

Центральный процессор

Материалы по дисциплине «Микропроцессорные системы»
Специальность «Компьютерные системы и комплексы»
Составитель: Торгашин Р.Г

ГБПОУ ВО "Борисоглебский техникум промышленных и информационных технологий"

2016 год

Оглавление

Характеристики CPU.....	3
Тактовая частота.....	4
TDP.....	5
Техпроцесс.....	6
Системная шина. Шина процессора.....	6
Архитектура процессора.....	11
Способы увеличения производительности процессоров.....	13
Конвейерная архитектура.....	13
Суперскалярность.....	16
Математический сопроцессор.....	17
КЭШ.....	17
Поддерживаемый набор расширений.....	19
Поддержка виртуализации.....	19
Архитектура NetBurst.....	20
Многоядерная архитектура.....	23
Архитектура K8.....	23
Архитектура Core.....	24
Архитектура K10.....	25
Архитектура Intel Core i7 (Nehalem).....	27
Вопросы для контроля.....	33
Источники.....	34

Характеристики CPU

Центральный процессор (ЦП, или центральное процессорное устройство — ЦПУ; англ. central processing unit, сокращенно — CPU, дословно — центральное обрабатывающее устройство) — электронный блок либо микросхема — исполнитель машинных инструкций (кода программ), главная часть аппаратного обеспечения компьютера или программируемого логического контроллера. Иногда называют микропроцессором или просто процессором.

Ранние ЦП создавались в виде уникальных составных частей для уникальных, и даже единственных в своём роде, компьютерных систем. Позднее от дорогостоящего способа разработки процессоров, предназначенных для выполнения одной единственной или нескольких узкоспециализированных программ, производители компьютеров перешли к серийному изготовлению типовых классов многоцелевых процессорных устройств.

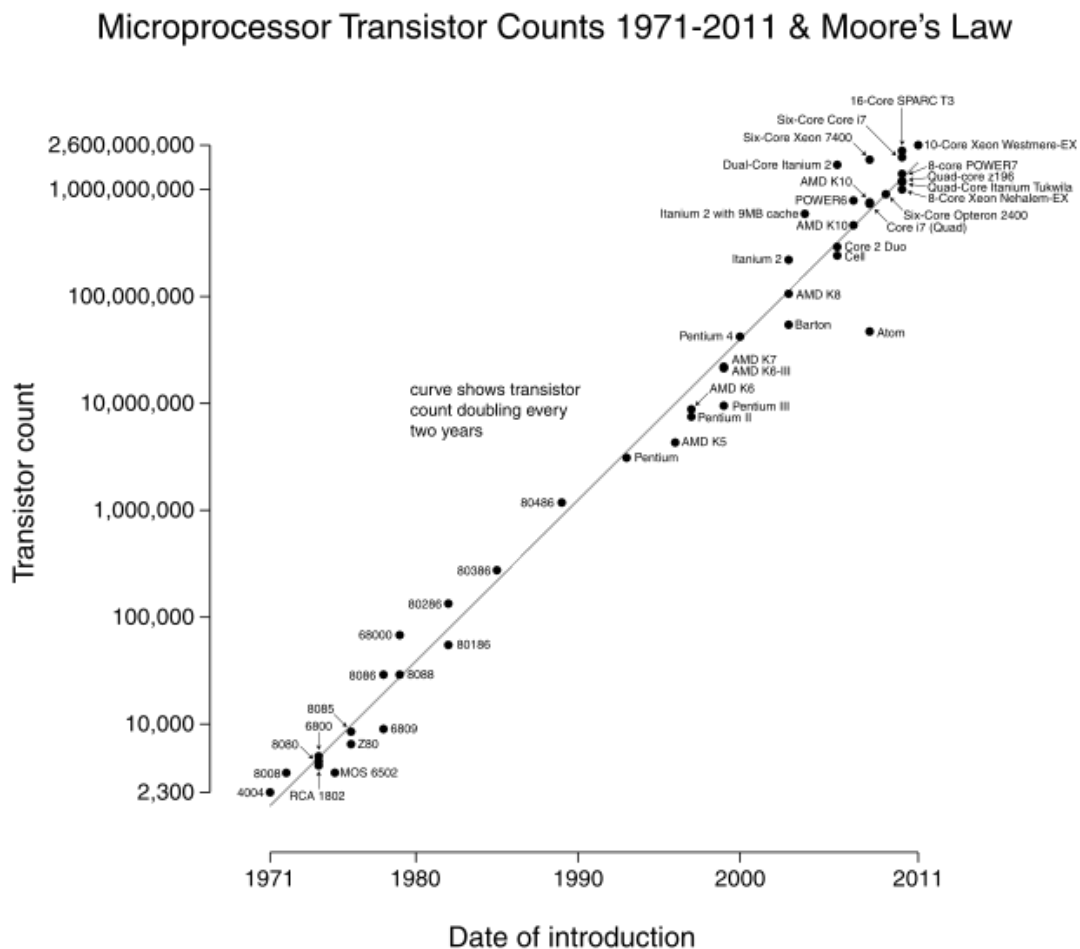


Рисунок 1 Число транзисторов в микропроцессорах 1971-2011 год. Закон Мура¹
Главными характеристиками ЦПУ являются:

- тактовая частота,
- энергопотребление,
- тепловыделение, TDP
- нормы литографического процесса используемого при производстве – техпроцесс
- архитектура,
 - число ядер,
 - объем и организация КЭШа,
 - поддерживаемая система команд

¹ Мур высказал предположение, что число транзисторов на кристалле будет удваиваться каждые 24 месяца

- производительность,

Ясно, что с точки зрения пользователя наибольший интерес представляет производительность процессора. Поэтому все остальные параметры рассматриваются обычно с точки зрения их влияния на производительность.

Тактовая частота

Тактовая частота — частота синхронизирующих импульсов синхронной электронной схемы, то есть количество синхронизирующих тактов, поступающих извне на вход схемы за одну секунду. Обычно термин употребляется применительно к компонентам компьютерных систем. В самом первом приближении тактовая частота характеризует производительность подсистемы (процессора, памяти и пр.), то есть количество выполняемых операций в секунду.

Однако системы с одной и той же тактовой частотой могут иметь различную производительность, так как на выполнение одной операции разным системам может потребоваться различное количество тактов (обычно от долей такта до десятков тактов), а кроме того, системы, использующие конвейерную и параллельную обработку, могут на одних и тех же тактах выполнять одновременно несколько операций.

Ранее тактовая частота была основным показателем производительности процессоров. Сейчас, из-за повсеместного внедрения многоядерных процессоров, кеширования операций, производители процессоров не рекомендуют использовать тактовую частоту как основной ориентир для определения производительности. В частности, утверждается, что при распараллеливании задачи на несколько ядер, каждой ядро может работать на своей тактовой частоте. То есть при выполнении различных задач, тактовая частота ядер различается. Следовательно, неясно, тактовую частоту какого ядра, в какой момент учитывать при анализе?

Поэтому такие фирмы как Intel и AMD предлагают синтетические рейтинги, которые не соответствуют тактовой частоте или какому-либо другому объективному параметру. Они лишь показывают место процессора среди остальных в рейтинге производительности. Фирма AMD для такой индексации ввела термин «PR-рейтинг»

Таблица 1 Тактовая частота, количество ядер и производительность процессоров

Model	МГц	Ядра	3DMark06 CPU
Intel Core i7-990X EE	3467	6	6971
Intel Core i7 (Desktop) 980X EE	3333	6	7122.5
Intel Core i7 (Desktop) 2600K	3400	4	6637.9
Intel Xeon X5670	2933	6	5794
Intel Core i7 (Desktop) 975	3333	4	5836.5
Intel Core i5 (Desktop) 2500K	3300	4	5853
Intel Core i5 (Desktop) 2400	3100	4	5715
Intel Core i7 (Desktop) 960	3200	4	5360
Intel Core i7 (Desktop) 950	3060	4	5219.6
Intel Core i7 (Desktop) 940	2930	4	5054
Intel Core i7 (Desktop) 920	2660	4	4727.5
Intel Core i5 (Desktop) 750	2660	4	4320
Intel Core 2 Extreme (Desktop) QX6850	3000	4	4174
Intel Core 2 Quad (Desktop) Q9550	2830	4	4233.3
AMD A8-3850	2900	4	4027
Intel Core 2 Extreme (Desktop) QX6700	2666	4	4021.5
Intel Core 2 Duo (Desktop) E6850	3000	2	2661
Intel Core 2 Extreme (Desktop) X6800	2930	2	2567.5
Intel Core 2 Quad (Desktop) Q6600	2400	4	3546.5
Intel Core 2 Duo (Desktop) E6700	2660	2	2354
Intel Core 2 Duo (Desktop) E6600	2400	2	1982
AMD Athlon 64 (Desktop) FX-62	2800	2	2149
AMD Athlon 64 X2 (Desktop) FX-60	2600	2	1959.5
AMD Athlon 64 X2 (Desktop) 4800+	2400	2	1800.7
AMD Athlon 64 X2 (Desktop) 4400+	2200	2	1639
Intel Pentium D 820	2800	2	1412
Intel Pentium 4 P4 3.0	3000	1	811

Сравним процессоры, показанные в таблице 1.

Видно, что хотя частота Intel Core 2 Duo (Desktop) E6850 равная 3000 МГц на 340 больше частоты Intel Core i7 (Desktop) 920, производительность Core i7 больше в 1,7 раз. Также, можно сделать вывод что немаловажную роль в таком росте производительности играет вдвое большее число ядер. Но несмотря на то, что у Intel Core i7 (Desktop) 920 и Intel Core i5 (Desktop) 750 число ядер одинаково и усредненная рабочая частота обоих процессоров 2660 МГц, производительность Core i7 выше на 10%.

Хотя тактовая частота до сих пор является важным параметром при анализе производительности процессора, ее увеличение далеко не всегда означает увеличение производительности.

TDP

TDP (thermal design power) - величина, показывающая, на отвод какой тепловой мощности должна быть рассчитана система охлаждения процессора или другого полупроводникового прибора. К примеру, если система охлаждения процессора рассчитана на TDP 30 Вт, она должна быть в состоянии отвести 30 Вт тепла при некоторых заданных «нормальных условиях».

TDP показывает не максимальное теоретическое тепловыделение процессора, а лишь требования к производительности системы охлаждения.

TDP рассчитан на определённые "нормальные" условия, которые иногда могут быть нарушены.

Например в случае поломки вентилятора или неверного охлаждения самого корпуса.

Современные процессоры при этом или дают сигнал выключения компьютера, или переходят в так называемый режим троттлинга (англ. throttling), когда процессор пропускает часть циклов.

Разные производители микросхем рассчитывают TDP по-разному, поэтому величина не может напрямую использоваться для сравнения энергопотребления процессоров. Никогда нельзя с уверенностью утверждать, что процессор с TDP 100 Вт потребляет больше энергии, чем процессор другого производителя с TDP 5 Вт.

Немного странно, но TDP часто заявляется для кристалла, который объединяет целое семейство процессоров, без учета тактовой частоты работы процессора, при этом младшие модели обычно потребляют меньше энергии и рассеивают меньше тепла, чем старшие.

Также некоторые специалисты переводят этот термин как «thermal design package» («термопакет») — проектирование устройства на основе температурного анализа конструкции.

Для современных процессоров TDP можно определить по их маркировке. Например

- Классификация TDP процессоров Intel Core i3, i5, i7 (Sandy Bridge)
 - Безиндексные модели — TDP 95 Вт
 - К — TDP 95 Вт для 4-ядерных моделей (индекс “К” отображает наличие у процессора разблокированного множителя)
 - S — TDP 65 Вт для 4-ядерных моделей
 - Т — TDP 45 Вт для 4-ядерных моделей, 35 Вт для 2-ядерных моделей
- Классификация TDP процессоров Intel Core 2 Duo
 - X — TDP более 75 Вт
 - E — TDP до 45 Вт
 - Т — TDP до 35 Вт
 - P — TDP до 25 Вт
 - L — TDP до 17 Вт
 - U — TDP до 10 Вт
 - SP — TDP до 25 Вт
 - SL — TDP до 17 Вт
 - SU — TDP до 10 Вт
- Классификация TDP процессоров AMD Athlon II & Phenom II
 - E — TDP до 45Вт
 - U — TDP до 25Вт

Важно, что TDP никак не связан с производительностью процессора. В основном он зависит от техпроцесса использованного при его изготовлении и архитектуры

Таблица 2 TDP, техпроцесс и производительность процессоров

Model	Техпроцесс		
	TDP (ватт)	(нм)	3DMark06 CPU
Intel Core i7-990X EE	130	32	6971
Intel Core i7 (Desktop) 980X EE	130	32	7122.5
Intel Core i7 (Desktop) 2600K	95	32	6637.9
Intel Xeon X5670	95	32	5794
Intel Core i7 (Desktop) 975	130	45	5836.5
Intel Core i5 (Desktop) 2500K	95	32	5853
Intel Core i5 (Desktop) 2400	95	32	5715
Intel Core i7 (Desktop) 960	130	45	5360
Intel Core i7 (Desktop) 950	130	45	5219.6
Intel Core i7 (Desktop) 940	130	45	5054
Intel Core i7 (Desktop) 920	130	45	4727.5
Intel Core i5 (Desktop) 750	95	45	4320
Intel Core 2 Extreme (Desktop) QX6850	130	65	4174
Intel Core 2 Quad (Desktop) Q9550	95	45	4233.3
AMD A8-3850	100	32	4027
Intel Core 2 Extreme (Desktop) QX6700	130	65	4021.5
Intel Core 2 Duo (Desktop) E6850	65	65	2661
Intel Core 2 Extreme (Desktop) X6800	75	65	2567.5
Intel Core 2 Quad (Desktop) Q6600	105	65	3546.5
Intel Core 2 Duo (Desktop) E6700	65	65	2354
Intel Core 2 Duo (Desktop) E6600	65	65	1982
AMD Athlon 64 (Desktop) FX-62	125	90	2149
AMD Athlon 64 X2 (Desktop) FX-60	110	90	1959.5
AMD Athlon 64 X2 (Desktop) 4800+	65	90	1800.7
AMD Athlon 64 X2 (Desktop) 4600+	65	90	
AMD Athlon 64 X2 (Desktop) 4400+	65	90	1639
Intel Pentium D 820	95	90	1412
Intel Pentium 4 660	115	90	
Intel Pentium 4 560	88	90	
Intel Pentium 4 630	95	90	
Intel Pentium 4 P4 3.0	89	130	811

Техпроцесс

Техпроцесс - это масштаб технологии, которая определяет размеры полупроводниковых элементов, составляющих основу внутренних цепей процессора (эти цепи состоят из соединенных соответствующим образом между собой транзисторов). Совершенствование технологии и пропорциональное уменьшение размеров транзисторов способствуют улучшению характеристик процессоров.

Чем меньше техпроцесс, тем больше транзисторов можно уместить в кристалле. У процессоров изготовленных по меньшему техпроцессу меньше длина «проводников», а значит, меньше сопротивление. Соответственно уменьшается сопротивление и тепловыделение процессора.

Как видно из таблицы 2 производители постоянно стремятся уменьшить техпроцесс.

Системная шина. Шина процессора

Когда говорят о шине процессора, чаще всего имеют в виду шину данных, представленную как набор соединений (или выводов) для передачи или приема данных.

Параметры шины:

Тактовая частота

Разрядность - Шина состоит из нескольких каналов для передачи электрических сигналов. Если говорят, что шина тридцатидвухразрядная, то это означает, что она способна передавать электрические сигналы по тридцати двум каналам одновременно. Здесь есть одна фишка. Дело в том, что шина любой заявленной разрядности (8, 16, 32, 64) имеет, на самом деле, большее количество каналов. То есть, если

взять ту же тридцатидвухразрядную шину, то для передачи собственно данных выделено 32 канала, а дополнительные каналы предназначены для передачи специфической информации.

Скорость передачи данных - Название этого параметра говорит само за себя. Он высчитывается по формуле: тактовая частота * разрядность = скорость передачи данных

Сделаем расчет скорости передачи данных для 64 разрядной системной шины, работающей на тактовой частоте в 100 МГц.

$$100 * 64 = 6400 \text{ Мбит/сек}$$

$$6400 / 8 = 800 \text{ Мбайт/сек}$$

Но полученное число не является реальным. В жизни на шины влияет куча всевозможных факторов: неэффективная проводимость материалов, помехи, недостатки конструкции и сборки а также многое другое. По некоторым данным, разность между теоретической скоростью передачи данных и практической может составлять до 25%.

За работой каждой шины следят специально для этого предназначенные контроллеры. Они входят в состав набора системной логики (чипсет).

До недавнего времени основной шиной процессора была FSB

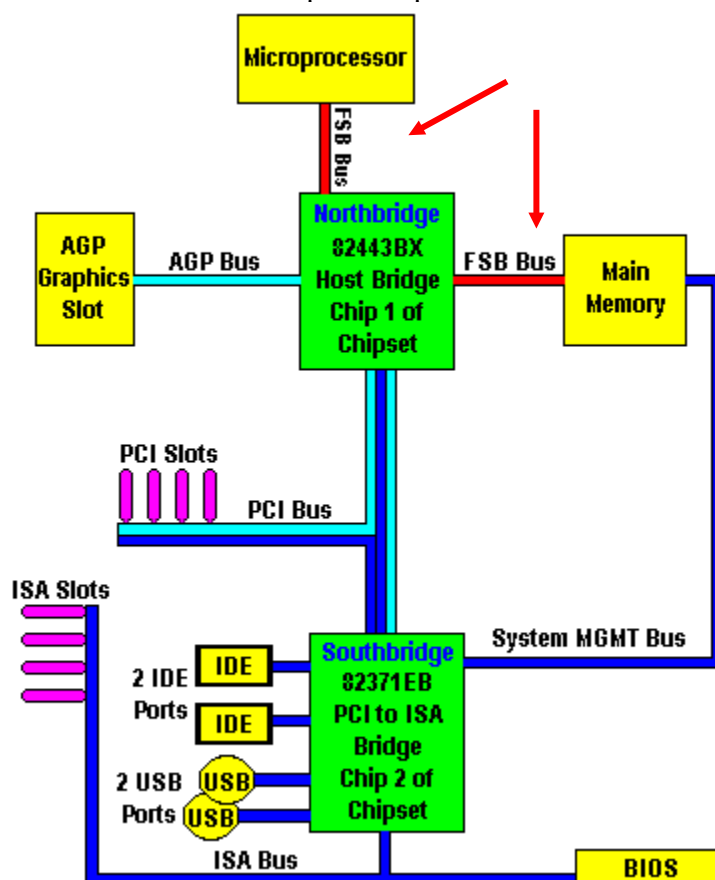


Рисунок 2 Место FSB в структуре компьютера. На примере чипсета Intel 440BX

Front Side Bus (FSB) — шина, обеспечивающая соединение между x86-совместимым центральным процессором и внутренними устройствами.

Как правило, в современном персональном компьютере на базе x86-совместимого микропроцессора микропроцессор через FSB подключается к северному мосту.

Шина часто делится на несколько – шину адреса и шину данных. Первая служит для передачи адреса ячейки памяти. Вторая же для передачи данных в эту ячейку.

Таким образом, FSB работает в качестве магистрального канала между процессором и чипсетом.

Некоторые компьютеры имеют внешнюю кэш-память, подключенную через «заднюю» шину (англ. back side bus), которая быстрее, чем FSB, но работает только со специфичными устройствами.

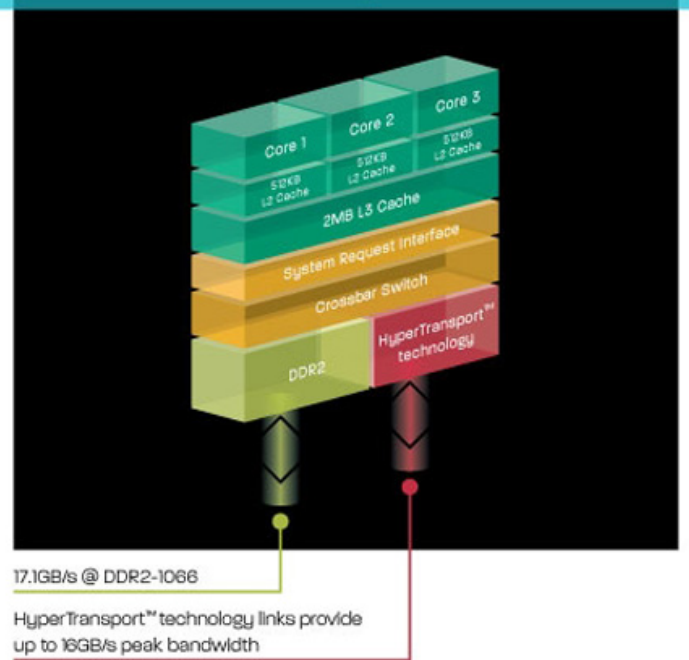
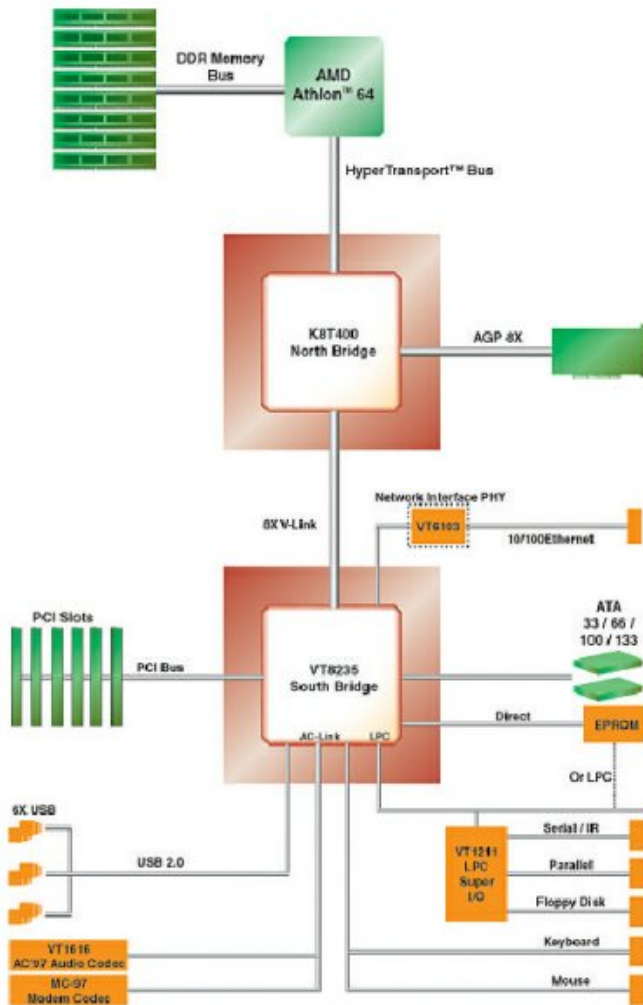


Рисунок 3 Шина HyperTransport в чипсете для процессора Athlon64 и в процессорах Phenom

Шина HyperTransport (HT), ранее известная как Lightning Data Transport (LDT), — это двунаправленная последовательно/параллельная компьютерная шина с высокой пропускной способностью и малыми задержками. Для разработки и продвижения данной шины был образован консорциум HyperTransport Technology. Технология используется компаниями AMD и Transmeta в x86-процессорах.

Шина HyperTransport нашла широкое применение, в основном, в качестве замены шины процессора.

Другое применение HyperTransport — шина многопроцессорных компьютеров. AMD использует HyperTransport как часть проприетарной архитектуры Direct Connect Architecture в своей линейке процессоров Opteron, Athlon 64 и Phenom. Технология шинного соединения HORUS компании Newisys расширяет концепцию до уровня кластерных систем.

HyperTransport работает на частотах от 200 МГц до 3,2 ГГц (у шины PCI — 33 и 66 МГц). Кроме того, она использует DDR, что означает, что данные посылаются как по фронту так и по срезу сигнала синхронизации, что позволяет осуществлять до 5200 миллионов посылок в секунду при частоте сигнала синхронизации 2,6 ГГц; частота сигнала синхронизации настраивается автоматически.

HyperTransport поддерживает автоматическое определение ширины шины, от 2-х до 32 бит. Полноразмерная, полноскоростная, 32-битная шина в двунаправленном режиме способна обеспечить пропускную способность до $41\,600 \text{ Мбайт/с} = 2 \text{ (DDR)} \times 2 \times 32/8 \text{ (байт)} \times 2600 \text{ (МГц)}$ (максимум в одном направлении — 20 800 Мбайт/с), являясь, таким образом, самой быстрой шиной среди себе подобных. Шина может быть использована как в подсистемах с высокими требованиями к пропускной способности (оперативная память и ЦПУ), так и в подсистемах с низкими требованиями (периферийные устройства). Данная технология также способна обеспечить низкие задержки для других применений в других подсистемах.

Шина HyperTransport основана на передаче пакетов. Пакеты HyperTransport передаются по шине последовательно. Увеличение пропускной способности влечёт за собой увеличение ширины шины. HyperTransport может использоваться для передачи служебных сообщений системы, для передачи прерываний, для конфигурирования устройств, подключённых к шине и для передачи данных.

Шина HyperTransport нашла широкое применение, в основном, в качестве замены шины процессора. AMD использует HyperTransport как часть проприетарной архитектуры Direct Connect Architecture в своей линейке процессоров Opteron, Athlon 64 и Phenom.

Шина HyperTransport поддерживает технологии энергосбережения, а именно ACPI. Это значит, что при изменении состояния процессора (C-state) на энергосберегающее изменяется также и состояние устройств (D-state). Например, при отключении процессора жёсткие диски также отключаются.

Электрический интерфейс HyperTransport/LDT — низковольтные дифференциальные сигналы с напряжением 1,2 В.

Версия	Год	максимальная частота	максимальная ширина	пиковая пропускная способность (в оба направления)
1.0	2001	800 МГц	32 бит	12,8 Гбайт/с ^[1]
1.1	2002	800 МГц	32 бит	12,8 Гбайт/с
2.0	2004	1,4 ГГц	32 бит	22,4 Гбайт/с
3.0	2006	2,6 ГГц	32 бит	41,6 Гбайт/с
3.1	2008	3,2 ГГц	32 бит	51,2 Гбайт/с

Рисунок 4: Версии Hyper Transport

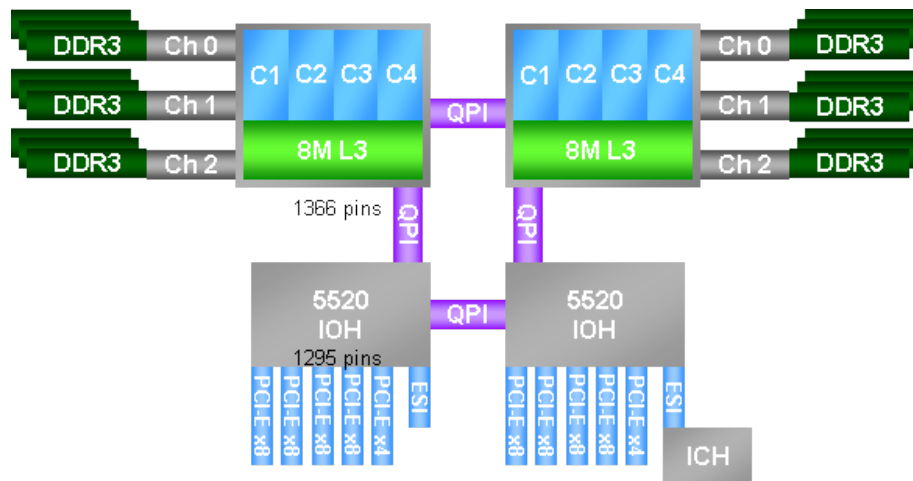


Рисунок 5 Шина QPI в архитектуре Nehalem. Процессоры Core i5, Core i7

Intel QuickPath Interconnect или просто QuickPath, сокр. QPI (ранее Common System Interface, CSI) — последовательная кэш-когерентная шина типа точка-точка для соединения процессоров между собой и с чипсетом, разработанная фирмой Intel. QPI создавался в ответ на разработанную ранее консорциумом во главе с фирмой AMD шину HyperTransport

Шина QuickPath была создана для замены применявшейся ранее шины Front Side Bus, которая осуществляла связь между центральным процессором и северным мостом материнской платы. Первые процессоры с интерфейсом QuickPath были выпущены на рынок в 2008 году. По состоянию на начало 2010 года, внешний интерфейс QuickPath используется только в сериях процессоров Xeon и Core i7 с ядром Nehalem для разъёма LGA 1366, а также будет использоваться в следующем поколении Itanium (ядро Tukwila). При этом чипсеты для разъёма LGA 1366 используют шину DMI для связи между северным и южным мостом. Процессоры для разъёма LGA 1156 не имеют внешнего интерфейса

QuickPath, поскольку чипсеты для данного разъёма поддерживают только однопроцессорную конфигурацию, а функциональность северного моста встроена в сам процессор (и следовательно, для связи процессора с аналогом южного моста используется шина DMI). Однако внутри процессора LGA 1156 связь между ядрами и встроенным контроллером PCIe осуществляется через встроенную шину QuickPath

Каждое соединение шины QuickPath состоит из пары односторонних каналов, каждый из которых физически реализован как 20 дифференциальных пар проводов. Данные передаются в виде пакетов (датаграмм). Пропускная способность одного канала составляет от 4,8 до 6,4 миллиарда передач в секунду. Одна передача содержит 16 бит полезной нагрузки, следовательно теоретическая суммарная пропускная способность одного соединения (в двух направлениях) — от 19,2 до 25,6 гигабайт в секунду (то есть от 9,6 до 12,8 гигабайт/с в каждую сторону); при этом один процессор может иметь несколько соединений.

Direct Media Interface

DMI — интерфейс, представленный Intel в 2004 году (начиная с Intel 915 Express), который служит мостом для соединения южного (ICH) и северного моста (MCH), либо для непосредственной шины «процессор – чипсет – процессор» в LGA 1156, LGA 775 и более ранние, кроме LGA 1366.

Шина DMI имеет свойства архитектуры, объединяющие ее с такой шиной для подключения периферийных устройств, как PCI Express. В частности, DMI использует линии с последовательной передачей данных, а также имеет отдельные проводники для передачи и приема данных.

Пропускная способность шины DMI первого поколения составляет 2 Гб/с. Вместе с этим полоса пропускания 2 Гб/с (по 1 Гб/с в каждом направлении) делится с другими устройствами (например, PCI Express x1, PCI, HD Audio, жесткие диски), так что фактически её ширина уже.

В материнских платах для процессоров с разъёмом LGA 1156 (то есть для Core i3, Core i5 и некоторых серий Core i7 и Xeon) и со встроенным контроллером памяти DMI используется для подсоединения чипсета (PCH) непосредственно к процессору. (Процессоры серии Core i7 для LGA 1366 подсоединяется к чипсету через шину QPI)

DMI является проприетарной технологией Intel. В 2009 году Intel отказалась лицензировать шину DMI фирме Nvidia. Поскольку поддержка DMI встроена в процессоры с ядрами Lynnfield и Clarkdale для разъёма LGA 1156 и используется для подсоединения к чипсету, Nvidia фактически потеряла право производить чипсеты для большей части новых процессоров Intel.

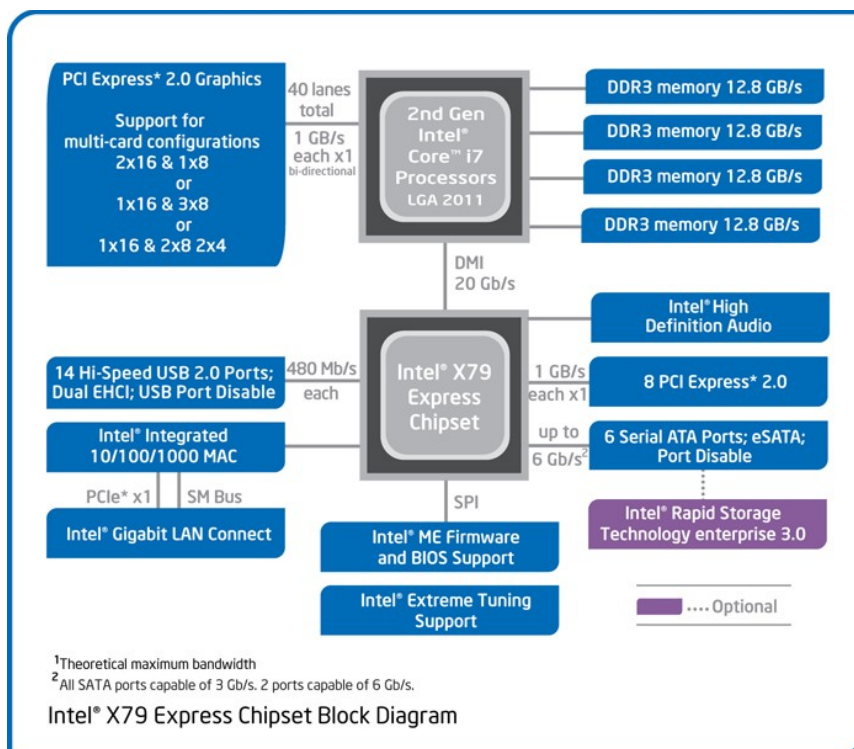


Рисунок 6: Шина DMI в чипсете Intel X79

В 2011 году было представлено второе поколение интерфейса, DMI 2.0, в котором скорость передачи данных увеличилась в 2 раза, до 2 ГБ/с в каждую сторону по DMI 2.0 на базе 4 линий. Данный вариант использовался для соединения центральных процессоров Intel 2011-2015 годов с микросхемой Platform Controller Hub (PCH), частично заменившей набор из южного и северного мостов.

DMI 3.0 был представлен в августе 2015. В третьем поколении скорость обменов была увеличена до 8 GT/s (гигатранзакций в секунду) на каждой линии. Интерфейс с 4 линиями позволяет передавать данные со скоростью до 3,93 ГБ/с между процессором и PCH. Используется в процессорах с микроархитектурой Skylake (варианты с 2 чипами) и чипсетах Intel серии 100, например Z170.

Архитектура процессора

Современные процессоры персональных компьютеров основываются на архитектуре¹ фон Неймана.

Такая архитектура основывается на следующих принципах:

- Принцип двоичного кодирования. Для представления данных и команд используется двоичная система счисления.
- Принцип однородности памяти. Как программы (команды), так и данные хранятся в одной и той же памяти (и кодируются в одной и той же системе счисления — чаще всего двоичной). Над командами можно выполнять такие же действия, как и над данными.
- Принцип адресуемости памяти. Структурно основная память состоит из пронумерованных ячеек; процессору в произвольный момент времени доступна любая ячейка; память внутренняя.
- Принцип последовательного программного управления. Все команды располагаются в памяти и выполняются последовательно, одна после завершения другой, в последовательности, определяемой программой.
- Принцип жесткости архитектуры. Неизменяемость в процессе работы топологии, архитектуры, списка команд.

Альтернативой является «Гарвардская» архитектура, в которой реализовано физическом разделении линий передачи команд и данных. Соответствующая схема реализации доступа к памяти имеет один очевидный недостаток — высокую стоимость. При разделении каналов передачи команд и данных на кристалле процессора последний должен иметь почти в два раза больше выводов (так как шины адреса и данных составляют основную часть выводов микропроцессора). Способом решения этой проблемы стала идея использовать общую шину данных и шину адреса для всех внешних данных, а внутри процессора использовать шину данных, шину команд и две шины адреса. Такую концепцию стали называть модифицированной Гарвардской архитектурой.

Такой подход применяется в современных сигнальных процессорах. Еще дальше по пути уменьшения стоимости пошли при создании однокристальных ЭВМ — микроконтроллеров. В них одна шина команд и данных применяется и внутри кристалла.

Существуют гибридные модификации архитектур, сочетающие достоинства как Гарвардской, так и фон Неймановской архитектур. Современные CISC-процессоры обладают отдельной кэш-памятью 1-го уровня для инструкций и данных, что позволяет им за один такт получать одновременно как команду, так и данные для её выполнения, то есть процессорное ядро, формально, является гарвардским, но с программной точки зрения выглядит как фон Неймановское, что упрощает написание программ. Обычно в данных процессорах одна шина используется и для передачи команд, и для передачи данных, что упрощает конструкцию системы. Современные варианты таких процессоров могут иногда содержать встроенные контроллеры сразу нескольких разнотипных шин для работы с различными типами памяти — например, DDR RAM и Flash. Тем не менее, и в этом случае шины, как правило, используются и для передачи команд, и для передачи данных без деления, что делает данные процессоры еще более близкими к фон Неймановской архитектуре при сохранении плюсов Гарвардской архитектуры..

Большинство современных процессоров для персональных компьютеров в общем основаны на той или иной версии циклического процесса последовательной обработки данных.

¹ Также называют «Принстонской» архитектурой

В различных архитектурах и для различных команд могут потребоваться дополнительные этапы. Например, для арифметических команд могут потребоваться дополнительные обращения к памяти, во время которых производится считывание операндов и запись результатов.

Этапы цикла выполнения:

1. Процессор выставляет число, хранящееся в регистре счётчика команд, на шину адреса и отдаёт памяти команду чтения.
2. Выставленное число является для памяти адресом; память, получив адрес и команду чтения, выставляет содержимое, хранящееся по этому адресу, на шину данных и сообщает о готовности.
3. Процессор получает число с шины данных, интерпретирует его как команду (машинную инструкцию) из своей системы команд и исполняет её.
4. Если последняя команда не является командой перехода, процессор увеличивает на единицу (в предположении, что длина каждой команды равна единице) число, хранящееся в счётчике команд; в результате там образуется адрес следующей команды.

Данный цикл выполняется неизменно, и именно он называется процессом (откуда и произошло название устройства).

Иначе процесс можно изобразить в виде последовательности действия

Получение инструкции	Раскодирование инструкции	Выполнение	Доступ к памяти	Запись в регистр
IF - Instruction Fetch	ID - Instruction Decode	EX - Execute	MEM – Memory access	WB - Register write back

Рисунок 7 Процесс

Во время процесса процессор считывает последовательность команд, содержащихся в памяти, и исполняет их. Такая последовательность команд называется программой и представляет алгоритм работы процессора. Очерёдность считывания команд изменяется в случае, если процессор считывает команду перехода, — тогда адрес следующей команды может оказаться другим. Другим примером изменения процесса может служить случай получения команды остановка или переключение в режим обработки прерывания.

Команды центрального процессора являются самым нижним уровнем управления компьютером, поэтому выполнение каждой команды неизбежно и безусловно. Не производится никакой проверки на допустимость выполняемых действий, в частности, не проверяется возможная потеря ценных данных. Чтобы компьютер выполнял только допустимые действия, команды должны быть соответствующим образом организованы в виде необходимой программы.

Скорость перехода от одного этапа цикла к другому определяется тактовым генератором.

Реализации архитектуры.

Существуют две глобальных реализации архитектуры процессора, на базе которых строятся все процессоры «персональных» компьютеров. Независимо от особенностей реализации процессора, если для него заявлена поддержка определенной архитектуры (IA-32 или x86-64) он должен выполнять все стандартные команды и обеспечивать работу программ рассчитанных на данную архитектуру. Это делает возможным запуск таких OS как MS-DOS и Unix на современных системах. В тоже время рекомендуется для новых процессоров использовать новые операционные системы и программы, так как они написаны с учетом новых расширений наборов команд. Например на процессорах Pentium IV и новее WindowsXP показывает большую производительность чем Windows98se. А на процессорах Phenome Windows 7 работает быстрее Windows XP.

x86 - архитектура процессора с одноименным набором команд, впервые реализованная в процессорах компании Intel.

Название образованно от двух цифр, которыми заканчивались названия процессоров Intel ранних моделей — 8086, 80186, 80286 (i286), 80386 (i386), 80486 (i486). За время своего существования набор команд постоянно расширялся, сохраняя совместимость с предыдущими поколениями.

Помимо Intel, архитектура также была реализована в процессорах других производителей: AMD, VIA, Transmeta, IDT и др. В настоящее время для этой архитектуры существует еще одно название — **IA-32** (Intel Architecture — 32).

Эта архитектура, точнее ее команды и расширения, поддерживается до сих пор вплоть до Core i7/i5/i3

x86-64 (также x64/AMD64/Intel64/EM64T) — 64-битная аппаратная платформа (чипсет, архитектура микропроцессора и команд), разработанная компанией AMD для выполнения 64-разрядных приложений.

Это расширение архитектуры x86 с полной обратной совместимостью. Корпорации Microsoft и Oracle используют для обозначения этого набора инструкций термин «x64», однако каталог с файлами для архитектуры в дистрибутивах Microsoft называется «amd64»

Набор команд x86-64 в настоящее время поддерживается:

- AMD — процессорами Athlon 64, Athlon 64 FX, Athlon 64 X2, Athlon II, Phenom, Phenom II, Turion 64, Turion 64 X2, Opteron, последними моделями Sempron;
- Intel (с незначительными дополнениями) под названием «Intel 64» (ранее известные как «EM64T» и «IA-32e») в поздних моделях процессоров Pentium 4, а также в Pentium D, Pentium Extreme Edition, Celeron D, Pentium Dual-Core, Core 2 Duo, Core 2 Quad, Core i3, Core i5, Core i7, Atom и Xeon;
- VIA — процессор Nano.

Способы увеличения производительности процессоров

Конвейерная архитектура

Конвейер — это способ организации вычислений, используемый в современных процессорах и контроллерах с целью повышения их производительности (увеличения числа инструкций, выполняемых в единицу времени), технология, используемая при разработке компьютеров и других цифровых электронных устройств.

Идея заключается в разделении обработки компьютерной инструкции на последовательность независимых стадий с сохранением результатов в конце каждой стадии. Это позволяет управляющим цепям процессора получать инструкции со скоростью самой медленной стадии обработки, однако при этом намного быстрее, чем при выполнении эксклюзивной полной обработки каждой инструкции от начала до конца.

В процессоре построенном по бесконвейерной архитектуре в каждый момент времени нагружен только один модуль. Процессор или получает инструкцию или декодирует или выполняет. Процессоры с конвейером внутри устроены так, что обработка инструкций разделена на последовательность стадий, предполагая одновременную обработку нескольких инструкций на разных стадиях. Результаты работы каждой из стадий передаются через ячейки памяти на следующую стадию, и так — до тех пор, пока инструкция не будет выполнена.

Подобная организация процессора, при некотором увеличении среднего времени выполнения каждой инструкции, тем не менее, обеспечивает значительный рост производительности за счёт высокой частоты завершения выполнения инструкций.

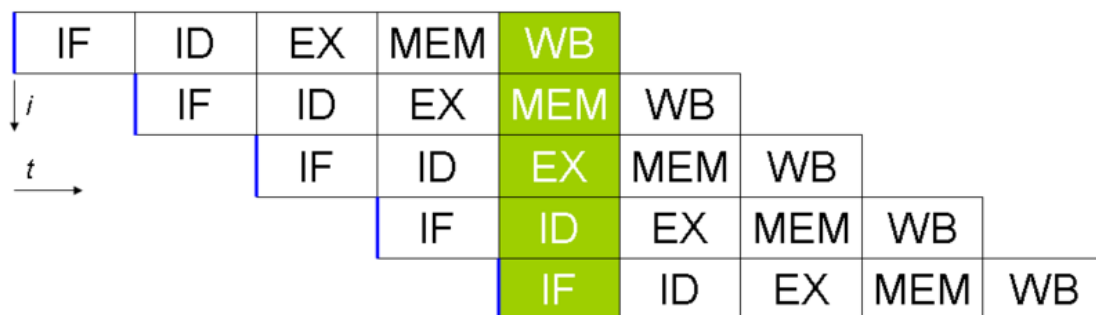


Рисунок 8 Простой пятиуровневый конвейер в RISC-процессорах. Вертикальная ось — это последовательные независимые инструкции, горизонтальная — время.

На рисунке 4 в зеленой колонке, которая описывает состояние процессора в один момент времени, самая ранняя, верхняя инструкция уже находится в состоянии записи в регистр, а самая последняя, нижняя инструкция только в процессе чтения. Очевидно, что в идеальном случае, когда все стадии конвейера заполнены, время простоя каждого модуля близко к нулю. А количество обработанных операций значительно больше, чем у бесконвейерного процессора.



Рисунок 9 Общий четырехуровневых конвейер; цветные квадраты символизируют независимые друг от друга инструкции

Не все инструкции являются независимыми. В простейшем конвейере, где обработка инструкции представлена пятью стадиями, для обеспечения полной загрузки, в то время пока заканчивается обработка первой инструкции, должно обрабатываться параллельно ещё четыре последовательных независимых инструкции. Если последовательность содержит инструкции, зависящие от выполняемых в данный момент, то конвейер не «знает» какую инструкцию поместить на начальные стадии. Управляющая логика простейшего конвейера приостанавливает несколько начальных стадий конвейера, помещая этим самым в конвейер пустую инструкцию («пузырёк»), иногда неоднократно, — до тех пор, пока зависимость не будет разрешена. Зависимость между инструкциями, одновременно обрабатываемыми процессором, не позволяет добиться увеличения производительности кратно количеству стадий конвейера в сравнении с бесконвейерным процессором.

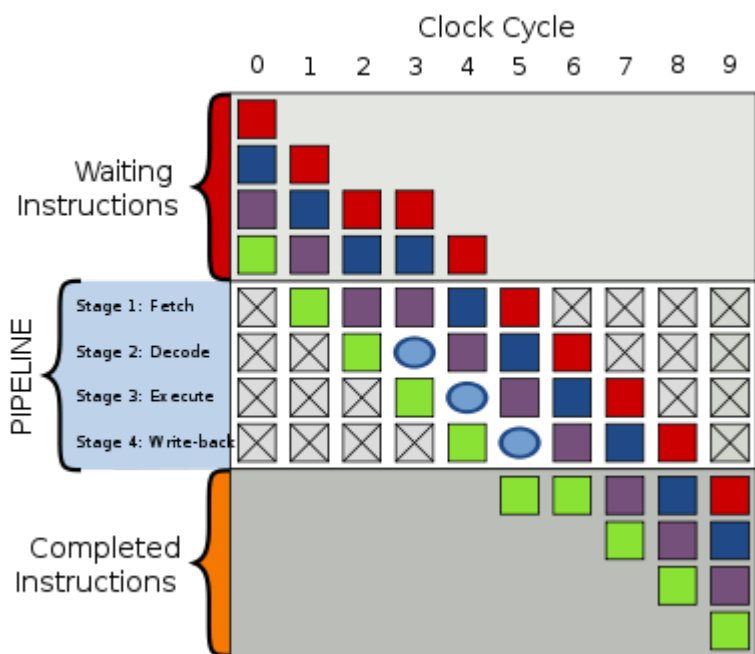


Рисунок 10 Пузырек в третьем такте обработки задерживает исполнение

Во втором такте обработка фиолетовой инструкции задерживается и вместо стадии расшифровки в третьем такте теперь находится пузырек. Всё, что находится «за» фиолетовой инструкцией, испытывает задержку в один такт, тогда как все, что находится «перед» фиолетовой инструкцией продолжает исполняться.

Очевидно, что наличие пузырька в конвейере дает суммарное время исполнения в 8 тактов вместо 7 на схеме исполнения, продемонстрированной на рисунке 5

Множество схем включают в себя конвейеры в 7, 10 или даже 20 уровней (как, например, в Intel Pentium 4). Поздние ядра Pentium 4 с кодовыми именами «Prescott» и «Cedar Mill» (и их Pentium D-производные) имеют 31-уровневый конвейер, самый длинный среди популярных процессоров. Xelerator X10q имеет конвейер длиной более, чем в тысячу шагов.

Обратной стороной медали в данном случае является необходимость сбрасывать весь конвейер в случае, если ход программы изменился (например, по условному оператору).

Предсказание переходов

Модуль предсказания условных переходов (Branch Prediction Unit) — устройство, входящее в состав микропроцессоров, имеющих конвейерную архитектуру, определяющее направление ветвлений (предсказывающее, будет ли выполнен условный переход) в исполняемой программе.

Предсказание ветвлений позволяет осуществлять предварительную выборку инструкций и данных из памяти, а также выполнять инструкции, находящиеся после условного перехода, до того, как он будет выполнен. Предсказатель переходов является неотъемлемой частью всех современных суперскалярных микропроцессоров, так как в большинстве случаев (*точность предсказания переходов в современных процессорах превышает 90 %*) позволяет оптимально использовать вычислительные ресурсы процессора

Существует два основных метода предсказания переходов: статический и динамический. Статические методы предсказания ветвлений являются наиболее простыми. Суть этих методов состоит в том, что различные типы переходов либо выполняются всегда, либо не выполняются никогда. В современных процессорах статические методы используются лишь в том случае, когда невозможно использование динамического предсказания.

Динамические методы, широко используемые в современных процессорах, подразумевают анализ истории ветвлений. Примером динамического предсказания может служить двухуровневый адаптивный исторический алгоритм (англ. Bimodal branch prediction), использовавшийся процессорами архитектуры P6¹ (анализируется таблица истории переходов, содержащая младшие значимые биты адреса инструкции и соответствующую им вероятность условного перехода: «скорее всего, будет выполнен», «возможно, будет выполнен», «возможно, не будет выполнен», «скорее всего, не будет выполнен» и обновляемая после каждого перехода).

Предсказание переходов само по себе может только усугубить ситуацию, если предсказание производится плохо. Если предсказание переходов не справляется и переход предсказан неверно — конвейер очищается, что создает задержку в работе процессора. В некоторых областях применения,

¹ Pentium Pro, Pentium II, Pentium III, Celeron и Xeon

таких как вычисления на суперкомпьютерах, программы специально пишутся так, чтобы как можно реже использовать условные операторы.

Высокая пропускная способность конвейеров оборачивается тормозами в случае, если в исполняемом коде содержится много условных переходов: процессор не знает, откуда читать следующую инструкцию, и поэтому вынужден ждать, когда закончится инструкция условного перехода, оставляя за ней пустой конвейер.

В крайнем случае, производительность конвейерного процессора может теоретически упасть до производительности бесконвейерного, или даже быть хуже за счет того, что будет занят только один уровень конвейера и между уровнями присутствует небольшая задержка.

Из-за конвейера процессора, код, который загружает процессор, не будет исполнен мгновенно. Из-за этого, обновления в коде, которые находятся очень близко к текущему месту исполнения программы, могут пройти незамеченными из-за того, что код уже предзагружен в Prefetch Input Queue. Кэш инструкций еще больше усугубляет эту проблему. Стоит учитывать, что данная проблема присутствует только в самомодифицирующихся программах, а также в упаковщиках исполняемых файлов.

Суперскалярность

Архитектура вычислительного ядра, использующая несколько декодеров команд, которые могут нагружать работой множество исполнительных блоков. Планирование исполнения потока команд является динамическим и осуществляется самим вычислительным ядром.

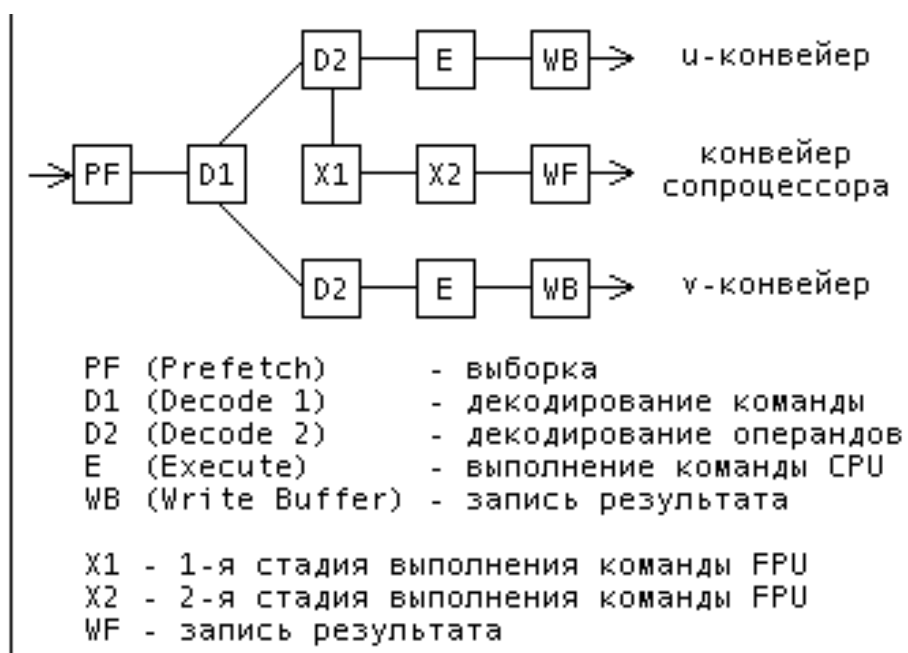


Рисунок 11 Реализация суперскалярности в процессорах Pentium.

Например, в процессорах Pentium впервые были использованы 3 конвейера. Основной, вспомогательный и конвейер математического сопроцессора. Главным считается u-конвейер, он поддерживает весь набор инструкций и работает на полную катушку. v-конвейер имеет некоторые ограничения и лишь помогает основному. Он умеет обрабатывать только "простые" инструкции - целочисленная арифметика, логические команды, команды перехода.

Если в процессе работы команды, обрабатываемые конвейером, не противоречат друг другу, и одна не зависит от результата другой, то такое устройство может осуществить параллельное выполнение команд. В суперскалярных системах решение о запуске инструкции на исполнение принимает сам вычислительный модуль, что требует много ресурсов¹.

Использование суперскалярности позволяет увеличить производительность за счет распараллеливания - выполнения инструкций на нескольких конвейерах одновременно.

¹ В более поздних системах, таких как Эльбрус-3 и Itanium, используется статпланирование, то есть параллельные инструкции объединяются компилятором в длинную команду, в которой все инструкции заведомо параллельные

Такая технология используется не только в CPU но и в GPU – процессорах видеокарт

Математический сопроцессор

Математический сопроцессор — сопроцессор¹ для расширения командного множества центрального процессора и обеспечивающий его функциональностью модуля операций с плавающей запятой, для процессоров, не имеющих интегрированного модуля.



Рисунок 12 Примеры сопроцессоров реализованных в виде отдельного модуля

Начиная с Intel 80486 DX сопроцессор интегрируют в основной процессор в виде дополнительного модуля. Такое устройство принято назвать модуль операций с плавающей запятой (floating point unit (FPU))

Простым «целочисленным» процессорам для работы с вещественными числами и математическими операциями требуются соответствующие процедуры поддержки и время для их выполнения. Модуль операций с плавающей запятой поддерживает работу с ними на уровне примитивов — загрузка, выгрузка вещественного числа (в/из специализированных регистров) или математическая операция над ними выполняется одной командой, за счёт этого достигается значительное ускорение таких операций.

С внедрением суперскалярной архитектуры для FPU выделили отдельный конвейер. Сам FPU был переработан и ускорен. Это сделало возможным появление целого ряда игр (Quake, MDK), сотворивших революцию в игровом мире. Благодаря новому FPU, мощность оказалась достаточной для показа Video CD без приобретения MPEG-карт, стоящих сотни долларов.

КЭШ

Кэш микропроцессора — кэш (сверхоперативная память), используемый микропроцессором компьютера для уменьшения среднего времени доступа к компьютерной памяти. Является одним из верхних уровней иерархии памяти. Кэш использует небольшую, очень быструю память (обычно типа SRAM), которая хранит копии часто используемых данных из основной памяти. Если большая часть запросов в память будет обрабатываться кэшем, средняя задержка обращения к памяти будет приближаться к задержкам работы кэша.

В современных многоядерных процессорах кэш может использоваться для обмена данными между ядрами

Большинство современных микропроцессоров для компьютеров и серверов имеют как минимум три независимых кэша: кэш инструкций для ускорения загрузки машинного кода, кэш данных для ускорения чтения и записи данных, и буфер ассоциативной трансляции (TLB) для ускорения трансляции виртуальных (математических) адресов в физические, как для инструкций, так и для данных. Кэш данных часто реализуется в виде многоуровневого кэша (L1, L2, L3).

¹ Сопроцессор — специализированный процессор, расширяющий возможности центрального процессора компьютерной системы, но оформленный как отдельный функциональный модуль.

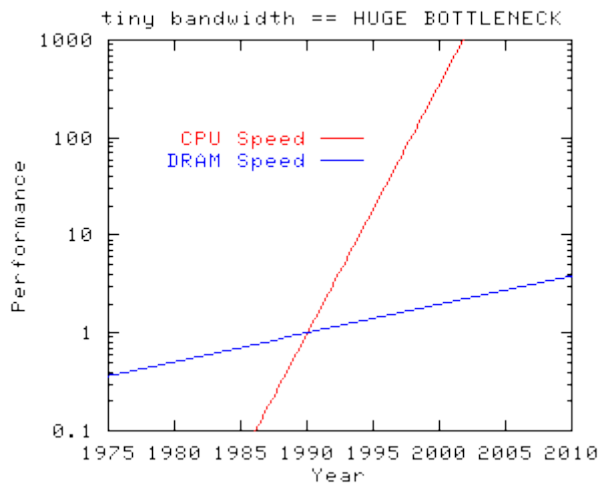


Рисунок 13 Соотношение скорости работы процессоров и оперативной памяти

Кэш состоит из собственно кэш-памяти, и кэш-контроллера. Кэш-контроллер управляет кэш-памятью: загружает в неё нужные данные из оперативной памяти, и возвращает, когда нужно, модифицированные процессором данные в оперативную память.

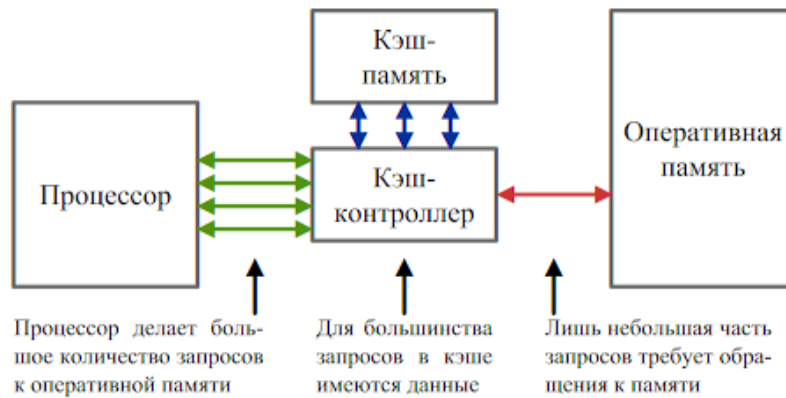


Рисунок 14 Принцип функционирования кэша

Эффективность кэша зависит от следующих факторов:

- Объём кэша. Чем больше объём кэша, тем большую часть требуемых программе данных он может в себе содержать, тем реже будут происходить обращения к оперативной памяти, и тем выше будет общее быстродействие системы.
- Алгоритм функционирования кэша. К сожалению, зачастую объёма кэш-памяти недостаточно для того, чтобы вместить все необходимые для вычислений данные. В этом случае кэш-контроллер должен «решить», какие именно данные следует держать в кэше. Поэтому кроме объёма кэша важным является алгоритм его функционирования: кэш, оснащённый хорошим алгоритмом, будет гораздо эффективнее использовать свой объём, храня меньше ненужных данных.
- Выполняемая процессором программа. Кэш оказывается эффективным потому, что большинство компьютерных программ обращаются к памяти не случайным образом, а закономерно. Чем лучше кэш-контроллер может «предсказать» обращения приложения к памяти, тем выше эффективность.

Одной из проблем является фундаментальная проблема баланса между задержками кэша и интенсивностью попаданий. Большие кэши имеют более высокий процент попаданий но, вместе с тем, и большую задержку. Чтобы ослабить противоречие между этими двумя параметрами, большинство компьютеров использует несколько уровней кэша, когда после маленьких и быстрых кэшей находятся более медленные большие кэши (в настоящий момент — суммарно до 3 уровней в иерархии кэшей)

Поддерживаемый набор расширений¹

Расширения архитектуры x86 (x86 extended features) — инженерные решения, разрабатываемые различными производителями микропроцессоров архитектуры x86 для собственных процессоров, которые, зачастую, присутствуют в одних моделях, и отсутствуют в других. Разработчики программного обеспечения, прежде чем использовать их в собственных программах, должны выполнить специальную команду CPUID, которая сообщает о наличии тех или иных расширений. Если расширение не поддерживается, программа должна либо завершить работу, либо использовать другие алгоритмы, не использующие расширений.

Иногда расширения становятся часто используемыми в популярных программах, тогда их начинают включать в свои процессоры практически все производители.

Наиболее известные из расширений MMX, SSE, 3DNow!, Hyper-threading. В основном, новые расширения разрабатываются фирмой Intel и, затем, подхватываются другими производителями. Но иногда бывает так, что сама Intel включает в свои процессоры расширения, созданные иными фирмами. К примеру, так было с архитектурой x86-64, впервые разработанной компанией AMD, а уже впоследствии реализованная Intel как Intel64/EM64T.

Поддержка виртуализации

В основе виртуализации лежит возможность одного компьютера выполнять работу нескольких компьютеров благодаря распределению его ресурсов по нескольким средам. С помощью виртуальных серверов и виртуальных настольных компьютеров можно разместить несколько ОС и несколько приложений в едином местоположении.

Сейчас возможность запуска нескольких виртуальных машин на одной физической вызывает большой интерес среди компьютерных специалистов, не только потому, что это повышает гибкость ИТ-инфраструктуры, но и потому, что виртуализация, на самом деле, позволяет экономить деньги.



Рисунок 15: Виртуальные машины с различными версиями Windows

Аппаратная виртуализация — виртуализация с поддержкой специальной процессорной архитектуры. Аппаратная виртуализация обеспечивает производительность, сравнимую с производительностью неvirtуализованной машины, что дает виртуализации возможность практического использования и влечет её широкое распространение. Наиболее распространены технологии виртуализации Intel-VT и AMD-V.

¹ SSE | SSE2 | SSE3 | SSSE3 | SSE4 | ATA | 3DNow! | 3DNowExt | SSE5 | AVX | AES

В **Intel VT** (Intel Virtualization Technology) реализована виртуализация режима реальной адресации (режим совместимости с 8086). Соответствующая аппаратная виртуализация ввода-вывода — VT-d (кодовое название — Vanderpool). Часто обозначается аббревиатурой VMX (Virtual Machine eXtension). AMD-V часто обозначается аббревиатурой SVM (Secure Virtual Machines). Кодовое название — Pacifica. Соответствующая технология виртуализации ввода-вывода — IOMMU. AMD-V проще и эффективнее, чем Intel VT.

По состоянию на 2015 год не все процессоры Intel поддерживали VT-x, что используется компанией Intel для сегментирования своего рынка. Поддержка VT-x может различаться даже между различными версиями (которые идентифицируются по sSpec Number) одной и той же модели.

На некоторых материнских платах пользователи должны вручную включить виртуализацию VT-x в настройках BIOS.

VT-d (Virtualization technology for directed I/O) — технология виртуализации ввода-вывода, созданная корпорацией Intel в дополнение к её технологии виртуализации вычислений (VT), известной под кодовым названием Vanderpool. Виртуализация ввода-вывода позволяет пропускать (pass-through) устройства на шине PCI (и более современных подобных шинах) в гостевую ОС, таким образом, что она может работать с ним с помощью своих штатных средств.

AMD virtualization (AMD-V) AMD разработала свои расширения виртуализации первого поколения под кодовым названием «Pacifica», и первоначально опубликовала их как AMD Secure Virtual Machine (SVM), но позже, на рынке, — под торговой маркой «AMD Virtualization», сокращенно «AMD-V».

Как и у Intel не все процессоры AMD поддерживают виртуализацию.

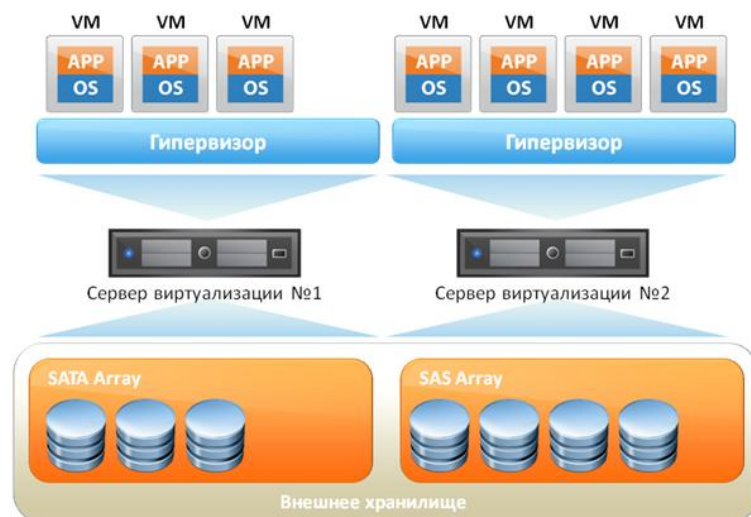


Рисунок 16: Структура виртуализации серверов

Архитектура NetBurst

NetBurst (рабочее название — P68) — суперскалярная гиперконвейерная микроархитектура, разработанная компанией Intel и лежавшая в основе микропроцессоров Pentium 4, Pentium D, Celeron и Xeon. Архитектура NetBurst пришла на смену архитектуре P6 на рынке настольных и серверных процессоров. Она не являлась развитием архитектуры P6, а представляла собой принципиально новую по сравнению со всеми предшественниками архитектуру. Первые процессоры архитектуры NetBurst были анонсированы 20 ноября 2000 года, а 8 августа 2007 года компания Intel объявила о начале действия программы по снятию с производства всех процессоров этой архитектуры.

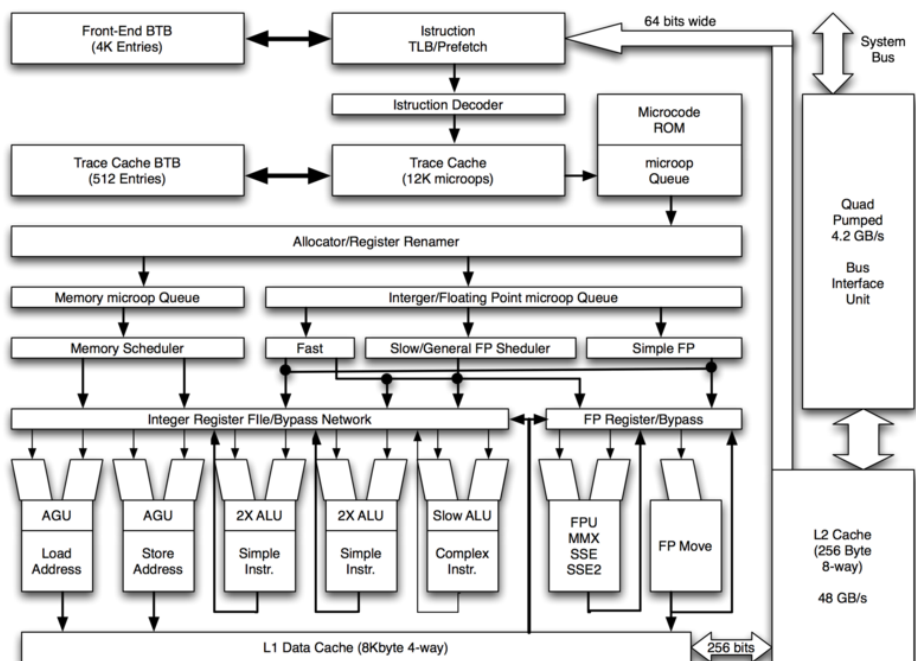


Рисунок 17 Структура процессора PentiumIV Willamette

Архитектура NetBurst разрабатывалась, в первую очередь, с целью достижения высоких тактовых частот процессоров. Характерными особенностями архитектуры NetBurst являются гиперконвейеризация и применение кэша последовательностей микроопераций вместо традиционного кэша инструкций. ALU процессоров архитектуры NetBurst также имеет существенные отличия от ALU процессоров других архитектур.

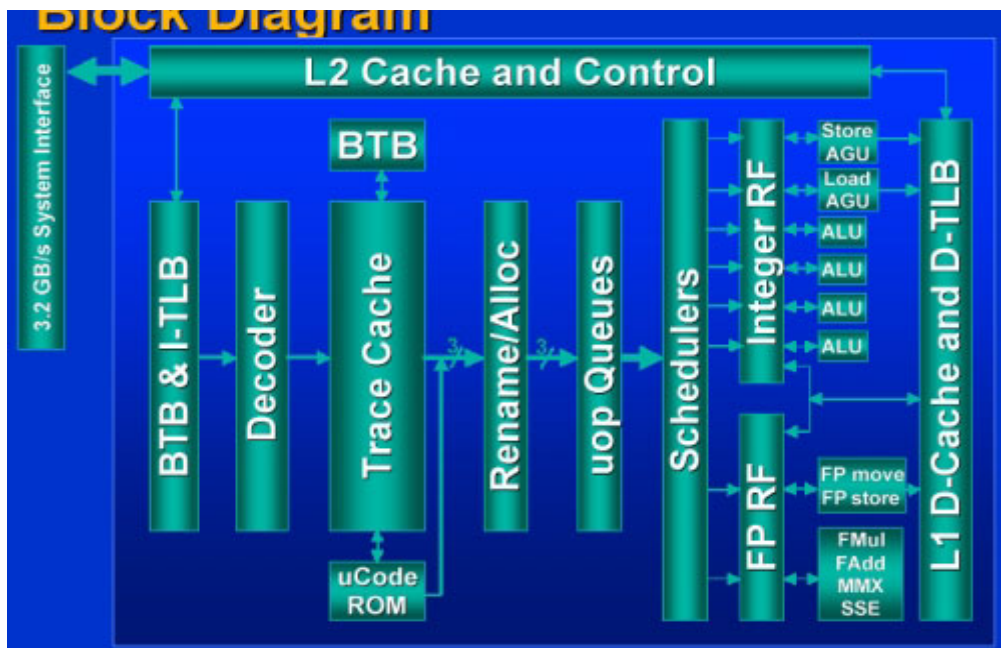


Рисунок 18 Структура ядра NetBurst

Гиперконвейеризация. Процессоры Pentium 4 на ядрах Willamette и Northwood имеют конвейер глубиной 20 стадий, а процессоры на ядрах Prescott и Cedar Mill — 31 стадию. При этом стадии декодирования инструкций не учитываются: в связи с применением кэша последовательностей микроопераций, декодер вынесен за пределы конвейера. Это позволяет процессорам Pentium 4 достигать более высоких тактовых частот по сравнению с процессорами, имеющими более короткий конвейер при одинаковой технологии производства.

Для минимизации влияния неверно предсказанных переходов, в процессорах архитектуры NetBurst используются увеличенный по сравнению с предшественниками буфер предсказания ветвлений (BTB, англ. branch target buffer) и новый алгоритм предсказания ветвлений, что позволило

достичь высокой точности предсказания (около 94%) в процессорах на ядре Willamette. В последующих ядрах механизм предсказания ветвлений подвергался модернизациям, повышавшим точность предсказания.

Кэш последовательностей микроопераций (Execution Trace Cache) - Процессоры архитектуры NetBurst, как и большинство современных x86-совместимых процессоров, являются CISC-процессорами с RISC-ядром: перед исполнением сложные инструкции x86 преобразуются в более простой набор внутренних инструкций (микроопераций), что позволяет повысить скорость обработки команд. Однако, вследствие того, что инструкции x86 имеют переменную длину и не имеют фиксированного формата, их декодирование связано с существенными временными затратами.

В связи с этим, при разработке архитектуры NetBurst было принято решение отказаться от традиционной кэш-памяти инструкций первого уровня, хранящей команды x86, в пользу кэша последовательностей микроопераций, хранящего последовательности микроопераций в соответствии с предполагаемым порядком их исполнения. Емкость trace cache составляла около 12 тыс. микроопераций. Такая организация кэш-памяти позволила также снизить временные затраты на выполнение условных переходов и на выборку инструкций.

АЛУ и механизм ускоренного исполнения целочисленных операций (Rapid Execution Engine)- Так как основной целью разработки архитектуры NetBurst было повышение производительности за счёт достижения высоких тактовых частот, возникла необходимость увеличения темпа выполнения основных целочисленных операций. Для достижения этой цели АЛУ процессоров архитектуры NetBurst разделено на несколько блоков: «медленное АЛУ» (англ. slow ALU), способное выполнять большое количество целочисленных операций, и два «быстрых АЛУ» (англ. 2X ALU), выполняющих только простейшие целочисленные операции (например, сложение). Выполнение операций на «быстрых АЛУ» происходит последовательно в три этапа: сначала вычисляются младшие разряды результата, затем старшие, после чего могут быть получены флаги. «Быстрые АЛУ», обслуживающие их планировщики (англ. Fast scheduler), а также регистровый файл (англ. Integer register file) синхронизируются по половине такта процессора, таким образом, эффективная частота их работы вдвое превышает частоту ядра. Эти блоки образуют механизм ускоренного выполнения целочисленных операций.

Система повторного исполнения микроопераций (англ. Replay System) - Основной задачей планировщиков микроопераций является определение готовности микроопераций к исполнению и передача их на конвейер. Вследствие большого числа стадий конвейера, планировщики вынуждены отправлять микрооперации на исполнительные блоки до того, как завершится выполнение предыдущих микроопераций. Это обеспечивает оптимальную загрузку исполнительных блоков процессора и позволяет избежать потери производительности в том случае, если данные, необходимые для выполнения микрооперации, находятся в кэш-памяти первого уровня, регистровом файле, или могут быть переданы минуя регистровый файл.

Достоинства

Основным достоинством процессоров архитектуры NetBurst является возможность работы на высоких тактовых частотах. Это позволяет достичь высокой производительности в оптимизированных задачах и компенсировать низкую удельную производительность. Кроме того, высокая тактовая частота даёт маркетинговые преимущества: потребители склонны выбирать процессоры с большей тактовой частотой («покупают мегагерцы»). К достоинствам процессоров архитектуры NetBurst можно также отнести высокую пропускную способность памяти.

Поддержка технологии HyperThreading некоторыми процессорами архитектуры NetBurst позволяла поднять производительность в задачах, поддерживающих многопроцессорность, однако существуют некоторые задачи, при выполнении которых производительность может снижаться.

Недостатки

Основными недостатками длинного конвейера являются уменьшение удельной производительности по сравнению с коротким конвейером (за один такт выполняется меньшее количество инструкций), а также серьёзные потери производительности при некорректном выполнении инструкций (например, при неверно предсказанном условном переходе или кэш-промахе). Так, например, процессор Pentium 4 с частотой 1700 МГц в неоптимизированных под архитектуру NetBurst задачах уступал процессорам с частотой 1333 МГц

Кроме того, работа процессоров на высоких частотах была связана с высоким тепловыделением. Несмотря на то, что процессоры на ядре Cedar Mill были способны работать на частотах, превышавших 7 ГГц, с использованием экстремального охлаждения (обычно использовался стакан с жидким азотом),

максимальная тактовая частота серийных процессоров Pentium 4 составила 3800 МГц. При этом типичное тепловыделение превышало 100 Вт, а максимальное — 150 Вт.

Двухъядерные процессоры архитектуры NetBurst для настольных компьютеров (Pentium D) представляли собой два ядра Prescott (процессоры на ядре Smithfield), находящиеся на одном кристалле, или Cedar Mill (Presler), находящиеся в одном корпусе (по сути два отдельных процессора в одном корпусе). Так как процессоры архитектуры NetBurst изначально разрабатывались как одноядерные, обмен данными между ядрами осуществлялся через оперативную память, что приводило к потерям производительности.

Многоядерная архитектура

Многоядерный процессор — центральный процессор, содержащий два и более вычислительных ядра на одном процессорном кристалле или в одном корпусе.

Многоядерные процессоры можно классифицировать по наличию поддержки когерентности кешей между ядрами. Бывают процессоры с такой поддержкой и без нее.

Способ связи между ядрами: Разделяемая шина, Сеть (Mesh) на каналах точка-точка, сеть с коммутатором, общие кеши,

Кеши: Во всех существующих на сегодняшний день многоядерных процессорах кэш-память первого уровня у каждого ядра своя, а кэш 2-го уровня существует в нескольких вариантах:

разделяемый — кэш расположен на одном кристалле с ядрами и доступен каждому из них в полном объеме. Используется в процессорах семейств Intel Core.

индивидуальный — отдельные кэши равного объема, интегрированные в каждое из ядер. Обмен данными из кэшей L2 между ядрами осуществляется через контроллер памяти — интегрированный (Athlon 64 X2, Turion X2, Phenom) или внешний (использовался в Pentium D, в настоящее время Intel отказалась от использования такого подхода).

В приложениях, оптимизированных под многопоточность, наблюдается прирост производительности на двухъядерном процессоре. Однако, если приложение не оптимизировано, то оно не будет получать практически никакой выгоды от дополнительных ядер, а может даже выполняться медленнее, чем на процессоре с меньшим количеством ядер, но большей тактовой частотой. Это в основном старые приложения, либо приложения, не использующие многопоточность.

Архитектура K8

Многие производимые в настоящее время процессоры AMD базируются на микроархитектуре K8, представлена она была 23 сентября 2003 года. По архитектуре K8 производятся современные настольные процессоры Athlon и Sempron, мобильные Turion и серверные Opteron.

Процессоры AMD K8 — 64-разрядные, но прекрасно работают и с 32-разрядными приложениями, демонстрируя при этом заметно более высокое быстродействие, чем их предшественники при тех же тактовых частотах.

Наиболее важными факторами улучшения производительности процессоров с этой технологией, даже на прежних 32-разрядных приложениях являются:

- Поддержка 64-битной адресации памяти при полной совместимости с 32-битными приложениями
- Встроенный контроллер двухканальной памяти. Не только то, что данный контроллер работает на частоте ядра процессора, но также и то, что из связки процессор-память исчезло лишнее звено — северный мост, позволило существенно уменьшить задержки при обращении к ОЗУ.
- Athlon 64 также обладает технологией изменения тактовой частоты процессора, названной Cool'n'Quiet. Если пользователь запускает приложения, не требующие от процессора большой вычислительной мощности, то процессор самостоятельно понижает свою тактовую частоту, а также напряжение питания ядра. Применение данной технологии позволяет снизить

тепловыделение при максимальной нагрузке с 89 Вт до 32 Вт (степпинг C0, частота ядра понижена до 800 МГц), и даже до 22 Вт (степпинг CG, частота ядра снижена до 1 ГГц).

- Увеличенная до 1 мегабайта (на ядро) кэш-память второго уровня (L2)
- Возросла скорость работы с кэш-памятью (до 12,8 Гбит/сек), благодаря вдвое (с 64 до 128 бит) увеличенной разрядности шины
- Поддержка набора инструкций SSE, SSE2, SSE3 и 3DNow!

Двухядерные процессоры, основанные на данном ядре (Windsor), компания AMD выпустила во втором квартале 2006 года. Процессоры, выпущенные на данном ядре, предназначены для Socket AM2 и имеют тип корпуса OmPGA. Оснащены двухканальным контроллером памяти типа DDR2 (предположительно PC2-5300). Частота шины HyperTransport увеличилась до 333 МГц. Выпускаются процессоры по 90 нм техпроцессу. Размер кеша L2 по 1Мб на каждое ядро.

Высокоскоростная шина HyperTransport обеспечивает между процессором и чипсетом поток данных до 8 Гбит/сек, имеется электронная схема для защиты кристалла от перегрева.



Рисунок 19 Структура ядра Windsor (K8)

Каждое ядро имеет отдельную кэш-память L2. Оба ядра обслуживаются единым для двух ядер контроллером памяти и контроллером шины, распределяющим потоки данных между каналами HyperTransport и контроллером памяти.

Архитектура Conroe ¹

Впервые детальные подробности об архитектуре процессоров Core 2 Duo были озвучены на весенней сессии Форума Intel для разработчиков - IDF Spring 2006. Впрочем, названия новых процессорных брендов Intel для настольных ПК - Core 2 Duo и Core 2 Extreme, стали известны несколько позже, в мае 2006.

¹ Intel Core 2 Duo и Intel Core 2 Extreme

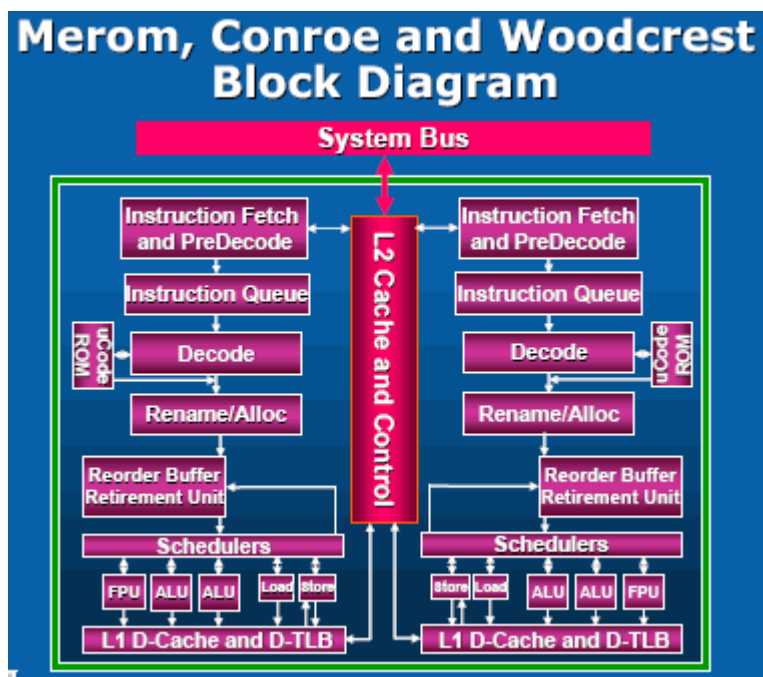


Рисунок 20 Блок - схема ядер Merom, Conroe, Woodcrest

Эта процессорная архитектура наследует философию эффективного энергопотребления, впервые реализованную в процессорах Intel Pentium M для мобильных ПК с рабочим названием Banias. В новом поколении возможности процессоров улучшены не только благодаря совершенно новым технологиям, но также за счет использования наработок, с успехом применявшихся в ядрах с архитектурой Intel NetBurst. И всё же ключевая роль отводится инновациям, впервые реализованным в новом поколении процессорной архитектуры Intel:

- Intel Wide Dynamic Execution - технология выполнения большего количества команд за каждый такт, повышающая эффективность выполнения приложений и сокращающая энергопотребление. Каждое ядро процессора может выполнять до четырех инструкций одновременно с помощью 14-стадийного конвейера
- Intel Intelligent Power Capability - технология, с помощью которой для исполнения задач активируется работа отдельных узлов чипа по мере необходимости, что значительно снижает энергопотребление системы в целом
- Intel Advanced Smart Cache - технология использования общей для всех ядер кэш-памяти L2, что снижает общее энергопотребление и повышает производительность, при этом, по мере необходимости, одно из ядер процессора может использовать весь объём кэш-памяти при динамическом отключении другого ядра
- Intel Smart Memory Access - технология оптимизации работы подсистемы памяти, сокращающая время отклика и повышающая пропускную способность подсистемы памяти
- Intel Advanced Digital Media Boost - технология обработки 128-разрядных команд SSE, SSE2 и SSE3, широко используемых в мультимедийных и графических приложениях, за один такт

Архитектура K10

K10 — поколение архитектуры микропроцессоров x86 компании AMD. Процессоры этой архитектуры появились в продаже в конце 2007 года.

На ее базе построены процессор Phenom для настольных систем, а также Opteron серий 13xx для сокета Socket AM2+. Все процессоры серии Phenom построены на Socket AM2+ обратно совместимом с Socket AM2. При использовании процессоров Phenom на материнских платах с поддержкой Socket AM2 он лишается поддержки шины Hyper-Transport 3.0, отдельного тактования контроллера памяти (северного моста), кэша L3 и ядер, а также некоторых энергосберегающих функций.

В процессорах K10 все ядра выполнены на одном кристалле и снабжены выделенной кэш-памятью L2. В чипах Phenom/Phenom 2 и серверных Opteron также предусмотрена общая для всех ядер кэш-память L3, объём которой составляет от 2 до 6 Мбайт.

Второе важное преимущество K10 – новая системная шина HyperTransport 3.0 с пиковой пропускной способностью до 41,6 Гбайт/с в обоих направлениях в 32-битном режиме или до 10,4 Гбайт/с в одном направлении в 16-битном режиме и частотой до 2,6 ГГц. Напомним, что максимальная рабочая частота предыдущей версии HyperTransport 2.0 составляет 1,4 ГГц, а пиковая пропускная способность – до 22,4 или 5,6 Гбайт/с.

Широкая шина особенно важна для многоядерных процессоров, при этом в HyperTransport 3.0 предусмотрена возможность конфигурации канала, что позволяет предоставить каждому ядру собственную независимую линию. Кроме того, процессор K10 способен динамически изменять ширину и рабочую частоту шины пропорционально собственной частоте.

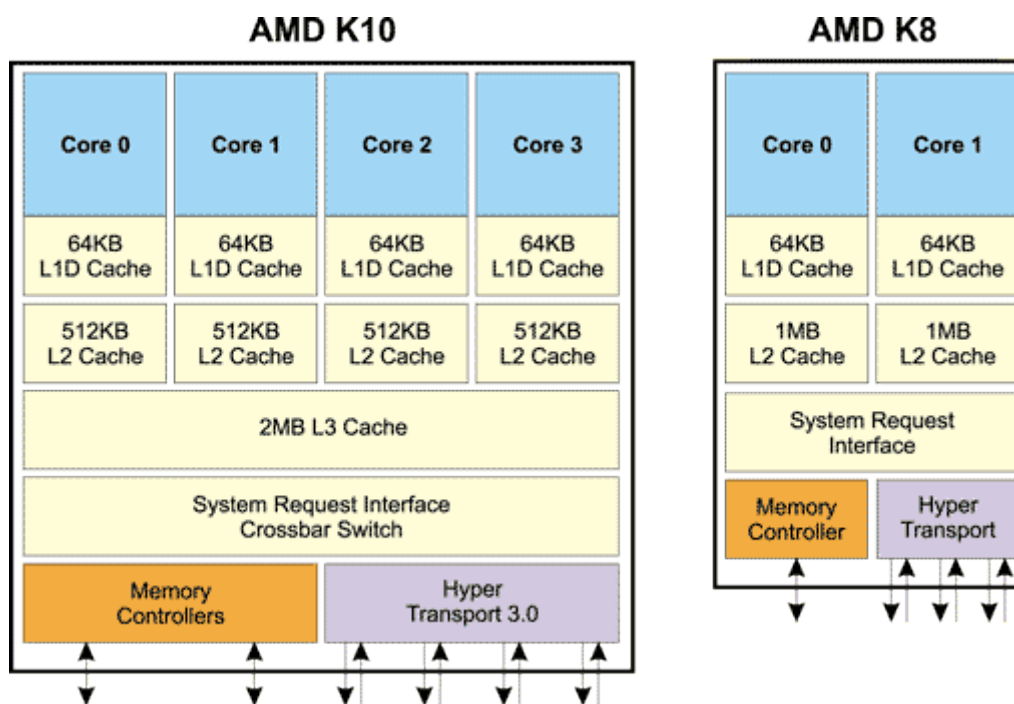


Рисунок 21 Структура ядра процессоров на архитектурах K10 и K8

В процессоры K10 встраиваются два независимых контроллера оперативной памяти, что ускоряет доступ к модулям в реальных условиях эксплуатации. Контроллеры способны работать с памятью *DDR2-1066 (модели для разъёма AM2+ и AM3) или DDR3 (чипы для разъёма AM3)*. Поскольку интегрированный в Phenom II и Athlon II для Socket AM3 контроллер поддерживает оба типа оперативной памяти, а разъём AM3 обратно совместим с AM2+, новые ЦП могут устанавливаться на старые платы для AM2+ и работать с памятью DDR2. Это означает, что при покупке Phenom II для апгрейда вам не придётся сразу же менять и системную плату, а также приобретать оперативную память другого типа – как, например, в случае с чипами Intel i3/i5/i7.

В микропроцессорах с архитектурой K10 реализован целый набор модернизированных технологий энергосбережения – AMD Cool'n'Quiet, CoolCore, Independent Dynamic Core и Dual Dynamic Power Management. Эта сложная система позволяет автоматически снижать энергопотребление всего чипа в режиме простоя, обеспечивает независимое управление питанием контроллера памяти и ядер и способна отключать неиспользуемые элементы процессора.

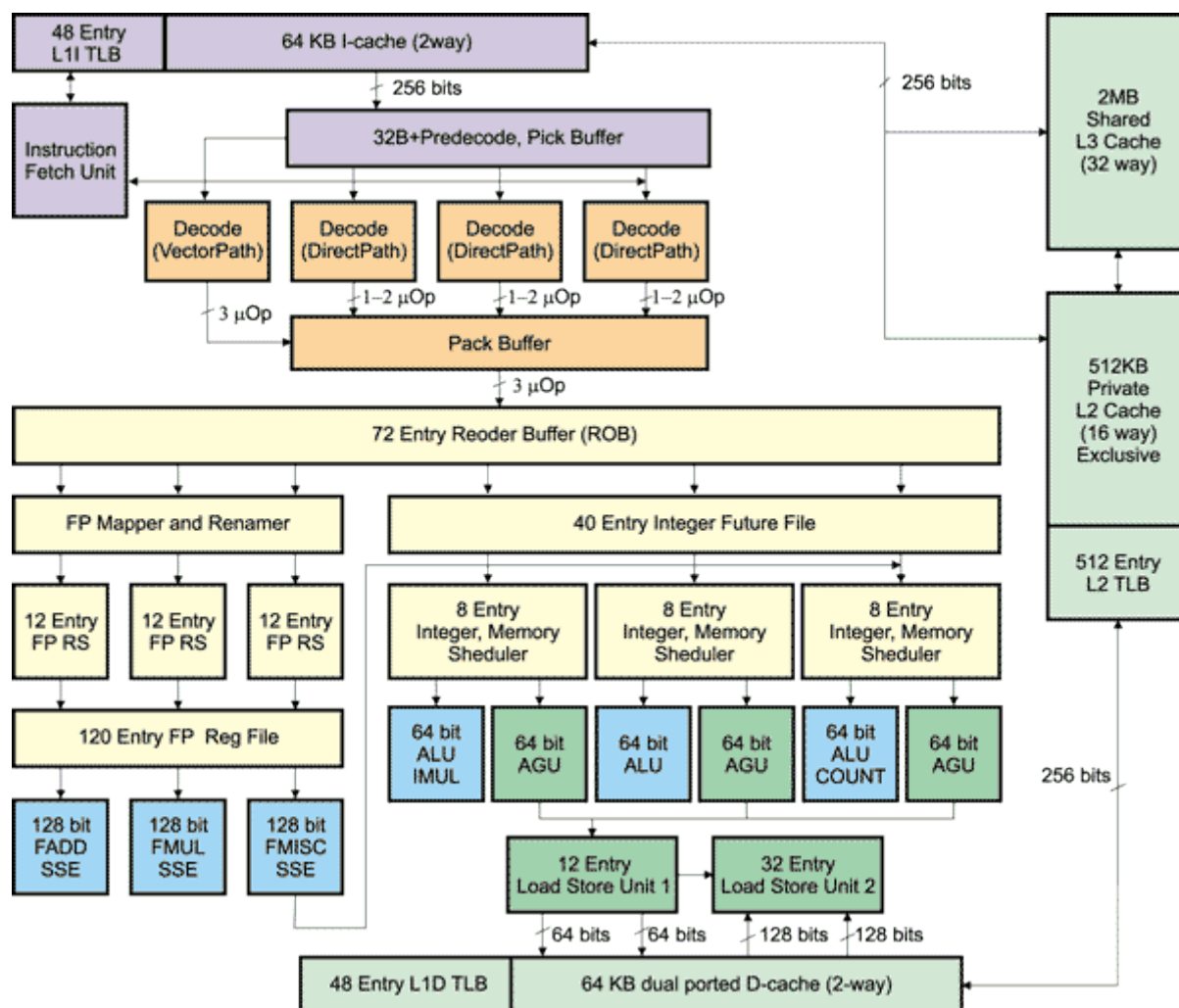


Рисунок 22 Структура одного ядра архитектуры K10

Наконец, сами ядра также были существенно усовершенствованы. Была переработана конструкция блоков выборки, предсказания переходов и ветвлений, диспетчеризации, что позволило оптимизировать загрузку ядра и в конечном итоге повысить производительность. Разрядность блоков SSE была увеличена с 64 до 128 бит, появилась возможность выполнять 64-разрядные инструкции как одну, была добавлена поддержка двух дополнительных инструкций SSE4a (не путать с наборами инструкций SSE4.1 и 4.2 в процессорах Intel Core).

Архитектура Intel Core i7 (Nehalem)

Инженеры Intel хотели создать набор базовых "кирпичей, которые можно собирать как блоки конструктора Lego, чтобы создавать разные версии архитектуры.

Впрочем, в то же время вполне реально рассмотреть флагмана новой архитектуры - high-end версию, которая будет использоваться в серверах и производительных рабочих станциях. На первый взгляд, спецификации очень напоминают архитектуру Barcelona (K10) от AMD. То есть это "родной" четырёхядерный процессор, который использует три уровня кэш-памяти, встроенный контроллер памяти, а также высокопроизводительную систему интерфейсов "точка-точка" для связи с периферией и другими CPU в многопроцессорной конфигурации. Это доказывает, что технологические решения AMD были не такими плохими, но пострадали из-за технической реализации, которая не была достаточно хорошо адаптирована для текущего дизайна.

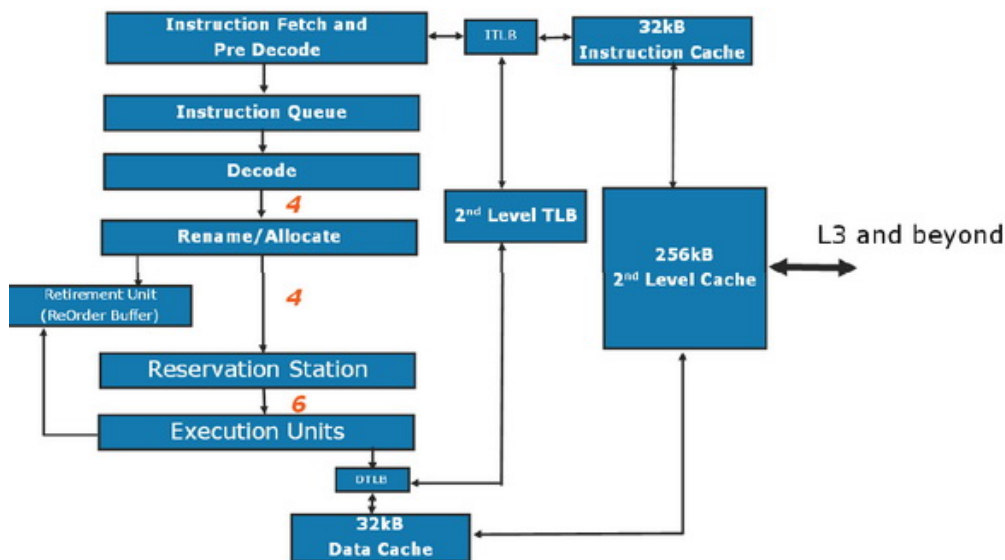
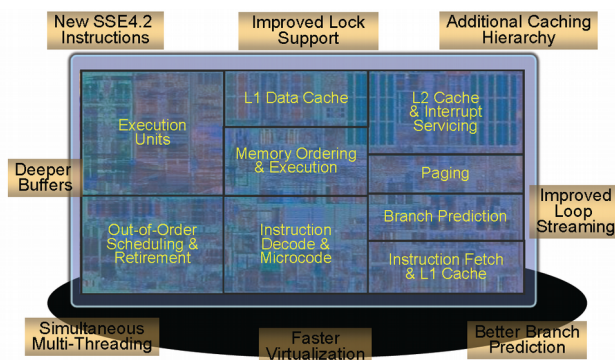


Рисунок 23 Архитектура ядра Nehalem

Но Intel не только обновила свою архитектуру, взяв у конкурента интересные инновации. С бюджетом более 700 млн. транзисторов (731 млн., если быть точным), инженеры смогли серьёзно улучшить основные характеристики исполнительного ядра, добавив в то же время новую функциональность. Например, поддержка многопоточности (simultaneous multi-threading, SMT), которая впервые появилась в Pentium 4 "Northwood" под названием Hyper-Threading, вновь вернулась. Поскольку физических ядер на кристалле четыре, некоторые версии Nehalem, которые используют два ядра в одной упаковке, смогут выполнять до 16 потоков одновременно. Изменения на первый взгляд кажутся простыми, но, как мы увидим позже, они привели к серьёзным переменам на разных уровнях конвейера; многие буферы пришлось изменить, чтобы данный режим не влиял на производительность. Как уже случалось с каждой последней архитектурой на протяжении последних лет, Intel добавила к Nehalem новые инструкции SSE. Архитектура поддерживает набор SSE 4.2, некоторые компоненты которого были, как может показаться, взяты из микроархитектуры AMD K10.

Intel реализовала довольно оригинальную технологию под названием режима Turbo. Если процессор работает ниже стандартного теплового пакета, например, режим Turbo повышает частоту используемых ядер, вместе с тем оставаясь в пределах теплового пакета.

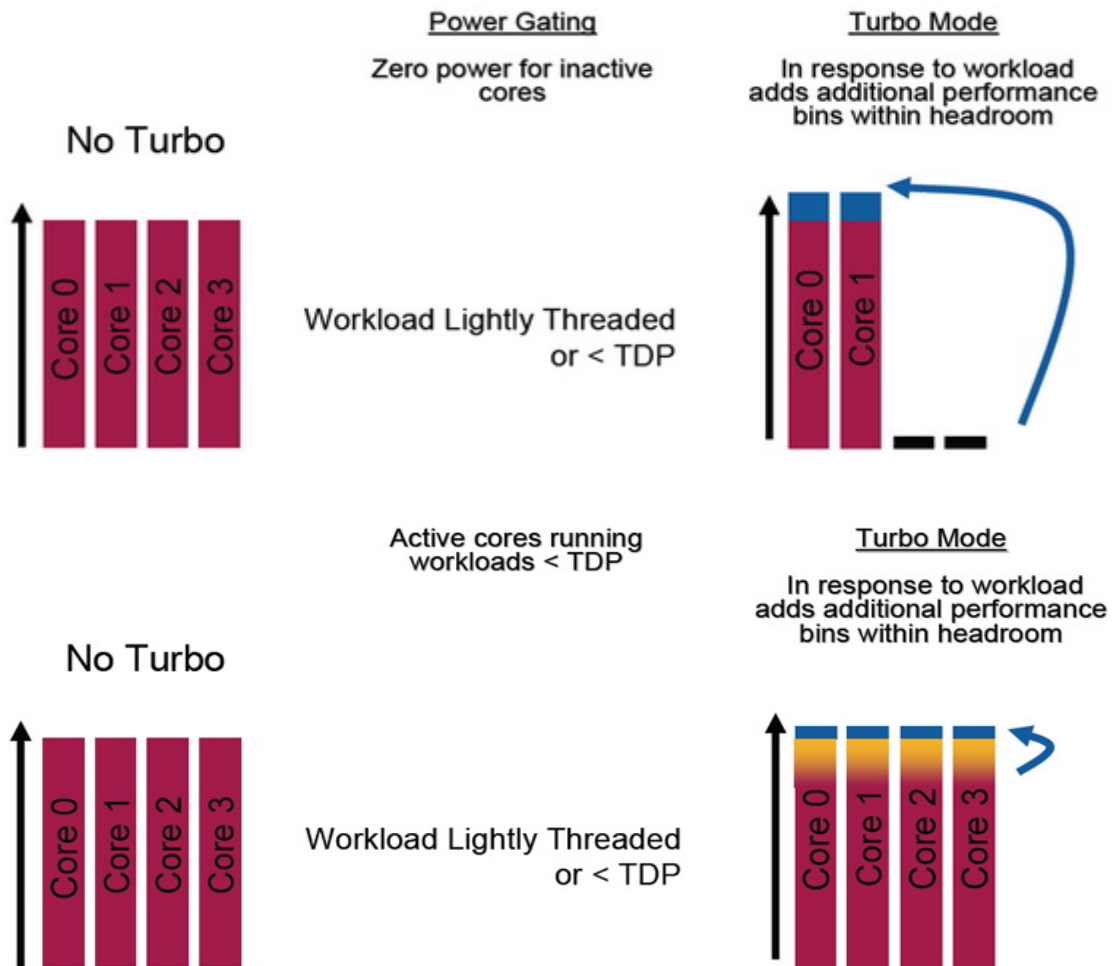


Рисунок 24 Turbo Boost

Хотя архитектура Core была очень эффективной, некоторые детали дизайна начали показывать свой возраст, и первой среди них является шина Front Side Bus (FSB).

Решение, выбранное Intel под названием QuickPath Interconnect (QPI), не является чем-то новым; оно представляет собой встроенный контроллер памяти и очень быструю последовательную шину "точка-точка". Подобная технология была представлена пять лет назад в процессорах AMD, но на самом деле она ещё старше.

С технической точки зрения интерфейс QPI является двунаправленным с двумя 20-битными шинами, по одной на каждое направление, из которых 16 зарезервировано под данные, а оставшиеся четыре - под функции исправления ошибок или служебную информацию протокола. Это даёт максимальную скорость 6,4 GT/s (млрд. передач в секунду) или полезную пропускную способность 12,8 Гбайт/с, как на чтение, так и на передачу. Для сравнения, FSB на самых современных процессорах Intel работает с максимальной тактовой частотой 400 МГц, при этом для передачи адресов требуется два такта (200 MT/s), а данные передаются в режиме QDR, с пропускной способностью 1,6 GT/s. При 64-битной ширине FSB даёт суммарную пропускную способность 12,8 Гбайт/с, но она доступна только для чтения или записи.

Поэтому интерфейс QPI даёт пропускную способность вплоть до двух раз выше, если запись и чтение сбалансированы должным образом. В теоретической ситуации, когда есть только операции чтения или записи, пропускная способность будет идентична FSB. Но следует помнить, что шина FSB использовалась как для доступа к памяти, так и для передачи всех данных на периферию или между процессорами. В случае Nehalem, интерфейс QPI исключительно предназначается для передачи данных на периферию, а за работу с памятью отвечает интегрированный в процессор контроллер. Связь между несколькими CPU в многосокетной конфигурации осуществляется ещё одним интерфейсом QPI. Даже в самой тяжёлой ситуации QPI должен показать лучшую производительность, чем FSB.

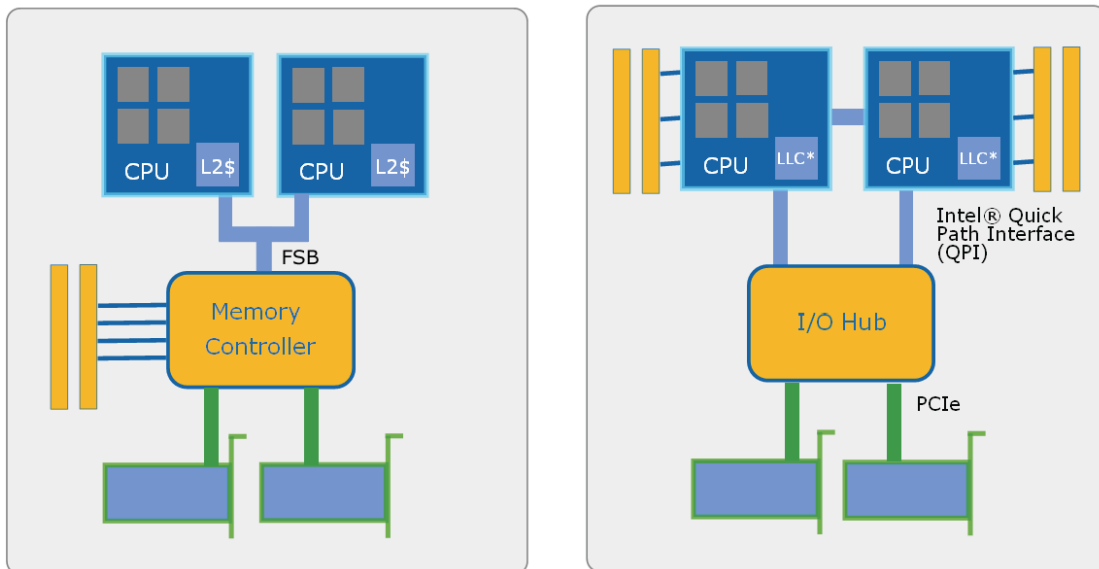


Рисунок 25 Шины FSB и QPI

Nehalem был разработан с учётом гибкой и масштабируемой архитектуры, поэтому число доступных интерфейсов QPI меняется в зависимости от ориентации на тот или иной сегмент рынка - от одного интерфейса для связи с чипсетом в односокетных конфигурациях до целых четырёх для четырёхсокетных серверов. Это позволяет создавать полносвязные четырёхпроцессорные системы, когда каждый процессор может получать доступ к любой области памяти через один хоп QPI, поскольку каждый процессор напрямую подключён к трём остальным.

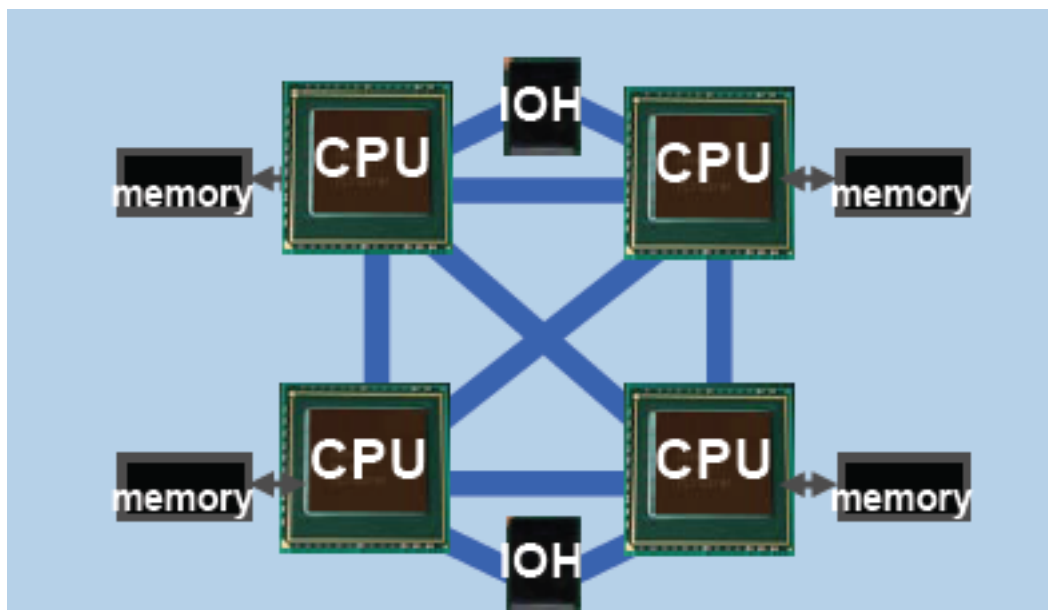


Рисунок 26 Структура четырёхядерной системы на базе Nehalem

Архитектура Nehalem использует встроенный контролер памяти.

Local Memory Access

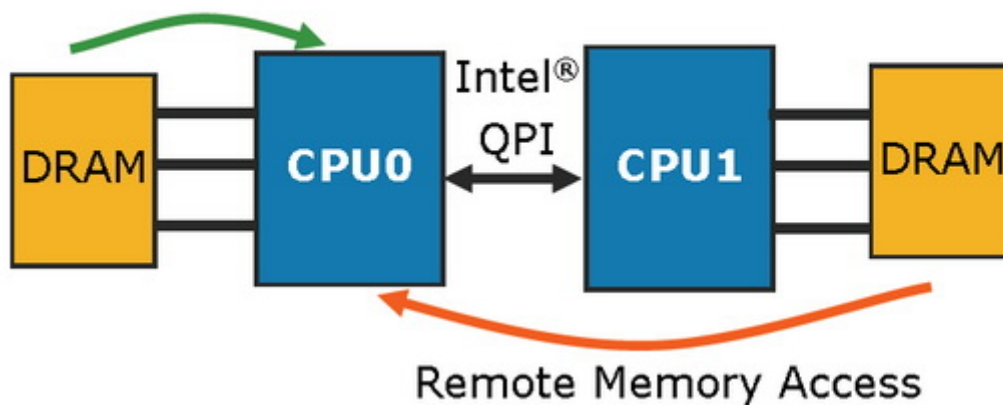


Рисунок 27 Интеграция контроллера памяти

Иерархия памяти в Core2 была очень простой; Intel сконцентрировалась на производительности общего кэша L2, который стал лучшим решением для архитектуры, которая нацеливалась, главным образом, на двухядерные конфигурации. Но в случае с Nehalem инженеры начали с нуля и пришли к такому же заключению, что и конкуренты: общий кэш L2 не очень хорошо подходит для "родной" четырехъядерной архитектуры. Они оснастили каждое ядро собственным кэшем L2. Поскольку он выделен на каждое ядро и относительно мал (256 кбайт), получилось обеспечить кэш очень высокой производительностью; в частности, задержки существенно улучшились по сравнению с Penryn - с 15 тактов до, примерно, 10 тактов.

Затем есть огромная кэш-память третьего уровня (8 Мбайт), отвечающая за связь между ядрами. На первый взгляд архитектура кэша Nehalem напоминает Barcelona, но работа кэша третьего уровня очень отличается от AMD - она инклюзивная для всех нижних уровней иерархии кэша. Это означает, что если ядро попытается получить доступ к данным, и они отсутствуют в кэше L3, то нет необходимости искать данные в собственных кэшах других ядер - там их нет. Напротив, если данные присутствуют, четыре бита, связанные с каждой строчкой кэш-памяти (один бит на ядро) показывают, могут ли данные потенциально присутствовать (потенциально, но без гарантии) в нижнем кэше другого ядра, и если да, то в каком.

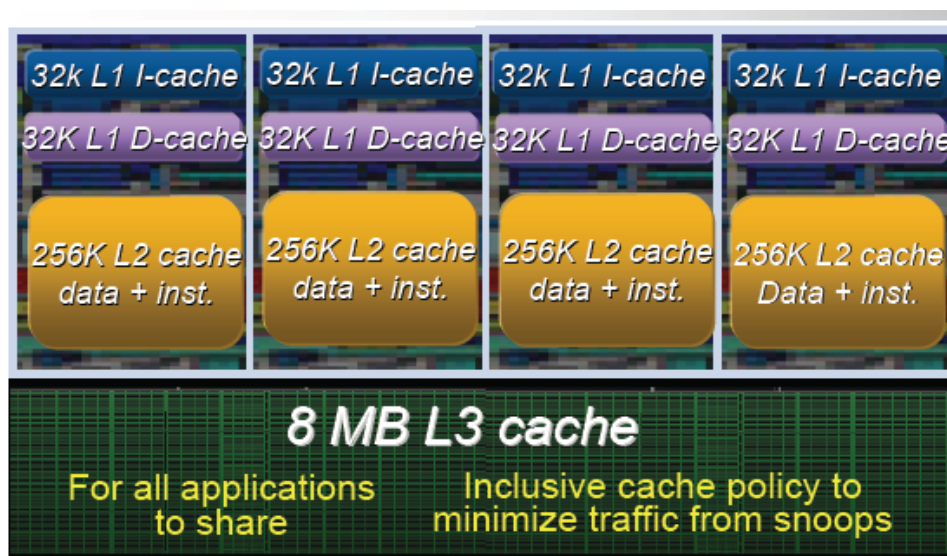


Рисунок 28 Структура памяти в процессорах Nehalem

Эта техника весьма эффективна для обеспечения когерентности персональных кэшей каждого ядра, поскольку она уменьшает потребность в обмене информацией между ядрами. Есть, конечно, недостаток в виде потери части кэш-памяти на данные, присутствующие в кэшах других уровней. Впрочем, не всё так страшно, поскольку кэши L1 и L2 относительно маленькие по сравнению с кэшем

L3 - все данные кэшей L1 и L2 занимают, максимум, 1,25 Мбайт в кэше L3 из доступных 8 Мбайт. Как и в случае Barcelona, кэш третьего уровня работает на других частотах по сравнению с самим чипом. Следовательно, задержка доступа на данном уровне может меняться, но она должна составлять около 40 тактов.

Некоторые модели процессоров на базе архитектуры Nehalem имеют не только интегрированный контроллер памяти, но и интегрированный графический контроллер.

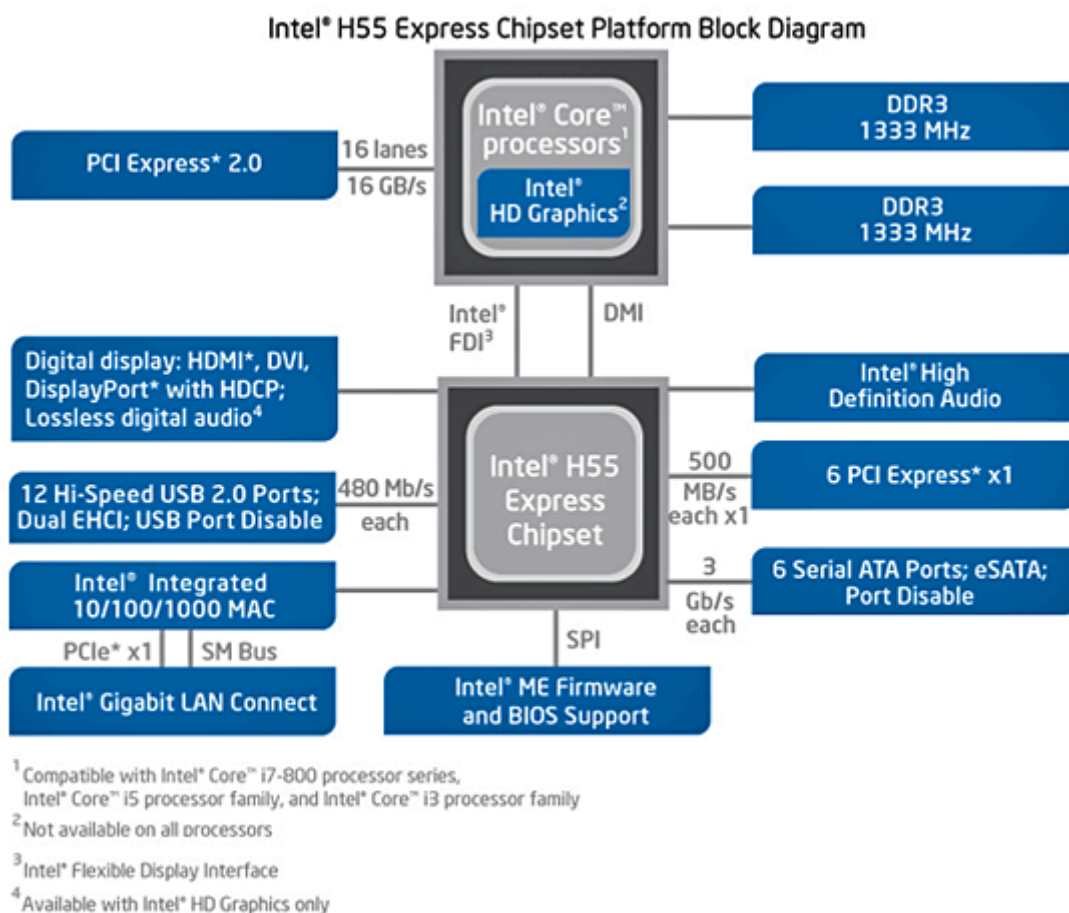


Рисунок 29 Структура системы с процессором на базе архитектуры Nehalem

Вопросы для контроля

Обязательный уровень

1. Что такое - центральный процессор? В чем его функции?
2. Перечислите главные характеристики процессора
3. Что такое тактовая частота? Как она влияет на производительность процессора?
4. Что такое TDP? Как он влияет на работу процессора?
5. Что такое техпроцесс? Как он влияет на работу процессора?
6. Перечислите характеристики системной шины процессора и опишите их влияние на работу процессора.
7. Опишите шину FSB.
8. Опишите шину HyperTransport Ее характеристики.
9. Опишите шину Intel QuickPath Interconnect. Ее характеристики.
10. Перечислите по порядку этапы цикла выполнения.
11. Опишите реализации архитектуры процессоров x86 и x86-64
12. Как организована конвейерная архитектура?
13. Какие трудности возникают при реализации конвейера? Как их решают?
14. Что такое Branch Prediction Unit? Как он работает?
15. Как организована суперскалярная архитектура?
16. Зачем нужен математический сопроцессор?
17. Зачем нужен КЭШ? Что такое уровни КЭША?
18. Что такое «набор расширений»? В чем его роль?
19. Какие технологические решения были использованы в NetBurst
20. Чем характерна архитектура многоядерного процессора?
21. Какие технологические решения были использованы в K8?
22. Какие технологические решения были использованы в Conroe?
23. Какие технологические решения были использованы в K10?
24. Какие технологические решения были использованы в Nehalem?

Источники

<http://ru.wikipedia.org/>
<http://www.notebookcheck-ru.com/Reiting-mobilnykh-processorov-sravnenie-proizvoditelnosti.14099.0.html>
<http://iproc.ru/parallel-programming/lection-7/>
<http://www.what-is.ru/hard/mainb2.shtml>
<http://www.compress.ru/article.aspx?id=18196&iid=842>
<http://www.fcenter.ru/online.shtml?articles/hardware/processors/22080>
http://www.thg.ru/cpu/intel_nehalem_theory/index.html