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 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

<table>
<thead>
<tr>
<th></th>
<th>Merom(^1)</th>
<th>Penryn</th>
<th>Nehalem</th>
<th>Westmere</th>
<th>Sandy Bridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW</td>
<td>NEW</td>
<td>NEW</td>
<td>NEW</td>
<td>NEW</td>
<td>NEW</td>
</tr>
<tr>
<td>Microarchitecture</td>
<td>Microarchitecture</td>
<td>Microarchitecture</td>
<td>Microarchitecture</td>
<td>Microarchitecture</td>
<td></td>
</tr>
<tr>
<td>65nm</td>
<td>45nm</td>
<td>45nm</td>
<td>32nm</td>
<td>32nm</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)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

Consumer Desktop
- Learn
- Manageability
- Security
- Collaboration
- Cable free computing
- Connectivity

Office Desktop
- Play & Enjoy
- Multi-core
- Cell
- SMCA
- Seamless Wireless
- Increase Battery Life
- Anywhere, anytime computing

Genoa
- Service ability
- Reliability
- Availability

Desktop
- Media Processing
- Power/Perf Efficiency
- SMCA
- Reduce Form Factor

Tech Trends
- Graphics Processing
- 4/3/2/1 Core Scalability
- RAS features
- Threading
- TPCC Perf

Usage Trends
- Health & Wellness
- TPCC Perf
- RAS features
- Threading

Server
- Nehalem:
- Lots of competing aspects
- Learn
- Play & Enjoy
- Multi-core
- SMCA
- Seamless Wireless
- Anywhere, anytime computing

Mobile
- Nehalem:
- Lots of competing aspects
- Learn
- Play & Enjoy
- Multi-core
- SMCA
- Seamless Wireless
- Anywhere, anytime computing
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 envelopes

- 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

Workload Lightly Threaded or < TDP
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

Workload Lightly Threaded or < TDP

No Turbo

Core 0
Core 1
Core 2
Core 3

Frequency (F)
Intel® Core™ Microarchitecture (Nehalem) Turbo Mode

**No Turbo**

- Workload Lightly Threaded or < TDP

**Power Gating**
- Zero power for inactive cores

**Turbo Mode**
- In response to workload adds additional performance bins within headroom
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
- Execution Units
- L1 Data Cache
- L2 Cache & Interrupt Servicing
- Memory Ordering & Execution
- Paging
- Out-of-Order Scheduling & Retirement
- Instruction Decode & Microcode
- Branch Prediction
- Instruction Fetch & L1 Cache
- Deeper Buffers
- Improved Loop Streaming
- Simultaneous Multi-Threading
- Faster Virtualization
- Better Branch Prediction
The First Intel® Core™ Microarchitecture (Nehalem) Processor

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

Relative Performance
Higher is better

Xeon 5500 vs Xeon 5400 on Server Benchmarks

Baseline
 SPECjbb* 2005
 SPECint*_rate base2006
 SPECpower*_ssj2008
 SPECjvm* 2008
 SPECApp*_Server2004
 SAP-SD*_2-Tier
 SPECfp*_rate_base2006
 TPC*-C
 TPC*-E
 SPECWeb* 2005
 VMmark*

Xeon 5400 series
 Server Side Java
 Integer
 Energy Efficiency
 Java Apps
 App Server
 ERP
 Floating point
 Database
 Database
 Web
 Virtualization

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](http://www.intel.com/performance/resources/limits.htm) Copyright © 2009, Intel Corporation. * Other names and brands may be claimed as the property of others.
### Intel® Xeon® Processor 5500 series based Server platforms

#### HPC Performance comparison to Xeon 5400 Series

#### Xeon 5500 vs Xeon 5400 on HPC Benchmarks

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Relative Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1.00</td>
</tr>
<tr>
<td>Weather</td>
<td>1.94</td>
</tr>
<tr>
<td>FEA</td>
<td>2.12</td>
</tr>
<tr>
<td>FEA</td>
<td>2.15</td>
</tr>
<tr>
<td>CFD</td>
<td>2.17</td>
</tr>
<tr>
<td>CFD</td>
<td>2.27</td>
</tr>
<tr>
<td>Energy</td>
<td>2.39</td>
</tr>
<tr>
<td>Open MP</td>
<td>2.55</td>
</tr>
<tr>
<td>Energy</td>
<td>2.54</td>
</tr>
<tr>
<td>Open MP</td>
<td>2.92</td>
</tr>
<tr>
<td>Weather</td>
<td>2.95</td>
</tr>
<tr>
<td>Energy</td>
<td>2.96</td>
</tr>
</tbody>
</table>

- **Baseline**: MM5* v4.7.4 t3a - LS-DYNA* - 3 Vehicle Collision - ANSYS* - Distributed - Star-CD* A-Class - ANSYS* FLUENT* 12.0 bmk - CMG IMEX* - SPECompM* base2001 - Eclipse* - 300/2mm - SPECompL* base2001 - WRF* v3.0.1 - 12km CONUS - Landmark* Nexus

**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](http://www.intel.com/performance/resources/limits.htm) Copyright © 2009, Intel Corporation. * Other names and brands may be claimed as the property of others.
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](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](http://www.intel.com).
### ISV Application Performance

<table>
<thead>
<tr>
<th>ISV Application</th>
<th>Xeon 5500 vs Xeon 5400</th>
<th>Performance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ansys* CFX11* – CFD Simulation</td>
<td>ANSYS</td>
<td>+88%</td>
</tr>
<tr>
<td>ERDAS* ERMapper Suite*</td>
<td>erdas</td>
<td>+69%</td>
</tr>
<tr>
<td>ESI Group* PAM-CRASH*</td>
<td>ESI GROUP</td>
<td>+50%</td>
</tr>
<tr>
<td>ExitGames* Neutron* – Service Platform</td>
<td></td>
<td>+80%</td>
</tr>
<tr>
<td>Epic* – EMR Solution</td>
<td>Epic</td>
<td>+82%</td>
</tr>
<tr>
<td>Giant* Juren* – Online Game</td>
<td></td>
<td>+160%</td>
</tr>
<tr>
<td>IBM* DB2* 9.5 – TPoX XML</td>
<td>IBM Information Management software</td>
<td>+60%</td>
</tr>
<tr>
<td>IBM* Informix* Dynamic Server</td>
<td>IBM Information Management software</td>
<td>+84%</td>
</tr>
<tr>
<td>IBM* solidDB* – In Memory DB</td>
<td>IBM Information Management software</td>
<td>+87%</td>
</tr>
<tr>
<td>Image Analyzer* – Image Scanner</td>
<td></td>
<td>+100%</td>
</tr>
<tr>
<td>Infowrap* – Small Data Set</td>
<td>Infowrap</td>
<td>+129%</td>
</tr>
<tr>
<td>Infowrap* – Weather Forecast</td>
<td>Infowrap</td>
<td>+155%</td>
</tr>
<tr>
<td>Intense* IECCM* – Output Mgmt.</td>
<td>VMware, Lotus Notes, SIS</td>
<td>+150%</td>
</tr>
<tr>
<td>Intersystems* Cache* – EMR</td>
<td>Intersystems</td>
<td>+63%</td>
</tr>
<tr>
<td>Kingdee* APUSIC* – Middleware</td>
<td>Kingdee</td>
<td>+93%</td>
</tr>
<tr>
<td>Kingdee* EAS* – ERP</td>
<td>Kingdee</td>
<td>+142%</td>
</tr>
<tr>
<td>Kingdom* – Stock Transaction</td>
<td></td>
<td>+141%</td>
</tr>
</tbody>
</table>

### ISV Application Performance

<table>
<thead>
<tr>
<th>ISV Application</th>
<th>Xeon 5500 vs Xeon 5400</th>
<th>Performance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kingsoft* JXIII Online* – Game Server</td>
<td></td>
<td>+98%</td>
</tr>
<tr>
<td>Mediarware* Instream* – Video Conv.</td>
<td>Mediarware, Instream</td>
<td>+73%</td>
</tr>
<tr>
<td>Neowiz* Pmang* – Game Portal</td>
<td>Xen, Microsoft Software</td>
<td>+100%</td>
</tr>
<tr>
<td>Neusoft* – Healthcare</td>
<td>Neusoft</td>
<td>+131%</td>
</tr>
<tr>
<td>Neusoft* – Telecom BSS VMware*</td>
<td>VMware, Neusoft</td>
<td>+115%</td>
</tr>
<tr>
<td>NHN Corp* Cubrid* – Internet DB</td>
<td>NHN Corporation</td>
<td>+44%</td>
</tr>
<tr>
<td>QlikTech* QlikView* – BI</td>
<td>QlikView</td>
<td>+36%</td>
</tr>
<tr>
<td>SAS* Forecast Server*</td>
<td>SAS</td>
<td>+80%</td>
</tr>
<tr>
<td>SAP* NetWeaver* – BI</td>
<td>SAP</td>
<td>+51%</td>
</tr>
<tr>
<td>SAP* ECC 6.0* – ERP Workload</td>
<td>VMware, SAP</td>
<td>+71%</td>
</tr>
<tr>
<td>Schlumberger* Eclipse300*</td>
<td>Schlumberger</td>
<td>+213%</td>
</tr>
<tr>
<td>SunGard* BancWare Focus ALM*</td>
<td>SONGARD</td>
<td>+38%</td>
</tr>
<tr>
<td>Supcon* – APC Intel. Sensor</td>
<td>Supcon, SUPCON</td>
<td>+191%</td>
</tr>
<tr>
<td>TongTech* – Middleware</td>
<td>TongTech</td>
<td>+95%</td>
</tr>
<tr>
<td>UFIDA* NC* – ERP Solution</td>
<td>UFIDA</td>
<td>+230%</td>
</tr>
<tr>
<td>UFIDA* Online – SaaS Hosting</td>
<td>UFIDA</td>
<td>+237%</td>
</tr>
<tr>
<td>Vital Images* – Brain Perfusion 4DCT*</td>
<td></td>
<td>+77%</td>
</tr>
</tbody>
</table>

**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 the performance of Intel products, visit [http://www.intel.com/technology/resources/limits.htm](http://www.intel.com/technology/resources/limits.htm) Copyright © 2009, Intel Corporation. * Other names and brands may be claimed as the property of others.
Execute 6 operations/cycle
• 3 Memory Operations
  • 1 Load
  • 1 Store Address
  • 1 Store Data
• 3 “Computational” Operations

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

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

### Increased Resources for Higher Performance

<table>
<thead>
<tr>
<th>Structure</th>
<th>Intel® Core™ microarchitecture (formerly Merom)</th>
<th>Intel® Core™ microarchitecture (Nehalem)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reservation Station</td>
<td>32</td>
<td>36</td>
<td>Dispatches operations to execution units</td>
</tr>
<tr>
<td>Load Buffers</td>
<td>32</td>
<td>48</td>
<td>Tracks all load operations allocated</td>
</tr>
<tr>
<td>Store Buffers</td>
<td>20</td>
<td>32</td>
<td>Tracks all store operations allocated</td>
</tr>
</tbody>
</table>

1Intel® 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 2\textsuperscript{nd} level TLB

<table>
<thead>
<tr>
<th></th>
<th># of Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1\textsuperscript{st} Level Instruction TLBs</td>
<td></td>
</tr>
<tr>
<td>Small Page (4k)</td>
<td>128</td>
</tr>
<tr>
<td>Large Page (2M/4M)</td>
<td>7 per thread</td>
</tr>
<tr>
<td>1\textsuperscript{st} Level Data TLBs</td>
<td></td>
</tr>
<tr>
<td>Small Page (4k)</td>
<td>64</td>
</tr>
<tr>
<td>Large Page (2M/4M)</td>
<td>32</td>
</tr>
<tr>
<td>New 2\textsuperscript{nd} Level Unified TLB</td>
<td></td>
</tr>
<tr>
<td>Small Page Only</td>
<td>512</td>
</tr>
</tbody>
</table>
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

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

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

<table>
<thead>
<tr>
<th>Hyper-Threading</th>
<th>Multi-cores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shared or Partitioned</td>
</tr>
<tr>
<td>Register State</td>
<td>X</td>
</tr>
<tr>
<td>Return Stack</td>
<td>X</td>
</tr>
<tr>
<td>Reorder Buffer</td>
<td>X</td>
</tr>
<tr>
<td>Instruction TLB</td>
<td>X</td>
</tr>
<tr>
<td>Reservation Stations</td>
<td>X</td>
</tr>
<tr>
<td>Cache (L1, L2)</td>
<td>X</td>
</tr>
<tr>
<td>Data TLB</td>
<td>X</td>
</tr>
<tr>
<td>Execution Units</td>
<td>X</td>
</tr>
</tbody>
</table>

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

Intel® Microarchitecture codenamed Nehalem
Caches

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

- Shared across all cores
- Size depends on \# of cores
  - Quad-core: Up to 8MB (16-ways)
  - \textit{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 \textbf{performance}
  - Address residing in L1/L2 \textit{must} be present in 3\textsuperscript{rd} 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 prefetcheds can conduct page walks
- SW prefetcheds can spawn HW prefetcheds
  - 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

**Massive memory BW provides performance and scalability**

---

Source: Intel Internal measurements – August 2008

1HTN: 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

---

1Next generation Quad-Core Intel® Xeon® processor (Harpertown)
Intel® Core™ 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 (VPIID) 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

Higher Virtualization Performance Through Lower Transition Latencies

1. Intel® Core™ microarchitecture (formerly Merom)
2. 45nm next generation Intel® Core™ microarchitecture (Penny)
3. Intel® Core™ microarchitecture (Nehalem)
Extended Page Tables (EPT) Motivation

- 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

---

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:

VMM maintains the active page table, which is used by the CPU.
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 Next-Generation Mainstream Processors

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

## Bringing Intel® Core™ i7 Benefits into Mainstream

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