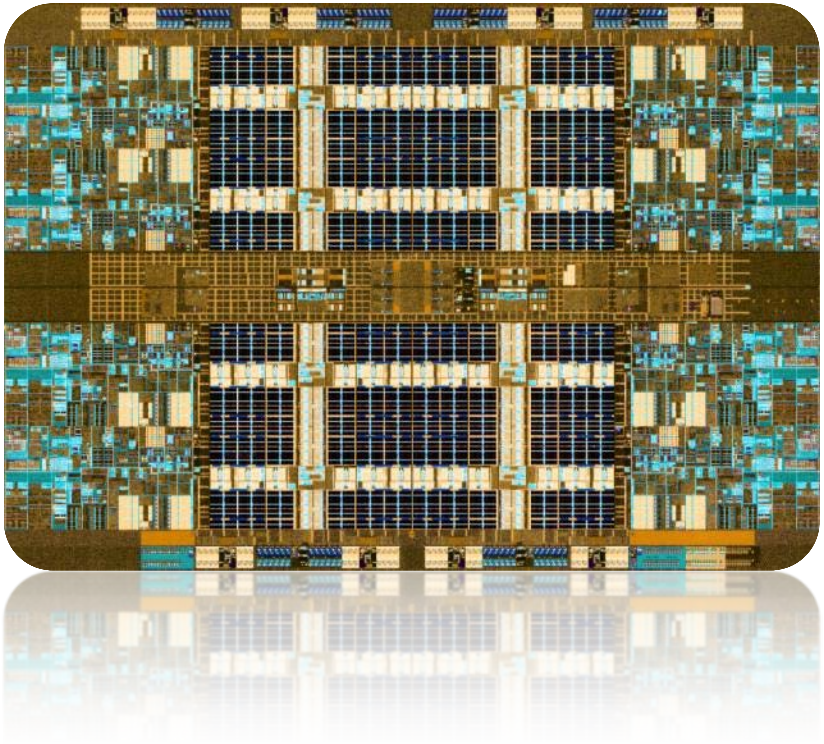


Intel® QuickPath Interconnect

Architectural Features Supporting Scalable System Architectures

Dimitrios Ziakas, Allen Baum,
Robert A. Maddox, Robert J. Safranek

Legal Disclaimers

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.

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, Go to: http://www.intel.com/performance/resources/benchmark_limitations.htm.

Intel does not control or audit the design or implementation of third party benchmarks or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmarks are reported and confirm whether the referenced benchmarks are accurate and reflect performance of systems available for purchase.

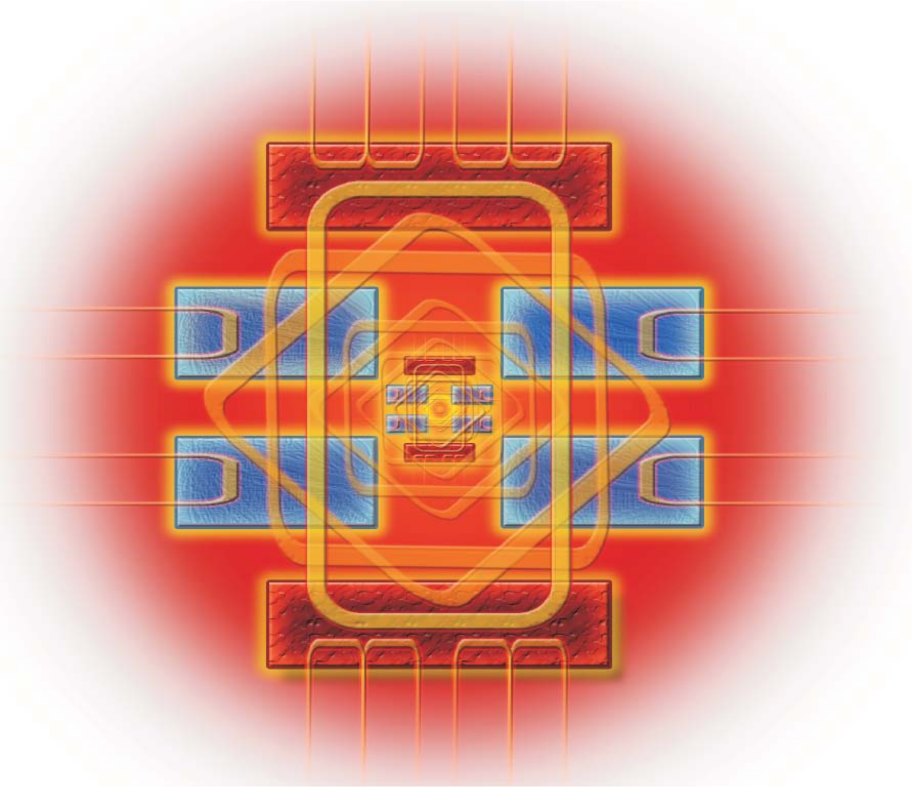
Relative performance is calculated by assigning a baseline value of 1.0 to one benchmark result, and then dividing the actual benchmark result for the baseline platform into each of the specific benchmark results of each of the other platforms, and assigning them a relative performance number that correlates with the performance improvements reported.

SPEC, SPECint, SPECfp, SPECrate. SPECpower, SPECjAppServer, SPECjEnterprise, SPECjbb, SPECompM, SPECompL, and SPEC MPI are trademarks of the Standard Performance Evaluation Corporation. See http://www.spec.org for more information.
TPC-C, TPC-H, TPC-E are trademarks of the Transaction Processing Council. See http://www.tpc.org for more information.
SAP and SAP NetWeaver are the registered trademarks of SAP AG in Germany and in several other countries. See http://www.sap.com/benchmark for more information.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor series, not across different processor sequences. See http://www.intel.com/products/processor_number for details. Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications. All dates and products specified are for planning purposes only and are subject to change without notice

* Other names and brands may be claimed as the property of others.

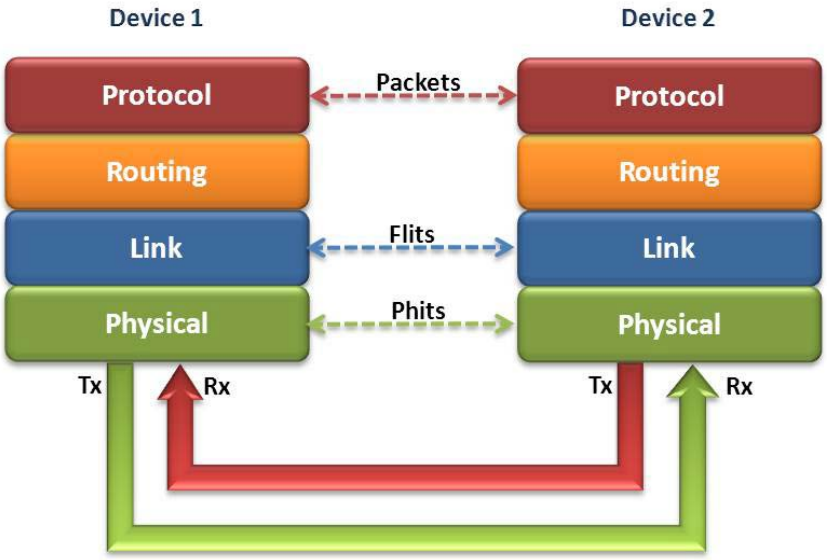
Agenda



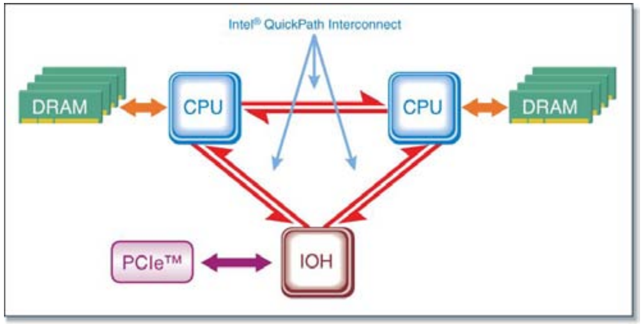
- Overview
- Scalability
- Error Handling
- Future Extensions

Intel® QuickPath Interconnect

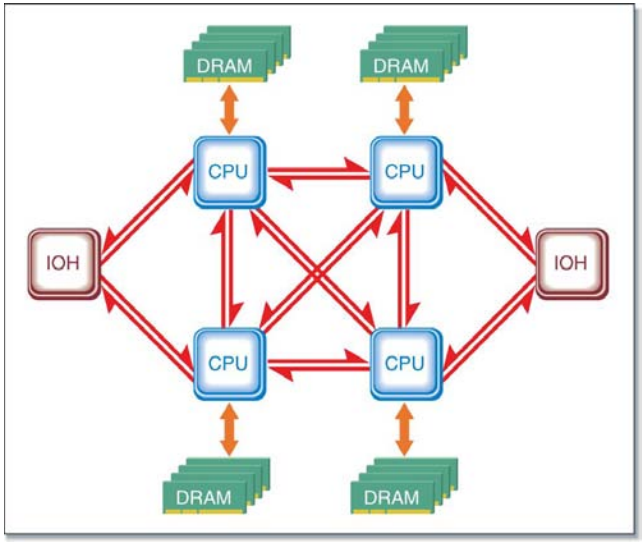
- High speed, packetized, point to point, coherent system interconnect
  - Handles two socket servers and up
  - Architecture built for scaling efficiently
  - RAS features to support large systems
- Layered Architecture
  - Modularity & Flexibility
  - Rich Link layer features
    - MCs and Virtual Networks
  - Coherent and Non-Coherent Protocols
  - Applied to both Intel® Xeon® and Itanium® processor based systems



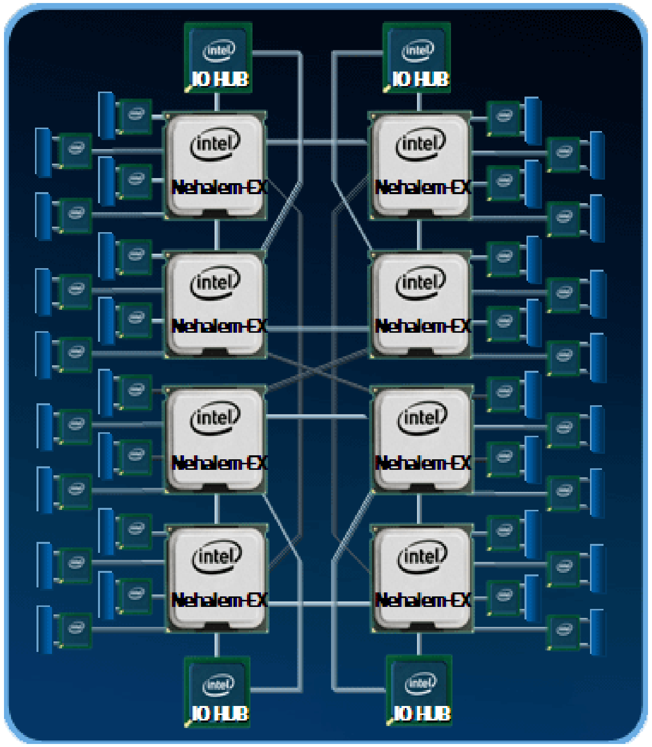
Typical Platform Topologies



2 socket server



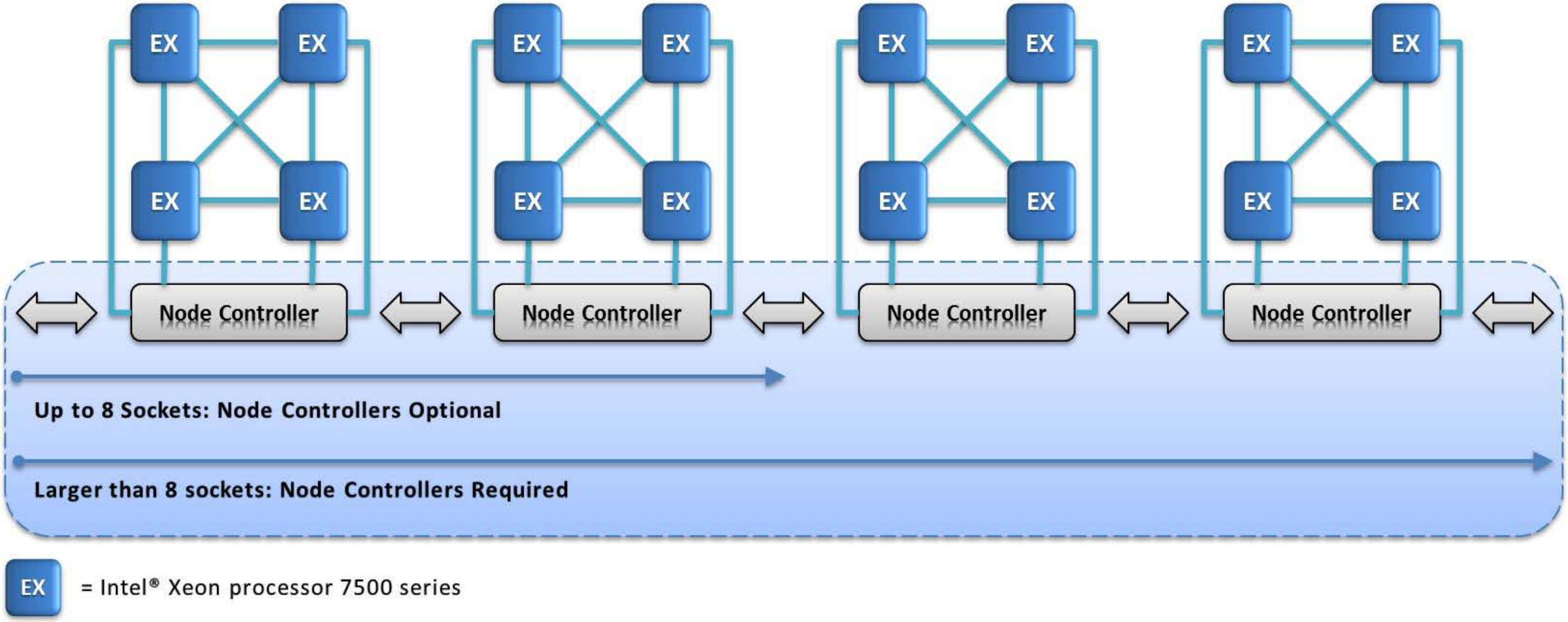
4 socket server



8 socket server

- •Multiple distributed memory controllers
- •Platform wide coherency protocol
- •Routing functions for message delivery
- •Flexible configurations

Very Large Systems



Up to 8 Sockets: Node Controllers Optional

Larger than 8 sockets: Node Controllers Required

EX = Intel® Xeon processor 7500 series

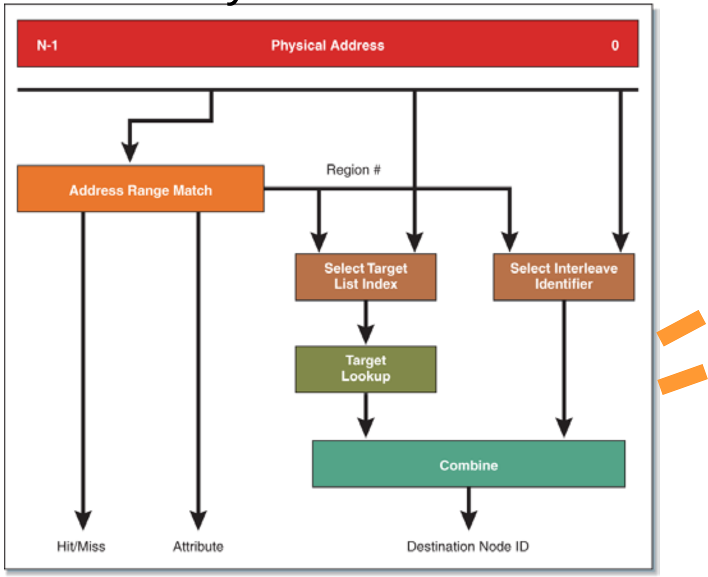
Intel® QPI Features for Scalable Systems

- MESIF Coherency Protocol
  - F State => Clean line forwarding
  - Single round-trip delay for cached data
- SADs, TADs, and RTs
  - SADs determine where to send request
    - Socket interleaving
  - TADs determine who gets it
    - Memory Channel interleaving
  - RTs determine how it gets there
    - Local socket, or out another link
  - Configuration done through firmware at boot
  - Can be tailored to NUMA or UMA configurations

SADs and TADs - Interleaving

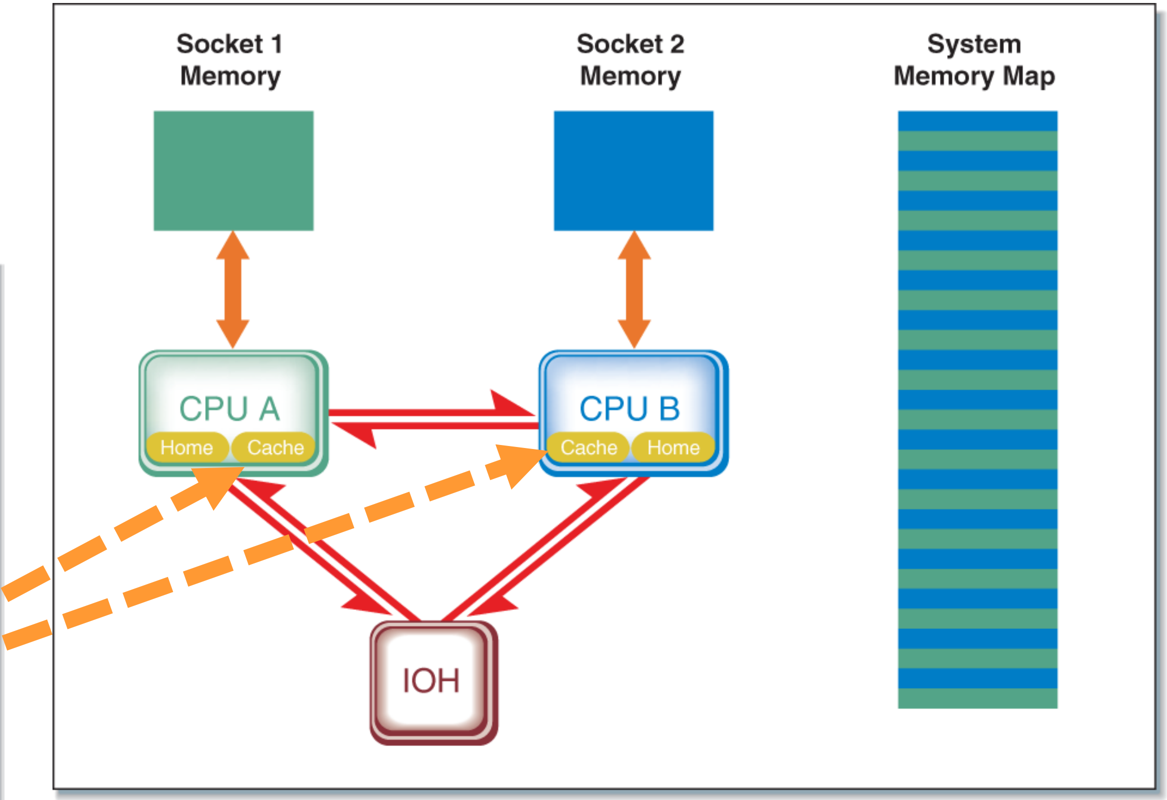
SAD Conceptual Diagram

Physical Address



Attributes   Destination NodeID

Socket Interleaving Example



SADs and TADs - Interleaving

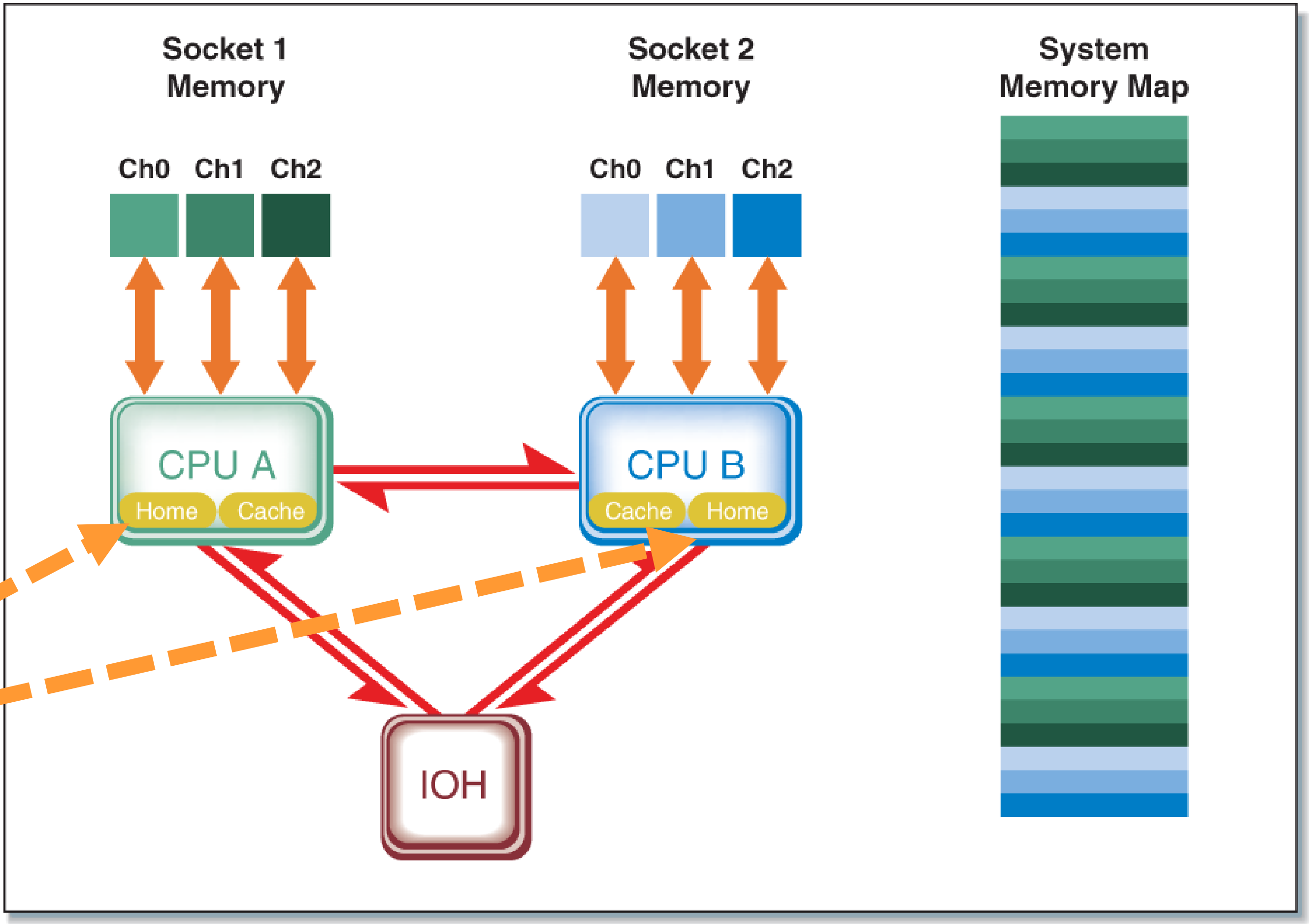
Destination NodeID

Attributes

TADs

Targeted Function

Channel Interleaving Example

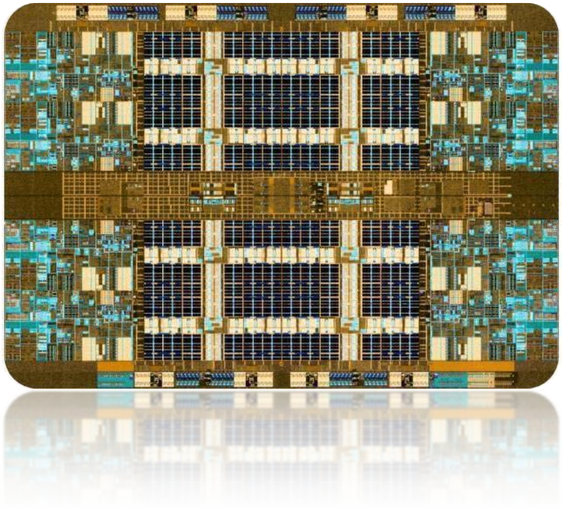


Efficient Bandwidth Utilization

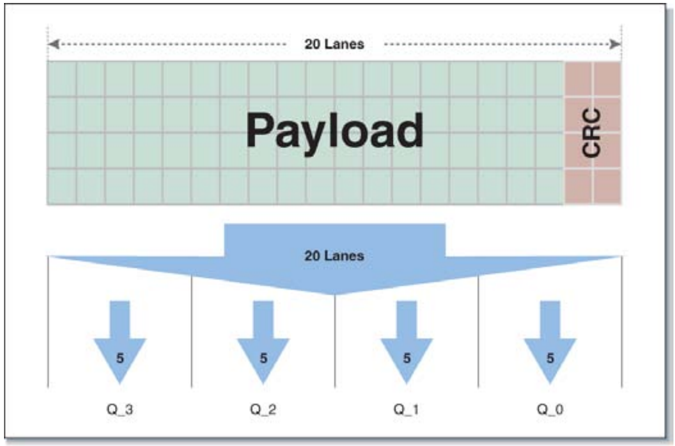
- Variable length messages
  - All messages built from one or more ‘Flits’
  - Most common messages are encoded into shortest format
- Source snoopy protocol
  - Good for transaction processing, 2-8 sockets
- Minimizing the number of messages
  - Source snoop and/or directory controlled caching agents
  - No need to send snoops to the home agent getting the request
  - Multiple caching and home agents in a single socket
  - Router broadcast of snoops

Reliability in Scalable Systems

- Routing layer allows system partitioning to smaller clusters for application isolation & reliability
- System reconfiguration without bringing the system down
- Intel® QPI also allows memory writes and reads to be mirrored



Error Handling

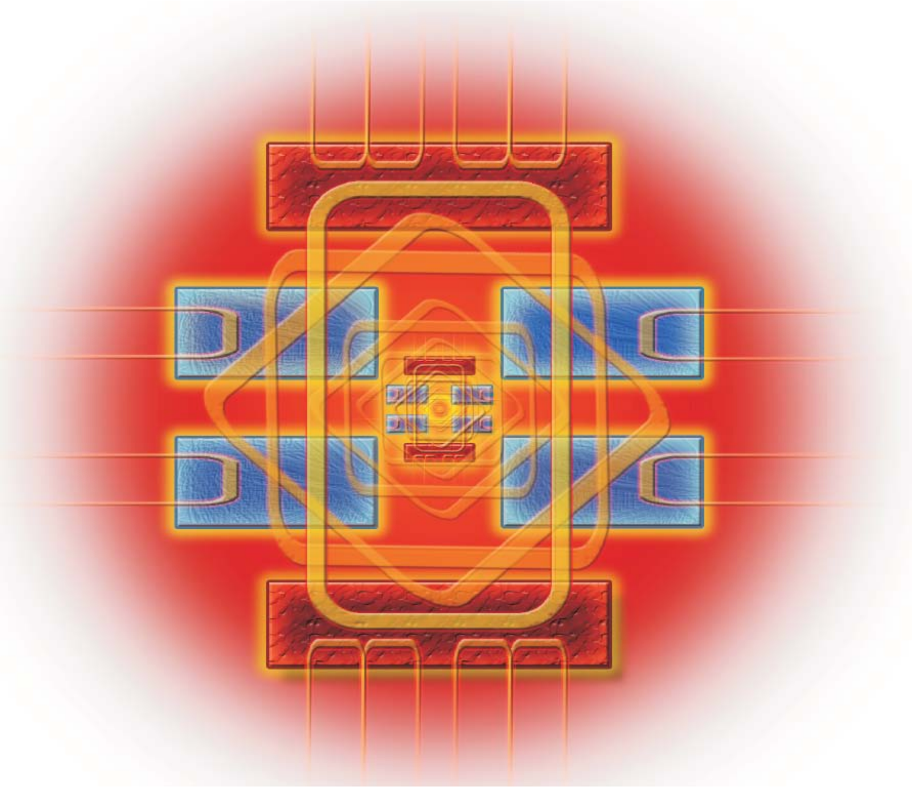


- Detection
  - All messages built from 80 bit Flits
  - 8 bits of CRC in every Flit
  - Optional 16 bit rolling CRC
- Recovery
  - Link layer retry process
  - Link in-band reset on too many failed retries
- Reduce single points of failure
  - Link reconfiguration on data lane failure
  - Link reconfiguration can also substitute for a failed clock using one of the data lanes
- Poison and Viral Responses

Future Extensions

- Increase speed on the link
  - Similar to FSB history, growth in transfer rates over time
- Improve Power Management
  - Integrated into the overall CPU and platform power management policies
- Improve messaging efficiency
  - In terms of link utilization or latency
  - Reduced implementation and validation complexity

Summary



- Looked at basic Intel® QPI functions and topologies
- Described some scalability features
- Reviewed Intel® QPI RAS Features
- Touched on the future

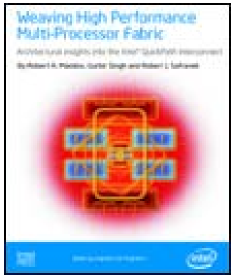
THANK YOU!

Q&A

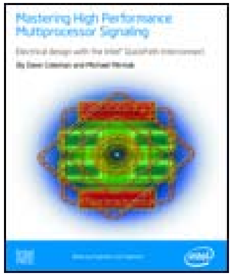
Additional sources of information on Intel® QuickPath Interconenct:

- Web info:
  - http://www.intel.com/technology/quickpath/index.htm
  - http://www.intel.com/technology/quickpath/whitepaper.pdf

- Books: Published by **Intel Press**



Weaving High Performance Multiprocessor Fabric
Architectural Insights into the Intel® QuickPath Interconnect
By Robert A. Maddox, Gurbir Singh and Robert J. Safranek
http://www.intel.com/intelpress/sum_qpi.htm



Mastering High Performance Multiprocessor Signaling
Electrical Design with the Intel ® QuickPath Interconnect
By Dave Coleman, Michael Mirmak
http://www.intel.com/intelpress/sum_qpied.htm