Nehalem-EX CPU Architecture

Sailesh Kottapalli, Jeff Baxter

NHM-EX Architecture

Intel Confidential

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. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.
- Intel may make changes to specifications and product descriptions at any time, without notice.
- All products, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.
- Intel, processors, chipsets, and desktop boards 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.
- Nehalem, Merom, Boxboro, Millbrook, Penryn, Westmere, Sandy Bridge and other code names featured are used internally within Intel to identify products that are in development and not yet publicly announced for release. Customers, licensees and other third parties are not authorized by Intel to use code names in advertising, promotion or marketing of any product or services and any such use of Intel's internal code names is at the sole risk of the user
- 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.
- Intel, Intel Inside, Intel Core, Intel Xeon, Intel Core2 and the Intel logo are trademarks of Intel Corporation in the United States and other countries.
- *Other names and brands may be claimed as the property of others.
- Copyright © 2008 Intel Corporation.

Risk Factors

This presentation contains forward-looking statements that involve a number of risks and uncertainties. These statements do not reflect the potential impact of any mergers, acquisitions, divestitures, investments or other similar transactions that may be completed in the future. The information presented is accurate only as of today's date and will not be updated. In addition to any factors discussed in the presentation, the important factors that could cause actual results to differ materially include the following: 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's results could be affected by the timing of closing of acquisitions and divestitures. 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. 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; Intel's ability to respond quickly to technological developments and to incorporate new features into its products; and the availability of sufficient supply of components from suppliers to meet demand. The gross margin percentage could vary significantly from expectations based on changes in revenue levels; product mix and pricing; capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; excess or obsolete inventory; 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, including start-up costs. Expenses, particularly certain marketing and compensation expenses, vary depending on the level of demand for Intel's products, the level of revenue and profits, and impairments of long-lived assets. Intel is in the midst of a structure and efficiency program that is resulting in several actions that could have an impact on expected expense levels and gross margin. Intel's results could be impacted by adverse economic, social, political and physical/infrastructure conditions in the countries in which 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 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.

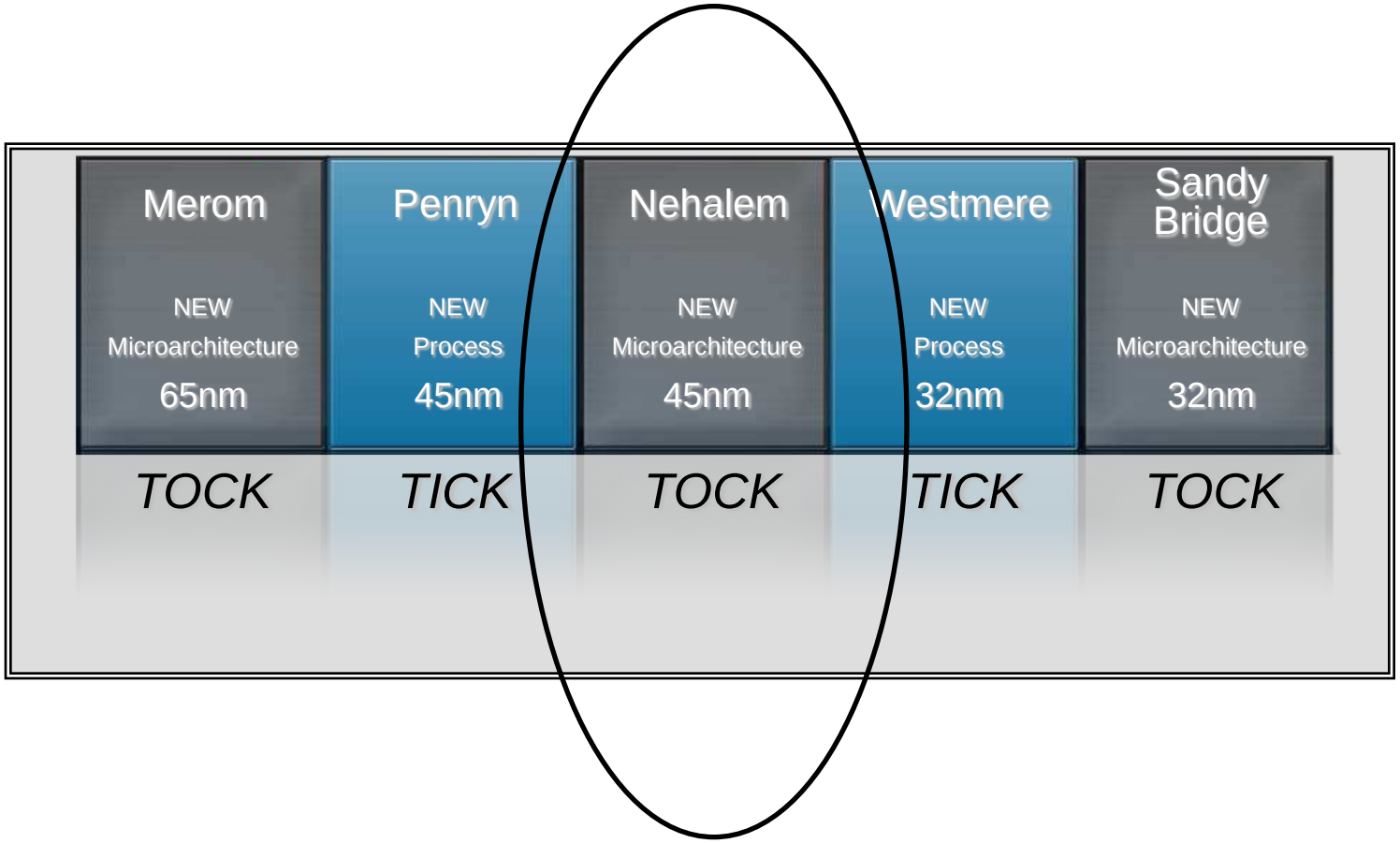
Today's Talk

- Nehalem-EX (NHM-EX) Processor Architecture
  - Focus on dataflow throughout the CPU
  - Focus on the bandwidth/latency/scalability aspects
- Things we are not disclosing today
  - Product clock Speeds
  - Detailed performance
  - Overall Intel Server Roadmap
- Minimize overlap with existing public information
  - Pointers to website links

Agenda: Building the Beast

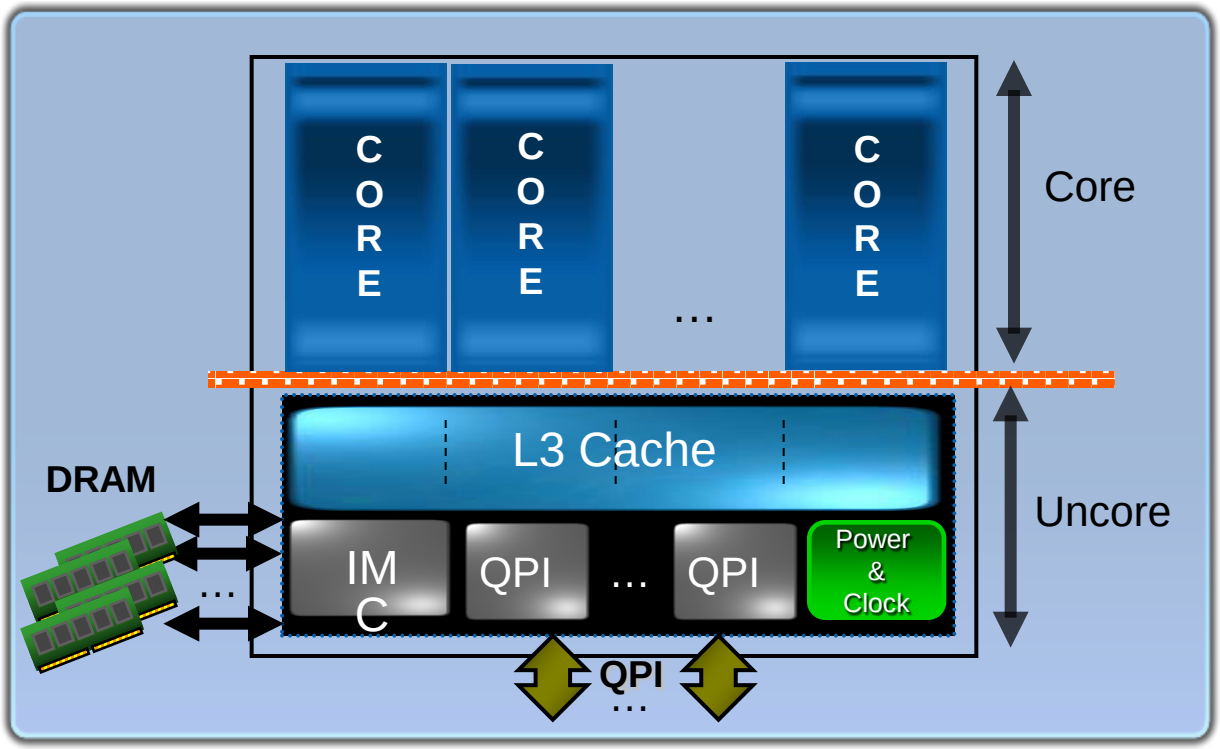
- Start at the core and build outward
- Last Level Cache
- System Agent
- Coherence Agent
- Memory Controller
- System Interconnect
- Performance
- Conclusion

Tick-Tock Development Model



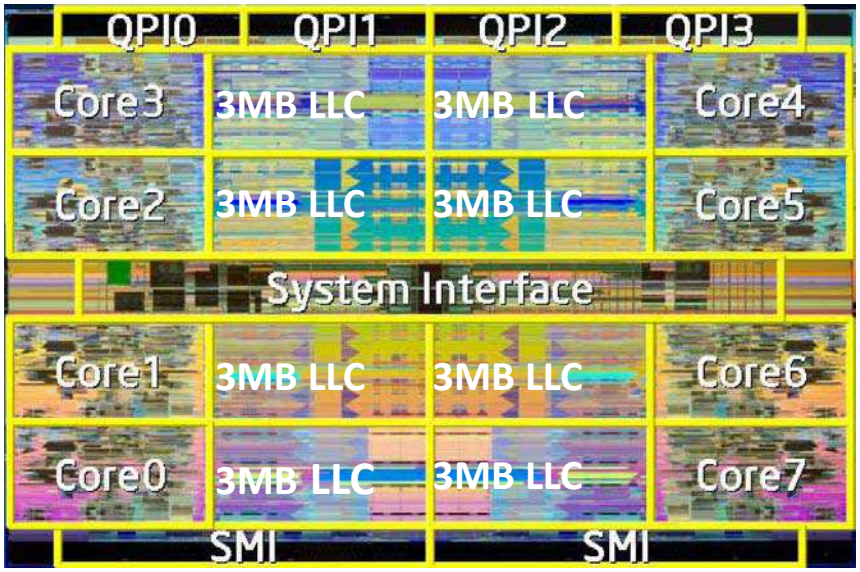
All products, dates, and figures are preliminary and are subject to change without notice.

Nehalem Core/Uncore Modularity



- Common core for client and server CPUs
  - http://www.intel.com/technology/architecture-silicon/next-gen/whitepaper.pdf
  - Some unique features only on NHM-EX
- Uncore differentiates different segment specific CPUs
  - Scalable Core/Uncore gasket interface
  - Decouples core and uncore operation

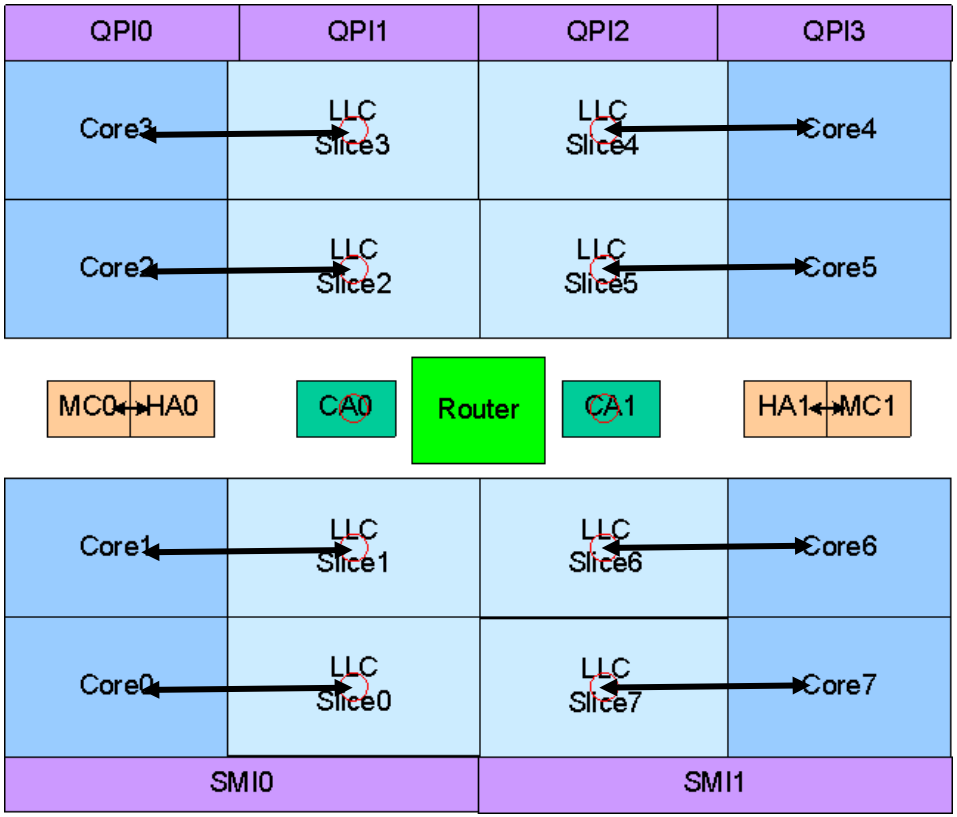
Nehalem-EX CPU



- Monolithic single die CPU
- 8 Nehalem cores, 16 threads
- 24MB shared L3 cache
- 2 integrated memory controllers
- Scalable Memory Interconnect (SMI) with support for up to 8 DDR channels
- 4 Quick Path Interconnect (QPI) links with up to 6.4GT/s
- Supports 2, 4 and 8 socket in glueless configs and larger systems using Node Controller (NC)
- Intel 45nm process technology
- 2.3 Billion transistors

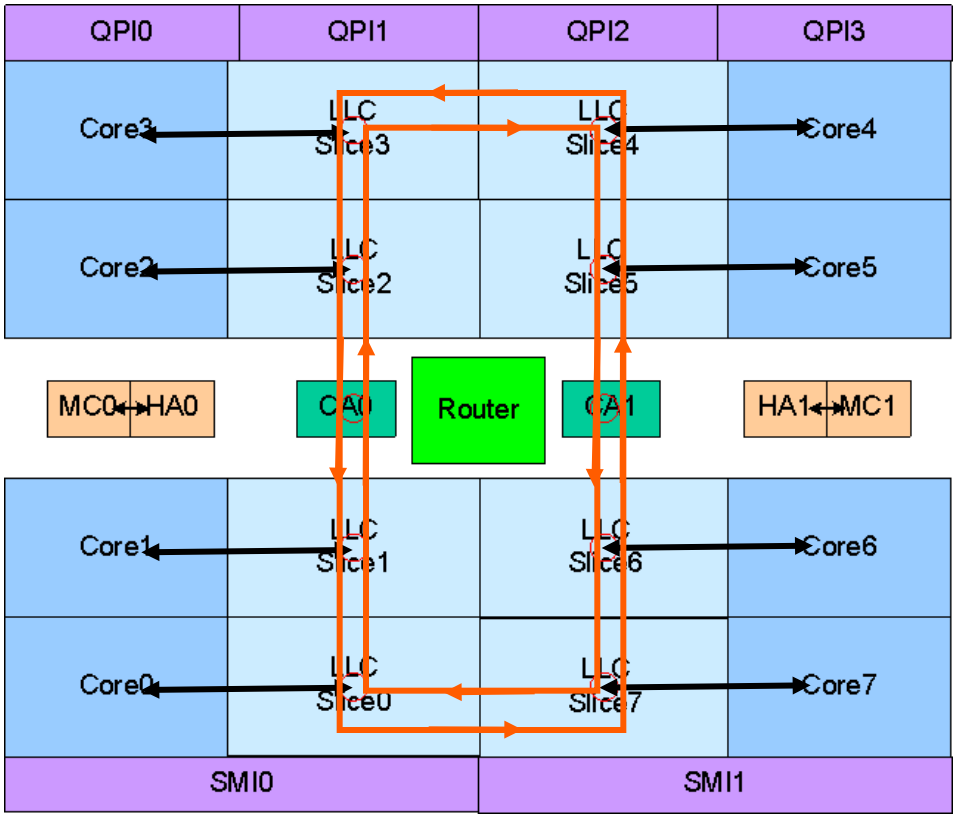
Cores and Last Level Caches

- Start with 8 Nehalem Cores
  - Modular and high throughput
  - 16 outstanding requests
- Large shared LLC
  - 24MB optimal mix of cache hierarchy capacity scaling as well as memory BW reduction
- Distributed, partitioned LLC for high BW
  - One cache slice per core
  - Multiple independent accesses in parallel
    - 4 per clock
  - PA hashed across slices to avoid hot-spots
- Inclusive LLC
  - Per-core valid bit tracks each line
  - Minimizes the amount of back snooping for last level cache WBs
  - Minimizes core snoops on external requests
  - Improves scalability and snoop latency



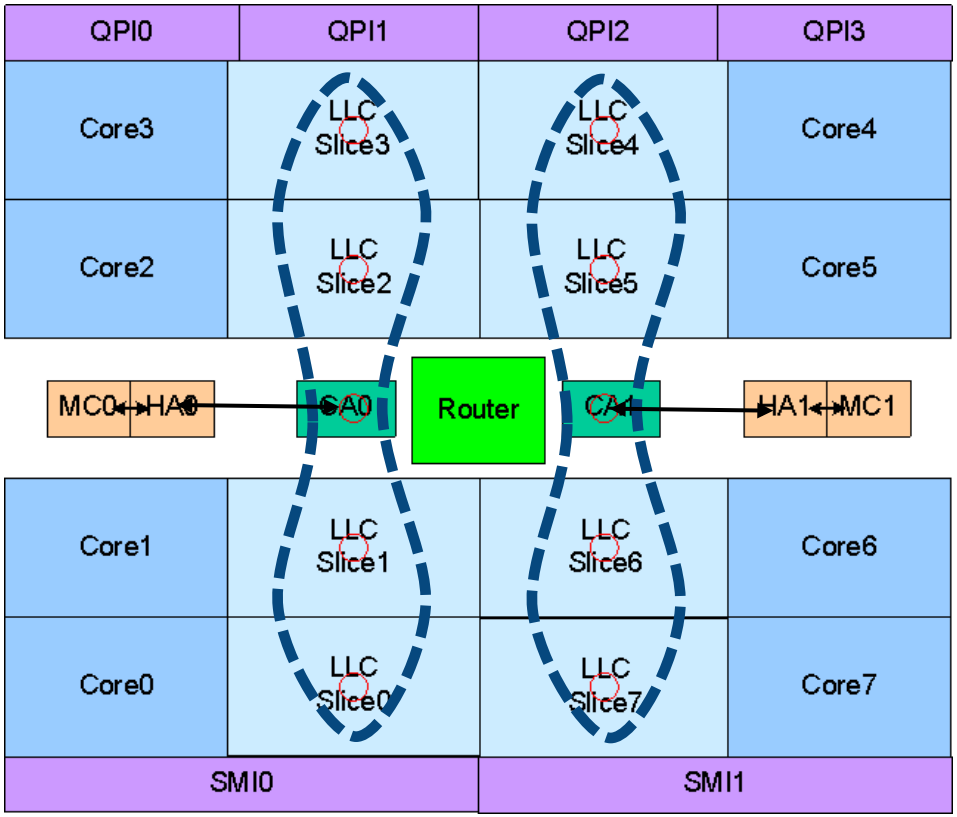
On-die network: Ring Interconnect

- High BW Ring Interconnect
  - Two counter rotating rings
  - Average latency ½ of unidirectional ring
  - 4X the BW of unidirectional ring
  - 4 protocol rings
    - Data ring 32Bytes in each direction
- Simple arbitration rules
  - Ring advances one stop per clock
  - "rotary rules": traffic on the ring wins arbitration
  - Ring stops shared by core/cache
  - Ring stops tagged as "even"/"odd" polarities
  - Ring stops only sink from a single ring/polarity per cycle
  - Ring traffic injector responsible for polarity match at target polarity
- Scalable fabric
  - BW scales with ring stops
  - Simulated >250GB/s of interconnect BW



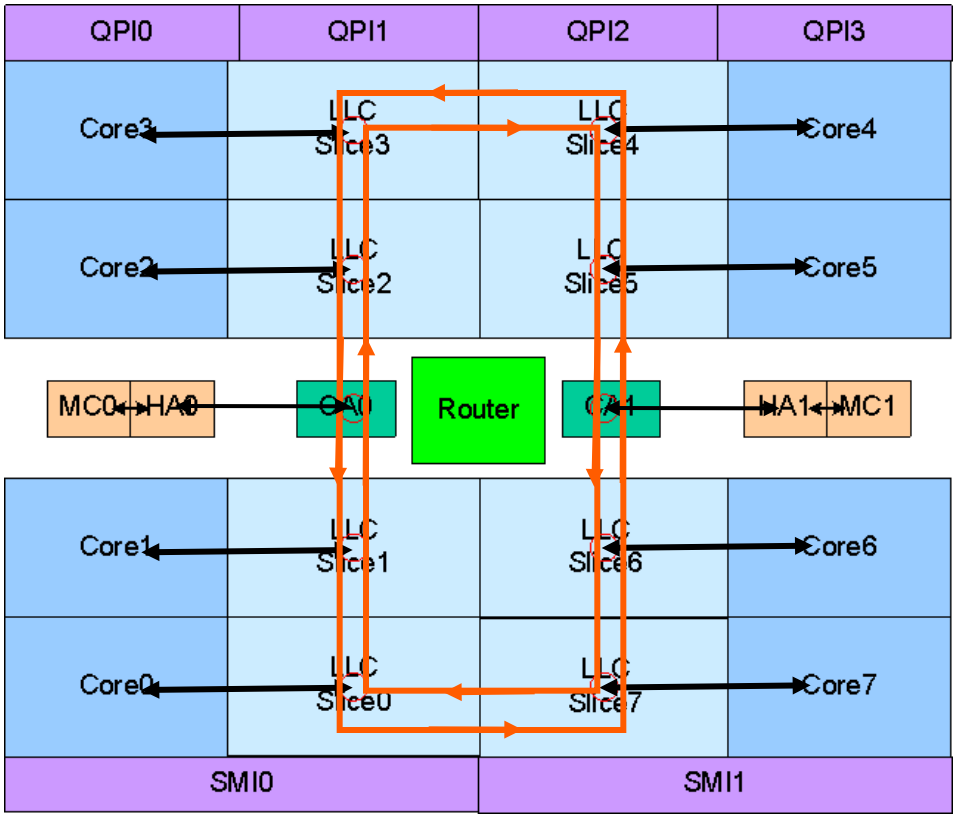
QPI Caching agent

- Dual Caching agent
  - QPI source snoop protocol
  - CA 0/CA1 act as caching agent hubs to the rest of the system
    - QPI link layer end points
  - Processes inbound snoops
- Local memory bypass
  - Request to local memory controller
  - Direct fill to core for latency reduction
- Supports high throughput
  - 48 (96 total) requests outstanding.
  - All outstanding requests can be mapped to local socket
  - Up to 120 snoops buffered beyond QPI Link Layer



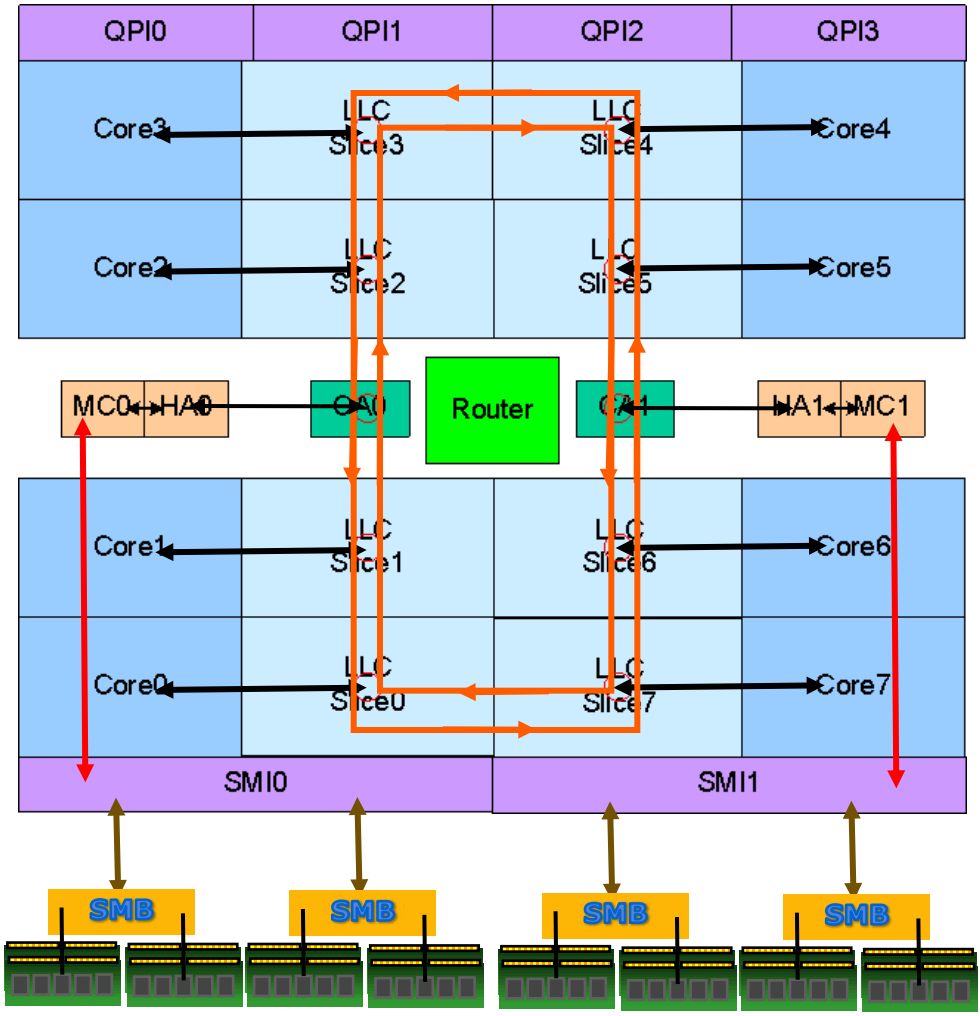
QPI Home Agent

- Dual Home agents
  - System coherence manager for memory associated with the adjacent memory controller
  - Tracks requests, resolves conflicts, returns memory data and completes transactions
  - Supports snoopy and hybrid coherence protocol.
    - IO agents tracked by in-memory directory
- Built to support large systems
  - Supports 256 outstanding requests
    - 48 requests from any single caching agent
  - Multiple modes for flexible topologies
    - Manages up to 8 other sockets and up to 4 IO/NC hubs
- Latency optimizations
  - Memory prefetch
  - Write-posting



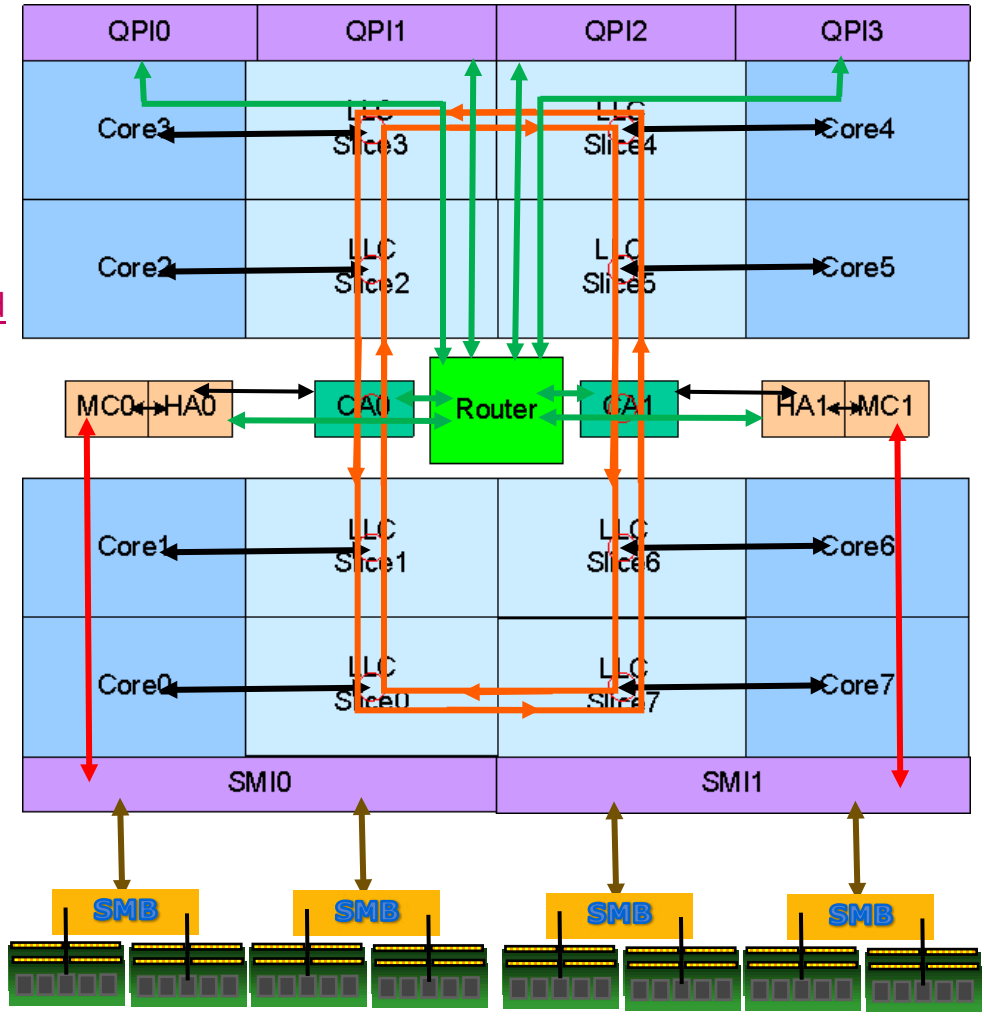
Memory Controller and Memory subsystem

- Dual memory controllers
- Buffered interface
  - Two Scalable Memory Interface (SMI) interfaces per memory controller
    - Run in lockstep mode to minimize latencies & enable RAS
  - High speed serial links for maximum bandwidth (6:1 ratio)
  - Each channel connects to an Scalable Memory Buffer (SMB)
    - 2 DDR3 channels per buffer
    - Up to 4 DIMMs per buffer chip
- Large capacity/bandwidth per controller
  - Up to 16 DIMMS/64 Ranks per socket
  - Up to 8 DDR channels per socket
- Scheduler optimized for latency and BW
  - 32 (64 total) simultaneous requests can bid
  - Open, close and adaptive page policy
  - Out of order, rank granular scheduling

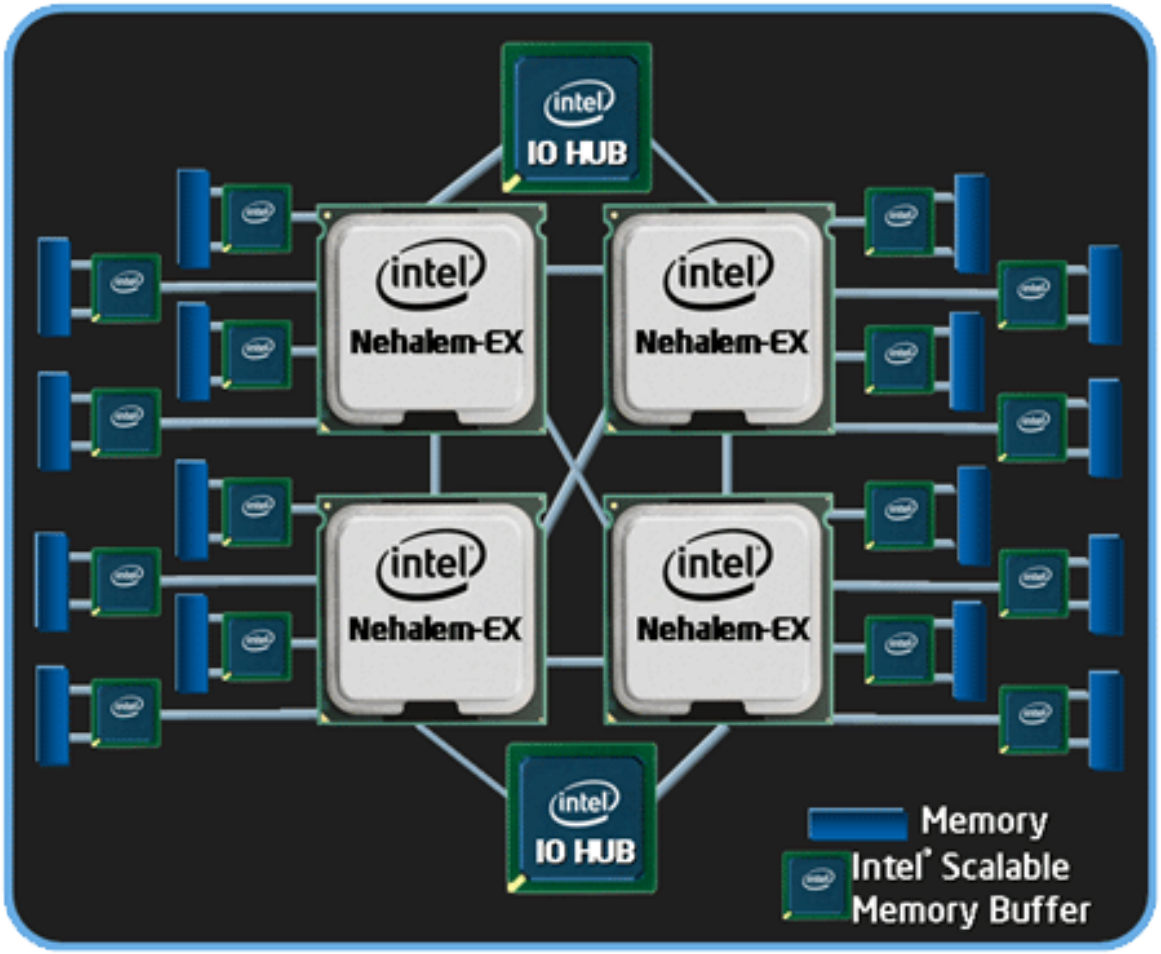


System Interconnect: Router and QPI

- Nehalem-EX supports 4 QPI link interfaces
  - High speed (6.4GT/s), packetized, P2P bidirectional link
  - Enables optimal, scalable and flexible system topologies
  - 25.6 GB/s of peak bandwidth per link
  - http://www.intel.com/technology/quickpath/introduction.pdf
- Router glues external and internal interfaces
  - High BW, full crossbar connection with simultaneous 8 way routing
  - Route thru to support sparse topologies
  - Fully programmable routing capability for Efficient, flexible and broadcast messaging
  - Two deadlock free virtual channels to support complex deadlock free topologies



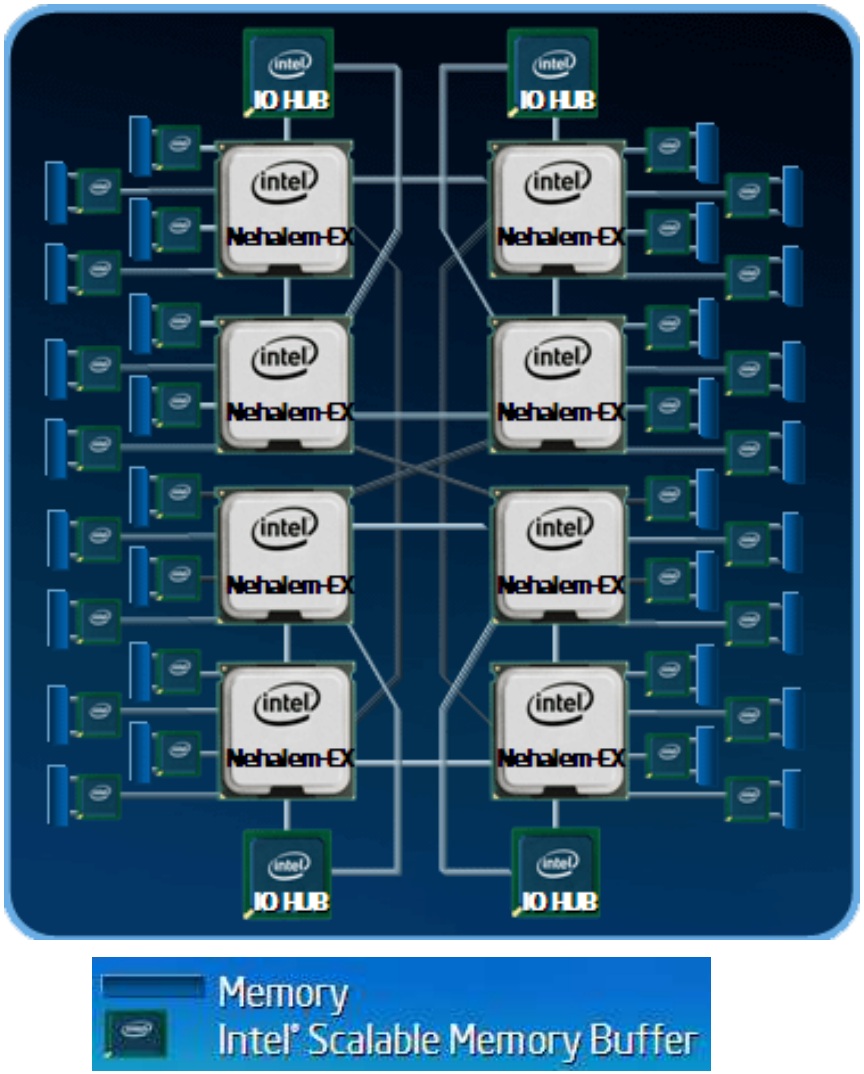
Nehalem-EX 4S Platform



Fully connected 4S system

– 1 link for the IO hub, 3 links for inter-socket

Platform Configurations



- Intel Architecture capable of QPI connected 8-Sockets / 128 threads
- Scalable systems and >8-socket capability with OEM node controllers
- Twisted Hypercube Interconnection

Max 2 QPI hops between two sockets

Nehalem-EX Performance

Xeon® 5500 vs. Xeon® 5400		*Nehalem-EX vs. Xeon® 7400*	
Up to 3.5x	Memory Bandwidth	Up to 9x	Memory Bandwidth[2]
Up to 2.5x	Database Performance	> 2.5x	Database Performance[1]
Up to 1.7x	Integer Throughput	> 1.7x	Integer Throughput
Up to 2.2x	Floating Point Throughput	> 2.2x	Floating Point Throughput

Expecting larger gains from Nehalem Architecture in MP

[1]Based on May'09 internal measurement using OLTP workload.

[2]Based on May'09 internal measurement using Intel internal workload

Nehalem-EX Summary

- Next generation of enterprise server CPU
- Focused on per socket throughput/bandwidth
  - 8 multi-threaded cores
  - Large LLC with high BW ring interconnect
  - Dual Caching agent, home agent/memory controllers
  - 4 high speed SMI memory links per socket
  - Support for large memory capacity
- Designed for highly scalable systems
  - Support high socket count as well as flexible MP configurations
  - 4 QPI links per socket
  - Large amount of on die buffering/queuing

Acknowledgements
“Beckton Team”