

Информационно-аналитическая система моделирования химико-технологических процессов с использованием параллельных вычислений

И.В. Ахметов¹, И.М. Губайдуллин^{1,2}

¹Уфимский государственный нефтяной технический университет, Космонавтов 1, Уфа, Россия, 450062

²Институт нефтехимии и катализа РАН, проспект Октября 141, Уфа, Россия, 450075

Аннотация. В данной работе рассмотрено распараллеливание вычислительного процесса для решения обратных задач химической кинетики на трех уровнях: распараллеливание по экспериментальной базе; использование внутреннего параллелизма задачи; декомпозиция алгоритма решения обратной задачи.

1. Введение

При изучении сложных химических реакций возникают следующие трудности: 1) существует несколько гипотетических механизмов протекания реакции, из которых необходимо выбрать лучший в смысле некоторого критерия расхождения опытных и расчетных данных; 2) существует несколько экспериментов, проведенных при различных условиях и имеющих различную погрешность экспериментов, поэтому необходимо выбрать те опыты, по которым расчетные значения наиболее близки к экспериментальным; 3) каждый из кинетических параметров определяется неоднозначно, кроме того, существует зависимость решения от правильного выбора начального приближения, основанного на некоторых физико-химических предположениях [1].

Поэтому для последовательного решения всех этих задач иногда необходимы существенные затраты времени (от нескольких месяцев до года), что ставит вопрос о сокращении времени расчетов. В настоящее время, при энергичном развитии вычислительной техники возникла идея последовательно-параллельного проведения расчетов [2].

Участвующие в реакциях металлокомплексного катализа вещества часто имеют сложную структуру и представляют собой большие макромолекулярные комплексы. Натурные эксперименты для таких процессов проводятся в несколько взаимосвязанных этапов с расщеплением на независимые частные реакции. Для полного понимания природы взаимодействия веществ, участвующих в реакциях металлокомплексного катализа, необходимо рассмотреть большое количество олефинов, ацетиленов, алленов, спиртов. Чтобы установить более достоверный механизм реакции приходится моделировать многочисленные варианты предполагаемых механизмов, которые включают в себя большое количество параллельных стадий, как в виде итоговых уравнений, так и в виде уравнений элементарных стадий. Параллельное изучение подобных сложных механизмов на основе натуральных и вычислительных экспериментов требует обработки большого количества информации. Определение параметров

кинетических моделей, варьирование входных данных при проведении вычислительных экспериментов на основе этих моделей относится к классу многопараметрических задач. На современном этапе такие задачи целесообразно решать с использованием технологии параллельных вычислений, позволяющей обрабатывать большие объемы данных и вести параллельный расчет при решении обратных задач химической кинетики на многопроцессорных вычислительных системах для независимых между собой реакций [3].

Многим реальным многопараметрическим задачам, для решения которых необходимо использование вычислительной техники, свойственен естественный внутренний параллелизм, то есть возможность в той или иной форме распараллелить действия, связанные с решением задачи, на качественном уровне. Решение многопараметрических обратных задач при построении кинетических моделей сложных реакций металлокомплексного катализа представляет собой многоуровневый, многократно вложенный, последовательно-параллельный процесс.

Существуют разные способы организации параллельных вычислений. Основная последовательность действий для определения эффективных способов организации параллельных вычислений состоит в следующем:

- анализ исследуемого объекта на внутренний параллелизм;
- анализ экспериментальной информации, проектирование баз данных и реализация параллельного доступа к ним;
- анализ используемых вычислительных схем и осуществление их разделения (декомпозиции) на части (подзадачи), которые могут быть реализованы в значительной степени независимо друг от друга;
- выделение для выделенного набора подзадач информационных взаимодействий, которые должны осуществляться в ходе решения исходной поставленной задачи;
- распределение выделенного набора подзадач между процессорами используемой вычислительной системы.

Очевидно, объем вычислений для каждого используемого процессора должен быть примерно одинаковым, что позволит обеспечить равномерную вычислительную загрузку процессоров, а распределение подзадач между процессорами должно быть выполнено таким образом, чтобы количество информационных взаимодействий между подзадачами было минимальным. Также обязательными требованиями являются целостность, универсальность и гибкость разрабатываемых баз данных.

Наиболее простым способом организации параллельных вычислений является распараллеливание по экспериментальной базе, т.е. распределение однотипных вычислений для разных наборов начальных данных. Лучше всего подходят для распараллеливания задачи, обладающие внутренним параллелизмом, например, включающие протекающие параллельно процессы. Наиболее эффективным, но и трудоемким способом распараллеливания является распараллеливание численных методов и алгоритмов решения задач. Рассматриваемые задачи химической кинетики относятся к классу задач, решение обратной задачи в которых осуществляется на основе решения серии прямых задач с минимизацией выбранного критерия отклонения расчетных данных от экспериментальных. Это создает основу для эффективного распараллеливания вычислительного процесса с использованием принципа геометрического параллелизма. Таким образом, распараллеливание вычислительного процесса может быть осуществлено на трех уровнях:

- распараллеливание по экспериментальной базе;
- использование внутреннего параллелизма задачи;
- декомпозиция алгоритма решения обратной задачи.

Имеется трехуровневая модель распараллеливания, сочетающая все эти подходы (рисунок 1) [4,5]:

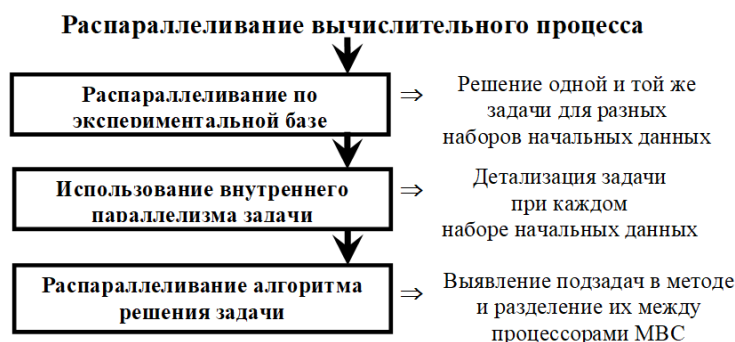


Рисунок 1. Трехуровневая модель распараллеливания.

2. Трехуровневая модель распараллеливания вычислительного процесса при решении обратной задачи

2.1. Распараллеливание по экспериментальной базе

Решение многопараметрических задач, таких как задача идентификации кинетической модели сложных химических реакций, состоит в циклическом чередовании натурального и вычислительного экспериментов. Натурный эксперимент требует больших материальных и энергетических затрат, особенно для сложных реакций, и его информативность напрямую зависит от качества и скорости проведения вычислительного эксперимента. Объем данных вычислительного эксперимента зависит от следующих факторов:

- для каждой реакции необходимо обработать несколько предполагаемых схем химических превращений, чтобы выбрать лучшую схему;
- для каждой реакции проводится ряд натуральных экспериментов при разных условиях; при обработке опытных данных выбираются несколько лучших, по которым расчетные значения имеют наименьшее отклонение от экспериментальных данных;
- недостаточная информативность эксперимента, что является одной из причин неоднозначного определения кинетических параметров.

Поскольку размерность систем обыкновенных дифференциальных уравнений (ОДУ) достигает 30 уравнений, при решении обратной задачи необходимо решать более 3 миллионов систем нелинейных ОДУ (рисунок 2), что определяет целесообразность использования технологии параллельных вычислений.

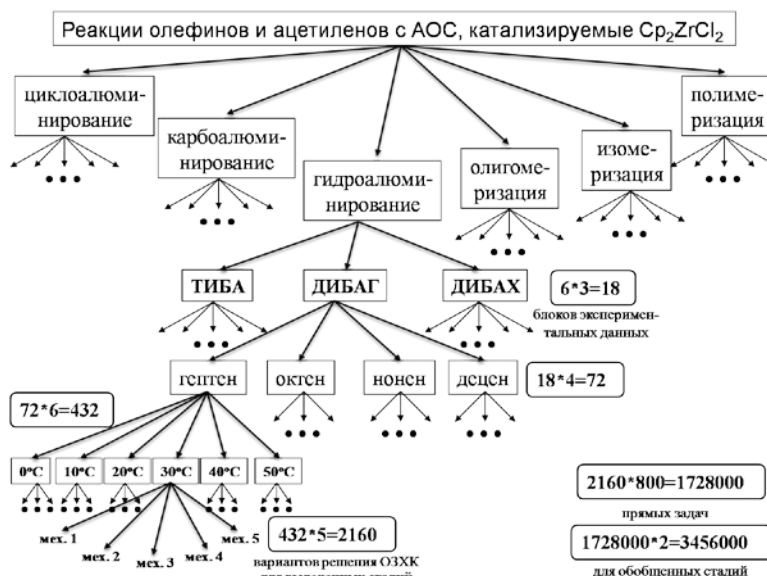


Рисунок 2. Число решаемых прямых задач.

Таким образом, на первом уровне все множество процессоров многопроцессорной вычислительной системы разбивается на подмножества для решения обратной задачи при конкретном наборе начальных данных. При этом организация взаимодействия с базой данных осуществляется по принципу master-slave, при котором выбирается один главный процессор, имеющий доступ к базе данных и выполняющий распределение данных между всеми подчиненными процессорами [6,7]. Такая организация доступа к базе данных (БД) позволяет избежать конфликтов при доступе к памяти, а также «гонок данных», когда результаты взаимодействия с БД зависят от того, в каком порядке процессоры получили к ней доступ.

2.2. Использование внутреннего параллелизма задачи

На втором уровне распараллеливания каждое подмножество процессоров относят к различным коммуникаторам в соответствии с внутренним параллелизмом задачи, который для рассматриваемой задачи заключается в возможности независимого решения задачи для выделенных, частных, детализированных и общих реакций. Распараллеливание на основе использования внутреннего параллелизма задачи при построении кинетической модели реакции гидроалюминирования олефинов с алюминий-органическими соединениями (АОС) иллюстрирует рисунок 3.

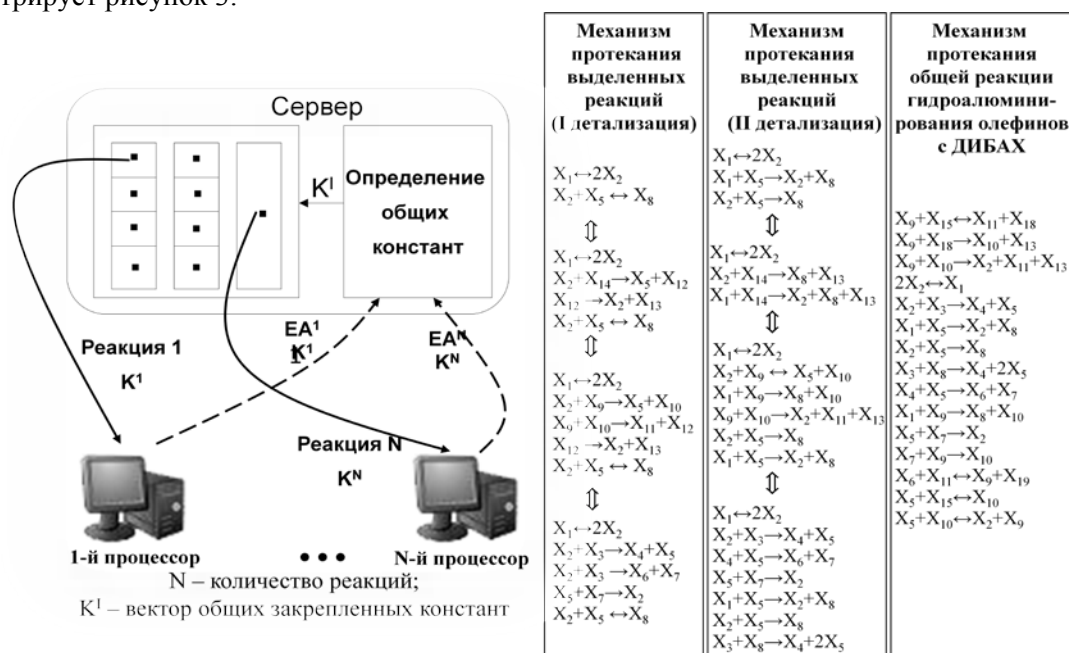


Рисунок 3. Второй уровень распараллеливания решения обратной задачи химической кинетики.

3. Распараллеливание алгоритма решения задачи

Третий уровень распараллеливания включает декомпозицию алгоритма решения обратной задачи по числу процессоров, входящих в созданные процессорные коммуникаторы.

3.1. Генетический алгоритм решения обратной задачи

Для параллельного решения обратной задачи химической кинетики наиболее эффективным является генетический алгоритм, основу которого составляет заимствованная из биологии идея селекции, то есть преимущественного размножения наиболее приспособленных особей [8]. Практическое применение генетического алгоритма во всех известных случаях приводило к положительным результатам [9,10]. Показано, что генетический алгоритм, в отличие от градиентных методов минимизации, является универсальным методом для поиска оптимума независимо от сложности функций [10]. Последовательность операций, составляющих основу генетического алгоритма, следующая.

На первом шаге алгоритма случайным образом создается начальная популяция, состоящая из N особей (N точек в пространстве кинетических параметров, каждая точка имеет m координат – значений параметров).

На этапе мутации особи популяции изменяются в соответствии с заранее определенной операцией мутации, в качестве которой был взят покоординатный/параболический спуск из точек пространства.

На этапе селекции из всей популяции выбирается определенная ее доля, которая останется «в живых» на этом этапе эволюции. Вероятность выживания особи зависит от значения функции приспособленности для этой особи; в качестве функции приспособленности выступает функционал невязки. Доля выживших s является параметром генетического алгоритма, и по итогам отбора из N особей популяции в итоговую популяцию войдут sN особей. В рассматриваемом случае $s=1/2$.

При формировании нового поколения используется скрещивание – чтобы произвести потомка, нужны два родителя. Для формирования новой точки в пространстве параметров в качестве родителей выбирается одна точка из «выживших» и одна из «погибающих», и скрещивание производится путем выбора $m/2$ координат от первой точки и оставшихся – от второй; при этом потомок наследует черты обоих родителей. Особи для размножения выбираются из всей популяции, а не из выживших на первом шаге элементов с целью исключения возможности деградации популяции.

Этот набор действий повторяется итеративно, так моделируется «эволюционный процесс», продолжающийся несколько жизненных циклов (поколений), пока не будет выполнен критерий остановки алгоритма, в качестве которого выступает любое из условий:

- нахождение глобального либо субоптимального решения;
- исчерпание числа поколений, отпущенных на эволюцию;
- исчерпание времени, отпущенного на эволюцию.

Распараллеливание вычислительного процесса производится на стадии начального заполнения, когда заданные псевдослучайно точки в пространстве параметров равномерно распределяются по процессам МВС. Мутация осуществляется каждым процессом независимо; обмен данными производится на этапе селекции. При этом время автономной работы процессов значительно превышает время межпроцессорных взаимодействий, что обуславливает эффективность данного алгоритма.

Оценка эффективности распараллеливания при тестировании программы на вычислительном кластере суперкомпьютера МВС-100К показала, что параллельная программа работает достаточно эффективно при увеличении числа процессов. При решении обратной задачи для реакции гидроалюминирования олефинов по всем экспериментальным данным общее время расчета составило: на персональном компьютере – 360 часов, на суперкомпьютере МВС-100К (с использованием 120 процессоров Intel Xeon E5450) – 15 минут.

3.2. Геометрический параллелизм по кинетическим параметрам

Принцип геометрического параллелизма предполагает декомпозицию расчетной области пространства кинетических параметров на подобласти соответственно числу процессоров МВС. Для двух констант строится двумерная плоскость (для n констант – соответственно, n -мерная). Далее эта область делится на подобласти по числу процессоров, и каждый процессор решает обратную задачу только в своей подобласти.

Определив наборы констант, соответствующие минимальному значению функционала в заданной подобласти, каждый процессор передает найденные значения серверу. Сервер исключает подобласти, в которых функционал имеет наибольшие значения, и из полученной области снова формирует подобласти по количеству процессоров. Данный циклический процесс продолжается до тех пор, пока не будут определены константы, описывающие эксперимент с заданной точностью. При этом каждый процессор минимизирует невязку в своей области методом покоординатного или параболического спуска (рис. 4). Здесь L – количество подобластей [11, 12].

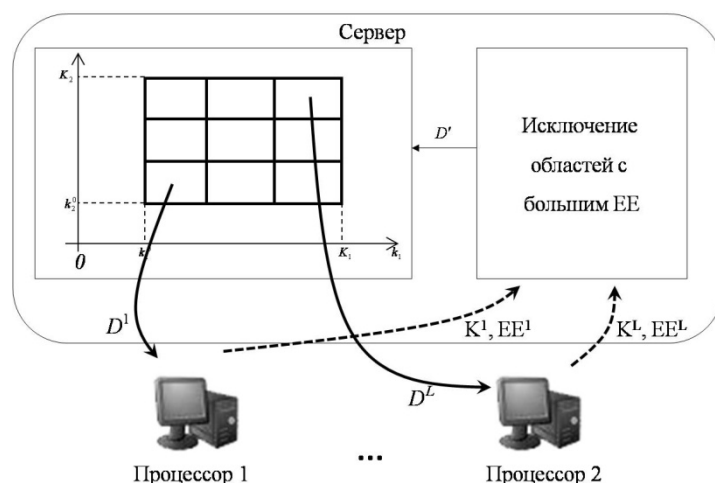


Рисунок 4. Геометрический параллелизм по кинетическим параметрам.

Благодаря использованию значительного вычислительного ресурса становится возможным резко увеличить количество подобластей, и, соответственно, с требуемой точностью решить поставленную задачу без риска остановиться на локальном минимуме оптимизируемого функционала.

4. Выводы

Выявлен и рассмотрен внутренний параллелизм задачи на примере реакции гидроалюминирования олефинов, заключающийся в том, что общая схема реакции включает протекающие параллельно стадии, некоторые из них совпадают для реакций с алюминийорганическими соединениями и олефинами. При осуществлении вычислительного процесса решение обратной задачи осуществляется независимо для параллельно протекающих стадий, что позволяет существенно сократить время расчета.

Для организации вычислительного процесса рассмотрена трехуровневая модель распараллеливания, объединяющая распараллеливание по экспериментальной базе, в соответствии с внутренним параллелизмом задачи и на основе декомпозиции метода решения обратной задачи.

Программный комплекс, реализующий расчеты по этой модели, протестирован на суперкомпьютере МВС-100К Межведомственного суперкомпьютерного центра РАН. При решении обратной задачи для реакции гидроалюминирования олефинов по всем экспериментальным данным на основе генетического алгоритма общее время расчета на суперкомпьютере МВС-100К (с использованием 120 процессоров Intel Xeon E5450) составило 15 минут против 360 часов на персональном компьютере с использованием стандартных алгоритмов.

5. Литература

- [1] Ахметов, И.В. Разработка кинетической модели реакции получения метилового эфира 5-ацетил-2-пирролкарбоновой кислоты / И.В. Ахметов, И.М. Губайдуллин, А.В. Балаев // Журнал Средневолжского математического общества. – 2010. – Т. 12, № 3. – С. 50-54.
- [2] Ахметов, И.В. Разработка кинетических моделей с использованием параллельных вычислений на многоядерных системах / И.В. Ахметов, И.М. Губайдуллин // Вестник Омского университета. – 2012. – № 2(64). – С. 172-174.
- [3] Ахметов, И.В. Построение кинетических моделей химических реакций на основе многоядерных систем / И.В. Ахметов, И.М. Губайдуллин // Журнал Средневолжского математического общества. – 2012. – Т. 14, № 3. – С. 38-42.

- [4] Линд, Ю.Б. Методология параллельных вычислений для решения задач химической кинетики и буровой технологии / Ю.Б. Линд, И.М. Губайдуллин, Р.А. Мулюков // Системы управления и информационные технологии. – 2009. – №2/36. – С. 44-49.
- [5] Губайдуллин, И.М. Современные технологии высокопроизводительных вычислений при моделировании детального механизма реакции каталитического гидроалюминирования олефинов / И.М. Губайдуллин, К.Ф. Коледина, Ю.Б. Линд // Наука и образование. – 2011. – № 6. – URL: <http://technomag.edu.ru/doc/187631.html> (дата обращения: 20.11.2011).
- [6] Линд, Ю.Б. О применении параллельных вычислительных технологий при нахождении кинетических параметров общего механизма Zr катализа в реакциях карбо-, гидро- и циклометаллирования олефинов в присутствии катализатора $\text{Cr}_2\text{ZrCl}_{12}$ / Ю.Б. Линд, И.М. Губайдуллин, Л.В. Парфенова, М.Д. Рамазанов, С.И. Спивак // Параллельные вычислительные технологии (ПаВТ'2007): Труды международной научной конференции. – Челябинск: Издательский центр ЮУрГУ, 2007. – С. 128-133.
- [7] Губайдуллин, И.М. Реляционная система управления базой данных для реакции гидроалюминирования олефинов в присутствии циркониевого катализатора, реализующая динамическое распределение данных между процессорами многопроцессорной вычислительной системы / И.М. Губайдуллин, Ю.Б. Линд, Э.Р. Ахматсафина, С.И. Спивак // Параллельные вычислительные технологии (ПаВТ'2008): Труды международной научной конференции. – Челябинск: Издательский центр ЮУрГУ, 2008. – С. 370-375.
- [8] Holland, J.H. Adaptation in natural and artificial systems / J.H. Holland // University of Michigan Press, Ann Arbor. – 1975. – 96 p.
- [9] Никитин, А.В. Эволюционная модель оптимизации модульной ассоциативной памяти для машин потока данных на основе генетического алгоритма / А.В. Никитин, Л.И. Никитина // Программирование. – 2002. – № 6. – С. 31-42.
- [10] Чернышев, О. Сравнительный анализ решения задач оптимизации генетическими и градиентными методами / О. Чернышев, А. Борисов // Transport and Telecommunication. – 2007. – Vol. 8(1). – P. 40-52.
- [11] Линд, Ю.Б. Параллельные вычисления при решении обратных задач физической химии / Ю.Б. Линд, И.М. Губайдуллин, М.Д. Рамазанов // Параллельные вычислительные технологии (ПаВТ'2010): Труды международной научной конференции. – Челябинск: Издательский центр ЮУрГУ, 2010. – С. 507-518.
- [12] Линд, Ю.Б. Параллельные вычисления при построении кинетической модели реакции гидроалюминирования олефинов / Ю.Б. Линд, А.В. Аристархов, И.М. Губайдуллин // Научный сервис в сети Интернет: суперкомпьютерные центры и задачи. Труды Международной суперкомпьютерной конференции. – М.: Изд-во МГУ, 2010. – С. 231-237.

Information-analytical system for modeling chemical-technological processes using parallel computations

I.V. Akhmetov¹, I.M. Gubaydullin^{1,2}

¹Ufa State Petroleum Technological University, Kosmonavtov street 1, Ufa, Russia, 450062

²Institute of Petrochemistry and Catalysis of RAS, Prospect Oktyabrya 141, Ufa, Russia, 450075

Abstract. In this paper we consider parallelization of the computational process for solving inverse problems of chemical kinetics on three levels: parallelization on experimental basis; the use of internal concurrency problems; de-composition algorithm solving the inverse problem.

Keywords: modeling, chemical processes, optimization, high-performance computing.