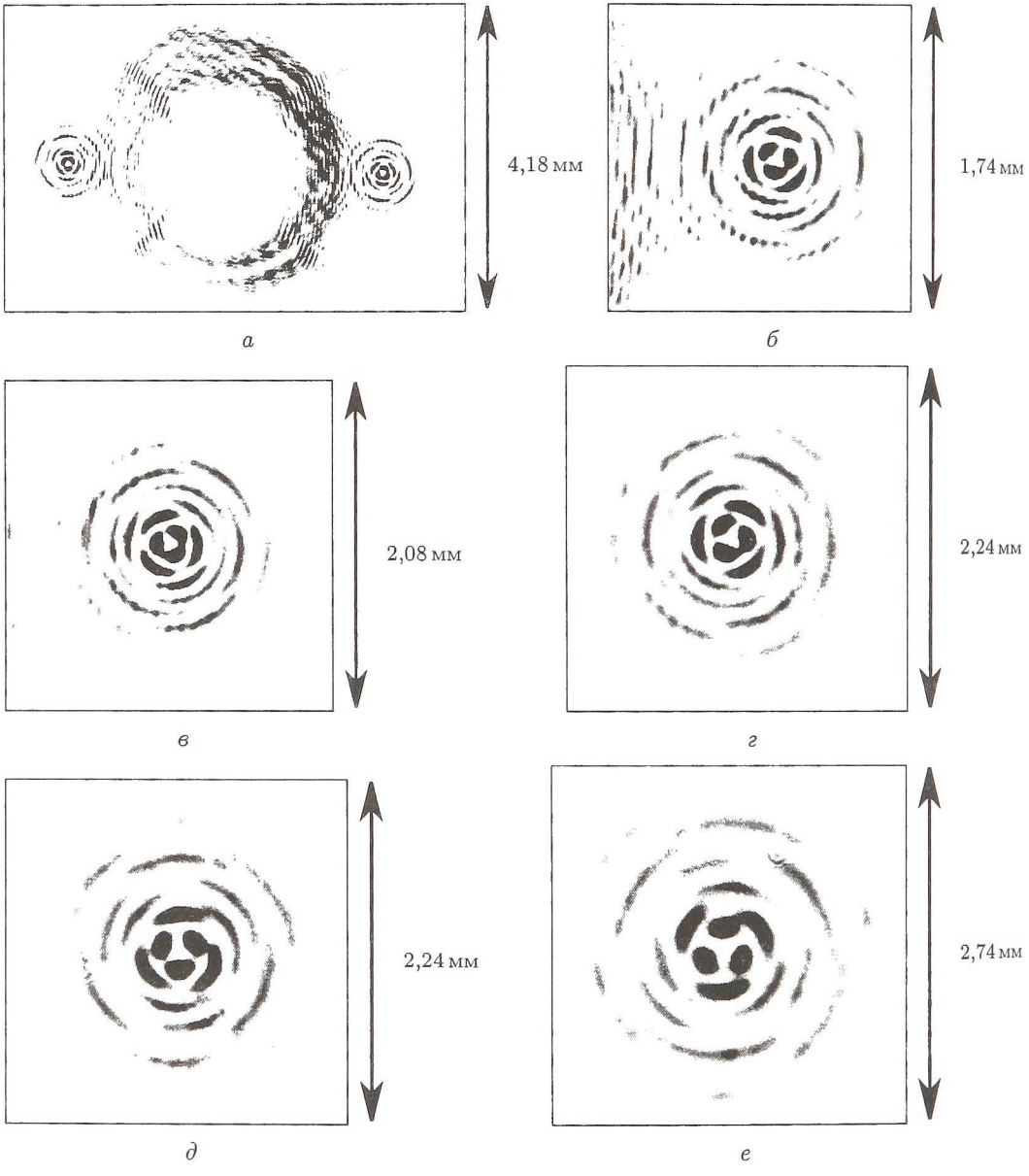


Рис. 7.23. Эксперимент: распределения интенсивности в поперечном сечении пучков на разных расстояниях z: 167 мм (*а*), 190 мм (*б*), 225 мм (*в*), 295 мм (*г*), 335 мм (*д*), 465 мм (*е*)

На рис. 7.21*а* и 7.23*а* показана картина интерференции недифрагированной части пучка с ± 1-ми порядками дифракции: видны смещения полос на полпериода при переходе от кольца к кольцу в сечении модовых пучков.

7.8. Обобщенные эрмитовы световые пучки в свободном пространстве

Интерес к эрмитовым пучкам или, в частности, модам Гаусса–Эрмита (ГЭ), обусловлен тем, что в свободном пространстве они распространяются без изменения своей структуры, изменяясь только масштабно, а в световых волокнах с параболическим профилем показателя преломления они распространяются и без изменения масштаба. В работах [60, 61] приведены разные аналитические выражения для мод ГЭ. Это послужило толчком к тому, чтобы выяснить степень различия и сходства двух разных форм эрмитовых пучков. Оказалось, что оба типа пучков являются частными случаями более общих модовых эрмитовых пучков. Известны итеративные алгоритмы [50, 51, 53] для расчета ДОЭ, генерирующего одномодовый пучок с использованием вспомогательной области или на основе аппроксимации функции пропускания транспаранта конечной суммой ортогональных мод.

В данном разделе показано, что с энергетической точки зрения оптимальным ДОЭ, формирующим одномодовый эрмитов пучок является транспарант, функция пропускания которого равна знаковой функции соответствующего многочлена Эрмита. Приведены также аналитические выражения, описывающие дифракцию Фраунгофера на таких фазовых элементах.

Известно, что частным решением параболического уравнения распространения

$$\left(2ik\frac{\partial}{\partial z}+\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)E\left(x,y,z\right)=0, \tag{7.142}$$

где $E(x,y,z)$ — медленно меняющаяся вдоль z комплексная амплитуда светового поля, являются модовые функции вида [60]

$$E^{(1)}_{m,n}\left(x,y,z\right)=\left[\frac{\omega_0}{\omega\left(z\right)}\right]^{m+n+2}\exp\left[-\frac{x^2+y^2}{\omega^2\left(z\right)}\right]H_n\Big(\frac{x}{\omega\left(z\right)}\Big)H_m\Big(\frac{y}{\omega\left(z\right)}\Big), \tag{7.143}$$

где

$$\omega\left(z\right)=\omega_0\left(1+iz/z_0\right)^{1/2}, \tag{7.144}$$

$$z_0=k\omega_0^2/2, \tag{7.145}$$

$H_n(x)$ — многочлен Эрмита, ω_0 — минимальный радиус пучка при $z=0$, z_0 — конфокальный параметр.

Функция $E^{(1)}_{m,n}$, при изменении расстояния z, изменяется только масштабно, сохраняя свой вид. Поэтому решение (7.143) уравнения (7.142) можно назвать световой эрмитовой модой. Заметим, что моды (7.143) взаимно не ортогональны, что следует из справочного интеграла [54]

$$\int\limits_{-\infty}^{\infty}e^{-2x^2}H_n(x)H_m(x)\,\mathrm{d}x=(-1)^{\frac{m+n}{2}}2^{\frac{m+n-1}{2}}\Gamma\left(\frac{m+n+1}{2}\right), \tag{7.146}$$

где $\Gamma(x)$ — гамма-функция.

С другой стороны, известны взаимно ортогональные моды ГЭ, которые также являются частными решениями уравнения (7.142) [61]:

$$E^{(2)}_{m,n}(x,y,z) = \frac{\omega_0}{a(z)} \exp\left[i(n+m+1)\eta(z)\right] \times \\ \times \exp\left[\frac{-ik(x^2+y^2)}{2R(z)}\right] \exp\left[-\frac{x^2+y^2}{a^2(z)}\right] H_n\left(\frac{\sqrt{2}x}{a(z)}\right) H_m\left(\frac{\sqrt{2}y}{a(z)}\right), \quad (7.147)$$

где

$$a(z) = \omega_0\left(1 + z^2/z_0^2\right)^{1/2}, \quad (7.148)$$

$$\eta(z) = \operatorname{arctg}(z/z_0), \quad (7.149)$$

$$R(z) = z\left(1 + z_0^2/z^2\right), \quad (7.150)$$

$R(z)$ — радиус кривизны модового пучка.

Функции $E^{(1)}_{m,n}$ и $E^{(2)}_{m,n}$ описывают разные световые пучки, так как порождаются разными граничными условиями. При $z = 0$ вместо (7.143) и (7.147) получим

$$E^{(j)}_{m,n}(x,y,0) = \exp\left[-\frac{x^2+y^2}{\omega_0}\right] H_n\left(\frac{\theta_j x}{\omega_0}\right) H_m\left(\frac{\theta_j y}{\omega_0}\right), \quad j = 1,2, \quad (7.151)$$

где $\theta_j = \begin{cases} 1, j = 1, \\ \sqrt{2}, j = 2. \end{cases}$

Эти два типа модовых эрмитовых пучков отличаются также разным видом дифракции Фраунгофера. Световое поле $E^{(2)}_{m,n}$ является собственной функцией оператора Фурье [18]:

$$\int_{-\infty}^{\infty} e^{-x^2} H_n\left(\sqrt{2}x\right) e^{-i2x\xi}\,\mathrm{d}x = \sqrt{\pi}(-i)^n e^{-\xi^2} H_n\left(\sqrt{2}\xi\right). \quad (7.152)$$

Это свойство мод ГЭ сохранять свою структуру не только в ближней зоне дифракции Френеля, но и в дальней зоне Фраунгофера (или в фокальной плоскости линзы), используется для эффективного ввода лазерного излучения в световые волокна [50].

Световое поле $E^{(1)}_{m,n}$, сохраняя свою структуру в зоне дифракции Френеля, в дальней зоне «вырождается» в поле с нулевой интенсивностью в центре. Это следует из выражения

$$\int_{-\infty}^{\infty} e^{-x^2} H_n(x)\, e^{-i2x\xi}\,\mathrm{d}x = \sqrt{\pi}(-i)^n e^{-\xi^2} (2\xi)^n. \quad (7.153)$$

Можно показать, что функции $E^{(1)}_{m,n}$ и $E^{(2)}_{m,n}$ из уравнений (7.143) и (7.147) являются частным случаем более общего решения уравнения (7.142), которое имеет вид:

$$E^{(3)}_{m,n}(x,y,z) = A_{m,n}(z) \exp\left[-B(z)\left(x^2+y^2\right)\right] H_n(x/C(z)) H_m(y/C(z)), \quad (7.154)$$

где

$$A_{m,n}(z) = (1 + i\gamma z/z_0)^{\frac{m+n}{2}} (1 + iz/z_0)^{-\frac{m+n}{2}-1}, \tag{7.155}$$

$$B(z) = \left[\omega_0^2 \varepsilon^2 (1 + iz/z_0)(1 + i\gamma z/z_0)\right]^{-1} + k\gamma \left[2z_0 (1 + i\gamma z/z_0)\right]^{-1}, \tag{7.156}$$

$$C(z) = \omega_0 \varepsilon \left[(1 + iz/z_0)(1 + i\gamma z/z_0)\right]^{1/2}, \tag{7.157}$$

$\gamma = 1 - \varepsilon^{-2}$, ε — модовый параметр. При $\varepsilon = 1$ решение (7.154) переходит в решение $E^{(1)}_{m,n}$, а при $\varepsilon = \left(\sqrt{2}\right)^{-1}$ решение $E^{(3)}_{m,n}$ переходит в $E^{(2)}_{m,n}$.

Обобщенные эрмитовы световые пучки (7.154) формируются с помощью комплексной амплитуды при $z = 0$ вида

$$E^{(3)}_{m,n}(x, y, 0) = \exp\left[-\frac{x^2 + y^2}{\omega_0^2}\right] H_n\left(\frac{x}{\omega_0 \varepsilon}\right) H_m\left(\frac{y}{\omega_0 \varepsilon}\right). \tag{7.158}$$

Выражение для комплексной амплитуды светового пучка (7.154) в дальней зоне следует из соотношения:

$$\int_{-\infty}^{\infty} e^{-x^2} H_n(x/\varepsilon) e^{-i2x\xi}\, \mathrm{d}x = \sqrt{\pi}(-i)^n (-\gamma)^{n/2}\, e^{-\xi^2} H_n\left(\frac{\xi}{\sqrt{1 - \varepsilon^2}}\right), \tag{7.159}$$

где $\varepsilon < 1$. При $\varepsilon = \left(\sqrt{2}\right)^{-1}$ (7.159) переходит в (7.152). При $\varepsilon = 1$, используя ассимптотику $H_n(\omega\xi) \xrightarrow{\omega \to \infty} 2^n (\omega\xi)^n$, получаем: $(-\gamma)^{n/2} H_n\left(\xi/\sqrt{1 - \varepsilon^2}\right) = (2\xi)^n$. Т. е. (7.159) переходит в (7.153).

Для энергетически эффективного формирования эрмитовых пучков обоих типов предлагается в плоскости $z = 0$ располагать фазовый ДОЭ с функцией пропускания

$$\tau_{m,n}(x, y) = \operatorname{sgn}\left[H_n(x) H_m(y)\right] = \operatorname{sgn} H_n(x) \operatorname{sgn} H_m(y), \tag{7.160}$$

где $\operatorname{sgn}(x)$ — знаковая функция.

Задача поиска оптимального фазового фильтра, формирующего эрмитовы пучки, сходна с аналогичной задачей для формирования бесселевых пучков [62].

Покажем в одномерном случае, что транспарант (7.160) эффективно формирует эрмитовый пучок n-го порядка. Для этого разложим функцию $\tau_n(x)$ по ортогональным многочленам:

$$\operatorname{sgn} H_n(x) = \sum_{p=0}^{\infty} C_p^{(n)} H_p(x), \tag{7.161}$$

где

$$C_p^{(n)} = A \int_{-\infty}^{\infty} \operatorname{sgn}\left[H_n(x)\right] e^{-x^2} H_p(x)\, \mathrm{d}x, \tag{7.162}$$

$A = (2^n n!\sqrt{\pi})^{-1}$. Для четных номеров $n = 2l$ нули $x_{k,l}$ многочлена: $H_{2l}(x_{k,l}) = 0$, будут расположены симметрично относительно точки $x = 0$: $x_{-k,l} = -x_{k,l}$. Тогда интеграл в уравнении (7.162) можно представить в виде суммы интегралов:

$$\int_{-\infty}^{\infty} \operatorname{sgn}[H_{2l}(x)]\, e^{-x^2} H_p(x)\, \mathrm{d}x = \int_{-\infty}^{\infty} e^{-x^2} H_p(x)\, \mathrm{d}x + 2(-1)^l \sum_{k=1}^{l} (-1)^{k-1} \int_{-x_{k,l}}^{x_{k,l}} e^{-x^2} H_p(x)\, \mathrm{d}x. \quad (7.163)$$

С учетом справочного интеграла [54]

$$\int_0^{\xi} e^{-x^2} H_n(x)\, \mathrm{d}x = H_{n-1}(0) - e^{-\xi^2} H_{n-1}(\xi), \quad (7.164)$$

вместо (7.163) получим следующее выражение для коэффициентов разложения (7.162):

$$C_{2s}^{(2l)} = \left[2^{2s-1}(2s)!\sqrt{\pi}\right]^{-1} (-1)^l \sum_{k=1}^{l} (-1)^k e^{-x_{k,l}^2} H_{2s-1}(x_{k,l}). \quad (7.165)$$

При этом нечетные коэффициенты равны нулю.

Аналогично можно показать, что для нечетных многочленов Эрмита $n = 2l+1$ имеет место следующее выражение для коэффициентов разложения (7.162):

$$C_{2s+1}^{(2l+1)} = \left[2^{2s}(2s+1)!\sqrt{\pi}\right]^{-1} (-1)^l \left[H_{2s}(0) + 2\sum_{k=1}^{l} (-1)^k e^{-x_{k,l}^2} H_{2s}(x_{k,l})\right], \quad (7.166)$$

где $H_{2l+1}(x_{k,l}) = 0$, $H_{2s}(0) = (-1)^s \cdot 2^s \cdot 3 \cdot 5 \cdot 7 \ldots$, четные коэффициенты равны нулю.

Таким образом, освещая ДОЭ с пропусканием (7.160) коллимированным гауссовым пучком с амплитудой $\exp[-(x^2+y^2)/2]$, получим поле за элементом в виде суперпозиции эрмитовых пучков, один из которых будет энергетически преобладающим с эффективностью около 70%:

$$e^{-x^2/2} \operatorname{sgn}[H_n(x)] = \sum_{p=0}^{\infty} C_p^{(n)} \left(e^{-x^2/2} H_p(x)\right). \quad (7.167)$$

Для расчета дифракции Фраунгофера плоской волны на участке $[-a, a]$ оптического элемента с пропусканием $\operatorname{sgn}[H_n(x)]$, где a должно быть больше максимального корня многочлена $H_n(x)$, удобно использовать представления вида ($n = 2l$) [63]:

$$\operatorname{sgn}[H_{2l}(x)] = \operatorname{rect}(x/a) + 2(-1)^l \sum_{k=1}^{l} (-1)^{k-1} \operatorname{rect}(x/x_{k,l}), \quad (7.168)$$

где

$$H_{2l}(x_{k,l}) = 0, \quad \operatorname{rect}\left(\frac{x}{x_0}\right) = \begin{cases} 1, & |x| \leqslant x_0, \\ 0, & |x| > x_0. \end{cases}$$

Тогда преобразование Фурье от выражения (7.168) будет иметь вид

$$P_{2l}(\xi) = \mathcal{F}\{\operatorname{sgn}[H_{2l}(x)]\} = 2a \operatorname{sinc}(a\xi) + 2(-1)^l \sum_{k=1}^{l} (-1)^k 2x_{k,l} \operatorname{sinc}(x_{k,l}\xi), \qquad (7.169)$$

где

$$\mathcal{F}\{f(x)\} = \int_{-\infty}^{\infty} f(x) \exp(-ix\xi)\, \mathrm{d}x, \quad \operatorname{sinc}(x) = \sin(x)/x. \qquad (7.170)$$

Параметр диафрагмы a выбирается из энергетических соображений.

Аналогично получаются выражения для нечетных полиномов Эрмита:

$$\operatorname{sgn}[H_{2l+1}(x)] = \left[\operatorname{rect}(x/a) + 2(-1)^l \sum_{k=1}^{l} (-1)^{k-1} \operatorname{rect}(x/x_{k,l})\right] \operatorname{sgn} x, \qquad (7.171)$$

$$P_{2l+1}(\xi) = \mathcal{F}\{\operatorname{sgn}[H_{2l+1}(x)]\} = \\ = -i\left\{2a \operatorname{cosc}(a\xi) + 2(-1)^l \sum_{k=1}^{l} (-1)^k 2x_{k,l} \operatorname{cosc}(x_{k,l}\xi)\right\}, \qquad (7.172)$$

где $\operatorname{cosc}(x) = (1 - \cos x)/x$.

Пример 7.15. Пусть требуется сформировать первую моду ГЭ при $n = 1$, $m = 0$. Для этого в плоскости $z = 0$, следует расположить транспарант с пропусканием $H_1(x)\exp\left(-x^2/2\right)$. Тогда в фокальной плоскости линзы согласно (7.152) сформируется световое поле с комплексной амплитудой, пропорциональной выражению

$$-i\sqrt{\pi} H_1(\xi) \exp\left(-\xi^2/2\right), \text{ где } H_1(\xi) = 2\xi.$$

Если в плоскости $z = 0$ разместить энергетически более выгодный фазовый элемент с пропусканием $\operatorname{sgn}[H_1(x)] = \operatorname{sgn}(x)$ [64], то согласно уравнению (7.166) в пространстве за фильтром сформируется суперпозиция световых мод:

$$\operatorname{sgn} x = \sum_{s=0}^{\infty} C^{(1)}_{2s+1} H_{2s+1}(x) = \frac{1}{\sqrt{\pi}}\left(H_1(x) - \frac{1}{12}H_3(x) + \frac{1}{160}H_5(x) - \ldots\right). \qquad (7.173)$$

Освещая такой фазовый элемент коллимированным гауссовым пучком, получим, что согласно (7.173), в частотной плоскости сформируется световое поле с амплитудой, пропорциональной

$$\int_{-\infty}^{\infty} \operatorname{sgn}(x)\, e^{-x^2/2} e^{-ix\xi}\, \mathrm{d}x = \sqrt{\pi} \sum_{s=0}^{\infty} C^{(1)}_{2s+1} (-i)^{2s+1} e^{-\xi^2/2} H_{2s+1}(\xi) = \\ = -ie^{-\xi^2/2}\left(H_1(\xi) - \frac{1}{12}H_3(\xi) + \frac{i}{160}H_5(\xi) - \ldots\right). \qquad (7.174)$$

Отношение энергий пучков, описываемых 2-м и 1-м слагаемыми (7.174) равно

$$\frac{\int\limits_{-\infty}^{\infty} H_3^2(\xi) e^{-\xi^2}\,\mathrm{d}\xi}{(12)^2 \int\limits_{-\infty}^{\infty} H_1^2(\xi) e^{-\xi^2}\,\mathrm{d}\xi} \approx 0{,}167. \tag{7.175}$$

Таким образом, в первом слагаемом суммы (7.174), описывающем первую моду ГЭ, содержится более 80% всей энергии светового пучка.

Еще более эффективным способом формирования первой эрмитовой моды оказывается способ освещения фазового фильтра с пропусканием

$$\tau_1(x) = \operatorname{sgn}\left[H_1(x)\right] \operatorname{rect}\left(x/a\right) \tag{7.176}$$

не гауссовым пучком, а плоской световой волной. В этом случае в фокальной плоскости линзы сформируется световое поле с амплитудой, пропорциональной

$$\int\limits_{-a}^{a} \operatorname{sgn}(x)\, e^{-ix\xi}\,\mathrm{d}x = -2ai \operatorname{cosc}(a\xi), \tag{7.177}$$

функция $\operatorname{cosc}(x)$ определена (7.172).

Взаимная корреляция светового поля (7.177) и первой эрмитовой моды оценивается выражением

$$\eta = \frac{\left| \int\limits_{-\infty}^{\infty} \left[2a \operatorname{cosc}(a\xi)\right] H_1(\xi) e^{-\xi^2/2}\,\mathrm{d}\xi \right|^2}{\int\limits_{-\infty}^{\infty} \left|2a \operatorname{cosc}(a\xi)\right|^2 \mathrm{d}\xi \cdot \int\limits_{-\infty}^{\infty} \left|H_1(\xi) e^{-\xi^2/2}\right|^2 \mathrm{d}\xi} = \frac{4}{a\sqrt{\pi}} \left(1 - e^{-a^2/2}\right)^2. \tag{7.178}$$

Размер апертуры фазового элемента (7.176) $[-a, a]$ выбирается из условия максимума выражения (7.178), который достигается при $a = 2{,}2$. При этом корреляция (7.178) будет равна $\eta = 0{,}85$. Таким образом, первая мода ГЭ содержится в световом поле (7.177) с эффективностью около 85%.

При численном моделировании использовались следующие значения параметров: размерность массивов $N = 256$, область изменения аргумента $x \in [-8\,\text{мм}, 8\,\text{мм}]$, длина волны $\lambda = 0{,}63$ мкм, фокусное расстояние $f = 100$ мм.

На рис. 7.24*а* показан эрмитов пучок

$$E_3^{(1)}(x) = H_3(x) \exp\left(-x^2\right). \tag{7.179}$$

На рис. 7.24*б,в* показаны соответственно интенсивность (квадрат модуля) и фаза (аргумент) фурье-преобразования от функции (7.179). Для сравнения на рис. 7.25 показан результат дифракции Фраунгофера для второго типа эрмитовых пучков. На рис. 7.25*а* показан сам пучок

$$E_3^{(2)}(x) = H_3(x) \exp\left(-x^2/2\right), \tag{7.180}$$

а на рис. 7.25*б,в* показаны интенсивность и фаза его фурье-образа, соответственно.

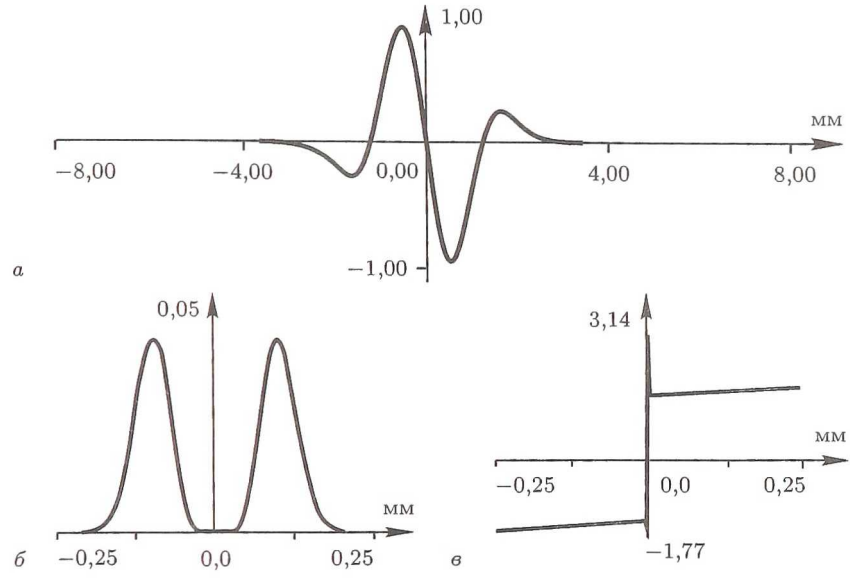


Рис. 7.24. Результат моделирования: мода $E_3^{(1)}(x)$ ($а$), квадрат модуля ($б$) и фаза ($в$) ее фурье-спектра

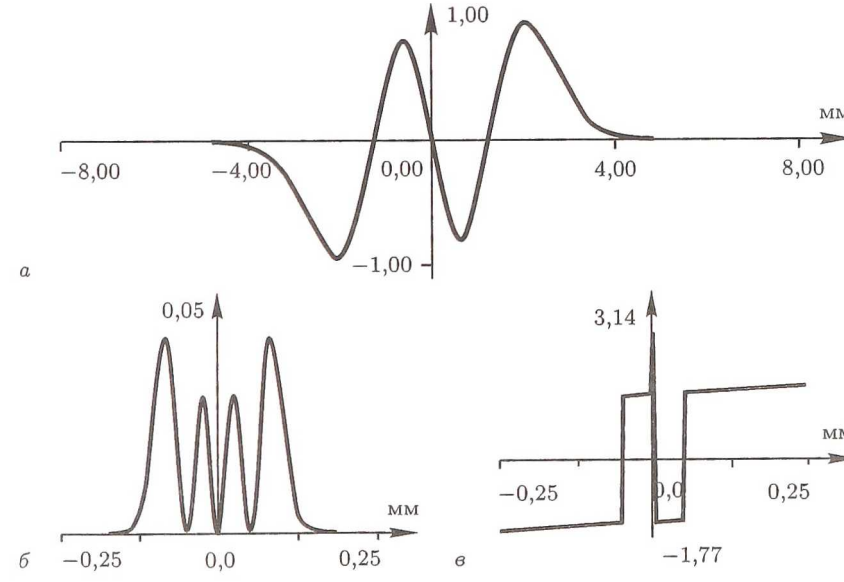


Рис. 7.25. Результат моделирования: мода $E_3^{(2)}(x)$ ($а$), квадрат модуля ($б$) и фаза ($в$) ее фурье-спектра

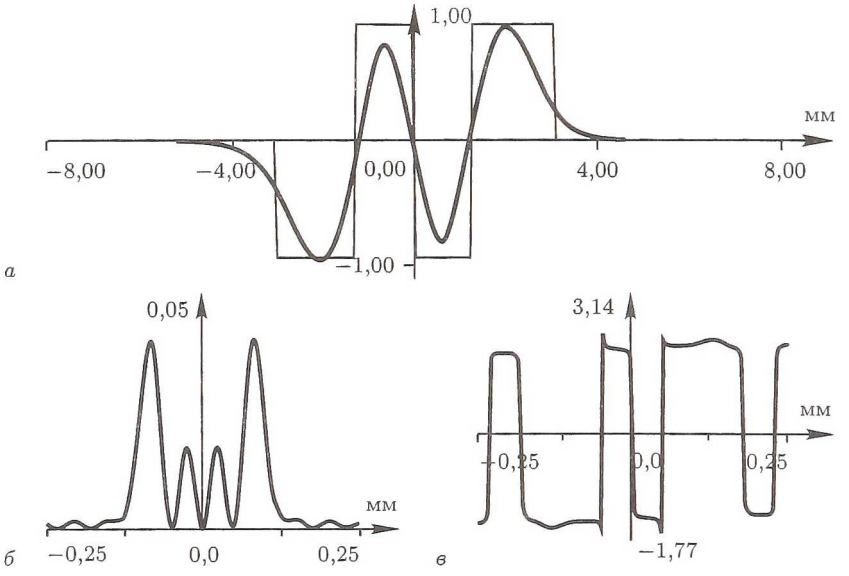


Рис. 7.26. Результат моделирования: бинарная фаза ДОЭ (*а*, кривая *1*), согласованная с $E_3^{(2)}(x)$ (*а*, кривая *2*), интенсивность (*б*) и фаза (*в*) фурье-спектра от ДОЭ

На рис. 7.26*а* показан результат бинаризации эрмитова пучка (7.180): кривая *1* — это функция

$$\tau_3(x) = \operatorname{sgn}\left[H_3(x)\right], \tag{7.181}$$

а кривая *2* — это функция (7.180). На рис. 7.26*б,в* показаны, соответственно, интенсивность и фаза фурье-преобразования от бинарной функции (7.181) при $x \in \in [-3\,\text{мм}, 3\,\text{мм}]$. Диапазон изменения аргумента x функции (7.181) выбран таким образом, чтобы корреляция η, определенная в (7.178), была максимальной. В этом случае корреляция между функциями интенсивности, рис. 7.25*б* и рис. 7.26*б*, составляет 0,84. Поэтому можно заключить, что 84% световой энергии плоского пучка, освещающего фазовый элемент (7.181), будет идти на формирование эрмитова пучка (7.180).

На рис. 7.27 показана зависимость взаимной корреляции η от полудлины отрезка $[-a, a]$ для эрмитова пучка третьего порядка $n = 3$. В таблице 7.3 приведены оптимальные размеры отрезков $[-a, a]$, на которых заданы бинарные

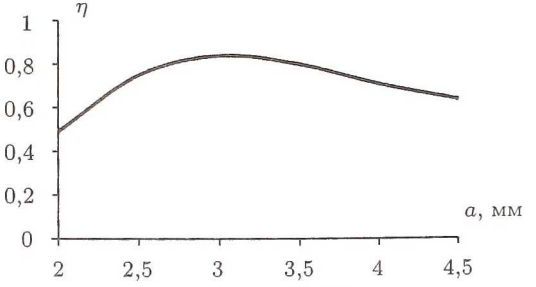


Рис. 7.27. Зависимость взаимной корреляции (7.178) от полудлины апертуры для эрмитова пучка третьего порядка

Таблица 7.3. Оптимальные размеры апертуры a для знаковых функций и значения взаимной корреляции η

Номер моды, n	1	2	3	4	5	6	7
Оптимальный размер апертуры, a (мм)	2,1	2,6	3,0	3,4	3,7	4,0	4,2
Коррелция, η	0,86	0,84	0,84	0,82	0,82	0,81	0,81

функции $\operatorname{sgn}[H_n(x)]$, $n = \overline{1,7}$ и значения взаимной корреляции между фурье-образами от функций (7.180) и (7.181) при $n = \overline{1,7}$.

Из таблицы 7.3 видно, что с ростом номера эрмитова пучка энергетическая эффективность η медленно спадает с 86% до 81% ($n \leqslant 7$).

7.9. Формирование мод Гаусса–Эрмита с помощью бинарных ДОЭ

Известны различные методы расчета ДОЭ, формирующих одно- или многомодовые пучки ГЭ. Одни из них позволяют добиваться высокой эффективности создаваемого пучка за счет не очень точного соблюдения модового состава, другие, наоборот, обеспечивая высокую точность формируемой моды, не позволяют достигнуть желаемой эффективности.

Так, в [65] для формирования лазерной моды ГЭ $(1,0)$ был изготовлен и исследован экспериментально фазовый 16-ти уровневый ДОЭ. Он был изготовлен по технологии электронной литографии с помощью травления полиметилметакрелата, нанесенного на подложку из кварцевого стекла. Этот элемент был рассчитан с помощью итеративной процедуры, предложенной в [51]. Размерность массива была 2048×2048 пикселов. Теоретическая эффективность равнялась 45,5%, а экспериментально измеренная — 37,7%.

В [64, 66] теоретически показано, что для эффективного формирования одномерных мод ГЭ можно использовать фазовые бинарные ДОЭ, которые получаются с помощью знаковой функции от соответствующего многочлена Эрмита. При освещении такого ДОЭ плоским пучком света с оптимальной апертурой, в дальней зоне дифракции формируется требуемая мода с эффективностью более 80%.

В данном разделе приведены результаты расчета и экспериментального исследования двухуровневых (бинарных) ДОЭ, формирующих моды ГЭ с номерами $(1,0)$, $(1,1)$ и $(1,2)$ с теоретической эффективностью более 63,8%. Кроме того, результаты работ [64, 66] обобщаются на двумерный случай формирования мод ГЭ. С помощью технологии электронной литографии получены несколько фазовых бинарных ДОЭ, согласованных с модами ГЭ, и проведена экспериментальная апробация их работоспособности.

Для эффективного формирования отдельных мод ГЭ предлагается использовать фазовый элемент с функцией пропускания

$$\tau_{m,n}(x,y) = \operatorname{sgn} H_m(x) \operatorname{sgn} H_n(y) \operatorname{rect}(x/a) \operatorname{rect}(y/a), \tag{7.182}$$

где $H_m(x)$, $H_n(y)$ — многочлены Эрмита m-го и n-го порядков,

$$\operatorname{sgn}(x) = \begin{cases} 1, & x \geqslant 0, \\ -1, & x < 0, \end{cases} \tag{7.183}$$

$$\operatorname{rect}\left(\frac{x}{a}\right)=\begin{cases}1, & |x|\leqslant a,\\ 0, & |x|>a.\end{cases} \tag{7.184}$$

С помощью уравнения (7.166) можно получить конкретные значения коэффициентов для разложения в ряд (7.161) для каждой моды ГЭ. Например, для моды ГЭ $(1,1)$

$$\operatorname{sgn} H_{1,1}(x,y)=\operatorname{sgn}(x)\cdot\operatorname{sgn}(y), \tag{7.185}$$

используя представление

$$\operatorname{sgn}(x)=\frac{1}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-1)^n}{2^{2n}(2n+1)n!}H_{2n+1}(x), \tag{7.186}$$

можно записать

$$\operatorname{sgn} H_{1,1}(x,y)=\frac{1}{\pi}\left(H_1(x)-\frac{1}{12}H_3(x)+\frac{1}{160}H_5(x)-\ldots\right)\times \\ \times\left(H_1(y)-\frac{1}{12}H_3(y)+\frac{1}{160}H_5(y)-\ldots\right). \tag{7.187}$$

Из (7.186) получается, что с эффективностью около 86%, бинарный ДОЭ с пропусканием $\operatorname{sgn} H_1(x)$ сформирует моду ГЭ, пропорциональную многочлену $H_1(x)$.

Для двумерных ДОЭ с пропусканием (7.182) эффективность, очевидно, можно оценить как произведение одномерных эффективностей: $0{,}86\times0{,}86\times100\%=73{,}96\%$. Эта оценка подтверждается на основании результатов моделирования.

В таблице 7.4 приведены значения среднеквадратичного отклонения δ и эффективности η, рассчитанных по следующим формулам:

$$\delta=\sqrt{\frac{\int\limits_{-b}^{b}\int\limits_{-b}^{b}\left\{\left|\mathcal{F}\left[\tau(x,y)\right]\right|^2-\left|\Psi_{m,n}(x,y)\right|^2\right\}^2\,\mathrm{d}x\,\mathrm{d}y}{\int\limits_{-b}^{b}\int\limits_{-b}^{b}\left\{\left|\Psi_{m,n}(x,y)\right|^2\right\}^2\,\mathrm{d}x\,\mathrm{d}y}}, \tag{7.188}$$

$$\eta=\frac{\left|\int\limits_{-b}^{b}\int\limits_{-b}^{b}\mathcal{F}\left[\tau(x,y)\right]\Psi^*_{m,n}(x,y)\,\mathrm{d}x\,\mathrm{d}y\right|^2}{\int\limits_{-b}^{b}\int\limits_{-b}^{b}\left|\mathcal{F}\left[\tau(x,y)\right]\right|^2\,\mathrm{d}x\,\mathrm{d}y\cdot\int\limits_{-b}^{b}\int\limits_{-b}^{b}\left|\Psi_{m,n}(x,y)\right|^2\,\mathrm{d}x\,\mathrm{d}y}, \tag{7.189}$$

Таблица 7.4. Рассчитанные средние ошибки и эффективность (корреляция) при формировании мод ГЭ с помощью бинарных знаковых ДОЭ с квадратной апертурой оптимального размера

ТЕМ (m,n)	$(1,0)$	$(1,1)$	$(1,2)$	$(7,0)$
$\delta,\%$	28,8	14,4	30,6	62,4
$\eta,\%$	63,8	72,3	68,5	33,5

где

$$\Psi_{m,n}(x,y) = \exp\left[-(x^2+y^2)/2\right] H_m(x) H_n(y). \tag{7.190}$$

Область интегрирования $[-b,b]\times[-b,b]$ в (7.188) и (7.189) выбиралась из условия $|\Psi_{m,n}(x,y)|^2 \leqslant 10^{-4}$. В (7.188) и (7.189) использовано обозначение преобразования Фурье от функции пропускания ДОЭ:

$$\mathcal{F}[\tau(x,y)] = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \tau(x,y) \exp\left[-\frac{ik}{f}(ux+vy)\right] \mathrm{d}x\,\mathrm{d}y. \tag{7.191}$$

Однако, на практике часто используют не квадратную, а круглую диафрагму, в этом случае результат получается примерно такой же (см. таблицу 7.5) Бинарные фазовые ДОЭ, функция пропускания которых удовлетворяет уравнению (7.182), были изготовлены с помощью низковольтового электронного генератора изображений Leica LION LV1 в Университете Йоенсуу, Финляндия (см. главу 4). На рис. 7.28 показан увеличенный под микроскопом участок получившегося фазового бинарного ДОЭ.

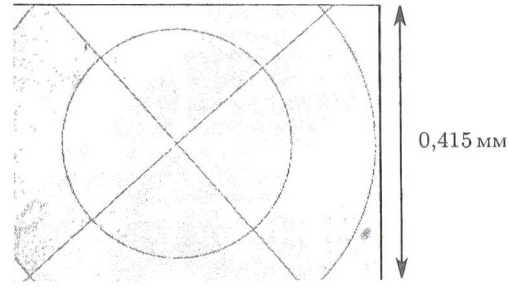


Рис. 7.28. Вид участка фазового бинарного ДОЭ под микроскопом.

Изготовленные по литографической технологии бинарные фазовые ДОЭ исследовались на установке, оптическая схема которой показана на рис. 7.29. Луч света от He-Ne лазера малой мощности (*1*) расширялся микрообъективом (*2*) и фильтровался с помощью малой круглой диафрагмы (*3*). После коллимирования расходящегося пучка линзой (*4*), он ограничивался другой круглой диафрагмой (*5*) радиусом примерно 1,1–1,2 мм. Ограниченный по радиусу плоский пучок попадал затем на бинарный ДОЭ (*6*). Распределение интенсивности в ближней зоне Френеля регистрировалось с помощью ПЗС-камеры (*8*). Приемное окно камеры имело размеры 6,6 мм×4,78 мм, число фотоячеек 768×567. Зарегистрированные картины дифракции запоминались в компьютере (*9*). Для исследования дифракции Фраунгофера на ДОЭ, в пространство за элементом вводилась линза с фокусным расстоянием 500 мм (*7*).

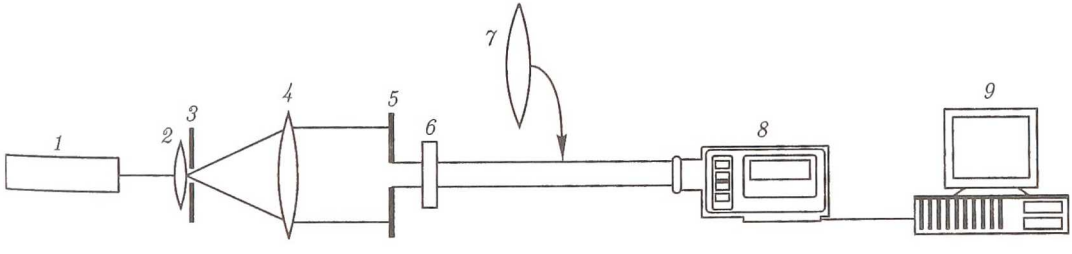


Рис. **7.29.** Экспериментальная установка: He-Ne лазер (*1*), микрообъектив (*2*), фильтрующая диафрагма (*3*), коллимирующая линза (*4*), ограничивающая диафрагма (*5*), бинарный ДОЭ (*6*), дополнительная линза для выполнения фурье-преобразования (*7*), ПЗС-камера (*8*), компьютер (*9*)

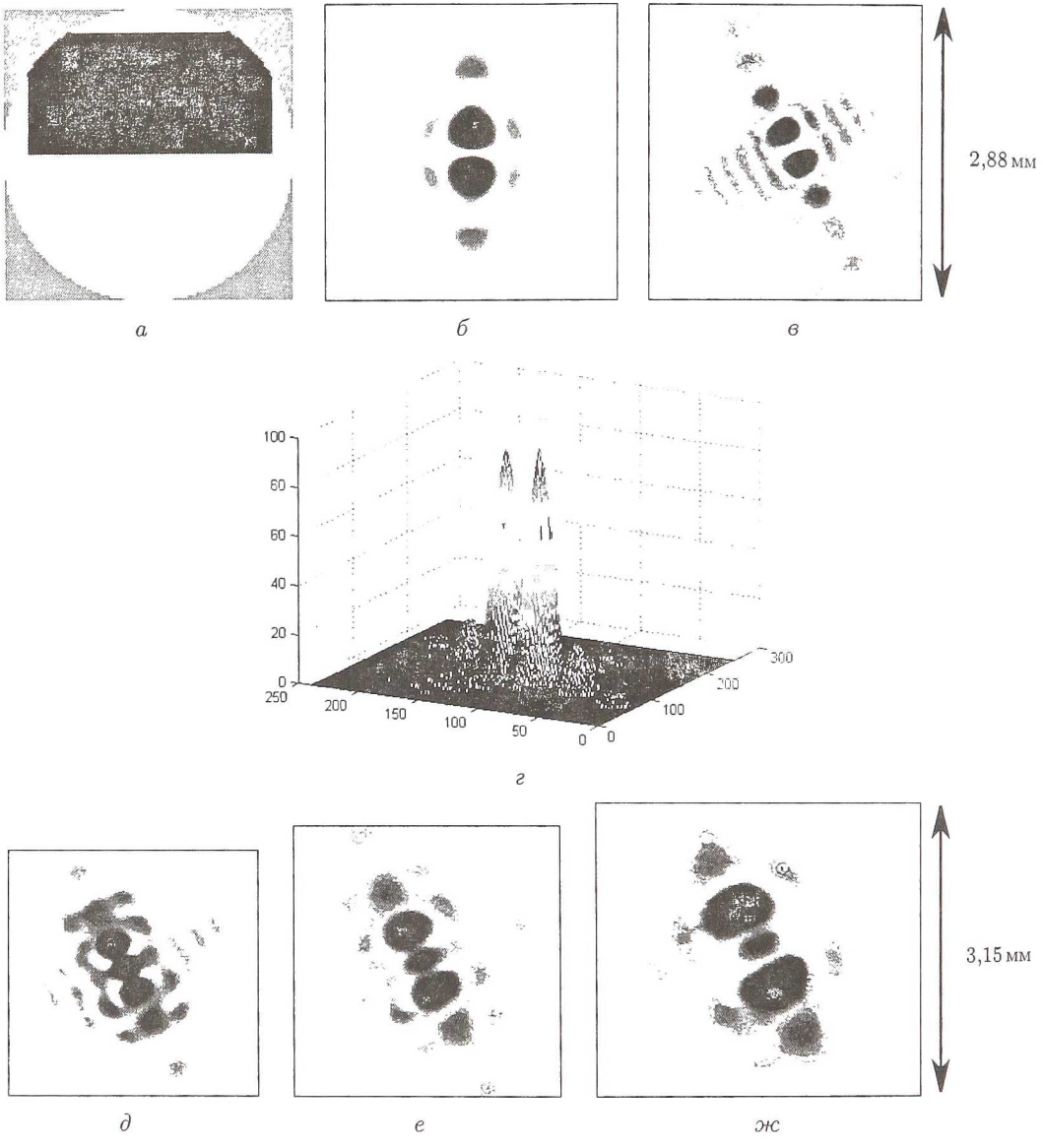


Рис. 7.30. Результаты численного моделирования и эксперимента для моды ГЭ (1,0): фотомаска, ограниченная круглой апертурой освещающего пучка (*а*); рассчитанная картина дифракции в дальней зоне (*б*); экспериментальная картина дифракции в дальней зоне (*в*) и ее трехмерное представление (*г*); экспериментальные распределения интенсивности на расстояниях 370 мм (*д*), 480 мм (*е*) и 620 мм (*ж*) от ДОЭ

Таблица 7.5. Рассчитанные средние ошибки и эффективность (корреляция) при формировании мод ГЭ с помощью бинарных знаковых ДОЭ с круглой апертурой оптимального размера

ТЕМ (m,n)	$(1,0)$	$(1,1)$	$(1,2)$	$(7,0)$
$\delta,\%$	26,9	15,6	31,8	66,9
$\eta,\%$	70,1	69,5	65,4	40,7

Таблица 7.6. Рассчитанные средние ошибки и эффективность при формировании мод ГЭ с помощью бинарных знаковых ДОЭ с круглой апертурой немного больше оптимального размера

ТЕМ (m,n)	$(1,0)$	$(1,1)$	$(1,2)$	$(7,0)$
$\delta, \%$	24,7	42,7	31,5	64,2
$\eta, \%$	55,6	62,2	59,4	31,7

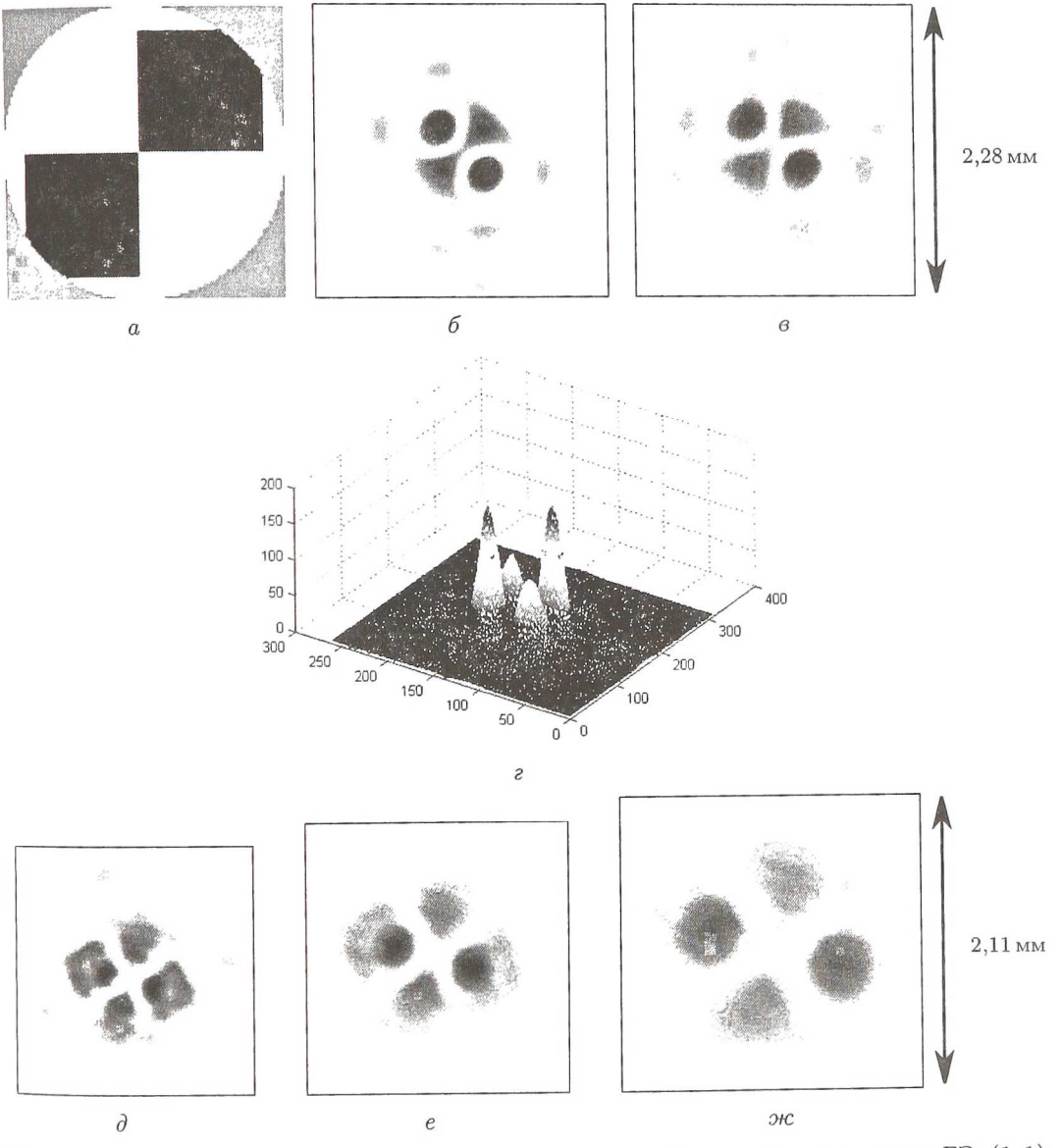


Рис. 7.31. Результаты численного моделирования и эксперимента для моды ГЭ $(1,1)$: фотомаска, ограниченная круглой апертурой освещающего пучка (*а*); рассчитанная картина дифракции в дальней зоне (*б*); экспериментальная картина дифракции в дальней зоне (*в*) и ее трехмерное представление (*г*); экспериментальные распределения интенсивности на расстояниях 340 мм (*д*), 455 мм (*е*) и 640 мм (*ж*) от ДОЭ

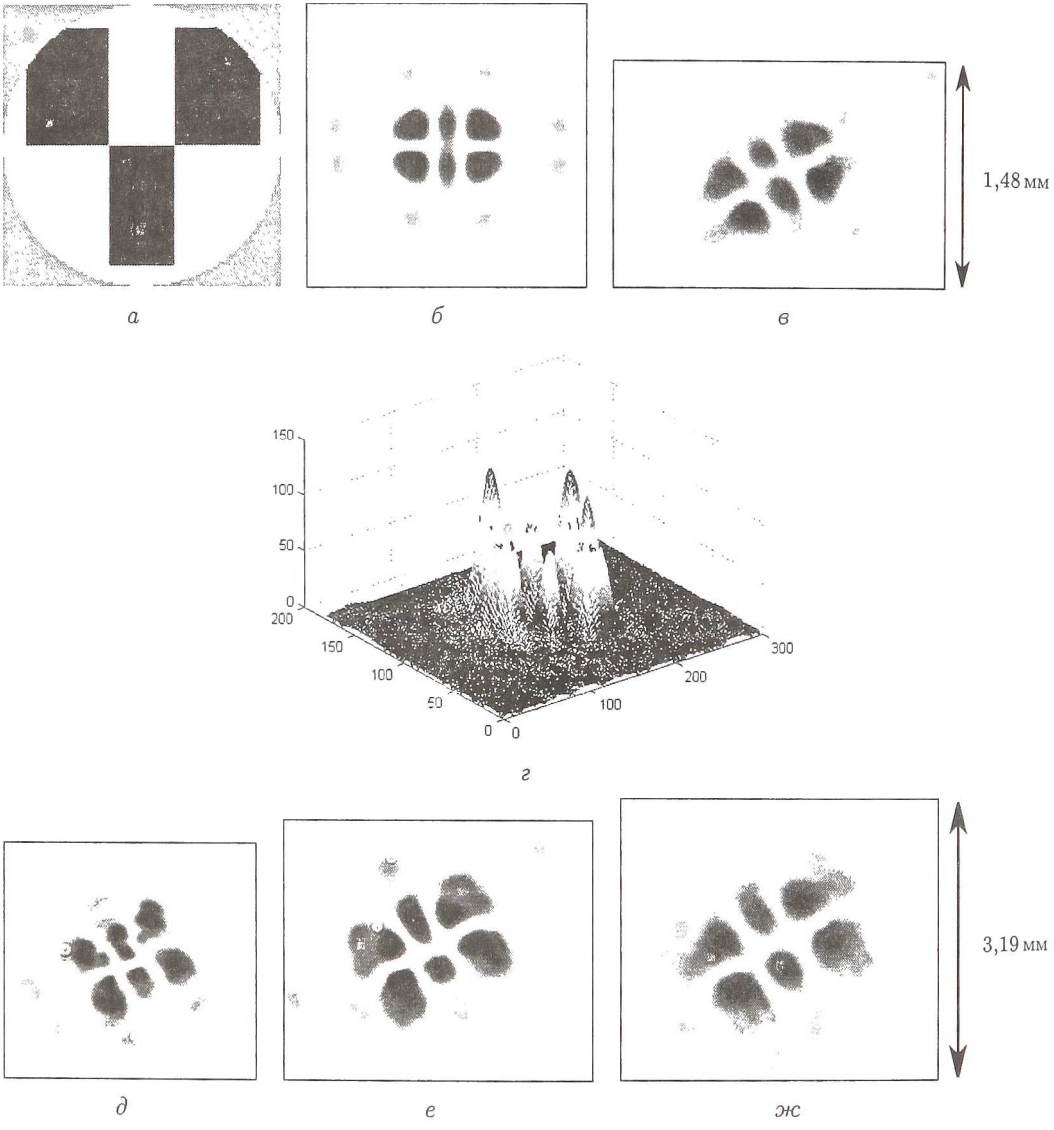


Рис. 7.32. Результаты численного моделирования и эксперимента для моды ГЭ (1,2): фотомаска, ограниченная круглой апертурой освещающего пучка (*а*); рассчитанная картина дифракции в дальней зоне (*б*); экспериментальная картина дифракции в дальней зоне (*в*) и ее трехмерное представление (*г*); экспериментальные распределения интенсивности на расстояниях 260 мм (*д*), 400 мм (*е*) и 580 мм (*ж*) от ДОЭ

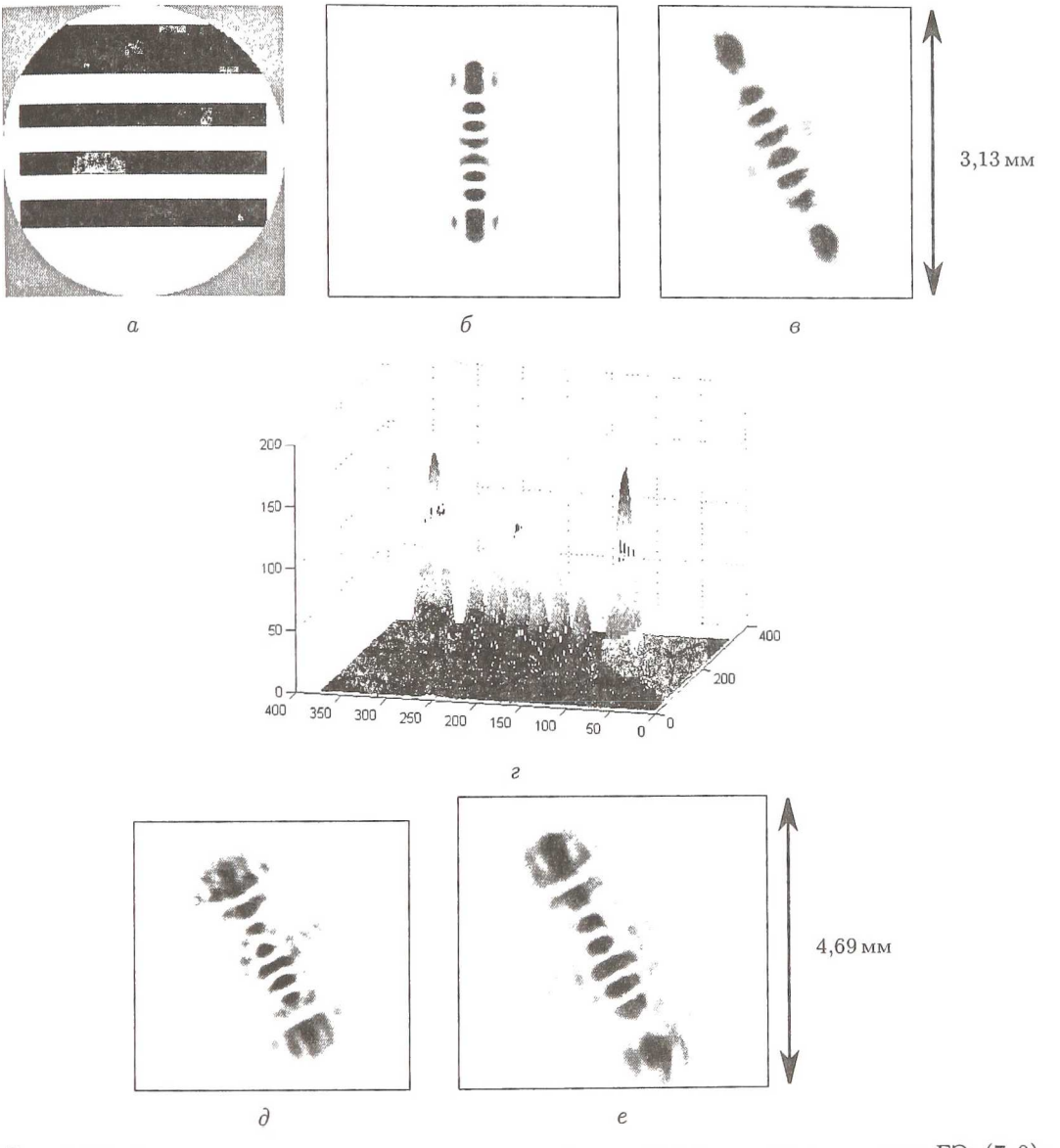


Рис. 7.33. Результаты численного моделирования и эксперимента для моды ГЭ (7,0): фотомаска, ограниченная круглой апертурой освещающего пучка (*а*); рассчитанная картина дифракции в дальней зоне (*б*); экспериментальная картина дифракции в дальней зоне (*в*) и ее трехмерное представление (*г*); экспериментальные распределения интенсивности на расстояниях 360 мм (*д*) и 470 мм (*е*) от ДОЭ

Пример 7.16. На рис. 7.30–7.33 показаны результаты исследования для мод ГЭ с номерами $(1,0)$, $(1,1)$, $(1,2)$ и $(7,0)$, соответственно. На рис. 7.30*а*–7.33*а* показаны амплитудные бинарные фотошаблоны ДОЭ, ограниченные круглыми апертурами с диаметром 1,2 мм. Фотошаблоны рассчитывались в соответствии с формулой (7.182). Однако, апертура была круглой и немного больше оптимального размера, что ухудшило результаты (в основном эффективность), которые могли быть получены в оптимальном случае (см. таблицу 7.6). На рис. 7.30*б*–7.33*б* показаны рассчитанные по формуле (7.190) картины дифракции Фраунгофера на соответствующих фазовых бинарных ДОЭ. На рис. 7.30*в*–7.33*в* показаны экспериментальные картины дифракции Фраунгофера (дальняя зона), которые были зарегистрированы в задней фокальной плоскости линзы с фокусным расстоянием 500 мм, дополнительно введенной в оптическую схему (рис. 7.29, позиция 7). На рис. 7.30*г*–7.33*г* показаны трехмерные представления экспериментальных картин дифракции Фраунгофера (рис. 7.30*в*–7.33*в*). Сравнивая рис. 7.30*б*–7.33*б* и рис. 7.30*в*–7.33*в*, можно убедиться, что имеет место хорошее согласие теории и эксперимента.

На рис. 7.30*д,е,ж*–7.33*д,е,ж* показаны экспериментальные картины дифракции Френеля (ближняя зона) на разных расстояниях от ДОЭ. Видно, что лазерный пучок проявляет модовые свойства — сохраняет свою структуру.

7.10. Самовоспроизведение многомодовых пучков Гаусса–Эрмита

При распространении в однородном пространстве некоторых типов когерентных световых полей, могут наблюдаться эффекты самовоспроизведения, то есть повторения распределения интенсивности в поперечном сечении. Одномодовые световые пучки являются примерами самовоспроизводящихся пучков с периодом равным нулю. В предыдущих разделах получены условия для вращения модовых световых пучков, которые также являются примерами полей с продольной периодичностью (с точностью до масштаба). В [67] получено условие, при котором многомодовый пучок Гаусса–Лагерра будет с точностью до масштаба самовоспроизводиться по интенсивности на некоторых расстояниях.

В данном разделе аналогичное условие получено для многомодовых пучков Гаусса–Эрмита. С помощью численного моделирования показан эффект самовоспроизведения для конкретных модовых пучков.

Рассмотрим многомодовый пучок ГЭ, распространяющийся в свободном пространстве. Комплексная амплитуда такого пучка может быть записана в виде:

$$E(x,y,z) = \sum_{m,n=0}^{N} C_{m,n}\Psi_{m,n}(x,y,z), \tag{7.192}$$

где $\Psi_{m,n}(x,y,z)$ — функции вида (7.147) — (7.150), z — продольная, а (x,y) — поперечные декартовы координаты, $C_{m,n}$ — комплексные коэффициенты.

Поскольку

$$\Phi_{m',n'}^{m,n} = \arg C_{m,n} - \arg C_{m',n'} + [(m-m') + (n-n')]\,\eta(z), \tag{7.193}$$

то выражение для интенсивности света в поперечном сечении пучка ГЭ примет вид:

$$I(x,y,z)=\frac{\omega_0^2}{a^2(z)}\exp\left[-\frac{2(x^2+y^2)}{a^2(z)}\right]\left\{\sum_{m,n=0}^{N}|C_{m,n}|^2\,H_m^2\left(\frac{\sqrt{2}x}{a(z)}\right)H_n^2\left(\frac{\sqrt{2}y}{a(z)}\right)+\right.$$
$$\left.+\sum_{m,n}^{N}\sum_{m',n'}^{N}|C_{mn}C_{m'n'}|\,H_m\left(\frac{\sqrt{2}x}{a(z)}\right)H_{m'}\left(\frac{\sqrt{2}x}{a(z)}\right)H_n\left(\frac{\sqrt{2}y}{a(z)}\right)H_{n'}\left(\frac{\sqrt{2}y}{a(z)}\right)\cos\Phi_{m'n'}^{mn}\right\}, \tag{7.194}$$

Чтобы распределение интенсивности (7.194) на расстоянии z_1 с точностью до масштаба повторилось на расстоянии z_2, нужно, чтобы функции (7.193) в каждом слагаемом в уравнении (7.194) удовлетворяли условию

$$\Phi_{m',n'}^{m,n}(z_2)=\Phi_{m',n'}^{m,n}(z_1)+2\pi l,\quad l=1,2,3,\ldots \tag{7.195}$$

Подставив в уравнение (7.195) функцию (7.193), получим выражение для расстояния z_2:

$$z_2=\frac{z_1+z_0\,\mathrm{tg}\,(2\pi l/p)}{1-(z_1/z_0)\,\mathrm{tg}\,(2\pi l/p)}, \tag{7.196}$$

где

$$p=(m-m')+(n-n'). \tag{7.197}$$

Присутствие в формуле параметра l показывает, что для пары мод с номерами (m,n), (m',n') может существовать несколько расстояний, на которых будет повторяться поперечное распределение интенсивности их суммы, зафиксированное на некотором z:

$$z_{l,p}(z)=\frac{z+z_0\,\mathrm{tg}\,(2\pi l/p)}{1-(z/z_0)\,\mathrm{tg}\,(2\pi l/p)},\quad l=1,2,3,\ldots \tag{7.198}$$

Рассмотрим пример для пары мод с $p=20$ и найдем расстояния, на которых повторится распределение интенсивности, зафиксированное для $z=0$:

$$z_{l,p}(0)=z_{l,p}=z_0\,\mathrm{tg}\,(2\pi l/p),\quad l=1,2,3,\ldots \tag{7.199}$$

В этом случае получается пять повторений с возрастающим периодом $z_{1,20}=0{,}3249z_0$, $z_{2,20}=0{,}7265z_0$, $z_{3,20}=1{,}3764z_0$, $z_{4,20}=3{,}0777z_0$, $z_{5,20}=\infty$, одно из которых повторяется на бесконечности, или фокальной плоскости линзы. Понятно, что все изменения, возникающие на интервале $[z,z_{1,p}(z))$, будут повторяться на последующих периодах $[z_{1,p}(z),z_{l+1,p}(z))$ с замедляющейся с ростом l скоростью.

Далее рассмотрим способы выбора номеров мод для саморепродуцирующегося пучка (7.192), содержащего более чем два слагаемых. Если выбрана пара мод с номерами (m,n), (m',n'), $p_0=(m-m')+(n-n')$, то добавление мод с теми же скоростями, а именно, с номерами (m'',n''), такими, что

$$m''+n''=m+n\quad\text{или}\quad m''+n''=m'+n', \tag{7.200}$$

позволит сформировать пучок самовоспроизводящийся на тех же расстояниях (7.198), что и для первоначальной пары мод. Это обусловлено тем, что дополнительные моды будут давать только два значения $p'=p_0$ и $p'=0$. При $p=0$

Таблица 7.7. Расположение точек самовоспроизведения

Значение $\lvert p_0\rvert$	Точка самовоспроизведения z_{l,p_0}
$\lvert p_0\rvert < 4$	Нет
$\lvert p_0\rvert = 4$	$z_{1,p_0} = \infty$
$4 < \lvert p_0\rvert < 8$	$z_0 < z_{1,p_0} < \infty$
$\lvert p_0\rvert = 8$	$z_{1,p_0} = z_0$, $z_{2,p_0} = \infty$
$8 < \lvert p_0\rvert < 12$	$0 < z_{1,p_0} < z_0 < z_{2,p_0} < \infty$
$\lvert p_0\rvert = 12$ *)	$0 < z_{1,p_0} < z_0 < z_{2,p_0} < \infty$, $z_{3,p_0} = \infty$

в (7.193) пропадает зависимость от z, что характерно для стабильного пучка. Примером такого многомодового пучка с $p_0 = 20$ может служить композиция из 7-мод $(0,1) + (1,0) + (10,11) + (11,10) + (9,12) + (8,13) + (0,20)$.

На рис. 7.34 приведены результаты численного моделирования для 4-х модового пучка ГЭ: $(0,1) + (1,0) + (10,11) + (11,10)$. Все моды входят в состав пучка с одинаковыми весами. При этом использовались следующие параметры: размерность массивов 512×512, область изменения аргументов $x, y \in [-0{,}5\,\text{мм}, 0{,}5\,\text{мм}]$, длина волны $\lambda = 0{,}63\,\text{мкм}$, радиус гауссова пучка $\omega_0 = 0{,}1\,\text{мм}$, конфокальный параметр $z_0 = k\omega_0^2/2 = 49{,}63\,\text{мм}$. Поперечное распределение интенсивности такого пучка, зафиксированное на любом расстоянии внутри интервала $[0, z_{1,20}(z)) = [0, 16{,}24)$ должно повториться еще 4 раза до $z = \infty$. Действительно, для $z = 12\,\text{мм}$ формула (7.198) дает следующие возможные повторения: $z_{1,20}(12) = 30{,}63\,\text{мм}$, $z_{2,20}(12) = 58{,}53\,\text{мм}$, $z_{3,20}(12) = 120{,}69\,\text{мм}$, $z_{4,20}(12) = 634{,}72\,\text{мм}$. На рис. 7.34 показана бинарная фаза (*а*) и интенсивность (*б*) в плоскости $z = 0$ (размер соответствует размеру апертуры 1 мм) для 4-х модового пучка ГЭ: $(0,1) + (1,0) + (10,11) + (11,10)$. На последующих рисунках показаны распределения интенсивности на разных расстояниях: $z = 12\,\text{мм}$ (*в*), $z_{1,20}(0) = 16{,}24\,\text{мм}$ (*г*), $z_{1,20}(12) = 30{,}63\,\text{мм}$ (*д*), $z_{2,20}(0) = 36{,}33\,\text{мм}$ (*е*), $z_{2,20}(12) = 58{,}53\,\text{мм}$ (*ж*) (размер каждого кадра равен 0,97 мм).

Из рис. 7.34 видно, что картины интенсивности при $z = 0$ (*б*), $z = 16{,}24\,\text{мм}$ (*г*) и $z = 36{,}33\,\text{мм}$ (*е*) совпадают с точностью до масштаба. Такое же совпадение наблюдается для расстояний $z = 12\,\text{мм}$ (*в*), $z = 30{,}63\,\text{мм}$ (*д*) и $z = 58{,}53\,\text{мм}$ (*ж*).

Еще одна степень свободы в выборе номеров дополнительных мод появляется, если рассмотреть следующее условие повторения для выражения (7.198):

$$z_{l,p}(z) = z_{l',p'}(z). \tag{7.201}$$

Так, если выбрана пара мод с номерами (m,n), (m',n'), $p_0 = (m-m') + (n-n')$, то для повторения на тех же расстояниях, разность номеров дополнительных мод должна быть следующей:

$$p = |p_0|\,q, \quad q = 0,1,2,\ldots, \tag{7.202}$$

и номера дополнительных мод находятся из следующего соотношения:

$$m'' + n'' = m + n + |p_0|q, \quad q = 0,1,2,\ldots \tag{7.203}$$

Например, если пучок $(1,1)+(5,5)$ повторяет свое начальное (при $z = 0$) поперечное распределение интенсивности на расстоянии (см. (7.199)) $z_{1,8} = z_0\,\mathrm{tg}\,(\pi/4) = z_0$, то и пучок $(1,1) + (0,2) + (5,5) + (4,6) + (9,9) + (8,10)$ повторится на этом расстоянии:

$$z_{1,8} = z_{2,16} = z_0.$$

Интересно, что в зависимости от значения $|p_0|$ в (7.203), можно предсказать расположение повторений начального (при $z = 0$) распределения интенсивности относительно z_0 (см. таблицу 7.7).

Если расстояния z_1 и z_2 в (7.196) заданы, то номера мод ГЭ, которые должны входить в пучок, чтобы он имел подобные поперечные распределения интенсивности на этих расстояниях, будут удовлетворять условию

$$\frac{2\pi l}{p} = \operatorname{arctg}\left[\frac{z_0(z_2 - z_1)}{z_2 z_1 + z_0^2}\right], \quad |p| > 4l. \tag{7.204}$$

Из уравнения (7.201) следует, что расстояния z_1 и z_2 нельзя задавать произвольно.

Кроме расстояний самовоспроизведения распределения интенсивности, зафиксированной на расстоянии z, могут быть интересны расстояния «частичного» периода (то есть части периода) (7.194). Так, выражение

$$z^q_{l,p}(z) = \frac{z + z_0 \operatorname{tg}(2\pi/q \cdot l/p)}{1 - z/z_0 \operatorname{tg}(2\pi/q \cdot l/p)}, \quad l = 1, 2, 3, \ldots \tag{7.205}$$

позволяет вычислять расстояния, соответствующие q-ой части периода саморепродукции. Здесь можно провести аналогию с «частичным» периодом Тальбота для решеток [7]. Частичные расстояния для начального распределения интенсивности (7.199) будут выглядеть следующим образом:

$$z^q_{l,p} = z_0 \operatorname{tg}(2\pi/q \cdot l/p), \quad l = 1, 2, 3, \ldots \tag{7.206}$$

Значение $q = 2$ соответствует «половинному» периоду. В этом случае при нечетных $l = 2s + 1$, $s = 0, 1, 2, \ldots$ в выражении (7.194) вместо добавления перекрестных членов происходит их вычитание: $\cos(x \pm \pi l) = -\cos(x)$, что можно определить как квазиконтрастирование. Аналогичная ситуация наблюдается на половине периода Тальбота для решеток [7], когда сдвиг первоначального изображения решетки выглядит как ее контрастирование. Для частичных периодов имеем следующее выражение для аргументов косинусов в (7.194):

$$\cos\left[x + p \operatorname{arctg}\left(z^q_{l,p} \Big/ z_0\right)\right] = \cos[x + 2\pi l/q], \tag{7.207}$$

где $x = \arg C_{m,n} - \arg C_{m',n'}$.

На рис. 7.35 приведены результаты численного моделирования для 3-х модового пучка ГЭ $(1,1) + (5,5) + (9,9)$ с одинаковыми весами.

Используя выражение (7.206), можно получить следующие расстояния для саморепродукции (при $q = 2, l$ — четные) и квазиконтраста (при $q = 2, l$ — нечетные) начального распределения интенсивности:

$$z^2_{2,8} = z_0 \operatorname{tg}(\pi/4) = z_0 \approx 49{,}63, \quad z^2_{4,8} = z_0 \operatorname{tg}(\pi/2) = \infty,$$
$$z^2_{1,8} = z_0 \operatorname{tg}(\pi/8) \approx 20{,}71, \qquad z^2_{3,8} = z_0 \operatorname{tg}(\pi/8) \approx 120{,}71.$$

Для четвертей периода будут следующие расстояния (при $q = 4$, $l = 2s + 1$, $s = 0, 1, 2, \ldots$):

$$z^4_{1,8} = z_0 \operatorname{tg}(\pi/2 \cdot 1/8) \approx 9{,}87, \quad z^4_{3,8} = z_0 \operatorname{tg}(\pi/2 \cdot 3/8) \approx 33{,}16,$$
$$z^4_{5,8} = z_0 \operatorname{tg}(\pi/2 \ \ 5/8) \approx 74{,}27, \quad z^4_{7,8} = z_0 \operatorname{tg}(\pi/2 \cdot 7/8) \approx 249{,}51.$$

*) Далее картина повторяется с кратным увеличением количества точек самовоспроизведения.

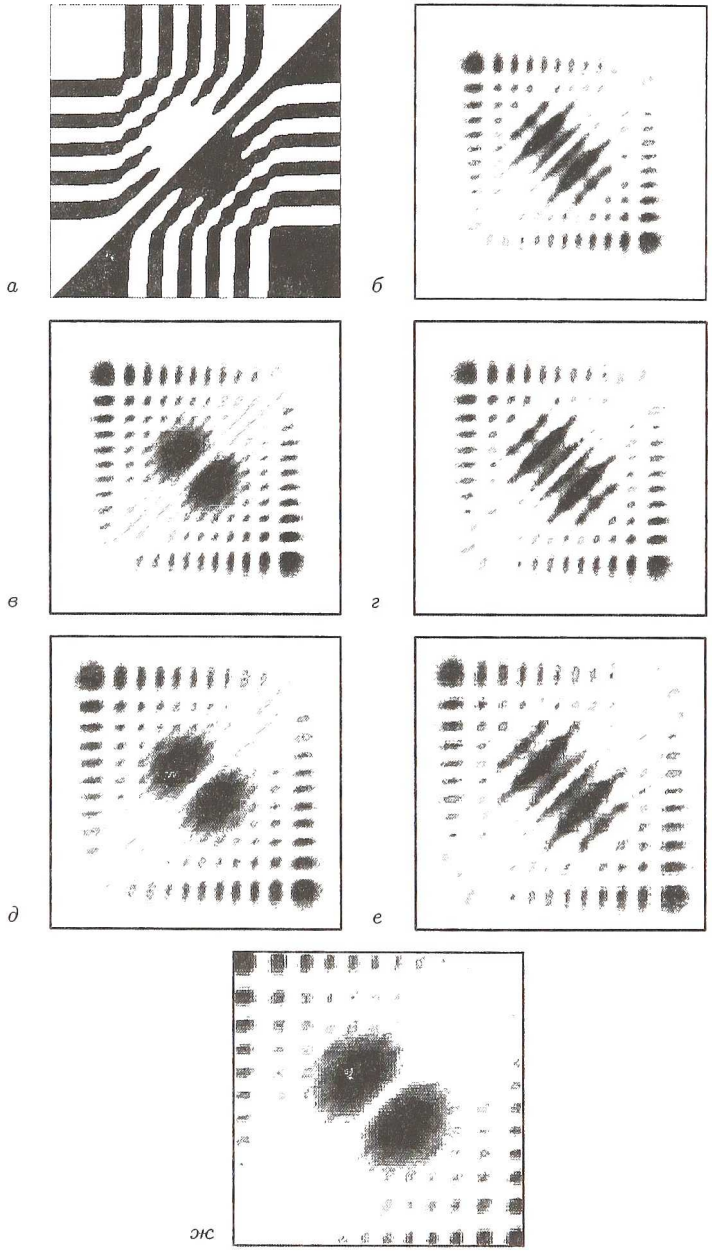


Рис. 7.34. Бинарная фаза ($а$) (белый — 0, черный — π), и интенсивность ($б$) в плоскости $z = 0$, и распределения интенсивности на расстояниях z: 12 мм ($в$), 16,24 мм ($г$), 30,63 мм ($д$), 36,33 мм ($е$), 58,53 мм ($ж$)

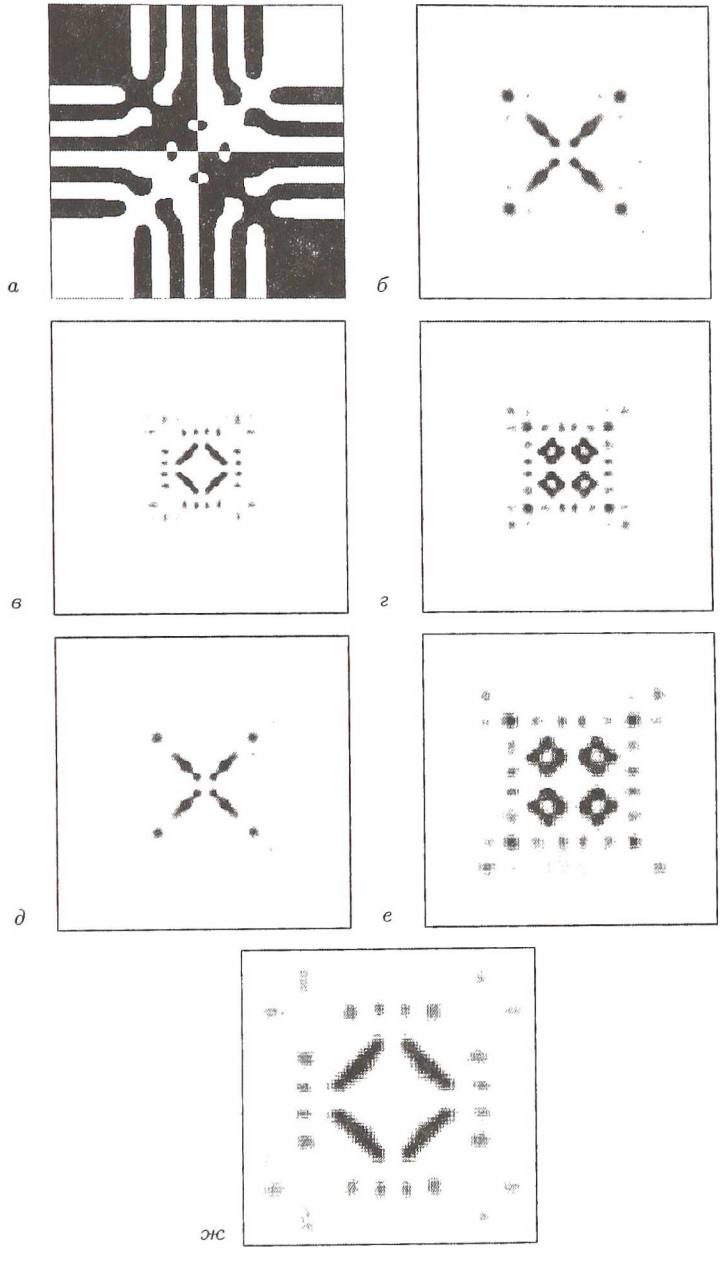


Рис. 7.35. Бинарная фаза (*а*) (белый — 0, черный — π), и интенсивность (*б*) в плоскости $z = 0$, и распределения интенсивности на расстояниях z: 20,71 мм (*в*), 33,16 мм (*г*), 49,63 мм (*д*), 74,27 мм (*е*), 120,71 мм (*ж*)

На рис. 7.35*а* показана бинарная фаза, а на рис. 7.35*б* — интенсивность в плоскости $z = 0$ (размер соответствует размеру апертуры 1 мм). На следующих рисунках показаны распределения интенсивности на разных расстояниях: $z^2_{1,8} = 20{,}71$ мм (*в*), $z^4_{3,8} = 33{,}16$ мм (*г*), $z^2_{2,8} = z_0 = 49{,}63$ мм (*д*), $z^4_{5,8} = 74{,}27$ мм (*е*), $z^2_{3,8} = 120{,}71$ мм (*ж*) (размер каждого кадра равен 1,6 мм).

Из рис. 7.35 видно, что картины интенсивности при $z = 0$ (*б*) и $z = z_0$ (*д*) совпадают с точностью до масштаба. Такое же совпадение наблюдается для расстояний полупериода $z = 20{,}71$ мм (*в*) и $z = 120{,}71$ мм (*ж*). Их совпадение можно предсказать и с помощью (7.205) $z^2_{1,8}(20, 71) = 120{,}71$.

Для четверти периода в выражении (7.207) получим $\cos(x + \pi l/2)$, $l = 2s + 1$, $s = 0, 1, 2, \ldots$, и подобие картины распределения интенсивности на рис. 7.35*г* и 7.35*е* возможно только при $x = \arg C_{m,n} - \arg C_{m',n'} = 0$ (что и было в нашем случае).

7.11. Заключение

Исследования, проведенные в этой главе, позволяют сделать вывод о возможности управления продольно-периодическими свойствами основных типов световых мод свободного пространства: Бесселя, Гаусса–Эрмита, Гаусса–Лагерра. Хотя эти моды описываются амплитудно-фазовыми функциями, методы, рассмотренные в этой главе, предлагают способы получения чисто фазовой (наиболее энергетически выгодной) функции пропускания ДОЭ, формирующих многомодовые световые пучки. Серии тестов с изготовленными ДОЭ показали хорошее согласие теории и эксперимента.

Подкласс световых полей, обладающих коническим спектром плоских волн и названных многомодовыми пучками Бесселя, обладает свойством распространяться в свободном пространстве практически без дифракции. Рассмотренные в главе ДОЭ работают как винтовой аксикон, обеспечивая инвариантные свойства сформированного пучка на расстоянии, пропорциональном радиусу ДОЭ и обратно пропорциональном углу наклона плоских волн пространственного спектра данного поля (или масштабу функции Бесселя). Дифракционное расширение диаметра пучка компенсируется за счет притока энергии из периферийных областей ДОЭ. То есть, с увеличением расстояния z от ДОЭ до рассматриваемой плоскости растет и радиус зоны (кольца) ДОЭ, которая отвечает за формирование светового поля на этой плоскости.

Пучки с продольной периодичностью возникают в случае, когда несколько (не менее двух) конических световых волн с различными углами при вершине конуса распространяются вдоль оптической оси. В результате их интерференции образуется интерференционная картина, амплитуда модуляции которой примерно постоянна на отрезке оптической оси от ДОЭ до расстояния инвариантности, о котором говорилось выше. При этом волна, распространяясь вдоль оптической оси, в своем поперечном сечении испытывает периодические изменения. Минимальный период интерференционной картины будет определяться максимальной разностью углов наклона различных конических волн.

Если многомодовый пучок состоит из композиции бесселевых функций с различными масштабами, но с одинаковыми индексами, а, следовательно, и одинаковыми винтовыми составляющими, то такое поле распространяется без дифракции и без изменения вида как Бесселева мода. Эффект вращения картины в поперечном сечении бесселевого пучка возникает при наличии в пучке составляющих с различными масштабами (угол наклона плоских волн) и индексами.

Условия вращения, полученные в этой главе, позволяют получать вращающиеся с различной скоростью пучки Гаусса–Лагерра как в волокне, так и в свободном пространстве. При распространении в свободном пространстве такие пучки вращаются все более замедляясь, и на расстоянии от ДОЭ до бесконечности совершают небольшое число оборотов (для мод с низшими номерами 1–3 оборота). При распространении в градиентном оптическом волокне вращение происходит с постоянной скоростью и число оборотов — больше (для длины волокна 1 мм около 100 оборотов).

Одномодовые световые пучки являются примерами самовоспроизводящихся пучков с периодом равным нулю. В данной главе показано, что с энергетической точки зрения оптимальным дифракционным элементом, формирующим одномодовый гауссов пучок, является фазовый транспарант, функция пропускания которого равна знаковой функции соответствующего многочлена Лагерра или Эрмита.

Проведенные исследования дают возможность формирования многомодовых пучков Гаусса–Эрмита, самовоспроизводящихся на определенных расстояниях.

Литература к главе 7

1. Talbot H. F. *Phil. Mag. and J. Sci. London* **9** 401 (1836)
2. Lau E. *Annln. Phys.* **6** 417 (1948)
3. Lord Rayleigh *Phil. Mag.* **2**(5) 196 (1881)
4. Winthrop J. T., Worthington C. R. *J. Opt. Soc. Am.* **55** 373 (1965)
5. Leger J. R., Swanson G. J. *Opt. Lett.* **15** 288 (1990)
6. Arrizon V., Ojeda-Castaneda J. *J. Opt. Soc. Am.* **A9**(10) 1801 (1992)
7. Westerholm J., Turunen J., Huttunen J. *J. Opt. Soc. Am.* **A11**(4) 1283 (1994)
8. Montgomery W. D. *J. Opt. Soc. Am.* **57**(6) 772 (1967)
9. Montgomery W. D. *J. Opt. Soc. Am.* **58**(8) 1112 (1968)
10. Slepian D. *Bell System Tech. J.* **43** 3009 (1964)
11. Indebetouw G. *Opt. Acta* **31**(5) 531 (1984)
12. Ceperley P. H. *Am. J. Phys.* **60**(10) 938 (1992)
13. Chavez-Cerda S., McDonald G. S. *New Opt. Commun.* **123** 225 (1996)
14. Paterson C., Smith R. *Opt. Commun.* **124** 131 (1996)
15. Basistiy I. V., Bazhenov V. Yu., Soskin M. S. *Opt. Commun.* **103** 422 (1993)
16. Indebetouw G. *J. Mod. Opt.* **40**(1) 73 (1993)
17. Abramochkin E., Volostnikov V *Opt. Commun.* **102** 336 (1993)
18. Смирнов В. И. *Курс высшей математики* (М.: Наука, т. IV, 1974)
19. Лебедев В. В., Лукьянов Ю. Н., Орлов М. И., Преображенский Н. Г., Соколовский Р. И. *Препринт института теоретической и прикладной механики СО РАН* **16-89** (1989)
20. Durnin J. *J. Opt. Soc. Am.* **4** 651 (1987)
21. Durnin J., Miceli J. J., Eberby J. H. *Opt. Lett.* **13** 79 (1988)
22. Gori F., Guattari G., Padovani C. *Opt. Commun.* **64**(6) 491 (1987)
23. Валиев А. Б., Кривошлыков С. Г. *Квантовая электроника* **16** 1047 (1989)
24. Vasara A., Turunen J., Friberg A. T. *J. Opt. Soc. Am.* **6** 1748 (1989)
25. Унгер Х. Г. *Планарные и волоконные оптические волноводы* (М.: Мир, 1980)
26. Бельский А М *Оптика и спектроскопия* **73**(5) 947 (1992)
27. Kotlyar V. V., Khonina S. N., Soifer V. A. *J. Mod. Opt.* **42**(6) 1231 (1995)
28. Golub M. A., Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Soifer V. A. *J. Mod. Opt.* **39**(6) 1245 (1992)
29. Khonina S. N., Kotlyar V V., Soifer V. A. *Proceedings of SPIE* **1780** 263 (1993)
30. Soifer V. A., Doskolovich L. L., Kazanskiy N. L. *Opt. Eng.* **33**(11) 3611 (1994)
31. Корн Г., Корн Т. *Справочник по математике* (М.: Наука, 1968)

32. Котляр В. В., Сойфер В. А., Хонина С. Н. *Оптика и спектроскопия* **84**(5) 853 (1998)
33. Кузнецов Д. С. *Специальные функции* (М.: Высшая школа, 1962)
34. Fienup J. R. *Appl. Opt.* **21**(15) 2758 (1982)
35. Szwaykowski P *J. Opt. Soc. Am.* **5**(2) 185 (1988)
36. Ananev Yu. Yu. *Opt. Spectrosc.* **64** 722 (1988)
37. Gori F., Guattari G., Padovani C. *Optics Communications* **64**(4) 311 (1987)
38. Turunen J., Vasara A., Friberg A. T. *J. Opt. Soc. Am.* **A8**(2) 282 (1991)
39. Overfelt P. L., Kenney C. S. *J. Opt. Soc. Am.* **A8**(5) 732 (1991)
40. Indebetouw G. *J. Opt. Soc. Am.* **A6**(1) 150 (1988)
41. Turunen J., Vasara A., Friberg A. T. *Appl. Opt.* **27**(19) 3959 (1988)
42. Paterson C., Smith R. *Optics Communications* **124** 123 (1996)
43. Khonina S. N., Kotlyar V. V., Soifer V. A., Shinkaryev M. V., Uspleniev G. V. *Optics Communications* **91**(34) 158 (1992)
44. Andreev N. E., Bychkov S. S., Kotlyar V. V., Margolin L. Ya., Pyatnitskii L. N., Serafimovich P. G. *Quantum Electronics* **23**(2) 130 (1996)
45. Ruschin S. *J. Opt. Soc. Am.* **A11**(12) 3224 (1994)
46. Spektor B., Piestun R., Shamir J. *Opt. Lett.* **21**(7) 456 (1996)
47. Kotlyar V V., Soifer V. A., Khonina S. N. *J. Modern Opt.* **44**(7) 1409 (1997)
48. EOS Topical meeting on Diffractive Optics (Savonlinna, 1997, c. 34)
49. Goodman J. W., Silvestri A. M. *IBM J. Res. Devel.* **9** 478 (1970)
50. Soifer V. A., Golub M. A. *Laser beam mode selection by computer-generated holograms.* (Boca Raton: CRC Press U.S 1994)
51. Голуб М. А., Павельев В. С., Сойфер В. А. *Компьютерная оптика* **14–15**(2) 85 (1995)
52. Kotlyar V. V., Nikolsky I. V., Soifer V. A. *Optik* **98**(1) 26 (1994)
53. Soifer V. A., Kotlyar V. V., Doskolovich L. L. *Iterative methods for diffractive optical elements computation.* (London: Taylor & Francis, 1997)
54. Прудников А. П., Брычков Ю. А., Марычев О. И. *Интегралы и ряды. Специальные функции.* (М.: Наука, 1983)
55. Хонина С. Н., Котляр В. В., Сойфер В. А. *Компьютерная оптика* **17** 25 (1997)
56. Ярив А. *Оптическая электроника* (М.: Советское радио, 1986)
57. Adams M. J. *An introduction to optical waveguides* (J. Wiley & Sons, Chichster, 1981)
58. Schechner Y. V., Piestun R., Shamir J. *Physical Review* **E54**(1) 50 (1996)
59. Котляр В. В., Сойфер В. А., Хонина С. Н. *Письма в ЖТФ* **23**(17) 1 (1997)
60. Виноградова М. Б., Руденко О. В., Сухоруков А. П. *Теория волн* (М.: Наука, 1979)
61. Ярив А. *Введение в оптическую электронику* (М.: Высшая школа, 1983)
62. Fedotovsky A., Lehovec K. *Appl. Opt.* **13**(12) 2919 (1974)
63. Mait J. *J. Opt. Soc. Am.* **A7**(8) 1514 (1990)
64. Котляр В. В., Сойфер В. А., Хонина С. Н. *Компьютерная оптика* **17** 31 (1997)
65. Duparre M., Pavelyev V., Luedge B., Kley B., Kowarschik R., Soifer V *Proceedings of SPIE* **3134** 357 (1997)
66. Kotlyar V. V., Khonina S. N., Soifer V. A. *Optik* **108**(1) 20 (1998)
67. Piestun R., Schechner Y. V., Shamir J. *Opt. Lett.* **22**(4) 200 (1997)

ГЛАВА 8
КОРРЕКЦИЯ ВОЛНОВЫХ ФРОНТОВ

8.1. Проблема создания волновых фронтов

Для решения широкого класса задач астрономии, оптического приборостроения и оптической обработки информации необходимы высококачественные асферические поверхности, например, параболическая оптика для больших телескопов. Получение асферических поверхностей высокого качества зависит от разработки эффективных методов их аттестации [1–4]. Методы аттестации асферических поверхностей основаны на создании эталонных волновых фронтов, форма которых соответствует контролируемой поверхности. Сферические и плоские волновые фронты естественным образом формируются в классических оптических системах, состоящих из линз, призм, сферических зеркал, а также пробными стеклами. Создание эталонов волновых фронтов более сложной формы наталкивается на значительные трудности.

В классической оптике проблема создания волновых фронтов требуемой формы решается с помощью компенсационных объективов [1, 2], позволяющих получать волновые фронты в виде поверхностей вращения второго порядка. Однако создание компенсационного объектива является уникальной задачей для каждого типа волнового фронта. Асферические волновые фронты высших порядков, а также фронты без круговой симметрии, по-видимому, вообще не могут быть сформированы компенсационными объективами приемлемой сложности. Поэтому для решения проблем формирования сложных волновых фронтов в 70-е годы активно использовались методы цифровой голографии [8–15]. В [3–7] для создания осесимметричных асферических волновых фронтов использовались специальные кольцевые дифракционные решетки с дифракционной эффективностью не выше 40%. В [8–15] показана возможность создания асферических фронтов с помощью бинарных синтезированных на ЭВМ голограмм. Однако при этом в требуемый волновой фронт дифрагируется лишь незначительная часть энергии освещающего пучка, соответствующая первому порядку дифракции, большая доля элементов разрешения голограммы идет на передачу несущей пространственной частоты, а относительное отверстие волнового фронта ограничено наложением высших дифракционных порядков. Настоящая глава посвящена анализу методов компьютерной оптики, используемых для создания волновых фронтов сложной формы [16–24]. Созданию этого направления послужила публикация в «Докладах Академии наук СССР» в 1980 году статьи М. А. Голуба, Е .С. Живописцева, С. В. Карпеева, А. М. Прохорова, И. Н. Сисакяна и В. А. Сойфера [16]. При этом в настоящей работе мы не касаемся специфических проблем коррекции аберраций оптических систем с применением ДОЭ и граданов, подробно рассмотренных в монографиях [25, 26].

8.2. Оптические схемы с ДОЭ для анализа асферических поверхностей

Рассмотрим некоторые оптические схемы контроля оптических поверхностей с дифракционными оптическими элементами, синтезированными на компьютере [3–30]. Функциональное назначение ДОЭ состоит либо в создании эталонного асферического волнового фронта из плоского, либо в преобразовании (компенсации) одного волнового фронта в другой. В последнем случае ДОЭ называется оптическим компенсатором. В оптических контрольных схемах происходит анализ волнового фронта от исследуемого асферического зеркала или линзы. В интерференционных методах такой анализ осуществляется путем сравнения исследуемого фронта с эталонным. В теневом методе анализируется теневая картина волнового фронта, отраженного (или прошедшего) от контролируемого элемента и преобразованного оптическим компенсатором. Теневой метод с использованием ножа Фуко более прост при реализации, но обладает низким отношением сигнал/шум, так как нулевая пространственная частота, наиболее энергетическая, не используется для анализа.

Рассмотрим оптическую схему интерферометра со сферическими пучками, показанную на рис. 8.1. В этой схеме луч света от лазера Л преобразуется микрообъективом МО в расходящийся сферический пучок света. Этот пучок с помощью полупрозрачного зеркала РП (расщепитель пучка) расщепляется на два пучка. Отраженный пучок направляется на эталонное сферическое зеркало ЭЗ, а прошедший пучок, проходя через ДОЭ, попадает на контролируемое асферическое зеркало КАЗ. Компенсатор ДОЭ в этом случае должен быть рассчитан так, чтобы он преобразовывал сходящийся асферический волновой фронт, отраженный от идеального асферического зеркала, в сходящийся сферический фронт. Такой ДОЭ будет работать следующим образом. Расходящийся сферический фронт от МО преобразуется ДОЭ в расходящийся асферический фронт, лучи которого будут падать на контролируемое асферическое зеркало КАЗ по нормали и отражаться от него тоже по нормали (если оно идеальное). На обратном пути отраженный от КАЗ сходящийся асферический волновой фронт преобразуется ДОЭ в сходящийся сферический. В опорной ветви интерферометра для формирования эталонного сходящегося сферического волнового фронта используется эталонное сферическое зеркало ЭЗ. Два когерентных между собой сходящихся сферических фронта пересекаются в плоскости регистрации ПР и создают интерференционную картину, которая далее анализируется. Таким образом, проблема изготовления асферического зеркала сводится к более простой проблеме изготовления сферического эталонного зеркала. Однако высокие требования к устойчивости оптических элементов между собой и по отношению к КАЗ при микровибрациях не позволяют использовать этот интерферометр в процессе технологического контроля зеркал. Если расположить компенсатор ДОЭ в опорной ветви интерферометра [28], то он будет выполнять функцию, в некотором смысле, обратную схеме рис. 8.1. Он будет преобразовывать расходящийся асферический фронт в расходящийся сферический, а сходящийся сферический фронт в сходящийся асферический.

Для контроля асферики применяются также интерферометры Тваймана–Грина и модифицированный интерферометр Маха–Цендера [5, 8, 10–12, 27, 28]. Компенсатор работает в них так же, как и в рассмотренной выше схеме на рис. 8.1. Однако, интерференционная картина формируется не сферическими, а плоскими волнами, что существенно упрощает юстировку схемы и позволяет заменить эталонное сферическое зеркало на плоское. Но преобразование сферического пучка в плоский требует дополнительного объектива, вносящего и дополнительные погрешности.

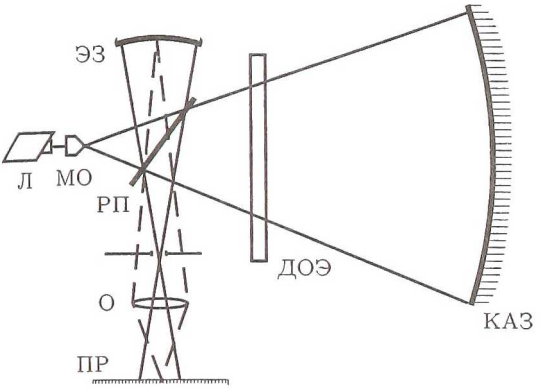


Рис. 8.1. Оптическая схема интерферометра с ДОЭ со сферическими пучками

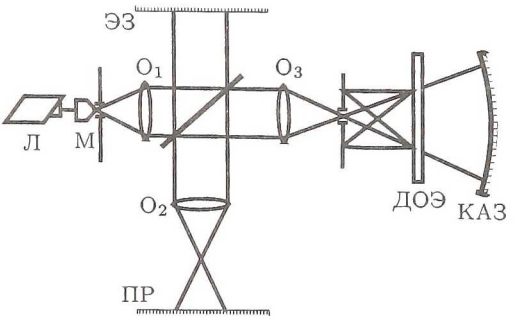


Рис. 8.2. Оптическая схема интерферометра Тваймана–Грина с компенсатором в объектной ветви

На рис 8.2 показана оптическая схема интерферометра Тваймана–Грина с компенсатором в объектной ветви. Интерферометр работает следующим образом. Луч света от лазера Л расширяется микрообъективом М и коллимируется объективом O_1. Далее плоский пучок света расщепляется полупрозрачным зеркалом на два пучка. Опорный пучок отражается от ЭЗ. Объектный пучок объективом O_3 преобразуется в сходящийся сферический пучок, который после прохождения диафрагмы, расположенной в фокальной плоскости, преобразуется в расходящийся сферический пучок. Этот пучок преобразуется ДОЭ в расходящийся асферический пучок, лучи которого падают и отражаются по нормали от КАЗ. На обратном пути два плоских пучка (объектный и опорный) пересекаются в ПР и формируют интерференционную картину в виде набора эквидистантных полос. Объектив O_2 строит изображение поверхности зеркал. Если КАЗ имеет дефекты геометрии, то это проявится в искривлении интерференционных полос. Компенсатор ДОЭ в интерферометре на рис. 8.2 может быть расположен в опорной ветви. При этом он должен преобразовывать плоский волновой фронт в некоторый «квазиплоский», который соответствует асферическому фронту таким же образом, как плоский пучок соответствует сферическому.

Известны оптические схемы интерферометров, в которых ДОЭ располагается на выходе [29]. На рис. 8.3 показана схема интерферометра типа Маха–Цендера с компенсатором на выходе. Опорная ветвь этого интерферометра включает два полупрозрачных зеркала для расщепления пучка и два плоских эталонных «глухих» зеркала для поворота излучения. В этом случае ДОЭ выполняет следующую функцию. Падающий на него плоский опорный пучок ДОЭ преобразует в «квазиплоский» пучок, а падающий на него «квазиплоский» объектный пучок, пришедший от КАЗ, ДОЭ преобразует в плоский пучок. Так как через компенсатор проходят оба пучка, то это приводит к взаимной компенсации технологических погрешностей изготовления ДОЭ и повышению точности контроля.

Оригинальное использование синтезированного компенсатора предложено в работе [30]. На рис. 8.4 показана оптическая схема интерферометра с компенсатором на входе. Луч света от лазера Л, совмещенного с коллиматором, падает на ДОЭ и преобразуется в сходящийся асферический пучок, лучи которого падают и отражаются от КАЗ по нормали. Оптическая схема рис. 8.4 объединяет достоинства схем рис. 8.2 и 8.3, так как погрешности, допущенные при изготовлении ДОЭ, взаимно компенсируются из-за того, что ДОЭ порождает оба пучка: и опорный, и объектный.

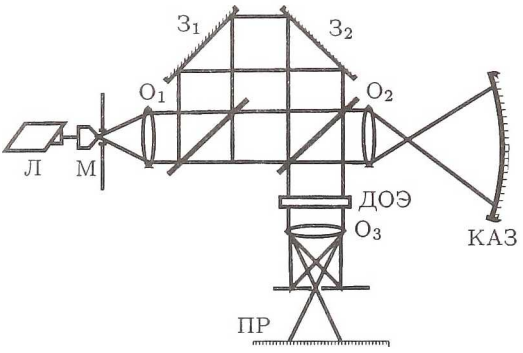


Рис. 8.3. Оптическая схема интерферометра Маха–Цендера с компенсатором на выходе

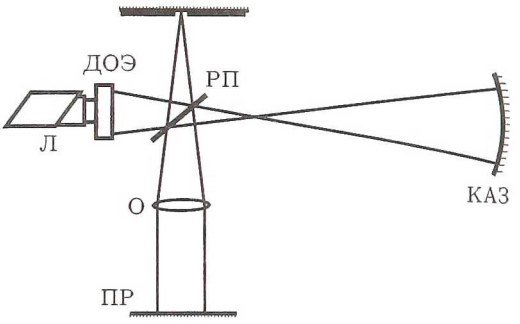


Рис. 8.4. Оптическая схема интерферометра с компенсатором на входе

Кроме того, число элементов в оптической схеме минимально по сравнению с рассмотренными выше, и не требуется плоское эталонное зеркало, поскольку отражение от опорного зеркала происходит в малой области. Диафрагма вблизи опорного зеркала пропускает только нулевую пространственную частоту. В отличие от всех вышерассмотренных типов преобразований (сферический в асферический, плоский в квазиплоский), в этой схеме компенсатор должен не только вносить заданные аберрации в волновой фронт, но и иметь некоторую положительную оптическую силу. Таким образом, ДОЭ на рис. 8.4 является аналогом асферической линзы, в то время как в схемах на рис. 8.1–8.3 — аналогом пластинки с заданными аберрациями. Из-за значительной разности хода лучей между плоским и асферическим фронтами, соответствующий компенсатор должен иметь несколько тысяч структурных зон (колец), изготовление которых является трудной технологической задачей.

Рассмотренные выше оптические схемы интерферометров, применимы для бесконтактного контроля широкого класса оптических поверхностей зеркал и линз. При этом вся специфика контролируемой поверхности учитывается при расчете компенсатора и не требуется изменение самой схемы. Контролироваться могут как вращательно-симметричные, так и несимметричные поверхности. Дополнительные варианты использования ДОЭ в контрольных оптических схемах приведены в [16].

8.3. Расчет плоского компенсатора

Формирование эталонного волнового фронта σ будем производить из светового пучка E согласно схеме (рис. 8.5), основным элементом которой является компенсатор K с зонированным прозрачным микрорельефом, выполненным по технологии компьютерной оптики на плоской подложке. Задачей компенсатора является создание постоянной фазы на поверхности σ, вершина которого отстоит на расстояние l от K (см. рис. 8.5). Зададим σ уравнением

$$H = f(\mathbf{x}), \quad \mathbf{x} \in D, \tag{8.1}$$

с гладкой (непрерывно дифференцируемой) функцией f, $f(0) = 0$. Здесь $\mathbf{x} = (x, y)$ — декартовы координаты проекции точки $Q \in \sigma$ на плоскость Π, проходящую через вершину D_0 параллельно компенсатору; D — область значений $\mathbf{x}$, соответствующая формируемому сегменту фронта σ. H — алгебраическое расстояние от $Q \in \sigma$ до

П, выбираемое отрицательным, если П лежит за Q по ходу лучей. Ось $z \perp K$ выходит из центра O по ходу лучей. Освещающий пучок E зададим длиной волны λ, распределением интенсивности $I_0(\mathbf{u})$ и эйконалом $\psi_0(\mathbf{u})$ в рабочей области G плоскости компенсатора K, $\mathbf{u} = (u, v)$ — декартовы координаты в плоскости компенсатора. Чисто фазовый плоский оптический элемент-компенсатор K будем характеризовать фазовой функцией $\varphi(\mathbf{u})$ [16, 17]. При создании компенсатора фазовая функция приводится к интервалу $[0, 2\pi(m)]$ и реализуется в виде фазового микрорельефа ($m = \text{const}$ — натуральное число; для киноформов $m = 1$).

Геометрооптическое уравнение компенсатора имеет вид

$$k\psi_0(\mathbf{u}) + \varphi(\mathbf{u}) + kL(\mathbf{x}) = \varphi_0 \equiv \text{const}, \quad k = \frac{2\pi}{\lambda}, \tag{8.2}$$

где $L(\mathbf{x})$ — расстояние от компенсатора до фронта σ вдоль луча с направлением

$$\mathbf{N} = \frac{(-\nabla_x f, 1)}{\sqrt{1 + (\nabla_x f)^2}}, \tag{8.3}$$

ортогональным поверхности (8.1), $L > 0$, если луч проходит σ после K . В силу (8.3)

$$L = [L + f]\sqrt{1 + (\nabla_x f)^2}, \tag{8.4}$$

а координаты точки $\mathbf{u}$ на K и точки $\mathbf{x}$ на σ связаны уравнениями соответствия:

$$\mathbf{u} = \mathbf{x} + [l + f(\mathbf{x})]\,\nabla_x f(\mathbf{x}). \tag{8.5}$$

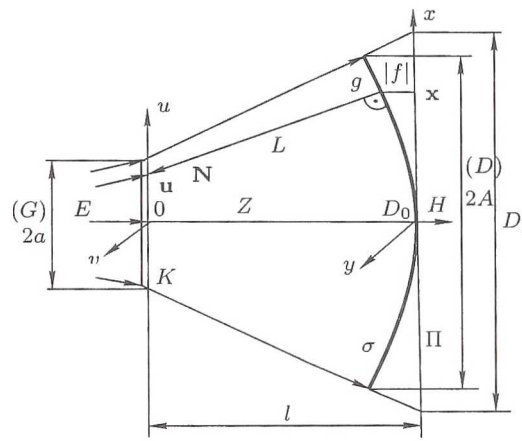


Рис. 8.5. Геометрия преобразования волнового фронта компенсатором

Области G и D должны быть также связаны соответствием, как $\mathbf{u}$ и $\mathbf{x}$; фазовая функция компенсатора согласно (8.2) и (8.4) определяется по формуле

$$\varphi(\mathbf{u}) = \varphi_0 - k\left\{[l + f(\mathbf{x})]\sqrt{1 + [\nabla_x f(\mathbf{x})]^2} + \psi_0(\mathbf{u})\right\}, \tag{8.6}$$

координату $\mathbf{x}$ следует находить путем решения (8.5) при заданной координате $\mathbf{u}$.

8.4. Спектральные свойства компенсаторов

Синтезированный по технологии компьютерной оптики компенсатор представляет собой криволинейную фазовую дифракционную решетку со сложными топологией и профилем штрихов.

Зафиксируем геометрию σ и изменим рабочую длину волны с расчетного значения λ_0 до значения $\lambda \neq \lambda_0$. Поскольку изменение эйконала компенсатором не зависит от λ, то фазовая функция $\varphi(\lambda)$ компенсатора, необходимая для формирования σ с длиной волны λ, связана с фазовой функцией φ для λ_0 соотношением $\varphi(\lambda) = (\lambda_0/\lambda)\,\varphi$, (пространственный аргумент $\mathbf{u}$ в этом параграфе опускаем).

При освещении синтезированного ДОЭ-компенсатора излучением с $\lambda \neq \lambda_0$ сформированный волновой фронт отличается от заданного. Погрешности обусловлены двумя основными факторами:

— изменением оптической разности хода через компенсатор вследствие дисперсии коэффициента преломления;

— нарушением условия соответствия фазы и синфазности зон ДОЭ.

Погрешности первого типа родственны хроматическим аберрациям обычных пропускающих оптических элементов. Если ДОЭ имеет одну зону, то и для него дисперсия показателя преломления является единственной причиной аберраций.

Второй фактор является специфичным именно для ДОЭ, в которых максимальное значение оптической разности хода должно соответствовать фазе $2\pi m$ для расчетной длины волны λ_0 ($m = 1, 2, \dots$). При изменении длины волны, даже при пренебрежимо малой дисперсии показателя преломления, условие соответствия фазы нарушается. Очевидно, что синфазность зон будет снова появляться, если максимальное значение разности фаз для используемой длины волны λ близко либо к $2\pi m$, либо к $4\pi m$, $6\pi m$ и т. д. Количественное рассмотрение влияния несоответствия фазы на функционирование киноформа проведено в [31] для $m = 1$. Однако качество работы оценивалось там по точности формирования заданной интенсивности в плоскости объекта. При исследовании же компенсатора существенна точность формирования фазового распределения, определяющего форму волнового фронта.

Поскольку изменение эйконала φ/k, обеспечиваемое компенсатором согласно (8.6) и (8.5), не зависит от длины волны, а фазовая функция приведена к интервалу $[0, 2\pi m)$ для λ_0, то фазовый сдвиг при $\lambda \neq \lambda_0$ равен в каждой точке $\mathbf{u}$

$$\Phi(\lambda) = c \bmod_{2\pi m} \varphi, \tag{8.7}$$

где

$$c = \lambda_0/\lambda \cdot \Delta n(\lambda)/\Delta n(\lambda_0), \quad \Delta n(\lambda) = n(\lambda) - 1, \tag{8.8}$$

$n(\lambda)$ — показатель преломления материала микрорельефа зон компенсатора; $\bmod_{2\pi m} \varphi$ — фаза φ по модулю $2\pi m$. Функция комплексного пропускания компенсатора

$$T(\lambda) = \exp(i\Phi(\lambda)), \tag{8.9}$$

периодична по φ с периодом $2\pi m$ и, следовательно, может быть разложена в ряд Фурье:

$$T(\lambda) = \sum_{l=-\infty}^{\infty} a_l \exp\left(i\frac{l}{m}\varphi\right), \tag{8.10}$$

$$a_l = \operatorname{sinc}(l - mc)(-1)^l \exp(i\pi mc), \tag{8.11}$$

соответствующий различным дифракционным порядкам [25, 32]. Причем согласно равенству Парсеваля можно записать

$$\sum_{l=-\infty}^{\infty} |a_l|^2 = 1. \tag{8.12}$$

Дифракционный порядок с номером l (8.10) обеспечивает фазовую функцию

$$\Phi^{(l)} = \frac{l}{m}\varphi = \frac{l\lambda}{m\lambda_0}\varphi(\lambda). \tag{8.13}$$

Наиболее интенсивным, согласно (8.11), является порядок с номером $l = \nu$, ближайшим к mc:

$$mc \equiv m\lambda_0/\lambda \cdot \Delta n(\lambda)/\Delta n(\lambda_0) = \nu \in \delta, \tag{8.14}$$

где $\delta \in [1/2, 1/2]$, причем ν зависит от λ. В порядке ν действует фазовая функция

$$\Phi(\nu) = (\nu/m)\varphi = \mu\varphi(\lambda), \tag{8.15}$$

где

$$\mu = \frac{\nu\lambda}{m\lambda_0} = \frac{\Delta n(\lambda)}{\Delta n(\lambda_0)} \cdot \frac{1}{1+\delta/\nu}. \tag{8.16}$$

Заметим, что при любых λ, (любых δ, ν)

$$\mu \in [2/3 \cdot \Delta n(\lambda)/\Delta n(\lambda_0), 2\Delta n(\lambda)/\Delta n(\lambda_0)]\,. \tag{8.17}$$

Таким образом, в самом интенсивном ν-м порядке, при измененной по сравнению с расчетной длине волны λ, вместо $\varphi(\lambda)$ действует линейно искаженное в μ раз фазовое пропускание, дающее деформированную форму волнового фронта.

8.5. Характеристика точности эталонного волнового фронта

В силу несовершенства технологии формирования микрорельефа, наличия дифракции и рассеяния света в среде компенсатора, ограничения числа уровней градаций фазы и разрешения по поверхности компенсатора, вместо требуемой фазовой функции φ (8.6) реализуется фазовая функция $\hat{\varphi}$. Соответственно вместо эталонного волнового фронта σ формируется волновая поверхность $\hat{\sigma}$ с некоторыми искажениями формы по сравнению с σ, определяющими качество $\hat{\sigma}$. Ниже формируются удобные при работе с ДОЭ количественные характеристики отличия $\hat{\sigma}$ от σ как в каждой точке, так и в целом. Хотя компенсатор для фронта σ рассчитывался методами геометрической оптики, волновая поверхность формируется дифракционно и может не быть геометрооптическим фронтом. Механизм формирования $\hat{\sigma}$ описывается в общем случае суперпозицией многих дифракционных порядков [25, 32].

В данной работе предлагается оценивать качество компенсатора прямым сравнением световых полей, соответствующих $\hat{\sigma}$ и σ, в виртуальном двухлучевом интерферометре, выставленном на полосы бесконечной ширины с разностью фаз π в плечах.

Введем обозначения w, I, ψ для комплексной амплитуды, интенсивности и эйконала эталонного светового поля, соответствующего волновому фронту σ, а также аналогичные обозначения $\hat{w}$, $\hat{I}$, $\hat{\psi}$, соответствующие $\hat{\sigma}$:

$$I = |w|^2\,, \quad \psi = \frac{1}{k}\arg w;$$
$$\hat{I} = |\hat{w}|^2\,, \quad \hat{\psi} = \frac{1}{k}\arg \hat{w}.$$

При заведении w и $\hat{w}$ в плечи виртуального интерферометра, сформируется разностное световое поле с комплексной амплитудой $\hat{w} - w$ и интенсивностью

$$i = |\hat{w} - w|^2 . \tag{8.18}$$

В качестве аргументов для w, I, $\hat{w}$, $\hat{I}$, $\hat{w} - w$, i будем использовать координаты точки в декартовой системе с центром 0 (см. рис. 8.5). При этом точки плоскости компенсатора K имеют координаты $(u, v, 0) = (\mathbf{u}, 0)$, точки волнового фронта σ — координаты $(x, y, l+f) = (\mathbf{x}, l+f(\mathbf{x}))$, а точки плоскости Π — координаты $(x_\Pi, y_\Pi, l) = (\mathbf{x}_\Pi, l)$. Причем по условию:

$$I(\mathbf{u}, 0) = I_0(\mathbf{u}), \quad \psi(\mathbf{u}, 0) = \psi_0(\mathbf{u}) + \frac{1}{k}\varphi(\mathbf{u}), \tag{8.19}$$

$$\hat{I}(\mathbf{u}, 0) = I_0(\mathbf{u}), \quad \hat{\psi}(\mathbf{u}, 0) = \psi_0(\mathbf{u}) + \frac{1}{k}\hat{\varphi}(\mathbf{u}), \tag{8.20}$$

$$i(\mathbf{u}, 0) = I(\mathbf{u}) \cdot 4 \sin^2 \left[\frac{\hat{\varphi}(\mathbf{u}) - \varphi(\mathbf{u})}{2} \right]. \tag{8.21}$$

Поскольку $i(\mathbf{x}_\Pi, l)$ обращаются в ноль при $\hat{w} = w$, то нормированная величина

$$\aleph(\mathbf{x}_\Pi) = \frac{i(\mathbf{x}_\Pi, l)}{I(\mathbf{x}_\Pi, l)} \in [0, 1] , \tag{8.22}$$

может служить характеристикой отличия $\hat{\sigma}$ от σ в каждой точке сегмента D волнового фронта. Для построения усредненных характеристик отличия $\hat{\sigma}$ от σ в целом могут использоваться интегральные световые потоки разностного поля $\hat{w} - w$ и поля $\hat{w}$, проходящие через сегмент D фронта σ:

$$\Phi_D = \int\limits_{D_\Pi} i(\mathbf{x}_\Pi, l)\, \mathrm{d}^2\mathbf{x}_\Pi, \tag{8.23}$$

$$E_D = \int\limits_{D_\Pi} I(\mathbf{x}_\Pi, l)\, \mathrm{d}^2\mathbf{x}_\Pi, \tag{8.24}$$

а также соответствующие полные световые потоки

$$\Phi = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} i(\mathbf{x}_\Pi, l)\, \mathrm{d}^2\mathbf{x}_\Pi, \tag{8.25}$$

$$E = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \hat{I}(\mathbf{x}_\Pi, l)\, \mathrm{d}^2\mathbf{x}_\Pi = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} I(\mathbf{x}_\Pi, l)\, \mathrm{d}^2\mathbf{x}_\Pi, \tag{8.26}$$

(D_Π — проекция D на плоскость Π, осуществляемая лучами).

Заметим, что комплексная амплитуда w близка к нулю вне области D_Π сегмента σ. Поэтому $i(\mathbf{x}_\Pi, l) = \hat{I}(\mathbf{x}_\Pi, l)$ при $\mathbf{x}_\Pi \bar{\in} D_\Pi$ и

$$\Phi = \Phi_D + E_{\bar{D}}, \tag{8.27}$$

где

$$E = E_D + E_{\bar{D}}, \tag{8.28}$$

а величина $E_{\bar{D}}$ дает долю светового потока E, проходящую вне D_Π и соответствующую высшим дифракционным порядкам. Усредненными характеристиками качества формирования эталонного волнового фронта σ могут служить относительные погрешности

$$\eta = \frac{\Phi}{E}, \quad \eta_D = \frac{\Phi_D}{E_D}, \tag{8.29}$$

подсчитываемые по формулам (8.23), (8.25), (8.26), (8.29) и изменяющиеся в интервале $[0, 1]$.

Погрешность η_D характеризует ошибку формирования сегмента D фронта σ. Погрешность η дополнительно включает долю энергии освещающего пучка, следующую в высшие дифракционные порядки вокруг σ.

Введем также понятия уклонений волнового фронта. Для этого заметим, что в частном случае гладкой геометрооптической поверхности $\hat{\sigma}$, мало отличающейся от σ, имеют место соотношения

$$\left|\hat{I}(\mathbf{x}_\Pi, l) - I(\mathbf{x}_\Pi, l)\right| \ll I(\mathbf{x}_\Pi, l), \tag{8.30}$$

$$i(\mathbf{x}_\Pi, l) = \hat{I}(\mathbf{x}_\Pi, l) \cdot 4\sin^2\left[\frac{\pi}{\lambda}\varepsilon(\mathbf{x}_\Pi)\right], \tag{8.31}$$

где величина

$$\varepsilon(\mathbf{x}_\Pi) = \hat{\psi}(\mathbf{x}_\Pi, l) - \psi(\mathbf{x}_\Pi, l), \tag{8.32}$$

представляет собой уклонение $\hat{\sigma}$ от σ, отсчитанное по нормали к σ, проходящей через точку $(\mathbf{x}_\Pi, l)$ (см. рис. 8.6). В общем случае для волновой поверхности $\hat{\sigma}$ формально введем понятия уклонения $\varepsilon(\mathbf{x}_\Pi)$ $\hat{\sigma}$ от σ формулами (8.31), (8.18) в каждой точке $\mathbf{x}_\Pi$. Понятие среднего уклонения $\bar{\varepsilon}_D$ $\hat{\sigma}$ от σ в пределах сегмента D введем по аналогии, заменяя в (8.31) интенсивности на соответствующие световые потоки:

$$\Phi_D = E_D \cdot 4\sin^2\left(\frac{\pi\bar{\varepsilon}_D}{\lambda}\right) \tag{8.33}$$

При геометрооптической поверхности $\hat{\sigma}$, мало отличающейся от σ, $\bar{\varepsilon}_D$ представляет среднеквадратичное нормальное уклонение $\hat{\sigma}$ от σ в пределах сегмента D.

В силу (8.27), (8.28) и (8.33) имеет место оценка

$$\bar{\varepsilon}_D \leqslant \bar{\varepsilon}, \tag{8.34}$$

где $\bar{\varepsilon}$ представляет собой характеристику, усредненную по всему волновому фронту:

$$\Phi = E \cdot 4\sin^2\left(\frac{\pi\bar{\varepsilon}}{\lambda}\right). \tag{8.35}$$

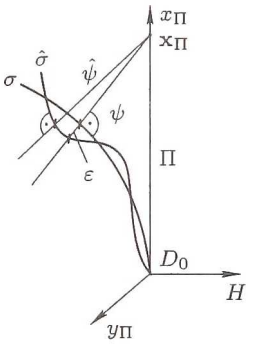


Рис. 8.6. К расчету нормального уклонения волновых фронтов

Для световых потоков Φ и E можно записать закон сохранения, следующий непосредственно из уравнения Гельмгольца:

$$\Phi \equiv \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} i(\mathbf{x}_П, l)\,\mathrm{d}^2\mathbf{x}_П = \int\limits_G i(\mathbf{u}, 0)\,\mathrm{d}^2\mathbf{u}, \tag{8.36}$$

$$E \equiv \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \hat{I}(\mathbf{x}_П, l)\,\mathrm{d}^2\mathbf{x}_П = \int\limits_G \hat{I}(\mathbf{u}, 0)\,\mathrm{d}^2\mathbf{u}. \tag{8.37}$$

Применение (8.36) и (8.37) к оценке (8.35) позволяет выразить уклонение $\bar{\varepsilon}$ непосредственно через фазовую функцию компенсатора:

$$\bar{\varepsilon}_D \leqslant \bar{\varepsilon} = \frac{\lambda}{\pi}\arcsin\left[\int\limits_G \sin^2\left(\frac{\hat{\varphi}-\varphi}{2}\right)\beta(\mathbf{u})\,\mathrm{d}^2\mathbf{u}\right]^{1/2}, \tag{8.38}$$

где

$$\beta(\mathbf{u}) = I_0(\mathbf{u}) \Big/ \int\limits_G I_0(\mathbf{u})\,\mathrm{d}^2\mathbf{u}. \tag{8.39}$$

Формула (8.38) дает оценку сверху точности уклонения $\bar{\varepsilon}_D$ фронта $\hat{\sigma}$ от эталона σ непосредственно через невязку фазовой функции компенсатора и верна для любой технологии его изготовления как тонкого оптического элемента.

8.6. Влияние дискретизации и квантования фазовой функции компенсатора на точность эталонного волнового фронта

Наличие конечного числа M градаций высоты микрорельефа и ограниченное пространственное разрешение $\delta u \times \delta v$ технологии компьютерной оптики [17] порождают квантование по уровням и дискретизацию по аргументам для фазовой функции компенсатора, приведенной к интервалу значений $[0, 2\pi m)$, $m = 1, 2, \ldots$. Влияние квантования фазовой функции обычно исследуется по интенсивности различных дифракционных порядков [25, 32]. Совместный учет дискретизации и квантования фазы в работах [27, 33] выполнялся лишь для внеосевых бинарных голограмм.

Оценим совместное воздействие дискретизации и квантования компенсатора с дифракционным фазовым микрорельефом на качество волнового фронта, пользуясь вышеопределенным, дополнительно усредненным по ансамблю шумов квантования, критерием уклонения $\bar{\varepsilon}$. Компенсатор, синтезированный методами компьютерной оптики, имеет дискретную структуру, содержащую не более $N_1 \times N_2$ ячеек разрешения G_{jk}, $(j,k) \in J$, размером $\delta u \times \delta v$ каждая. Здесь J — множество номеров (j,k) ячеек разрешения, попадающих в область G. В каждой ячейке G_{jk} фазовая функция $\hat{\varphi}$ принимает постоянное значение $\hat{\varphi}_{jk}$, получаемое путем квантования по M уровням отсчета $\varphi(\boldsymbol{\xi}_{jk})$ функции φ в центре $\boldsymbol{\xi}_{jk} = (\xi_i, \eta_k)$ ячейки G_{jk}. Причем значения $\hat{\varphi}_{jk}$ выбираются из конечного множества

$$\left\{\hat{\varphi}:\ \hat{\varphi}=j\cdot q,\ j=\overline{0,M-1}\right\},\quad q=\frac{2\pi m}{M}. \tag{8.40}$$

Оценку уклонения $\bar{\varepsilon}$ (8.38) будем производить в предположении, что:

— размеры ячейки малы по сравнению с характерным интервалом изменения фазы;

— шаг квантования q мал по сравнению с $2\pi m$, так что можно применить статистическую модель квантования [33].

Для фиксированной ячейки G_{jk} можно разложить φ по степеням $(\mathbf{u}-\boldsymbol{\xi}_{jk})$:

$$\varphi(\mathbf{u})\cong\varphi(\boldsymbol{\xi}_{jk})+(\mathbf{u}-\boldsymbol{\xi}_{jk})\nabla_u\varphi\left(\boldsymbol{\xi}_{jk}\right), \tag{8.41}$$

$$\mathbf{u}\in G_{jk},\quad \nabla_u=\left(\frac{\partial}{\partial u};\frac{\partial}{\partial v}\right),\quad \beta(\mathbf{u})\cong\beta(\boldsymbol{\xi}_{jk}),\quad (\mathbf{u}\in G_{jk}), \tag{8.42}$$

а величины шумов квантования,

$$\Theta_{jk}=\hat{\varphi}_{jk}-\varphi(\boldsymbol{\xi}_{jk}), \tag{8.43}$$

некоррелированы, имеют нулевое среднее и дисперсию

$$\left\langle\left|\Theta_{jk}\right|^2\right\rangle=\frac{q^2}{12}, \tag{8.44}$$

где $\langle\ldots\rangle$ — символ усреднения по ансамблю шумов квантования.

Производя вычисление интегралов по $\mathbf{u}$ и усреднение $\langle\ldots\rangle$ в каждой ячейке G_{jk}, переходя от полученных интегральных сумм к интегралам и подставляя результат в (8.38), получаем

$$\langle\bar{\varepsilon}\rangle=\lambda/\pi\arcsin\left[\pi/\lambda\left(\varepsilon_q^2+\varepsilon_d^2\right)^{1/2}\right], \tag{8.45}$$

где

$$\varepsilon_q^2=\lambda^2/2\pi^2\left(1-\operatorname{sinc}\frac{m}{M}\right), \tag{8.46}$$

$$\varepsilon_d^2=\frac{\lambda^2}{2\pi^2}\operatorname{sinc}\left(\frac{m}{M}\right)\int\limits_G\left[1-\operatorname{sinc}\left(\delta u\cdot\frac{\varphi_u(\mathbf{u})}{2\pi}\right)\operatorname{sinc}\left(\delta v\cdot\frac{\varphi_v(\mathbf{u})}{2\pi}\right)\right]\beta(\mathbf{u})\,\mathrm{d}^2\mathbf{u}, \tag{8.47}$$

$$\left(\varphi_u=\partial\varphi/\partial u,\ \varphi_v=\partial\varphi/\partial v\right).$$

Величина ε_q обращается в нуль при $M\to\infty$ и представляет собой компоненту среднего уклонения $\langle\bar{\varepsilon}\rangle$ из-за квантования фазы по уровням. Величина ε_d обращается в нуль при $\delta u,\delta v\to 0$ и представляет собой компоненту уклонения из-за дискретизации фазы в плоскости компенсатора.

В силу сделанных предположений о малости δu, δv и q, можно записать (8.46) и (8.47) в более простом виде при $\delta u=\delta v=\delta$:

$$\begin{aligned}\langle\bar{\varepsilon}\rangle&=\left(\varepsilon_q^2+\varepsilon_d^2\right)^{1/2},\\ \varepsilon_q^2&=1/12\,(m\lambda/M)^2,\end{aligned} \tag{8.48}$$

$$\varepsilon_d^2 = \frac{\delta^2}{12}\int\limits_G \left[\frac{1}{k}\nabla_u\varphi(\mathbf{u})\right]^2 \beta(\mathbf{u})\,\mathrm{d}^2\mathbf{u} = \frac{\delta^2}{12}\int |\chi(\mathbf{x})|^2\beta_0(\mathbf{x})\,\mathrm{d}^2\mathbf{x}, \tag{8.49}$$

где согласно (8.6), (8.3) и (8.5)

$$\chi(\mathbf{x}) = \frac{1}{k}\nabla_u\varphi(\mathbf{u}) = -\left\{\frac{\nabla_x f(\mathbf{x})}{\sqrt{1+(\nabla_x f(\mathbf{x}))^2}}\right\} + \nabla_u\psi_0(\mathbf{u}), \tag{8.50}$$

$$\beta_0(\mathbf{x}) = \beta(\mathbf{u})\partial\mathbf{u}/\partial\mathbf{x}, \quad \nabla_u = (\partial/\partial u,\ \partial/\partial v)\,, \tag{8.51}$$

$\partial\mathbf{u}/\partial\mathbf{x}$ — якобиан преобразования (8.5), а вектор $\mathbf{u}$ находится из уравнений (8.5). При малой асферичности поверхности σ, когда ее главные центры кривизны лежат вблизи центра ближайшей сферы, и при постоянной интенсивности I_0 освещающего пучка, весовые функции β и β_0 аппроксимируются константами

$$\beta(\mathbf{u}) \cong 1/|G|, \quad \beta(\mathbf{x}) \cong 1/|D|, \tag{8.52}$$

где $|G|$ — площадь области G.

Если связать параметры волнового фронта с характеристиками компенсатора, полученные оценки позволяют вычислить точность $\bar{\varepsilon}$ формирования волнового фронта имеющимся компенсатором, а также выяснить заранее, до осуществления попытки реального физического изготовления компенсатора, можно ли его получить с заданной точностью $\bar{\varepsilon} \leqslant \varepsilon_{\text{доп}}$. Кроме того, оценки (8.45)–(8.49) позволяют осуществить выбор проектных параметров технологии компьютерной оптики (δ, M и др.) в пределах допустимой погрешности $\varepsilon_{\text{доп}}$ создания эталонного волнового фронта.

8.7. Формирование волновых фронтов с малым относительным отверстием

Для гладких волновых фронтов σ (8.1), удовлетворяющих условию параксиального приближения

$$A/R_1 \ll 1, \quad B/R_2 \ll 1, \tag{8.53}$$

можно упростить уравнения для фазовой функции φ и точности ε путем разложения их в ряд Тейлора. Здесь $2A\times 2B$ — размеры прямоугольного сегмента D фронта σ по осям x, y соответственно; R_1, R_2 — главные радиусы кривизны в вершине D_0; введем также размеры $2a\times 2b$ светового отверстия G компенсатора K, связанные с A, B уравнениями типа (8.5). В данном разделе будут использованы скалярная форма записи $(u,v) = \mathbf{u}$; $(x,y) = \mathbf{x}$ и обозначения f_x, f_y, f_{xy} и т. п. для производных от f

Для упрощения записей направим оси x, y из вершины 0 по пересечению главных нормальных сечений σ с касательной плоскостью к σ в вершине D_0:

$$\begin{aligned} f(0,0) &= 0, \quad & f_x(0,0) &= 0, \\ f_y(0,0) &= 0, \quad & f_{xy}(0,0) &= 0. \end{aligned} \tag{8.54}$$

Если функция f имеет непрерывные производные по крайней мере до порядка n_0, то в силу (8.53) функции f и φ целесообразно разложить в ряды по степеням x, y и u, v соответственно:

$$f(x,y)=\sum_{n=0}^{n_0}\frac{1}{n!}\sum_{v=0}^{n}c_n^v f_{v,n-v}x^v y^{n-v}+O\left((x+y)^{n_0+1}\right), \tag{8.55}$$

$$\varphi(u,v)=\sum_{n=0}^{n_0}\frac{1}{n!}\sum_{v=0}^{n}c_n^v \varphi_{v,n-v}u^v v^{n-v}+O\left((u+v)^{n_0+1}\right), \tag{8.56}$$

где c_n^v — биномиальные коэффициенты; n_0 — порядок приближения; $O(\xi)$ — символ величины того же порядка малости, что и ξ. В силу условий (8.54) получаем

$$f_{00}=0,\quad f_{01}=f_{10}=0,\quad f_{11}=0. \tag{8.57}$$

С учетом (8.57) главные радиусы кривизны в точке $(x,y)=(0,0)$ равны

$$R_1=-\frac{1}{f_{20}},\quad R_2=-\frac{1}{f_{02}}. \tag{8.58}$$

Если, кроме того, имеет место разложение функции ψ_0,

$$\psi_0(u,v)=\sum_{n=0}^{n_0}\frac{1}{n!}\sum_{v=0}^{n}c_n^v \psi_{v,n-v}u^v v^{n-v}+O\left((u+v)^{n_0+1}\right), \tag{8.59}$$

то формулы (8.5), (8.6), (8.57) и (8.58) позволяют установить связь между коэффициентами:

$$\varphi_{00}=\varphi_0-k\left(\psi_{00}+l\right),\quad \varphi_{01}=-k\psi_{01},\quad \varphi_{10}=-k\psi_{10}, \tag{8.60}$$

$$\varphi_{20}=k\left(\frac{1}{R_1-l}-\psi_{20}\right),\quad \varphi_{02}=k\left(\frac{1}{R_2-l}-\psi_{02}\right), \tag{8.61}$$

$$\begin{aligned}
&\varphi_{11}=-k\psi_{11},\\
&\varphi_{21}=k\left[\frac{f_{21}}{(1-l/R_1)^2(1-l/R_2)}-\psi_{21}\right],\\
&\varphi_{12}=k\left[\frac{f_{12}}{(1-l/R_1)(1-l/R_2)^2}-\psi_{12}\right],
\end{aligned} \tag{8.62}$$

$$\varphi_{30}=k\left[\frac{f_{30}}{(1-l/R_1)^3}-\psi_{30}\right],\quad \varphi_{03}=k\left[\frac{f_{03}}{(1-l/R_2)^3}-\psi_{03}\right]. \tag{8.63}$$

Заметим, что решение уравнения (8.5) можно представить в виде

$$x=\frac{u}{1-l/R_1}+O((u+v)^3),\quad y=\frac{v}{1-l/R_2}+O((u+v)^3). \tag{8.64}$$

Уклонение $\bar{\varepsilon}$ волнового фронта также может быть выражено через коэффициенты параксиальной фазовой функции (8.56). Например, формулы (8.46), (8.49) в приближении (8.52) принимают вид

$$\bar{\varepsilon} = \left\{ \frac{1}{12}\left(\frac{m\lambda}{M}\right)^2 + \frac{\delta^2}{48abk^2} \int_{-a}^{a}\int_{-b}^{b} \left(\varphi_u^2 + \varphi_v^2\right) \mathrm{d}u\,\mathrm{d}v \right\}^{1/2}, \tag{8.65}$$

а в параксиальном приближении (8.53) получаем (при $n_0 = 3$) оценку

$$\bar{\varepsilon} = \left\{ \frac{1}{12}\left(\frac{m\lambda}{M}\right)^2 + \frac{\delta^2}{12k^2}\left[\varphi_{10}^2 + \varphi_{01}^2 + \frac{a^2}{3}\left(\varphi_{20}^2 + \varphi_{11}^2\right) + \frac{b^2}{3}\left(\varphi_{02}^2 + \varphi_{11}^2\right)\right] \right\}^{1/2}, \tag{8.66}$$

где коэффициенты φ_{01}, φ_0, φ_{10}, φ_{20}, φ_{02}, φ_{11} определяется по (8.60)–(8.61).

Заметим, что при используемом порядке аппроксимации ($n_0 = 3$) можно в формулах (8.65) и (8.66) заменить (см. рис. 8.5 и уравнения (8.5) и (8.64))

$$a \cong A\,(1 - l/R_1)\,, \quad b \cong B\,(1 - l/R_2)\,. \tag{8.67}$$

8.8. Осесимметричные компенсаторы

Осесимметричные компенсаторы, формирующие волновые фронты вращения, являются непосредственными аналогами компенсационных объективов и рассмотрены в [16, 18]. В данном разделе соответствующие формулы выводятся как частный случай изложенной выше общей теории. При этом для радиально-симметричных функций от $r = |\mathbf{u}| = \sqrt{u^2 + v^2}$ и $\rho = |\mathbf{x}| = \sqrt{x^2 + y^2}$ будем применять те же символы, что и для соответствующих функций от $\mathbf{u}, \mathbf{x}$.

Пусть волновой фронт σ является поверхностью вращения диаметра D с уравнением

$$z = f(\rho), \quad 0 \leqslant \rho \leqslant D/2, \tag{8.68}$$

а освещающий пучок имеет радиально-симметричную интенсивность $I_0(r)$ и эйконал $\psi_0(r)$. Переходя в уравнениях (8.5) и (8.6) к полярным координатам

$$x = \rho\cos\alpha, \quad y = \rho\sin\alpha, \tag{8.69}$$

$$u = r\cos\beta, \quad v = r\sin\beta, \tag{8.70}$$

нетрудно получить [16, 18] соотношения

$$\alpha = \beta, \tag{8.71}$$

$$r = \rho + [l + f(\rho)\cdot f'(\rho)]\,, \tag{8.72}$$

$$\varphi(r) = \varphi_0 - k\left[[l + f(\rho)]\cdot\sqrt{1 + [f'(\rho)]^2} + \psi_0(r)\right], \tag{8.73}$$

где $f'(\rho) = \mathrm{d}f/\,\mathrm{d}\rho$.

Таким образом, для нахождения фазовой функции в точке r достаточно решить относительно ρ одно нелинейное уравнение (8.72) и подставить результат в (8.73). Заметим, что согласно (8.72)

$$l = -f\,(D/2) - (D-d)/2f'\,(d/2). \tag{8.74}$$

В параксиальном приближении, когда

$$D/R \ll 1 \tag{8.75}$$

и соответственно

$$d/(R-l) \ll 1, \tag{8.76}$$

можно представить f и ψ_0 коэффициентами разложений в ряд Тейлора:

$$f(\rho) = -\frac{1}{2R}\rho^2 + \frac{1}{4!}f_4\rho^4 + \frac{1}{6!}f_6\rho^6 + O(\rho^8), \tag{8.77}$$

$$\psi_0(r) = \psi_0 + \frac{1}{2R}\psi_2 r^2 + \frac{1}{4!}\psi_4 r^4 + \frac{1}{6!}\psi_6 r^6 + O(r^8), \tag{8.78}$$

где R — радиус кривизны в вершине σ. При этом решение (8.72) представляется в виде

$$\rho = \rho(r) = \frac{1}{1-l/R}r + \frac{1}{3!}\rho_3 r^3 + \frac{1}{5!}\rho_5 r^5 + O(r^7), \tag{8.79}$$

где

$$\rho_3 = -\frac{3}{R^2\,(1-l/R)^4}\left(1 + \frac{1}{3}R^2 l f_4\right), \tag{8.80}$$

$$\rho_5 = \frac{90}{R^4\,(1-l/R)^7}\left(1 + \frac{1}{3}R^2 l f_4\right) + \frac{90}{(1-l/R)^6}\left(\frac{15f_4}{R} - l f_6\right). \tag{8.81}$$

Согласно (8.73)–(8.81) разложение фазовой функции

$$\varphi(r) = \varphi_0 + \frac{1}{2}\varphi_2 r^2 + \frac{1}{4!}\varphi_4 r^4 + \frac{1}{6!}\varphi_6 r^6 + O(r^8), \tag{8.82}$$

имеет коэффициенты

$$\begin{aligned}
\varphi_2 &= k\left[-\psi_2 + \frac{1}{R-l}\right],\\
\varphi_4 &= k\left[-\psi_4 - \frac{3\,(2-l/R)}{R^3\,(1-l/R)^4} - \frac{f_4}{(1-l/R)^4}\right],\\
\varphi_6 &= k\left[-\psi_6 + \frac{45}{R\,(1-l/R)^5} - \frac{90\,(2-l/R)}{R^5\,(1-l/R)^7} + \right.\\
&\quad \left. + \frac{15\,(5-l/R)}{R^2\,(1-l/R)^7}f_4 + \frac{10R\,(2-l/R)}{(1-l/R)^7}f_4^2 - \frac{f_6}{(1-l/R)^6}\right].
\end{aligned} \tag{8.83}$$

Точность $\bar{\varepsilon}$ создания волнового фронта вращения оценивается по формулам (8.49) и (8.82) и выражается через коэффициенты (8.83). Например, в приближении (8.52) получаем удобную расчетную формулу

$$\bar{\varepsilon} = \left\{ \frac{1}{12}\left(\frac{m\lambda}{M}\right)^2 + \frac{\delta^2 d^2}{96k^2}\left(\varphi_2^2 + \frac{d^2}{18}\varphi_2\varphi_4 + \frac{d^4}{9\cdot 2^7}\varphi_4^2 + \frac{d^4}{15\cdot 2^7}\varphi_2\varphi_6\right)\right\}^{1/2}. \tag{8.84}$$

Пример 8.1. Рассмотрим компенсатор, формирующий асферический волновой фронт вращения второго порядка, задаваемый уравнением

$$\rho^2 = -2Rf - (1-e^2)f^2, \tag{8.85}$$

где e — эксцентриситет. Освещающий пучок E создается точечным источником, расположенным в точке с координатой $(-S_0)$ на оптической оси. Здесь

$$f_4 = -3(1-e^2)/R^3, \quad f_6 = -45(1-e^2)^2/R^5, \tag{8.86}$$

$$\psi_0 = 0, \quad \psi_2 = 1/S_0, \quad \psi_4 = 3/S_0^3, \quad \psi_6 = 45/S_0^5, \tag{8.87}$$

и по формулам (8.82)–(8.84) можно подсчитать фазовую функцию φ и уклонение $\bar{\varepsilon}$ волнового фронта. Представляют интерес два случая: компенсатор «сфера — поверхность второго порядка» с источником с $S_0 = R - l$ и компенсатор «плоскость — поверхность второго порядка» с $S_0 = \infty$. Таблицы 8.1, 8.2 и 8.3 показывают зависимость $\bar{\varepsilon}$ от M и $N = d/\delta$ для случаев компенсаторов «сфера — параболоид» (гиперболоид, эллипсоид), соответственно, для $\lambda = 0{,}63$ мкм; $D = 0{,}2$ м; $R = 1$ м.

Графики и изолинии уклонения $\bar{\varepsilon}$ показаны на рисунках 8.7 и 8.8 для $S_0 = R - l$ и на рисунках 8.9 и 8.10 для плоского освещающего пучка с $S_0 = \infty$. На рисунках 8.8 и 8.10 штриховкой отмечены области, где требуемая точность не достигается.

8.9. Формирование волновых фронтов высших порядков

Уравнение поверхностей вращения n-го порядка, используемых в оптике, имеет вид

$$\rho^2 = -\sum_{p=1}^{n} a_p f^p. \tag{8.88}$$

Поскольку дифференцированием (8.88) нетрудно выразить производные f' и f'' как функции от f, то (8.72) превращается в уравнение связи $r = r(f)$, (8.73) дает фазовую функцию φ в зависимости от f. Таким образом, решая уравнение $r = r(f)$ относительно f, по заданному r можно определить $\varphi = \varphi(r(f))$.

Таблица 8.1. Значения $\bar{\varepsilon}$ для случая компенсатора «сфера — параболоид» ($e = 1$)

M	N				
	256	512	1024	2048	4096
4	$\lambda/8{,}5$	$\lambda/11{,}6$	$\lambda/13$	$\lambda/13{,}7$	$\lambda/14$
8	$\lambda/10$	$\lambda/17$	$\lambda/23$	$\lambda/26$	$\lambda/28$
256	$\lambda/11$	$\lambda/21$	$\lambda/43$	$\lambda/86$	$\lambda/173$

Таблица 8.2. Значения $\bar{\varepsilon}$ для случая компенсатора «сфера — гиперболоид» ($e = 1{,}3$)

M	N				
	256	512	1024	2048	4096
4	$\lambda/6$	$\lambda/9{,}5$	$\lambda/12$	$\lambda/13$	$\lambda/14$
8	$\lambda/6{,}2$	$\lambda/12$	$\lambda/19$	$\lambda/25$	$\lambda/30$
256	$\lambda/6{,}4$	$\lambda/13$	$\lambda/26$	$\lambda/51$	$\lambda/102$

Таблица 8.3. Значения $\bar{\varepsilon}$ для случая компенсатора «сфера — эллипсоид» ($e = 0{,}7$)

M	N				
	256	512	1024	2048	4096
4	$\lambda/11$	$\lambda/13$	$\lambda/13{,}7$	$\lambda/14$	$\lambda/14$
8	$\lambda/17$	$\lambda/23$	$\lambda/26$	$\lambda/27$	$\lambda/28$
256	$\lambda/22$	$\lambda/44$	$\lambda/91$	$\lambda/182$	$\lambda/364$

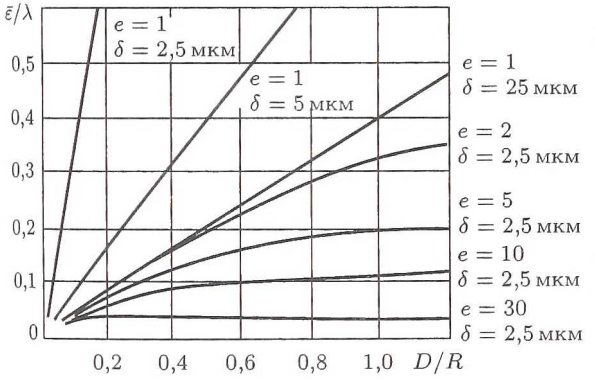


Рис. 8.7. Погрешность дискретизации компенсатора «сфера — поверхность вращения второго порядка» ($D = 0{,}2$ м; $\lambda = 0{,}6328$ мкм; $\delta = 25$ мкм; $M = 256$)

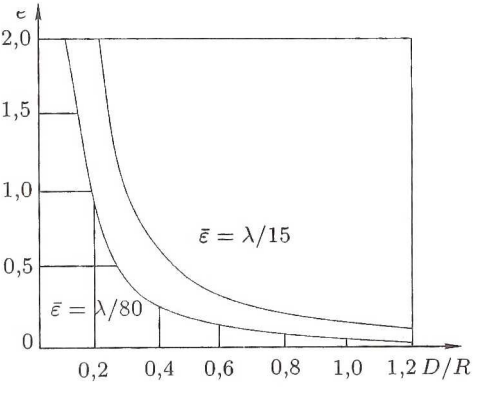


Рис. 8.8. Область значений параметров e и D/R волновых фронтов второго порядка, формируемых из сферического фронта ($D = 0{,}2$ м; $\lambda = 0{,}6328$ мкм; $M = 256$)

Таблица 8.4. Значения $\bar{\varepsilon}$ для различных δ и M

M	N				
	512	1024	2048	4096	8192
4	$\lambda/3$	$\lambda/5{,}7$	$\lambda/9$	$\lambda/12$	$\lambda/13$
8	$\lambda/3{,}1$	$\lambda/6$	$\lambda/11$	$\lambda/18$	$\lambda/22$
256	$\lambda/3{,}1$	$\lambda/6{,}1$	$\lambda/12$	$\lambda/25$	$\lambda/33$

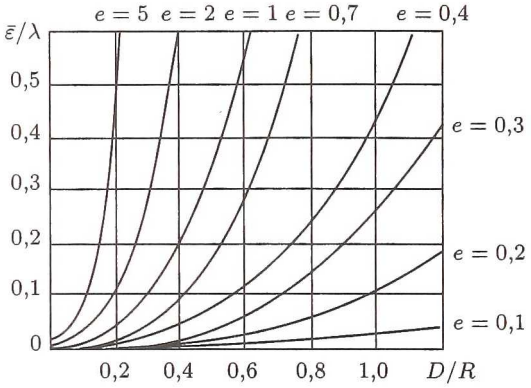


Рис. 8.9. Погрешность дискретизации компенсатора «плоскость — поверхность вращения второго порядка» ($D = 0{,}2$ м; $\lambda = 0{,}6328$ мкм; $M = 256$)

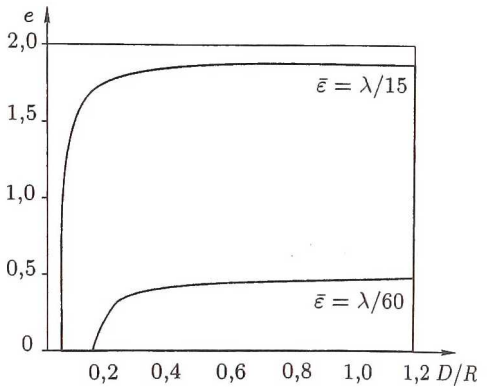


Рис. 8.10. Изолинии уклонения для волновых фронтов второго порядка, формируемых из плоского фронта ($D = 0{,}2$ м; $\lambda = 0{,}6328$ мкм; $\delta = 25$ мкм; $M = 256$)

Пример 8.2. В качестве примера неалгебраической поверхности вращения высших порядков рассмотрим поверхность зеркала, периферийные кольцевые зоны которого являются частями параболоидов вращения с различным фокусом. В [18] показано, что форма $z = f(\rho)$ такого зеркала, фокусирующего плоский осевой пучок света в тонкий осевой цилиндр длины $\aleph$, определяется дифференциальным уравнением

$$\frac{\rho}{\aleph\,[\rho/(D/2)]^2 + f(\rho) + \mathcal{F}_0} = \frac{-2f'(\rho)}{1 - [f'(\rho)]^2}, \tag{8.89}$$

где $\mathcal{F}_0$ — параксиальный фокус, D — диаметр зеркала. Очевидно, что при выполнении условия $D/\mathcal{F}_0 \ll 1$, решение уравнения (8.89) хорошо аппроксимировать несколькими членами разложения (8.77), в котором

$$f_4 = \frac{1}{2\mathcal{F}_0 D^2} \cdot \frac{\aleph}{\mathcal{F}_0}, \quad f_6 = -\frac{1}{24\mathcal{F}_0^3 D^2} \cdot \frac{\aleph}{\mathcal{F}_0} - \frac{4}{3\mathcal{F}_0 D^4} \cdot \left(\frac{\aleph}{\mathcal{F}_0}\right)^2. \tag{8.90}$$

Фазовая функция φ и уклонение $\bar{\varepsilon}$ могут оцениваться по формулам (8.82)–(8.84). В табл. 8.4 приведены значения $\bar{\varepsilon}$ для различных δ и M (при $d = 25{,}6$ мм, $\aleph = 5$ мм, $D = 130$ мм, $\mathcal{F}_0 = 500$ мм, $\lambda = 0{,}6328$ мкм и сферическом освещающем пучке с $S_0 = R - l$).

8.10. Формирование неосесимметричных волновых фронтов

Приведенные в разделах 8.3–8.4 общие соотношения позволяют рассчитать фазовую функцию φ компенсатора и уклонение $\bar{\varepsilon}$ для волновых фронтов, не являющихся поверхностями вращения.

Пример 8.3. В качестве простого примера поверхности волнового фронта, не имеющей вращательной симметрии, приведем эллиптический параболоид:

$$f(x, y) = -\left(x^2/2R_1 + y^2/2R_2\right). \tag{8.91}$$

Таблица 8.5. Значения $\bar{\varepsilon}$ ($2A = 0{,}2$; $R_1 = 2R_2 = 1$ м; $M = 256$; $m = 1$)

λ, мкм	δ, мкм					
	50	25	12	5	2,5	1
0,6328	–	$\lambda/2$	$\lambda/3$	$\lambda/6$	$\lambda/13$	$\lambda/31$
10,6	$\lambda/10$	$\lambda/21$	$\lambda/42$	$\lambda/100$	$\lambda/200$	–

Для него связь координат $\mathbf{x}$ и $\mathbf{u}$ дается уравнениями

$$\begin{aligned} u &= x(R_1 - l)/R_1 + x/R_1\left(x^2/2R_1 + y^2/2R_2\right), \\ v &= y(R_2 - l)/R_2 + y/R_2\left(x^2/2R_1 + y^2/2R_2\right). \end{aligned} \tag{8.92}$$

При сферическом освещающем пучке с

$$\psi_0(u,v) = \sqrt{S_0^2 + u^2 + v^2} - S_0, \tag{8.93}$$

фазовая функция определяется по формуле

$$\varphi(u,v,0) = \varphi_0 - k\left[\left(l - \frac{x^2}{2R_1} - \frac{y^2}{2R_2}\right)\sqrt{1 + \frac{x^2}{R_1^2} + \frac{y^2}{R_2^2}} + \sqrt{S_0^2 + u^2 + v^2}\right], \tag{8.94}$$

где x, y — решение системы уравнений (8.92), S_0 — координата точечного источника на оси z. Для плоского освещающего пучка следует положить $\psi_0 = 0$.

Рассмотрим оценку точности $\bar{\varepsilon}$ формирования из плоской волны асферического волнового фронта, представляющего сегмент эллиптического параболоида (8.91) с областью

$$D = \left\{(x,y):\ x^2/A^2 + y^2/B^2 \leqslant 1\right\}. \tag{8.95}$$

Размеры A, B согласованы с (8.91), т. е.

$$A/R_1 = B/R_2, \tag{8.96}$$

и соответственно $\bar{\varepsilon}$ зависит лишь от A/R_1. Согласно (8.50), (8.91) и (8.96), при $\varphi_0 = 0$,

$$\chi(x,y) = \left(\frac{x}{R_1},\ \frac{y}{R_2}\right) \Big/ \sqrt{1 + \frac{x^2}{R_1^2} + \frac{y^2}{R_2^2}}, \tag{8.97}$$

$$\bar{\varepsilon} = \left\{\frac{1}{12}\left(\frac{m\lambda}{M}\right)^2 + \frac{\delta^2}{12}\left[1 - \frac{R_1^2}{A^2}\ln\left(1 + \frac{A^2}{R_1^2}\right)\right]\right\}^{1/2}. \tag{8.98}$$

Значения $\bar{\varepsilon}$ в соответствии с (8.98) приведены в табл. 8.5.

Сравнивая строки таблицы 8.5, видим, что для изготовления хорошего компенсатора данного типа в видимом участке спектра, требуются устройства с разрешением 2,5 мкм и выше, в то время как в инфракрасной области уже при разрешении 25 мкм достигается приемлемое качество волнового фронта.

8.11. Формирование внеосевых сегментов волновых фронтов вращения

Контроль качества внеосевых сегментов параболических и других зеркал требует создания соответствующих эталонных волновых фронтов. Методы компьютерной оптики являются, по-видимому, единственным способом изготовления компенсаторов, формирующих необходимый сегмент без ненужной остальной части поверхности вращения [21–23].

Пусть асферическая поверхность вращения σ_1, задается уравнением

$$H_1 = F(\rho_1), \quad \rho_1 = \sqrt{x_1^2 + y_1^2}, \tag{8.99}$$

в декартовой системе координат $x_1 y_1 H_1$ (или $\rho_1 H_1$) с центром $D_1 \in \sigma_1$, лежащим на оси вращения $D_1 H_1$ (см. рис. 8.11), где $F(\rho_1)$ — гладкая (непрерывно дифференцируемая) функция с $F(0) = 0$ и постоянным знаком второй производной.

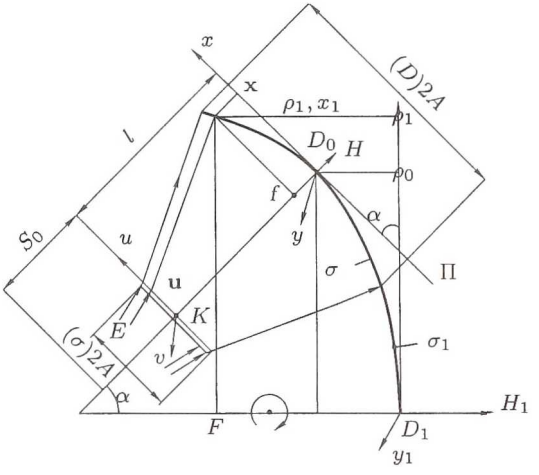


Рис. 8.11. Геометрия формирования эталона внеосевого сегмента асферической поверхности

Внеосевой сегмент σ поверхности σ_1 характеризуется своими размерами $2A \times 2B$ и центром D_0 с координатами ρ_0, $F(\rho_0)$ в системе $\rho_1 H_1$. Направим новую индивидуальную оптическую ось $D_0 H$ сегмента σ по нормали к σ_1 в точке D_0. Тогда σ характеризуется углом наклона $\alpha \in [0, \pi/2]$, представляющим угол нормали в точке D_0 с осью вращения $D_1 H_1$, и связанным с ρ_0 соотношением (см. рис. 8.11).

$$\operatorname{tg} \alpha = -F'(\rho_0), \quad \text{где } F' = \mathrm{d}F / \mathrm{d}\rho_1. \tag{8.100}$$

Рассматривая сегмент σ как самостоятельный волновой фронт независимый от σ_1, введем новую систему декартовых координат x, y, H с центром D_0, ось H которой направлена по оптической оси $D_0 H$ сегмента σ; оси x, y расположены в плоскости П, касательной к σ в точке D_0, причем ось y параллельна оси y_1. Функция

$$H = f(x, y), \quad (x, y) \in D, \tag{8.101}$$

в силу (8.99) и согласно построению осей координат, является гладкой и удовлетворяет условиям

$$f(0,0) = 0, \quad f_x(0,0) = f_y(0,0) = f_{xy}(0,0) = 0 \tag{8.102}$$

и определяется неявно уравнением

$$f \cos\alpha - x \sin\alpha - F(\rho_0) = F\sqrt{(f \sin\alpha + x \cos\alpha + \rho_0)^2 + y^2}, \quad (x, y) \in D \tag{8.103}$$

где

$$D = \{(x, y) : \ |x| \leqslant A, \ |y| \leqslant B\}, \tag{8.104}$$

f_x, f_y, f_{xy} — производные от f. Дифференцируя (8.103), получаем

$$f_x(x,y)=\frac{\rho_1\sin\alpha+F'(\rho_1)x_1\cos\alpha}{\rho_1\cos\alpha-F'(\rho_1)x_1\sin\alpha}, \tag{8.105}$$

$$f_y(x,y)=\frac{yF'(\rho_1)}{\rho_1\cos\alpha-F'(\rho_1)x_1\sin\alpha}, \tag{8.106}$$

где

$$x_1=f\sin\alpha+x\cos\alpha+\rho_0,\quad \rho_1=\sqrt{x_1^2+y_1^2},\quad y_1=y. \tag{8.107}$$

Поместим плоский фазовый компенсатор K перпендикулярно оптической оси D_0H на расстоянии l от плоскости П (см. рис. 8.11; $l>0$, если точка D_0 лежит за K по ходу лучей). Введем на плоскости K декартову систему координат u, v, оси которой параллельны x, y. Световое отверстие компенсатора выполним в виде прямоугольника:

$$G=\{(u,v):\ |u|\leqslant a,\ |v|\leqslant b\}. \tag{8.108}$$

Теперь общие формулы (8.5) и (8.6) расчета компенсатора применяются с учетом того, что в случае внеосевого сегмента поверхности (8.99), f определяется путем решения (8.103), f_x, f_y — (8.105) и (8.106). При условии (8.53) параксиального приближения и гладкой функции

$$F(\rho_1)=\sum_{n=0}^{n_0}\frac{1}{n!}F_n(\rho_1-\rho_0)^n+O\left((\rho_1-\rho_0)^{n_0+1}\right), \tag{8.109}$$

коэффициенты разложения (8.55) определяются по формулам (8.57) и

$$f_{20}=F_2\cos^3\alpha,\quad f_{02}=-1/\rho_0\cdot\operatorname{tg}\alpha, \tag{8.110}$$

$$\begin{aligned}&f_{30}=(F_3+3F_2^2\cos\alpha\cdot\sin\alpha)\cos^4\alpha,\\&f_{21}=0,\quad f_{12}=\left(1/\rho_0F_2\cos^3\alpha+1/\rho_0^2\sin\alpha\right)\cos\alpha,\\&f_{03}=0,\end{aligned} \tag{8.111}$$

$$\begin{aligned}&f_{40}=\left(F_4-10F_3F_2\cos\alpha\sin\alpha+15F_2^3\cos^2\alpha\sin^2\alpha\right)\cos^5\alpha,\\&f_{13}=0,\quad f_{31}=0,\\&f_{22}=-\left[F_3\cos^3\alpha+1/\rho_0F_2\cos\alpha\left(3\sin^2\alpha-2\right)+\right.\\&\quad\left.+3F_2^2\cos^4\alpha\sin\alpha-2/\rho_0^2\sin\alpha\right]\cdot1/\rho_0\cos^2\alpha,\\&f_{04}=3\left[F_2\cos^5\alpha+1/\rho_0\sin\alpha\left(\cos^2\alpha-\sin^2\alpha\right)\right]\cdot1/\rho_0^2.\end{aligned} \tag{8.112}$$

В соответствии с (8.100) и (8.58) имеем

$$f_{30}=3\sin\alpha\cos^5\alpha/R^2,\quad f_{21}=0,\quad f_{12}=\sin\alpha\cos^3\alpha/R^2,\quad f_{03}=0. \tag{8.113}$$

Таблица 8.6. Значения $\bar{\varepsilon}$ ($\alpha = 30°$; $4A/R_1 = 0{,}2$; $\lambda = 0{,}63$ мкм)

M	δ, мкм								
	0,2	0,5	1	2,5	5	10	25	50	100
4	$\lambda/14$	$\lambda/14$	$\lambda/13$	$\lambda/11$	$\lambda/8$	$\lambda/4$	$\lambda/2$	$1{,}1\lambda$	$2{,}2\lambda$
8	$\lambda/28$	$\lambda/27$	$\lambda/24$	$\lambda/15$	$\lambda/9$	$\lambda/4{,}4$	$\lambda/2$	$1{,}1\lambda$	$2{,}2\lambda$
16	$\lambda/54$	$\lambda/47$	$\lambda/35$	$\lambda/17$	$\lambda/9$	$\lambda/4{,}5$	$\lambda/2$	$1{,}1\lambda$	$2{,}2\lambda$
∞	$\lambda/227$	$\lambda/91$	$\lambda/45$	$\lambda/18$	$\lambda/9$	$\lambda/4{,}5$	$\lambda/2$	$1{,}1\lambda$	$2{,}2\lambda$

Таблица 8.7. Значения $\bar{\varepsilon}$ ($4A/R_1 = 0{,}2$; $\lambda = 0{,}63$ мкм; $\delta = 2{,}5$ мкм; $M = 8$)

α, град	0	10	20	30	40	50	60	70
$\bar{\varepsilon}$	$\lambda/17$	$\lambda/16{,}7$	$\lambda/16$	$\lambda/15$	$\lambda/13$	$\lambda/11$	$\lambda/7$	$\lambda/3$

Для задач фокусировки параллельных пучков света, а также в Фурье-каскадах оптической обработки информации [34] особый интерес представляют внеосевые сегменты параболоида вращения

$$F(\rho_1) = -\rho_1^2/2R. \tag{8.114}$$

Для него

$$\begin{aligned} F_0 &= -\rho_0^2/2R, \quad F_1 = -\rho_0/R, \\ F_2 &= -1/R, \quad F_m = 0, \quad m \geqslant 3, \end{aligned} \tag{8.115}$$

где R — радиус кривизны поверхности σ_1 в вершине D_1 на оси вращения. Согласно (8.113)

$$R_1 = R/\cos^3\alpha, \quad R_2 = R/\cos\alpha. \tag{8.116}$$

Уравнение (8.103) разрешается аналитически относительно f и

$$f(x,y) = \frac{\sqrt{R^2 + 2Rx\sin\alpha\cos^2\alpha - y^2\sin^2\alpha\cos^2\alpha}}{\sin^2\alpha\cos\alpha} - \frac{x\sin\alpha\cos^2\alpha + R}{\sin^2\alpha\cos\alpha}, \tag{8.117}$$

а центр сегмента имеет координаты

$$\rho_0 = R\,\mathrm{tg}\,\alpha, \quad F(\rho_0) = -R/2\,\mathrm{tg}^2\alpha. \tag{8.118}$$

Фазовая функция может определяться согласно (8.5) и (8.6). Далее по формулам (8.110)–(8.112) и (8.115) получаем

$$f_{20} = -\cos^3\alpha/R, \quad f_{02} = -\cos\alpha/R, \tag{8.119}$$

$$f_{30} = 3\sin\alpha\cos^5\alpha/R^2, \quad f_{21} = 0, \quad f_{12} = \sin\alpha\cos^3\alpha/R^2, \quad f_{03} = 0, \tag{8.120}$$

$$\begin{gathered} f_{40} = -15\sin^2\alpha\cos^7\alpha/R^3, \quad f_{31} = f_{13} = 0, \quad f_{22} = 3\sin^2\alpha\cos^5\alpha/R^3, \\ f_{04} = 3\sin^2\alpha\cos^3\alpha/R^3. \end{gathered} \tag{8.121}$$

Выберем сферический освещающий пучок E с центром $(-S_0)$ на оптической оси сегмента, для которого в параксиальном приближении

$$\psi_{00} = 0, \quad \psi_{10} = \psi_{01} = 0, \quad \psi_{20} = \psi_{02} = 1/S_0, \\ \psi_{30} = \psi_{03} = \psi_{12} = \psi_{21} = 0. \tag{8.122}$$

Подставляя (8.119)–(8.122) и (8.116) в (8.60)–(8.63), получаем для коэффициентов фазовой функции

$$\varphi_{01} = \varphi_{10} = 0, \quad \varphi_{11} = 0, \\ \varphi_{20} = k\left[-1/S_0 + 1/(R_1 - l)\right], \quad \varphi_{02} = k\left[-1/S_0 + 1/(R_2 - l)\right], \tag{8.123}$$

$$\varphi_{21} = 0, \quad \varphi_{30} = k\frac{3R_1 \sin\alpha}{(R_1 - l)^3 \cos\alpha}, \\ \varphi_{12} = k\frac{R_2 \sin\alpha}{(R_1 - l)(R_2 - l)^2 \cos\alpha}, \quad \varphi_{03} = 0. \tag{8.124}$$

Согласно (8.66) и (8.123) среднее уклонение в параксиальном приближении с $n_0 = 3$ оценивается по формуле

$$\bar{\varepsilon} = \left\{\frac{1}{12}\left(\frac{m\lambda}{M}\right)^2 + \frac{\delta^2}{36}\left[\left(\frac{1}{R_1 - l} - \frac{1}{S_0}\right)^2 a^2 + \left(\frac{1}{R_2 - l} - \frac{1}{S_0}\right)^2 b^2\right]\right\}^{1/2}, \tag{8.125}$$

где a, b определяются из (8.67).

Пример 8.4. Рассмотрим компенсатор, преобразующий плоскую волну $(1/S_0 = 0)$ во внеосевой сегмент квадратного сечения $(B = A)$. Зафиксируем размер a компенсатора. Согласно (8.67) имеем

$$b = \frac{A}{\cos^2\alpha}\left(\frac{a}{A} - \sin^2\alpha\right), \quad l = R_1\left(1 - \frac{a}{A}\right). \tag{8.126}$$

Из (8.125) и (8.126) получаем расчетную формулу

$$\bar{\varepsilon} = \left\{\frac{1}{12}\left(\frac{m\lambda}{M}\right)^2 + \frac{1}{2}\left(\frac{\delta^2}{12}\right)\cdot\left(\frac{4A}{R_1}\right)^2\cdot\left(\frac{1}{2} + \frac{1}{2\cos^4\alpha}\right)\right\}^{1/2}. \tag{8.127}$$

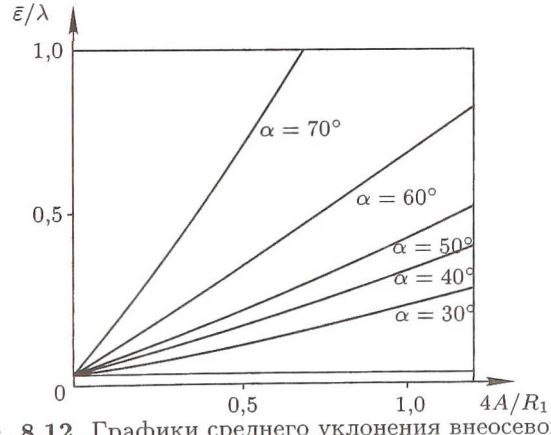


Рис. 8.12. Графики среднего уклонения внеосевого сегмента параболического волнового фронта ($\lambda = 0{,}63$ мкм; $\delta = 2{,}5$ мкм; $M = 8$)

Выражение (8.127) позволяет проанализировать зависимость $\bar{\varepsilon}$ от λ, угла наклона сегмента α и относительного отверстия сегмента $4A/R_1$ при фокусировке. Кроме того, (8.127) дает зависимость $\bar{\varepsilon}$ от числа ступенек микрорельефа M и разрешения в плоскости компенсатора δ (см. таблицы 8.6 и 8.7, рис. 8.11).

Из табл. 8.6 видно, что для формирования сегмента с $\alpha = 30°$, $4A/R_1 = 0{,}2$ при $\lambda = 0{,}63$ мкм с точностью $\sim \lambda/15$ нужно обеспечить разрешение фотошаблона $\delta \sim$ $\sim 2{,}5$ мкм при $M = 8$. При больших δ величина $\bar{\varepsilon}$ практически не зависит от M, в то время как для получения высококачественных сегментов с $\bar{\varepsilon} \sim \lambda/20 - \lambda/40$, выбор M уже существенен (см. табл. 8.6). Из табл. 8.7 и рис. 8.11 видно, что с увеличением α добиться высокой точности формирования сегмента все сложнее.

8.12. Формирование волновых фронтов с заданным распределением интенсивности

Широкий круг проблем оптического неразрушающего контроля, нелинейной оптики, оптической обработки информации подводит к постановке задачи формирования волнового фронта сложной формы с переменным распределением интенсивности по его поверхности. Достаточно упомянуть проблему создания световых реперных знаков или координатной сетки на криволинейных зеркальных или прозрачных поверхностях типа изогнутых лобовых стекол, роговицы исследуемого офтальмологами глаза и др. Другая важная проблема — формирование волны накачки при обращении волнового фронта, основанное на нелинейных эффектах вынужденного рассеяния либо на 3–4-волновом взаимодействии волн. Интересна также задача создания фазового оптического пространственного фильтра, согласованного с неплоским объектом при распознавании образов.

Рассмотренные выше компенсаторы позволяют создавать заданные волновые фронты. При расчете компенсаторов распределение интенсивности на волновом фронте не играет роли свободного параметра и не контролируется. Рассмотренные в главе 5 фокусаторы позволяют сформировать заданное распределение интенсивности. Для фокусаторов неконтролируемым параметром является распределение фазы в области фокусировки. Расчет ДОЭ для формирования заданных и амплитуды и фазы поля может быть проведен на основе методов кодирования, традиционно используемых в цифровой голографии [35]. Такие недостатки методов кодирования, как высокая частота микрорельефа и низкая энергетическая эффективность, могут быть частично компенсированы использованием более гибких итерационных методов. Одним из таких методов является рассмотренный в разделе 2.8.11 итерационный метод расчета квантованных ДОЭ. Платой за использование только одного ДОЭ для формирования как амплитуды, так и фазы поля является низкая энергетическая эффективность.

Целью данного раздела является расчет оптических систем с двумя «формирователями», являющимися ДОЭ с гладкими фазовыми функциями. Формирователи, во-первых, создают волновой фронт σ, заданный длиной волны λ, вершиной D_0, оптической осью OD_0 и функцией уклонения

$$H = f(\mathbf{x}), \quad \mathbf{x} \in D, \tag{8.128}$$

от ближайшей плоскости П, проходящей через вершину D_0 перпендикулярно оптической оси, и, во-вторых, обеспечивают требуемое распределение интенсивности $I_\sigma(\mathbf{x})$, $\mathbf{x} \in D$ по поверхности σ (рис. 8.13). Здесь $\mathbf{x} = (x, y)$ — декартовы координаты проекции точки волнового фронта σ на плоскость П; $H < 0$, если точка

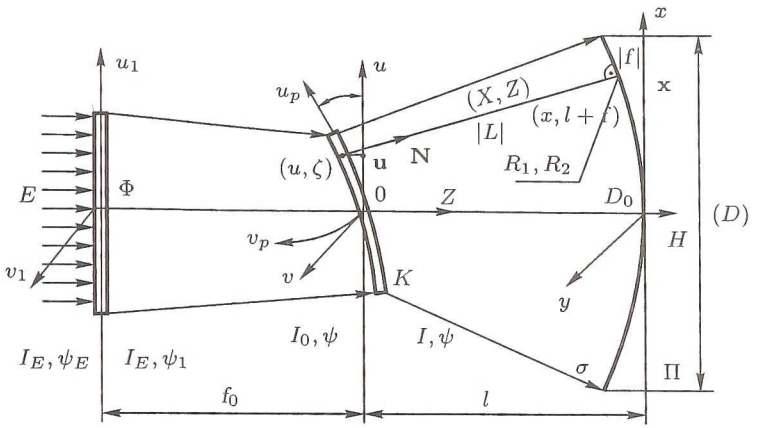


Рис. 8.13. Двухкаскадная оптическая система формирования волнового фронта с пространственно-модулированной интенсивностью

волнового фронта лежит перед П при распространении пучка; область D точек $\mathbf{x}$ соответствует световому отверстию волнового фронта σ; функция f непрерывная и гладкая, $f(0,0)=0$. Исходными данными задачи являются также распределение интенсивности $I_E(\mathbf{u}_1)$ и эйконал $\psi_E(\mathbf{u}_1)$ освещающего пучка во входной плоскости, расположенной на расстоянии (f_0+l) от вершины D_0 (рис. 8.13).

Постановка задачи расчета формирователя по f, I_σ, I_F, ψ_E является полной, так как вполне определяет световое поле в объеме однородной среды, включающей волновой фронт σ. Так, для геометрооптического светового поля можно написать эйконал и интенсивность в любой точке объема. Выберем трехмерную декартову систему координат с центром O на оптической оси на расстоянии Z от вершины D_0. Поскольку единичный вектор $\mathbf{N}$ луча, проходящего через точку $(\mathbf{x}, l+f(\mathbf{x})) \in \sigma$, определяется соотношением

$$\mathbf{N}=\frac{(-\nabla_\perp f,1)}{\left(1+|\nabla_\perp f|^2\right)^{1/2}},\quad \nabla_\perp=\left(\frac{\partial}{\partial x};\ \frac{\partial}{\partial y}\right),$$

то эйконал в произвольной точке $(\mathbf{X}, z)=(X, Y, z)$ равен

$$\Psi(\mathbf{X}, z)=-n_0 L(\mathbf{x}, z), \tag{8.129}$$

где

$$L(\mathbf{x}, z)=[l+f(\mathbf{x})-z]\cdot\sqrt{1+|\nabla_\perp f(\mathbf{x})|^2}, \tag{8.130}$$

а двумерные векторы $\mathbf{x}$, $\mathbf{X}$ связаны уравнением

$$\mathbf{X}=[l+f(\mathbf{x})-z]\cdot\nabla_\perp f(\mathbf{x}). \tag{8.131}$$

Интенсивность в произвольной точке $(\mathbf{X}, z)$ в геометрооптическом приближении легко определяется из уравнения сохранения светового потока вдоль соответствующей лучевой трубки:

$$I(\mathbf{X}, z)=I_\sigma(\mathbf{x})\,|1-L(\mathbf{x}, z)/R_1(\mathbf{x})|\cdot|1-L(\mathbf{x}, z)/R_2(\mathbf{x})|\,,$$

где $R_1(\mathbf{x})$, $R_2(\mathbf{x})$ — главные радиусы кривизны поверхности σ (8.128) в ее точке $(\mathbf{x}, l + f(\mathbf{x}))$, выбираемые положительными, если соответствующий центр кривизны лежит перед волновым фронтом σ по ходу световых лучей.

Фокусатор Ф двухкаскадной оптической системы для упрощения расчетов предполагаем выполненным на плоской поверхности, перпендикулярной оптической оси (рис. 8.13). Компенсатор K выполнен на криволинейной рабочей поверхности, задаваемой, например, параметрическими уравнениями

$$\mathbf{u} = \mathbf{u}(\mathbf{u}_p), \quad \zeta = \zeta(\mathbf{u}_p),$$

где $(\mathbf{u}, \zeta)$ — трехмерные декартовы координаты точки на рабочей поверхности компенсатора K, $\mathbf{u}_p = (u_p, v_p)$ — системы криволинейных координат на поверхности компенсатора. Расчет системы формирования проведем в 2 прохода. Начальный проход в направлении, обратном ходу лучей, позволяет заключить, что компенсатор K должен сформировать непосредственно за собой эйконал

$$\Psi\,|_K = \Psi\left(\mathbf{u}\left(\mathbf{u}_p\right), \zeta\left(\mathbf{u}_p\right)\right),$$

где ψ определяется из (8.129)–(8.131). Кроме того, фокусатор Ф должен создать на компенсаторе K распределение интенсивности, определяемое уравнением

$$I\,|_K = I_0(\mathbf{u}_p) = I(\mathbf{u}(\mathbf{u}_p), \zeta(\mathbf{u}_p)).$$

Знание $I_0(\mathbf{u}_p)$, $I_E(\mathbf{u}_1)$ и $\varphi_E(\mathbf{u}_1)$ позволяет рассчитать эйконал $\psi_1(\mathbf{u}_1)$ непосредственно за плоскостью фокусатора по его градиенту,

$$\nabla_\perp \psi_1(\mathbf{u}_1) = \frac{u(\mathbf{u}_p(\mathbf{u}_1)) - \mathbf{u}_1}{\left[f_0^2 + \left|\mathbf{u}\left(\mathbf{u}_p\left(\mathbf{u}_1\right)\right) - \mathbf{u}_1\right|^2\right]^{1/2}} n_0, \tag{8.132}$$

и гладкому лучевому отображению $\mathbf{u}_p = \mathbf{u}_p(\mathbf{u}_1)$ точек плоскости фокусатора Ф на точки поверхности компенсатора K. Определение гладкого отображения с требуемым якобианом I_0/I_F и решение уравнении вида (8.132) описано в главе 5.1. Фазовая функция фокусатора находится из уравнения

$$\varphi_1(\mathbf{u}_1) = k\left[\psi_1(\mathbf{u}_1) - \psi_E(\mathbf{u}_1)\right].$$

Последующий проход по оптической системе рис. 8.13 в прямом направлении дает возможность определить эйконал ψ_0 непосредственно перед поверхностью компенсатора K в виде

$$\psi_0(\mathbf{u}_p) = \psi_1(\mathbf{u}_1) + \left[f_0^2 + |\mathbf{u}_1 - \mathbf{u}|^2\right]^{1/2} n_0$$

и выписать фазовую функцию компенсатора:

$$\varphi(\mathbf{u}_p) = k\left[\Psi(\mathbf{u}(\mathbf{u}_p), \zeta(\mathbf{u}_p)) - \psi_0(\mathbf{u}_p)\right],$$

где $\mathbf{u}_1$ — точка на фокусаторе Ф, соответствующая точке на компенсаторе при определенном отображении $\mathbf{u}_p = \mathbf{u}_p(\mathbf{u}_1)$. Выписанные функции φ_1, φ определяют оба ДОЭ системы рис. 8.13 при известных параметрах f_0, l.

При радиально-симметричных функциях $f(\rho)$, $I_\sigma(\rho)$, $(\rho = |\mathbf{x}|)$, $\psi_E(r_1)$, $I_E(r_1)$ $(r_1 = |\mathbf{u}_1|)$, для плоской поверхности компенсатора удается существенно продвинуться в геометрооптических расчетах формирователя (рис. 8.14). Используя кольцеобразные лучевые трубки, нетрудно получить процедуру расчета, непосредственно следующую из приведенных общих соотношений.

По заданной форме волнового фронта $H = f(\rho)$ находятся эйконал

$$\psi(r) = -n_0\,[l + f(\rho)] \cdot \sqrt{1 + |f'(\rho)|^2}$$

и уравнение соответствия

$$r = \rho + [l + f(\rho)] \cdot f'(\rho), \qquad (8.133)$$

где f' — производная f по ρ, а r равно $|\mathbf{u}|$. По требуемой интенсивности $I_\sigma(\rho)$ и интенсивности I_E освещающего пучка из уравнения

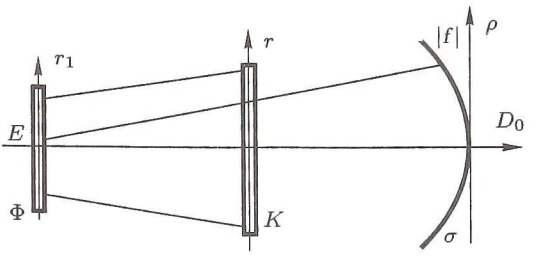


Рис. 8.14. Радиально-симметричная система формирования

$$\int_0^{r_1} I_E(r_1) \cdot 2\pi r_1\,\mathrm{d}r_1 = \int_{\rho(0)}^{\rho(r_1)} I_\sigma(\rho) \cdot 2\pi\rho\,\mathrm{d}\rho$$

находится соответствие $\rho = \rho(r_1)$, подстановка которого в (8.133) дает соответствие $r = r(r_1)$. Затем восстанавливается фазовая функция фокусатора

$$\varphi_1(r_1) = k\,[\psi_1(r_1) - \psi_E(r_1)]\,,$$

где

$$\psi_1(r_1) = \psi_1(0) + n_0 \int_0^{r_1} \frac{[r(r_1) - r_1]\,\mathrm{d}r_1}{[(r(r_1) - r_1)^2 + f_0^2]^{1/2}},$$

и эйконал перед плоскостью компенсатора

$$\psi_0(r) = \psi_0(0) + n_0 \int_{r_0}^{r} \frac{[r - r_1(r)]\,\mathrm{d}r}{[(r - r_1(r))^2 + f_0^2]^{1/2}},$$

где соответствие $r_1 = r_1(r)$ обратно соответствию $r = r(r_1)$.

И, наконец, можно рассчитать фазовую функцию компенсатора из соотношения

$$\varphi(r) = k\,[\psi(r) - \psi_0(r)]\,.$$

Пример 8.5. Продемонстрируем решение задачи формирования кольца радиусов a_1, a_2 с равномерным распределением интенсивности на параболическом волновом фронте,

$$f(\rho) = -\frac{\rho^2}{2R},$$

при плоской волне освещающего пучка диаметром $2a_0$. Для упрощения выкладок будем использовать параксиальное приближение, в котором предполагается

$$r/f_0 \ll 1, \quad r_1/f_0 \ll 1, \quad \rho/R \ll 1, \quad |f(\rho)|/l \ll 1.$$

В данном примере требуемая интенсивность

$$I_\sigma(\rho) = \begin{cases} I_\sigma,\ a_1 \leqslant \rho \leqslant a_2, \\ 0,\ \text{иначе},\ \ I_\sigma = \text{const},\ n_0 = 1, \end{cases}$$

а уравнения соответствия r_1, r и ρ имеют вид

$$r = \rho - \left(l - \frac{\rho^2}{2R}\right) \cdot \frac{\rho}{R} \cong \left(1 - \frac{l}{R}\right)\rho, \quad \frac{r_1^2}{a_0^2} = \frac{\rho^2 - a_1^2}{a_2^2 - a_1^2}.$$

Фазовые функции фокусатора и компенсатора здесь выписываются в явном виде

$$\varphi_1(r_1) = k\left[\frac{a_3 a_4}{f_0} q_1\left(\frac{r_1}{a_3}\right) - \frac{r_1^2}{2f_0}\right], \quad 0 \leqslant r_1 \leqslant a_0,$$

$$\varphi(r) = k\left[-\frac{r^2}{2f_0} + \frac{r^2}{2(R-l)} + \frac{a_3 a_4}{f_0} q_2\left(\frac{r}{a_4}\right)\right],$$

при $a_4 \leqslant r \leqslant a_5$, где

$$a_3 = a_1 a_0 \Big/ \sqrt{a_2^2 - a_1^2}, \quad a_4 = (1 - l/R)\, a_1, \quad a_5 = (1 - l/R)\, a_2,$$

$$q_1(t) = \frac{1}{2}\left[t\sqrt{t^2+1} + \ln\left(t + \sqrt{t^2+1}\right)\right], \quad q_2(t) = \frac{1}{2}\left[t\sqrt{t^2+1} - \ln\left(t + \sqrt{t^2+1}\right)\right].$$

8.13. Итеративные алгоритмы для расчета ДОЭ, формирующих заданные распределения фазы

При расчете дифракционных компенсаторов методами геометрической оптики не учитывается дифракция волн в свободном пространстве. Методы цифровой голографии [3–15] позволяют синтезировать голограмму, которая может формировать заданный волновой фронт в первом порядке дифракции, что уменьшает энергетическую эффективность оптического элемента. Ниже рассматриваются итеративные процедуры, использование которых позволяет формировать заданные волновые фронты, свободные от недостатков вышеназванных методов.

Пусть требуется рассчитать фазу $\varphi(u, v)$ оптического элемента, освещаемого световым пучком с амплитудой $A_0(u, v)$ и формирующего на расстоянии z световое поле с заданным произвольным распределением фазы $\psi_0(\xi, \eta)$. Такая задача может быть сведена к решению нелинейного интегрального уравнения [36, 37]

$$\psi_0(\xi, \eta) = \arg\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_0(u, v) \exp\left[i\varphi(u, v)\right] \times \right.$$

$$\left. \times \exp\left[\frac{ik}{2z}\left|(u - \xi)^2 + (v - \eta)^2\right|\right] du\, dv\right\}, \quad (8.134)$$

где $\arg(a + ib) = \text{tg}^{-1}(b/a)$.

Очевидно, что процедура итеративного решения (8.134) будет лишь незначительно отличаться от стандартного алгоритма Герчберга–Секстона (ГС) [38]: вместо замены рассчитанной амплитуды заданной амплитудой, мы будем заменять рассчитанную фазу заданной фазой. Шаги алгоритма будут следующие.

1. Начальная оценка фазы ДОЭ $\varphi(u,v)$ выбирается случайной.

2. Берется преобразование Френеля от функция $A_0(u,v)\exp[i\varphi(u,v)]$ (выражение в скобках в (8.134)) при помощи алгоритма быстрого преобразования Фурье.

3. Полученная функция $F_n(\xi,\eta)$ (n — номер итерации) заменяется на $\bar{F}_n(\xi,\eta)$ по правилу

$$\bar{F}_n(\xi,\eta) = |F_n(\xi,\eta)|\exp[i\psi_0(\xi,\eta)]. \tag{8.135}$$

4. Берется обратное преобразование Френеля от функции (8.135), и полученная функция $W_n(u,v)$ в плоскости ДОЭ заменяется на $\overline{W_n}(u,v)$ при помощи общепринятого правила:

$$\overline{W}_n(u,v) = \begin{cases} A_0(u,v)\,W_n(u,v)\,|W_n(u,v)|^{-1}, & (u,v)\in Q, \\ 0, & (u,v)\notin Q. \end{cases} \tag{8.136}$$

5. Переход к шагу 2 и т. д.

Степень сходимости рассчитанной фазы $\psi_n(\xi,\eta)\in[0,2\pi]$ к заданной фазе формируемого волнового фронта $\psi_0(\xi,\eta)\in[0,2\pi]$ проверяется по величине ошибки:

$$\delta_\psi^2 = \left[\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}|\psi_0-\psi|^2\,\mathrm{d}\xi\,\mathrm{d}\eta\right]\cdot\left[\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\psi_0^2\,\mathrm{d}\xi\,\mathrm{d}\eta\right]^{-1}. \tag{8.137}$$

Доказательство сходимости данного алгоритма отличается от доказательства, используемого в алгоритме ГС. Оно было получено только «в малом», т. е. при условии близости двух последовательных амплитуд света в плоскости наблюдения, и приведено в [36]. Из этого доказательства следует неравенство

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}|A_{n+1}-A_0|^2\,\mathrm{d}u\,\mathrm{d}v \leqslant \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}|A_n-A_0|^2\,\mathrm{d}u\,\mathrm{d}v, \tag{8.138}$$

где $A_{n+1}(u,v)$ и $A_n(u,v)$ — это амплитуды света в плоскости ДОЭ, рассчитанные на $(n+1)$-м и n-м шаге итерации соответственно, и $A_0(u,v)$ — заданная амплитуда света в плоскости ДОЭ. Из уравнения (8.138) следует, что в ходе итераций среднеквадратичное отклонение рассчитанной амплитуды от заданной в плоскости ДОЭ уменьшается. Не удается доказать аналогичное неравенство для фазы $\psi_n(\xi,\eta)$, но все численные эксперименты показывают, что ошибка (8.137) уменьшается с ходом итераций.

Пример 8.6. Чтобы продемонстрировать работу алгоритма (8.134)–(8.136), мы выбрали следующие параметры: число пикселов на ДОЭ и в плоскости наблюдения 128×128, радиус ДОЭ — 5,12 мм, дискретность отсчетов в этих плоскостях равна 0,04 мм; $k = 10^4\,\text{мм}^{-1}$; $z_0 = 325$ мм. Амплитуда пучка, освещающего ДОЭ, выбиралась гауссовой:

$$A_0(u,v) = \exp\left[-\frac{u^2+v^2}{w^2}\right]. \tag{8.139}$$

На рис. 8.15 показаны: (*а*) фаза формирователя волнового фронта, рассчитанная за 10 итераций; (*б*) ее сечение вдоль оси u; (*в*) фаза, сформированная на расстоянии z_0 от ДОЭ; (*г*) ее сечение вдоль оси ξ (сплошная линия соответствует заданной фазе, а пунктирная линия соответствует сформированной фазе). Заданная фаза определялась выражением

$$\psi_0(\xi,\eta) = -\alpha\left(\xi^4+\eta^4\right), \tag{8.140}$$

где $\alpha = 0{,}8$. Отклонение рассчитанной фазы $\psi_n(\xi,\eta)$ от заданной (8.140), которое определяется из уравнения (8.137), равно 12%.

Пример 8.7. Этот метод можно применять к расчету ДОЭ, способных формировать световой фронт с разной кривизной относительно осей u и v. Например, на рис. 8.16 показаны: (*а*) фаза ДОЭ; (*б*) ее сечение вдоль оси u; (*в*) фаза, формируемая на расстоянии z_0; (*г*) ее сечение вдоль оси ξ (сплошная линия соответствует заданной фазе, пунктирная линия — сформированной фазе). Последняя фаза (рис. 8.16*в*) отличается от заданной фазы, определяемой выражением

$$\psi_0(\xi,\eta) = -\beta\left(\xi^2-\eta^4\right), \tag{8.141}$$

где $\beta = 0{,}46$, в среднем на 20%.

Далее рассматривается задача расчета амплитудного транспаранта (АТ), который преобразует плоскую когерентную волну в заданное распределение фазы на некотором расстоянии вдоль оптической оси. В частном случае, такой АТ можно рассматривать как амплитудную линзу, которая преобразует падающий плоский волновой фронт в сходящийся волновой пучок, что обеспечивает концентрацию световой энергии на оптической оси на некотором расстоянии от АТ. Задача фокусировки лазерного света при помощи АТ рассматривается в работе [39].

При расчете ДОЭ такие параметры как амплитуды в плоскостях оптического элемента и наблюдения считаются заданным, а фаза ДОЭ находится. Ниже мы рассматриваем «обратную» задачу: фаза светового поля в плоскости АТ и в плоскости наблюдения считается известной, а ищется амплитудное пропускание АТ. Эта проблема оказывается формально эквивалентной решению интегрального уравнения

$$\psi_0(\xi,\eta) = \arg\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(u,v)\exp\left[i\varphi_0(u,v)\right]\exp\left[\frac{ik}{2z}\left|(u-\xi)^2+(v-\eta)^2\right|\right]\mathrm{d}u\,\mathrm{d}v\right\}, \tag{8.142}$$

где $\psi_0(\xi,\eta)$ и $\varphi_0(u,v)$ — заданные фазы в плоскости наблюдения, расположенной на расстоянии z от АТ и в плоскости транспаранта, $A(u,v)$ — искомая функция амплитудного пропускания транспаранта.

Разница между уравнениями (8.142) и (8.134) подразумевает, что в подынтегральном выражении в (8.134) считается неизвестной фаза, в то время как в (8.142) неизвестна амплитуда $A(u,v)$. Таким образом, алгоритм решения уравнения (8.142) будет содержать те же шаги, что и алгоритм решения (8.134), но вместо замены (8.136) следует использовать следующую замену [40]:

$$\overline{w}_n(u,v) = \begin{cases} |w_n(u,v)|\exp\left[i\varphi_0(u,v)\right], & (u,v)\in Q, \\ 0, & (u,v)\notin Q, \end{cases} \tag{8.143}$$

где Q — форма апертуры АТ.

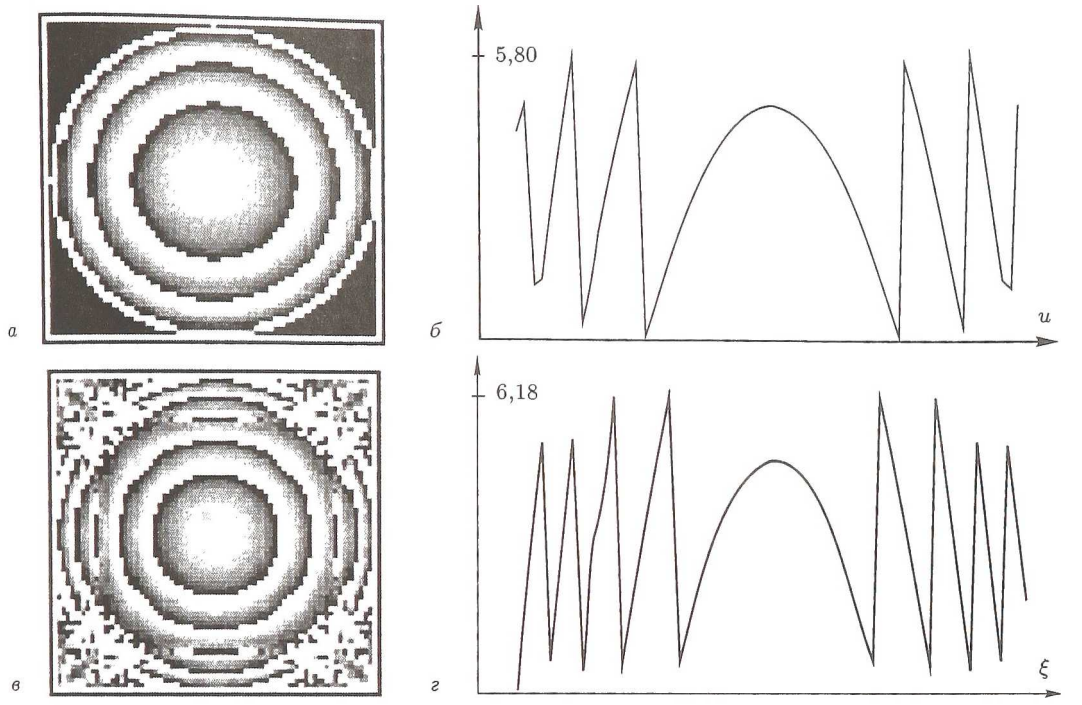


Рис. 8.15. Расчет фазового формирователя волнового фронта: (*а*) фаза ДОЭ, (*б*) ее горизонтальное сечение, (*в*) сформированная фаза, (*г*) ее горизонтальное сечение

Если транспарант освещается плоской волной, $\varphi_0(u,v)$ будет постоянной. Для случая такого итеративного алгоритма решения (8.142) мы можем доказать сходимость в среднем. Заметим, что для любого n (n — номер итерации) выполняется неравенство

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} |A_{n+1}(u,v) - A_n(u,v)|^2 \,\mathrm{d}u\,\mathrm{d}v \leqslant \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} |A_n(u,v) - A_{n-1}(u,v)|^2 \,\mathrm{d}u\,\mathrm{d}v. \tag{8.144}$$

Неравенство (8.144) показывает, что в ходе итераций среднеквадратичное отклонение для двух последовательных амплитуд в плоскости АТ не будет увеличиваться.

Аналогично мы можем показать, что для амплитуд поля в плоскости наблюдения выполняется следующее неравенство:

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} |B_{n+1}(\xi,\eta) - B_n(\xi,\eta)|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta \leqslant \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} |B_n(\xi,\eta) - B_{n-1}(\xi,\eta)|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta, \tag{8.145}$$

где B_{n+1}, B_n , B_{n-1} — это амплитуды светового поля в плоскости наблюдения на соответствующих шагах итераций.

Чтобы доказать неравенство (8.145), введем обозначения

$$B_n \exp[i\psi_n] = \mathrm{FR}\left[A_{n-1}\exp[i\varphi_0]\right], \tag{8.146}$$

$$A_n \exp[i\varphi_n] = \mathrm{FR}^{-1}[B_n \exp[i\psi_0]], \tag{8.147}$$

где B, A — амплитуды в плоскостях наблюдения и транспаранта, ψ и φ — фазы в плоскостях наблюдения и транспаранта, соответственно, FR и FR^{-1} — прямое и обратное преобразования Френеля, n — номер итерации.

Из условия сохранения световой энергии мы можем получить равенство Парсеваля. Основываясь на этом равенстве и на утверждении, что модуль разности двух комплексных чисел с разными аргументами больше чем разность двух комплексных чисел с одинаковыми аргументами, мы получим следующую цепочку неравенств:

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |B_{n+1} - B_n|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta = \\
&\qquad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|B_{n+1}\exp[i\psi_{n+1}] - B_n \exp[i\psi_{n+1}]\right|^2 \mathrm{d}\xi\,\mathrm{d}\eta \leqslant \\
&\leqslant \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|B_{n+1}\exp[i\psi_{n+1}] - B_n \exp[i\psi_n]\right|^2 \mathrm{d}\xi\,\mathrm{d}\eta = \\
&\qquad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|A_n\exp[i\varphi_0] - A_{n-1} \exp[i\varphi_0]\right|^2 \mathrm{d}u\,\mathrm{d}v \leqslant \\
&\leqslant \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|A_n\exp[i\varphi_n] - A_{n-1} \exp[i\varphi_{n-1}]\right|^2 \mathrm{d}u\,\mathrm{d}v = \\
&\qquad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|B_n\exp[i\psi_0] - B_{n-1} \exp[i\psi_0]\right|^2 \mathrm{d}\xi\,\mathrm{d}\eta = \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |B_n - B_{n-1}|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta.
\end{aligned} \tag{8.148}$$

Из сравнения первого и последнего выражения в этой цепочке мы получим окончательное неравенство

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |B_{n+1} - B_n|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta \leqslant \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |B_n - B_{n-1}|^2 \,\mathrm{d}\xi\,\mathrm{d}\eta. \tag{8.149}$$

Неравенство (8.149) показывает, что среднеквадратичная ошибка для двух последовательных оценок амплитуды светового поля в плоскости наблюдения не увеличивается в ходе итераций.

Не удается доказать, что в ходе итераций среднеквадратичная ошибка фазы в уравнении (8.137) уменьшается. Однако при помощи численных примеров удалось показать, что в тех точках (ξ, η), в которых амплитуда $B_n(\xi, \eta)$ оказывается близкой к 0, фаза $\psi_n(\xi, \eta)$ может быть существенно отличной от заданной фазы $\psi_0(\xi, \eta)$. Сходимость амплитуды, вытекающая из уравнения (8.145), приводит к сходимости фазы, в то же время функция, к которой сходится фаза $\psi_n(\xi, \eta)$, может существенно отличаться от заданной фазы в тех точках поля, где энергия света близка к нулю.

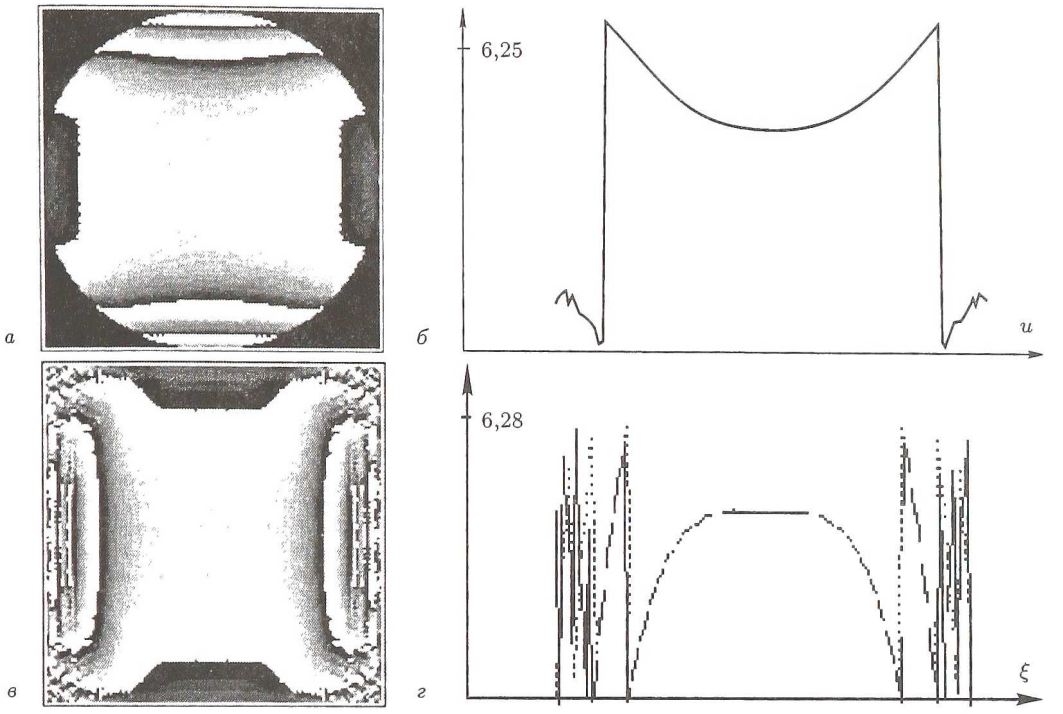


Рис. 8.16. Расчет фазового формирователя волнового фронта: (*а*) фаза ДОЭ, (*б*) ее горизонтальное сечение, (*в*) полученная фаза, (*г*) ее горизонтальное сечение

Пример 8.8. На рис. 8.17 показаны: (*а*) функция пропускания АТ, рассчитанная при помощи итеративного алгоритма за 800 итераций; (*б*) фаза, формируемая транспарантом на расстоянии $z_0 = 20$ мм (сплошная линия), которая отличается от заданной фазы (пунктирная линия) в точках малой интенсивности; (*в*) распределение интенсивности, формируемое данным транспарантом на расстоянии $z = 900$ мм. Параметры расчета следующие: $k = 10^4$ мм$^{-1}$, дискретность отсчетов в плоскости АТ равна 0,01 мм, число отсчетов 128.

Разница между синтезируемой фазой (рис. 8.17*б*) и заданной имеет специфический характер: в тех точках, где распределение интенсивности существенно не равно нулю (распределение интенсивности в плоскости наблюдения точно такое же, как в плоскости самого АТ, см. рис. 8.17*а*), соответствия между рассчитанной и заданной фазами вполне удовлетворительное, но в тех точках, где интенсивность поля близка к нулю, наблюдается существенная разница между фазами.

Заданная фазовая функция выбирается в виде

$$\psi_0(\xi) = -\alpha\xi^2, \tag{8.150}$$

где $\alpha = 2$ мм$^{-2}$.

Соотношение между распределением интенсивности на транспаранте (рис. 8.17*а*) и амплитудой поля на большом расстоянии от транспаранта (рис. 8.17*в*) можно понять, по крайней мере, качественно, из следующего известного интеграла:

$$\int_{-a}^{a} \left[a^2 - x^2\right]^{-1/2} \exp\left[ix\xi\right] \mathrm{d}x = \pi J_0(a\xi), \tag{8.151}$$

где $J_0(x)$ — функция Бесселя нулевого порядка.

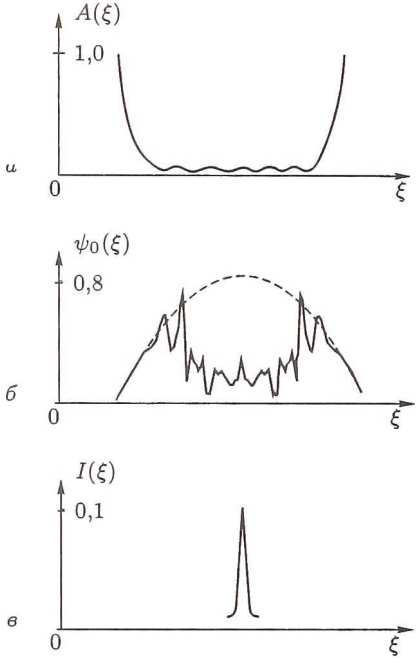


Рис. 8.17. Расчет амплитудного транспаранта: (*а*) амплитуда АТ, (*б*) сформированная фаза (сплошная линия) и заданная фаза (пунктирная линия), (*в*) интенсивность в дальней зоне дифракции

На рис. 8.18 показаны среднеквадратичное отклонение (8.145) в зависимости от числа итераций для амплитуды поля в плоскости формирования заданного волнового фронта. Видно, что с ростом числа итераций ошибка не растет, что находится в соответствии с доказательством, представленным выше.

Пример 8.9. С более очевидным выигрышем итеративный метод можно применить к расчету АТ, формирующих расходящиеся волновые фронты. На рис. 8.19 показаны: (*а*) функция пропускания АТ, полученная после 100 итераций (с теми же самыми параметрами расчета); (*б*) фаза (сплошная линия), формируемая транспарантом на расстоянии z_0 и почти идентичная заданной фазе (пунктирная линия),

$$\psi_0(\xi) = \beta\xi^4, \tag{8.152}$$

где $\beta = 17\,\text{мм}^{-4}$.

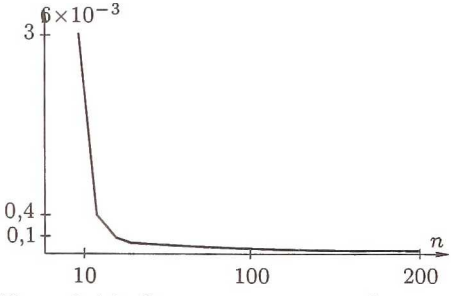
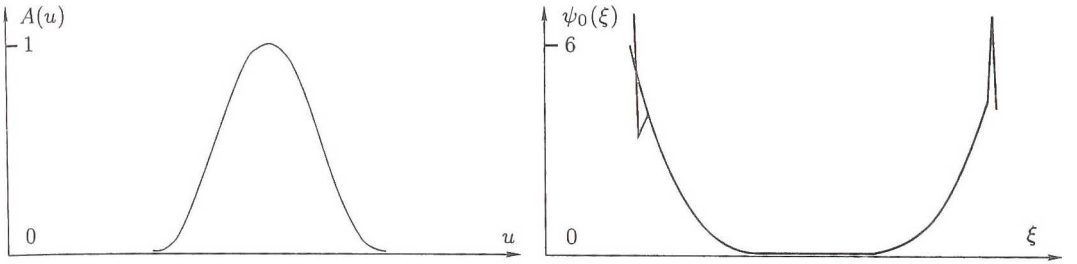


Рис. 8.18. Зависимость ошибки от числа итераций

Рассмотрим еще один частный случай модернизации итеративного алгоритма Герчберга–Секстона [38], который оказывается более приспособленным к расчету амплитудных линз. Под амплитудной линзой мы понимаем АТ, который обеспечивает концентрацию световой энергии в некоторой области, расположенной в плоскости наблюдения.

В этом случае, вместо уравнения (8.142) нам нужно итеративно решить следующее уравнение:



а *б*

Рис. 8.19. Расчет амплитудного транспаранта: (*а*) амплитуда АТ, (*б*) формируемая фаза

$$I_0(\xi,\eta)=\left|\frac{k}{z}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}A(u,v)\exp\left[\frac{ik}{2z}\left|(u-\xi)^2+(v-\eta)^2\right|\right]\mathrm{d}u\,\mathrm{d}v\right|^2, \tag{8.153}$$

где $I_0(\xi,\eta)$ — заданное распределение интенсивности в плоскости наблюдения, $A(u,v)$ — неизвестное амплитудное пропускание линзы. Чтобы обеспечить концентрацию световой энергии в такой области плоскости наблюдения, предлагается выбирать интенсивность в виде

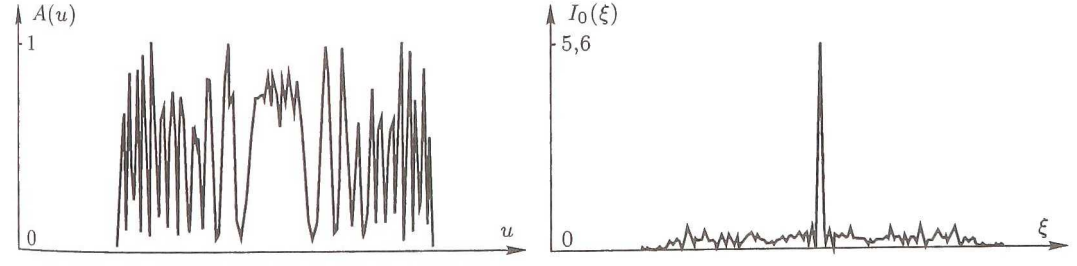
$$I_0(\xi,\eta)=p\,\mathrm{rect}\,(\xi/T_1)+q\,\mathrm{rect}\,(\xi/T_2), \tag{8.154}$$

где $T_1=\lambda z/2a$ — радиус малого дифракционного пятна в плоскости наблюдения, $T_2=a$, $2a$ — размер апертуры АТ, $p\gg q$. В (8.154) первый член аппроксимирует центральный пик интенсивности в плоскости фокусировки, а второй аппроксимирует неизбежный фоновый шум.

Уравнение (8.153) можно решить при помощи итераций используя алгоритм ГС, с помощью замены

$$\overline{W}_n(u,v)=\begin{cases}\left|W_n(u,v)\right|, & (u,v)\in Q,\\ 0, & (u,v)\notin Q,\end{cases} \tag{8.155}$$

где $W_n(u,v)$ — комплексная амплитуда света в плоскости АТ, полученная на n-м шаге итераций, Q — форма апертуры АТ.



а *б*

Рис. 8.20. Расчет амплитудной линзы: (*а*) амплитуда АТ, (*б*) формируемая интенсивность

Пример 8.10. На рис. 8.20 показаны: (*а*) функция амплитудного пропускания транспаранта, рассчитанная за 20 итераций; (*б*) распределение интенсивности, формируемое этим транспарантом на расстоянии $z_0 = 20$ мм. Методом подбора констант p и q мы можем добиться в уравнении (8.154) максимально возможной энергетической плотности в центральном пике для данных параметров. На рис. 8.20 интенсивность центрального пика в 5,6 раз больше максимальной относительной эффективности поля вблизи транспаранта и в 50 раз больше среднего значения фоновой интенсивности в плоскости фокусировки.

Эффективность фокусировки в нашем случае составляет 5%. Под эффективностью мы понимаем долю световой энергии освещающего пучка, которая идет на формирование центрального пика в плоскости фокусировки. Дифракционная эффективность равна отношению световой энергии, идущей на формирование центрального пика, ко всей энергии, прошедшей через АТ. В данном случае (см. рис. 8.20) дифракционная эффективность составляет 17%.

8.14. Практическое использование

Описанные в настоящей главе методы и алгоритмы реализованы в рамках программных продуктов "QUICK-DOE" и "ITER-DOE" для персонального компьютера на языке программирования C++ [41, 42]. Аналитические методы геометрооптического расчета реализованы в рамках "QUICK-DOE", а итерационные — в рамках "ITER-DOE"

В частности, при расчете внеосевых компенсаторов система (8.5) решается численно методом Ньютона–Рафсона на сетке (u_j, v_k), $j = \overline{1, N_1}$, $k = \overline{1, N_2}$, с шагом

$$\Delta u = u_{j+1} - u_j, \quad \Delta v = v_{k+1} - v_k. \tag{8.156}$$

Для нахождения следующей точки (x, y) по точке (x_j, y_k),

$$x = x_j + \Delta x, \quad y = y_k + \Delta y, \tag{8.157}$$

вначале, согласно (8.5), (8.103), (8.105) и (8.106), подсчитываются производные:

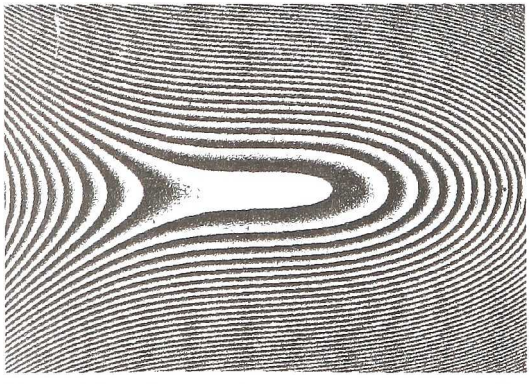


Рис. 8.21. Фотошаблон компенсатора «сфера–внеосевой сегмент параболоида»

$$v_x = u_y = (l + f)\, f_{xy} + f_x f_y, \tag{8.158}$$

$$\begin{aligned} u_x &= 1 + (l + f)\, f_{xx} + f_x^2, \\ v_y &= 1 + (l + f)\, f_{yy} + f_y^2, \end{aligned} \tag{8.159}$$

где функция f и ее производные берутся в точке (x_u, y_k).

Затем решается система линейных алгебраических уравнений

$$\begin{aligned} u_x \Delta x + u_y \Delta y &= \Delta u, \\ v_x \Delta x + v_y \Delta y &= \Delta v, \end{aligned} \tag{8.160}$$

и операции (8.157)–(8.160) повторяются итерационно.

При найденных точках (u, v) и (x, y) фазовая функция φ компенсатора подсчитывается по формуле (8.6), преобразованной к форме, не содержащей вычитания близких больших чисел.

На рисунке 8.21 представлен фотошаблон компенсатора, предназначенного для формирования из сферического пучка внеосевого сегмента параболоида вращения с параметрами $R = 303$ мм, $\alpha = 15°$, $2A = 2B = 40$ мм, $\lambda = 0{,}63$ мкм.

Развитие в последние годы техники записи фотошаблонов и технологии изготовления элементов компьютерной оптики позволило осуществить создание оптики для больших телескопов на основе контроля с помощью сверхточных ДОЭ [43].

8.15. Заключение

В главе систематически изложены теоретические основы формирования волновых фронтов с помощью ДОЭ, описаны перспективные оптические схемы контроля асферических поверхностей с помощью дифракционных компенсаторов. Приведенный набор средств оценки погрешностей формирования волновых фронтов с помощью ДОЭ позволяет оптимизировать выбор параметров оптической схемы, физических параметров и параметров дискретизации плоского компенсатора. Особого внимания заслуживает предложенный метод совместного формирования волнового фронта и заданного амплитудного распределения, близкий к идеям цифровой голографии, однако позволяющий реализовать амплитудно-фазовое распределение с помощью чисто фазовых ДОЭ.

Литература к главе 8

1. Пуряев Д. Т. *Методы контроля оптических асферических поверхностей* (М.: Машиностроение, 1976).
2. Offner A. *Applied Optics* **2**(2) 153-155 (1963)
3. Ларионов Н. П., Лукин А. В., Мустафин К. С. *Оптика и спектроскопия* **32**(2) 396 (1972)
4. Лукин А. В., Мустафин К. С. *Оптико-механическая промышленность* **4** 53 (1979)
5. Ichioka Y., Lohmann A. W. *Applied Optics* **11**(11) 2597 (1972)
6. Schwider J., Burov R. *Optica Applicata* **6**(6) 83 (1976)
7. Коронкевич В. П., Ленкова Г. А., Пальчикова И. Г., Полещук А. Г., Седухин А. Г., Чурин Е. Г., Юрлов Ю. И. *Автометрия* **1** 4 (1985)
8. McGovern A. J., Wyant J. C. *Applied Optics* **10**(3) 619 (1971)
9. Сойфер В. А., Голуб М. А., Храмов А. Г. *Материалы десятой Всесоюзной школы по голографии* (Л.: ЛИЯФ, 1978)
10. Birch K. G., Green F. J. *Appl. Phys.* **5**(11) 1982 (1972).
11. Takahashi T., Konno K., Kawai M., Isshiki M. *Applied Optics* **15**(2) 546 (1976)
12. Yatagai T., Saito H. *Optica Acta* **26**(8) 985 (1979)
13. Prowe B. *Optik* **63**(3) 203 (1982)
14. Tiziani H. J. *Laser and Optoelektron.* **15**(4) 315 (1983)
15. Ono A., Wayant J. C. *Applied Optics* **24**(4) 560 (1925).
16. Голуб М. А., Живописцев Т. С., Карпеев С. В., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Доклады Академии наук СССР* **253**(5) 1104 (1980)
17. Сисакян И. И., Сойфер В. А. *Тонкая оптика, синтезируемая на ЭВМ . Физические основы и прикладные вопросы голографии.* (Л.: ЛИЯФ, 1984)
18. Голуб М. А., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Машинный синтез оптических компенсаторов для получения асферических волновых фронтов.* Препринт

ФИАН СССР № 29 (М., 1981)
19. Golub M. A., Kazanskiy N. L., Sisakyan I. N., Soifer V. A. *Proceedings SPIE* **1183** 727 (1990)
20. Голуб М. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А. *Компьютерная оптика* (М.: МЦНТИ, 1990)
21. Голуб М. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А. *Оптика и спектроскопия* **68**(2) 461 (1990)
22. Голуб М .А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А. *Устройство для контроля оптических асферических поверхностей. А. с.* 1516767 *СССР* (*Бюлл. изобретений*, 1989) — №39
23. Голуб М. А., Казанский Н. Л., Сисакян И. Н., Сойфер В. А. *Способ изготовления асферических зеркал. А. с.* 1675812 *СССР* (*Бюлл. изобретений*, 1991) — №33
24. Голуб М. А., Сисакян И. Н., Сойфер В. А. *Оптика и спектроскопия* **69**(5) 1151 (1990)
25. Бобров С. Т., Грейсух Г. И., Туркевич Ю. Г. *Оптика дифракционных элементов и систем* (Л.: Машиностроение, 1986)
26. Greisukh G. I., Bobrov S. T., Stepanov S. A. *Optics of Diffractive and Gradient-Index Elements and Systems* (Bellingham, Washington: SPIE PRESS, 1997)
27. Wyant J. C., Bonnet V. P *Applied Optics* **11**(12) 2833 (1972)
28. Wyant J. C., O'Nelii P. K. *Applied Optics* **13**(12) 2762 (1974)
29. Handoro A., de Fong J. *Applied Optics* **16**(3) 546 (1977)
30. Biedermann K., Holngren O. *Applied Optics* **17**(8) 413 (1978)
31. Троицкий И. Н., Сафронов А. Н., Демин А. А. *Зарубежная радиоэлектроника* **9** 3 (1978)
32. Ган М. А. *Оптика и спектроскопия* **41**(4) 652 (1976)
33. Gabel R. A., Lin B. *Applied Optics* **9**(5) 1180 (1970)
34. Soifer V. A., Golub M. A. *Laser beam mode selection by computer generated holograms* (Boca Raton: CRC Press, 1994)
35. *Применение методов Фурье-оптики* Под ред. Г. Старка. *Пер. с англ. под ред. И. Н. Компанца* (М.: Радио и связь, 1988)
36. Soifer V. A., Kotlyar V. V., Doskolovich L. L. *Iterative Methods for Diffractive Optical Elements Computation* (London: Taylor & Francis, 1997)
37. Kotlyar V. V., Philippov S. V. *Opt. Las. Technol.* **27**(4) 229 (1995)
38. Gerchberg R. W., Saxton W. D. *Optik* **35** 237-246 (1972)
39. Волынкин Я. А., Ибрагимов Е. А., Усманов Т. *Оптика и спектроскопия* **72**(6) 1457 (1992)
40. Котляр В. В., Сойфер В. А., Филиппов С. В. *Оптика и спектроскопия* **75**(4) 923 (1993)
41. Волотовский С. Г., Голуб М. А., Досколович Л. Л., Казанский Н. Л., Павельев В. С., Серафимович П. Г., Сойфер В. А., Харитонов С. И., Царегородцев А. Е. *Компьютерная оптика* **14–15** 94 (1995)
42. Волотовский С. Г., Казанский Н. Л., Павельев В. С. *Компьютерная оптика* **17** 48 (1997)
43. Cherkashin V. V., Churin E. G., Burge J. H., Koronkevich V. P., Poleschuk A. G. *Proceedings SPIE* **3010** 168 (1997)

ГЛАВА 9

ПРИМЕНЕНИЕ ДОЭ В СВЕТОТЕХНИЧЕСКИХ УСТРОЙСТВАХ

9.1. Перспективы использования в светотехнических устройствах

Проектирование световых приборов всегда играло значительную роль в жизни человечества, начиная с распределения свечей по залу и кончая расчетом современных автомобильных фар с изменяемой диаграммой направленности излучения, которая адаптируется к состоянию окружающей атмосферы и скорости движения автомобиля. Не касаясь в данной работе методов создания источников света [1, 2], рассмотрим возможности использования ДОЭ в оптических трактах световых приборов.

Широкие функциональные возможности, минимальные массогабаритные характеристики, низкая себестоимость тиражирования обеспечивают хорошие перспективы использования ДОЭ в светотехнических устройствах. Имеются в виду в первую очередь устройства, формирующие монохроматическое излучение. Однако развитие современных технологий создания «квазидифракционных» (или «гармонических» [3, 4]) оптических элементов с высотой рельефа, пропорциональной большому количеству длин волн, позволяет не только сохранить преимущества дифракционной оптики, но и минимизировать хроматические аберрации ДОЭ.

Теоретической основой для использования ДОЭ в светотехнических приборах, можно считать работу [5]. В ней разрабатывалась теория расчета ДОЭ, формирующего требуемую диаграмму направленности излучения. Продолжением указанной темы является работа [6], посвященная созданию теоретических и программных средств для проектирования светотехнических устройств с ДОЭ.

В качестве примера, убеждающего в перспективности использования ДОЭ в светотехнике, рассмотрим один из вариантов компоновки светорассеивателя заднего фонаря автомобиля. Задачей каждой секции светорассеивателя является формирование расходящегося светового пучка прямоугольной формы, имеющего разные углы расходимости по вертикали и горизонтали. В некоторых европейских моделях легковых автомобилей решение этой задачи воплощается в виде «сэндвича» из двух (перпендикулярных друг другу) наборов цилиндрических линз (рис. 9.1).

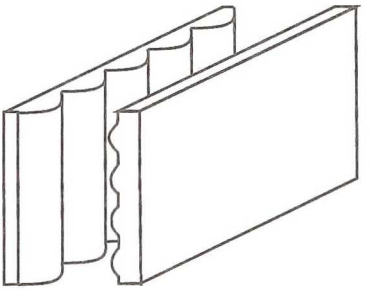
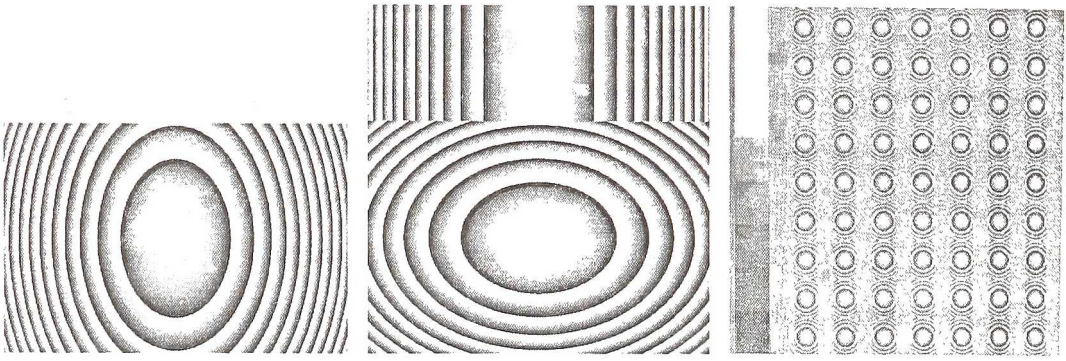


Рис. 9.1. Общий вид скрещенных массивов цилиндрических линз в заднем фонаре автомобиля

Каждый из наборов цилиндрических линз отвечает за обеспечение требуемого угла расходимости пучка в одном из направлений (по вертикали или горизонтали), а вся система в целом — за формирование



а *б* *в*

Рис. 9.2. Фотошаблоны ячеек секций стоп-сигнала (*а*) и поворота (*б*), а также общий вид секции стоп-сигнала (*в*)

требуемой диаграммы направленности излучения. При этом большинство секций (стоп-сигнал, сигнал поворота, противотуманный сигнал) характеризуются монохроматическим излучением, за исключением сигнала заднего хода. Это облегчает реализацию двух наборов (и соответственно двух оптических пластин) в виде одного оптического элемента с дифракционным микрорельефом, обеспечивающего фазовое пропускание равное сумме фазовых функций двух наборов цилиндрических линз. Расчет и изготовление такого оптического элемента не представляет никакой сложности, а улучшение конструктивного решения и технологии сборки задних огней очевидны. При тиражировании такого элемента средствами штамповки на ПММА, себестоимость новой пластины практически не отличается от себестоимости одного из двух используемых ранее наборов цилиндрических линз. На рисунке 9.2 приведены фотошаблоны сегментов скрещенных цилиндрических линз для стоп-сигнала (*а*) и сигнала поворота (*б*), а также общий вид фотошаблона секции стоп-сигнала (*в*) вместе с дизайнерскими линиями.

Испытания макета нового фонаря описаны в разделе 9.5.

9.2. Средства проектирования светотехнических устройств с ДОЭ

Рассмотрим сложную оптическую схему, состоящую из точечного источника, отражающих и преломляющих поверхностей и дифракционных оптических элементов. Оптическая установка может быть представлена как набор сред $S_1, \ldots, S_n$.

Каждая среда характеризуется показателем преломления и уравнением границы. ДОЭ вводится как фазовый разрыв вдоль элемента и приводит к существованию поверхности фазовых скачков. Для дальнейшего анализа укажем, что идеальная отражающая поверхность соответствует среде с показателем преломления -1.

Для моделирования работы светотехнического устройства необходимо определить распределение интенсивности, формируемое исследуемым устройством на трехмерной поверхности $\Phi(x_1, x_2, x_3) = 0$, где (x_1, x_2, x_3) — координаты на трехмерной поверхности.

Комплексная амплитуда поля E в скалярном приближении подчиняется уравнению Гельмгольца [7]

$$\nabla^2 E + k^2 n^2 E = \delta(\mathbf{r} - \mathbf{r}_0), \tag{9.1}$$

где $k = 2\pi/\lambda$ — волновое число, λ — средняя длина волны, $n(x_1, x_2, x_3)$ — показатель преломления в оптической системе, $\delta(\mathbf{r} - \mathbf{r}_0)$ — функция Дирака, описывающая точечный источник.

Для решения уравнения Гельмгольца необходимо определить граничные условия каждой среды. Кроме того, необходимо использовать информацию о поведении комплексной амплитуды светового поля вблизи ДОЭ.

Световое поле вдали от источников излучения может быть аппроксимировано более простыми уравнениями. Это упрощение базируется на предположении, что $kL \ll 1$, где L — характерная длина, описывающая расстояние наблюдения, размеры препятствий и неоднородностей среды.

Введем приближение Дебая [7]:

$$E(\mathbf{x}) = \sqrt{I(\mathbf{x})} \exp\{ik\Psi(\mathbf{x})\}, \tag{9.2}$$

где $\mathbf{x} = (x_1, x_2, x_3)$ — декартовы координаты, $I(x)$ — интенсивность светового поля, $\Psi(\mathbf{x})$ — эйконал светового поля. Подстановка уравнения (9.2) в уравнение Гельмгольца показывает, что функции I и Ψ подчиняются системе уравнений

$$\begin{cases} (\nabla\Psi)^2 = n^2(\mathbf{x}), \\ \operatorname{div}[I(\mathbf{x})\nabla\Psi(\mathbf{x})] = 0. \end{cases} \tag{9.3}$$

Граничные условия для уравнения Гельмгольца могут быть преобразованы:

1) для преломляющей поверхности

$$[\nabla\Psi_1, \mathbf{N}] = [\nabla\Psi_2, \mathbf{N}], \tag{9.4}$$

где $\mathbf{N}$ — нормаль к границе между двумя средами с различными показателями преломления;

2) для отражающей поверхности

$$[\nabla\Psi_1, \mathbf{N}] = -[\nabla\Psi_2, \mathbf{N}]; \tag{9.5}$$

3) для дифракционного оптического элемента

$$[(\nabla\Psi_2 - \nabla\Psi_1), \mathbf{N}] = [\nabla\varphi, \mathbf{N}]. \tag{9.6}$$

Наличие в оптической схеме источника светового поля описывается с помощью граничных условий на исходной поверхности.

Рассмотрим точечный источник, окруженный исходной поверхностью, которая задается следующей системой уравнений:

$$\begin{cases} x_1 = x_1(t_1, t_2), \\ x_2 = x_2(t_1, t_2), \\ x_3 = x_3(t_1, t_2), \end{cases} \tag{9.7}$$

где t_1, t_2 — криволинейные координаты на исходной поверхности.

Предположим, что $\Psi(t_1, t_2)$ и $I_0(t_1, t_2)$ — распределения эйконала и интенсивности на исходной поверхности, которые определяются диаграммой направленности и положением точечного источника. Вместо распределения эйконала возможно использование поля направлений лучей. Уравнения (9.3) включают частные производные, однако с помощью лучевых координат эти уравнения могут быть сведены к системе обыкновенных дифференциальных уравнений. Введем лучевые координаты

$$\begin{cases} x_1 = x_1(t_1, t_2, l), \\ x_2 = x_2(t_1, t_2, l), \\ x_3 = x_3(t_1, t_2, l), \end{cases} \tag{9.8}$$

или в векторной форме

$$\mathbf{r} = \mathbf{r}(t_1, t_2, l). \tag{9.9}$$

Функция $\mathbf{r}(t_1, t_2, l)$ подчиняется уравнению распространения луча

$$\frac{\mathrm{d}}{\mathrm{d}l}\left(n\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}l}\right) = \nabla n. \tag{9.10}$$

При $l = 0$ уравнение $\mathbf{r}_0 = \mathbf{r}(t_1, t_2, 0)$ описывает исходную поверхность. Функция $\mathbf{S}_0(t_1, t_2) = \mathrm{d}\mathbf{r}/\mathrm{d}l\,(t_1, t_2, 0)$ определяет направление лучей на исходной поверхности. Решение уравнения переноса в лучевых координатах принимает форму

$$I(t_1, t_2, l) = I_0(t_1, t_2)\,\frac{\partial(x_1, x_2, x_3)/\partial(t_1, t_2, 0)}{\partial(x_1, x_2, x_3)/\partial(t_1, t_2, l)}, \tag{9.11}$$

где $I_0(t_1, t_2)$ — распределение интенсивности на исходной поверхности, $\partial(x_1, x_2, x_3)/\partial(t_1, t_2, l)$ — якобиан преобразования декартовых координат к лучевым. Перепишем последнее уравнение в векторной форме, так как это более удобно для дальнейших рассуждений:

$$I(t_1, t_2, l) = I_0(t_1, t_2) = \frac{[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\,\mathbf{r}_l|_{l=0}}{[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\,\mathbf{r}_l}, \tag{9.12}$$

и введем криволинейную систему координат на поверхности регистрации

$$\mathbf{r} = \mathbf{r}(t_1, t_2, l(t_1, t_2)) = \mathbf{R}(t_1, t_2), \tag{9.13}$$

где $l(t_1, t_2)$ — решение уравнения

$$\Phi(\mathbf{r}(t_1, t_2, l)) = 0. \tag{9.14}$$

Дифференцируя по t_1 или t_2 , можно получить выражения для базисных векторов

$$\begin{aligned} \mathbf{R}_{t_1} &= \mathbf{r}_{t_1} - \mathbf{r}_l \cdot \frac{\nabla\Phi \cdot \mathbf{r}_{t_1}}{\nabla\Phi \cdot \mathbf{r}_l}, \\ \mathbf{R}_{t_2} &= \mathbf{r}_{t_2} - \mathbf{r}_l \cdot \frac{\nabla\Phi \cdot \mathbf{r}_{t_1}}{\nabla\Phi \cdot \mathbf{r}_l}. \end{aligned} \tag{9.15}$$

Принимая во внимание выражения для базисных векторов на исходной поверхности и поверхности регистрации, уравнение (9.12) преобразуется в следующее уравнение:

$$I(t_1, t_2, l) = I_0(t_1, t_2)\,\frac{[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\,\mathbf{r}_l|_{l=0}}{[\mathbf{R}_{t_1}, \mathbf{R}_{t_2}]\,\mathbf{r}_l}. \tag{9.16}$$

По сравнению с выражением (9.11) уравнение (9.16) содержит распределение интенсивности на исходной поверхности и метрические характеристики исходной поверхности и поверхности регистрации. В качестве поверхности регистрации может использоваться сфера с бесконечно большим радиусом. В этом случае направления лучей, падающих на поверхность регистрации, совпадают с нормалью к ней:

$$[\mathbf{R}_{t_1}, \mathbf{R}_{t_2}]\,\mathbf{r}_l = \big|[\mathbf{R}_{t_1}, \mathbf{R}_{t_2}]\big|. \tag{9.17}$$

То же самое соотношение возможно на исходной поверхности

$$[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\,\mathbf{r}_l = \left|[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\right|. \tag{9.18}$$

Можно применить эти уравнения для получения более простого выражения для распределения интенсивности на поверхности регистрации. Однако это выражение не подходит для расчета. Неудобство связано с использованием лучевой системы координат, так как выбор лучевых координат зависит от оптической схемы.

Введем новую криволинейную систему координат (T_1, T_2) на поверхности регистрации:

$$\mathbf{r} = \hat{\Phi}\,(T_1, T_2)\,. \tag{9.19}$$

Предположим, что криволинейные координаты (T_1, T_2) связаны с лучевыми координатами (t_1, t_2) с помощью следующих уравнений:

$$\begin{aligned} T_1 &= F_1\,(t_1, t_2)\,, \\ T_2 &= F_2\,(t_1, t_2)\,. \end{aligned} \tag{9.20}$$

Подставляя (9.18) и (9.19) в (9.16), получаем

$$I\,(t_1, t_2, l) = I_0\,(t_1, t_2)\,\frac{\left|[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\right|}{\left|\left[\hat{\Phi}_{F_1}, \hat{\Phi}_{F_2}\right]\right|\,\partial\,(F_1, F_2)/\partial\,(t_1, t_2)}. \tag{9.21}$$

Используя свойства функции Дирака можно переписать последнюю формулу в интегральной форме:

$$I = \int I_0\,(t_1, t_2)\,\frac{\left|[\mathbf{r}_{t_1}, \mathbf{r}_{t_2}]\right|}{\left|\left[\hat{\Phi}_{F_1}, \hat{\Phi}_{F_2}\right]\right|}\,\delta\left[T_1 - F_1\,(t_1, t_2)\,, T_2 - F_2\,(t_1, t_2)\right]\mathrm{d}t_1\,\mathrm{d}t_2. \tag{9.22}$$

Уравнение (9.22) обеспечивает базис для компьютерного моделирования и оптимизации работы проектируемых оптических схем.

Предположим, что исходная поверхность и поверхность регистрации описываются уравнениями

$$\begin{aligned} \mathbf{r}\,(t_1, t_2) &= \varepsilon\,(\sin t_1 \cos t_2 \cdot \mathbf{e}_1 + \sin t_1 \sin t_2 \cdot \mathbf{e}_2 + \cos t_1 \cdot \mathbf{e}_3)\,, \\ \hat{\Phi}\,(T_1, T_2) &= R_0\,(\sin T_1 \cos T_2 \cdot \mathbf{e}_1 + \sin T_1 \sin T_2 \cdot \mathbf{e}_2 + \cos T_1 \cdot \mathbf{e}_3)\,, \end{aligned} \tag{9.23}$$

где ε — радиус исходной поверхности, R_0 — радиус поверхности регистрации, $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ — базисные вектора прямоугольной системы координат. В этом случае распределение интенсивности на исходной поверхности имеет вид

$$I\,(t_1, t_2) = \frac{I_0}{\varepsilon^2}, \tag{9.24}$$

а распределение интенсивности на поверхности регистрации принимает вид

$$I\,(T_1, T_2) = \int \frac{I_0}{R_0^2} \cdot \frac{\sin t_1}{\sin F_1\,(t_1, t_2)}\,\delta\left[T_1 - F_1\,(t_1, t_2)\,, T_2 - F_2\,(t_1, t_2)\right]\mathrm{d}t_1\,\mathrm{d}t_2. \tag{9.25}$$

9.3. Моделирование устройств с квантованными ДОЭ

В предыдущем разделе был предложен метод моделирования светотехнических устройств с ДОЭ, имеющими непрерывный внутренний микрорельеф. Однако большинство разработанных технологий и существующего технологического оборудования не обеспечивают формирование ДОЭ с непрерывной фазовой функцией. Для изготовления многоуровневых ДОЭ успешно используется ряд технологий, подробно описанных в главе 4. В настоящем разделе предлагается метод расчета диаграммы направленности светотехнических устройств, включающих многоуровневые ДОЭ.

Моделирование аналогичных оптических систем основано на модифицированном методе расчета лучевых траекторий [4, 8]. Для понимания предлагаемого метода рассмотрим прохождение светового пучка через многоуровневый ДОЭ.

Эйконал ДОЭ со ступенчатым профилем микрорельефа может быть представлен в виде

$$\tilde{\psi} = \Phi\left(\mathrm{mod}_{\lambda}\,\psi\left(t_1, t_2\right)\right), \tag{9.26}$$

где (t_1, t_2) — декартовы координаты на дифракционном оптическом элементе, $\Phi(z)$ — функция предыскажения [9, 10]. Функция предыскажения описывает преобразование эйконала ДОЭ в результате использования конкретной технологии формирования ДОЭ и возникших при изготовлении технологических погрешностей микрорельефа.

Решение уравнения Гельмгольца может быть записано в форме

$$E\left(x_1, x_2, x_3\right) = \frac{i}{\lambda}\int E_0\left(t_1, t_2\right)\frac{z}{L}\cdot\frac{\exp\left\{ikL\right\}}{L}\exp\left\{ik\tilde{\psi}\left(t_1, t_2\right)\right\}\mathrm{d}t_1\,\mathrm{d}t_2. \tag{9.27}$$

Для вычисления интеграла Кирхгофа разложим функцию $\exp\left\{ik\tilde{\psi}\left(t_1, t_2\right)\right\}$ в ряд Фурье:

$$\exp\left\{ik\tilde{\psi}\left(t_1, t_2\right)\right\} = \sum C_n \exp\left\{ikn\psi\left(t_1, t_2\right)\right\}. \tag{9.28}$$

Подставляя формулу (9.28) в уравнение (9.27) и вычисляя все интегралы методом стационарной фазы [11], получаем выражение для комплексной амплитуды светового поля:

$$E = -\sum C_n E_0\left(t_1^n, t_2^n\right)\frac{\exp\left\{ikn\psi\left(t_1^n, t_2^n\right)\right\}}{\sqrt{J_n\left(t_1^n, t_2^n, l^n\right)}}, \tag{9.29}$$

где

$$J_n\left(t_1^n, t_2^n, l^n\right) = \frac{\partial\left(x_1, x_2, x_3\right)/\partial\left(t_1^n, t_2^n, 0\right)}{\partial\left(x_1^n, x_2^n, x_3^n\right)/\partial\left(t_1^n, t_2^n, l^n\right)} \text{ — лучевые координаты.}$$

Функция преобразования лучевых координат в декартовы подчиняется системе дифференциальных уравнений

$$\frac{\mathrm{d}}{\mathrm{d}l^n}\left(\frac{\mathrm{d}x_i}{\mathrm{d}l^n}\left(t_1^n, t_2^n, l^n\right)\right) = 0 \tag{9.30}$$

с начальными условиями

$$x_i\left(t_1^n, t_2^n, 0\right) = t_i, \quad x_3\left(t_1^n, t_2^n, 0\right) = 0,$$

$$\frac{\mathrm{d}x_i}{\mathrm{d}l}\left(t_1^n, t_2^n, 0\right) = \frac{\partial\left[n\psi\left(t_1^n, t_2^n\right)\right]}{\partial t_i^n}, \tag{9.31}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}l}\left(t_1, t_2, 0\right) = \sqrt{1 - \left(\frac{\mathrm{d}x_1}{\mathrm{d}l}\right)^2 - \left(\frac{\mathrm{d}x_2}{\mathrm{d}l}\right)^2}, \quad i = 1, 2.$$

На основе формулы (9.29) можно получить распределение интенсивности

$$I\left(x_1, x_2, x_3\right) = \sum \frac{\left|C_n\right|^2 I_0\left(t_1^n, t_2^n\right)}{J_n\left(t_1^n, t_2^n, l\right)} +$$

$$+ \sum_{m \neq n} \frac{C_n C_m E_0\left(t_1^n, t_2^n\right) E_0^*\left(t_1^m, t_2^m\right) \exp\left\{ik\left(n\psi\left(t_1^n, t_2^n\right) - m\psi\left(t_1^m, t_2^m\right)\right)\right\}}{\sqrt{J_n\left(t_1^n, t_2^n, l^n\right)}\sqrt{J_m\left(t_1^m, t_2^m, l^m\right)}}. \tag{9.32}$$

Проанализируем полученное выражение. Первый член в уравнении (9.32) представляет собой интенсивность светового поля, сформированного ДОЭ с эйконалом $n\psi$. Второй член этого выражения описывает интерференцию между световыми пучками. Обычно источник света в исследуемых светотехнических устройствах является некогерентным, поэтому интерференцией между различными световыми пучками можно пренебречь.

Описанные в разделах 9.2 и 9.3 методы реализованы в виде программного обеспечения, позволяющего моделировать и проектировать оптические схемы, которые содержат не только классические оптические элементы (линзы, зеркала, призмы), но и современные оптические структуры и элементы (волноводные, голографические, дифракционные, граданы и т. д.).

9.4. Проектирование автомобильных фар с ДОЭ

Пример 9.1. Рассмотрим задачу проектирования фары автомобиля. Международные требования к диаграмме направленности излучения такой оптической системы заключаются в формировании удаленного распределения интенсивности, схематично представленного на рис. 9.3. Построение такой формы диаграммы направленности излучения объясняется не только необходимостью освещения дороги перед автомобилем, но и задачей освещения части пространства, непосредственно прилегающей к шоссе по направлению движения, и требованием неослепления водителей движущегося навстречу транспорта.

В современном автомобилестроении задача формирования требуемой диаграммы направленности решается, в частности, за счет использования сложной формы отражателя (например, сплюснутый по вертикали эллипсоид), диафрагмы, которая расположена в передней фокальной плоскости мощной линзы и повторяет форму диаграммы направленности (см. рис. 9.4), и других ухищрений.

Использование уникальных функциональных возможностей дифракционных оптических элементов позволяет упростить форму отражателя, повысить энергетическую эффективность и технологичность фары, улучшить ее функциональные характеристики и уменьшить стоимость изготовления. Целью проводимых исследований может быть, например, замена отражателя, имеющего дорогостоящую в

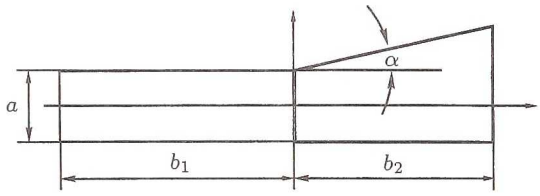


Рис. 9.3. Требуемая форма диаграммы направленности автомобильной фары

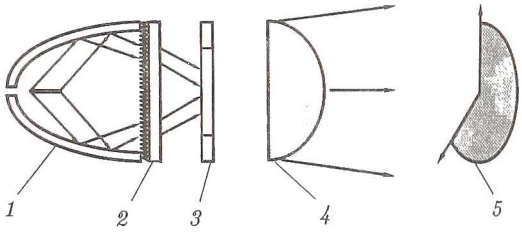


Рис. 9.4. Оптическая схема автомобильной фары: *1* — отражатель; *2* — ДОЭ; *3* — диафрагма; *4* — линза, *5* — регистратор

изготовлении форму сплюснутого по вертикали эллипсоида, на осесимметричный эллиптический отражатель, который значительно дешевле. Для этого рассмотрим возможность введения в оптическую схему фары (рис. 9.4) ДОЭ, перераспределяющего излучение из экранирующей части диафрагмы в открытую (пропускающую) часть. В качестве ДОЭ в оптической схеме (рис. 9.4) может использоваться, например, квазипериодический фокусатор в отрезок [12]. При этом фокальный отрезок целесообразно разместить в открытой части диафрагмы вблизи и вдоль нижней границы. Результаты исследования такой оптической схемы с ДОЭ и без него показаны на рис. 9.5. Параметры моделирования: длина излучающей спирали — 5 мм; радиус спирали — 0,5 мм; число витков спирали — 14; смещение центра спирали от задней стенки отражателя — 18 мм; размеры полуосей осесимметричного эллипсоида-отражателя вдоль оси Z — 50 мм, вдоль осей X и Y — 40 мм; диаметр круглого отверстия в задней части отражателя для установки источника света — 10 мм; глубина отражателя — 50 мм; период на дифракционном оптическом элементе — 1 мм; длина отрезка — 30 мм; расстояние между краем отражателя и диафрагмой — 20 мм; угол наклона половины диафрагмы (соответствующий углу α на формируемой диаграмме направленности, рис. 9.3) — 15 градусов; диаметр линзы — 59 мм; толщина линзы — 14 мм; показатель преломления материала линзы — 1,5; расстояние от диафрагмы до линзы — 52 мм; расстояние до плоскости регистрации — 25 м; число точек на стороне регистратора — 201; число просчитываемых лучей (обратно пропорционально параметру дискретизации на излучателе) для рис. 9.5*а* и рис. 9.5*в* — не менее 10^4, для рис. 9.5*б* — не менее 10^5. Использование большего количества просчитываемых лучей требует существенно большего времени, но дает более точный результат (сравните рис. 9.5*а* и рис. 9.5*б*). Выбранное расстояние до плоскости регистрации является стандартным при испытаниях в автомобилестроении.

Анализ рис. 9.5 подтверждает эффективность использования ДОЭ в светотехнических устройствах. Расчетные распределения интенсивности на рис. 9.5*а* и рис. 9.5*б*

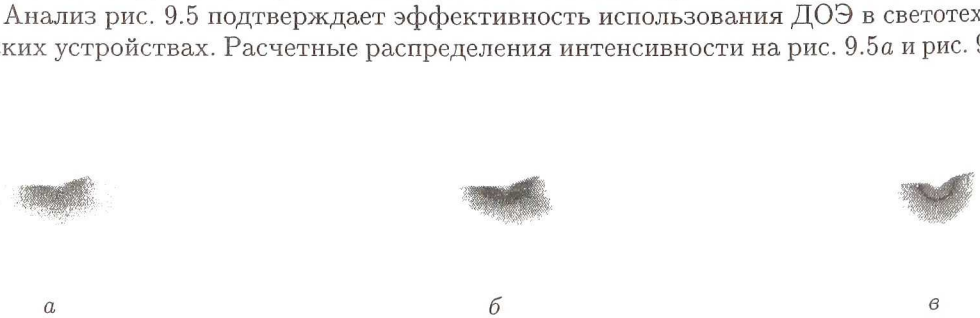


Рис. 9.5. Распределение интенсивности в плоскости регистрации излучения фары: (*а*) количество просчитываемых лучей не менее 10^4; (*б*) количество просчитываемых лучей не менее 10^5; (*в*) без ДОЭ

при использовании предлагаемого ДОЭ намного ближе к требуемому распределению (рис. 9.3), чем при использовании осесимметричного эллиптического отражателя без ДОЭ (рис. 9.5*в*). Дальнейшая оптимизация оптической схемы фары (рис. 9.4) возможна в случае использования более эффективных методов расчета применяемого ДОЭ. Например, в случае удачного применения метода согласованных прямоугольников [13] (с целью расчета ДОЭ для фокусировки излучения в область, вырезаемую диафрагмой), можно упростить оптическую схему рис. 9.4, исключив из нее диафрагму. При высокой эффективности работы такого ДОЭ качество фары и ее энергетическая эффективность только улучшатся.

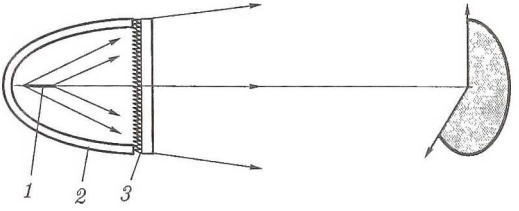


Рис. 9.6. Простейшая оптическая схема автомобильной фары с ДОЭ: *1* — источник света; *2* — параболический отражатель; *3* — ДОЭ

В наиболее простом варианте оптическая схема фары может быть основана на параболическом отражателе и ДОЭ, формирующем требуемую диаграмму направленности (рис. 9.6). Сложность расчета ДОЭ в оптических схемах такого рода заключается в большом количестве разнонаправленных лучей, приходящих в каждую точку дифракционного элемента от протяженного источника и параболического отражателя. В результате использование только одного оптического элемента, пускай и с широкими функциональными возможностями, не всегда позволяет сформировать требуемую диаграмму направленности с достаточной эффективностью.

9.5. Проектирование компланарных осветителей

Разработанный подход позволяет создавать и исследовать оптические схемы, основанные на применении волноводных и дифракционных структур. К новому классу исследованных оптических систем относятся компланарные осветители. Особенностью компланарных систем является то, что подводимое сбоку (в торец) излучение затем распределяется по выходной поверхности, создавая равномерно светящееся фронтальное изображение. Рассмотрим несколько оптических схем, реализующих идею такого устройства.

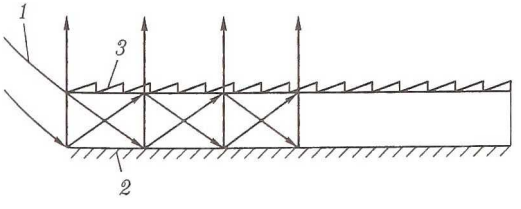


Рис. 9.7. Оптическая схема 1 компланарного осветителя: *1* — падающее излучение; *2* — отражающий дифракционный микрорельеф; *3* — пропускающий дифракционный микрорельеф

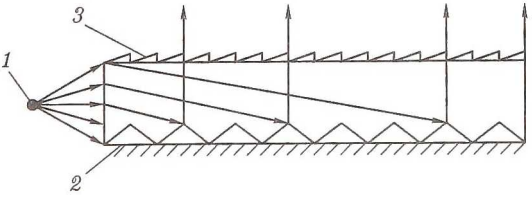


Рис. 9.8. Оптическая схема 2 компланарного осветителя: *1* — линейка светодиодов; *2* — отражающий рельеф; *3* — пропускающий дифракционный микрорельеф

Пример 9.2. Оптическая схема 1, изображенная на рис. 9.7, представляет собой плоскопараллельную пластину. На передней (верхней — рис. 9.7) стороне пластины находится дифракционный микрорельеф, на противоположной (нижней) стороне находится отражающая дифракционная решетка. Оптическая схема работает следующим образом. Квазиколлимированный световой пучок от источника падает на дифракционную решетку. Отраженный поток представляет собой суперпозицию нескольких световых пучков, распространяющихся в различных направлениях. Параметры отражающей дифракционной решетки выбираются таким образом, чтобы существовали только два отраженных пучка. Первый пучок распространяется вдоль пластины, второй пучок выходит перпендикулярно пластине через дифракционный микрорельеф наружу.

Оптическая схема 2, изображенная на рис. 9.8, также представляет собой плоскопараллельную пластину. На одной грани плоскопараллельной пластины, обращенной к наблюдателю, нанесен дифракционный микрорельеф, служащий для формирования заданной диаграммы направленности. На противоположной грани пластины нанесен рельеф с треугольным профилем, отражающим излучение. Оптическая схема работает следующим образом. Квазиколлимированный световой поток, формируемый источником света, падает на рельеф с отражающим покрытием. Отраженный свет проходит через дифракционный микрорельеф, нанесенный на передней стороне пластины, формируя требуемую диаграмму направленности.

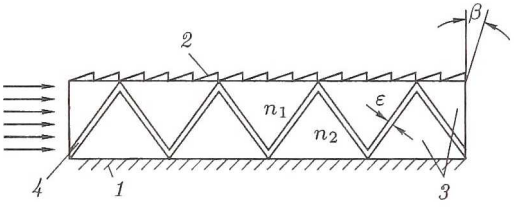


Рис. 9.9. Оптическая схема 3 компланарного осветителя: *1* — отражающее покрытие; *2* — пропускающий дифракционный микрорельеф; *3* — оптическая среда с показателем преломления n_1; *4* — оптическая среда с показателем преломления n_2

Оптическая схема 3, изображенная на рис. 9.9, представляет собой совмещенные призматические пластины с показателем преломления n_1, которые разделяет пилообразная щель, заполненная материалом с показателем преломления n_2. Нижняя часть пластины является отражающей и не выпускает излучение наружу, а на верхней части пластины располагается пропускающий дифракционный микрорельеф, формирующий требуемую диаграмму направленности.

Соотношение между показателями преломления n_1 и n_2 выбирается в оптической схеме 3 таким образом, чтобы равномерно распределить световой поток по глубине пластины при выбранной ширине осветителя. Это обеспечивает равномерное распределение интенсивности на фронтальной выходной поверхности за счет отражения нужной доли излучения на границе двух сред.

Если необходимо (например, с точки зрения дизайна), пластины осветителей могут быть плавно изогнуты, повторяя форму кузова легкового автомобиля или интерьера.

Целью моделирования был выбор работоспособных оптических схем (кроме представленных на рис. 9.7–9.9 рассматривались и другие) и определение оптимальных параметров исследуемой оптической схемы. В качестве критерия качества осветителя использовалось требование равномерности распределения интенсивности выходного светового потока. Такие требования предъявляются, в частности, при проектировании задних сигнальных огней легковых автомобилей. Некоторые интересные результаты моделирования представлены на рис. 9.10–9.12.

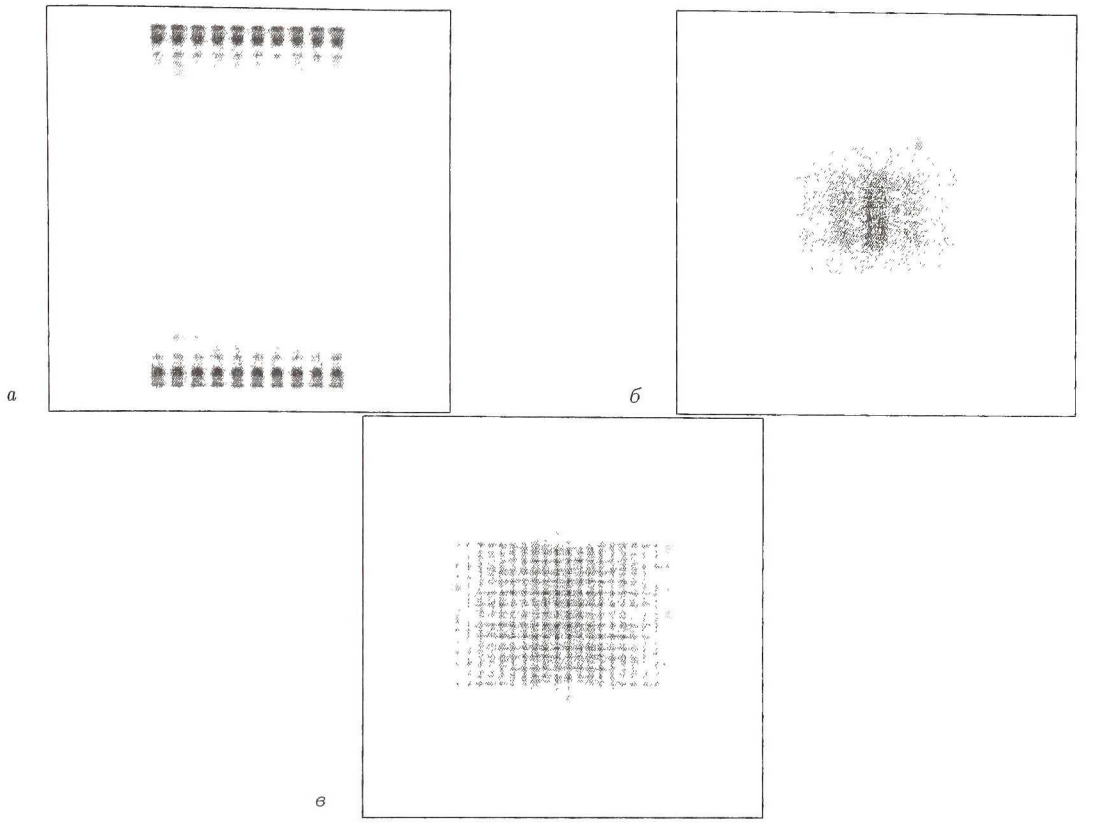


Рис. 9.10. Распределение интенсивности в плоскостях регистрации излучения компланарного осветителя, представленного на рис. 9.7: (*а*) на расстоянии 1 мм; (*б*) на расстоянии 1 м; (*в*) на расстоянии 10 м

На рис. 9.10 представлены результаты моделирования следующей модификации оптической схемы 1 (рис. 9.7): выходной дифракционный микрорельеф 3 отсутствует, вместо него на входе излучения в пластину стоит обычная бинарная пропускающая дифракционная решетка, которая формирует два пучка, падающих под углом на отражающий микрорельеф и на выходную плоскость, отражаясь от которой излучение проходит дальше вдоль пластины. Параметры моделирования: фокусное расстояние линзы светодиода — 5 мм; диаметр линзы светодиода — 5 мм; длина пластины — 100 мм; толщина пластины — 3 мм; расстояние между светодиодом и торцом пластины – 5 мм; число светодиодов на стороне — 10 (по 10 диодов светят с правой и левой сторон пластины); расстояние между диодами — 5 мм; микрорельеф отражающей решетки рассчитывается методами, изложенными в главе 3, таким образом, чтобы 95% падающей на решетку энергии уходило дальше (в нулевой порядок дифракции), а 5% энергии шло через выходную плоскость наружу (в минус первый порядок); число точек на стороне регистратора — 100; расстояние до первого регистратора (рис. 9.10*а*) — 1 мм; расстояние до второго регистратора (рис. 9.10*б*) — 1 м; расстояние до третьего регистратора (рис. 9.10*в*) — 10 м; угол на внешний край регистратора $\beta = 10°$; число просчитываемых лучей — не менее 5000. Используемая в расчетах модель светодиода предполагает, что в первой фокальной плоскости менисковой линзы (выходной линзы светодиода) расположена светящаяся квадратная

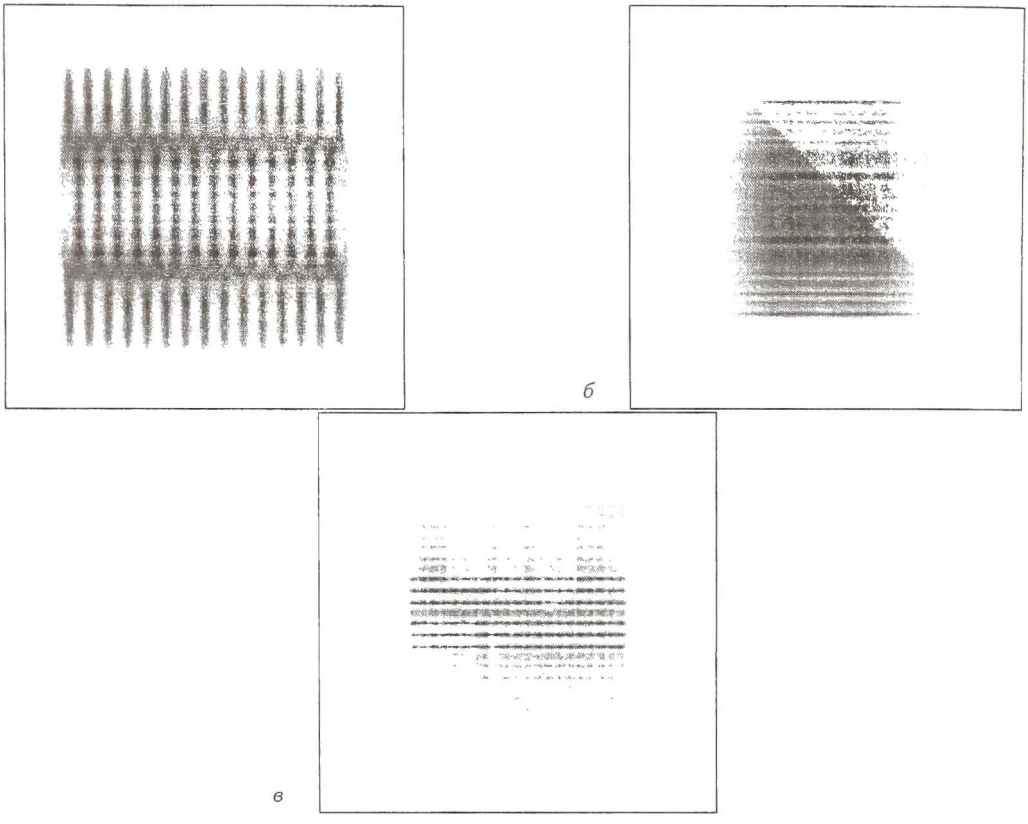


Рис. 9.11. Распределение интенсивности в плоскостях регистрации излучения компланарного осветителя, представленного на рис. 9.8: (*а*) на расстоянии 1 мм; (*б*) на расстоянии 1 м; (*в*) на расстоянии 10 м

площадка размером 1×1 мм. На рис. 9.10*а* представлено распределение интенсивности на расстоянии 1 мм от выходной плоскости осветителя, на рис. 9.10*б* — на расстоянии 1 м; на рис. 9.10*в* — на расстоянии 10 м. Для качественного моделирования работы осветителя на большем расстоянии необходимо просчитывать прохождение существенно большего количества лучей. На рис. 9.10*а* хорошо видны места расположения светодиодов, однако с увеличением расстояния (см. рис. 9.10*б* и 9.10*в*) поле от осветителя выглядит как светящаяся прямоугольная площадка. Разный угол расходимости излучения приводит к изменению соотношения размеров сторон светящегося прямоугольника с увеличением расстояния наблюдения (сравните рис. 9.10*а*, рис. 9.10*б* и рис. 9.10*в*). Анализ рис. 9.10 подтверждает работоспособность рассмотренной модификации оптической схемы 1 (рис. 9.7) при выбранных параметрах моделирования, даже в отсутствие выходного дифракционного микрорельефа.

На рис. 9.11 представлены результаты моделирования оптической схемы 2 (рис. 9.8) при следующих параметрах: фокусное расстояние линзы светодиода — 5 мм; диаметр линзы светодиода — 5 мм; расстояние между светодиодом и торцом пластины — 1 мм; число светодиодов на стороне — 15 (по 15 диодов светят с правой и левой сторон пластины); расстояние между диодами — 5 мм; высота отражающего рельефа — 0,5 мм; число периодов отражающего рельефа — 100 (длина пластины, соответственно, — 100 мм); толщина пластины — 8 мм; число точек на стороне реги-

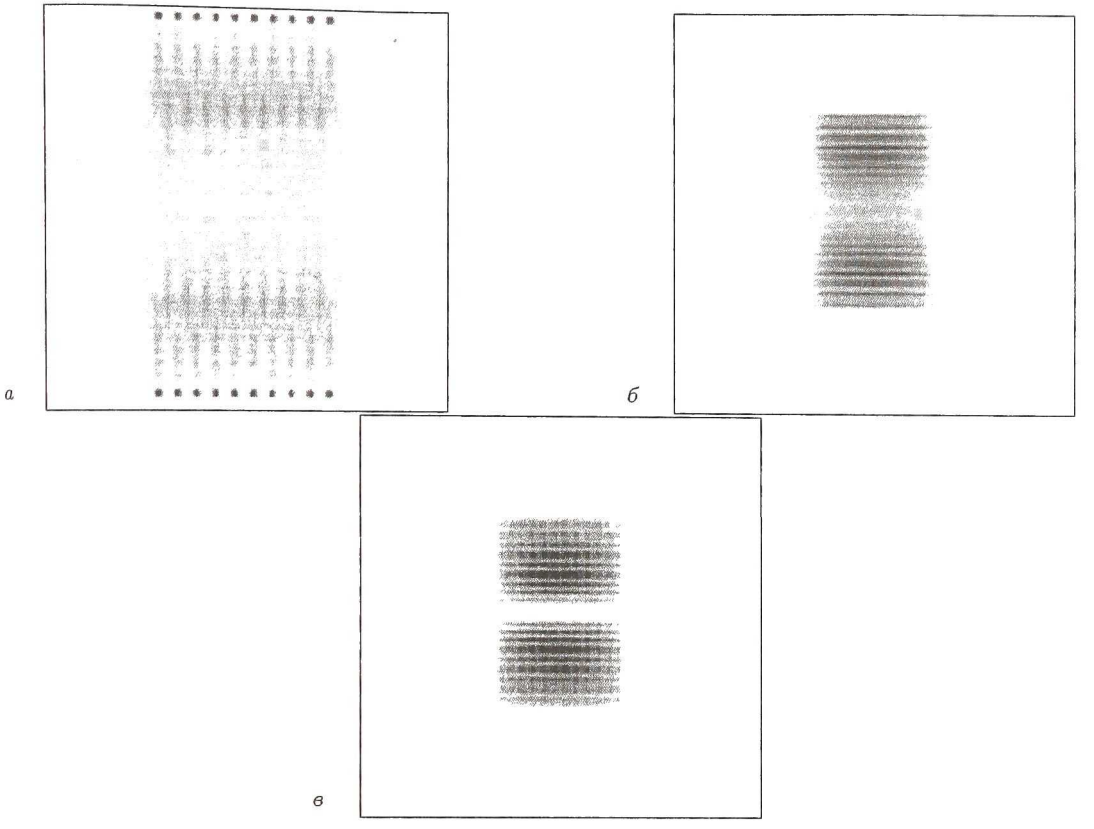


Рис. 9.12. Распределение интенсивности в плоскостях регистрации излучения компланарного осветителя, представленного на рис. 9.9: (*а*) на расстоянии 1 мм; (*б*) на расстоянии 1 м; (*в*) на расстоянии 5 м

стратора — 100; расстояние до первого регистратора (рис. 9.11*а*) — 1 мм; расстояние до второго регистратора (рис. 9.11*б*) — 1 м; расстояние до третьего регистратора (рис. 9.11*в*) — 10 м; угол на внешний край регистратора $\beta = 10°$; число просчитываемых лучей — не менее 5000.

На рис. 9.11*а* представлено распределение интенсивности на расстояние 1 мм от выходной плоскости осветителя, на рис. 9.11*б* — на расстоянии 1 м, а на рис. 9.11*в* — на расстоянии 10 м. Аналогично рис. 9.10 на рис. 9.11*а* хорошо видны места расположения светодиодов, однако с увеличением расстояния (см. рис. 9.11*б* и рис. 9.11*в*) осветитель выглядит как светящаяся прямоугольная площадка. Анализ рис. 9.11 подтверждает работоспособность оптической схемы 2 (рис. 9.8) при выбранных параметрах моделирования.

На рис. 9.12 представлены результаты моделирования упрощенного варианта оптической схемы 3 (рис. 9.9): выходной дифракционный микрорельеф отсутствует, в качестве среды с показателем преломления n_1 используется воздух, $n_1 = 1$, в качестве среды с показателем преломления n_2 используется плексиглас, $n_2 = 1{,}5$. Параметры моделирования: фокусное расстояние линзы светодиода — 5 мм; диаметр линзы светодиода — 5 мм; расстояние между светодиодом и краем гребенки — 5 мм; число светодиодов на стороне — 10 (по 10 диодов светят с правой и левой сторон пластины); расстояние между диодами — 5 мм; толщина слоя с показателем преломления n_2, $\varepsilon = 0{,}4$ мм; длина пластины — 100 мм; число точек на стороне регистратора —

101; расстояние до первого регистратора (рис. 9.12*а*) — 1 мм; расстояние до второго регистратора (рис. 9.12*б*) — 1 м; расстояние до третьего регистратора (рис. 9.12*в*) — 5 м; угол на внешний край регистратора $\beta = 30°$; число просчитываемых лучей — не менее 5000. На рис. 9.12*а* представлено распределение интенсивности на расстоянии 1 мм от выходной плоскости осветителя, на рис. 9.12*б* — на расстоянии 1 м, а на рис. 9.12*в* — на расстоянии 5 м.

Анализ рис. 9.12 подтверждает работоспособность оптической схемы 3 (рис. 9.9) и в упрощенном варианте. В то же время использование при моделировании большего угла на внешний край регистратора ($\beta = 30°$ вместо 10°, для рис. 9.10 и 9.11) приводит к более сильной расходимости выходного пучка света, а ослабление выходного излучения в центре выходного пучка осветителя (особенно заметное на рис. 9.12*в*) говорит о необходимости дальнейшей оптимизации параметров схемы 3 для более равномерного распределения излучения от светодиодов по глубине компланарного осветителя.

Проведенное исследование продемонстрировало работоспособность предложенных оптических схем и послужило основой для создания нескольких образцов компланарных осветителей. Выбор оптимальной оптической схемы может быть осуществлен по другим (не оптическим) критериям: технологичность, дешевизна выпуска, минимальные массогабаритные характеристики и т.п. По совокупности достоинств и результатам оптических испытаний наиболее перспективной представляется оптическая схема 1 (рис. 9.7).

9.6. Результаты испытаний светотехнических устройств с ДОЭ

Для исследования качества работы светотехнических устройств, описанных выше, использовалась автоматизированная установка, схематически представленная на рис. 9.13. Фотография установки для испытания заднего фонаря автомобиля приведена на рис. 9.14.

В качестве устройства ввода фотометрических данных вместо камеры целесообразно использование прецизионных откалиброванных люксметров класса LX-02 (рис. 9.14). Результаты измерения показателей работы светотехнических устройств автомобиля должны удовлетворять требованиям стандартов Европейского Сообщества (ЕС) на яркость центра картины и процент спада интенсивности по углам зрения на установленном расстоянии. В частности, требования по спаду интенсивности излучения для сигнала поворота заднего фонаря должны быть в процентах не ниже, чем представленные в табл. 9.1.

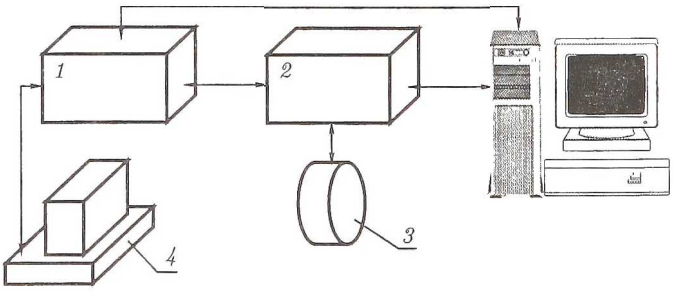


Рис. 9.13. Блок-схема автоматизированной системы обработки фотометрической информации: *1* — модель контроллера шаговых двигателей; *2* — модель аналого-цифрового преобразования; *3* — фотоприемник; *4* — электромеханический привод ориентации

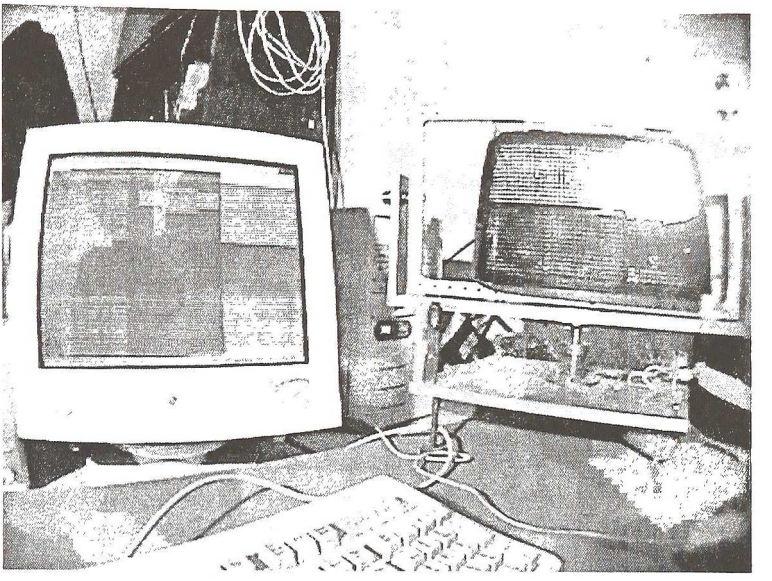


Рис. 9.14. Общий вид автоматизированной системы испытаний светотехнических устройств

Таблица 9.1. Требования стандарта ЕС к фотометрической картине сигнала поворота заднего фонаря автомобиля

Угол по вертикали	Угол по горизонтали						
	−20°	−10°	−5°	0°	5°	10°	20°
10°			20		20		
5°	10	20		70			
0°		35	90	100	90	35	
−5°	10	20		70		20	10
−10°			20		20		

Описанный в разделе 9.1 метод реализации светорассеивателя заднего фонаря автомобиля в виде единой пластинки с наборами скрещенных линз, оригинальных для каждой секции фонаря, был использован для формирования макета. При этом для изготовления микрорельефа с максимальной высотой 50–60 мкм (пропорциональной 50 длинам волн) использовалась технология формирования «квазидифракционного» микрорельефа на основе эффекта темнового роста в слоях жидких фотополимеризующихся композиций (ЖФПК) (см. главу 4). Результаты испытаний одной из секций (сигнал поворота) такого фонаря представлены в табл. 9.2. С учетом того, что яркость в центре картины удовлетворяет требованиям стандарта, общая равномерность засветки картины с большим запасом вписывается в стандарты ЕС.

В качестве дифракционного элемента в оптической схеме фары (рис. 9.4) при испытаниях использовался бинарный квазипериодический ДОЭ, фокусирующий излучение в область, приближенную по форме к форме отверстия диафрагмы. На рис. 9.15 представлена диаграмма направленности, формируемая ячейкой квазипериодического ДОЭ. Расчет ячейки проводился на основе метода согласованных прямоугольников [13], полученная фазовая функция использовалась затем для создания квазипериодического ДОЭ [12]. Использование для испытаний бинарного элемента обусловлено технологической простотой получения бинарного микрорельефа.

Таблица 9.2. Результаты испытаний сигнала поворота, квазидифракционный светорассеиватель которого изготовлен на основе ЖФПК

Угол по вертикали	Угол по горизонтали						
	−20°	−10°	−5°	0°	5°	10°	20°
10°			77,29	78,36	79,71		
5°	72,46	82,61	86,96	91,79	89,86	77,78	67,63
0°	72,46	88,89	97,10	100	94,20	82,13	69,08
−5°	73,43	82,13	91,79	93,72	90,82	79,23	65,22
−10°			79,71	79,71	77,78		

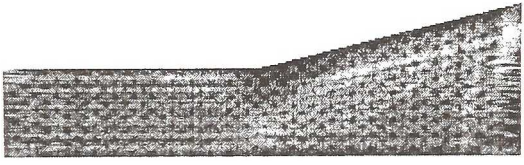


Рис. 9.15. Диаграмма направленности излучения, формируемого одиночной ячейкой квазипериодического ДОЭ

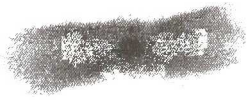


Рис. 9.16. Диаграмма направленности, формируемая оптической схемой фары (рис. 9.4, без диафрагмы) с бинарным квазипериодическим ДОЭ

На рис. 9.16 представлены результаты испытания фары (рис. 9.4) с бинарным квазипериодическим ДОЭ, но без диафрагмы. Симметрия полученной картины обусловлена бинарным характером микрорельефа ДОЭ. При введении в оптическую схему диафрагмы, симметрия диаграммы направленности устраняется, и получаемая картина практически полностью начинает соответствовать требуемой (рис. 9.3). При этом энергетическая эффективность работы фары увеличивается на 12–14% по сравнению с работой осесимметричного эллиптического рефлектора с диафрагмой, но без бинарного ДОЭ. При качественном изготовлении многоуровневого ДОЭ эффективность может быть дополнительно увеличена на 10–15%.

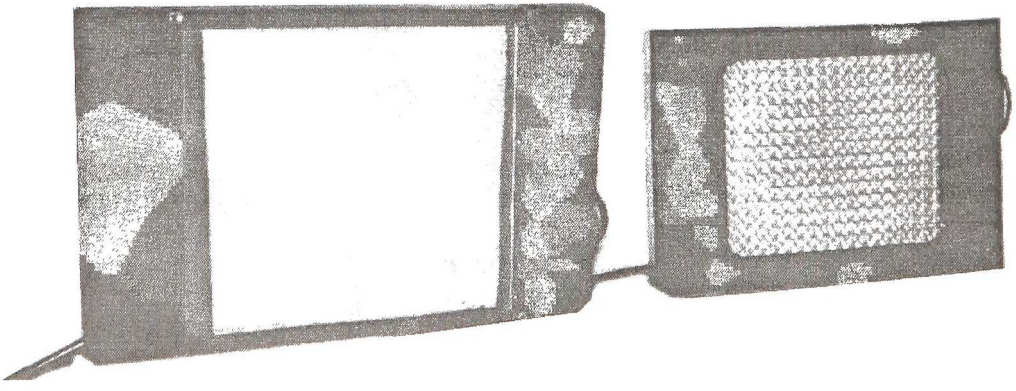


Рис. 9.17. Внешний вид работающих компланарных осветителей, изготовленных на основе оптических схем 1 (рис. 9.7) и 3 (рис. 9.9)

Таблица 9.3. Результаты тестирования компланарной секции поворота заднего фонаря автомобиля, реализованной в соответствии с оптической схемой 3 (рис. 9.9)

Угол по вертикали	Угол по горизонтали						
	−20°	−10°	−5°	0°	5°	10°	20°
10°			32,68	37,96	34,76		
5°	18,78	39,64	61,83	77,26	63,27	40,40	18,04
0°	20,53	46,16	78,67	100	77,06	45,17	18,83
−5°	18,06	34,85	52,84	60,93	50,27	34,29	16,50
−10°			28,63	30,79	27,50		

Для исследования работоспособности компланарных осветителей были изготовлены макетные образцы, реализующие оптические схемы, представленные на рис. 9.7 и рис. 9.9. Целью реализации было изготовление полного макета заднего фонаря автомобиля с минимальными массогабаритными характеристиками, поэтому для каждой секции фонаря был рассчитан оригинальный микрорельеф, обеспечивающий формирование требуемой диаграммы направленности секции. Дифракционный микрорельеф формировался средствами сухой литографии (плазмохимического травления) на основе фотошаблонов, полученных на электронном литографе ZBA-20 фирмы „Карл Цейс“(г. Йена, Германия) и круговой лазерной записывающей машины института автоматики и электрометрии СО РАН (г. Новосибирск). Внешний вид работающих секций, реализующих разные оптические схемы, представлен на рис. 9.17.

В табл. 9.3 представлено измеренное распределение от компланарной секции поворота заднего фонаря автомобиля, реализованного на основе оптической схемы 3 (рис. 9.9). Сравнения с требованиями (табл. 9.1) свидетельствуют об удовлетворении стандарта с большим запасом по равномерности формируемой диаграммы направленности. Характеристики реализации схемы 1 (рис. 9.7) несколько выше.

9.7. Заключение

Основными препятствиями в применении ДОЭ, для уменьшения массогабаритных характеристик и повышения эффективности светотехнических устройств, всегда считалось наличие хроматических аберраций ДОЭ и сложность проектирования оптических устройств с ДОЭ. В настоящее время развитие технологий формирования «квазидифракционного» (или «гармонического») микрорельефа, имеющего высоту, пропорциональную десяткам длин волн, позволяет обойти первое из перечисленных препятствий, сохраняя при этом достоинства ДОЭ. Изложенная в настоящей главе теория может служить основой для моделирования и оптимизации оптических устройств с ДОЭ. Программное обеспечение, разработанное на базе развитой теории, позволило проанализировать и реализовать в виде макетных образцов ряд оригинальных светотехнических устройств, имеющих благодаря использованию ДОЭ высокую эффективность и уникальные функциональные характеристики.

Проведенные натурные исследования макетных образцов подтверждают перспективность использования ДОЭ в светотехнике. Особый выигрыш в массогабаритных характеристиках и эффективности достигается при совместном использовании волноводных, дифракционных и других структур современной оптики так, как это происходит при проектировании компланарных осветителей.

Литература к главе 9

1. Трембач В. В. *Световые приборы: учебное пособие для вузов по специальности «Светотехника и источники света»* (М.: Высшая школа, 1990)
2. Kusch O. *Computer-Aided Optical Design of Illuminating and Irradiating Devices* (Moscow: "ASLAN" Publishing House, 1993)
3. Волков А. В., Волотовский С. Г., Гранчак В. М., Казанский Н. Л., Моисеев О. Ю., Сойфер В. А., Соловьев В. С., Якуненкова Д. М. *Журнал технической физики* **65**(9) 181 (1995)
4. Sweeney D. W., Sommargren G. E. *Applied Optics* **34**(14) 2469 (1995)
5. Голуб М. А., Казанский Н. Л., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Компьютерная оптика* **1** 35 (1989)
6. Казанский Н. Л., Сойфер В. А., Харитонов С. И. *Компьютерная оптика* **14-15** 107 (1995)
7. Борн М., Вольф Э. *Основы оптики* (М.: Наука, 1973)
8. Farn M. W. *OSA Proceedings of the International Optical Design Conference* **22** 246 (1994)
9. Soifer V. A., Doskolovich L. L., Kazanskiy N. L. *Optical Engineering* **33**(11) 3610 (1994)
10. Golub M. A., Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I, Soifer V. A. *Journal of Modern Optics* **39**(6) 1245 (1992)
11. Федорюк М. В. *Асимптотика: Интегралы и ряды* (М.: Наука, 1987)
12. Досколович Л. Л., Казанский Н. Л., Сойфер В. А. *Компьютерная оптика* **13** 16 (1993)
13. Doskolovich L. L., Kazanskiy N. L., Kharitonov S. I., Soifer V. A. *Journal of Modern Optics* **43**(7) 1423 (1996)

ГЛАВА 10

ОПТИЧЕСКАЯ ОБРАБОТКА ИНФОРМАЦИИ С ПРИМЕНЕНИЕМ ДОЭ

10.1. Оптическое формирование признаков изображения

При оптической обработке информации наиболее часто выполняется операция линейной пространственной фильтрации, описываемая уравнением

$$f(\mathbf{u}) = \int\limits_G h(\mathbf{x}, \mathbf{u}) w(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{10.1}$$

где $w(\mathbf{x})$, $f(\mathbf{u})$ — волновые поля на входе и выходе системы соответственно, $h(\mathbf{x}, \mathbf{u})$ — импульсный отклик или функция рассеяния системы, G — область интегрирования.

Когерентная оптика представляет широкие возможности для проведения в реальном времени пространственной фильтрации световых полей. Наиболее просто реализуются пространственно-инвариантные фильтры, для которых имеет место уравнение свертки:

$$f(\mathbf{u}) = \int\limits_G h(\mathbf{x} - \mathbf{u}) w(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{10.2}$$

Пример когерентно-оптической системы, реализующей пространственно-инвариантную фильтрацию, приведен на рис. 10.1.

Входная амплитуда модулируется транспарантом $w(\mathbf{x})$ в плоскости Π_1. Объектив O_1 с фокусным расстоянием f_0 формирует в частотной плоскости Ф двумерный

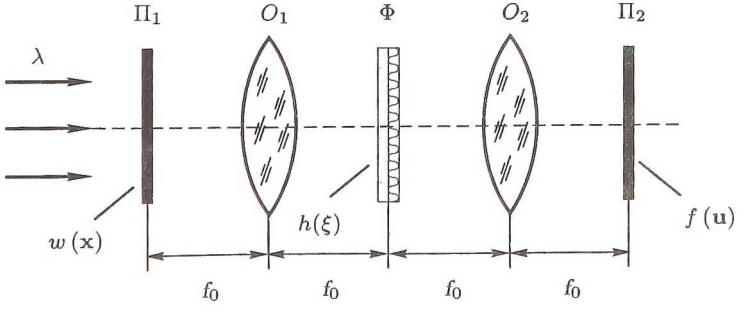


Рис. 10.1. Когерентно-оптическая система, реализующая пространственно-инвариантную фильтрацию

фурье–образ поля $w(\mathbf{x})$:

$$W(\nu) = \int\limits_G w(\mathbf{x}) \exp\left[-i2\pi(\mathbf{x}\boldsymbol{\nu})\right] \mathrm{d}\mathbf{x}. \tag{10.3}$$

Здесь $\boldsymbol{\nu} = (\nu_x, \nu_y)$ — пространственные частоты, связанные с координатами $\mathbf{u}$ в плоскости Ф формулой

$$\mathbf{u} = (u, v) = (\lambda f_0 \nu_x, \lambda f_0 \nu_y).$$

Фурье-образ $W(\boldsymbol{\nu})$ умножается на функцию пропускания транспаранта

$$H(\boldsymbol{\nu}) = \int\limits_G h(\mathbf{x}) \exp\left[-i2\pi(\mathbf{x}\boldsymbol{\nu})\right] \mathrm{d}\mathbf{x}, \tag{10.4}$$

установленного в частотной плоскости. Сразу за плоскостью Ф возникает поле

$$F(\boldsymbol{\nu}) = W(\boldsymbol{\nu}) H(\boldsymbol{\nu}). \tag{10.5}$$

Как видно из формулы (10.5), транспарант отображает частотную характеристику $H(\boldsymbol{\nu})$ фильтра (10.4). Далее объектив O_2 осуществляет обратное преобразование Фурье. Для этого на выходе фурье-коррелятора в плоскости Π_2, оси u и ν должны быть направлены в противоположные стороны по отношению к осям x и y в плоскости Π_1. По теореме о свертке в плоскости Π_2 формируется поле $f(\mathbf{u})$.

Описанная система фильтрации далеко не исчерпывает всего многообразия схем, реализующих операции (10.2) методами когерентной оптики. Однако общим и основным элементом таких систем является пространственный фильтр, характеризующийся функцией пропускания, зависящей от координат плоскости.

Операции фильтрации встречаются при решении задач согласованной фильтрации [1, 2], улучшения качества изображений [3], выделения сигналов на фоне шумов [4], распознавания изображений [5–7].

Излагаемый в данном разделе подход к распознаванию изображений ориентирован на когерентно-оптическое вычисление коэффициентов ортогонального разложения с помощью оптических пространственных фильтров, синтезированных на компьютере [1, 7–9] с последующим использованием коэффициентов в цифровой процедуре классификации [10] как набора признаков изображения [11–13].

Пусть $\{\psi_{\mathbf{p}}(\mathbf{x})\}$ — базисные функции разложения в ортогональный ряд. Задачу измерения коэффициентов разложения $\{w_{\mathbf{p}}\}$ пучка с комплексной амплитудой

$$w(\mathbf{x}) = \sum_{\mathbf{p}} w_{\mathbf{p}} \psi_{\mathbf{p}}(\mathbf{x}), \quad \mathbf{x} \in G, \tag{10.6}$$

предлагается решать в два этапа [14]. Первый этап представляет собой оптическое разложение поля по базисным функциям на основе их свойства ортогональности. При этом распределение мощности по базисным функциям преобразуется в распределение интенсивности света в выходной фокальной плоскости. Второй этап состоит в измерении фотоприемниками с точечными апертурами интенсивности света, которая пропорциональна квадрату модуля каждого коэффициента $\sim |w_{\mathbf{p}}|^2$.

Покажем, что при установке на вход фурье-каскада пространственного фильтра с комплексной функцией пропускания $\psi^*_{\mathbf{p}}(\mathbf{x})$ и подаче на вход поля $w(\mathbf{x})$, в центре выходной плоскости получается коэффициент $w_{\mathbf{p}}$ одной из базисных функций (рис. 10.2).

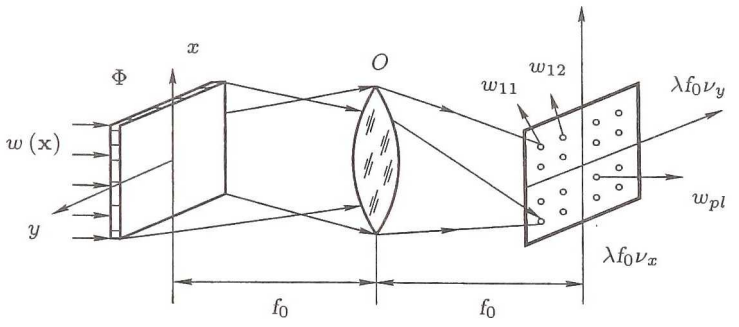


Рис. 10.2. Оптическая схема для параллельного вычисления набора коэффициентов разложения светового поля по ортогональному базису

Действительно, сразу за плоскостью (x, y) получим поле

$$w(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x}) = \sum_{\mathbf{p}'} w_{\mathbf{p}'}\psi_{\mathbf{p}'}(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x}). \tag{10.7}$$

Преобразование Фурье (10.3), взятое в центре $\boldsymbol{\nu} = 0$ выходной плоскости пространственных частот, дает операцию интегрирования комплексной амплитуды. В силу ортогональности взаимодействие базисных функций происходит так, что при интегрировании комплексной амплитуды (10.7) в точке $\boldsymbol{\nu} = 0$ фурье-плоскости фон от базисных функций с номерами $\mathbf{p}' \neq \mathbf{p}$ взаимно уничтожается и получаются значения комплексной амплитуды, равные коэффициенту разложения:

$$w_{\mathbf{p}} = \int\limits_G w(\mathbf{x})\psi^*_{\mathbf{p}}(\mathbf{x})\,\mathrm{d}^2\mathbf{x}. \tag{10.8}$$

Мощность $|w_{\mathbf{p}}|^2$ базисной функции $\psi_{\mathbf{p}}$ может быть измерена уже обычным фотоприемником с точечной апертурой. Последовательно меняя фильтры, формирующие пучки с амплитудой $\psi^*_{\mathbf{p}}$, можно измерить весь состав базисных функций. Для параллельного вычисления набора коэффициентов и их разнесения в выходной плоскости (см. рис. 10.2) используется метод пространственной несущей [14]. Далее этот метод будет рассмотрен для конкретных базисов (пункты 10.2.2–10.2.4).

10.2. Разложение светового поля по ортогональному базису

10.2.1. Оптимальный базис Карунена–Лоэва

Для того, чтобы представление анализируемых изображений при распознавании было наиболее информативно, коэффициенты разложения должны быть независимы (или хотя бы некоррелированы) в статистическом смысле [15]. Для случайного поля с заданной корреляционной функцией, как известно [16, 17], существует единственный ортогональный базис, обеспечивающий некоррелированные коэффициенты разложения. Это базис Карунена–Лоэва [15, 18], который состоит из собственных функций интегрального уравнения с ядром, являющимся корреляционной функцией рассматриваемого поля. Некоррелированные коэффициенты разложения изображения по базису Карунена–Лоэва могут эффективно использоваться в качестве признаков изображения.

При компьютерном синтезе оптических пространственных фильтров, согласованных с базисными функциями ортогонального разложения, важно иметь их аналитическое представление. Особенно это актуально для разложения Карунена–Лоэва (РКЛ), так как численное решение интегрального уравнения Карунена–Лоэва очень трудоемко и неустойчиво при размерах фильтра, применяемых на практике (более чем 103×103 пикселов). В работе [19] дан общий подход к аналитическому нахождению базисных функций Карунена–Лоэва для дробно-рационального спектра. Однако он требует выполнения значительного количества громоздких операций типа вычисления комплексных определителей и неудобен для алгоритмизации. Известны [20] простые формулы для случая экспоненциальной корреляционной функции, но, учитывая ее важность для построения базиса Карунена–Лоэва, желательно иметь более точную — экспоненциально-косинусную [21, 22] аппроксимацию корреляционной функции. В данном разделе рассматриваются аналитические формулы для базисных функций Карунена–Лоэва в случае экспоненциально-косинусной корреляционной функции.

При оптической реализации ортогональных разложений возникает неизбежное искажение базисных функций, вносимое дискретизацией по аргументу при компьютерном синтезе ДОЭ и квантованием по уровню при его последующем изготовлении. Поэтому нужно учесть влияние возмущений базисных функций на качество распознавания. При этом особенно важно выяснить, как влияет оптическая реализация РКЛ [23, 24] на сохранение его оптимальных свойств. Полученные в работе [25] теоретические оценки описанных возмущений требуют трудоемких вычислений норм бесконечных матриц и других сложных операций. В этом разделе проведены экспериментальные исследования и построены численные зависимости влияния таких возмущений на устойчивость РКЛ.

Базисные функции РКЛ находятся по корреляционной функции распознаваемых изображений путем решения интегрального уравнения на собственные значения:

$$\int\limits_G \bar{R}(\mathbf{x},\mathbf{u})\psi_{\mathbf{p}}(\mathbf{x})\,\mathrm{d}^2\mathbf{u} = \lambda_{\mathbf{p}}\psi_{\mathbf{p}}(\mathbf{x}), \tag{10.9}$$

где

$$\bar{R}(\mathbf{x},\mathbf{u}) = \sum_{i=1}^{N_{cl}} R_i(\mathbf{x},\mathbf{u})P_i, \quad \sum_{i=1}^{N_{cl}} P_i = 1, \tag{10.10}$$

$\overline{R}$ — усредненная [13] по N_{cl} классам корреляционная функция, $R_i(\mathbf{x},\mathbf{u})$ — корреляционная функция i-го класса распознаваемых изображений, P_i — вероятность появления класса, $\lambda_{\mathbf{p}}$ — собственные значения с номером $\mathbf{p} = (p,l)$.

Кроме того, РКЛ обладает еще целым набором замечательных свойств, ставящих его на особое место среди всевозможных разложений случайных полей

1) Коэффициенты разложения некоррелированны:

$$\left\langle w_{\mathbf{p}}, w^*_{\mathbf{p}'}\right\rangle = \left(\widehat{R}\left[\psi_{\mathbf{p}}\right],\psi_{\mathbf{p}}\right) = (\lambda_{\mathbf{p}}\psi_{\mathbf{p}},\psi_{\mathbf{p}}) = \lambda_{\mathbf{p}}\delta_{\mathbf{pp}}, \tag{10.11}$$

где

$$\hat{R}[\psi](\mathbf{x}) = \int\limits_G R(\mathbf{x},\mathbf{u})\psi(\mathbf{u})\,\mathrm{d}^2\mathbf{u} \tag{10.12}$$

— корреляционный оператор.

В [13] показано, что свойство некоррелированности линейных признаков, имеющее место при использовании базисных функций Карунена–Лоэва, благотворно сказывается на качестве распознавания и эффекте кластеризации [11].

2) Среднеквадратичная ошибка усечения бесконечного ряда минимальна среди всех ортогональных базисов [13]. Представим поле $w(\mathbf{x})$ приближенно конечным числом L членов ряда:

$$w(\mathbf{x}) \cong \widehat{w}_L(\mathbf{x}) \equiv \sum_{\mathbf{p}\in K} w_{\mathbf{p}}\psi_{\mathbf{p}}(\mathbf{x}) + \bar{w}(\mathbf{x}), \quad |K| = L. \tag{10.13}$$

Погрешность такой аппроксимации имеет мощность

$$e_0^2(L) = \left\langle \int\limits_G |w(\mathbf{x}) - \hat{w}_L(\mathbf{x})|\, \mathrm{d}^2\mathbf{x} \right\rangle. \tag{10.14}$$

Заметим, что собственные значения упорядочены в последовательности $\{\lambda_{\mathbf{p}}\}$ по убыванию, и соответствующая им последовательность базисных функций дает наименьшую погрешность аппроксимации конечным числом членов ряда.

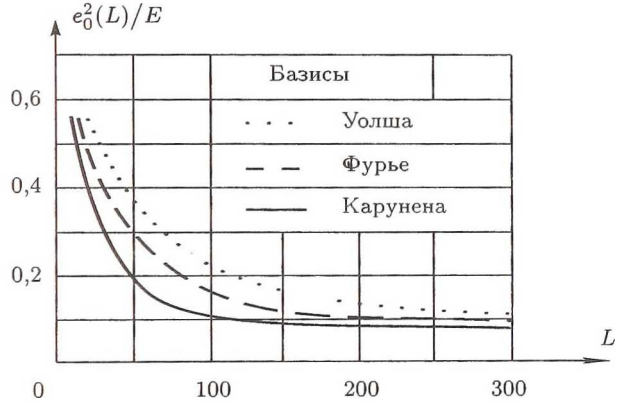


Рис. 10.3. Сравнение между некоторыми ортогональными базисами по погрешности аппроксимации конечным числом членов ряда

Сравнение между некоторыми ортогональными базисами, данное в [25], приведено на рис. 10.3 и подтверждает свойство минимальности погрешности.

3) Энтропия, определенная на множестве нормированных дисперсий коэффициентов РКЛ минимальна. Доказательство и строгая формулировка этого свойства дана в [15].

4) Наибольшее собственное значение λ удовлетворяет неравенству

$$\lambda \equiv \max_{\mathbf{p}} \lambda_{\mathbf{p}} \leqslant \sup_{\mathbf{p}\in G} S(\boldsymbol{\omega}), \tag{10.15}$$

где $S(\boldsymbol{\omega}) = \int\limits_G R(\mathbf{x}) \exp(-i\boldsymbol{\omega}\mathbf{x})\, \mathrm{d}^2\mathbf{x}$ — спектральная плотность, $\boldsymbol{\omega} = (\omega_1, \omega_2)$.

5) Монотонные собственные значения. Собственное значение $\lambda_{\mathbf{p}}(G)$ есть монотонно возрастающая функция размеров области, т. е.

$$\lambda_{\mathbf{p}}(G) \leqslant \lambda_{\mathbf{p}}(G'), \quad G \subset G'. \tag{10.16}$$

6) Асимптотическое поведение при расширении области G. На бесконечной области $G = R^n$ собственные функции и собственные значения принимают вид

$$\psi(\mathbf{x}) = \exp(i\boldsymbol{\omega}\mathbf{x}), \quad \lambda = S(\boldsymbol{\omega}). \tag{10.17}$$

7) Факторизация. Пусть корреляционная функция факторизуется, т. е. $R(\mathbf{x}, \mathbf{u}) = R^{\mathrm{I}}(\mathbf{x}^{\mathrm{I}}, \mathbf{u}^{\mathrm{I}}) \cdot R^{\mathrm{II}}(\mathbf{x}^{\mathrm{II}}, \mathbf{u}^{\mathrm{II}})$. Легко видеть, что собственные числа и собственные функции также факторизуются и могут быть упорядочены так, что $\lambda_{\mathbf{k}_m \mathbf{l}_m} \geqslant \lambda_{\mathbf{k}_{m+1} \mathbf{l}_{m+1}}$, $m = 0, 1, 2, \ldots$:

$$\begin{aligned} \lambda_{\mathbf{p}_m} &= \lambda^{\mathrm{I}}_{\mathbf{k}_m} \cdot \lambda^{\mathrm{II}}_{\mathbf{l}_m}, \\ \psi_{\mathbf{p}_m}(\mathbf{x}) &= \psi^{\mathrm{I}}_{\mathbf{k}_m}(\mathbf{x}^{\mathrm{I}}) \cdot \psi^{\mathrm{II}}_{\mathbf{l}_m}(\mathbf{x}^{\mathrm{II}}), \quad m = 0, 1, 2, \ldots \end{aligned} \tag{10.18}$$

Для оптического вычисления некоррелированных коэффициентов РКЛ методами цифровой голографии [1, 7, 8] синтезируются ДОЭ, согласованные с базисными функциями Карунена–Лоэва. Применение оптического РКЛ [23, 26] ставит задачу исследования устойчивости РКЛ к возмущениям, обусловленным дискретизацией базисных функций при компьютерном синтезе оптического фильтра, а также квантованием по уровню при его последующем литографическом изготовлении [27].

При компьютерном синтезе оптических пространственных фильтров, согласованных с РКЛ, важно иметь аналитические представления базисных функций.

Известно [21, 22], что корреляционная функция стационарных случайных процессов с дробно-рациональным спектром имеет общий вид

$$R(\tau) = \sigma^2 \sum_k \exp(-\alpha_k |\tau|) \cdot [A_k(\tau) \cos \beta_k \tau + B_k(\tau) \sin \beta_k |\tau|], \tag{10.19}$$

где $A_k(\tau)$, $B_k(\tau)$ — многочлены относительно τ; σ^2 α_k, β_k — параметры.

Как видно из (10.19) экспоненциальная аппроксимация представляет собой только первое приближение общего вида корреляционной функции. Учитывая важность корреляционной функции для построения РКЛ, желательно иметь ее более точную аппроксимацию. Принимая во внимание (10.19), предлагается описывать корреляционную функцию более точной экспоненциально-косинусной формулой и найти для этого случая такое же простое аналитическое представление РКЛ, как и для экспоненциального [20].

В случае анализа изображений применяются двумерное РКЛ. При гипотезе о факторизации корреляционной функции [14] можно рассматривать одномерное интегральное уравнение на собственные значения на симметричном интервале $x \in [-A, A]$:

$$\int_{-A}^{A} R(x - u)\psi(u)\,\mathrm{d}u = \lambda\psi(x). \tag{10.20}$$

Экспоненциально-косинусная корреляционная функция

$$R(\tau) = \sigma^2 \exp(-\alpha |\tau|) \cos \beta\tau, \tag{10.21}$$

имеет спектр

$$G(\omega) = \frac{2\alpha\sigma^2 \left(\alpha^2 + \beta^2 + \omega^2\right)}{\omega^4 + 2\omega^2 \left(\alpha^2 - \beta^2\right) + \left(\alpha^2 + \beta^2\right)^2}, \tag{10.22}$$

где ω — пространственная частота.

В соответствии с алгоритмом решения интегрального уравнения II-го рода для ядра с дробно-рациональным спектром [19], интегральное уравнение (10.20) сводится к линейному дифференциальному уравнению с постоянными коэффициентами:

$$\psi^{(4)}(x) - 2\left(\alpha^2 - \beta^2 - \frac{\alpha\sigma^2}{\lambda}\right)\psi^{(2)}(x) + \left(\alpha^2 + \beta^2\right)\left(\alpha^2 + \beta^2 - \frac{2\alpha\sigma^2}{\lambda}\right)\psi(x) = 0. \tag{10.23}$$

Решением этого уравнения являются собственные функции $\psi_k(x)$ вида

$$\psi_k(x) = \begin{cases} d_{1k}\cos\left(\varepsilon_k\dfrac{x}{A}\right) + d_{2k}\operatorname{ch}\left(\gamma_k\dfrac{x}{A}\right), & k = 0, 2, 4, \ldots \\ c_{1k}\sin\left(\varepsilon_k\dfrac{x}{A}\right) + c_{2k}\operatorname{sh}\left(\gamma_k\dfrac{x}{A}\right), & k = 1, 3, 5, \ldots \end{cases} \tag{10.24}$$

Здесь ε_k и γ_k являются решениями системы трансцендентных уравнений [28]:

$$\left.\begin{aligned} G_1(\gamma)G_2(i\varepsilon)\left\{\left(\frac{\alpha}{\varepsilon}F_1(i\varepsilon)\right)^P \pm \operatorname{tg}\varepsilon\right\}\left\{\left(F_2(\gamma)\right)^P + \operatorname{th}\gamma\right\} - \\ -G_1(i\varepsilon)G_2(\gamma)\left\{\left(\frac{\alpha}{\gamma}F_1(\gamma)\right)^P - \operatorname{th}\gamma\right\}\left\{\left(F_2(i\varepsilon)\right)^P \mp \operatorname{tg}\varepsilon\right\} = 0 \\ \gamma^2 = \left(\alpha^2 + \beta^2\right)\frac{\alpha^2 - 3\beta^2 + \varepsilon^2}{\alpha^2 + \beta^2 + \varepsilon^2}, \end{aligned}\right\} \tag{10.25}$$

где $F_1(x) = \dfrac{\beta^2 + \left(\alpha^2 - x^2\right)}{\beta^2 - \left(\alpha^2 - x^2\right)}$, $F_2(x) = \dfrac{\alpha^2 + \beta^2 + x^2}{2\alpha x}$, для $k = 0, 2, 4, \ldots$: $G_1(x) = 1$, $G_2(x) = \beta^2 - \left(\alpha^2 - x^2\right)$, $P = 1$, и используется верхний знак; для $k = 1, 3, 5, \ldots$: $G_1(x) = \alpha^2 + \beta^2 + x^2$, $G_2(x) = \alpha^2 + \beta^2 - x^2$, $P = -1$, и используется нижний знак.

Решения системы (10.25) могут быть отделены на непересекающихся интервалах:

$$\left(0, \frac{\pi}{2}\right),\ k = 0; \quad \left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \ldots, \left(\frac{\pi(k-1)}{2}, \frac{\pi(k+1)}{2}\right),\ k = 2, 4, 6, \ldots$$

$$\left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \left(\frac{3\pi}{2}, \frac{5\pi}{2}\right), \ldots, \left(\frac{\pi(k)}{2}, \frac{\pi(k+2)}{2}\right),\ k = 1, 3, 5, \ldots$$

Коэффициенты d_{1k}, d_{2k} и c_{1k}, c_{2k} определяются из условия нормировки:

$$\left\{\begin{array}{l} d_{1k},\ k = 0, 2, 4, \ldots \\ c_{1k},\ k = 1, 3, 5, \ldots \end{array}\right\} = \left\{A\left[\left(1 \pm \frac{\sin 2\varepsilon_k}{2\varepsilon_k}\right) + 4D_k^2\left(1 \pm \frac{\operatorname{sh} 2\gamma_k}{2\gamma_k}\right) - \frac{4D_kB_k}{\varepsilon_k^2 + \gamma_k^2}\right]\right\}^{-1/2},$$

$$\left\{\begin{array}{l} d_{2k},\ k = 0, 2, 4, \ldots \\ c_{2k},\ k = 1, 3, 5, \ldots \end{array}\right\} = -D_k\left\{\begin{array}{l} d_{1k},\ k = 0, 2, 4, \ldots \\ c_{1k},\ k = 1, 3, 5, \ldots \end{array}\right\}, \tag{10.26}$$

где $D_k = 2\dfrac{E_{1k}\left(\alpha^2 + \beta^2 - \gamma_k^2\right)}{E_{2k}\left(\alpha^2 + \beta^2 + \varepsilon_k^2\right)}$,

$$B_k = \left\{\begin{array}{ll} e^{\gamma_k}(\gamma_k\cos\varepsilon_k + \varepsilon_k\sin\varepsilon_k) - e^{-\gamma_k}(\gamma_k\cos\varepsilon_k - \varepsilon_k\sin\varepsilon_k), & k = 0, 2, 4, \ldots, \\ e^{\gamma_k}(\gamma_k\sin\varepsilon_k + \varepsilon_k\cos\varepsilon_k) - e^{-\gamma_k}(\gamma_k\sin\varepsilon_k - \varepsilon_k\cos\varepsilon_k), & k = 1, 3, 5, \ldots, \end{array}\right.$$



Рис. 10.4. Полутоновое тестовое изображение (отпечаток пальца)

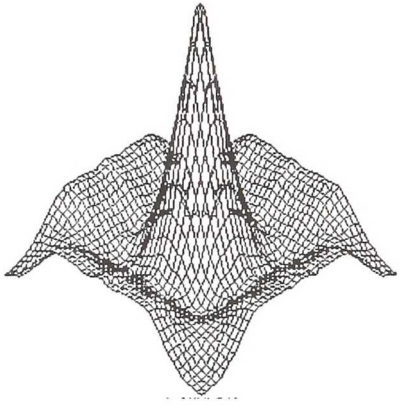


Рис. 10.5. Трехмерный вид корреляционной функции изображения, представленного на рис. 10.4

$$E_{1k}\left\{\begin{array}{l} k=0,2,4,\ldots \\ k=1,3,5,\ldots \end{array}\right\} = \left(\alpha^2+\beta^2-\varepsilon_k^2\right)\left\{\begin{array}{l} \cos\varepsilon_k \\ \sin\varepsilon_k \end{array}\right\} + 2\alpha\varepsilon_k\cos\varepsilon_k,$$

$$E_{2k}\left\{\begin{array}{l} k=0,2,4,\ldots \\ k=1,3,5,\ldots \end{array}\right\} = \left(\alpha^2+\beta^2+\gamma_k^2\right)\left\{\begin{array}{l} \left(e^{\gamma_k}+e^{-\gamma_k}\right) \\ \left(e^{\gamma_k}-e^{-\gamma_k}\right) \end{array}\right\} + 2\alpha\gamma_k\left(e^{\gamma_k}+e^{-\gamma_k}\right).$$

При этом собственные значения находятся из выражения

$$\lambda_k = \frac{2A\alpha\sigma^2\left(\alpha^2+\beta^2+\varepsilon_k^2\right)}{\varepsilon_k^4+2\varepsilon_k^2\left(\alpha^2-\beta^2\right)+\left(\alpha^2+\beta^2\right)^2}. \tag{10.27}$$

Анализ полученных результатов показывает, что монотонно возрастающие $\{\varepsilon_k\}$ и $\{\gamma_k\}$ определяют строго убывающую последовательность собственных значений (10.27), что в соответствии с теорией РКЛ [13] позволяет обеспечить выполнение одного из важнейших оптимальных свойств РКЛ — минимальность среднеквадратичной ошибки усечения бесконечного ряда.

При $\beta = 0$ приходим к случаю экспоненциальной корреляционной функции и формулы (10.24)–(10.27) сводятся к ранее полученным в [20] решениям. Таким образом, при $\beta \neq 0$ получено обобщение на случай экспоненциально-косинусной корреляционной функции.

Пример 10.1. Рассмотрим двумерное полутоновое изображение. Для представленного на рис. 10.4 центрированного изображения $w(x, y)$ была рассчитана двумерная корреляционная функция $R(\Delta x, \Delta y)$ (рис. 10.5) по формуле [29]

$$R(\Delta x, \Delta y) = \frac{1}{4AB} \int_{-A}^{A} \int_{-B}^{B} w(x, y) w^*(x + \Delta x, y + \Delta y)\, \mathrm{d}x\, \mathrm{d}y, \tag{10.28}$$

причем, при выходе аргументов $x + \Delta x$, $y + \Delta y$ за пределы области изображения полагалось $w(x + \Delta x, y + \Delta y) = 0$.

Затем двумерная корреляционная функция $R(\Delta x, \Delta y)$ была факторизована на две одномерные корреляционные функции $R_1(\Delta x)$ и $R_2(\Delta x)$, для которых мето-

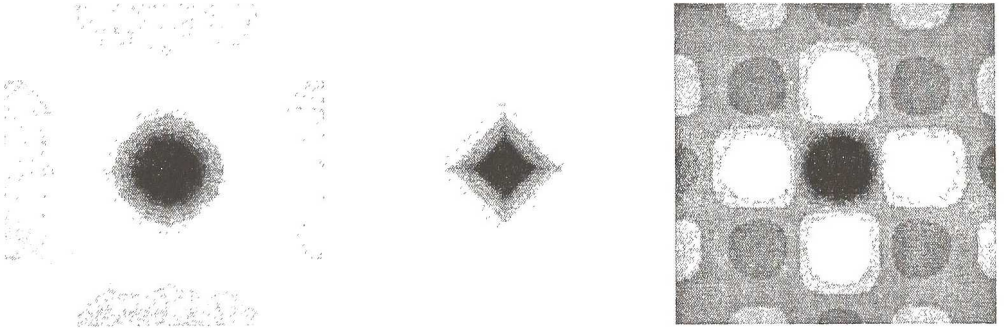


Рис. 10.6. Рассчитанная корреляционная функция (*а*), ее экспоненциальная (*б*) и экспоненциально-косинусная аппроксимации (*в*), соответственно

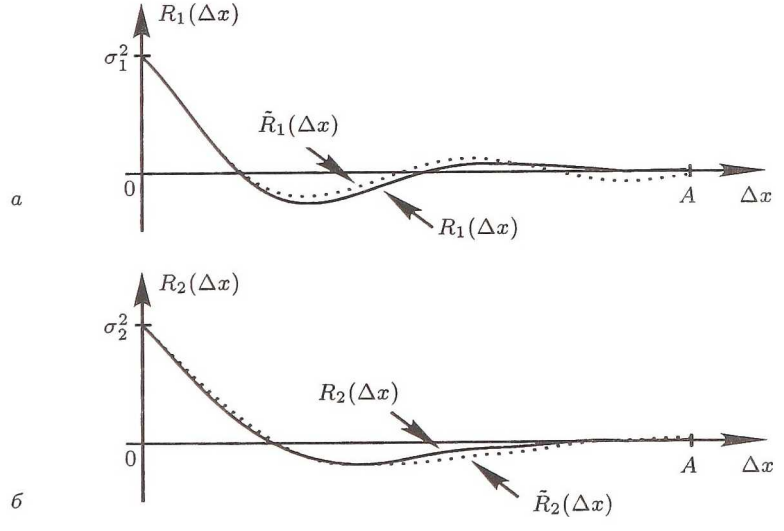


Рис. 10.7. Одномерные сечения рассчитанных корреляционных функций (сплошные линии) и их экспоненциально-косинусные аппроксимации (пунктирные линии)

дом наименьших квадратов были построены экспоненциальная и экспоненциально-косинусная аппроксимации. Для сравнения на рис. 10.6*а* приведена рассчитанная корреляционная функция, а на рис. 10.6*б* и 10.6*в* показаны ее экспоненциальная и экспоненциально-косинусная аппроксимации соответственно.

На рис. 10.6 видно, что экспоненциально-косинусная аппроксимация лучше описывает центральную часть, чем экспоненциальная. А на рис. 10.7 видно, что она еще и отражает осцилляции корреляционной функции.

Для аппроксимирующих корреляционных функций изображения, представленного на рис. 10.4, были получены следующие оценки параметров экспоненциально-косинусной аппроксимации: $\tilde{R}_1(\Delta x)$: $\sigma = 14{,}82$, $\alpha = 2{,}87$, $\beta = 0{,}41$, $\tilde{R}_2(\Delta x)$: $\sigma = 7{,}85$, $\alpha = 2{,}52$, $\beta = 0{,}86$.

Результаты решения интегрального уравнения для ядра $\tilde{R}_1(\Delta x)$ и $\tilde{R}_2(\Delta x)$ приведены в табл. 10.1 и табл. 10.2 соответственно.

Таблица 10.1. Параметры базисных функций для корреляционной функции $\tilde{R}_1(\Delta x)$

Номер базисной функции, k	ε_k	γ_k	Собственное значение, λ_k
0	1,495	17,368	64,726
1	2,987	17,383	64,088
2	4,483	17,406	63,069
3	5,981	17,437	61,730
4	7,480	17,473	60,147

Таблица 10.2. Параметры базисных функций для корреляционной функции $\tilde{R}_2(\Delta x)$

Номер базисной функции, k	ε_k	γ_k	Собственное значение, λ_k
0	1,513	13,349	38,387
1	3,024	13,440	37,341
2	4,529	13,582	35,620
3	6,030	13,759	33,268
4	7,529	13,960	30,363

Вид базисных функций Карунена–Лоэва $\psi_k(x)$ на интервале $x \in [-A, A]$ приведен на рис. 10.8.

С ростом индекса k собственные числа становятся приблизительно кратными $\pi/2$, а собственные функции хорошо аппроксимируются периодическими синусоидами и косинусоидами, то есть имеет место асимптотическое поведение. Однако, учитывая информационную важность первых коэффициентов РКЛ, асимптотическую аппроксимацию представляется возможным использовать только для достаточно высоких индексов.

При оптической реализации РКЛ наиболее общие методы синтеза пространственных фильтров дает цифровая голография. Функция $H(\nu)$ из (10.3), описывающая фильтр, является, вообще говоря, комплексной. Соответственно, для синтеза комплексных пространственных фильтров применимы методы синтезированных голо-

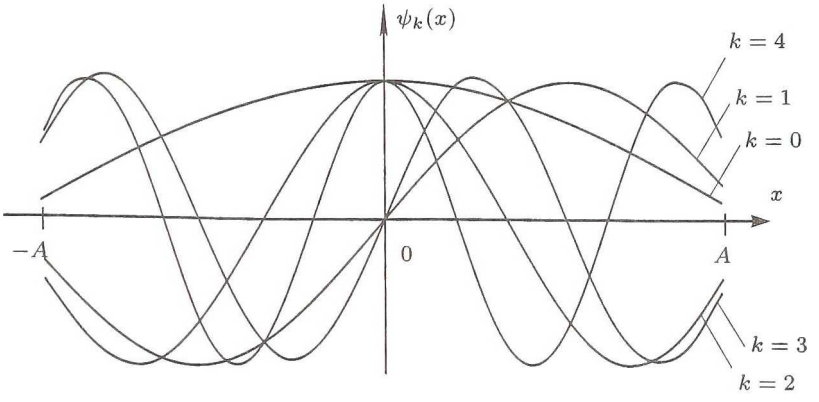


Рис. 10.8. Вид одномерных базисных функций Карунена–Лоэва

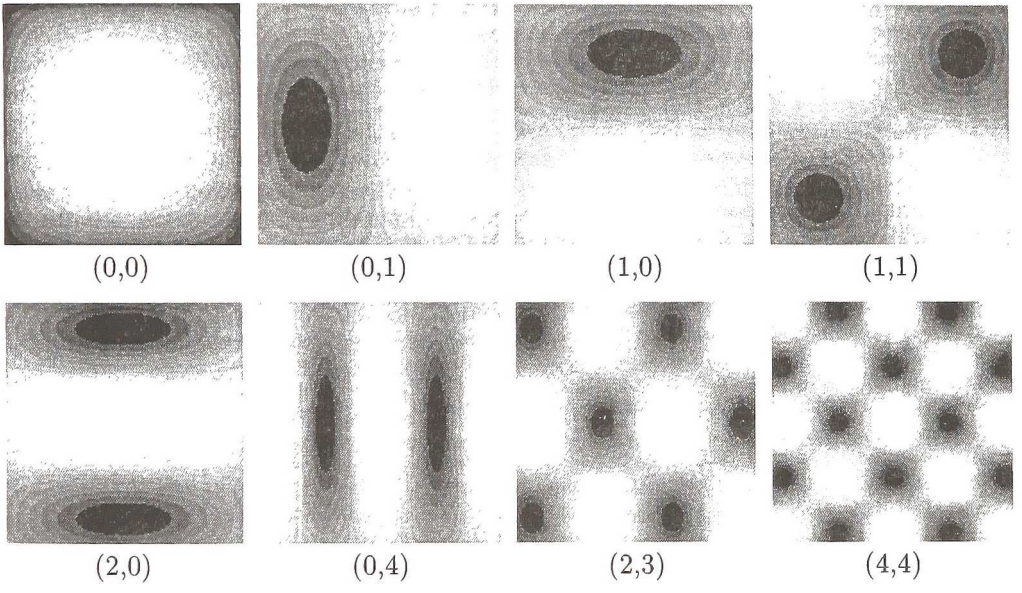


Рис. 10.9. Вид двумерных базисных функций Карунена–Лоэва

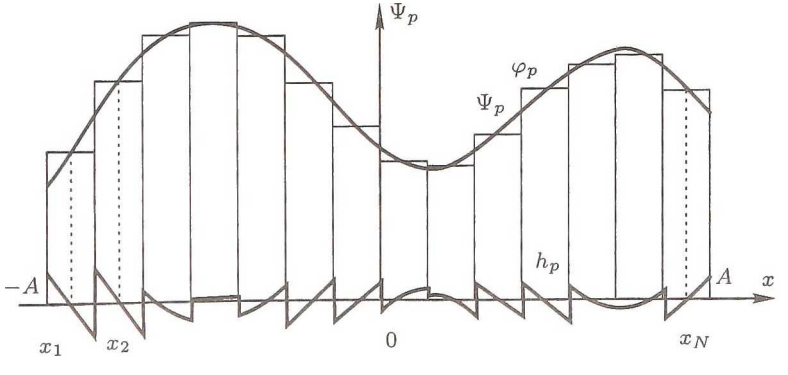


Рис. 10.10. Влияние дискретизации и квантования на возмущение базисных функций

грамм. Фотокопии пространственных фильтров, реализующих базисные функции Карунена–Лоэва для экспоненциально-косинусного вида корреляционной функции ($\alpha_x = \alpha_y = 3$, $\beta_x = \beta_y = 1$) приведены на рис. 10.9.

В связи с применением оптического РКЛ при анализе изображений необходимо учитывать влияние ошибок оптической реализации на устойчивость РКЛ.

При дискретизации и квантовании по уровню собственные функции $\psi_{\mathbf{p}}$ подвергаются воздействию возмущений $h_{\mathbf{p}}$ и преобразуются в «возмущенные» базисные функции $\varphi_{\mathbf{p}}$ (см. рис. 10.10):

$$\varphi_{\mathbf{p}}(\mathbf{x}) = \psi_{\mathbf{p}}(\mathbf{x}) + h_{\mathbf{p}}(\mathbf{x}). \tag{10.29}$$

При наличии возмущений базисных функций, вместо коэффициентов $w_{\mathbf{p}}$ РКЛ получим «возмущенные» коэффициенты

$$\hat{w}_{\mathbf{p}} = \int\limits_G w(\mathbf{x})\varphi^*_{\mathbf{p}}(\mathbf{x})\,\mathrm{d}^2\mathbf{x}, \quad \mathbf{p} \in L, \tag{10.30}$$

которые являются уже слабокоррелированными.

В [14] получены теоретические оценки сверху для возмущений базисных функций Карунена–Лоэва, однако, они требуют трудоемких вычислений норм бесконечных матриц и других операций. Поэтому в данном разделе для определения диапазона устойчивости РКЛ в конкретно поставленных задачах были построены экспериментальные зависимости влияния таких возмущений.

Так как выбор РКЛ среди других базисов определялся его оптимальными свойствами, естественно проследить влияние возмущений, возникающих при его оптической реализации, именно на эти свойства. Сохранение оптимальных свойств РКЛ при внесении возмущений будет говорить об устойчивости РКЛ к данным возмущениям. Среди всех свойств РКЛ выделим два основных: ортонормированность базисных функций и некоррелированность коэффициентов разложения.

Условие ортонормированности непрерывных базисных функций записывается следующим образом:

$$\int\limits_{-A}^{A} \psi_i(t)\psi^*_j(t)\,\mathrm{d}t = \delta_{ij} = \begin{cases} 1, i = j, \\ 0, i \neq j, \end{cases} \quad t \in [-A, A]. \tag{10.31}$$

При дискретизации базисных функций по аргументу путем замены интеграла на сумму, получим численную оценку условия (10.31) в следующем виде:

$$\sum_{n=-N/2}^{N/2} \psi_i(n\Delta t)\psi^*_j(n\Delta t) = E^{KL}_{ij}, \quad N = \frac{2A}{\Delta t}, \tag{10.32}$$

где Δt — шаг дискретизации.

В идеале матрица E^{KL} должна быть единичной:

$$E^{KL}_{ij} = \delta_{ij} = \begin{cases} 1, i = j, \\ 0, i \neq j. \end{cases} \tag{10.33}$$

Рассмотрим базисные функции Карунена–Лоэва, построенные по экспоненциально–косинусной корреляционной функции

$$R(\tau) = \sigma^2 \cdot \exp(-\alpha\,|\tau|)\cos\beta\tau, \quad |\tau| \leqslant 2A, \tag{10.34}$$

с конкретными значениями параметров $\sigma^2 = 1$, $\alpha = 3$, $\beta = 1$, $A = 1{,}28$.

Для $M = 10$ первых базисных функций с шагом дискретизации $\Delta t = 0{,}01$ (число пикселов $N = 2A \cdot \Delta t = 256$) были получены следующие результаты (приводится часть матрицы):

$$E^{KL} = \begin{bmatrix} 1{,}0014 & 0{,}0000 & -0{,}0011 & 0{,}0000 & 0{,}0018 \\ 0{,}0000 & 1{,}0015 & 0{,}0000 & -0{,}0023 & -0{,}0000 \\ -0{,}0011 & 0{,}0000 & 1{,}0017 & -0{,}0000 & -0{,}0036 \\ 0{,}0000 & -0{,}0023 & -0{,}0000 & 1{,}0020 & 0{,}0000 \\ 0{,}0018 & -0{,}0000 & -0{,}0036 & 0{,}0000 & 1{,}0026 \end{bmatrix}.$$

Относительное отклонение от идеальной матрицы ортонормированности (10.32)

$$\varepsilon\left[E^{KL}\right] = \sqrt{\frac{\sum\limits_{i,j=1}^{M}\left(E_{ij}^{KL} - \delta_{ij}\right)^2}{\sum\limits_{i,j=1}^{M}\delta_{ij}^2}} = \sqrt{\frac{1}{M}\sum_{i,j=1}^{M}\left(E_{ij}^{KL} - \delta_{ij}\right)^2}, \tag{10.35}$$

составило величину $\varepsilon\left[E^{KL}\right] = 0{,}016$, то есть 1,6%.

Отклонение $\varepsilon\left[E^{KL}\right]$ зависит от относительных параметров α/A и β/A, так как всегда можно привести интервал анализа к единичному отрезку. С ростом α/A и уменьшением β/A величина $\varepsilon\left[E^{KL}\right]$ уменьшается. Оставляя параметры α/A и β/A неизменными, проследим детальнее влияние дискретизации $\Delta t = 2A/N$.

Уменьшение шага дискретизации в два раза $\Delta t = 0{,}005$ ($N = 512$) приводит к снижению относительного отклонения примерно в два раза $\varepsilon\left[E^{KL}\right] = 0{,}0083$, а увеличение шага дискретизации в два раза $\Delta t = 0{,}02$ ($N = 128$) соответственно к увеличению относительного отклонения в два раза $\varepsilon\left[E^{KL}\right] = 0{,}0321$. Такая же линейная зависимость сохраняется и для других значений Δt (см. рис. 10.11) вплоть до критического увеличения $\Delta t_0 = 0{,}32$ ($N = 8$), когда происходит резкий скачок $\varepsilon\left[E^{KL}\right]$ и линейная зависимость нарушается. Это связано с тем, что данный шаг дискретизации не позволяет адекватно «прописывать» все колебания базисных функций. Таким образом, при $\Delta t \leqslant 0{,}16$ ($N \geqslant 16$) можно говорить о линейной зависимости $\varepsilon\left[E^{KL}\right](\Delta t) \approx c \cdot \Delta t$, $c \approx 1{,}6$.

При этом зависимость $\varepsilon\left[E^{KL}\right]$ от числа пикселов N будет обратная (см. рис. 10.12). Из рисунка 10.12 хорошо видно, что при $N \geqslant 64$ ($\Delta t \leqslant 0{,}04$) обеспечивается уровень относительного отклонения $\varepsilon\left[E^{KL}\right](\Delta t)$ не выше 10%, при котором можно говорить о сохранении свойства ортонормированности при дискретизации базисных функций.

Сохранение условия ортонормированности (10.32) можно проследить и при других возмущениях базисных функций. Известно [30], что квантование по уровню, возникающее при литографическом изготовлении оптических элементов, может быть описано добавлением равномерно распределенного белого шума. В вычислительном эксперименте базисные функции Карунена–Лоэва подвергались искажению путем добавления аддитивного белого шума $\xi(n\Delta t)$:

$$\tilde{\psi}(n\Delta t) = \psi(n\Delta t) + \xi(n\Delta t), \quad n = \overline{-N/2, N/2}, \tag{10.36}$$

и вычислялась величина $\varepsilon\left[\tilde{E}^{KL}\right]$, аналогичная (10.35), где

$$\tilde{E}_{ij}^{KL} = \sum_{n=-N/2}^{N/2} \tilde{\psi}_i(n\Delta t)\tilde{\psi}_j^*(n\Delta t). \tag{10.37}$$

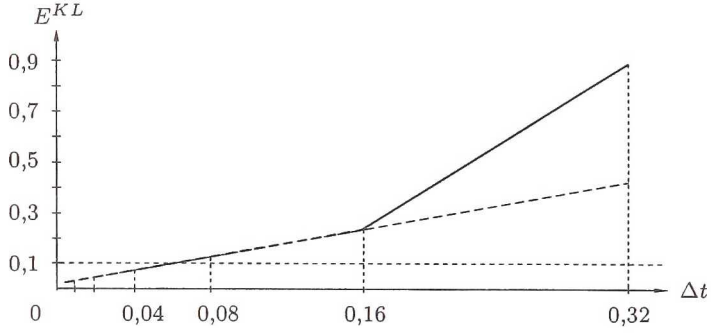


Рис. 10.11. Зависимость отклонения $\varepsilon\left[E^{KL}\right]$ от шага дискретизации Δt

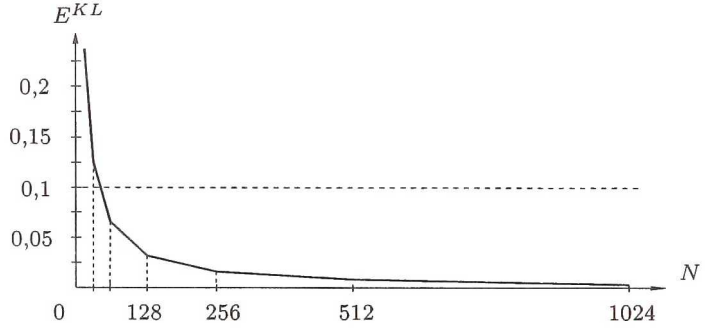


Рис. 10.12. Зависимость отклонения $\varepsilon\left[E^{KL}\right]$ от числа пикселов N на отрезке $[-A, A]$

Так как шум $\xi(n\Delta t)$ является случайной величиной, то и величина $\varepsilon\left[\tilde{E}^{KL}\right]$ будет случайной. Следовательно, определить ее значение можно лишь с некоторой доверительной вероятностью на некотором интервале в результате $N_э$ экспериментальных наблюдений. В этом случае оценка $\hat{\varepsilon}\left[\tilde{E}^{KL}\right]$ определяется как среднее над результатами наблюдений:

$$\hat{\varepsilon}\left[\tilde{E}^{KL}\right] = \frac{1}{N_э}\sum_{i=1}^{N_э}\varepsilon_э^i\left[\tilde{E}^{KL}\right]. \tag{10.38}$$

Из центральной предельной теоремы следует [31], что с увеличением размера выборки $N_э$ распределение выборочного среднего $\hat{\varepsilon}\left[\tilde{E}^{KL}\right]$ стремится к нормальному распределению независимо от вида распределения исходной случайной величины $\varepsilon\left[\tilde{E}^{KL}\right]$. Практически, распределение $\hat{\varepsilon}\left[\tilde{E}^{KL}\right]$ можно считать нормальным при $N_э > 10$ [31]. Следовательно, для построения доверительных интервалов можно воспользоваться следующей формулой [31]:

$$P\left\{\left|\widehat{\varepsilon}\left[\tilde{E}^{KL}\right] - \varepsilon\left[\tilde{E}^{KL}\right]\right| \leqslant \frac{\hat{\sigma}_\varepsilon \cdot t_{N_э-1;\alpha/2}}{\sqrt{N_э}}\right\} = 1-\alpha, \tag{10.39}$$

где $\hat{\sigma}_\varepsilon^2$ — выборочное значение дисперсии $\hat{\varepsilon}\left[\tilde{E}^{KL}\right]$, $t_{N_э-1;\alpha/2}$ — значение t-распределения Стьюдента, $1-\alpha$ — доверительная вероятность.

Для различных значений отношения сигнал/шум (S/N)

$$S/N = \sum_{n=-N/2}^{n=N/2} [\psi(n\Delta t)]^2 \Big/ \sum_{n=-N/2}^{n=N/2} [\xi(n\Delta t)]^2, \tag{10.40}$$

и числа пикселов N, по $N_э = 1000$ наблюдениям были получены зависимости $\varepsilon\left[\tilde{E}^{KL}\right]$, представленные на рис. 10.13.

Задавшись доверительной вероятностью 95% ($\alpha = 0{,}05$), для проведенных $N_э = 1000$, получаем $t_{999;\,0{,}025} = 1{,}98$ и

$$P\left\{\left|\hat{\varepsilon}\left[\tilde{E}^{KL}\right] - \varepsilon\left[\tilde{E}^{KL}\right]\right| \leqslant \hat{\sigma}_\varepsilon \cdot 0{,}06\right\} = 0{,}95.$$

Для удобства можно рассматривать относительную ошибку:

$$\mu_{0,95} = \frac{\left|\hat{\varepsilon}\left[\tilde{E}^{KL}\right] - \varepsilon\left[\tilde{E}^{KL}\right]\right|}{\varepsilon\left[\tilde{E}^{KL}\right]} \approx \frac{\left|\hat{\varepsilon}\left[\tilde{E}^{KL}\right] - \varepsilon\left[\tilde{E}^{KL}\right]\right|}{\hat{\varepsilon}\left[\tilde{E}^{KL}\right]} = \frac{\hat{\sigma}_\varepsilon \cdot 0{,}06}{\hat{\varepsilon}\left[\tilde{E}^{KL}\right]},$$

которая для полученных значений $\hat{\varepsilon}\left[\tilde{E}^{KL}\right]$ и $\hat{\sigma}_\varepsilon^2$ оказалась не выше 2%: $\mu_{0,95}^{\max} \leqslant 0{,}02$.

Из рис. 10.13 видно, что кривые для $N = 256 \div 1024$ очень близки даже при больших уровнях зашумления: $S/N \geqslant 2{,}5$, следовательно, дальнейшее увеличение количества отсчетов при достижении $N = 256$ становится малоэффективным.

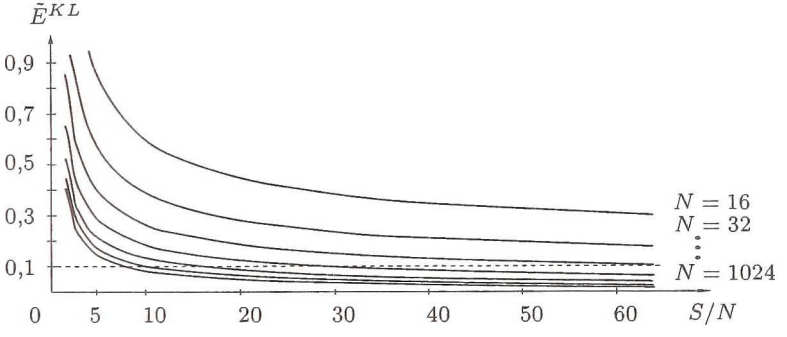


Рис. 10.13. Зависимость отклонения $\varepsilon\left[\tilde{E}^{KL}\right]$ от отношения сигнал/шум (S/N) при различном количестве пикселов N

Основываясь на полученных результатах можно определить диапазон устойчивости РКЛ в зависимости от величины отношения сигнал/шум: уровень относительного отклонения $\varepsilon\left[\tilde{E}^{KL}\right]$ не превышает 10% для $S/N \geqslant 16$ при $N \geqslant 256$, а для $S/N \geqslant 30$ при $N \geqslant 128$.

Другим, важным для РКЛ, критерием влияния возмущений базисных функций на его устойчивость является некоррелированность коэффициентов разложения функции $x(t)$:

$$\langle x_i x_j \rangle = \lambda_i \delta_{ij} = \begin{cases} \lambda_i, i = j, \\ 0, i \neq j. \end{cases} \tag{10.41}$$

Здесь $\langle \ldots \rangle$ означает усреднение по ансамблю реализаций, x_i — коэффициенты разложения

$$x_i = \int_{-A}^{A} x(t)\psi_i(t)\,\mathrm{d}t, \quad |t| \leqslant A, \tag{10.42}$$

по собственным функциям корреляционного оператора:

$$\lambda_i \psi_i(t) = \int_{-A}^{A} R(t-u)\psi_j(u)\,\mathrm{d}u, \quad |t| \leqslant A, \tag{10.43}$$

где λ_j — собственные значения, $R(t-u)$ — корреляционная функция процесса $x(t)$.

Для проведения эксперимента необходимо генерировать стационарные случайные процессы, соответствующие заданной корреляционной функции. Стационарный случайный процесс для экспоненциально-косинусной корреляционной функции $R(\tau)$ (10.36) с теми же параметрами, что и в предыдущем разделе ($\sigma^2 = 1, \alpha = 3, \beta = 1, A = 1{,}28$) моделировался методом рекуррентных разностных уравнений [21].

По $K = 500$ дискретным реализациям такого случайного процесса проводилось вычисление матрицы корреляции коэффициентов разложения:

$$\rho_{ij} = \frac{1}{K}\sum_{k=1}^{K} x_i^k x_j^k, \tag{10.44}$$

где $x_i^k = \sum\limits_{n=-N/2}^{N/2} x^k(n\Delta t)\psi_i(n\Delta t)$, $k = \overline{1,K}$, $i = \overline{1,M}$ — коэффициент разложения k-й дискретной реализации случайного процесса $x(t)$ по i-й базисной функции.

При $\Delta t = 0{,}01$ ($N = 256$) была получена следующая матрица корреляции коэффициентов РКЛ ($M = 10$, приведена только часть):

$$\rho^{KL} = \begin{bmatrix} 0{,}5723 & -0{,}0011 & -0{,}0052 & -0{,}0009 & 0{,}0069 \\ -0{,}0011 & 0{,}3956 & -0{,}0031 & -0{,}0091 & 0{,}0040 \\ -0{,}0052 & -0{,}0031 & 0{,}2522 & 0{,}0020 & -0{,}0070 \\ -0{,}0009 & -0{,}0091 & 0{,}0020 & 0{,}1705 & 0{,}0038 \\ 0{,}0069 & 0{,}0040 & -0{,}0070 & 0{,}0038 & 0{,}1235 \end{bmatrix}.$$

Причем, собственные значения λ_i, $i = \overline{1,M}$, находим по формуле (10.31): $\lambda_1 = 0{,}5757$, $\lambda_2 = 0{,}3992$, $\lambda_3 = 0{,}2552$, $\lambda_4 = 0{,}1649$, $\lambda_5 = 0{,}1112$.

Из приведенных результатов видно, что возмущение корреляционной функции внесло корреляцию между коэффициентами разложения $\rho_{ij}^{KL} \neq 0$, $i \neq j$ и нарушило равенство диагональных элементов собственным числам. Однако, сохранилось соотношение упорядоченности $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \ldots$ [13] для собственных чисел.

Чтобы оценить внесенное возмущение, такой же эксперимент был проведен для базиса Фурье:

$$f_m = \begin{cases} \dfrac{1}{\sqrt{A}} \cos\left(\dfrac{m}{2} \cdot \dfrac{\pi n \Delta t}{A}\right), m = 0, 2, 4, \ldots \\ \dfrac{1}{\sqrt{A}} \sin\left(\dfrac{m+1}{2} \cdot \dfrac{\pi n \Delta t}{A}\right), m = 1, 3, 5, \ldots, \end{cases} \tag{10.45}$$

и была получена соответствующая матрица корреляции коэффициентов разложения Фурье:

$$\rho^F = \begin{bmatrix} 0,5866 & 0,0039 & -0,0074 & -0,0025 & 0,0154 \\ 0,0039 & 0,4118 & -0,0022 & -0,0061 & 0,0043 \\ -0,0074 & -0,0022 & 0,2715 & 0,0019 & 0,0149 \\ -0,0025 & -0,0061 & 0,0024 & 0,1793 & -0,0057 \\ 0,0154 & 0,0043 & 0,0149 & -0,0057 & 0,1312 \end{bmatrix}$$

Для оценки степени корреляции коэффициентов разложения вычислялась следующая величина:

$$\varepsilon[\rho] = \sqrt{\frac{\sum\limits_{i,j=1}^{M} (\rho_{ij} - \lambda_i \delta_{ij})^2}{\sum\limits_{i,j=1}^{M} (\lambda_i \delta_{ij})^2}} = \sqrt{\frac{\sum\limits_{i,j=1}^{M} (\rho_{ij} - \lambda_i \delta_{ij})^2}{\sum\limits_{i=1}^{M} \lambda_i^2}}, \tag{10.46}$$

аналогичная (10.35) и означающая относительное отклонение полученной матрицы корреляции коэффициентов от идеальной. Для РКЛ она составила $\varepsilon\left[\rho^{KL}\right] = 0{,}0092$, а для разложения Фурье $\varepsilon\left[\rho^F\right] = 0{,}0186$, что в два раза больше. Зависимость относительного отклонения $\varepsilon[\rho]$ от шага дискретизации Δt базисных функций для РКЛ и разложения Фурье представлена на рис. 10.14, откуда видно, что кривая РКЛ лежит ниже кривой Фурье для всего рассмотренного диапазона.

На основании полученных результатов можно сделать вывод, что в условиях дискретизации базисных функций РКЛ сохраняет свои оптимальные свойства и обеспечивает меньшую степень корреляции коэффициентов разложения, чем базис Фурье. Влияние зашумленности базисных функций Карунена–Лоэва, моделирующих квантование по уровню (10.36), на матрицу корреляции коэффициентов разложения

$$\tilde{\rho}_{ij}^{KL} = \frac{1}{K} \sum_{k=1}^{K} \tilde{x}_i^k \tilde{x}_j^k, \tag{10.47}$$

где $\tilde{x}_i^k = \sum\limits_{n=-N/2}^{N/2} x^k(n\Delta t)\tilde{\psi}_i(n\Delta t)$, $k = \overline{1,K}$, $i = \overline{1,M}$.

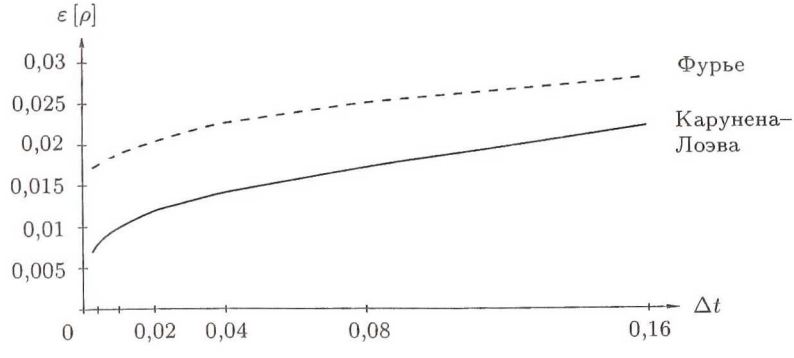


Рис. 10.14. Зависимость отклонения $\varepsilon[\rho]$ от шага дискретизации Δt базисных функций для РКЛ и разложения Фурье

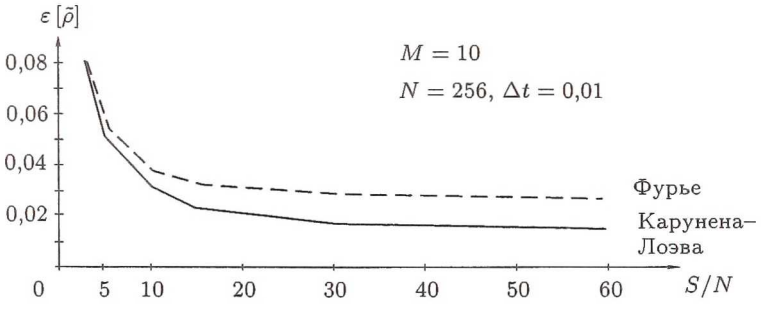


Рис. 10.15. Зависимость отклонения $\varepsilon[\tilde{\rho}]$ для РКЛ и разложения Фурье от величины зашумления базисных функций S/N при $\Delta t = 0{,}01$ ($N = 256$)

Зависимость относительного отклонения матрицы корреляции коэффициентов $\varepsilon\left[\tilde{\rho}^{KL}\right]$, $\varepsilon\left[\tilde{\rho}^{F}\right]$ для РКЛ и разложения Фурье от величины зашумления базисных функций S/N (10.39) при $\Delta t = 0{,}01$ ($N = 256$) представлена на рис. 10.15. Из полученных результатов видно, что при $S/N \geqslant 15$ соотношение степени корреляции коэффициентов для РКЛ и Фурье остается таким же, как и в отсутствии шума, однако, при дальнейшем увеличении зашумленности базисных функций кривые начинают быстро сближаться, и при достижении $S/N = 2{,}5$ оптимальность РКЛ нарушается.

Таким образом, при шаге дискретизации $\Delta t \leqslant 0{,}01$ ($N \geqslant 256$) базисных функций и уровне их зашумления $S/N \geqslant 15$ обеспечивается сохранение основных свойств РКЛ, а значит, и его устойчивость к возмущениям, возникающим при оптической реализации.

Пример 10.2. Для исследования устойчивости системы распознавания изображений с помощью РКЛ проведена серия вычислительных экспериментов на компьютере по распознаванию тестовых изображений с использованием коэффициентов разложения по собственным функциям интегрального уравнения (10.10). При этом каждое изображение интерпретировалось как реализация случайного поля. Корреляционная функция статистически однородных эталонных изображений определялась по одной реализации w_{ij} [29]. Усредненная корреляционная функция $\bar{R}$ вычислялась как среднее арифметическое от всех R_{ij}.

В вычислительном эксперименте использовалось $N_{cl} = 16$ классов изображений. Эталонные представители классов представляют собой треугольники, каждый из которых имеет вложенный треугольник, причем внешний и внутренний треугольники могут занимать одно из четырех положений (см. рис. 10.16).

Вычисленная для представленных на рис. 10.16 эталонных изображений двумерная корреляционная функция $\bar{R}(\Delta x, \Delta y)$ была факторизована на две одномерных $R^{I}(\Delta x)$ и $R^{II}(\Delta y)$, для каждой из которых была построена экспоненциально-косинусная аппроксимация. С использованием аппроксимирующих корреляционных функций $\hat{R}^{I}(\Delta x)$ и $\hat{R}^{II}(\Delta y)$ аналитически решались одномерные интегральные уравнения (10.10) с экспоненциально-косинусным ядром и находились одномерные базисные функции. Двумерные базисные функции, рассчитанные в виде произведения одномерных в соответствии с (10.18), изображены на рис. 10.17.

В эксперименте вычислялось $L = 25$ признаков (10.9) по упорядоченным двумерным базисным функциям $(0,0)$, $(0,1)$, $(1,0)$, $\ldots$, $(4,4)$ для эталонов и объектов, по которым затем проводилась классификация объектов методом среднего для класса вектора признаков. Объектами расознавания служили слегка искаженные эталоны — подвергнутые масштабированию (рис. 10.18), повороту (рис. 10.19) и зашумлению аддитивным белым шумом (рис. 10.20).

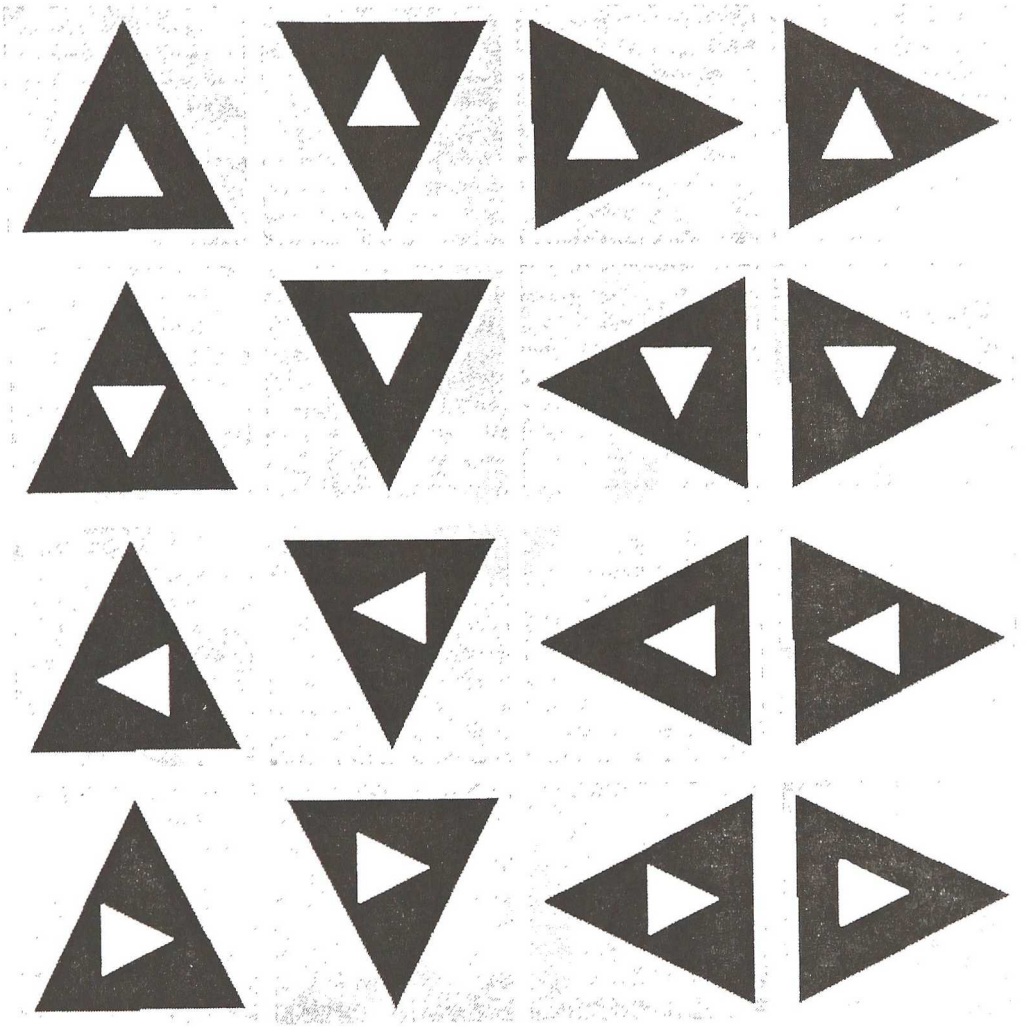


Рис. 10.16. Эталонные представители 16 классов

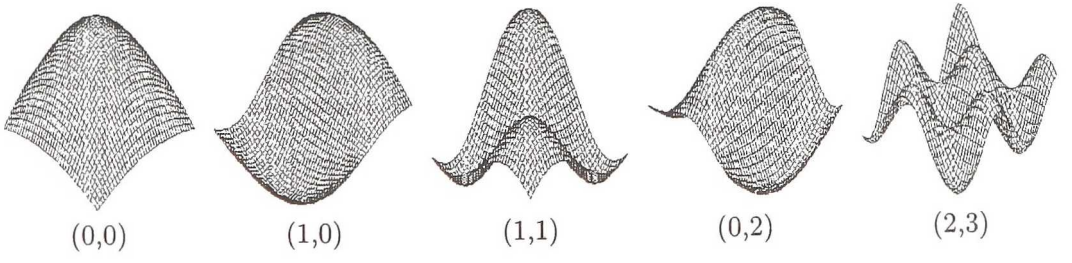
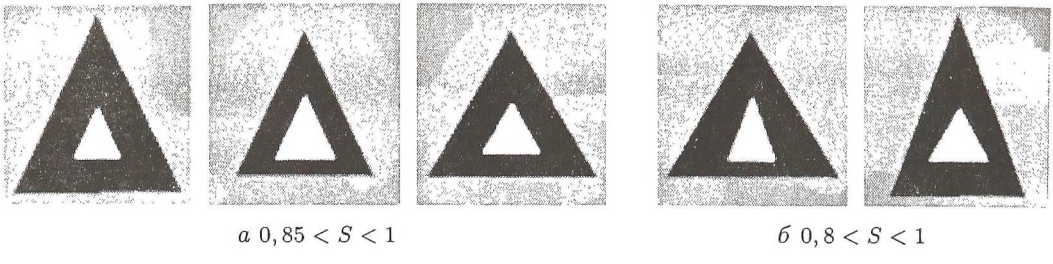


Рис. 10.17. Трехмерный вид базисных функций Карунена–Лоэва



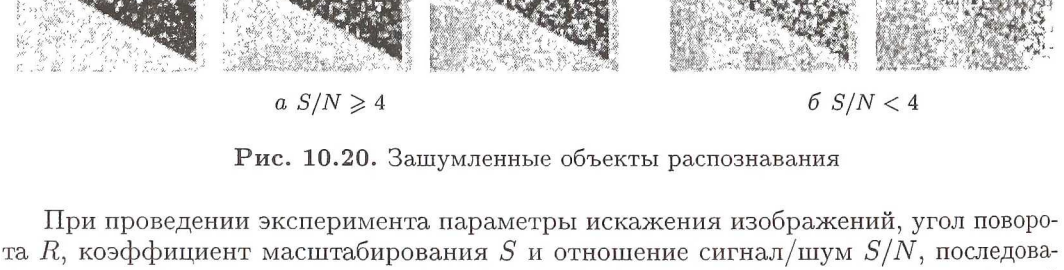
а $0,85 < S < 1$ *б* $0,8 < S < 1$

Рис. 10.18. Масштабированные объекты распознавания



а $-45 < P < 45$ *б* $p < -45 \quad p > 45$

Рис. 10.19. Повернутые объекты распознавания



а $S/N \geqslant 4$ *б* $S/N < 4$

Рис. 10.20. Зашумленные объекты распознавания

При проведении эксперимента параметры искажения изображений, угол поворота R, коэффициент масштабирования S и отношение сигнал/шум S/N, последовательно увеличивались до некоторых значений, определяющих пределы устойчивой работы системы распознавания. На рис. 10.18*а*, 10.19*а*, 10.20*а* представлены объекты со степенью искажения, позволяющей устойчиво распознавать объекты, в то время как на рис. 10.18*б*, 10.19*б*, 10.20*б* показаны объекты со степенью искажения, выходящей за рамки устойчивого распознавания с помощью РКЛ.

В случае поворота и масштабирования параметры искажения являются неслучайными величинами, поэтому точность получаемых результатов не зависит от количества проведенных экспериментов $N_э$. При зашумлении искажение вносится случайной величиной и, следовательно, необходимо оценить точность получаемых в эксперименте результатов. Воспользуемся элементами математической статистики. В случае 95% доверительного интервала получаем оценку частоты успеха $W \in [\hat{W} - \mu_{0,95}, \hat{W} + \mu_{0,95}]$ с точностью

$$\mu_{0,95} = 2\sqrt{\frac{1-W}{W \cdot N_э}} \approx 2\sqrt{\frac{1-\hat{W}}{\hat{W} \cdot N_э}}, \tag{10.48}$$

где $\hat{W} = N_у/N_э$ — оценка частоты успеха в $N_э$ проведенных экспериментах.

Результаты эксперимента по распознаванию искаженных изображений приведены в таблице 10.3. Эксперимент с модельными изображениями показал хорошую устойчивость рассматриваемой системы распознавания к зашумлению и повороту изображений (объектов распознавания) и значительную чувствительность к масштабированию.

Таблица 10.3. Результаты эксперимента по распознаванию искаженных изображений

Вид искажения	Параметры искажения	$N_э$	$\hat{W} = N_у/N_э$	Точность $\mu_{0,95}$
Масштабирование	$S = 1 \div 0{,}85$		1	
	$S = 0{,}85 \div 0{,}8$	25	0,95	неслуч.
Поворот	$R = -45° \div 45°$		1	
	$R < 45°$, $R > 45°$ *)	100	1	неслуч.
Зашумление	$S/N = 30 \div 4$		1	
	$S/N = 2{,}5 \div 1{,}5$	200	0,65	0,1

Пример 10.3. Рассматривается задача распознавания натурных изображений — кристаллограмм слезной жидкости.

В настоящее время в офтальмологии изучена возможность диагностики различных глазных заболеваний по кристаллограмме слезы [32, 33]. Основой данного метода является различие в процессах кристаллизации препарата слезы, полученных у пациентов с нормальным состоянием здоровья и с различными отклонениями от нормы. Однако, проведенный анализ показал, что визуально очень сложно, а иногда и просто невозможно выделить ведущие патологические признаки. Поэтому было бы желательно формализовать признаки и автоматизировать процесс диагностики.

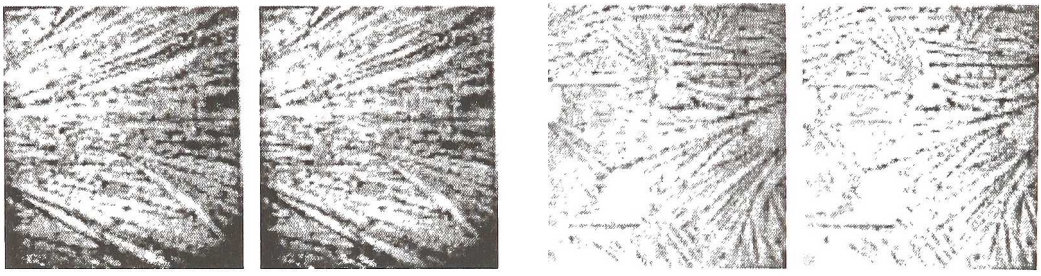
Для предварительного исследования возможности применения рассматриваемой процедуры, в этой задаче были отобраны фотографии кристаллограмм, где разделение образцов между нормой и патологией было предварительно задано на основе обследования пациентов по другим методикам. На рис. 10.21*а* приведены образцы кристаллограмм, соответствующих норме, а на рис. 10.21*б* — образцы кристаллограмм, соответствующих патологии.

В [34] было выделено несколько формальных признаков, по которым отличаются препараты, характерные для нормы и патологии. Среди таких признаков можно назвать «прозрачность» кристалла (в норме линии должны быть тонкими, без бахромы), количество центров, из которых исходят лучи кристалла (в норме он должен быть один), глобальность направления (в норме линии не должны иметь разветвлений), частоту линий на изображении кристалла (в норме расстояние между линиями должно быть максимальным).

В поставленной задаче обнаружения нарушений структуры микрокристаллов из эталонных изображений формируются два класса — норма и патология, для каждого из которых по признакам эталонов, представляющих собой набор коэффициентов ортогонального разложения, строятся центры классов. Классификация вновь поступившего изображения осуществляется по минимуму расстояния его вектора до центров классов.

Для проведения вычислительного эксперимента было отобрано 17 эталонных образцов кристаллограмм слезы (3 из них норма и 14 — патология). Эксперимент

*) В этом случае объект переходил в другой класс и правильно распознавался как представитель другого класса.



а *б*

Рис. 10.21. Натурные изображения кристаллограмм слезы

проводился в различных вариантах. Распознаванию подвергались исходные изображения кристаллограмм на основе коэффициентов, вычисленных с помощью двух ортогональных базисов: Фурье и Карунена–Лоэва. Результаты распознавания 15-ти вновь поступивших изображений приведены в таблице 10.4 (W — частота успеха, L — количество коэффициентов разложения)

Таблица 10.4. Результаты распознавания кристаллограмм слезы

Базис Фурье	Базис Карунена–Лоэва
$W = 0{,}94$ при $L \geqslant 30$	$W = 0{,}94$ при $L \geqslant 19$

Из таблицы 10.4 видно, что для достижения одного и того же результата коэффициентов РКЛ понадобилось в полтора раза меньше, чем коэффициентов разложения Фурье, что наглядно подтверждает оптимальность РКЛ и эффективность его применения.

10.2.2. ДОЭ для разложения функции интенсивности по базису Адамара

Базисные функции Карунена–Лоэва — многоградационные и знакопеременные, поэтому для их оптической реализации требуется изготовление полутоновых амплитудно-фазовых фильтров.

Рассмотрим базис Адамара, который, как показано в [35], хорошо подходит для анализа отпечатков пальцев. Преимущества базиса Адамара в его универсальности и простоте оптической реализации с помощью бинарных фазовых фильтров.

Функции Адамара определяются следующим образом [36, 37]:

$$\mathrm{had}_m(t) = \frac{1}{\sqrt{T}} \prod_{k=0}^{\infty} \left[\mathrm{rad}_{k+1}(t)\right]^{m_k^{\Gamma}}, \quad t \in [0, T], \tag{10.49}$$

где m_k^{Γ} — k-й разряд кода Грея номера m, $rad_m(t)$ — функции Радемахера. Код Грея образуется из двоичного номера m по следующему правилу:

$$m_k^{\Gamma} = m_k \oplus m_{k+1}, \tag{10.50}$$

где k — номер двоичного разряда (справа налево); m_k — двоичный разряд в двоичной записи номера m: $m = \sum\limits_{k=0}^{\infty} m_k 2^k$, а знак $\oplus$ означает сложение по модулю 2.

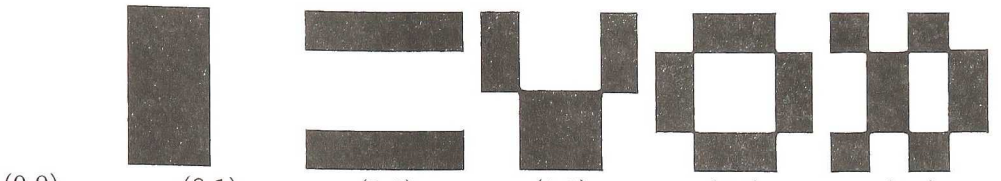


Рис. 10.22. Примеры некоторых функций Адамара (белый цвет соответствует значению +1, а черный — значению −1)

Функции Радемахера определяются следующим образом:

$$\mathrm{rad}_m(t) = \mathrm{sgn}\left[\sin\left(2^m \pi t/T\right)\right]. \tag{10.51}$$

Двумерные функции Адамара определяются как произведение одномерных:

$$\mathrm{had}_{m,n}(x,y) = \mathrm{had}_m(x)\,\mathrm{had}_n(y). \tag{10.52}$$

Базис Адамара обладает следующими интересными свойствами: 1) функции Адамара удовлетворяют условию нуля, которое означает, что их фурье–образы равны нулю в центральной точке; 2) они принимают только два значения: +1 или −1; и 3) определены на ограниченном интервале. На рис. 10.22 представлены примеры некоторых функций Адамара.

В этом разделе рассмотрен итеративный метод расчета фазовых ДОЭ, функция пропускания которых представляет собой произвольную линейную комбинацию конечного числа функций Адамара с различными несущими пространственными частотами. Оптическая схема вычисления признаков изображения $w(x,y)$ показана на рис. 10.2.

Если ДОЭ, имеющий функцию пропускания в виде линейной комбинации конечного числа базисных функций $\{\Psi_{m,n}(x,y)\}$ с различными несущими пространственными частотами

$$\sum_{m=0}^{M}\sum_{n=0}^{N} C_{m,n}\Psi^*_{m,n}(x,y)\cdot\exp\left[i\left(\alpha_m x+\beta_n y\right)\right], \tag{10.53}$$

дополнить сферической линзой O и осветить пучком с амплитудой $w(x,y)$, то интенсивность света в заданных точках фокальной плоскости линзы ($u_m = (t/k)\alpha_m$, $v_n = (t/k)\beta_n$) будет пропорциональна квадрату модуля коэффициентов разложения $C_{m,n}$ амплитуды этого светового поля по базису $\{\Psi_{m,n}(x,y)\}$:

$$C_{m,n} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} w(x,y)\Psi^*_{m,n}(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{10.54}$$

Пусть необходимо рассчитать фазу ДОЭ $\varphi(x,y)$, для вычисления (при освещении волной с амплитудой $A_0(x,y)$) коэффициентов разложения Адамара в заданных точках (α_m,β_n) частотной плоскости. Комплексная функция пропускания такого элемента должна удовлетворять следующему соотношению:

$$\tau(x,y) = A_0(x,y)\exp\left[i\varphi(x,y)\right] = \sum_{m=0}^{M}\sum_{n=0}^{N} C_{m,n}\,\mathrm{had}_{m,n}(x,y)\cdot\exp\left[i\left(\alpha_m x+\beta_n y\right)\right]. \tag{10.55}$$

Если разность между соседними частотами $|\alpha_{m+1}-\alpha_m|$ выбрана достаточно большой, т. е. такой, чтобы пространственные спектры соседних функций Адамара разделялись между собой по некоторому критерию, то коэффициенты $C_{m,n}$ в (10.55) приближенно равны

$$C_{m,n} \cong \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_0(x,y)\exp\left[i\varphi(x,y)\right]\mathrm{had}_{m,n}(x,y)\exp\left[-i(\alpha_m x+\beta_n y)\right]\,\mathrm{d}x\,\mathrm{d}y. \quad (10.56)$$

Коэффициенты $C_{m,n}$ могут быть вычислены с помощью преобразования Фурье:

$$F(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_0(x,y)\exp\left[i\varphi(x,y)\right]\tau(x,y)\,\mathrm{d}x\,\mathrm{d}y. \quad (10.57)$$

В этом случае функция $F(\xi,\eta)$ в точках плоскости (ξ,η) с частотами (α_m,β_n) равна соответствующим коэффициентам

$$F(\xi=\alpha_m,\eta=\beta_n) \cong C_{m,n}. \quad (10.58)$$

Предыдущие выкладки позволяют предложить следующий итеративный алгоритм [38] нахождения фазы $\varphi(x,y)$ оптического элемента. В качестве начальной оценки фазы выбирается $\varphi_0(x,y)$. Предположим, что на k-ой итерации мы получили k-ю оценку фазы: $\varphi_k(x,y)$. Используя функцию $\varphi_k(x,y)$, вычисляются коэффициенты $C^{(k)}_{m,n}$ из (10.56) или (10.57), (10.58). Затем они заменяются на коэффициенты

$$\hat{C}^{(k)}_{m,n} = B_{m,n}C^{(k)}_{m,n}\left|C^{(k)}_{m,n}\right|^{-1} \quad (10.59)$$

с заданными модулями $B_{m,n}$. Далее, коэффициенты $\hat{C}^{(k)}_{m,n}$ подставляются в (10.55). В результате суммирования (10.55) получается функция $E_{k+1}(x,y)$, аргумент которой служит последующей оценкой фазы:

$$\varphi_{k+1}(x,y) = \arg E_{k+1}(x,y), \quad (10.60)$$

и так далее.

Пример 10.4. Предложенный алгоритм использовался для расчета фазового оптического элемента согласованного с базисом Адамара. Как видно из рис. 10.23*а* фаза пропускающей функции такого ДОЭ, учитывая бинарный характер функции Адамара и согласованность несущей с шагом дискретизации, может иметь всего 4–5 уровней, что упрощает его изготовление методами микролитографии. После 10 итераций отклонение квадратов модулей коэффициентов, вычисленных с помощью рассчитанного ДОЭ в частотной плоскости (см. рис. 10.23*б*), от идеальных составило менее 9%.

Также предложенный алгоритм применялся для расчета ДОЭ, предназначенного вычислять первые 25 коэффициентов двумерного разложения Адамара $C_{m,n}$, $m,n=\overline{0,4}$. В этом случае были выбраны следующие параметры. Общее число пикселов 512×512. Количество членов суммы (10.55) $N=M=10$. Амплитуда освещающего пучка $A_0(x,y)$ была выбрана постоянной.

На рис. 10.24*а* показана рассчитанная за 5 итераций фаза ДОЭ, который вычисляет первые 25 коэффициентов двумерного разложения Адамара в плоскости

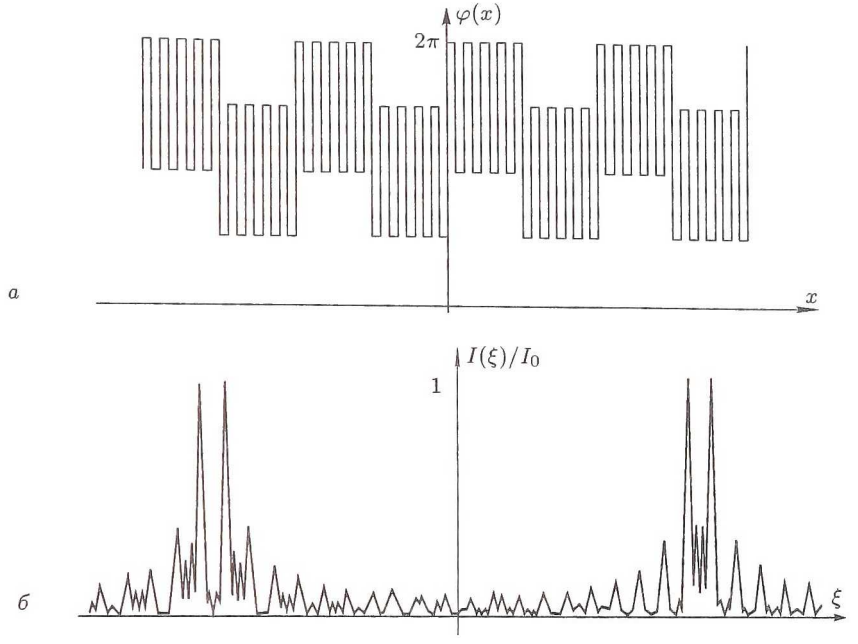


Рис. 10.23. Фаза оптического элемента, вычисляющего 3-й и 4-й коэффициенты одномерного разложения Адамара (*а*) и нормированная интенсивность в плоскости пространственного спектра (*б*) (плоскость формирования квадратов модулей коэффициентов)

пространственного спектра. На рисунках 10.24*б* и 10.24*в* представлены распределения интенсивности в фокальной плоскости при освещении фазового ДОЭ пучком с амплитудой $A_0(x,y) = \mathrm{had}_{0,0}(x,y)$ (пик в точке (α_0, β_0) плоскости (ξ, η)) и $A0(x,y) = \mathrm{had}_{3,2}(x,y)$ (пик в точке (α_3, β_2)), соответственно.

Около 80% энергии падающего света идет на оптическое вычисление заданных коэффициентов. Среднеквадратичная ошибка расчета коэффициентов, получаемых с помощью такого ДОЭ не превышает 20% и сравнима с технологическими ошибками изготовления ДОЭ.

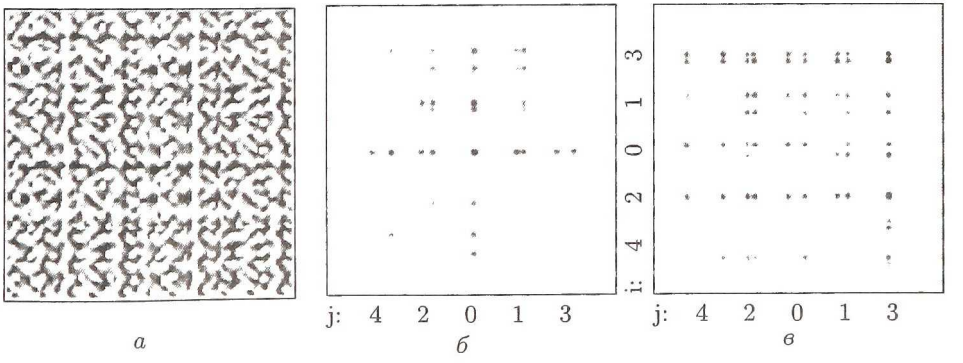


Рис. 10.24. Фаза ДОЭ, обеспечивающего вычисление первых 25 коэффициентов двумерного разложения Адамара (*а*), распределение интенсивности в плоскости пространственного спектра при освещении ДОЭ световым пучком с амплитудой $A_0(x,y) = \mathrm{had}_{0,0}(x,y)$ и (*в*) с амплитудой $A_0(x,y) = \mathrm{had}_{3,2}(x,y)$ (*б*)

Рассчитанный ДОЭ может использоваться в задаче анализа изображений. В работе [35] показано, что коэффициенты разложения Адамара полей направлений (см. раздел 10.3) отпечатков пальцев хорошо подходят для распознавания таких изображений. Заметим, что без существенных изменений предложенный алгоритм может быть применен для расчета ДОЭ, выполняющих оптическое вычисление коэффициентов разложения Уолша и Хаара.

10.2.3. ДОЭ для разложения амплитуды волны с угловыми гармониками

В задаче распознавания изображений инвариантно к их повороту в плоскости наблюдения целесообразно использовать пространственные фильтры, разделяющие амплитуду когерентного светового поле на отдельные дифракционные составляющие специальных ортогональных базисов, содержащих угловые гармоники. Под угловыми гармониками понимаются комплексные функции с единичным модулем и линейной зависимостью от полярного угла. Такие гармоники появляются, например, в бессель-оптике [39] при оптическом выполнении преобразования Ханкеля высшего порядка, или при генерации бездифракционных пучков [40], бездифракционных изображений [41], бесселевых пучков с продольной периодичностью [42], многомодовых вращающихся пучков Гаусса–Лагерра [43].

В этом разделе рассматривается итеративный алгоритм расчета фазовых ДОЭ, которые могут быть названы угловыми спектральными анализаторами, служащими для разложения амплитуды когерентного светового поля по ортогональному базису с угловыми гармониками. Сферическая линза фактически играет роль фурье-анализатора, так как она раскладывает светового поля на плоские волны или пространственные фурье-гармоники. Аналогично, комбинация «линза + ДОЭ» может быть названа анализатором Бесселя, Гаусса–Лагерра, или Цернике если данный оптический элемент раскладывает лазерный свет по соответствующему базису. Разложение по модам Гаусса–Лагерра используется при селекции поперечных мод на выходе многомодового волокна с параболическим профилем показателя преломления [44]. Базис круговых полиномов Цернике используется при анализе аберраций волновых фронтов [45].

Комплексная амплитуда $F(x, y)$ монохроматической световой волны, представленная в полярных координатах (r, φ) всегда допускает разложение в функциональный ряд Фурье по угловым гармоникам. Такой функциональный ряд записывается следующим образом:

$$F(r, \varphi) = \sum_{n=-\infty}^{\infty} F_n(r)\, e^{in\varphi}, \tag{10.61}$$

где

$$F_n(r) = \frac{1}{2\pi} \int_0^{2\pi} F(r, \varphi)\, e^{-in\varphi}\, \mathrm{d}\varphi. \tag{10.62}$$

Ряд (10.63) дает аналогичное разложение по гармоникам $e^{in\theta}$ пространственного спектра амплитуды данного светового пучка. Действительно, если мы применим

преобразование Фурье функции $F(r,\varphi)$, то получим

$$G(\rho,\theta)=\mathcal{F}\left[F(r,\varphi)\right]=\sum_{n=-\infty}^{\infty}G_n\left(\rho\right)e^{in\theta}, \tag{10.63}$$

где

$$G_n\left(\rho\right)=\frac{2\pi k i^n}{f}\int\limits_0^{2\pi}F_n\left(r\right)J_n\left(\frac{k}{f}r\rho\right)r\,\mathrm{d}r, \tag{10.64}$$

$$\mathcal{F}\left[F(r,\varphi)\right]=\frac{k}{f}\int\limits_0^{\infty}\int\limits_0^{2\pi}F(r,\varphi)\exp\left[-i\frac{k}{f}r\rho\cos(\varphi-\theta)\right]r\,\mathrm{d}r\,\mathrm{d}\varphi, \tag{10.65}$$

k — волновое число, f — фокусное расстояние линзы и $J_n(x)$ — функция Бесселя первого рода и n-го порядка. В уравнении (10.64)–(10.65) предполагается, что преобразование Фурье выполняется оптически с помощью сферической линзы с фокусным расстоянием f (см. рис. 10.2).

Видно, что выражения в (10.61) и (10.63) похожи по структуре, в то время как функции $G_n(\rho)$ и $F_n(r)$ связаны преобразованием Ханкеля n-го порядка в (10.64).

Разложение в уравнении (10.61) удобно использовать при инвариантном к повороту распознавании изображений, так как поворот анализируемого изображения $F(r,\varphi)$ на угол $\Delta\varphi$ не приведет к изменению экспериментально измеренных модулей коэффициентов $|F_n(r)|$. В связи с этим задача инвариантного к повороту распознавания изображений может быть сведена к измерению коэффициентов разложения изображений по следующим, предлагаемым ниже, базисам.

Рассмотрим функции, описывающие отдельные бездифракционные моды Бесселя, распространяющиеся в свободном пространстве [41, 46]:

$$Q_{n,m}\left(r,\varphi\right)=A_{n,m}J_m\left(\gamma_{n,m}\frac{r}{r_0}\right)e^{im\varphi}, \tag{10.66}$$

$$A_{n,m}=\left[r_0\sqrt{\pi}J'_m\left(\gamma_{n,m}\right)\right]^{-1}, \tag{10.67}$$

где r_0 — радиус апертуры, ограничивающей пучок Бесселя, $\gamma_{n,m}$ — корни функции Бесселя: $J_m(\gamma_{n,m})=0$, $n,m=0,1,2,\ldots$; $J'_m(\gamma_{n,m})$ — производная от функции Бесселя в точке $r=r_0$.

Множество функций в (10.66) является полным ортогональным базисом внутри круга радиуса r_0. Подставляя их в уравнение (10.61) и (10.63), получаем

$$F(r,\varphi)=\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}C_{n,m}Q_{n,m}\left(r,\varphi\right), \tag{10.68}$$

$$G(\rho,\theta)=\frac{2\pi k}{f}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}i^m C_{n,m}A_{n,m}e^{im\theta}\int\limits_0^{r_0}J_m\left(\gamma_{n,m}\frac{r}{r_0}\right)J_m\left(\frac{k}{f}r\rho\right)r\,\mathrm{d}r. \tag{10.69}$$

При $\rho = \gamma_{m,n}\,(f/kr_0)$, некоторые члены в (10.69) равны нулю в соответствии со свойством ортогональности функций Бесселя:

$$\int_0^{r_0} J_m\left(\gamma_{n,m}\frac{r}{r_0}\right) J_m\left(\gamma_{p,m}\frac{r}{r_0}\right) r\,\mathrm{d}r = \frac{r_0^2}{2}\left[J'_m\left(\gamma_{n,m}\right)\right]^2 \delta_{n,p}. \tag{10.70}$$

Базисные функции в уравнении (10.62) удобно использовать для селекции отдельных бесселевых мод в многомодовом пучке. Такой многомодовый бесселевый пучок может быть на выходе светового волновода со ступенчатым профилем показателя преломления.

Существует и другое полное множество ортогональных функций с угловыми гармониками в круге радиуса r_0. Это круговые полиномы Цернике [45], которые подробно описываются в разделе 10.2.4.

Базисные функции Цернике удобно использовать при анализе небольших аберраций волнового фронта. Анализатор, основанный на суперпозиции этих функций позволяет пространственно разделять вклад в световой пучок отдельных аберраций, описываемых одиночными полиномами Цернике. Сигнал на выходе такого анализатора может быть использован для управления адаптивным зеркалом в целях компенсации аберраций.

Обобщенные моды Гаусса–Лагерра [43, 44] представляют полное множество ортогональных функций на всей плоскости:

$$S_{n,m}(r,\varphi) = D_{n,m}\left(\frac{r}{a}\right)^{|m|} \exp\left[-\frac{1}{2}\left(\frac{r}{a}\right)^2\right] L_n^{|m|}\left(\frac{r^2}{a^2}\right) e^{im\varphi}, \tag{10.71}$$

где

$$D_{n,m} = \frac{2}{a}\sqrt{\frac{\pi(n-m)}{(n!)^3}}, \tag{10.72}$$

$$L_n^m(x) = (-1)^m \frac{\mathrm{d}^m}{\mathrm{d}x^m}\left[L_{n+m}(x)\right], \tag{10.73}$$

a — радиус перетяжки гауссового пучка, $L_n^m\,(x)$ и $L_n\,(x)$ — обобщенные и обычные полиномы Лагерра, соответственно.

В этом случае, разложения (10.61) и (10.63) по угловым гармоникам будет следующим:

$$F(r,\varphi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} C_{n,m} S_{n,m}\,(r,\varphi), \tag{10.74}$$

$$G(\rho,\theta) = \frac{2\pi k}{f}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} (-1)^n (-i)^m C_{n,m} S_{n,m}\,(r,\varphi). \tag{10.75}$$

Уравнение (10.75) следует из того, что функции в (10.71) являются собственными функциями преобразования Фурье:

$$\int_0^{\infty}\int_0^{2\pi} S_{n,m}(r,\varphi)\exp\left[-i\frac{k}{f}r\rho\cos(\varphi-\theta)\right] r\,\mathrm{d}r\,\mathrm{d}\varphi = 2\pi(-1)^n(-i)^m S_{n,m}(\rho,\theta). \tag{10.76}$$

Базисные функции из уравнения (10.71) используются при расчете анализатора, согласованного с многомодовым пучком на выходе из светового волокна с параболическим профилем показателя преломления или лазерного резонатора со сферическими зеркалами. Такой анализатор позволяет пространственно разделять цилиндрические моды пучка.

Реализация ДОЭ для углового спектрального анализа. В [44] вводится понятие моданов — оптических элементов, используемых в качестве пространственных фильтров для анализа поперечно-модового состава когерентного лазерного пучка. Аналогичным образом можно рассматривать оптические элементы, служащие для разложения амплитуды светового поля по любому ортогональному базису, как спектральные анализаторы. На рис. 10.2 показана оптическая схема для спектрального анализатора светового пучка. Предположим, что пропускающая функция ДОЭ такого анализатора представляет собой линейную комбинацию конечного набора базисных функций $\psi_{n,m}(x,y)$, выбранных с заданными наклонами (10.53). Если такой фильтр поместить рядом со сферической линзой и осветить световой волной с амплитудой $F(x,y)$, то интенсивность света в точках $(u_{n,m}, v_{n,m})$ фокальной плоскости

$$u_{n,m} = \frac{f}{k}\alpha_{n,m} \text{ и } v_{n,m} = \frac{f}{k}\beta_{n,m}, \tag{10.77}$$

где $(\alpha_{n,m}, \beta_{n,m})$ — несущие пространственные частоты, будет приближенно пропорциональна квадратам модулей коэффициентов $C_{n,m}$ разложения комплексной амплитуды:

$$F(x,y) = \sum_{n=0}^{N}\sum_{m=0}^{M} C_{n,m}\psi_{n,m}(x,y). \tag{10.78}$$

Действительно,

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)\tau(x,y)\exp\left[-i\frac{k}{f}(ux+vy)\right]\mathrm{d}x\,\mathrm{d}y = \\
&= \frac{k}{f}\sum_{n=0}^{N}\sum_{m=0}^{M}\sum_{p=0}^{N}\sum_{q=0}^{M} C_{n,m}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_{n,m}(x,y)\psi^{*}_{p,q}(x,y)\exp\left[i\left(\alpha_{p,q}x+\beta_{p,q}y\right)\right]\times \\
&\qquad\qquad \times\exp\left[-i\frac{k}{f}(ux+vy)\right]\mathrm{d}x\,\mathrm{d}y \approx \\
&\approx \sum_{n=0}^{N}\sum_{m=0}^{M}\sum_{p=0}^{N}\sum_{q=0}^{M} C_{n,m}\delta_{n,p}\delta_{m,q}\delta\left(\frac{k}{f}u-\alpha_{p,q},\frac{k}{f}v-\beta_{p,q}\right) = \\
&\qquad\qquad = \sum_{n=0}^{N}\sum_{m=0}^{M} C_{n,m}\delta\left(\frac{k}{f}u-\alpha_{n,m},\frac{k}{f}v-\beta_{n,m}\right).
\end{aligned} \tag{10.79}$$

В уравнении (10.79) приближенное равенство возникает, если пренебречь значениями быстро осциллирующих интегралов при $p \neq n$, $q \neq m$:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_{nm}(x,y)\psi^{*}_{pq}(x,y)\exp\left[i\left(\alpha_{p,q}-\alpha_{n,m}\right)x+i\left(\beta_{pq}-\beta_{nm}\right)y\right]\mathrm{d}x\,\mathrm{d}y. \tag{10.80}$$

Чем более эффективно выполняется разделение пучков, распространяющихся под различными углами, тем более точно будет выполняться приближенное равенство (10.79).

Для повышения энергетической эффективности пространственного фильтра (10.53) желательно, чтобы он был фазовым. Задача расчета фазовых ДОЭ, выполняющих разложение светового поля в различных дифракционных порядках для мод Гаусса–Эрмита и Гаусса–Лагерра без угловых гармоник, рассматривалась в [47, 48]. Уравнение (10.53) для функции пропускания обобщенного фазового анализатора можно модифицировать следующим образом:

$$\exp\left[i\Omega(r,\varphi)\right] = \sum_{n=0}^{N}\sum_{m=0}^{M} E_{n,m}\psi^{*}_{n,m}(r,\varphi)\exp\left[ir\rho_{n,m}\cos\left(\varphi-\theta_{n,m}\right)\right], \tag{10.81}$$

где $(\rho_{n,m},\theta_{n,m})$ — компоненты векторов несущих пространственных частот в полярных координатах, $E_{n,m}$ — комплексные коэффициенты, модули которых задаются произвольно, а аргументы являются свободными параметрами задачи, и которые подбираются так, чтобы уравнение (10.81) было точным равенством.

Для эффективного пространственного разделения отдельных дифракционных порядков в фурье-плоскости предлагается считать функции

$$\psi^{*}_{nm}(r,\varphi)\exp\left[ir\rho_{nm}\cos\left(\varphi-\theta_{nm}\right)\right] \tag{10.82}$$

приближенно ортогональными и вычислять коэффициенты в (10.81) по формуле

$$E_{nm} = \int_0^{\infty}\int_0^{2\pi} \exp\left[i\Omega(r,\varphi)\right]\psi_{nm}(r,\varphi)\exp\left[-ir\rho_{nm}\cos\left(\varphi-\theta_{nm}\right)\right] r\,\mathrm{d}r\,\mathrm{d}\varphi. \tag{10.83}$$

Предлагаемый итеративный алгоритм расчета фазовой функции фильтра $\Omega(r,\varphi)$ основан на последовательном вычислении сумм (10.81) и интегралов (10.83) с помощью алгоритма быстрого преобразования Фурье, с учетом определенных ограничений. Например, на p-й итерации коэффициенты $E^{(p)}_{nm}$ находятся из уравнения (10.83) и заменяются на $\bar{E}^{(p)}_{nm}$ следующим образом:

$$\bar{E}^{(p)}_{nm} = \begin{cases} \left[(1+\alpha)T_{n,m} - \alpha\left|E^{(p)}_{n,m}\right|\right] E^{(p)}_{n,m}\left|E^{(p)}_{n,m}\right|^{-1}, & (n,m)\in W, \\ E^{(p)}_{n,m}\left|E^{(p)}_{n,m}\right|^{-1}, & (n,m)\notin W. \end{cases} \tag{10.84}$$

где $T_{n,m} > 0$ заданные числа, характеризующие отклик каждого канала разложения, W — множество пар индексов с ненулевыми значениями $T_{n,m}$, и $0 < \alpha \leqslant 2$ — адаптивный или релаксационный коэффициент, позволяющий контролировать сходимость алгоритма.

Следующая, $(p+1)$-я, оценка фазовой функции фильтра находится из уравнения

$$\Omega_{p+1}(r,\varphi) = \arg\left\{\sum_{n=0}^{N}\sum_{m=0}^{M}\bar{E}^{(p)}_{n,m}\psi^{*}_{n,m}(r,\varphi)\exp\left[ir\rho_{n,m}\cos\left(\varphi-\theta_{n,m}\right)\right]\right\}. \tag{10.85}$$

Далее функцию $\Omega_{p+1}(r,\varphi)$ подставляют в (10.83) и на следующей итерации находят коэффициенты $E^{(p+1)}_{n,m}$, и так далее. Аналогично [48] можно доказать релаксационное свойство ошибки такого итеративного процесса:

$$\varepsilon^{(p+1)}_{L} \leqslant \varepsilon^{(p)}_{L}, \tag{10.86}$$

где

$$\varepsilon_L^{(p)} = \left\{ \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \left[T_{n,m} - \left| E_{n,m}^{(p)} \right| \right]^2 \right\}^{1/2} \cdot \left\{ \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} T_{n,m}^2 \right\}^{-1/2}. \qquad (10.87)$$

Пример 10.5. При численном моделировании для мод Гаусса–Лагерра использовались следующие параметры: 128 пикселов по радиусу r и 128 пикселов по угловой составляющей φ, диапазон изменения аргументов $r \in [0, 7\,\text{мм}]$, $\varphi \in [0, 2\pi]$, длина волны $\lambda = 0{,}63\,\text{мкм}$, фокусное расстояние $f = 100\,\text{мм}$, радиус гауссового пучка в перетяжке $a = 1\,\text{мм}$. В формуле (10.81) рассматривались члены ряда с номерами $n, m \leqslant N = 7$. Действие рассчитанных ДОЭ [49] моделировалось с помощью численного преобразования Фурье.

На рис. 10.25 представлено формирование 9-ти мод Гаусса–Лагерра ((0,0), (1,0), (1,1), (2,0), (2,1), (2,2), (3,0), (3,1), (3,2)) в различных дифракционных порядках: фаза ДОЭ (*а*), сформированные порядки: интенсивность (*б*) и фаза (*д*), эталонное распределение: интенсивность (*в*) и фаза (*е*) и схема, показывающая расположение номеров мод по порядкам (*г*).

Рассчитанные ДОЭ могут использоваться не только для формирования пучков Гаусса–Лагерра с заданным модовым составом, но и как пространственные фильтры для определения поперечно-модового состава исследуемого пучка.

На рис. 10.26 показано действие ДОЭ, согласованного с модой Гаусса–Лагерра (5,3), в качестве фильтра (рис. 10.26*б*). Из рисунка видно, что при прохождении

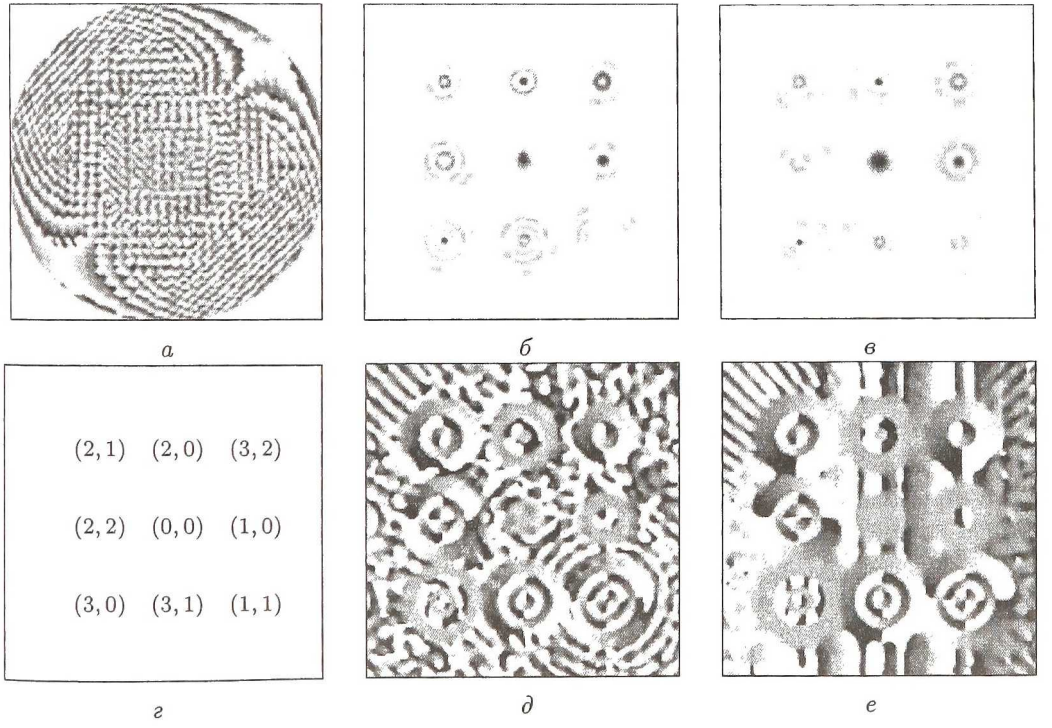


Рис. 10.25. Многомодовый пучок в различных дифракционных порядках: фаза ДОЭ (*а*); номера сформированных мод (*г*); распределения интенсивности в фокальной плоскости: полученное при использовании ДОЭ (*б*) и эталонное (*в*); распределения фазы в фокальной плоскости: полученное при использовании ДОЭ (*д*) и эталонное (*е*)



Рис. 10.26. Действие одномодового фильтра: интенсивность на входе фурье-каскада, соответствующая модам Гаусса–Лагерра $(5,3)$ (*а*) и $(1,0)$ (*г*); фаза фильтра, согласованного с модой $(5,3)$ (*б*); интенсивность на выходе фурье-каскада для моды $(5,3)$ (*в*) и моды $(1,0)$ (*д*)

моды $(5,3)$ (рис. 10.26*а*) через фурье-каскад, на входе которого установлен соответствующий фильтр, в центре фокальной плоскости формируется резкий «всплеск» интенсивности. Если же фильтр согласованный с модой Гаусса–Лагерра $(5,3)$ осветить другим пучком, например, модой $(1,0)$ (рис. 10.26*г*), то в фокальной плоскости сформируется пятно с четким «провалом» интенсивности в центре (рис. 10.26*д*).

Аналогичное действие, но для 9-модового фильтра, представлено на рис. 10.27. Пучки, состоящие из одной моды $(2,1)$ (рис. 10.27*а*) и двух мод $(2,2)+(3,1)$ (рис. 10.27*б,в*), соответственно, проходят через фильтр, показанный на рис. 10.25*а* (соответствие порядков модам см. на рис. 10.25*г*) и производят «всплески» интенсивности в соответствующих точках фокальной плоскости (рис. 10.27*г* и 10.27*д,е*) и пятна с «провалами» интенсивности в центрах других порядков. Пучки, представленные на рис. 10.27*б* и 10.27*в*, отличаются поворотом на 90 градусов. Результаты, приведенные на рис. 10.27*д* и 10.27*е*, подтверждают инвариантность к повороту обобщенных анализаторов с угловыми гармониками.

Таблица 10.5. Нормированные значения интенсивности в дифракционных порядках

Номер базисной функции	$(0,0)$	$(1,0)$	$(1,1)$	$(2,0)$	$(2,1)$	$(2,2)$	$(3,0)$	$(3,1)$	$(3,2)$
Пучок $(2,1)$	0,002	0,001	0,039	0,046	1,000	0,066	0,042	0,025	0,003
Пучок $(2,2)+(3,1)$	0,008	0,001	0,001	0,009	0,053	0,826	0,041	1,000	0,106

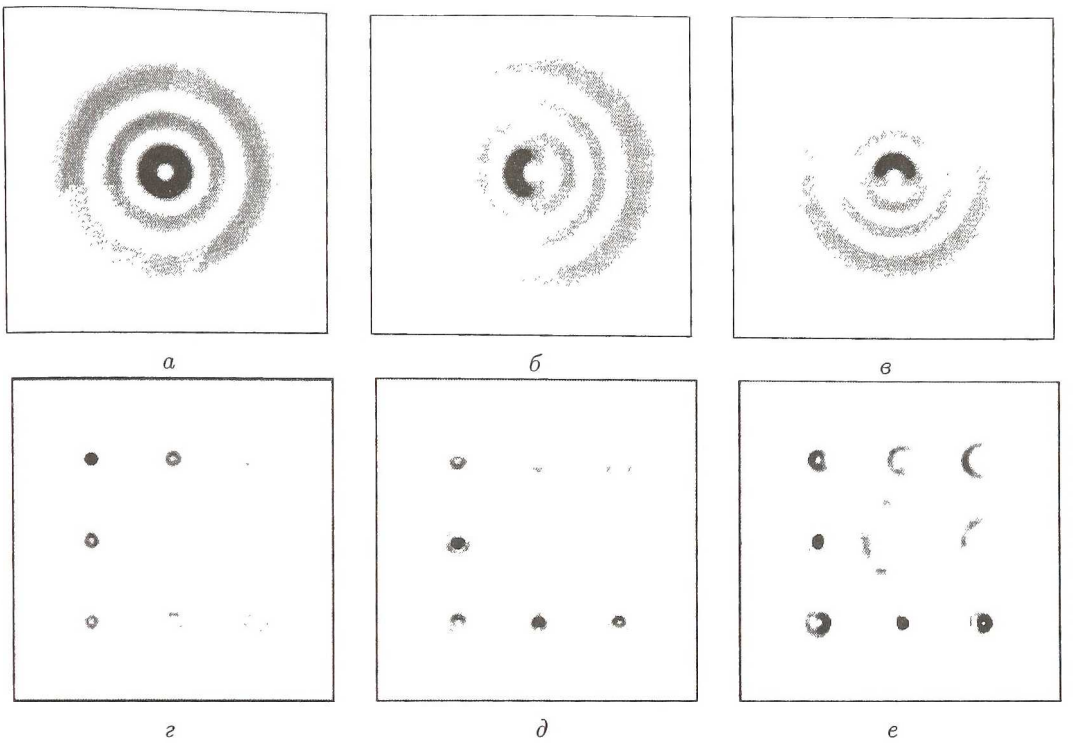


Рис. 10.27. Действие многомодового фильтра: интенсивность на входе фурье-каскада, соответствующая пучкам Гаусса–Лагерра с модами $(2,1)$ $(а)$ и $(2,2)+(3,1)$ $(б,в)$; интенсивность на выходе фурье-каскада для пучка с модами $(2,1)$ $(г)$ и $(2,2)+(3,1)$ $(д,е)$

В таблице 10.5 приведены нормированные значения интенсивности в каждом из дифракционных порядков, усредненные по 5×5 центральным пикселам каждого порядка, что соответствует площадке минимального дифракционного пятна.

10.2.4. ДОЭ для разложения пучка по базису Цернике

Восстановление фазы светового поля — одна из основных задач обработки сигналов. Не имея возможности прямым способом измерять фазу нужно определять ее косвенно, через измерения интенсивности света. Например, волновой фронт может быть восстановлен по интерферограмме [50], по измерениям распределения интенсивности пространственного спектра [51]. Датчик волнового фронта Гартмана–Шаке, состоящий из массива одинаковых узких отверстий или матрицы микролинз [52], также служит для восстановления фазы. Восстановить фазу также можно с использованием амплитудно-фазовых фильтров, формирующих порядки с ортогональным базисом [53, 54, 44, 48].

В данном разделе с помощью фазового пространственного фильтра анализируются аберрации волнового фронта, с использованием разложения амплитуды пучка по базису ортогональных круговых полиномов Цернике [45]. При этом рассматривается разложение по полиномам Цернике комплексной амплитуды, а не сами фазовые поля. В этом случае интенсивность, пропорциональная коэффициентам разложения поля, будет формироваться в пространственной плоскости фурье-спектра. Далее, измеренные модули коэффициентов используются для вычисления аргумента ком-

плексной амплитуды светового поля. Также рассматриваются итеративные алгоритмы расчета фазы фильтра Цернике и вычисления фазы амплитуды поля.

Существует полное множество ортогональных функций с угловыми гармониками в круге радиуса r_0. Это круговые полиномы Цернике [45]:

$$\Psi_{nm}(r,\varphi) = A_n R_n^m(r)\exp(im\varphi), \tag{10.88}$$

где

$$A_n = \sqrt{\frac{n+1}{\pi r_0^2}}, \tag{10.89}$$

$$R_n^m(r) = \sum_{p=0}^{(n-m)/2} (-1)^p (n-p)! \left[p!\left(\frac{n+m}{2}-p\right)!\left(\frac{n-m}{2}-p\right)!\right]^{-1} \left(\frac{r}{r_0}\right)^{n-2p},$$

$R_n^m(r)$ — радиальные полиномы Цернике:

$$R_n^{-m}(r) = R_n^m(r), \quad R_n^{\pm 1}(r_0) = 1, \quad R_{2k+1}^{2l}(r) = 0, \quad R_{2k}^{2l+1}(r) = 0, \quad R_0^0(r) = 1,$$

причем $|m| \leqslant n$, и (r,φ) — полярные координаты.

Разложение комплексной амплитуды $E(r,\varphi)$ в ряд по функциям (10.88) записывается в виде

$$E(r,\varphi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} C_{n,m}\Psi_{n,m}(r,\varphi), \tag{10.90}$$

$$\int_0^{r_0} R_n^m(r)R_p^m(r)r\,\mathrm{d}r = \frac{r_0^2}{2(n+1)}\delta_{np}, \tag{10.91}$$

$$C_{n,m} = \int_0^{r_0}\int_0^{2\pi} E(r,\varphi)\Psi_{n,m}^*(r,\varphi)r\,\mathrm{d}r\,\mathrm{d}\varphi. \tag{10.92}$$

В плоскости пространственного фурье-спектра, который может быть сформирован с помощью сферической линзы с фокусным расстоянием f, комплексная амплитуда светового поля $F(\rho,\varphi)$ будет иметь вид

$$F(\rho,\theta) = \frac{k}{2\pi f}\int_0^{r_0}\int_0^{2\pi} E(r,\varphi)\exp\left[-i\frac{k}{f}r\rho\cos(\varphi-\theta)\right] r\,\mathrm{d}r\,\mathrm{d}\varphi, \tag{10.93}$$

где (ρ,θ) — полярные координаты. Основываясь на (10.90), можно представить разложение (10.93) по полиномам Цернике из (10.88) в виде

$$F(\rho,\theta) = \frac{k}{f}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(-i)^m C_{n,m}A_{n,m}e^{im\theta}\int_0^{2\pi} R_n^m(r)J_m\left(\frac{k}{f}r\rho\right) r\,\mathrm{d}r. \tag{10.94}$$

При выводе уравнения (10.94) использовалось интегральное представление функции Бесселя первого рода m-го порядка:

$$J_m(x) = \frac{i^m}{2\pi} \int_0^{2\pi} \exp\left[-ix\cos t + imt\right] \mathrm{d}t.$$

Интеграл из уравнения (10.94) берется в явном виде [45]:

$$W_{n,m}(\rho) = \int_0^{r_0} R_n^m(r) J_m\left(\frac{k}{f} r\rho\right) r\,\mathrm{d}r = (-1)^{(n-m)/2} r_0^2 \frac{J_{n+1}(kf^{-1}r_0\rho)}{(kf^{-1}r_0\rho)}. \tag{10.95}$$

Из (10.95) видно, что при $n > 0$ комплексная амплитуда в центральных точках $\rho = 0$ равна нулю:

$$W_{n,m}(\rho = 0) = \begin{cases} 0, & n > 0, \\ 1/2r_0^2, & n = 0. \end{cases} \tag{10.96}$$

Следовательно, при $n > 0$ распределение интенсивности в дифракционных порядках фурье-плоскости будет иметь кольцевую структуру.

На рис. 10.28 показана оптическая схема (спектральный анализатор Цернике), демонстрирующая использование фазового фильтра Цернике для анализа волнового фронта с комплексной амплитудой $E(r, \varphi)$. Аналогично датчику волнового фронта Гартмана–Шаке [52], фильтр Цернике FZ устанавливается непосредственно в плоскости исследуемого волнового фронта. Сферическая линза L с фокусным расстоянием f устанавливается сразу за ним. В задней фокальной плоскости линзы L устанавливается матрица фотоприемников PA, сопряженная с компьютером PC.

Чтобы пропускающая функция FZ была фазовой,

$$\tau(r, \varphi) = \exp[iS(r, \varphi)], \tag{10.97}$$

искать ее следует в виде

$$\tau(r, \varphi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \Psi^*_{n,m}(r, \varphi) \exp\left[ikf^{-1} r\rho_{n,m} \cos(\varphi - \theta_{n,m}) + \nu_{n,m}\right], \tag{10.98}$$

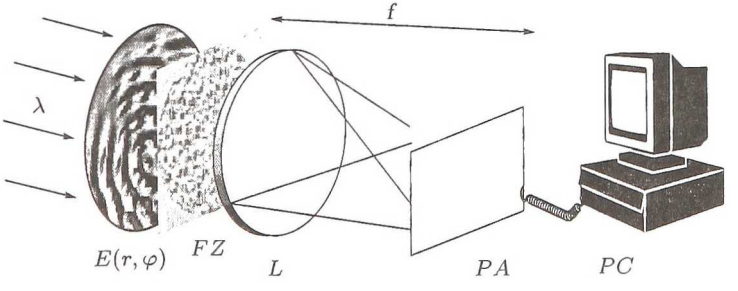


Рис. 10.28. Оптическая схема анализатора Цернике: FZ — фильтр Цернике, L — сферическая линза, PA — матрица фотоприемников, PC — компьютер

где $(\rho_{n,m}, \theta_{n,m})$ — компоненты векторов несущих пространственных частот в полярных координатах, $\nu_{n,m}$ — свободные параметры задачи, которые подбираются так, чтобы уравнение (10.98) было точным равенством.

После того как интенсивность света, пропорциональная квадрату модуля коэффициентов разложения (10.90),

$$I_{nm} = |C_{nm}|^2, \tag{10.99}$$

измерена в дискретных точках фурье-плоскости (рис. 10.28), нужно сделать дополнительные вычисления, чтобы найти фазу волнового фронта из (10.90):

$$Q(r, \varphi) = \arg E(r, \varphi). \tag{10.100}$$

Для этого можно использовать алгоритм, похожий на алгоритм уравнений (10.93) и (10.94). В этом случае оценка фазы на $(k+1)$-й итерации будет иметь вид

$$Q_{k+1}(r, \varphi) = \arg \left\{ \sum_{n=0}^{N} \sum_{m=-n}^{n} \sqrt{I_{n,m}} \Psi_{n,m}(r, \varphi) \exp \left[i \nu_{n,m}^{(k)} \right] \right\}, \tag{10.101}$$

где $\nu_{n,m}^{(k)}$ — свободные параметры на k-й итерации, полученные из уравнения

$$\nu_{n,m}^{(k)} = \arg \left\{ \int_0^{r_0} \int_0^{2\pi} \exp \left[i Q_k(r, \varphi) \right] \Psi_{n,m}^*(r, \varphi) r \, \mathrm{d}r \, \mathrm{d}\varphi \right\}, \tag{10.102}$$

где $Q_k(r, \varphi)$ — оценка искомой фазы на k-й итерации.

Так как аберрации волнового фронта, встречающиеся в оптических системах, описываются четными функциями по отношению к азимутальному углу φ [45], мы можем представить $E(r, \varphi)$ в виде

$$E(r, \varphi) = \exp \left\{ i \sum_{n=0}^{N} \sum_{m=-n}^{n} B_{n,m} R_n^m(r) \cos(m\varphi) \right\}. \tag{10.103}$$

В связи с этим, вместо общего разложения (10.90) следует использовать разложение по четным функциям:

$$E(r, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} C_{n,m} \overline{\Psi}_{n,m}(r, \varphi), \tag{10.104}$$

$$\overline{\Psi}_{n,m}(r, \varphi) = \varepsilon_m \sqrt{\frac{n+1}{\pi r_0^2}} R_n^m(r) \cos(m\varphi,) \tag{10.105}$$

где $\varepsilon_m = \{2, \, m \neq 0; \; 1, \, m = 0\}$.

Для малых аберраций соотношение между коэффициентами разложения $B_{n,m}$ и $C_{n,m}$ — линейное:

$$1 + iB_{00} = \frac{C_{0,0}}{\sqrt{\pi r_0^2}}, \quad iB_{nm} = \varepsilon_m \frac{C_{n,m}}{\sqrt{\pi r_0^2}}. \tag{10.106}$$

Заметим, что базисные функции (10.88) из-за нормировки (10.89) имеют размерность мм^{-1}, поэтому коэффициенты (10.106) имеют такую же размерность. Для произвольных аберраций соотношение между $B_{n,m}$ и $C_{n,m}$ нелинейное, и после измерения модулей $|C_{n,m}|^2$ необходимо использовать алгоритм (10.101) и (10.102), чтобы получить фазу $Q(r,\varphi)$ из (10.100). После чего следует найти коэффициенты волновой аберрации $B_{m,n}$, используя (10.103).

Заметим, что так как $R_0^0(r) = 1$, базис полиномов Цернике содержит единицу в качестве члена разложения. Поэтому если осветить фильтр Цернике плоской волной с амплитудой $E(r,\varphi) = \text{const}$, окажется, что только один коэффициент разложения (10.90) будет ненулевым:

$$|C_{0,0}|^2 \neq 0.$$

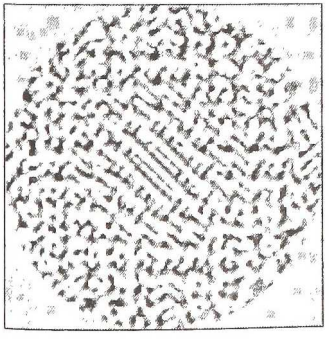
Из уравнения (10.95) также следует, что дифракционные порядки, соответствующие базисным функциям с различными номерами m, но одинаковыми номерами n, будут иметь похожие дифракционные картины (кольцевые структуры при $n > 0$) в фурье-плоскости:

$$|W_{n,m}(\rho)| = r_0^2 \frac{\left|J_{n+1}(kf^{-1}r_0\rho)\right|}{(kf^{-1}r_0\rho)}. \tag{10.107}$$

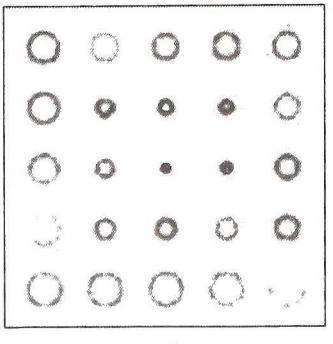
Пример 10.6. Параметры моделирования были выбраны следующие: 256 пикселов по радиусу r и 256 пикселов по углу φ, $r_0 = 1\,\text{мм}$, $k = 104\,\text{мм}^{-1}$, $f = 100\,\text{мм}$. Был рассчитан 25-канальный фильтр [55], формирующий дифракционные порядки для базисных функций с номерами (n,m): $m \leqslant 8$ и $n \leqslant 8$, распространяющиеся под некоторыми углами к оптической оси.

На рис. 10.29 показаны: (*а*) полутоновая фаза фильтра Цернике (черный цвет соответствует значению фазы 0, а белый — 2π), (*б*) 25 дифракционных порядков, сформированных в частотной плоскости линзы (негатив), (*в*) соответствие номеров (n,m) дифракционным порядкам.

Предполагается, что фильтр освещался плоской волной. В этом случае, анализатор «расщепляет» падающий пучок на 25 пучков, имеющих примерно одинаковые энергии. На эти дифракционные порядки приходится более 80% всей энергии пучка. Из рис. 10.29*б* видно, что интенсивность в центральных точках всех порядков фурье-плоскости, кроме нулевого порядка $(0,0)$, нулевая. Это означает, что освещающий волновой фронт не имеет аберраций.



а

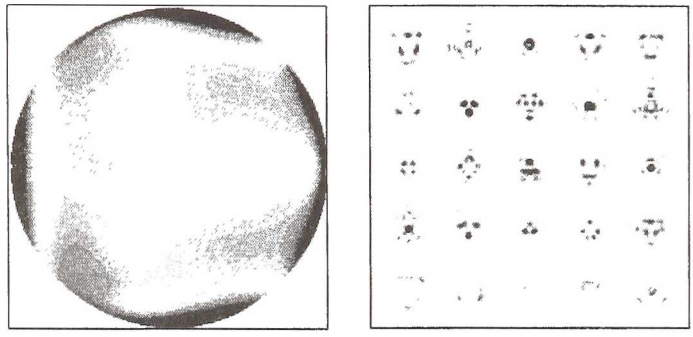


б

(7,1)	(6,6)	(6,4)	(6,2)	(6,0)
(7,3)	(3,1)	(2,2)	(2,0)	(5,5)
(7,5)	(3,3)	(0,0)	(1,1)	(5,3)
(7,7)	(4,0)	(4,2)	(4,4)	(5,1)
(8,0)	(8,2)	(8,4)	(8,6)	(8,8)

в

Рис. 10.29. Полутоновая фаза фильтра Цернике (*а*), распределение интенсивности в частотной плоскости линзы (*б*), соответствие номеров (n,m) дифракционным порядкам (*в*)



а *б*

Рис. 10.30. Действие фильтра Цернике (рис. 10.29*а*): интенсивность освещающего пучка (*а*), дифракционная картина в фурье-плоскости (*б*)

На рис. 10.30 представлен результат действия того же самого 25-канального фильтра Цернике, освещенного пучком, состоящим из трех мод с одинаковыми весами и номерами (n,m): $(2,0)+(5,3)+(7,7)$. На рис. 10.30 показаны: (*а*) интенсивность освещающего пучка, (*б*) дифракционная картина в фурье-плоскости.

Сравнивая со схемой распределения номеров мод по порядкам (рис. 10.29*в*) из рис. 10.30*б* видно, что интенсивность ненулевая (черные точки на рис. 10.30*б*) в центральных точках дифракционных порядков с номерами $(2,0)$, $(5,3)$, и $(7,7)$. В таблице 10.6 приведены соответствующие усредненные значения в окрестности центральных точек для всех порядков (размер окрестности 3×3 отсчета).

Как видно из таблицы 10.6, коэффициенты, имевшие одинаковые веса в пучке, поданном на вход анализатора Цернике, на выходе имеют различные значения:

$$|C_{2,0}|^2 = 0{,}977, \quad |C_{5,3}|^2 = 1{,}000, \quad |C_{7,7}|^2 = 0{,}683.$$

Кроме того, интенсивность в других порядках не является нулевой. Это связано с тем, что при стремлении получить фазовый фильтр, амплитуда в правой части уравнения (10.98) заменяется постоянным значением. Таким образом можно видеть, что квадраты модулей коэффициентов на выходе анализатора Цернике измерены с относительной ошибкой около 20%. Ошибку можно понизить двумя способами: улучшение расчета фазы фильтра Цернике и уменьшение величины окрестности (3×3 отсчетов) центров дифракционных порядков, по которым измеряются модули коэффициентов.

Если в задаче *распознавания изображений* достаточно вычисления модулей коэффициентов разложения по некоторому ортогональному базису, то для однозначного *восстановления комплексной амплитуды* этого недостаточно. Надо также знать фазы коэффициентов разложения.

Фазы коэффициентов $\phi_{n,m} = \arg(C_{n,m})$ могут быть восстановлены, если добавить в фильтр (10.98) члены в виде линейных комбинаций базисных функций [54]:

$$\begin{aligned} s_{n,m}(r,\varphi) &= \left\{\Psi^*_{n,m}(r,\varphi) + \Psi^*_{n',m'}(r,\varphi)\right\} \times \\ &\quad \times \exp\left[ikf^{-1}r\rho''_{n,m}\cos\left(\varphi-\theta'_{n,m}\right)+\nu'_{n,m}\right], \\ p_{n,m}(r,\varphi) &= \left\{\Psi^*_{n,m}(r,\varphi) + i\Psi^*_{n',m'}(r,\varphi)\right\} \times \\ &\quad \times \exp\left[ikf^{-1}r\rho''_{n,m}\cos\left(\varphi-\theta''_{n,m}\right)+\nu''_{n,m}\right]. \end{aligned} \tag{10.108}$$

В этом случае интенсивность света в дополнительных каналах, соответствующих точкам в фурье-спектре с пространственными частотами $(\rho'_{n,m},\theta'_{n,m})$ и $(\rho''_{n,m},\theta''_{n,m})$,

Таблица 10.6. Усредненные значения интенсивности в окрестности центральных точек для порядков в фурье-плоскости при освещении фильтра Цернике (рис. 10.29*а*) пучком, состоящим из трех мод $(2,0)+(5,3)+(7,7)$

$(7,1)$ 0,011	$(6,6)$ 0,028	$(6,4)$ 0,005	$(6,2)$ 0,006	$(6,0)$ 0,035
$(7,3)$ 0,103	$(3,1)$ 0,013	$(2,2)$ 0,018	$(2,0)$ 0,977	$(5,5)$ 0,008
$(7,5)$ 0,004	$(3,3)$ 0,007	$(0,0)$ 0,059	$(1,1)$ 0,003	$(5,3)$ 1,000
$(7,7)$ 0,683	$(4,0)$ 0,056	$(4,2)$ 0,000	$(4,4)$ 0,001	$(5,1)$ 0,001
$(8,0)$ 0,004	$(8,2)$ 0,014	$(8,4)$ 0,018	$(8,6)$ 0,006	$(8,8)$ 0,026

будет следующей:

$$\begin{aligned} S_{n,m} &= |C_{n,m}|^2 + |C_{n',m'}|^2 + 2\,|C_{n,m}|\,|C_{n',m'}|\cos\left(\phi_{n',m'} - \phi_{n,m}\right), \\ P_{n,m} &= |C_{n,m}|^2 + |C_{n',m'}|^2 + 2\,|C_{n,m}|\,|C_{n',m'}|\sin\left(\phi_{n',m'} - \phi_{n,m}\right), \end{aligned} \tag{10.109}$$

что позволяет найти фазы $\phi_{n,m}$, полагая, например, $\phi_{0,0}=0$.

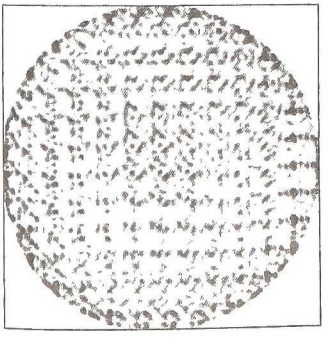
Рекурсивное соотношение для искомых фаз может быть записано в форме

$$\phi_{n',m'} - \phi_{n,m} = \mathrm{tg}^{-1}\left(\frac{P_{n,m} - I_{n,m} - I_{n',m'}}{S_{n,m} - I_{n,m} - I_{n',m'}}\right). \tag{10.110}$$

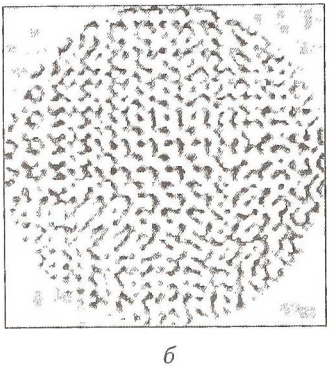
Таким образом, рассматриваемый оптический метод позволяет находить комплексные коэффициенты разложения волнового фронта по ортогональному базису и восстанавливать волновой фронт.

Пример 10.7. Был рассчитан 25-канальный фильтр Цернике, формирующий моды в различных дифракционных порядках с номерами (n,m): $n \leqslant 4$ и $m \leqslant 4$ — всего 9 и их линейные комбинации из уравнения (10.108) — всего $8+8$.

На рис. 10.31 показаны: полутоновая амплитуда (*а*) и фаза (*б*) фильтра Цернике и распределение мод с номерами (n,m) и их линейных комбинаций между порядками (*в*).



а



б

S8	S7	S6	S5	S4
P1	(3,1)	(2,2)	(2,0)	(5,5)
P2	(3,3)	(0,0)	(1,1)	(5,3)
P3	(4,0)	(4,2)	(4,4)	(5,1)
P4	P5	P6	P7	P8

в

Рис. 10.31. Полутоновая амплитуда (*а*) и фаза (*б*) фильтра Цернике и распределение мод с номерами (n,m) и их линейных комбинаций между порядками (*в*)

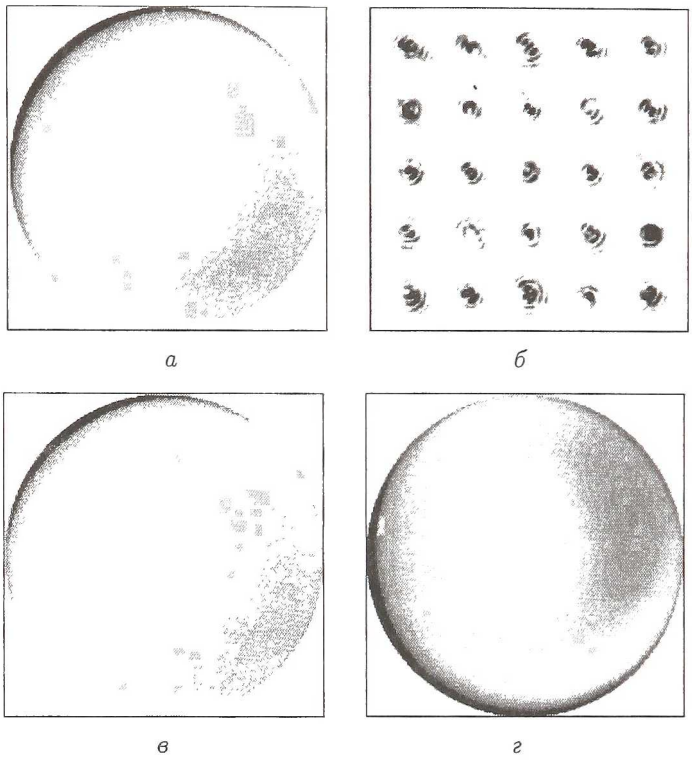


Рис. 10.32. Интенсивность анализируемого пучка (*а*), дифракционная картина в фурье-плоскости для чисто фазового фильтра Цернике, представленного на рис. 10.31*а,б* (*б*), световое поле, восстановленное с помощью амплитудно-фазового фильтра (*в*) и световое поле, восстановленное с помощью чисто фазового фильтра (*г*)

На рис. 10.32 представлен результат действия такого 25-канального фильтра Цернике, освещенного пучком, состоящим из трех мод с коэффициентами $C_{n,m}$: $C_{1,1} = \exp(i0)$, $C_{3,3} = \exp(i\pi/2)$, $C_{4,2} = \exp(i\pi)$. На рис. 10.32 показаны: интенсивность анализируемого пучка (*а*), дифракционная картина в фурье-плоскости для фазового фильтра (*б*), интенсивность, восстановленная с помощью амплитудно-фазового фильтра (*в*) и интенсивность, восстановленная с помощью фазового фильтра (*г*).

Из этого примера видно, что для *распознавания изображений*, когда достаточно измерения модулей коэффициентов разложения, можно использовать чисто фазовый фильтр (ошибка в измерении модулей коэффициентов разложения менее 9%).

Таблица 10.7. Результаты восстановления комплексных коэффициентов амплитудно-фазовым и чисто фазовым фильтрами

	$\lvert C_{1,1}\rvert^2$	$\lvert C_{3,3}\rvert^2$	$\lvert C_{4,2}\rvert^2$	Ошибка	$\phi_{3,3} - \phi_{1,1}$	$\phi_{4,2} - \phi_{3,3}$
Исходные коэффициенты	1	1	1		1,57	1,57
Восстановленные с помощью *амплитудно-фазового* фильтра	0,272	0,274	0,279	3,1%	1,46	1,54
Восстановленные с помощью *фазового* фильтра	12,12	14,83	14,59	8,6%	0,95	1,53

При *восстановлении полной информации* о световом поле требуется знать также фазы коэффициентов разложения. В этом случае следует использовать амплитудно-фазовый фильтр.

10.3. Оптическое построение поля направлений и поля пространственных частот

В соответствии с определением [56] полем направлений называется геометрическая интерпретация множества линейных элементов, соответствующих обыкновенному дифференциальному уравнению, которое в случае двух переменных имеет следующий вид [57]:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y). \tag{10.111}$$

Линейным элементом называется набор чисел x, y, $f(x, y)$, который можно представить себе как совокупность координат точки $(x, y) \in G \subset \mathbf{R}^2$ и соответствующего ей направления с направляющими косинусами

$$\left(\frac{1}{\sqrt{1 + f^2(x, y)}}, \frac{f(x, y)}{\sqrt{1 + f^2(x, y)}} \right),$$

которое отображается отрезком малой длины, проходящим через эту точку параллельно вектору $(1, f(x, y))$.

Поле направлений может быть вычислено для изображений с ярко выраженным контурным характером. Рассмотрим произвольную функцию распределения интенсивности света в плоскости изображения $I(x, y)$, которую будем считать гладкой, т.е. имеющей непрерывные частные производные первого порядка. Рассмотрим множество кривых на плоскости (x, y), соответствующих постоянным значениям интенсивности изображения:

$$I(x, y) = I_0. \tag{10.112}$$

Легко видеть, что множество направлений, касательных к линиям (10.112), в соответствии с определением (10.111), образует поле направлений при

$$f(x, y) = -\frac{\partial I(x, y)/\partial x}{\partial I(x, y)/\partial y}.$$

Таким образом, классическое определение поля направлений приводит к функции $\hat{\varphi}(x, y)$, которая имеет физический смысл угла наклона касательной к линии постоянных значений интенсивности изображения и задается следующим уравнением:

$$\operatorname{tg} \hat{\varphi}(x, y) = -\frac{\partial I(x, y)/\partial x}{\partial I(x, y)/\partial y}, \quad 0 \leqslant \hat{\varphi}(x, y) < \pi. \tag{10.113}$$

Формула вычисления поля направлений (10.113) может быть непосредственно применена к классу изображений со структурной избыточностью [57], функция интенсивности которых должна обладать достаточной гладкостью для возможности ее дифференцирования. Таким классом изображений являются, в частности, интерферограммы, для которых интуитивное понятие поля направлений связывается с направлением интерференционных полос и совпадает с определением (10.113).

В произвольной локальной области достаточно малых размеров (окно на рис. 10.33) интенсивность таких изображений может быть описана гармонической функцией

$$I(x,y) = A\cos[\omega_x x + \omega_y y] + B, \tag{10.114}$$

где $\omega_x = 2\pi/T_x$, $\omega_y = 2\pi/T_y$ — локальные пространственные частоты, T_x, T_y — периоды полос контурного изображения.

Впервые идея оптического вычисления поля направлений была предложена в [34]. Если интенсивность в некоторой локальной области Ω, которая задается функцией

$$q(x,y) = \begin{cases} 1, & (x,y) \in \Omega, \\ 0, & (x,y) \notin \Omega, \end{cases}$$

аппроксимируется косинусной решеткой (10.114), то ее можно представить в виде двух экспонент:

$$\cos(\omega_x x + \omega_y y) = \frac{1}{2}\left\{\exp[i(\omega_x x + \omega_y y)] + \exp[-i(\omega_x x + \omega_y y)]\right\}.$$

Тогда преобразование Фурье от функции интенсивности имеет вид

$$A\cos(\omega_x x + \omega_y y) + B \overset{F}{\to} \frac{A}{2}\{\delta(\xi - \omega_x, \eta - \omega_y) + \\ + \delta(\xi + \omega_x, \eta + \omega_y)\} + B\delta(\xi,\eta), \tag{10.115}$$

где (x,y) и (ξ,η) — декартовы координаты во входной и частотных плоскостях, F — двумерное преобразование Фурье, $\delta(x,y)$ — функция Дирака.

Таким образом, решетка дает δ-импульс в центре фурье-плоскости и два δ-импульса, расположенные симметрично на линии, проходящей через центр фурье-плоскости под углом θ к оси x:

$$\operatorname{tg}\theta = -\frac{\omega_y}{\omega_x}.$$

Тогда θ и угол направления несущей $\hat{\varphi}$ (см. рис. 10.33) связаны следующим образом:

$$\theta = \hat{\varphi} + \frac{\pi}{2}.$$

Если решетка ограничена некоторой областью $q(x,y)$, то в частотной плоскости получается свертка фурье-образа от функции, описывающей эту область $Q(\xi,\eta)$, с δ-импульсами.

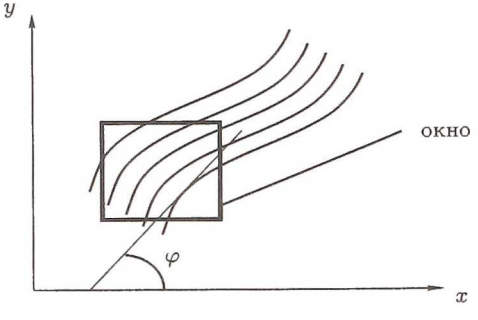
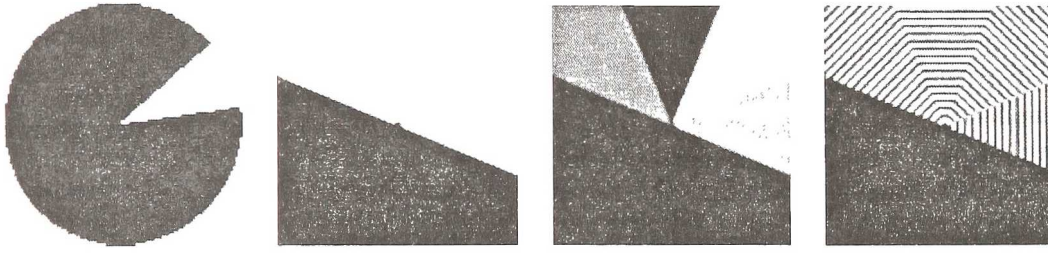


Рис. 10.33. Угол наклона линии в контурном изображении

Известно, что свертка δ-функции с любой функцией дает ту же функцию с соответствующим смещением:

$$\delta(\xi - \omega_x, \eta - \omega_y) \otimes Q(\xi, \eta) \overset{\mathcal{F}}{\to} q(-x, -y), \tag{10.116}$$

где $\otimes$ — обозначение свертки. То есть, в выходной плоскости отличная от нуля интенсивность будет только в области, соответствующей решетке с углом наклона штрихов $\hat{\varphi}$. Этот эффект имеет место независимо от частоты полос, так как им будут соответствовать δ-функции, находящиеся на одном луче — ближе или дальше от центра (нулевой пространственной частоты). Следовательно, «вырезая» сегмент вдоль направления $\theta = \mathrm{mod}_\pi(\hat{\varphi} + \pi/2)$, «захватываются» спектры от всех решеток с направлением штрихов $\hat{\varphi}$ независимо от частоты штрихов и областей, их ограничивающих.



а *б* *в* *г*

Рис. 10.34. Пространственные фильтры, используемые для оптического вычисления поля направлений

Вращая щель (рис. 10.34*а*) и «вырезая» сегменты в верхней части фурье-плоскости, можно просканировать все направления линий, имеющиеся на изображении в диапазоне $[0, \pi)$. Наличие механически подвижного элемента — вращающейся щели, — усложняет реализацию этого метода и увеличивает время формирования поля направлений. Целесообразно исключить подвижный элемент, как это предложено в [58]. Для этого попеременно вводятся в схему два пространственных фильтра и осуществляются простые вычисления.

1) В фурье-плоскость помещается амплитудный фильтр, изображенный на рис. 10.34*в*, функция пропускания $t(\hat{\varphi})$ которого имеет вид

$$t(\hat{\varphi}) = \begin{cases} \displaystyle\sum_{l=0}^{N-1} \mathrm{rect}\left(\hat{\varphi} + \frac{\pi}{2N} - \frac{\pi}{N} l\right) \sqrt{\left[\frac{1}{2} + \mathrm{sgn}\left(l - \frac{N-1}{2}\right) \frac{l}{N}\right]}, \\ \qquad -\dfrac{\pi}{2N} < \hat{\varphi} < \pi - \dfrac{\pi}{2N}, \\ 0, \quad \pi - \dfrac{\pi}{2N} < \hat{\varphi} < 2\pi - \dfrac{\pi}{2N}, \end{cases} \tag{10.117}$$

где

$$\mathrm{rect}(x) = \begin{cases} 1, & x \in [0, 1), \\ 0, & x \notin [0, 1), \end{cases}$$

а $\hat{\varphi}$ — полярный угол в полярной системе координат $(\rho, \hat{\varphi})$ в частотной плоскости:

$$\begin{cases} \xi = \rho \cos \widehat{\varphi}, \\ \eta = \rho \sin \widehat{\varphi}, \end{cases} \tag{10.118}$$

N — количество выделяемых направлений. Затем с помощью линзы делается еще одно фурье-преобразование; распределение интенсивности в выходной плоскости через телекамеру вводится в компьютер в виде массива чисел.

2) В фурье-плоскость помещается амплитудный фильтр, изображенный на рис. 10.34*б*, функция пропускания которого имеет вид

$$t(\hat{\varphi}) = \begin{cases} 1, \ -\dfrac{\pi}{2N} < \hat{\varphi} < \pi - \dfrac{\pi}{2N}, \\ 0, \ \pi - \dfrac{\pi}{2N} < \hat{\varphi} < 2\pi - \dfrac{\pi}{2N}, \end{cases} \tag{10.119}$$

и вновь выполняется фурье-преобразование; интенсивность в выходной плоскости запоминается в памяти компьютера в виде массива чисел.

3) Интенсивность в выходной плоскости, получаемая после первого фильтра, поточечно делится на интенсивность, получаемую после второго фильтра. Результатом этого деления будет аппроксимация поля направлений.

Наличие процедуры деления в методе налагает жесткие требования к юстировке оптической системы при проведении натурного эксперимента. Поэтому данный метод мало пригоден для практического использования.

В [59] был предложен метод построения поля направлений с использованием пространственного фильтра из бинарных дифракционных решеток. Суть метода заключается в следующем.

В частотной плоскости установки находится оптический фильтр, который представляет собой набор амплитудных или фазовых секторных дифракционных решеток (рис. 10.34*г*). Если это набор амплитудных дифракционных решеток, то его функция пропускания имеет вид

$$t(\rho, \hat{\varphi}) = \sum_{l=0}^{N-1} \operatorname{rect}\left(\hat{\varphi} + \frac{\pi}{2N} - \frac{\pi}{N} l\right) \sum_{n=0}^{M} \operatorname{rect}\left[2\frac{\rho \cos\left(\hat{\varphi} - \pi l/N\right)}{T} - nT\right]. \tag{10.120}$$

Если это набор фазовых решеток, то его функция пропускания имеет вид

$$t(\rho, \hat{\varphi}) = \exp\left(i \sum_{l=0}^{N-1} \operatorname{rect}\left(\hat{\varphi} + \frac{\pi}{2N} - \frac{\pi}{N} l\right) \sum_{n=0}^{M} \operatorname{rect}\left[2\frac{\rho \cos\left(\hat{\varphi} - \pi l/N\right)}{T} - nT\right]\right). \tag{10.121}$$

В выходной плоскости фурье-коррелятора с таким фильтром в 1 и -1 порядках дифракции решеток фильтра получается набор изображений (см. рис. 10.57), количество которых равно удвоенному числу направлений. Будем в дальнейшем называть их «парциальными» изображениями. В каждом из этих парциальных изображений отличная от нуля интенсивность должна наблюдаться только в тех местах, где линии контурного изображения имели соответствующее направление, но так как реальное изображение помимо основных полос содержит еще шум, то отличная от нуля интенсивность наблюдается на всей площади парциального изображения. Изображения с помощью телекамеры вводятся в компьютер в виде массива чисел размерностью 512×512, которые могут принимать значения от 0 до 255. Чтобы выделить

в этом массиве парциальные изображения, предварительно необходимо определить их положение; для этого во входной плоскости фурье-коррелятора размещается точечная диафрагма, а в выходной плоскости регистрируется набор светлых точек, координаты которых соответствуют координатам центров парциальных изображений. Размер парциальных изображений известен, он определяется формулой

$$S_c = \frac{f_2}{f_1} S, \tag{10.122}$$

где S_c — размер парциального изображения, S — размер исходного контурного изображения на входе, f_1 — фокусное расстояния первого плеча фурье-коррелятора, f_2 — фокусное расстояния второго плеча фурье-коррелятора. Координаты точки в общем массиве (l, j), координаты точки в k-м парциальном изображении (l'_k, j'_k). Они связаны между собой формулами

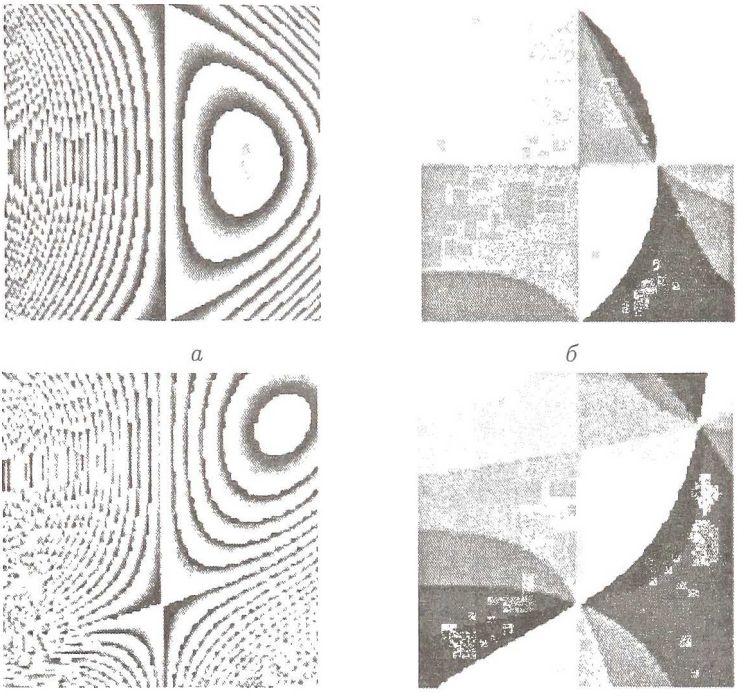
$$l'_k = l - L_k + \frac{S_c}{2\Delta l}, \tag{10.123}$$

$$j'_k = j - J_k + \frac{S_c}{2\Delta l}, \tag{10.124}$$

где (L_k, J_k) — координаты центра k-го парциального изображения, Δl — размер минимального чувствительного элемента телекамеры.

Далее парциальные изображения обрабатываются в компьютере. Чтобы выделить на парциальных изображениях зоны с одинаковым направлением, проводится процедура бинаризации. Каждое парциальное изображение сравнивается поточечно с другими, и если интенсивность в данной точке парциального изображения максимальна по отношения к интенсивностям соответствующих точек других парциальных изображений, то в данной точке этого парциального изображения значение интенсивности принимается равным 1, если это условие не соблюдается, то значение интенсивности принимается равным 0. Поле направлений получается из набора парциальных изображений простым сложением с умножением на коэффициент равный соответствующим значениям углов наклона сектора решетки.

Поле направлений, сформированное оптическими методами, формируется с некоторой погрешностью, которая может сильно изменяться с изменением параметров оптической установки. В частности, на качество поля направлений влияет количество выделяемых направлений, соотношение между периодом дифракционной решетки и шириной спектра исходного изображения, качество линий дифракционной решетки. Для выявления зависимости точности построения поля направлений, а также для определения оптимальных параметров оптической установки был проведен ряд вычислительных экспериментов. В этих экспериментах использовался набор из 10 тестовых изображений, два из которых изображены на рис. 10.35*а,в*; поля направлений которых можно вычислить аналитически (рис. 10.35*б,г*). Период полос на тестовых изображениях менялся от 1/50 до 1/10 размера изображения. Такой подбор был использован из соображений практического применения оптического поля направлений для идентификации дактилограмм. Именно в этом диапазоне лежат значения периодов полос дактилограмм. Изображения были записаны в числовые массивы размерностью 512×512 с 256 уровнями квантования. Затем поля направлений вычислялись численным моделированием работы оптической установки и сравнивались с полями, полученными аналитически. С помощью алгоритма быстрого преобразования Фурье (БПФ) вычислялся фурье-образ исходного изображения, затем фурье-образ умножался на функцию пропускания фазового пространственного



а *б* *в* *г*

Рис. 10.35. Изображения и вычисленные аналитически их поля направлений

фильтра (10.120), которая была записана в бинарном числовом массиве размерностью 512×512. Затем с помощью алгоритма БПФ осуществлялось второе преобразование Фурье, результат которого фиксировался в цифровом массиве. Далее по методике, описанной выше, выделялись парциальные изображения, и формировалось поле направлений. В качестве критерия оценки использовалась величина среднеквадратичного отклонения полей направлений сформированных в вычислительном эксперименте от полей направлений вычисленных аналитически. Обозначим эту величину буквой ε.

Согласно [60] ε угловой величины определяется формулой

$$\varepsilon = \sqrt{\frac{2}{n^2} \sum_{i,j=1}^{n} \sin^2 \frac{\hat{\varphi}_{ij} - \hat{\varphi}^a_{ij}}{2}}, \tag{10.125}$$

где n — размер массива данных, в котором записано поле направлений, $\hat{\varphi}_{ij}$ — направление в точке (i,j) полученное оптическим методом, $\hat{\varphi}^a_{ij}$ — направление в точке (i,j) вычисленное аналитически.

В этом эксперименте в фильтре, изображенном на рис. 10.34*в*, менялось количество секторов от 2 до 16, с каждым из этих фильтров вычислялось поле направлений для 10 тестовых изображений. На рис. 10.36*а,б,в,г* изображены соответственно поля направлений, которые были построены с выделением 2, 4, 8, 16 направлений для изображения на рис. 10.35*а*.

На рис. 10.37 изображен график зависимости погрешности вычисления поля направлений ε от количества выделяемых направлений n.

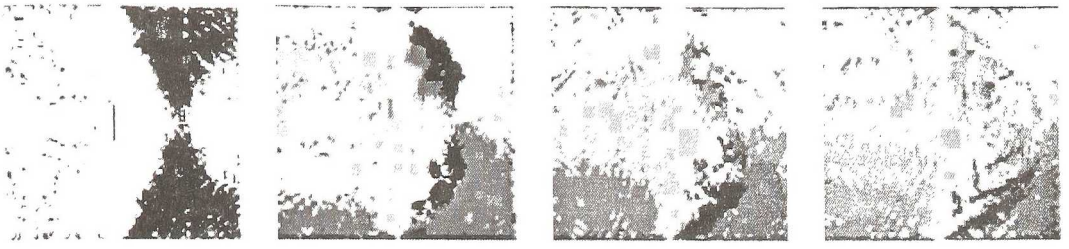


Рис. 10.36. Поля направлений, которые были построены для изображения показанного на рис. 10.35*а* в ходе вычислительного эксперимента с выделением: 2 (*а*), 4 (*б*), 8 (*в*) и 16 (*г*) направлений

Как видно из графика, существует минимум погрешности вычисления при $n = 7 - 8$. При увеличении n среднеквадратичное отклонение сначала уменьшается в результате увеличения числа выделяемых направлений, но затем в результате того, что в спектре «вырезаются» все более узкие сектора, форма локальных областей одинакового направления начинает все более искажаться, из-за влияния стыков дифракционных решеток, и ε возрастает.

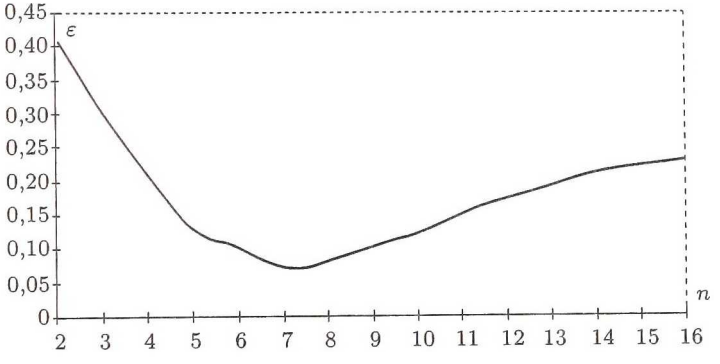


Рис. 10.37. Зависимость ε поля направлений от количества выделяемых направлений n

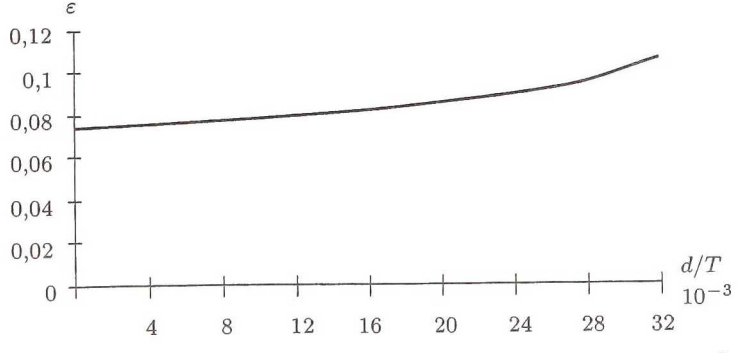


Рис. 10.38. Зависимость ε поля направлений от отношения периода дифракционной решетки пространственного фильтра d к ширине пространственного спектра T контурного изображения

В эксперименте у фильтра, изображенного на рис. 10.34*в*, менялся период дифракционных решеток; с каждым из этих фильтров вычислялось поле направлений от 10 тестовых изображений. На рис. 10.38 изображен график зависимости ошибки ε вычисления поля направлений от отношения периода дифракционной решетки пространственного фильтра d к ширине пространственного спектра T контурного изображения. Под шириной пространственного спектра, в данном случае, понимается минимальный диаметр круга с центром на оптической оси системы, который полностью покрывает области в частотной плоскости, где интенсивность отлична от нуля. Так как в эксперименте использовалось изображение, квантованное по 256 уровням, то условие отличия интенсивности от нуля заменялось на условие превышения интенсивности $1/256$ от максимальной интенсивности в спектре.

Видно, что чем меньше отношение d/T, тем меньше погрешность вычисления поля направлений для данного метода вычисления. Конечно при этом, существует ограничение на уменьшение периода дифракционной решетки, вызванное появлением эффектов квантования и дискретизации при синтезе пространственного фильтра, которые экспериментально не исследованы. Также уменьшение периода решетки ограничено технологическими возможностями, а расширение спектра ограничено геометрическими размерами установки. Поэтому в оптической установке отношение периода дифракционной решетки d к ширине спектра T следует выбрать, исходя из возможностей технической реализации в пределах от 0,0015 до 0,003.

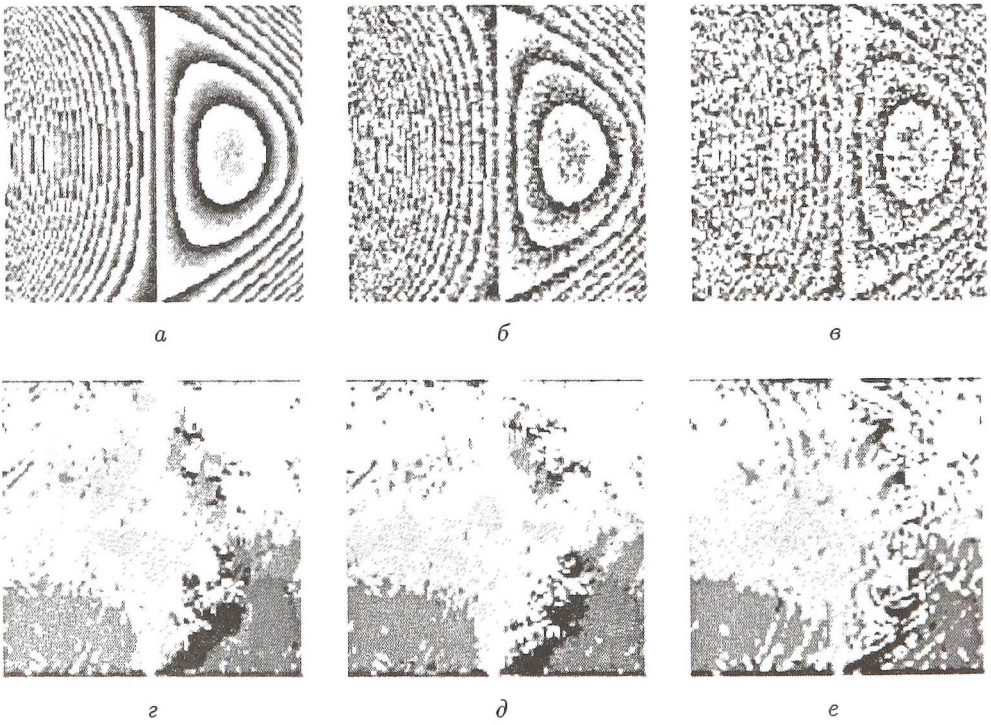


Рис. 10.39. Зашумленные изображения и их поля направлений, построенные с выделением 8 направлений: (*а*) изображение — сигнал/шум = 10, (*г*) его поле направлений, (*б*) изображение — сигнал/шум = 2, (*д*) его поле направлений, (*в*) изображение — сигнал/шум = = 1, (*е*) его поле направлений

Точность вычисления поля направлений оптическими методами может сильно изменяться не только с изменением параметров оптической установки, но и с изменением параметров самого изображения. В частности на качество поля направлений влияет отношение сигнал/шум, а также отношение периода полос на изображении к размеру зоны для одного направления. Для выявления множества изображений, для которого поле направлений может быть вычислено оптическим методом, был проведен ряд экспериментов, по результатам которых были определены интервалы изменения параметров изображений. При описании исходного изображения использовалась аппроксимация (10.123). Но большинство реальных контурных изображений имеют довольно существенные отклонения от этой аппроксимации, которые можно считать результатом действия шума. В качестве исходных были использованы 10 тестовых изображений. Каждое из 10 тестовых изображений зашумлялось, причем отношение сигнал/шум изменялось от 11 до 1. Затем был проведен вычислительный эксперимент по определению ошибки оптического формирования поля направлений зашумленных изображений. На рис. 10.39*а,в,д* показаны изображения с отношением сигнал/шум 10, 2, 1 соответственно; на рис. 10.39*б,г,е* их поля направлений.

Из рисунка 10.39 видно, что шум в изображении слабо влияет на структуру поля направлений. Шум генерировался с помощью датчика псевдослучайных чисел.

Так как ошибка вычисления, очевидно, должна изменяться не только от отношения сигнал/шум, но и от количества выделяемых направлений в изображении, исследования устойчивости метода вычисления поля направлений к шуму были проведены для случаев, когда количество выделяемых направлений изменялось от 2 до 16. На рис. 10.40 изображены графики зависимости ошибки вычисления поля направлений ε от отношения сигнал/шум для $n = 4, 8, 16$. В дальнейшем будем обозначать отношение сигнал/шум как S/N.

Из рис. 10.40 видно, что при $S/N > 4$ среднеквадратичное отклонение поля направлений практически постоянно при фиксированном n и почти не увеличивается, по сравнению с среднеквадратичным отклонением поля направлений от незашумленного изображения. При $S/N < 4$ среднеквадратичное отклонение начинает резко возрастать и в среднем для данного набора тестовых изображений при $S/N = 1$ достигает 0,35 (при $n = 8$).

Ошибка вычисления поля направлений, очевидно должна увеличиваться с увеличением периода полос и с уменьшением размера зон одинаковых направлений. Это связано с тем, что происходит «размазывание» фурье-спектра зоны изображения од-

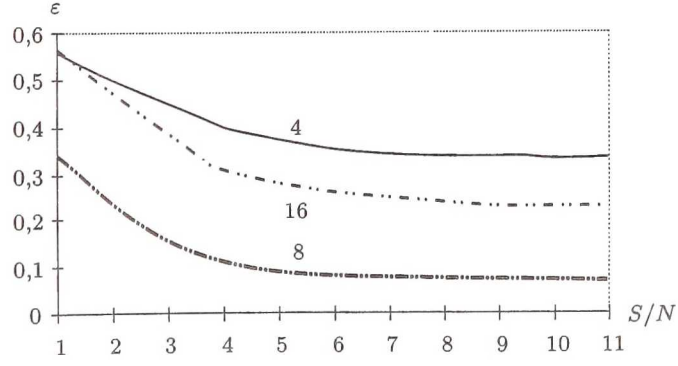
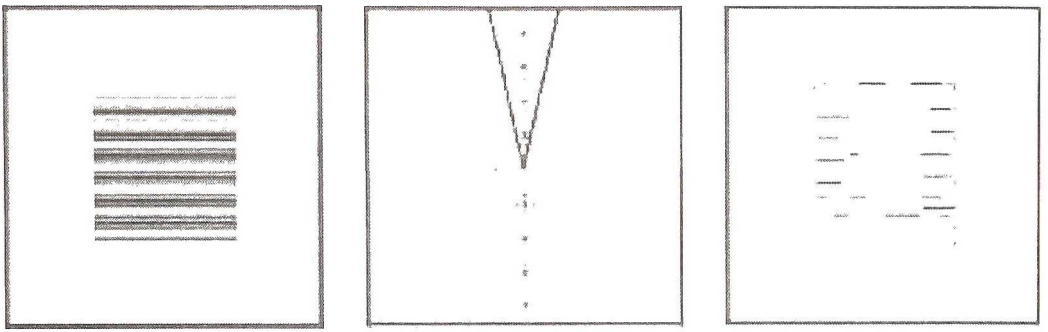


Рис. 10.40. Графики зависимости ошибки вычисления поля направлений ε от отношения S/N для $n = 4, 8, 16$



а *б* *в*

Рис. 10.41. Контурное изображение (*а*), результат секторной сегментации его пространственного спектра (*б*) и поле направлений (*в*)

ного направления так, что спектр попадает сразу в несколько угловых секторов. Для иллюстрации этого свойства проведем простой эксперимент. На рис. 10.41*а* показано изображение, представляющее собой набор горизонтальных полос ограниченных квадратом. На рис. 10.41*б* представлен результат фурье-преобразования от этого изображения и сектор, который «вырезается» из спектра в методе для выделения 4-х направлений [59].

Видно, что не весь спектр попадает в сектор, в результате чего в поле направлений (рис. 10.41*в*) появляются полосы, с другими направлениями. Очевидно, что чем меньше размер зоны одинакового направления, тем более широким будет ее спектр, и тем большая его часть не попадет в сектор, и тем больше будут искажения в поле направлений. Также, если период полос изображения будет увеличиваться, то спектр будет ближе к оптической оси и меньшая его часть будет попадать в тот же сектор. В результате увеличатся искажения формы локальной области одного направления в поле направлений. Под зоной одного направления здесь понимается область изображения, в которой направление полос принадлежит интервалу $\hat{\varphi} \in (\hat{\varphi}_i - \pi/2n, \hat{\varphi}_i + \pi/2n)$, где $\hat{\varphi}_i$ — выделенное направление, n — количество выделяемых направлений. В качестве характерного параметра, определяющего точность вычисления поля направлений, будет выступать отношение периода полос к размеру локальной области одного направления. В дальнейшем будем обозначать это отношение буквой Q.

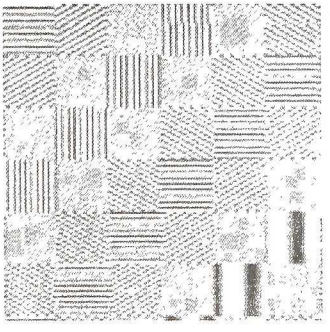


Рис. 10.42. Пример тестового изображения

На рис. 10.42 представлен пример одного из 10 тестовых изображений, для которых определялась зависимость ошибки ε от отношения Q. В экспериментах, как в численном, так и в натурном, выделялось 4 направления. Для каждой зоны одного направления ошибка ε считалась отдельно.

На рис. 10.43 представлен график зависимости ошибки вычисления поля направлений ε от Q. Кривая *1* соответствует численному эксперименту, кривая *2* соответствует натурному эксперименту. В натурном эксперименте тестовые изображения с помощью лазерного принтера выводились на прозрачную пленку. Для каждого из изображений, с помощью оптической установки, которая подробно описана в разделе 10.3.1, формировалось поле направлений.

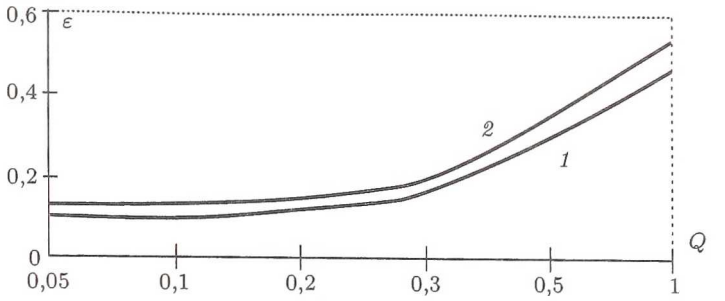


Рис. 10.43. График зависимости среднеквадратичного отклонения поля направлений, полученного в численном эксперименте (*1*) и полученного в натурном эксперименте (*2*), от отношения периода полос к размеру зоны одного направления

Как видно из графика для вычисления поля направлений с погрешностью меньше 10% необходимо, чтобы значение Q не превышало 0,2. Небольшое различие между графиками объясняется техническими погрешностями натурного эксперимента: неточное совмещение парциальных изображений, погрешности в изготовлении дифракционных решеток, небольшая расфокусировка парциальных изображений.

10.3.1. Оптическое распознавание дактилограмм

Для идентификации дактилограмм цифровыми методами [61] требуются значительные затраты машинного времени. Применение оптической предобработки позволяет, сократив время, достичь высокой степени сжатия информации для последующей цифровой обработки.

Известные оптико-цифровые системы [62, 63, 64, 65–68] идентификации дактилограмм основаны либо на корреляционном анализе, либо на анализе пространственного спектра. В работе [62] описаны два варианта (когерентный и некогерентный) оптико-цифровой системы идентификации дактилограмм, основанные на оптическом построении пространственного фурье-спектра отпечатка пальца и регистрации его с помощью сегментно-кольцевого фотоприемника. В результате этого в компьютере вместо матрицы отсчетов изображения отпечатка пальца находится вектор признаков из 64 компонент (фотоприемник разделен на 32 угловых сегмента и 32 полукольцевых).

Существуют гибридные оптико-цифровые устройства, в которых распознавание контурных изображений основано на оптическом выполнении преобразования Хоу–Радона [69, 70]; на использовании совместной корреляции двух изображений с помощью фоторефрактивных кристаллов [71]. В некоторых существующих системах отпечатки пальцев вводятся в компьютер в реальном времени с помощью призмы, используемой в режиме полного внутреннего отражения [71–73]. Быстрый оптический ввод отпечатка пальца в память компьютера и оптическая предобработка изображения существенно ускоряют процесс распознавания, и в этом преимущество гибридных систем перед цифровыми.

Однако существующие гибридные системы идентификации дактилограмм несвободны от недостатков. Метод анализа дактилограмм по фурье-спектру [62] будет интерпретировать изображения, отличающиеся перестановкой фрагментов, как одинаковые. Кроме того, дактилоскопические изображения являются структурно избыточными. Т.е. представляют собой совокупность линий с характерным перио-

дом, который составляет 0,5–0,7 мм. Пространственный спектр таких изображений имеет специфику: он преимущественно расположен в узкой полосе частот в окрестности несущей пространственной частоты. Именно несущая частота воспринимается визуально как множество контурных линий и является основным фактором структурной избыточности дактилоскопического изображения. В связи с этим, наличие 32 полукольцевых фотоприемников в фурье-спектре, измеряющих радиальное распределение интенсивности света [62], информационно избыточно.

Анализ и классификация структуры полос дактилограммы (папиллярного узора) с помощью оптической нейронной сети, предназначенной для выделения участков прямых линий на изображении методом преобразования Радона [70], или с помощью подсчета числа Галтона (число бифуркаций и центров полос дактилограммы) [69], являются операциями со значительными затратами машинного времени. Это связано с наличием высокой пространственной несущей частоты на дактилограмме, которая затрудняет поэлементную обработку изображения. В [74] предложен метод оптического построения поля направлений структурно избыточных изображений. Этот метод позволяет эффективно сжимать информацию, содержащуюся в изображениях с несущей пространственной частотой. Он основан, фактически, на замене участка изображения одним числом, равным среднему углу наклона полос на данном участке. В [76] описан численный эксперимент по идентификации дактилограмм (для 10 различных пальцев) с помощью полей направлений (выделялось 8 направлений), который продемонстрировал высокую надежность идентификации по сравнению с методом фурье-спектра [62].

Рассмотрим несколько реализаций оптико-цифрового метода распознавания отпечатков пальцев по их полю направлений, с помощью компьютерного моделирования, для определения наиболее эффективной из них, и реализации на экспериментальной установке [76].

В первом методе для построения поля направлений моделируется оптическая схема, изображенная на рис. 10.44*а*, где L — лазер, K — коллиматор, LA, LB — сферические линзы, TV — телекамера.

Отпечаток пальца на стекле F помещен на вход когерентного фурье-коррелятора (рис. 10.44*а*), в частотной плоскости которого расположен пространственный фильтр DF в виде амплитудного транспаранта из бинарных дифракционных решеток, делящих 180° на 8 секторов, другая часть фильтра не пропускает свет (рис. 10.44*б*). На выходе коррелятора регистрируется набор изображений, отличающихся от входного

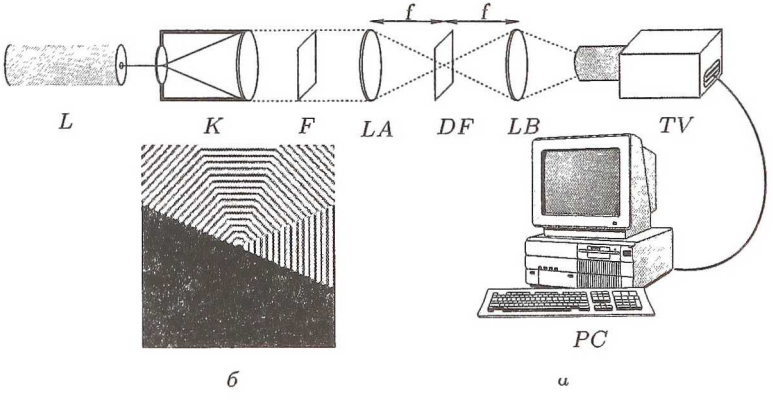


Рис. 10.44. Оптическая схема фурье-коррелятора для построения поля направлений с пространственным фильтром (*а*), из амплитудных секторных дифракционных решеток (*б*)

изображения тем, что каждое из них имеет только часть спектра, попадающую в соответствующий сектор. Так как каждый сегмент фильтра DF представляет собой амплитудную бинарную дифракционную решетку, то для каждого сектора получается пара симметричных изображений. Таким образом, из 16 полученных изображений, неповторяющихся — только 8. Далее эти 8 изображений в компьютере совмещаются в одно изображение с весами, соответствующими сектору и равными углу наклона полос на изображении отпечатка пальца.

Поле направлений в данном случае аппроксимируется выражением

$$\hat{\varphi}(x,y) = \frac{\sum\limits_{k=1}^{N} \hat{\varphi}_k I_k(x,y)}{\sum\limits_{k=1}^{N} I_k(x,y)}, \tag{10.126}$$

где $I_k(x,y)$ — функция интенсивности k-го неповторяющегося изображения, сформированного на выходе коррелятора в оптической схеме на рис. 10.44*а*, после прохождения светом k-го сектора решетки фильтра с центральным углом

$$\theta_k = \begin{cases} \dfrac{\pi}{2} - \hat{\varphi}_k, \, 0 < \hat{\varphi}_k < \dfrac{\pi}{2} \\ \dfrac{3\pi}{2} - \hat{\varphi}_k, \, \dfrac{\pi}{2} < \hat{\varphi}_k < \pi \end{cases}, \; k = \overline{1,N}, \tag{10.127}$$

где $\hat{\varphi}_k$ — отсчеты угла наклона полос на отпечатке, $N = 8$ — число секторов решетки.

Поле направлений, построенное по этому методу, будем в дальнейшем называть «многоканальным».

Заметим, что совмещение нескольких парциальных изображений, при получении поля направлений в ходе натурного эксперимента, происходит цифровым методом в компьютере. При этом необходима большая точность позиционирования парциальных изображений, что ведет к увеличению затрат времени на обработку. Поэтому рассматривается более простой в реализации метод получения поля направлений, но ошибка построения поля направлений ε в этом методе больше чем в методе (10.126). Оптическая схема для получения альтернативного поля направлений показана на рис. 10.45*а*.

Отпечаток пальца на стекле F помещен на вход когерентного фурье-коррелятора (рис. 10.45*а*), в частотной плоскости которого попеременно располагаются пространственные фильтры SF в виде амплитудных транспарантов. Один из них состоит из 8 секторов, коэффициент поглощения которого T пропорционален корню квадратному из угла. Эти сектора занимают 180°, другая часть фильтра не пропускает свет (рис. 10.45*б*). Другой фильтр просто закрывает половину частотной плоскости (рис. 10.45*в*). После последовательного применения обоих фильтров, на выходе коррелятора регистрируются два изображения, которые используются далее для вычисления в компьютере их отношения, аппроксимирующего поле направлений исходного изображения:

$$\hat{\varphi}(x,y) \cong \frac{|f(x,y) \otimes g_1(\xi,\eta)|^2}{|f(x,y) \otimes g_2(\xi,\eta)|^2 + \alpha^2}, \tag{10.128}$$

где $f(x,y)$ — функция исходного изображения, $g_1(\xi,\eta)$, $g_2(\xi,\eta)$ — фурье-образы от функции пропускания фильтров, изображенных на рисунках 10.45*б* и 10.45*в*, соответственно, α — регуляризующая константа. Поле направлений, полученное этим методом, будем называть «одноканальным».

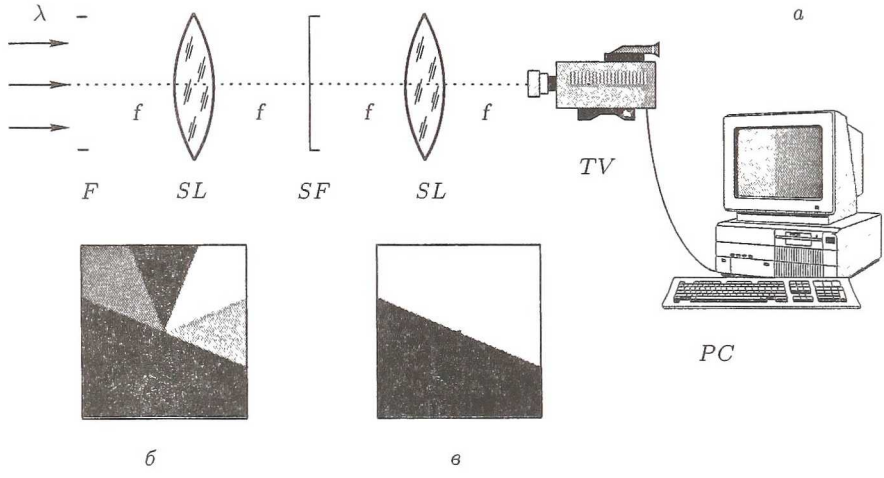


Рис. 10.45. Оптическая схема фурье-коррелятора (*а*) для построения поля направлений с секторным амплитудным пространственным фильтром (*б*) и амплитудным пространственным фильтром, который обрезает половину спектра (*в*)

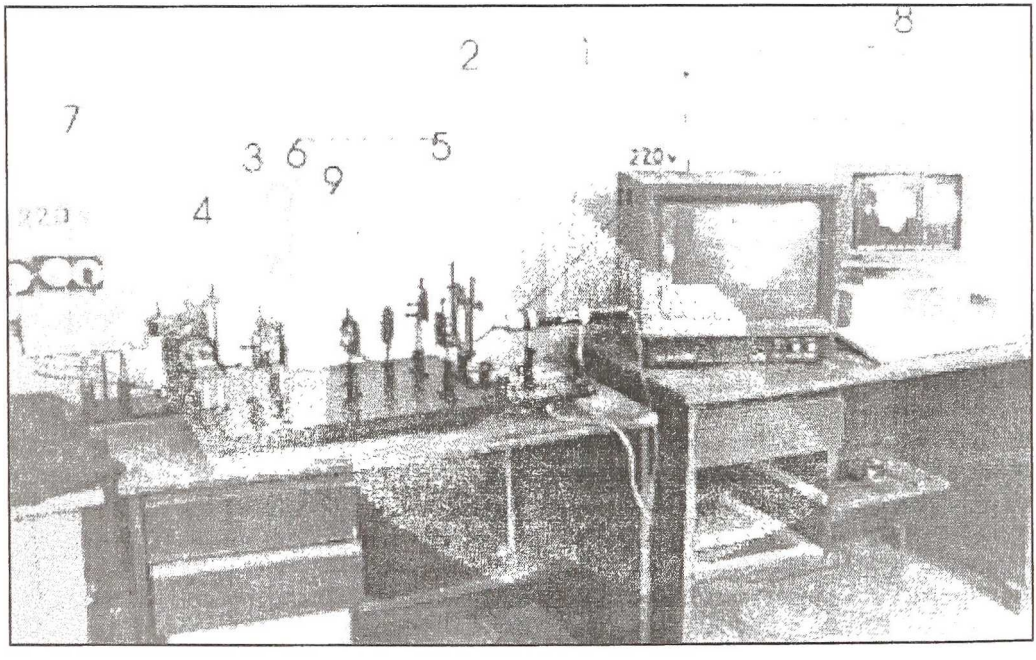


Рис. 10.46. Оптическая установка для вычисления поля направлений: *1* — He-Ne лазер, *2* — коллиматор, *3* — дактилограмма, *4* — призма, *5* — первая линза фурье-коррелятора, *6* — фазовый пространственный фильтр, *7* — телекамера, *8* — компьютер, *9* — система поворотных зеркал

В обоих рассмотренных методах, (10.126) и (10.128), присутствует операция деления на функцию, которая может приводить к заметным ошибкам ε вычисления поля направлений. Далее рассматривается метод поля направлений, в котором операция деления отсутствует.

В связи с наличием ряда технических ограничений, оптическая установка для вычисления поля направлений (рис. 10.46), на которой проводились натурные эксперименты имеет более сложную оптическую схему, чем схема которая моделировалась в численном эксперименте [79]. Внешний вид оптической установки для построения полей направлений показан на рис. 10.46.

На рис. 10.47*а* показана схема оптической установки для вычисления поля направлений. Амплитудный транспарант F с отпечатком пальца освещается пучком когерентного света от He-Ne лазера L. Пучок формируется коллиматором K. Призма P позволяет вводить отпечатки пальцев в реальном времени, если отпечаток вводится с амплитудного транспаранта, то призма играет роль поворотного зеркала. Система поворотных зеркал M_1, M_2, M_3 позволила смонтировать оптическую установку на небольшом столе. SL_1 и SL_2 — линзы фурье-коррелятора. SF — фазовый пространственный фильтр, который показан на рис. 10.47*б*. Так как размер поля дактилограммы — 20 мм, а размер поля камеры — 10 мм, линзы фурье-коррелятора имеют разные фокусные расстояния. Фокусные расстояния линз должны подбираться с учетом двух условий:

1) парциальные изображения должны быть разделены в пространстве;
2) все парциальные изображения должны умещаться в поле телекамеры.

Размер парциального изображения r определяется формулой (10.131), величина

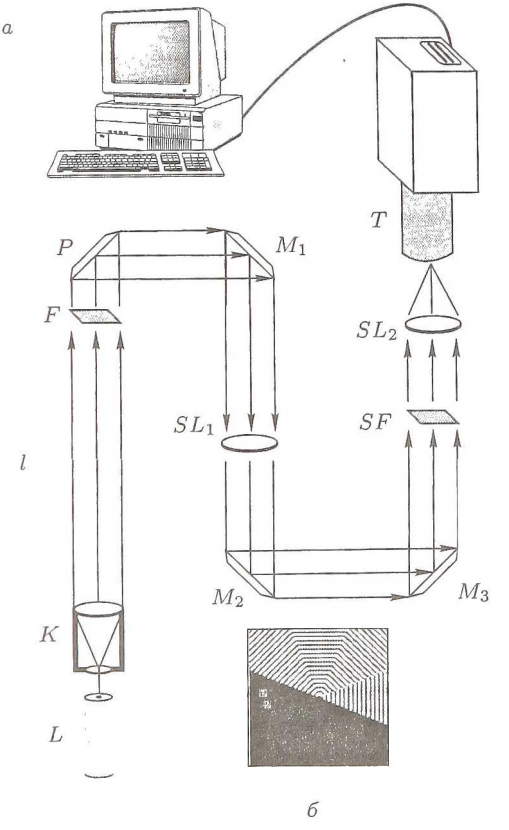


Рис. 10.47. Оптическая схема установки для вычисления поля направлений (*а*), фазовый пространственный фильтр (*б*)

отклонения частичного изображения от оптической оси r определяется формулой

$$r = f_2 \operatorname{tg}\left[\arcsin\left(\frac{\lambda}{d}\right)\right],$$

где λ — длина волны света, d — период решетки, f_2 — фокусное расстояние второго плеча фурье-коррелятора. Пространственный фильтр SF, представляет собой набор фазовых дифракционных решеток с периодом 20 мкм.

Для случая выделения 4 направлений первое условие (разделение парциальных изображений) можно записать в виде $r > \sqrt{2}S_c$, где S_c — размер парциального изображения.

Второе условие (размещение всех изображений в фотоприемном окне телекамеры) $L > 4S_c$, где L — размер поля телекамеры.

Из второго условия и формулы (10.122) следует условие для отношения фокусных расстояний плеч фурье-коррелятора:

$$\frac{f_2}{f_1} < \frac{L}{4S},$$

где S — размер дактилограммы. Подставляя реальные значения получим, что отношение фокусных расстояний должно быть меньше 0,125.

Первое условие дает нам минимальное значение $f_2 \approx 110$ мм, тогда минимальное значение $f_1 \approx 880$ мм.

Камера позволяет считывать изображение квантованное по 256 уровням в графический файл размерностью 512×512. Линза SL_1 имеет фокусное расстояние 1000 мм, линза SL_2 имеет фокусное расстояние 138 мм.

На рис. 10.48 показаны два варианта секторов в частотной плоскости коррелятора: бесконечно узкий сектор, заданный углом θ_0 ($а$), и сектор конечной ширины, ограниченный углами θ_1 и θ_2 ($б$).

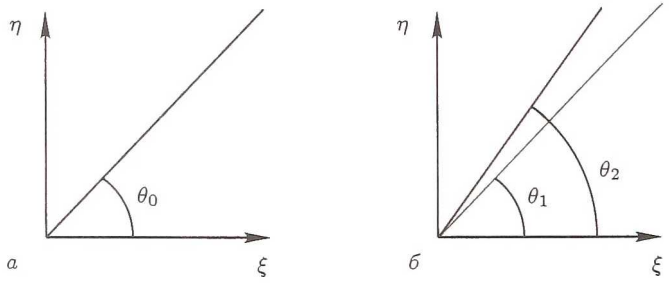


Рис. 10.48. Вид угловых секторов, располагаемых в частотной плоскости: бесконечно узкий сектор ($а$) и сектор конечной ширины ($б$)

Амплитуда светового поля $A(x', y')$ на выходе коррелятора с фильтрами, показанными на рис. 10.48, может быть выражена через амплитуду входного изображения $A_0(x, y)$ известным образом [80]:

$$A(x', y') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_0(x, y)\, h(x + x', y + y')\, \mathrm{d}x\, \mathrm{d}y, \qquad (10.129)$$

где $h(x,y)$ — частотная характеристика пространственного фильтра, являющаяся фурье-образом функции пропускания фильтра. Для узкого сектора, с заданным углом θ_0 (рис. 10.48*а*), функция пропускания $H(r,\theta)$ и частотная характеристика равны

$$H_1(r,\theta) = \delta(\theta - \theta_0), \tag{10.130}$$

$$h_1(x,y) = [x\cos\theta_0 + y\sin\theta_0]^{-2}, \tag{10.131}$$

где $\delta(x)$ — дельта-функция, а для конечного сектора, показанного на рис. 10.48*б*, соответствующие функции примут вид

$$H_2(r,\theta) = \begin{cases} 1, \theta_1 \leqslant \theta \leqslant \theta_2, \\ 0, 0 \leqslant \theta \leqslant \theta_1,\ \theta_2 < \theta < 2\pi, \end{cases} \tag{10.132}$$

$$h_2(x,y) = \sin(\theta_1 - \theta_2)\left[x^2\cos\theta_1\cos\theta_2 + y^2\sin\theta_1\sin\theta_2 + xy\sin(\theta_1+\theta_2)\right]^{-1}. \tag{10.133}$$

Если таких секторов N в плоскости фильтрации, и в каждом из них имеется дифракционная решетка, то на выходе коррелятора образуется последовательность парциальных изображений:

$$I_n(x',y') = \left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_0(x,y)\, h_n(x+x', y+y')\, \mathrm{d}x\, \mathrm{d}y \right|^2, \quad n = \overline{1,N}, \tag{10.134}$$

где функция $h_n(x,y)$ определяется из уравнений (10.130) или (10.132) при следующем выборе углов: $\theta_k + \Delta n$, $k = 0,1,2$, $n = 1,2,\ldots,N$, Δ — дискретность отсчетов по углу.

С помощью линейной комбинации парциальных изображений, взятых с весами, пропорциональными центральным углам щелей, можно определить функцию поля направления по формуле (10.126).

Заметим, что так как исходное изображение $A_0(x,y)$ описывается действительной функцией, то модуль фурье-образа такой функции будет центрально-симметричным, и поэтому диапазон изменений угла $\hat{\varphi}$ лежит в интервале $[0,\pi)$.

Хотя сектора решетки, формирующие парциальные изображения не пересекаются (рис. 10.44*б*), каждая точка исходного изображения, в принципе, вносит вклад в каждую точку парциального изображения. Изображения со структурной избыточностью (контурные изображения типа отпечатков пальцев) состоят из большого числа линий с характерным периодом или из участков дифракционных решеток, пространственный спектр каждой из которых вносит вклад только в один из угловых секторов фильтра. Поэтому следует ожидать, что парциальные изображения будут почти ортогональны. То есть «интеграл перекрытия» двух парциальных изображений будет много меньше суммарной энергии каждого из них:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_n(x,y)\, I_m(x,y)\, \mathrm{d}x\, \mathrm{d}y \ll \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_n^2(x,y)\, \mathrm{d}x\, \mathrm{d}y. \tag{10.135}$$

Если неравенство (10.135) не выполняется, то можно осуществить операцию ортогонализации парциальных изображений по следующему правилу:

$$\tilde{I}_k(x,y) = \begin{cases} 1, I_k(x,y) = \max\limits_n \{I_n(x,y)\}, \\ 0, I_k(x,y) \neq \max\limits_n \{I_n(x,y)\}, \end{cases} \tag{10.136}$$

где $\max\limits_n \{\ldots\}$ — максимальное значение функции в точке среди функций с номерами $n = \overline{1,N}$.

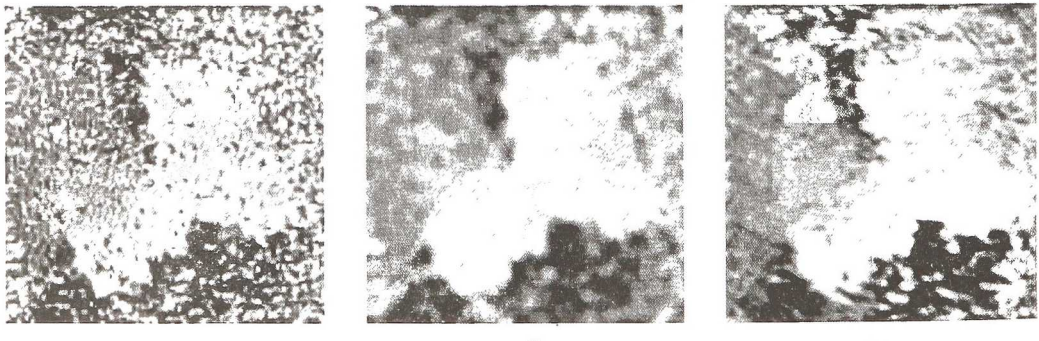
Из уравнения (10.136) видно, что парциальные изображения $\tilde{I}_k(x,y)$ представляют собой почти ортогональный набор бинарных функций. Ортогональность нарушается лишь в точках, где интенсивности парциальных изображений совпадают. Как показывает вычислительный эксперимент, количество этих точек не превышает 0,001 от общего количества точек изображения. Это позволяет вместо уравнения (10.126) рассчитывать поле направлений по более простой формуле без операции деления:

$$\hat{\varphi}(x,y) = \sum_{n=1}^{N} \hat{\varphi}_n \tilde{I}_n(x,y), \quad \hat{\varphi}_n = \frac{\pi n}{N}. \tag{10.137}$$

Формула (10.137), благодаря ортогонализации и бинаризации парциальных изображений, как видно из рис. 10.49*в*, содержит в себе процедуры частичной регуляризации и фильтрации получаемого поля направлений. По этой же причине, формулу (10.137) нельзя свести путем предельного перехода к формуле (10.113).

На рис. 10.49*а* показано поле направлений (128×128 пикселов), полученное путем численного моделирования коррелятора изображенного на рис. 10.45*а*, и вычисленное по формуле (10.128) для отпечатка пальца, изображенного на рис. 10.50*а*; на рис. 10.49*б* показано то же самое поле направлений, но обработанное медианным фильтром; на рис. 10.49*в* показано поле направлений от того же отпечатка пальца, рассчитанное по формуле (10.126). Как видно из рис. 10.50 эти поля направлений совпадают в крупных деталях.

На рис. 10.50*б,д,з,л* показаны поля направлений (128×128 пикселов), рассчитанные на компьютере по формуле (10.126).



а *б* *в*

Рис. 10.49. Поле направлений (*а*), полученное путем численного моделирования работы коррелятора изображенного на рис. 10.45*а*, поле направлений (*б*), полученное путем численного моделирования работы коррелятора изображенного на рис. 10.45*а* и обработанное медианным фильтром, и поле направлений (*в*), рассчитанное на компьютере по формуле (10.137)

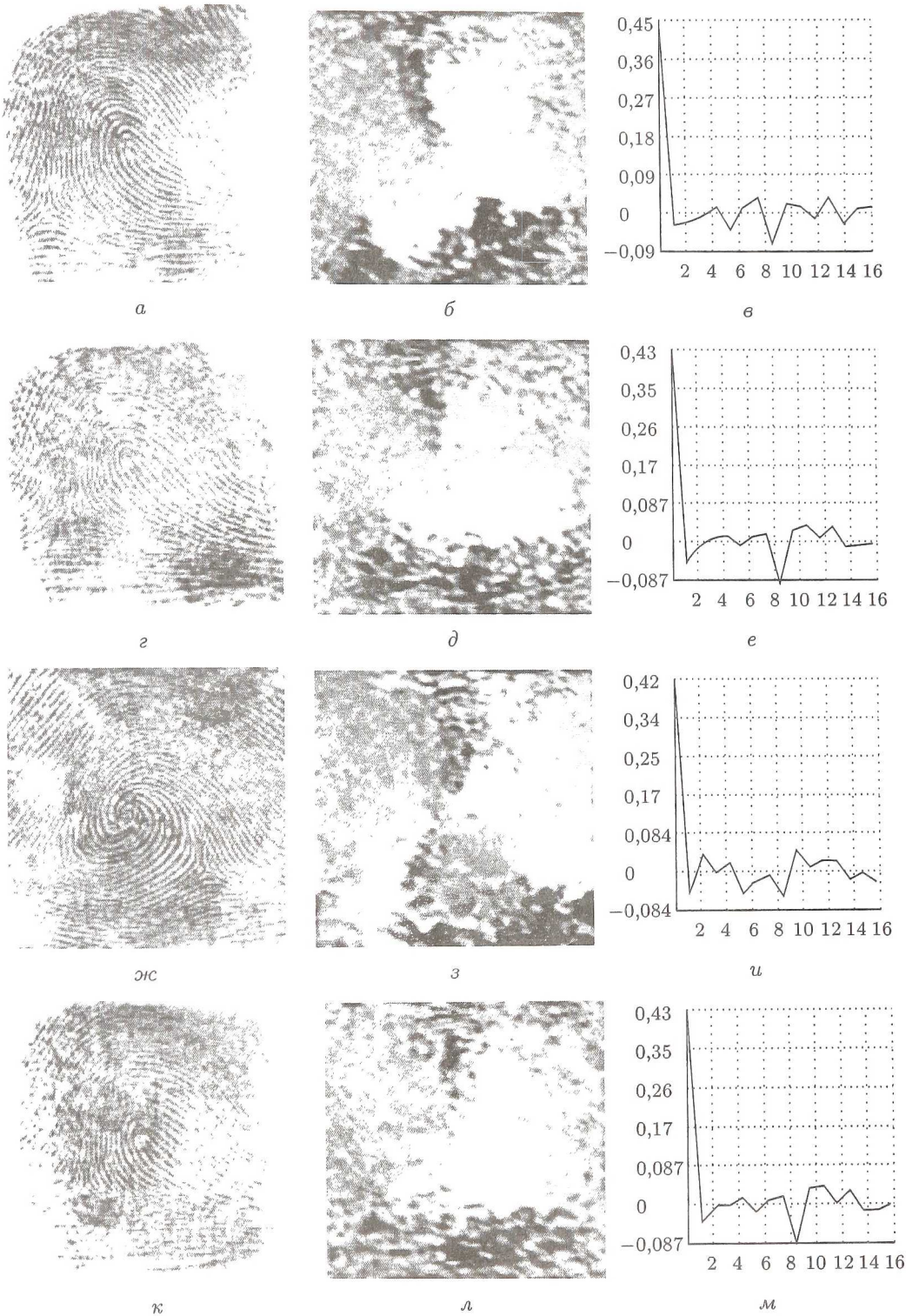


Рис. 10.50. Отпечатки пальцев (*а,г,ж,к*), их поля направлений (*б,д,з,л*) и вектора признаков из 16 первых коэффициентов разложения Адамара (*в,е,и,м*)

Рассмотрим как влияет на процесс идентификации дактилограмм способ построения векторов признаков для полей направлений. Будем формировать вектор признаков с помощью коэффициентов разложения полей направлений по ортогональному базису. Для этого будем использовать базисы Адамара (10.58) и Фурье (10.54) [57, 58].

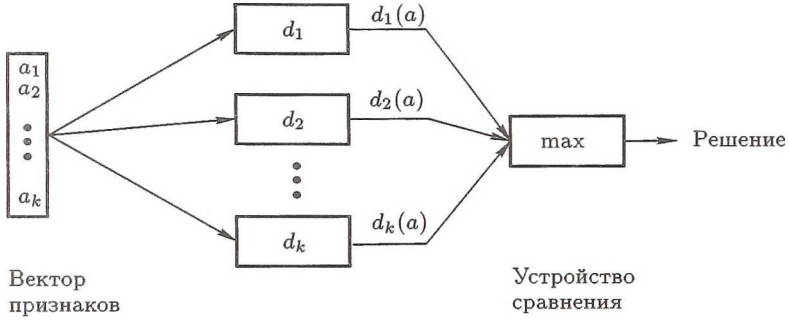


Рис. 10.51. Схема классификатора образов

В таблице 10.8 представлены значения евклидового расстояния d_{mn} между векторами признаков из $K = 16$ компонент для 10 базовых и 10 анализируемых отпечатков 10 разных пальцев. Расстояние d_{mn} вычислялось по формуле

$$d_{mn} = \sqrt{\sum_{k=1}^{K} \left[b_m^{(k)} - a_n^{(k)} \right]^2}, \tag{10.138}$$

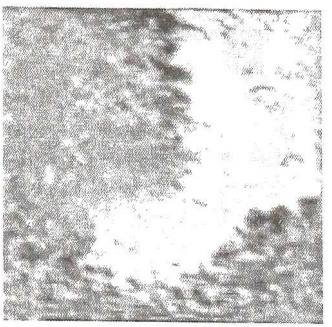
где $b_m^{(k)}$ и $a_n^{(k)}$ — k-е коэффициенты m-го базового и n-го анализируемого векторов, соответственно. По сути своей $d_{mn}\,(a_n)$, $m = \overline{1, N}$, представляют собой систему разделяющих функций, которые описывают классификатор, блок-схема которого изображена на рис. 10.52. Говорят, что классификатор ставит в соответствие вектор

Таблица 10.8. Расстояния между векторами признаков из 16 коэффициентов разложения по ортогональному базису Адамара для полей направлений, построенных по 10 базовым $\{b_n\}$ и 10 анализируемым $\{a_n\}$ отпечаткам 10 различных пальцев

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	R
a_1	5,13	13,00	9,73	8,83	11,07	12,56	7,99	10,12	11,42	13,39	0,22
a_2	13,37	3,97	9,52	10,20	10,03	7,96	12,14	10,13	7,39	8,17	0,30
a_3	8,70	7,09	4,56	8,21	6,05	10,65	8,24	7,64	7,66	9,29	0,13
a_4	8,31	8,84	5,90	4,58	6,79	9,78	6,61	8,16	8,75	11,41	0,12
a_5	9,42	9,53	6,18	6,36	2,07	9,15	7,58	5,94	8,21	9,71	0,48
a_6	11,74	7,76	9,86	9,22	9,03	3,47	10,01	7,23	7,84	9,32	0,35
a_7	8,92	10,54	5,99	5,64	4,71	10,52	4,22	7,34	10,61	12,52	0,06
a_8	10,94	9,25	9,04	9,42	7,09	8,53	9,78	3,52	5,31	6,10	0,20
a_9	11,11	7,82	9,56	9,69	9,80	9,63	10,85	7,75	3,60	5,25	0,18
a_{10}	12,74	8,52	11,63	11,90	11,17	10,15	12,74	8,94	4,36	3,58	0,10
R	0,24	0,28	0,12	0,10	0,38	0,39	0,22	0,15	0,10	0,18	0,21



а



б

$$\begin{bmatrix} a_1 & a_6 & a_{11} & a_{16} & a_{21} \\ a_2 & a_7 & a_{12} & a_{17} & a_{22} \\ a_3 & a_8 & a_{13} & a_{18} & a_{23} \\ a_4 & a_9 & a_{14} & a_{19} & a_{24} \\ a_5 & a_{10} & a_{15} & a_{20} & a_{25} \end{bmatrix}$$

в

Рис. 10.52. отпечаток пальца (*а*), его поле направлений (*б*) и матрица коэффициентов, полученная с помощью рассчитанного ДОЭ (рис. 10.24*а*) (*в*)

признаков a_n классу b_m, если справедливо неравенство

$$d_{mn}(a_n) > d_{ln}(a_n),\ m \neq l. \tag{10.139}$$

Значения расстояний в каждой строке таблицы 10.8 показывают результат идентификации вектора $\mathbf{a}_m$ среди набора базовых векторов $\{\mathbf{b}_m\}$, $m = \overline{1, M}$, а значения в столбцах показывают результат идентификации вектора $\mathbf{b}_m$ среди набора анализируемых векторов $\{\mathbf{a}_n\}$, $n = \overline{1, N}$. Количество распознаваемых дактилограмм невелико, поэтому прямо определить надежность метода, как отношение числа правильно распознанных дактилограмм к общему их количеству, нельзя, но можно ввести параметр, который характеризует относительную удаленность классов друг от друга [76], и тем самым позволяет косвенно судить о надежности метода:

$$R_i = \frac{d_{\min} - d_{mm}}{d_{\min} + d_{mm}}, \quad i = \begin{cases} m, d_{\min} \text{ определено в строке,} \\ M + m, d_{\min} \text{ определено в столбце,} \end{cases} \tag{10.140}$$

где d_{mm} — диагональный элемент матрицы, $d_{\min}$ — минимальный недиагональный элемент.

Среднее значение параметра $\langle R \rangle$, характеризует среднюю удаленность классов для метода распознавания в целом и определяется формулой

$$\langle R \rangle = (2M)^{-1} \sum_{i=1}^{2M} R_i, \tag{10.141}$$

где $2M$ — число анализируемых и базовых дактилограмм.

Таким образом, рассматриваемый метод идентификации дактилограмм состоит из последовательности шагов.

1) Для каждого из анализируемых или базовых отпечатков пальцев, описываемых функцией пропускания $I(x, y)$, оптическими средствами с помощью коррелятора формируются N парциальных изображений $I_n(x, y)$, имеющих пространственный спектр, заключенный в узком угловом секторе. Набор парциальных изображений регистрируется и вводится в виде массива чисел в память компьютера.

2) Цифровая обработка парциальных изображений состоит из этапов:

а) формируется набор бинарно-ортогональных парциальных изображений $\tilde{I}_n(x, y)$;

б) по формуле (10.126) или (10.137) рассчитывается поле направлений $\hat{\varphi}(x, y)$;

в) рассчитывается K коэффициентов $\left(\left\{a_k^{(n)}\right\}, \left\{b_k^{(n)}\right\}\right)$ разложения поля направлений $\hat{\varphi}(x, y)$ по базису Адамара;

г) определяются евклидовы расстояния $d_{m,n}$ и ищется минимальное среди них.

Так как вектора $\mathbf{b}_n$ и $\mathbf{a}_n$ с одинаковыми номерами соответствуют разным отпечаткам одного пальца, то идентификация будет успешной, если минимальные значения расстояний $d_{m,n}$ в каждой строке и в каждом столбце находятся на диагонали, при этом величина R должна быть положительной.

В результате распознавания с помощью векторов признаков, построенных из первых 16 коэффициентов разложения полей направлений по базису Фурье, не удалось добиться полной идентификации отпечатков, и величина $\langle R \rangle$ меньше ($\langle R \rangle = 0{,}19$), чем аналогичная величина при распознавания по коэффициентам разложения в базисе Адамара ($\langle R \rangle = 0{,}21$). В данном случае базис Адамара является более эффективным.

Заметим также, что если использовать для распознавания тех же отпечатков пальцев поля направлений, полученные с помощью упрощенной «одноканальной» оптической схемы, поле направлений для которой показано на рис. 10.49*а*, и вектор признаков выбирать из 16 первых коэффициентов разложения Адамара, то результаты так же ухудшаются. Полная идентификация имеет место и в этом случае, но из-за наличия мелкомасштабной структуры в полях направлений (рис. 10.49*а*), надежность распознавания ухудшается.

Надежность распознавания по этим полям направлений можно улучшить, если обработать исходные поля направлений медианным фильтром.

Тогда имеет место полная идентификация, среднее относительное расстояние между классами увеличивается $\langle R \rangle = 0{,}19$.

Общее сравнение среднего относительного расстояния между классами для всех рассмотренных методов приведено в таблице 10.9. Из таблицы видно, что наилучшие характеристики имеет метод, в котором вектор признаков набирается из первых коэффициентов разложения по базису Адамара с использованием «многоканального» поля направлений.

В разделе 10.2.2 описывается ДОЭ, который может использоваться в задаче анализа изображений. На рисунке 10.52 представлены: (*а*) отпечаток пальца, (*б*) его поле направлений и (*в*) матрица коэффициентов, полученных с помощью ДОЭ, показанного на рис. 10.24*а*. Эти коэффициенты пропорциональны модулям коэффициентов разложения дактилограммы по 25 базисным функциям Адамара.

Общая схема распознавания на основе оптического разложения Карунена–Лоэва представлена на рис. 10.53.

Эффективность данной схемы состоит в том, что наиболее трудоемкая часть («начальное обучение») проделывается один раз на начальном этапе. Основная же часть — анализ поступающего на вход изображения — обладает всеми желательны-

Таблица 10.9. Значения показателя надежности для различных методов идентификации

Метод обработки изображения	Метод набора вектора признаков	$\langle R \rangle$
«Многоканальное» поле направлений	16 коэффициентов разложения по базису Адамара	0,21
	16 коэффициентов разложения по базису Фурье	0,19
«Одноканальное» поле направлений	16 коэффициентов разложения по базису Адамара	0,17
«Одноканальное» отфильтрованное поле направлений	16 коэффициентов разложения по базису Адамара	0,19

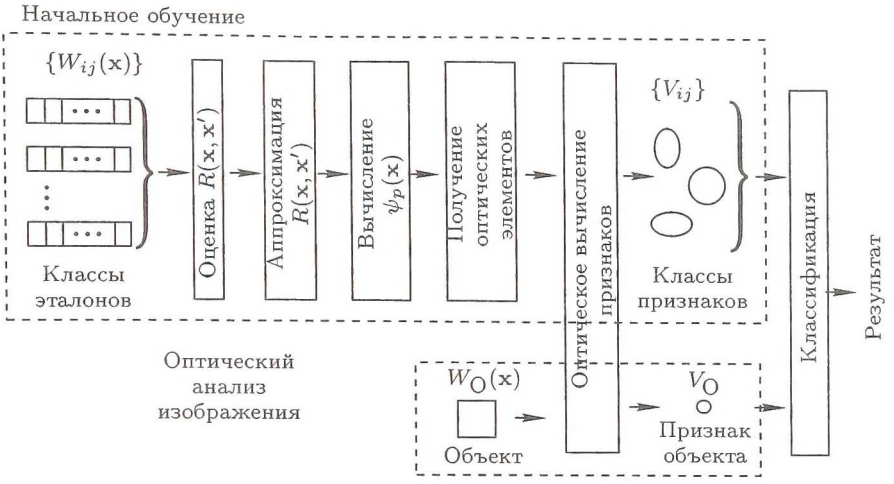


Рис. 10.53. Общая схема распознавания на основе оптического РКЛ

ми характеристиками: быстродействием, благодаря оптической реализации, и эффективностью, благодаря выбору разложения, имеющего оптимальные свойства. Все это, естественно, отражается на усложнении начального этапа, который, однако, проделывается всего один раз, а следовательно, целесообразность такого подхода налицо.

При классификации в пространстве признаков необходимо выбрать решающую функцию, которая будет классифицировать объект в зависимости от «расстояния» r_i от объекта до i-го класса. Некоторые методы классификации приведены в [11].

Например, метод среднего для класса вектора признаков заключается в следующем.

Для каждого класса по векторам V_{ij} находится средний вектор признаков — центр класса

$$V_i = \frac{1}{N_s}\sum_{j=1}^{N_s} V_{ij}, \qquad (10.142)$$

затем вычисляется евклидово расстояние r_i^c от объекта до центров классов:

$$r_i^c = |V_o - V_i|, \qquad (10.143)$$

и выбирается минимальное.

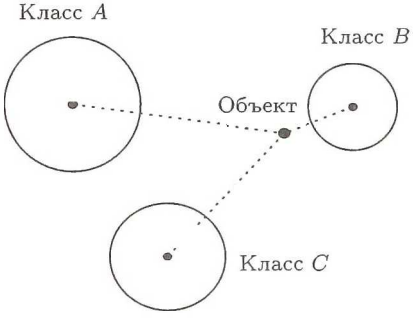


Рис. 10.54. Схема для определения расстояния от объекта до различных классов

Номер минимального расстояния будет считаться номером класса, к которому принадлежит объект (см. рис. 10.54).

Результаты распознавания 10 классов близких по рисунку дактилоскопических изображений для выборки из 15 представителей на основе РКЛ (см. раздел 10.2.1) с различным типом аппроксимации корреляционной функции приведены в табл. 10.10 (W — частота успеха, K — количество коэффициентов РКЛ). Поле направлений строилось как оптическим методом, так и дисперсионным методом [74]. Поле направлений, построенное дисперсионным методом, подвергалось операции фильтрации.

Таблица 10.10. Результаты распознавания дактилоскопических изображений

		РКЛ с экспоненциальной корреляционной функцией	РКЛ с экспоненциально-косинусной корр. функцией
Исходные изображения		$W = 0{,}3$ при $K \geqslant 18^*$	$W = 0{,}3$ при $K \geqslant 9^*$
Поля направлений	Дисперсионные	$W = 1$ при $K \geqslant 34$	$W = 1$ при $K \geqslant 31$
	Фильтр. дисперсионые	$W = 0{,}8$ при $K \geqslant 22^*$	$W = 0{,}8$ при $K \geqslant 15^*$
	Оптические	$W = 1$ при $K \geqslant 12$	$W = 1$ при $K \geqslant 11$

* — увеличение K до 100 не улучшает результата.

Тот факт, что исходные дактилоскопические изображения практически не поддавались распознаванию [78], не только оправдывает, но и делает необходимым использование поля направлений при анализе таких сложных изображений. Ситуацию с ухудшением качества распознавания для отфильтрованных полей направлений можно объяснить подавлением некоторых мелких, но существенные деталей при фильтрации, а также фильтрующим свойством РКЛ, позволяющим опустить этот достаточно сложный этап. Оптический метод показал в этом эксперименте высокую эффективность.

При использовании уравнения (10.126), поле направлений отражает не только структуру основных полос дактилограммы, но и побочные структурные элементы, возникающие из-за колебаний контраста полос изображения. Но хотя их вклад в частичное изображение относительно невелик, в целом это может искажать структуру поля направлений. На рис. 10.55 изображен график нормированной интенсивности для одной характерной точки в 4-х разных «парциальных» изображениях.

Как видно из рис. 10.55 наибольшая интенсивность в произвольно выбранной точке (x, y) наблюдается в частичном изображении под номером 1. Интенсивность в этой же точке в других парциальных изображениях существенно меньше, но сумма интенсивностей в точке в трех остальных парциальных изображениях уже сравнима с интенсивностью в этой точке в первом частичном изображении. Подобное отношение интенсивностей наблюдается в большинстве точек. Поэтому в эксперименте использовали для получения поля направлений формулы (10.136), (10.137).

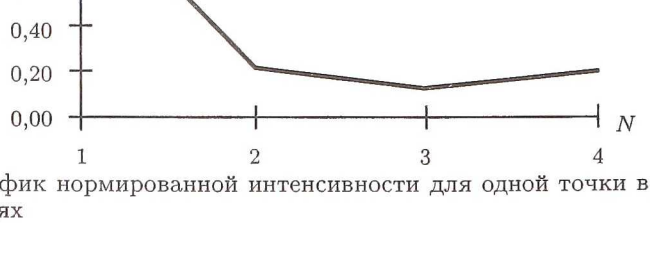


Рис. 10.55. График нормированной интенсивности для одной точки в разных парциальных изображениях



а

б *в*

Рис. 10.56. Экспериментальные парциальные изображения (*а*) и поля направлений, полученные из них по формулам (10.126) (*б*) и (10.137) (*в*) (изображения в рамках участвовали в построении поля направлений)

В эксперименте с помощью оптической схемы, показанной на рис. 10.47*а*, получены парциальные изображения, показанные на рис. 10.56*а*.

Четыре из них участвовали в построении поля направлений, рассчитанного по формулам (10.126) и (10.137). Эти поля показаны на рис. 10.56*б,в*, соответственно. Размер парциальных изображений был 100×100 пикселов, все поле изображения имеет размер 512×512 пикселов. Изображения квантованы по 256 уровням. Геометрический размер поля телекамеры составлял 10×10 мм, геометрический размер частичного изображения составляет около 2 мм. Как и в численном эксперименте в качестве базисных выбирались 10 отпечатков. Базисные вектора признаков составлялись из коэффициентов разложения по базису Адамара полей направлений базовых отпечатков. Выбирались 10 анализируемых отпечатков, и определялось евклидово расстояние d_{mn} от вектора каждого анализируемого отпечатка до вектора каждого базового отпечатка по формуле (10.147); во всех экспериментах использовались первые 16 коэффициентов разложения Адамара.

На рис. 10.57*а* изображен график евклидового расстояния d_{mn} между вектором признаков $\mathbf{a}_5$ и 10 базовыми векторами 10 разных пальцев. На рис. 10.57*б* изображен

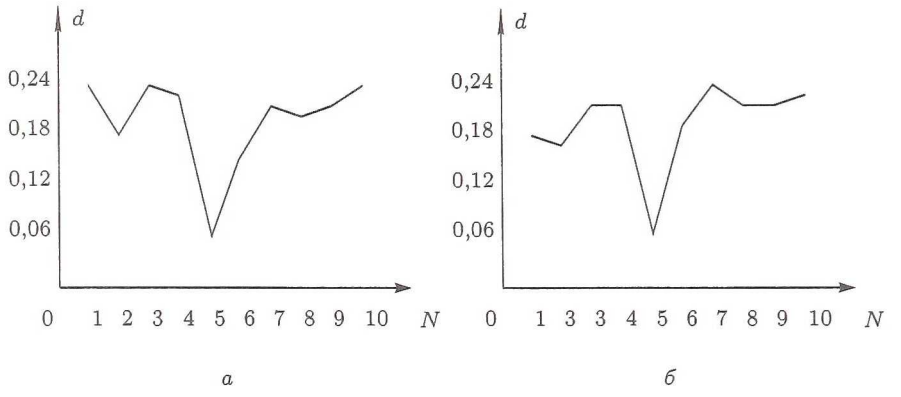


Рис. 10.57. Расстояния между вектором анализируемого отпечатка $\{\mathbf{a}_5\}$ и векторами 10 базовых отпечатков $\{\mathbf{b}_n\}$ ($а$) и расстояния между вектором базового отпечатка $\{\mathbf{b}_5\}$ и векторами 10 анализируемых отпечатков $\{\mathbf{a}_n\}$ ($б$)

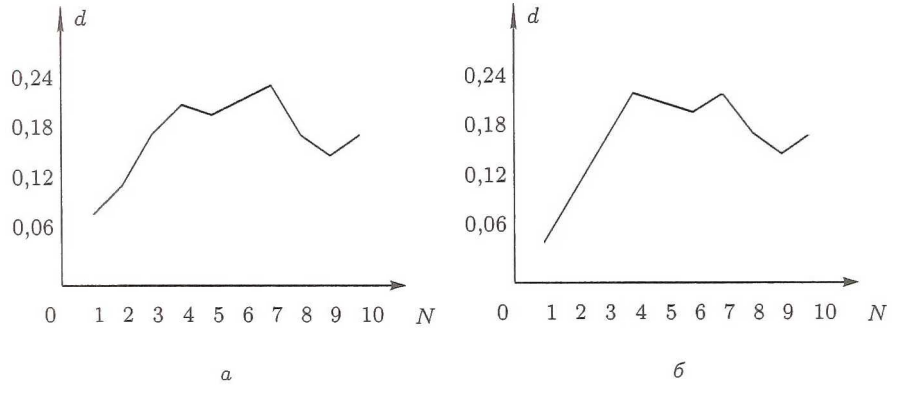


Рис. 10.58. Расстояния между вектором анализируемого отпечатка $\{\mathbf{a}_1\}$ и векторами 10 базовых отпечатков $\{\mathbf{b}_n\}$ ($а$) и расстояния между вектором $\{\mathbf{a}_1\}$ и векторами 10 базовых отпечатков $\{\mathbf{c}_n\}$, которые получены путем усреднения векторов от двух вариантов отпечатков одного пальца

график евклидового расстояния d_{mn} между вектором признаков $\mathbf{b}_5$ и 10 анализируемыми векторами 10 разных пальцев.

Как видно из рис. 10.57 расстояние минимально между векторами $\mathbf{a}_5$ и $\mathbf{b}_5$, т. е. происходит распознавание. То же самое происходит с любой другой парой векторов. При распознавании отпечатков по разложению Адамара их оптических полей направлений, полученных по формуле (10.137), значение $\langle R \rangle$ равно 0,24, это несколько выше чем значение аналогичного параметра, при распознавании того же набора отпечатков по разложению Адамара их полей направлений, но полученных путем численного моделирования [74] данной оптической установки по формуле (10.126): $\langle R \rangle = 0{,}21$.

Проводился также эксперимент по распознаванию дактилограмм по полям направлений, которые были получены оптическим путем с использованием формулы (10.126). В этом случае не удалось добиться распознавания всех отпечатков, и среднее относительное расстояние между классами для этого метода оказалось существенно меньше: $\langle R \rangle = 0{,}14$. Ухудшение надежности, по сравнению с результатом

Таблица 10.11. Расстояния между векторами признаков из 16 компонент для 10 базовых $\{\mathbf{b}_n\}$ векторов, которые составлены из двух разных отпечатков одного пальца и 10 анализируемых $\{\mathbf{a}_n\}$ векторов отпечатков 10 различных пальцев; последний столбец и последняя строка состоят из значений величины R

	$\mathbf{b}_1$	$\mathbf{b}_2$	$\mathbf{b}_3$	$\mathbf{b}_4$	$\mathbf{b}_5$	$\mathbf{b}_6$	$\mathbf{b}_7$	$\mathbf{b}_8$	$\mathbf{b}_9$	$\mathbf{b}_{10}$	R
$\mathbf{a}_1$	4,8	10,0	15,0	20,0	18,0	18,0	21,0	16,0	13,0	15,0	0,35
$\mathbf{a}_2$	12,0	5,6	14,0	18,0	17,0	16,0	19,0	17,0	12,0	16,0	0,36
$\mathbf{a}_3$	13,0	15,0	4,9	11,0	21,0	19,0	14,0	14,0	11,0	12,0	0,38
$\mathbf{a}_4$	16,0	18,0	10,0	5,3	21,0	14,0	8,8	13,0	10,0	12,0	0,25
$\mathbf{a}_5$	22,0	14,0	22,0	21,0	4,0	13,0	21,0	19,0	20,0	22,0	0,53
$\mathbf{a}_6$	16,0	17,0	17,0	12,0	17,0	6,1	12,0	11,0	13,0	16,0	0,29
$\mathbf{a}_7$	17,0	21,0	15,0	8,7	24,0	16,0	6,0	8,5	9,8	9,1	0,17
$\mathbf{a}_8$	15,0	18,0	13,0	7,9	21,0	14,0	7,2	5,1	7,8	7,0	0,16
$\mathbf{a}_9$	13,0	17,0	11,0	8,9	21,0	15,0	7,7	6,4	4,8	6,6	0,14
$\mathbf{a}_{10}$	14,0	19,0	15,0	12,0	22,0	17,0	11,0	6,7	10,0	6,1	0,05
R	0,43	0,28	0,34	0,20	0,62	0,36	0.09	0,11	0,24	0,04	0,27

численного моделирования [59], связано с тем, что в натурном эксперименте выделялось 4 направления, а в численном — 8 направлений.

Для увеличения надежности распознавания необходимо, чтобы базовый вектор был составлен как среднее от векторов полей направлений нескольких вариантов отпечатка одного пальца. На рис. 10.58*а* представлены расстояния между вектором $\{\mathbf{a}_1\}$ и 10 базовыми векторами, на рис. 10.58*б* показаны расстояния между тем же вектором $\{\mathbf{a}_1\}$ и 10 базовыми векторами, каждый из которых получен путем усреднения по двум векторам от разных отпечатков одного пальца.

Видно, что разность расстояний $d_{1,1}$ и $d_{2,1}$, на рис. 10.58*б* больше. Следовательно увеличивается параметр $\langle R\rangle$.

В табл. 10.11 базовые вектора получены с усреднением, т. е. центром класса считается не координаты базового вектора, а усредненные координаты двух базовых векторов.

Из табл. 10.11 видно, что минимальные расстояния располагаются на диагонали, т. е. имеет место полное распознавание дактилограмм. Видно также, что операция усреднения для базовых отпечатков повысила надежность с $\langle R\rangle = 0{,}24$ до $\langle R\rangle = 0{,}27$.

Для исследования устойчивости полей направлений к сдвигу и повороту в [82] был проведен ряд натурных экспериментов. Вектора признаков составлялись из коэффициентов разложения по базису Адамара полей направлений отпечатков, и определялось евклидово расстояние $r_{m,n}$ от вектора каждого сдвинутого отпечатка до вектора каждого не сдвинутого отпечатка по формуле (10.138); во всех экспериментах использовались первые 16 коэффициентов разложения Адамара. Сдвиг осуществлялся с интервалом 0,5 мм на расстояние до 4 мм.

На рис. 10.59 изображен график зависимости евклидового расстояния r между вектором признаков $\mathbf{a}_3$ и векторами признаков сдвинутых на расстояние l отпечатков одного пальца

Для того, чтобы определить устойчивость к сдвигу, необходимо знать среднее минимальное расстояние между векторами признаков и дисперсию. Среднее минимальное расстояние между 100 не сдвинутыми векторами в 16-мерном пространстве



Рис. 10.59. График зависимости расстояния между векторами признаков сдвинутого на расстояние l и не сдвинутого отпечатков пальца от значения сдвига

признаков $\langle r \rangle = 0{,}085$, дисперсия: $D = 0{,}00044$. Для 50 векторов было получено: $\langle r \rangle = 0,089$ и $D = 0{,}00045$. Вероятность того, что $r < \langle r \rangle / 2$ согласно [83] равна

$$P\left\{r < \frac{\langle r \rangle}{2}\right\} = 0{,}5 - \Phi\left(\frac{r - \langle r \rangle / 2}{\sqrt{D}}\right), \tag{10.144}$$

где r — расстояние от вектора до центра класса.

На рис. 10.60 изображен усредненный график зависимости расстояния вектора сдвинутого отпечатка от вектора не сдвинутого отпечатка от значения сдвига.

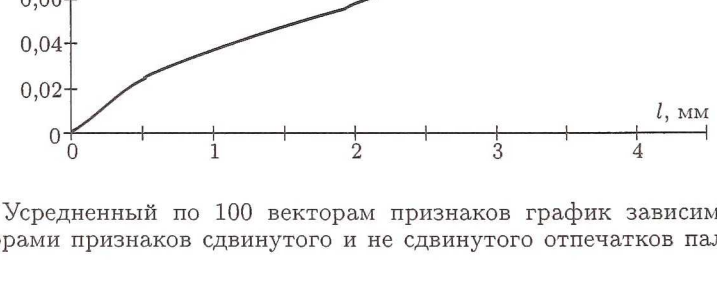


Рис. 10.60. Усредненный по 100 векторам признаков график зависимости расстояния между векторами признаков сдвинутого и не сдвинутого отпечатков пальца от значения сдвига

Зная эту зависимость, можно по формуле (10.144) рассчитать зависимость вероятности правильного распознавания P от значения сдвига. График этой зависимости представлен на рис. 10.61.

На рис. 10.62 изображена зависимость параметра $\langle R \rangle$ от значения сдвига.

Из сравнения рисунков 10.61 и 10.62 видно, что при $\langle R \rangle = 0$, $P = 0{,}5$, т. е., как и предполагалось, положительное значение параметра $\langle R \rangle$ позволяет утверждать, что правильная идентификация преобладает над ошибочной.

В качестве базовых в эксперименте используются не сдвинутые отпечатки, а в качестве анализируемых — сдвинутые. На рис. 10.63 изображена зависимость надежности распознавания K от значения сдвига, полученная обработкой экспериментальных данных, где K определяется по формуле

$$K = N_r / N, \tag{10.145}$$

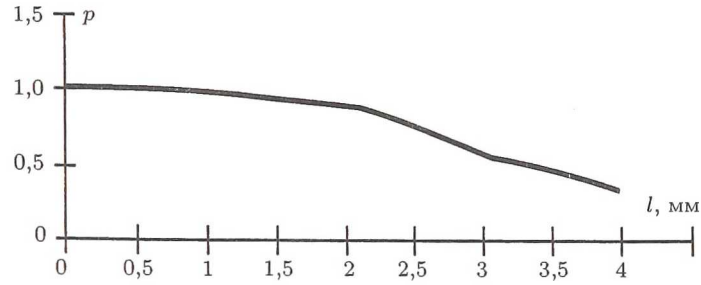


Рис. 10.61. Зависимость вероятности правильного распознавания P от значения сдвига исходного изображения, рассчитанная по формуле (10.144)

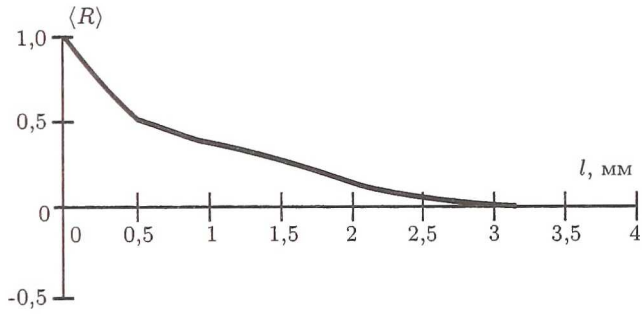


Рис. 10.62. Зависимость параметра $\langle R \rangle$ от значения сдвига исходного изображения

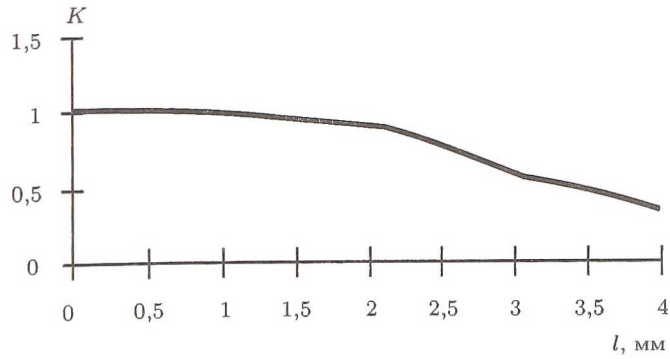


Рис. 10.63. Зависимость надежности распознавания K от значения сдвига исходного изображения, полученная экспериментально

где N — общее количество анализируемых отпечатков, N_r — количество отпечатков правильно идентифицированных.

Для эксперимента с большим количеством дактилограмм величины K и P должны совпадать. Из рисунков 10.61 и 10.63 видно, что для эксперимента со 100 дактилограммами значения K и P совпадают с погрешностью (5%). Так как при уменьшении количества векторов в эксперименте не произошло существенного изменения значений $\langle r \rangle$ и D, то можно будет использовать формулу (10.144) для приблизительного расчета надежности метода распознавания для большого количества отпечатков (> 100), используя при этом данные ($\langle r \rangle$, D), полученные в эксперименте с небольшим количеством базовых отпечатков.

Аналогичный эксперимент был проведен для исследования устойчивости метода идентификации к повороту дактилограммы. Был использован тот же набор из 100 дактилограмм. Дактилограмма последовательно поворачивалась с интервалом 5° в диапазоне от 0° до 20°. График зависимости среднего расстояния между повернутыми и не повернутыми векторами от значения поворота показан на рис. 10.64. Экспериментально была определена надежность распознавания в зависимости от поворота. Эта зависимость показана на рис. 10.65, из которого видно что для достижения высокой надежности распознавания ($> 0{,}9$) необходимо, чтобы поворот дактилограммы не превышал 10°.

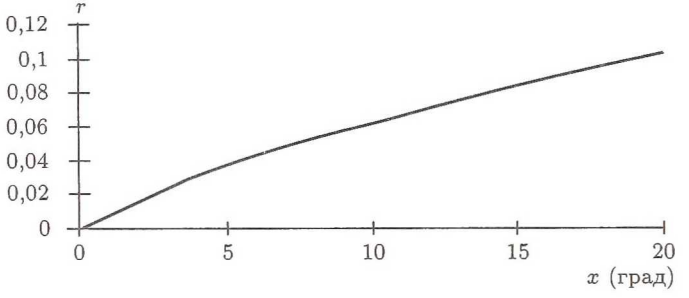


Рис. 10.64. Усредненный по 100 векторам признаков график зависимости расстояния между векторами признаков повернутого и не повернутого отпечатков пальца от значения поворота

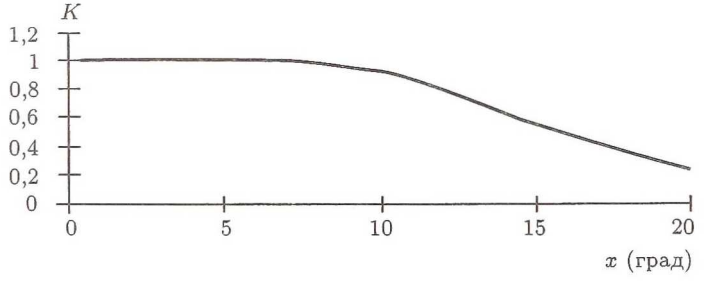


Рис. 10.65. Полученный экспериментально график зависимости надежности распознавания K от значения поворота исходного изображения

Суммируя все вышесказанное можно утверждать, что оптически построенные поля направлений могут быть эффективно использованы в системах идентификации дактилограмм, которые предназначены для работы с относительно небольшим набором базовых дактилограмм (10^3–10^4).

10.3.2. Оптическая расшифровка интерферограмм.

Как правило, интерферограмма обрабатывается с помощью компьютера методом фурье-преобразования [84] или методом выделения центров полос [50]. Основной недостаток этих методов в том, что с увеличением уровня шума в обрабатываемом изображении и при большом числе разрывов полос из-за спекл-эффекта они теряют надежность и могут приводить к большим искажениям восстановленной фазы.

Существуют итеративные методы, позволяющие с помощью процедуры регуляризации добиться устойчивого восстановления фазы по зашумленной интерферограмме с отношением сигнал/шум не меньше единицы. Однако для использования таких методов требуются значительные затраты машинного времени.

Оптические методы восстановления фазы световых полей основаны на использовании датчиков волнового фронта. В этом случае не требуется формирования интерферограммы, и датчик непосредственно размещается в область анализируемого светового поля. Самым распространенным анализатором волнового фронта является датчик Гартмана–Шаке [52].

В работе [87] рассмотрен гибридный оптико-цифровой метод обработки интерферограмм, основанный на оптическом построении поля направлений и поля частот. Значения двух соответствующих отсчетов в поле частот (ПЧ) и поле направлений (ПН), аналогично датчику Гартмана–Шаке, задают координаты вектора наклона плоскости касательной в данной точке к искомой функции фазы.

Пусть интерферограмма описывается функцией вида

$$I(x,y) = \left|\exp\left[i\Phi(x,y)\right] + \exp\left[i\left(\alpha_0 x + \beta_0 y\right)\right]\right|^2, \tag{10.146}$$

где $\Phi(x,y)$ — искомая фаза светового поля, (α_0, β_0) — вектор несущей частоты опорной плоской волны. Рассмотрим два скалярных поля $\hat{\varphi}(x,y)$ и $\omega(x,y)$ определенных по следующим формулам:

$$\operatorname{tg}\hat{\varphi}(x,y) = -\frac{\partial I/\partial x}{\partial I/\partial y}, \quad 0 \leqslant \hat{\varphi}(x,y) < \pi, \tag{10.147}$$

$$\omega(x,y) = \left[\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2\right]^{1/2}. \tag{10.148}$$

Функции $\hat{\varphi}(x,y)$ и $\omega(x,y)$ называются, соответственно, полем направлений и полем частот.

Из ур. (10.147) следует, что ПН определено с точностью до π. Из сравнения (10.146), (10.147) и (10.148) также следует связь ПН и ПЧ с функцией фазы $\Phi(x,y)$:

$$\operatorname{tg}\hat{\varphi}(x,y) = -\frac{\partial\Phi/\partial x - \alpha_0}{\partial\Phi/\partial y - \beta_0}, \tag{10.149}$$

$$\omega(x,y) = A_0(x,y)\left[\left(\frac{\partial\Phi}{\partial x} - \alpha_0\right)^2 + \left(\frac{\partial\Phi}{\partial y} - \beta_0\right)^2\right]^{1/2}, \tag{10.150}$$

где

$$A_0(x,y) = 2\left|\sin\left[\Phi(x,y) - \alpha_0 x - \beta_0 y\right]\right|. \tag{10.151}$$

Будем называть нормированным полем частот (НПЧ) функцию $\omega'(x,y)$ вида

$$\omega'(x,y) = \frac{\omega(x,y)}{A_0(x,y)}. \tag{10.152}$$

Можно утверждать, что по функциям $\hat{\varphi}(x,y)$ и $\omega'(x,y)$ можно восстановить фазу светового поля $\Phi(x,y)$. Однако это восстановление будет неоднозначным, т. к. неизвестен знак $\partial\Phi/\partial x$ (или $\partial\Phi/\partial y$). Если дополнительно потребовать, чтобы пространственные частоты опорной волны были выбраны достаточно большими, чтобы выполнялось условие

$$\left|\frac{\partial\Phi}{\partial x}\right|_{\max} < \alpha_0, \quad \left|\frac{\partial\Phi}{\partial y}\right|_{\max} < \beta_0, \tag{10.153}$$

то фазу можно восстановить однозначно по ПН и НПЧ. Чтобы однозначно восстановить фазу $\Phi(x,y)$ при любых значениях α_0 и β_0, в том числе и при $\alpha_0 = \beta_0 = 0$, необходимо привлекать дополнительную априорную информацию о фазе.

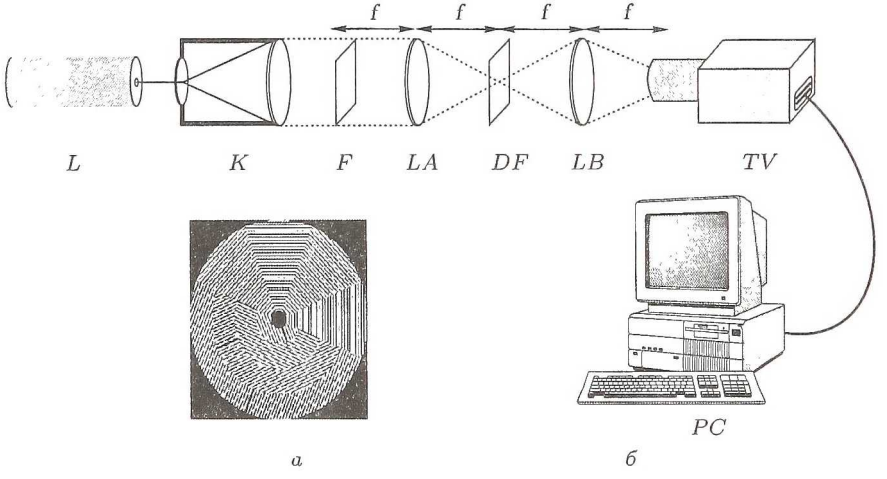


Рис. 10.66. Схема оптической установки для получения полей направлений и полей частот от интерферограмм (*а*) и амплитудный пространственный фильтр (*б*)

Пусть интерферограмма F выполнена в виде прозрачного транспаранта и помещена на вход когерентного фурье-коррелятора (рис. 10.66), в частотной плоскости которого расположен фазовый бинарный пространственный фильтр DF. Функция пропускания фильтра имеет вид

$$H_{mn}^{(S)(C)}(\xi,\eta) = \begin{cases} S_n(\hat{\varphi})\, R_n(\xi,\eta), & \hat{\varphi}_{n-1} < \hat{\varphi} < \hat{\varphi}_n,\ n = \overline{1,N} \\ C_m(r)\, R_m(\xi,\eta), & r_{m-1} < r < r_m,\ m = \overline{1,M}, \end{cases} \tag{10.154}$$

где

$$S_n(\hat{\varphi}) = \frac{1}{2}\left[\operatorname{sgn}(\hat{\varphi} - \hat{\varphi}_{n-1}) - \operatorname{sgn}(\hat{\varphi} - \hat{\varphi}_n)\right], \tag{10.155}$$

$$C_m(r) = \operatorname{circ}\left(\frac{r}{r_{m+1}}\right) - \operatorname{circ}\left(\frac{r}{r_m}\right), \tag{10.156}$$

$$R_n(\xi,\eta) = \operatorname{sgn}\left[\cos\left(\alpha_n\xi + \beta_n\eta\right)\right], \tag{10.157}$$

$S_n(\hat{\varphi})$, $C_m(r)$ — секторные и кольцевые участки фильтра, в которых остается неизменной бинарная фазовая дифракционная решетка $R_n(\xi,\eta)$, (α_n,β_n) — вектора несущей пространственной частоты решетки, задающей координаты центра частичного изображения на выходе коррелятора.

На выходе коррелятора сформируются эффективно $2(M+N)$ изображений. Это следует из того, что фазовая бинарная решетка со скачком фазы, равным π, формирует с эффективностью 80% всего два порядка дифракции. Бинарная решетка выбрана вместо дифракционных клиньев, как более простая при изготовлении. И если пространственные частоты (α_n,β_n) подобраны определенным образом, то N секторов и M колец сформируют как раз $2N+2M$ парциальных изображений, интенсивности которых описываются выражениями

$$I_m^{(c)}(u,v) = \left|\mathcal{F}\left\{C_m\right\} * I\right|^2, \tag{10.158}$$

$$I_n^{(s)}(u,v) = \left|\mathcal{F}\left\{S_n\right\} * I\right|^2, \tag{10.159}$$

где $\mathcal{F}$ — операция преобразования Фурье, выполняемая линзой коррелятора. В дальнейшем будет обрабатываться только половина из всех парциальных изображений. Парциальное изображение $I_n^{(s)}(u,v)$ имеет узкий радиально-щелевой спектр, т. к. сформировано с помощью фильтрации исходного изображения секторным фильтром (10.155). В это изображение вносят вклад те области исходного изображения $I(x,y)$, на которых полосы интерферограммы наклонены под углом $\hat{\varphi}+\pi/2$ из диапазона углов: $\hat{\varphi}_{n-1} < \hat{\varphi} < \hat{\varphi}_n$. Аналогично этому, частичное изображение $I_m^{(c)}(u,v)$ имеет узкий кольцевой спектр, т. к. сформировано с помощью фильтрации исходного изображения кольцевым фильтром (10.156). В него вносят вклад те точки интерферограммы $I(x,y)$, пространственная частота полос в окрестности которых равна $2\pi r/\lambda f$, где $r_{m-1} < r < r_m$.

Разнесенные пространственно дифракционными решетками и непересекающиеся на выходе коррелятора парциальные изображения регистрируется телекамерой и, в виде массива цифровых данных, вводятся в компьютер. Центры всех парциальных изображений опять сводятся в одну точку.

При выборе N одинаковых секторов (10.155) фильтра (10.154) центральные углы будут равны

$$\overline{\hat{\varphi}_n} = \frac{\hat{\varphi}_{n-1}+\hat{\varphi}_n}{2} = \frac{\pi n}{N}.$$

а также при M одинаковых кольцах (10.156) фильтра (10.154) центральные радиусы колец будут равны

$$\overline{r_m} = \frac{r_{m-1}+r_m}{2} = \frac{Rm}{M},$$

R — радиус апертуры фильтра.

Алгоритм аппроксимации аналогичен преобразованию Хоу–Радона [87], используемому в томографии и заключается в следующем. Точке выходного изображения с координатами (u,v) присваивается значение того угла $\overline{\hat{\varphi}_n}$, величина интенсивности частичного изображения которого в этой точке больше $I_n^{(s)}(u,v)$. Зависимость углов

$\overline{\hat{\varphi}_n}$ от координат (u, v) и будет аппроксимацией ПН:

$$\hat{\varphi}(u,v) = \sum_{n=1}^{N} \overline{\hat{\varphi}_n} \theta \left[I_n^{(c)}(u,v) - \overline{I^{(s)}} \right], \tag{10.160}$$

где

$$\overline{I^{(s)}}(u,v) = \max_{n \in [1,N]} \left\{ I_n^{(s)}(u,v) \right\}, \tag{10.161}$$

$\theta(x)$ — функция Хевисайда: $\theta(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0. \end{cases}$

Аналогично определяется аппроксимация НПЧ:

$$\omega'(u,v) = \sum_{m=1}^{M} \overline{r_m} \theta \left[I_m^{(c)}(u,v) - \overline{I^{(c)}}(u,v) \right], \tag{10.162}$$

где

$$\overline{I^{(c)}}(u,v) = \max_{m \in [1,M]} \left\{ I_m^{(c)}(u,v) \right\}. \tag{10.163}$$

Продолжая аналогию с томографией, можно заметить, что парциальные изображения являются как бы сечениями интерферограммы в пространстве направлений и частот. После вычисления ПН и НПЧ, для каждой точки интерферограммы с координатами (u, v) имеются два числа (φ и ω), которые с точностью до константы определяют вектор наклона (направление и величину наклона) плоскости касательной в данной точке к искомой функции фазы:

$$\Phi_0(x,y) = \omega'(u,v) \sin\left[\hat{\varphi}(u,v)\right](x-u) + \omega'(u,v) \cos\left[\hat{\varphi}(u,v)\right](y-v) + C. \tag{10.164}$$

На практике интерферограмма разбивается на k квадратных участков, в каждом из которых находится среднее ПН $\overline{\hat{\varphi}_{ij}}$; $i, j = \overline{1,k}$ и среднее НПЧ $\overline{\omega'_{ij}}$.

Тогда искомая функция фазы аппроксимируется набором плоскостей (10.164) в каждом квадрате средними $\overline{\hat{\varphi}_{ij}}$ и $\overline{\omega'_{ij}}$:

$$\Phi_{ij}(x,y) = \overline{\omega'_{ij}} \sin\left[\overline{\hat{\varphi}_{ij}}\right] x + \overline{\omega'}_{ij} \cos\left[\overline{\hat{\varphi}_{ij}}\right] y + C_{ij}, \quad i,j = \overline{1,k}. \tag{10.165}$$

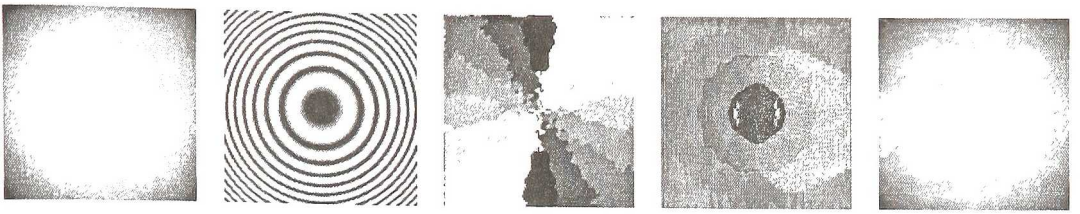
Постоянные C_{ij} выбираются из условия непрерывности функции (10.165) на границах соседних квадратов.

Для сравнения был проведен численный эксперимент по реконструкции фазы по данным измерения интенсивности в датчике Шаке–Гартмана (рис. 10.69).

Пример 10.8. На рис. 10.67*а* показана исходная фаза

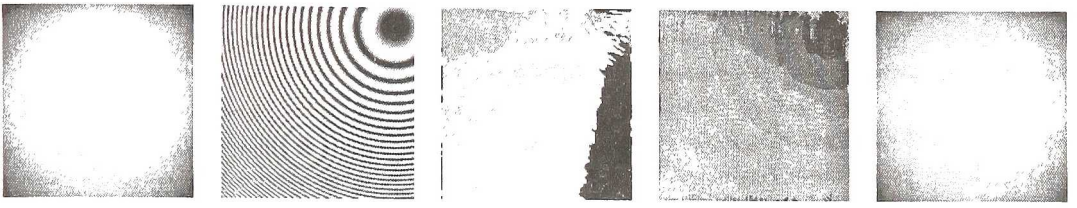
$$\Phi(x,y) = x^2 + y^2. \tag{10.166}$$

На рис. 10.67*б* показана интерферограмма, полученная в результате интерференции световой волны с фазой (10.166) и плоской световой волны, падающей перпендикулярно плоскости регистрации. Размер области регистрации 2 мм. Затем методом, описанным выше, были получены поле направлений интерферограммы, показанное на рис. 10.67*в*, и поле частот, рис. 10.67*г*. На последней стадии эксперимента из полученных ПН и ПЧ восстанавливается фаза, рис. 10.67*д*.



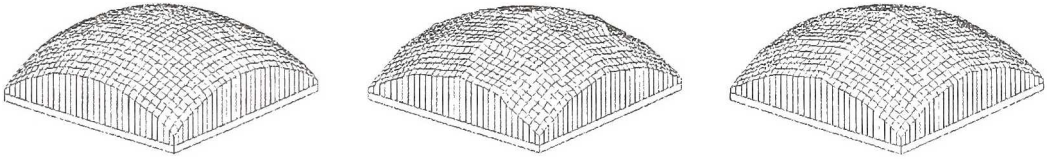
а *б* *в* *г* *д*

Рис. 10.67. Исходная фаза (*а*), интерферограмма (*б*), поле направлений (*в*), поле частот (*г*), восстановленная фаза (*д*)



а *б* *в* *г* *д*

Рис. 10.68. Исходная фаза (*а*), интерферограмма (*б*), поле направлений (*в*), поле частот (*г*), восстановленная фаза (*д*)



а *б* *в*

Рис. 10.69. Исходная фаза (*а*), фаза восстановленная по полю направлений и полю частот интерферограммы (*б*) и фаза восстановленная методом Гартмана–Шаке (*в*)

На рис. 10.68*б* показана интерферограмма, полученная в результате интерференции световой волны с фазой (10.166) и плоской световой волны, падающей наклонно к плоскости регистрации. Затем из интерферограммы были получены ее ПН, рис. 10.68*в*, и ПЧ, рис. 10.68*г*. Восстановленная фаза показана на рис. 10.68*д*.

На рис. 10.69 в трехмерном виде представлена исходная фаза (*а*), восстановленная фаза из интерферограммы по методу ПН, и для сравнения приведена фаза восстановленная по гартманограмме (*в*).

Для определения ошибки восстановления фазы использовалась величина

$$\Delta^2 = \frac{\int\limits_0^a \int\limits_0^a |\Phi(x,y) - \Phi_r(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y}{\int\limits_0^a \int\limits_0^a |\Phi(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y}, \tag{10.167}$$

где $\Phi(x,y) \in [0, 2\pi]$ — исходная фаза, $\Phi_r(x,y) \in [0, 2\pi]$ — восстановленная фаза, a — размер изображения.

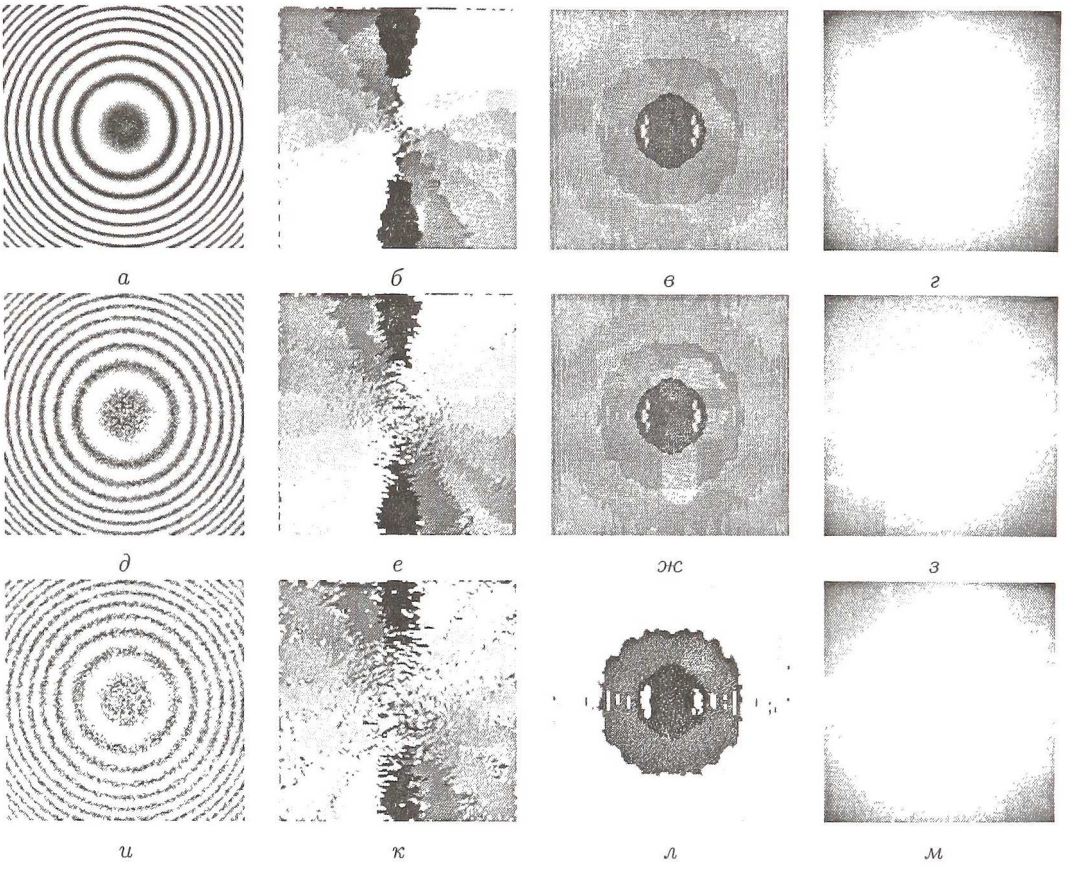


Рис. 10.70. Зашумленные интерферограммы: $S = 10$ (*а*), $S = 2$ (*д*), $S = 1$ (*к*), и соответствующие им поля направлений (*б, е, л*), поля частот (*в, ж, м*) и восстановленные фазы (*г, з, н*). $S = 10$, $\Delta = 0{,}047$ (верхняя строчка); $S = 2$, $\Delta = 0{,}061$ (средняя строчка); $S = 1$, $\Delta = 0{,}147$ (нижняя строчка)

При восстановлении фазы (10.166) методом ПН, $\Delta = 0{,}045$, при восстановлении этой же фазы из гартманограммы $\Delta = 0{,}039$. При восстановлении фазы (10.166) методом ПН из интерферограммы полученной с наклонной опорной волной, $\Delta = 0{,}041$, при восстановлении фазы методом ПН из светового поля $\Delta = 0{,}041$, т. е. погрешность не изменилась. Как видно из этих примеров точность метода близка к точности метода Гартмана, но рассмотренный здесь метод позволяет с хорошей точностью восстанавливать фазу, используя сильно зашумленные интерферограммы.

На рис. 10.70 показаны зашумленные интерферограммы с разными значениями отношения сигнал/шум S, поля направлений, поля частот и восстановленная фаза.

Как видно из рис. 10.70, наибольшие изменения, которые появляются в поле направлений и в поле частот в результате зашумления, находятся там, где период полос интерферограммы больше. Эти участки и вносят основной вклад в увеличение погрешности. Если интерферограмма не содержит участков с большим периодом полос (наклонное падение опорной волны), то погрешность метода может быть уменьшена.

На рисунке 10.71 показана зависимость ошибки восстановления фазы Δ от отношения сигнал/шум S для интерферограммы, полученной в результате интерференции световой волны с фазой (10.166) и плоской световой волны, падающей пер-

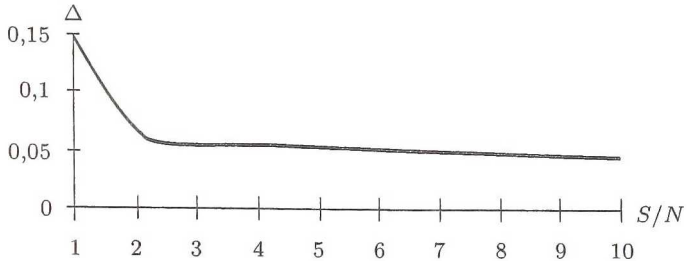


Рис. 10.71. Зависимость Δ от отношения сигнал/шум S, при восстановлении фазы по интерферограмме, полученной в результате интерференции световой волны с фазой (10.166) и плоской световой волны, падающей перпендикулярно плоскости регистрации

пендикулярно плоскости регистрации. Как видно из рис. 10.71 величина ошибки восстановления фазы при $S > 2$ практически не изменяется. При приближении S к единице значение ошибки резко возрастает. Был проведен эксперимент с набором из 10 интерферограмм, каждая из которых была получена в результате интерференции световой волны с фазой (10.166) и плоской световой волны, падающей перпендикулярно плоскости регистрации, с изменением геометрического размера области регистрации от 2 мм до 3 мм.

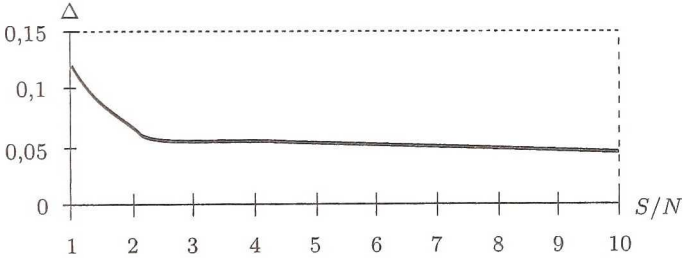


Рис. 10.72. Зависимость Δ от отношения сигнал/шум S/N для набора интерферограмм

Как видно из рис. 10.72, характер зависимости не меняется для набора интерферограмм. В среднем величина ошибки для группы интерферограмм даже несколько уменьшилась.

10.4. Оптическое выполнение преобразования Хоу–Радона

Определять направления линий на изображении можно, не только ПН, но и осуществляя преобразование Хоу–Радона от изображения. Оптико-цифровое устройство для выполнения преобразования Хоу–Радона (ПХР) [89], основано на реализации свойства ПХР переводить неосевую точку в сдвинутую синусоиду. Это устройство содержит матрицу $N \times N$ микро-голограмм, каждая из которых формирует, при освещении плоским пучком, участок кривой равной одному периоду синусоиды, сдвинутой от центра координат пропорционально смещению микро-голограммы

от центра матрицы. В результате, если осветить такую матрицу отрезком прямой, то на выходе сформируется яркая точка, как результат пересечения многих синусоид.

Недостатком такой реализации ПХР является трудоемкость расчета и изготовления матрицы $N\times N$ голограмм, каждая из которых имеет размерность $M\times M$ и рассчитывается за число итераций L. Чтобы рассчитать всю матрицу нужно выполнить $L\times N\times N$ итераций, каждая из которых включает выполнение двух преобразований Фурье размерности $M\times M$.

В [90] рассматривается модифицированная оптическая реализация ПХР. Такое преобразование осуществляется с помощью N пар скрещенных цилиндрических линз. Одна линза из пары формирует изображение линии в поперечном направлении, а вторая — формирует пространственный спектр в продольном направлении линии (собирает линию в точку). Набор из N пар нужен, чтобы осуществить фокусировку прямых линий с разной ориентацией во входной плоскости. Недостатком этого способа является то, что каждая прямая линия будет переводиться в точку с некоторой эффективностью и другими $(N-1)$ парами линз. Это приводит к уменьшению отношения сигнал/шум всего устройства.

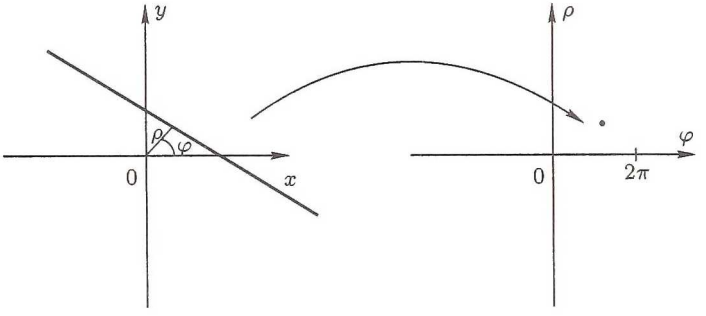


Рис. 10.73. Преобразование Хоу–Радона

В [91] был предложен более простой способ оптического выполнения ПХР. Он основан на свойстве преобразования переводить прямую линию в точку (рис. 10.73). В этом случае устройство содержит коррелятор с одним входом и с N выходами. В задней фокальной плоскости сферической линзы коррелятора расположен фазовый пространственный фильтр, состоящий из N угловых секторов с углом при вершине каждого $\Delta\hat{\varphi}=2\pi/N$.

Каждый такой угловой сектор имеет функцию пропускания цилиндрической линзы

$$\tau(x,y)=\exp\left[-\frac{ik}{2f}\left(x\cos\hat{\varphi}_n-y\sin\hat{\varphi}_n-p_n\right)^2\right], \tag{10.168}$$

где

$$(x,y)\in\left\{\varphi_n-\frac{\Delta\varphi}{2}<\operatorname{arctg}\left(\frac{y}{x}\right)\leqslant\varphi_n+\frac{\Delta\varphi}{2}\right\},\quad \varphi_n=n\Delta\varphi,\quad n=0,1,\ldots,N-1.$$

На фокусном расстоянии f от фильтра расположен фотоприемник, регистрирующий излучение только вдоль лучей, исходящих из центра под соответствующими углами. Оптическая схема устройства показана на рис. 10.74, на котором изображены коллимированный пучок, освещающий объект-изображение, сферическая линза,

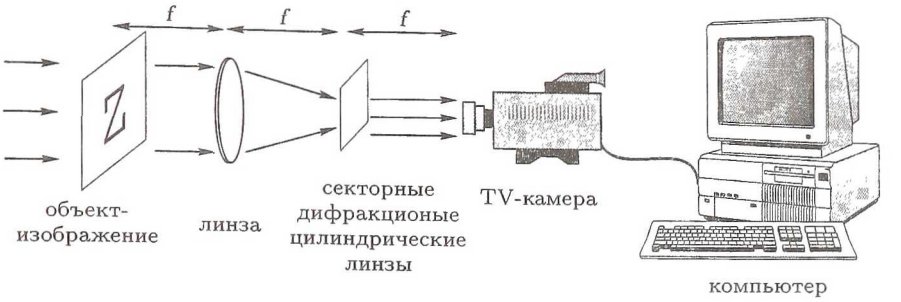


Рис. 10.74. Оптическая схема для реализации преобразования Хоу–Радона

секторный пространственный фильтр с цилиндрическими линзами, фотоприемное устройство и компьютер.

Работа устройства основана на реализации интегральной формы, связывающей ПХР и Фурье. Преобразование Хоу–Радона от функции $f(x,y)$ имеет вид

$$R(\rho,\varphi)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)\,\delta(\rho-x\cos\hat{\varphi}-y\sin\hat{\varphi})\,\mathrm{d}x\,\mathrm{d}y, \tag{10.169}$$

где δ — дельта функция Дирака, ρ — величина перпендикуляра, проведенного из центра координат к текущей прямой линии, $\hat{\varphi}$ — угол между перпендикуляром к текущей прямой линии и осью OX (см. рис. 10.73). Нетрудно показать, что существует уравнение связи между двумя интегральными преобразованиями:

$$R(\rho,\hat{\varphi})=\int\limits_{0}^{\infty} F(t,\hat{\varphi})\,e^{2\pi i\rho t}\,\mathrm{d}t, \tag{10.170}$$

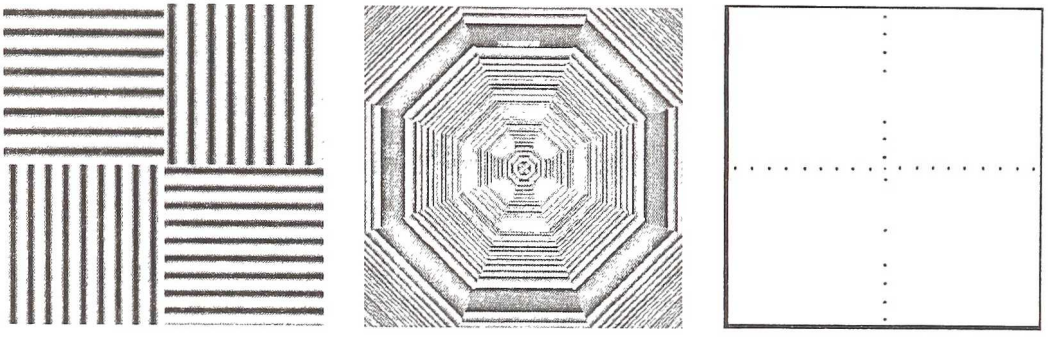
где

$$F(t,\hat{\varphi})=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x,y)\exp\left[-2\pi i\,(xt\cos\hat{\varphi}+yt\sin\hat{\varphi})\right]\,\mathrm{d}x\,\mathrm{d}y. \tag{10.171}$$

Из уравнений (10.170) и (10.171) видно, что ПХР является результатом выполнения одномерного преобразования Фурье по радиальной координате от двумерного фурье-образа исходной функции.

Устройство, показанное на рис. 10.74, реализует уравнения (10.170) и (10.171) приближенно с дискретностью по координате $\hat{\varphi}$ равной $\Delta\hat{\varphi}=2\pi/N$, и по координате ρ с дискретностью $\Delta\rho$ равной разрешению фотоприемника.

На рис. 10.75 показан пример численного моделирования работы оптической схемы, рис. 10.74. Тест-объект (рис. 10.75*а*) состоит из набора прямых отрезков, расположенных под углами 0 и $\pi/2$ к осям координат. Амплитудная маска размерностью 512×512 пространственного фильтра, который состоит из 8 цилиндрических линз, расположенных в угловых секторах, показана на рис. 10.75*б*. На рис. 10.75*в* показан результат, появляющийся на выходе оптической установки: набор светлых пятен, расположенных на осях OX и OY, число которых совпадает с числом отрезков прямых на входном изображении.



а *б* *в*

Рис. 10.75. Тестовое изображение (*а*), фазовый пространственный фильтр(*б*), хоу–радон-образ изображения (*в*)



а *б*

Рис. 10.76. Изображение отпечатка пальца (*а*) и его хоу–радон-образ (*б*)

На рис. 10.76 показаны изображение отпечатка пальца (*а*) и его хоу–радон-образ (*б*), полученный с помощью численного моделирования того же устройства. Каждой линии исходного изображения соответствует точка в хоу–радон-образе. Происходит сжатие информации аналогичное тому, которое имеет место при построении поля направлений.

10.5. Заключение

Существует ряд актуальных задач обработки изображений, таких как выделение признаков при распознавании образов, которые наиболее оперативно решаются гибридными оптико-цифровыми методами. В данной главе разработаны оптико-цифровой метод построения поля направлений, и его возможные применения для идентификации дактилограмм и реконструкции фазы светового поля по интерферограмме. В многочисленных вычислительных и натурных экспериментах исследованы свойства этого метода, и доказана эффективность его применения в указанных выше задачах.

Метод позволяет существенно сократить затраты машинного времени для получения поля направлений, за счет оптической реализации наиболее емкой по количеству вычислений стадии работы. Оптическое вычисление поля направлений реализуется за время, которое необходимо для их ввода в компьютер с помощью

телекамеры (около 0,1 с). На эту же операцию, при цифровой обработке, современному компьютеру требуется время около 100 с, что существенно замедляет обработку. Известная проблема метода оптических вычислений здесь решается за счет сжатия информации: при вычислении поля направлений на изображении исчезает высокая пространственная частота, играющая роль несущей, наличие которой характерно для дактилограмм и интерферограмм.

Так как количество дактилограмм в системах персонального доступа и мировых базах данных постоянно увеличивается, то очевидно, что цифровые методы анализа дактилограмм могут оказаться недостаточно быстрыми для практического использования. В этом случае гибридные оптико-цифровые методы, основанные на построении поля направлений, дают эффективное решение проблемы.

Приложение П1 к разделу 10.2.2

Решение интегрального уравнения на собственные значения с экспоненциально-косинусным ядром. Рассмотрим одномерное интегральное уравнение на собственные значения на симметричном интервале $x \in [-A, A]$:

$$\int_{-A}^{A} R(x-u)\psi(u)\,\mathrm{d}u = \lambda\psi(x), \tag{10.172}$$

с экспоненциально-косинусным ядром

$$R(\tau) = \sigma^2 \exp\left(-\alpha\,|\tau|\right)\cos\beta\tau, \tag{10.173}$$

имеющим спектр

$$G(\omega) = \frac{2\alpha\sigma^2\left(\alpha^2+\beta^2+\omega^2\right)}{\omega^4 + 2\omega^2\left(\alpha^2-\beta^2\right) + \left(\alpha^2+\beta^2\right)^2}, \tag{10.174}$$

где ω — пространственная частота.

В соответствии с алгоритмом решения интегрального уравнения II-го рода для ядра с дробно-рациональным спектром [19], интегральное уравнение (10.172) сводится к линейному дифференциальному уравнению с постоянными коэффициентами:

$$\psi^{(4)}(x) - 2\left(\alpha^2-\beta^2-\frac{\alpha\sigma^2}{\lambda}\right)\psi^{(2)}(x) + \left(\alpha^2+\beta^2\right)\left(\alpha^2+\beta^2-\frac{2\alpha\sigma^2}{\lambda}\right)\psi(x) = 0. \tag{10.175}$$

Решение этого уравнения можно записать в виде

$$\psi(x) = b_1\exp\left(-\mathrm{E}\frac{x}{A}\right) + b_2\exp\left(\mathrm{E}\frac{x}{A}\right) + b_3\exp\left(-\Gamma\frac{x}{A}\right) + b_4\exp\left(\Gamma\frac{x}{A}\right), \tag{10.176}$$

где $b1 \ldots b4$ — постоянные коэффициенты, а E и Γ — корни соответствующего (10.176) характеристического уравнения, удовлетворяющие по теореме Виета условиям

$$\begin{aligned} \mathrm{E}^2+\Gamma^2 &= 2\left(\alpha^2-\beta^2-\frac{A\alpha\sigma^2}{\lambda}\right), \\ \mathrm{E}^2\Gamma^2 &= \left(\alpha^2+\beta^2\right)\left(\alpha^2+\beta^2-\frac{2A\alpha\sigma^2}{\lambda}\right). \end{aligned} \tag{10.177}$$

Переходя от комплексных уравнений (10.177) к вещественным, представив E и Γ в комплексном виде: $\mathrm{E} = \varepsilon_0 + i\varepsilon$, $\Gamma = \gamma + i\gamma_0$, получаем пять возможных случаев решения (10.176):

$$1:\ \begin{cases} \varepsilon_0 = 0, \quad \gamma = 0, \\ \varepsilon^2 + \gamma_0^2 = -2\left(\alpha^2 - \beta^2 - \dfrac{A\alpha\sigma^2}{\lambda}\right), \\ \varepsilon^2\gamma_0^2 = \left(\alpha^2 + \beta^2\right)\left(\alpha^2 + \beta^2 - \dfrac{2A\alpha\sigma^2}{\lambda}\right); \end{cases}$$

$$2:\ \begin{cases} \varepsilon = 0, \quad \gamma_0 = 0, \\ \varepsilon_0^2 + \gamma^2 = 2\left(\alpha^2 - \beta^2 - \dfrac{A\alpha\sigma^2}{\lambda}\right), \\ \varepsilon_0^2\gamma^2 = \left(\alpha^2 + \beta^2\right)\left(\alpha^2 + \beta^2 - \dfrac{2A\alpha\sigma^2}{\lambda}\right); \end{cases}$$

$$3:\ \begin{cases} \varepsilon_0 = 0, \quad \gamma_0 = 0, \\ \varepsilon^2 - \gamma^2 = 2\left(\alpha^2 - \beta^2 - \dfrac{A\alpha\sigma^2}{\lambda}\right), \\ \varepsilon^2\gamma^2 = -\left(\alpha^2 + \beta^2\right)\left(\alpha^2 + \beta^2 - \dfrac{2A\alpha\sigma^2}{\lambda}\right); \end{cases}$$

$$4:\ \begin{cases} \varepsilon = 0, \quad \gamma = 0, \\ \varepsilon_0^2 - \gamma_0^2 = -2\left(\alpha^2 - \beta^2 - \dfrac{A\alpha\sigma^2}{\lambda}\right), \\ \varepsilon_0^2\gamma_0^2 = -\left(\alpha^2 + \beta^2\right)\left(\alpha^2 + \beta^2 - \dfrac{2A\alpha\sigma^2}{\lambda}\right); \end{cases}$$

$$5:\ \begin{cases} \varepsilon_0 = \gamma, \quad \varepsilon = -\gamma_0, \\ \varepsilon_0^2 + \varepsilon^2 = \left(\alpha^2 + \beta^2\right)\left(\alpha^2 + \beta^2 - \dfrac{2A\alpha\sigma^2}{\lambda}\right), \\ \varepsilon_0^2 - \varepsilon^2 = \alpha^2 - \beta^2 - \dfrac{A\alpha\sigma^2}{\lambda}. \end{cases}$$

Все случаи, кроме 2 и 3, вступают в противоречие при $\lambda \to 0$. При этом случаи 2 и 3 эквивалентны. Примем случай 3.

Подставив (10.173) и (10.176) в интегральное уравнение (10.172) и приравнивая коэффициенты при одинаковых линейно-независимых функциях, получаем систему уравнений, из которой видно, что возможны два варианта:

$$\text{B.1: } b_1 = b_2,\ b_3 = b_4, \quad \text{B.2: } b_1 = -b_2,\ b_3 = -b_4,$$

каждому из которых соответствует система трансцендентных уравнений [28]:

$$\begin{cases} G_1(\gamma)G_2(i\varepsilon)\left\{\left(\dfrac{\alpha}{\varepsilon}F_1(i\varepsilon)\right)^P \pm \operatorname{tg}\varepsilon\right\}\left\{(F_2(\gamma))^P + \operatorname{th}\gamma\right\} - \\ -G_1(i\varepsilon)G_2(\gamma)\left\{\left(\dfrac{\alpha}{\gamma}F_1(\gamma)\right)^P - \operatorname{th}\gamma\right\}\left\{(F_2(i\varepsilon))^P \mp \operatorname{tg}\varepsilon\right\} = 0, \\ \gamma^2 = \left(\alpha^2+\beta^2\right)\dfrac{\alpha^2-3\beta^2+\varepsilon^2}{\alpha^2+\beta^2+\varepsilon^2}, \end{cases} \tag{10.178}$$

где $F_1(x) = \dfrac{\beta^2+\left(\alpha^2-x^2\right)}{\beta^2-\left(\alpha^2-x^2\right)}$, $\quad F_2(x) = \dfrac{\alpha^2+\beta^2+x^2}{2\alpha x}$,

для В.1: $G_1(x) = 1$, $G_2(x) = \beta^2 - \left(\alpha^2 - x^2\right)$, $P = 1$, и используется верхний знак;
для В.2: $G_1(x) = \alpha^2+\beta^2+x^2$, $G_2(x) = \alpha^2+\beta^2-x^2$, $P = -1$, и используется нижний знак.

Каждый из вариантов системы (10.178) имеет бесконечное множество решений, которые могут быть отделены на непересекающихся интервалах, соответствующих $\psi_k(x)$:

$$\begin{aligned} &\text{В.1: } \left(0,\frac{\pi}{2}\right),\ k=0; \quad \left(\frac{\pi}{2},\frac{3\pi}{2}\right),\dots,\left(\frac{\pi(k-1)}{2},\frac{\pi(k+1)}{2}\right), \quad k=2,4,6,\dots \\ &\text{В.2: } \left(\frac{\pi}{2},\frac{3\pi}{2}\right),\ \left(\frac{3\pi}{2},\frac{5\pi}{2}\right),\dots,\left(\frac{\pi k}{2},\frac{\pi(k+2)}{2}\right), \quad k=1,3,5,\dots \end{aligned} \tag{10.179}$$

Итак, решением интегрального уравнения для экспоненциально-косинусной корреляционной функции являются собственные функции $\psi_k(x)$ вида

$$\psi_k(x) = \begin{cases} d_{1k}\cos\left(\varepsilon_k\dfrac{x}{A}\right) + d_{2k}\operatorname{ch}\left(\gamma_k\dfrac{x}{A}\right),\ k=0,2,4,\dots \\ c_{1k}\sin\left(\varepsilon_k\dfrac{x}{A}\right) + c_{2k}\operatorname{sh}\left(\gamma_k\dfrac{x}{A}\right),\ k=1,3,5,\dots \end{cases} \tag{10.180}$$

где ε_k и γ_k с четными k являются решениями системы (10.178) для В.1, с нечетными k — решениями системы (10.178) для В.2, а коэффициенты d_{1k}, d_{2k} и c_{1k}, c_{2k} определяются из условия нормировки, соответственно, по формулам

$$\begin{aligned} K_1 &= \left\{A\left[\left(1\pm\frac{\sin 2\varepsilon}{2\varepsilon}\right) + 4D^2\left(1\pm\frac{sh2\gamma}{2\gamma}\right) - \frac{4DB}{\varepsilon^2+\gamma^2}\right]\right\}^{-1/2}, \\ K_2 &= -DK_1, \quad D = 2\frac{E_1\left(\alpha^2+\beta^2-\gamma^2\right)}{E_2\left(\alpha^2+\beta^2+\varepsilon^2\right)}, \end{aligned} \tag{10.181}$$

где В.1: $d_{1k} = K_1$, $d_{2k} = K_2$, $k = 0,2,4,\dots$;

$$\begin{aligned} B &= e^{\gamma}(\gamma\cos\varepsilon + \varepsilon\sin\varepsilon) - e^{-\gamma}(\gamma\cos\varepsilon - \varepsilon\sin\varepsilon), \\ E_1 &= \left(\alpha^2+\beta^2-\varepsilon^2\right)\cos\varepsilon + 2\alpha\varepsilon\cos\varepsilon, \\ E_2 &= \left(\alpha^2+\beta^2+\gamma^2\right)\left(e^{\gamma}+e^{-\gamma}\right) + 2\alpha\gamma\left(e^{\gamma}+e^{-\gamma}\right); \end{aligned}$$

В.2: $c_{1k} = K_1$, $c_{2k} = K_2$, $k = 1, 3, 5, \ldots$;

$$B = e^{\gamma}(\gamma \sin\varepsilon + \varepsilon\cos\varepsilon) - e^{-\gamma}(\gamma\sin\varepsilon - \varepsilon\cos\varepsilon),$$
$$E_1 = \left(\alpha^2 + \beta^2 - \varepsilon^2\right)\sin\varepsilon + 2\alpha\varepsilon\cos\varepsilon,$$
$$E_2 = \left(\alpha^2 + \beta^2 + \gamma^2\right)\left(e^{\gamma} - e^{-\gamma}\right) + 2\alpha\gamma\left(e^{\gamma} + e^{-\gamma}\right).$$

При этом собственные значения

$$\lambda_k = \frac{2A\alpha\sigma^2\left(\alpha^2 + \beta^2 + \varepsilon_k^2\right)}{\varepsilon_k^4 + 2\varepsilon_k^2\left(\alpha^2 - \beta^2\right) + \left(\alpha^2 + \beta^2\right)^2}. \tag{10.182}$$

При $\beta = 0$ приходим к случаю экспоненциальной корреляционной функции. При этом в системе (10.178) $\gamma^2 = \alpha^2$, то есть становится константой, и трансцендентная система превращается в уравнение. Для обоих вариантов (10.178) принимает вид

$$\begin{aligned}\left(\alpha^2 + \varepsilon_k^2\right)\left(\frac{\varepsilon_k}{\alpha} + \operatorname{tg}\varepsilon_k\right) = 0, \quad k = 0, 2, 4, \ldots,\\ \left(\alpha^2 + \varepsilon_k^2\right)\left(\frac{\alpha}{\varepsilon_k} - \operatorname{tg}\varepsilon_k\right) = 0, \quad k = 1, 3, 5, \ldots\end{aligned} \tag{10.183}$$

Нормированные коэффициенты описываются формулами

$$\begin{aligned}d_{1k} &= \frac{1}{\sqrt{A\left(1 + \operatorname{sinc} 2\varepsilon_k\right)}}, \quad D = 0, \quad d_{2k} = 0, \quad k = 0, 2, 4, \ldots,\\ c_{1k} &= \frac{1}{\sqrt{A\left(1 - \operatorname{sinc} 2\varepsilon_k\right)}}, \quad D = 0, \quad c_{2k} = 0, \quad k = 1, 3, 5, \ldots\end{aligned} \tag{10.184}$$

Собственные значения принимают вид

$$\lambda_k = \frac{2A\alpha\sigma^2}{\alpha^2 + \varepsilon_k^2}, \tag{10.185}$$

а базисные функции

$$\psi_k(x) = \begin{cases} d_{1k}\cos\left(\varepsilon_k\dfrac{x}{A}\right), & k = 0, 2, 4, \ldots\\ c_{1k}\sin\left(\varepsilon_k\dfrac{x}{A}\right), & k = 1, 3, 5, \ldots\end{cases} \tag{10.186}$$

Формулы (10.183)–(10.186) совпадают с ранее полученным в [20] решением. Таким образом при $\beta \neq 0$ получено обобщение на случай экспоненциально-косинусной корреляционной функции.

Как видно из системы (10.177), полученное решение имеет место при выполнении условия

$$\alpha^2 - 3\beta^2 + \varepsilon^2 > 0. \tag{10.187}$$

При достаточно больших ε, то есть для больших номеров базисных функций: $\varepsilon_k \xrightarrow[k\to\infty]{} \infty$, условие (10.187) будет всегда выполнено.

Приложение П2 к разделу 10.2.2.

Пусть ξ — случайная величина, принимающая значение 1, если некоторое событие, вероятность которого p, происходит, и значение 0, если это событие не происходит.

Пусть проведено N опытов и получена серия значений ξ: $\xi_1, \xi_2, \ldots, \xi_N$, каждое из которых равно либо 0, либо 1. В этом случае оценкой вероятности p будет служить величина [22]

$$\hat{p} = \frac{1}{N}\sum_{n=1}^{N}\xi_i. \tag{10.188}$$

Оценка (10.188) является несмещенной и ее точность характеризуется шириной интервала, в который она попадает с заданной доверительной вероятностью γ:

$$P\left\{|\hat{p}-p| \leqslant C_\gamma\sqrt{\frac{p(1-p)}{N}}\right\} = 1-\gamma, \tag{10.189}$$

где C_γ — коэффициент, определяемый по заданной γ с помощью таблиц математической статистики (обычно полагается $C_\gamma = 2$ или $C_\gamma = 3$, что при $N \gg 1$ соответствует доверительной вероятности $\gamma \approx 0{,}95$ и $\gamma \approx 0{,}997$).

Из формулы (10.189) следует выражение для относительной ошибки оценивания:

$$\mu_\gamma = \frac{|\hat{p}-p|}{p} = C_\gamma\sqrt{\frac{1-p}{pN}}. \tag{10.190}$$

Литература к главе 10

1. Сойфер В. А. *Цифровая голография и ее применение* (Куйбышев: КуАИ, 1978)
2. Lee W. H., Greer M. O. *Appl. Opt.* **13**(4) 925 (1974)
3. Stroke G. W., Haliowa M., Srinivasar V. *Phys. Letters.* **A51**(7), 383 (1975)
4. Сойфер В. А. *Моделирование процесса голографирования в случайной среде. Моделирование многолучевых радиоканалов для анализа и синтеза систем передачи информации* (М.: Наука, 1978)
5. Веряскин Ф. Ф., Выдрин Д. В., Давыдов В. Г., Мантуш Т. Н., Нежевенко Е. С., Панков В. Н., Твердохлеб П. Н. *Автометрия* (3), 73 (1975)
6. Голуб М. А., Сойфер В. А. *Тез. докл. IV Междунар. симпоз. по теории информации* (Ч. I. — М.-Л., 31, 1976).
7. Сойфер В. А. *Матер. IX Всесоюзной школы по автоматизации научных исследований* (Л.: ЛИЯФ, 350 1977)
8. Сойфер В. А. *Матер. IX Всесоюзной школы по голографии и когерентной оптике* (Л.: ЛИЯФ, 199, 1977)
9. Сойфер В. А. *Автореферат диссерт. на соиск. ученой степени д. т. н.* (Л.: ЛЭТИ, 32 1984)
10. Голуб М. А., Сойфер В. А. *Труды МФТИ. Серия: радиотехника и электроника* (Долгопрудный, 120, 1977)
11. Ту Дж, Гонсалес Р. *Принципы распознавания образов* (М.: Мир, 1978)
12. Фор А. *Восприятие и распознавание образов* (М.: Машиностроение, 1989)
13. Фу К. С. *Последовательные методы в распознавании образов и обучении машин* (М.: Наука, 1971)
14. Голуб М. А. *Диссертация на соискание ученой степени д. ф.-м. н.* (Куйбышев, 1990)

15. Ватанабе С. *Автоматический анализ сложных изображений* (М.: Мир, 254, 1969)
16. Давенпорт Б. В., Рут В. Л. *Введение в теорию случайных сигналов и шумов* (М.: Изд-во иностр. лит., 1960)
17. Karhunen K. *Ann. Acad. Sci. Fennicae. Helsinki* 1947, *Ser. A137*
18. Кловский Д. Д., Сойфер В. А. *Обработка пространственно-временных сигналов* (М.: Связь, 1976)
19. Youla D. C., *IRE Trans. on Inf. Theory* **IT-3**(3) 187 (1957)
20. Van Trees H. L. *Detection, Estimation, and Modulation Theory. Part 2. Nonlinear Modulation Theory* (New York, 1971)
21. Быков В. В. *Цифровое моделирование в статистической радиотехнике* (Изд-во „Советское радио", 1971)
22. Ермаков С. М., Михайлов Г. А. *Статистическое моделирование* (М.: Наука, 1982)
23. Голуб М. А., Сойфер В. А. *Тезисы докладов III всесоюзной конференции: Математические методы распознавания образов* (Львов, ФМИ АН УССР, 141, 1987)
24. Soifer V. A., Khonina S. N. *Pattern Recognition and Image Analysis* **4**(2) 137 (1994)
25. Голуб М. А., Сойфер В. А. *Спектральные методы обработки информации в научных исследованиях: Труды Всесоюзного семинара* (Пущино, 108, 1980)
26. Soifer V. A., Golub M. A., Khonina S. N. *Pattern Recognition and Image Analysis* **3**(3) 289 (1993)
27. Сойфер В. А. *Введение в дифракционную микрооптику* (Самара, 1996)
28. Голуб М. А., Хонина С. Н. *Компьютерная оптика* вып. 13, 49 (1993)
29. Свешников А. А. *Прикладные методы теории случайных функций* (М.: Наука, 1968)
30. Сойфер В. А. *Теория информации* (Куйбышев: КуАИ, 1975)
31. Бендат Дж., Пирсол А. *Прикладной анализ случайных данных* (М.: Мир, 1989)
32. Ченцова О. Б., Тедор И. Л., Прокофьева Г. Л., Маркушева Л. И., Прошина О. И. *Кристаллографический метод обследования при некоторых заболеваниях глаз. Методические рекомендации* (М., 1988)
33. Чухман—Дворянова Т. П. *VI съезд офтальмологов России. Тезисы докладов* (М., 395, 1994).
34. Khonina S. N., Kotlyar V. V., Soifer V. A., Dvoryanova T. P. *Proceedings of SPIE* **2363** 249 (1994)
35. Soifer V. A., Kotlyar V V., Khonina S. N., Skidanov R. V. *Proceedings of the 4-th Russian-German Workshop "Pattern Recognition and Image Analysis"* (Valday, 139, 1996)
36. Ярославский Л. П. *Введение в цифровую обработку изображений* (М.:Сов. радио, 1979)
37. Pratt W. K. *Digital image processing* (Wiley-Interscience publication, New York, 1978)
38. Gerchberg R. W., Saxton W. O. *Optik* **35** 237 (1972)
39. Березный А. Е., Прохоров А. М., Сисакян И. Н., Сойфер В. А. *Бессель-оптика* (ДАН СССР, **234**(4) 802, 1984)
40. Kotlyar V V., Khonina S. N., Soifer V. A. *J. Mod. Opt.* **42**(6) 1231 (1995)
41. Kotlyar V V., Khonina S. N., Soifer V. A. *Optik* **102**(2) 45 (1996)
42. Kotlyar V. V., Soifer V. A., Khonina S. N. *J. Mod. Opt.* **44**(7) 1403 (1997)
43. Schechner Y. Y., Piestun R., Shamir J. *Phys. Rev. E* **54**(1) 51 (1996)
44. Soifer V. A., Golub M. A. *Laser beam mode selection by computer-generated holograms* (CRC Press, Boca Raton, 1994)
45. Born M., Wolf E. *Principlies of Optics* (Pergamon Press, Oxford, 1968)
46. Vasara A., Turunen J., Friberg A. T. *J. Opt. Soc. Am. A* **6**(11) 1748 (1989)
47. Kotlyar V. V., Nikolsky I. V., Soifer V. A. *Optik* **98**(1) 26 (1994)
48. Soifer V. A., Kotlyar V. V., Doskolovich L. L. *Iterative methods for diffractive optical elements computation* (Taylor & Francis, London, 1997)
49. Kotlyar V. V., Khonina S. N., Soifer V. A. *Journal of Modern Optics* **45**(7) 1495 (1998)
50. Yatagai T., Nakadata S., Idesawa M., Saito H. *Opt. Eng.* **21**(2) 432 (1982)
51. Fienup J. R. *Appl. Opt.* **21**(15) 2758 (1982)
52. Artzner G. *Opt. Eng.* **31**(6) 1311 (1992)

53. Golub M. A., Prokhorov A. M., Sisakian I. N., Soifer V. A. *Sov. J. Quant. Electron.* **12**(9), 1208 (1982)
54. Котляр В. В. *Компьютерная оптика* (Москва, **5** 31, 1989)
55. Kotlyar V. V., Khonina S. N., Soifer V. A., Wang Y., Zhao D. *Компьютерная оптика* (Самара-Москва, вып. 17 43, 1997)
56. Виноградов И. М. *Математическая энциклопедия* (Москва: Советская энциклопедия, 1982, т. 3)
57. Soifer V. A., Kotlyar V. V., Khonina S. N., Khramov A. G., Skidanov R. V. *Proceeding of SPIE* **3346** 238 (1997)
58. Soifer V. A., Kotlyar V. V., Khonina S. N., Skidanov R. V. *The 13th Inter. Conf. on Pattern Recognition, Technical Univer.* (Vienna, 25-29 Aug. **2** 586, 1996)
59. Сойфер В. А., Котляр В. В., Хонина С. Н., Скиданов Р. В. *Компьютерная оптика* (вып. 16, 78, 1996)
60. Мардиа К. *Статистический анализ угловых наблюдений* (М.: Наука, 1978)
61. Novicov S. O., Kot V. S. *Proceeding of SPIE* **3346** 258 (1997)
62. Chen Z., Sun Y., Zhang Y., Mu G. *Opt. Eng.* **34**(3) 834 (1995)
63. Gamble F. T., Frye L. M., Grieser D. R. *Applied Optics* **31**(5) 652 (1992)
64. McMahon D. H., Johnson G. L., Teeter S. L., Whitheg C. G. *IEEE Trans. Comput.* **C-24**(5) 358 (1975)
65. Johnson E. G., Brasher J. D., Gregory D., Erbach P., Duingnan M., Behrmann G., Lee S. H., Dasghner W., Long P. *Opt. Eng.* **37**(1) 18 (1998)
66. Javidi J. D., Sergent A. *Opt. Eng.* **36**(3) 935 (1997)
67. Brasher J. D., Johnson E. G. *Opt. Eng.* **36**(9) 2409 (1997)
68. Grycewicz T. J., Javidi B. *Opt. Eng.* **35**(9) 2519 (1996)
69. Huh H., Pan J. K. *Opt. Eng.* **35**(4) 997 (1996)
70. Seth M., Datta A. K. *Opt. Eng.* **35**(6), 1779 (1996)
71. Rodolfo J., Rajbenbach H., Haignard J. P. *Opt. Eng.* **34**(4) 1166 (1995)
72. Dawagoe M., Tojo A. *Pattern Recognition* **17** 295 (1984)
73. Drake M. D., Lidd M. L., Fiddy M. A. *Opt. Eng.* **35**(9) 2499 (1996)
74. Сойфер В. А., Котляр В. В., Хонина С. Н. *Автометрия* вып. 1 31 (1996)
75. Дуда Р. О., Харт П. Е. *Распознавание образов и анализ сцен* (М.: Мир, 1976)
76. Загоруйко Н. Г. *Методы распознавания и их применение* (М.:Советское радио, 1972)
77. Soifer V. A., Kotlyar V. V., Khonina S. N., Skidanov R. V. *The 13th Inter. Conf. on Pattern Recognition* (Technical Univer. of Vienna, 25-29 Aug. **2** 586, 1996)
78. Soifer V. A., Kotlyar V. V., Khonina S. N., Skidanov R. V. *Proceedings of SPIE* **3238** 66 (1997)
79. Soifer V A., Khonina S. N., Ilyasova N. Yu., Kotlyar V. V. *Proceedings of SSPR'94* (1994)
80. Khonina S. N., Kotlyar V. V., Soifer V A., Dvoryanova T. P *Proceedings of SPIE* **2363** 249 (1994)
81. *Применение методов фурье-оптики* /Под ред. Старка Г. — (М.: Радио и связь, 1988)
82. Скиданов Р. В. *Компьютерная оптика* вып. 17 130 (1997)
83. Вентцель Е. С., Овчаров Л. А. *Теория вероятностей и ее инженерные приложения* (М.: Наука, 1988)
84. Takeda M., Ina H., Kabayashi S. *J. Opt. Soc. Am.* **72**(1) 157 (1982)
85. Kotlyar V. V., Seraphimovich P. G., Zalyalov O. K. *Opt. and Laser Technology* **27**(4) 251 (1995)
86. Котляр В. В., Залялов О. К. Компьютерная оптика (Самара, СГАУ, вып. 16 71, 1996)
87. Скиданов Р. В., Сойфер В. А., Котляр В. В. *Труды XXV школы-симпозиума по когерентной оптике и голографии* (Ярославль, ЯГПУ, 102, 1997)
88. Aksenov V. P., Isaev Yu. P *Optics Letters* **17** 1180 (1992)
89. Ambs P., Lee S. H., Tian Q., Fainman Y. *Appl. Opt.* **25**(22) 4039 (1986)
90. Woodford P., Casasent D. *Opt. Mem. and Neur. Net.* **1** 1 (1997)
91. Сойфер В. А., Котляр В. В., Скиданов Р. В. *Компьютерная оптика* вып. 17, 143 (1997)

ЗАКЛЮЧЕНИЕ

Малые апертуры и ограниченная оптическая сила ДОЭ в значительной мере определяет сферу их практического применения:

— волоконно-оптические линии связи;
— интегрально-оптические приборы;
— научные приборы;
— микроинтерферометры;
— безаберрационные объективы;
— контактные линзы и глазные хрусталики;
— дифракционные зеркала для приборов и установок;
— лазерные технологические установки ИК-диапазона;
— лазерные медицинские установки ИК-диапазона;
— офтальмологические приборы;
— оптические компьютеры;
— пространственные фильтры в процессорах оптических изображений;
— матрицы микролинз;
— светоделители и ответвители пучка;
— коллиматоры лазерных пучков и т.д.

Сфера приложений микрооптических дифракционных элементов, таким образом, является достаточно широкой благодаря широким функциональным возможностям ДОЭ по сравнению с традиционной оптикой. При этом, как правило, речь не идет о замене рефракционного оптического элемента на дифракционный: новая элементная база оптики позволяет решать новый класс прикладных задач.

Несомненным достоинством ДОЭ являются их улучшенные массогабаритные характеристики. Для искусственных хрусталиков и контактных линз эти характеристики ставят ДОЭ вне конкуренции. Эти же характеристики могут определить выбор ДОЭ для бортовых оптических систем.

Еще одним достоинством ДОЭ является их легкая тиражируемость и соответственно низкая стоимость в продукции массового назначения.

Ключевыми направлениями развития теории компьютерной оптики являются:

- разработка асимптотических и численных методов решения уравнений Максвелла;
- разработка итерационных методов расчета ДОЭ с учетом векторного характера оптических полей;
- решение класса задач на собственные функции и собственные значения оператора распространения света, позволяющие осуществлять селекцию мод и создавать пучки с инвариантными свойствами;
- развитие метода поля направлений и других методов интегральных преобразований в задачах обработки и распознавания оптических изображений;

Практические успехи компьютерной оптики будут зависеть от создания специальных технологических атоматов с разрешающей способностью не хуже 0,1 мкм и высоким быстродействием для изготовления ДОЭ.

Учебное издание

ВОЛКОВ Алексей Васильевич
ГОЛОВАШКИН Димитрий Львович
ДОСКОЛОВИЧ Леонид Леонидович
КАЗАНСКИЙ Николай Львович
КОТЛЯР Виктор Викторович
ПАВЕЛЬЕВ Владимир Сергеевич
СКИДАНОВ Роман Васильевич
СОЙФЕР Виктор Александрович
СОЛОВЬЕВ Владимир Степанович
УСПЛЕНЬЕВ Глеб Витальевич
ХАРИТОНОВ Сергей Иванович
ХОНИНА Светлана Николаевна

Редактор *М. Б. Козинцова*
Технический редактор *Г. Н. Морозова*
Оригинал-макет *С. М. Донского, И. И. Каледина, Е. Ю. Морозова, И. В. Шутова*

ЛР №071930 от 06.07.99
Подписано в печать 23.05.03. Формат $70\times100^{1}/_{16}$
Бумага офсетная № 1. Печать офсетная.
Усл. печ. л. 55,73. Уч.-изд. л. 61,28.

Издательская фирма
«Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117864 Москва В-485, Профсоюзная ул., 90

Отпечатано с оригинал-макета
в ППП «Типография «Наука» РАН
121099 Москва Г-99, Шубинский пер., 6
Заказ тип. № 8554

Издательская фирма
«Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117864 Москва, Профсоюзная ул., 90

В издательстве «ФИЗМАТЛИТ» вышли из печати (2000–2003 гг.)

Аракин В.Д. История английского языка: Учеб. пособие. — Изд. 2-е. — 2001. — 272 с.

Аракин В.Д. Сравнительная типология английского и русского языков: Учеб. пособие. — Изд. 3-е. — 2000. — 256 с.

Афанасьева О.В., Резвецова М.Д., Самохина Т.С. Сравнительная типология английского и русского языков (практикум): Учеб. пособие. — Изд. 2-е, испр. — 2000. — 208 с.

Беклемишев Д.В. Курс аналитической геометрии и линейной алгебры: Учеб. для вузов. — Изд. 8-е. — 2000. — 320 с.

Беклемишева Л.А., Петрович А.Ю., Чубаров И.А. Сборник задач по аналитической геометрии и линейной алгебре / Под ред. Д.В. Беклемишева: Учеб. пособие. — Изд. 2-е, перераб. — 2001. — 496 с.

Бендриков Г.А., Буховцев Б.Б. и др. Физика. Задачи для поступающих в вузы: Учеб. пособие для подготов. отделений вузов. — Изд. 9-е. — 2000. — 400 с.

Бугров С.Я., Никольский С.М. Сборник задач по высшей математике. — 3-е изд. — 2001. — 304 с.

Бутиков Е.И., Кондратьев А.С. Физика. В 3-х кн.: Учеб. пособие.
Кн. 1. Механика. — 2000. — 352 с.
Кн. 2. Электродинамика и оптика. — 2000. — 336 с.
Кн. 3. Строение и свойства вещества. — 2000. — 336 с.

Бутузов В.Ф., Крутицкая Н.Д., Шишкин А.А. Линейная алгебра в вопросах и задачах / Под ред. В.Ф. Бутузова: Учеб. пособие. — Изд. 2-е, испр. — 2002. — 248 с.

Бутузов В.Ф., Крутицкая Н.Ч., Медведев Г.Н., Шишкин А.А. Математический анализ в вопросах и задачах / Под ред. В.Ф. Бутузова: Учеб. пособие. — Изд. 5-е, испр. — 2002. — 480 с.

Буховцев Б.Б., Кривченков В.Д. и др. Сборник задач по элементарной физике / Под ред. В.Ф. Бутузова. — Изд. 6-е — 2000.

Владимиров В.С., Жаринов В.В. Уравнения математической физики: Учеб. для вузов. — Изд. 2-е., стереотип. — 2003. — 560 с.

Волковыский Л.И., Лунц Г.Л., Араманович И.Г. Сборник задач по теории функций комплексного переменного: Учеб. пособие для вузов. — Изд. 4-е, перераб. — 2002. — 312 с.

Гантмахер Ф.Р. Лекции по аналитической механике / Под ред. Е.С. Пятницкого: Учеб. пособие для вузов. — Изд. 3-е. — 2001. — 264 с.

Гильденбург В.Б., Миллер М.А. Сборник задач по электродинамике. — Изд. 2-е, доп. — 2001. — 168 с.

Гольдштейн Б.В., Городцов В.А. Механика сплошных сред. Ч. 1. — 2000. — 256 с.

Елютин П.В., Кривченков В.Д. Квантовая механика (с задачами) / Под ред. Н.Н. Боголюбова. — Изд. 2-е, перераб. — 304 с.

Задачи по общей физике / *Белонучкин В.Е., Заикин Д.А., Кингсеп А.С., Локшин Г.Р., Ципенюк Ю.М.* — 2001. — 336 с.

Зайцев В.Ф., Полянин А.Д. Справочник по обыкновенным дифференциальным уравнениям. — 2001. — 576 с.

Зимина О.В., Кириллов А.И., Сальникова Т.А. Решебник. Высшая математика — 2000. — 368 с.

Ильин В.А., Позняк Э.Г. Основы математического анализа. В 2-х ч. — 2001. — 648 с.; 464 с.

Ильин В.А., Позняк Э.Г. Линейная алгебра: Учеб. для вузов. — Изд. 5-е — 2001. — 320 с.

Ильин В.А., Позняк Э.Г. Аналитическая геометрия: Учеб. для вузов. — Изд. 5-е. — 2001. — 240 с.

Кадомцев С.Б. Аналитическая геометрия и линейная алгебра. — 2001. — 160 с.

Карманов В.Г. Математическое программирование: Учеб. пособие. — Изд. 5-е, стереотип. — 2001. — 264 с.

Кингсеп А.С., Локшин Г.Р., Ольхов О.А. Основы физики. Курс общей физики. В 2 т. Т. 1. Механика, электричество и магнетизм, колебания и волны, волновая оптика / Под ред. А.С. Кингсепа: Учебник. — 2001. — 560 с.

Кострикин А.И. Введение в алгебру. В 3-х кн.: Учеб. для вузов.
Кн. 1. Основы алгебры. — Изд. 2-е. — 2000. — 352 с.
Кн. 2. Линейная алгебра. — Изд. 2-е. — 2000. — 352 с.
Кн. 3. Основные структуры алгебры. — Изд. 2-е. — 2000. — 336 с.

Кудрявцев Л.Д. Краткий курс математического анализа. В 2-х т.: Учеб. пособие. — Изд. 3-е, перераб. — 2002. — 400 с.; 424 с.

Лавров И.А., Максимова Л.Л. Задачи по теории множеств, математической логике и теории алгоритмов. — 4-е изд. — 2001. — 256 с.

Ландау Л.Д., Лифшиц Е.М. Теоретическая физика. В 10 т.: Учеб. пособие для вузов. — 2001.

Лебедев В.И. Функциональный анализ и вычислительная математика.: Учеб. пособие. — Изд. 4-е, перераб. и доп. — 2001. — 296 с.

Лурье М.В. Геометрия. Техника решения задач: Учеб. пособие. — 2001. — 240 с.

Медведев Г.Н. Абитуриенту о письменном экзамене по математике. Физический факультет МГУ. 1997–2000 гг. — 2001. — 64 с.

Назаретов А.П., Пигарев Б.П., Садовничая И.В., Симонов А.А. Задачи и варианты их решения на вступительных экзаменах в московских вузах (экономические специальности): Учеб. пособие. — Изд. 2-е, перераб. и доп. — 2001. — 458 с.

Никольский С.Н. Курс математического анализа: Учеб. для вузов — Изд. 5-е. — 2000. — 640 с.

Пинский А.А. Задачи по физике / Под ред. Ю.И. Дика. — 2000. — 336 с.

Полянин А.Д. Справочник по линейным уравнениям математической физики. — 2001. — 576 с.

Потапов М.К., Олехник С.Н., Нестеренко Ю.В. Конкурсные задачи по математике: Справочное пособие. — 2001. — 400 с.

Пугачев В.С. Теория вероятностей и математическая статистика: Учебник для вузов. — Изд. 2-е, испр. и доп. — 2002. — 496 с.

Рябенький В.С. Введение в вычислительную математику. — Изд. 2-е. — 2000. — 296 с.

Сборник задач по алгебре / Под ред. А.И. Кострикина: Учеб. для вузов. — Изд. 3-е, испр. и доп. — 2001. — 464 с.

Сборник задач по уравнениям математической физики / *Владимиров В.С., Вашарин А.А., Каримова Х.Х., Михайлов В.П., Сидоров Ю.И., Шабунин М.И.* — Изд. 4-е, стереотип. — 2003. — 288 с.

Сигал И.Х., Иванова А.П. Введение в прикладное дискретное программирование. Учеб. пособие. — 2002. — 240 с.

Треногин В.А. Функциональный анализ: Учеб. пособие. — Изд. 3-е, испр. — 2002. — 488 с.

Треногин В.А., Писаревский Б.М., Соболева Т.С. Задачи и упражнения по функциональному анализу: Учеб. пособие. — Изд. 2-е, испр. и доп. — 2002. — 240 с.

Шклярский Д.О., Ченцов Н.Н., Яглом И.М. Избранные задачи и теоремы элементарной математики. Геометрия. Стереометрия. — 2000. — 248 с.

Шклярский Д.О., Ченцов Н.Н., Яглом И.М. Избранные задачи и теоремы элементарной математики. Геометрия. Планиметрия. — 2000. — 248 с.
Шклярский Д.О., Ченцов Н.Н., Яглом И.М. Избранные задачи и теоремы элементарной математики. Арифметика и алгебра. — 2000. — 320 с.
Элементарный учебник физики / Под ред. Г.С. Ландсберга. В 3-х кн.: Учеб. пособие. — 2000. — 512 с.; 400 с.; 400 с.
Яворский Б.М., Пинский А.А. Основы физики / Под ред. Ю.И. Дика. В 2-х кн. — 2000. — 624 с.; 576 с.
Яворский Б.М., Селезнев Ю.А. Физика: Справочное пособие для поступающих в вузы. — Изд. 5-е, перераб. — 2000. — 592 с.

По вопросам приобретения книг обращаться:
Издательская фирма
«Физико-математическая литература»
117864 Москва, Профсоюзная ул., 90
тел./факс (095) 334-7421, e-mail: fizmat@maik.ru