

На правах рукописи

ЕВСЕЕВ Александр Владимирович

СОВЕРШЕНСТВОВАНИЕ МЕТОДОВ МОДЕЛИРОВАНИЯ
ТЕЧЕНИЙ С ХИМИЧЕСКИМИ ПРЕВРАЩЕНИЯМИ
И ИХ РЕАЛИЗАЦИЯ НА ГРАФИЧЕСКИХ ПРОЦЕССОРАХ

Специальность 05.13.18 – "Математическое моделирование,
численные методы и комплексы
программ"

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Иваново 2011

Работа выполнена в ФГБОУВПО "Ивановский государственный энергетический университет им. В.И. Ленина"

Научный руководитель д-р физ.-мат. наук,
профессор Ф.Н. Ясинский

Официальные оппоненты: д-р техн. наук, проф. Е.Н. Калинин
д-р физ.-мат. наук, проф. А.И. Григорьев

Ведущая организация ФГБОУВПО "Ивановский государственный
химико-технологический университет"

Защита диссертации состоится 21 декабря 2011 г. в 11-00 часов в ауд. Б-237 на заседании диссертационного совета Д 212.064.03 при ИГЭУ (153003, г. Иваново, ул. Рабфаковская, 34)

С диссертацией можно ознакомиться в библиотеке ИГЭУ.

Автореферат разослан 21 ноября 2011 г.

Учёный секретарь
диссертационного совета

А.А. Шульпин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Применение численных методов становится всё более и более востребованным в связи с увеличивающейся трудоёмкостью проведения натуральных экспериментов и сложностью получения детальной информации о течениях и химических процессах, происходящих в них, например: при исследовании химических лазеров, моделировании процессов, происходящих при пожаре, изучении распространения загрязнений.

Использование высокопроизводительных вычислительных систем принципиально изменило возможности теоретического анализа сложных процессов. Стало доступным оперативно анализировать сложные модели с большим объёмом данных. При этом точность моделирования определяется математической моделью и точностью необходимых данных, учитывающих физические условия проведения процесса. В настоящее время в большинстве исследований перед проведением натурального эксперимента проводится его математическое моделирование для определения наиболее оптимальных параметров. Этим достигается экономия сил, средств и времени. Другой несомненный плюс от использования вычислительных экспериментов заключается в возможности проведения испытаний в критических условиях или в тех областях параметров, где нельзя провести реальные эксперименты на имеющемся оборудовании.

Моделирование течений, осложнённых химическими процессами, является сложной математической и вычислительной задачей.

В большинстве случаев уравнения Навье – Стокса, описывающие поведение течения вязкой несжимаемой жидкости, аппроксимируются различными сеточными методами. Основным требованием, предъявляемым к таким методам, является обеспечение высокой точности получаемых результатов при минимально необходимых вычислительных затратах (времени, объёме памяти).

При решении уравнений Навье – Стокса требуется многократно использовать уравнение Пуассона. Оно является вычислительно сложным и составляет основную вычислительную нагрузку.

Моделирование активно реагирующей среды требует надёжного метода решения жёстких дифференциальных уравнений. В качестве такого метода может выступать метод Гира на основе формул дифференцирования назад (ФДН). Его преимуществом по сравнению с многошаговыми методами является возможность использования стратегий с переменным шагом интегрирования и простое построение параллельной схемы.

К настоящему времени существует множество работ по моделированию как обычных течений, так и осложнённых различными физико-химическими реакциями.

Основы методов моделирования были заложены в работах С.К. Годунова, Н.Н. Калиткина, А.А. Самарского, Р.П. Федоренко, Г.И. Марчука, Л.Г. Лойцянского, Н.Н. Яненко, Р. Хокни, Дж. Бетчелора, Г. Биркгофа, Д. Андерсона.

Значительный вклад в моделирование турбулентных течений внесли А.Дж. Рейнольдс, В.М. Ковеня, В.А. Перминов, А.М. Гришин, С.А. Лосев, А.Е. Алоян, П.Г. Фрик, Б. Сполдинг, С. Патанкар.

Проблемы интегрирования уравнений химической кинетики описываются в работах У. Гира, Л.С. Полака, Э. Хайрера.

В сфере научных вычислений всё большую популярность набирают вычисления с использованием графических ускорителей. Современные видеоакселераторы представляют собой массивно-параллельные процессоры с общей памятью. В отличие от центрального процессора с несколькими ядрами, один графический процессор содержит от нескольких десятков до одной тысячи ядер, которые могут проводить вычисления параллельно. Наиболее развитой на данный момент технологией является система CUDA (Compute Unified Device Architecture), предложенная компанией NVidia в 2007 году.

Использование кластерных систем с системой параллельного программирования MPI (Message Passing Interface) и графических ускорителей для решения описанных задач приводится в работах В.Э. Витковского, Н.А. Сахарных, Дж. Коуэна (J.Cohen), Я. Джао (Y. Zhao), Т. Брандвика (T. Brandvik).

Таким образом, тема диссертации является **актуальной**, так как она посвящена разработке в системе CUDA быстрых параллельных алгоритмов решения сеточных уравнений (Навье – Стокса, Пуассона) и химической кинетики, получаемых при решении задач моделирования течений с протекающими в них реакциями множества веществ.

Цель работы

Целью работы является разработка усовершенствованных методов моделирования течений, осложнённых химическими превращениями, с использованием графических процессоров, позволяющих значительно повысить скорость вычислений, эффективно использовать ресурсы гетерогенного суперкомпьютера с несколькими графическими ускорителями, а также разработка комплекса программ для решения класса таких задач.

Задачи исследования

Поставлены следующие задачи исследования:

1. Исследовать существующие способы решения уравнений Навье – Стокса, Пуассона, химической кинетики, массообмена.
2. Разработать и реализовать общий алгоритм решения задачи расчёта течения жидкости или газа, осложнённого химическими реакциями на основе методов вычислительной гидродинамики и химической кинетики.
3. Разработать методику оптимизации вычислительного процесса, позволяющую значительно сократить затраты машинного времени и повысить точность решения данной задачи за счёт использования многосеточных методов с функциями перехода на основе интерполяционных полиномов Лагранжа.
4. Разработать стратегию распараллеливания для решения поставленной задачи с учётом предложенной методики оптимизации на гетерогенном супер-

компьютере с несколькими графическими ускорителями, позволяющую значительно уменьшить время решения задачи.

5. Использовать разработанные методы для решения ряда задач: модельное плоское ламинарное движение, расчёт поведения воздушных масс в районе над лесным пожаром, окислительные превращения метана.

Методы исследования

Для решения поставленных задач используются методы гидродинамики, химической кинетики, решения систем жёстких дифференциальных уравнений, вычислительной математики, теории распараллеливания вычислений, теории алгоритмов, теории разностных схем.

Научная новизна

1. Разработаны новые параллельные вычислительные алгоритмы методов решения уравнений Навье – Стокса и Пуассона для системы CUDA.

2. Разработана новая параллельная реализация метода Гира для системы CUDA.

3. Предложен и разработан способ интерполяции многосеточного метода на основе полиномов Лагранжа для системы CUDA.

4. Разработано новое алгоритмическое и программное обеспечение, позволяющее моделировать течения, осложнённые химическими превращениями, с использованием суперкомпьютера с несколькими графическими ускорителями CUDA.

Практическая ценность работы

В результате исследования разработан программный комплекс, позволяющий моделировать движение среды, осложнённой химическими превращениями.

Он может работать с любым количеством графических ускорителей, имеющихся в кластерной системе, оптимальным образом распределяя объём вычислений между ними, позволяет повысить скорость и точность вычислений.

Разработанные программы могут быть использованы для быстрого решения ряда практически значимых задач. Например, моделирование распространения воздушных масс при пожарах, воздействие лазерного излучения на вещество, получение новых топлив из продуктов окисления алканов. Ценность комплекса в том, что решения можно получать оперативно, в режиме реального времени для больших объёмов данных.

Реализация и внедрение результатов работы

Предложенные параллельные схемы и алгоритмы были внедрены в процесс выполнения комплексного проекта по созданию высокотехнологичного производства "Разработка программно-технического комплекса обнаружения и прогнозирования крупномасштабных природных пожаров" (2010-218-02-139)

при поддержке Министерства образования и науки РФ (ГК № 13.G25.31.0077) сотрудниками ГОУВПО "Ивановский институт ГПС МЧС России".

Составные части представленного программного комплекса зарегистрированы в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам.

Апробация работы

Основные положения диссертационной работы докладывались и обсуждались на следующих конференциях: межвузовская научно-техническая конференция аспирантов и студентов "Молодые учёные – развитию текстильной и лёгкой промышленности" (ПОИСК-2009) (Иваново, 2009); IX международная конференция "Высокопроизводительные параллельные вычисления на кластерных системах" (Владимир, 2009); X международная научно-практическая конференция "Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности" (Санкт-Петербург, 2010); X международная конференция "Высокопроизводительные параллельные вычисления на кластерных системах" (НПС-2010) (Пермь, 2010); V международная научно-практическая конференция "Пожарная и аварийная безопасность" (Иваново, 2010); межвузовская научно-техническая конференция аспирантов и студентов "Молодые учёные – развитию текстильной и лёгкой промышленности" (ПОИСК-2011) (Иваново, 2011); семинар "Физико-химическая кинетика в газовой динамике" НИИ механики МГУ (Москва, 6 октября 2011 года).

Публикации

По теме диссертации опубликовано 14 работ, среди которых 4 публикации в ведущих рецензируемых изданиях, рекомендованных в действующем перечне ВАК, 6 публикаций в трудах и материалах конференций, 2 монографии, 2 свидетельства о государственной регистрации программы для ЭВМ.

Личный вклад автора

Основные научные и практические результаты диссертации получены автором лично. В работе [1] автору принадлежит методика численного моделирования задачи динамики вязкой несжимаемой жидкости на графических ускорителях с системой CUDA. В [2, 6] автором предложен способ организации вычислений указанной выше задачи на системе с несколькими графическими ускорителями. В работе [3] автором описывается способ решения задачи моделирования течения, осложнённого физико-химическими преобразованиями в системе параллельного программирования CUDA. В работах [4, 5, 10, 11] автором рассматриваются и сравниваются различные способы решения уравнения Пуассона как в однопроцессорном варианте, так и в многопроцессорном. В работах [8, 9] автором предлагается методика организации вычислительного процесса на гетерогенном суперкомпьютере для задач моделирования течения среды, с проходящими в ней физико-химическими процессами. В монографии [12] автором описывается способ решения жёстких задач химической кинетики с

использованием метода Гира. Автор принимал участие в подготовке и представлении докладов на конференциях.

Структура и объём работы

Работа состоит из введения, 4-х глав, заключения и списка цитируемой литературы из 145 наименований. Полный объём диссертации составляет 198 страниц, включая 37 рисунков, 5 таблиц и 2 приложения. Каждая глава разбита на параграфы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность исследования, ставятся цели и задачи исследования, раскрывается его научная новизна и практическая ценность, излагаются основные положения, выносимые на защиту.

В **первой главе** ставится задача моделирования течений жидкости или газа с протекающими в них многочисленными реакциями множества веществ.

В *параграфе 1.1* приводятся и раскрываются все уравнения, необходимые для описания математической модели поставленной задачи. Уравнения Навье – Стокса позволяют приближённо описывать движение реальной жидкости путём решения краевых задач для соответствующих систем дифференциальных уравнений в частных производных:

$$\begin{cases} \frac{\partial \vec{U}}{\partial t} = -(\vec{U} \cdot \nabla) \vec{U} + \nu \Delta \vec{U} - \frac{1}{\rho} \nabla P + \vec{F}; \\ \operatorname{div} \vec{U} = 0, \end{cases} \quad (1)$$

где \vec{U} – вектор скорости; P – давление среды; \vec{F} – вектор массовых сил; ν – кинематическая вязкость.

Для решения задачи физико-химической гидродинамики, а именно изучения процесса переноса вещества при химических и фазовых превращениях, необходимо добавить уравнения массообмена. Перенос вещества в движущейся среде обусловлен двумя различными механизмами. Наличие разности концентраций в жидкости (газе) вызывает молекулярную диффузию. Кроме того, частицы вещества, растворенного в жидкости, увлекаются последней при её движении и переносятся вместе с ней. Совокупность обоих процессов называется конвективной диффузией:

$$\frac{\partial C_j}{\partial t} + \sum_{i=1}^2 U_i \frac{\partial C_j}{\partial x_i} = D_{C_j} \sum_{i=1}^2 \frac{\partial^2 C_j}{\partial x_i^2} + W_j; \quad j = \overline{1, N}, \quad (2)$$

где C_j – концентрация j -го вещества; D_{C_j} – коэффициент диффузии j -го вещества; N – число веществ. Член W_j определяет изменение концентрации j -го вещества в результате химических реакций с течением времени.

Система дифференциальных уравнений химической кинетики определяется набором химических реакций. Это система обыкновенных дифференциальных уравнений первого порядка следующего вида:

$$W_i = \frac{dC_i}{dt} = \sum_{j=1}^K R_{ij} A_j \prod_{m=1}^N C_m^{L_{mj}} - \sum_{j=1}^K L_{ij} A_j \prod_{m=1}^N C_m^{L_{mj}}; i = \overline{1, N}, \quad (3)$$

где L_{ij} — стехиометрический коэффициент при i -м веществе в левой части j -й реакции; R_{ij} — стехиометрический коэффициент при i -м веществе в правой части j -й реакции; A_j — константа скорости j -й реакции.

В *параграфе 1.2* проводится исследование работ авторов, изучающих различные способы и методы решения предложенных уравнений, способы построения их параллельных схем. Также описываются существующие библиотеки и комплексы программ для моделирования течений, указываются их достоинства и недостатки.

В *параграфе 1.3* ставятся цели исследования, описываются возможные пути решения поставленной задачи, способы улучшения достигнутых результатов.

Во **второй главе** приводится теоретическое описание существующих способов решения уравнений Навье – Стокса как в естественных переменных "Скорость – Давление", так и в математических – "Вихрь – Функция тока".

Указанные уравнения предлагается решать с помощью локально-одномерных методов расщепления или метода переменных направлений для двумерного случая. Данные методы позволяют разбить сложную многомерную задачу на ряд одномерных систем трёхдиагональных уравнений.

Подобные системы уравнений удобно решать с помощью метода прогонки (Томаса). Суть его состоит в решении системы линейных уравнений, заданной в виде $Ax=F$, где A – трёхдиагональная матрица, построением рекуррентных соотношений, основанном на методе исключений Гаусса.

Параграф 2.4 посвящён исследованию различных способов решения уравнения Пуассона. Все численные методы решения уравнения можно условно разделить на две группы: итерационные и прямые. Среди итерационных методов рассматриваются метод верхней релаксации, метод Ричардсона, метод Дугласа – Писмана – Рекфорда и попеременно-треугольный метод. Из прямых методов исследуется метод неполной редукции FACR.

Метод верхней релаксации имеет достаточно простой итерационный процесс. Влияющим фактором на этот процесс является параметр релаксации Θ .

Идея Л. Ричардсона состоит в том, чтобы на каждом шаге менять итерационный параметр τ . Показано, что при специальном выборе последовательности шагов τ_k ($k = 1, 2, \dots, n$), называемой далее циклом, метод будет сходиться к стационарному решению в $O(M)$ раз быстрее, чем при итерациях с постоянным оптимальным шагом τ .

Главное отличие метода Дугласа – Писмана – Рекфорда состоит в том, что он основан на схеме переменных направлений. В нем также используется последовательность шагов τ_k ($k = 1, 2, \dots, n$), зависящая от номера итерации.

В попеременно-треугольном методе используется расщепление исходного сеточного оператора на два треугольных оператора, для которых потом используется схема бегущего счёта.

Алгоритм FACR, предложенный Р. Хокни, основывается на алгоритме циклической редукции. В методе используется то обстоятельство, что редукцию можно остановить на любом уровне r и решить полученные уравнения с помощью преобразования Фурье. Уровень r , на котором редукция останавливается, служит теперь параметром. Этот параметр можно варьировать для минимизации числа арифметических операций.

Теоретическое и экспериментальное сравнение последовательных версий указанных методов решения уравнения Пуассона позволяет выделить прямой метод FACR как наиболее быстрый и точный. Однако необходимо помнить и учитывать ограничения в применении данного метода, которые выражаются в жёстких условиях на размеры расчётной области и виды граничных условий.

Среди итерационных методов наилучшие показатели выявились у попеременно-треугольного метода и метода Дугласа – Писмана – Рекфорда. Они показали высокую скорость сходимости, устойчивость и точность. Метод Ричардсона не имеет высоких скоростных характеристик, но обладает достаточно устойчивой и простой схемой, которая может быть использована в качестве оператора сглаживания в многосеточном методе.

Параграф 2.5 посвящён методам решения жёстких систем дифференциальных уравнений. Обычно жёсткие задачи описывают как задачи, моделирующие физический процесс, в котором присутствуют компоненты с несоизмеримыми временными характеристиками, т.е. когда в системе присутствуют как быстро меняющиеся, так и медленно текущие процессы. Подобные процессы наиболее часто встречаются в задачах химической кинетики, когда в реакции участвуют и активные частицы (например, ионы), и стабильные молекулы.

В качестве способов решения данных систем уравнений часто используются многошаговые методы, например метод Адамса – Бэшфорта – Моултона.

Но более точным и устойчивым является многозначный метод Гира, использующий подход, получивший название ФДН-метод в представлении Нордсика, поскольку в нём используется одноимённый вектор:

$$\mathbf{z} = (y, h \cdot y', \frac{h^2}{2} y'', \dots, \frac{h^p}{p!} y^{(p)})^T. \quad (4)$$

где y – неизвестная функция; h – шаг по времени; p – порядок метода Гира.

Суть метода состоит в нахождении приближения к точному значению вектора Нордсика, аппроксимирующего функцию, путём вычисления "предиктора" и "корректора". Значение предиктора находится на основе текущей точки интегрирования, а значение корректора вычисляется исходя из текущей погрешности метода с помощью итераций Ньютона.

Третья глава посвящена построению параллельных схем методов для графических ускорителей CUDA, а также для объединений этих ускорителей в рамках одной системы или кластера компьютерных систем.

Отдельно в параграфе 3.1 предлагается способ повышения точности сеточных методов с использованием многосеточного метода. Он состоит в последовательном укрупнении шага расчётной сетки и вычислении оператора сглаживания сетки до достижения максимально допустимой грубой сетки, на которой достигается минимальная необходимая точность решения. После этого вычисляется поправка и происходит процесс её обратного распространения на более точные сетки. Графически это представлено на рис.1.

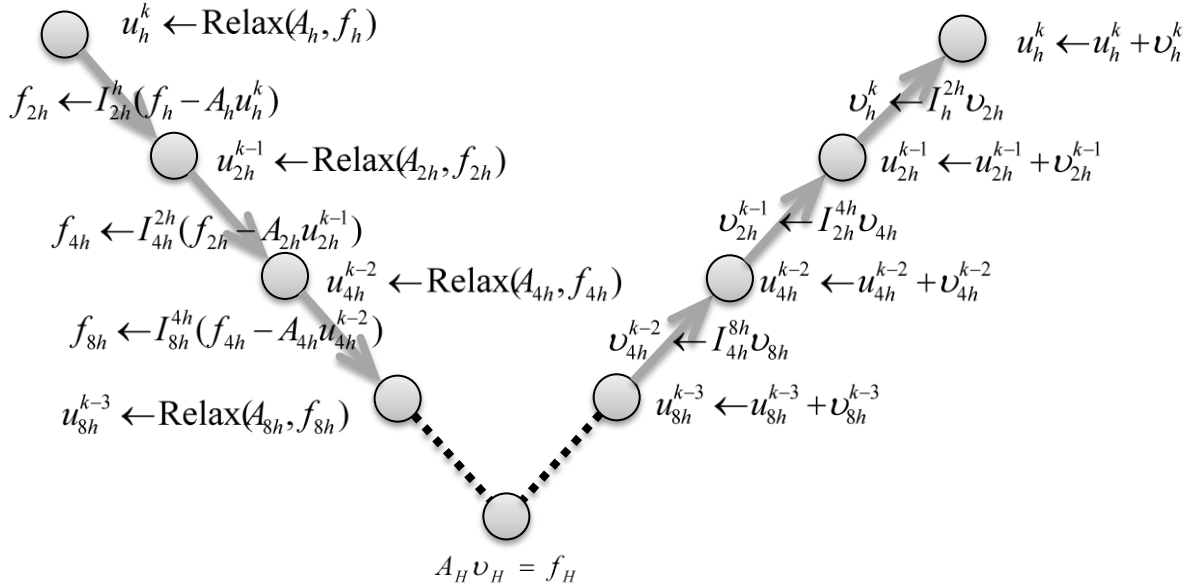


Рис. 1. Алгоритм многосеточного метода

При переходах от сетки к сетке используются операторы перехода, которые на основе некоторых шаблонов, основанных на степени двойки, вычисляют промежуточные значения. Многосеточный метод позволяет эффективно гасить высокочастотные составляющие спектра используемого метода релаксации. Однако при определённом сочетании частоты гармоник и величины шага по времени эти гармоники не будут погашены. При этом изменение величины шага в 2 раза на каждом шаге многосеточного метода, как это обычно принято, приводит к той же проблеме.

Чтобы исключить данный негативный фактор, используем функции перехода и интерполяции на основе полиномов Лагранжа с произвольным коэффициентом перехода между сетками.

Интерполяционный полином в форме Лагранжа имеет вид

$$L_n(x) = \sum_{i=0}^n y_i l_i(x), \quad l_i(x) = \prod_{\substack{p=0, \\ p \neq i}}^n \frac{x - x_p}{x_i - x_p}, \quad \begin{cases} l_i(x_i) = 1; \\ l_i(x_s) = 0, s \neq i, \end{cases} \quad (5)$$

где n – степень полинома; y_i – исходная функция; $l_i(x)$ – базисные полиномы.

Двумерный интерполяционный полином строится следующим образом:

$$l_{ij}(x, y) = l_i(x) \cdot l_j(y), \quad \begin{cases} l_{ij}(x_i, y_j) = 1; \\ l_{ij}(x_s, y_t) = 0, s \neq i, t \neq j; \end{cases} \quad (6)$$

$$l_i(x) = \prod_{\substack{p=0, \\ p \neq i}}^n \frac{x - x_p}{x_i - x_p}, \quad l_j(y) = \prod_{\substack{q=0, \\ q \neq j}}^m \frac{y - y_q}{y_j - y_q}, \quad (7)$$

$$L_{nm}(x, y) = \sum_{i=0}^n \sum_{j=0}^m z_{ij} l_{ij}(x, y), \quad (8)$$

где m – степень полинома по второй переменной; z_{ij} – исходная табличная двумерная функция.

Пусть имеется некоторая сеточная функция z_{ij} и сетка $[x_i, y_j]$, $0 \leq i \leq n$, $0 \leq j \leq m$. Исходя из вышеизложенного, для того, чтобы перейти от одной сетки к другой, необходимо в первую очередь построить новую сетку $[x'_p, y'_q]$, $0 \leq p \leq n'$, $0 \leq q \leq m'$, а затем по полученным узлам сетки построить новую сеточную функцию z'_{pq} , используя (8).

В параграфе 3.2 приведены разработанные параллельные алгоритмы для систем с разделённой памятью, использующие MPI.

Главными вопросами, которые необходимо было решить при построении параллельных алгоритмов, были способ разбиения задачи между процессорами и определение эффективного взаимодействия между ними при передачах данных. Замечательным свойством решения уравнения Пуассона является то, что при любом числе измерений с использованием геометрического параллелизма исходную задачу можно разбить на ряд аналогичных задач меньшего объёма. Подобное разбиение легко реализуется и требует в общем случае пересылок данных только на граничных слоях.

При построении параллельной схемы метода Дугласа – Писмана – Рекфорда, а также для интегрирования уравнений Навье – Стокса методами расщепления и переменных направлений был разработан алгоритм двухкаскадного конвейера для вычисления метода прогонки на системах с разделённой памятью в случае распределения данных по узлам. Принцип этого метода представлен на рис. 2 (сплошными линиями показаны вычисления прямых прогоночных коэффициентов, пунктирными – обратных).



Рис. 2. Сокращение времени вычисления прогоночных коэффициентов при использовании алгоритма двухкаскадного конвейера

В алгоритме оптимизированной прогонки вычисления прямых коэффициентов проводятся без ожидания получения данных обратной прогонки.

При интегрировании уравнений химической кинетики факторами, определяющими скорость вычислений, являются векторные и матричные операции, поскольку каждый этап алгоритма метода Гира в той или иной степени использует эти операции с векторами получаемого решения, векторами Нордсика, погрешности, поправки и т.д. Поэтому основной упор в построении параллельной схемы для метода Гира делается на распараллеливание векторных операций. В этом случае оптимальным является разделение векторных данных и операций равномерно по всем процессорам, когда каждый из них имеет свою локальную часть общего массива данных.

В параграфах 3.3 – 3.5 исследуется создание параллельных алгоритмов решения поставленной задачи с использованием графических ускорителей CUDA и систем таких ускорителей.

Количество потоков, одновременно исполняемых на графическом ускорителе, измеряется сотнями, что, согласно закону Амдала, позволяет достичь высокого ускорения при условии, что существенная часть вычислительного процесса будет выполняться на устройстве CUDA, максимально и эффективно используя его ресурсы. Кроме того, необходимо учитывать накладные расходы на работу с памятью, поскольку в некоторых случаях это может стать узким местом алгоритма. Именно поэтому крайне важно создавать хорошо распараллеленные алгоритмы, эффективно использующие все виды доступа к памяти, представляемые системой CUDA.

В отличие от MPI, единственным подходящим видом параллелизма является геометрический параллелизм, когда каждому блоку назначается свой сектор исходных данных, а каждый поток выполняет вычисления только с одним элементом данных. Но в тоже время граничные условия становятся существенной проблемой, т.к. они необходимы для вычислений, но не должны вычисляться по общему алгоритму. Следовательно, требуется, чтобы каждый расчётный блок использовал разделяемую память с учётом граничных условий и обновлял их на каждой итерации.

Отличительной особенностью CUDA является то, что в ней нет возможности осуществить глобальную синхронизацию вычислений между блоками иным способом, кроме как закончить вычисления во всех блоках и произвести обмен данными на компьютере. В связи с этим каждый этап вычислительного алгоритма может использовать собственное, оптимальное для него разделение задачи между блоками. Однако эта особенность не позволяет полностью перенести итерационные алгоритмы на графические ускорители, поскольку каждый блок после проведения итерации на своём участке данных и вычисления погрешности должен соотнести их с другими блоками, а это не представляется возможным. Таким образом, для итерационных алгоритмов естественным решением является вычисление одной итерации метода на устройстве, а проверка сходимости и принятие решения о продолжении расчёта принимает поток CPU. Кроме того, данный подход решает проблему граничных значений для каждого блока, т.к. они обновляются при каждом запуске итерации.

При построении параллельных схем методов верхней релаксации и попеременно-треугольного для избавления связанности данных по времени используется схема вычисления по разделению точек сетки в шахматном порядке на "чёрные" и "белые". Процесс вычисления разделяется на два этапа: сначала на основе старых исходных значений вычисляются "белые" точки, затем на основе новых значений рассчитываются "чёрные" точки. Таким образом, достигается ускорение сходимости метода и массовость параллелизма – одновременный расчёт как можно большего количества точек.

В конце каждой итерации по времени требуется проверка сходимости метода, которая заключается в выполнении редакционного суммирования по массиву сходимости. В результате получается количество элементов, не достигших сходимости.

Существует два основных подхода к организации параллельного вычислительного алгоритма метода прогонки: применение специализированных алгоритмов параллельной прогонки, основанных на методах циклической редукции, и использование обычного последовательного алгоритма прогонки одним потоком.

Кроме того, необходимо учитывать, что обычно требуется решение не одной трёхдиагональной системы, а целого набора таких систем. В трёхмерном случае это будет $N_x \times N_y \times N_z$ систем. При этом каждая из этих систем независима от других, что позволяет вычислять их параллельно.

Реализуя алгоритмы прогонки в системе CUDA, необходимо учитывать доступ к памяти, поскольку он может существенно повлиять на производительность методов. Если удаётся использовать возможности связанной загрузки данных, то сокращается общее время ожидания загрузки и возрастает эффективность. При организации массивов данных (a, b, c, f, x) для наборов систем уравнений можно располагать данные либо последовательно (система за системой), либо чередовать между собой (первые элементы всех систем, вторые элементы и т.д.). Первый способ подходит для алгоритмов, основанных на циклической редукции, поскольку в них данные каждым блоком считываются только для одной системы. Второй способ эффективен при использовании обычного алгоритма прогонки, так как в данном случае один блок решает несколько систем одновременно и каждый поток считывает один и тот же элемент данных только у разных систем. Таким образом, достигается связанная загрузка данных для каждого элемента.

Сравнение результатов алгоритмов и ограничений позволяет выбрать обычный последовательный алгоритм с перемежающимся расположением данных как наиболее эффективный.

Одним из преимуществ системы CUDA является возможность одновременного использования нескольких ускорителей, установленных в компьютер. В ряде задач это позволяет существенно повысить скорость вычислений. Однако имеется одно важное архитектурное ограничение – не существует иного способа обмена данными между ускорителями, кроме как через ОЗУ компьютера в той части программы, которая выполняется на центральном процессоре.

Таким образом, при проектировании алгоритмов решения необходимо учитывать эти ограничения и сводить количество пересылок данных между устройствами CUDA к минимуму.

Следует заметить, что в данном случае можно рассматривать систему с несколькими графическими ускорителями CUDA как систему с разделённой памятью. Подобная аппроксимация позволяет применять методы, рассчитанные на системы параллельного программирования MPI. Так, например, решение уравнения Пуассона можно разделить на отдельные независимые участки со своими граничными условиями, обновляемыми на каждой итерации. В тоже время эти независимые участки можно решать стандартными способами на каждой графической карте в отдельности.

Также, как и в случае с системами MPI, нетривиальным является решение трёхдиагональных систем, когда данные распределены между графическими ускорителями. Как уже указывалось выше, в этом случае пересылки данных между ними необходимо осуществлять через центральный процессор и ОЗУ, следовательно, эффективные параллельные алгоритмы, например параллельная циклическая редукция, становятся крайне неэффективными, если прогонку необходимо провести на данных, находящихся на разных картах. Также неэффективным оказывается и каскадный алгоритм, предложенный для систем MPI, поскольку он содержит большое число обменов данными.

В этом случае предлагается использовать транспонирование данных. Транспонирование может быть эффективно выполнено с использованием графических ускорителей, поскольку каждый из них может повернуть свой участок данных в своей памяти, а затем они передают свои транспонированные участки в ОЗУ, где центральный процессор записывает их вертикальными полосками. В итоге получается транспонированное общее поле данных, которое затем снова разделяется на горизонтальные полоски и передаётся на карты, где снова выполняется алгоритм обычной прогонки.

Использование гетерогенного суперкомпьютера оправдано в случае моделирования полей большой размерности, поскольку в этом случае расчётные сетки не смогут поместиться в памяти одной системы.

Гетерогенный кластер можно создать на основе любых обычных компьютеров с графическими ускорителями CUDA, объединённых в сеть. В этом случае вычислительная среда является комбинацией технологий параллельного программирования и разделяется на три уровня: MPI – отвечает за обмен данными между узлами кластерной системы; OpenMP – обеспечивает создание многопоточного окружения на каждом узле кластера, необходимого для работы с несколькими ускорителями и для полноценного использования многоядерных процессоров; CUDA – выполняет основные вычисления.

Основные результаты исследования приведены в *параграфе 3.6*, описывающем сравнение предложенных параллельных схем, оценки ускорения и достигнутые результаты.

В первую очередь, для более корректного сравнения, было решено сравнить производительность и масштабируемость методов в среде параллельного

программирования MPI для кластерных систем и для систем с графическими ускорителями CUDA (рис. 3).

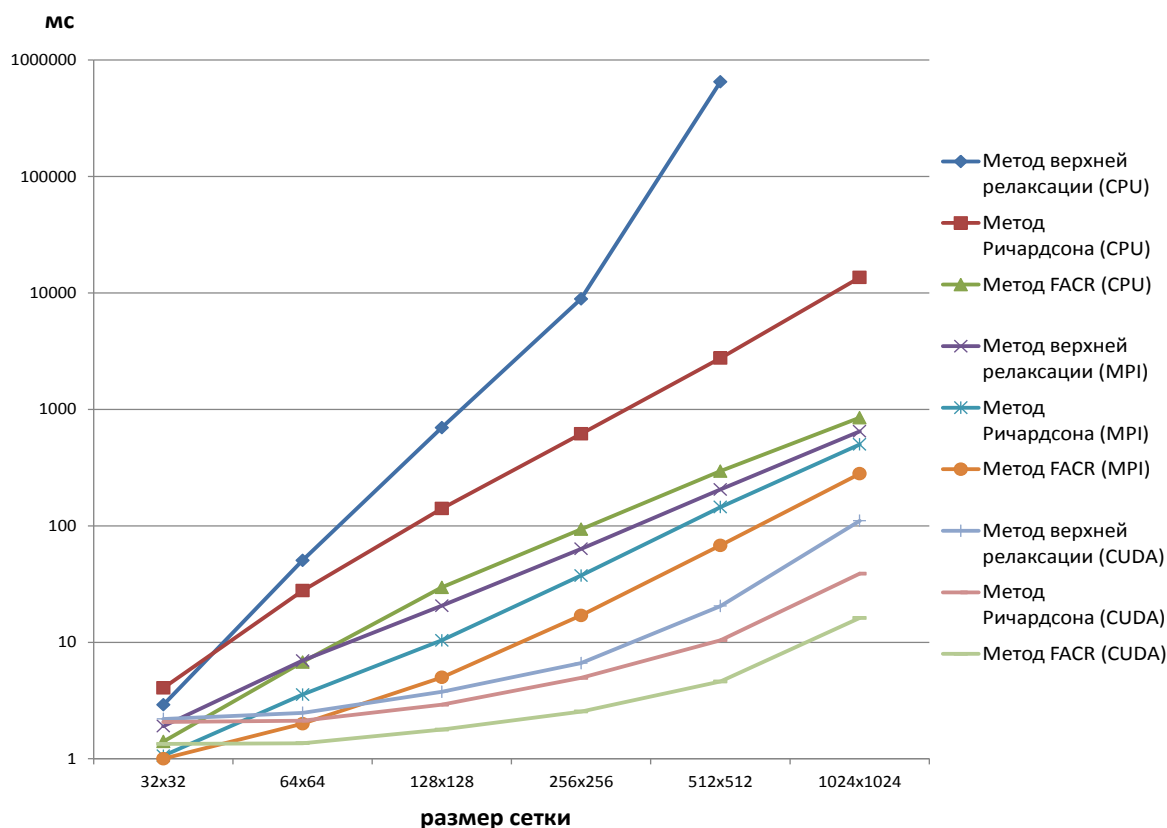


Рис. 3. Сравнение методов решения уравнения Пуассона на различных платформах

Как видно из приведённых графиков, производительность параллельных алгоритмов в системе CUDA начинает преобладать над системами MPI при увеличении расчётных областей. Это связано с возрастанием загрузки устройства. Алгоритмы для CUDA рассчитаны так, чтобы иметь большой объём данных для обработки, и в этом их главное преимущество перед традиционными системами MPI. При увеличении объёмов данных разрыв увеличивается, поскольку в этом случае графические ускорители не будут работать вхолостую, ожидая данных, а будут максимально загружены вычислениями.

Производительность методов решения трёхдиагональных систем в CUDA значительно превосходит MPI. Так, производительность MPI зависит от направления, в котором производится прогонка. В случае, когда данные распределены по узлам, используется каскадный алгоритм, обладающий заведомо меньшей производительностью. Кроме того, алгоритмы на CUDA выигрывают в том, что они рассчитаны на одновременное решение набора трёхдиагональных систем. В случае использования обычного алгоритма прогонки, вычисляемого одним потоком, каждый мультипроцессор графического ускорителя может вычислять 128 систем одновременно, а с учётом того, что на любой графической карте имеется N таких мультипроцессоров, то в целом одна карта может производить вычисления $N \times 128$ систем.

При вычислении методом Гира с использованием библиотеки CVODE были протестированы три доступные реализации векторных операций: последовательная (CPU) и параллельные в системах MPI и CUDA. Результаты сравнения производительности векторных операций представлены на рис. 4. Как видно из графиков, CUDA-реализация векторных операций начинает выигрывать в производительности при увеличении размерности векторов, характеризующих количество веществ и реакций между ними. Таким образом, это позволит строить химические модели многокомпонентных смесей со сложными реакциями между ними, и их число может измеряться тысячами и сотнями тысяч.

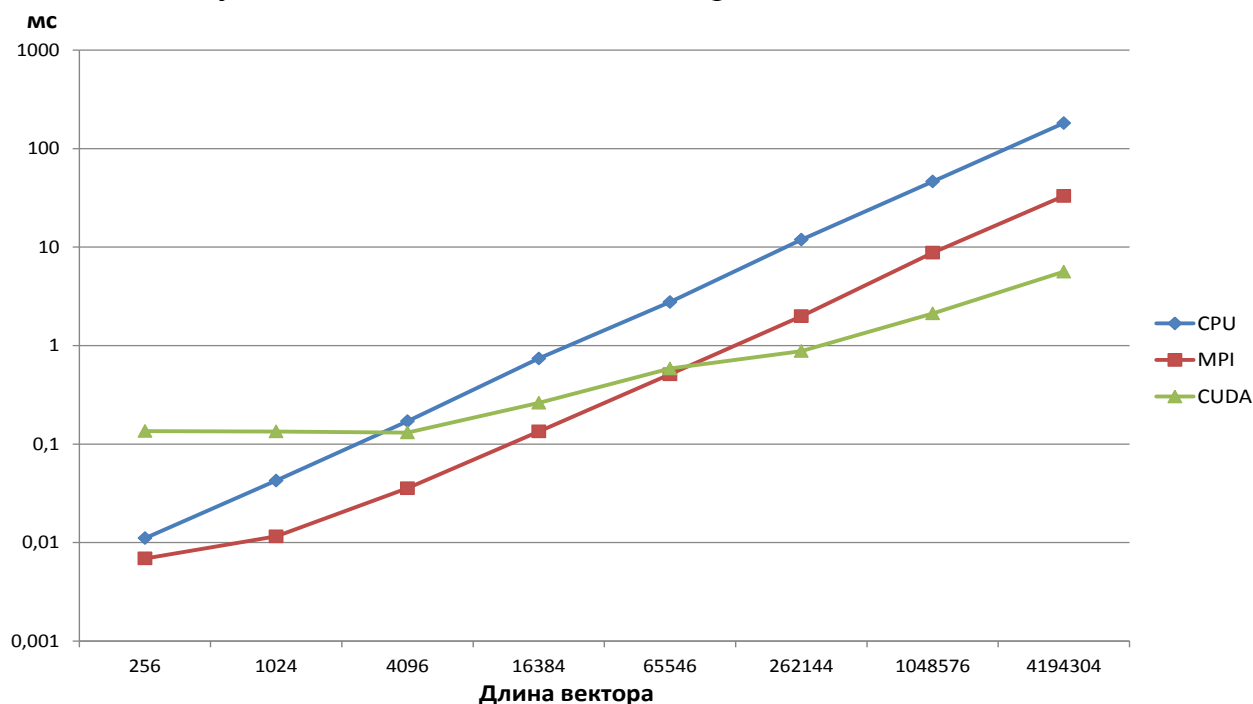


Рис. 4. Сравнение векторных операций на различных платформах

Моделирование движения среды, осложнённой множеством химических реакций большого числа веществ, изначально предполагает наличие больших расчётных сеток для более детального изучения. Кроме того, число веществ и реакций между ними может исчисляться десятками тысяч. В этом случае использование графических ускорителей становится оправданным, поскольку позволяет их полностью загрузить вычислениями. Однако на очень больших сетках или при переходе к моделированию в трёх измерениях становится очевидным и необходимым использование системы с несколькими графическими ускорителями или даже кластера таких систем.

Предложенная методика позволяет заметно сократить накладные расходы на рассмотрение одного варианта течения и при поиске наилучшего инженерного решения, например в области лазерных технологий, существенно увеличить число исследуемых вариантов. Также в некоторых случаях становится возможным моделирование процесса в режиме реального времени с интерактивным взаимодействием.

В **четвертой главе** предлагается решение ряда задач гидродинамики и химической кинетики с использованием полученных алгоритмов. Показывается их точность и достоверность.

В *параграфе 4.1* предлагается модельная задача гидродинамики нахождения плоского ламинарного течения. Она позволяет проверить достоверность и точность предложенных быстрых численных алгоритмов. Ускорение вычислений составляет более 20 раз при размерах области 256×256 по сравнению с обычным последовательным вариантом.

В *параграфе 4.2* ставится задача моделирования распространения воздушных масс в районе над лесным пожаром. В ней применяется модель турбулентности, основанная на методе осреднения по Рейнольдсу (RANS). Идея этого метода заключается в замене случайно изменяющихся характеристик потока (скорость, давление, плотность) суммами осреднённых и пульсационных составляющих. Важно отметить, что одной из составляющих предложенной модели турбулентности является температура среды и её усреднённые и пульсационные величины. Это позволяет учитывать сложный характер конвективного движения нагретых воздушных масс. При использовании расчётной сетки 512×256 получаемое ускорение составляет более 12 раз по сравнению с последовательным вариантом, что позволяет моделировать поведение в реальном времени, оперативно получая результаты.

Параграф 4.3 посвящён моделированию задачи химической кинетики по газофазному окислению метана. В рассматриваемой задаче используется достаточно подробная модель окисления метана, включающая 63 реакции и 23 вещества. Она позволяет определить точность вычисления метода Гира с помощью разработанных параллельных алгоритмов. Ускорение получается в 4 раза при использовании одного CUDA устройства. Выигрыш не настолько значителен, как при сравнении векторных операций, но это объясняется прежде всего тем, что число задействованных веществ и реакций было не достаточно велико и не позволило полностью раскрыть потенциал системы CUDA.

В **заключении** приводятся выводы и результаты по работе в целом.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

1. Подробно исследовано уравнение Пуассона и различные методы его численного решения. Произведено сравнение методов друг с другом. Рассмотрены параллельные варианты на предмет быстроты и точности в системах MPI и CUDA. (Получено свидетельство о государственной регистрации программы.)

2. Разработаны и реализованы усовершенствованные алгоритмы решения задачи гидродинамики в переменных "Скорость – Давление" и "Вихрь – Функция тока". Предложены как двумерные, так и трёхмерные варианты.

3. Предложена методика оптимизации вычислительного процесса с использованием многосеточного подхода с функциями перехода между сетками на основе интерполяционных полиномов Лагранжа.

4. Подробно исследован метод Гира и предложена его параллельная реализация в системе CUDA, которая позволяет проводить моделирование десятков тысяч химических реакций с высокой скоростью. Достигаемое ускорение составляет более 5 раз по сравнению с традиционными системами.

5. Построен программный комплекс для решения гидродинамических задач с физико-химическими превращениями, сочетающий в себе решение уравнений Навье – Стокса, Пуассона, диффузии, химической кинетики в системе параллельного программирования CUDA. (Получено свидетельство о государственной регистрации программы.)

6. Показано применение разработанных алгоритмов для изучения течений, сопровождающихся большим количеством химических реакций. Доказана возможность применения ускоренных алгоритмов на графических ускорителях. Достигаемое ускорение по сравнению с обычными системами составляет более 20 раз. Это позволяет быстрее получать новые результаты, рассматривать больше различных вариантов, условий протекания процессов. Кроме того, становится возможным получение результатов в режиме реального времени.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ РАБОТЫ

Публикации в изданиях, рекомендованных ВАК РФ:

1. Ясинский, Ф.Н. О решении уравнения Навье – Стокса в переменных "Функция тока – Вихрь" на многопроцессорной вычислительной машине с использованием системы CUDA / Ф.Н. Ясинский, А.В. Евсеев // Вестник ИГЭУ. – 2010. – Вып. 3. – С. 73-75.

2. Ясинский, Ф.Н. Использование системы с несколькими ускорителями CUDA для решения уравнения Навье – Стокса в переменных "Функция тока – Вихрь" / Ф.Н. Ясинский, А.В. Евсеев // Вестник ИГЭУ. – 2010. – Вып. 4. – С. 86-89.

3. Евсеев, А.В. О моделировании физико-химических течений при большом числе реагирующих веществ [Электронный ресурс] // Физико-химическая кинетика в газовой динамике. – 2011. – Т.12. – Режим доступа <http://www.chemphys.edu.ru/pdf/2011-05-05-001.pdf>

4. Евсеев, А.В. О моделировании течения воздушных масс, осложнённых физико-химическими процессами, при большом числе реагирующих веществ с использованием графических ускорителей / А.В. Евсеев // Вестник ИГЭУ. – 2011. – Вып. 3. – С. 40-43.

Труды конференций:

5. Евсеев, А.В. Распараллеливание методов решения уравнения Пуассона / А.В. Евсеев, Ф.Н. Ясинский // Высокопроизводительные параллельные вычисления на кластерных системах: материалы Девятой международной конференции-семинара (Владимир, 2-3 ноября 2009 г.). – Владимир: Изд-во ВГУ, 2009. – С. 167-168.

6. Евсеев, А.В. Вопросы распараллеливания уравнения Пуассона и сравнение эффективности различных вариантов / А.В. Евсеев // Высокие технологии, исследования, промышленность: сб. тр. Девятой международной научно-практической конференции "Исследование, разработка и применение высоких технологий в промышленности". – СПб.: Изд-во Политехн. ун-та, 2010. – Т.3. – С. 46-52.

7. Евсеев, А.В. О решении уравнения Навье – Стокса в переменных "Функция тока – Вихрь" с использованием системы с несколькими графическими ускорителями / А.В. Евсеев, Ф.Н. Ясинский // Высокопроизводительные параллельные вычисления на кластерных системах (НРС-2010): материалы X международной конференции (Пермь, 1-3 ноября 2010 г.). – Пермь: Изд-во ПГТУ, 2010. – Т.1. – С. 245-251.

Тезисы:

8. Евсеев, А.В. О моделировании свойств тканей, содержащих магнитные наночастицы / А.В. Евсеев, Ф.Н. Ясинский // Молодые учёные – развитию текстильной и лёгкой промышленности (ПОИСК-2009): сб. материалов межвузовской научно-технической конференции аспирантов и студентов / ИГТА. – Иваново, 2009. – С. 175-176.

9. Евсеев, А.В. Моделирование течения жидкости на гетерогенном кластере с использованием MPI, OpenMP и CUDA / А.В. Евсеев // Молодые учёные – развитию текстильной и лёгкой промышленности (ПОИСК-2011): сб. материалов межвузовской научно-технической конференции аспирантов и студентов / ИГТА. – Иваново, 2011. – С. 333-334.

10. Евсеев, А.В. О моделировании течения жидкости с помощью уравнения Навье – Стокса в переменных "Функция тока – Вихрь" на гетерогенном суперкомпьютере / А.В. Евсеев, Ф.Н. Ясинский // Сборник трудов V Международной научно-практической конференции "Пожарная и аварийная безопасность". – Иваново, 2010. – Ч.1. – С. 169-170.

Прочие издания:

11. Евсеев А.В. Методы решения уравнения Пуассона / А.В. Евсеев, Ф.Н. Ясинский / ИГТА. – Иваново, 2009. – 108 с.

12. Малый И.А. Математическое моделирование горения и взрыва. Ч.1. Метод Гира. Химическая кинетика. Течения, сопровождающиеся горением / И.А. Малый, Ф.Н. Ясинский, О.В. Потемкина, С.Г. Сидоров, А.В. Евсеев, И.Ф. Ясинский. – Иваново, 2011. – 92с.

13. Евсеев А.В. Система численного параллельного решения уравнения Пуассона в системах MPI и CUDA "ParallelPoissonSolvers"/ А.В. Евсеев // Свидетельство о государственной регистрации программы для ЭВМ №2011617028 от 9.09.2011.

14. Евсеев А.В. Система моделирования распространения воздушных масс в районе над лесным пожаром "ForestFireDynamics"/ А.В. Евсеев // Свидетельство о государственной регистрации программы для ЭВМ №2011616155 от 5.08.2011.

