Node Hardware

- Improved microprocessor performance means availability of desktop PCs with performance of workstations (and of supercomputers of 10 years ago) at significanty lower cost
- Parallel supercomputers are now equipped with COTS components, especially microprocessors
- Increasing usage of SMP nodes with two to four processors
- The average number of transistors on a chip is growing by about 40% per annum
- The clock frequency growth rate is about 30% per annum

Three Basic Operations

- Instruction execution
  - Involves only CPU and registers
- Register loading
  - Load data from cache or memory into registers
  - Involves CPU, front-side bus, cache, memory
- Peripheral usage
  - Copying data through I/O bus from peripheral to memory
  - Involves peripheral, I/O bus, interface from I/O bus into peripheral and memory, memory

Processor

- Binary encoding determined by Instruction Set Architecture (ISA)
- Processors can share part of ISA but not have identical ISAs due to addition of features (instructions such as Streaming SIMD Extensions)
  - SSE, SSE2, SSE3 are numerical instructions for PIII and P4
- Processor clock rate in MHz or GHz is number of clock ticks per second (up to 3.8GHz in 2005 or 4.2GHz)
  - CPUs with different clock rates can perform equivalently
  - CPUs with same rate can perform differently
- Instructions per second / Floating point instructions per second (fps) depend also on ISA, and components on chip

Processor

- Cache mitigates the effect of much slower memory
- CPUs can have cache from kilobytes to 4 to 8 megabytes
- Cache can be of different speeds (L1, L2, L3)
  - Faster is more expensive, therefore less used

Processors

- Intel IA32 (x86) Processors
  - Pentium 3, Pentium4, *Pentium Pro* and Pentium Xeon
  - Athlon, *AMD x86, Cyrix x86*, etc.
- Digital Alpha 21364
  - Alpha 21364 processor integrates processing, memory controller, network interface into a single chip
- IBM PowerPC G5
- IA64 (Itanium, Athlon 64. Opteron)
- Sun SPARC
- SGI MIPS
- HP PA-RISC
- Berkeley Intelligent RAM (IRAM) integrates processor and DRAM onto a single chip

IA32

- 32 bit instruction set
- Binary compatibility specification
  - Hardware may be very different but instruction set is the same
  - Pentium III, 4 and Athlon
- Additions to ISA include SSE, SSE2 and SSE3 (streaming SIMD extensions)
  - Can substantially increase performance
  - Important to consider
- Hyperthreading : multiple threads per CPU
  - Negatively impacts performance
  - Can be turned off

IA32

- Pentium III
  - Has SSE and L2 cache on chip
  - Can be used in 2 CPU SMPs
  - Xeon can be used in 4 CPU SMPs
- Pentium 4
  - Designed for higher clock cycles, but less computing power per cycle
  - Also has SSE2 (SSE3 since Prescott) and Hyperthreading
- Athlon
  - Processor architecture like PIII, bus like Compaq Alpha
  - Two 64KB L1 caches and one 256 KB L2 cache
  - Has SSE but not SSE2
  - Can be used in 2 CPU SMPs

HP/Compaq/DEC Alpha 21264

- True 64 bit architecture
- RISC (Reduced Instruction Set Computer)
  - Simple instructions at high clock rate
- Fastest for a long time
- Used in Cray T3D and T3E
- Popular in early and large clusters due to superior fp performance e.g. Los Alamos NL ASCI Q

Power PC G5

- IBM and Apple Mac
- 64 bit CPU running at over 2GHz (2003), 2.5GHz (2005)
- Up to 1.25GHz front-side bus
- Multiple functional units

IA64 Itanium

- New IS, cache design, fp processor
- Clock rates up to 1.6 GHz (2004), multiway fp instruction issue
- Aimed at 1 to 2 Gflops performance
  - HPServer rx 4610, 800 Mhz Itanium SPecfp2000 of 701
  - HP rx2600, 1.5 GHz I2, SPecfp2000 of 2119
  - I2 is significantly faster
- Both need efficient compilers to exploit EPIC (Explicitly Parallel Instruction Computing)

AMD Opteron

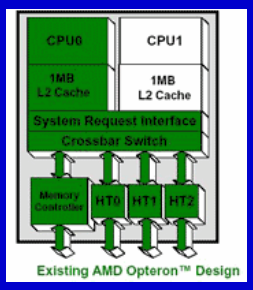
- Supports IA32 and IA64 ISA
- Can run legacy 32 bit codes
- Can access in excess of 4GB memory with new 64 bit instructions
- Integrated DDR memory controller
- Up to 3 high-performance "Hypertransport" interconnects with 6.4GB/sec bandwidth per CPU
- Early Opterons had SPECfp2000 of 1154
- Can have 2 CPU SMPs each with separate memory busses
- More popular than I2 for clusters (cost-performance)

Athlon 64

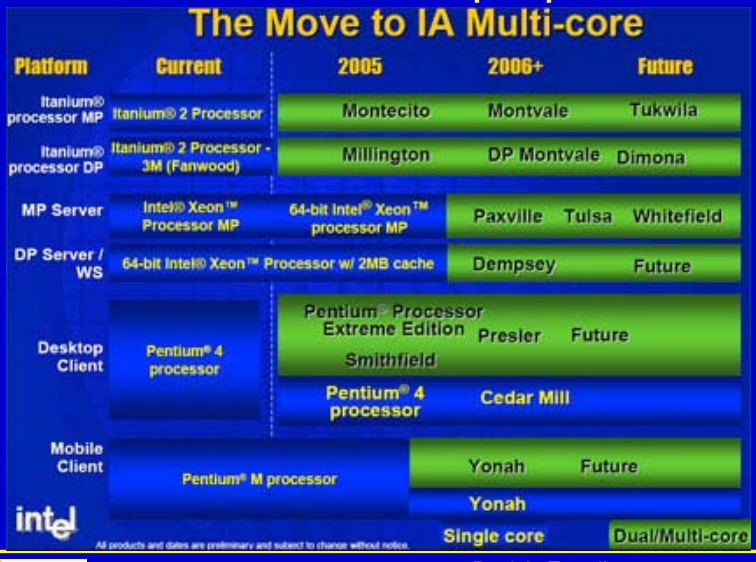
- HyperTransport™ technology with up to 14.4 GB/sec total processor-to-system bandwidth
- integrated DDR memory controller
- Supports SSE2
- Later versions (since Venice Stepping E3 and San Diego Stepping E4) support SSE3

Dual Core Processors

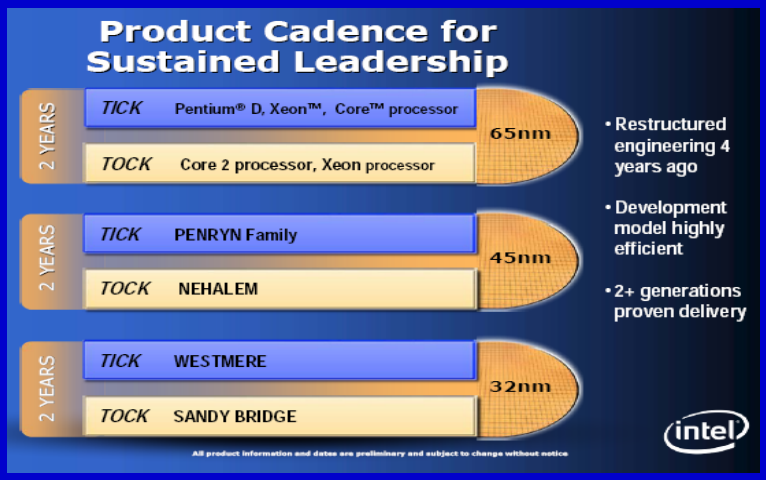
- Opteron
  - for servers by mid-2005
  - for clients late-2005
  - In laptops in 2006
- One die with 2 CPU cores, each core has its own 1MB L2 cache
- Drops into existing AMD Opteron 940-pin sockets
- 2 CPU cores share the same memory and HyperTransport™ technology resources found in single core AMD Opteron processors
- Followed by P4 EE (4/05), Athlon 64, laptops



Intel Dual Core Roadmap April 2005



Intel Tick-tock development cycle



Tick = refresh ; tock = new architecture

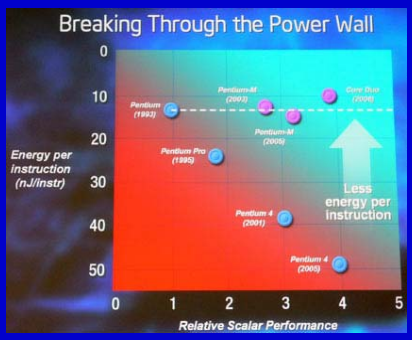
Dual Core Pentium

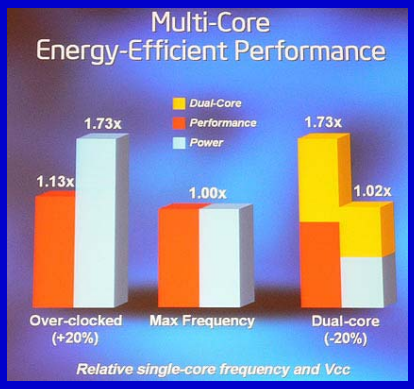
- Dec 2005 Intel Pentium Extreme Edition 840 at 3.2GHz
  - Multiple cores in single processor
  - Can use hyperthreading to give 4 threads
  - Aimed at reducing power consumption
  - 65nm technology
  - 2MB L2 cache

Intel Core MicroArchitecture

- July 2006 Core 2 Duo
  - New Micro Architecture
    - http://www.intel.com/technology/architecture/coremicro/
  - Unified (shared) L2 cache (2MB or 4MB)
  - **Wide Dynamic Execution**
  - **Advanced Smart Cache**
  - **Smart Memory Access**
  - **Advanced Prefetch**
  - **Memory Disambiguation**
  - **Advanced Digital Media Boost**
  - **Intelligent Power Capability**
  - **Presentation:**
  - http://www.intel.com/technology/architecture/coremicro/demo/demo.htm

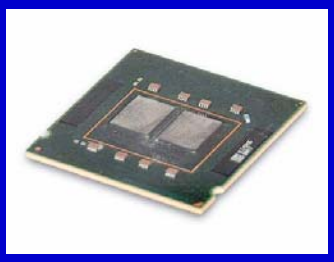
Energy Issue





Intel Quad Core

- **Nov 2006 : Core 2 Quad Core - Kentsfield**
  - **two Core 2 Duo E6700 cores in one socket package**
- **Dec 2006: Xeon Quad Core Clovertown**
  - **two dual-core Woodcrest-type Xeons in a single processor package**
  - 2x 4 MB L2 cache and FSB1066 or FSB1333



Nehalem

- Intel's next-generation microarchitecture
- hafnium-based Intel® 45nm hi-k metal gate silicon technology
- Intel® QuickPath technology
- **Dynamic scalability**, managed cores, threads, cache, interfaces, and power for energy-efficient performance on demand.
- **Design and performance scalability** with support for 2-8+ cores and up to 16+ threads with simultaneous multi-threading (SMT), and scalable cache sizes, system interconnects, and integrated memory controllers.
- **Simultaneous multi-threading** brings high-performance applications into mainstream computing with 1-16+ threads optimized for a new generation multi-core processor architecture.
- **Scalable shared memory** of Intel QuickPath technology features memory distributed to each processor with integrated memory controllers and high-speed point-to-point interconnects to unleash the performance of next-generation Intel® multi-core processors.
- **Multi-level shared cache** improves performance and efficiency by reducing latency to frequently used data.

Intel® QuickPath Technology

- Each processor core will feature an integrated memory controller and high-speed interconnect, linking processors and other components to deliver:
  - **Dynamically scalable interconnect bandwidth** designed to set loose the full performance of Nehalem, Tukwila, and future generations of Intel® multi-core processors
  - **Outstanding memory performance and flexibility** to support leading memory technologies
  - **Tightly integrated interconnect reliability, availability, and serviceability (RAS)** with design-scalable configurations for optimal balance of price, performance, and energy efficiency
- http://www.intel.com/technology/quickpath/index.htm?iid=tech_arch_nextgen+body_quickpath_bullet

Streaming SIMD Extensions 4

- 54 instructions divided into two major categories: Vectorizing Compiler and Media Accelerators, and Efficient Accelerated String and Text Processing.
  - **Vectorizing Compiler and Media Accelerators** will benefit audio, video, and image editing applications, video encoders, 3-D applications, and games.
  - **Efficient Accelerated String and Text Processing** will benefit database and data mining applications, and those that utilize parsing, search, and pattern matching algorithms like virus scanners and compilers.
- 47 in 45nm Intel® Core™ microarchitecture – rest in Nehalem in 2008.
- http://www.intel.com/technology/architecture-silicon/sse4-instructions/index.htm?iid=tech_arch+body_sse4

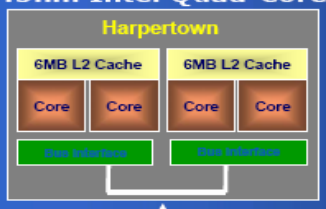
Harpertown 2007

- refresh to the quad-core Xeon
- *Harpertown* will launch with clock speeds all the way up to 3.16GHz, and will also ship two low voltage parts (2.3GHz and 2.6GHz).
- The rumor mill speculates that Intel may be able to reach 3.4GHz with the new 45nm process shrink.
- **45nm**
- **1600MHz FSB**
- **12MB L2**
- **New SSE4 Instructions**
- **High-K Process Technology**

In order to extend Moore's Law, Intel uses a new material in their transistors which is a combination of high-k gate dielectrics and metal gates. This new technology increases the switching speed of the transistors and helps reduce power consumption to allow Intel to continue to deliver faster processors that consume less power.

2nd Generation Intel Quad-Core

Mainstream 45nm Intel Quad-Core



1600/1333 MHz FSB[1], 12M L2 (2x6M)
80W mainstream performance
120W accelerated performance
50W ultra-dense
Intel® 64, Intel® VT, FBD, EIST, Execute Disable
New SSE4 instructions

- **2P Server and Workstations**
- **Improved Performance**
  - 45nm manufacturing
  - Higher core speeds
  - Larger caches
  - 1600 / 1333 MHz[1]
- **Energy Efficiency**
  - Intel Core Microarchitecture
  - Consistent power levels
- **Ease of IT Qualification**
  - Socket, Software and Technology compatible with Quad-Core Intel Xeon 5300 and Dual-Core Intel Xeon 5100 Series

More Performance, Improved Energy Efficiency

1 1600MHz FSB only available with new Stoakley platform, Seaburg chipset
A Dual-Core 45nm version (codename 'Wolfdale-DP') also planned for availability early 2008

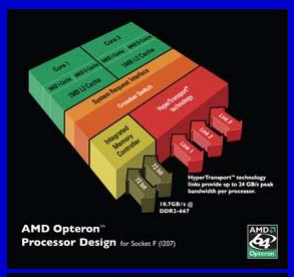
intel

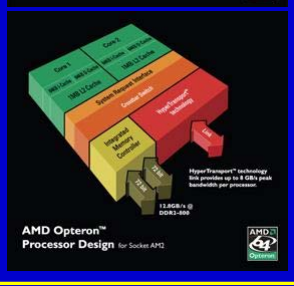
AMD and Intel Differences

- Shared L2 cache v separate caches
- Shared memory v separate memory
- AMD processors : built-in DDR or DDR2 memory controller
- Intel : memory controller, which is part of the platform core logic
  - data path is still shared by all available cores

Next Generation Opteron

- Direct Connect Architecture
  - •HyperTransport™ technology provides a scalable bandwidth interconnect between processors, I/O subsystems, and other chipsets, with up to three coherent HyperTransport technology links providing up to 24.0 GB/s peak bandwidth per processor
- AMD Virtualization™ (AMD-V™)
- energy efficient DDR2 memory
  - Integrated Memory Controller on-die DDR2 DRAM memory controller
- Up to 8 processor (16 core) systems
- quad-core upgradeability in 2007





AMD Barcelona plan

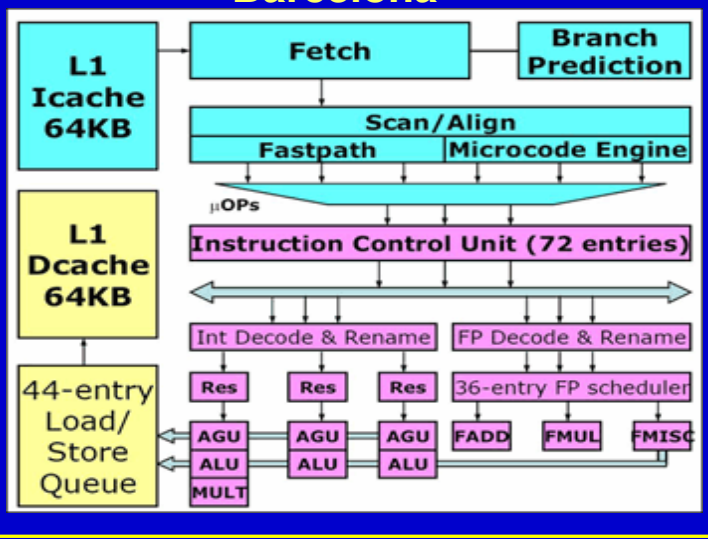
- Quad core Opteron

	Current AMD CPUs	Barcelona
SSE execution width	64 bits wide	128 bits wide
Instruction fetch bandwidth	16 bytes/cycle	32 bytes/cycle
Data cache bandwidth	2 x 64 bit loads/cycle	2 x 128 bits loads/cycle
L2 cache/memory controller bandwidth	64 bits/cycle	128 bits/cycle
Floating-point scheduler depth	36 dedicated x 64-bit ops	36 dedicated x 128-bit ops

Instructions per Cycle in AMD's Barcelona

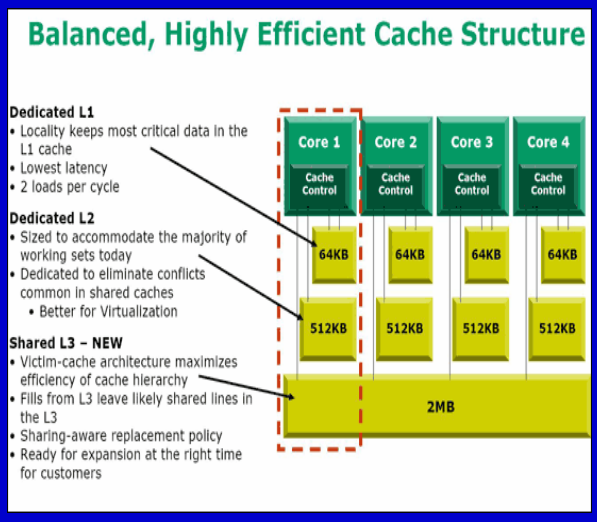
- Enhancements
- **Advanced branch prediction.** AMD's architects have doubled the return stack size, added more branch history bits, and built in a 512-entry indirect branch predictor.
- **32-byte instruction fetch.** Increases efficiency by reducing split-fetch instruction cases.
- **Sideband stack optimizer.** Adjustments to the stack don't take up functional unit bandwidth.
- **Out-of-order load execution.** Load instructions can actually bypass other loads in some cases, as well as stores that are not dependent on the load in question. This minimizes the effect of L2 cache latency

Instructions per Cycle in AMD's Barcelona



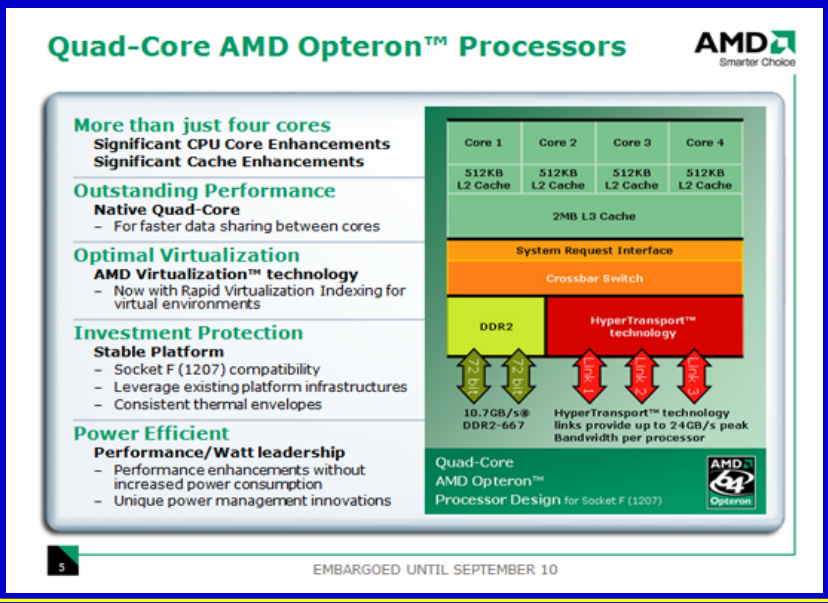
Memory Bandwidth

- •Independent memory controllers
- • full 48-bit hardware addressing
- • 1GB memory page size in addition to the common 4KB and 2MB page sizes.
- •L1 cache is 64KB, the L2 cache is 512KB and the L3 cache is 2MB
- • virtualized address translation, instead of the current shadow paging



Power Management in AMD's Barcelona

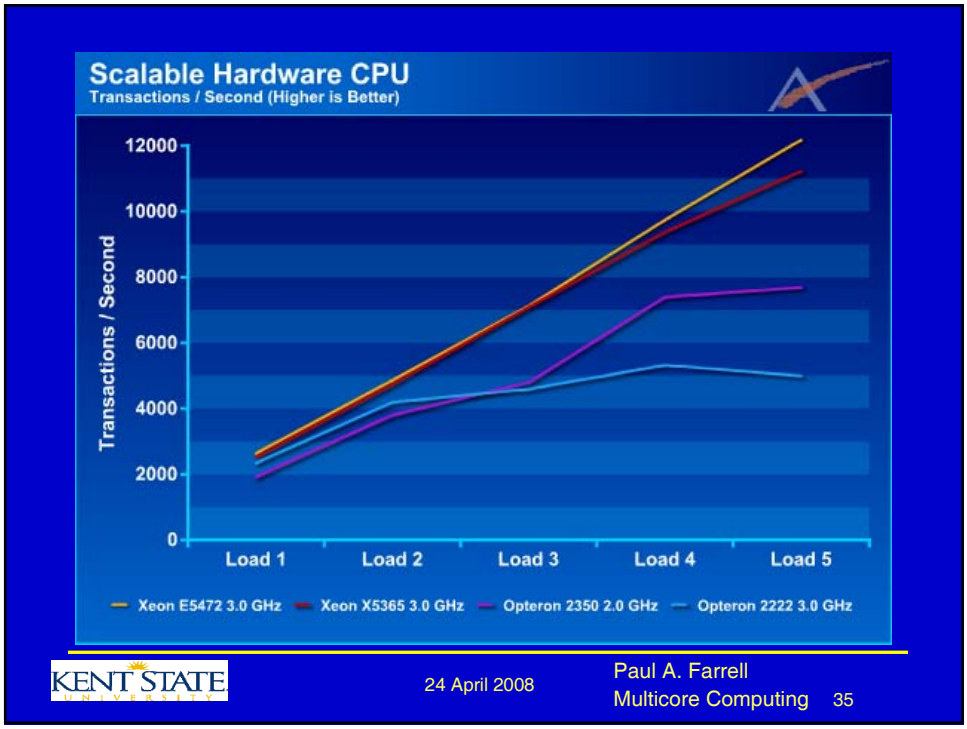
- AMD's 65nm SOI process, which allows for lower voltage and thermal output while increasing circuit densities
- separate CPU core and memory controller power
- enhanced version of AMD's PowerNow
  - will allow individual core frequencies to lower while other cores may be running full bore
- two-socket system, plus memory, plus the chipset, is estimated at about 240W (not counting graphics and storage.)
- shipping by mid-2007



AMD Quad-Core Opteron (Barcelona)

- first true (x86) quad-core processor that features one die with four cores
- based on a 65nm fabrication process
- several micro-architectural changes
  - **Independent Dynamic Core Technology** : can alter the frequency of each core individually.
  - **AMD CoolCore Technology** : can cut power to unused transistor areas to reduce power consumption and lower heat generation.
  - **SSE128 :** can now execute 128-bit SSE operations in a single clock cycle.
  - **2MB L3 Shared Cache :** all four cores share The cache breakdown for the new *Barcelona* is as follows: 64KB L1 cache (64K each for data and instructions) and 512KB L2 cache per core, and then the 2MB L3 shared cache.
  - **Drop-in to existing Socket-F**

- AMD Barcelona
- http://www.anandtech.com/IT/showdoc.aspx?i=3091
- Comparison at AMD
- http://www.amd.com/us-en/Processors/ProductInformation/0,,30_118_8796_15225,00.html



Motherboard

- PCB (Printed Circuit Board)
- Next to CPU most important component for performance
- Sockets/connectors include:
  - CPU, Memory, PCI/PCI-X, AGP, Floppy disk
  - ATA and/or SCSI
  - Power
  - LEDs, speakers, switches, etc
  - Externel I/O
- Chips
  - System bus to memory
  - Peripheral bus to system bus
  - PROM with BIOS software

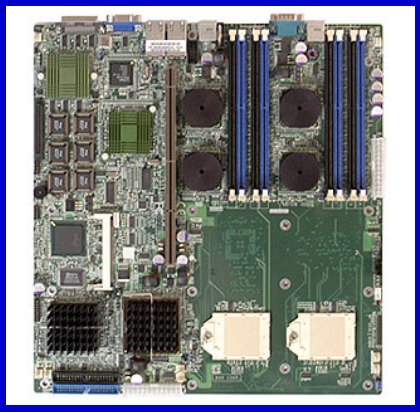
Motherboard

- Choice restricts
  - CPU
  - Clock speed
  - # of CPUs
  - Memory capacity, type
  - Disk interfaces
  - Number and types of I/O busses

Motherboards – I2

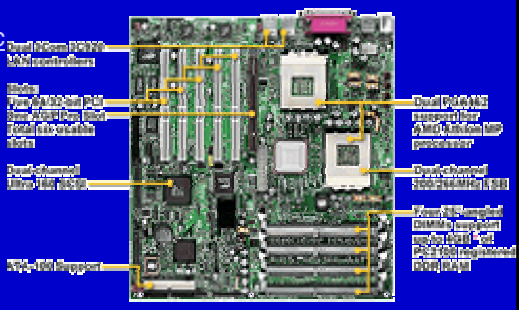
- Supermicro i2DML-8G2
- Dual I2
- Intel E8870 chipset
- **400 MHz** FSB
- 64-bit 133/100 MHz PCI-X

- http://www.supermicro.com/PRODUCT/MotherBoards/E8870/i2DML-8G2.htm



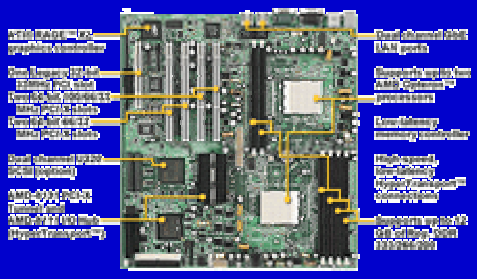
Motherboards - Opteron

- Tyan Thunder K7
- http://www.tyan.com/products/html/thunderk7.html
- Dual Opteron
- AMD's new 760MP chipset
- , DDR memory support,
- 64-bit PCI slots,
- AGP Pro slot (and integrated VGA),
- dual LAN controllers,
- dual-channel Ultra160 SCSI,



Motherboards - Opteron

- Tyan Thunder K8S
- Dual Opteron
- Two 128-bit DDR memory buses
- Two independent PCI-X buses
- Two 64-bit 66/33 MHz (3.3-volt) PCI-X slots - from PCI-X bridge A
- Two 64-bit 133/100/66/33 MHz (3.3-volt) PCI-X slots - from PCI-X bridge B (closer to CPUs)
- One Legacy 32-bit 33MHz (5-volt) PCI slot 64-bit PCI slots
- dual LAN controllers,
- dual-channel Ultra160 SCSI,



Motherboards - Opteron

- Tyan Thunder K8WE
- Dual Opteron
- Three HyperTransport links support up to 6.4GB/s data transfer rate each link
- 144-bit DDR interface (128-bit data + 16 bit ECC)
- Supports up to 16GB of Registered DDR 400/333/266 meory
- dual LAN controllers,
- dual-channel Ultra160 SCSI,

Two x16 PCI Express FULL SPEED slots
- Slot1 PCI-E x16 from nForce Prof. 2200
- Slot3 PCI-E x16 from nForce Prof. 2050

• Two independent 64-bit PCI-X buses
- Slot 4 & 5: PCI-X 100 MHz max. (Bridge B)
- Slot 6: PCI-X 133 MHz max. (Bridge A)

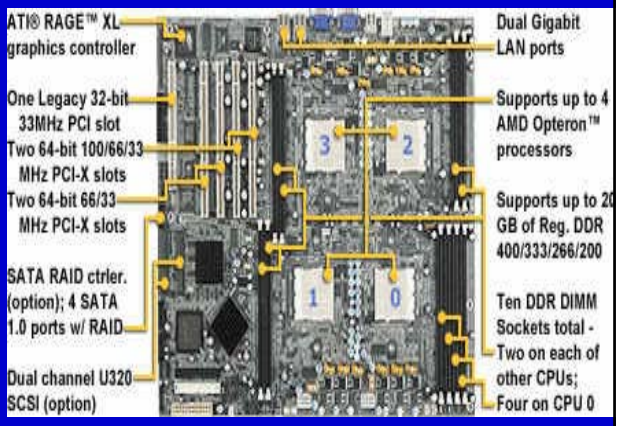
• One 32-bit 33MHz PCI v2.3 (Slot 2)

Motherboard – quad opteron

Thunder K8QS - S4880

128-bit dual channel memory bus
- Total ten DDR DIMM sockets (4 on CPU 0 and 2 each on CPU1, CPU2, CPU3 respectively)
- Supports up to 20 GB Registered DDR
- - Supports PC1600, PC2100, & PC2700 DDR

Two independent PCI-X buses
- Two 64-bit 66/33 MHz (3.3-Volt) PCI-X slots from PCI-X bridge A



Two 64-bit 100/66/33 MHz (3.3-Volt) PCI-X slots from PCI-X bridge B
- One legacy 32-bit 33MHz (5-Volt) PCI slot

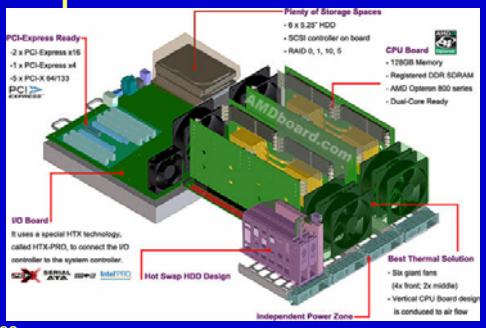
8-way Opteron

- **IWILL H8502**
- **8-way computing power**
  Linpack benchmark raised 80% from 4-way to 8-way system. Dual-core ready

  128G Memory

  PCI-Express Ready
  IWILL Octet-way server adopts nForce Professional 2200 plus 2050. It could provide up to 40 lanes PCI-Express bandwidth.

  High-performance I/O
  IWILL Octet-way server architecture uses a special HTX technology to connect the I/O controller to the system controller. The Serial ATA controller, USB 2.0, and serial ports are all integrated through a bidirectional 1000MHz HyperTransport interconnects for a maximum throughput of 1.6GBps.



Cluster in Computer Science

- Fianna 32 dual PIII-450 nodes with Gigabit & 100Mbs Ethernet
- Cluster Ohio cluster : 8 nodes each with four 550 Megahertz Intel Pentium III Xeon with Myrinet
- RocketCalc clusters
  - 4 boards each with dual Intel Pentium Xeon 2.4 GHz processors with Gigabit Ethernet

Cluster in Testing

Cluster in Testing – Aspen Systems

- 13 Nodes
  - AMD, 2.8GHz, 2220, dual-core, socket F, 2MB cache
- 13 Nodes
  - AMD, 2.0GHz, 2350, Quad-core, socket F, 4MB cache
- All with
  - 16GB, 256x72, 128x4, 667, DDR2, Registered, 240 Pin
  - 2 x EVGA, NVIDIA, 768MB, 8800GTX, PCI-E x16, Superclocked
  - SilverStorm, 9000, Dual port, Low Profile Infiniband NIC
- SilverStorm 24 port Infiniband DDR Switch
- HP, Ethernet Switch, 1U, 48, 100/1000, Layer 3

Clusters at OSC 2004

- **OSC IA32 Cluster**
  - **128 compute and 16 I/O nodes each with**
    - **Two gigabytes of RAM**
    - **Two, 1.4 GHz AMD Athlon MPs, each with 256kB of secondary cache**
- **OSC Itanium 2 Cluster (includes SGI Altix 3000)**
  - **32 processor SGI Altix 3000 for SMP and large memory applications**
  - 128 parallel compute nodes
    - **900 Megahertz Intel Itanium 2 processors with 1.5MB of tertiary cache**, 4GB RAM
  - **20 serial compute nodes with 20GB RAM**
  - **http://www.osc.edu/supercomputing/hardware/**

KENT STATE UNIVERSITY 24 April 2008 Paul A. Farrell Multicore Computing 47

Clusters at OSC 2005-2007

- **OSC Pentium 4 Cluster**
  - **252 compute and 16 I/O nodes each with**
    - **Four gigabytes of RAM**
    - **Two 2.4 GHz Pentium Xeon, each with 512kB of secondary cache**
    - 100Base-T and 1000Base-T Ethernet
    - 112 with Infiniband 10Gb interface
- **OSC Itanium 2 Cluster (includes SGI Altix 3000)**
  - **32 processor SGI Altix 3000 for SMP and large memory applications 64GB**
  - 128 parallel compute nodes
    - **900 Megahertz Intel Itanium 2 processors with 1.5MB of tertiary cache**, 4GB RAM, Myrinet 2000, Gigabit Ethernet
  - **20 serial compute nodes with 12GB RAM**

KENT STATE UNIVERSITY 24 April 2008 Paul A. Farrell Multicore Computing 48

Additions in 2005

- Itanium Cluster
- Two 16 processor SGI Altix 350 systems for SMP and large memory applications configured with:
  - 32 GB of memory
  - 16 1.4 Gigahertz Intel Itanium 2 processors
  - 4 Gigabit Ethernet interfaces
  - 2 Gigabit Fibre Channel interfaces
  - approximately 250 GB of temporary disk (/tmp)
- 110 compute nodes for parallel jobs configured with:
  - Four Gigabytes of RAM
  - Two, 1.3 Gigahertz Intel Itanium 2 processors
  - 80 Gigabytes, ultra-wide SCSI hard drive
  - One Myrinet 2000 interface card
  - One Gigabit Ethernet interface

IBM 1350

- **Glenn, the Ohio Supercomputer Center IBM Cluster 1350**
- **877 System x3455 compute nodes**
  - **Dual socket, dual core 2.6 GHz Opterons**
  - **8 GB RAM**
  - **48 GB local disk space in /tmp**
- **86 System x3755 compute nodes**
  - **Quad socket, dual core 2.6 GHz Opterons**
  - **16 GB RAM**
  - **1.8 TB (10 nodes) or 218 GB (76 nodes) local disk space in /tmp**
- **2 System x3755 accelerated compute nodes**
  - **Dual socket 2 dual core 2.6 GHz Opterons**
  - **8 GB RAM**
  - **Voltaire 10 Gbps PCI Express adapter**
- **1 e1350 Blade Center**
  - **4 Dual Cell based QS20 blades**
  - **Voltaire 10 Gbps PCI Express adapter**
- **4 System x3755 login nodes**
  - **Quad socket 2 dual core 2.6 GHz Opterons**
  - **8 GB RAM**
- **All connected together by 10 Gbps Infiniband**

OSC – BALE Cluster

- 50 node dual purpose system
- Used to run classroom and when not in use to run graphics rendering jobs
  - Dual processor AMD system, Tyan Tiger MPX 760MPX Motherboard
  - 1.53GHz AMD Athlon 1800MP CPUs
  - Myrinet 2000 interface
  - Quadro4 900 XGL graphics board

Apple G5 Cluster

- 1 front-end node configured with:
  - 4 Gigabytes of RAM
  - 2 - 2.0 Gigahertz PowerPC G5 processors
  - 2 Gigabit Fibre Channel interfaces
  - Approximately 750 GB of local disk
  - Approximately 12TB of Fibre Channel attached storage
- 32 compute nodes configured with:
  - Four Gigabytes of RAM
  - Two, 2.0 Gigahertz PowerPC G5 processors
  - Approximately 80 GB of local disk space (/tmp)
  - Dual Gigabit Ethernet interfaces