

МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

С.Н. Хонина

**Оптическая информатика.
Конспект лекций.**

Электронное учебное пособие

САМАРА

2011

Автор: **Хонина Светлана Николаевна**

Хонина, С. Н. Оптическая информатика [Электронный ресурс] : электрон. учеб. пособие / С. Н. Хонина; Минобрнауки России, Самар. гос. аэрокосм. ун-т им. С. П. Королева (нац. исслед. ун-т). - Электрон. текстовые и граф. дан. (0,43 Мбайт). - Самара, 2011. - 1 эл. опт. диск (CD-ROM).

Лекционный материал содержит описание частных решений волнового уравнения Гельмгольца для монохроматического излучения. Рассмотрены такие частные решения как плоские и сферические волны, используемые в оптике как базисы для описания распространения произвольных полей в среде с постоянным показателем преломления. Детально показано построение оператора распространения произвольного поля, заданного в некоторой плоскости, на основе разложения по базису плоских волн. Показана его связь с преобразованием Фурье, что существенно облегчает программную реализацию данного оператора распространения для моделирования. Использование базиса сферических волн позволяет получить более удобный для аналитических исследований вид оператора распространения. Среди решений параксиального волнового уравнения, кроме классических мод Гаусса-Эрмита, Гаусса-Лагерра и Бесселя, рассмотрено новое решение – гипергеометрические пучки. Лекционный материал соответствует программе курса «Оптическая информатика» 010400.62 Прикладная математика и информатика (бакалавриат), 6 факультет, кафедра технической кибернетики.

© Самарский государственный
аэрокосмический университет, 2011

ОПТИЧЕСКАЯ ИНФОРМАТИКА.

Введение.

Данный курс основывается на разделе "дифракционная оптика", синонимом этого термина является "компьютерная оптика", "плоская оптика".

Дифракция – по Зоммерфельду – любое отклонение световых лучей от прямой линии, если оно не может быть объяснено преломлением на границе раздела двух сред или рефракцией, т.е. изгибанием лучей в средах с непрерывно меняющимся показателем преломления. Т.е. если не может быть объяснено лучевой теорией (геометрическим приближением), а требует волновой теории, основой которой служит теории дифракции Кирхгофа—Зоммерфельда.

Первое объяснение дифракции было сделано Гюйгенсом, который предположил, что если окружить все источники излучения замкнутой поверхностью, то каждую точку этой поверхности можно рассматривать как источник вторичных волн, распространяющихся во всех направлениях. Тогда в следующий момент времени волновой фронт можно найти как огибающую вторичных сферических волн (рис. 1) – **принцип Гюйгенса**.



Рис. 1. Волновой фронт как огибающая вторичных сферических волн.

Затем к принципу Гюйгенса было добавлено **Френелем** предположение о том, что при нахождении нового светового распределения необходимо учитывать также и интерференцию вторичных волн, принимая во внимание когерентность вторичных источников, так как они возбуждаются одними и теми же первичными источниками. Подобное дополнение привело к значительному уточнению картины дифракции, однако при этом были сделаны достаточно произвольные предположения об амплитудах и фазах вторичных источников.

В **математическом виде** эта теория была сформулирована в 1883 году **Кирхгофом**, который показал, что при учете волновой природы света

особенности амплитуд и фаз, введенные Френелем, легко объясняются. Но и Кирхгоф допустил некий произвол, связанный с определением граничных условий при дифракции, т.е. компонент поля на границе объекта, на котором происходит дифракция.

Окончательно теория была сформулирована **Зоммерфельдом** на рубеже XIX и XX вв. при использовании метода функций Грина. Ему удалось добиться внутренней непротиворечивости теории. Именно этот вариант теории дифракции Кирхгофа—Зоммерфельда будет рассматриваться.

Следует помнить, что и эта прекрасно работающая теория является **приближенной**, так как она носит **скалярный характер**. Теория рассматривает только одну компоненту поперечного электромагнитного поля и считает, что другие компоненты можно рассматривать независимо. Так как уравнения Максвелла являются векторными, то в общем случае это утверждение несправедливо и следовало бы оперировать с векторными потенциалами.

Скалярная теория дифракции дает очень **хорошие результаты**, подтвержденные экспериментами, но при следующих **условиях**:

- 1) отверстия в экранах велики, по сравнению с длиной волны λ ;
- 2) дифрагированные волны наблюдаются не слишком близко от экрана.

В качестве примера, когда скалярная теория не работает, можно привести теорию дифракционных решеток высокого разрешения (с количеством штрихов более 1000 на 1 мм). В этом случае угловые соотношения в дифрагированном поле описываются скалярной теорией дифракции, но она дает неверные значения амплитудного распределения.

Тем не менее, ограничимся рассмотрением только скалярной теорией дифракции, так как она описывает большинство наблюдаемых явлений.

Основная часть курса посвящена решению **обратной задачи** дифракции: по известному (заданному) распределению света на некоторой поверхности определить функцию источника.

Т.е. определить (рассчитать) какое должно быть комплексное распределение во входной плоскости (плоскости дифракционного оптического элемента – ДОЭ), чтобы на некоторой поверхности формировалось желаемое распределение. На основе полученного решение обратной задачи определяется функция комплексного пропускания ДОЭ:

$$\tau(x, y) = A(x, y) \exp[i\varphi(x, y)], \quad (1)$$

где $A(x, y)$ – амплитуда, $\varphi(x, y)$ – фаза. В определении $\tau(x, y)$ состоит синтез (расчет) ДОЭ.

Наибольший интерес с точки зрения энергетической эффективности представляют собой **фазовые ДОЭ**:

$$\tau(x, y) = \exp[i\varphi(x, y)], \quad (2)$$

которые изготавливаются в виде тонких прозрачных пластинок с нанесенным на них микрорельефом. Фазовый рельеф получается путем химического травления материала пластинки пропорционально значению фазовой функции.

Фазовые ДОЭ осуществляют требуемые преобразования волновых полей теоретически без потери световой энергии (практическая дифракционная эффективность 60-95%). Действие фазового ДОЭ в приближении тонкой оптики сводится к определенному набегу фазы падающего излучения в каждой точке области ДОЭ. Затем, после прохождения некоторого расстояния, фазовые изменения проявляются в изменениях амплитуды падающего света.

Отличие ДОЭ от традиционной оптики в том, что они тонкие, «плоские». Толщина традиционного оптического элемента может составлять тысячи длин волн. При этом значения фазы φ лежат в интервале от 0 до тысяч единиц 2π . С учетом того, что комплексная экспонента в выражении (2) является периодической функцией с периодом 2π , фаза может быть приведена к интервалу $[0, 2\pi]$.

В качестве примера рассмотрим приведение фазовой функции сферической линзы к диапазону $[0, 2\pi)$ (см. рис. 2)

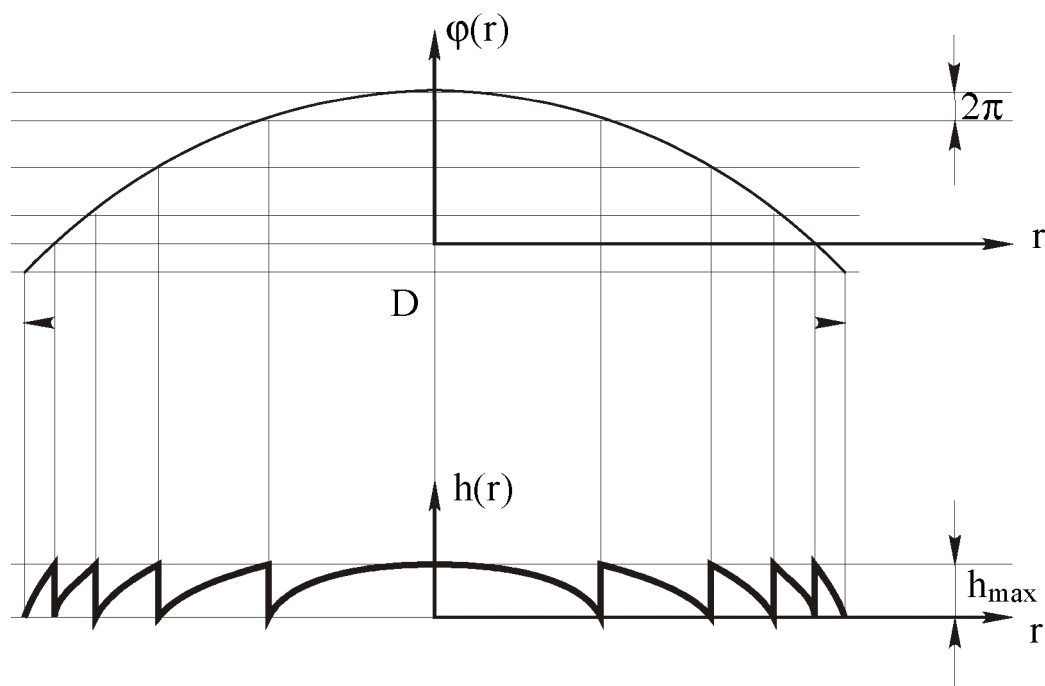


Рис. 2. Приведение фазовой функции сферической линзы к диапазону $[0, 2\pi)$

Высота микрорельефа, создаваемого на прозрачной или отражающей пластинке определяется по формуле

$$h(x, y) = \frac{\lambda}{(n-1)} \frac{\text{mod}_{2\pi} \varphi(x, y)}{2\pi}, \quad (3)$$

где λ – длина волны освещающего ДОЭ источника, n – коэффициент преломления материала пластинки.

Максимальная высота рельефа составляет

$$h_{\max} = \frac{\lambda}{n-1} \quad (4)$$

и имеет порядок длины волны (например, для стекла $n=1,5$ и $h_{\max}=2\lambda$).

Качество расчета и изготовления ДОЭ характеризуется степенью отличия сформированного распределения света от заданного и величиной световой энергии, идущей на формирование этого распределения. Для задач обработки информации, в которых требуется высокая точность формирования волновых фронтов или распределений интенсивности, основной характеристикой качества ДОЭ является ошибка отклонения. А для задач обработки материалов мощным лазерным излучением требуется высокая степень концентрации энергии в области заданного изображения. Для таких задач требуются фазовые

ДОО, обладающие высокой энергетической эффективностью. Поэтому различаются и методы, с помощью которых рассчитываются соответствующие ДОО.

Уравнения Максвелла. Волновое уравнение Гельмгольца.

В то время, когда стало ясно, что свет – это электромагнитное поле, появился универсальный математический аппарат, связывающий между собой функции изменения во времени и пространстве электрического и магнитного полей – **уравнения Максвелла**.

Уравнения Максвелла в классических обозначениях имеют следующий вид:

$$\text{rot}\mathbf{E} = -\dot{\mathbf{B}}, \quad (1)$$

$$\text{rot}\mathbf{H} = \dot{\mathbf{D}} + \mathbf{J}, \quad (2)$$

$$\text{div}\mathbf{B} = 0, \quad (3)$$

$$\text{div}\mathbf{D} = \rho, \quad (4)$$

и связывают следующие функции времени t и координат в пространстве $\mathbf{r}=(x,y,z)$:

$\mathbf{E}$ – напряженность электрического поля (вольт/метр),

$\mathbf{H}$ – напряженность магнитного поля (ампер/метр),

$\mathbf{D}$ – электрическая индукция (кулон/метр²),

$\mathbf{B}$ – магнитная индукция (вебер/метр²),

$\mathbf{J}$ – поверхностная плотность электрического тока (ампер/метр²),

ρ – объемная плотность электрического заряда (кулон/метр³).

Скалярная функция $U(\mathbf{r}, t) = U(x, y, z, t)$

Векторная функция $\mathbf{F}(\mathbf{r}, t) = \begin{pmatrix} F_x(x, y, z, t) \\ F_y(x, y, z, t) \\ F_z(x, y, z, t) \end{pmatrix} = F_x \mathbf{i} + F_y \mathbf{j} + F_z \mathbf{k}$

Дифференцирование по времени $\dot{\mathbf{B}} = \frac{\partial \mathbf{B}}{\partial t}$

Оператор дифференцирования по координатам: $\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} = \frac{\partial}{\partial x} \mathbf{i} + \frac{\partial}{\partial y} \mathbf{j} + \frac{\partial}{\partial z} \mathbf{k}$

$$\nabla \cdot U = \text{grad}U = \begin{pmatrix} \frac{\partial U}{\partial x} \\ \frac{\partial U}{\partial y} \\ \frac{\partial U}{\partial z} \end{pmatrix} = \frac{\partial U}{\partial x} \mathbf{i} + \frac{\partial U}{\partial y} \mathbf{j} + \frac{\partial U}{\partial z} \mathbf{k}$$

$$\nabla \cdot \mathbf{F} = \text{div} \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

$$\nabla \times \mathbf{F} = \text{rot} \mathbf{F} = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{pmatrix} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \mathbf{i} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \mathbf{j} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathbf{k}$$

Дифференциальной форме уравнения Максвелла:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (1')$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (2')$$

$$\nabla \cdot \mathbf{B} = 0, \quad (3')$$

$$\nabla \cdot \mathbf{D} = \rho, \quad (4')$$

Величины $\mathbf{J}$ и ρ связаны уравнением непрерывности:

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0. \quad (5)$$

Векторы поля, удовлетворяющие уравнениям Максвелла (1)-(2), д.б. непрерывными функциями пространственных и временной координат с непрерывными производными, однозначно определенными и ограниченными повсюду, кроме точек резкого изменения в распределении тока или заряда. Обычно такие скачки происходят на границе двух сред.

Нужно отметить, что система уравнений (1)-(4) не является полной, т.к. в нее входят пять векторов, а векторных уравнений всего два, причем два скалярных являются следствием уравнений (1), (5) и (2). Для расчета электромагнитных полей в различных средах уравнения (1)-(4) дополняются системой материальных уравнений:

$$\mathbf{D} = D(\mathbf{E}), \quad \mathbf{B} = B(\mathbf{H}), \quad \mathbf{J} = J(\mathbf{D}). \quad (6)$$

Для **линейной изотропной однородной** среды известны следующие соотношения:

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}, \quad \mathbf{J} = \sigma \mathbf{E}, \quad (7)$$

где ε – электрическая проницаемость (фарад/метр),

μ – магнитная проницаемость (генри/метр),

σ – удельная электрическая проводимость (сименс/метр).

Для **непоглощающих диэлектрических** сред:

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0}, \quad \mu_r = \frac{\mu}{\mu_0} = 1, \quad \sigma = 0, \quad (8)$$

где ε_r и μ_r относительные диэлектрическая и магнитная проницаемости, соответственно, ε_0 и μ_0 – параметры свободного пространства:

$$\varepsilon_0 = \frac{1}{36\pi \cdot 10^9} \approx 8.854 \cdot 10^{-12}, \quad \mu_0 = 4\pi \cdot 10^{-7}.$$

Если речь идет о работе в оптической части частотного спектра и распространении в диэлектрической среде, то часто вместо диэлектрической проницаемости используют показатель преломления среды:

$$n = \sqrt{\varepsilon_r}. \quad (9)$$

Так, например, для стекла, часто используемого в качестве материала для диэлектрических волноводов, показатель преломления равен $n_{gl} \approx 1.5$.

В общем случае материальные уравнения (6) можно выразить через тензорные соотношения, т.е.:

$$\mathbf{D} = \boldsymbol{\varepsilon} \mathbf{E}, \quad (10)$$

где

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix}, \quad (11)$$

тогда

$$\begin{aligned} D_x &= \varepsilon_{11}E_x + \varepsilon_{12}E_y + \varepsilon_{13}E_z, \\ D_y &= \varepsilon_{21}E_x + \varepsilon_{22}E_y + \varepsilon_{23}E_z, \\ D_z &= \varepsilon_{31}E_x + \varepsilon_{32}E_y + \varepsilon_{33}E_z. \end{aligned} \quad (12)$$

Такое тензорное соотношение существует для анизотропных материалов, например, для кристаллов. В то время как изотропные материалы демонстрируют одинаковые свойства во всех направлениях.

Кроме того, если среда неоднородная, то диэлектрическая проницаемость является функцией, зависящей от положения, т.е. $\varepsilon(x,y,z)$. Одним из примеров неоднородной среды является градиентное оптическое волокно.

В вакууме и диэлектриках, считая среду свободной от зарядов и токов, а также линейно изотропной, можно использовать уравнения Максвелла (1)-(4) в следующем виде:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (13)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \quad (14)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (15)$$

$$\nabla \cdot \mathbf{D} = 0, \quad (16)$$

и материальные уравнения (6) для такой среды:

$$\mu_r \approx 1, \quad \mathbf{D} = \boldsymbol{\varepsilon} \mathbf{E}, \quad \mathbf{B} = \mu_0 \mathbf{H}. \quad (17)$$

Уравнения Максвелла описывают поле, в котором вектор электрической напряженности перпендикулярен вектору магнитной напряженности, и оба они перпендикулярны направлению распространения света, поэтому поле называется поперечным:

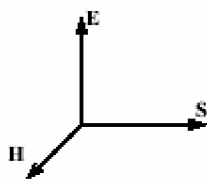


Рис. Взаимное расположение векторов электрической ($\mathbf{E}$) и магнитной ($\mathbf{H}$) напряженности и направления распространения света ($\mathbf{S}$).

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}, \quad (17)$$

$\mathbf{S}$ – вектор Умова-Пойнтинга, это вектор плотности потока энергии электромагнитного поля, по модулю равен количеству энергии, переносимой через единичную площадь, нормальную к $\mathbf{S}$, в единицу времени, а своим направлением вектор определяет направление переноса энергии волны.

Уравнения Максвелла являются парными и их сложно использовать при решении задачи с граничными условиями. Из уравнений Максвелла можно получить волновое уравнение, в котором участвует только один вектор. Решение волнового уравнения позволяет получить описание распространения энергии в данной среде при выполнении граничных условий. Для электрического вектора (для магнитного вектора вид аналогичный) волновое уравнение.

Возьмем уравнение для ротора электрического поля $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$ и материальное уравнение $\mathbf{B} = \mu \mathbf{H}$, тогда:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu \frac{\partial \mathbf{H}}{\partial t},$$

Векторно домножим это уравнение на ∇ :

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E} &= -\mu \left(\nabla \times \frac{\partial \mathbf{H}}{\partial t} \right) = -\mu \left(\frac{\partial}{\partial t} (\nabla \times \mathbf{H}) \right) = -\mu \left(\frac{\partial}{\partial t} \left(\frac{\partial \mathbf{D}}{\partial t} \right) \right) = \\ &= -\mu \left(\frac{\partial^2}{\partial t^2} (\epsilon \mathbf{E}) \right) = -\mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \end{aligned}$$

Оператор Лапласа $\Delta = \nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$

$$\Delta U = \nabla^2 U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2}$$

$$\Delta \mathbf{F} = \nabla^2 \mathbf{F} = (\nabla^2 F_x, \nabla^2 F_y, \nabla^2 F_z) = \nabla^2 F_x \mathbf{i} + \nabla^2 F_y \mathbf{j} + \nabla^2 F_z \mathbf{k}$$

$$\text{div grad} U = \nabla \cdot (\nabla U) = \nabla^2 U$$

$$\begin{aligned} \text{rot grad} U &= \nabla \times (\nabla U) = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial U}{\partial x} & \frac{\partial U}{\partial y} & \frac{\partial U}{\partial z} \end{pmatrix} = \\ &= \left(\frac{\partial}{\partial y} \frac{\partial U}{\partial z} - \frac{\partial}{\partial z} \frac{\partial U}{\partial y} \right) \mathbf{i} + \left(\frac{\partial}{\partial x} \frac{\partial U}{\partial z} - \frac{\partial}{\partial z} \frac{\partial U}{\partial x} \right) \mathbf{j} + \left(\frac{\partial}{\partial y} \frac{\partial U}{\partial x} - \frac{\partial}{\partial x} \frac{\partial U}{\partial y} \right) \mathbf{k} = 0 \end{aligned}$$

$$\begin{aligned} \text{grad div} \mathbf{F} &= \nabla \cdot (\nabla \cdot \mathbf{F}) = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \cdot \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) = \\ &= \left(\frac{\partial^2 F_x}{\partial x^2} + \frac{\partial^2 F_y}{\partial x \partial y} + \frac{\partial^2 F_z}{\partial x \partial z} \right) \mathbf{i} + \left(\frac{\partial^2 F_x}{\partial y \partial x} + \frac{\partial^2 F_y}{\partial y^2} + \frac{\partial^2 F_z}{\partial y \partial z} \right) \mathbf{j} + \left(\frac{\partial^2 F_x}{\partial z \partial x} + \frac{\partial^2 F_y}{\partial z \partial y} + \frac{\partial^2 F_z}{\partial z^2} \right) \mathbf{k} = \\ &= \nabla^2 \mathbf{F} + \nabla \times (\nabla \times \mathbf{F}) \end{aligned}$$

$$\text{rot rot} \mathbf{F} = \nabla \times (\nabla \times \mathbf{F}) = \nabla \cdot (\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$$

$$\text{div rot} \mathbf{F} = \nabla \cdot (\nabla \times \mathbf{F}) = \frac{\partial}{\partial x} \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) + \frac{\partial}{\partial y} \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) + \frac{\partial}{\partial z} \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) = 0$$

Из $\nabla \times \nabla \times \mathbf{E} = -\mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}$, используя $\nabla \times (\nabla \times \mathbf{F}) = \nabla \cdot (\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}$, получаем:

$$\nabla \cdot (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

В неоднородной среде $\varepsilon(x, y, z)$ **волновое уравнение** выглядит следующим образом ($\mathbf{D} = \varepsilon \mathbf{E}$):

$$\nabla^2 \mathbf{E} - \mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = -\nabla \left(\mathbf{E} \cdot \frac{\nabla \varepsilon}{\varepsilon} \right). \quad (18)$$

Если же среда однородная, то градиент диэлектрической проницаемости равен нулю $\nabla\epsilon=0$, и тогда правая часть тоже будет равна нулю ($\nabla \cdot \mathbf{D} = 0$):

$$\nabla^2 \mathbf{E} - \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \quad (19)$$

и аналогично
$$\nabla^2 \mathbf{H} - \mu\epsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0. \quad (20)$$

Уравнения (19) и (20) являются векторными, имеющими по три скалярных компоненты каждый. Уравнение (19) можно представить в декартовых координатах следующим образом:

$$\begin{cases} \nabla^2 E_x - \mu\epsilon \frac{\partial^2 E_x}{\partial t^2} = 0 \\ \nabla^2 E_y - \mu\epsilon \frac{\partial^2 E_y}{\partial t^2} = 0 \\ \nabla^2 E_z - \mu\epsilon \frac{\partial^2 E_z}{\partial t^2} = 0 \end{cases} \quad (21)$$

Т.е. каждая компонента электромагнитного поля, если обозначить ее через $\psi(x, y, z, t)$, удовлетворяет **скалярному волновому уравнению**:

$$\nabla^2 \Psi(x, y, z, t) - \frac{1}{v^2} \frac{\partial^2 \Psi(x, y, z, t)}{\partial t^2} = 0, \quad (22)$$

где $v = \frac{1}{\sqrt{\mu\epsilon}}$ – скорость волны в среде.

Учитывая, что скорость волны в воздухе $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$, показатель преломления

данной среды $n = \frac{c}{v} = \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_0 \epsilon_0}}$.

Для постоянных гармонических по времени – **монохроматических** - полей временная зависимость имеет определенный вид:

$$\Psi(x, y, z, t) = \psi(x, y, z) \exp(i\omega t + i\varphi), \quad (23)$$

где ω – частота излучения, φ – начальная фаза.

В частности,

$$\begin{aligned} E_x(x, y, z, t) &= E_x(x, y, z) \exp(i\omega t + i\varphi_x), \\ E_y(x, y, z, t) &= E_y(x, y, z) \exp(i\omega t + i\varphi_y), \\ E_z(x, y, z, t) &= E_z(x, y, z) \exp(i\omega t + i\varphi_z). \end{aligned} \quad (24)$$

Тогда, беря производную по времени, уравнения Максвелла (13)-(16) можно переписать:

$$\nabla \times \mathbf{E} = -i\omega\mu\mathbf{H}, \quad (25)$$

$$\nabla \times \mathbf{H} = i\omega\epsilon\mathbf{E}, \quad (26)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (27)$$

$$\nabla \cdot \mathbf{E} = 0, \quad (28)$$

где $\mathbf{E}(x, y, z) = \mathbf{i}E_x(x, y, z) + \mathbf{j}E_y(x, y, z) + \mathbf{k}E_z(x, y, z)$ и т. д.

Откуда можно получить (будет подробно рассмотрено в третьей лекции) **волновое уравнение Гельмгольца**:

$$(\nabla^2 + k^2)\mathbf{E}(x, y, z) = 0, \quad (29)$$

$$(\nabla^2 + k^2)\mathbf{H}(x, y, z) = 0, \quad (30)$$

где $k = \omega\sqrt{\mu\epsilon} = \frac{\omega}{v} = \frac{2\pi}{\lambda}$ – волновое число, $\lambda = \frac{\lambda_0}{n}$ – длина волны в среде, λ_0 – длина волны в воздухе.

Фазовая скорость одной монохроматической гармонической волны:

$$v = \frac{\omega}{k}. \quad (31)$$

Для квазимонохроматической волны

$$\omega \pm \Delta\omega, \quad \omega \gg \Delta\omega,$$

выражение для фазовой скорости совпадает с (31), но вводится также понятие групповой скорости:

$$v_g = \frac{\Delta\omega}{\Delta k}. \quad (32)$$

В пределе при $\Delta\omega \rightarrow 0$

$$v_g = \lim_{\Delta\omega \rightarrow 0} \frac{\Delta\omega}{\Delta k} = \frac{d\omega}{dk}. \quad (33)$$

Фактически, групповая скорость представляет собой фазовую скорость огибающей.

Волновое уравнение. Частные решения волнового уравнения. Интегральные преобразования в оптике.

В волновой теории света предполагается, что свет распространяется в виде волны. Оптическая волна описывается скалярной **комплексной волновой функцией** пространственных (x, y, z) и временной t координат:

$$V(x, y, z, t)$$

Точное физическое значение этой функции не определено. По этому поводу можно только сказать, что она представляет одну из компонент электрического или магнитного поля.

Распространение светового поля в пространстве описывается **волновым уравнением**

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) V(x, y, z, t) = 0 \quad (1.1)$$

где c - скорость света в свободном пространстве,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

- оператор Лапласа или лапласиан.

Любая функция, удовлетворяющая (1.1) представляет собой возможную оптическую волну. А так как волновое уравнение (1.1) линейное, то применим **принцип суперпозиции**, то есть, композиция двух функций, удовлетворяющих (1.1) будет также представлять собой возможную оптическую волну.

В свободном пространстве свет распространяется с одинаковой скоростью $c_0 = 3 \cdot 10^8$ м/сек. При этом свет может иметь различную длину волны λ .

Оптические длины волн делятся на 3 группы: инфракрасные, видимые и ультрафиолетовые. Длина волны видимого диапазона 390-760 nm, т.е. ~ 0.5 мкм ($5 \cdot 10^{-7}$ м).

Для каждой длины волны λ определена частота электромагнитных колебаний ν , удовлетворяющая следующему соотношению $\nu \lambda = c$.

Как правило свет представляет собой сумму световых компонент с различными длинами волн. Однако, далее, пока не будет специально оговорено, будем рассматривать световое поле одной монохроматической составляющей, то есть будем считать, что свет имеет одну длину волны λ и одну соответствующую ей частоту электромагнитных колебаний ν .

Источником такого монохроматического света является, например, лазер. Лазеры бывают разных цветов: синие, зеленые, красные. Наиболее распространен He-Ne лазер с длиной волны 633 nm (красный).

Для такой монохроматической волны временная зависимость может быть выражена в явном виде

$$V(x, y, z, t) = U(x, y, z) \exp(-i2\pi\nu t). \quad (1.2)$$

Подставим (1.2) в (1.1):

$$\begin{aligned} \left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) U(x, y, z) \exp(-i2\pi\nu t) &= \exp(-i2\pi\nu t) \nabla^2 [U(x, y, z)] - U(x, y, z) \frac{1}{c^2} \frac{\partial^2}{\partial t^2} [\exp(-i2\pi\nu t)] = \\ &= \exp(-i2\pi\nu t) \left\{ \nabla^2 [U(x, y, z)] - U(x, y, z) \frac{1}{c^2} (-i2\pi\nu)^2 \right\} = 0 \end{aligned}$$

$$\nabla^2 [U(x, y, z)] + U(x, y, z) \left(\frac{2\pi\nu}{c} \right)^2 \stackrel{\nu = \frac{1}{\lambda}}{=} \left[\nabla^2 + \left(\frac{2\pi}{\lambda} \right)^2 \right] U(x, y, z) = 0$$

обозначив $k = 2\pi/\lambda$ - **волновое число света** (для оптических волн $\sim 10^7$ м), получим **уравнение Гельмгольца**:

$$[\nabla^2 + k^2] U(x, y, z) = 0 \quad (1.3)$$

описывающее распространение пространственной части.

$$U(x, y, z) = A(x, y, z) \exp[i\varphi(x, y, z)]$$

называется **комплексной амплитудой**. При этом величина $A(x, y, z) = |U(x, y, z)|$ называется **амплитудой** волны, $\exp[i\varphi(x, y, z)] = \arg[U(x, y, z)]$ **фазой**.

С комплексной амплитудой $U(x, y, z)$ можно связать конкретную физическую величину, которая измеряется оптическими приборами. Такой величиной является оптическая **интенсивность** света:

$$I(x, y, z) = |U(x, y, z)|^2 = U(x, y, z) U^*(x, y, z)$$

Интенсивность монохроматической волны не зависит от времени.

Волновым фронтом называются поверхности равной фазы:

$$\varphi(x, y, z) = \text{const} = 2\pi q, \quad q - \text{целое.}$$

Нормаль волнового фронта в точке (x, y, z) параллельна **вектору градиента фазы**

$$\nabla\varphi(x, y, z) = \left(\frac{\partial\varphi(x, y, z)}{\partial x}, \frac{\partial\varphi(x, y, z)}{\partial y}, \frac{\partial\varphi(x, y, z)}{\partial z} \right),$$

который указывает направление наибольшей скорости изменения фазы.

Частные решения уравнения Гельмгольца: плоская и сферическая волны

Плоская волна

Частным решением уравнения Гельмгольца являются плоские и сферические волны, а также моды Бесселя (будут рассмотрены позже).

Плоская волна имеет следующую комплексную амплитуду:

$$F(x, y, z) = A \exp(i\mathbf{k}\mathbf{x}) = A \exp[i(k_x x + k_y y + k_z z)]. \quad (1.4)$$

где A - комплексная константа, называемая **комплексной оболочкой**, $\mathbf{k} = (k_x, k_y, k_z)$ - **волновой вектор**.

Так как фаза $\arg\{F(\mathbf{x})\} = \arg\{A\} + \mathbf{k}\mathbf{x}$, то волновой фронт определяется из уравнения $\mathbf{k}\mathbf{x} = k_x x + k_y y + k_z z = 2\pi q - \arg\{A\}$, q - целое, и будет представлять собой параллельные плоскости, перпендикулярные волновому вектору $\mathbf{k}$ (поэтому и называется плоская волна). Расстояние между двумя плоскостями:

$$\mathbf{k}\mathbf{x} = k_x x + k_y y + k_z z = 2\pi q - \arg\{A\} \quad \text{и} \quad \mathbf{k}\mathbf{x}' = k_x x' + k_y y' + k_z z' = 2\pi(q+1) - \arg\{A\}$$

$$\|\mathbf{k}(\mathbf{x} - \mathbf{x}')\| = 2\pi \Rightarrow \|\mathbf{x} - \mathbf{x}'\| = \frac{2\pi}{k} = \lambda \quad \text{- длина волны}$$

Таким образом, волновая функция (1.4) является периодической в пространстве с периодом λ . Интенсивность плоской волны $I(\mathbf{x}) = |A|^2 = \text{const}$ везде в пространстве. Ясно, что эта волна является абстракцией.

Подставим (1.4) в (1.3):

$$(\nabla^2 + k^2) A \exp[ik(\alpha x + \beta y + \gamma z)] = A \exp[ik(\alpha x + \beta y + \gamma z)] \{ (i^2 k^2 (\alpha^2 + \beta^2 + \gamma^2) + k^2 \} = 0$$

равенство выполняется при условии $\alpha^2 + \beta^2 + \gamma^2 = 1 \Rightarrow \gamma = \pm \sqrt{1 - \alpha^2 - \beta^2}$.

Следовательно, плоская волна представима в виде:

$$F(x, y, z) = A \exp \left[ik \left(\alpha x + \beta y \pm z \sqrt{1 - \alpha^2 - \beta^2} \right) \right] = F(x, y, z=0) \Phi(\alpha, \beta, z) = F_0(x, y) \Phi(\alpha, \beta, z) = \\ = A \exp[ik(\alpha x + \beta y)] \exp \left[ik \pm z \sqrt{1 - \alpha^2 - \beta^2} \right] = F_{\alpha, \beta}(x, y, z)$$

где
$$\Phi(\alpha, \beta, z) = A \exp \left(\pm ikz \sqrt{1 - \alpha^2 - \beta^2} \right) \quad (1.5)$$

- уравнение распространения плоских волн в пространстве вдоль оси z .

Знак "+" перед квадратным корнем выбирается в случае, если волна приходит к плоскости z слева, знак "-", если волна приходит справа. В дальнейшем будем выбирать знак "+".

Из (1.5) следует условие на параметры α и β :

$$\alpha^2 + \beta^2 < 1 \quad (1.6)$$

В случае, если $\alpha^2 + \beta^2 > 1$, то вместо (1.5) будем иметь

$$\Phi(\alpha, \beta, z) = \exp \left(- kz \sqrt{\alpha^2 + \beta^2 - 1} \right) \quad (1.7)$$

Амплитуда (1.7) описывает затухающие по экспоненте световые волны, которые не распространяются далее в пространстве по оси z , и мы исключим их из рассмотрения.

Базис плоских волн

Наша цель теперь состоит в получении связи между некоторой амплитудой света на плоскости z : $U(x, y, z)$ и амплитудой света на плоскости $z=0$: $U_0(x, y)$ [1].

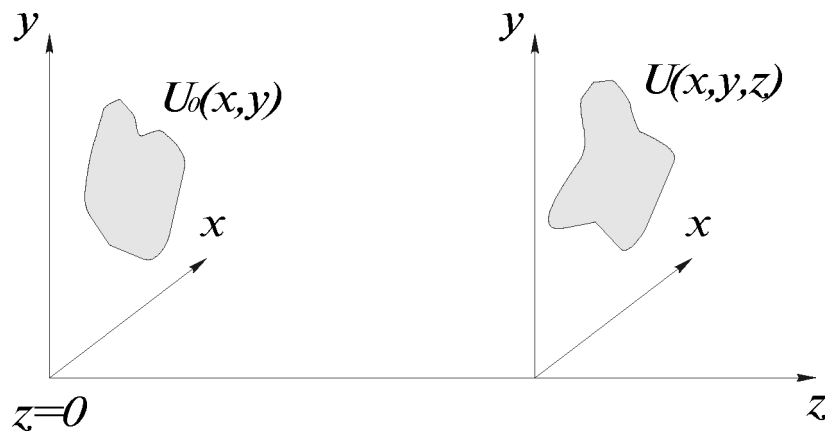


Рис. 4

Представим функцию $U(x, y, z)$ в плоскости z в виде разложения по базису плоских волн:

$$\begin{aligned}
U(x, y, z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\alpha, \beta) F_{\alpha, \beta}(x, y, z) d\alpha d\beta = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\alpha, \beta) \exp[ik(\alpha x + \beta y)] \exp[ikz\sqrt{1-\alpha^2-\beta^2}] d\alpha d\beta
\end{aligned} \tag{1.8}$$

Такое разложение выглядит как результат обратного преобразования Фурье. Действительно, физически, преобразование Фурье - это разложение поля по плоским волнам, а переменные α и β , пропорциональные углам Эйлера, задающим направление распространения плоской волны, являются частотными переменными, $C(\alpha, \beta)$ - веса соответствующих плоских волн.

При $z=0$ имеем:

$$U(x, y, z=0) = U_0(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\alpha, \beta) \exp[ik(\alpha x + \beta y)] d\alpha d\beta. \tag{1.9}$$

Следовательно

$$C(\alpha, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(x, y) \exp[-ik(\alpha x + \beta y)] dx dy. \tag{1.10}$$

Далее, подставив (1.10) в (1.9), получим

$$\begin{aligned}
U(x, y, z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(x', y') \exp[-ik(\alpha x' + \beta y')] dx' dy' \right\} \exp[ikz\sqrt{1-\alpha^2-\beta^2}] \exp[ik(\alpha x + \beta y)] d\alpha d\beta = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(x', y') \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[ikz\sqrt{1-\alpha^2-\beta^2}] \exp[ik(\alpha(x-x') + \beta(y-y'))] d\alpha d\beta \right\} dx' dy' = \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(x', y') H(x-x', y-y', z) dx' dy'
\end{aligned} \tag{1.11}$$

где
$$H(x-x', y-y', z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[ikz\sqrt{1-\alpha^2-\beta^2}] \exp[ik(\alpha(x-x') + \beta(y-y'))] d\alpha d\beta$$

Интеграл (1.11) является конечной целью. Он связывает комплексную амплитуду

$U_0(x, y)$ в плоскости $z=0$ и комплексную амплитуду $U(x, y, z)$ в плоскости z (рис. 4).

Сферическая волна

Рассмотрим комплексную амплитуду сферической волны

$$S(x, y, z) = A \frac{\exp\left(ik\sqrt{x^2 + y^2 + z^2}\right)}{\sqrt{x^2 + y^2 + z^2}} \stackrel{R=\sqrt{x^2+y^2+z^2}}{=} A \frac{\exp(ikR)}{R}. \quad (2.1)$$

которая также является частным решением уравнения Гельмгольца (1.3), (что мы покажем на практике), $R = \sqrt{x^2 + y^2 + z^2}$ - расстояние от начала координат до текущей точки. Волновой фронт представляет собой поверхности постоянной фазы $kR = 2\pi q - \arg\{A\}$, считая для простоты, что $\arg\{A\} = 0$, можно записать:

$$R = \lambda q, \quad q - \text{целое.}$$

Это уравнение концентрических сфер, разделенных по радиальному расстоянию на величину λ (длина волны). Интенсивность $I(x, y, z) = \frac{|A|^2}{R^2}$ - обратно пропорциональна квадрату расстояния.

Разложение сферической волны по плоским волнам имеет вид (будет показано на практике)

$$\begin{aligned} \frac{\exp(ikR)}{R} &= \iint_{-\infty}^{\infty} \frac{1}{ik\sqrt{1-\alpha^2-\beta^2}} F_{\alpha,\beta}(x, y, z) d\alpha d\beta = \\ &= -2\pi \iint_{-\infty}^{\infty} \frac{\exp\left(ikz\sqrt{1-\alpha^2-\beta^2}\right)}{ik\sqrt{1-\alpha^2-\beta^2}} \exp[ik(\alpha x + \beta y)] d\alpha d\beta \end{aligned} \quad (2.8)$$

Ранее мы получили связь комплексной амплитуды поля $U_0(x, y)$ в плоскости $z=0$ с комплексной амплитудой $U(x, y, z)$ в плоскости z (рис. 4):

$$U(x, y, z) = \iint_{-\infty}^{\infty} U_0(x', y') H(x-x', y-y', z) dx' dy'$$

где
$$H(x, y, z) = \iint_{-\infty}^{\infty} \exp\left[ikz\sqrt{1-\alpha^2-\beta^2}\right] \exp[ik(\alpha x + \beta y)] d\alpha d\beta \quad (2.9)$$

Учитывая связь правых частей формул (2.8) и (2.9), можно показать (будет на практике), что

$$H(x, y, z) = -\frac{z \exp(ikR)}{2\pi R^2} \left(ik - \frac{1}{R} \right) \quad (2.10)$$

Т.е. получен более простой, чем для разложения по плоским волнам, вид оператора распространения.

Условие сведения интеграла связывающего комплексную амплитуду поля $U_0(x,y)$ в плоскости $z=0$ с комплексной амплитудой $U(x,y,z)$ в плоскости z с использованием плоских волн к интегралу Кирхгофа (интегралу сферических волн) следующее:

$$k \gg \frac{1}{R} \quad \text{или} \quad \frac{2\pi}{\lambda} \gg \frac{1}{R} \Rightarrow R \gg \lambda \quad (2.11)$$

Тогда $H(x, y, z) = -\frac{ikz}{2\pi R} \frac{\exp(ikR)}{R}$ и получаем:

$$U(x, y, z) = \iint_{-\infty}^{\infty} U_0(x', y') H(x-x', y-y', z) dx' dy' = -\frac{ik}{2\pi} \iint_{-\infty}^{\infty} U_0(x', y') \frac{\exp(ikR)}{R} \frac{z}{R} dx' dy', \quad (2.12)$$

где $R = \sqrt{(x-x')^2 + (y-y')^2 + z^2}$.

Параболические волны

Рассмотрим сферическую волну:

$$S(x, y, z) = A \frac{\exp\left(ik\sqrt{x^2 + y^2 + z^2}\right)}{\sqrt{x^2 + y^2 + z^2}}$$

в точке (x, y, z) , достаточно близкой к оси z , но далекой от начала координат (см. Рис. 6).

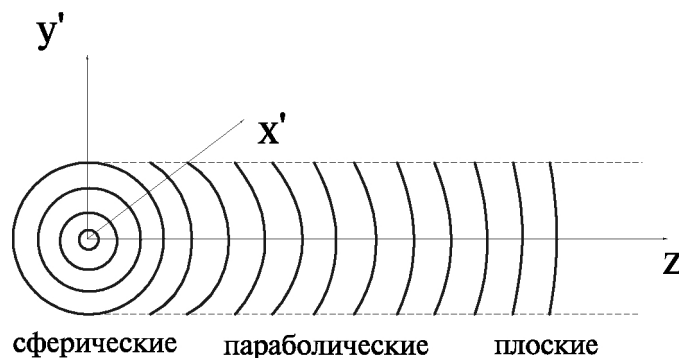


Рис. 6

В этом случае можно считать, что $\sqrt{x^2 + y^2} \ll z$. Тогда

$$R = \sqrt{x^2 + y^2 + z^2} = z\sqrt{1 + \frac{x^2 + y^2}{z^2}} = z\sqrt{1 + \theta}, \quad \text{где } \theta = \frac{x^2 + y^2}{z^2} \ll 1$$

можно разложить в ряд Тейлора:

$$R = z\sqrt{1+\theta} = z\left(1 + \frac{\theta}{2} + \frac{\theta^2}{8} + \dots\right) \approx z\left(1 + \frac{\theta}{2}\right) = z + \frac{x^2 + y^2}{2z} \quad (2.14)$$

Подставляя в сферической волне в фазу $R = z + \frac{x^2 + y^2}{2z}$, а в амплитуду $R=z$ (в фазу подставляем более точную аппроксимацию, так как она гораздо сильнее меняется от z), получаем:

$$P(x, y, z) \approx \frac{A}{z} \exp(ikz) \exp\left(ik \frac{x^2 + y^2}{2z}\right) \quad (2.15)$$

аппроксимацию Френеля сферической волны – параболическую волну. Приближение Френеля играет важную роль в теории дифракции.

Параксиальное (параболическое) волновое уравнение.

При рассмотрении пучков когерентного света от лазера, мод свободного пространства или гауссовых пучков, или любых других световых полей с небольшой расходимостью всегда можно выделить из комплексной амплитуды быстро меняющуюся от z составляющие. Считаем что ось z выбрана вдоль направления распространения:

$$U(x, y, z) = E(x, y, z) \exp(ikz). \quad (3.1)$$

где $E(x, y, z)$ - медленная меняющаяся от z составляющая, $\exp(ikz)$ - быстро меняющаяся от z составляющая, k - волновое число.

Если далее подставить (3.1) в уравнение Гельмгольца (1.3) (будет рассмотрено на практике):

$$[\nabla^2 + k^2]U(x, y, z) = 0,$$

то получим уравнение для медленно меняющейся по z амплитуды $E(x, y, z)$:

$$\nabla_{\perp}^2 E(x, y, z) + 2ik \frac{\partial E(x, y, z)}{\partial z} + \frac{\partial^2 E(x, y, z)}{\partial z^2} = 0 \quad (3.2)$$

где $\nabla_{\perp}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ - поперечный лапласиан.

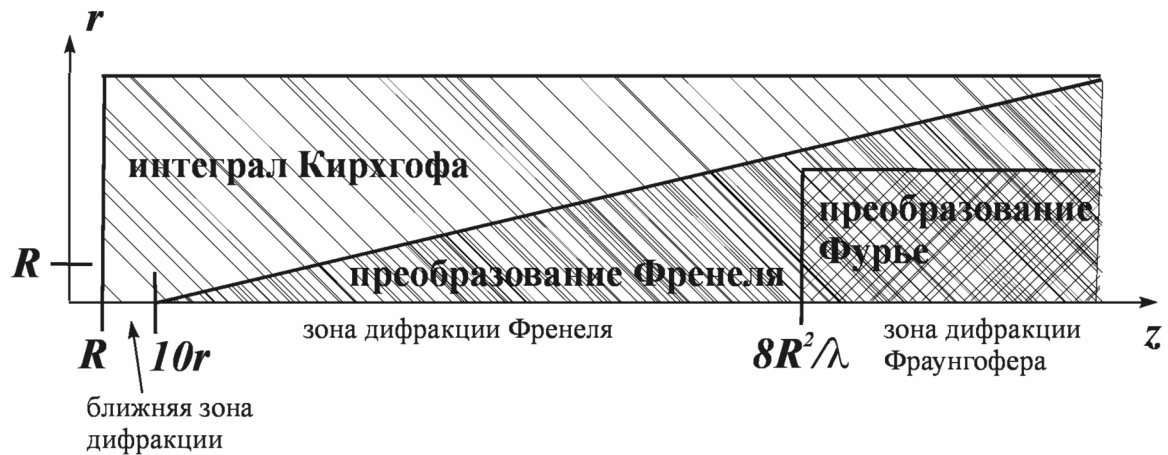
Пусть медленно меняющаяся амплитуда удовлетворяет условию

$$\left| \frac{\partial^2 E(x, y, z)}{\partial z^2} \right| \ll 2k \left| \frac{\partial E(x, y, z)}{\partial z} \right| \quad (3.3)$$

тогда вместо (3.2) получим параболическое (или параксиальное) уравнение распространения света

$$\left[\nabla_{\perp}^2 + 2ik \frac{\partial}{\partial z} \right] E(x, y, z) = 0. \quad (3.4)$$

Операторы распространение в свободном пространстве.



1) Разложение по плоским волнам (точное)

$$F(u, v, z) = \iint_{-\infty}^{\infty} f(x, y) \left\{ \iint_{-\infty}^{\infty} \exp(ikz\sqrt{1-\alpha^2-\beta^2}) \exp(ik[\alpha(u-x) + \beta(v-y)]) d\alpha d\beta \right\} dx dy$$

выполняется при условии $\alpha^2 + \beta^2 < 1$.

М.б. реализовано через два преобразования Фурье:

$$\begin{aligned} F(u, v, z) &= \iiint_{-\infty}^{\infty} \iiint_{-\infty}^{\infty} f(x, y) \exp(-ik[\alpha x + \beta y]) \exp(ikz\sqrt{1-\alpha^2-\beta^2}) \exp(ik[\alpha u + \beta v]) d\alpha d\beta dx dy = \\ &= \iint_{-\infty}^{\infty} \left\{ \iint_{-\infty}^{\infty} f(x, y) \exp(-ik[\alpha x + \beta y]) dx dy \right\} \exp(ikz\sqrt{1-\alpha^2-\beta^2}) \exp(ik[\alpha u + \beta v]) d\alpha d\beta = \\ &= \mathfrak{I}^{-1} \left\{ g(\alpha, \beta) \cdot \exp(ikz\sqrt{1-\alpha^2-\beta^2}) \right\} \quad \text{где } g(\alpha, \beta) = \mathfrak{I} \{ f(x, y) \} \end{aligned}$$

2) Разложение по сферическим волнам (интеграл Кирхгофа)

(точное)
$$F(u, v, z) = -\frac{z}{2\pi} \iint_{-\infty}^{\infty} f(x, y) \frac{\exp(ikR)}{R^2} \left(ik - \frac{1}{R} \right) dx dy, \quad \text{где } R = \sqrt{(x-u)^2 + (y-v)^2 + z^2}$$

выполняется при условии $\alpha^2 + \beta^2 < 1$.

При $k \gg \frac{1}{R}$

$$F(u, v, z) = -\frac{ik}{2\pi} \iint_{-\infty}^{\infty} f(x, y) \frac{\exp(ikR)}{R} \frac{z}{R} dx dy$$

3) Разложение по параболическим волнам (преобразование Френеля):

$$F(u, v, z) = -\frac{ik}{2\pi z} \exp(ikz) \iint_{-\infty}^{\infty} f(x, y) \exp\left[\frac{ik}{2z}((x-u)^2 + (y-v)^2)\right] dx dy,$$

выполняется при условии $\sqrt{(x-u)^2 + (y-v)^2} \ll z$.

М.б. реализовано через преобразование Фурье:

$$\begin{aligned} F(u, v, z) &= -\frac{ik}{2\pi z} \exp(ikz) \exp\left[\frac{ik}{2z}(u^2 + v^2)\right] \iint_{-\infty}^{\infty} f(x, y) \exp\left[\frac{ik}{2z}(x^2 + y^2)\right] \exp\left[-\frac{ik}{z}(xu + yv)\right] dx dy = \\ &= \Im\left\{f(x, y) \exp\left[\frac{ik}{2z}(x^2 + y^2)\right]\right\} \end{aligned}$$

4) Преобразование Фурье:

$$G(u, v) = \frac{k}{f} \iint_{-\infty}^{\infty} g(x, y) \exp\left[-\frac{ik}{f}(xu + yv)\right] dx dy = \Im\{g(x, y)\}$$

1D случай

Точное:

$$F(u, z) = \int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{\infty} \exp(ikz\sqrt{1-\alpha^2}) \exp[ik\alpha(u-x)] d\alpha \right\} dx, \quad \alpha^2 < 1$$

или

$$F(u, z) = \frac{ikz}{2} \int_{-\infty}^{\infty} f(x) \frac{H_1^{(1)}(kR)}{R} dx \xrightarrow{kR \rightarrow \infty} z \sqrt{-\frac{ik}{2\pi}} \int_{-\infty}^{\infty} f(x) \frac{\exp(ikR)}{R\sqrt{R}} dx, \quad R = \sqrt{(x-u)^2 + z^2},$$

$H_1^{(1)}(x)$ - функция Ханкеля

Интеграл Френеля:

$$F(u, z) = \sqrt{-\frac{ik}{2\pi z}} \exp(ikz) \int_{-\infty}^{\infty} f(x) \exp\left[\frac{ik}{2z}(x-u)^2\right] dx, \quad x-u \ll z$$

Преобразование Фурье:

$$G(u) = \sqrt{\frac{k}{f}} \int_{-\infty}^{\infty} g(x) \exp\left[-\frac{ik}{f}xu\right] dx = \Im\{g(x)\}$$

Итерационный подход к расчету ДОЭ

Развитие данного подхода связано с постоянным повышением быстродействия средств вычислительной техники, а также с высоким качеством ДОЭ, рассчитанных с помощью данного подхода. Хорошо исследованы вопросы, связанные с условиями сходимости и применимости итерационных процедур. На рис. 1 представлена в самом общем виде схема организации итерационного процесса расчета функции пропускания ДОЭ:

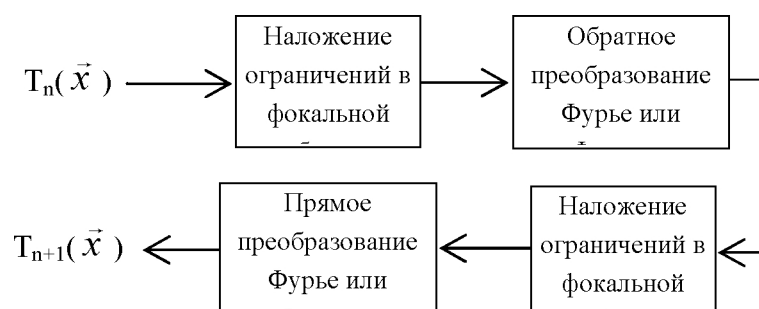


Рис. 1 Схема организации итерационного процесса

Математическим обоснованием построения итерационного процесса расчета дифракционных оптических элементов служит, как правило, либо последовательное построение проекций на замкнутые множества, соответствующие ограничениям, накладываемым на функцию пропускания элемента в фокальной плоскости и плоскости элемента (H.Stark, A.Levi, 1984), либо широко известная градиентная процедура поиска функции (Воронцов, Сивоконь).

В наиболее простом и часто используемом случае, когда объект задан своим амплитудным распределением, для расчета голографического элемента может быть применена итерационная процедура, известная как процедура уменьшения ошибки Фьенапа (J.R. Fienup, 1984). 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 невозрас-

тания ошибки.

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

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

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

1. 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.

Свет считается монохроматическим и когерентным, если его можно описать некоторой комплексной функцией, удовлетворяющей уравнению Гельмгольца.

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 рамках скалярной теории дифракции комплексная амплитуда света в

плоскости оптического элемента

$$W(u, v) = A(u, v)e^{ip(u, v)}$$

связана с комплексной амплитудой света

$$F(\xi, \eta) = B(\xi, \eta)e^{Ap(\xi, \eta)}$$

в плоскости наблюдения, в которой формируется требуемое распределение интенсивности $I_0(\xi, \eta)$, через интегральное преобразование [2.70]

$$F(\xi, \eta) = \frac{ik}{2\pi z} e^{ikz} \iint_{-\infty}^{\infty} W(u, v) H(u - \xi, v - \eta, z) du dv, \quad (2.1)$$

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

это функция импульсного отклика свободного пространства в приближении Френеля, z - это расстояние между ДОЭ и плоскостью наблюдения, $k = 2\pi/\lambda$ - волновое число света для длины волны λ . 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 $W_0(u, v)$ освещающего света на собственную функцию пропускания $\tau(u, v)$ ДОЭ:

$$W(u, v) = W_0(u, v)\tau(u, v). \quad (2.3)$$

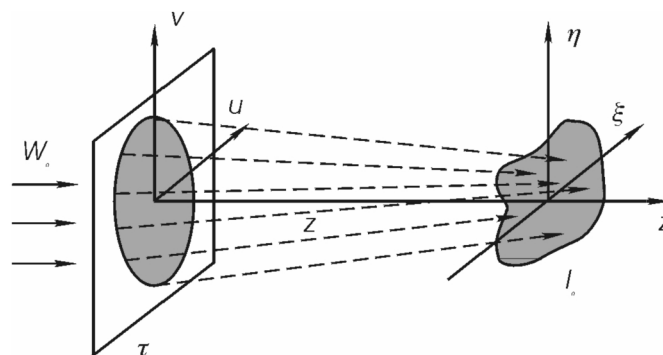


Рис. 2.1 Схема формирования изображения при помощи ДОЭ

Ввиду того, что в последующем мы предполагаем рассматривать только фазовые оптические элементы (если не указано иначе), функция пропускания ДОЭ выбрана в виде:

$$\tau(u, v) = e^{ig(u, v)}, \quad (2.4)$$

где $g(u, v)$ - заданная фаза ДОО. Для иллюстрации вышеуказанных обозначений следует обратиться к рис. 2.1. Задачу расчета фазовой функции ДОО $g(u, v)$ можно свести к решению нелинейного интегрального уравнения

$$I_0(\xi, \eta) = |F(\xi, \eta)|^2 = \left| \iint_{-\infty}^{\infty} A_0(u, v) e^{i\varphi(u, v)} H(u - \xi, v - \eta, z) du dv \right|^2, \quad (2.5)$$

где $I_0(\xi, \eta)$ - заданная интенсивность в области изображения, $A_0(u, v)$ - амплитуда освещающего пучка, $\varphi(u, v) = g(u, v) + g_0(u, v)$, где $g_0(u, v)$ - фаза освещающего пучка. Итеративный метод расчета фазы $\varphi(u, v)$, а в месте с ней фазы $g(u, v)$, состоит в решении уравнения (2.5) методом последовательных приближений. Алгоритм Герчберга-Секстона (ГС), или алгоритм уменьшения ошибки [2.14], включает следующие шаги (см. его схему на рис. 1):

- 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

$$\bar{F}(\xi, \eta) = B_0(\xi, \eta) F(\xi, \eta) |F(\xi, \eta)|^{-1}, \quad (2.6)$$

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

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

$$W(u, v) = \frac{ik}{2\pi z} e^{-ikz} \iint_{-\infty}^{\infty} \bar{F}(\xi, \eta) H^*(\xi - u, \eta - v, z) d\xi d\eta; \quad (2.7)$$

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

$$\bar{W}(u, v) = \begin{cases} A_0(u, v) W(u, v) |W(u, v)|^{-1}, & (u, v) \in Q, \\ 0, & (u, v) \notin Q \end{cases}, \quad (2.8)$$

где Q - форма апертуры ДОО;

б) переход к шагу 2).

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

$$\delta_F^2 = \frac{\iint_{-\infty}^{\infty} [F(\xi, \eta) - B_0(\xi, \eta)]^2 d\xi d\eta}{\iint_{-\infty}^{\infty} B_0^2(\xi, \eta) d\xi d\eta}, \quad (2.9)$$

$$\delta_W^2 = \frac{\iint_{-\infty}^{\infty} [W(u, v) - A_0(u, v)]^2 du dv}{\iint_{-\infty}^{\infty} A_0^2(u, v) du dv}. \quad (2.10)$$

Алгоритм ГС называют алгоритмом уменьшения ошибки потому, что было показано [4], что ошибки (2.9) и (2.10) 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

$$\varepsilon_0 = \iint_{-\infty}^{\infty} [F(\xi, \eta) - B_0(\xi, \eta)]^2 d\xi d\eta. \quad (2.11)$$

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 (2.11) 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 Алгоритм входа-выхода

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

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

$$F(\xi, \eta) = \iint_{-\infty}^{\infty} W(u, v) \exp \left[-i \frac{k}{f} (u\xi + v\eta) \right] du dv, \quad (2.12)$$

где f - фокальное расстояние тонкой линзы, формирующей Фурье-спектр.

Переход от преобразования Френеля, (2.1), (2.2), к преобразованию Фурье (2.12) означает, что в последнем случае следует использовать комбинацию ДОО плюс линза.

Итеративное решение проблемы восстановления фазы светового поля, т.е. аргумента функции $F(\xi,\eta)$ на основе известной амплитуды, т.е. модуля $|F(\xi,\eta)|$, при помощи алгоритма входа-выхода (одной из его модернизаций) имеет вид:

$$\bar{F}_{n+1}(\xi, \eta) = \bar{F}_n(\xi, \eta) + \beta \delta F_n(\xi, \eta), \quad (2.13)$$

где

$$\delta F_n(\xi, \eta) = \begin{cases} 0, & (\xi, \eta) \notin \gamma \\ -F_n(\xi, \eta), & (\xi, \eta) \in \gamma \end{cases} \quad (2.14)$$

$$F_n(\xi, \eta) = \mathfrak{Z} D_A \mathfrak{Z}^{-1} \{ \bar{F}_n(\xi, \eta) \} \quad (2.15)$$

является выходной функцией, $\mathfrak{Z}$ и $\mathfrak{Z}^{-1}$ - прямое и обратное Фурье-преобразования, D_A - ограничивающий оператор в области ДОО, который может быть представлен в виде:

$$D_A W_n = A_0 \frac{W_n}{|W_n|}, \quad (2.16)$$

где $0.5 \leq \beta \leq 1$ - некоторый параметр, γ - область нарушения ограничений, накладываемых на заданную функцию поля.

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

$$\delta F = [B_0 F |F|^{-1} - F] + [B_0 F |F|^{-1} - B_0 \bar{F} |\bar{F}|^{-1}], \quad (2.17)$$

где $B_0^2(\xi, \eta)$ - заданное распределение интенсивности в Фурье-плоскости, $\bar{F}$ и F - это 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 $(\bar{F}(\xi, \eta) = \bar{F})$:

$$\begin{cases} |\bar{F}_{n+1}| = B_0 + \beta [|F_n| - B_0] \\ \arg \bar{F}_{n+1} = \arg F_n + \alpha [\arg F_n - \arg \bar{F}_n] \end{cases}. \quad (2.18)$$

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 (2.18) 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 амплитуду в Фурье-плоскости B_0 .

3. 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 (2.18) (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 (2.6) для алгоритма Герчберга-Секстона. Эта замена предполагает, что амплитуда света $|F_n(\xi, \eta)|$ 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 (2.6) попробовать использовать замену, в которой обе функции (заданная и рассчитанная аналитическая) участвуют в виде линейной комбинации с разными весами [12]:

$$|\bar{F}_n(\xi, \eta)| = |\lambda B_0(\xi, \eta) + (1 - \lambda) F_n(\xi, \eta)|, \quad (2.19)$$

В этом случае замена (2.6) в итеративном ГС-алгоритме принимает вид:

$$\bar{F}_n(\xi, \eta) = |\bar{F}_n(\xi, \eta)| F_n(\xi, \eta) |F_n(\xi, \eta)|^{-1}. \quad (2.20)$$

Диапазон значений параметра λ находится из условия, что среднее отклонение ε_0 , определяемое соотношением (2.11), не увеличивается при использовании замены (2.19). То есть мы предполагаем, что выполняется следующее условие:

$$\begin{aligned} \bar{\varepsilon}_0 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [|\bar{F}_n| - B_0]^2 d\xi d\eta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\lambda B_0 + (1 - \lambda) |F_n| - B_0]^2 d\xi d\eta = \\ &= (1 - \lambda)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [|F_n| - B_0]^2 d\xi d\eta \leq \varepsilon_0 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F_n - B_0]^2 d\xi d\eta. \end{aligned} \quad (2.21)$$

Из уравнения (2.21) следует, что весовой коэффициент следует выбирать из условия $0 \leq \lambda \leq 2$. Заметим, что для $\lambda = 1$ замена (2.19) переходит в уравнение (2.6) алгоритма Герчберга-Секстона, в то время как $\lambda = 2$ приводит к замене “зеркального изображения”:

$$|\bar{F}_n(\xi, \eta)| = |2 B_0(\xi, \eta) - F_n(\xi, \eta)|. \quad (2.22)$$

В последнем случае ($\lambda = 2$) 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 (2.22) имела следующий вид [10, 11]:

$$\bar{I}_n(\xi, \eta) = \begin{cases} |2 I_0(\xi, \eta) - I_n(\xi, \eta)|, & (\xi, \eta) \in \Omega \\ I_n(\xi, \eta), & (\xi, \eta) \notin \Omega \end{cases}, \quad (2.23)$$

где I_n - 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 $I_n \geq 2I_0$, то $\bar{I}_n = 0$.

Основная литература

1. Дифракционная компьютерная оптика /Под ред. В.А. Сойфера, М.: Физматлит, 2007. <http://www.ipsi.smr.ru/Books/DKO/Titul.html>
2. Методы компьютерной оптики //Под ред. В.А. Сойфера, М.: Физматлит, 2003.
<http://www.ipsi.smr.ru/Books/MKO/TitulPol.html>
3. Хонина С.Н. Методы расчета дифракционных оптических элементов на основе функционального представления // Учебное пособие, Самара, СГАУ, ИСОИ РАН, 2006.
4. Хонина С.Н., Котляр В.В., Сойфер В.А. Лазерные модовые пучки с замечательными свойствами // Учебное пособие, Самара, СГАУ, 2006.

Дополнительная литература.

1. М. Борн, Э. Вольф "Основы Оптики ", М: Мир, 1989.
2. Зверев В.А. Радиооптика, М., Сов. радио, 1975
3. Ярославский Л.П., Мерзляков Н.С. Цифровая голография. – М.: Наука, 1982
4. Ландсберг Г. С. "Оптика", М: Мир, 1976.
5. Бобров С.Т., Грейсух Г.И, Туркевич Ю.Г "Оптика дифракционных элементов и систем", Ленинград, Машиностроение, 1986.
6. Слюсарев Г.Г. "Расчет оптических систем", Л.: Машиностроение, 1973.
7. Гудмен Дж. Введение в Фурье-оптику. М., Наука, 1973.
8. Информационная оптика, под ред. Н.Н. Евтихеева, М.: Изд. МЭИ, 2000.
9. А.В. Гончаровский, В.В. Попов, В.В. Степанов, Введение в компьютерную оптику, М.: Изд. МГУ, 1991.