

INSIDE: Криптоаналитические рекорды «Ломоносова»

с. 43-46

Стратегическая цель

ТРИ ПУТИ ВЗАИМОДЕЙСТВИЯ

Параллельные вычисления в профильных лицах

с. 22

www.supercomputers.ru

3 (18) лето 2012

СУПЕР КОМПЬЮТЕРЫ

Не революция, но переворот

Торговля
с Гамбургом с. 6

с. 55

Математическое моделирование транспортных потоков

Особое мнение:
Сложные задачи
аэрогазодинамики

с. 59



T-ПЛАТФОРМЫ

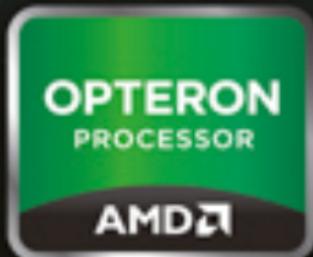
T-Blade v-class

Высвобождение ядерной энергии



- Шасси 5RU для установки в стандартные 19" шкафы
- До 10 узлов x86 или до 5 узлов x86+GPU
- Конфигурации на базе 4-, 8-, 16-, 24- или 32-ядерных процессоров Intel и AMD
- До 256 Гб оперативной памяти на узел
- До 20 процессоров на шасси с вычислительными модулями S
- До 5-10 ускорителей и 10 процессоров на шасси с вычислительными модулями F
- Интегрированный в шасси модуль управления
- Оптимизированное энергопотребление и пониженный шум.

Система T-Blade V-Class
будет представлена
на международной конференции
«Параллельные вычислительные
технологии (ПаВТ) 2012»,
проходящей 26-30 марта 2012 г.
в Новосибирске.



Информацию о серверах T-Платформы на базе процессоров AMD Opteron™ смотрите на www.t-platform.ru

©2012 Advanced Micro Devices, Inc.
AMD, логотип стрелы AMD, AMD-Opteron и логотип на комбинации являются товарными знаками компании Advanced Micro Devices, Inc.

СУПЕР КОМПЬЮТЕРЫ

ЛЕТО 2012

Председатель редакционного совета

Владимир ВОЕВОДИН

Vladimir.voevodin@supercomputers.ru

Над номером работали:

Выпускающий редактор

Игорь ЛЁВШИН

Igor.levshin@supercomputers.ru

Арт-директор

Виктория ИВАШКОВА

Корректор

Юлия ГОЛОМАЗОВА

Тексты:

Андрей АДИНЕЦ

С. И. БАСТРАКОВ

А. В. ГИЛЛАВЫЙ

Евгений ГРЕЧНИКОВ

Игорь ЛЁВШИН

Дмитрий МУРАШОВ

Игорь ОБУХОВ

Даниэль ОРЛОВ

Эдуард ПРОЙДАКОВ

Андрей СЁМИН

В. А. СОЛОВЬЁВ

Сергей СЫСОВЕВ

Александр ФРОЛОВ

Андрей ЧЕЛЫШЕВ

В. И. ШАЛАЕВ

Иллюстрации:

Александр ЖЕЛОНКИН

Владимир КАМАЕВ

Учредитель

Даниэль ОРЛОВ

Издатель

ООО «Издательство СКР-Медиа»

SCR-MEDIA

Генеральный директор

Даниэль ОРЛОВ

Daniel.orlov@supercomputers.ru

Директор по развитию

Андрей ЧЕЛЫШЕВ

Andrei.chelyshev@supercomputers.ru

Адрес редакции и Издателя:

117342, Москва, ул. Бултерова, 17Б

www.supercomputers.ru

Издание «СУПЕРКОМПЬЮТЕРЫ» зарегистрировано
Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций
(Роскомнадзор). Свидетельство о регистрации

СМИ ПИ №ФС77-38346 от 10.12.2009

Тираж 5000 экз.

Отпечатано

Типография ООО «Вива-Стар»

107023, Россия, Москва,

ул. Электрозаводская, дом 20, стр. 3

www.vivastar.ru

Редакция не несет ответственности за достоверность
информации, содержащейся в опубликованных
рекламных материалах. Мнение редакции может не
совпадать с мнением авторов статей.

Присланные материалы не рецензируются.

Цена свободная

© ООО «Издательство СКР-Медиа» 2012

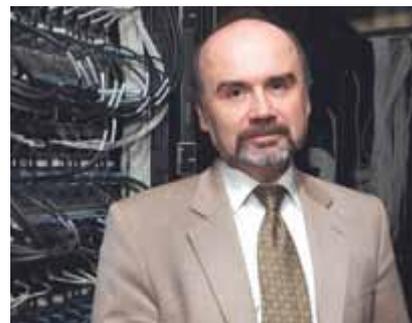
© «СУПЕРКОМПЬЮТЕРЫ» 2010-2012

От редакции

Еще памятливы не столь давние времена, когда в академической среде отношение к «прикладникам», а особенно к «численникам», было покровительственно ироничным: дескать, «мы занимаемся высокой наукой, а тут дырки в перфокартах пробивают». Сейчас это забавный курьез, а тогда людям, начинавшим отечественную вычислительную отрасль, подобное казалось обидным. Времена другие, но эхо былого снобизма нет-нет да оттолкнется от стен студенческих курилочек, пролетит по форумам, осядет на задних рядах аудиторий. Сейчас это скорее защитная реакция молодых на «модную тему», нонконформизм, бунтарство наоборот. Суетный век порождает ломателей машин. А мы их строители.

Мы не занимаемся вечностью, зато мы занимаемся будущим, которое слишком рьяно стремится стать прошлым. И приходится нагонять и нагонять, увеличивать скорости, прилагать все больше усилий. Такая у нас работа. Такое дело, которому мы посвятили себя.

Используя открытые фундаментальной наукой высочайшие абстракции,



*...Я сделался ремесленник: перестал
Придал послушную, сухую беглость
И верность уху. Звуки мертвев,
Музыку я развлял, как труп. Поверил
Я алгеброй гармонию. Тогда
Уже дерзнул, в науке искушенный,
Предаться неге творческой мечты.*

А. С. Пушкин, «Моцарт и Сальери»

мы те, кто помогает математике проникать в реальную жизнь, в то, что становится материальным миром вокруг каждого. Идея обретает силу в формулах, обрастает, как плотью, строчками программного кода – и вот оно, рождение новой сущности, того, чего не было до и что есть теперь. Это не просто решение задач, игра ума и чистое творчество – это созидание.

Дырки в перфокартах... Листки картона, как продолжение дерева, древа познания, начало чего-то нового. Какая же это была прекрасная метафора для всей вычислительной отрасли! Почему-то мне кажется, что люди, начинавшие нашу науку, чувствовали это. Еще можно успеть спросить их учеников, они помнят.

А нам надо запомнить то, что происходит сейчас. Мы перешли порог нового. Надо только внимательнее смотреть по сторонам. Мы уже в неизведанном.

Владимир В. Воеводин,
Председатель редакционного совета
журнала «Суперкомпьютеры»

В номере



- 3 Сибирский федеральный университет**
О работе Центра высокопроизводительных вычислений СФУ
- 6 Не революция, но переворот**
ISC2012. Прощание с Гамбургом
- 22 Три пути взаимодействия**
Параллельные вычисления в профильных лицах
- 27 Подмосковные вечера**
Семинар компании STORUS
- 28 С играми не шутят**
GTC в Сан-Хосе
- 30 Квантовые вычисления**
От бита к кубиту
- 34 Появление машин шестого поколения затянулось**
Пора избавляться от засилья алгоритмов
- 36 YarcData uRiKA**
Эврика в семантическом анализе
- 40 Бесплатное охлаждение**
в суперкомпьютерах и центрах обработки данных
- 43 Криптоаналитические рекорды «Ломоносова»**
Трудовые будни самого мощного в России суперкомпьютера
- 47 Численное моделирование плазмы на суперкомпьютерных системах**
- 51 Системы управления экспериментом**
- 55 Математическое моделирование транспортных потоков**
- 59 Применение параллельных вычислительных технологий**
для решения сложных задач аэрогазодинамики в ВУ ФАЛТ МФТИ

ISC2012
Прощание
с Гамбургом

Редакционный совет:

Владимир Воеводин, д. ф.-м. н., чл.-корр. РАН, НИВЦ МГУ, г. Москва
 Виктор Гергель, д. т. н., ННГУ, г. Нижний Новгород
 Юрий Зеленков, к. ф.-м. н., НПО Сатурн, г. Рыбинск
 Вячеслав Ильин, д. ф.-м. н., НИИЯФ МГУ, г. Москва
 Леонид Соколинский, д. ф.-м. н., ЮУрГУ, г. Челябинск

Михаил Токарев, НОЦ «Нефтегазовый центр МГУ», г. Москва
 Александр Томилин, д. ф.-м. н., ИСП РАН, г. Москва
 Борис Четверушкин, д. ф.-м. н., чл.-корр. РАН, ИПМ им. М. В. Келдыша, РАН, г. Москва
 Борис Шабанов, к. ф.-м. н., МСЦ РАН, г. Москва

Сибирский Федеральный Университет (СФУ)

Член Суперкомпьютерного консорциума университетов России

www.sfu-kras.ru

С момента основания СФУ одной из приоритетных задач развития была задача создания мощного центра высокопроизводительных вычислений для обеспечения современными инструментами научных и инженерно-технических расчетов. В 2007 году в университете появляется суперкомпьютер, вошедший в список TOP500. В настоящее время суперкомпьютерный комплекс СФУ входит в престижный список TOP50 российских суперкомпьютеров.

В 2010 году СФУ возглавил рейтинг классических и национальных исследовательских университетов по уровню развития инноваций и коммерциализации разработок.

Центр высокопроизводительных вычислений

<http://clustersfu-kras.ru> (<http://clustersfu-kras.ru/?lang=en>)

Руководитель – к. т. н. Д. А. Кузьмин

Центр обслуживает научные коллективы СФУ, взаимодействует с наукоемкими предприятиями Сибири, отделениями РАН, университетами России и Европы. В составе центра три высокопроизводительных системы, объединяющих 280 счетных вычислительных серверов, SAN-сеть 8 Гбит/с, InfiniBand 20 Гбит/с. Гибридная составляющая комплекса – платформа NextIO Vcore Express с 4 GPU Tesla M2090.

Установлены программные комплексы, позволяющие проводить ресурсоемкие расчеты в различных областях науки. Среди установленного ПО – ANSYS, Matlab, GAMESS-US, CFOUR, MRCC, NWChem, ORCA, AMBER 11, SigmaFlow. Все программное обеспечение имеет академические лицензии или является продуктами с открытым кодом.

Для предоставления удаленного

доступа к высокопроизводительным ресурсам СФУ имеет в своем распоряжении оптоволоконные каналы связи, обеспечивающие передачу информации со скоростью до 10 Гбит/с, и канал доступа в общие сети Internet со скоростью 200 Мбит/с. Центр активно сотрудничает с компаниями IBM, Intel, NVIDIA, AMD, Cisco, VMware, Novell.

Сотрудники центра специализируются в области организации управления высокопроизводительными комплексами. Силами центра реализуется проект «Высокопроизводительные вычисления как сервис». Проект обеспечивает создание универсальной инфраструктуры обработки данных, в которой функционирует множество сервисов, решающих конкретные прикладные задачи. Доступ пользователей к распределенным ресурсам центра осуществляется через единый веб-портал. Реализована комплексная система управления, позволяющая администраторам эффективно управлять вычислительными ресурсами, вводить новое

специализированное ПО, осуществлять мониторинг ресурсов и задач пользователей.

На базе центра реализован сервис 3D-рендеринга www.rendermama.com, ориентированный на широкий круг пользователей, работающих с 3D-графикой, – дизайнеров, аниматоров, архитекторов. Система поддерживает загрузку проектов, выполненных в пакетах 3DMax, Maya, Cinema4D. Рендеринг проектов выполняется в автоматическом режиме на распределенных суперкомпьютерных ресурсах центра с использованием ПО MentalRay и V-Ray. Центр высокопроизводительных вычислений СФУ является базой для обучения магистров и бакалавров в области суперкомпьютерных технологий (СКТ).

Базовые и специальные курсы по СКТ читаются ведущими специалистами Института Математики (ИМ) и Института космических и информационных технологий (ИКИТ) СФУ. Аккредитована магистратура по профилю «Высокопроизводительные вычислительные системы». Работает программа повышения квалификации «Параллельное программирование».

Научными группами СФУ ведутся фундаментальные и прикладные исследования в области суперкомпьютерных технологий, среди которых:

– **Инструментальная и языковая поддержка архитектурно-независимого параллельного программирования.**

Научной группой под руководством

профессора А. И. Легалова решается проблема создания архитектурно-независимого параллельного программного обеспечения. Предлагаемый подход опирается на функционально-потокую парадигму параллельного программирования, обеспечивающую написание, отладку, верификацию программ без привязки к конкретным вычислительным ресурсам. Инструментом исследований является созданный группой язык программирования «Пифагор». В настоящее время проводятся исследования, связанные с использованием данного языка в качестве высокоуровневой надстройки при проектировании ПЛИС, а также для выполнения параллельных программ в многоядерных процессорах.

– **Технологии и инструментальные средства архитектурно-независимого проектирования высокопроизводительных однокристалльных систем.**

Одним из основных направлений деятельности коллектива специализированной научно-учебной лаборатории «Микропроцессорных систем» под руководством профессора О. В. Непомнящего является создание новых технологий проектирования реконфигурируемых и динамически реконфигурируемых систем на кристалле (System on Chip SoC and Dynamically Reconfigurable System on Chip – DRSoC). Разрабатываемые в рамках научных исследований методы и инструментальные средства сквозного высокоуровневого проектирования и верификации высокопроизводительных систем на кристалле позволяют эффективно работать над проектами SoC, создавая готовые решения, функционалы и топологии однокристалльных систем, которые, с одной стороны, не зависят от конечной реализации и позволяют обеспечить максимальную абстрагируемость алгоритмов функционирования от архитектуры целевого кристалла, а с другой стороны, максимально приближены к представлению на языках описания аппаратуры и имеют возможность



к. т. н. Д. А. Кузьмин

адекватного отображения на нижние слои в иерархии проектирования SoC.

При таком подходе разработчики оперируют не на уровне готовых аппаратных платформ или ранее разработанных и верифицированных блоков (IP block), а на уровне принципов системной организации вычислительного процесса с эффективным переносом полученной модели на целевой кристалл.

Результаты исследований применяются при разработке систем для освоения космического пространства, в том числе систем управления энергообеспечением и исполнительной автоматикой космических аппаратов.

– **Компьютерное моделирование больших алгебраических систем, вычислительная теория групп (ВТГ).**

Исследования научного коллектива под руководством докторов наук А. К. Шлепкина и А. А. Кузнецова включают в себя проектирование, анализ алгоритмов и структур данных для вычисления различных характери-

стик (чаще всего – конечных) групп. Область интересна исследованием важных с различных точек зрения групп, данные о которых невозможно получить вычислениями вручную. Разработан комплекс алгоритмов и программ для решения задач теории групп. В результате на суперкомпьютере СФУ решены проблема Ч. Симса о структуре соотношений в свободной двупорожденной бернсайдовой группе периода пять – группе $B(2,5)$ и проблема о распознаваемости по спектру проективной специальной линейной группы размерности два над полем из семи элементов – группе $L_2(7)$. Изучена структура ряда групп с различными условиями насыщенности.

– **Параллельные вычисления в задачах механики сплошных сред.**

Институт математики СФУ активно сотрудничает с Институтом вычислительного моделирования СО РАН. На базовой кафедре вычислительных и информационных технологий, которой руководит член-корреспондент РАН В. В. Шайдулов, студенты и магистранты участвуют в нескольких интеграционных проектах СО РАН по решению задач механики сплошной среды с использованием высокопроизводительных вычислений. Так, группа ИМ СФУ и ИВМ СО РАН под руководством профессора В. М. Садовского разрабатывает и исследует математические модели механики деформируемых неклассических сред. Модели учитывают различное сопротивление растяжению и сжатию, структурные неоднородности среды (грунты, горные породы, углеграфиты, полимеры, пористые среды). Разработан комплекс прикладных параллельных программ для решения задач о распространении волн напряжений и деформаций в средах со сложными реологическими свойствами на суперкомпьютерах кластерной архитектуры. На рис. 1 приведены результаты расчетов пространственной задачи о периодическом воздействии сосредоточенного момента на

границе образца из синтетического полиуретана. Впервые установлено, что в модели моментной упругой среды существует резонансная частота вращательного движения частиц микроструктуры материала, которая не зависит от размеров и формы исследуемого образца.

– **Развитие теории фотофоретического движения аэрозольных частиц и кластеров.**

Научная группа под руководством профессора А. А. Черемисина ведет исследование фотофоретического движения аэрозоля. С этой целью были разработаны соответствующие алгоритмы и осуществлено компьютерное моделирование.

Установлено, что для аэрозольных частиц, находящихся в атмосфере Земли, существуют зоны устойчивого и неустойчивого движения против силы тяжести. Выявлены динамические эффекты, приводящие к образованию аэрозольных слоев, рассчитано движение сложных аэрозольных кластеров. Установлено, что существуют кластеры, способные двигаться против силы тяжести. Обнаружен новый тип взаимодействия аэрозольных частиц, вызываемый фотофорезом.

Разработанный алгоритм позволяет рассчитать фотофоретическую силу, действующую на аэрозольную частицу, кластер или группу частиц, а также силу, возникающую между частицами, между частицами и кластером. Также рассчитывается момент фотофоретической силы и четыре тензора вязких сил, что позволяет провести расчет движения частицы.

Разрабатывается версия алгоритма под CUDA, которая позволит строить модели тепловой коагуляции с учетом фотофоретического взаимодействия и газокинетических эффектов, что до сих пор плохо изучено. Также открывается возможность моделирования транспорта нанотрубок и наноструктур лазерными пучками.

– **Разработка методов и алгоритмов анализа и интерпретации космоснимков**

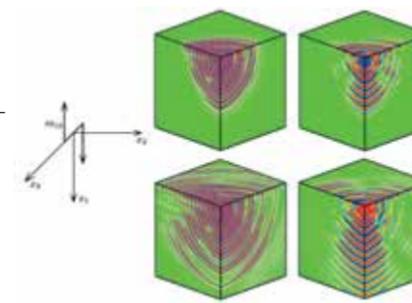


Рис. 1. Схема нагружения и поверхности уровня угловой скорости для нерезонансной (слева) и резонансной (справа) частот в различные моменты времени

Для решения задач предварительной и тематической обработки космоснимков установлен специализированный программный комплекс на базе MATLAB, создана система приема, обработки и хранения космических снимков, обеспечивающая полное покрытие территории Красноярского края. Решаются задачи сегментации облачности на спутниковых данных, расчет вегетационных индексов для агроландшафтного районирования.

Работы ведутся творческим коллективом ИКИТ при поддержке Министерства информатизации Красноярского края.

– **SigmaFlow – решение задач гидродинамики, теплообмена и горения.**

Специалистами Красноярского филиала института теплофизики СО РАН, кафедры теплофизики СФУ и фирмы ООО «ТОРИНС» создана программа SigmaFlow, предназначенная для решения широкого класса задач гидродинамики, теплообмена и горения.

SigmaFlow позволяет формировать геометрию расчетного объекта, готовить расчетные сетки, выполнять собственно вычисления и проводить анализ результатов моделирования с использованием графических средств. Программа дает возможность выполнять расчеты на кластерах под управлением Windows или Linux. Численная методика, заложенная в программу, основывается

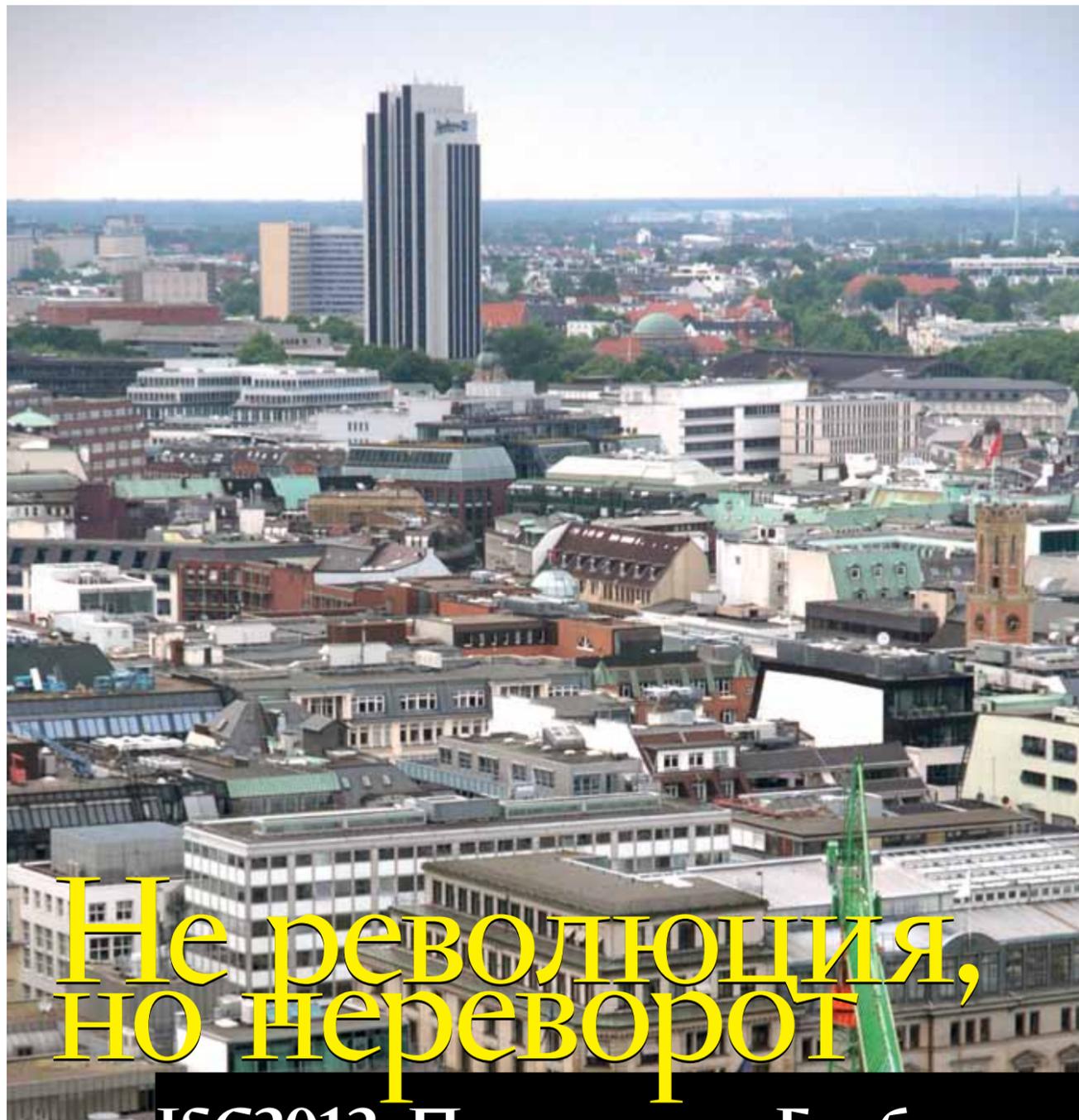
на методе конечного объема для неструктурированных сеток, что обеспечивает консервативность алгоритма и позволяет моделировать процессы в геометрически сложных объектах.

Для аппроксимации дифференциальных уравнений используются устойчивые схемы высокого порядка точности. Связь между полями скорости и давления реализуется при помощи процедуры расщепления. Системы разностных уравнений решаются итерационным способом с применением многосеточных методов. Коллективом разработчиков программы SigmaFlow накоплен большой опыт расчета «реальных» прикладных задач. Имеется учебная версия программы.

– **Моделирование механизмов химических и адсорбционных процессов.**

Научная группа в составе О. Б. Гаджиева (Институт химии высококачественных веществ им Г. Г. Девятовых РАН) и А. И. Петрова (СФУ) является примером академической мобильности нового поколения. Работа ведется в дистанционном режиме, объединяя научные ресурсы СФУ, институтов РАН и ведущих зарубежных вузов.

Решаются задачи в области моделирования механизмов химических и адсорбционных процессов, которые открывают широкие перспективы для расширения представлений о сущности химических процессов, приближают к ответам на вопросы о зарождении жизни как открытой неравновесной химической системы. Проведено полномасштабное моделирование адсорбционных процессов на поверхности льда важных для химии атмосферы и моделирования пребиотического периода эволюции Земли. Достигнут вычислительный рекорд для квантово-химического изучения механизма реакции $2NO + O_2 \rightarrow 2NO_2$. Установлен механизм и причина стереоселективности каталитической реакции конденсации в системе, состоящей из 100 атомов с участием производного L-пролина. ■■■



Не революция, но переворот

ISC2012. Прощание с Гамбургом

Текст Игорь Лёвшин, Даниэль Орлов, Андрей Сёмин, Андрей Чельшев

В чудесный город Гамбург деятели HPC-индустрии вернутся разве что в качестве туристов: следующую международную конференцию ISC проведут уже в Лейпциге. Ну а в этом году все шло по плану: немного подросло количество участников и стран, те, кого ждали, прочитали доклады и поучаствовали в дискуссиях.

Экзаскейл по-прежнему у всех на уме и на языке. Но ключевое слово теперь «энергопотребление». Разговоры о флопсах звучат уже как-то несовременно. То ли дело флопсы на ватт. Не скажешь, что ISC2012 – парад сенсаций, но ощущение перемен определенно есть.

Гамбит IBM

Не то чтобы триумф IBM в TOP500 стал полным сюрпризом. Суперкомпьютер не иголка в стоге сена, его не соберешь в гараже втайне от соседей. Но массивное наступление Blue Gene/Q поразило многих своей наглядностью. И дело тут отнюдь не только в первой и прочих престижных строчках в TOP500. Они и в GRAPH500 на первых ролях, а в GREEN500 все первые 20 мест занимают Blue Gene/Q!

Шокирующий отказ IBM от поставки Blue Waters в Университет Иллинойса похож задним числом на гамбит: IBM жертвует слона и выигрывает партию. На наш взгляд, это неплохие новости для HPC-бизнеса. Если говорить о процессорах, то с HPC-индустрией происходит, кажется, лучшее, что с ней вообще может произойти: состязание разных архитектур. Наконец-то мы видим в списках рядышком и x86, и PowerPC (в списках они фигурируют под загадочным даже для многих сотрудников IBM именем Power V90), и SPARC. Кстати говоря, и величественный IBM POWER7 отнюдь не сошел со сцены: в списке он представлен 15 системами – от Хоккайдо до Словакии и Польши, ну и, конечно, США. Причем не только в системах Power 775, но и в Hitachi SR16000. В списке есть и 6 машин с архитектурой SPARC. Все они в Японии, а три из них – PRIMEHPC FX10 – младшие братья сенсационного K Computer. Fujitsu удается все-таки тиражировать сложные системы со SPARC и интерконнектом TOFU, хотя и не в таких масштабах, как IBM Blue Gene/Q.

В России нелегко найти предпринимателя или руководителя IT-отдела или института, сильно интересующегося, скажем, тем же PowerPC или SPARC. Жаль. Нам кажется, это почти традиция. Для нашей страны характерны были крайности: то строят уникальные машины усилиями сотни НИИ (я говорю об «Эльбрусах»), то спохватываются и копируют IBM и PDP, плодя бесчисленные «ЕС» и «СМ». Существует выражение «зоопарк компьютеров». Зоопарк – роскошь для страны, но поощать его полезно и интересно.

TOP500

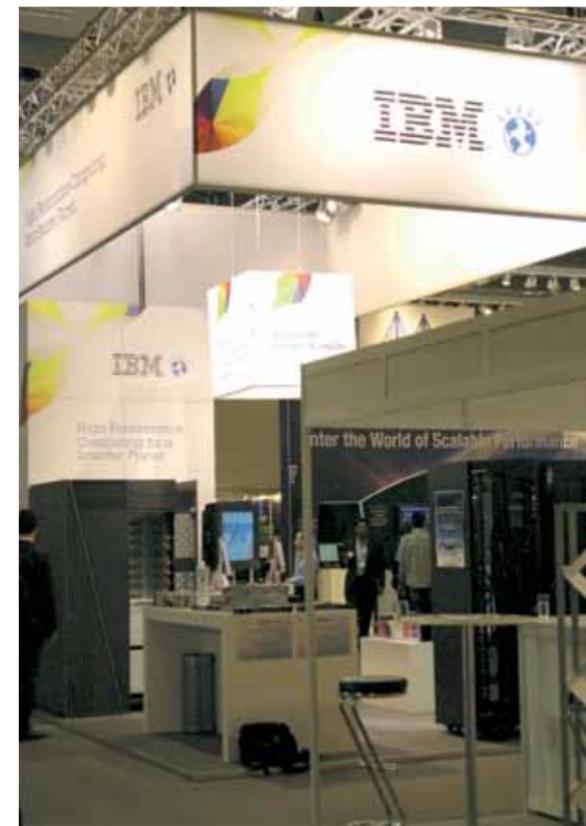
Итак, в первой десятке 39-го списка TOP500 произошли революционные

изменения: 6 из 10 систем – новые, да и список первых 50 тоже обновился более чем на 60%. Возглавила список IBM Sequoia с рекордной производительностью 16.3 ПФлопс на тесте Linpack. Система установлена в LLNL – Lawrence Livermore National Laboratory (США).

Sequoia, суперкомпьютер семейства IBM BlueGene/Q состоит из 96 стоек, которые вмещают 1.6 миллиона процессорных ядер (впервые за всю историю TOP500) и 1.6 Петабайта памяти. Пиковая производительность вычислительного гиганта составляет 20.1 ПФлопс. Sequoia будет использоваться в основном для моделирования ядерных взрывов и заменит суперкомпьютеры ASC Purple и Blue Gene/L Ливерморской национальной лаборатории. Когда система не будет загружена приоритетными вычислениями, на ней будут решать задачи в области астрономии, энергетики, изучения человеческого генома и изменения климата.

Sequoia также поставила рекорд по энергосбережению: потребляемая ею мощность составляет всего 7.9 Мегаватта.

Вторая строчка досталась бывшему лидеру рейтинга – японской системе K Computer (производительность 10.5 ПФлопс). Дальше идет еще одна новая инсталляция: IBM BlueGene/Q system (Mira, 8.2 ПФлопс) в другой национальной лаборатории США – Argonne National Laboratory. Самая мощная европейская система – SuperMUC архитектуры IBM iDataPlex в Leibniz Rechenzentrum (Германия), она занимает 4-ю позицию с производительностью 2.9 ПФлопс. Система исследовательского центра в Юлихе (FZJ) – JuQUEEN BlueGene/Q – находится на восьмой позиции (1.4 ПФлопс), пропуская вперед итальянскую систему Fermi вычислительного центра CINECA – опять архитектуры IBM BlueGene/Q (1.7 ПФлопс). Суперкомпьютер Jaguar, установленный в Окриджской национальной лаборатории США, увеличил производительность до 1.9 ПФлопс за счет перехода на более быстрые процессоры и занял 6-е место списка. Супер-ЭВМ Tianhe-1A, принадлежащая Национальному суперкомпьютерному центру в г. Тяньцзинь сохранила производительность 2.6 ПФлопс и опустилась





Александр Чичковский (Storus):

Увеличение количества ядер и рост частоты неизбежно влечет за собой повышение теплового пакета процессора. Практически все производители признали, что эффективно отводить тепло возможно только применяя охлаждение на основе воды. Речь идет о системах, в которых полностью исключается промежуточное воздушное охлаждение. На конференции были представлены как готовые решения, так и концептуальные разработки в этой области. Такие системы помимо повышения производительности открывают возможности использования технологий фрикулинга и охлаждения перегретой жидкостью, а также позволяют значительно повысить плотность вычислительных ресурсов. Вторым направлением, которое активно обсуждалось участниками конференции, стали системы хранения данных. В последние годы мы наблюдаем активный рост производительности вычислительных узлов, вся ИТ-индустрия работает в этом направлении. Проблемы хранения данных в области HPC оказались отодвинутыми на второй план, и развитие в этой отрасли не было столь динамичным. Существующие на рынке решения не могли обеспечить растущие требования потребителей в полном объеме, как по емкости, так и по производительности. Ответом на сложившуюся ситуацию стали анонсы новых СХД для HPC сразу от нескольких производителей. Новые системы представили CRAY, Xyratex, DDN, а наиболее интересными мне показались решения от компании Rapasas.

со второго места на пятое. Московский суперкомпьютер «Ломоносов» отступил на 22-е место, на данный момент это самая мощная российская система в списке. Всего отечественных супер-ЭВМ в рейтинге пять, на один компьютер меньше, чем у Австралии, и в два раза меньше, чем у Канады. По-прежнему лидером по количеству систем в списке остаются США с 252 компьютерами, правда, это на 11 ЭВМ меньше, чем шесть месяцев назад. Несмотря на то, что США вернули лидерство в TOP500 впервые после 2009 года, количество их систем в первой десятке уменьшилось с пяти до трех. Первая десятка 39-го списка TOP500 сильно расширила свою географию, помимо США туда входят Китай, Япония, Германия, Франция и Италия. Обладатель второго места тоже не изменился – это

Китай с 68 системами, его потери составляют минус шесть компьютеров по сравнению с ноябрем 2011 г. Далее идут страны, имеющие более десяти супер-ЭВМ в списке, – Япония (35), Великобритания (25), Франция (22) и Германия (20).

Для того чтобы попасть в новый рейтинг, потребовалась производительность не менее 60 ТФлопс, в 2004 году системы с такими показателями занимали первые места TOP500. Суммарная производительность систем списка составила 123.4 ПФлопс, что практически вдвое больше аналогичных показателей предыдущей редакции рейтинга – 74.2 ПФлопс.

Наибольшее количество систем в списке у компании IBM – 213, то есть практически половина рейтинга. Второе место по количеству супер-ЭВМ в списке принадлежит HP (138), далее идут Cray (26), Appro (19), Bull (16), SGI (16) и Dell (12). У остальных производителей по одной системе в рейтинге. Что касается следующей редакции рейтинга TOP500, которая выйдет в ноябре 2012 года, то большинство аналитиков склонны считать, что мы сможем опять наблюдать серьезные изменения в первой десятке списка. Планируется, что к этому времени будут введены в эксплуатацию как минимум три суперкомпьютера производительностью в десятки петафлопс. Это суперкомпьютер Stampede для Техасского вычислительного центра (TACC) и системы Titan и Blue Waters в Окриджской Национальной Лаборатории (ORNL) и Национальном Центре Суперкомпьютерных Приложений (NCSA) соответственно. Примечательно, что в Stampede будет использоваться сопроцессор Xeon Phi (Knights Corner), а в Titan и Blue Waters будет использована новая технология NVIDIA Kepler.

Три богатыря в младенчестве

Не только у нас было ощущение, что важнейшее событие ISC этого года – сессия «Европейский путь к экзаскейлу!» Три доклада о трех важнейших проектах собрали вместе: Алекс Рамирез рассказывал о *Mont Blanc* в Суперкомпьютерном центре Барселоны, Ульрих Брюнинг – о разработках DEEP в Гейдельберге, а Марк Парсонс – о CRESTA в Эдинбурге.

Девиз, который обнародовал руководитель проекта *Mont Blanc* – «Чтобы стать лучшим, надо быть другим», – можно отнести ко всем трем. Все три проекта поддерживает Европейская комиссия.

Mont Blanc

Проект *Mont Blanc* интересует очень многих, хотя впечатляющих достижений – числа узлов, петафлопсности, мест в списках – нет и в ближайшее время они вряд ли будут. Даже прототип энергоэффективной системы запланирован на 2014 год, а система большой производительности должна появиться в 2017 году. Сейчас даже не приняты решения о том, какие конкретно процессоры будут в системе. Этим в том числе проект и интересен: главное в нем – исследование возможностей создания экономичных (с точки зрения энергии и не только) HPC из стандартных недорогих компонент, способных эффективно работать с петафлопсного масштаба приложениями (список которых, кстати, тоже еще не определен). Ну а прототипы незаменимы для разработки и испытания программного кода – это понимают и партнеры проекта (Bull, ARM, Gnodal и шесть европейских ведущих исследовательских центров), и другие разработчики.

Координатор проекта Алекс Рамирез считает: вполне можно ожидать, что мобильные процессоры вытеснят «обычные» в HPC, как в свое время микропроцессоры «убили» векторные компьютеры. Сейчас об этом говорить может, и рано, поскольку при типичной разнице в цене в 100 раз производительность мобильных меньше в 8, но можно ожидать, что через некоторое время они будут отставать всего лишь вдвое при том же соотношении цен. Прежде всего, *Mont Blanc* – это система на ARM. Рассматриваются и испытываются и Tegra 3, и ARM-процессоры от AMD, и Calxeda. И, конечно, в фокусе интереса *Mont Blanc* – будущие 64-битные 16-ядерные ARM Cortex-A15. Их Рамирез оценивает как потенциальных конкурентов архитектур BlueGene/Q. Однако идеологи *Mont Blanc* не считают получить всю мощность из одних лишь ARM-CPU. Основную работу будут выполнять GPU-ускорители, то есть система будет гетерогенным кластером с управляющими легкими CPU

архитектуры ARM и ускорителями GPU. Прототипирование именно такой архитектуры сейчас особенно важно. Но интегрированные GPU+CPU тоже важны и нуждаются в прототипах для исследования. В таких системах не надо тратить энергию на пересылку данных по PCIe и на графическую память GDDR5.

Ethernet выглядит органично в такой экономной системе. Один из партнеров проекта – компания Gnodal, производящая коммутаторы Ethernet. Рассматриваются и ARM SoC с интегрированными Ethernet-контроллерами. Это позволит более эффективно расходовать энергию на обмен данными с другими узлами. Главное, чтобы как можно больше производителей проявляли себя, вписывая в общие открытые стандарты. Но на интегрированный GPU при этом рассчитывать уже не приходится. Что касается программной части межузлового взаимодействия, то в *Mont Blanc* делают ставку на OpenMP, которая поддерживает асинхронные вычисления. Пересылка данных и счет должны происходить параллельно, а не последовательно, чтобы система могла эффективно работать даже при не очень высокой пропускной способности интерконнектов.

Экономить на памяти получается плохо. В системах с тысячами узлов трудно обойтись без ECC. Разве что стоит исследовать эффективность контроля

ошибок при помощи таких средств, как Tibidabo или Pedraforca, разработанных в рамках того же проекта.

DEEP (EXTOLL)

О другом интереснейшем проекте рассказывал Ульрих Брюнинг из Университета Гейдельберга. Идея проекта – в соединении не слишком хорошо масштабируемых архитектур (констелляций и кластеров графических ускорителей) с очень хорошо масштабируемыми системами (концепция «кластер-бустер»). В реальности речь идет о скоростных интерконнектах EXTOLL для узлов бустера с Intel MIC (теперь Xeon Phi), с кластерной частью (на Intel Sandy Bridge) все более-менее традиционно. Система будет работать с охлаждением горячей водой. Первая большая система такой архитектуры будет построена в 2014 году. Пока, как и в случае *Mont Blanc*, речь идет о прототипах.

Для эффективной работы узлов будет использоваться программная архитектура OpenMP, разработанная коллегами из Барселонского Суперкомпьютерного центра. Программы с MPI-кодом должны работать в такой архитектуре. Поддерживается и обмен данными по PGAS. Сеть EXTOLL, которая соединит все узлы бустера с Xeon Phi, гарантирует малые задержки. Это новая разработка с топологией 3D-тор как основной, но в связях при желании можно

связать другой топологией. EXTOLL рассчитана на малые накладные расходы даже при обмене очень короткими сообщениями. Это сверхмасштабируемая скоростная сеть (проектная пропускная способность – 120 Гб/с). Она поддерживает до 64 000 узлов. Сеть EXTOLL соединяет узлы бустера, а не кластеров. Узлы кластеров соединены с интерфейсами бустера сетью InfiniBand (сейчас там InfiniBand HCA Mellanox), при этом в EXTOLL есть механизмы SMFU для бесшовной работы



Виктор Гергель (ННГУ):

Прежде всего, заметен общий уверенный рост производительности, что отражено в новой редакции TOP500. Все прорывные технологии на выставке прекрасно представлены. Очень интересны сообщения и доклады сотрудников Intel, посвященные Xeon Phi (MIC). Порадовали отечественные «Т-Платформы», участвующие в совместном с европейцами проекте по созданию прототипа ExaFlop-системы. Общий тренд в развитии гибридных архитектур заметен и устойчив. Вместе с тем следует отметить, что ISC во многом «железная» конференция. Нашим университетам (в Гамбурге три стенда!) имеет смысл, не теряя позиций на этой конференции, подумать о консолидированном участии в международных мероприятиях, где больше алгоритмов и ПО.





Умные процессоры сделают больше с меньшими затратами.

Процессоры Intel® Xeon®



TB2-XN

на базе процессора Intel® Xeon® X5600

- Специализированная суперкомпьютерная блейд-система на базе процессоров Intel® Xeon® X5600
- 192 узла, 2304 ядра
- Интегрированные сети Infiniband QDR и Gigabit Ethernet
- Сверхвысокая вычислительная плотность - 27 TFlops в стандартной 19-дюймовой стойке
- Основа самого мощного суперкомпьютера в СНГ «Ломоносов»
- Выделенные сети барьерной синхронизации и глобальных прерываний
- Уникальная система воздушного охлаждения
- Максимальное энергопотребление 10.5 КВт на шасси



OS ClustrX T-Platforms Edition

- Суперкомпьютерная ОС нового поколения
- Представляет вычислительный кластер и его инфраструктуру как единый ресурс
- Поддержка петафлопс-систем
- Масштабируемая до десятков тысяч узлов система управления и мониторинга в реальном времени
- Поддержка популярных ОС вычислительных узлов
- Бинарная совместимость с приложениями под RHEL
- Поддержка гетерогенных систем
- Проактивный режим экономии электроэнергии

T-PLATFORMS

- Разработка и производство суперкомпьютерных платформ;
- Разработка системного программного обеспечения для высокопроизводительных систем;
- Аутсорсинг полного цикла высокопроизводительных вычислений;
- Проектирование, создание и управление суперкомпьютерными центрами.

«Т-Платформы» - международный поставщик
полного спектра решений и услуг
для высокопроизводительных вычислений.



**Мощный.
Интеллектуальный.**

с IB. За коммуникации по принципу zero-copy MPI отвечает механизм VELO, за доступ к удаленной памяти – механизм RMA. Пока сеть реализована на FPGA-чипах, но в конце 2012 года в узлах бустера она будет уже на ASIC, сеть тогда будет примерно в 2 раза быстрее. Система работает под Linux, оптимизированный под HPC, некоторые операции могут выполняться в обход операционной системы. Этот проект интересен как сам по себе, так еще и тем, что в альянс разработчиков входит и российская компания «Т-Платформы».

CRESTA

В свою очередь Марк Парсонс из Университета Эдинбурга рассказывал о проекте CRESTA. Это альянс для реше-

ния софтверных проблем, связанных с экзаскейлом. Собственно, большую часть презентации Парсонс живописал софтверные ужасы масштабирования грядущей эпохи и довольно плачевное нынешнее состояние отрасли – в противоположность вполне сноской ситуации у конструкторов самих систем. CRESTA занимается и системными разработками (systemware), и приложениями. Шесть приложений заслужили особое внимание CRESTA: GROMACS – биомолекулярные вычисления, пакет, разработанный KTH (Швеция) ELMFIRE Fusion – энергетика, ABO/JYU (Финляндия)

NeteLB – физиология человека, UCL (Великобритания)

IFS – предсказание погоды, ECMWF

(интернациональная разработка)

OpenFOAM – инженерный пакет, EPCC/HLR/ECF

Nek5000 – инженерный пакет, KTH (Швеция)

Эти приложения отобраны по принципу европейского происхождения и популярности в Европе.

Системное ПО, в разработках которого участвуют команды CRESTA, – это весь спектр, от операционных систем до инструментов.

Парсонс относится к тем, кто призывает переписывать по возможности алгоритмы с нуля, сразу имея в виду распараллеливание вычислений. Но и приспособлять существующий код разработчикам CRESTA тоже прихо-

дится. В качестве примера, где тупик распараллеливания близок, Парсонс привел быстрое преобразование Фурье. Но делать с этим что-то надо, потому что БПФ используется в огромном количестве популярных приложений. Значит, надо искать новые пути. Девиз CRESTA: давайте переписывать софт уже сейчас, а не в 2019 году, когда сами системы экзамасштаба появятся.

ВЕНДОРЫ в зрелости

ACER

Компания Acer совместно с Национальным Центром Высокопроизводительных Вычислений Тайваня (NCHC) наглядно продемонстрировали на своем стенде, что при помощи даже одного сервера можно эффективно решать сложные интерактивные HPC-задачи. Для этого они использовали сервер Acer AT350 F2, оснащенный GPU NVIDIA Tesla и Quadro и процессорами Intel Xeon серии E5. С такой конфигурацией вычислителя Acer и NCHC добились прекрасных результатов работы CFD-приложений, моделирующих в реальном времени обтекание воздушным потоком различных объектов. При моделировании специалисты использовали технологии дополнительной реальности, что позволило им в режиме реального времени наблюдать и оптимизировать работу приложений. Впечатляющих результатов команда из NCHC добилась также за счет того, что изначально оптимизировала алгоритмы расчетов под графические ускорители.

AMD

В новой редакции TOP500 у компании AMD хорошие показатели – 24 системы в первой сотне списка построены на базе ее процессоров. Правда, если оценивать весь рейтинг, то преимущество окажется на стороне их давнего конкурента, на процессорах которого работают 78% систем от всего списка. В этом году AMD не смогла порадовать нас на выставке большим количеством новых продуктов, зато наглядно и убедительно продемонстрировала, насколько эффективно используются технологии AMD ее партнерами и конечными пользователями. Наиболее интересными были примеры сотру-

ничества с такими всемирно известными компаниями, как LSTC, Mellanox и CAPS. Итак, компания Livermore Software Technology Corporation (LSTC) выпустила новую версию ПО LS-DYNA, специально оптимизированную под процессоры серии AMD Opteron 6200. Напомним, что программа LS-DYNA предназначена для анализа высокочастотных и быстротекущих процессов в задачах механики твердого и жидкого тела. Как заверяют представители AMD, новый адаптер Mellanox InfiniBand Connect-IB, который мы упоминаем в нашем обзоре, поможет значительно ускорить скорость передачи данных у серверов на базе процессоров AMD Opteron серии 6200.

Известный компилятор CAPS HMPP Workbench теперь поддерживает возможность использования интегрированного графического ядра в процессорах компании AMD для параллельной обработки данных и повышения собственного быстродействия. Работа компилятора CAPS HMPP Workbench была продемонстрирована на стенде AMD. В качестве демо-сервера была использована система производства компании Supermicro, укомплектованная еще не вышедшей на рынок картой AMD FirePro W8000, созданной по технологии AMD Tahiti.

Mellanox

Компания Mellanox представила по-

сетителям выставки новый адаптер InfiniBand Connect-IB, обеспечивающий скорость передачи данных 100 Гбит/с. Новые адаптеры предназначены для соединения серверов и системы хранения данных, отличаются высочайшей производительностью и наибольшей масштабируемостью. Как отмечают разработчики, использование этой новинки в областях высокопроизводительных вычислений, Web 2.0, облачных вычислений, финансовых сервисов, виртуализованных центров данных и приложений хранения, а также при работе с большими данными значительно улучшит производительность вычислительной инфраструктуры, что позволит сократить время завершения задач и снизить стоимость одной операции.

MEGWARE

Компания MEGWARE на момент объявления нового рейтинга TOP500 стала единственной немецкой HPC-компанией, представленной в нем тремя проектами. В список вошли HPC-кластер MOGON, установленный в университете Майнца, – № 81, 33 792 процессорных ядра, производительность 204.99 ТФлопс. MOGON занимает 8-е место среди HPC-систем, установленных в Германии. Кластер ABEL Университета Осло занял 96-е место с производительностью 178.6 ТФлопс. Третья система

MEGWARE в последнем TOP500 – VSC-2. Это самый мощный суперкомпьютер Австрии и находится он в Техническом университете в Вене с мая 2011 года. В прошлом году этот компьютер занимал 56-е место, в последней редакции рейтинга он опустился на 117-е (Linpack – 152.90 ТФлопс). На стенде MEGWARE было представлено большое количество интересных технологических новинок компании, в том числе система прямого охлаждения тепловой водой ColdCon, которая в настоящее время используется в суперкомпьютере CoolMUC Вычислительного центра имени Лейбница (Гархинг, Германия).



Даниэль Орлов (СКР-Медиа):

По сравнению с прорывными 2010 и 2011 годами может возникнуть ощущение, что отрасль топчется на месте. Но на самом деле есть вполне осознанное поступательное движение в логике развития технологий. Это касается как гибридных систем, так и общей направленности на «зеленые технологии». Происходящее отвечает общеевропейской концепции энергетической безопасности и тенденциям сбережения ресурсов. Радует, что вендоры идут вслед европейским инициативам, а не наоборот. Intel здесь, на мой взгляд, смотрит минимум лет на десять вперед, хотя, как опытный игрок, открывать все свои карты не торопится. Вообще, выставка полезна для российских компаний тем, что на ней собираются все, кто более-менее заметен в отрасли. У всех есть свои экспертные мнения. Общение с европейскими коллегами для нас очень важно. Одно жаль – что в Гамбурге ISC проводится в последний раз.



Ханс Вернер Мейер (ISC):

Больше всего участников по-прежнему из Германии, но относительная доля их за последние годы стремительно уменьшается. Растет заинтересованность представителей Азии, особенно Китая. Число их увеличивается темпами сильно выше среднего. Но еще быстрее растет доля России: теперь это пятая страна-участник после Германии, США, Великобритании и Франции. И то, что на этот раз ученые и инженеры из России взяли себе уже 6 стендов, не может нас не радовать. Еще интересный факт: первый раз в выставке приняла участие Индия.



«РСК Торнадо» – Вашему ЦОД



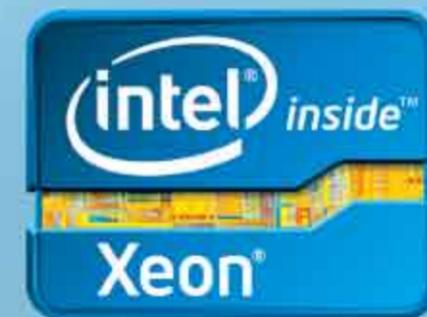
МиниЦОД на базе архитектуры «РСК Торнадо» в Росгидромете, Федеральной службе по гидрометеорологии и мониторингу окружающей среды РФ (22 место в Top50)

Суперкомпьютерный комплекс на базе архитектуры «РСК Торнадо» в Южно-Уральском государственном университете, Челябинск (4 место в Top50)



Группа компаний РСК, ведущий в России и СНГ интегратор полного цикла НРС-решений нового поколения

Передовая энергоэффективная масштабируемая архитектура



Энергоэффективность и плотность:

- Жидкостное охлаждение для стандартных серверных плат и процессоров
- Лучший в индустрии показатель PUE 1,06 → низкая стоимость владения (TCO)
- Высокая вычислительная плотность → 150 TFLOPS на 0,64 кв. м.
- Еще больше производительности на новом процессоре Intel® Xeon® E5-2600

Масштабируемость и готовность к расширению:

- От миниЦОД до суперкомпьютера уровня десятков PFLOPS
- Выдающаяся вычислительная эффективность: 92% на тесте LINPACK с Intel® Turbo Boost
- До 256 ГБ ОЗУ на каждом узле, поддержка процессоров с термопакетом 150 Вт (TDP)
- Поддержка до 3-х карт PCIe Gen3 x16 на узел (вкл. будущие Intel® Xeon® Phi™)

Проверено на практике во многих проектах:

- Системы РСК с жидкостным охлаждением работают у заказчиков с 2009 г.
- #1 — самый энергоэффективный кластер в России согласно рейтингу Green500*
- Более 4 500 000 кВт*ч электроэнергии уже сэкономлено благодаря использованию систем РСК с жидкостным охлаждением**

Всеволод Опанасенко (Т-Платформы): Что было заметно на прошедшем ISC? Существенно увеличилось присутствие систем IBM BlueGene в объявленной на конференции новой редакции списка TOP500 – только новых BlueGene/Q в списке порядка 20, из них 4 – в первой десятке (например, «Секвойя»). Это говорит прежде всего о том, что суперкомпьютерное сообщество живо интересуется технологическими новинками, вне зависимости от их стоимости и уровня «проверенности», т. е. консервативность ему не свойственна. Особенно остра тема появления кардинально новых (disruptive) технологий сейчас, когда ясно, что без принципиально новых способов решения проблем энергопотребления, времени наработки компонентов на отказ и параллелизма дальнейшее движение в сторону увеличения реальной производительности систем невозможно. Однако помимо архитектуры BlueGene/Q, которая, строго говоря, была создана и объявлена еще год назад, так же как и процессоры Intel MIC, переименованные в Phi, текущая ISC не продемонстрировала никаких технологических прорывов. Видно, что в ближайшем будущем количество систем с ускорителями (как от NVIDIA, так и от Intel) будет расти, так же как и интерес к разработке проприетарных процессоров с более энергоэффективной архитектурой. Новых интерконнектов не появилось и не появится, хотя в связи с последними приобретениями Intel, видимо, нужно ожидать появления интерконнекта, интегрированного на чип. В течение следующего года вряд ли стоит ожидать существенных изменений – мощности будут расти, но машина с уровнем производительности, скажем, в 50 Петафлопс не появится, так же как и принципиально новые процессорные или коммуникационные технологии.

Intel на конференции ISC'12

В этом году представительство Intel на конференции ISC'12 было в значительной степени посвящено нашей новой архитектуре MIC. На конференции был не только представлен первый экспериментальный кластер с MIC, но и показаны приложения независимых разработчиков ПО, использующих всю вычислительную мощь нового сопроцессора.

В первый день конференции мы объявили новое имя для первого коммерческого продукта на базе архитектуры Intel Many Integrated Cores (MIC) – Intel Xeon Phi, который также известен под кодовым именем Knights Corner. Коммерческие отгрузки Intel Xeon Phi начнутся уже в 2012 году. В настоящий момент мы работаем с более чем 40 OEM – производителями для реализации готовых решений на базе Intel Xeon Phi, среди которых есть такие авторитетные компании на рынке HPC, как Cray.

Для демонстрации степени готовности продукта мы создали в лаборатории Intel экспериментальный кластер Discovery с пиковой производительностью 180 ТФлопс и поместили его в рейтинге TOP500: в текущем списке он занял 150-е место с производительностью 118 ТФлопс. Если в ноябре 2011 года на конференции ISC'11 мы продемонстрировали ранний образец карточки Knights Corner, показавшей реальную производительность более 1 ТФлопс

на функции DGEMM (произведение матриц), то сейчас мы показали успешную работу сопроцессора в кластере с FDR InfiniBand и результат на тесте Linpack одного узла, превышающий 1 ТФлопс. Это еще одно подтверждение действительности Закона Мура: система № 1 в TOP500 1997 года ASCI Red, построенная на базе Intel Pentium Pro, занимала 72 кабинета и выдавала такую же производительность, как 1 узел с Intel Xeon E5-2670 и Intel Xeon Phi сегодня.

На стенде Intel (№ 540) были продемонстрированы приложения независимых разработчиков ПО, работающих на Intel Xeon Phi. Особо стоит отметить решатель компании Altair, использующий метод конечных элементов: в качестве демонстрации решалась задача оптимизации конструкции корпуса гоночной яхты. Американский NICS (National Institute for Computational Sciences) показал эффективную работу метода Больцмана для решения задачи вычислительной гидродинамики (CFD). Следует также отметить исследовательский проект DEEP, представленный Европейской Экзаскельной лабораторией Intel, которая работает над созданием инновационной архитектуры кластерного ускорителя с использованием Intel Xeon Phi и оптимизацией для него широкого спектра приложений: от симуляции работы мозга человека до прогнозов космической погоды.



Важно также отметить успехи Intel Xeon, представленные на конференции в Гамбурге. Из всех новых систем в списке – более 77% построено на базе Intel Xeon. Среди них – самая быстрая система в Европе: SuperMUC, созданная в Вычислительном Центре имени Лейбница (LRZ) в Мюнхене. Общее место № 4 в мировом списке TOP500 и 2.9 ПФлопс производительности SuperMUC получил исключительно за счет эффективного использования Intel Xeon E5-2680 (2.7GHz, Sandy Bridge) без применения каких-либо ускорителей. При этом система является одной из самых энергоэффективных – в ней используется непосредственное жидкостное охлаждение большинства электронных компонент серверов суперкомпьютера, которое для российских пользователей HPC уже стало достаточно привычным.

НАШИ

RSC

Велико было наше удивление, когда, получив на рецепции «конференц-бэг» (сумку с материалами, выдаваемую всем участникам), мы увидели на ней огромный логотип компании RSC. Так

серебряный партнер конференции решил привлечь внимание посетителей и участников к своей экспозиции. Надо сказать, что за все время, пока была открыта выставка, при всей скромности стенда недостатка в посетителях не замечалось. Даже на третий день конференции, когда кто-то из редакции оказывался возле стойки RSC, то видел изрядно уставших, но вдохновенных Егора Дружинина и Александра Московского, окруженных заинтересованной аудиторией. Мы думаем, что немалую роль сыграло неожиданное и изящное решение – представить пульт управления для территориально распределенного ЦОД в виде некоего конечного устройства со стильной голубой подсветкой, навевающего ассоциации с компонентами WiFi-системы. Две ручки прибора позволяли «настраивать» в реальном времени энергопотребление в зависимости от времени суток, доступности и стоимости электроэнергии как для отдельных элементов ЦОД, так и всего ЦОД. Подобное решение дает возможность пользователю следовать выбранной политике управления и существующим запросам на вычислительные ресурсы. Ну а выглядит это очень эффектно. Вообще, исходя из последних заявлений и пресс-релизов RSC, можно

сделать вывод, что компания позиционирует себя на рынке как поставщика именно энергоэффективных решений, делая упор на экономии электричества. Среди характеристик архитектуры «РСК Торнадо» и решений на ее основе обращает внимание величина PUE (Power Usage Effectiveness) в 1.06, т. е. не более 5.7% энергопотребления расходуется на охлаждение всей системы. По оценкам самих специалистов РСК, с момента начала эксплуатации в 2009 году первого вычислительного кластера в ЮУрГУ, а также проектов в Росгидромете и МФТИ, суммарная экономия электроэнергии за счет использования жидкостного охлаждения составила 4.5 млн кВт/ч. Система «СКИФ-Аврора» ЮУрГУ (117 ТФлопс) – самый энергоэффективный суперкомпьютер из представленных российских в рейтинге GREEN500. Вообще, то, что делает RSC, находится сейчас в тренде политики Intel. Именно по этой причине производитель процессоров оказывает компании заметную поддержку. На ISC'12 RSC продемонстрировала последнее обновление архитектуры «РСК Торнадо», в котором была реализована поддержка новых серверных процессоров Intel Xeon E5-2600 еще до официального их представления. При этом проекты Росгидромета и МФТИ стали



первыми в СНГ внедрениями кластерных систем на базе этого нового серверного семейства процессоров Intel Xeon и серверных плат Intel S2600JF, осуществленными уже в марте 2012 года. Соответственно группа компаний РСК была отмечена наградой «За разработку и внедрение 1-го HPC-решения с жидкостным охлаждением на базе серверных плат Intel в регионе EMEA». Специалисты РСК планируют разработать и внедрить в будущих проектах у своих заказчиков дальнейшее расширение архитектуры «РСК Торнадо» с поддержкой сопроцессоров семейства Intel Xeon Phi. На ISC'12 анонсировано, что новое решение РСК будет поддерживать планируемые к выпуску сопроцессоры семейства Intel Xeon Phi, содержащие более 50 ядер.

Т-Платформы

Шумит или не шумит? Вроде не шумит. Поразительно, что почти совсем не шумит, хотя и греется. Для мини-кластера это очень хорошо. Наконец-то можно «потрогать руками» действующий T-Mini P-Class. Формально ТП представляют свой мини-кластер уже третий «Гамбург» подряд, но только в этом году посетители могли оценить его в реальности. В прошлом году Алексей Комков – бывший заместитель гендиректора компании по маркетингу и R&D – заявил, что ТП готовы начать

поставки мини-кластеров в коробочном решении с предустановленной системой Microsoft HPC-Server 2008, чем спровоцировал общее недоумение журналистов, присутствующих на пресс-ланче в гамбургском отеле Redisson Blue. Продукт от Microsoft традиционно не пользуется популярностью на рынке параллельных вычислений, вызывает массу нареканий и в целом отношение к этому детищу компании Уильяма Генри Гейтса Третьего более чем ироничное. За прошедший год стало ясно, что Microsoft окончательно провалил свое HPC-направление и теперь сконцентрировался на облаках, ну а Алексей Комков покинул компанию «Т-Платформы». Представленный в серию мини-кластер T-Mini P-Class «всеяден», всепогоден и не шумит. Для системы, которую можно поставить под стол в обычном офисе вместо системного блока и воткнуть в десятиамперную розетку, уровень шума – чрезвычайно важный показатель. Система T-Mini P ориентирована на рабочие группы, традиционно использующие графические станции, и включает в себя встроенный узел управления, отсек для размещения двухпроцессорных вычислительных узлов и подсистему хранения данных на 16 дисков горячей замены. Вычислительные узлы объединены высокоскоростной сетью Gigabit Ethernet или QDR InfiniBand.

Шасси с воздушным охлаждением подключается к сетям переменного тока 220 В, не требуя дополнительного охлаждения. Головной узел с четырьмя слотами расширения PCI Express позволяет установить до двух профессиональных видеокарт, RAID-контроллер и сетевые адаптеры, расширяя возможности визуализации, хранения информации и удаленного доступа. Мини-кластер доступен с вычислительными модулями серии P200 на базе процессоров Intel Xeon E5 2600 и серии P205 на базе процессоров AMD Opteron 6200. Модуль двойной толщины позволяет установить ускоритель NVIDIA Tesla серии С и большее количество локальных дисков. Высокопроизводительные конфигурации с восемью модулями могут устанавливаться в стойку (8U) и наращиваться в более крупный кластер. Пиковая производительность системы на данный момент составляет более 2.6 ТФлопс в конфигурации с процессорами архитектуры x86 и более 3.38 Тфлопс в конфигурации с ускорителями NVIDIA Tesla C2075. Цена полностью укомплектованного T-Mini P в России составляет всего 1,5 млн рублей. На наш взгляд, это вполне «подъемная» сумма для потенциального заказчика. Наверное, было бы неплохо поддержать эту цену и некой лизинговой схемой, позволяющей заказчику оптимизировать затраты, растягивать их во времени и получать



возможность «оставлять» систему на балансе лизингодателя до полного начисления амортизации. Для малых и средних компаний, являющихся целевой аудиторией данного продукта, наличие подобного сервиса может стать решающим при принятии решения о покупке. Вообще, в России не развито ни страхование IT-проектов, ни кредитование. На выставке мы разговаривали с европейскими интеграторами и как раз обсуждали тему работы с клиентами. В основном в Европе конкуренция проходит на поле сервиса, на поле удобства покупки, выгоды владения. При всем различии европейских и отечественного налоговых законодательств можно найти массу схожих черт. Учитывая особенности налогового обслуживания инсталляции заказчика, можно и должно предлагать рынку цивилизованные решения, позволяющие заказчику либо экономить, либо выгодно перераспределять налоговую нагрузку. Для тяжелых систем речь может идти о весьма ощутимых цифрах. Те лизинговые компании, которые первые предложат отечественному рынку внятный и масштабируемый лизинговый продукт под IT-отрасль, могут получить хорошую прибыль. На стенде ТП посетители могли увидеть и обновленную T-Blade V-Class, представляющую собой универсальную основу для построения высоко-

производительных систем начального и среднего уровней. О нем «Суперкомпьютеры» уже писали. Сейчас продукт доделали, появился интеловский узел, тогда как на ISC'2011 был представлен только AMD, «дорисовали» систему IMU – встроенную утилиту управления. В нынешнем виде система вполне конкурентоспособна. По крайней мере посетители, фигурально выражаясь, «паслись на стенде Платформ табунами». Традиционно ни RSC, ни ТП не комментируют стратегические и коммерческие результаты выставки, хотя представители обеих компаний проводили переговоры с зарубежными и российскими партнерами и, судя по косвенным признакам, весьма продуктивные. Придется ждать официальных пресс-релизов.

ttgLabs

Приятно видеть стенд российского стартапа, возглавляемого давним автором журнала «Суперкомпьютеры» и нашим хорошим знакомым Максимом Кривовым. Это достаточно смелый и амбициозный шаг. Компания занимается переносом существующего ПО заказчиков на графические ускорители, и на выставке демонстрировала как уже успешно завершенные проекты, так и собственные технологии для оптимизации GPU-программ. В большей степени интерес посетителей стенда вызывали

технические вопросы – какой формат используется для разреженных матриц, сколько гигафлопс удалось достичь на такой-то задаче или как все это будет работать на Кеплере? Мы надеемся, что участие в выставке в качестве экспонента поможет молодой российской компании в налаживании контактов с зарубежными коллегами.

МГУ

МГУ на выставке ISC уже в четвертый раз. Третий год москвичам удается выбрать для стенда весьма выигрышное место – на перекрестке основных потоков, недалеко от стендов крупных компаний и от чая-кофе. Посетителей на стенде много: кого-то интересуют вопросы суперкомпьютерного образования (это традиционно одна из центральных тем экспозиции), кто-то заходил узнать о том, какие задачи считаются на «Ломоносове», кому-то просто приятно было поговорить по-русски, а кто-то буквально замирал перед видами Главного здания МГУ, которыми традиционно оформлен стенд.

ННГУ

Коллеги из Нижнего Новгорода впервые участвовали в работе конференции самостоятельным стендом. По некоторым данным, университету в этом помогла компания «Ниагара Компью-



терс» – давний партнер нижегородцев. Стенд в основном был посвящен образованию в области НРС. Были представлены проекты по оптимизации системного ПО и по управлению кластерами. Оптимизация ПО для процессоров Cell BE, моделирование сердечной активности, моделирование мозга, глобальная оптимизация, библиотека методов для вероятностных сетей, моделирование плазмы и др. – все это в обиходе решаемых и решенных задач научного коллектива ННГУ. Интерес посетителей вызвали работы по компьютерному зрению и специфические НРС-приложения для медицины. Особо следует отметить представление программы сертификации по параллельному программированию. На стенде ННГУ «Ниагара Компьютерс» и НИИ «Квант» демонстрировали совместную разработку «Высокоплотный вычислительный модуль МВС Орион в блейд-конструктиве с графическими ускорителями и интерконнектом PCIe». Архитектура вычислительного модуля МВС Орион минимизирует влияние ограничений и позволяет строить системы масштаба десятков Петафлопс с высокой плотностью компоновки графических ускорителей, что

уменьшает занимаемую площадь, снижает стоимость интерконнекта, повышает эффективность системы охлаждения и уменьшает стоимость обслуживания всей системы. Обратила на себя внимание архитектура компактного вычислительного узла с соотношением универсальных и графических процессоров «два к трем», что увеличивает энергоэффективность вычислений, позволяя строить системы петафлопсного класса на стандартных энергетических возможностях дата-центров. Помимо этого в модуле Орион используется интегрированный скоростной интерконнект «МВС-Экспресс», который позволяет без сверхвысоких затрат увеличить скорость межпроцессорного обмена, благодаря возможности использовать прямой доступ к памяти, уменьшению задержек обмена, настраиваемой геометрии связи между узлами и расширению обменной пропускной способности. Редакция радуется, что разработка модуля является отечественной, конструкторская документация и ноу-хау – собственность российских предприятий, а производство может быть в любое время перенесено на территорию России.

ЮУрГУ

На стенде челябинцев демонстрировались кейсы по решению производственных и научных задач. Многочисленные китайские посетители оживленно обсуждали моделирование производственных и технологических процессов, выполненные в рамках контрактов с Emerson, Челябинским трубопрокатным заводом, УРАЛТРА-Ком и др. Вообще, работы по моделированию химических наносоединений, процессов внутри человеческого тела, химических наносоединений, различных экономических процессов, проблем промышленной безопасности соответствуют мировым стандартам и вписываются в международные программы по развитию фундаментальной и прикладной науки. Особое внимание на стенде ЮУрГУ уделялось энергоэффективности и высокой компактности сборки суперкомпьютера «СКИФ-Аврора» ЮУрГУ. Для России это неплохой результат, но, строго говоря, 158-е место в списке GREEN500 (<http://www.green500.org/lists/2012/06/top/list.php?from=101&to=200>) – как говорится, «есть куда расти». Забавно, что это достижение ставят



себе в заслугу как инженеры Eurotech, так и инженеры RSC. Впрочем, и 165-е место у «Ломоносова» – не «ах, какой результат». Что касается «Ломоносова», то инфраструктура под него делалась при активном участии APC (Schneider Electric), и их специалисты в разговорах с журналистами «Суперкомпьютеров» уверяют, что решение для «Ломоносова» до сих пор считается непревзойденным внутри компании. Учитывая, каким «бронепоездом» является «Ломоносов», охотно верится. Сказанное ни в коей мере не умаляет заслуги инженеров и архитекторов российских сборщиков суперкомпьютеров, а лишь иллюстрирует этапы становления самостоятельной российской НРС-отрасли, которая, как нам видится, за последние несколько лет достаточно успешно развивается от блоковой сборки к собственной компонентной базе, накапливая know how, опыт и демонстрируя внятную постановку стратегических целей. Одно печалит – медленно. Мир опять развивается быстрее. Отметим, что тема НРС-образования – de-facto флагманская тема для стендов российских университетов. В отличие от отечественных, университетские центры Европы и Азии

делают упор не на образование (оно традиционно считается «само собой разумеющимся»), а на перспективные разработки в области программирования и аппаратной части НРС, на уникальные алгоритмы в решении прикладных задач. Впрочем, на стендах всех университетов так или иначе обсуждаются проблемы организации доступа к вычислительным ресурсам, вопросы сервиса. Так что тут российские коллеги несомненно в тренде.

ISC Student Cluster Competition

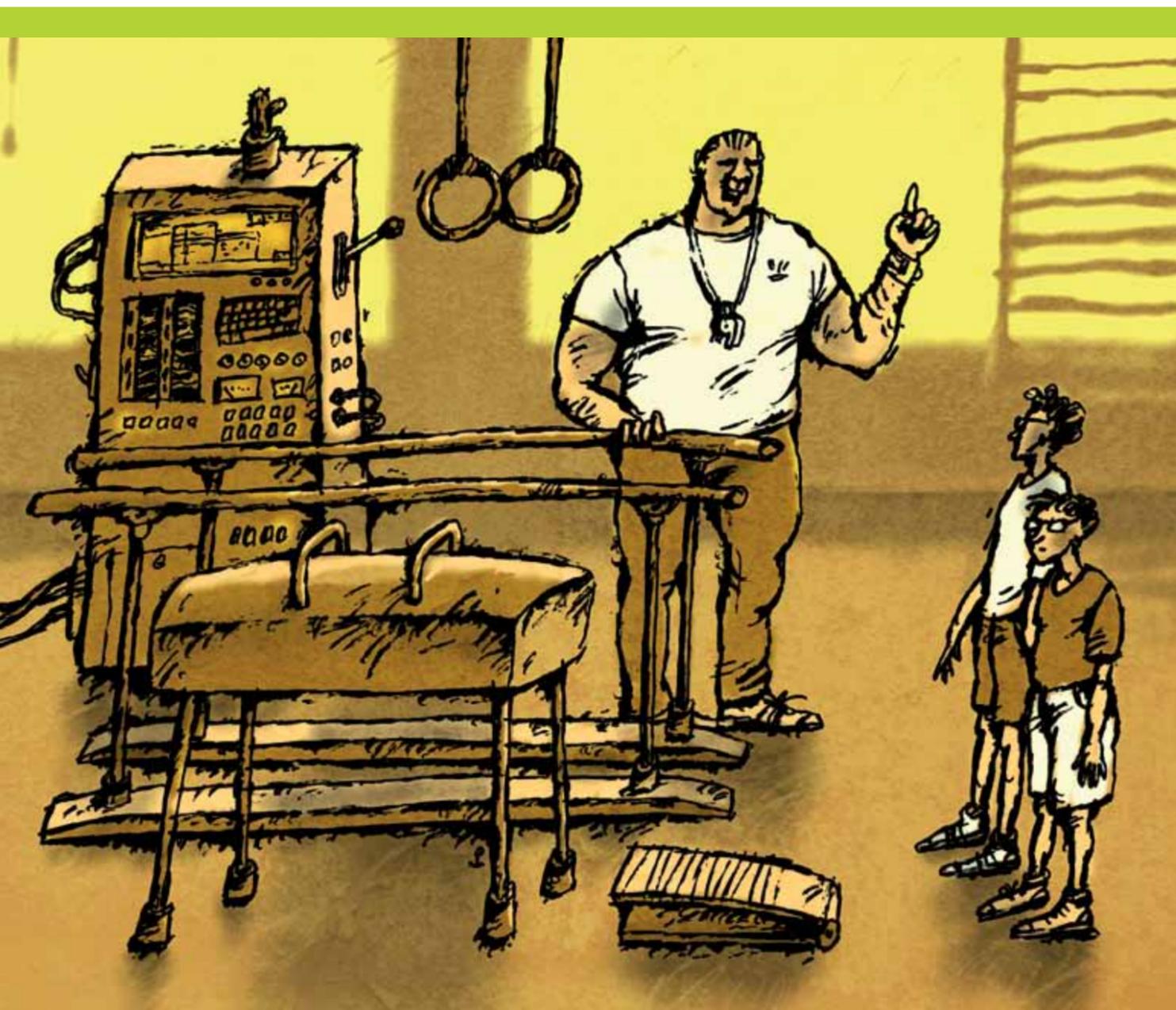
В 2012 году на выставке ISC впервые прошел кластерный конкурс среди студентов. Участники соревнования должны были сами собрать, настроить вычислительный кластер и запустить на нем ряд приложений, предложенных организаторами конкурса. Награды получили команды, на чьих системах приложения выполнялись с наибольшей эффективностью. В финал вышли пять участников: «Абсолютным победителем» была признана команда университета Цинхуа (Китай), набравшая наи-

большее количество баллов по всем бенчмаркам соревнования. Студенты из Национального университета оборонных технологий Китая построили гибридную систему с наибольшим количеством ядер. Команда также получила дополнительную награду за достижение на своем кластере самого высокого результата теста Linpack – 2.65 ТФлопс. Группа из Университета Стоуни Брук в Нью-Йорке (США) была отмечена за впечатляющие успехи в устранении ошибок в работе программного и аппаратного обеспечения своего детища. Молодые люди из Университета Колорадо (США) отличились тем, что стали первыми, кто запустил все приложения конкурса, в том числе секретные. Представители Технологического института Карлсруэ (Германия) собрали традиционную кластерную систему на базе x86-процессоров и добились наибольшей производительности на тесте Linpack в своем классе – 2.37 ТФлопс. Конкурс финансировался Советом по Высокопроизводительным Вычислениям (HPC Advisory Council) и компанией Airbus.



Три пути взаимодействия

Текст А.В. Гиглавый, Лицей информационных технологий № 1533



Цель этой статьи – показать необходимость и возможность погружения проблематики параллельных вычислений в учебный план профильной школы – московского лицея информационных технологий (ЛИТ) № 1533. На наш взгляд, к решению этой задачи ведут три пути, и сегодня ни один из них не может считаться предпочтительным – они прекрасно дополняют друг друга.

Еще в докладе профессора Сеймура Пейперта на конгрессе ИФИП-80 (1980 г., Токио/Мельбурн) была четко сформулирована центральная проблема продвижения компьютеров в школы – проблема превращения компьютера в интеллектуальный инструмент познания, доступный ребенку во всем многообразии своих функций. Академик А. П. Ершов отмечал, что в работе С. Пейперта была доказана возможность «создания новой, ранее невиданной операционной обстановки, которая потребует новых представлений в психологии развития, поможет преодолению инфантилизма и чувства зависимости». Работая вместе с лицеистами, мы ищем пути к решению этих проблем. Единая методическая основа предвуниверситетского (в лицее) и университетского образования дает

нам возможность в полной мере использовать те концептуальные связи, которыми соединены разделы современной информатики. Лицеисты раньше своих сверстников получают возможность заниматься программистскими проектами. Тема выпускного проекта выбирается в первом семестре 10-го класса; собственно работа над проектом в 10-м классе длится с января до середины июня (важную роль приобрела трехнедельная летняя практика), а в начале учебного года 11-го класса проводится защита. Уровни сложности проектов различаются. Одним ученикам больше нравится работа по детальной спецификации, предложенной руководителем проекта; другие отстаивают право заниматься задачей, которую они придумали сами, третьи готовы принять вызов и заняться поиском сложной задачи в определенной руководителем предметной области. Право заниматься сложной тематикой необходимо доказать, и лучший способ это сделать – заняться стайерским, исследовательским проектом. «Вступительный взнос» здесь высок (освоение теории, выбор инструментов и т. п.), но результат неизменно оказывается интересным.

Спортивная терминология здесь неслучайна. В сегодняшней практике поиска талантливых школьников, избравших предметом своего увлечения программирование, ведущие позиции занимает пирамида олимпиад – от школьных до всемирных. Такой широтный (breadth-first) метод имеет главное достоинство – в идеале ни один юный талант не останется незамеченным. Спринтерский дух, необходимый для участника школьных олимпиад, хорошо согласуется со спортивной по своей сути методикой подготовки к олимпиадам по программированию (тренировки, изучение причин ошибок и слабых мест в работах предшественников и т. п.).

Сегодняшние подростки, обнаружившие в себе интерес к математике и программированию, имеют весьма разнообразный опыт использования цифровой техники. Стимулом к погружению в сложную тему для них может стать и несложная программка на Паскале (lingua franca учителей информатики в российских школах), генерирующая ковер Серпинского, и самостоятельно написанная игра не сложнее «Арканоида», и желание написать что-нибудь такое, что обеспечит автору популярность в магазине приложений для смартфонов. И здесь важно предложить начинающему исследователю путь погружения в тему – depth-first. Итак, вот первые два пути работы с параллельными вычислениями в профильной школе – либо анализ качества распараллеливания на разных компьютерных архитектурах для хорошо распараллеливаемых алгоритмов, либо исследования, суть которых сформулирована в работе В. В. Воеводина [1]: «До сих пор специалистов в области вычислительной математики учили, как решать задачи математически правильно. Теперь надо к тому же учить, как решать задачи эффективно на современной вычислительной технике. А это совсем другая

наука, математическая по своей сути, но которую пока почти не изучают в вузах». Решение задач первой группы вполне по силам старшеклассникам, как показывает опыт двух проектов, разделенных десятилетием (см. врезки).

Начало проектам второй группы было положено в лицее профессором Александром Львовичем Брудно. Предложенная им еще в 60-х годах прошлого века модель изучения основ программирования [2] опиралась на модель простой трехадресной виртуальной машины MM. Лицейист Александр Шипилов решил реализовать под MS DOS среду программирования для MM. Языком программирования (псевдокодом ассемблерного уровня) стала нотация «содержательных обозначений» по Брудно. Модель-эмулятор MM выдержала испытание временем – сотни выпускников ЛИТ вспоминают о вводном курсе «по MM» с теплым чувством. В кулуарах ЛИТ не раз обсуждалась идея реализации мультипроцессорной MM для работы с векторами и матрицами. Однако до поры до времени развитие идей Александра Львовича в ЛИТ идет по другому пути. Человек, хорошо знавший Александра Львовича, Евгений Беркович, вспоминает: «Первые шахматные программы, первые алгоритмы поиска (перебора), например алгоритм Брудно «границы и оценки» [3], на Западе известный как «альфа-бета», – все это вышло из кружка (семинара) Брудно-Кронрода». Интерес к красивым алгоритмам решения ресурсоемких задач, неизменно возникавший у лицейистов в разговорах с преподавателями ЛИТ, стал отправной точкой для многих выпускных проектов. Лицейист Александр Кириллов занялся реализацией метода ветвей и границ, демонстрацией применения которого стала «задача о рюкзаке»¹

¹ Задача об отборе из неограниченного множества предметов со свойствами «ценность» и «вес»; требуется отобрать предметы таким образом, чтобы получить максимальную суммарную их ценность при одновременном соблюдении ограничения на суммарный вес.

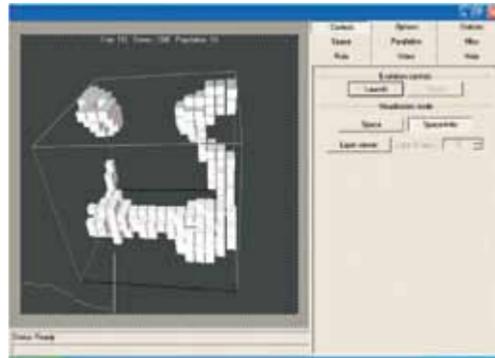
(вполне актуальная для опытных лицейских туристов). «При продолжении разработок в этой области можно порекомендовать реализацию совместного

Лицейисты раньше своих сверстников получают возможность заняться программистскими проектами. Тема выпускного проекта выбирается в первом семестре 10-го класса; собственно работа над проектом в 10-м классе длится с января до середины июня (важную роль приобрела трехнедельная летняя практика), а в начале учебного года 11-го класса проводится защита



В 2001 году лицейисты Михаил Козлов и Александр Ивлев подготовили выпускной проект «Исследование трехмерных клеточных автоматов (КА) для решения задач моделирования физических процессов началось в 80-х годах прошлого века (работы Э. Фредкина, Х. де Гариса, О. Бандман и др.). Ресурсоемкость расчетов для таких КА растет быстро, и при выполнении их на обычных для того времени ПК объем рабочего пространства КА оказывался весьма ограниченным. Это ограничение не позволяло выявлять интересные примеры эволюционного поведения популяций. Разделение работы между лицейистами было следующим: Михаил сосредоточился на оптимизации «движка» КА по быстродействию, а Александр проводил эксперименты по применению протокола TCP/IP для организации обмена данными между ПК в локальной сети, на каждом из которых выполнялась копия «движка» с одним из фрагментов рабочего пространства КА. Визуализация хода эволюции выполнялась на одном из выделенных

применения нескольких комбинаторных методов, например – метода ветвей и границ в сочетании с методом динамического программирования» (из записки к выпуск-



ПК в локальной сети. Исследовались режимы как автоматического, так и ручного распределения нагрузки между ПК в локальной сети. Было отмечено, что визуализация (рендеринг) на «ведущем» ПК обладает большей ресурсоемкостью, чем расчет каждого шага эволюции для заданной популяции КА. В дальнейшем исследования были сосредоточены на выборе эффективного представления данных о структуре популяции на границах между фрагментами рабочего пространства КА. Решалась задача минимизации трафика между «ведомыми» ПК. Была найдена еще одна возможность оптимизации: для некоторых интересных популяций визуализация выполнялась «скачками» через несколько поколений эволюции. В ходе тестирования получившейся инструментальной среды были обнаружены весьма примечательные примеры эволюции исходных популяций. Так, для определенных правил результатом эволюции становились плотные устойчивые конфигурации «живых» клеток, среди которых была конфигурация, напоминающая двойную спираль. Поиск интересных итоговых конфигураций для трехмерных КА стал одним из основных стимулов к исследованию.



В 2011 году учащиеся лицея Александр Царьков и Елизавета Трухова подготовили выпускной проект «Модель гравитационного линзирования». С помощью этого физического эффекта было открыто существование темной материи – гипотетической формы материи, которая не испускает электромагнитного излучения и не взаимодействует с ним. Выявить ее присутствие оказалось возможным с помощью гравитационных эффектов. Проект был задуман как своеобразный тренажер, с помощью которого студенты-физики могут изучать на модельных примерах наблюдаемую картину линзирования. Для реализации этой цели разработанная программа обладает значительным количеством настроечных параметров (масса дефлектора, радиус источника, масштаб изображения и т. п.), а также содержит аналог журнала астрометрических наблюдений. Вот выдержки из пояснительной записки к этому выпускному проекту: «Искривление лучей света в гравитационном поле аналогично действию линзы на световые лучи. Поэтому возникло понятие гравитационной линзы – гравитирующего объекта, создающего путем искривления лучей света изображения («духи») далекого объекта (галактики, квазара). Важно отметить, что при этом блеск «духов» может быть много больше блеска самого линзируемого объекта. Массивные объекты, выполняющие роль линз, представлены большими скоплениями массы – такими как галактики или скопления галактик. Такие линзы способны существенно исказить изображение фонового объекта. В проекте предпочтение отдано именно массивным гравитационным линзам. Серьезной проблемой стал вопрос преобразования физических формул для корректной работы в 3D. Однако было найдено оригинальное решение проблемы: вместо того чтобы усложнять алгоритмы трехмерными расчетами, перед началом вычислений плоскость, в которой находились источник, дефлектор и наблюдатель, поворачивалась так, чтобы их координаты Z были равны. Это позволило опустить их при расчетах, следовательно, задача сводилась к уже решенной. Расположение объектов восстанавливалось после вычислений обратным поворотом этой плоскости. После успешной реализации модели выяснилось, что программа стала ресурсоемкой и медленной из-за большого объема данных, требующих обработки за каждый полный расчет. В ходе поиска решения данной проблемы было решено распределить вычисления на несколько равных частей, выполняющихся параллельно. Сегодня большинство компьютеров имеют

многоядерные процессоры, поэтому оптимизация работы программы путем распараллеливания вычислений является актуальной задачей. Поиски руководств по реализации многопоточности дали положительный результат: в сети был найден класс MultiThreading, позволяющий запускать алгоритм параллельно в нескольких потоках, число которых определяется числом ядер процессора. Проблемой стало разделение вычислительного алгоритма, так как выполняемые потоки конфликтовали друг с другом (например, обращаясь к одному и тому же объекту). Решением стало клонирование требуемых объектов для корректного использования их каждым из потоков. В результате продленной работы скорость работы программы с большим объемом данных увеличилась в 1.5–2 раза. После оптимизации работы программы и более детального изучения предметной области стало очевидно, что наравне со сферическими объектами (звездами) можно моделировать линзирование объектов произвольной формы. Первым шагом на пути к этому стала разработка графического редактора формы источника. Реализована также возможность линзирования изображений, представленных в виде графических файлов (т. е. фотографий реальных астрономических объектов). Затем была изучена технология распараллеливания вычислений на видеокартах NVIDIA CUDA (Computing Unified Device Architecture), позволяющая существенно увеличить производительность программы за

счет распределения вычислений на множестве графических процессоров CUDA. Подключение созданной библиотеки DLL к модулю CUDA-вычислений позволило значительно увеличить скорость моделирования гравитационных линз с большими изображениями источника (в 5–10 раз). В ходе этой разработки были использованы CUDA SDK 4.0 и NVIDIA Parallel NSight 2.1. Многопоточные вычисления позволяют значительно ускорить расчеты при больших объемах данных (фотографии больших скоплений массы). Это достигается путем разделения одинаковых и независимых друг от друга вычислений для работы с равными частями данных. Такие вычисления запускаются параллельно в отдельных потоках, выполняющихся на отдельных ядрах CPU компьютера или процессорах CUDA. Схема выполнения программы и скриншот представлены на рис. 1. При малых объемах данных наиболее эффективно однопоточное вычисление, но эффективность вычислений в два и четыре потока быстро обгоняет однопоточное с ростом объема данных. Такие результаты связаны с ростом трафика «процессор–память» для нескольких потоков. Однако чем больше данных для обработки, тем больше требуется ресурсов процессора, что и становится слабым местом однопоточности. В то же время вычисления с использованием CUDA всегда остаются наиболее эффективными благодаря гораздо большему количеству процессоров по сравнению с количеством ядер CPU.

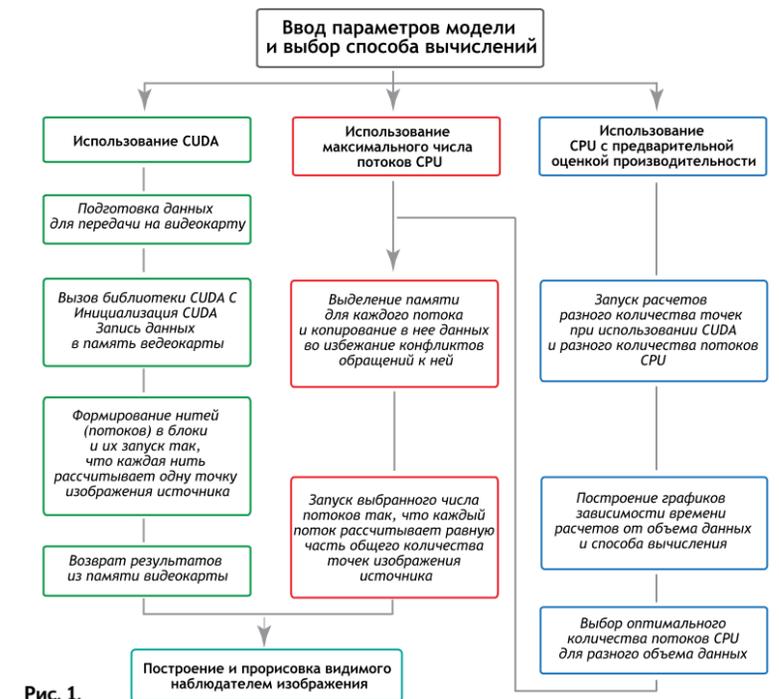


Рис. 1.



В работе лицейстов Александра Бокова и Дмитрия Ульянова «Моделирование организации сообществ роботов в целевые структуры» получила развитие актуальная для микроэлектроники и мехатроники концепция «программируемой материи».

Суть этой концепции в том, чтобы синтезируемые на микро- или наноуровне материалы или конструкции могли изменять свои физические свойства (например, форму или плотность) в ответ на изменение состояния среды или в ответ на некоторую информацию, полученную извне.

Еще одной областью применения концепции «программируемой материи» является круг задач конструкторской визуализации и систем виртуальной реальности. В проекте предложена реализация инструментальной программной среды для исследования одной из таких моделей [5], развиваемой в исследовательском центре Clautronics (университет Карнеги-Меллона) при поддержке корпорации Intel.

Среда содержит средства как для работы с самими элементами программируемой материи («атомарными» роботами, atoms), так и для задания механизмов их поведения. Обеспечено также физически достоверное моделирование поведения конфигурации программируемой материи в целом с некоторыми априорными параметрами. Назначение разработанной программной среды состоит в предоставлении пользователю средств для отбора наиболее эффективных параметров элементов программируемой материи и алгоритмов их поведения.

Главным препятствием для реализации программной среды остается алгоритмическая проблема организации всех ее элементов таким образом, чтобы любая возможная цель (формирование конфигурации) была достигнута максимально эффективным путем. Решение этой задачи достигнуто в разработанной программной среде.

Разработана программа, позволяющая:

- редактировать архитектуру робота – элемента программируемой материи;
- задавать алгоритм поведения робота;
- задавать начальное и конечное расположения роботов в сообществах;
- моделировать процесс организации сообществ роботов при заданных параметрах.

Первоначально было проведено макетирование среды в системе визуального программирования StarLogo TNG. Затем для итоговой версии среды, реализованной на языке C#, было проведено исследование, направленное на отбор наиболее стабильных конфигураций сообществ и наиболее эффективных алгоритмов коллективного поведения роботов.

Проект реализован в среде Microsoft Visual Studio 2008 (язык C#). Используются также библиотека Open Dynamics Engine (ODE) для создания физически достоверной модели и компилятор компиляторов Coco/R. С помощью последнего реализован компилятор языка, предложенного одним из авторов (А. Боков) для задания алгоритмов поведения элемента программируемой материи.

ному проекту А. Кириллова). Стоит отметить, что в такой постановке задача привлекает внимание многих современных исследователей, причем при ее решении широко используются методы распараллеливания [4]. К задачам третьей группы следует отнести проекты, основанные на применении мультиагентного подхода к проектированию моделей различного назначения. Среди работ выпускников лица последние лет здесь стоит привести два примера (см. врезки). В проектах третьей группы, как показывает наш опыт последних лет, скрыт значительный потенциал для решения еще одной важной задачи. Преподаватели во многих школах мира ведут педагогические исследования,



В работе лицейстов Сергея Леоненкова и Сергея Деревянкина «Программная среда для моделирования чрезвычайных ситуаций»

предложена расширяемая программная среда, использование которой целесообразно при обучении, планировании и реализации действий спасателей. В проекте реализованы модели развития и оценки последствий землетрясений, лесных пожаров и цунами. Разработанная инструментальная среда позволяет накапливать библиотеку данных об исследуемых стихийных бедствиях этих трех категорий. Особое внимание уделено детальной визуализации динамики развития двух моделируемых процессов (лесные пожары и землетрясения). В результате создан набор моделей стихийных бедствий, которые не только помогают воссоздать условия природного катаклизма, но и оценить масштабы

выясняя возможности визуальных языков программирования [6]. Особый интерес представляет язык Starlogo TNG, предыстория которого указывает на связь между проблематикой «школьного программирования» и магистральным направлением развития технологии параллельных вычислений – архитектурой массивно-параллельных вычислительных систем. Суперкомпьютер 1980-х Connection Machine, разработанный Дэниелом Хиллисом, стал первой аппаратной платформой для изначального диалекта StarLogo, разработанного бессменным руководителем проектов развития языков этого семейства, профессором Митчелом Резником из MIT Media Lab.

разрушений (в частности, оценить расположение построек и возможность выбрать для новых построек наиболее безопасное место расположения).

Модели написаны в среде Starlogo TNG; мультиагентный язык визуального программирования в этой среде удобен для разработки моделей.

В среде Microsoft Visual Studio 6.0 на языке Visual C# была написан внешний пользовательский интерфейс, из которого можно запустить любую из моделей и где можно посмотреть весь справочный материал.

Почему именно Starlogo TNG? Мы использовали некоторые специфические свойства этой среды разработки – автономность (агенты, хотя бы частично, независимы), ограниченность представления о мире (ни у одного из агентов нет представления обо всей системе, или система слишком сложна, чтобы знание о ней имело практическое значение для агента), децентрализация (нет агентов, управляющих всей системой).

Подмосковные вечера

На границе весны и лета компания «Сторус» проводит свои традиционные партнерские семинары. «Сторус» знаком участникам российского рынка высокопроизводительных вычислений как поставщик систем хранения и передачи данных различного уровня сложности. На этот раз семинар был посвящен конвергентным и гибридным технологиям Истринского водохранилища.

Издатель журнала «Суперкомпьютеры» Даниэль Орлов беседовал с генеральным директором компании «Сторус» Александром Чичковским.

Даниэль Орлов («Суперкомпьютеры»): Эффективность системы во многом определяется правильной организацией межзвонных соединений, устойчивой совместной работой сетей и средств хранения. Мы понимаем, что зачастую конечный заказчик оказывается в ситуации бюджетного ограничения и не может позволить себе самые дорогие решения, однако при этом не желает потерять в эффективности и надежности. Это общая проблема. Какие возможности для маневра вы даете своим партнерам?

Александр Чичковский («Сторус»): Чем компетентнее наши партнеры, тем взвешенней и адекватней предложения конечным заказчикам. Потому на своих семинарах мы стараемся не рассказывать об отдельных продуктах, а стремимся продемонстрировать комплексный

подход к построению СХД. Предоставляя общий анализ рынка, мы говорим о перспективах развития систем хранения, сетей передачи данных, программного обеспечения. Развитие СХД тесно связано с серверными технологиями, именно поэтому мы ежегодно приглашаем представителей компании Intel для обзора основных тенденций в этой области. Таким образом, проявляется общая картина IT-рынка в разрезе систем хранения данных. Традиционно семинар посещают представители производителей, оборудование которых мы представляем на российском рынке. Презентации построены по принципу открытой дискуссии, это позволяет задать интересующие вопросы и получить информацию о продуктах, так сказать, из первых уст.

Д. О.: Давайте вернемся к проблемам конечного заказчика, единожды установившего у себя на производстве или в институте кластер и стоящего перед необходимостью апгрейда системы за сравнительно малые средства. На рынке множество предложений по фактическому реинжинирингу систем. Есть ли что-то, что можно предложить партнерам в качестве разумной альтернативы, что-то, что будет выгодно и партнеру, и конечному заказчику?

А. Ч.: Тема «Конвергентные и гибридные технологии» была выбрана для нашего семинара неслучайно. На сегодняшний день это одни из самых актуальных технологий, в том числе и при построении СХД. Применение таких технологий позволяет существенно упростить управление и внедрение новых систем, а также расширить функционал уже существующего оборудования. Немаловажным фактом является то, что внедрение конвергентных и гибридных сетей дает возможность значительно сократить эксплуатационные расходы, связанные с поддержкой IT-инфраструктуры.

Д. О.: Мода на облака против ожиданий

не прошла, а облачные сервисы развиваются чуть ли не быстрее других технологий. Есть решения в этой области?

А. Ч.: Верно, системы хранения данных уверенно «мигрируют в облако». Рост данных в облаках задает новые стандарты доступа, записи, обработки и хранения информации. Такие системы должны отвечать требованиям высокой надежности, обладать интеллектуальным управлением и самое главное – обеспечивать множественный параллельный доступ к данным. Нельзя сказать, что задача по построению таких систем является для нас новой, на протяжении многих лет мы занимаемся внедрением СХД для высокопроизводительных кластеров. Требования, предъявляемые к СХД, в обоих случаях очень похожи, на сегодняшний день мы имеем готовые решения для развертывания СХД в составе частного облака.

Д. О.: На ваш взгляд, исходя из сказанного, на решения каких производителей стоит обратить внимание отечественным строителям кластеров?

А. Ч.: Конкурентоспособность бизнеса зачастую напрямую зависит от скорости и точности расчетов. Это касается геофизических компаний, компаний, занимающихся инженерными расчетами. Любые отрасли промышленности, где используются большие и сверхбольшие объемы первичных данных, предполагают распараллеливание вычислений, использование той или иной HPC-платформы. Важно, на чем строить новые и с помощью чего апгрейдить существующие системы. Я полагаю, что на настоящий момент весьма адекватные предложения есть у Panasas и DDN, нам их продукты интересны, а вашей аудитории понятны.



С играми не шутят. GTC в Сан-Хосе

Текст Игорь Лёвшин

Строго говоря, GTC – не суперкомпьютерная конференция, хотя добрая половина докладов и была посвящена именно НРС. Ученых там много, а геймера не сыщешь, разве что он при этом еще и ученый или президент NVIDIA. Но в НРС-бизнесе с играми шутить не стоит. Печальная судьба интереснейшего детища IBM – процессора Cell – тому доказательство.

Игроки кормят NVIDIA, питая ее разработки. «У NVIDIA есть хорошо работающая бизнес-модель, позволяющая сбалансировать доходы от массовых продаж и дорогие исследовательские работы. Сейчас мы надеемся, что архитектура Kepler будет всю использоваться в облачных игровых сервисах, а мировой рынок игр огромен, сейчас он больше, чем, скажем, кинорынок. А это значит, что есть деньги на новые исследования». Так говорит Сумит Гупта, глава направления Tesla, отвечающий в компании за все, что связано с НРС. На предыдущем GTC, то есть GTC Asia в Пекине, президент NVIDIA Дженсен Хуанг провозгласил конвергенцию игровых и конструкторских вычислений: в играх нового поколения выгодней вычислять картинку, а не прокручивать ранее смоделированную. А это значит, что обрушение угла

дома из-за попадания гиперболюида или бластера надо рассчитывать примерно так же, как это делали бы строительные фирмы, только в играх надо делать это в реальном времени. Слово «игры» звучало постоянно и на GTC в Сан-Хосе в этом году. Но теперь речь шла в основном об игровых облаках, которые заволокут весь горизонт будущего потребительского бизнеса. Рынок игр колоссален. NVIDIA верит, что компьютер опять превратится из персоналки в консоль, на которую будут закачиваться игровые сцены с огромных облачных дата-центров, оснащенных, разумеется, GPU от NVIDIA.

Кеплер

Чаще, чем слово «игра», звучало только имя астронома и астролога Иоганна Кеплера. Собственно,

главное событие GTC 2012 в Сан-Хосе – представление архитектуры Kepler, воплощенной в процессорах GK104 и GK110: 2 процессора GK104 вошли в GPU-модуль Tesla K10 пиковой производительностью 4.58 TFlopс и суммарной пропускной способностью памяти 320 ГБ/с, а GK110 войдут в Tesla K20, которые появятся в последнем квартале этого года. Младший и старший Кеплеры не слишком похожи друг на друга: GK104 строго ориентирован на вычисления с одинарной. Огромная производительность достигается в том числе за счет отказа от «излишеств» – ECC для кэша и некоторого уменьшения производительности двойной точности по сравнению с Fermi. Но гордость NVIDIA – GK110. Говорят, что это самый сложный чип в истории. Во всяком случае, на нем более 7 млрд транзисторов – он первый по этому показателю среди коммерчески доступных чипов (а сразу за ним идут 2 чипа FPGA, увеличивая тем самым отрыв от следующих собственно процессоров). В нем будет 2 важнейшие технологии, недоступные в GK104: управление очередями – Hyper-Q и динамическое распараллеливание – Dynamic Parallelism. Hyper-Q дает возможность передавать данные и команды от нескольких (до 32) физических ядер одновременно. В Fermi и в GK104 команды передаются последова-

тельно. Динамическое распараллеливание – это возможность GPU динамически (в соответствии с результатами очередных вычислений) порождать новые потоки, не обращаясь к CPU. А общее у двух Кеплеров – новые мощные вычислительные блоки: SMX вместо SM в Fermi. Кроме того, благодаря GPUDirect два GPU внутри одного модуля теперь могут обмениваться данными, вообще не обращаясь к CPU. Этот модуль может пересылать данные и по сети напрямую, не через CPU. То есть новые ускорители, хотя и не могут совсем обойтись без координирующей деятельности центрального процессора, гораздо более «самостоятельны». Итого: у K110 более полная функциональность и в 3 раза большая производительность на Вт, чем у Fermi. Как удалось втиснуть сложную логику в процессор, у которого и количество ядер резко возросло? Ведь в каждом модуле SMX, которых на чипе может уместиться 15 штук, по 192 CUDA-ядра одинар-

ной точности, 64 блока двойной точности, 32 блока специальных функций (в 8 раз больше, чем в модуле SM в Fermi GF110) и 32 блока load/store. Увеличением площади чипа этого уже не объяснишь. Кое-какие объяснения чуду были получены от Сумита Гупты. Первый секрет: тактовую частоту ядер уменьшили с 1.3 ГГц до 700-750 МГц – то есть чуть ли не в два раза. Это дало возможность избавиться от многих регистров, необходимых для быстрой работы, и высвободить место под логику. Второй секрет – был полностью переработан механизм получения ядрами инструкций. Получилось очень удачно, и благодаря этому удалось упростить логику управления потоками и высвободить еще место под арифметику.

Последствия

Процессоры GK110 будут установлены в суперкомпьютерах Cray: в Titan Национальной Лаборато-

рии Окриджа и в систему Blue Waters в Национальном Центре Суперкомпьютерных Приложений Университета Иллинойса, которые, видимо, мы увидим вверху списка TOP500. Появились первые официальные отзывы о K10 – от нефтяников, как и намечалось. В бразильской Petrobras получили ускорение на приложениях в 1.8 раз по сравнению с GPU предыдущего поколения. Schumberger цифр не обнародовала, но засвидетельствовала, что результаты их тестов воодушевляют. На очереди компании, предоставляющие игры на своих облаках, такие как OnLive, Otoy или Gaikai. Это только в мире НРС. Очевидно, что удачные архитектурные находки попадут и в процессоры для потребительского рынка. Там тоже есть что распараллеливать. Иногда, кстати, кажется (на GTC, например) что цель NVIDIA – не продажа своих ускорителей, а распараллеливание всего мира. ■■■

ИЗ РАЗГОВОРА С ДЭВИДОМ КИРКОМ, ПОЧЕТНЫМ СОТРУДНИКОМ NVIDIA, ДИРЕКТОРОМ ПО НАУКЕ NVIDIA ДО 2009 ГОДА:

Игорь Лёвшин: GK110, можно сказать, без пяти минут CPU. Во всяком случае, он способен сам создавать потоки, работать с очередями, обмениваться данными внутри узла и по сети – все без обращения к CPU. Осталось сделать маленький шагочок. Отчего не шагнули?
Дэвид Кирк: Да, перейти к CPU и GPU на одном чипе для нас было бы не то чтобы очень сложно. Мы пока просто не видим в этом острой необходимости. Со временем мы, безусловно, сделаем это. Пока мы больше озабочены проблемами распределения вычислительных нагрузок, планировкой процессов, созданием условий для того, чтобы можно было создать больше задач одновременно. Сейчас индустрия НРС в той ситуации, когда нет смысла думать о 2-м и 3-м препятствии на дороге, пока мы не разобрались с первым: до следующих еще надо доехать, чтобы понять, что это за препятствия. Некоторые производители говорят о том, что их цель – дать возможность операционной системе работать на GPU.

Но разве это цель? Такая интеграция – просто решение определенной проблемы. А я не уверен, что эта проблема останется актуальной после того, как индустрия освоит возможности Kepler. Я не сомневаюсь, что ОС будут работать на CPU, но сейчас для нас главное увидеть, как пойдут дела у разработчиков приложений для Kepler, получить, так сказать, реакцию Реальности, а не теоретиков. Может статься, что появится необходимость в какой-то другой, более важной и срочной инновации.
И. Л.: У интеграции CPU и GPU на чипе немало плюсов. Каковы минусы?
Д. К.: Я думаю, что интеграция всегда должна быть нацелена на решение какой-то проблемы. Если есть проблема, которую не решить без интеграции, то тогда надо интегрировать. В конечном счете речь идет об эффективности и удобстве, о гибкости. Во многих очень эффективных системах отношение GPU/CPU довольно высокое. Но интеграция на чипе фиксирует это отношение, лишает разработчиков свободы в создании сбалансированных систем.

В будущем, конечно, появятся системы с GPU и CPU на одном чипе, и они будут хороши для тех, кому нужны высокая энергоэффективность и кому нужна большая вычислительная плотность, чтобы было побольше процессоров внутри одной стойки. Но при этом системы с отдельными GPU и CPU продолжат существовать. Мы не стоим перед выбором «или/или» и это вообще не самый принципиальный из возникающих вопросов.
И. Л.: Стоит ли ожидать, что в будущих системах CPU получит контроль над многими GPU?
Д. К.: Такие системы уже есть у HP, Supermicro и у других производителей. Дело тут в том, что CPU должен формировать задания для GPU, загружать их работой, поэтому, чем больше отношение GPU/CPU, тем сложнее корректно загрузить работой все ускорители. Все это легко может превратиться из системы для вычислений в систему приготовления данных и заданий для вычислений.

От бита к кубиту

Текст Сергей Сысоев

Иллюстрация Владимир Камаев



Математики и программисты привыкли воспринимать вычисления как абстрактный процесс, не имеющий физической основы. Тем не менее любое устройство, выполняющее вычисление, основано на каком-либо физическом процессе.

Принятие этого факта приводит к отказу от платоновского мира идей и потере веры в бессмертную душу (вычислительные способности которой, по-видимому, не должны иметь материальной основы). Зато открываются великолепные перспективы для экспериментов с различными физическими системами.

Поколения ЭВМ являются наглядной иллюстрацией истории таких экспериментов. До ЭВМ вычисления выполнялись механическими системами. Потом появились устройства, основанные на электромагнитных реле, радиолампах, транзисторах... Каждый новый тип устройства был эффективнее предыдущего – естественно было бы ожидать появления новых, еще более эффективных устройств. Вместе с устройствами эволюционировали и алгоритмы. В 1936 году Алан Тьюринг предложил математическую модель вычислений, названную впоследствии машиной Тьюринга (Deterministic Turing Machine – DTM). Модель Тьюринга формализовала понятие алгоритма (способа вычисления некоторой функции на физическом устройстве), что позволило ввести понятие вычислимости и, в дальнейшем, сложности вычислений. Тогда же, в 1936 году Тьюринг показал, что существуют невычислимые (undecidable) функции – такие функции, которые можно формально определить, но для вычисления

которых невозможно предложить алгоритм, работающий конечное время.

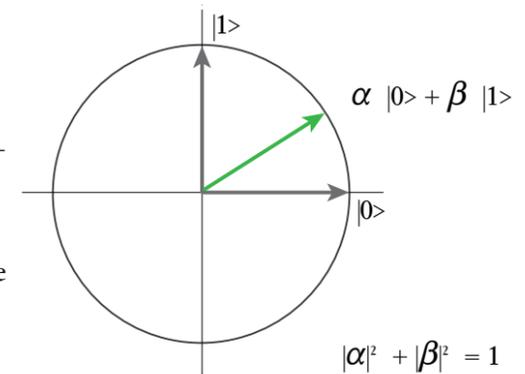
Понятие сложности вычислений (количества ресурсов, необходимых для выполнения алгоритма) позволило выделить классы задач, названных неразрешимыми (intractable). В отличие от невычислимых функций, для них существуют алгоритмы, работающие за конечное время. Однако ресурсы, необходимые для выполнения этих алгоритмов, растут слишком быстро (например, экспоненциально) с увеличением размера входных данных. Открытым является вопрос, являются ли неразрешимыми задачи из класса NP-complete (знаменитая гипотеза P! = NP).

С появлением неразрешимых задач возникла потребность в физических устройствах, вычислительные возможности которых зависели бы экспоненциально от размеров (ресурсов) устройства. В 1981 году нобелевский лауреат по физике Ричард Фейнман предложил построить вычислительное устройство на основе квантовой системы. Причин для оптимизма относительно этого направления у Фейнмана было две:

1. Квантовый параллелизм.
2. Квантовая запутанность (entanglement).

Квантовый параллелизм эксплуатирует тот факт, что простейшая квантовая система намного сложнее

простейшей классической системы – математически это единичный вектор в двумерном гильбертовом пространстве над комплексными числами. Раз пространство двумерное, то базис в нем состоит из двух векторов. Результат измерения (наблюдения) системы – всегда один из базисных векторов (выбор базиса определяется наблюдаемой). Одному из векторов дают название $|0\rangle$, второму – $|1\rangle$, таким образом, в результате измерения получается классический бит информации. Преимущество заключается в том, что **до момента измерения система может находиться в любом из бесконечного количества состояний – суперпозиции базисных векторов. Такая простейшая квантовая система называется «кубит» (quantum bit – qubit).**



Квантовая запутанность – явление намного более загадочное и вместе с тем перспективное с точки зрения эффективности вычислений. Дело в том, что **система из n кубитов описывается вектором в тензорном произведении пространств своих подсистем.** Таким образом, состояние n-кубитной системы является вектором в пространстве размерности 2^n . Для классического описания состояния 10-кубитной системы потребуется 1024 комплексных числа, а для 30 кубит – уже более миллиарда комплексных чисел.

Моделирование квантового компьютера на классическом требует экспоненциального роста ресурсов при линейном увеличении квантового устройства. Этот факт позволил Ричарду Фейнману предположить, что квантовый компьютер окажется эффективным в решении неразрешимых задач.

Оракул и квантовые алгоритмы

Эволюция квантовой системы унитарна. Это означает, что любое преобразование над системой кубитов является унитарным оператором в пространстве, соответствующем этой системе. Унитарный оператор как минимум линеен, а значит, обратим. В то же время классические вычисления часто необратимы. Для того чтобы разрешить это противоречие, квантовую систему расширяют дополнительными регистрами – регистрами исходных данных, который в процессе вычислений остается неизменным. Рассмотрим классическую функцию f , преобразующую n -битную строку в m -битную:

$$f: \{0,1\}^n \rightarrow \{0,1\}^m$$

Определим унитарный оператор на пространстве размерности 2^{n+m} следующим образом:

$$U_f |x\rangle |y\rangle \rightarrow |x\rangle |y \oplus f(x)\rangle,$$

где $|x\rangle$ – регистр из n кубит, а $|y\rangle$ – регистр из m кубит.

Такое определение полезно нам по следующим причинам:

1. U_f линеен.
2. U_f задан на всех базисных векторах, следовательно, задан на всем пространстве.
3. На векторах вида $|x\rangle |0\rangle$ оператор позволяет получить во втором регистре значение $f(x)$.

Сразу бросается в глаза факт, что оператором U_f можно подействовать на суперпозицию (сумму) всех базисных векторов и получить таким образом значения функции f сразу от всех аргументов из области определения в результирующей системе. Однако напрямую

До ЭВМ вычисления выполнялись механическими системами. Потом появились устройства, основанные на электромагнитных реле, радиолампах, транзисторах... Каждый новый тип устройства был эффективнее предыдущего – естественно было бы ожидать появления новых, еще более эффективных устройств. Вместе с устройствами эволюционировали и алгоритмы

воспользоваться этой возможностью нам не удастся – измерение результата уничтожит суперпозицию, дав нам одно из значений. Тем не менее, пока результат не измерен, квантовая система может хранить информацию обо всех возможных значениях функции, что полезно для определения ее свойств.

Предположим, что мы ничего не знаем о свойствах функции f , но можем ее вычислять (похожая ситуация, например, имеет место для всех NP-полных задач. Мы можем посчитать длину конкретного пути в задаче коммивояжера, но обычно ничего не можем сказать о существовании других путей короче данного). Предположим также, что в нашем распоряжении находится квантовое устройство, способное вычислять U_f . Имея два этих оракула (классический и квантовый), можем ли мы предложить задачи, на которых преимущества квантового оракула станут очевидны?

Задача Дойча

Первый теоретический результат, использующий преимущества квантового вычислителя, был получен Дэвидом Дойчем в 1985 году. Задача, которую решал Дойч, заключалась в следующем. Есть функция $f: \{0,1\} \rightarrow \{0,1\}$. Алгоритм, выполняющий функцию, нам неизвестен, функция является оракулом, «черным ящиком».

Мы можем использовать оракул, передавая аргумент и получая зна-

чение функции от этого аргумента. Требуется определить, является ли функция f константой. Классическое решение задачи требует двух обращений к оракулу – один раз от 0, второй – от 1. Но оракул в задаче Дойча обладает неприятным свойством – одно вычисление функции f занимает целых 24 часа!

Поможет ли нам квантовый компьютер решить задачу быстрее, чем за 48 часов? Дэвид Дойч показал, что да.

Устройство Дойча помимо оракула использует оператор Адамара H . Суперпозиция оракула и оператора Адамара приводит систему в два ортогонально противоположных состояния в зависимости от свойств оракула. Если функция f – константа, то оператор HU_fH не изменяет входной вектор $|0\rangle|1\rangle$. Если же f константой не является, входной вектор преобразуется этой последовательностью операторов в вектор $|1\rangle|1\rangle$. Таким образом, не получая ни одного значения функции f , за одно обращение к квантовому оракулу мы определяем ее свойства.

Другие задачи

Аналогичное устройство (последовательность операторов) позволяет решить за одно обращение к оракулу и более интересные задачи.

1. Задача Дойча-Джозсы (определение того, является ли f константой, когда аргументом для f является n -битный регистр).
2. Задача Берништейна-Вазирани (поиск

вида линейной функции в Z_2).

Кроме того, используя оператор Адамара, можно с экспоненциальным ускорением найти период функции в Z_2 (задача Саймона).

Алгоритм Гровера

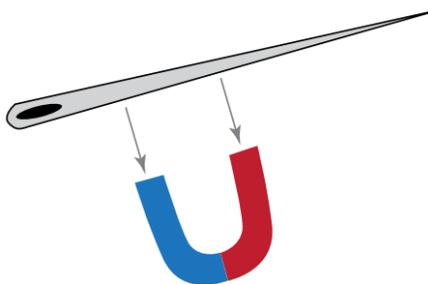
Поговорка об иголке в стоге сена часто применяется в быту как пример сложной поисковой задачи. При рассмотрении NP-полных задач, однако, о стоге сена придется только мечтать.

В сущности, все NP-полные задачи могут быть сведены к поиску прообраза некоей легко вычисляемой функции.

Функция принимает различные значения на обширном пространстве аргументов, и только на некотором малом множестве аргументов она принимает нужные нам особые значения. В качестве примера можно рассмотреть задачу о поиске подмножества с нулевой суммой.

Есть конечный набор n целых чисел. Существует ли подмножество этого набора, такое, что сумма всех чисел в нем равна нулю? Сформулированная задача сложна. С ростом величины n количество вариантов подмножеств растет экспоненциально – 2^n . Подмножества можно перенумеровать n -битным регистром. Функция f отображает номер подмножества в сумму его членов. Нам нужно найти число w , такое, что $f(w) = 0$.

Если бы существовала физическая система, способная «вместить в себя» все значения f от всех аргументов, то это и был бы наш «стог сена» (а число w – «иголка»). Имея такую систему, можно попытаться искать «иголку», не перебирая сено. Проблема в том, что размер такой физической системы в классическом случае растет экспоненциально при увеличении n , а значит, уже при некоторых небольших n достигнет грандиозных значений. Но квантовый компьютер легко строит суперпозицию всех значе-



манипуляции с ней. Если убрать необходимость для квантовых операторов быть линейными и унитарными, возможны были бы значительно более эффективные алгоритмы. Таким образом, для того, чтобы эффективно решать задачи о поиске прообраза f , требуется, как и в классическом случае, получить дополнительную информацию о свойствах оракула. К сожалению, классически неразрешимые задачи остаются таковыми и для квантовых вычислений.

Алгоритм Шора

Стихийный интерес к квантовым вычислениям спровоцировал Питер Шор, разработав в 1996 году квантовый алгоритм поиска периода функции.

Алгоритм Шора использовал квантовый аналог быстрого преобразования Фурье, и позволил, помимо прочего, быстро находить

ний функции – квантовый «стог сена»! В 1996 году Лов Гровер разработал квантовый алгоритм, дающий квадратичное (по сравнению с классическим) ускорение при поиске прообраза любого оракула. Доказано, что алгоритм оптимален и большего ускорения для квантового алгоритма получить не удастся. Квантовый мир позволяет смоделировать грандиозную по сложности систему, но накладывает ограничения на

множители больших чисел. Задача разложения произвольного большого числа N на множители в классических вычислениях считается сложной. Существующие алгоритмы имеют надполиномиальную сложность от $\log(N)$, в то время как алгоритм Шора на квантовом устройстве справляется с задачей за $\log^3(N)$ операций. Сложность факторизации больших чисел лежит в основе криптостойкости асимметричной схемы шифрования RSA. Алгоритм RSA был предложен в 1978 году Рональдом Ривестом, Ади Шамиром и Леонардом Адельманом. Революционная идея алгоритма заключалась в возможности использования открытого ключа для шифрования информации. Знание открытого ключа не давало возможности расшифровать зашифрованную им информацию. Для расшифровки использовался закрытый ключ, хранимый в секрете.

Возможность быстрой факторизации больших чисел позволяет, зная открытый ключ, вычислить закрытый и, следовательно, сводит к нулю эффективность шифрования. Квантовые вычисления представляются очень перспективным направлением в общей эволюции вычислений.

Квантовый мир со своей возможностью «вместить в себя» столь грандиозные конструкции на сегодняшний день не имеет конкурентов. Проблемы с quantum hardware сейчас уже не выглядят такими устрашающими, как, например, еще десять лет назад. В начале века с огромным трудом «воздвигались» квантовые компьютеры на нескольких кубитах. В 2012 году заявлено об успешных вычислениях на нескольких десятках (и даже сотнях) кубит.

Локомотив науки несется вперед. Квантовый компьютер – уже не экзотическое словосочетание. Кто знает, может, через несколько лет можно будет арендовать квантовые процессоры на Amazon EC2? ■

Появление машин шестого поколения затянuloсь

Текст Эдуард Пройдаков

Не могу сказать, что современная тенденция наращивать число ядер процессора мне нравится. Благо былых ограничений по числу вентиляей на кристалле, которые лимитировали развитие внутренней архитектуры ЦП, сейчас практически нет. Не то чтобы были какие-то серьезные возражения, но мне такой стиль кажется недостаточно интеллектуальным, и более того, он воплощает американский подход к решению сложных задач, который именуется «метод грубой силы», далее МГС.

Российская инженерная школа всегда отличалась тем, что искала оригинальное элегантное решение, позволяющее сделать то же самое, но, благодаря некоторой выдумке, гораздо более скромными средствами и более изящно. Разумеется поиск такого решения весьма трудоемок, и возникает вопрос: а стоят ли того потраченные усилия? Очевидно, что МГС применим для решения разовых задач, имеющих серьезные временные ограничения. Применим он и в поисковых задачах, чтобы показать саму возможность решения. Однако для задач массовых МГС не сильно подходит в силу своей неэффективности. Так и с многоядерностью. Видна ли за наращиванием числа процессоров хоть какая-то «тропинка к храму»? Трудности с параллельностью при наличии тысяч ядер отчетливо видны, а вот каких-то серьезных перспектив, кроме поддержки пресловутого МГС, что-то не видно. Здесь возникает еще один вопрос: а что, собственно, мы ожидаем от развития вычислительной техники, где та земля обетованная, к которой должны стремиться разработчи-

ки? Рост производительности для задач, в которых есть внутренний параллелизм, – это, несомненно, хорошо, и на этом еще какое-то время VT может развиваться, но понятно, что возможности такого развития ограничены. Поэтому хотелось бы, чтобы VT ближайшего будущего обеспечивало бы нам принципиально новые возможности, которые лежат в первую очередь в развитии встраиваемых систем, робототехники и систем искусственного интеллекта. Давайте посмотрим, возможны ли другие пути. Генетические, квантовые и нейрокомпьютеры мы рассматривать здесь не будем, поскольку это машины достаточно специализированные. Нас интересуют машины общего назначения. Первое, на что стоит обратить внимание, – это неизбежность фон-неймановской (престонской) архитектуры процессора. С момента рассылки в 1945 г. Фон-Нейманом известного научного отчета, в котором описывалась эта архитектура, прошло почти 70 лет, а серьезных конкурентов у нее не образовалось. Так называемая гарвардская архитектура недалеко ушла от фон-неймановской – она обеспечивает только раздельное хранение программ и данных, а сам принцип работы процессора никак не изменился. Несколько специфичнее стековая архитектура, но если присмотреться, то это тоже разновидность фон-неймановской архитектуры, только операнды и результаты хранятся в стеке, что налагает некоторые требования к программе. Более интересны нефон-неймановские архитектуры, в первую очередь машины потоков данных. Это некоторая прелюдия, а вопрос собственно в том, что стоило бы сделать в части компьютерных архитектур машин следующего поколения. Мне как бывшему разработчику представляются интересными следующие направления:

1. Прогресс в компьютерных архи-

тектурах определяется прогрессом в организации компьютерной памяти, поэтому важным направлением представляется создание машин с ассоциативной памятью, доступной программисту (АПДП). Поскольку большая часть времени программы тратится на операции поиска и сортировок, то наличие ассоциативной памяти дает громадный выигрыш в производительности. Понятно, что такая память должна быть поддержана программным обеспечением, компиляторами, библиотеками времени исполнения и т. п. Разработка архитектур с АПДП может привести к ряду существенных и пока еще не опробованных решений, таких как многослойная ассоциативная память, когда поиск в ней идет не по одному ассоциативному полю, а по нескольким. Наилучшие области применения таких решений – базы данных, системы искусственного интеллекта, лингвистические программы типа систем автоматического перевода и т. п. Частично в этом направлении двигался академик В. С. Бурцев, который предлагал в своем суперкомпьютере использовать ассоциативную память для ускорения переключения задач.

2. Объектно-ориентированные архитектуры (ООА). В 1980-х годах корпорация Intel выпустила совершенно революционный объектно-ориентированный процессор iAPX-432, однако ряд ошибок в реализации сделал его недостаточно эффективным по производительности, и эта идея была отложена в долгий ящик. Однако сама идея была блестящей, и я полагаю, что сейчас время таких процессоров пришло. С интересом жду момента, когда Intel жестом фокусника снова явит ее миру. Вполне естественно объединить ООА с АПДП.

3. Важным направлением, совсем не исключающим многоядерность, является развитие архитектур, поддерживающих так называемое программирование в ограничениях. Это новая математика, которая показала, что

достаточно сложные задачи, для решения которых использовались суперкомпьютеры, могут быть решены на обычном ПК. Пионером здесь, безусловно, является А.С. Нариньяни, разработавший методы своей недоопределенной математики лет на двадцать раньше, чем они стали развиваться на Западе под названием «программирование в ограничениях». Нариньяни называл компьютер будущего «безалгоритмическим». Здесь можно только посоветовать, что этот талантливый ученый был сильно недооценен в России, и страна упустила шанс развить его совершенно прорывные решения.

4. Для машин следующего поколения что-то нужно делать с видеопамью и операциями над графическими объектами. Нужны новые виды графической памяти. Здесь есть надежды на вытеснение из компьютеров электронных компонентов и замену их на оптические и голографические. Человеческое мышление хорошо умеет делить объекты на части и работать с этими кусочками, битами и пикселями, гораздо хуже у него получается работа с целым объектом. Между тем существует математика, позволяющая производить логические операции сразу над всем объектом, поэтому хотелось бы найти решения, в которых компьютер будет способен без тысячи ядер в GPU работать со сложными графическими объектами. Такую математику многие годы развивал в своей докторской диссертации О. А. Ханджян. Новая графика должна работать за счет свойств графической среды, а не исполнения процессором многочисленных алгоритмов обработки видео. Подводя некоторый итог, видно, что в рассмотренных здесь на концептуальном уровне направлениях в целом наиболее перспективны архитектуры, избавляющие компьютер от засилья алгоритмов. Основную работу должна брать на себя его архитектура. ■

YarcData uRiKA – эврика в семантическом анализе

Текст Александр Фролов

В настоящее время мир стоит на пороге радикального изменения информационных систем, связанного с масштабным внедрением технологий, основанных на семантическом анализе данных. Весной 2012 года Google объявил о запуске Google Knowledge Graph – технологии поиска с использованием семантической базы данных, а производитель суперкомпьютеров Cray – о создании YarcData uRiKA, первой высокопроизводительной системы, ориентированной на анализ семантических баз данных, в том числе и в реальном времени.

Cray и YarcData

В феврале 2012 года компания Cray объявила о создании нового подразделения YarcData, целью которого является внедрение и развитие продукта под названием uRiKA (произносится [juːˈri:kə]), ориентированного на рынок систем анализа семантических баз данных. Потенциальными областями применения таких систем являются бизнес-аналитика, биоинформатика, медицина, телекоммуникации, логистика, анализ социальных сетей, поисковые системы. В целом рынок семантических баз данных в США оценивается в 40 млрд долларов.

Семантические (графовые) базы данных

В отличие от реляционных баз данных, где объекты и отноше-

ния между ними задаются в виде таблиц, в семантических (или графовых) базах данных объекты представляются в виде вершин с некоторыми атрибутами, а отношения – в виде дуг, соединяющих вершины графа. Такой подход к хранению данных является естественным для отображения отношений между объектами. Соответственно, анализ отношений в графовых базах данных реализуется значительно проще, а следовательно, эффективнее, чем в реляционных.

Модель RDF

Одним из стандартов, разработанных консорциумом W3C для реализации концепции Web 3.0 (его еще называют семантическим Web'ом), является RDF (Resource Description Framework). Модель RDF позволяет строить ориентированный граф, состоящий из объектов

и отношений между ними. Для обработки данных, представленных в RDF, разработан специальный язык запросов SPARQL, также стандартизированный W3C.

Информативность графа (реальные графы)

Существует гипотеза о том, что информативность семантического графа прямо пропорциональна квадрату количества его вершин. Другими словами, чем больше размер графа, тем более высока степень полезности извлекаемой из него информации. Таким образом, задача графового анализа может быть охарактеризована как проблема класса Big Data, или «Больших Данных».

Особенности параллельного анализа графов

Параллельный анализ графов имеет ряд особенностей. Во-первых, граф крайне сложно, а в общем случае невозможно равномерно распределить между узлами параллельной вычислительной системы. Как следствие, во время выполнения анализа графа возникает интенсивный обмен сообщениями между узлами, характеризующийся коммуникационным паттерном «все-всем». Во-вторых, связи между вершинами графа нерегулярны, что не позволяет использовать преднакачку данных из памяти (как локальной, так и удаленной). В-третьих, графы могут динамически изменять свою структуру, при

этом эти изменения невозможно предсказать заранее, вследствие чего необходимо обеспечение высокоскоростного ввода-вывода данных ко всем вычислительным узлам системы, что особенно критично для анализа графовых баз данных в режиме реального времени. В итоге традиционные подходы к построению систем для анализа «Больших Данных» приводят к низкой производительности на графовых задачах из-за неадекватности архитектуры вышеперечисленным проблемам.

Общая память, Shmem, MPI

Идеальным способом программирования при решении графовых задач является использование общей памяти с единым глобальным адресным пространством (в стиле OpenMP) – в этом случае не приходится решать задачу распределения графа между вычислительными узлами. Менее удобным является использование программной модели PGAS (распределенное глобальное адресное пространство) с односторонними коммуникациями и ее реализаций, таких как Shmem или UPC. Ну, и самое неудобное – это библиотека MPI с двусторонним обменом сообщениями. Поэтому наиболее распространенной платформой для графовых приложений являются системы с общей памятью типа SMP. Недостатками SMP-систем являются ограниченная масштабируемость и высокая стоимость.

YarcData uRiKA Graph Appliance

YarcData uRiKA – высокопроизводительный аналитический программно-аппаратный комплекс, ориентированный на выявление зависимостей в семантических базах данных сверхбольшого объема. Расшифровывается uRiKA как «universal RDF integration Knowledge



Appliance», из чего следует, что основным способом предоставления данных является RDF-формат. Уникальность uRiKA заключается в применении в качестве аппаратной платформы мультитредового суперкомпьютера с общей памятью Cray XMT2.

Cray XMT2: где установлены системы

На сегодняшний день Cray XMT2 является последним из серии массово-мультитредовых суперкомпьютеров, разработанных фирмой Cray: Tera MTA, MTA-2 и XMT. В настоящий момент суперкомпьютеры Cray XMT2 были установлены в Швейцарском Национальном Суперкомпьютерном Центре (CSCS), Центре Прикладных Высокопроизводительных Вычислений (CAHPC) в Денвиле в США, медицинском центре Mayo Clinic, а также в одной из государственных структур, относящихся к правительству США.

Cray XMT2: архитектура, Бартон Смит

Архитектура Cray XMT была разработана еще в 1990-х годах выдающимся ученым и инженером Бартоном Смитом. Первая реализация появилась в 1998 году. Cray XMT2, как и предшествующие ей системы, является специализированной системой, ориентированной на эффективное выполнение задач, характеризующихся интенсивным нерегулярным доступом к памяти. Частным примером таких задач являются задачи графового анализа.

Cray XMT2: мультитредовый процессор ThreadStorm

В основе Cray XMT2 лежит микропроцессор ThreadStorm4, поддерживающий одновременное выполнение 128 аппаратных потоков (тредов). Таким образом, процессор может выполнять одновременно до 128 программных потоков. Для того чтобы быстро переключать контекст, каждый аппаратный тред имеет свой собственный набор регистров, как управляющих, так и общего назначения. Кроме того, каждое тредовое устройство может выдать до восьми одновременно выполняющихся команд обращений к памяти. Такой подход позволяет обеспечить толерантность к задержкам выполнения команд обращений к памяти за счет большого количества одновременно выполняющихся обращений к памяти и лучшего использования пропускной способности памяти и сети.

Cray XMT2: глобальная общая память

Другой важной особенностью Cray XMT2 является поддержка общей памяти с глобальным адресным пространством, что позволяет вычислительному узлу обратиться к памяти любого другого вычислительного узла, выполнив обычную команду чтения или записи. В

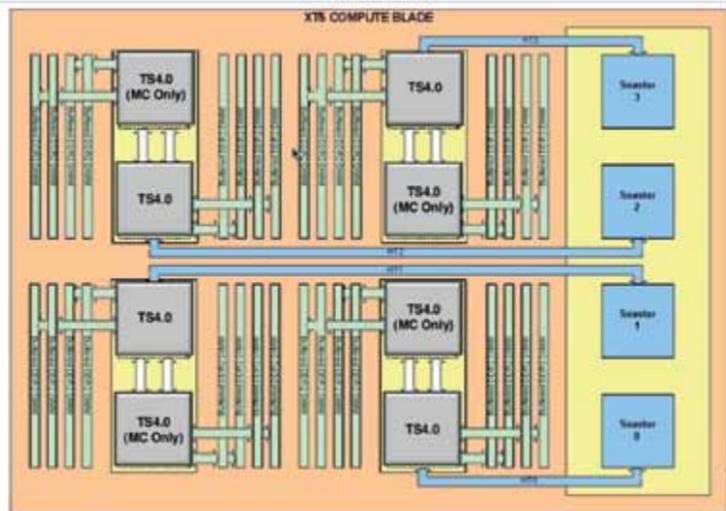


Рис. 2. Вычислительный модуль Cray XMT2

максимальной конфигурации (8192 узла) размер доступной памяти составляет 512 Тбайт. Кроме того, виртуальное адресное пространство скремблируется, то есть адреса равномерно распределяются по вычислительным узлам, использующимся задачей. Такой прием позволяет работать с моделью памяти с псевдоравномерным доступом (UMA).

Cray XMT2: синхронизация, f/e-биты

Важнейшим элементом в архитектуре Cray XMT2 является поддержка синхронизации с использованием теговых битов в ячейках памяти. Каждая ячейка памяти, представленная 64-разрядным словом, имеет два дополнительных бита – f/e-бит и ext-бит. В зависимости от режима обращения к ячейке памяти (Normal, Sync и Future) и значения f/e-бита (full или empty) контроллер памяти выполняет команду обращения к памяти или отклоняет ее, вызывая тем самым повторную выдачу обращения. При превышении заданного предела повторений происходит вызов исключительной ситуации, снятие

программного треда с тредового устройства и передача управления тредовым устройством другому программному треду. Данная возможность позволяет осуществить эффективную синхронизацию программных тредов, избегая активного ожидания. В системе команд Cray XMT поддерживаются атомарные операции обращения к памяти, которые также позволяют эффективно реализовать основные синхронизационные примитивы.

Cray XMT2: инфраструктура XT5 и сеть SeaStar2, сравнение SS2 и Gemini

В качестве сокета процессор ThreadStorm4 использует AMD Socket F, поэтому ThreadStorm4 может быть установлен в вычислительные модули Cray XT4 и Cray XT5. В качестве сети используется Cray SeaStar2+ с топологией 3D-top.

Cray XMT2: вычислительный модуль (TS4 и TS4 (MC only))

Вычислительный модуль Cray

XMT2 состоит из 4 узлов, каждый из которых состоит из двух микросхем: процессора ThreadStorm4 и внешнего контроллера памяти (который является тем же процессором, только с отключенным ядром), соединенного с этим процессором через линк NPL (Node Pair Link). В качестве контроллера памяти используется DDR2 с шириной канала 16 байт и частотой 300 MHz, что дает пропускную способность памяти в 9,6 Гб/с на один процессор (троекратное увеличение пропускной способности относительно Cray XMT).

DDR2 и DDR3

Выбор DDR2 был также обусловлен тем, что DDR2 имеет вдвое меньшую длину вектора (или burst-a), выкачиваемого за одно обращение к памяти из модулей DRAM, по сравнению с DDR3. Поскольку основным режимом работы с памятью является однословный доступ по случайным адресам, то уменьшение длины выкачиваемого из памяти вектора увеличивает долю полезных данных, что повышает производительность задач.

uRiKA

Программное обеспечение uRiKA состоит из клиентской и серверной частей. Клиентская часть предоставляет пользователю набор инструментальных средств для построения SPARQL-запросов и отображения результатов. В основе клиентской части uRiKA лежит программное обеспечение с открытым исходным кодом, в том числе используются такие пакеты, как Apache Tomcat, Apache Jena-Fuseki, а также средства визуализации WS02, Google Gadgets, Relfinder.

Cray Query Engine (CQE)

На вычислительных узлах анализ графов выполняется Cray Query Engine (CQE). Граф в CQE пред-

ставлен в виде списковых структур, работа с которыми аппаратно поддерживается в Cray XMT. Ядро CQE является параллельной мульти-тредовой библиотекой работы с графами, реализующей такие функции, как поиск вширь, кластеризация графа, проверка изоморфизма графов и т. п. При необходимости CQE можно использовать совместно с другими графовыми библиотеками, такими как MTGL или PBGL, для чего предусмотрен специальный интерфейс.

Совместимость uRiKA с другими СУБД

Также предполагается, что uRiKA может быть использована совместно с другими СУБД, как реляционными, так и NoSQL-типа, например Hadoop. Такая возможность позволит пользователям создавать сложные неоднородные системы обработки и анализа

данных с возможностью адаптации к конкретным нагрузкам.

uRiKA 1.0

В настоящий момент выпущена uRiKA с версией 0.9, выход uRiKA 1.0 запланирован на сентябрь этого года. В uRiKA 1.0 планируется добавить поддержку спецификации SPARQL 1.1, что позволит работать с графовой базой данных не только в режиме read-only, но и выполнять запросы по добавлению и удалению данных. Кроме того, планируется расширить набор форматов импортируемых данных, а также улучшить интерфейс администратора базы данных.

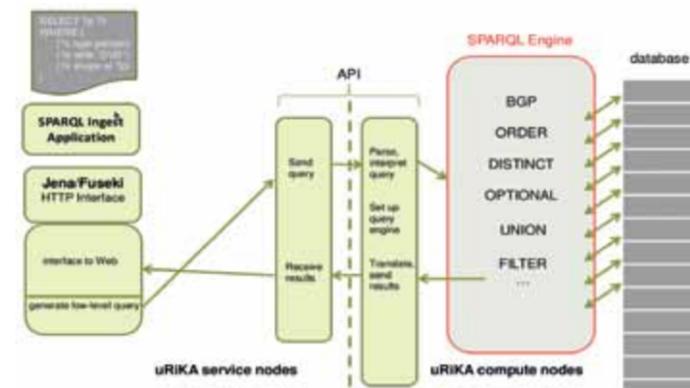


Рис. 4. Взаимодействие вычислительных и сервисных узлов в uRiKA

Перспективы

Задачи с интенсивной нерегулярной работой с памятью, в частности графовые задачи, становятся все более востребованы. Наиболее ярким примером графовых приложений являются семантические базы данных. Развитие технологий высокопроизводительного семантического анализа связано с созданием новых

перспективных архитектур, учитывающих специфику графовых задач. Пример компании Cray показывает, каким образом аппаратная поддержка общей памяти и массово-мультитредовой вычислительной модели позволяет получить качественно новый результат, не имеющий аналогов в своем сегменте.

Весьма интересен вопрос, сохранит ли Cray разработку собственных специализированных процессоров, особенно в свете недавней продажи подразделения по разработке высокоскоростных интерконнектов. Представители Cray не спешат раскрывать свои планы по реализации XMT3. ■

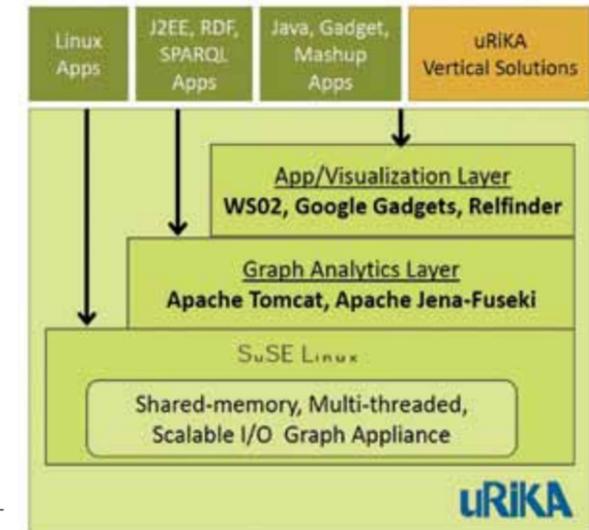


Рис. 3. Программный стек uRiKA

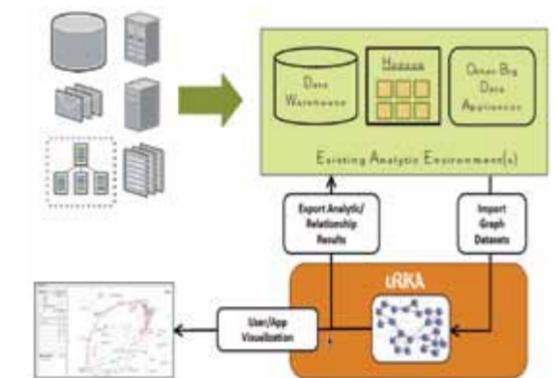


Рис. 5. Взаимодействие uRiKA с другими СУБД

Бесплатное охлаждение

Текст Игорь Обухов
Иллюстрация Владимир Камаев



Как уже писал журнал «Суперкомпьютеры», водяное охлаждение в суперкомпьютерах и центрах обработки данных сопряжено с разнообразными проблемами с обслуживанием и дополнительными тратами из-за необходимости в любом случае обеспечивать часть охлаждения за счет традиционных кондиционеров. В своих статьях мы много говорили о развитии систем воздушного охлаждения за последние несколько лет.

Сейчас в некоторых регионах вполне реально получить эффективность использования электроэнергии на системах охлаждения ниже чем 1.1, что уже достаточно близко к показателям систем водяного охлаждения без учета «традиционной» части. Некоторые производители показывают расчет коэффициента PUE системы охлаждения для Москвы ниже 1.03. Это уже меньше, чем PUE систем водяного охлаждения на такую же мощность: в системе водяного охлаждения на 1 МВт к 30 кВт, которые расходуются вентиляторами градирен, добавляются 10–15 кВт расходуемых циркуляционными насосами жидкостных контуров. После публикации работы «Failure Trends in a Large Disk Drive Population» («Тенденции отказов в большой группе дисковых накопи-

телей», <http://research.google.com/pubs/pub32774.html>), опубликованной группой исследователей компании Google, а также последующих работ, относящихся к другим компонентам серверов, выяснилось, что фразы типа «повышение температуры на каждые X градусов приводит к снижению надежности в Y раз» не совсем соответствуют действительности. Для жестких дисков зависимость скорее обратная, для микросхем памяти заметной корреляции между температурой и частотой отказов просто не обнаружилось. После чего началось своеобразное соревнование – кто загонит сервер в более ужасные условия. Предельным случаем был эксперимент группы сотрудников Microsoft с установкой серверной стойки в палатке во дворе ЦОД. В том эксперименте пять «пожилых» серверов проработали 8 месяцев без единого отказа. За это время на стойку падал забор, проливалась вода, вентиляционные решетки сервера забивались опавшей листвой.

Одним из результатов всех этих исследований и экспериментов стало существенное расширение границ допустимых климатических условий в центрах обработки данных в последней версии рекомендаций ASHRAE (Американское общество инженеров по нагреванию, охлаждению и кондиционированию воздуха).

Конечно же, серьезные вычисления или ответственные задачи в палатке решать не стоит. Потому в любом случае потребуется помещение, а в этом помещении – охлаждение.

Итак, какие сейчас есть варианты «бесплатного» охлаждения? Первый и самый очевидный: *прямое свободное охлаждение*, когда наружный воздух фильтруется, увлажняется и подается в машинный зал. Казалось бы, все просто, надежно и хорошо. Но условия, когда можно просто подавать воздух в зал, бывают далеко не всегда.

В результате появляются камеры смешения с регулирующими заслонками, где в наружный воздух может быть подмешано некоторое количество нагретого воздуха из машзала в холодное время.

Обязательно ставятся датчики задымления и запыления, которые нужны, чтобы перекрыть подачу наружного воздуха, если на улице слишком пыльно или есть задымление (иначе произойдет срабатывание автоматики пожаротушения в машзале, что весьма недешево и приведет к простоям). Как только появляется необходимость работы в замкнутом цикле, возникает и необходимость иметь резервную систему охлаждения на полную требуемую мощность. Эта система охлаждения может быть встроена в установку прямого свободного охлаждения, а может быть и обычной системой охлаждения, использующей стандартные кондиционеры. Одна проблема: мощность такой системы охлаждения должна на все 100% соответствовать мощности машинного зала, а отказоустойчивость системы – соответствовать требованиям отказоустойчивости остальной инфраструктуры.

В результате «бесплатное» охлаждение начинает требовать стандартных затрат на закупку, монтаж и пусконаладку «обычной» системы охлаждения, плюс на установку «свободного» охлаждения. Добавим сюда и автоматику, которая увяжет эти системы в единое целое.

Вода, распыляемая в потоке воздуха, идущего непосредственно в машинный зал, значит еще и необходимость специальной антибактериальной обработки воды и периодической обработки «внутренностей» установки, чтобы избежать риска развития бактериальных культур на деталях системы охлаждения (в том числе и знаменитой легионеллы).

С одной стороны, увлажнение наружного воздуха приводит к снижению его температуры за счет эффекта адиабатического охлажде-

ния, а с другой стороны, подавать слишком влажный воздух в машинный зал тоже не очень хорошо, а значит, и снижение температуры будет не таким большим. Следовательно, достаточно часто будет требоваться доохлаждение воздуха за счет традиционных кондиционеров.

Косвенное свободное охлаждение является логическим развитием идеи прямого свободного охлаждения. Ставится теплообменник-рекуператор между потоком внутреннего воздуха и потоком наружного воздуха. Вариантов теплообменников-рекуператоров существует великое множество – от роторных систем (которые считаются самыми эффективными рекуператорами) до конструкций с использованием тепловых трубок, проходящих через обычные пластинчатые теплообменники.

воздуха. Основное преимущество такого охлаждения в том, что есть возможность добиться насыщения воздуха влагой и таким образом максимально снизить температуру наружного воздуха. С другой стороны, можно «упереться» в количество воды, которую установка способна распылить.

Таким образом, большую часть года в системе охлаждения работают только вентиляторы и иногда небольшой насос в системе распыления воды. Казалось бы, чудеса и панацея от всех проблем с охлаждением. Но чудес не бывает (по крайней мере, не в таких приземленных вещах). За все приходится чем-то расплачиваться. Установки свободного охлаждения полагаются на традиционные системы в случае неблагоприятных атмосферных условий. Установки адиабатическо-

вод еще большего сечения.

Плюс расход воды. Да, средний расход за год может быть порядка 4–6 литров в час. Однако в жаркие дни расход воды будет составлять больше 160 литров в час на мегаватт.

Еще один важный момент – часть системы находится в потоке наружного воздуха. Это значит, что в потоке внешнего воздуха должен стоять фильтр, который может достаточно быстро загрязняться. Как правило, эффективность всех установок свободного охлаждения очень сильно зависит от потока наружного воздуха и при достижении критического уровня загрязненности фильтров эффективность практически всех установок резко снижается. Чистота фильтров должна непрерывно отслеживаться, и запас фильтров для замены всегда должен быть в комплекте ЗИП. В некоторых установках для поддержания чистоты теплообменников используется вода, но в этом случае резко возрастает расход воды и возникают трудности с зимней эксплуатацией.

Все это создает дополнительные финансовые трудности. В результате плата за электроэнергию действительно существенно снижается, но могут потребоваться дополнительные капитальные вложения, появляется существенный расход пресной воды (которая в некоторых регионах является дефицитом и стоит дорого), а затраты на обслуживание зачастую не ниже, чем на обслуживание традиционных систем. Выводы делать пока рано. На охлаждении суперкомпьютеров и ЦОД системы «свободного охлаждения» работают относительно недавно, у производителей очень разная ценовая политика, а у владельцев ЦОД и суперкомпьютеров большое влияние на выбор решения оказывают не технические факторы, а социальные факторы, а именно престижа или, например, возможности получения сертификата энергетической эффективности. 

го охлаждения все равно требуют использования дополнительных систем охлаждения при высокой температуре и/или относительной влажности окружающего воздуха, с одной стороны, и расходуют весьма большое количество воды, с другой стороны.

В результате в большинстве случаев выигрыша по занимаемому пространству не получается. Если кажется, что установка охлаждения занимает мало места, – «значит, вы чего-то не заметили». Например, того, что для отвода мегаватта тепла нужно обеспечить через установку поток воздуха в 200 000 м³/ч. Это 55 м³ в секунду. Это значит, что для систем внутреннего исполнения нужно привести к установке воздуховод сечением никак не меньше 5 м² и отвести от установки воздухо-

Логическим развитием систем косвенного свободного охлаждения является использование адиабатического охлаждения наружного воздуха в теплое время года. Однако не всякие системы косвенного свободного охлаждения могут использоваться совместно с адиабатическим охлаждением. Например, роторные теплообменники не очень сочетаются с адиабатикой из-за неполной герметичности на стыке ротора с воздуховодами и риском коррозии ротора, если в воду попадут ионы «активных» металлов.

Дальше возникает разнообразие методов адиабатического охлаждения, используемых в разных установках, – где-то теплообменник просто поливают водой, где-то вода мелко распыляется в потоке

Предельным случаем был эксперимент группы сотрудников Microsoft с установкой серверной стойки в палатке во дворе ЦОД. В том эксперименте пять «пожилых» серверов проработали 8 месяцев без единого отказа. За это время на стойку падал забор, проливалась вода, вентиляционные решетки сервера забивались опавшей листвой.

Криптоаналитические рекорды «Ломоносова»

Текст Андрей Адинец, Евгений Гречников

Криптографические хэш-функции, в частности SHA-1, широко используются для обеспечения информационной безопасности. Важным свойством таких хэш-функций является сложность построения коллизии, т. е. пары различных сообщений с одинаковым хэшем. Нам удалось развить метод дифференциальных атак и при помощи использования ГПУ на суперкомпьютере «Ломоносов» построить коллизию для урезанной 75-раундовой версии SHA-1, что в настоящее время является мировым рекордом.

Современные информационные технологии, и в частности информационная безопасность, немыслимы без криптографии. Когда браузер проверяет подлинность пароля при помощи сертификата, он использует криптографические алгоритмы, в данном случае – цифровую подпись и асимметричную криптографию. А когда пароль введен и кнопка «Войти» нажата, пароль передается в зашифрованном виде с использованием симметричного шифрования, чтобы избежать его перехвата. Если приватность особо критична, зашифрован будет весь поток данных между пользователем и сайтом. Наконец, если требуется возможность независимого доказательства факта запроса от пользователя, например в банковских операциях, – используется цифровая подпись и на стороне пользователя.

Вся современная криптография основана на решении задач, которые в одну сторону решаются очень просто, а в другую сторону – очень сложно. От сложности решения обратной задачи зависит криптостойкость используемого алгоритма. Например, если есть ключ, то любой смартфон может шифровать или расшифровывать мегабайты сообщений в секунду при помощи алгоритма AES. Но если ключ неизвестен, то его подбор, даже если известны и открытый текст, и шифротекст, – задача очень сложная, для ее решения

не хватит даже мощностей существующих суперкомпьютеров. Если есть два больших простых числа, то посчитать их произведение легко, но вот разложение большого числа на простые множители – вычислительно сложная задача; на этом держится криптостойкость алгоритма асимметричного шифрования RSA. В свою очередь, этот алгоритм используется в цифровых сертификатах, цифровой подписи и для проверки подлинности. Одной из разновидностей криптографических алгоритмов являются криптографические хэш-функции. Криптографическая хэш-функция $h(M)$ – это функция, вычисляющая по сообщению M произвольной

длины хэш-значение H фиксированной длины, при этом по сообщению очень просто вычислить хэш, а вот подобрать сообщение с заданным хэшем – сложно. Таким образом, хэш-значение является как бы «цифровым отпечатком пальца» сообщения и в этом смысле идентифицирует сообщение. Так, криптографические хэш-функции используются в алгоритмах цифровой подписи – подписывается в конечном итоге именно хэш сообщения, а не само сообщение. Также они используются для идентификации блоков в системах управления версиями и некоторых файловых системах. А в базах данных пользователей хранятся не сами пароли, а их хэш-значения.

Важным аспектом криптостойкости хэш-функций является устойчивость к коллизиям. Коллизия – это пара различных сообщений M и M' , таких, что $h(M) = h(M')$. Если для хэш-функции найдена коллизия, она считается скомпрометированной и не может дальше использоваться в практических приложениях криптографии. Таким образом, анализ коллизий является важной частью криптоанализа хэш-функций. Одной из самых широко используемых криптографических хэш-функций является SHA-1 (Secure Hash Algorithm-1), предложенная Национальным институтом стандартов США (NIST) в 1995 году. В настоящее время она использу-

ется как важная составная часть различных государственных и промышленных стандартов безопасности, таких как электронная цифровая подпись, аутентификация пользователей, обмен ключами и построение псевдослучайных последовательностей. SHA-1 внедрена почти во все коммерческие системы безопасности, а также является частью реализации OpenSSL. Хэш-функция SHA-1 устроена по схеме Дамгарда-Меркля. Она оперирует в терминах 32-битных слов и обрабатывает блоки сообщения по 16 слов. В конец исходного сообщения добавляется единичный бит, а затем несколько нулевых бит так, чтобы получилось целое число полных блоков и последний блок без двух слов. В оставшиеся 2 слова записывается длина исходного сообщения. После этого для каждого блока выполняется сжимающая функция. На вход эта функция, вместе с сообщением, принимает результат работы предыдущей такой функции, состоящий из 5 слов. На вход сжимающей функции первого блока подается стандартный вектор начальных значений, а результат сжимающей функции для последнего блока является хэш-значением сообщения. Сжимающая функция состоит из 80 раундов, на каждом вычисляется слово состояния и слово расширенного сообщения (для первых 16 раундов оно совпадает с самим сообщением). Более подробно в популярном виде устройство хэш-функции SHA-1 описано здесь: <http://en.wikipedia.org/wiki/Sha-1>. Мы построили коллизию для урезанной 75-раундовой версии хэш-функции SHA-1, что в настоящее время является мировым рекордом по числу раундов для данной хэш-функции. Для поиска коллизий использовался метод дифференциального криптоанализа, или разностных атак, который был впервые использован для хэш-функций Ван Сяюнь в 2005 году. В этом методе поиск коллизии ведется только среди пар сообщений с фиксированной

разностью (т. е. побитовой суммой по модулю 2), отсюда название. На основании разности сообщений можно дополнительно фиксировать условия на слова расширенных сообщений и состояний, порождаемых исходной парой сообщений. Условия задаются индивидуально для каждой пары битов, связь с другими битами слов состояния и расширенного сообщения определяется только алгоритмом вычисления хэш-функции. Полученный набор условий называется характеристикой. Сам по себе алгоритм построения характеристики достаточно сложен, однако для характеристик, построенных определенным образом, если пара сообщений удовлетворяет характеристике, пара сообщений является коллизией. На раундах 1–16 выполняется подбор слов пары сообщений, а на оставшихся раундах – только проверка сообщения. В таблице 1 приведены условия на биты характеристик и их значения. Условия 10up жестко фиксируют значение пары бит, в то время как «x» и «<» позволяют два возможных значения; биты с такими условиями будем называть свободными. На раундах подбора требуется подобрать значения свободных бит таким образом, чтобы получить значения фиксированных бит, а также бит на раундах проверки, удовлетворяющие характеристики. Для поиска сообщения используется алгоритм поиска с возвратом: сначала подбирается слово сообщения на 1-м раунде, потом на 2-м и т. д. Если на каком-то раунде ни один из вариантов не оказывается успешным, то производится откат на предыдущий раунд, и перебор продолжается на нем. После подбора слова на 16-м раунде выполняется проверка сообщения. Перебор продолжается до нахождения коллизии; используемый алгоритм построения характеристики гарантирует, что коллизия будет найдена. Процесс перебора можно представить в виде дерева, в котором каждая вершина

значению слова для определенного раунда перебора или проверки; таким образом, скорость перебора измеряется в вершинах в секунду. Мы выполняем построение двублочной коллизии, т. е. каждое сообщение коллизии состоит из двух блоков. Это позволяет минимизировать вычислительную сложность задачи. Таким образом, мы используем две характеристики, по одной на каждый блок. Кроме того, на раундах 1–16 можно подбирать или слово сообщения, или слово состояния – одно однозначно вычисляется по другому. Мы делаем выбор, исходя из степени свободы, т. е. количества свободных бит в соответствующей характеристике, – это также позволяет минимизировать объем перебора. Но, несмотря на это, вычислительная сложность задачи все равно остается очень большой. Из-за особенностей алгоритма наиболее сложным является подбор 2-го блока коллизии. Для 75 раундов оценка объема перебора составляет 263 вершины. Обработка одной вершины занимает в среднем 50 целочисленных операций. Таким образом, общее количество целочисленных операций составляет 268.64 – это количество настолько велико, что оно не помещается в 64-битное беззнаковое целое, самый большой фиксированный целочисленный тип, доступный в стандарте языка C. А для выполнения такого количества операций на обычном двухядерном процессоре потребуется... 500 лет. К счастью, данная задача обладает не только большой вычислительной сложностью, но и высокой степенью параллелизма: можно полностью раскрыть дерево перебора до нужного уровня, а после этого вести поиск по всем поддеревьям параллельно. Высокий ресурс параллелизма позволяет использовать для решения задачи суперкомпьютеры. Кроме того, задачу можно ускорить и с использованием графических процессоров. Для этого необходимо немного модифицировать алгоритм: до определенного раунда дерево

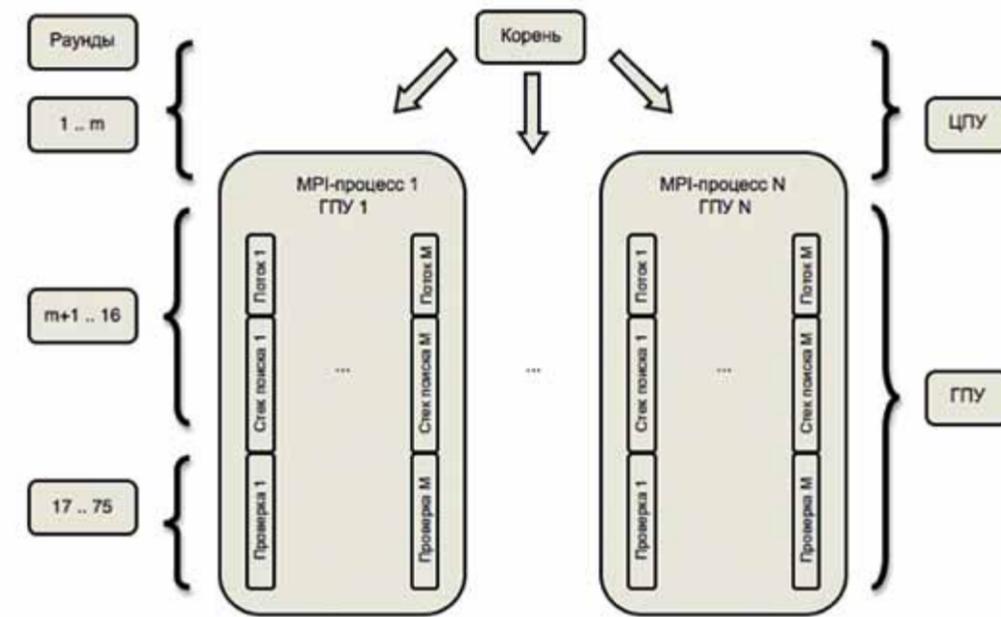


Рис. 1. Схема алгоритма подбора сообщения по характеристике с использованием ГПУ

раскрывается полностью на ЦПУ, а после полученные поддеревья, или стеки поиска, обрабатываются исключительно на ГПУ; ЦПУ служит лишь как вспомогательный процессор для запуска ядер, сбора статистики и записи контрольных точек на диск. Общая схема алгоритма с использованием ГПУ приведена на рис. 1. Алгоритм был реализован на языке Nemerle с использованием системы расширений NUDA для программирования ГПУ. Система расширений NUDA использует OpenCL для генерации кода и взаимодействия с ГПУ. Благодаря особенностям задачи, а

именно минимальному взаимодействию между процессами, обрабатываемыми различными поддеревьями, задача легко масштабируется на любое количество узлов и ГПУ. Это означает, что если на одном узле достигается эффективность 15% (от пиковой производительности ГПУ на целочисленных операциях), то и весь кластер будет использоваться с эффективностью 15%. Разумеется, столь низкая эффективность неприемлема для решения таких больших задач, хочется ее как-то повысить. Наши исследования показывают, что основной причиной снижения производительности ГПУ на данной задаче является расхождение потоков управления между нитями в одном варпе, т. е. низкая когерентность соседних потоков. Расхождение, в свою очередь, вызвано большим количеством ветвлений в алгоритме. А избавиться от них, увы, нельзя, так как при этом будет потеряна суть алгоритма поиска с возвратом. Поэтому мы применим ряд оптимизаций, с целью организовать соседние потоки таким образом, чтобы когерентность между ними в одном варпе была выше. Первой оптимизацией является

сортировка стеков после каждого вызова ГПУ-ядра поиска с возвратом. Наши эксперименты показывают, что устойчивая сортировка (например, поразрядная) оказывается эффективнее, чем неустойчивая (например, быстрая). Мы исследовали различные ключи поиска, и наиболее эффективным среди них оказалось значение поиска (слово сообщения или состояния) для текущего раунда. Кроме того, мы использовали снаппинг: выход из ядра теперь возможен не в любой момент, а только при смене раунда; это позволяет сократить расхождение между потоками. Сортировка и снаппинг дают ускорение (за счет повышения когерентности) в 1.87x раз, при этом сортировка стеков на ГПУ занимает менее 1% общего времени счета. Второй оптимизацией была модификация алгоритма подбора характеристики. На определенных этапах работы этого алгоритма выполняется фиксация бит характеристики, и фиксация различных свободных бит равновероятна. Мы модифицировали алгоритм таким образом, чтобы с большей вероятностью фиксировались биты около «середины»

Обозначение условия	Разрешенные значения бит
x	(0, 1), (1, 0)
-	(0, 0), (1, 1)
1	(1, 1)
0	(0, 0)
u	(1, 0)
n	(0, 1)

Таблица 1. Используемые условия на биты характеристики

раундов подбора. За счет грамотного выбора начальных данных удается ускорить перебор в 1.8x раз. Третья оптимизация связана с модификацией алгоритма поиска. Вместо одного цикла по вершинам дерева теперь используется гнездо из 3 циклов. В самом внутреннем подбирается очередное слово состояния, в самом внешнем выполняется переключение раундов и завершение ядра, а промежуточный цикл используется для эффективного перебора сообщений с проверкой на 16-м раунде. Подобное преобразование позволило ускорить работу алгоритма в 1.25x раз. Выполнялись также прочие оптимизации, такие как использование локальной (`__shared__` в CUDA) памяти, использование константной памяти для хранения характеристик, а также сокращение количества вычислений за счет проверки только состояний (без сообщений). Последнее работает корректно в силу особенностей нашего алгоритма. Ис-

итогам очень длительного расчета 2-го блока эффективность оказалась ниже, но все равно оставалась на уровне 50%, что мы считаем хорошим результатом. С учетом всех оптимизаций использование ГПУ позволяет получить ускорение в 39x раз по сравнению с использованием одного ядра центрального процессора.

Результаты расчетов на «Ломоносове» приведены в таблице 3. Расчет первого блока коллизии завершился за 3 часа. Расчет второго блока занял почти 22 дня. Поскольку в процессе расчета графическая часть «Ломоносова» находилась в режиме бета-тестирования, количество доступных ГПУ колебалось во время расчетов, от примерно 360 в самом начале до 512 в конце.

Но даже если для задачи была бы доступна вся графическая часть «Ломоносова», все равно потребовалась бы как минимум неделя на завершение расчета. И это с учетом того, что реальный объем перебора оказался

Оптимизация	Ускорение
Сортировка + снимки	1.87x
Выбор характеристики	1.8x
Тройной цикл	1.25x
Прочее	1.2x
Итого	4.2x

Таблица 2. Эффекты различных оптимизаций

даже самых мощных современных компьютеров. Да и построение даже 76-раундовой коллизии с помощью текущего метода потребует занять весь ГПУ-раздел «Ломоносова» на два месяца, что, разумеется, невозможно.

Но развитие суперкомпьютеров не стоит на месте; по данным списка

Блок	Оценка объема перебора	Реальный объем перебора	Среднее (максимальное) количество ГПУ	Время расчета
1	2**58	2**54.06	264 (264)	3 часа
2	2**63.01	2**61.92	445 (512)	22 дня

Таблица 3. Характеристики расчетов на «Ломоносове» для поиска коллизии

пользование локальной памяти дало прирост производительности всего лишь 2%, что говорит о том, что предыдущие оптимизации хорошо справляются с повышением когерентности. В целом прочие оптимизации дают ускорение примерно в 1.12x раз, главным образом за счет исключения проверки расширенного сообщения.

Общие результаты оптимизации приведены в таблице 2. В процессе тестирования ускорение и эффективность измерялись на получасовом интервале расчета на ГПУ; на заданном интервале была достигнута эффективность 63%. По

в 2 раза меньше, чем предварительная оценка!

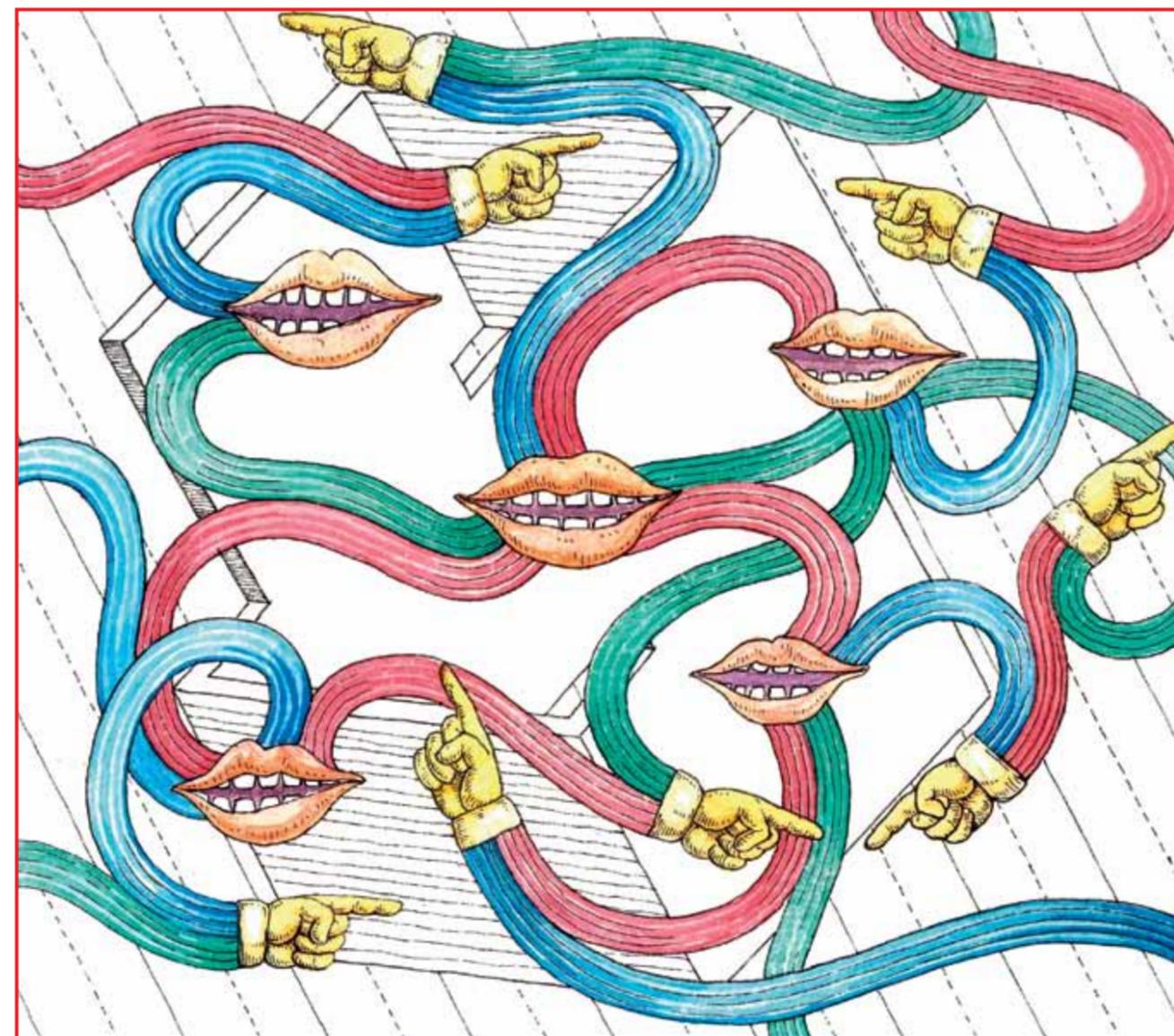
В ходе расчета была построена коллизия для урезанной 75-раундовой версии хэш-функции SHA-1, что в настоящее время является мировым рекордом для данной хэш-функции по числу раундов.

А сколько потребовалось бы ресурсов, чтобы построить коллизии для полной версии SHA-1? По нашим оценкам, объем перебора увеличивается в 8 раз при увеличении на 1 числа раундов; таким образом, для 80 раундов потребуются перебрать примерно 2^{28} вершин, что лежит далеко за пределами возможностей

TOP500, производительность самого мощного суперкомпьютера удваивается в среднем каждые 13 месяцев. Да и криптоанализ развивается, появляются более новые и более совершенные методы, требующие меньших вычислительных ресурсов. Скорее всего, построение коллизии для полной SHA-1 не за горами. Поэтому уже сейчас идут работы по разработке новых криптографических хэш-функций, а Агентство национальной безопасности США проводит конкурс на замещение семейства хэш-функций SHA. Победитель должен быть объявлен в конце 2012 года. 📊

Численное моделирование плазмы на суперкомпьютерных системах

Текст С. И. Бастраков, А. А. Гоносков, Р. В. Донченко, Е. С. Ефименко, А. В. Коржиманов, И. Б. Мееров



Появление мощных суперкомпьютерных систем ознаменовало начало новой эры в вычислительных науках. Многие задачи, о решении которых ранее можно было только мечтать, в настоящее время с успехом решаются на современных вычислительных системах.

Важное достижение состоит в принципиальной возможности замены натурального эксперимента компьютерным, что во много раз сокращает дистанцию между научными гипотезами и их практическим воплощением. Известно, что в науке не все удачные идеи получают признание легко и непринужденно. Некоторым из них приходится пройти долгий путь, ожидая своего звездного часа. Именно так случилось с одним из самых мощных методов современной физики взаимодействия лазерного излучения с веществом, методом моделирования плазмы частицами в ячейках (так называемым методом Particle-In-Cell, или просто PIC-методом). Метод частиц в ячейках известен относительно давно. Он был впервые предложен еще в середине 50-х годов XX века в работах американского ученого Харлоу и его сотрудников. Метод использовался для анализа различных задач газодинамики и физики плазмы. Он постепенно развивался, обрстал соответствующей теорией, став со временем рутинным инструментом в работе многих ученых. В начале 1990-х годов на стыке лазерной физики и физики плазмы появилось новое активно развивающееся направление – физика взаимодействия мощного лазерного излучения с веществом. Хорошо известный и проработанный к этому времени PIC-метод, однако, практически не использовался работавшими в новом направлении исследователями. Господствовало убеждение, что полученные им результаты не обладали достаточной степенью достоверности. Безусловное преимущество в теоретических

исследованиях отдавалось аналитическим методам. Ситуация изменилась в середине 2000-х годов в связи с изучением актуальной проблемы ускорения электронных сгустков при взаимодействии лазерного импульса с плазмой. Лазерный импульс, распространяясь в прозрачной плазме, возбуждает позади себя плазменную волну, подобно тому как возбуждают кильватерные волны корабли, идущие по воде. Эти волны так и называются – кильватерные плазменные волны. Интересно, что эти волны могут быть использованы для ускорения некоторой части электронов до очень высоких энергий. Механизм такого ускорения похож на механизм ускорения серфера на гребне волны. Подобные ускорители в перспективе могут заменить традиционные многокилометровые и очень дорогие линейные ускорители электронов. Лазерные ускорители отличаются компактностью и, как следствие, значительно более низкой стоимостью. Однажды, проводя рутинные расчеты по моделированию процесса распространения короткого лазерного импульса в прозрачной плазме, молодой физик-теоретик Александр Пухов неожиданно для себя обнаружил совершенно новый режим взаимодействия. В этом режиме не возбуждалась плазменная волна в привычном виде. Вместо этого позади лазерного импульса наблюдалась плазменная структура, по форме напоминающая пузырек. Замечательным оказалось то, что новый режим, получивший название bubble-режима, обладал целым рядом преимуществ по сравнению с традиционной схемой. Поначалу многие эксперты сочли

полученный режим артефактом численного моделирования. Только спустя два года напряженной работы было получено экспериментальное подтверждение результатов компьютерного моделирования, а в 2004-м они были помещены на обложку журнала Nature с говорящим названием «Dream Beam»! Этот триумф стал поворотным моментом в изменении отношения научного сообщества к методу частиц в ячейках. Стало очевидным, что численное моделирование может предсказывать результаты эксперимента не хуже аналитических методов. В то же время провести один, два или даже сто расчетов значительно проще, чем получить удобную для использования формулу. Число ученых, использующих PIC-метод для решения задач взаимодействия мощного лазерного излучения с веществом, стало резко расти. Сегодня PIC-моделирование успешно применяется для анализа сложнейших процессов во многих областях современной физики, включая проблему управляемого термоядерного синтеза, астрофизику и квантовую электродинамику. Появились научные группы, практически исключительно специализирующиеся на усовершенствовании метода частиц в ячейках. И, пожалуй, вершиной признания стала публикация в журнале Nature Physics в 2010 году статьи, полностью основанной на численном расчете методом частиц в ячейках. Идея метода заключается в том, что моделируемый газ (или плазма, которая является просто ионизированным газом) представляется как набор некоторого достаточно большого количества взаимодействующих частиц. Вообще говоря, данная идея является вполне естественной, поскольку настоящий газ также состоит из большого коли-

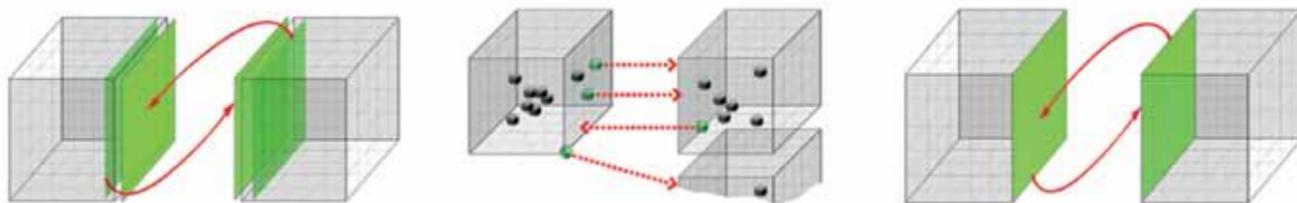


Рис. 1. Схема организации обменов (поля, частицы, токи)

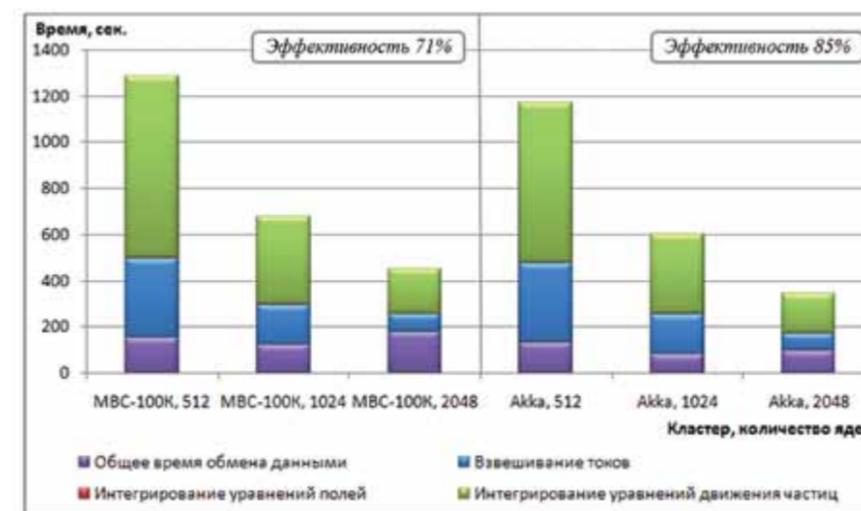


Рис. 2. Масштабируемость параллельной версии

чества частиц. Однако в реальности число частиц столь велико (около 1018 в кубическом сантиметре), что промоделировать их движение на компьютере в полном объеме невозможно. Поэтому в PIC-методе каждая частица – это так называемая макрочастица, соответствующая сразу большому количеству настоящих частиц с близкими параметрами. Точность моделирования при этом, фактически, определяется тем, насколько много частиц удастся включить в расчет, на современных суперкомпьютерах это число может достигать нескольких миллиардов. Кроме данных о частицах метод работает с информацией об электромагнитном поле и плотностях токов и зарядов, заданных на пространственной сетке. При этом частицы взаимодействуют не непосредственно, а только через сеточные значения.

Классическая схема метода, применяемая для моделирования плазмы, может быть приставлена в виде четырех этапов. На первом этапе происходит расчет движения частиц в заданных электрических и магнитных полях. На втором этапе происходит процедура вычисления распределения токов и зарядов, создаваемых частицами (так называемое «взвешивание» частиц). На третьем этапе по вычисленным на предыдущем этапе токам и зарядам

происходит расчет новой конфигурации электромагнитных полей. И наконец, на четвертом этапе происходит «взвешивание» полей, то есть определение их величины в точках, где располагаются частицы, путем интерполяции. На следующем шаге по времени все четыре этапа повторяются. В результате рассчитывается самосогласованная динамика плазмы с учетом собственных полей. При компьютерной реализации удобно и выгодно с точки зрения производительности объединить интерполяцию полей и расчет движения

частиц в один этап. Таким образом, в упрощенном виде схема принимает следующий вид: расчет полей, расчет частиц, расчет токов.

В 2010 году была сформирована исследовательская группа, объединившая сотрудников Института прикладной физики РАН (специалистов в области физики плазмы) и Нижегородского государственного университета (специалистов в области высокопроизводительных вычислений). Наличие в группе ученых разного профиля позволило в относительно короткие сроки разработать новый PIC-код – реализацию метода PIC, ориентированную на эффективное использование современных традиционных и гетерогенных суперкомпьютерных систем. Потребность в использовании таких систем обусловлена исключительной вычислительной трудоемкостью многих задач, представляющих практический интерес.

В настоящее время научное сообщество широко использует несколько PIC-кодов (VLPL, OSIRIS, VPIC и другие), демонстрирующих хорошую масштабируемость на кластерах традиционной архитектуры. В течение последних нескольких лет ведутся активные исследования, посвященные эффективной реализации метода для графических ускорителей (GPU). Рассмотрим основные особенности

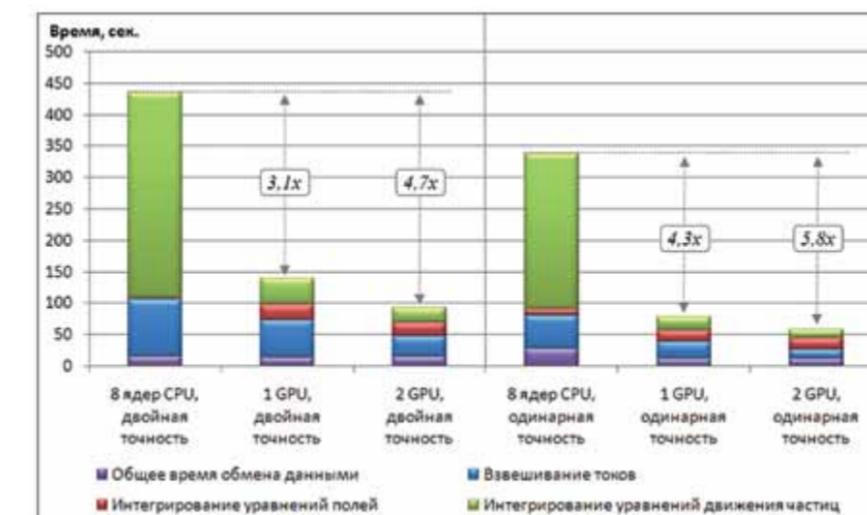


Рис. 3. Производительность на GPU

параллельной реализации, разрабатываемой для использования на гетерогенных кластерных системах. Она ориентирована на случай, когда узлы кластера содержат центральные и графические процессоры (рис. 2а). Используется следующая схема работы:

- *Распределение задачи по узлам вычислительного кластера осуществляется при помощи интерфейса MPI.*
- *На каждом из узлов кластера запускается несколько MPI-процессов. Каждый процесс использует для вычислений либо одно или несколько ядер центральных процессоров, либо один из графических процессоров.*
- *Распараллеливание на ядра центральных процессоров внутри узла производится при помощи технологии OpenMP. Программирование графических процессоров выполнено при помощи технологии OpenCL.*

Декомпозиция задачи осуществляется по территориальному принципу: расчетная область разбивается на подобласти (домены), операции над которыми выполняются параллельно разными MPI-процессами. Процесс, на котором производятся операции над определенным доменом, хранит данные об электромагнитных полях и частицах, находящихся в соответствующей части физического пространства. Данные, относящиеся к узлам, попадающим на границы между двумя или более доменами, хранятся во всех вычислительных процессах, обрабатывающих такие домены. Кроме того, процесс хранит данные о магнитных полях в центрах ячеек, граничащих с обрабатываемым им доменом. Схема работы параллельной версии требует выполнения обменов по окончании каждого из основных этапов (рис. 1). Схема организации обменов устроена следующим образом (рис. 4). При обменах токами и полями каждый домен взаимодействует с 6 соседями, при обменах частицами – с 26 соседями (в случае, когда частица покидает пределы домена, она попадает в один из 26 соседних доменов). Таким образом, все обмены данными осущест-

вляются между процессами, обрабатывающими соседние домены. Основной идеей реализации для GPU является организация схемы хранения и обхода частиц для эффективной работы с памятью: использование «быстрой» памяти (shared memory в терминах CUDA, local в OpenCL), где это возможно, и эффективной схемы доступа к глобальной памяти, при которой соседние потоки GPU обращаются к соседним ячейкам памяти. Для обеспечения высокой степени загрузки GPU, также существенно влияющей на производительность, используется объединение соседних ячеек в так называемые «суперячейки», что сейчас является достаточно популярной идеей реализации метода для GPU.

Рассмотрим результаты расчетов на суперкомпьютерных системах MBC-100K (МСЦ РАН), Akka (High Performance Computing Center North), «Ломоносов» (НИВЦ МГУ), кластер ННГУ. Анализ масштабируемости производился на тестовой задаче моделирования Ленгмюровских колебаний плазмы с использованием 503 миллионов частиц и 17 миллионов ячеек пространственной сетки, двойная точность. На каждое ядро создавался отдельный MPI-процесс. Результаты для систем MBC-100K и Akka приведены на рис. 2. При использовании 2048 ядер на рассматриваемых системах достигается эффективность 71% и 85% относительно 512 ядер. Эксперименты по сравнению производительности CPU и GPU производились на задаче моделирования



Рис. 4а. Слабая масштабируемость CPU-версии на системе Akka

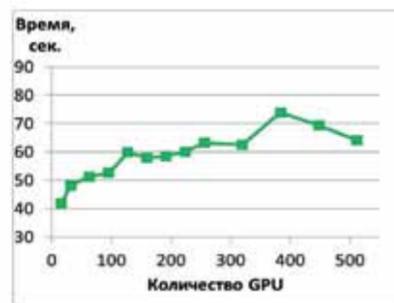


Рис. 4б. Слабая масштабируемость GPU-версии на системе «Ломоносов»

Ленгмюровских колебаний плазмы с использованием 983 тысяч частиц, 33 тысяч ячеек пространственной сетки. Результаты для одинарной и двойной точности, полученные на кластере ННГУ, приведены на рис. 3, времена обменов между оперативной памятью и памятью GPU включены во времена соответствующих этапов вычислений. В двойной точности при использовании 1 GPU достигает ускорения около 3,1 раза относительно версии с использованием 8 ядер CPU, при использовании 2 GPU – около 4,7 раза.

В одинарной точности ускорения при использовании 1 и 2 GPU относительно 8 ядер CPU составляют 4,3 и 5,8 раза соответственно.

При этом достигается около 15% от пиковой производительности GPU, что для данной задачи является неплохим результатом.

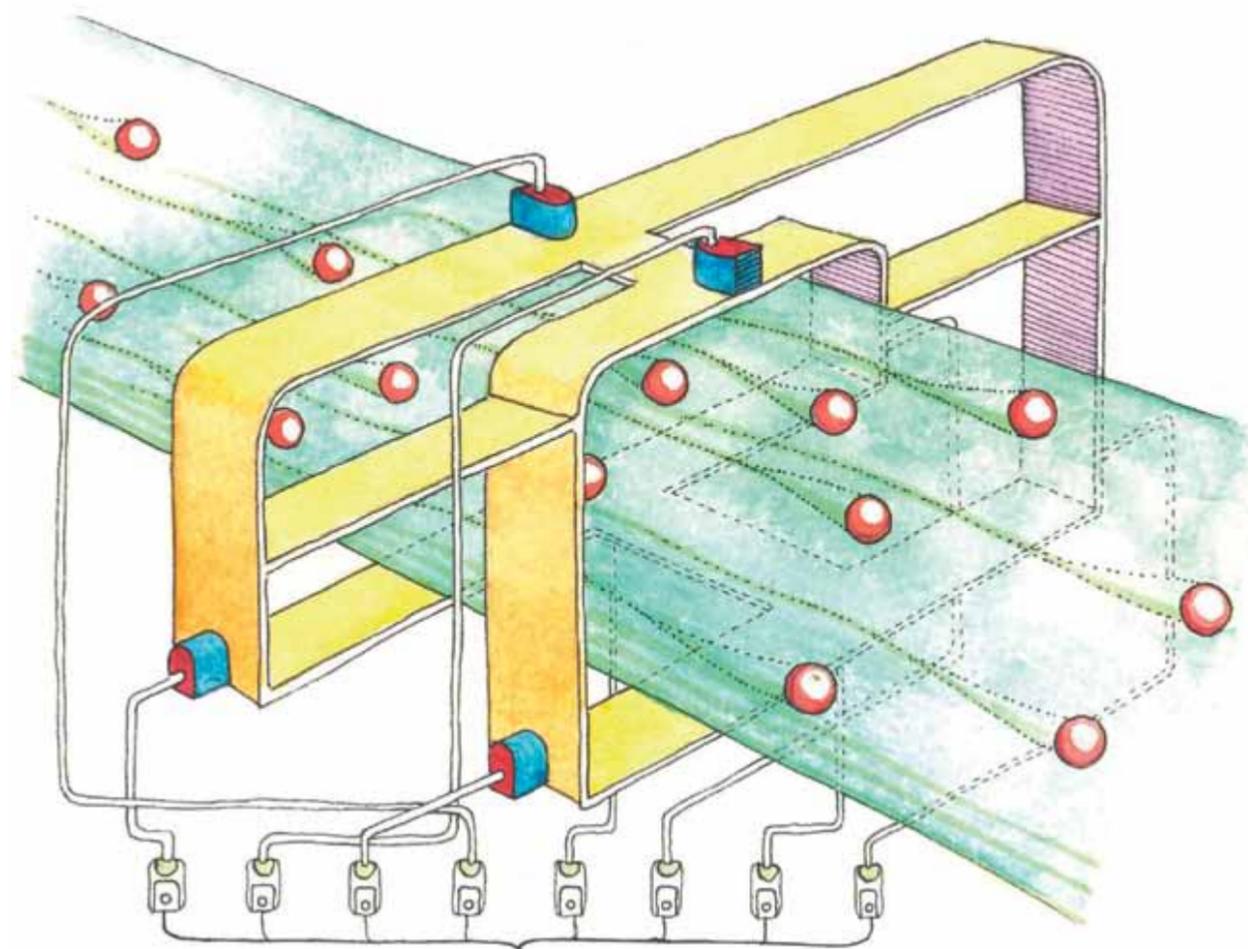
Еще одной важной характеристикой является так называемая слабая масштабируемость – в этом случае вычислительная трудоемкость задачи масштабируется пропорционально количеству используемых вычислительных устройств (GPU либо ядер CPU). Результаты с использованием только CPU, полученные на системе Akka, приведены на рис. 4а.

Результаты с использованием GPU, полученные на системе «Ломоносов», приведены на рис. 4б.

В дальнейшем планируется улучшение производительности и масштабируемости за счет асинхронной схемы передач, а также адаптация существующего кода для перспективных вычислительных систем на базе Intel MIC. ■

Системы управления экспериментом

Текст Дмитрий Мурашов Иллюстрация Александр Желонкин



Стендовые (модельные) испытания – неотъемлемая часть любого современного производства. Они начинаются на стадии научно-исследовательского эксперимента и заканчиваются выводом агрегата в эксплуатацию. Для решения современных задач проведение таких испытаний без применения высокоточных производительных систем обработки данных и контрольно-измерительных систем невозможно. Такие системы обычно являются рабочим инструментом научно-исследовательских институтов и промышленных предприятий.

Программный пакет LabVIEW компании National Instruments (NI) относится к числу тех, которые могут решать одновременно задачи получения данных и управления экспериментом, что дает инженеру функциональный и надежный инструмент, обеспечивающий решение поставленных перед ним задач.

Благодаря современным компьютерным технологиям, таким как высокоскоростные шины PXI и PXIe, USB, платформы на базе ПЛИС, операционные системы реального времени, алгоритмы обработки данных и управления, инженер-расчетчик получает производительность, функциональность и необходимую гибкость системы, которая приводит к снижению затраченного времени, а соответственно, и денежных затрат. В основе технологий National

Instruments лежат модули для подключения различных типов датчиков, контроллеров для создания систем управления в режиме жесткого реального времени, систем хранения данных с интерфейсами обмена данными со скоростью свыше 1200 МБ/с.

При создании систем автоматического управления модули, отвечающие за работу в режиме реального времени и за правильное функционирование ПЛИС, позволяют реализовывать алгоритмы с обратной связью и минимальным фазовым колебанием сигнала на ПЛИС и контроллерах реального времени. Для управления в режиме реального времени используется специальный модуль операционной системы реального времени, который обеспечивает работу с точностью 5 мкс, а также распределяет выполнение задач по установленным

приоритетам. Контрольно-измерительные системы могут использоваться в качестве вспомогательного инструмента при обработке данных, полученных в пакетах инженерного анализа (CAE). Производители прикладного программного обеспечения не всегда уделяют достаточно внимания функциональности постобработки данных, например отсутствия возможности подстраивать заложенные в постобработку алгоритмы. Создание графиков зависимостей и распределений, а в некоторых случаях пространственная визуализация решения также нередко требуют корректировки.

Компания NI предлагает использовать среду постобработки данных NI DIAdem, которая позволяет упростить и ускорить обработку больших объемов данных. NI DIAdem позволяет быстро подготавливать систему для обработки сигналов с десятками тысяч каналов в случае проведения натурного испытания или порядка 10 000 функций в случае обработки численной модели. При этом в случае использования продукции NI как дополнения к CAE-расчетам для каждого объекта возможно обрабатывать порядка 2 млн точек на каждую вычисляемую функцию, применяя следующие параметры функций обработки: проведение статического, корреляционного анализа, наложение фильтров, БПФ (быстрое преобразование Фурье), интерполяция и т. п. и анализ данных. С учетом вышеописанных возможностей и времени выполне-

ния операций идея использования связки CAE+LabVIEW+HPC, а в некоторых случаях LabVIEW+HPC является более чем приемлемым, однако нестандартным подходом. Возможность создания скриптов позволяет полностью автоматизировать процесс сбора, обработки данных по заданному алгоритму и сформировать отчет согласно шаблону, а также использовать инструмент на вычислительных машинах, использующих систему очередей заданий. Таким образом, с помощью среды NI возможно выделить необходимое из огромного количества данных, а взаимодействие с функциями моделирования и проектирования позволяет значительно сократить время. Теоретическое исследование физических явлений в некоторых случаях может иметь некоторые преимущества перед экспериментальным.

Компьютерное моделирование является одним из эффективных методов изучения сложных систем. Компьютерные модели проще и удобнее исследовать в силу их возможности проводить вычислительные эксперименты, которые по сравнению с реальным экспериментом затруднены из-за финансовых и физических препятствий или могут дать непредсказуемый результат.

Логичность и формализованность компьютерных моделей позволяет выявить основные факторы, определяющие свойства изучаемого объекта-оригинала (или целого класса объектов), в частности,

исследовать отклик моделируемой физической системы на изменения ее параметров и начальных условий.

Метод компьютерного моделирования в пакетах инженерного анализа STAR-CCM+ предполагает предварительную верификацию разрабатываемых моделей. Верификация осуществляется на примере трехмерной модели, а в качестве основы используются экспериментальные данные. К сожалению, при проведении эксперимента на измерения накладывается ряд ограничений и допущений. Ввиду того что любой датчик имеет свою геометрию и размеры, их количество весьма ограничено, а соответственно, необходимую характеристику невозможно получить с максимальной точностью, а последующую или смежную точку можно получить только с помощью интерполяции. Также во многих случаях неразрешимой проблемой является его установка в необходимое место, ввиду сложности геометрии модели, или физическая недоступность расположения. Также не стоит забывать, что, если датчик будет располагаться в потоке, он неизбежно будет создавать препятствие течению и, следовательно, и большие погрешности (в лучшем случае). В худшем – измерения не будут соответствовать моделям течения или обтекания объекта. Таким образом, единственным верным решением данной задачи является установка датчиков в характеристические зоны, которые представляют собой опорные показания. При совпадении показаний датчика и решения при компьютерном моделировании в той же локации и при тех же начальных и граничных условиях можно говорить о верном и точном решении в пакете STAR-CCM+. Поэтому с помощью STAR-CCM можно резко сократить количество датчиков на модели без потери качества результата эксперимента, дополняя их результатами численного

При создании систем автоматического управления модули, отвечающие за работу в режиме реального времени и за правильное функционирование ПЛИС, позволяют реализовывать алгоритмы с обратной связью и минимальным фазовым колебанием сигнала на ПЛИС и контроллерах реального времени

моделирования. На этапе создания численной модели в ряде случаев удобно реализовано изменение координат точек численной модели, например, в случае моделирования взаимодействующих объектов можно создавать несколько сот или тысяч файлов с координатами, которые затем имеют возможность поочередно импортироваться в среду CAE, а результаты обработки каждого импортированного файла «накладываются» на последующий. Это является аналогом некото-

рых алгоритмов, реализованных в CAE-пакетах, однако на данном этапе развития кодов такой подход является гораздо более быстрым и простым в освоении, нежели изучение сложных алгоритмов трансформации CAE-пакетов.

Использование такой связки целесообразно, например, для решения таких задач, как пластинчатый насос или обтекание подвижного элемента (крыло).

Сегодня продукция NI использует параллельные вычисления, однако характер их специфичен и отличается от стандартных понятий. Во-первых, это система PXI, которая представляет собой высокопроизводительные вычислительные системы.

Контрольно-измерительные системы в этом случае построены на базе модульной архитектуры с возможностью масштабирования, синхронизации измерений и управления в режиме реального времени. В такую систему входят контроллеры реального времени, каждый из которых – с процессором обычно до 4 ядер с частотой 2,5 ГГц, ОЗУ с 8 ГБ и пропускной способностью контроллера до 8 ГБ/с. Контроллер управляется ОС Windows/Linux. Для систем на базе PXI продукция NI за счет наглядного блок-схемного подхода и функциональности полноценного языка программирования является достойным средством создания масштабируемых приложений обработки больших данных ин-

Компьютерное моделирование является одним из эффективных методов изучения сложных систем. Компьютерные модели проще и удобнее исследовать в силу их возможности проводить вычислительные эксперименты, которые по сравнению с реальным экспериментом затруднены из-за финансовых и физических препятствий или могут дать непредсказуемый результат

женерного назначения. Особенно показательным является использование систем NI RIO, имеющих встроенную ПЛИС с параллельной обработкой данных на частоте от 40 МГц (*основной источник информации – Журнал с конференции от 26 мая 2012 г.*)

Концепция графического программирования NI LabVIEW дает серьезные преимущества разработчикам систем с многоядерными процессорами и прочими системами, позволяющими вести параллельные вычисления. Одно из этих преимуществ заключается в возможности автоматической адаптации приложения для работы в режиме параллельной версии. Для решения задач параллельно можно использовать различные инструменты, например создавать для определенных задач отдельные потоки, которым большинство операционных систем назначат разные ядра процессора.

Однако управление потоками – весьма трудоемкая задача. Несмотря на то, что некоторые среды разработки обеспечивают поддержку управления потоками, они все равно требуют явного указания, какие задачи следует распараллеливать. Программирование на LabVIEW подразумевает создание графического кода (G) в виде диаграммы. Подобный подход называется потоковым программированием. Вместо последовательности выполняющихся друг за другом команд код на LabVIEW представляет

данные и операции, которые соединяются между собой, а компилятор LabVIEW автоматически определяет последовательность выполнения команд. Это также означает, что два параллельных участка кода (ветви диаграммы) независимы и могут быть выполнены на разных ядрах процессора. Способность LabVIEW автоматически создавать и управлять потоками, используя преимущества многоядерных систем, не означает, что у разработчика не остается контроля над потоками. Структура Timed Loop, осуществляющая хронометраж выполняемого в ее теле кода, создает один поток для содержащегося в ней кода и позволяет указать конкретное ядро для выполнения кода в процессе разработки или выполнения программы.

Помимо программирования для многоядерных систем LabVIEW можно использовать для работы с другими платформами, поддерживающими параллельные вычисления. Это, например ПЛИС, процессоры графических акселераторов и даже инфраструктуры облачных вычислений. У каждой из этих платформ есть свои преимущества и недостатки, и LabVIEW позволяет максимально эффективно использовать каждую из них для соответствующих задач. При этом концепция графического программирования позволяет легко объединять компоненты с различными требованиями к организации параллельных вычислений. ■

Математическое моделирование транспортных потоков

Текст В. А. Соловьёв, И. С. Сунгуров, Р. Т. Файзуллин
Омский государственный технический университет

Количество транспортных средств на дорогах городов стремительно увеличивается с каждым годом, и транспортной инфраструктуре все труднее справляться с этим. Поэтому наряду с экстенсивными способами улучшения ситуации (добавление новых элементов инфраструктуры: дороги, развязки и т. п.) возникает необходимость оптимизации работы уже существующих элементов.

Здесь математическое моделирование оказывается крайне полезным, поскольку позволяет просчитать эффективность ввода новых элементов еще на стадии планирования, а также оптимизировать использование уже имеющихся элементов. К естественным возникающим задачам, которые ставят перед математическим моделированием, относятся оптимальное планирование транспортных сетей, увеличение пропускной способности магистралей, оптимизация маршрутов общественного транспорта и другие. Существуют различные подходы к анализу транспортных сетей и транспортных потоков в математических

моделях. Эти различия выражены в используемом математическом аппарате, исходных данных, в объекте движения. Математическое моделирование транспортных потоков возникает в 50-е годы XX века как приложение для столь популярных и усиленно изучаемых в то время начально-краевых задач для уравнений типа закона сохранения. Так, в 1955 г. независимо в работах Лайтхилла, Уизема, Ричардса была предложена, по-видимому, первая макроскопическая модель однополосного транспортного потока, названная впоследствии моделью Лайтхилла-Уизема-Ричардса, в которой поток транспортных средств рассма-

тривается как поток одномерной сжимаемой жидкости. После был предложен ряд модификаций данной модели (модель Танака, модель Уизема, модель Пэйна и другие). Микроскопические модели транспортного потока появились несколько позднее. В основе подходов лежит концепция «о желании придерживаться при движении безопасной дистанции до лидера». Так, в 1961 году Ньюэллом была предложена модель, которую можно считать первой микроскопической моделью. В ней постулируется, что для каждого водителя существует «безопасная» скорость движения, зависящая от дистанции до лидера. Другим важным классом микроскопических моделей, наряду с моделями оптимальной скорости, являются модели следования за лидером. В 1959 году сотрудники концерна «Дженерал Моторс» Д. Газис, Р. Херман, Р. Потс предложили одну из первых (хотя ранее похожие результаты были в моделях А. Рашеля (1950) и Л. Пайпса (1953)) нетривиальных микроскопических моделей однополосного транспортного потока, с помощью

которой можно получить фундаментальную диаграмму – зависимость между интенсивностью потока транспортных средств и плотностью. Ф. Хейт был первым, кто выделил исследование транспортных потоков в отдельный, самостоятельный раздел математики. Решение частных задач моделирования и оптимизации естественным образом приводит к требованию максимального повышения производительности вычислений. На практике движение в режиме реального времени транспортных средств в количестве 10^6 единиц моделируется одним процессором, но задачи моделирования большей размерности и оптимизационные задачи требуют повышения скорости вычислений на порядки. Например, если речь идет об управлении движением транспорта, то число расчетов неизбежно будет велико, но чтобы сохранить актуальность получаемых результатов, они должны происходить существенно быстрее. Таким образом, возникает закономерная необходимость в параллельных вычислениях. Как может быть реализована микроскопическая математическая модель транспортного потока? Будем считать, что в качестве топологической основы используется граф дорог. Тогда дорогами будут ребра графа, перекрестками – вершины, а транспортные средства примем точками на ребрах графа. В качестве источника картографических данных можно использовать сервис OpenStreetMap, который предоставляет свободный доступ к информации о дорогах и других объектах практически любых городов. Далее выбирается интервал времени, соответствующий одному шагу (такту) модели. Каждому ребру графа ставится в соответствие список машин, находящихся в текущий момент на данном ребре (дороге). Разумеется, задается уравнение движения, которое связывает положение и скорость автомобиля в конкретный момент времени.

Также необходимо продумать условия добавления и удаления машин, выбрать алгоритм следования для автомобилей, согласно которому они будут выбирать, на какую дорогу повернуть. Выбор следующего ребра может осуществляться как случайно, так и прямым указанием ребра из списка, заранее полученного в ходе алгоритма, например поиска кратчайшего пути. После выбора всех условий и алгоритмов получаем модель, которая на каждом шаге для каждого ребра просчитывает перемещения автомобилей. Сразу становится очевидно, что с ростом количества автомобилей растет трудоемкость вычислений. И при попытке производить практически сразу столкнемся с тем, что его возможностей не хватает для моделирования городского миллиона в реальном времени при загруженности, сравнимой с той, которую имеем в действительности. Таким образом, приходим к концепции распараллеливания вычислений. Какие мысли возникают по этому поводу? Можно пробовать делить список автомобилей на равные части так, чтобы каждая часть обрабатывалась на отдельном процессоре. Но ввиду того, что каждое транспортное средство рассматривается как отдельный элемент, при моделировании следующего шага необходимо учитывать положение других транспортных средств. Что, если машины находятся рядом, но при этом обрабатываются разными процессорами? Возникнет ситуация, в которой по необходимости присутствуют частые пересылки данных между процессорами, что сразу скажется на производительности всех вычислений. Попытки сгруппировать машины, находящиеся рядом, для того чтобы они обрабатывались одним процессором, ни к чему не приведут, так как транспортные средства не стоят на месте, а передвигаются, следовательно, через какое-то достаточно

короткое время они полностью перемешаются, и мы опять получим ситуацию, которой пытались избежать. Другой очевидный способ заключается в группировке ребер (дорог). При таком подходе считается, что все автомобили, находящиеся на одной дороге, «управляются» этой дорогой. Передача управления осуществляется при пересечении автомобилем перекрестка, т. е. смене ребра, которому принадлежит автомобиль. Следовательно, если два смежных ребра обрабатываются на разных процессорах, при смене ребра транспортным средством необходимо будет передать информацию об этом событии от одного процессора другому. Таким образом, если имеем много «точек соприкосновения» между наборами ребер, управляемыми разными процессорами, то мы должны ожидать и большого количества пересылок данных, что неизбежно приведет к падению производительности. Однако способ группировки по ребрам имеет существенный потенциал, если выделять группы таким образом, чтобы связи между ними были минимальны. Проводя параллель с реальностью,

Рис. 1. Визуализация загруженности дорог (зеленый – слабая, красный – сильная)



Рис. 2. Визуализации движения транспортных средств



обратим внимание на тот факт, что в городах есть районы, связи между которыми осуществляются всего по паре улиц. Другой пример – город разделен рекой, через которую имеется всего несколько мостов. В таком случае, поделив город по мостам и ключевым магистралям, можно получить регионы (регион – группа ребер (дорог), обрабатываемых одним процессором), связи между которыми минимальны, а следовательно, вычисления, производимые над такими структурами, будут более эффективными. Допустим, мы определились, что группировать ребра в регионы – это хорошо, тогда встает закономерный вопрос: «Как получить искомую группировку?». Можно

сформулировать ряд критериев, соблюдение которых позволит получить качественное разбиение: фиксированное количество регионов (по количеству вычислительных узлов); примерно одинаковая общая длина дорог в каждом регионе, наименьшее количество связей между регионами. Можно поделить ребра на группы «на глаз». Так как речь идет все же о карте города, то сделать это вполне реально. На это будет потрачено немало времени, но результат будет неплохо себя показывать, так как количество связей мы регулируем сами. Другой подход заключается в использовании алгоритмов кластеризации, которые как раз направлены



Рис. 3. Разбиение на 2 и 8 регионов (г. Омск)

на группировку, разбиение множества объектов на непересекающиеся подмножества, кластеры (в нашей терминологии – регионы), состоящие из схожих объектов. Существуют различные типы подходов к кластеризации: вероятностные, иерархические, логические, теоретико-графовые, подходы на основе систем искусственного интеллекта и другие. Рассмотрим один из самых популярных и легко программируемых вероятностных подходов, который был изобретен в 1950-х математиком Г. Штейнгаузом, – *k-means*. Алгоритм устроен так, что суммарное квадратичное отклонение от центров кластеров стремится к минимуму. Если кратко описать алгоритм, то он будет состоять из четырех этапов. Во-первых, случайным образом разделить объекты на *k* равных групп. Во-вторых, вычислить центроиды для каждой из групп. В-третьих, отнести каждый объект к кластеру с ближайшим «центром масс». В-четвертых, если хоть один объект был перемещен в другой кластер, повторить, начиная с шага 2. Так как при использовании алгоритма *k-means* в расчет берутся только первый и косвенно второй критерии из описанных выше, в качестве характеристик объектов логично выбрать координаты дороги, а в качестве метрики, или того самого признака разбиения, – евклидово расстояние между координатами дорог. Примеры разбиений, полученных с помощью алгоритма *k-means* для нескольких крупных городов России, можно наблюдать на рисунках 3, 4, 5, где каждый регион (кластер) окрашен в свой цвет. После того как продумана алгоритмическая часть модели, встает вопрос о реализации программной части модели с учетом параллельной архитектуры. Создание одного приложения, производящего расчеты и отображающего информацию, например, в виде карты с

движущимися автомобилями, при параллельной архитектуре довольно затруднительно. Логично будет выделить вычислительную часть и часть, отвечающую за предоставление информации, в отдельные приложения. Таким образом, получаем два приложения: имитатор, который представляет собой распределенное приложение, может быть реализован, например, с использованием протокола обмена сообщениями MPI и производить все вычисления. И графический интерфейс, который получает от имитатора только результаты вычислений на каждом шаге. Физически приложения имитатора располагаются на различных узлах:



Рис. 4. Разбиение на 2 и 8 регионов (г. Новосибирск)

Рис. 5. Разбиение на 2 и 8 регионов (г. Москва)

центральном узле и некотором количестве вычислительных узлов, каждый из которых обрабатывает определенный участок города (рис. 6).

На компьютере пользователя установлен только графический интерфейс, который запрашивает информацию для отображения с центрального узла.

Предлагаемая архитектура позволяет гибко настраивать как графическую часть, так и отдельно вычислительную.

На сегодняшний день имеется обширная литература по изучению и моделированию автотранспортных

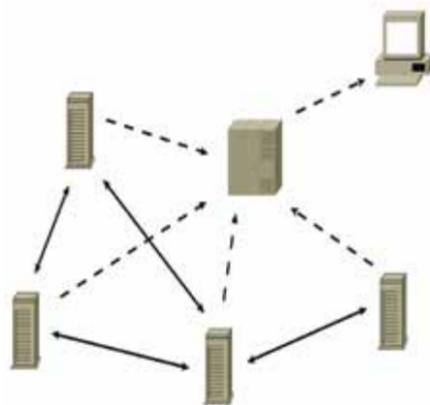


Рис. 6. Схема программного комплекса

потоков. Несколько академических журналов посвящены исключительно изучению проблем автомобильного движения.

Наиболее крупными являются Transportation Science, Transportation Research, Mathematical Computer Simulation, Operation Research, Automatica, Physical Review E, Physical Reports. Количество публикуемых статей исчисляется сотнями.

На сегодняшний день проблемы исследования транспортных систем являются крайне важными и в ряде стран возведены в ранг проблем национальной безопасности.

Применение параллельных вычислительных технологий для решения сложных задач аэрогазодинамики в ВЦ ФАЛТ МФТИ

Текст В. И. Шалаев, А. В. Ваганов, И. В. Воронич, Л. Ф. Ивчик, В. Н. Коньшин, С. А. Рыжов, М. А. Стародубцев, В. В. Ткаченко, В. Л. Юмашев

Факультет аэромеханики и летательной техники (ФАЛТ) МФТИ находится в г. Жуковский, но удаленность от метрополии никак не сказывается на качестве образования, поскольку весь основной цикл ведется общеинститутскими кафедрами: физики, математики, информатики, вычислительной математики, иностранных языков, гуманитарных наук. Это даже имеет свои преимущества, поскольку студентам и преподавателям не нужно тратить время на поездки – базовые организации, ЦАГИ, ЛИИ и ЦИАМ, расположены рядом с факультетом. Именно эти институты, и главным образом ЦАГИ, лидер российской авиационной науки, – определяют направления подготовки и научных исследований на факультете.

Основная тенденция последних двадцати лет – перемещение центра тяжести фундаментальных и прикладных научных работ в сторону численных исследований физических объектов и явлений – связана с быстрым развитием технических и программных средств и в значительной степени относится к задачам внутренней и внешней аэродинамики летательных аппаратов (ЛА). Проблемы, с которыми приходится сталкиваться при решении задач аэродинамики, являются одними из самых сложных в современной математической физике и математическом моделировании, что обусловлено присутствием таких явлений, как отрыв и ламинарно-

турбулентный переход, нелинейность уравнений Навье-Стокса и их сингулярные свойства при больших числах Рейнольдса. Поэтому численные модели должны иметь высокий порядок аппроксимации, и их приходится строить на сетках в сотни и более миллионов узлов, а решение систем такой огромной размерности за разумное время можно получить только с использованием параллельных вычислений на суперкомпьютерах. По этим причинам на ФАЛТ, наряду с факультетом управления и прикладной математики, наиболее хорошо поставлено обучение студентов методам математического моделирования и численного анализа, а начиная с четвертого

курса они непосредственно вовлекаются в сферу решения весьма сложных прикладных задач в базовых организациях. Большая часть представленных ниже результатов получена при участии студентов и аспирантов.

Параллельные вычисления в ВЦ ФАЛТ начались в 2002 году, когда благодаря настойчивости Ю. И. Хлопкова, тогдашнего декана и заведующего кафедрой вычислительной аэродинамики, было приобретено 17 вычислительных узлов Supermicro, из которых был собран кластер и поставлена управляющая программа – планировщик заданий MVS-1000, предоставленная ИПМ имени М. В. Келдыша РАН. На этом маленьком по сегодняшним меркам кластере был разработан пакет программ АРГОЛА-2, реализующий параллельные алгоритмы решения нестационарных трехмерных уравнений Эйлера первого (метод Годунова) и второго (метод Годунова-Колгана) порядка на многоблочных сетках (В. Л. Юмашев и др.). Этот кластер до сих пор находится в рабочем состоянии и используется для обучения студентов параллельным вычислениям и программированию.

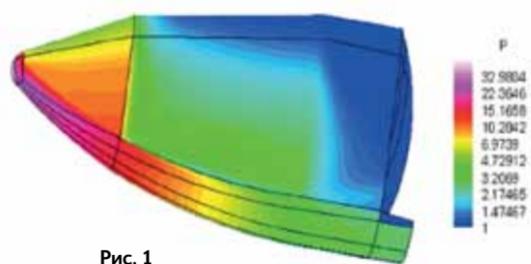


Рис. 1

Моделирование течений невязкого газа

Решение уравнений Эйлера в настоящее время – стандартное средство параметрических исследований аэродинамических характеристик ЛА, поскольку объем сетки и время расчетов на порядки меньше, чем при решении уравнений Навье-Стокса или Рейнольдса с моделями замыкания. Среди множества задач, решенных с помощью пакета АРГОЛА-2, можно выделить параметрические расчеты аэродинамических характеристик проектов многоразовых воздушно-космических аппаратов «Клипер» (РКК «Энергия») и МВКА ЦАГИ вдоль траектории спуска в диапазоне от малых ($M = 1.1$) до больших гиперзвуковых ($M = 25$) чисел Маха, проведенные совместно с ЦАГИ. Все расчеты выполнены на 16-процессорном кластере; для расчета вязкого сопротивления использовалась приближенная методика. Рис. 1 демонстрирует форму аппарата «Клипер» и распределение давления на его поверхности при $M = 6$, $\alpha = 10^\circ$, а графики – хорошее согласие результатов расчетов аэродинамического качества по углам атаки с экспериментальными данными ЦАГИ и расчетами по другим методикам при $M = 4$ и числе Рейнольдса $Re = 4.1 \cdot 10^6$. Математическая модель МВКА ЦАГИ и распределение давления на ее поверхности представлено на рис. 2. Рис. 1 демонстрирует влияние физических свойств воздуха при $M = 16.5$, $\alpha = 40^\circ$ и высоте полета $H = 63$ км на распределение

давления в потоке: верхний рисунок соответствует обтеканию совершенным газом, нижний – воздухом при наличии ионизации и диссоциации за ударной волной. Все расчеты проведены на многоблочных сетках с общим количеством узлов около $7 \cdot 10^6$, время расчетов стационарного течения методом установления – около 24 часов.

Моделирование течений вязкого газа

Задачи внутренней аэродинамики, например расчет течения около компрессора авиационного двигателя, невозможно решить в рамках модели невязкого газа, в связи с присутствием таких явлений, как отрыв турбулентного пограничного слоя и взаимодействие с ним скачков уплотнения. Первые численные решения уравнений Рейнольдса с моделями замыкания SST и k- ϵ в ВЦ ФАЛТ были получены совместно с ОАО «Тесис» для течения около ступени компрессора низкого давления, математическая модель которого и экспериментальные данные были представлены НТЦ им. А. И. Льюльки, где работал один из членов команды ВЦ Л. Ф. Ивчик.

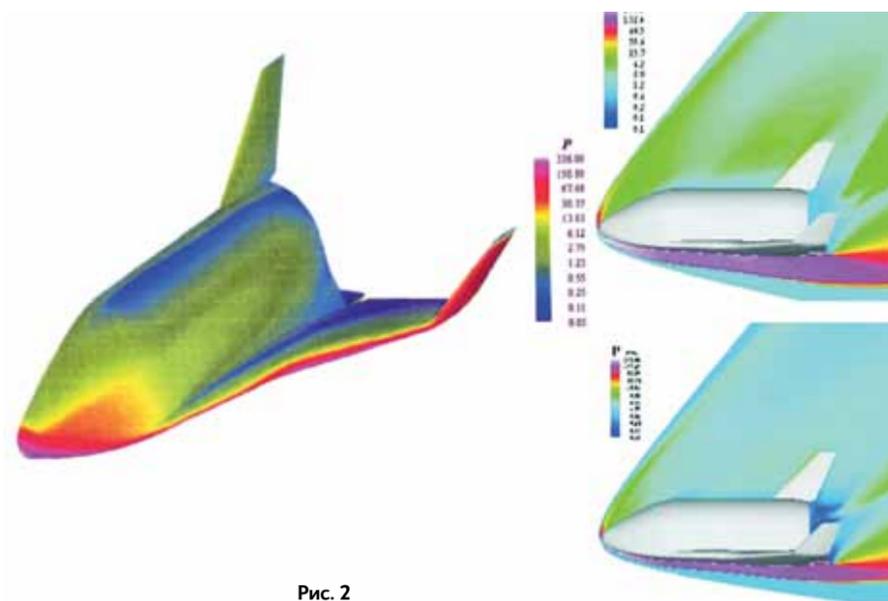


Рис. 2

На рис. 3 показана математическая модель ступени. Результаты расчетов, представленные справа, демонстрируют влияние упругой деформации лопаток компрессора под действием центробежных сил при скорости вращения 28 000 об./мин на результаты расчетов степени повышения давления, полученных с помощью пакета CFX. Качественное согласие с экспериментом (красные точки) было достигнуто только в результате решения сопряженной задачи аэродинамики и упругой деформации (желтые точки) при совместном использовании пакетов CFX и NASTRAN. Эти исследования показали, что расчетная сетка размером около 10^6 , которая допустима на 16-процессорном кластере, явно недостаточна для описания столь сложного течения и получения удовлетворительного согласия с экспериментом даже по интегральным характеристикам. Кроме этого, проводить параметрические расчеты на кластере весьма затруднительно, поскольку время решения для одного варианта превышало 1 месяц. Тем не менее двумерные задачи на таком кластере решать можно, и в качестве примера ниже приводятся результаты численного

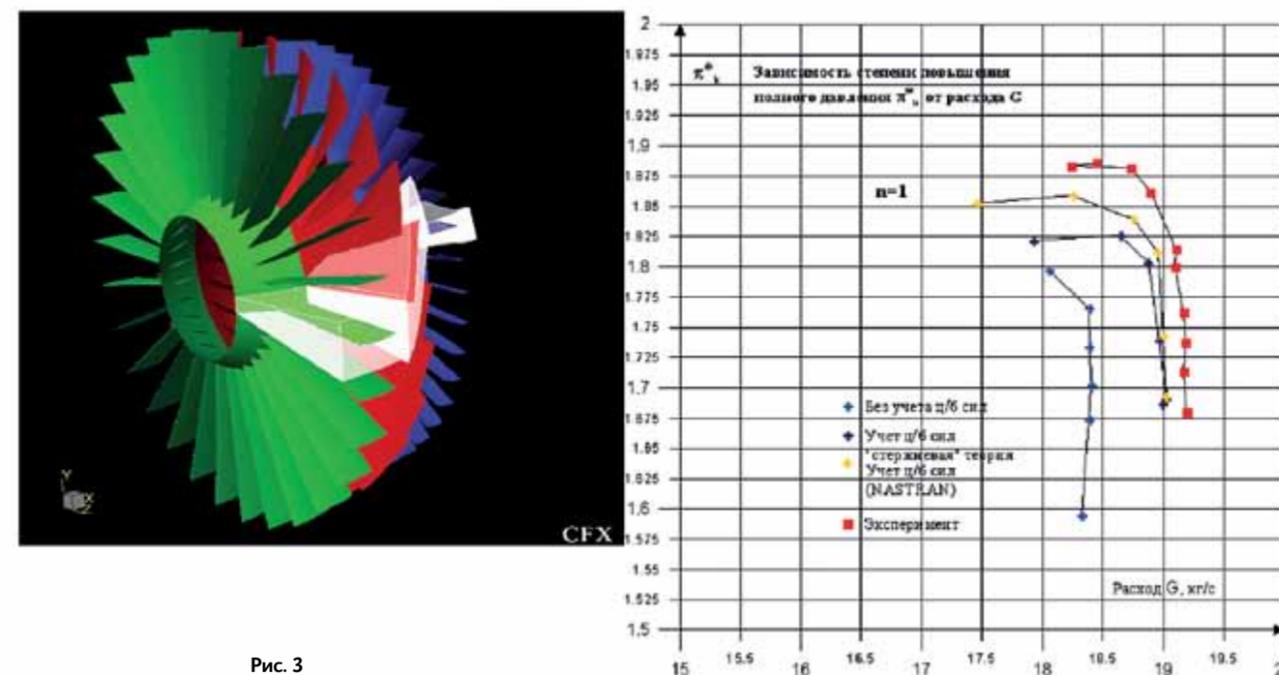


Рис. 3

исследования явления бафтинга – автоколебаний замыкающего сверхзвуковую зону скачка уплотнения – на теплоизолированном профиле NASA0012 длиной 1 м, при угле атаки $\alpha = 4^\circ$, числе Маха $M = 0.75$ и числе Рейнольдса $Re = 10^7$; параметры невозмущенного потока: температура 293°K, давление 57800 Па, скорость 266 м/с. Расчеты проведены с помощью пакета CFX с моделью турбулентности k- ϵ на сетке размером 10^6 узлов, с шагом по времени $0.5 \cdot 10^{-3}$ с. Работа выполнена совместно с ЦАГИ при поддержке МНТЦ по гранту № 2633 (М. А. Стародубцев). В центре рис. 4 приведена кривая изменения коэффициента подъемной силы от времени. Номера точек на графике соответствуют номерам рисунков, показывающих распределение чисел Маха в потоке около профиля в указанные моменты времени. Рис. 2–6 показывают развитие отрыва пограничного слоя из-под скачка уплотнения при его движении вверх по потоку: вначале образуется замкнутая отрывная зона, затем образуется глобальный

отрыв. Отрыв исчезает при изменении направления движения скачка (рис. 8, 1). Появление такого нестационарного режима течения весьма опасно для конструкции и представляет реальную проблему для современной транзвуковой авиации. Изучение этого явления позволяет разработать методы борьбы с ним и выявить безопасные режимы полета. Возможность исследования трехмерных течений вязкого теплопроводного газа в ВЦ ФАЛТ появилась после приобретения при активной поддержке декана ФАЛТ Г. Н. Дудина 96-процессорного кластера ОАО «Т-Платформы» с максимальной производительностью 1 Тфлопс в рамках ФЦП «Инновационные университеты России» – в 2008 году он даже входил в TOP50, хотя в настоящее время, по видимому, не принадлежит даже к первой сотне наиболее мощных кластеров России. Методика расчетов трехмерных течений с помощью пакета CFX была отлажена и верифицирована сравнением с экспериментальными данными ЦАГИ для схематизированной

модели крылатого МВКА М. А. Стародубцевым. Пример расчетов поверхностных линий тока и сравнение с данными экспериментов представлены на рис. 5 для числа Маха 7.5 и числа Рейнольдса 10^6 . Отлаженная методика расчетов позволила решать весьма сложные задачи, связанные прежде всего с разработкой нового поколения высокоскоростных летательных аппаратов. Даже на большом кластере ВЦ время решения уравнений Навье-Стокса или Рейнольдса превышает две недели, вследствие большой размерности сетки и системы конечно-разностных уравнений. Поэтому проведение полномасштабных параметрических расчетов, как в случае модели невязкого газа, не представляется возможным, расчеты проводились только для некоторых условий полета. Одним из критических режимов полета МВКА является транзвуковой режим в силу возникновения нестационарных явлений, подобных бафтингу. Для иллюстрации на рис. 5 представлен вариант расчета вектора сдвиговых напряжений

на поверхности МВКА ЦАГИ при $M = 0.95$, $Re = 7.56 \cdot 10^7$, $H = 10$ км, $\alpha = 5^\circ$ на сетке около $5 \cdot 10^7$ узлов. Исследования выявили наличие перестройки течения при $\alpha \geq 5^\circ$, связанной с образованием больших зон нестационарного отрыва пограничного слоя, что приводит к снижению аэродинамического качества и неустойчивости полета. Область отрыва пограничного слоя в кормовой части сочленения крыла и фюзеляжа хорошо заметна на рисунке. Как показали результаты расчетов, разработанная методика хорошо подходит для анализа деталей течения около ЛА и служит хорошим дополнением к ограниченному эксперименту. Такой комплексный подход позволяет выявить физические механизмы многих сложных явлений, наблюдаемых при испытаниях в аэродинамических трубах. Другим примером комплексного исследования перспективного ЛА является изучение обтекания многоблочной модели ракетоплана, выполненное совместно с ЦАГИ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России», ГК № 14.740.11.0150. Проблема здесь связана с возможным перегревом поверхности аппарата в областях отрыва потока. На рис. 7 приведено сравнение данных визуализации поверхностных линий тока в аэродинамической трубе ЦАГИ Т-117

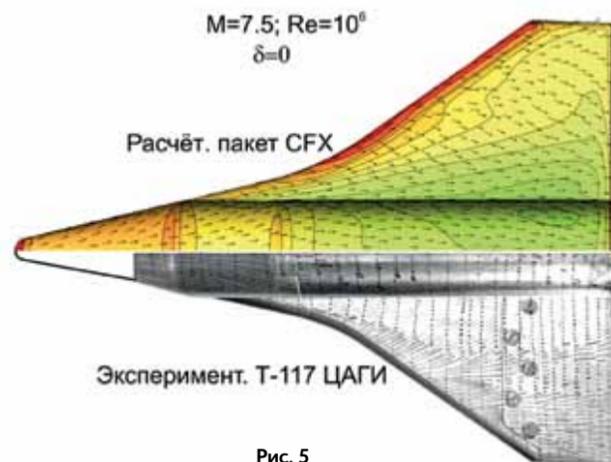


Рис. 5

(рис. 7а) с результатами расчетов (рис. 7б) ламинарного обтекания модели при $M = 7.5$, $Re = 2.76 \cdot 10^6$, $\alpha = 10^\circ$ с помощью пакета CFX на сетке, содержащей $32 \cdot 10^6$ узлов. Цифрами на экспериментальной картинке показано: 1 – линия отрыва, 2 – линия присоединения, 3 – трехмерное циркуляционное течение

в отрывной области, 4 – области пикового усиления нагрева в зоне присоединения потока. Нетрудно убедиться, что результаты расчетов и визуализации согласуются, положение критических точек и линий хорошо моделируется, так же как и локальные области максимального теплового потока; отрыв в данном случае является нестационарным и численные расчеты также воспроизводят это явление.

Моделирование аэродинамики прототипов реальных высокоскоростных ЛА сложной геометрической конфигурации – важная прикладная задача математического моделирования, позволяющая ускорить развитие аэрокосмической техники. Однако следует отметить, что более или менее точное воспроизведение деталей обтекания реального ЛА возможно лишь на суперкомпьютерах класса «Ломоносов» при использовании чис-

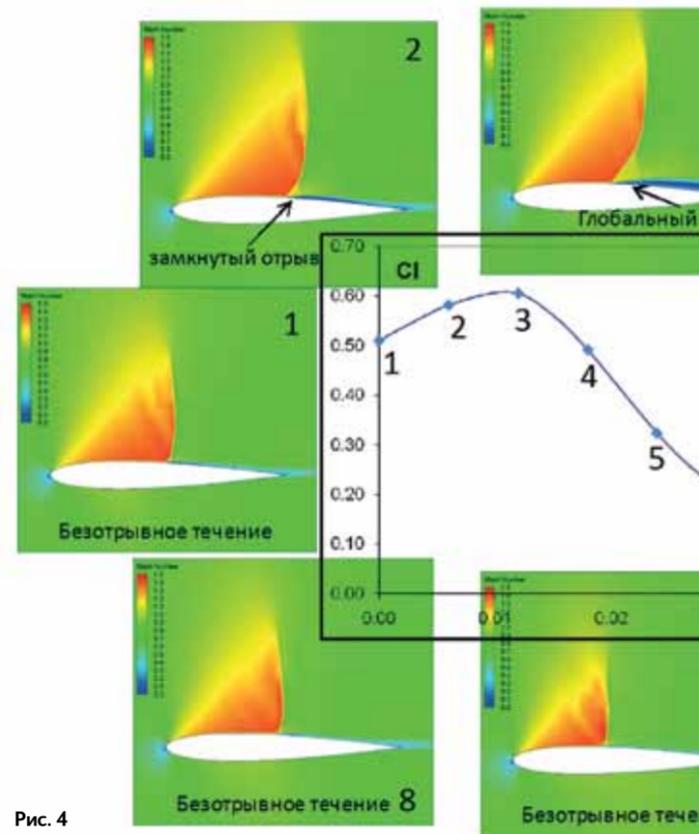


Рис. 4

ленных схем высокого порядка на сетках в триллионы узлов, и только для определенного физического состояния потока – ламинарного или турбулентного – адекватной модели для ламинарно-турбулентного перехода в настоящее время не существует.

Хорошим аргументом в пользу этого положения могут служить проводимые несколько лет экспериментальные и численные исследования гиперзвукового течения около наветренной поверхности плоского треугольного крыла с затупленными передними кромками – классического объекта простой геометрии. На рис. 8 представлена фотография конечного состояния термоиндикаторного покрытия на наветренной (нижней) поверхности такого крыла при испытаниях в аэродинамической трубе Т117 ЦАГИ, условия эксперимента указаны на фото. Более темные области соответствуют большим тепловым потокам на поверхности, причем локальный нагрев обусловлен как ламинарно-турбулентным переходом (красные клинья средней части крыла), так и явлениями, связанными со структурой головной ударной волны (вытянутые вдоль потока линзообразные светлые области пониженного теплового потока в средней по размаху части крыла) и эволюцией высокоэнтропийного этропийного слоя, образующегося перед затупленной

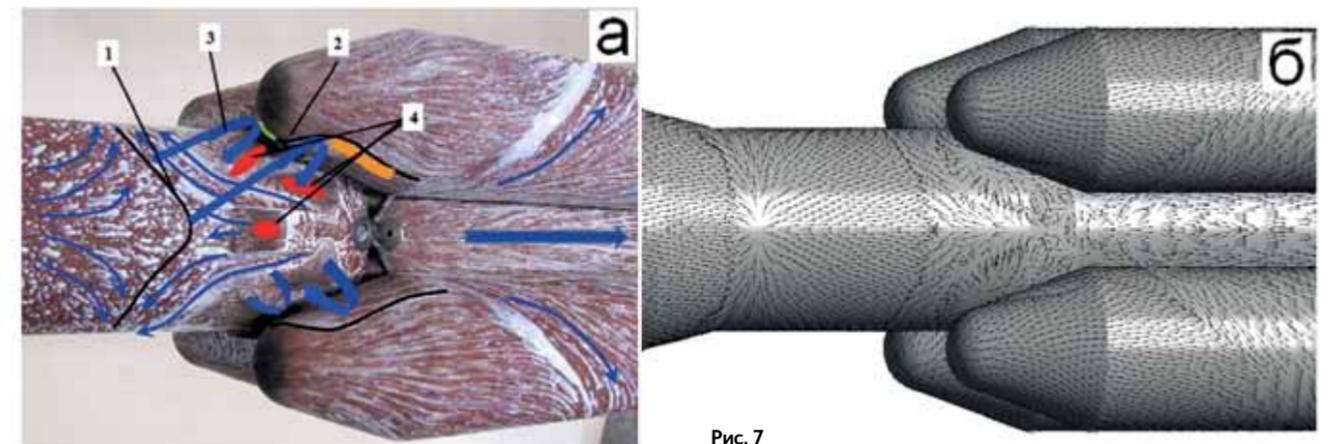


Рис. 7

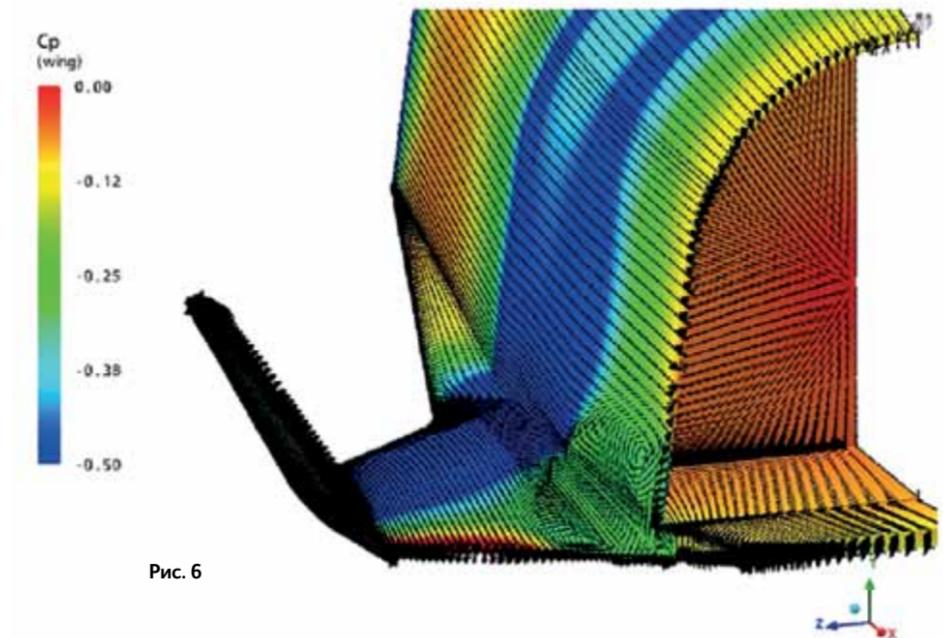


Рис. 6

вершиной (светлые, вытянутые образования около плоскости симметрии). Отметим, что подобная структура тепловых потоков и ламинарно-турбулентного перехода имеет место и при нулевом угле атаки, она наблюдалась во многих экспериментах, но до сих пор не получила адекватного толкования. Некоторые объяснения экспериментальным данным дают результаты численного моделирования тонкой структуры течения около крыла, проведенного при условиях, указанных на рис. 8 на сетке размерностью $30-50 \cdot 10^6$. На рис. 9 сверху приведено поле тепловых

потоков на фронтальной проекции передней кромки крыла, левая граница соответствует плоскости симметрии, максимум теплового потока около нее (красный цвет) – окрестности критической точки в вершине крыла. Недалеко от вершины располагается еще одна область высоких тепловых потоков, наблюдаемая также в экспериментах, появление которой связано с S-образным искривлением головной ударной волны в переходной области к режиму обтекания скользящего крыла. На нижней половине рис. 9 цветом представлено поле статической энтропии, а

кривыми – линии тока в плоскости поперечного сечения, удаленного от вершины на расстояние $X = 300$ мм. Из представленных результатов можно заключить, что высокоэнтропийный слой за ударной волной около плоскости симметрии сворачивается в явно выделяемый продольный вихрь, а светлые области вблизи плоскости симметрии на экспериментальных данных соответствуют области пониженных тепловых потоков под этим вихрем. Правая граница вихря в плоскости крыла – линия растекания поперечного течения, показанная на рис. 8 кружками, – совпадает с правой границей области пониженных тепловых потоков. В средней по размаху части крыла образуются два менее интенсивных, вытянутых в поперечном направлении продольных вихря, левый из которых расположен над светлой областью пониженных тепловых потоков в средней по размаху части крыла – линия растекания попереч-

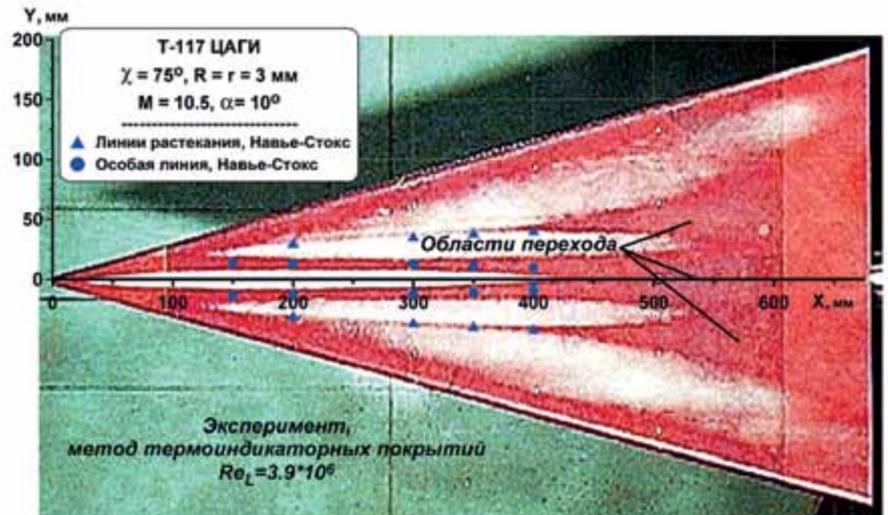


Рис. 8

Сложные вихревые образования, по-видимому, индуцируют механизм перехода в средней по размаху части крыла (клинья между светлыми областями на рис. 8), обусловленный неустойчивостью поперечного течения, что при небольших числах Маха не наблюдается.

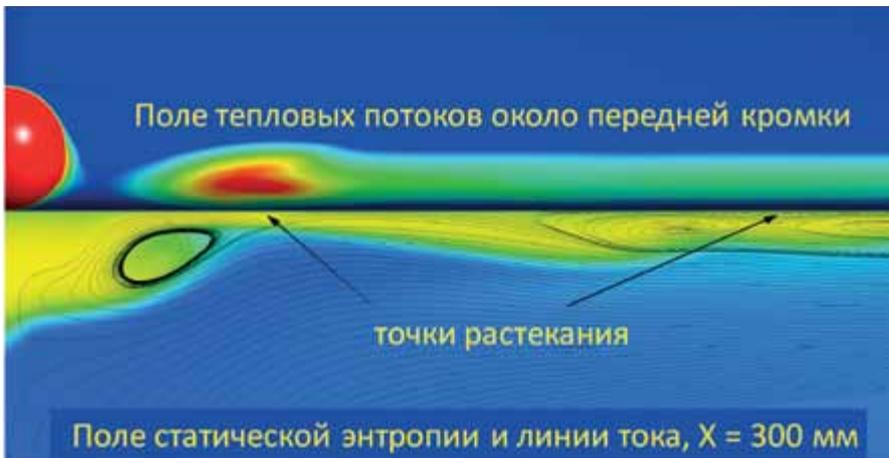


Рис. 9

ного течения между двумя вихрями совпадает с правой границей этой области (треугольники на рис. 8). В этой же задаче в явном виде встает и другая проблема современной аэродинамики – проблема предсказания положения ламинарно-турбулентного перехода в трехмерном гиперзвуковом пограничном слое.

Следует отметить, что поперечная скорость составляет всего несколько процентов от продольной скорости, но тем не менее приводит к весьма значительным эффектам. Для моделирования развития неустойчивых волн и ламинарно-турбулентного перехода в рамках теории гидродинамической неустойчивости или

с помощью прямого численного моделирования необходимо, чтобы поперек пограничного слоя и энтропийного слоя располагалось примерно 10^3 узлов сетки, и, кроме этого, во втором случае хотя бы 10 узлов должны располагаться на длине волны. Выполнение этих требований в случае трехмерного течения приводит к увеличению размерности численной задачи на порядки, что реализуемо только на самых больших суперкомпьютерах. Опыт численного моделирования течения около треугольного крыла на сетке с размерностью 10^8 показал, что этого недостаточно для моделирования возмущений. В заключение хотелось бы сказать, что наш взгляд на будущее супервычислительных технологий оптимистичен.

Появление в ведущих научных и образовательных центрах России больших многоядерных компьютеров и развитие образования в этом направлении несомненно принесет свои плоды. Однако наш опыт показывает, что успешное решение сложных задач фундаментального и прикладного характера обусловлено не только применением современных математических методов, но и ясным пониманием сути физических проблем в анализируемых явлениях. 



СУПЕРКОМПЬЮТЕРНЫЙ КОНСОРЦИУМ УНИВЕРСИТЕТОВ РОССИИ

УЧРЕДИТЕЛИ: Московский государственный университет имени М.В.Ломоносова
Нижегородский государственный университет имени Н.И.Лобачевского
Томский государственный университет
Южно-Уральский государственный университет

ПРЕЗИДЕНТ КОНСОРЦИУМА — Ректор МГУ имени М.В. Ломоносова, академик,
вице-президент РАН В.А.Садовничий.

Цель:

Использование потенциала высшей школы для развития и активного внедрения суперкомпьютерных технологий в российском образовании, науке и промышленности.

Как принять участие:

Постоянным членом Консорциума может стать любое высшее учебное заведение России, имеющее государственную лицензию на ведение образовательной деятельности и заинтересованное в расширении своей деятельности в области суперкомпьютерных технологий и высокопроизводительных вычислений.

Ассоциированным членом Консорциума может стать любая организация вне зависимости от формы собственности и сферы деятельности, заинтересованная в активном использовании и развитии суперкомпьютерных технологий как в университетах России, так и в стране в целом.

В настоящее время в составе Консорциума
25 постоянных и 10 ассоциированных членов.

ПОДРОБНЕЕ НА

HPC-RUSSIA.RU





Суперкомпьютер «Ломоносов»
создан с использованием
инновационных решений
Mellanox на основе технологии
InfiniBand.

Более подробную информацию
можно найти на веб-сайте
www.mellanox.com

ISS600

Матричный
компьютер
InfiniBand
с 648 портами
по 40Gb/s

ConnectX-2



Плата
контролера
InfiniBand
40Gb/s/10x