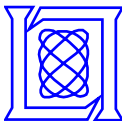


3-D Graph Processor

William S. Song, Jeremy Kepner, Huy T. Nguyen, Joshua I. Kramer, Vitaliy Gleyzer, James R. Mann, Albert H. Horst, Larry L. Retherford, Robert A. Bond, Nadya T. Bliss, Eric I. Robinson, Sanjeev Mohindra, Julie Mullen

HPEC 2010

16 September 2010

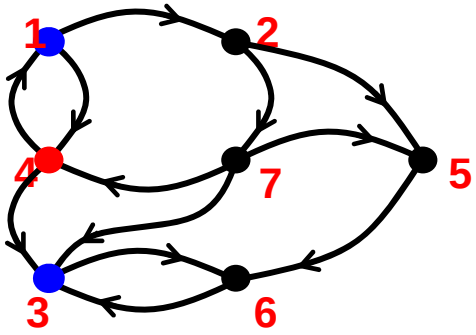
This work was sponsored by DARPA under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

Outline

- Introduction
  - Graph algorithm applications
    - Commercial
    - DoD/intelligence
  - Performance gap using conventional processors
- 3-D graph processor architecture
- Simulation and performance projection
- Summary

Graph Algorithms and Ubiquitous Commercial Applications

Graph Representation

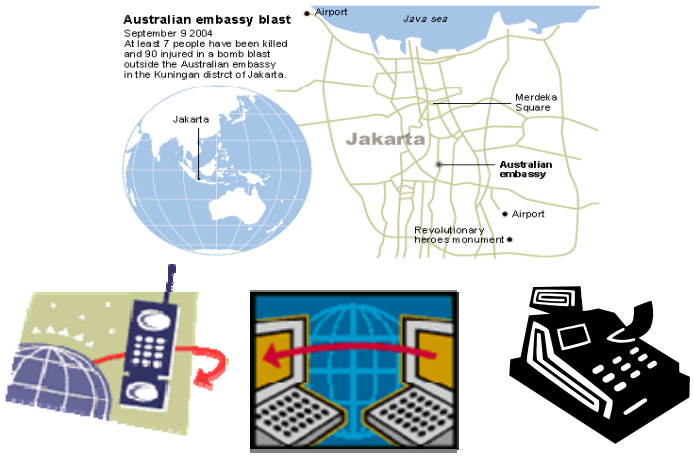


Applications

- Finding shortest or fastest paths on maps
- Communication, transportation, water supply, electricity, and traffic network optimization
- Optimizing paths for mail/package delivery, garbage collection, snow plowing, and street cleaning
- Planning for hospital, firehouse, police station, warehouse, shop, market, office and other building placements
- Routing robots
- Analyzing DNA and studying molecules in chemistry and physics
- Corporate scheduling, transaction processing, and resource allocation
- Social network analysis

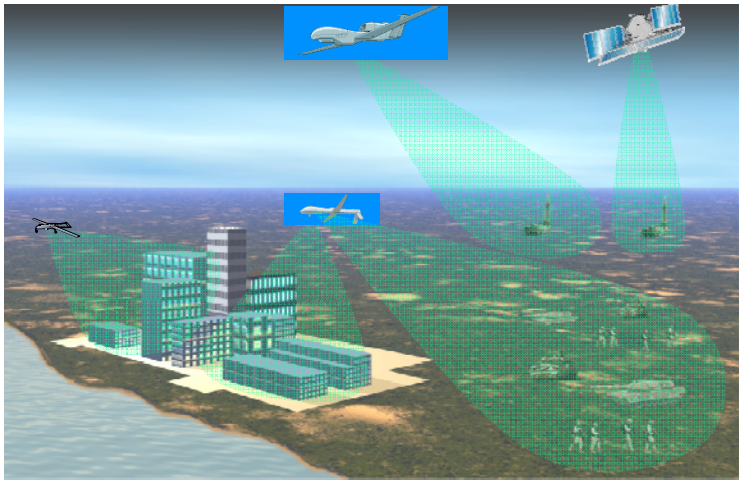
DoD Graph Algorithm Applications

Intelligence Information Analysis



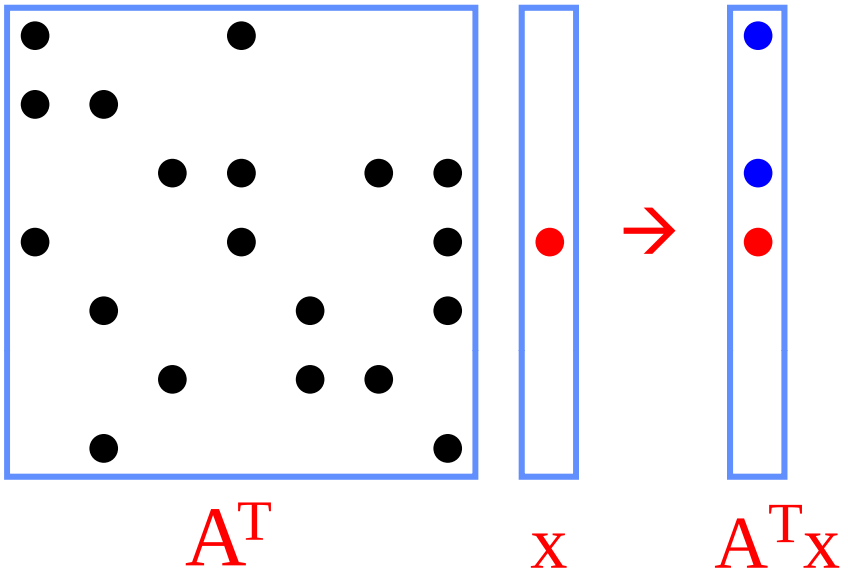
- Analysis of email, phone calls, financial transactions, travel, etc.
  - Very large data set analysis
  - Established applications

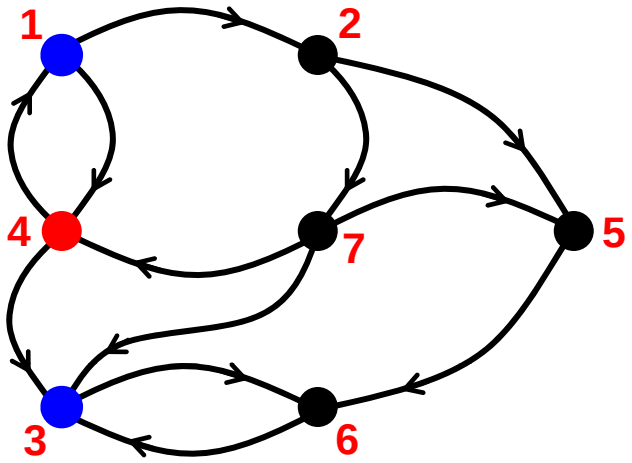
ISR Sensor Data Analysis



- Post-detection data analysis for knowledge extraction and decision support
  - Large data set
  - Real time operations
  - Small processor size, weight, power, and cost
  - New applications

Sparse Matrix Representation of Graph Algorithms

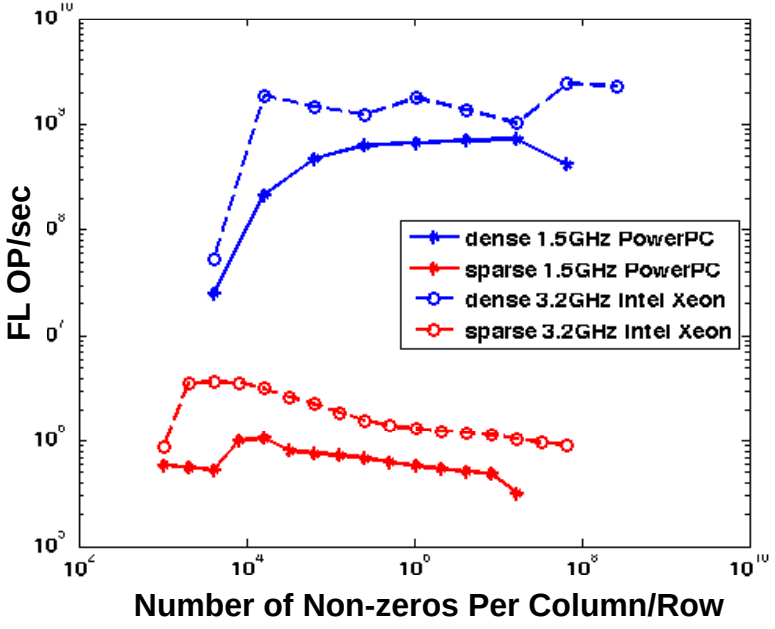




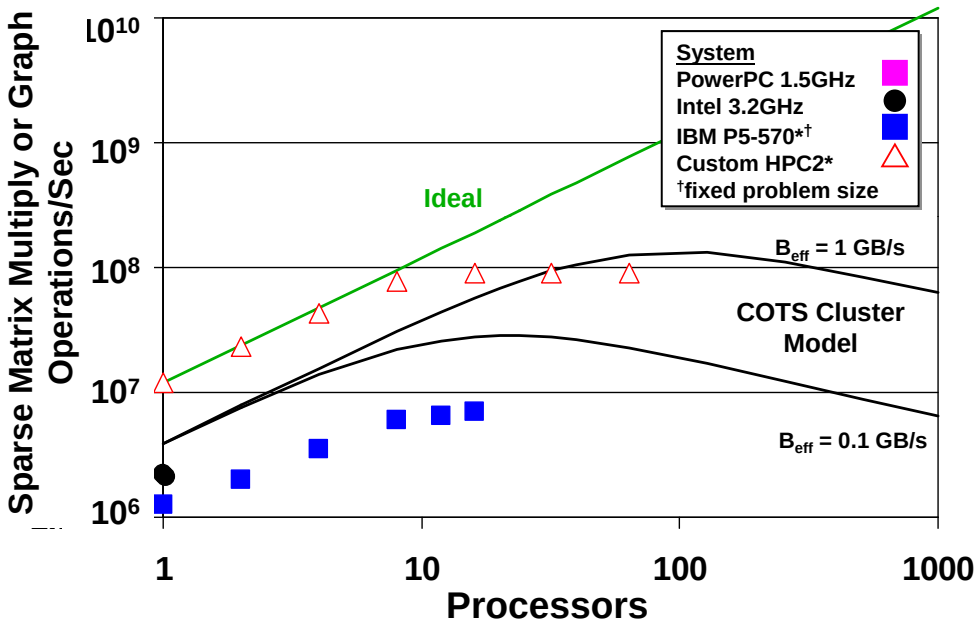
- Many graph algorithms may be represented and solved with sparse matrix algorithms
- Similar speed processing
  - Data flow is identical
- Makes good graph processing instruction set
  - Useful tool for visualizing parallel operations

Dense and Sparse Matrix Multiplication on Commercial Processors

Sparse/Dense Matrix Multiply Performance on One PowerPC



Sparse Matrix Multiply Performance on COTS Parallel Multiprocessors



Commercial microprocessors 1000 times inefficient in graph/sparse matrix operations in part due to poorly matched processing flow.

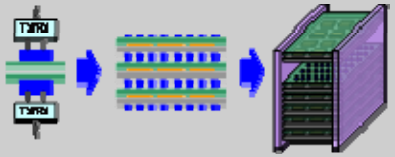
Communication bandwidth limitations and other inefficiencies limit the performance improvements in COTS parallel processors.

Outline

- Introduction
- 3-D graph processor architecture
  - Architecture
  - Enabling technologies
- Simulation and performance projection
- Summary

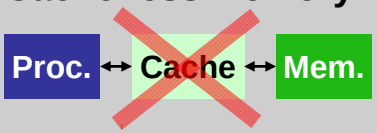
3-D Graph Processor Enabling Technology Developments

High Bandwidth 3-D Communication Network



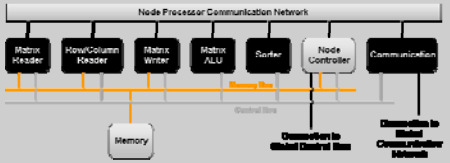
- 3D interconnect (3x)
- Randomized routing (6x)
- Parallel paths (8x)
- 144x combined bandwidth while maintaining low power

Cache-less Memory



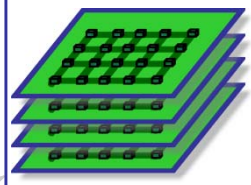
- Optimized for sparse matrix processing access patterns

Accelerator Based Architecture



- Dedicated VLSI computation modules
- Systolic sorting technology
- 20x-100x throughput

3-D GRAPH PROCESSOR

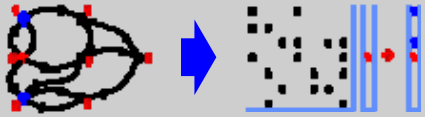


- **1024 Nodes**
- **75000 MSOPS***
- **10 MSOPS/Watt**

Data/Algorithm Dependent Multi-Processor Mapping

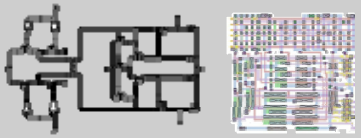
- Efficient load balancing and memory usage

Sparse Matrix Based Instruction Set



- Reduction in programming complexity via parallelizable array data structure

Custom Low Power Circuits



- Full custom design for critical circuitry (>5x power efficiency)

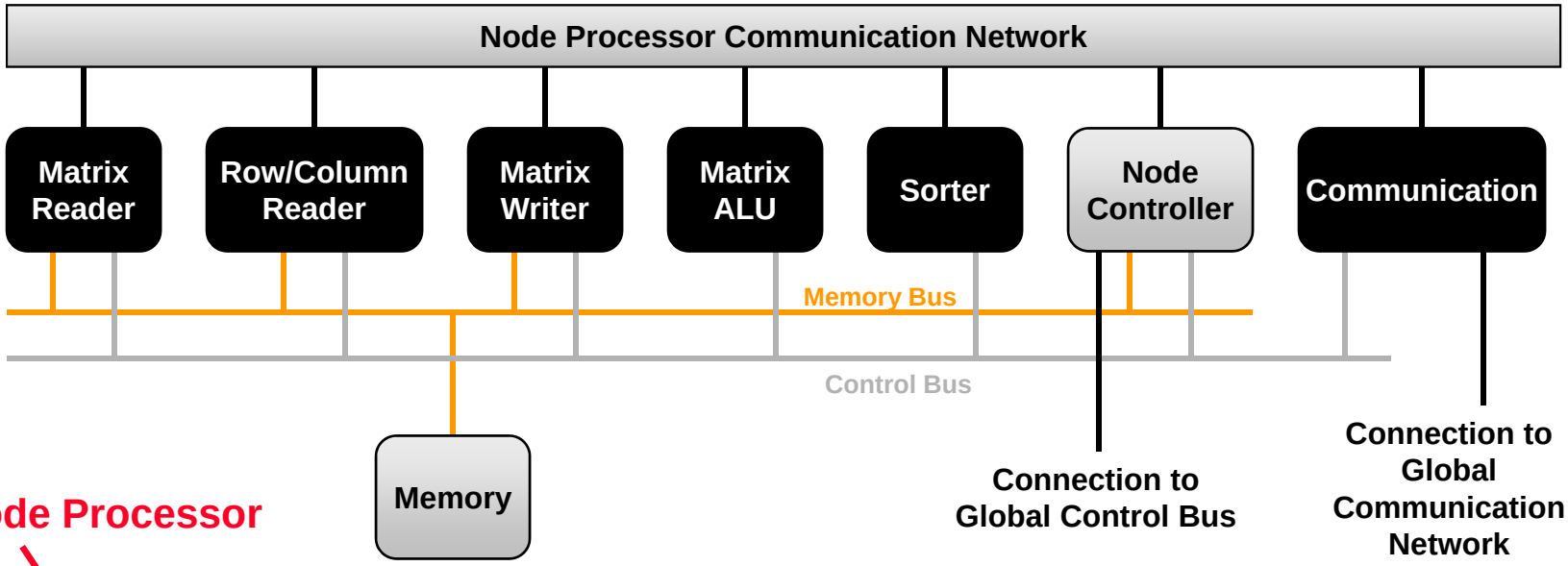
*Million Sparse Operations Per Second

Sparse Matrix Operations

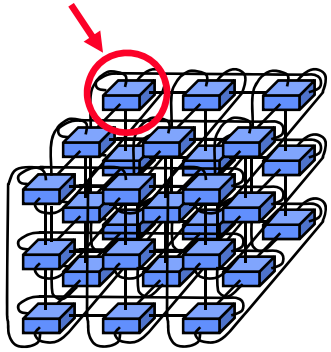
Operation	Distributions	Comments
C = A +.* B	Works with all supported matrix distributions	Matrix multiply operation is the throughput driver for many important benchmark graph algorithms. Processor architecture highly optimized for this operation.
C = A .± B C = A .* B C = A ./ B	A, B, and C has to have identical distribution	Dot operations performed within local memory.
B = op(k,A)	Works with all supported matrix distributions	Operation with matrix and constant. Can also be used to redistribute matrix and sum columns or rows.

- The +, -, *, and / operations can be replaced with any arithmetic or logical operators
  - e.g. max, min, AND, OR, XOR, …
- Instruction set can efficiently support most graph algorithms
  - Other peripheral operations performed by node controller

Node Processor Architecture

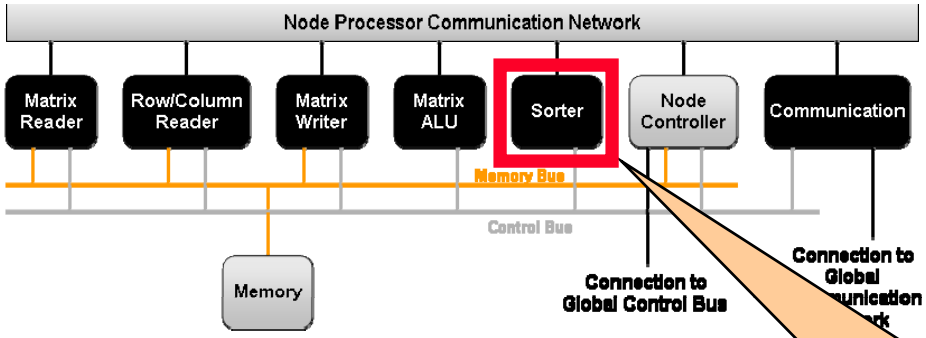


Node Processor

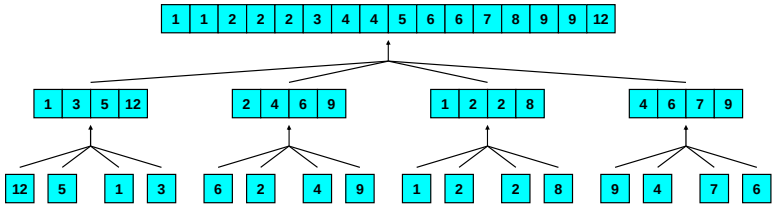


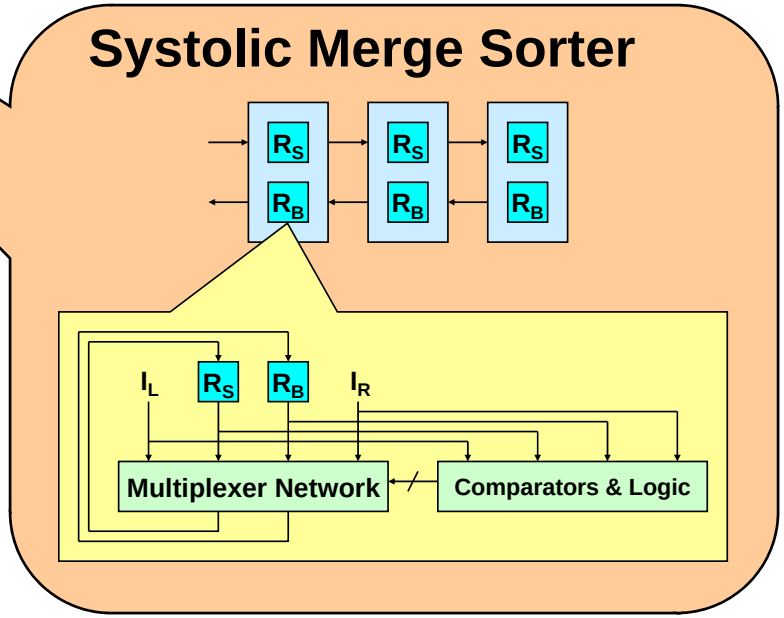
Accelerator based architecture for high throughput

High Performance Systolic k-way Merge Sorter



4-way Merge Sorting Example

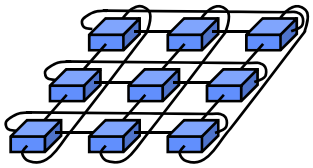


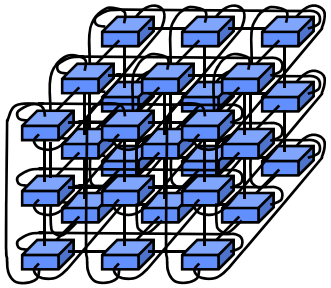


- Sorting consists of >90% of graph processing using sparse matrix algebra
  - For sorting indexes and identifying elements for accumulation
- Systolic k-way merge sorter can increase the sorter throughput >5x over conventional merge sorter
  - 20x-100x throughput over microprocessor based sorting

Communication Performance Comparison Between 2-D and 3-D Architectures

2-D





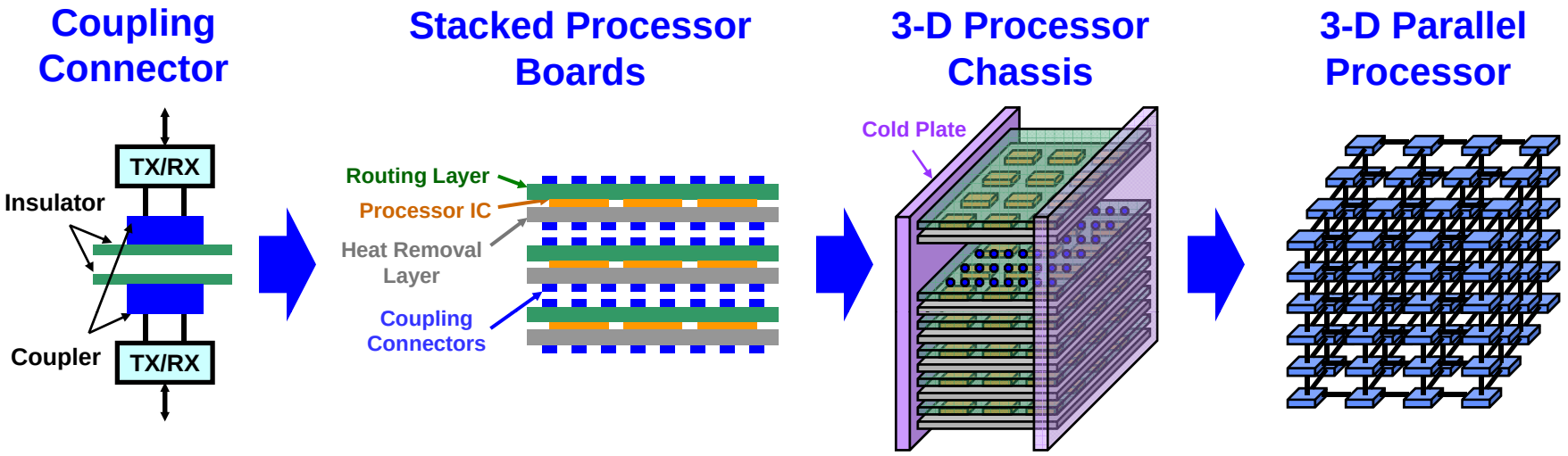
3-D

Normalized Bisection Bandwidth

P	2	10	100	1,000	10,000	100,000	1,000,000
2-D	1	3.2	10	32	100	320	1,000
3-D	1	4.6	22	100	460	2,200	10,000
3-D/2-D	1	1.4	2.2	3.1	4.6	6.9	10

- Significant bisection bandwidth advantage for 3-D architecture for large processor count

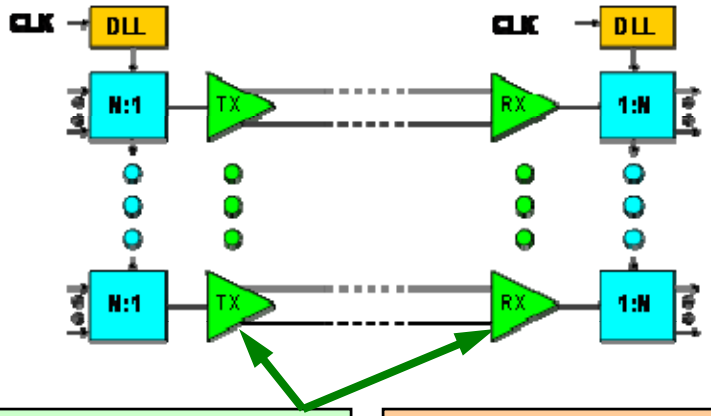
3-D Packaging



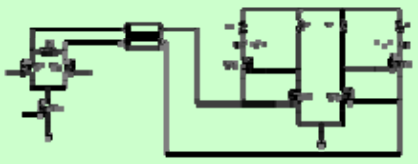
- **Electro-magnetic coupling communication in vertical dimension**
  - **Enables 3-D packaging and short routing distances**
- **8x8x16 planned for initial prototype**

2-D and 3-D High Speed Low Power Communication Link Technology

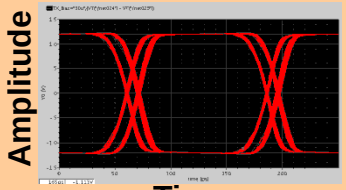
2-D Communication Link



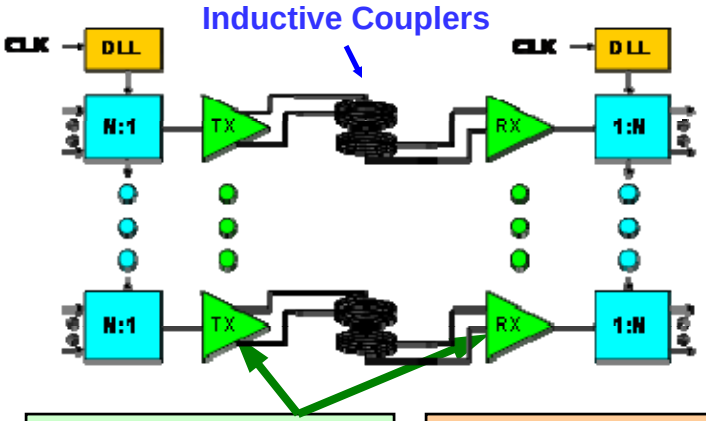
Current Mode TX/RX



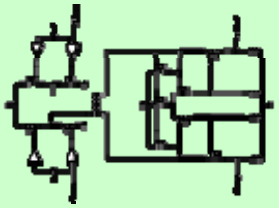
Signal Simulation



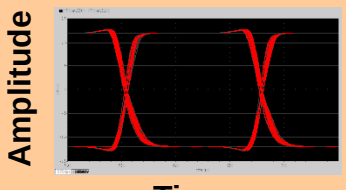
3-D Communication Link



Coupler TX/RX



Signal Simulation

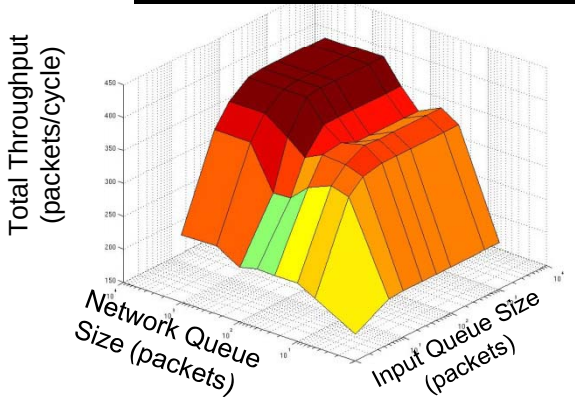


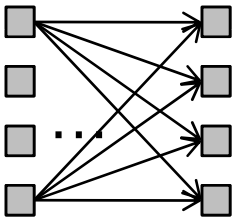
- High speed 2-D and 3-D communication achievable with low power
  - >1000 Gbps per processor node compared to COTS typical 1-20 Gbps
  - Only 1-2 Watts per node communication power consumption
  - 2-D and 3-D links have same bit rate and similar power consumption
    - Power consumption is dominated by frequency multiplication and phase locking for which 2-D and 3-D links use common circuitry
- Test chip under fabrication

Randomized-Destination 3-D Toroidal Grid Communication Network

Simulation Results

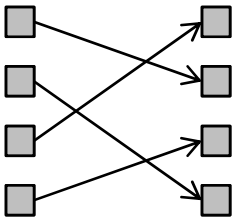
Input Queue Size vs. Network Queue Size vs. Total Throughput





Randomized Destination

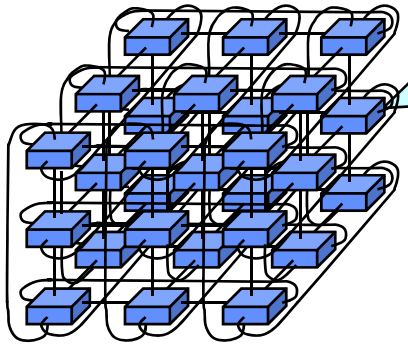
- Randomized destination packet sequence
- 87% full network efficiency achieved



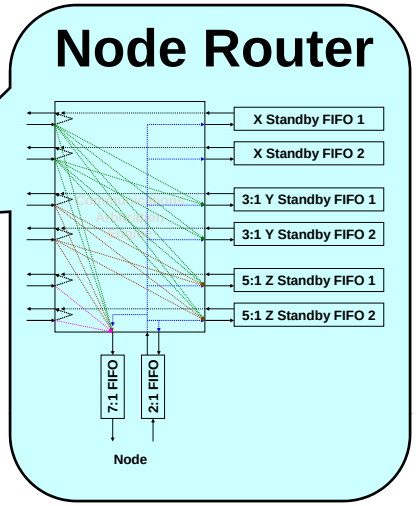
Unique Destination

- Unique destination for all packets from one source
- 15% full network efficiency achieved

3-D Toroidal Grid

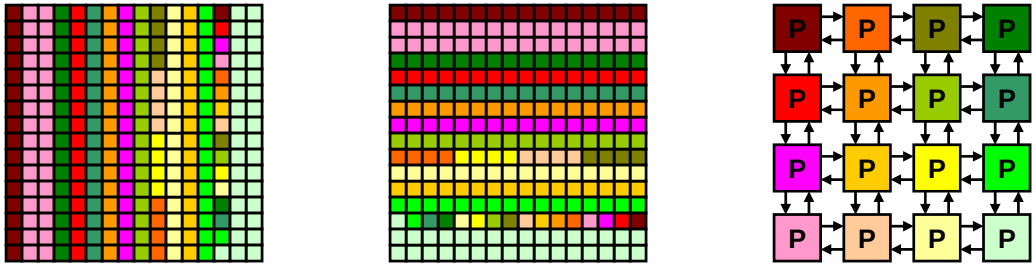


Node Router



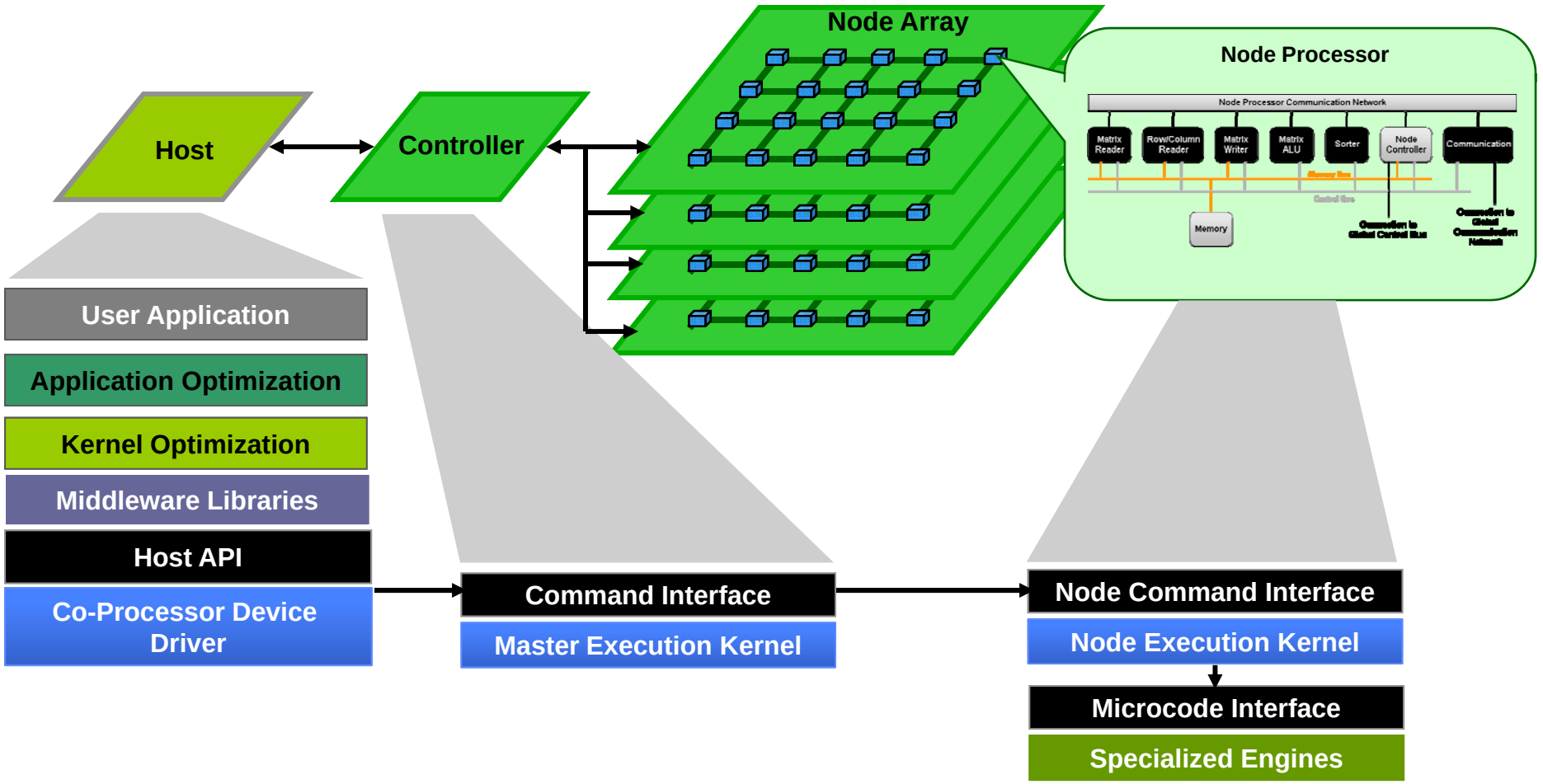
- **Randomizing destinations of packets from source nodes dramatically increases network efficiency**
  - **Reduces contention**
  - **Algorithms developed to break up and randomize any localities in communication patterns**
- **6x network efficiency achieved over typical COTS multiprocessor networks with same link bandwidths**

Hardware Supported Optimized Row/Column Segment Based Mapping



- Efficient distribution of matrix elements necessary
  - To balance processing load and memory usage
- Analytically optimized mapping algorithm developed
  - Provides very well balanced processing loads when matrices or matrix distributions are known in advance
  - Provides relatively robust load balancing when matrix distributions deviate from expected
  - Complex mapping scheme enabled by hardware support
- Can use 3-D Graph Processor itself to optimize mapping in real time
  - Fast computing of optimized mapping possible

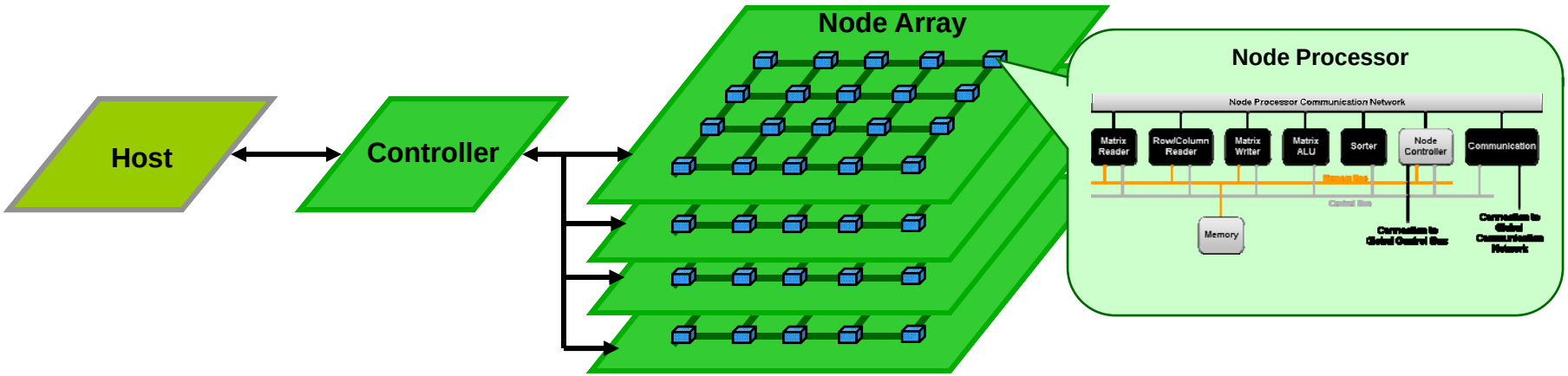
Programming, Control, and I/O Hardware and Software



Outline

- Introduction
- 3-D graph processor architecture
- Simulation and performance projection
  - Simulation
  - Performance projection
    - Computational throughput
    - Power efficiency
- Summary

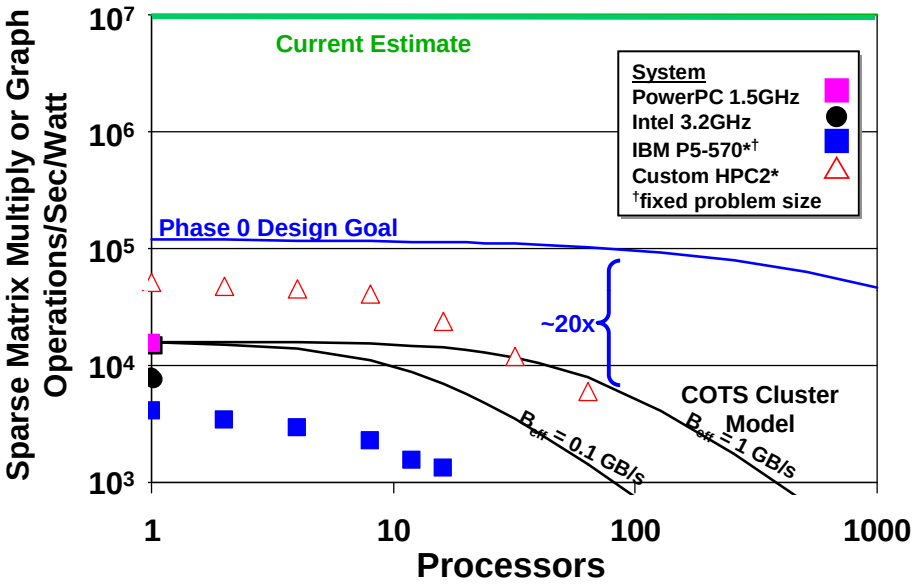
Simulation and Verification



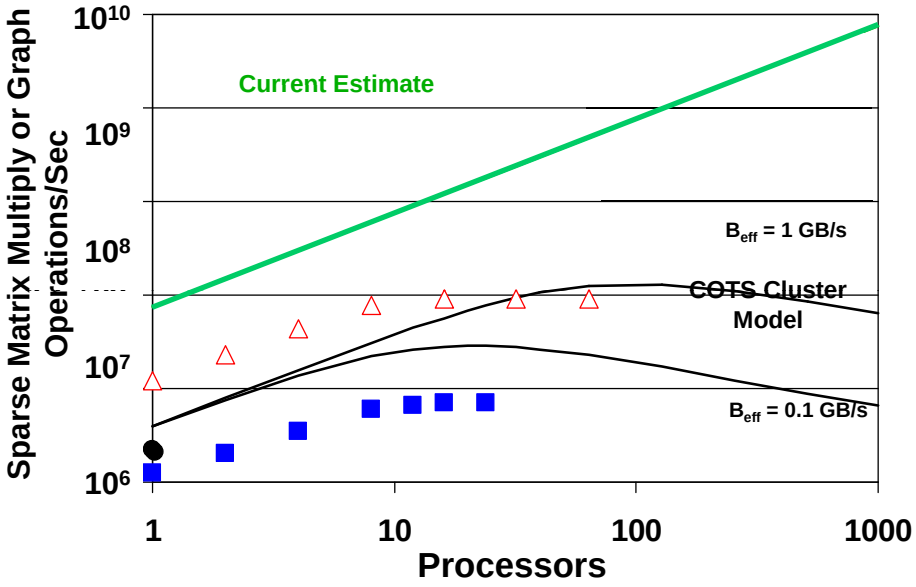
- Bit level accurate simulation of 1024-node system used for functional verifications of sparse matrix processing
- Memory performance verified with commercial IP simulator
- Computational module performance verified with process migration projection from existing circuitries
- 3-D and 2-D high speed communication performance verified with circuit simulation and test chip design

Performance Estimates (Scaled Problem Size, 4GB/processor)

Performance per Watt



Computational Throughput



Close to 1000x graph algorithm computational throughput and 100,000x power efficiency projected at 1024 processing nodes compared to the best COTS processor custom designed for graph processing.

Summary

- Graph processing critical to many commercial, DoD, and intelligence applications
- Conventional processors perform poorly on graph algorithms
  - Architecture poorly match to computational flow
- MIT LL has developed novel 3-D processor architecture well suited to graph processing
  - Numerous innovations enabling efficient graph computing
    - Sparse matrix based instruction set
    - Cache-less accelerator based architecture
    - High speed systolic sorter processor
    - Randomized routing
    - 3-D coupler based interconnect
    - High-speed low-power custom circuitry
    - Efficient mapping for computational load/memory balancing
  - Orders of magnitude higher performance projected/simulated