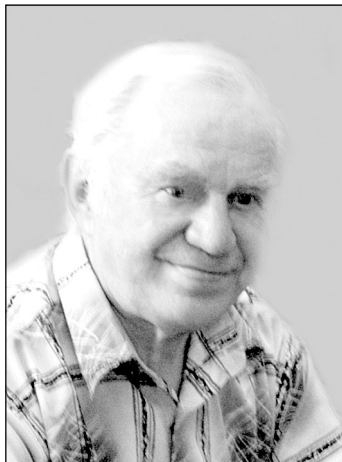


О НАУКЕ — ДОСТУПНО

Ухабы на дорожной карте математического моделирования



В.П. Ильин
главный научный сотрудник ИВМиГ СО РАН,
доктор физико-математических наук,
профессор НГУ, руководитель проекта
РФФИ № 11-01-00205

Введение.

Успехи порождают проблемы

Согласно закону Мура о компьютерной эволюции, 2013 год является приблизительно «медианой» постпетафлопного периода: 2008 г. — время появления первого в мире «петафлопника», а через 11 лет, в 2019 г., прогнозируется рождение экзафлопной машины (1 экзафлопс = 10^3 петафлопс (пфлопс) = 10^6 терафлопс (тфлопс) = 10^9 гигафлопс (гфлопс) = 10^{18} «усреднённых» арифметических операций с плавающей запятой в секунду). На текущий момент всё идёт по плану. Согласно объявленному в ноябре 2012 г. списку суперкомпьютеров TOP-500, в мире тогда имелось 23 многопроцессорные вычислительные системы (MBC) с производительностью выше 1 пфлопс: 10 — в США, 4 — в Японии, по 2 — в Германии, Китае, Великобритании и Франции, 1 — в Италии.

Вопрос измерения быстродействия MBC сам по себе является дискуссионным, но единственной общепринятой на сегодняшний день меркой является средняя скорость выполнения арифметических операций при решении алгебраических задач с помощью широко известного пакета прикладных программ (ППП) LINPACK. Другие ключевые характеристики — это теоретическое, или пиковое, быстродействие, объём оперативной памяти, количество вычислительных ядер и потребляемая электрическая мощность. Например, самый мощный в мире компьютер ТИТАН, расположенный в Окриджской Национальной лаборатории (США), имеет теоретическую скорость 27,1 пфлопс и 17,6 пфлопс — «в линпаках», 560 640 ядер, память — 0,71 петабайт = $7,1 \times 10^{14}$ байт (один байт состоит из восьми двоичных разрядов) и энергопотребление 8,2 мегаватт. Для сравнения приведём показатели находящегося на 26-й позиции в списке TOP-500 российского компьютера ЛОМОНОСОВ из МГУ: скорость в «линпаках» — 0,9 пфлопс, теоретическое быстродействие — 1,7 пфлопс, число ядер — 78669, что свидетельствует об отставании наших темпов роста, в том числе за последние 3–4 года.

На 2014 г. в мире уже запланировано появление компьютера с реальным быстродействием более 100 пфлопс, так что «пришествие» экзафлопника в 2019 г. наверняка состоится, и закон Мура не будет нарушен. При этом кардинального изменения архитектуры не ожидается, да и частота передачи сигналов остается по-прежнему на уровне нескольких гигагерц. В ближайшем будущем MBC — это кластерная сеть из многопроцессорных узлов гетерогенного типа, в которых процессоры, в свою очередь, состоят из множества вычислительных ядер и делятся на два основных вида: универсальные (центральные процессорные устройства) или специализированные ускорители (графические процессорные устройства, а также ПЛИС — программируемые логические интегральные схемы). Наибольшие хлопоты представляет оперативная память, которая распределена по узлам, обеспечивает общий доступ из «своих» процессоров и имеет многоуровне-

вую иерархическую структуру. В развитии суперкомпьютеров наблюдается закономерность — с увеличением быстродействия почти пропорционально возрастает объём хранимых данных, и эта полумистическая связь между совсем разными, казалось бы, категориями оказывается практически всегда оправданной.

Взрывной количественный рост технического оборудования приносит качественно новую ситуацию для математика-программиста. Во-первых, согласно теории вероятности становится вполне реальным отказ какого-либо устройства во время исполнения алгоритма, и это необходимо учитывать при программировании. Во-вторых, стоимость электроэнергии оказывается очень существенной в общих расходах на эксплуатацию суперкомпьютера. Серьёзность ситуации совсем не шуточная, так как при сохранении нынешнего отношения «быстродействие / мощность» на уровне 2 гфлопс/вт экзафлопная машина потребует более 100 мегаватт, а это уже небольшая, но электростанция. Снижение энергопотребления в несколько раз — одна из главных технических целей компьютерных конструкторов, а самые оптимистические прогнозы здесь — уменьшить количество мегаватт до 20.

В этой связи важно отметить, что коммуникационные операции — намного более энергозатратные, чем арифметические. Отсюда возникает новая проблема построения наиболее «дешёвых» алгоритмов в необычном смысле этого слова, т.е. требующих наименьших обменов данными. Вообще говоря, сокращение коммуникационных потерь — это традиционная задача теории и практики параллельных высокопроизводительных вычислений, но формулируемая в плане минимизации времени решения задачи или выполнения алгоритма на MBC. А это уже другая оптимизационная проблема для многопроцессорного компьютера. Например, если межпроцессорный обмен какого-то объёма данных совместить во времени с реализацией арифметических действий, то ускорения вычислений мы достигнем, но энергозатраты при этом останутся неизменными, так как сами коммуникации не уменьшаются.

Наука «теоретическое программирование» пока далека от возможности дать практические рекомендации по эффективному программированию на сложнейших компьютерных устройствах, поскольку сама модель таких вычислений фактически отсутствует (сапожник без сапог). Поэтому качественная реализация параллельных алгоритмов — это сложная эмпирическая работа, требующая большого объёма технических знаний, опыта, терпения и не предназначенная для слабонервных.

Естественно, что суперкомпьютер петафлопного уровня должен становиться центром коллективного пользования, другими словами — центром обработки данных (ЦОД, или Data Center), а сами расчёты могут проводиться в режиме удалённого доступа через Интернет. Такие технологии «облачных» вычислений (Cloud Computing) уже становятся реальностью, и при этом пользователю зачастую предлагается не просто машинное время, а вычислительные услуги с программным обеспечением для решения требуемых задач (SaaS — Software as a Service). Теперь какой-либо отраслевой организации или университету нет смысла заводить собственный дорогостоящий вычислительный центр и структурную службу, а лучше обращаться к «облаку».

Следует отметить, что если за последние 11 лет компьютерные мощности увеличились в 1000 раз, то отнюдь нельзя сказать, что за этот период выросла массовость эффективного применения математического моделирования хотя бы в 100 раз, в том числе и в России. В дальнейшем мы как раз будем интересоваться отечественной ситуацией, которую можно оценить и по уровню публикаций на соответствующие темы, и по характеру конференций, и по профессиональному общению. Сейчас в России несколько десятков «суперкомпьютеров» с быстродействием от 10 до 100 тфлопс, которые публикуются в списке TOP-500, включающем страны СНГ. Вопрос количественной оценки «роста моде-

лирования» очень непростой. Здесь можно даже вспомнить печальный опыт общесоюзной кампании «АСУ-низация всей страны» в 1970—1980 годы, когда директор завода или предприятия ставил у себя ЭВМ и рапортовал о внедрении автоматизированной системы управления производством, хотя как раз последнее на 90 % было блефом.

Сейчас в мире существует многомиллиардный рынок прикладных программных пакетов для математического моделирования самых разных процессов и явлений. Как правило, такие ППП разрабатываются коллективами из многих десятков человек в течение нескольких десятилетий. Широко известны, например, коммерческие продукты MSC NASTRAN, ANSYS, COMSOL, которые успешно используются в задачах гидро-газодинамики, упруго-пластичности, электрофизики и т.д. Как правило, идеологически — это закрытые инновации, почти не приспособленные к модификации пользователями, желающими их адаптировать к своим конкретным условиям. Кроме того, в Интернете имеется большое количество общедоступных программных библиотек и вычислительно-информационных инструментариев, решающих некоторые необходимые, но частные проблемы моделирования. Грубо говоря, имеется огромный программный «зоопарк», и конечному пользователю надо потратить слишком много времени для изучения объёмных технических материалов, что практически невозможно для эксперта в предметной профессиональной области.

Ситуация усугубляется, когда речь идет о необходимости эффективного распараллеливания алгоритмов на современных MBC сложнейшей архитектуры. С одной стороны, это неизбежно повышает требования к уровню знаний пользователя, но главное — разработчикам фактически приходится создавать новое поколение прикладного математического и программного обеспечения, рассчитанного на эффективное использование миллионов вычислительных процессоров и ядер. Именно данная проблема ставится во главу угла в дорожной карте IESP — Международного проекта экзамасштабного программирования, возглавляемого общепризнанным мировым авторитетом Джеком Донгаррой. Сейчас эта деятельность расширяется за счёт EESI — Европейской экзапроблемной инициативы, а также подключения национальных экзамасштабных программ Японии и Китая. На сегодняшний день проведено около десятка рабочих совещаний (Workshops) мировых экспертов. В Интернете сложно найти результаты обсуждений ими актуальных вопросов создания удобного программного окружения для пета- и экза-систем, развития модели кооперации вычислительного сообщества, обучения новым парадигмам программирования и т.д.

В Российской Федерации наибольшую организационную активность осуществляет возглавляемый академиком В.А. Садовничим Суперкомпьютерный консорциум университетов России, основанный в 2008 г. и включающий около 60 постоянных и ассоциированных членов, среди которых ведущие университеты РФ, институты РАН, а также коммерческие компании, активно работающие в области суперкомпьютерных технологий.

По инициативе Консорциума проводятся научные конференции, молодёжные школы и рабочие совещания, выпускаются книги и аналитические обзоры, а в 2010—2012 гг. был осуществлён проект «Суперкомпьютерное образование» Комиссии при Президенте РФ по модернизации и технологическому развитию экономики России. В рамках этого проекта создана сеть научно-образовательных центров суперкомпьютерных технологий (НОЦ СКТ) на базе восьми университетов по всем федеральным округам. Разработан Свод знаний и умений (компетенций), а также сформированы учебные курсы, программы повышения квалификации и проведены многочисленные образовательные мероприятия с вовлечением средств массовой информации.

Надо сказать, что эти безусловно полезные мероприятия оставляют ряд вопросов. Почему в сеть НОЦ СКТ не включен Новосибирск — столица Сибирского отделения РАН

и колыбель отечественной информатики? Где «потерялась» научная фундаментальная составляющая высокопроизводительных вычислений? Что будет после завершения разового трёхлетнего проекта Комиссии? И где вообще активная организующая роль Министерства образования и науки РФ в плане создания российского экзафлопника, о котором сообщалось в СМИ ещё пару лет назад?

Конечно, проблема кадров в области высокопроизводительных вычислений является критической, поскольку цивилизованное сообщество оказалось не готово к техническому прорыву в производстве суперкомпьютеров. Для преодоления этой болезни роста нужны не только адекватные фундаментальные достижения в математике и информатике, но и целенаправленные скоординированные усилия в создании новых массовых профессий — как разработчиков математического и прикладного программного обеспечения, так и его квалифицированных пользователей, включая вооружённых современными информационными технологиями физиков, химиков, биологов, машиностроителей и т.д.

Четыре

ипостаси математики

Главная особенность развития высокопроизводительных вычислений (по-английски HPC — High Performance Computing) заключается в том, что человечество получило доселе невиданные возможности решения сверхзадач — прямых и обратных, междисциплинарных, с реальной адекватностью моделей и высоким разрешением результатов, с обработкой огромных объёмов данных, — которые открывают новые пути к прорывным фундаментальным знаниям и технологическим инновациям.

В этом историческом процессе ключевую роль играет математизация всех научно-производственных сфер, причём в большинстве случаев технологический прогресс в какой-то отрасли тесно связан именно с тем, насколько успешно здесь применяются компьютерные расчёты.

Понятно, что сегодня роли и технологии внедрения математики должны быть переосмыслены. В США именно с этой целью создана Рабочая группа по «эксаматематике» (EMWG — Exascale Mathematics Working Group). Перед входящими в неё экспертами поставлены следующие задачи:

- проанализировать потенциальные расхождения в понимании миссии прикладной математики для экзавычислений;
- идентифицировать новые алгоритмические подходы, которые ориентированы на экзамасштабные вызовы;
- осмыслить математические возможности для решения новых фундаментальных проблем, посредством интегрирования различных математических дисциплин;
- сформировать целостную кооперативную концепцию эксаматематических исследований, которая включала бы диалог учёных из прикладных областей с учёными-вычислителями.

Чтобы вникнуть в суть поставленных вопросов, мы представим четыре ипостаси математики, которая, как известно, является и царицей наук, и слугой всех наук.

Начнём рассмотрение с теоретической математики, которая закладывает для нас основания о сути изучаемых проблем. Речь идёт о формировании, изучении и обосновании математической модели процесса или явления: естественного происхождения, научно-технического или социального. Надо сказать, что за последние десятилетия произошло идеологическое сближение теоретических физиков и математиков, и здесь на стыке наук получены замечательные результаты в области топологических методов, гомотопий, внешней алгебры и дифференциальных форм, гамильтонова формализма, прикладной теории групп, дифференциально-геометрического исчисления и т.д. Эти достижения только относительно недавно начали внедряться в приложения, и их широкое воплощение в алгоритмы как раз совпадает исторически с переходом на пета- и экза-компьютеры.