

Анализ методов решения обратных задач химической кинетики с применением параллельных вычислений

И.В. Ахметов^{1,2}, И.М. Губайдуллин^{1,2}

ФГБУН Институт нефтехимии и катализа РАН¹, ФГБОУ ВО Уфимский государственный нефтяной технический университет²

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.

Ключевые слова: обратные задачи, кинетические константы, физическая химия, многопроцессорные системы, MPI.

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:

- 1) анализ исследуемого объекта на внутренний параллелизм;
- 2) анализ экспериментальной информации, проектирование баз данных и реализация параллельного доступа к ним;
- 3) анализ используемых вычислительных схем и осуществление их разделения (декомпозиции) на части (подзадачи), которые могут быть реализованы в значительной степени независимо друг от друга;
- 4) выделение для выделенного набора подзадач информационных взаимодействий, которые должны осуществляться в ходе решения исходной поставленной задачи;
- 5) 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:

- 1) распараллеливание по экспериментальной базе;
- 2) использование внутреннего параллелизма задачи;
- 3) декомпозиция алгоритма решения обратной задачи.

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



Рис.1. Трехуровневая модель распараллеливания вычислительного процесса при решении обратной задачи

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: 1) для каждой реакции необходимо обработать несколько предполагаемых схем химических превращений, чтобы выбрать лучшую схему; 2) 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; 3) 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 (рис. 2), что определяет целесообразность использования технологии параллельных вычислений.

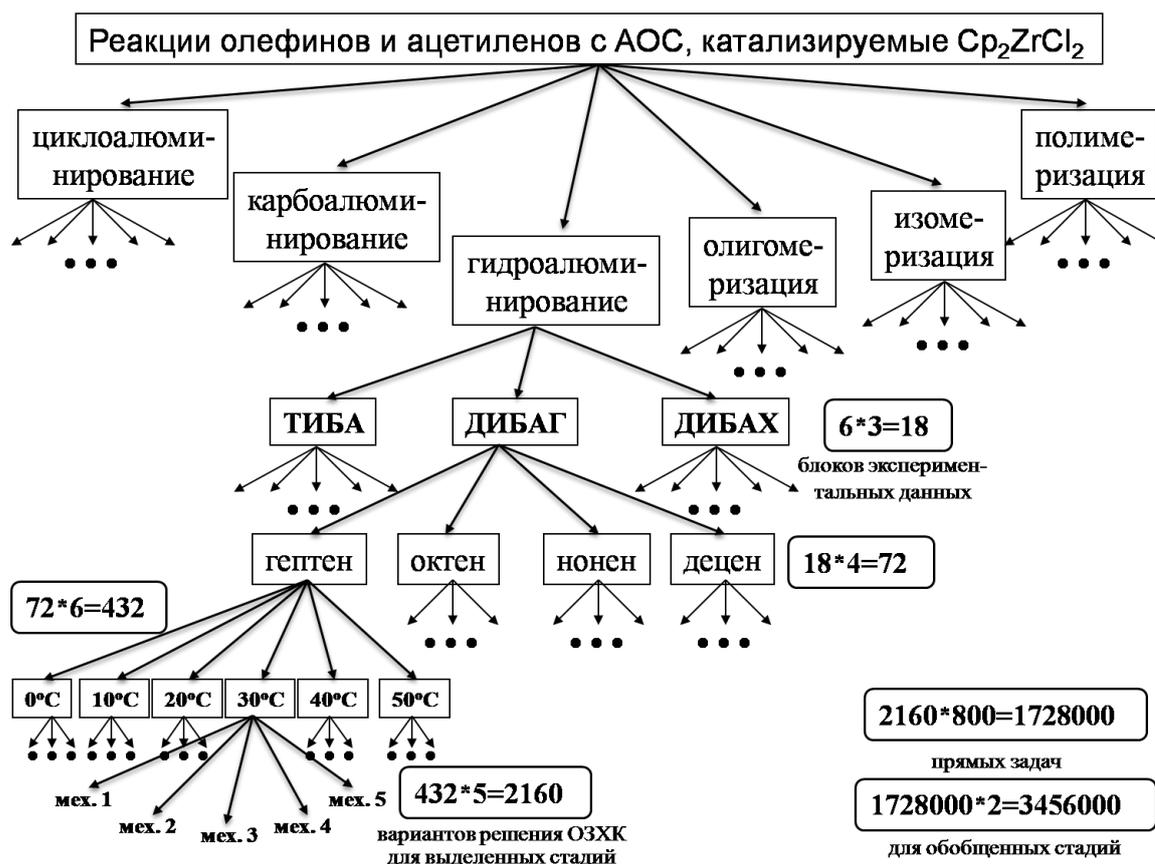


Рис. 2. Число решаемых прямых задач при построении кинетической модели реакции гидроалюминирования олефинов

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, име-

ющий 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 рис. 3.

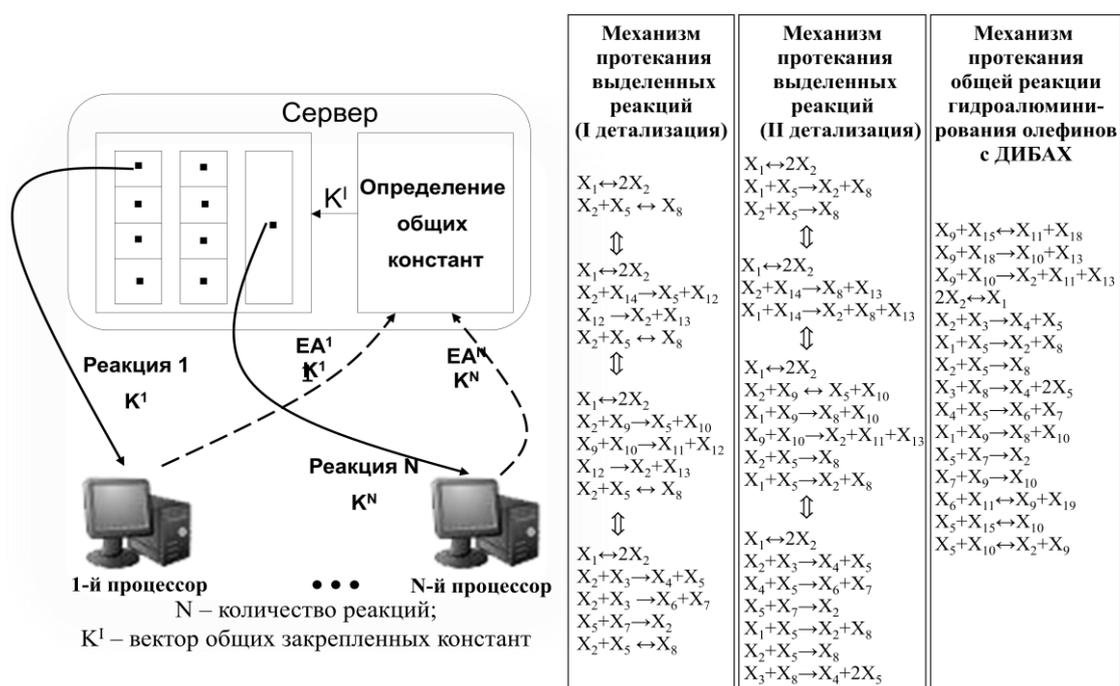


Рис. 3. Второй уровень распараллеливания решения обратной задачи химической кинетики

2.3 Распараллеливание алгоритма решения задачи

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

2.3.1 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: 1) нахождение глобального либо субоптимального решения; 2) исчерпание числа поколений, отпущенных на эволюцию; 3) 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 Башкирского государственного университета и 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) – 15 минут.

2.3.2 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].

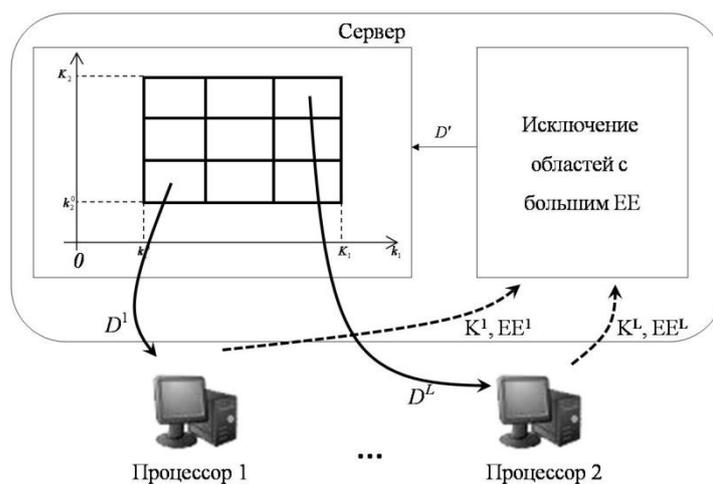


Рис. 4. Геометрический параллелизм по кинетическим параметрам

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.

3. Заключение

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, тестирован на вычислительном кластере Башкирского государственного университета и 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) составило 15 минут против 360 часов на персональном компьютере с использованием стандартных алгоритмов.

Литература

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

4. Линд Ю.Б., Губайдуллин И.М., Мулюков Р.А. Методология параллельных вычислений для решения задач химической кинетики и буровой технологии // Системы управления и информационные технологии. 2009. №2/36. С. 44–49.
5. Губайдуллин И. М., Коледина К. Ф., Линд Ю. Б. Современные технологии высокопроизводительных вычислений при моделировании детального механизма реакции каталитического гидроалюминирования олефинов // Наука и образование. 2011. № 6. URL: <http://technomag.edu.ru/doc/187631.html> (дата обращения: 20.11.2011).
6. Линд Ю.Б., Губайдуллин И.М., Парфенова Л.В., Рамазанов М.Д, Спивак С.И. О применении параллельных вычислительных технологий при нахождении кинетических параметров общего механизма Zr катализа в реакциях карбо-, гидро- и циклометаллирования олефинов в присутствии катализатора Cr_2ZrCl_2 . Параллельные вычислительные технологии (ПаВТ'2007): Труды международной научной конференции (Челябинск, 29 января - 2 февраля 2007 г.). Челябинск: Издательский центр ЮУрГУ, 2007. С. 128–133.
7. Губайдуллин И.М., Линд Ю.Б., Ахматсафина Э.Р., Спивак С.И. Реляционная система управления базой данных для реакции гидроалюминирования олефинов в присутствии циркониевого катализатора, реализующая динамическое распределение данных между процессорами многопроцессорной вычислительной системы. Параллельные вычислительные технологии (ПаВТ'2008): Труды международной научной конференции (Санкт-Петербург, 28 января - 1 февраля 2008 г.). Челябинск: Издательский центр ЮУрГУ, 2008. С. 370–375.
8. Holland J. H. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor. 1975. 96 p.
9. Никитин А.В., Никитина Л.И. Эволюционная модель оптимизации модульной ассоциативной памяти для машин потока данных на основе генетического алгоритма // Программирование. 2002. №.6. С. 31–42.
10. Чернышев О., Борисов А. Сравнительный анализ решения задач оптимизации генетическими и градиентными методами // *Transport and Telecommunication*. 2007. V. 8, № 1. P. 40–52.
11. Линд Ю.Б., Губайдуллин И.М., Рамазанов М.Д. Параллельные вычисления при решении обратных задач физической химии. Параллельные вычислительные технологии (ПаВТ'2010): Труды международной научной конференции (Уфа, 29 марта - 2 апреля 2010 г.). Челябинск: Издательский центр ЮУрГУ, 2010. С. 507–518.
12. Линд Ю.Б. , Аристархов А.В. , Губайдуллин И.М. Параллельные вычисления при построении кинетической модели реакции гидроалюминирования олефинов. Научный сервис в сети Интернет: суперкомпьютерные центры и задачи. Труды Международной суперкомпьютерной конференции (Новороссийск, 20-25 сентября 2010 г.). М.: Изд-во МГУ, 2010. С. 231–237.

Analysis of methods for solving inverse problems of chemical kinetics with the use of parallel computing

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

Institute of Petrochemistry and Catalysis of RAS¹, Ufa State Petroleum Technological University²

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: inverse problems, kinetic constants, physical chemistry, the set-goprotsessornye system, the MPI.

References

1. Akhmetov I.V., Gubaydullin I.M., Balayev A.V. Razrabotka kineticheskoy modeli reaktsii polucheniya metilovogo efira 5-atsetil-2-pirrolkarbonovoy kisloty [Development of a kinetic model of the reaction preparation of methyl 5-acetyl-2-pyrrolicarboxylic acid]. Zhurnal Srednevolzhskogo matematicheskogo obshchestva [Journal of Middle Volga Mathematical Society]. 2010. Vol. 12. No. 3. P. 50–54.
2. Akhmetov I.V., Gubaydullin I.M. Razrabotka kineticheskikh modeley s ispol'zovaniem parallel'nykh vychisleniy na mnogoyadernykh sistemakh [Development of kinetic models using parallel computing on multicore systems]. Vestnik Omskogo universiteta [Herald of Omsk University]. 2012. No. 2(64). P. 172–174.
3. Akhmetov I.V., Gubaydullin I.M. Postroenie kineticheskikh modeley khimicheskikh reaktsiy na osnove mnogoyadernykh sistem [Construction of kinetic models of chemical reactions based on multi-core systems]. Zhurnal Srednevolzhskogo matematicheskogo obshchestva [Journal of Middle Volga Mathematical Society]. 2012. Vol. 14. No. 3. P. 38–42.
4. Lind YU.B., Gubaydullin I.M., Mulyukov R.A. Metodologiya parallel'nykh vychisleniy dlya resheniya zadach khimicheskoy kinetiki i burovoy tekhnologii [The methodology of parallel computations to solve the problems of chemical kinetics and drilling technology]. Sistemy upravleniya i informatsionnye tekhnologii [Control systems and information technologies]. 2009. Vol. 2/36. P. 44–49.
5. Gubaydullin I.M., Koledina K.F., Lind YU.B. Sovremennye tekhnologii vysokoproizvoditel'nykh vychisleniy pri modelirovanii detal'nogo mekhanizma reaktsii kataliticheskogo gidroalyuminirovaniya olefinov [Modern technologies of high performance computing for modeling the detailed mechanism of the catalytic reaction of olefins hydroaluminizing]. URL: <http://technomag.edu.ru/doc/187631.html> (accessed: 20.11.2011).
6. Lind YU.B., Gubaydullin I.M., Parfenova L.V., Ramazanov M.D., Spivak S.I. O primeneni parallel'nykh vychislitel'nykh tekhnologiy pri nakhozhenii kineticheskikh parametrov obshchego mekhanizma Zr kataliza v reaktsiyakh karbo-, gidro- i tsiklometallirovaniya olefinov v prisutstvii katalizatora Cp2ZrCl2 [On the application of parallel computing technologies in finding kinetic parameters of the total Zr mechanism of catalysis in the reactions of carbonyl, hydro and cyclometallation Cp2ZrCl2 olefins in the presence of a catalyst]. Parallelnye vychislitelnye tekhnologii (PaVT'2007): Trudy mezhdunarodnoj nauchnoj konferentsii (Chelyabinsk, 29 yanvarya – 2 fevralya 2007) [Parallel Computational Technologies (PCT'2007): Proceedings of the International Scientific Conference (Chelyabinsk, Russia, January, 29 – February, 2, 2007)]. Chelyabinsk, Publishing of the South Ural State University, 2007. P. 128–133.

7. Gubaydullin I.M., Lind YU.B., Akhmatsafina E.R., Spivak S.I. Relyatsionnaya sistema upravleniya bazoy dannykh dlya reaktsii gidroaluminirovaniya olefinov v prisutstvii tsirkonievogo katalizatora, realizuyushchaya dinamicheskoe raspredelenie dannykh mezhd protsessorami mnogoprotsessornoy vychislitel'noy sistemy [A relational database management system for the reaction of olefins hydroaluminizing in the presence of a zirconium catalyst, implementing dynamic data distribution among the processors of multiprocessor computing system]. Parallelnye vychislitelnye tekhnologii (PaVT'2008): Trudy mezhdunarodnoj nauchnoj konferentsii (Sankt-Peterburg, 28 yanvarya – 1 fevralya 2008) [Parallel Computational Technologies (PCT'2008): Proceedings of the International Scientific Conference (St. Petersburg, January, 28 – February, 1, 2008)]. Chelyabinsk, Publishing of the South Ural State University, 2008. P. 370–375.
8. Holland J. H. Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor. 1975. 96 P.
9. Nikitin A.V., Nikitina L.I. Evolutionary optimization model modular associative memory based on the data flow machine genetic algorithm // Programming and Computer Software. 2002. No. 6. P. 31–42.
10. Chernyshev A., Borisov A. Sravnitel'nyy analiz resheniya zadach optimizatsii geneticheskimi i gradientnymi metodami [Comparative analysis of solving optimization problems genetic and gradient methods]. Transport and Telecommunication. 2007. Vol. 8, No. 1. P. 40–52.
11. Lind YU.B., Gubaydullin I.M., Ramazanov M.D. Parallel'nye vychisleniya pri reshenii obratnykh zadach fizicheskoy khimii [Parallel computing for solving inverse problems of physical chemistry]. Parallelnye vychislitelnye tekhnologii (PaVT'2009): Trudy mezhdunarodnoj nauchnoj konferentsii (Ufa, 29 marta – 2 aprelya 2010) [Parallel Computational Technologies (PCT'2010): Proceedings of the International Scientific Conference (Ufa, March, 29 – April, 2, 2010)]. Chelyabinsk, Publishing of the South Ural State University, 2010. P. 507–518.
12. Lind YU.B., Aristarkhov A.V., Gubaydullin I.M. Parallel'nye vychisleniya pri postroenii kineticheskoy modeli reaktsii gidroaluminirovaniya olefinov [Parallel computing in the construction of a kinetic model of the reaction of olefins hydroaluminizing]. Nauchnyy servis v seti Internet: superkomp'yuternye tsentry i zadachi. Trudy mezhdunarodnoy superkomp'yuternoy konferentsii (Novorossiysk, 20–25 sentyabrya 2010) [Scientific Services & Internet: Supercomputing Centers and Applications. Proceedings of the International Supercomputer Conference (Novorossiysk, September, 20–25, 2010)]. Moscow, Moscow State University, 2010. P. 231–237.