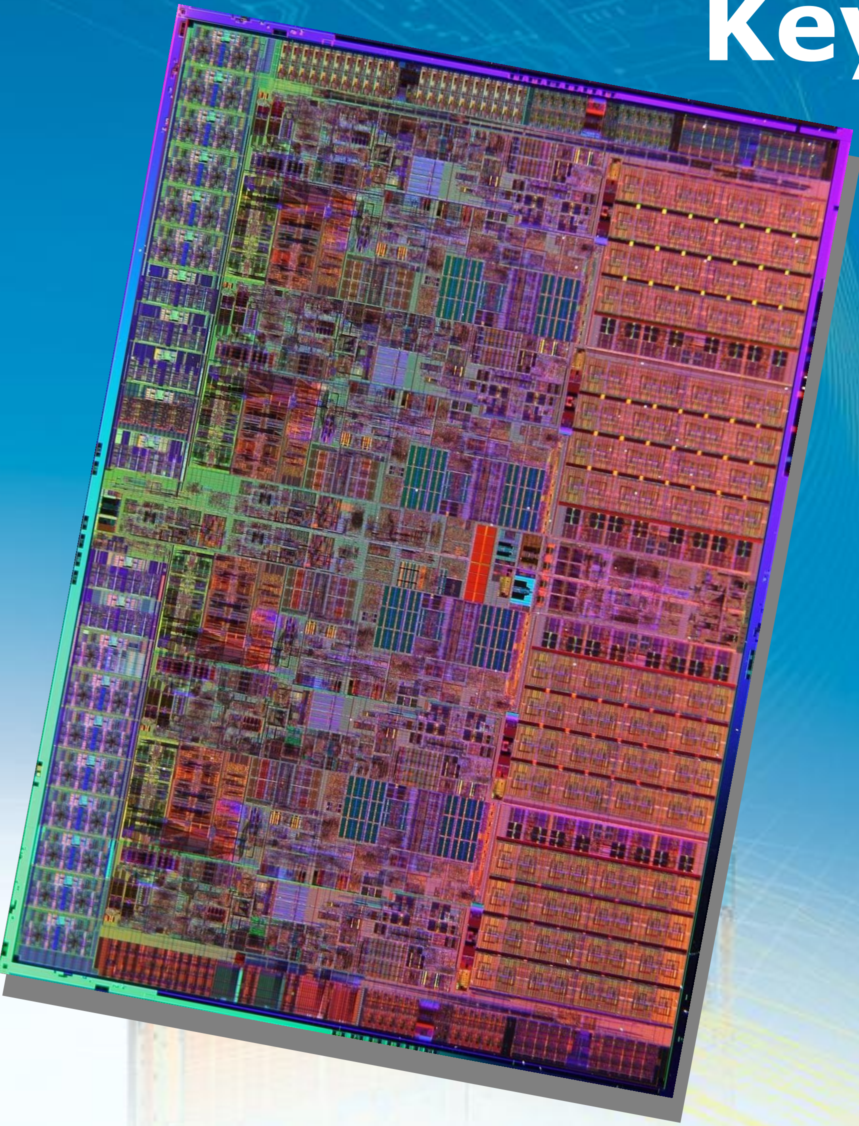
Key Nehalem Choices



Glenn Hinton

Intel Fellow

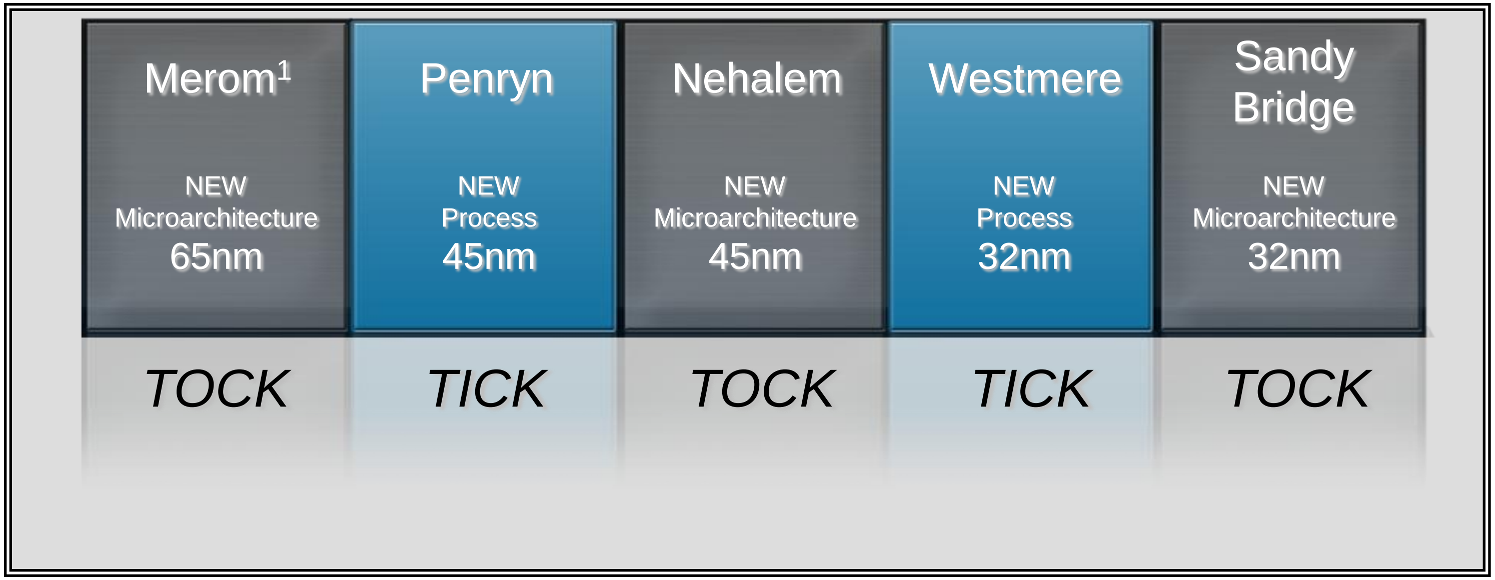
Nehalem Lead Architect

Feb 17, 2010

Outline

- Review NHM timelines and overall issues
- Converged core tensions
- Big debate – wider vectors vs SMT vs more cores
- How decided features
- Power, Power, Power
- Summary

Tick-Tock Development Model:
Pipelined developments – 5+ year projects



Merom[1]	Penryn	Nehalem	Westmere	Sandy Bridge
NEW Microarchitecture 65nm	NEW Process 45nm	NEW Microarchitecture 45nm	NEW Process 32nm	NEW Microarchitecture 32nm
TOCK	*TICK*	*TOCK*	*TICK*	*TOCK*

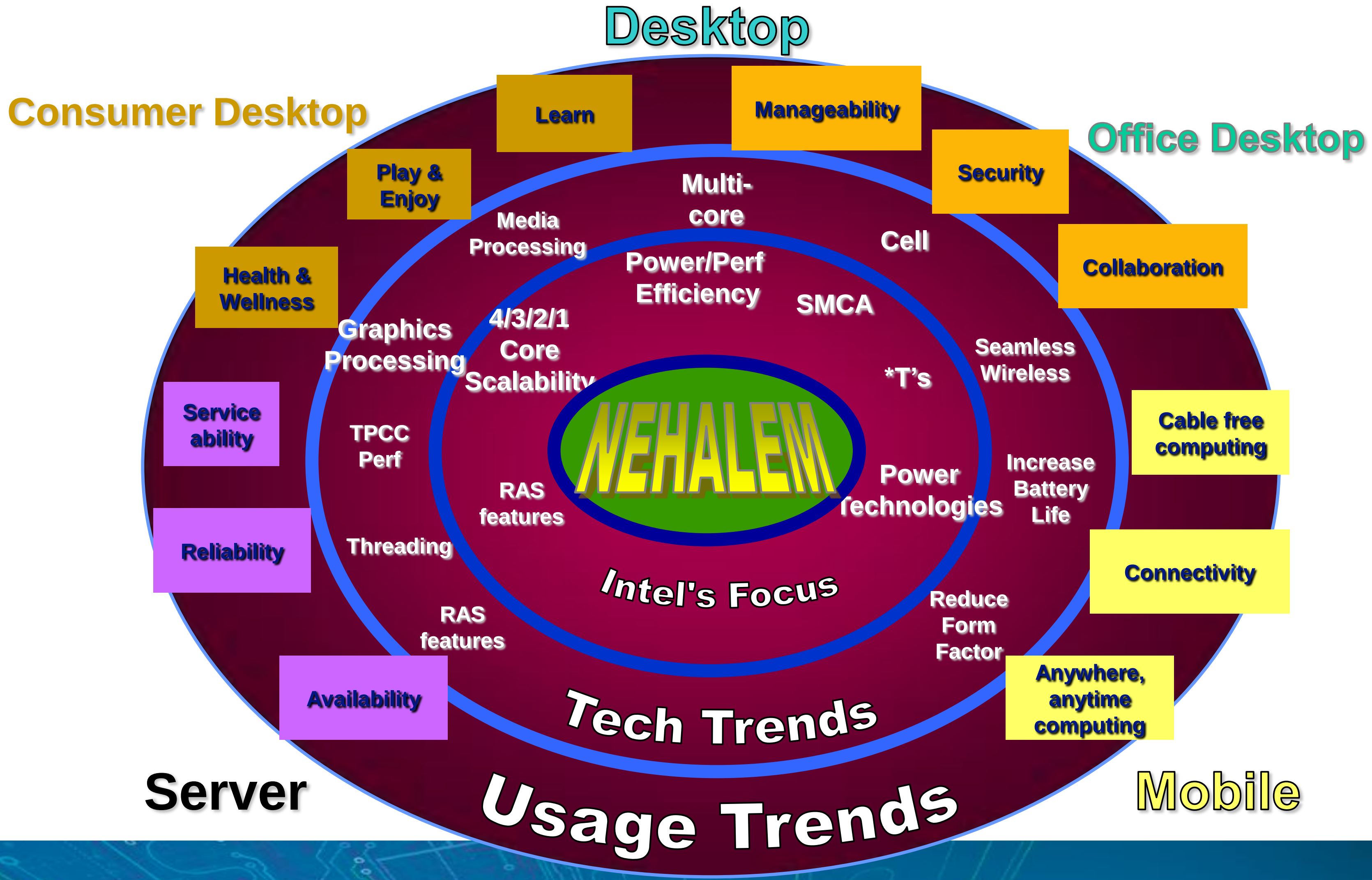
- Merom started in 2001, released 2006
- Nehalem started in 2003 (with research even earlier)
- Sandy bridge started in 2005
- Clearly already working on new Tock after Sandy Bridge
- Most of Nehalem uArch was decided by mid 2004
  - But most detailed engineering work happened in 2005/06/07

[1]Intel® Core™ microarchitecture (formerly Merom)
45nm next generation Intel® Core™ microarchitecture (Penryn)
Intel® Core™ Microarchitecture (Nehalem)
Intel® Microarchitecture (Westmere)
Intel® Microarchitecture (Sandy Bridge)

All dates, product descriptions, availability and plans are forecasts and subject to change without notice.

Nehalem:

Lots of competing aspects



Desktop

Consumer Desktop

Office Desktop

Server

Mobile

Learn

Manageability

Play & Enjoy

Security

Health & Wellness

Collaboration

Service ability

Reliability

Availability

Cable free computing

Connectivity

Anywhere, anytime computing

Multi-core

Media Processing

Cell

Power/Perf Efficiency

SMCA

Graphics Processing

4/3/2/1 Core Scalability

*T's

Seamless Wireless

TPCC Perf

NEHALEM

RAS features

Power Technologies

Increase Battery Life

Threading

Intel's Focus

RAS features

Reduce Form Factor

Tech Trends

Usage Trends

The Blind Men and the Elephant

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind
...
...

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!
by John Godfrey Saxe

“Converged Core” tradeoffs

Common CPU Core for multiple uses

- Mobile (Laptops)
- Desktop
- Server/HPC
- Workloads?
- How tradeoff?

“Converged Core” tradeoffs

- Mobile
  - 1/2/4 core options; scalable caches
  - Low TDP power CPU and GPU
  - Very low “average” power (great partial sleep state power)
  - Very low sleep state power
  - Low V-min to give best power efficiency when active
  - Moderate DRAM bandwidth at low power
  - Very dynamic power management
  - Low cost for volume
  - Great single threaded performance
    - Most apps single threaded

- Desktop
- Server
- Workloads – Productivity, Media, ISPEC, FSPEC, 32 vs 64 bit
- How tradeoff?

“Converged Core” tradeoffs

- Mobile

- Desktop
  - 1/2/4 core options; scalable caches
  - Media processing, high end game performance
  - Moderate DRAM bandwidth
  - Low cost for volume
  - Great single threaded performance
    - Most apps single threaded

- Server

- Workloads – Productivity, Media, Games, ISPEC, FSPEC, 32 vs 64 bit
- How tradeoff?

“Converged Core” tradeoffs

- Mobile
- Desktop
- Server
  - More physical address bits (speed paths, area, power)
  - More RAS features (ECC on caches, TLBs, Metc)
  - Larger caches, TLBs, BTBs, multi-socket snoop, etc
  - Fast Locks and multi-threaded optimizations
  - More DRAM channels (BW and capacity) & more external links
  - Dynamic power management
  - Many cores (4, 8, etc) so need low power per core
  - SMT gives large perf gain since threaded apps
  - Low V-min to allow many cores to fit in low blade power envelops

- Workloads – Workstation, Server, ISPEC, FSPEC, 64 bit
- How tradeoff?

“Converged Core” tradeoffs

- Mobile
  - 1/2/4 core options; scalable caches
  - Low TDP power CPU and GPU
  - Very low “average” power (great partial sleep state power)
  - Very low sleep state power
  - Low V-min to give best power efficiency when active
  - Moderate DRAM bandwidth at low power
  - Very dynamic power management
  - Low cost for volume
  - Great single threaded performance (most apps single threaded)
- Desktop
  - 1/2/4 core options; scalable caches
  - Media processing, high end game performance
  - Moderate DRAM bandwidth
  - Low cost for volume
  - Great single threaded performance (most apps single threaded)
- Server
  - More physical address bits (speed paths, area, power)
  - More RAS features (ECC on caches, TLBs, Metc)
  - Larger caches, TLBs, BTBs, multi-socket snoop, etc
  - Fast Locks and multi-threaded optimizations
  - More DRAM channels (BW and capacity) and more external links
  - Dynamic power management
  - Many cores (4, 8, etc) so need low power per core
  - SMT gives large perf gain since threaded apps
  - Low V-min to allow many cores to fit in low blade power envelops
- Workloads – Productivity, Media, Workstation, Server, ISPEC, FSPEC, 32 vs 64 bit, etc
- How tradeoff?

Early Base Core Selection - 2003

- Goals
    - Best single threaded perf?
    - Lowest cost dual core?
    - Lowest power dual core?
    - Best laptop battery life?
    - Most cores that fit in server size?
    - Best total throughput for cost/power in multi-core?
    - Least engineering costs?
- Major options
    - Enhanced Northwood (P4) pipeline?
    - Enhanced P6 pipeline (like Merom – Core 2 Duo)?
    - New from scratch pipeline?
- Why went with enhanced P6 (Merom) pipeline?
    - Lower power per core, lower per core die size, lower total effort
    - Better SW optimization consistency
- Likely gave up some ST perf (10-20%?)
    - But unlikely to have been able to do ‘bigger’ alternatives

2004 Major Decision: Cores vs Vectors vs SMT

- Just use older Penryn cores and have 3 or 4 of them?
  - No single threaded performance gains
- Put in wider vectors (like recently announced AVX)?
  - 256bit wide vectors (called VSSE back then)
  - Very power and area efficient, ***if*** doing wide vectors
  - Consumes die size and power when not using
- Add SMT per core and have fewer cores?
  - Very power and area efficient
  - Adds a lot of complexity; some die cost and power when not using
- What do servers want? Lots of cores/threads
- Laptops? Low power cores
- HE Desktops? Great media/game performance
- Options with similar die cost:
  - 2 enhanced cores + SMT + Wide Vectors?
  - 3 enhanced cores + SMT?
  - 4 simplified cores?

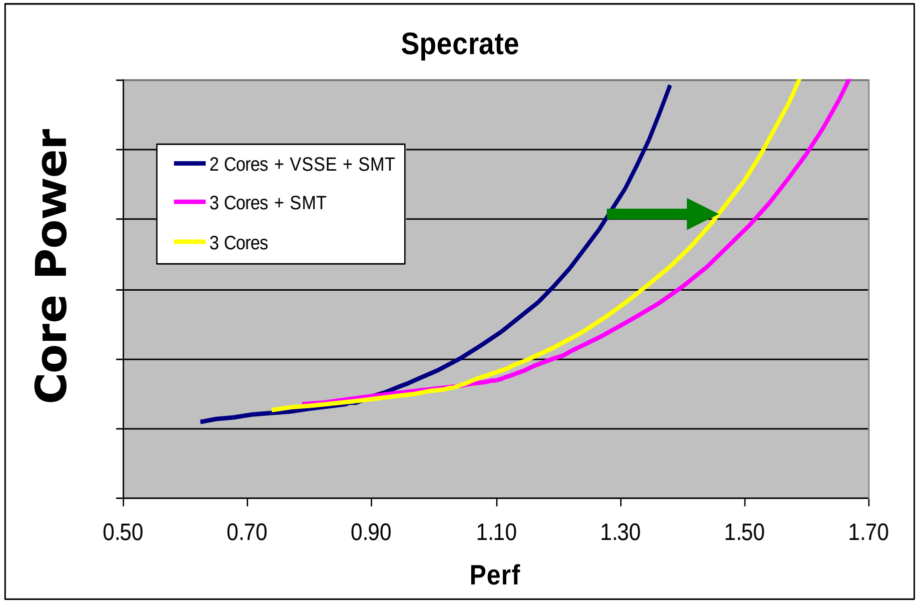
2 cores vs 4 Cores pros/cons

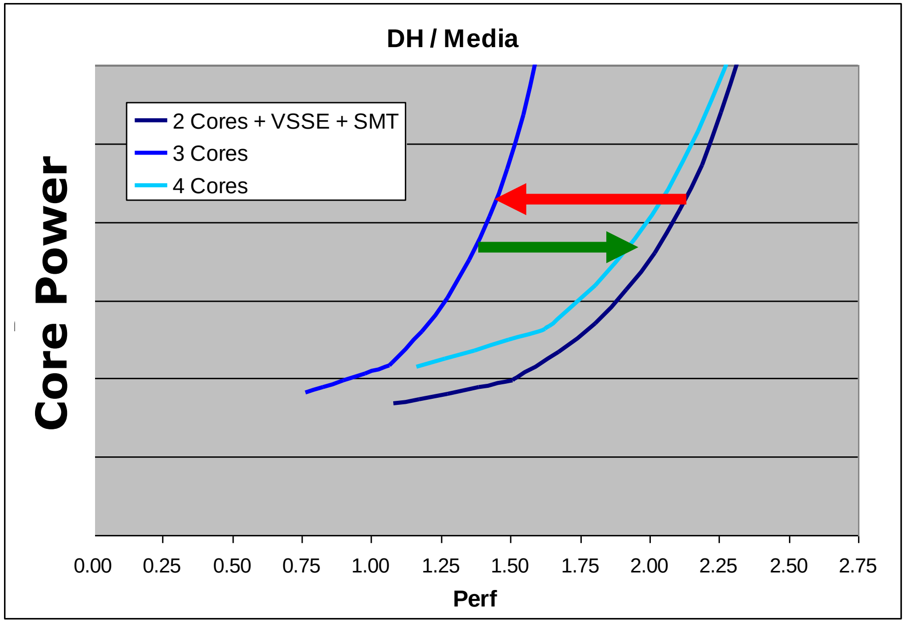
- 2 Cores+VSSE+SMT
  - Somewhat smaller die size
  - Lower power than 4 core
  - Better ST perf
  - Best media if use VSSE?
    - Not clear – looks like a wash
  - Specialized MIPS sometimes unused
  - VSSE gives perf for apps not easily threaded
    - Is threading really mainly for wizards?
  - New visible ISA feature like MMX, SSE

- 4 Cores
  - Better generic 4T perf
    - TPPC, multi-tasking
  - Best media perf on legacy 4T-enabled apps
  - Simpler HW design
  - Die size somewhat bigger
  - Simpler/harder SW enabling
    - Simpler since no VSSE
    - No SMT asymmetries
    - Harder since general 4T
  - 4T perf is also specialized
    - But probably less than VSSE
  - TDP Power somewhat higher
  - Somewhat lower average power
    - Since smaller single core
  - More granular to hit finer segments (1/2/3/4 core options)
  - More complex power management
  - Trade uArch change resources for power reduction

Nehalem Direction

- Tech Readiness Direction
  - 4/3/2/1 cores supported
  - VSSE dropped
    - Saves power and die size
  - SMT maintained
- VSSE in core less valued by servers and mobile
  - Casualty of Converged Core
  - Utilize scaled 4 core solution to recover media perf
- SMT to increase threaded perf
  - Initially target servers
- Spend more effort to reduce power

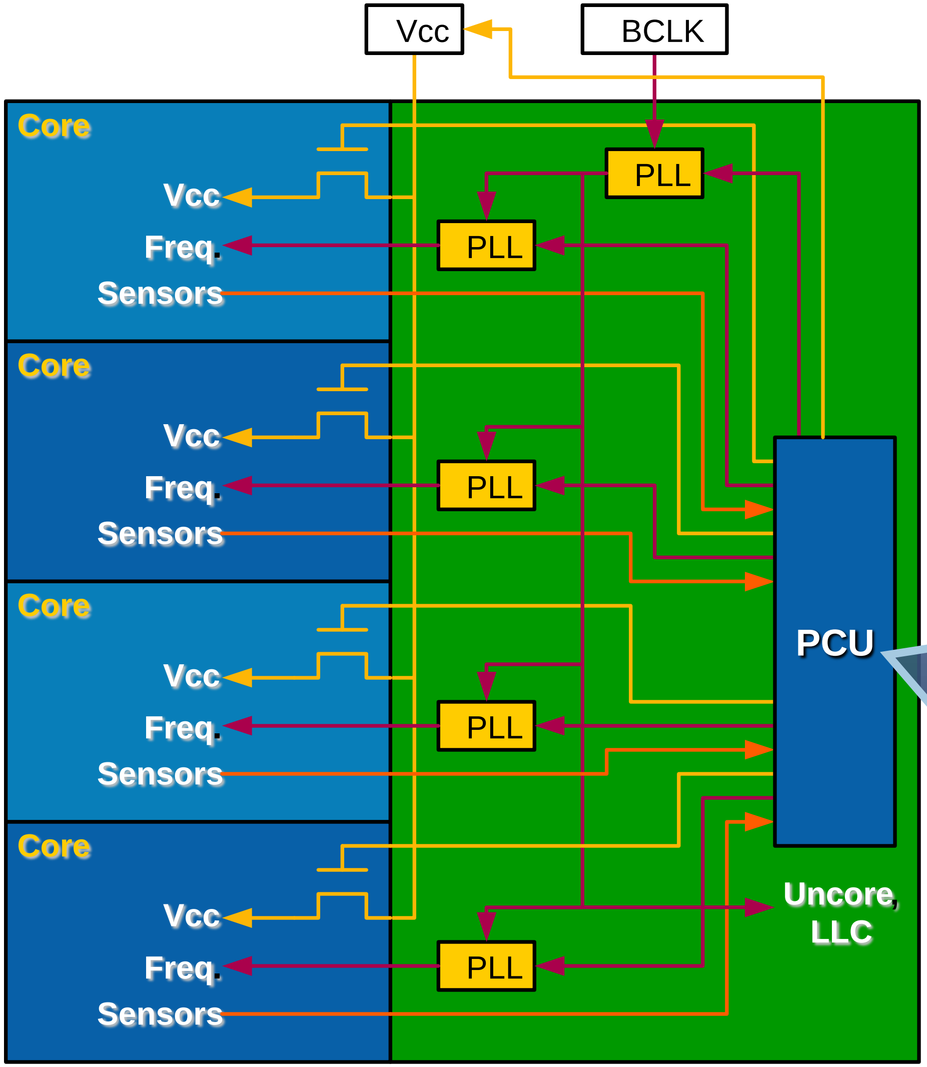




Early goal: Remove Multi-Core Perf tax
(In power constrained usages)

- Early Multi-cores lower freq than single core variants
- When started Nehalem dual cores still 2-3 years away…
  - Wanted 4 cores in high-end volume systems – A big increase
- Lower power envelopes planned for all Nehalem usages
  - Thinner laptops, blade servers, small-form-factor desktops, etc
- Many apps still single threaded
- All cores can have a lot of power - limits highest TDP freq
  - If just had one core the highest frequency could be a lot higher
- Turbo Mode/Power Gates removed this multi-core tax
  - One of biggest ST perf gains for mobile/power constrained usages
- Considered ways to do Turbo Mode
  - Decided must have flexible means to tune late in project
  - Added PCU with micro-controller to dynamically adapt

Power Control Unit



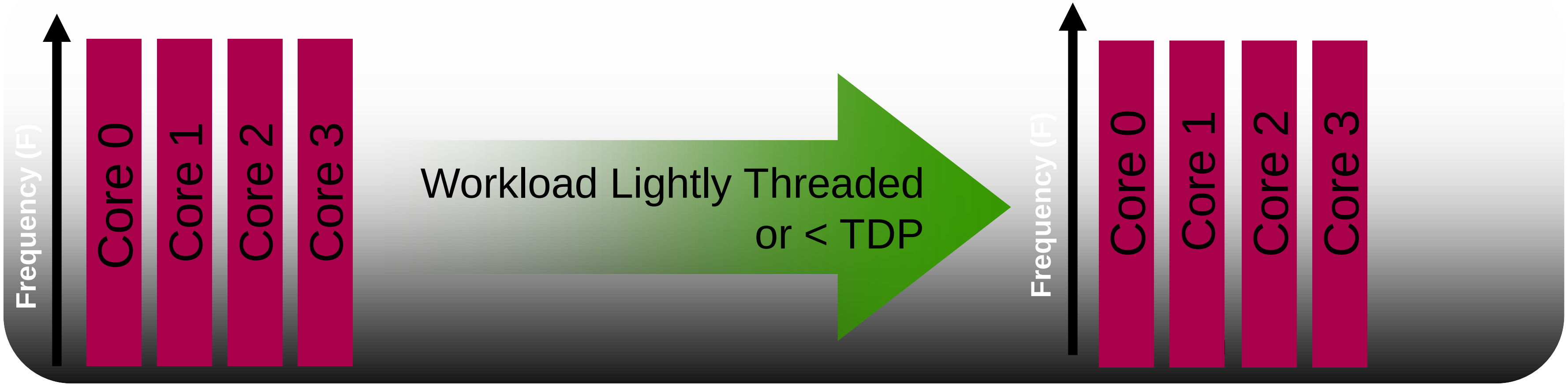
- Integrated proprietary microcontroller
- Shifts control from hardware to embedded firmware
- Real time sensors for temperature, current, power
- Flexibility enables sophisticated algorithms, tuned for current operating conditions

Intel® Core™ Microarchitecture (Nehalem) Turbo Mode

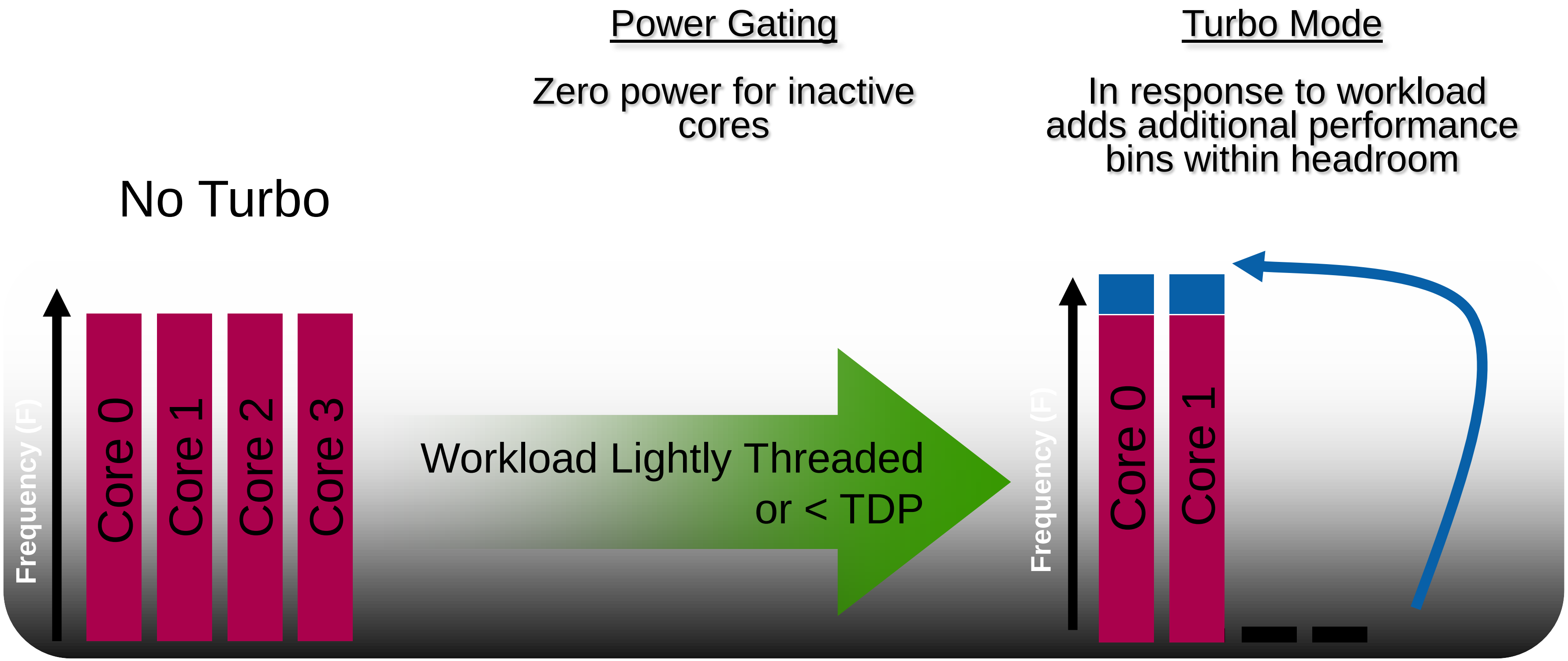
Power Gating

Zero power for inactive cores (C6 state)

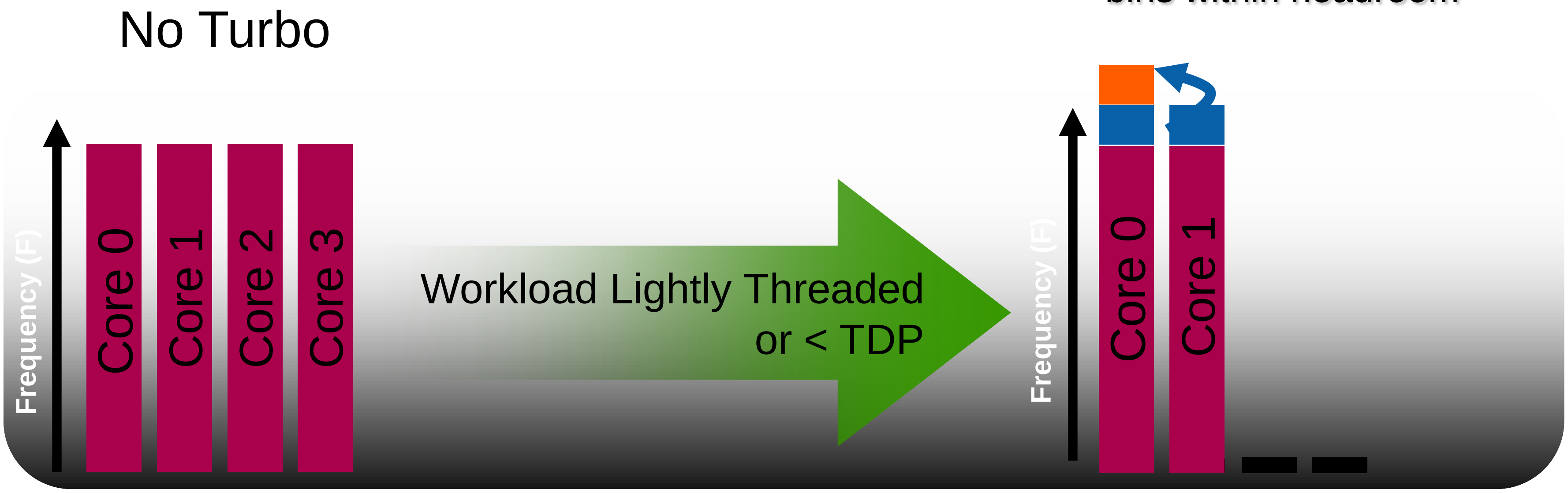
No Turbo



Intel® Core™ Microarchitecture (Nehalem) Turbo Mode


Power Gating
Zero power for inactive cores
Turbo Mode
In response to workload adds additional performance bins within headroom
No Turbo
Frequency (F)
Core 0
Core 1
Core 2
Core 3
Workload Lightly Threaded or < TDP
Frequency (F)
Core 0
Core 1

Intel® Core™ Microarchitecture (Nehalem) Turbo Mode



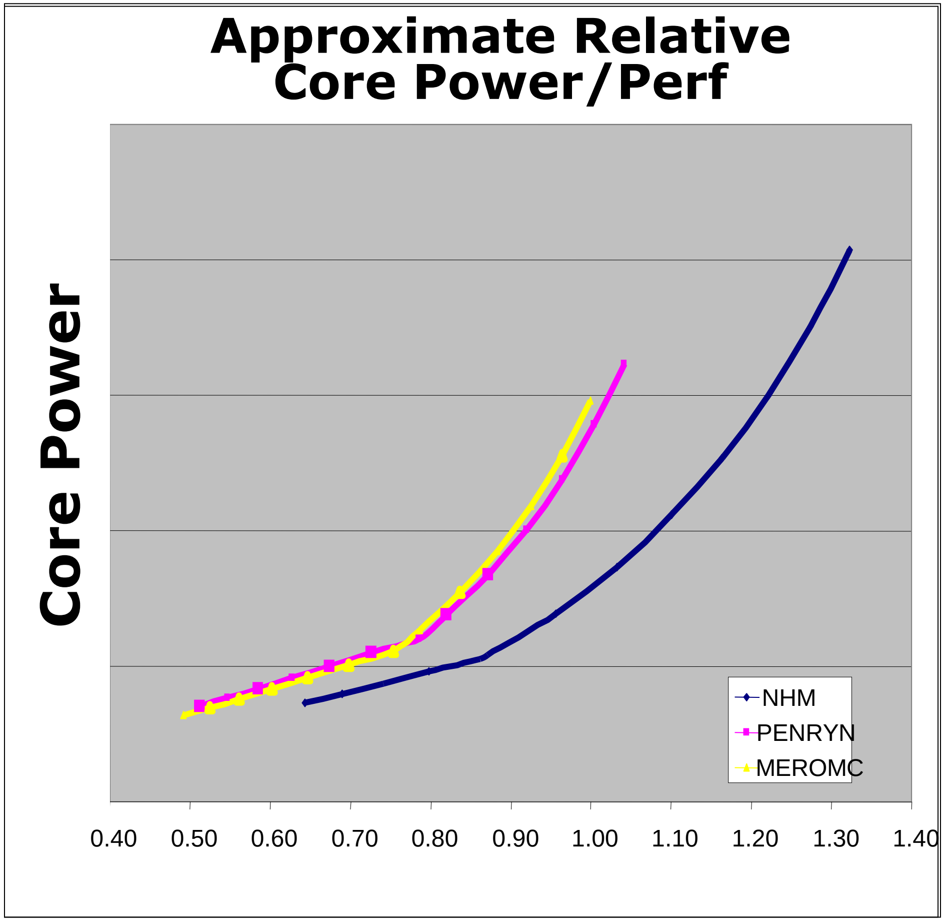
Power Gating
Zero power for inactive cores
Turbo Mode
In response to workload adds additional performance bins within headroom
No Turbo
Frequency (F)
Core 0
Core 1
Core 2
Core 3
Workload Lightly Threaded or < TDP
Frequency (F)
Core 0
Core 1

uArch features – Converged Core

- Difficult balancing the needs of the 3 conflicting requirements
- All CPU core features must be very power efficient
  - Helps all segments, especially laptops and multi-core servers
  - Requirement was to beat a 2:1 power/perf ratio
  - Ended up more like 1.3:1 power/perf ratio for perf features added
- Segment specific features can’t add much power or die size
- Initial Uncore optimized for 4 core DP server but had to scale down well to 2 core volume part
- Many things mobile wanted also helped servers
  - Low power cores, lower die size per core, etc
  - Active power management
  - More synergy than originally thought

Nehalem Power Efficiency Features

- Only adding power efficient uArch features
  - Net power : performance ratio of Nehalem core ~1.3 : 1
    - Far better than voltage scaling
- Reducing min operating voltage with linear freq decline
  - Cubic power reduction with ~linear perf reduction
- Implementing C6/Power Gated low-power state
  - Provides significant reduction in average mobile power
- Turbo mode
  - Allows processor to utilize entire available power envelope
  - Reduces performance penalty from multi-core on ST apps



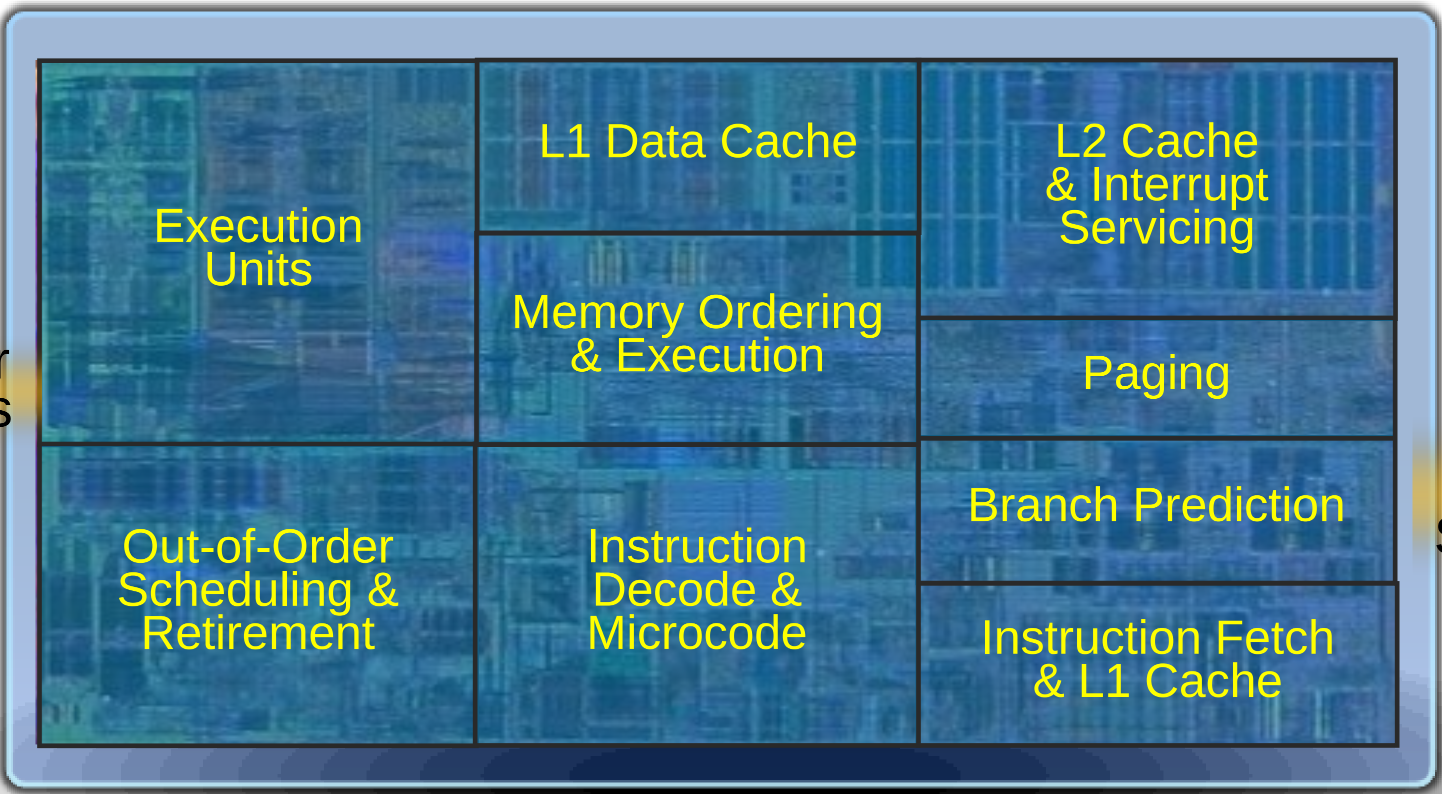
Designed for Performance

New SSE4.2 Instructions

Improved Lock Support

Additional Caching Hierarchy

Deeper Buffers


Execution Units
L1 Data Cache
L2 Cache & Interrupt Servicing
Memory Ordering & Execution
Paging
Branch Prediction
Out-of-Order Scheduling & Retirement
Instruction Decode & Microcode
Instruction Fetch & L1 Cache

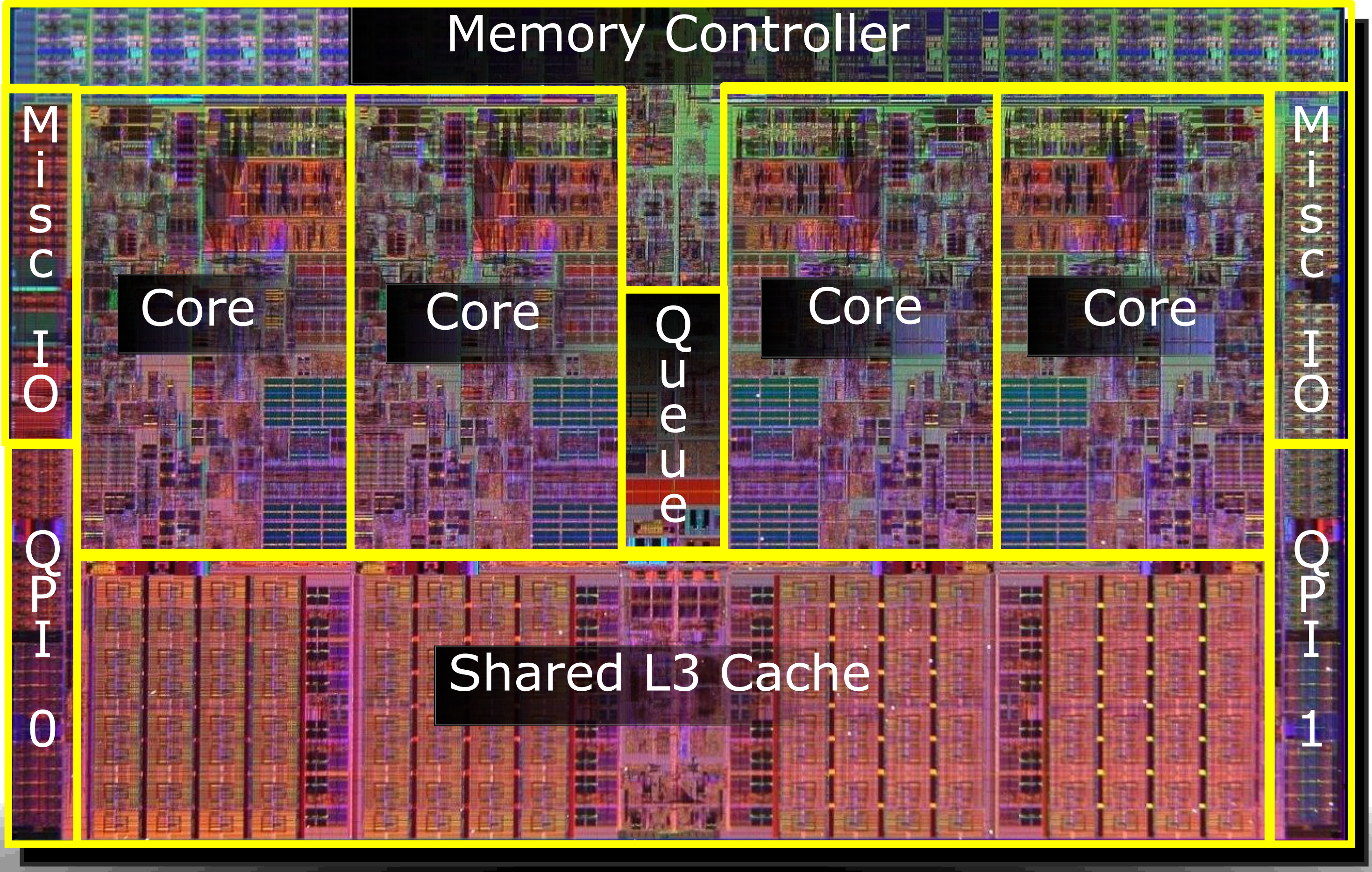
Improved Loop Streaming

Simultaneous Multi-Threading

Faster Virtualization

Better Branch Prediction

The First Intel® Core™ Microarchitecture (Nehalem) Processor



QPI: Intel® QuickPath Interconnect (Intel® QPI)

A Modular Design for Flexibility

Scalable Cores

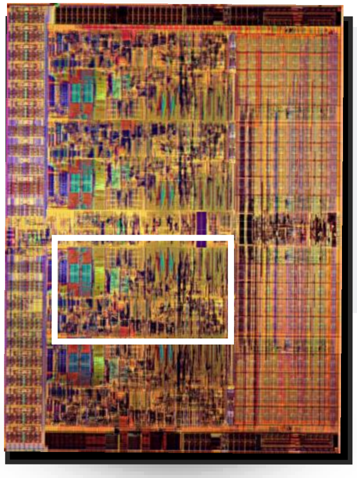
Same core for all segments

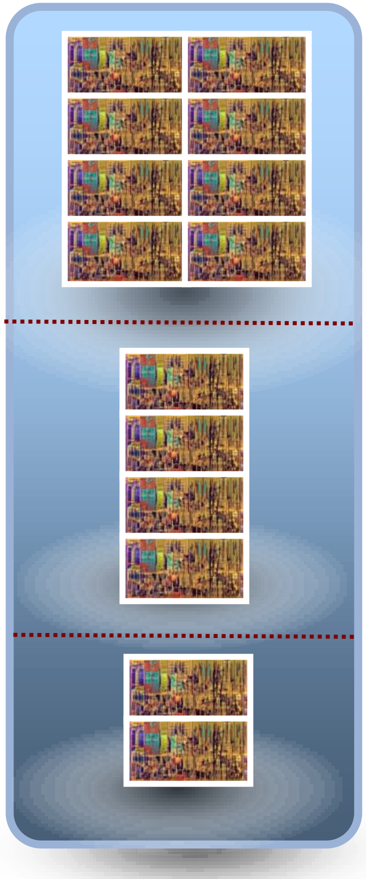
Common software optimization

Common feature set

Intel® Core™ Microarchitecture (Nehalem)

45nm





Servers/Workstations

Energy Efficiency, Performance, Virtualization, Reliability, Capacity, Scalability

Desktop

Performance, Graphics, Energy Efficiency, Idle Power, Security

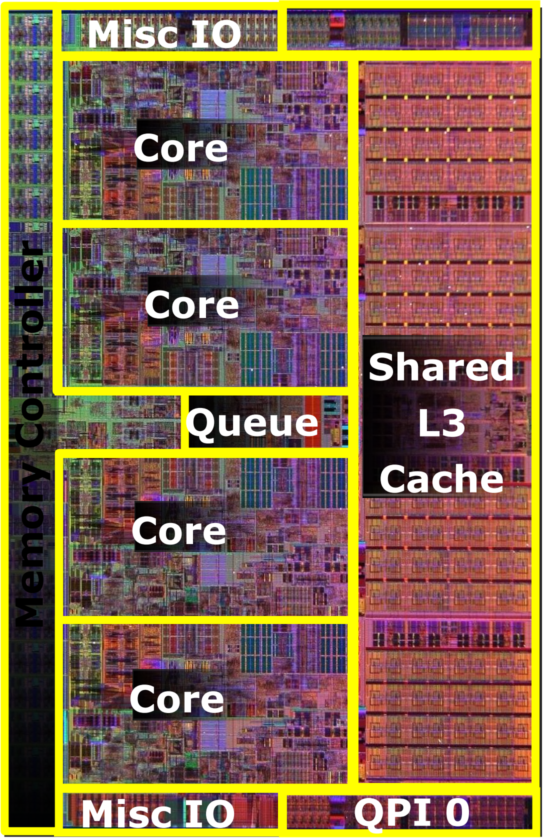
Mobile

Battery Life, Performance, Energy Efficiency, Graphics, Security

Optimized cores to meet multiple market segments

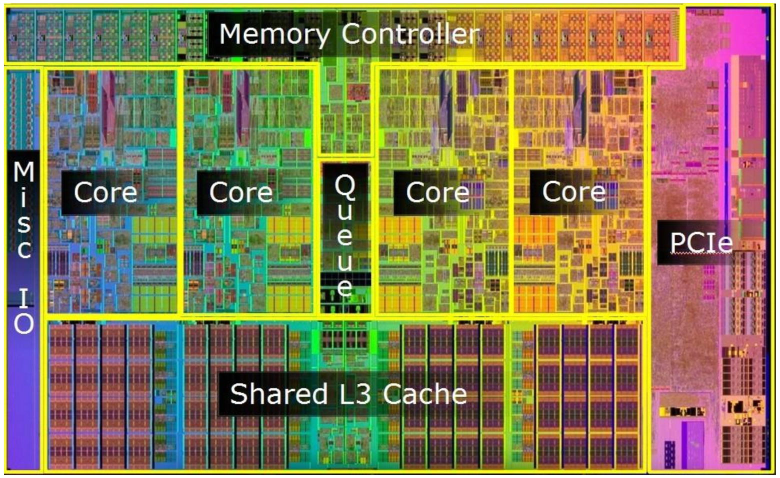
Modularity

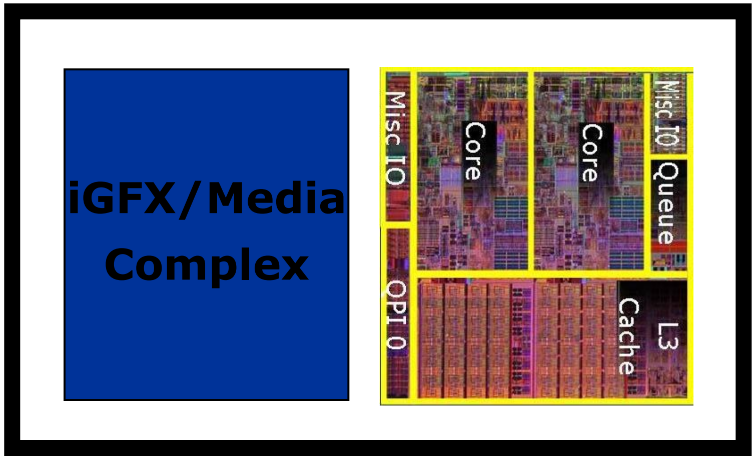
45nm Nehalem Core i7



Converged Core Architecture

45nm Lynnfield/Clarksfield

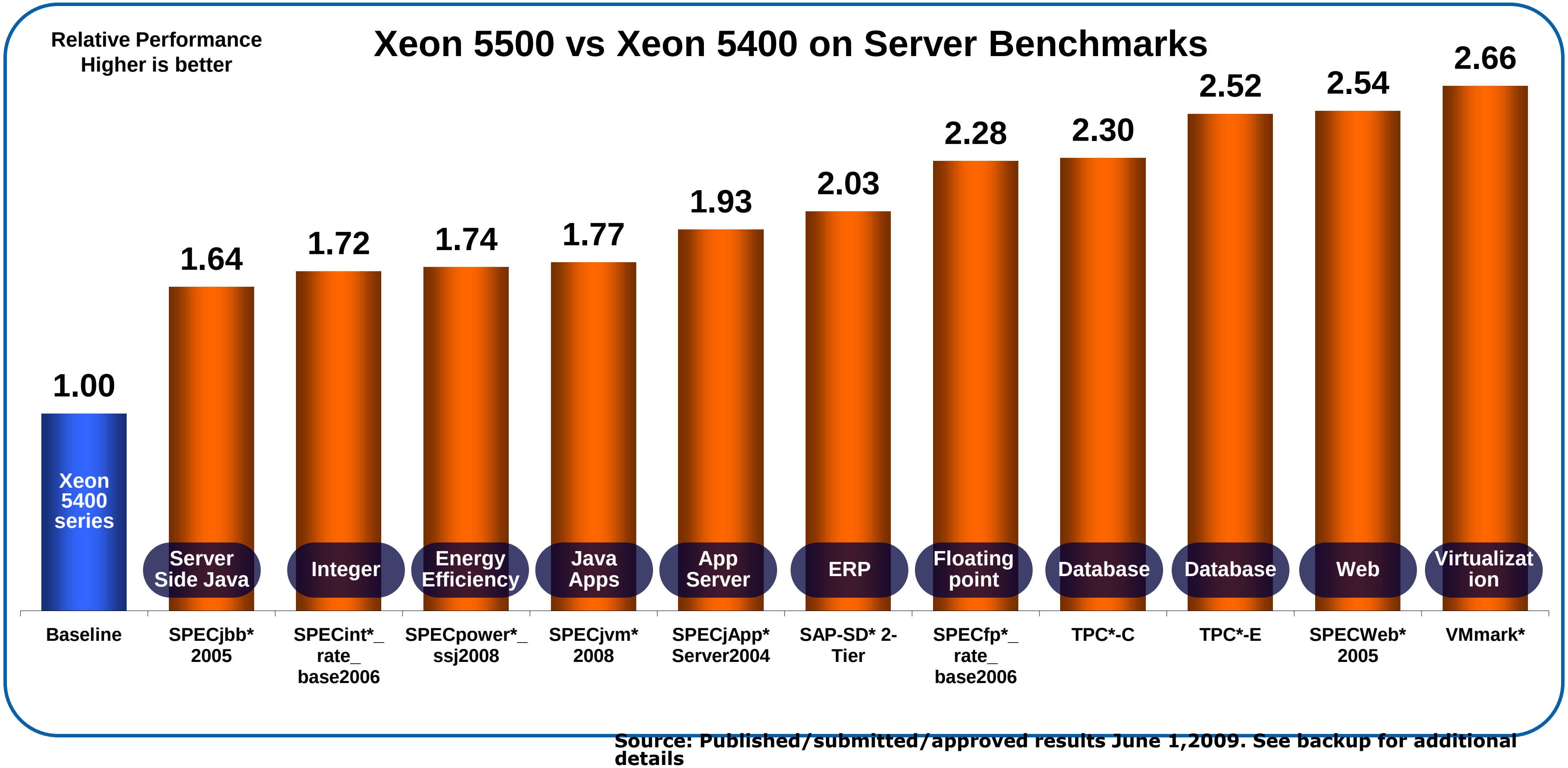




32nm Clarkdale/Arrandale

Intel® Xeon® Processor 5500 series based Server platforms

Server Performance comparison to Xeon 5400 Series



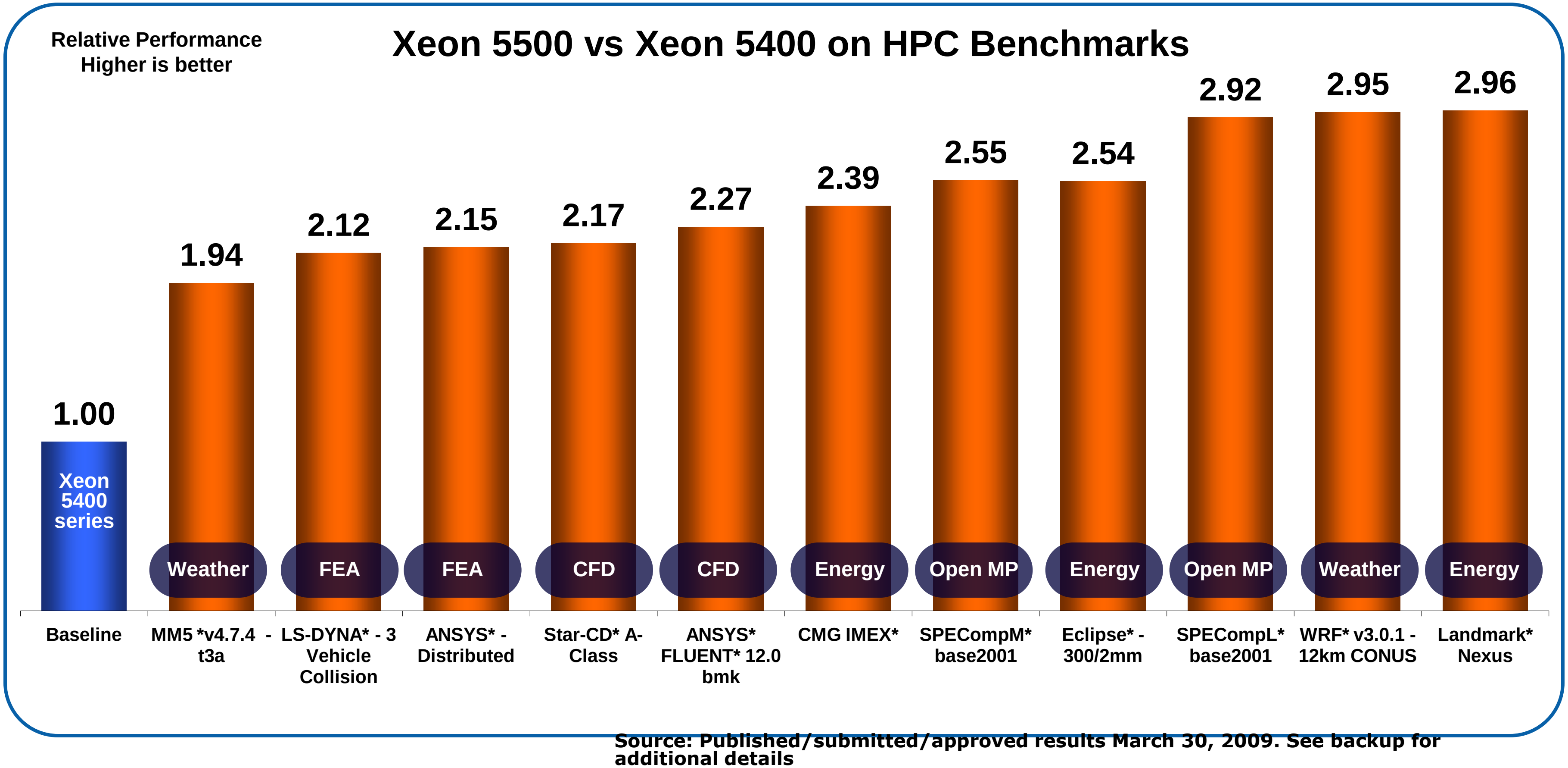
Source: Published/submitted/approved results June 1,2009. See backup for additional details

Leadership on key server benchmarks

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit http://www.intel.com/performance/resources/limits.htm Copyright © 2009, Intel Corporation. * Other names

Intel® Xeon® Processor 5500 series based Server platforms

HPC Performance comparison to Xeon 5400 Series



Source: Published/submitted/approved results March 30, 2009. See backup for additional details

Exceptional gains on HPC applications

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit http://www.intel.com/performance/resources/limits.htm Copyright © 2009, Intel Corporation. * Other names

Summary

- Nehalem was Intel’s first truly ‘converged core’
- Difficult tradeoffs between segments
- Result: Outstanding Server, DT, and mobile parts
- Acknowledge the outstanding Architects, Designers, and Validators that made this project a great success
  - A great team overcomes almost all challenges

Intel® Xeon® 5500 Performance Publications

SPECint*_rate_base2006

241 score (+72%)

SPECpower*_ssj2008

1977 ssj_ops/watt (+74%)

IBM J9* JVM

SPECfp*_rate_base2006

197 score (+128%)

SPECjAppServer*2004

3,975 JOPS (+93%)

Oracle WebLogic* Server

TPC*-C

631,766 tpmC (+130%)

Oracle 11g* database

SAP-SD* 2-Tier

5,100 SD Users (+103%)

SAP* ERP 6.0/IBM DB2*

VMmark*

24.35 @17 tiles(+166%)

VMware* ESX 4.0

TPC*-E

800 tpsE (+152%)

Microsoft SQL Server* 2008

SPECWeb*2005

75023 score (+150%)

Rock Web* Server

Fluent* 12.0 benchmark

Geo mean of 6 (+127%)

ANSYS FLUENT*

SPECjbb*2005

604,417 BOPS (+64%)

IBM J9* JVM

SPECapc* for Maya 6.5

7.70 score (+87%)

Autodesk* Maya

Over 30 New 2S Server and Workstation World Records!

Percentage gains shown are based on comparison to Xeon 5400 series; Performance results based on published/submitted results as of April 27, 2009. Platform configuration details are available at http://www.intel.com/performance/server/xeon/summary.htm *Other names and brands may be claimed as the property of others

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations

Intel® Xeon® Processor 5500 Series based Platforms

ISV Application Performance

ISV Application	Xeon 5500 vs Xeon 5400
Ansys* CFX11* – CFD Simulation	+88%
ERDAS* ERMapper Suite*	+69%
ESI Group* PAM-CRASH*	+50%
ExitGames* Neutron* – Service Platform	+80%
Epic* – EMR Solution	+82%
Giant* Juren* – Online Game	+160%
IBM* DB2* 9.5 – TPoX XML	+60%
IBM* Informix* Dynamic Server	+84%
IBM* solidDB* – In Memory DB	+87%
Image Analyzer * – Image Scanner	+100%
Infowrap* – Small Data Set	+129%
Infowrap* – Weather Forecast	+155%
Intense* IECCM* – Output Mgmt.	+150%
Intersystems* Cache* – EMR	+63%
Kingdee* APUSIC* – Middleware	+93%
Kingdee* EAS* – ERP	+142%
Kingdom* – Stock Transaction	+141%

ISV Application	Xeon 5500 vs Xeon 5400
Kingsoft* JXIII Online* – Game Server	+98%
Mediaware* Instream* – Video Conv.	+73%
Neowiz* Pmang* – Game Portal	+100%
Neusoft* – Healthcare	+131%
Neusoft* – Telecom BSS VMware*	+115%
NHN Corp* Cubrid* – Internet DB	+44%
QlikTech* QlikView* – BI	+36%
SAS* Forecast Server*	+80%
SAP* NetWeaver* – BI	+51%
SAP* ECC 6.0* – ERP Workload	+71%
Schlumberger* Eclipse300*	+213%
SunGard* BancWare Focus ALM*	+38%
Supcon* – APC Intel. Sensor	+191%
TongTech* – Middleware	+95%
UFIDA* NC* – ERP Solution	+230%
UFIDA* Online – SaaS Hosting	+237%
Vital Images* – Brain Perfusion 4DCT*	+77%

Exceptional gains (1.6x to 3x) on ISV applications

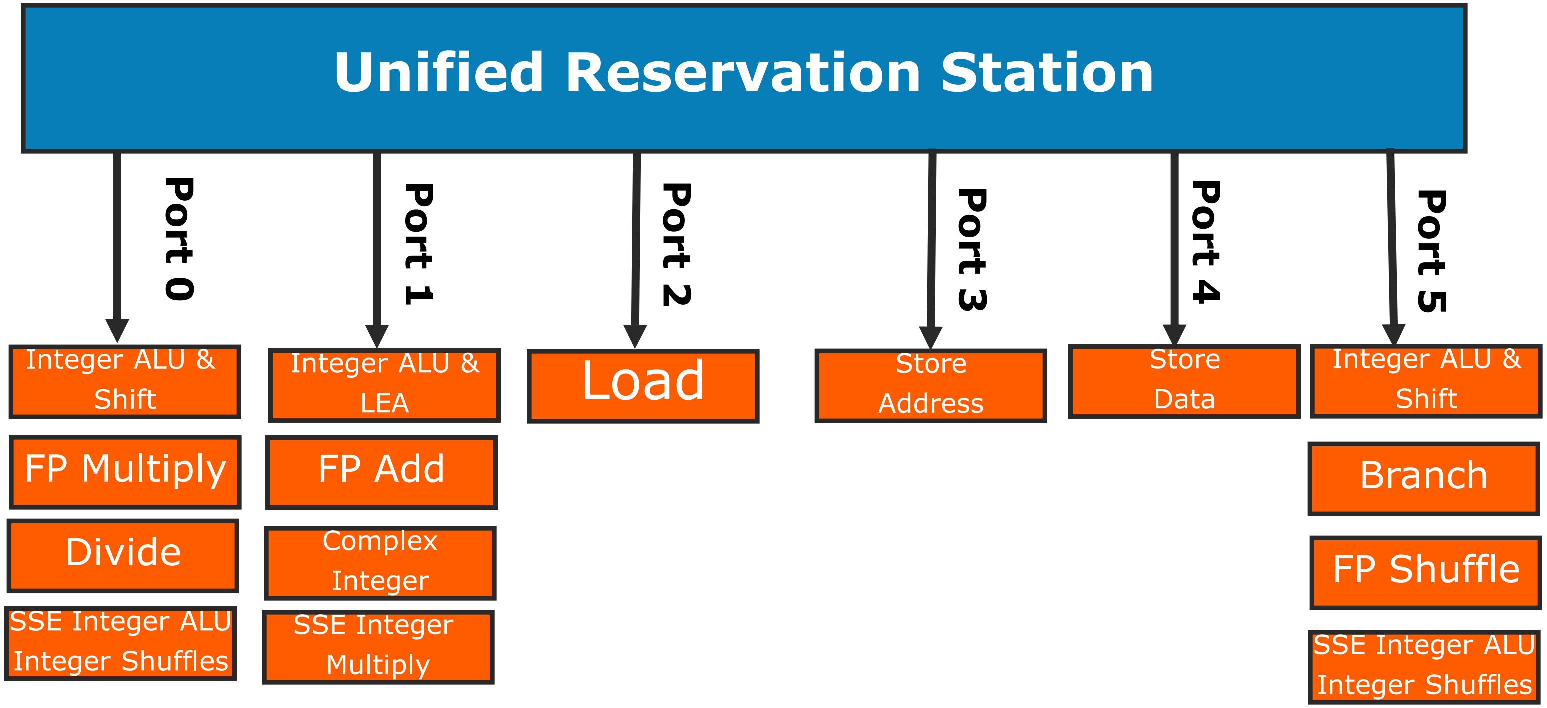
Source: Results measured and approved by Intel and ISV partners on pre-production platforms. March 30, 2009.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit http://www.intel.com/performance/resources/limits.htm Copyright © 2009, Intel Corporation. * Other names

Execution Unit Overview

Unified Reservation Station
- Schedules operations to Execution units
- Single Scheduler for all Execution Units
- Can be used by all integer, all FP, etc.

Execute 6 operations/cycle
- 3 Memory Operations
  - 1 Load
  - 1 Store Address
  - 1 Store Data
- 3 "Computational" Operations



Unified Reservation Station

Port 0
- Integer ALU & Shift
- FP Multiply
- Divide
- SSE Integer ALU Integer Shuffles

Port 1
- Integer ALU & LEA
- FP Add
- Complex Integer
- SSE Integer Multiply

Port 2
- Load

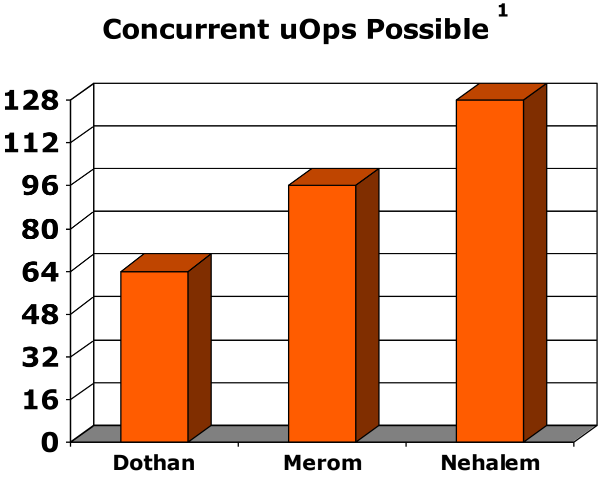
Port 3
- Store Address

Port 4
- Store Data

Port 5
- Integer ALU & Shift
- Branch
- FP Shuffle
- SSE Integer ALU Integer Shuffles

Increased Parallelism

- Goal: Keep powerful execution engine fed
- Nehalem increases size of out-of-order window by 33%
- Must also increase other corresponding structures



Structure	Intel® Core™ microarchitecture (formerly Merom)	Intel® Core™ microarchitecture (Nehalem)	Comment
Reservation Station	32	36	Dispatches operations to execution units
Load Buffers	32	48	Tracks all load operations allocated
Store Buffers	20	32	Tracks all store operations allocated

Increased Resources for Higher Performance

[1] Intel® Pentium® M processor (formerly Dothan)
Intel® Core™ microarchitecture (formerly Merom)
Intel® Core™ microarchitecture (Nehalem)

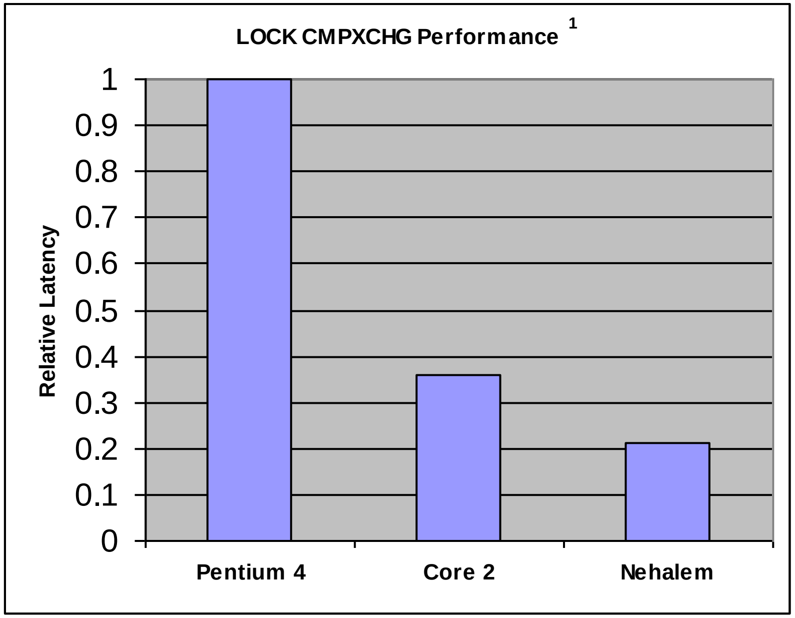
New TLB Hierarchy

- Problem: Applications continue to grow in data size
- Need to increase TLB size to keep the pace for performance
- Nehalem adds new low-latency unified 2nd level TLB

	# of Entries
1st Level Instruction TLBs	
Small Page (4k)	128
Large Page (2M/4M)	7 per thread
1st Level Data TLBs	
Small Page (4k)	64
Large Page (2M/4M)	32
New 2nd Level Unified TLB	
Small Page Only	512

Faster Synchronization Primitives

- Multi-threaded software becoming more prevalent
- ***Scalability*** of multi-thread applications can be limited by synchronization
- Synchronization primitives: LOCK prefix, XCHG
- Reduce synchronization latency for legacy software

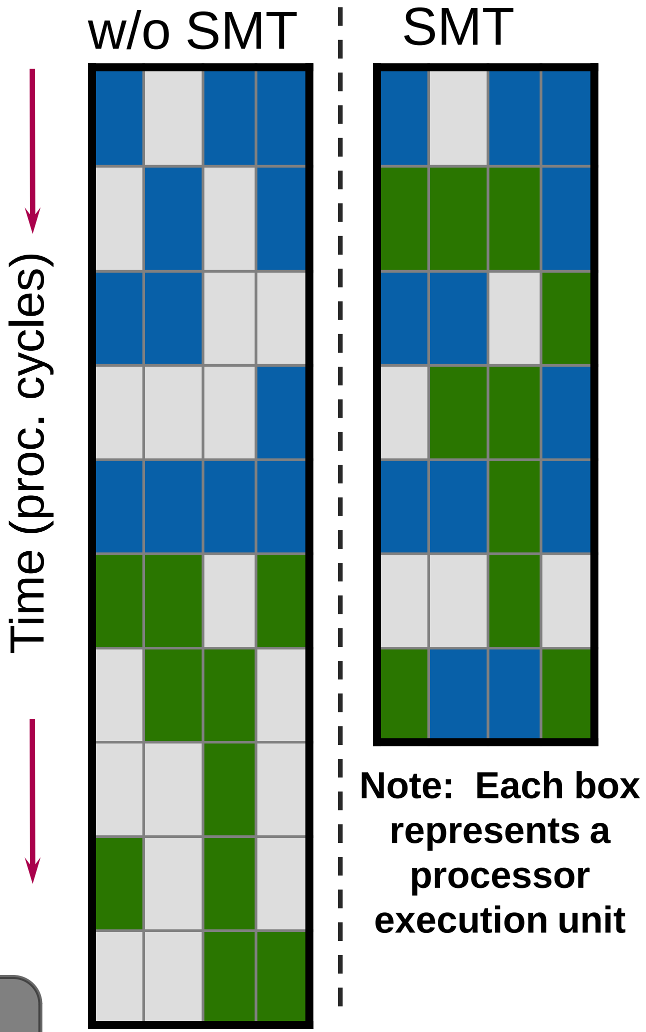


*Greater thread **scalability** with Nehalem*

[1] Intel® Pentium® 4 processor
Intel® Core™2 Duo processor
Intel® Core™ microarchitecture (Nehalem)-based processor

Intel® Hyper-Threading Technology

- Also known as Simultaneous Multi-Threading (SMT)
  - Run 2 threads at the same time per core
- Take advantage of 4-wide execution engine
  - Keep it fed with multiple threads
  - Hide latency of a single thread
- Most ***power efficient*** performance feature
  - Very low die area cost
  - Can provide significant performance benefit depending on application
  - Much more efficient than adding an entire core
- Intel® Core™ microarchitecture (Nehalem) advantages
  - Larger caches
  - Massive memory BW



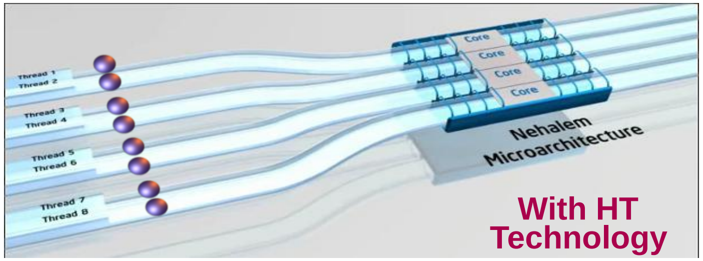
Note: Each box represents a processor execution unit

Simultaneous multi-threading enhances performance and energy efficiency

Intel® Hyper-Threading Technology

- Nehalem is a scalable multi-core architecture
- Hyper-Threading Technology augments benefits
  - Power-efficient way to boost performance in all form factors: higher multi-threaded performance, faster multi-tasking response

Core
Core
Core
Core
Nehalem
Microarchitecture
Without HT
Technology



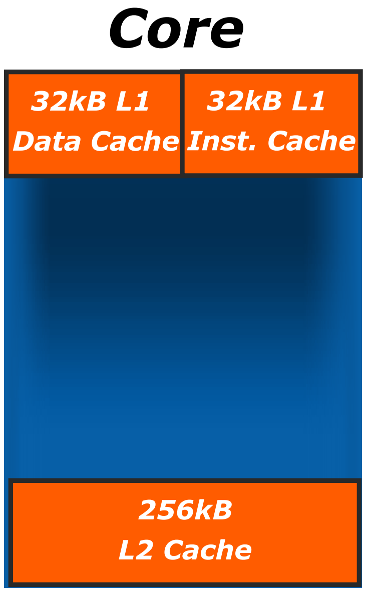
	Hyper-Threading		Multi-cores
	Shared or Partitioned	Replicated	Replicated
Register State		X	X
Return Stack		X	X
Reorder Buffer	X		X
Instruction TLB	X		X
Reservation Stations	X		X
Cache (L1, L2)	X		X
Data TLB	X		X
Execution Units	X		X

- Next generation Hyper-Threading Technology:
  - Low-latency pipeline architecture
  - Enhanced cache architecture
  - Higher memory bandwidth

Enables 8-way processing in Quad Core systems, 4-way processing in Small Form Factors

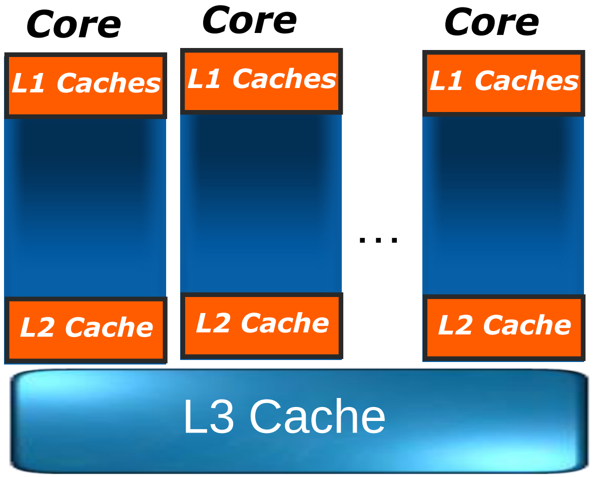
Caches

- New 3-level Cache Hierarchy
- 1st level caches
  - 32kB Instruction cache
  - 32kB, 8-way Data Cache
    - Support more L1 misses in parallel than Intel® Core™2 microarchitecture
- 2nd level Cache
  - New cache introduced in Intel® Core™ microarchitecture (Nehalem)
  - Unified (holds code and data)
  - 256 kB per core (8-way)
  - ***Performance:*** Very low latency
    - 10 cycle load-to-use
  - ***Scalability:*** As core count increases, reduce pressure on shared cache



3rd Level Cache

- Shared across all cores
- Size depends on # of cores
  - Quad-core: Up to 8MB (16-ways)
  - ***Scalability:***
    - Built to vary size with varied core counts
    - Built to easily increase L3 size in future parts
- Perceived latency depends on frequency ratio between core & uncore
- Inclusive cache policy for best ***performance***
  - Address residing in L1/L2 ***must*** be present in 3rd level cache



Hardware Prefetching (HWP)

- HW Prefetching critical to hiding memory latency
- Structure of HWPs similar as in Intel® Core™2 microarchitecture
  - Algorithmic improvements in Intel® Core™ microarchitecture (Nehalem) for higher performance
- L1 Prefetchers
  - Based on instruction history and/or load address pattern
- L2 Prefetchers
  - Prefetches loads/RFOs/code fetches based on address pattern
  - Intel Core microarchitecture (Nehalem) changes:
    - **Efficient Prefetch** mechanism
      - Remove the need for Intel® Xeon® processors to disable HWP
    - Increase prefetcher **aggressiveness**
      - Locks on address streams quicker, adapts to change faster, issues more prefetchers more aggressively (when appropriate)

SW Prefetch Behavior

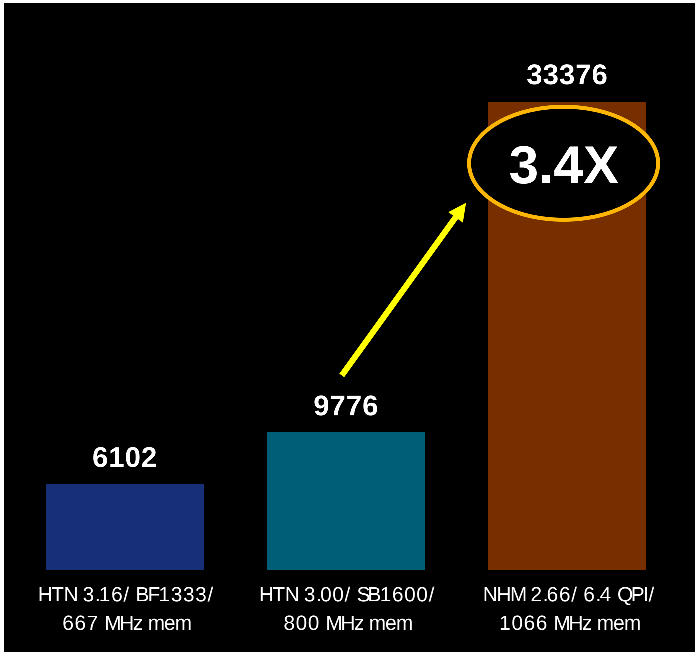
- PREFETCHT0: Fills L1/L2/L3
- PREFETCHT1/T2: Fills L2/L3
- PREFETCHNTA: Fills L1/L3, L1 LRU is not updated

- SW prefetches can conduct page walks
- SW prefetches can spawn HW prefetches
  - SW prefetch caching behavior not obeyed on HW prefetches

Memory Bandwidth – Initial Intel® Core™ Microarchitecture (Nehalem) Products

- 3 memory channels per socket
- ≥ DDR3-1066 at launch
- Massive ***memory BW***
- ***Scalability***
  - Design IMC and core to take advantage of BW
  - Allow performance to scale with cores
    - Core enhancements
      - Support more cache misses per core
      - Aggressive hardware prefetching w/ throttling enhancements
    - Example IMC Features
      - Independent memory channels
      - Aggressive Request Reordering

Stream Bandwidth – Mbytes/Sec (Triad)



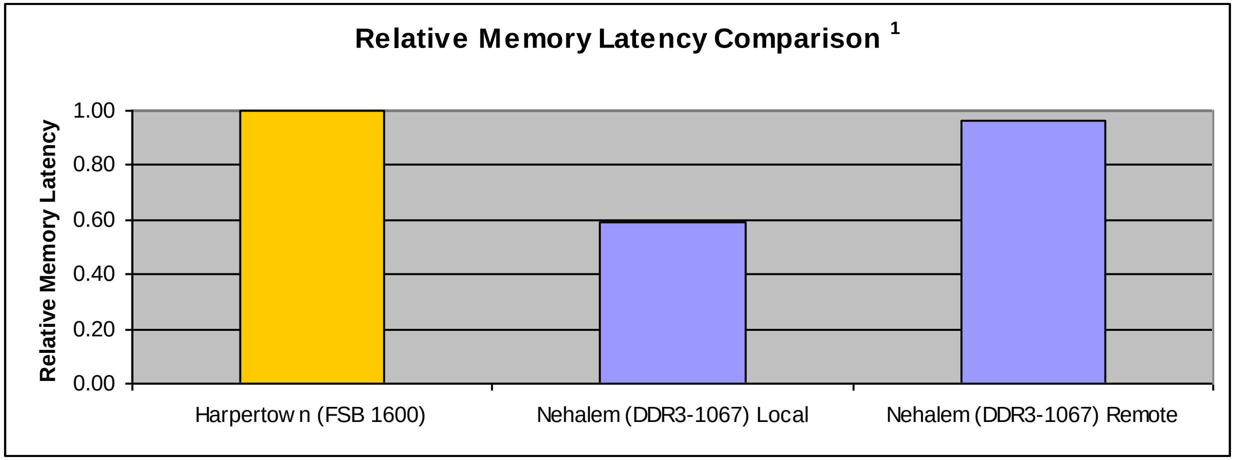
Source: Intel Internal measurements – August 2008[1]

Massive memory BW provides performance and scalability

[1]HTN: Intel® Xeon® processor 5400 Series (Harpertown)
NHM: Intel® Core™ microarchitecture (Nehalem)

Memory Latency Comparison

- ***Low memory latency*** critical to high performance
- Design integrated memory controller for low latency
- Need to optimize both local and remote memory latency
- Intel® Core™ microarchitecture (Nehalem) delivers
  - Huge reduction in local memory latency
  - Even remote memory latency is fast
- Effective memory latency depends per application/OS
  - Percentage of local vs. remote accesses
  - Intel Core microarchitecture (Nehalem) has lower latency regardless of mix



[1]Next generation Quad-Core Intel® Xeon® processor (Harpertown)
Intel® CoreTM microarchitecture (Nehalem)

Virtualization

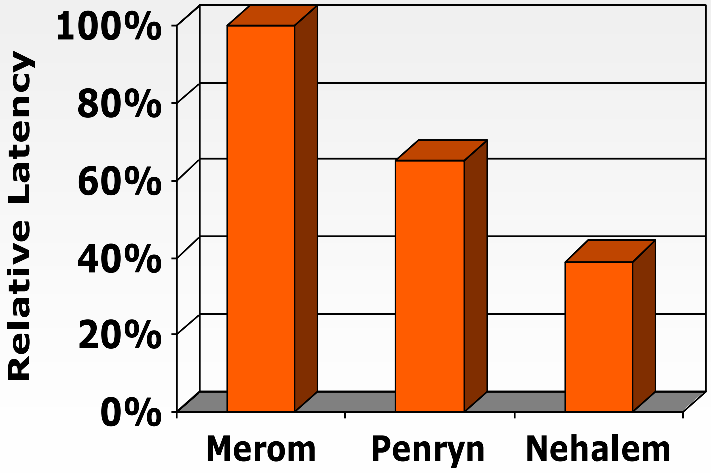
- To get best virtualized ***performance***
  - Have best native performance
  - Reduce:
    - # of transitions into/out of virtual machine
    - Latency of transitions
- Intel® Core™ microprocessor (Nehalem) virtualization features
  - Reduced latency for transitions
  - Virtual Processor ID (VPID) to reduce effective cost of transitions
  - Extended Page Table (EPT) to reduce # of transitions

Great virtualization performance with Intel® Core™ microarchitecture (Nehalem)

Latency of Virtualization Transitions

- Microarchitectural
  - Huge latency reduction generation over generation
  - Nehalem continues the trend
- Architectural
  - Virtual Processor ID (VPID) added in Intel® Core™ microarchitecture (Nehalem)
  - Removes need to flush TLBs on transitions

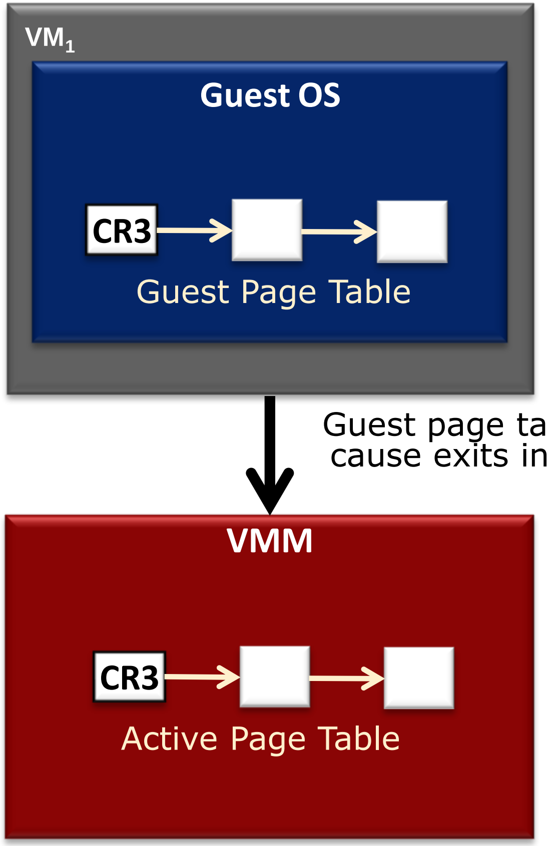
Round Trip Virtualization Latency [1]



Higher Virtualization Performance Through Lower Transition Latencies

[1] Intel® Core™ microarchitecture (formerly Merom)
45nm next generation Intel® Core™ microarchitecture (Penryn)
Intel® Core™ microarchitecture (Nehalem)

Extended Page Tables (EPT) Motivation

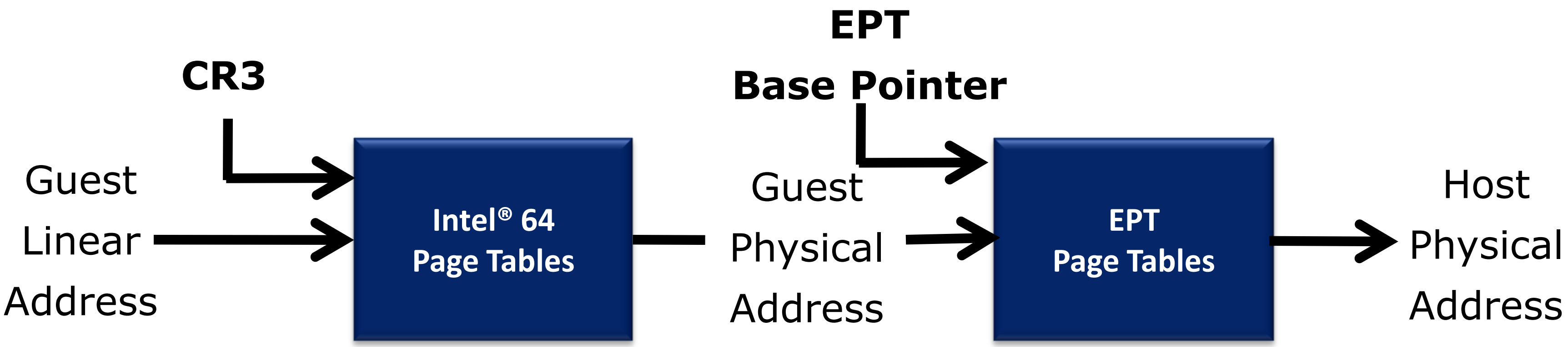


Guest page table changes cause exits into the VMM

VMM maintains the active page table, which is used by the CPU

- **A VMM needs to protect physical memory**
  - Multiple Guest OSs share the same physical memory
  - Protections are implemented through page-table virtualization
- **Page table virtualization accounts for a significant portion of virtualization overheads**
  - VM Exits / Entries
- **The goal of EPT is to reduce these overheads**

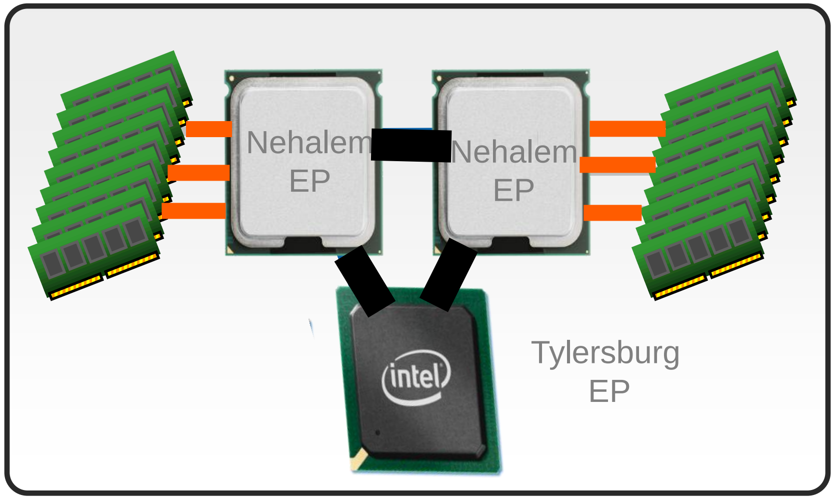
EPT Solution



- **Intel® 64 Page Tables**
  - Map Guest Linear Address to Guest Physical Address
  - Can be read and written by the guest OS
- **New EPT Page Tables under VMM Control**
  - Map Guest Physical Address to Host Physical Address
  - Referenced by new EPT base pointer
- **No VM Exits due to Page Faults, INVLPG or CR3 accesses**

Intel® Core™ Microarchitecture (Nehalem-EP) Platform Architecture

- Integrated Memory Controller
  - 3 DDR3 channels per socket
  - Massive memory ***bandwidth***
  - Memory Bandwidth scales with # of processors
  - Very ***low memory latency***
- Intel® QuickPath Interconnect (Intel® QPI)
  - New point-to-point interconnect
  - Socket to socket connections
  - Socket to chipset connections
  - Build ***scalable*** solutions



Significant performance leap from new platform

Intel® Core™ microarchitecture (Nehalem-EP)
Intel® Next Generation Server Processor Technology (Tylersburg-EP)

Intel Next-Generation Mainstream Processors[1]

Feature	Core™ i7	Lynnfield	Clarkdale	Clarksfield	Arrandale
Processing Threads [via Intel® Hyper-Threading Technology (HT)]	8	Up to 8	Up to 4	Up to 8	Up to 4
Processor Cores	4	4	2	4	2
Shared Cache	8MB	Up to 8MB	Up to 4MB	Up to 8MB	Up to 4MB
Integrated Memory Controller Channels	3 ch. DDR3	2 ch. DDR3			
DDR Freq Support (sku dependent)	800, 1066	1066, 1333			800, 1066
# DIMMs/Channels	2	2		1	
PCI Express* 2.0	2x16 or 4x8, 1x4 (via X58)	1x16 or 2x8	1x16 or 2x8	1x16 or 2x8	1x16 (1.0)
Processor Graphics	No	No	Yes	No	Yes
Processor Package TDP	130W	95W	73W	55W and 45W	35W, 25W, 18W
Socket	LGA 1366	LGA 1156		rPGA, BGA	
Platform Support	X58 & ICH10	Intel® 5 series Chipset			
Processor Core Process Technology	45nm	45nm	32nm	45nm	32nm

Bringing Intel® Core™ i7 Benefits into Mainstream

[1]Not all features are on all products, subject to change