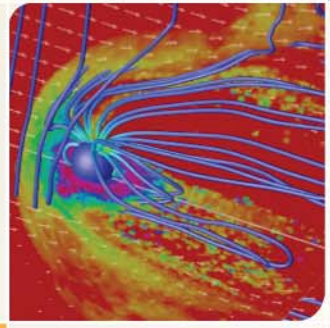
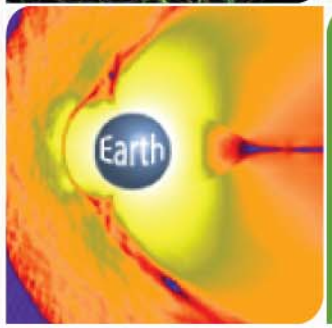
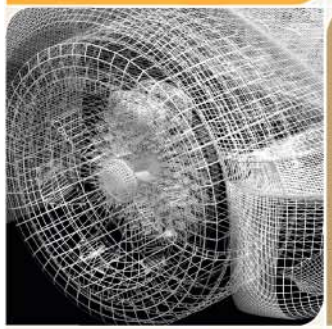





DOSA: Design Optimizer for Scientific Applications

David A. Bader and Viktor K. Prasanna

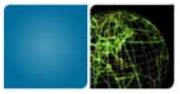
Georgia Tech College of Computing
Computational Science and Engineering

USC UNIVERSITY OF SOUTHERN CALIFORNIA

Acknowledgment of Support: National Science Foundation

- CSR: A Framework for Optimizing Scientific Applications
  - CNS-0614915 (Bader)
  - CNS-0613376 (Prasanna)

USC UNIVERSITY OF SOUTHERN CALIFORNIA

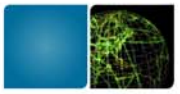




DOSA Motivation

- HPC systems are growing in architectural complexity
  - requires application programmers and compiler writers to perform the challenging task of optimizing the computation in order to achieve high performance.
- In the past decade:
  - Caches and high-speed networks
  - scientists and engineers must carefully lay out data and partition work to reduce communication, maintain a load balance, and expose locality for better cache performance.
- New architectural innovations now include
  - hardware accelerators (e.g., reconfigurable logic such as FPGAs, SIMD/vector processing units such as in IBM Cell, and graphics processing units (GPUs)),
  - adaptable general-purpose processors,
  - run-time performance advisors,
  - capabilities for processing in the memory subsystem and transactional memory,
  - power optimizations.

- With these innovations, the multidimensional design space for optimizing applications is huge.
  - Software must be sensitive to data layout, cache parameters, and data reuse, as well as dynamically changing resources, for highest performance.

USC UNIVERSITY OF SOUTHERN CALIFORNIA


DOSA Objectives

- *Design of a dynamic* ***application composition system*** *that provides*
  - high-performance computing
  - increased productivity
- *Develop a* ***component library representation*** *and example libraries*
  - These are used by the **automatic composition system** to compose optimized applications in the face of dynamically changing resources.
- *Develop the* ***Design Optimizer for Scientific Applications (DOSA),*** *a semi-automatic framework for software optimization.*
  - DOSA will allow rapid, high-level performance estimation and detailed low-level simulation.
- *Demonstrate the DOSA framework on applications from several computationally-intensive areas*
  - This demonstration will include three representative compact applications in molecular dynamics (floating-point arithmetic), dynamic programming (integer), and graph theory (sparse data structures); and one complex, full application for computational chemistry.

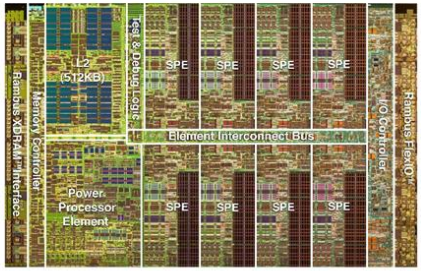
USC UNIVERSITY OF SOUTHERN CALIFORNIA

Example: Multicore Computing

IBM Research

Cell BE Architecture

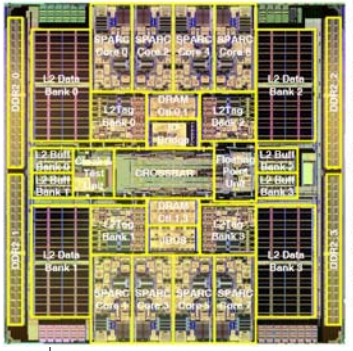
- Combines multiple high performance processors in one chip
  - 9 cores, 10 threads
  - A 64-bit Power Architecture™ core (PPE)
  - 8 Synergistic Processor Elements (SPEs) for data-intensive processing
- Current implementation—roughly 10 times the performance of Pentium for computational intensive tasks
  - Clock: 3.2 GHz (measured at >4GHz in lab)

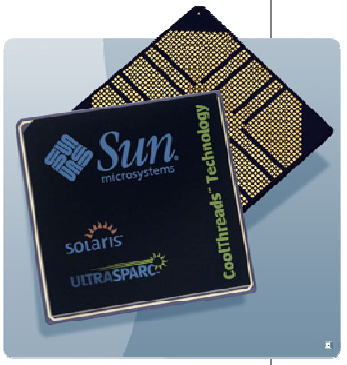


	Cell	Pentium D
Peak I/O BW	75 GB/s	~6.4 GB/s
Peak SP Performance	>200 GFLOPS	~30 GFLOPS
Area	221 mm²	206 mm²
Total Transistors	234M	~230M

4

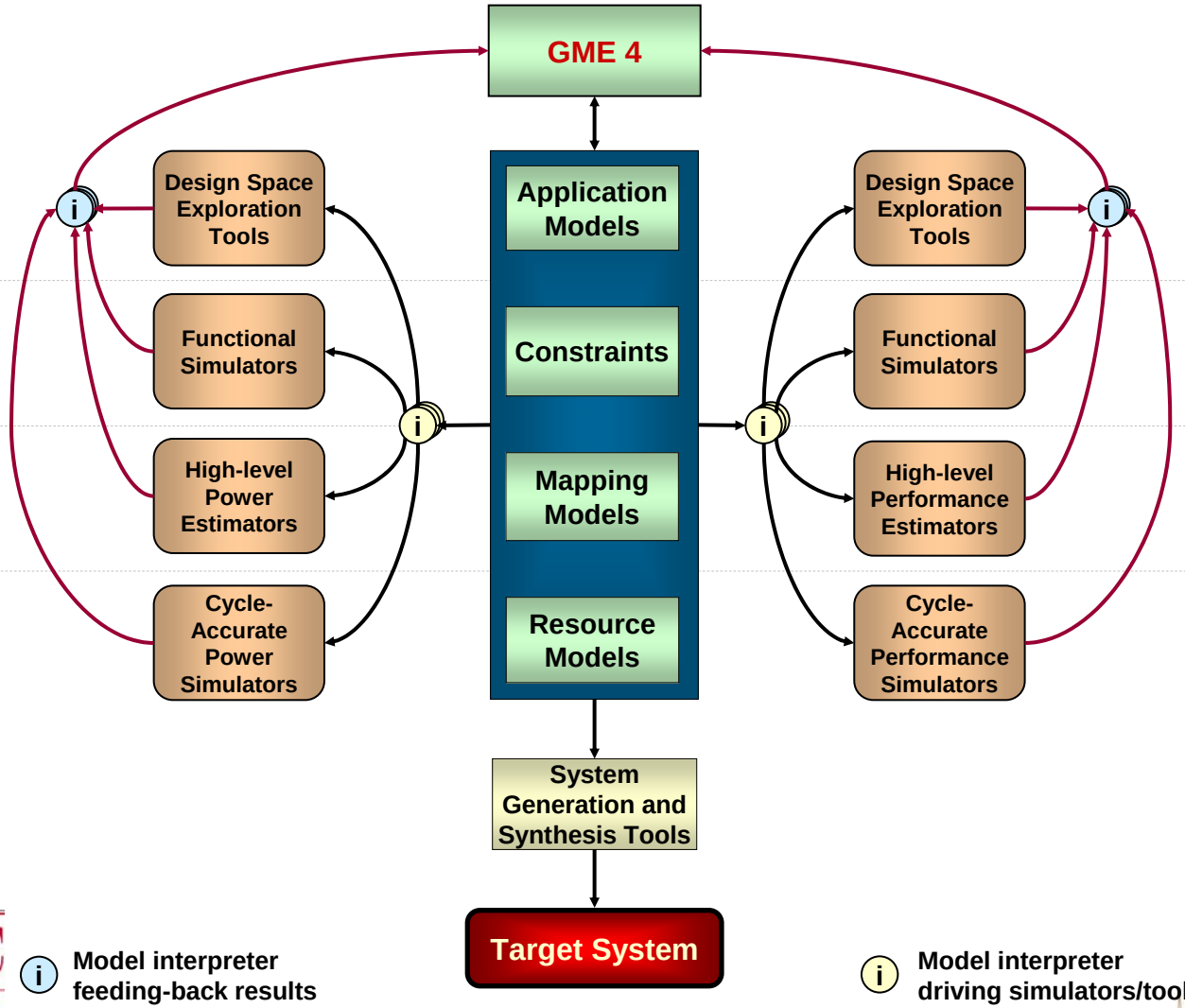
- Sun Fire T2000 Servers
- UltraSPARC T1 "Niagara" processor
- "the highest-throughput and most **eco-responsible** processor ever created"®





DOSA will provide techniques to optimize algorithms to take full advantage of parallelism and concurrency.

MILAN: Model-based Integrated Simulation



MILAN v1.0

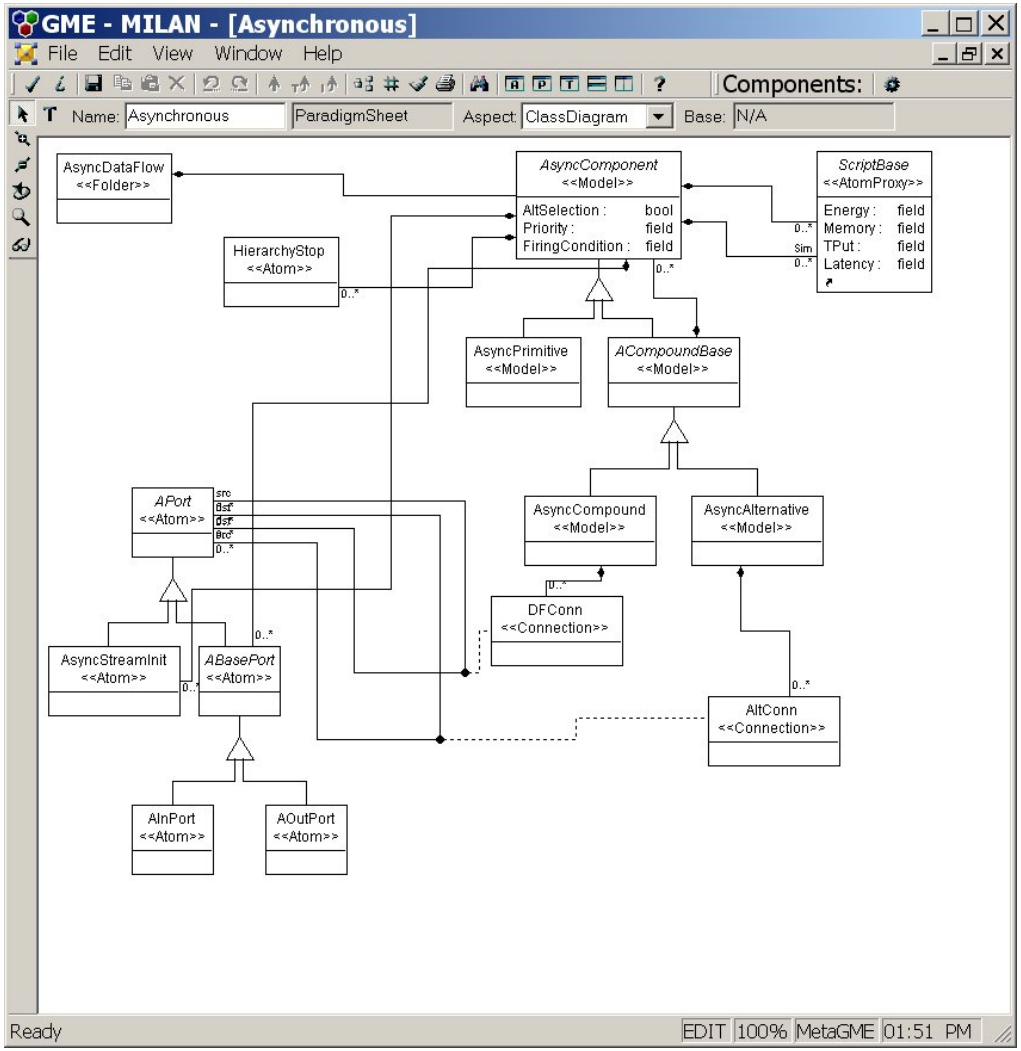
- ***DESERT***
- ***DesignBrowser***
- ***Optimization Tool***

- ***Matlab***
- ***SystemC***
- ***ActiveHDL***

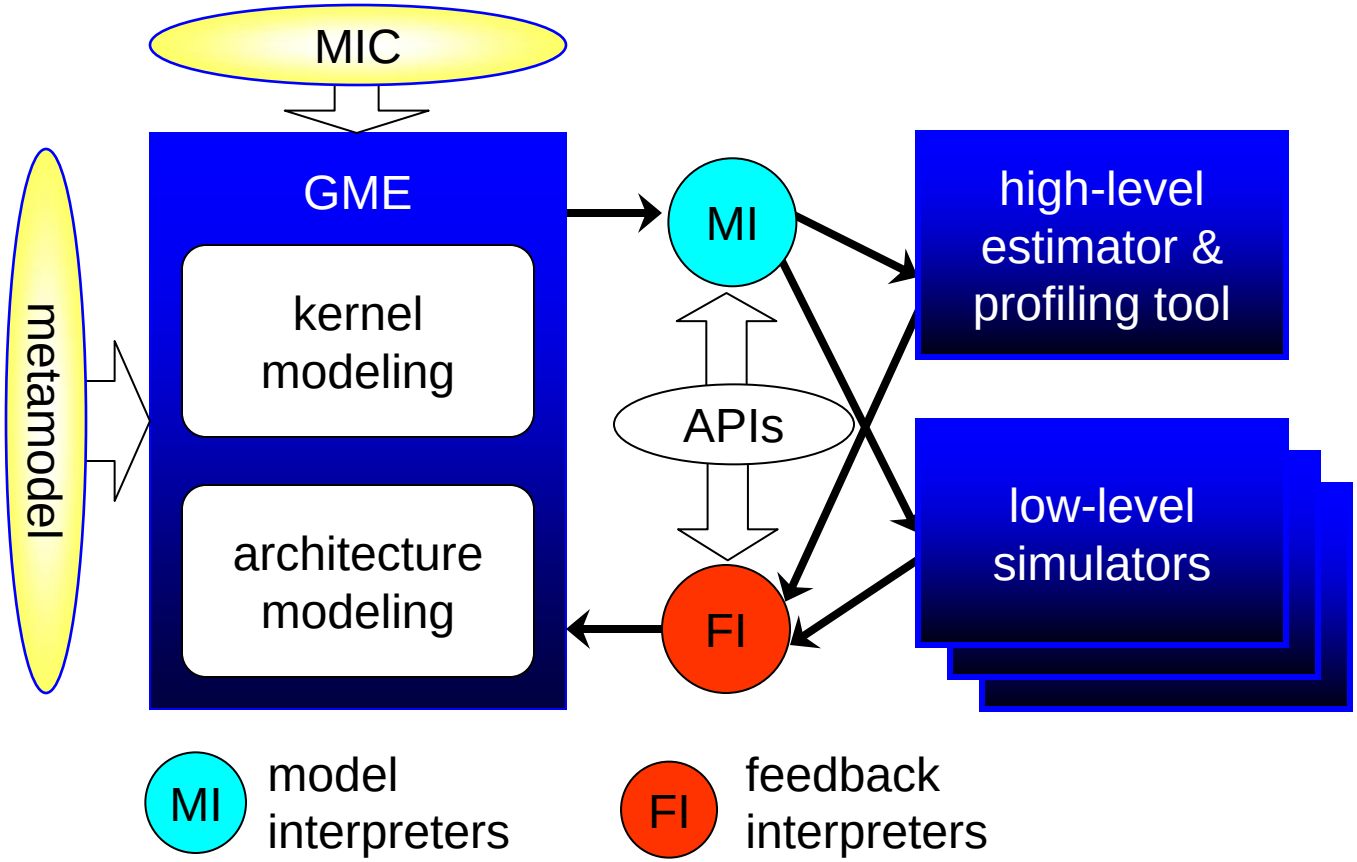
- ***HiPerE***
- ***Kernel Perf. Est.***

- ***Mambo***
- ***SimpleScalar***
- ***PowerAnalyzer***
- ***SimplePower***
- ***Armulator***
- ***CodeComposer***
- ***Jouletrack***
- ***EMSIM***

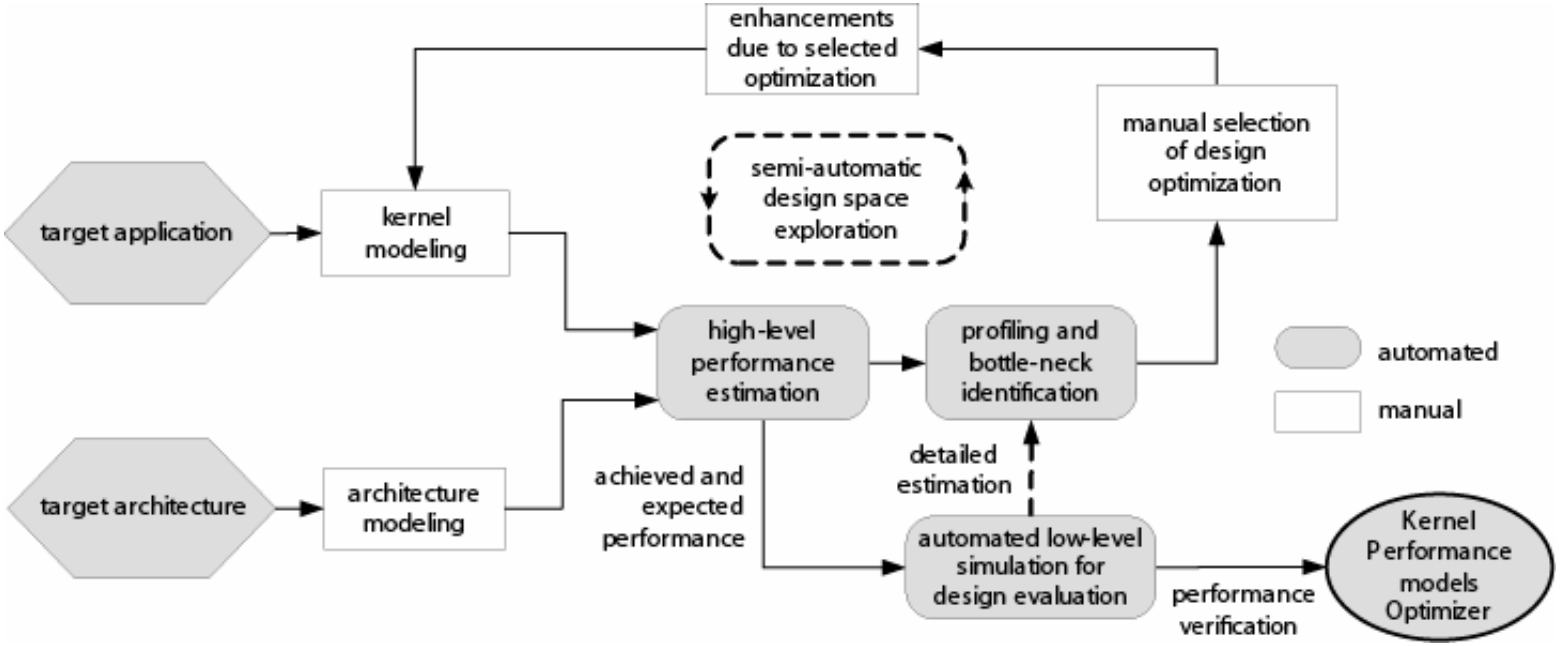
Generic Modeling Environment (GME) – MILAN



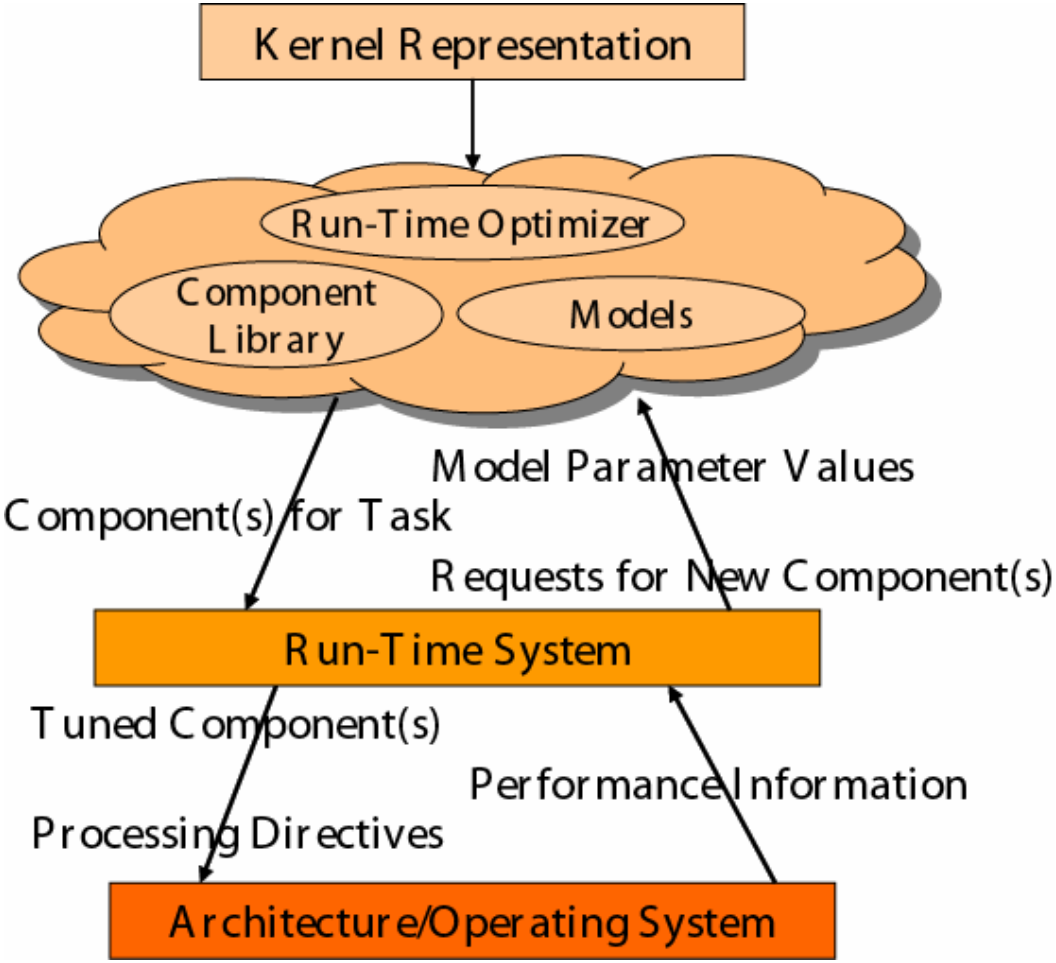
Design of the DOSA Framework

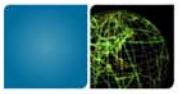


Design Flow using the DOSA Framework



Run-Time Optimizer and Run-Time System

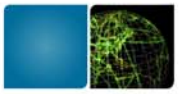



Recent Publications (Model/Framework)

- Low-level modeling
  - On the Design and Analysis of Irregular Algorithms on the Cell Processor: A case study on list ranking, David A. Bader, Virat Agarwal, and Kamesh Madduri, *21th IEEE International Parallel and Distributed Processing Symposium* (IPDPS), Long Beach, CA, March 26-30, 2007.
  - Performance Model for Heterogeneous Systems with Reconfigurable Hardware, Ling Zhuo and Viktor Prasanna, submitted to *International Conference on Field Programmable Logic* (FPL '07), August 2007.
- High-level framework
  - SWARM: A Parallel Programming Framework for Multi-Core Processors, David A. Bader, Varun N. Kanade, and Kamesh Madduri, *First Workshop on Multithreaded Architectures and Applications* (MTAAP), Long Beach, CA, March 30, 2007.
  - DOSA: Design Optimizer for Scientific Applications, David A. Bader and Viktor K. Prasanna, *NSF Next Generation Workshop*, Long Beach, CA, March 25-26, 2007.

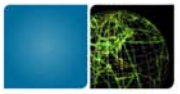
USC UNIVERSITY OF SOUTHERN CALIFORNIA


Recent Publications (Applications)

- Applications
  - Scalable Parallel Implementation of Bayesian Network to Junction Tree Conversion for Exact Inference, Vasanth Krishna Namasivayam, Animesh Pathak and Viktor Prasanna, *18th International Symposium on Computer Architecture and High Performance Computing* (SBAC-PAD 2006), October 2006.
  - An Experimental Study of A Parallel Shortest Path Algorithm for Solving Large-Scale Graph Instances, Kamesh Madduri, David A. Bader, Jonathan W. Berry, and Joseph R. Crobak, *Workshop on Algorithm Engineering and Experiments* (ALENEX), New Orleans, LA, January 6, 2007.
  - A Graph-Theoretic Analysis of the Human Protein-Interaction Network Using Multi-core Parallel Algorithms, David A. Bader and Kamesh Madduri, *Sixth IEEE International Workshop on High Performance Computational Biology* (HiCOMB), Long Beach, CA, March 26, 2007.

Sony-Toshiba-IBM Cell Center of Competence @ Georgia Tech

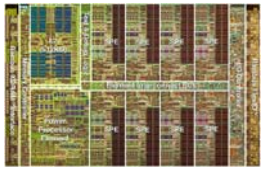

D TUESDAY, NOV. 14, 2006

The Atlanta Journal-Constitution

Business

ajc

CONTACT US: Mark Braykovich, Business editor / mbraykovich@ajc.com / 404-526-5869



"Georgia, not Austin, gets chip center," Bob Keefe, Austin American-Statesman, November 14, 2006

Ga. Tech lands research facility

By BOB KEEFE
bkeefe@ajc.com

Three of the biggest names in technology plan to announce today they will start a research center at Georgia Tech to explore ways to expand the reach of a promising new semiconductor design.

Sony Corp., IBM Corp. and Toshiba Corp. compare their new "Cell" microprocessor to a supercomputer on a chip that can handle some applications 10 times faster than traditional computer chips.

The technology that the companies jointly developed in Austin, Texas, over five years at a cost of $400 million is debuting in Sony's new PlayStation3 video game console. The $500 console went on sale in Japan last week and hits U.S. store shelves on Friday.

Now, the companies want to take the Cell technology much further.

With funding from the three Cell partners and additional money from Georgia Tech and outside grants, researchers at Tech's new STI Center of Competence will explore ways to adapt the technology for other industries, including biotech, finance and digital media creation.

Sony, Toshiba and IBM are providing an initial investment of $320,000, while Tech is putting in $230,000 and another $100,000 is coming from a National Science Foundation grant.

At the center, to be located in the school's new Christopher W. Klaus

➤ Please see RESEARCH, D6

SONY TOSHIBA
USC UNIVERSITY OF SOUTHERN CALIFORNIA
IBM

Research: Tech wins center

➤ Continued from D1

Advanced Computing Building, researchers will also teach students and outside companies how to program computers and write software for the new type of chip. There will be four faculty members involved in the project.

Landing the center puts Georgia Tech at the forefront of a groundbreaking new type of semiconductor design. David Bader, executive director of the school's high-performance computing program, said he believes the center will be the only one of its kind in the United States.

"We really see this as the future of technology and innovation," Bader said. "This is so high-impact."

Austin bypassed

In picking Georgia Tech for the Center of Competence, the Cell partners sidestepped Austin as well as other high-tech hubs across the country. In addition to the University of Texas, more than a dozen schools around the country were vying to land the center, according to officials involved.

"Texas universities were absolutely part of the consideration," said Hina Shah, the Austin-based Cell development program director at IBM. But Georgia Tech won out in the end, she said, partly because its curriculum and areas of expertise matched up better with the interests of the three companies involved.

For Georgia Tech, the center is the latest in a series of big wins and increased prominence for the College of Computing.

In part, the school benefited from its extensive programs in high-performance computing, digital media and video game design.

But since the 2002 arrival as dean of Rich DeMillo, the former chief technical officer for Hewlett-Packard Co., the school has redesigned its curriculum to focus less on computer science theory and more on real-world applications.

"In many ways, we found them to be much more grounded about focusing on what's needed, not 10 years from now, but what's needed today and tomorrow," Shah said.

"That made a huge difference."

Masa Chatani, Sony's senior general manager for Cell development, said in a statement that the "collaboration with the College of Computing at Georgia Tech will create innovative applications for Cell processors."

Georgia Tech computing director **David Bader** believes the center will be the only one of its kind in the country.

The Cell chip design is only in its infancy and has a lot to prove. The chip isn't expected to make a big dent in the traditional semiconductor market controlled by Intel Corp. and Advanced Micro Devices Inc. anytime soon. Reaching into other markets won't be easy either.

Still, what makes Cell so promising is its potential power, especially when it comes to graphics-intensive programs like video games, broadband Internet video processing and other digital media applications.

Just recently, Intel released its first "dual core" and "quad core" microprocessors that essentially put two or four processors on one chip.

Cell chips have already leapfrogged that capability. The chips in Sony's PlayStation3, for instance, essentially have nine cores – eight unique sub-processors that work in connection with a central processor.

16 cores a possibility

Future Cell designs could have as many as 16 sub-processing cores, which could dramatically increase the speed and the number of applications Cell-equipped computers could handle.

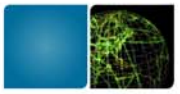
"This really is a new era in performance," Jim Kahle, an IBM fellow who oversaw the chip's design in Austin, said in announcing the first Cell chips in San Francisco last year.

Sony has the most riding on Cell. The Japanese giant is counting on the chip to help it regain ground in new technology development that it lost in areas like digital music.

Along with its video game machines, Sony is exploring putting Cell processors into a wide array of products, including personal computers, televisions and mobile phones.

Toshiba plans to use Cell processors in its TV sets and in other products.

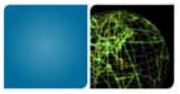
IBM already has introduced powerful computer servers based on the design.


Backup slides

USC UNIVERSITY OF SOUTHERN CALIFORNIA

Georgia Tech College of Computing


DOSA Goals

- Design of a dynamic **application composition system** that provides
  - high-performance computing
  - increased productivity
- Previous methodologies relied on static HPC systems, homogeneous resources, and a well-understood model of execution at each processor
  - design-time analysis
  - optimizing compilers
- DOSA must use **run-time optimizations** that depend on the dynamic nature of the computation and resources
  - load-balancing and job migration
  - readily-accessible grid computing
  - complex reconfigurable architectures
  - adaptive processors
- In a large, complex computing system, the available resources may change during iterations, and the run-time system must monitor, select, and tune new components to maintain or increase performance.

Detailed Design Flow in DOSA