Intelligent and Expandable High-End Intel® Server Platform, Codenamed Nehalem-EX

Steve Pawlowski
Intel Senior Fellow
General Manager, Central Architecture and Planning
Intel Architecture Group
Agenda

- Scalable Nehalem-EX Architecture
- Reliability, Availability and Serviceability
- The Platform for High-End Computing
Intel® Xeon® Processor - EXpandable (EX) Segment Roadmap

2007: Intel Xeon 7300
2008: Intel Xeon 7400
2009: Nehalem-EX
Future: Westmere-EX

Expandability and Scalability
Reliability, Availability, and Serviceability (RAS)

Designed for Demanding Apps and Large-Scale Consolidation

Westmere = Intel® microarchitecture, codenamed Westmere
Nehalem-EX (NHM-EX) = High-End Intel® Server Platform, Codenamed Nehalem-EX
Scalable Processor Cores

- Up to 8 cores per socket
- Intel® Hyper-Threading Technology (2 threads/core)
- 24MB shared last level cache
- Intel® Turbo Boost Technology for performance boost
- CPU TDP: 130W; 105W; 95W

Dynamically Scalable Micro-architecture
Scalable Sockets 2, 4, and up to 8+

Sample Platform Configurations

2 Processors

4 Processors

8 Processors

Glueless
3 peers at 1 hop
4 peers at 2 hops
Scalable Memory Subsystem

**Integrated Memory Controller (IMC):**
- 2 IMCs per socket
- High capacity / bandwidth per controller:
  - Up to 8 DIMMs per memory controller
  - 4 DDR Channels for two SMI interface

**Scalable Memory Interface (SMI):**
- High speed serial links for maximum bandwidth
- Run in lockstep mode to minimize latencies & enable RAS

**Scalable Memory Buffer (SMB):**
- Memory buffer is on board or memory riser
- 2 DDR channels & up to 4 DIMMs per buffer

**Memory Capacity:** Max population with 16GB DIMMS = 256 GB / Socket

**Theoretical Peak Bandwidth:** 50 GB/s/Socket
Scalable I/O PCIe Interconnect

PCI Express* (PCIe) Technology Configurations

Scalability:
- 2 Sockets – up to 2 I/O Hubs = (72+10) Lanes
- 4 Sockets – 2 or 3 I/O Hubs = (72+10) or (108+14) Lanes
- 8 Sockets – 4 I/O Hubs = (144 + 18) Lanes
Platform Capability Enhancement

Architectural Innovations:
- Front Side Bus (FSB) to Intel® QuickPath Interconnect with integrated memory controller
- FB-DIMM to DDR3
- PCI Express* Technology (PCIe) 1.0 to 2.0...

Note: Peak Memory BW represents theoretical peak; not actual measurement
Mission Critical Platform Requirement

Systems that

- Are always available
- Never corrupt data
- Deliver consistent performance
- Allow maintenance while operating

Reliability, Availability and Serviceability (RAS) are Crucial ...
RAS at Every Part of the Platform

Intel® QPI RAS

IO Hub (IOH) RAS

Memory RAS

Socket RAS
- Corrected Machine Check Interrupts (CMCI)
- Recoverable Machine Check Architecture (MCA)

Note: These are only examples, not comprehensive
Hardware Correctable Errors

Signaling Corrected Errors Using Corrected Machine Check Interrupts (CMCI)

Architected CMCI Support

Enables software value add through predictive failure solutions

Intel® QPI = Intel® QuickPath Interconnect
Nehalem = Intel® microarchitecture, codenamed Nehalem
Un-core Uncorrected Error Containment and Recovery

Recovery is for the system to work with the OS/firmware to recover from errors detected in the platform uncorrectable by hardware. The ability to recover depends on the type of error and state of the system.

QPI = Intel® QuickPath Interconnect (Intel® QPI)
PCIe = PCI Express* Technology
Recoverable Machine Check Architecture

System works in conjunction with firmware and OS to recover or restart processes to continue normal operation.

- **System Recovery**
  - Isolate/replace defective HW or add new resources without bringing down a system.
- **Service Reconfiguration**
- **Error Detected**
  - Un-correctable Errors
  - HW Correctable Errors

- **Normal Status With Error Prevention**
- **Error Corrected**
- **Error Contained**

First Machine Check Recovery in Intel® Xeon® Processor Based Systems
User Benefits

Increase system up time – important for servers
- Allows for OS to terminate/restart an application mapped to that address or the VMM to terminate the guest OS
- System remains active running other applications or guest OSs

Errors are detected ahead of software consumption
Provide software an opportunity to attempt to recover from an uncorrected error before the error brings down the machine.

Uncorrected errors detected outside of program execution have potential for error recovery.
The Platform for High-end Computing

Compute Virtualization
Intel® VT-x, Intel® VT-d

New Challenges in Computing
Massive Scaling
Multi-tenancy
Low Total Cost of Ownership (TCO)

Network Virtualization
Intel® VT-c

Power Optimization

Intel® Virtualization Technology (Intel® VT) for IA-32, Intel® 64 and Intel® Architecture (Intel® VT-x)
Intel® Virtualization Technology (Intel® VT) for Directed IO (Intel® VT-d)
Intel® Virtualization Technology (Intel® VT) for Connectivity (Intel® VT-c)
Nehalem-EX Virtualization Features

Virtual Machines (VMs)

VMM (a.k.a., hypervisor)

Processor Virtualization
- Virtual CPU Configuration
- Virtual CPU Configuration

Memory Virtualization
- EPT Configuration

I/O Device Virtualization
- DMA and Interrupt Remapping Configuration

Virtual CPUs

Higher-level VMM Functions:
- Resource Discovery / Provisioning / Scheduling / User Interface

Physical Platform Resources

EPT – VT-x Enhanced Page Tables

CPU

VT-x

EPT

VT-d

VT-c

Storage

Network

Intel® Virtualization Technology (Intel® VT) for IA-32, Intel® 64 and Intel® Architecture (Intel® VT-x)
Intel® Virtualization Technology (Intel® VT) for Directed IO (Intel® VT-d)
Intel® Virtualization Technology (Intel® VT) for Connectivity (Intel® VT-c)

Nehalem-EX (NHM-EX) = High-End Intel® Server Platform, Codenamed Nehalem-EX

IDF2009
Compute Virtualization (VT-x)

Core CPU: Context Switching

VM Context Switch Latencies (Cycles)

With Virtualization: Entirely new state to switch...
Network Virtualization (Intel® VT-c)

Virtual Machine Device Queues (VMDq)
- Data packets sorted in hardware
- Packets sent to their respective VMs
- Round-robin servicing on transmit

Throughput (GB)

<table>
<thead>
<tr>
<th>Throughput (GB)</th>
<th>w/o VMDq</th>
<th>w/ VMDq</th>
<th>w/ VMDq Jumbo Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>2x throughput</td>
<td>9.2</td>
<td>9.5</td>
</tr>
<tr>
<td>8.0</td>
<td>w/o VMDq</td>
<td>w/ VMDq</td>
<td>w/ VMDq Jumbo Frames</td>
</tr>
<tr>
<td>6.0</td>
<td>2x throughput</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>2x throughput</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>2x throughput</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>2x throughput</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Tests measure Wire Speed Receive (Rx) Side Performance With VMDq on Intel® 82598 10 Gigabit Ethernet Controller

VMDq + VMware* NetQueue

Intel® Virtualization Technology (Intel® VT) for Connectivity (Intel® VT-c)
Power Optimization

Leakage Control at Core and Cache Level

New to NHM-EX: Power gating disabled cores

Infrared back-side emission shows core leakage suppressed and cache leakage reduced

Core Level: 40x leakage reduction when shut off
Cache Level: 35% during sleep & 83% when shut off

QPI = Intel® QuickPath Interconnect (Intel® QPI)
SMI = Scalable Memory Interconnect
Nehalem-EX (NHM-EX) = High-End Intel® Server Platform, Codenamed Nehalem-EX
Socket Level Power Savings

Power-up detector senses unused links and disables them to save power (~2W per port)

- Shut-off driver bias current
- Turn-off the PLL to stop the clock
Intelligent and Expandable High-End Intel® Server Platform

<table>
<thead>
<tr>
<th>Business Driver</th>
<th>Performance</th>
<th>Nehalem-EX vs. Intel Xeon® 7400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Memory Bandwidth</td>
<td>Up to 9X</td>
</tr>
<tr>
<td></td>
<td>Database Performance</td>
<td>2.5X</td>
</tr>
<tr>
<td></td>
<td>Integer Throughput</td>
<td>1.7X</td>
</tr>
<tr>
<td></td>
<td>FP Throughput</td>
<td>2.2X</td>
</tr>
</tbody>
</table>

Note: The comparison is based on 6-socket Intel Xeon®7400 and 8-socket Nehalem-EX

Nehalem-EX (NHM-EX)= High-End Intel® Server Platform, Codenamed Nehalem-EX
Risk Factors

The above statements and any others in this document that refer to plans and expectations for the third quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Many factors could affect Intel’s actual results, and variances from Intel’s current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be the important factors that could cause actual results to differ materially from the corporation’s expectations. Ongoing uncertainty in global economic conditions pose a risk to the overall economy as consumers and businesses may defer purchases in response to tighter credit and negative financial news, which could negatively affect product demand and other related matters. Consequently, demand could be different from Intel’s expectations due to factors including changes in business and economic conditions, including conditions in the credit market that could affect consumer confidence; customer acceptance of Intel’s and competitors’ products; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel operates in intensely competitive industries that are characterized by a high percentage of costs that are fixed or difficult to reduce in the short term and product demand that is highly variable and difficult to forecast. Additionally, Intel is in the process of transitioning to its next generation of products on 32nm process technology, and there could be execution issues associated with these changes, including product defects and errata along with lower than anticipated manufacturing yields. Revenue and the gross margin percentage are affected by the timing of new Intel product introductions and the demand for and market acceptance of Intel’s products; actions taken by Intel's competitors, including product offerings and introductions, marketing programs and pricing pressures and Intel’s response to such actions; and Intel’s ability to respond quickly to technological developments and to incorporate new features into its products. The gross margin percentage could vary significantly from expectations based on changes in revenue levels; capacity utilization; start-up costs, including costs associated with the new 32nm process technology; variations in inventory valuation, including variations related to the timing of qualifying products for sale; excess or obsolete inventory; product mix and pricing; manufacturing yields; changes in unit costs; impairments of long-lived assets, including manufacturing, assembly/test and intangible assets; and the timing and execution of the manufacturing ramp and associated costs. Expenses, particularly certain marketing and compensation expenses, as well as restructuring and asset impairment charges, vary depending on the level of demand for Intel's products and the level of revenue and profits. The current financial stress affecting the banking system and financial markets and the going concern threats to investment banks and other financial institutions have resulted in a tightening in the credit markets, a reduced level of liquidity in many financial markets, and heightened volatility in fixed income, credit and equity markets. There could be a number of follow-on effects from the credit crisis on Intel’s business, including insolvency of key suppliers resulting in product delays; inability of customers to obtain credit to finance purchases of our products and/or customer insolvencies; counterparty failures negatively impacting our treasury operations; increased expense or inability to obtain short-term financing of Intel’s operations from the issuance of commercial paper; and increased impairments from the inability of investee companies to obtain financing. The majority of our non-marketable equity investment portfolio balance is concentrated in companies in the flash memory market segment, and declines in this market segment or changes in management’s plans with respect to our investments in this market segment could result in significant impairment charges, impacting restructuring charges as well as gains/losses on equity investments and interest and other. Intel’s results could be impacted by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Intel's results could be affected by adverse effects associated with product defects and errata (deviations from published specifications), and by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust and other issues, such as the litigation and regulatory matters described in Intel’s SEC reports. A detailed discussion of these and other risk factors that could affect Intel’s results is included in Intel’s SEC filings, including the report on Form 10-Q for the quarter ended June 27, 2009.
Legal Disclaimer

• INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

• UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

• Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

• The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

• Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

• Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web Site.