ГЛАВА 1
ОБЗОР АРХИТЕКТУРЫ СОВРЕМЕННЫХ МНОГОЯДЕРНЫХ ПРОЦЕССОРОВ

«Citius, Altius, Fortius» [1] – девиз Олимпийских игр современности, как ни к какой другой области, применим к вычислительной технике. Воплощение в жизнь не раз видоизменявшего свою исходную формулировку, но до сих пор действующего эмпирического закона, сформулированного в 1965 г. Гордоном Муром, похоже, стало «делом чести» производителей аппаратного обеспечения. Из всех известных формулировок этого закона точку зрения потребителя/пользователя наилучшим образом отражает вариант: «производительность вычислительных систем удваивается каждые 18 месяцев». Мы сознательно не использовали термин «процессор», поскольку конечного пользователя вовсе не интересует, кто обеспечивает ему повышение мощности: процессор, ускоритель, видеокарта – ему важен лишь сам факт роста возможностей «за те же деньги».

Правда, в последние несколько лет возможности увеличения мощности процессоров на основе повышения тактовой частоты оказались фактически исчерпаны, и производители, выбрав в качестве магистрального пути развития увеличение числа ядер на кристалле, были вынуждены призвать на помощь разработчиков программного обеспечения. Старые последовательные программы, способные использовать лишь одно ядро, теперь уже не будут работать быстрее на новом поколении процессоров «задаром» – требуется практически повсеместное внедрение программирования параллельного.

Кроме представленной выше, известна и другая формулировка закона Мура: «доступная (человечеству) вычислительная мощность удваивается каждые 18 месяцев». Зримое свидетельство этого варианта формулировки – список Top500 [106] самых высокопроизводительных вычислительных систем мира, обновляемый дважды в год. В 31-м списке Top500 (июнь 2008 г.) впервые в истории был преодолен петафлопный порог производительности – суперкомпьютер Roadrunner [112] производства компании IBM показал на тесте LINPACK 1,026 петафлопс (предыдущий «психологический» барьер в один терафлопс был преодолен системой ASCI Red [111] производства компании Intel в 1997 г.; как видим, всего за 11 лет пик мощности вырос на три порядка). А суммарная мощность систем, представленных в 31-м списке Top500, составила 11,7 петафлопс. Много это или мало? Если взять за основу, что реальная производительность хорошей «персо-

[1] «Быстрее, Выше, Сильнее» – лат.

налки» на четырехъядерном процессоре составляет порядка 20 гигафлопс, то весь список Top500 будет эквивалентен половине миллиона таких персоналок. Очевидно, что это лишь вершина айсберга. По данным аналитической компании Gartner общее число используемых в мире компьютеров превысило в 2008 г. 1 миллиард.

Представленные в списке Top500 данные позволяют проследить характерные тенденции развития индустрии в сфере суперкомпьютерных вычислений. Первый список Top500 датирован июнем 1993 г. и содержал 249 многопроцессорных систем с общей памятью и 97 суперкомпьютеров, построенных на основе единственного процессора; более 40% всех решений в нем были созданы на платформе, разработанной компанией Cray. Уже четырьмя годами позже в Top500 не осталось ни одного суперкомпьютера на основе единственного процессора, а взамен появилась первая система с производительностью всего в 10 гигафлопс (в 100 раз меньше, чем у лидера списка системы ASCI Red), относящаяся к довольно новому тогда виду кластерных вычислительных систем, которые сегодня занимают в Top500 80% списка и являются фактически основным способом построения суперкомпьютеров.

Основным преимуществом кластеров, предопределившим их повсеместное распространение, было и остается построение из стандартных массово выпускающихся компонент, как аппаратных, так и программных. Сегодня 75% систем из списка построены на основе процессоров компании Intel, чуть больше 13% – на процессорах компании IBM и 11% – компании AMD (на двух оставшихся производителей – NEC и Cray – приходится по одной системе соответственно); 81% систем используют всего два типа сетей передачи данных: Gigabit Ethernet или Infiniband; 85% систем работают под управлением операционной системы из семейства Linux. Как видим, разнообразием список не блещет, что является несомненным плюсом с точки зрения пользователей.

Однако для массового пользователя еще большим плюсом была бы возможность иметь персональный суперкомпьютер у себя на столе или, на худой конец, стоящий под столом. И кластеры, принесшие в индустрию высокопроизводительных вычислений идею «собери суперкомпьютер своими руками», как нельзя лучше отвечают этой потребности. Сейчас трудно достоверно установить, какая система может быть названа первым в мире «персональным кластером». Во всяком случае, уже в начале 2001 г. компания RenderCube [109] представила одноименный мини-кластер из 4-х двухпроцессорных систем, заключенных в кубический корпус со стороной всего в 42 см.

Тенденция «персонализации» супервычислений в последнее время развивается все активнее, и недавно была подхвачена в том числе и производителями видеокарт, мощности которых возросли настолько, что возникло естественное желание использовать их не только в графических рас-

четах, но и в качестве ускорителей вычислений общего назначения. Соответствующие решения представлены в настоящее время компанией NVIDIA (семейство NVIDIA® Tesla™) и компанией AMD (семейство ATI FireStream™), и демонстрируют – в силу специфики внутреннего устройства – потрясающую (в сравнении с универсальными процессорами) пиковую производительность, превышающую 1 терафлопс.

Данная глава посвящена рассмотрению современных многоядерных процессоров, которые являются основой для построения самых быстродействующих вычислительных систем. Для полноты картины приводится также описание ряда аппаратных устройств (видеокарт и вычислительных сопроцессоров), которые могут быть использованы для существенного ускорения вычислений. И для завершения рассматриваемой темы в конце данной главы дается краткая характеристика «персональных» мини-кластеров, которые позволяют при достаточно «экономных» финансовых затратах приступить к решению имеющихся вычислительно-трудоемких задач с использованием высокопроизводительных вычислительных систем.

1.1. Параллелизм как основа высокопроизводительных вычислений

Без каких-либо особых преувеличений можно заявить, что все развитие компьютерных систем происходило и происходит под девизом «Скорость и быстрота вычислений». Если быстродействие первой вычислительной машины ENIAC составляло всего несколько тысяч операций в секунду, то самый быстрый на данный момент времени суперкомпьютер RoadRunner может выполнять уже квадриллионы (10^{15}) команд. Темп развития вычислительной техники просто впечатляет – увеличение скорости вычислений в триллионы (10^{12}) раз немногим более чем за 60 лет! Для лучшего понимания необычности столь стремительного развития средств ВТ часто приводят яркие аналогии, например: если бы автомобильная промышленность развивалась с такой же динамикой, то сейчас автомобили весили бы порядка 200 грамм и тратили бы несколько литров бензина на миллионы километров!

История развития вычислительной техники представляет увлекательное описание замечательных научно-технических решений, радости побед и горечи поражений. Проблема создания высокопроизводительных вычислительных систем относится к числу наиболее сложных научно-технических задач современности и ее разрешение возможно только при всемерной концентрации усилий многих талантливых ученых и конструкторов, предполагает использование всех последних достижений науки и техники и требует значительных финансовых инвестиций. Здесь важно отметить, что при общем росте скорости вычислений в 10^{12} раз быстродействие самих технических средств вычислений увеличилось всего в несколько миллио-

нов раз. И дополнительный эффект достигнут за счет введения *параллелизма* буквально на всех стадиях и этапах вычислений.

Не ставя целью в рамках данной книги подробное рассмотрение истории развития компьютерного параллелизма, отметим, например, организацию независимости работы разных устройств ЭВМ (процессора и устройств ввода-вывода), появление многоуровневой памяти, совершенствование архитектуры процессоров (суперскалярность, конвейерность, динамическое планирование). Дополнительная информация по истории параллелизма может быть получена, например, в [67]; здесь же выделим, как принципиально важный итог – многие возможные пути совершенствования процессоров практически исчерпаны (так, возможность дальнейшего повышения тактовой частоты процессоров ограничивается рядом сложных технических проблем) и наиболее перспективное направление на данный момент времени состоит в явной организации многопроцессорности вычислительных устройств.

Ниже будут более подробно рассмотрены основные способы организации многопроцессорности – *симметричной мультипроцессорности* (*Symmetric Multiprocessor, SMP*), *одновременной многопотоковости* (*Simultaneous Multithreading, SMT*) и *многоядерности* (*multicore*).

1.1.1. Симметрическая мультипроцессорность

Организация симметричной мультипроцессорности (*Symmetric Multiprocessor, SMP*), когда в рамках одного вычислительного устройства имеется несколько полностью равноправных процессоров, является практически первым использованным подходом для обеспечения многопроцессорности – первые вычислительные системы такого типа стали появляться в середине 50-х – начале 60-х годов, однако массовое применение SMP систем началось только в середине 90-х годов.

Следует отметить, что SMP системы входят в группу MIMD (*multi instruction multi data*) вычислительных систем в соответствии с классификацией Флинна. Поскольку эта классификация приводит к тому, что практически все виды параллельных систем (несмотря на их существенную разнородность) относятся к одной группе MIMD, для дальнейшей детализации класса MIMD предложена практически общепризнанная структурная схема [47,99] – см. рис. 1.1. В рамках данной схемы дальнейшее разделение типов многопроцессорных систем основывается на используемых способах организации оперативной памяти в этих системах. Данный поход позволяет различать два важных типа многопроцессорных систем – *multiprocessors* (*мультипроцессоры* или системы с общей разделяемой памятью) и

multicomputers (*мультикомпьютеры* или системы с распределенной памятью).

Для дальнейшей систематики мультипроцессоров учитывается способ построения общей памяти. Возможный подход – использование единой (централизованной) *общей памяти* (*shared memory*) – см. рис. 1.2. Такой

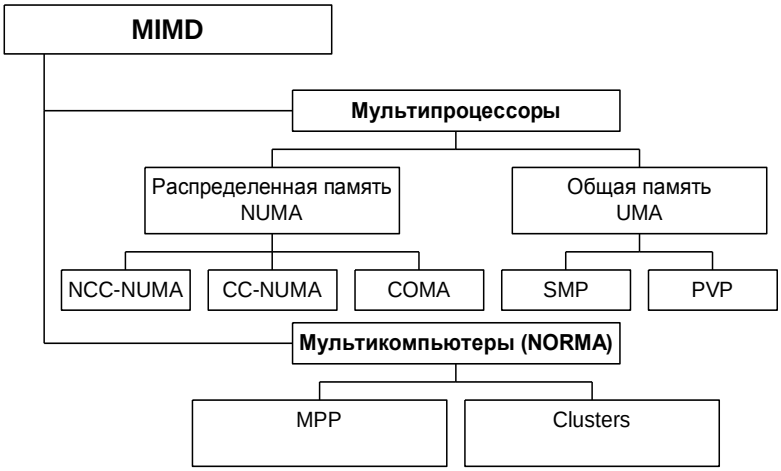


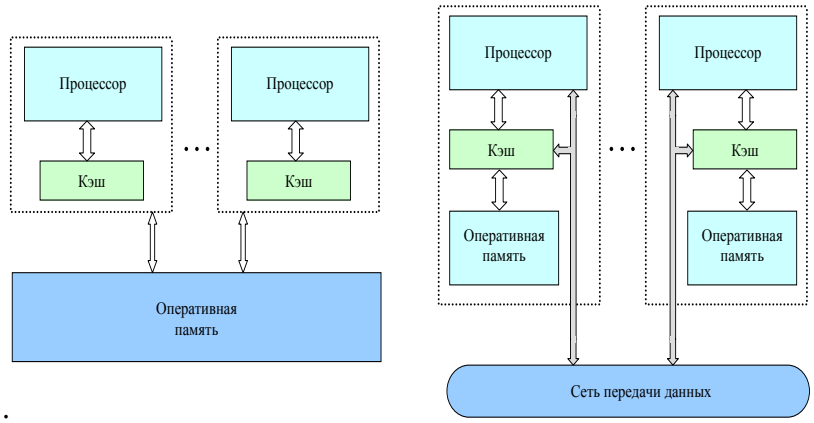
Рис. 1.1. Классификация многопроцессорных вычислительных систем

подход обеспечивает *однородный доступ к памяти* (*uniform memory access* or *UMA*) и служит основой для построения *векторных параллельных процессоров* (*parallel vector processor* или *PVP*) и *симметричных мультипроцессоров* (*symmetric multiprocessor* или *SMP*). Среди примеров первой группы – суперкомпьютер Cray T90, ко второй группе относятся IBM eServer, Sun StarFire, HP Superdome, SGI Origin и др.

Одной из основных проблем, которые возникают при организации параллельных вычислений на такого типа системах, является доступ с разных процессоров к общим данным и обеспечение, в этой связи, *однозначности (когерентности) содержимого разных кэшей* (*cache coherence problem*).

Дело в том, что при наличии общих данных копии значений одних и тех же переменных могут оказаться в кэшах разных процессоров. Если в такой ситуации (при наличии копий общих данных) один из процессоров выполнит изменение значения разделяемой переменной, то значения копий в кэшах других процессорах окажутся не соответствующими действительности и их использование приведет к некорректности вычислений. Обеспечение однозначности кэшей обычно реализуется на аппаратном уровне – для этого после изменения значения общей переменной все копии этой переменной в кэшах отмечаются как недействительные и последующий доступ к переменной потребует обязательного обращения к основной памяти.

Следует отметить, что необходимость обеспечения когерентности приводит к некоторому снижению скорости вычислений и затрудняет создание систем с достаточно большим количеством процессоров.



а) б)

Рис. 1.2. Архитектура многопроцессорных систем с общей (разделяемой) памятью: системы с (а) однородным и (б) неоднородным доступом к памяти

Наличие общих данных при выполнении параллельных вычислений приводит к необходимости *синхронизации взаимодействия* одновременно выполняемых потоков команд. Так, если изменение общих данных требует для своего выполнения некоторой последовательности действий, то необходимо обеспечить *взаимоисключение* (*mutual exclusion*) с тем, чтобы эти изменения в любой момент времени мог выполнять только один командный поток. Задачи взаимоисключения и синхронизации относятся к числу классических проблем, и их рассмотрение при разработке параллельных программ является одним из основных вопросов параллельного программирования.

Общий доступ к данным может быть обеспечен и при физически распределенной памяти (при этом, естественно, длительность доступа уже не будет одинаковой для всех элементов памяти) – см. рис. 1.2. Такой подход именуется как *неоднородный доступ к памяти* (*non-uniform memory access* или *NUMA*). Среди систем с таким типом памяти выделяют:

- Системы, в которых для представления данных используется только локальная кэш-память имеющихся процессоров (*cache-only memory architecture* или *COMA*); примерами таких систем являются KSR-1 и DDM.
- Системы, в которых обеспечивается когерентность локальных кэшей разных процессоров (*cache-coherent NUMA* или *CC-NUMA*); среди сис-

тем данного типа – SGI Origin 2000, Sun HPC 10000, IBM/Sequent NUMA-Q 2000.

- Системы, в которых обеспечивается общий доступ к локальной памяти разных процессоров без поддержки на аппаратном уровне когерентности кэша (*non-cache coherent NUMA* или *NCC-NUMA*); к данному типу относится, например, система Cray T3E.

Использование распределенной общей памяти (*distributed shared memory* или *DSM*) упрощает проблемы создания мультипроцессоров (известны примеры систем с несколькими тысячами процессоров), однако возникающие при этом проблемы эффективного использования распределенной памяти (время доступа к локальной и удаленной памяти может различаться на несколько порядков) приводят к существенному повышению сложности параллельного программирования.

1.1.2. Одновременная многопотоковость

Организация симметричной мультипроцессорности позволяет достаточно легко увеличивать производительность вычислительных устройств. Однако такое решение при увеличении числа процессоров обладает плохой масштабируемостью из-за проблем с обеспечением когерентности кэш-памяти разных процессоров – SMP системы содержат, как правило, 2 или 4, реже – 8, и совсем редко – большее количество процессоров. Кроме того, такой подход является сравнительно дорогим решением.

С другой стороны, проанализировав эффективность современных сложных процессоров, насчитывающих в своем составе десятки и сотни миллионов транзисторов, можно обратить внимание на то, что при выполнении большинства операций оказываются полностью задействованными не все составные компоненты процессоров (по имеющимся оценкам средняя загрузка процессора составляет всего лишь около 30%). Так, если в данный момент времени выполняется операция целочисленной арифметики, то блок процессора для выполнения вещественных операций окажется простаивающим. Для повышения загрузки процессора можно организовать *спекулятивное* (*опережающее*) *исполнение* операций, однако воплощение такого подхода требует существенного усложнения логики аппаратной реализации процессора. Гораздо проще было бы, если в программе заранее были бы предусмотрены последовательности команд (*потоки*), которые могли быть выполнены параллельно и независимо друг от друга. Такой подход тем более является целесообразным, поскольку поддержка многопоточного исполнения может быть обеспечена и на аппаратном уровне за счет соответствующего расширения возможностей процессора (и такая доработка является сравнительно простой).

Данная идея поддержки *одновременной многопоточности* (*simultaneous multithreading, SMT*) была предложена в 1995 г. в университете

Вашингтона Дином Тулсеном (Dean Tullsen) и позднее активно развита компанией Интел под названием технологии *гиперпоточности* (*hyper threading*, *HT*). В рамках такого подхода процессор дополняется средствами запоминания состояния потоков, схемами контроля одновременного выполнения нескольких потоков и т. д. За счет этих дополнительных средств на активной стадии выполнении может находиться несколько потоков; при этом одновременно выполняемые потоки конкурируют за исполнительные блоки единственного процессора и, как результат, выполнение отдельных потоков может блокироваться, если требуемые в данный момент времени блоки процессора оказываются уже задействованными. Как правило, число аппаратно-поддерживаемых потоков равно 2, в более редких случаях этот показатель достигает 4 и даже 8. Важно при этом подчеркнуть, что аппаратно-поддерживаемые потоки на логическом уровне операционных систем Linux и Windows воспринимаются как отдельные процессоры, т. е., например, единственный процессор с двумя аппаратно-поддерживаемыми потоками в менеджере Task Manager операционной системы Windows диагностируется как два отдельных процессора.

Использование процессоров с поддержкой многопотоковости может приводить к существенному ускорению вычислений (важно отметить – при надлежащей реализации программ). Так, имеется большое количество демонстраций, показывающих, что на процессорах компании Интел с поддержкой технологии гиперпоточности достигается повышение скорости вычислений около 30%.

1.1.3. Многоядерность

Технология одновременной многопоточности позволяет достичь многопроцессорности на логическом уровне. Еще раз отметим, что затраты на поддержку такой технологии являются сравнительно небольшими, но и получаемый результат достаточно далек от максимально-возможного – ускорение вычислений от использования многопоточности оказывается равным примерно 30%. Дальнейшее повышение быстродействия вычислений при таком подходе по-прежнему лежит на пути совершенствования процессора, что – как было отмечено ранее – требует решения ряда сложных технологических проблем.

Возможное продвижение по направлению к большей вычислительной производительности может быть обеспечено на основе «парадоксального», на первый взгляд, решения – возврат к более «простым» процессорам с более низкой тактовой частотой и с менее сложной логикой реализации! Такой неординарный ход приводит к тому, что процессоры становятся менее энергоемкими[2)], более простыми для изготовления и, как результат, более

[2] Проблема энергопотребления является одной из наиболее сложных для процессоров с высокой тактовой частотой

надежными. А также – что является чрезвычайно важным – «простые» процессоры требуют для своего изготовления меньшее количество логических схем, что приводит к освобождению – в рамках кремниевого кристалла, используемого для изготовления процессоров – большого количества свободных транзисторов. Эти свободные транзисторы, в свою очередь, могут быть использованы для реализации дополнительных вычислительных устройств, которые могут быть добавлены к процессору. Фактически, данный подход позволяет реализовать в единственном кремниевом кристалле несколько *вычислительных ядер* в составе одного многоядерного процессора, при этом по своим вычислительным возможностям эти ядра могут не уступать обычным (одноядерным) процессорам. Поясним сказанное на примере рис. 1.3. В центре рисунка (рис. 1.3б) приведены показатели исходного процессора, принятые за 1 для последующего сравнения.

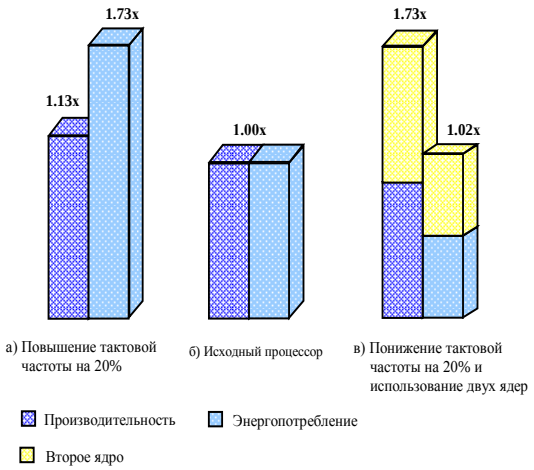


Рис. 1.3. Демонстрация зависимости между тактовой частотой, энергопотреблением и производительностью процессора

Пусть для повышения быстродействия процессора его тактовая частота увеличена на 20% (см. рис. 1.3а), тогда производительность процессора увеличится – однако не на 20%, а, например, на 13% (приводимые здесь числовые значения имеют качественный характер), в то время как энергопотребление возрастет существенно – в приведенном примере на 73%. Данный пример является, на самом деле, очень характерным – увеличение тактовой частоты процессора приводит в большинстве случаев к значительному росту энергопотребления. Теперь уменьшим тактовую частоту процессора – опять же на 20% (см. рис. 1.3в). В результате снижения тактовой частоты производительность процессора, конечно, уменьшится, но, опять же, не на 20%, а примерно на ту же величину 13% (т. е. станет равной 87% от производительности исходного процессора). И опять же, энер-

гопотребление процессора уменьшится, причем достаточно значительно – до уровня порядка 51% энергопотребления исходного процессора. И тогда, добавив в процессор второе вычислительное ядро за счет появившихся свободных транзисторов, мы можем довести суммарные показатели процессора по энергопотреблению до уровня 1.02 энергопотребления исходного процессора, а производительность – до уровня 1.73!!!

На логическом уровне архитектура многоядерного процессора соответствует практически архитектуре симметричного мультипроцессора (рис. 1.2 и 1.4). На рис. 1.4 приведена возможная архитектура двуядерного процессора – различия для разных многоядерных процессоров могут состоять в количестве имеющихся ядер и в способах использования кэш-памяти ядрами процессора – кэш-память может быть как общей, так и распределенной для разных ядер. Так, на рис. 1.4 кэш-память первого уровня L1 локальна для каждого ядра, в то же время кэш-память всех последующих уровней и оперативная память является общей.

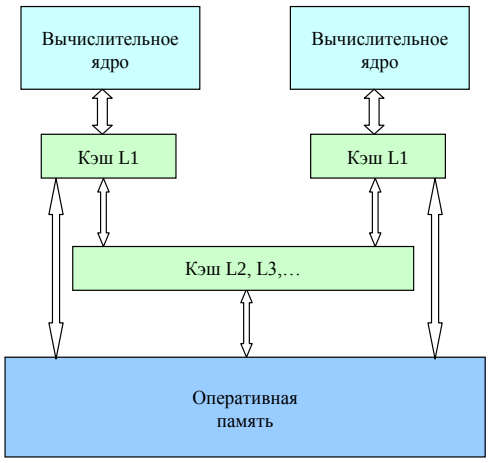


Рис. 1.4. Архитектура двухъядерного процессора

Как следует из проведенного рассмотрения, многоядерность позволяет повышать производительность процессоров и данный подход обладает целым рядом привлекательных моментов (уменьшение энергопотребления, снижение сложности логики процессоров и т. п.). Все сказанное приводит к тому, что многоядерность становится одной из основных направлений развития компьютерной техники.

Значимость такого подхода привела даже к тому, что известный закон Мура теперь формулируется в виде «Количество вычислительных ядер в процессоре будет удваиваться каждые 18 месяцев». В настоящее время для

массового использования доступны двух- и четырех- ядерные процессоры, компании-разработчики объявили о подготовке шести-ядерных процессоров. В научно-технической литературе наряду с рассмотрением обычных многоядерных (*multi-core*) процессоров начато широкое обсуждение процессоров с массовой многоядерностью (*many-core*), когда в составе процессоров будут находиться сотни и тысячи ядер!

И в заключение следует отметить еще один принципиальный момент – потенциал производительности многоядерных процессоров может быть задействован только при надлежащей разработке программного обеспечения – программы должны быть очень хорошо распараллелены. А, как известно, сложность разработки параллельных программ значительно превышает трудоемкость обычного последовательного программирования. Тем самым, проблема обеспечения высокопроизводительных вычислений перемещается теперь из области компьютерного оборудования в сферу параллельного программирования. И здесь нужны новые идеи и перспективные технологии для организации массового производства параллельных программ.

1.2. Многоядерность – два, четыре, восемь – кто больше?

Дополним теперь общее рассмотрение многоядерного направления развития компьютерной техники характеристикой ряда конкретных широко применяемых в настоящее время многоядерных процессоров основных компаний-разработчиков – Intel, AMD, IBM, Sun.

1.2.1. Процессоры Intel® Core™ и Intel® Xeon®

Как просто было когда-то сравнивать процессоры компании Intel между собой. Все знали: есть Pentium, есть его «урезанный» вариант Celeron, а в остальном, чем выше частота, тем лучше. Эта простота была следствием того факта, что в формуле, определяющей производительность вычислительной системы, «тактовая частота процессора × число инструкций, выполняемых за один такт (Instructions Per Cycle, IPC)» переменной величиной была только частота. Необходимо, конечно, отметить, что получаемая по этой формуле величина дает только так называемую «пиковую производительность», приблизиться к которой на практике можно лишь на отдельных специально подобранных задачах. Именно поэтому сравнение вычислительных систем, в том числе из списка Top500 и аналогичных им, выполняется на основе производительности, показанной на стандартном тесте, в качестве которого повсеместно используется LINPACK [110]. Как только наращивание тактовой частоты прекратилось, компании Intel понадобился другой способ описания и градации выпускаемых процессоров.

Сегодня все процессоры производства Intel делятся, прежде всего, по назначению: для настольных систем, ноутбуков, серверов и рабочих станций и т. д. Затем в каждом классе выделяют серии процессоров, отличающиеся между собой по некоторым ключевым характеристикам.

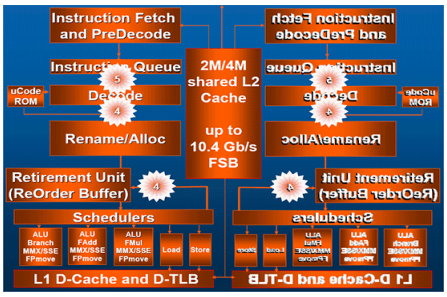


Рис. 1.5. Архитектура Intel® Core™

В классе настольных и мобильных систем сегодня «царствуют» представители семейства Intel® Core™2, в серверном сегменте – процессоры Intel® Xeon®, при этом и те, и другие построены на микроархитектуре Intel® Core™, пришедшей в 2006 г. на смену архитектуре Intel® NetBurst™.

Особенностям микроархитектуры Intel® Core™, позволившей компании Intel существенно потеснить своего основного конкурента – компанию AMD, посвящено множество материалов и публикаций. Не ставя перед собой задачу подробного ее обсуждения, кратко отметим лишь ключевые моменты, выделяемые самими разработчиками.

- Wide Dynamic Execution. Если основой повышения производительности процессоров архитектуры NetBurst была тактовая частота, то в архитектуре Core на первое место вышло число инструкций за такт (с учетом увеличения этого показателя за счет наращивания числа ядер): IPC каждого ядра в этой архитектуре равно 4, таким образом пиковая производительность четырехъядерных процессоров, например, равна «16 × на тактовую частоту».
- Advanced Smart Cache. Кэш второго уровня в архитектуре Core является общим на каждую пару ядер (четырехъядерные процессоры Intel сегодня фактически представляют собой два двухъядерных, размещенных на одном кристалле), что позволяет как динамически менять его «емкость» для каждого ядра из пары, так и использовать преимущества совместного использования ядрами данных, находящихся в кэше. Кроме того, в случае активного использования всего одного ядра, оно «задаром» получает кэш вдвое большего размера, чем было бы в случае отдельного кэша второго уровня на каждое ядро.
- Advanced Digital Media Boost. По сравнению с NetBurst в архитектуре Core была значительно улучшена работа с векторными расширениями SSE. С точки зрения конечного пользователя основным из этих улучшений, помимо добавления новых команд, стала способность процессоров выполнять SSE-инструкции за один такт вместо двух в NetBurst.

- Intelligent Power Capacity. Процессоры на архитектуре Core получили возможность как интерактивного отключения незадействованных в данный момент подсистем, так и «динамического» понижения частоты ядер, что дало возможность существенно снизить тепловыделение (Thermal Design Power, TDP), что особенно положительно сказалось на процессорах для настольных и мобильных систем. Так, двухъядерный Pentium D с частотой 2,8 ГГц имел TDP 130 Вт, тогда как четырехъядерный Core 2 Quad Q9300 с частотой 2,5 ГГц – всего 95 Вт.

Кроме того, необходимо отметить существенно уменьшившийся по сравнению с 31-стадийным в последних процессорах архитектуры NetBurst конвейер – его длина в архитектуре Core составляет 14 стадий, плюс «честную» 64-разрядность, плюс в очередной раз доработанное предсказание ветвлений, плюс многое, оставшееся за кадром…

Приведем технические данные текущих лидеров в классе настольных и серверных процессоров.

Процессор Intel® Core™2 Quad Q9650

Тактовая частота: 3 ГГц.

Число ядер: 4.

Кэш второго уровня: 12 Мб (по 6 Мб на каждую пару ядер).

Частота системной шины: 1333 МГц.

Технологический процесс: 45 нанометров.

Процессор Intel® Xeon® X7460

Тактовая частота: 2,66 ГГц.

Число ядер: 6.

Кэш второго уровня: 9 Мб (по 3 Мб на каждую пару ядер).

Кэш третьего уровня: 16 Мб.

Частота системной шины: 1066 МГц.

Технологический процесс: 45 нанометров.

В заключение отметим еще один весьма важный факт – помимо пиковой производительности той или иной архитектуры и соответственно процессоров, построенных на ее основе, значимым обстоятельством для конечного потребителя является процент мощности, который можно «отжать от пика». Для систем в Top500, построенных на процессорах компании Intel, этот показатель составляет в 31-м списке 61%, при этом «удельная мощность» в расчете на один процессор/ядро равна 6,23 гигафлопс (необходимо заметить, конечно, что значительная часть этих систем введена в строй уже несколько лет назад и построена не на новейших процессорах).

1.2.2. Процессоры AMD Phenom™ и AMD Opteron™

Компания AMD основана в 1969 г. (всего на год позже, чем Intel) и в сознании рядового пользователя прочно занимает место главного конку-

рента Intel на рынке процессоров для настольных систем и отчасти на рынке серверных, при этом практически всегда выступая в роли догоняющего. Если принимать во внимание только «внешние» факторы, вроде рыночной доли, то ситуация, действительно, может быть воспринята именно так. И в этом свете основной успех компании за последнее десятилетие связан с выпуском в 2003 г. 64-битных процессоров AMD Opteron™, быстро завоевавших популярность и позволивших AMD значительно упрочить свое положение, в том числе в сегменте высокопроизводительных решений. Достаточно отметить, что в 28-м списке Top500 (ноябрь 2006 г.) доля систем, построенных на основе процессоров AMD, достигла своего исторического максимума и составила 22,6%, против 52,6% у компании Intel и 18% у компании IBM. Однако кроме такого чисто количественного сравнения, в котором AMD неизменно проигрывает своим конкурентам, есть еще показатели качественные, и тут компания за прошедшие годы нередко бывала первопроходцем и реализовывала действительно интересные архитектурные решения.

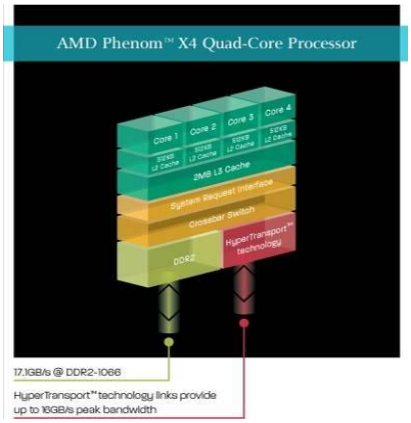


Рис. 1.6. Архитектура AMD Phenom™

Среди прочего это и интеграция в процессор северного моста, что дает более быстрый доступ к оперативной памяти, и использование Direct Connect Architecture для взаимодействия процессоров между собой посредством высокоскоростной шины HyperTransport™, позволяющей без существенных потерь в производительности объединять в рамках одной системы до 8 процессоров Opteron. Кроме того, нужно отметить, что в процессорах Opteron реализована «честная» четырехъядерность (Native Quad-Core Design), двухпотоковое управление 128-битными SSE-инструкциями, выполнение до четырех операций с плавающей точкой двойной точности за такт, расширенная технология оптимизации энергопотребления (Enhanced AMD PowerNow!) и многое другое. Последними серверными процессорами компании AMD являются четырехъядерные модели Opteron 3G на ядре Barcelona.

На рынке настольных систем основное оружие компании AMD сегодня – процессоры AMD Phenom™ (рис. 1.6). Процессоры Phenom построены на той же микроархитектуре (AMD K10), что и серверные Opteron. Помимо уже отмеченных особенностей, можно упомянуть наличие в процессорах Phenom кэша третьего уровня, пиковую пропускную способность шины HyperTransport до 16 Гб/с, поддержку 128-битных операций SSE, работу кэша второго уровня на частоте ядра, технологию улучшенной защиты от вирусов (NX бит / Enhanced Virus Protection). Текущее поколение процессоров Phenom выпускается по технологии 65 нм.

Приведем технические данные текущих лидеров в классе настольных и серверных процессоров.

Процессор AMD Phenom™ X4 Quad-Core 9950

Тактовая частота: 2,6 ГГц.

Число ядер: 4.

Кэш второго уровня: 2 Мб (по 512 Кб на каждое ядро).

Кэш третьего уровня: 2 Мб (общий на все ядра).

Частота интегрированного контроллера памяти: 1066 МГц.

Технологический процесс: 65 нанометров.

Процессор Third-Generation AMD Opteron™ 8360 SE

Тактовая частота: 2,5 ГГц.

Число ядер: 4.

Кэш второго уровня: 2 Мб (по 512 Кб на каждое ядро).

Кэш третьего уровня: 2 Мб (общий на все ядра).

Частота интегрированного контроллера памяти: 2000 МГц.

Технологический процесс: 65 нанометров.

В заключение, как и для процессоров компании Intel, приведем усредненные данные из списка Top500. Для систем в Top500, построенных на процессорах компании AMD, отношение «показанная мощность/пиковая мощность» составляет в 31-м списке 71%, при этом «удельная мощность» в расчете на один процессор/ядро равна 4,48 гигафлопс. Как и ранее, отметим, что значительная часть этих систем построена не на новейших процессорах.

1.2.3. Процессоры IBM Power6

История компании IBM значительно длиннее, чем у Intel и AMD, и, в отличие от последних, IBM никогда не производила только процессоры. Фактически, компания, говоря сегодняшним языком, всегда пыталась поставлять «готовые решения». Однако обсуждения всего списка продукции IBM выходит за рамки данного материала, и мы остановимся только на процессорах, которые выпускает компания сегодня и на основе которых строит как серверы «начального» уровня, так и суперкомпьютеры вроде Roadrunner или BlueGene.

Микропроцессорная архитектура Power (расшифровывается как Performance Optimization With Enhanced RISC) имеет не менее богатую историю, чем сама компания IBM. Начиная с 1990 г., когда были выпущены

первые компьютеры на основе процессоров Power, и по сегодняшний день архитектура постоянно развивается, с каждым поколением процессоров привнося значительные новшества. Текущая версия процессоров Power – Power6 (рис. 1.7) выпущена в середине 2007 г., тем не менее, уже 7 систем в 31-м списке Top500 построено на основе этих процессоров.

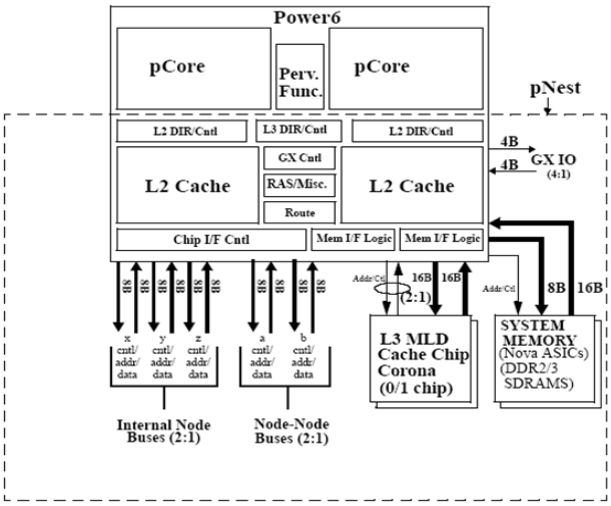


Рис. 1.7. Процессор IBM Power6

Процессор Power6 выпускается по 65 нм технологическому процессу. Максимальная частота серийно выпускаемых образцов на сегодня равна 4,7 ГГц.

Процессоры Power6 имеют два ядра, способных выполнять по два потока команд одновременно, по 4 Мб кэша второго уровня на каждое ядро, 32 Мб кэша третьего уровня на отдельном кристалле, присоединенном к шине с пропускной способностью 80 Гб/с. Каждое ядро содержит по два блока работы с целыми числами и числами с плавающей точкой соответственно. Однако главной изюминкой процессора Power6 является блок десятичных вычислений с плавающей точкой, аппаратно реализующий более 50 команд для выполнения математических операций над вещественными числами в десятичном представлении и перевода из двоичной системы счисления в десятичную и обратно. Второе, не менее важное отличие Power6 от процессоров предыдущих серий, – отказ IBM от внеочередного (out-of-order) исполнения команд, что стало одним из основных факторов, позволившим поднять частоту процессоров выше 4 ГГц.

Процессоры Power6 поставляются в многочиповом корпусе, аналогично Power5, вмещающем до 4 процессоров и общий кэш третьего уровня. В многопроцессорной конфигурации возможна «связка» из 32 процессоров

посредством двух шин межпроцессорного обмена с пропускной способностью 50 Гб/с.

В заключение, как и ранее, приведем усредненные данные из списка Top500. Для систем в Top500, построенных на процессорах IBM семейства Power, отношение «показанная мощность/пиковая мощность» составляет в 31-м списке 77%, при этом «удельная мощность» в расчете на один процессор/ядро равна 3,71 гигафлопс. Как и ранее, отметим, что значительная часть этих систем построена не на новейших процессорах.

Если же брать в расчет только системы на процессорах Power6, то картина несколько меняется: отношение «показанная мощность/пиковая мощность» для этих систем составляет 65%, а «удельная» на одно ядро – 12,14 гигафлопс. Существенно меньший показатель по показанной мощности не в последнюю очередь объясняется отказом от внеочередного исполнения команд, затрудняющего и без того непростую задачу достижения пиковой производительности. Что касается «удельной мощности», то здесь с наилучшей стороны проявляет себя высокая тактовая частота процессоров Power6.

1.2.4. Процессоры PowerXCell™ 8i

Рассказ о процессоре PowerXCell™ 8i, конечно же, нужно начать с его прямого предка – процессора Cell (рис. 1.8), разработанного альянсом STI (Sony, Toshiba, IBM) в первую очередь для использования в игровых приставках Sony PlayStation 3. В процессе создания этого процессора были приняты весьма интересные решения, дающие в итоге очень высокую пиковую производительность (более 200 гигафлопс, правда, только для вещественной арифметики одинарной точности), но требующие в качестве платы более сложного программирования.

Прежде всего, отметим, что процессор Cell имеет существенно «неоднородное» устройство. Он состоит из одного двухъядерного Power Processor Element (PPE) и 8 Synergistic Processor Element (SPE). PPE построен на архитектуре PowerPC и «отвечает» в процессоре Cell за исполнение кода общего назначения (операционной системы в частности), а также контролирует работу потоков на сопроцессорах SPE. Ядра PPE – 64-разрядные и, так же, как и Power6, используют поочередный (in-order) порядок исполнения команд. PPE имеет блок векторных операций Vector Multimedia eXtensions (VMX), кэш первого уровня размеров 64 Кб (по 32 Кб на кэш инструкций и данных) и кэш второго уровня размером 512 Кб.

В отличие от PPE, SPE-ядра представляют собой специализированные векторные процессоры, ориентированные на быструю потоковою работу с SIMD-инструкциями. Архитектура SPE довольно проста: четыре блока для работы с целочисленными векторными операциями и четыре блока для работы с числами с плавающей запятой. Большинство арифметических инст-

рукций представляют данные в виде 128-разрядных векторов, разделённых на четыре 32-битных элемента. Каждый SPE оснащён 128 регистрами, разрядность которых – 128-бит. Вместо кэша первого уровня SPE содержит 256 Кб собственной «локальной памяти» (local memory, также называемой local store или LS), разделённой на четыре отдельных сегмента по 64 Кб каждый, а также DMA-контроллер, который предназначен для обмена данными между основной памятью (RAM) и локальной памятью SPE (LS), минуя PPE. Доступ к LS составляет 6 тактов, что больше, чем время обращения к кэшу первого уровня, но меньше, чем к кэшу второго уровня для большинства современных процессоров. SPE-ядра, также как и PPE, используют упорядоченную схему (in-order) исполнения инструкций.

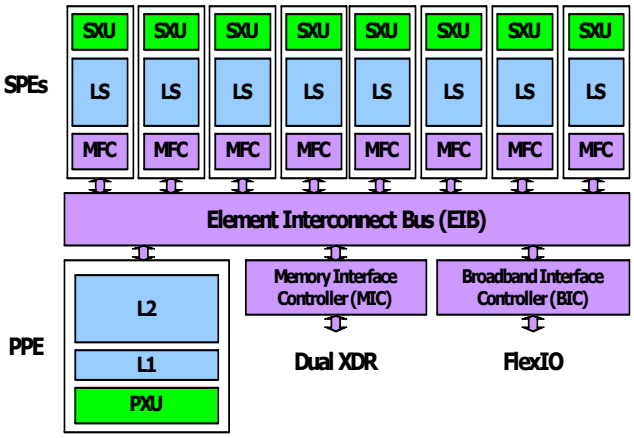


Рис. 1.8. Процессор Cell

Частота всех ядер в процессоре Cell составляет 3,2 ГГц, что дает производительность одного SPE в 3,2 × 4 × 2 = 25,6 гигафлопс (последняя двойка в произведении за счет двух конвейеров, позволяющих за один такт выполнять операции умножения и сложения над вещественными числами). Таким образом, пиковая производительность всего процессора Cell получается превышающей 200 гигафлопс.

Модель программирования для процессора Cell «изначально» многопоточная, поскольку на SPE могут выполняться только специализированные потоки. Данные, с которыми они работают, должны располагаться в LS, соответственно типичным подходом является их предвыборка. В целом Cell весьма эффективно справляется с «потоковой» обработкой, характерной для мультимедиа, для задач кодирования, сжатия и т. д.

Основное отличие процессора PowerXCell™ 8i от своего «предка» состоит в значительном улучшении работы с вещественными числами двойной точности, что позволило довести пиковую производительность на них

до уровня в 100 гигафлопс. Кроме того, PowerXCell™ 8i производится по 65 нм технологии, в отличие от 90 нм, использующихся в Cell. Наконец, в PowerXCell™ 8i был кардинально (до 32 Гб) увеличен объем поддерживаемой памяти.

В настоящий момент в Top500 три системы построены на процессорах PowerXCell 8i, в том числе лидер списка. Как и ранее, приведем усредненные данные из списка Top500 по этим трем системам. Отношение «показанная мощность/пиковая мощность» систем на основе процессоров PowerXCell 8i составляет в 31-м списке 74%, при этом «удельная мощность» в расчете на одно ядро равна 8,36 гигафлопс (т. е. порядка 80 гигафлопс на процессор).

1.2.5. Процессоры Sun UltraSPARC T1 и Sun UltraSPARC T2

Начиная разработку микроархитектуры UltraSPARC Architecture, компания Sun Microsystems подошла к процессу с позиций, существенно отличающихся от остальных производителей. В многоядерных процессорах Intel, AMD и IBM каждое ядро фактически является полноценным исполнительным устройством, ориентированным на выполнение кода общего назначения, а в процессорах семейства Cell SPE-ядра, напротив, в принципе не могут исполнять такой код и, по сути, являются сопроцессорами.

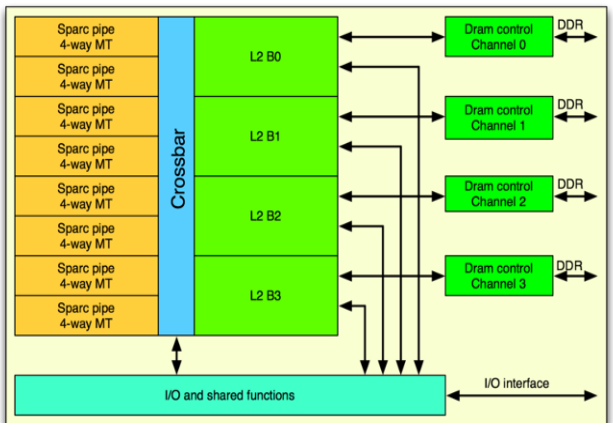


Рис. 1.9. Процессор Sun UltraSPARC T1

В основу процессоров UltraSPARC T1 (кодовое имя Niagara), см. рис. 1.9, выпущенных на рынок в 2005 г., и UltraSPARC T2 (кодовое имя Niagara-2), выпущенных в 2007 г., положена идея «многопоточности» для достижения высокой производительности не путем ускорения выполнения одного потока команд, а за счет обработки большого числа потоков в единицу времени. В результате процессоры UltraSPARC T1 способны выполнять 32 потока одновременно (на восьми «четырехпоточных» ядрах), а процессоры

UltraSPARC T2 – 64 потока (на восьми «восьмипоточных» ядрах). Эта многопоточность аппаратная (как, например, HyperThreading у компании Intel), т. е. операционная система воспринимает UltraSPARC T1 и UltraSPARC T2 как 32 и 64 процессора соответственно.

В обоих процессорах компания Sun реализовала технологию, названную ими CoolThreads, позволяющую значительно снизить энергопотребление – у процессоров UltraSPARC T1 TDP не превышает 79 Вт (по 2,5 ватта на поток), а у процессоров UltraSPARC T2 – 123 Вт (всего 2 ватта на поток).

Технические характеристики процессора Sun UltraSPARC T1:

- Тактовая частота: 1,0 или 1,2 ГГц.
- Число ядер: 8 (по 4 потока на каждое).
- Кэш инструкций первого уровня: 16 Кб на каждое ядро.
- Кэш данных первого уровня: 8 Кб на каждое ядро.
- Кэш второго уровня: 3 Мб (общий на все ядра).
- Интерфейс JBUS с пиковой пропускной способностью 3,1 Гб/с, 128-битной шиной частотой от 150 до 200 МГц.
- 90-нм технологический процесс.
- Энергопотребление: 72 Вт, пиковое – 79 Вт.

Переключение между потоками в процессоре UltraSPARC T1 осуществляется по циклической схеме на каждом такте, т. е. в каждый конкретный момент времени активен только один из четырех потоков ядра. Однако в случае, если в потоке возникает простой (например, при кэш-промахе), ядро переключается на работу с другим потоком. Такая стратегия позволяет скрыть возникающие задержки доступа к памяти при наличии достаточного количества потоков исполнения. Ядра UltraSPARC T1 по функциональности аналогичны процессорам предыдущего поколения UltraSPARC III, но существенно упрощены архитектурно, например, сокращены возможности прогноза ветвлений и спекулятивного выполнения команд, а число стадий конвейера уменьшено до шести (14 в UltraSPARC III).

Интересная особенность процессоров UltraSPARC T1 и T2 – наличие встроенного в ядро криптографического модуля (сопроцессора), реализующего на аппаратном уровне алгоритм RSA с 2048-разрядными ключами. Сопроцессор ядер в UltraSPARC T2 дополнительно поддерживает алгоритмы шифрования DES, 3DES, RC4, AES, SHA, MD5, CRC, а также алгоритм генерации случайных чисел.

Основной недостаток процессора UltraSPARC T1 – наличие в процессоре только одного блока вычислений над числами с плавающей точкой, доступного для всех потоков всех ядер. В процессоре UltraSPARC T2 эту проблему решили – у каждого ядра есть собственный модуль для выполнения операций с вещественными числами. Также в UltraSPARC T2 была

поднята максимальная тактовая частота – до 1,4 ГГц. Плюс к этому – увеличен объем кэша второго уровня до 4 Мб, однако в отличие от UltraSPARC T1 кэш не общий, а раздельный – по 512 Кб на каждое ядро. Кроме того, в процессор интегрированы два 10-Gbit контроллера Ethernet и контроллер шины PCI Express.

1.3. Ускорители вычислений

Технологический мир сегодня пронизан конвергенцией – взаимным влиянием и даже взаимопроникновением технологий, стиранием границ между ними, возникновением многих интересных результатов на стыке областей в рамках междисциплинарных работ. Одно из проявлений этого явления – «игра» основных производителей аппаратных составляющих компьютеров на «чужих полях». Так, компании NVIDIA и ATI (последняя теперь – в составе компании AMD), накопив опыт и поняв, что пиковая мощность их продуктов уже стала сравнима с кластерами «средней руки», от выпуска графических ускорителей начали движение на рынок высокопроизводительных решений, представив соответствующие продукты (семейства NVIDIA® Tesla™ и ATI FireStream™). Напротив, компании, традиционно выпускавшие процессоры и серверные решения, взялись осваивать область мультимедиа: компания Intel разрабатывает многопоточные векторные графические устройства, компания IBM в составе альянса STI создала, как мы уже обсуждали, процессор Cell, изначально ориентированный именно на быструю обработку мультимедиа информации. Никуда не делись и типичные ускорители вычислений, способные существенно добавить мощности даже обычным «персоналкам» – на этом направлении работает, например, компания ClearSpeed Technology [108].

Как следствие, перед обычными пользователями, желающими попробовать, «с чем едят» суперкомпьютерные технологии, возникает большой и не всегда просто осуществляемый выбор. Попытаемся немного прокомментировать возможности двух из представленных выше систем: ClearSpeed™ Advance™ X620 и NVIDIA® Tesla™ D870.

1.3.1. Ускоритель ClearSpeed™ Advance™ X620

ClearSpeed™ Advance™ X620 (см. рис. 1.10) – это ускоритель операций над данными с плавающей запятой, представленных в формате с двойной точностью. Ускоритель является сопроцессором, разработанным специально для серверов и рабочих станций, которые основаны на 32-х или 64-х битной x86 архитектуре, и построен на базе двух процессоров CSX600 со 194 вычислительными ядрами. X620 подключается к PCI-X разъему на материнской плате. Среда разработки под X620 основана на языке C и включает SDK, а также набор инструментов для написания и отладки программ.

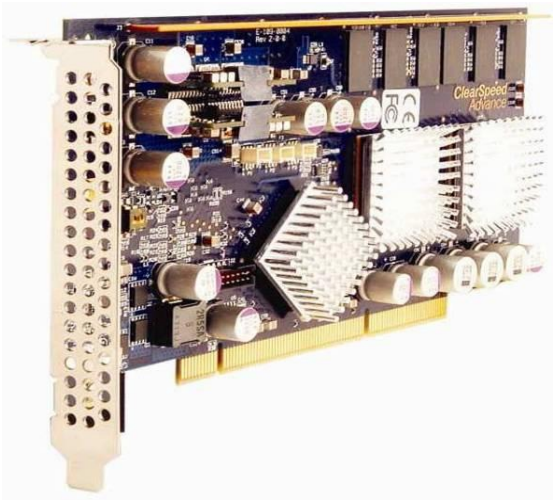


Рис. 1.10. Ускоритель ClearSpeed Advance X260

Технические характеристики X620:

- 2 процессора CSX600, работающих на частоте 210 МГц.
- Каждый CSX600 содержит 97 ядер (из них 96 в poly-исполнительном блоке и 1 в mono-исполнительном блоке).
- Пиковая производительность X620 – 66 гигафлопс.
- Память 1 Гб DDR2 DRAM 64-bit (по 0,5 Гб на каждый процессор).
- Пиковая пропускная способность шины памяти 3,2 Гб/с.
- Подключается к ПК через PCI-X разъем материнской платы.
- Максимальная потребляемая мощность: 43 Вт.
- Рабочий диапазон температур: 10–50ºC.

Ускоритель X260 поддерживает операционные системы семейств Linux (Red Hat, SUSE) и Windows (XP, 2003 Server). Для работы с X620 необходимо установить специальный драйвер.

При работе с ускорителем программист может использовать несколько уровней памяти:

- каждый mono-процессор обладает локальной памятью (mono local memory) размером 128 Кб;
- каждый poly-процессор обладает локальной памятью (poly local memory) размером 6 Кб;
- все процессоры во всех блоках имеют доступ к общей памяти устройства (global memory или device memory) размером 1024 Мб.

Главное (хост) приложение, использующее возможности X620, состоит из двух частей:

- программа, выполняющаяся на главном процессоре;
- программа, которая выполняется на ускорителе.

ClearSpeed™ Advance™ X620 обычно используется в качестве сопроцессора для ускорения внутренних циклов программы, достигающегося за счет использования стандартных математических библиотек, разработанных ClearSpeed. Если в приложении происходит вызов функции, поддерживаемой математическими библиотеками ClearSpeed, то библиотека сама анализирует возможность ее ускорения. В зависимости от выполненного анализа функция начинает выполняться либо на главном процессоре, либо перехватывается ускорителем. Этот механизм широко используется в ряде математических приложений, например в MATLAB.

Если в приложении происходит вызов функции, не поддерживаемой библиотеками ClearSpeed, то программист, в случае необходимости, может сам реализовать ее, используя имеющийся инструментарий.

1.3.2. Настольный суперкомпьютер NVIDIA® Tesla™ D870

В 2007 г. компания NVIDIA представила продукты семейства Tesla™ для построения высокопроизводительных вычислительных систем. Семейство включает вычислительный процессор NVIDIA® Tesla™ C870, приставной суперкомпьютер NVIDIA® Tesla™ D870 и вычислительный сервер NVIDIA® Tesla™ S870. Два последних решения построены на двух и четырех процессорах C870 соответственно. Сервер S870 предназначен для построения на его основе кластерных решений (его обсуждение выходит за рамки данной книги).

Рис. 1.11. Внешний вид NVIDIA® Tesla™ D870

Настольный (в терминологии компании NVIDIA – приставной) суперкомпьютер NVIDIA® Tesla™ D870 (см. рис. 1.11) конструктивно представляет собой два процессора C870, объединенных в небольшом корпусе. Необходимым условием для подключения D870 к компьютеру является наличие не менее двух PCI Express разъемов – в первом должна быть установлена видеокарта NVIDIA не менее чем 8xxx серии, а во второй ставится специальная плата-переходник, к которой и подключается D870.

Технические характеристики D870:

- 2 платы NVIDIA® Tesla™ C870.
- Каждый процессор C870 содержит 128 скалярных процессора с частотой 1,35 ГГц.
- Пиковая производительность D870 – 1 терафлопс.
- Память 3 Гб GDDR3 384-bit (по 1,5 Гб на каждый процессор).
- Пиковая пропускная способность шины памяти 76,8 Гб/с.
- Подключается кабелем к специальной плате-переходнику, установленной в разъем PCI Express x8 или x16.
- Максимальный уровень шума 40 дБ.
- Максимальная потребляемая мощность 520 Вт.

Суперкомпьютер D870 поддерживает операционные системы семейств Windows, Linux, Mac OS. В комплект поставки входят следующие компоненты (необходимо устанавливать в указанном порядке): CUDA Driver, CUDA Toolkit, CUDA SDK.

CUDA (Compute Unified Device Architecture) – программно аппаратное решение, позволяющее использовать видеопроцессоры для вычислений общего назначения. Возможность программирования видеопроцессоров компании NVIDIA на CUDA существует, начиная с семейства видеокарт 8xxx. Разработка программ для выполнения на D870 происходит таким же способом.

С точки зрения программиста, D870 представляет собой набор независимых мультипроцессоров. Каждый мультипроцессор состоит из нескольких независимых скалярных процессоров, двух модулей для вычисления математических функций, конвейера, а также общей памяти.

CUDA позволяет создавать специальные функции (ядра, kernels), которые выполняются параллельно различными блоками и потоками, в отличие от обычных C-функций. При запуске ядра блоки распределяются по доступным мультипроцессорам. Мультипроцессор занимается распределением, параллельным выполнением потоков внутри блока и их синхронизацией. Каждый поток независимо исполняется на одном скалярном процессоре с собственным стеком инструкций и памятью.

CUDA предоставляет программисту доступ к нескольким уровням памяти:

- каждый поток обладает локальной памятью;
- все потоки внутри блока имеют доступ к быстрой общей памяти блока, время жизни которой совпадает со временем жизни блока; память блока разбита на страницы, доступ к данным на разных страницах осуществляется параллельно;
- все потоки во всех блоках имеют доступ к общей памяти устройства.

Всем потокам также доступны два вида общей памяти для чтения: константная и текстурная, они кэшируются. Так же, как и в общей памяти устройства, данные сохраняются на протяжении работы приложения.

1.4. Персональные мини-кластеры

Были времена, когда кластеры собирали на основе обычных рабочих станций. Таков, например, был первый известный кластер Beowulf, собранный летом 1994 г. в научно-космическом центре NASA из 16 компьютеров на базе процессоров 486DX4 с тактовой частотой 100 MHz. Связь узлов в этом кластере осуществлялась посредством 10 Мбит/с сети, невероятно медленно по сегодняшним меркам. Однако довольно быстро стало понятно, что каждый отдельный узел кластера не нуждается во всем многообразии комплектующих, из которых состоит обычный «отдельно стоящий» компьютер.

Рис. 1.12. Мини-кластер Orion Multisystems

В действительности все, что должно быть в каждом узле, – процессор, память, жесткий диск с размером, достаточным, чтобы установить на него операционную систему, и сетевой интерфейс. Это понимание привело к тому, что характерным форм-фактором для стоечных серверов, из которых собирают современные кластеры, стал системный блок с высотой порядка 4,5 см. (форм-фактор 1U). Идея «упаковать» некоторое количество узлов в корпус небольших размеров, так, чтобы кластер не требовал отдельного помещения, что называется, лежала на поверхности. Мы уже упоминали про мини-кластер RenderCube. Еще одним примером подобного подхода является мини-кластер, представленный в 2005 г. стартап-компанией Orion Multisystems, включающий до 96 процессоров и до 192 Гб памяти [103] (рис. 1.12).

В 2006 г. свои решения в этой нише начала поставлять российская компания «Т-Платформы», сначала на основе процессоров компании AMD, а позднее и компании Intel.

1.4.1. Персональный суперкомпьютер T-Forge Mini

T-Forge Mini – компактный суперкомпьютер, габаритные размеры (360×321×680 мм) и небольшая масса которого позволяют установить его непосредственно на рабочем месте сотрудника (см. рис. 1.13). Уровень шума T-Forge Mini не превышает 45 децибел, что позволяет использовать мини-кластер в том числе и в офисных условиях.

Технические характеристики мини-кластера T-Forge Mini:

- До 4-х двухпроцессорных узлов на базе двухъядерных процессоров AMD Opteron™ или AMD Opteron™ HE с низким энергопотреблением, объединенных сетью Gigabit Ethernet.
- Память до 64 Гб.
- До 4-х устройств HDD SATA общим объемом до 2 Тб.
- 4 блока питания мощностью 350 Вт.
- До 9-ти портов Gigabit Ethernet.
- Контроль частоты вращения вентиляторов.
- Видео-карта ATI Rage XL 8Mb.
- Плата удаленного управления сервером с поддержкой стандарта IPMI (опционально).

Операционная система: ОС SUSE Linux Enterprise Server 9, RedHat Enterprise Linux 4 или Microsoft

- Windows Compute Cluster Server 2003.

Рис. 1.13. Мини-кластер T-Forge Mini

Один из первых мини-кластеров T-Forge Mini в максимальной конфигурации (не считая объема памяти) с пиковой производительностью в 70 гигафлопс был приобретен ННГУ в рамках нацпроекта «Образование» в 2006 г. [105] и используется при выполнении многих образовательных и научных проектов.

1.4.2. Мини-кластер T-Edge Mini

Мини-кластер T-Edge Mini (рис. 1.14) несколько крупнее, чем T-Forge Mini (габаритные размеры T-Edge Mini 530×360×700 мм), да и масса под 100 кг уже не позволяет назвать его «настольным», но под столом он вполне способен разместиться. Дополнительный объем был использован компанией-разработчиком для расширения возможностей кластера – во-первых, размещен дополнительный вычислительный узел, во-вторых, поддерживаются четырехъядерные процессоры, в-третьих, в качестве интерконнекта может быть использован не только Gigabit Ethernet, но и Infiniband.

Рис. 1.14. Мини-кластер T-Forge Mini

Технические характеристики мини-кластера T-Edge Mini:

- До 5-ти двухпроцессорных узлов на базе четырехъядерных процессоров Intel® Xeon®, объединенных сетью Gigabit Ethernet или Infiniband.
- Память до 64 Гб.

- 3 блока питания мощностью 600 Вт.
- Адаптер сервисной сети, осуществляющий мониторинг и администрирование управляющего узла по протоколу RS 485 ServNET v.2.0.
- Порты ввода/вывода на передней панели: VGA port, USB ports, Keyboard и Mouse.
- Порты ввода/вывода на задней панели: 2 RJ-45 GbE ports, 1 RJ-45 FE port, 4 порта для мониторинга узлов по протоколу RS 485.
- Встроенный DVD привод.
- Встроенный KVM и GbE коммутатор.
- Операционная система: ОС SUSE Linux Enterprise Server 9, RedHat Enterprise Linux 4 или Microsoft Windows Compute Cluster Server 2003.

В 2007 г. ННГУ в рамках нацпроекта «Образование» приобрел два мини-кластера T-Edge Mini, построенных на процессорах Intel® Xeon® 5320 (1,86 ГГц) с 20 Гб оперативной памяти и дисковой подсистемой на 1,25 Тб. В качестве интерконнекта в одном из мини-кластеров использован Gigabit Ethernet, во втором – Infiniband. Пиковая производительность каждого – 297 гигафлопс.

1.5. Краткий обзор главы

Данная глава посвящена рассмотрению компьютерных вычислительных устройств, построенных на базе многоядерных процессоров для организации высокопроизводительных вычислений.

С этой целью в главе дана краткая характеристика класса многопроцессорных вычислительных систем MIMD по классификации Флинна. С учетом характера использования оперативной памяти в составе этого класса выделены две важных группы систем с общей разделяемой и распределенной памятью – *мультипроцессоры* и *мультикомпьютеры*.

Далее отмечено, что наиболее перспективное направление для достижения высокой производительности вычислений на данный момент времени состоит в явной организации многопроцессорности вычислительных устройств. Для обоснования данного утверждения приведено подробное рассмотрение основных способов организации многопроцессорности – *симметричной мультипроцессорности* (*Symmetric Multiprocessor*, *SMP*), *одновременной многопотоковости* (*Simultaneous Multithreading*, *SMT*) и *многоядерности* (*multicore*). Сравнение перечисленных подходов позволяет сделать вывод, что многоядерность процессоров становится одним из основных направлений развития компьютерной техники. Значимость такого подхода привела даже к тому, что известный закон Мура теперь формулируется в новом «многоядерном» виде: «Количество вычислительных ядер в процессоре будет удваиваться каждые 18 месяцев».

В продолжение темы многоядерности в разделе 1.2 дана характеристика ряда конкретных широко применяемых в настоящее временя многоядерных процессоров основных компаний-разработчиков Intel, AMD, IBM, Sun.

Для полноты картины в разделе 1.3 приведено описание ряда аппаратных устройств (видеокарт и вычислительных сопроцессоров), которые могут быть использованы для существенного ускорения вычислений. И для завершения рассматриваемой темы в разделе 1.4 дана краткая характеристика «персональных» мини-кластеров, которые позволяют при достаточно «экономных» финансовых затратах приступить к решению имеющихся вычислительно-трудоемких задач с использованием высокопроизводительных вычислительных систем.

1.6. Обзор литературы

Дополнительная информация об архитектуре параллельных вычислительных систем может быть получена, например, [8,18,49,67,78]; полезная информация содержится также в [47, 100].

Информация по новейшим разработкам многоядерных процессоров содержится на официальных сайтах компаний-производителей компьютерного оборудования Intel, AMD, IBM, Sun, Nvidia, ClearSpeed и др.

Подробное рассмотрение вопросов, связанных с построением и использованием кластерных вычислительных систем, проводится в [47, 100]. Практические рекомендации по построению кластеров для разных систем платформ могут быть найдены в [19, 92–93].

1.7. Контрольные вопросы

1. В чем заключаются основные способы достижения параллелизма?
2. В чем могут состоять различия параллельных вычислительных систем?
3. Что положено в основу классификации Флинна?
4. В чем состоит принцип разделения многопроцессорных систем на мультипроцессоры и мультикомпьютеры?
5. Какие классы систем известны для мультипроцессоров?
6. В чем состоят положительные и отрицательные стороны симметричных мультипроцессоров?
7. Чем обосновывается целесообразность аппаратной поддержки одновременной многопотоковости (simultaneous multithreading)?
8. Какое ускорение вычислений может достигаться для процессоров с поддержкой одновременной многопотоковости?

9. Чем вызывается необходимость разработки многоядерных процессоров? Приведите положительные и отрицательные стороны многоядерности.
10. Какие требования должны быть предъявлены к программам для их эффективного выполнения на многоядерных процессорах?
11. В чем могут состоять различия конкретных многоядерных процессоров? Приведите несколько примеров.
12. Чем вызвана необходимость разработки ускорителей вычислений общего назначения?
13. Каким образом графические процессоры могут быть использованы для организации высокопроизводительных вычислений?
14. Какие характерные признаки отличают персональные мини-кластеры?

1.8. Задачи и упражнения

1. Приведите дополнительные примеры параллельных вычислительных систем.

2. Выполните рассмотрение дополнительных способов классификации компьютерных систем.

3. Рассмотрите способы обеспечения когерентности кэшей в системах с общей разделяемой памятью.

4. Приведите дополнительные примеры многоядерных процессоров.

5. Изучите и дайте общую характеристику способов разработки программ для графических процессоров (на примере настольного суперкомпьютера NVIDIA Tesla).

6. Изучите и дайте общую характеристику способов разработки программ для ускорителей вычислений общего назначения (на примере вычислительных сопроцессоров компании ClearSpeed)?