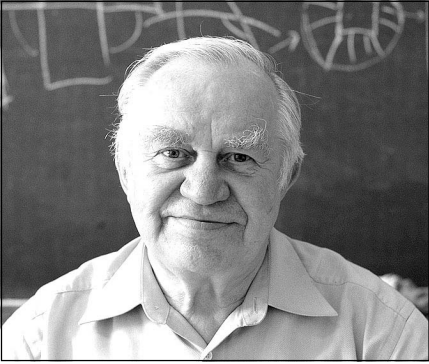


БЕСЕДЫ О НАУКЕ



В.П.Ильин, д.ф.-м.н.

Стратегии и тактики экстремального параллелизма

В 2008 году просвещённое человечество вступило в новую эру — эпоху петафлопных вычислений. Именно тогда в США был запущен первый в мире суперкомпьютер петафлопной (от английского flops — аббревиатура от flow point per second) производительности, что означает 10^{15} арифметических операций с плавающей запятой (при стандартном 32-битовом представлении вещественного числа) в секунду. Для сравнения — это примерно соответствует суммарному быстродействию одного миллиона обычного однопроцессорного персонального компьютера со скоростью около одного гигафлопа (10^9 флопс).

Надо сказать, что появление «петафлопника» было предсказано с высокой точностью ещё 20 лет назад, и тогда это казалось чем-то фантастическим, поскольку составляло суммарную мощность всего компьютерного парка США.

В соответствии с законом Мура (одного из основателей компании Intel), который достаточно строго выполняется около 50 лет, между появлением первого в мире «гигафлопника» и «терафлопника» (1000 гигафлоп) прошло 11 лет, между появлением «терафлопника» и «петафлопника» — тоже 11 лет. Так вот — в 2019 году грядёт пришествие эксафлопного компьютера с быстродействием 10^{18} флопс. И это уже совсем новый качественный уровень, который направляется называть компьютерной сингулярностью в истории человечества.

На пороге новой эры

Согласно текущим технологическим тенденциям, «эксафлопник» — это многопроцессорная компьютерная система (МКС) с сотнями миллионов вычислительных узлов и миллиардами ядер (арифметические устройства без большой собственной памяти), с рабочими площадями в несколько футбольных полей и с энергопотреблением в десятки мегаватт.

Важно отметить, что в последние два-три года темпы развития компьютеризации отнюдь не снижаются. Список Топ-500 мощнейших суперкомпьютеров мира на сентябрь 2011 г. уже насчитывает 10 петафлопников, причем на первые два места вышли японский и китайский МКС с быстродействием более 8 и 3 петафлопс соответственно. Отрадно отметить, что в клуб суперкомпьютеров держат уже вошла и Россия: отечественный петафлопник работает в Российском федеральном ядерном центре (ВНИИЭФ, г. Саров). Согласно интернетовскому порталу NEWSLAND, имеются планы запуска российского эксафлопника в 2020 году.

Признанный лидер мирового вычислительно сообществом Джека Донгарра (Окридская лаборатория, США) в 2008 г. организовал международную программу IESP (International Exascale Software Project, www.exascale.org), возглавляемую комитетом, в который вошли более 50 ведущих специалистов из разных стран. Под эгидой IESP проводятся регулярные рабочие совещания, на которых выработан и непрерывно развивается объёмный коллективный документ — Roadmap (дорожная карта), что-то вроде Библии, излагающий насущные компьютерные проблемы и пути их решения. Суть излагаемых скрижалей сводится к следующему.

Масштабы грядущего параллелизма МКС означают переход количества в качество и требуют кардинальной смены парадигм и вычислительно-информационных технологий.

Хотя сейчас алгоритмы реализуются на десятках, сотнях и тысячах процессоров, существующее программное обеспечение своими корнями остаётся однопроцессорным и использующим искусственные приёмы распараллеливания вычислений.

Именно сейчас мировое сообщество столкнулось с глобальной проблемой смены всего программного оснащения компьютеров: операционных систем, компиляторов, инструментов и приложений. Эта задача по своей грандиозности не под силу одному государству и срочно требует всеобщей координации и интеграции работ.

На основе данных положений формулируются концепции будущего «софта» (общепринятый русскоязычный жаргон от короткого английского software), главная из которых состоит в том, что все программы общего назначения должны быть открытыми и доступными (Open Source). Это не исключает возможность существования «закрытых» продуктов и права на интеллектуальную собственность, но главным теперь становится продажа не программного кода, а услуг по его использованию (термин SaS — Software as Service).

Новые условия меняют и принципы использования суперкомпьютерных мощностей. Эксафлопник должен стать базой Центра обработки данных (ЦОД), или Data Center, являющимся надёжным хранилищем огромных объёмов информации, в том числе всех необходимых системных и прикладных программ, и обеспечивающим в удалённом доступе необходимые пользовательские расчёты с предоставлением всех заказываемых ресурсов МКС. Такой режим эксплуатации получил название «облачные вычисления» (Cloud Computing) и направлен фактически на замену традиционных собственных вычислительных центров в небольших или средних организациях.

Следует сказать, что эти технологии в большей степени обсуждаются методически, а на практике могут сильно эволюционировать и привести к чему-то совсем новому как по форме, так и по содержанию. Организация работ вычислительных центров коллективного пользования является насыщенной многие десятилетия, а проходящие сейчас научно-технические дискуссии по этим вопросам подтверждают известный тезис «новое — это хорошо забытое старое».

Хотя стоимость вычислительных работ в абсолютном выражении стремительно дешевеет, сама разработка и эксплуатация суперкомпьютеров — достаточно дорогая производственная сфера. Чтобы по крайней мере окупить такие суперрасходы, пета- и эксафлопные МКС должны быть загружены расчётами актуальных приложений прорывного характера, так что концепция каждого типового ЦОДа — это мегапроект национального уровня.

Стремительный прогресс наукоемких технологий действительно выдвигает новые суперзадачи, которые могут быть решены только на суперкомпьютерах: био- и нанотехнологии, прогноз погоды, климата и экологических процессов, физика высоких энергий, финансовые задачи и т.д. Математическое моделирование становится третьим путём познания, играя роль связующего звена между теоретическими и экспериментальными исследованиями. И современные вычислительно-информационные технологии дают такие немалые возможности по проникновению в тайны природы, по оптимизации промышленных производств, что даже ставят новые философско-методологические проблемы о когнитивном потенциале человеческого общества.

Однако «дьявол таится в деталях», и надо пройти ещё длинный путь, чтобы построить виртуальные суперкомпьютерные миры, которые могли бы стать главным и повседневным орудием инженера, учёного, топ-менеджера, врача, госчиновника и любого профессионала в своей созидательной деятельности.

Если «нулевые» годы XXI века можно ассоциировать с победным шествием никем не прогнозируемых заранее Интернета и мобильной связи, в корне изменивших информационную сферу общества, то следующее десятилетие нас ожидает вторжение вычислительно-информационного моделирования в масштабах, последствия которых также непредсказуемы.

Результаты этих инноваций, конечно, будут зависеть в первую очередь от действий лидеров и ведущих экспертов компьютерного сообщества. Стоящие перед ними проблемы можно разбить на три основные части: brainware (в терминологии академика А.А. Дородницына, много лет возглавлявшего Вычислительный центр РАН) — это алгоритмы и совокупный математический интеллект, hardware (архитектура «железного» оборудования МКС, фантастический рост которых и служит главной движущей силой) и software — программное обеспечение, являющееся общепризнанным «слабым звеном» в данной триаде по росту производительности труда.

Возникшие «эксапроблемы» математического моделирования можно коротко сформулировать следующим образом: как решать сверхзадачи на супер-МКС пост-петафлопной производительности? Здесь надо пройтись по всей классической технологической цепочке — модели, алгоритмы, программы, расчёты, анализ, — и в первую очередь с точки зрения полномасштабного распараллеливания, которое как раз и составляет вызов нового поколения.

Задачи, модели, алгоритмы

Вычислительная математика после появления ЭВМ была обречена на бурное развитие, ознаменовавшееся созданием универсальных подходов к решению всех основных уравнений матфизики, что сейчас более полно называется задачами математического моделирования.

Несмотря на неисчерпаемое многообразие окружающего нас мира, все явления и процессы описываются достаточно ограниченным набором математических моделей. Можно перечислить системы дифференциальных уравнений: Максвелла — для электромагнетизма, Ламе — для упруго-пластичности, Навье–Стокса — для гидро-газодинамики, Шредингера — для квантовой физики, уравнения химической кинетики и др., — пальцев двух рук вполне хватит. Конечно, задачи могут формулироваться и посредством интегральных уравнений, и с помощью обобщенных вариационных постановок. Однако элементарных дифференциальных операторов, к счастью, совсем немного: градиент, ротор, дивергенция, внутренние и внешние производные, — а на основе их комбинирования можно представить всевозможные природные, технические, социальные, финансовые и даже общественные явления и процессы.

Каждая из таких задач может иметь свои сложности, обусловленные или пространственной сингулярностью в окрестности особых точек, или нелинейными динамическими эффектами типа предельных циклов, бифуркаций и странных аттракторов. Без понимания этих важных специальных свойств — а это прерогатива теоретиков — невозможно построить высокоточный численный метод, надёжно рассчитывающий практические ситуации с такой сложной спецификой.

Наращивание компьютерных мощностей в последние два-три десятилетия позволило серьёзно подступить к решению междисциплинарных задач, которые связаны с моделированием взаимосвязанных процессов различной природы и описываются большими системами дифференциальных и/или интегральных уравнений со многими неизвестными функциями. Хорошей иллюстрацией может служить проблема комплексного моделирования технологических процессов в электрометаллургии.

Например, алюминиевый электролизёр представляет собой сложнейшее техническое устройство с сотнями разномасштабных деталей из различных материалов, в котором надо рассчитывать огромные токи и электромагнитные поля, процессы теплопереноса с фазовыми переходами, прочность конструкций, циркуляция жидкого металла и электролита, т.е. решать совместно почти весь джентльменский набор уравнений матфизики. Сюда можно добавить и такое обстоятельство, что в одном заводском цехе находится около сотни электролизёров, которые последовательно соединены электрическими шинами, а также взаимодействуют между собой посредством магнитных полей. В общем, модель «виртуальный электролизёр» или «виртуальный цех» вполне подходит под статус суперзадачи для суперкомпьютера.

Однако этим потребная ресурсоёмкость ЭВМ далеко не исчерпывается. Рассмотренные до сих пор формулировки относятся к классу прямых задач, в которых решение вычисляется при полностью заданных граничных и начальных условиях. В жизни наиболее важно решать обратные задачи, включающие формальные параметры, которые надо оптимизировать по условию минимизации некоторого целевого функционала при

известных дополнительных линейных и/или нелинейных ограничениях на входные данные. Например, в описанном электролизёре параметры и ограничения могут быть связаны с его геометрическими или физическими характеристиками, а оптимизируемый показатель — «выход металла по току», т.е. вес производимого алюминия на единицу затраченной электроэнергии.

Решение обратной задачи методами оптимизации производится с помощью направленного перебора прямых задач, количество которых в «тяжёлых» случаях может измеряться сотнями и тысячами. Ситуация значительно усложняется при «многообразной» зависимости функционала от параметров, а особенно — при наличии большого количества локальных минимумов и необходимости поиска глобального минимума. Практический результат решения обратной задачи в данном случае — это оптимизация эксплуатационного режима электролизера, уменьшение количества аварий в цехе и, наконец, увеличение производственной прибыли.

Другой тип обратной задачи — это идентификация неизвестных параметров физико-математической модели на основе сопоставления результатов расчётов и натуральных измерений. Актуальный междисциплинарный пример можно привести из геофизики — это комплексная разведка полезных ископаемых с использованием сейсмических исследований и вибропросвечивания Земли, электромагнитных полей от разнообразных источников, магнитотеллурического зондирования и гравиметрических технологий. Результат такого крупномасштабного эксперимента — надёжное прогнозирование залежей полезных ископаемых, в частности углеводородов.

Список таких компьютерных сверхпроектов можно продолжить. В ближайшие годы следует ожидать создание виртуальной модели живой клетки, что представляет собой очень сложную проблему, но обещает сделать новый прорыв в биологии и фундаментальной медицине. Одна из продвинутых отраслей — авиастроение. «Виртуальный самолёт» уже помогает значительно сокращать многолетние аэродинамические продувки моделей, неизбежные прочностные и лётные испытания. И конечно, нельзя не упомянуть о моделировании ядерного оружия, которое наверняка активно развивается в странах «атомного клуба» при строжайшем моратории на любые натурные эксперименты.

Если теперь обратиться к численным методам, то здесь надо отметить, на первый взгляд, неожиданную тенденцию последних десятилетий — активное проникновение инноваций из теоретической математики. Такие, казалось бы, заоблачные абстракции, как теория групп, уравнения на многообразиях, внешние произведения и дифференциальные формы, гамильтоновы формализм, теория хаоса становятся базой для конструирования алгоритмов нового поколения. Отсюда следует необходимость создания качественно нового базового математического и прикладного программного обеспечения, способного к перманентному развитию, легко адаптируемого к новым платформам динамически развивающихся компьютерных систем и ориентированного на длительный жизненный цикл.

Разумеется, эффективность алгоритма в пост-петафлопную эру не котируется без такого качества, как высокопроизводительные вычисления, что однозначно воспринимается как масштабируемый параллелизм. Последний термин можно понимать в том смысле, что время решения задачи на P процессорах в P раз меньше, чем на одном процессоре. Другое возможное определение (не всегда эквивалентное): если общий объём арифметических операций для решения задачи увеличивается в P раз и число арифметических устройств (AU) МКС тоже увеличивается в P раз, то время расчёта остаётся неизменным. Строго говоря, эти качественные суждения надо уточнять, но главное уже