NoC’s at the Center of Chip Architecture:
Urgent Needs Today and for the Future

Andrew A. Chien
VP Research
Director, Future Technologies Research
Corporate Technology Group

May 12, 2009
Agenda

- Brief history of Parallel Interconnects
- On-chip Interconnection Networks
- SoC Interconnects
- What NoC’s should become
- A few Interesting Parallel Machines
Early Interconnection Networks

**ILLIAC IV 1964-1972**
- 256 processors, 1GFlop
- 2-D mesh (16x16)
- 800Mbps links, 13Gbps Bisection BW

**Cray 1 1975**
- 80Mhz, 250 MFLOPS
- 4GB/s Memory BW

**Caltech cosmic cube 1983**
- 64 Intel 8086/87 procs, 3 Mflops (32bit)
- 6-d hypercube
- 2Mbps bit serial links
- 32Mbps Bisection BW

**Intel iPSC/860 1990**
- 128 i860 processors, 80 Mflops
- 7-D hypercube, 20Mbps p2p, 1.2Gbps Bisection BW

wormhole routing

ed-thelen.org, wikipedia.org, ECE 669: Parallel Computer Architecture University of Massachusetts, 2004
http://www2.udec.cl/~marlagos/supercomputadoras.htm

"NoCs at the Center..."
Modern Interconnection Networks

Cray Jaguar XT5 2008
- 66,427 quad-core processors
- 1 Cray SeaStar2+ routing and comms ASIC/node, 57.6 GB/s switching
- ~1 PFLOPS
- 3D torus, 786 TB/s global BW
- Bisection BW 11.5 TB/s
- 6,329 miles of interconnect cables

IBM Roadrunner 2008
- 12,240 procs (control/network) & 12,240 Cell sockets (compute)
- ~1 PFLOPS
- 2:1 Fat Tree, Infiniband
- 55 miles of all-optical interconnect
- Bisection bandwidth = 3.5 TBps
Technology Trends...

(Dally, NoCs 2007 Keynote)
Summary of Off-Chip Networks*

• Topology
  – Fit to packaging and signaling technology
  – High-radix - Clos or FlatBfly gives lowest cost

• Routing
  – Global adaptive routing balances load w/o destroying locality

• Flow control
  – Virtual channels/virtual cut-through

*oversimplified

(Dally, NoCs 2007 Keynote)
Agenda

- Brief history of Parallel Interconnects
- On-chip Interconnection Networks
- SoC Interconnects
- What NoC’s should become
- A few Interesting Parallel Machines
Are On-chip networks different? How?

- We are applying the tools we’ve developed
  - Design space: topology, switching, routing, flow control, virtualization
  - Cost metrics: wire bisection, switch size, total power, total devices
  - Performance Metrics: throughput, latency, traffic structure independence/robustness, hot spot performance, quality of service
- Important differences -- Integration, costs, technology, workloads, energy... tradeoffs reopened
Ambric

- 45 Brics (2 compute units – streaming RISC, SRD, 2 RAM);
  - => Peak 1.2 Teraops/chip
- Mesh of Brics are nearest neighbor connected
  - Four channels each way up to 9.6 Gbps
  - Bisection bandwidth = 792 Gbps (~100GB/s)
- Programmed with “Structured Object Programming Model”... streaming objects mapped to fabric
Tilera

- 64 cores/tiles; 192 Gigaops
- On chip interconnect bandwidth 32 Tbps
- Five 2D Mesh networks, Bisection bandwidth: 2 Tbps total
- Packet switched, wormhole routed, p2p
Intel Polaris NoC

- **8x10 mesh**
  - Bisection BW = 320GB/s
  - 4-byte bidirectional links
  - 6 port non-blocking crossbar
  - Crossbar/Switch double-pumped to reduce area

- **Router Architecture**
  - Source routed
  - Wormhole switching
  - 2 virtual lanes
  - On/off flow control

- **Activity-based Power Management**
  - 7:1 ratio active to idle, per port
  - 980mW interconnect power/node
  - 80% in routers – xbar, buffers

An 80-Tile 1.28TFLOPS Network-on-Chip in 65nm CMOS, S. Vangal et al.
**Intel Larrabee**

- **Processing:** Scalar + Vector + Special Fcn units
- **Topology:** Bi-directional ring networks, 512 bits each way
  - Large port-count Interconnect
  - Extremely simple, low latency
  - Many CPU cores – 16, 32, 48...+ Cache, Memories, special units

Spectrum of complexity, energy, topology, speed, are broad. It’s still early and our understanding is evolving rapidly.
On-Chip Networks (NoC’s) Challenges

- Re-exploration of tradeoffs and needs [topology, routing, flow control]
  - Costs and design constraints
  - Needs and opportunities
- Flexible and systematic support of irregularity
  - Node types and communication needs
  - Traffic structures (static and dynamic)
  - Are there optimal topologies and designs? Or is that the wrong formulation?
- Energy efficiency and beyond
  - Network, Cores, memories, and implications
Agenda

- Brief history of Parallel Interconnects
- On-chip Interconnection Networks
- SoC Interconnects
- What NoC’s should become
- A few Interesting Parallel Machines
SoC Interconnects – an Evolution of Needs

- Traditional SoC Design
  - Processor Core and I/O as key drivers, low-level integration with NxN challenges
- Modern SoC Design (~50 blocks)
  - Standard format interfaces (OCP, VCI, AHB, APB, etc.) enable model-based NoC generation
  - Intelligent agents to interface IP blocks and standard format interfaces
  - Integrate, performance tune design to meet needs (high speed blocks and flows)
- Future SoC Design (100’s of blocks)
  - Numerous high speed blocks and flows
  - Complex and dynamic multitasking and demand-driven workloads
  - Growing system-level management challenges

TI OMAP processor, http://www.ti.com/
Intel Future SoC’s, ISSCC 2009.

“NoCs at the Center…”
Modern SoC Design: NoC Interconnect Generation

- Standard interfacing enables NoC generation
- NoC’s glue the design together @ comm level
- Enable ecosystem of reuse; rapid SoC SW/HW codesign
Flexibility & Optimization in Generating NoC’s

- What might you know?
  - Block types and placement
  - Speed and protocol needs
  - Likely communication partners

- What constraints?
  - Topology, wires, progress/deadlock avoidance, etc.

- Many examples of beating hand-integrated, optimized versions (success!)
  - Why?
  - Pooling, systematic optimization, small/simpler application needs
Future SoC’s will require deep understanding and customizability...

• How does understanding/specification scale?
  – System design problem, multitasking and irregular workloads, application/software behavior complexity increasingly variable
  – Workload ambiguity: Concurrent activities, interactions, content-based decisions

• Increasing system pressure (flexibility and x-system coordination)
  – Workload dynamics unknown and varied (what multi-tasking apps, external stimuli – with what timing)
  – Overall energy and heat, response time, real-time elements of communication, display, sensing

Decreasing ability to optimize? Different notion of what it means to exploit known behavioral properties.
Agenda

- Brief history of Parallel Interconnects
- On-chip Interconnection Networks
- SoC Interconnects
- What NoC’s should become
- A few Interesting Parallel Machines
NoC’s should become an Active Partner

- NoC’s are the distributed presence, the “uncore”, the “Si backplane”
- Become Energy-efficient – “Energy proportional”
- Enable efficient parallelism
- Enable energy efficient parallelism

- Shift: From pipes to intelligent, distributed management
  - ... and a complex partnership with software
Energy-Proportional Communication in NoC’s

- Energy-efficiency is a major issue
  - Polaris Interconnect Power ~35% of Chip [Kundu, OCIN 06]
  - Exascale HW Study, On-Chip Interconnect Power not budgeted [Kogge 2008]

- Challenge: NoC’s must achieve energy-proportional communication
  - ~0 power baseline
  - low energy/bit for wire/channel transit & minimum transits
  - Efficient routers – simple, switched only when necessary
  - Adapt to the changes imposed by power-management.
NoC’s should support Efficient Regular Parallelism

- Synchronization and coordination: Barriers, Fence, Fetch-n-Op
- Consistency: Ordering, Tagging, QoS, Priority, Coherence
- How to do this flexibly and efficiently?
  - partitioning, overlap, virtualization
  - notification, migration, energy efficiency
  - …
- + Efficient, high bandwidth, low-latency comms
NoC’s should support Efficient Irregular Parallelism (Advanced)

- Multicore systems => general-purpose parallelism; Challenge is to support irregular applications!
  - Irregular Memory Access, Communication, Work Balance
- Architectural support in Cores, what help does the NoC provide?
  - Hardware task queuing. Distributed load information and work stealing?
  - Load gradient-based task routing? Work stealing?
  - vs. OCP notion of a “computation-less” NoC...
- What are the software-core-NoC partnerships?
NoC’s should support Energy Efficient Parallelism (Radical)

- Traditional models of energy management (software, hw centralized) are too inflexible
  - Intelligent distributed energy management in HW? Blocks, NoC
    - How to do the most energy-efficient routing / network management?
  - How to support energy efficient software execution?
    - Information, adapting HW to workload state
    - Identify opportunities to schedule, migrate for EE
    - Find threads with data locality
    - Energy-based energy-based routing/swapping/migration
    - Energy limiting / voltage scaling, energy-based scheduling, balancing, work-stealing (turn it around)

- Other forms of chip-level management? Resilience?
Why the NoC Community?

- A core set of issues about how to design and manage the NoC?
  - NoC = unCore.... Isn’t this “unNoC”?
- SoC designers => the system architects
  - tools for NoC generation and overall DA are the key leverage points
  - Floorplan, interconnection, XX management all happen at this level.
- NoC-style methodology needed -- fundamental generators and distributed “uncore” presence to do all of these things.
- You’re where the leverage is for solving these new problems and challenges!
Agenda

• Brief history of Parallel Interconnects
• On-chip Interconnection Networks
• SoC Interconnects
• What NoC’s should become
• A few Interesting Parallel Machines
Interesting Parallel Machines

**Thinking Machines CM-2 1987**
- 12-dimensional hypercube, bit serial
- $2^{16}$ Bit serial ALU’s
- Integrated comp & comm (parallel sorting, prefix, etc.)
- 256kb/processor, 7Mhz clock
- SIMD, globally synchronous

**Thinking Machines CM-5 1991**
- 65.5 GFLOPS, 16K SPARC + Weitek, 160GB/s Bisection BW
- Fat tree network: data, Tree network: control (broadcast, reduction, synchronization). Global and segmented reduction/parallel-prefix/barrier
- Protection and context-switching

**MIT J-machine 1991**
- 1024 nodes
- 3-dimensional deterministic wormhole-routed mesh
- Bisection bandwidth = 14.4 GBps

http://www.cs.cmu.edu/~scandal/alg/
http://cva.stanford.edu/projects/j-machine/
Interesting Parallel Machines - today

**IBM Blue Gene/L-P:** 65,536 nodes, 478 TFLOPS

**Data Network: 3 D Torus**
- Interconnects all compute nodes; Virtual cut-through
- 1.7/3.9 TB/s bisection bandwidth, 188TB/s total bandwidth

**Collective Network**
- Interconnects all compute and I/O nodes (1152)
- One-to-all broadcast functionality, Reduction operations
- 6.8 Gb/s of bandwidth per link, ~62TB/s total binary tree bandwidth

**Low Latency Global Barrier and Interrupt**
- Latency of broadcast to all 72K nodes 0.65 μs, MPI 1.6 μs
Summary

• Building on foundation and tools of parallel interconnect research.
  – Reexamine tradeoffs for NoCs and *invent* new solutions
  – Leap to irregularity [many dimensions] and optimization
  – Energy efficient and energy proportional
  – Parallelism and work management
  – Energy efficiency, work, and system management

Exciting opportunities! The path forward is littered with major challenges for the NoC community.
Q&A

For more information:
http://www.intel.com/research/
Backup
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.
- 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 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 © 2009 Intel Corporation.
Risk Factors

The above statements and any others in this document that refer to plans and expectations for the first 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. Current 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. 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; capacity utilization; excess or obsolete inventory; product mix and pricing; variations in inventory valuation, including variations related to the timing of qualifying products for sale; 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, 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 recent financial crisis 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 extreme 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. 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.

Rev. 1/15/09