Computer Systems Research in India

R. Govindarajan
Supercomputer Centre
Indian Institute of Science
Bangalore, India
govind@serc.iisc.ernet.in

Overview

- Computer Systems Research in
  - My Group
  - Institute
  - Other Institutions in India
- Opportunities
- Conclusions

Research Focus of our Group

- Processor & Memory Architecture
- Compiler Analysis & Optimization
- High Performance Computing

Processor & Memory Arch.

- Performance and energy efficient structures for superscalar and multi-core processors
- Novel Cache architecture design for General Purpose and Application Specific Architectures
  - Shepherd Cache -- Emulating optimal replacement
  - Novel Two-level cache mapping – Cache placement

- **Performance-centric memory hierarchy design and Prefetching**
  - **Focused Prefetching**
  - **Prefetch Design Space Exploration**
  - **Efficient Cache Architecture for Multi-Cores**
  - **Efficient Memory controller Design**

Memory System for Multi-cores: Objectives

- Improve LLC Performance
  - Make LLC locality and sharing aware
    - Caches closer to the core filter locality
    - Interference in locality due to sharing
    - ⊙ NUCache Organization [HPCA 2011]
  - Make LLC performance predictable – impact on co-running applications
    - Enable fine-grain sharing of cache across applications
    - ⊙ Probabilistic Shared Cache Management (PriSM) [ISCA 2012]

Memory System for Multi-cores: Objectives

- Improve Memory performance
  - Make memory system performance sharing aware (in multi-core)
    - ⊙ Multiple Small Row Buffer (MSRB) Organization [ICS-2011]
    - ⊙ ANATOMY: Analytical Memory Modeling [Under Submission]
    - ⊙ Strawman Model for DRAM Cache Design [Under Submission – PACT 2013]

Compiler Analysis & Optimizations

- ILP Compilation Techniques
  - Software Pipelining, Instrn. Scheduling, Register Alloc. …
- Compiling Techniques for Embedded Systems
- Power-Aware Compiling Techniques
- Path Sensitive Dataflow Analysis

- **Efficient Points-to Analysis**
  - **Linear System of Equations [SAS 2010]**
  - **Probabilistic Points-to Analysis**
  - **Graph-Rewriting for Flow-sensitive [PACT 2013 submission]**
- **Compiler Optimizations for Software Transactional Memory systems**

High Performance Computing

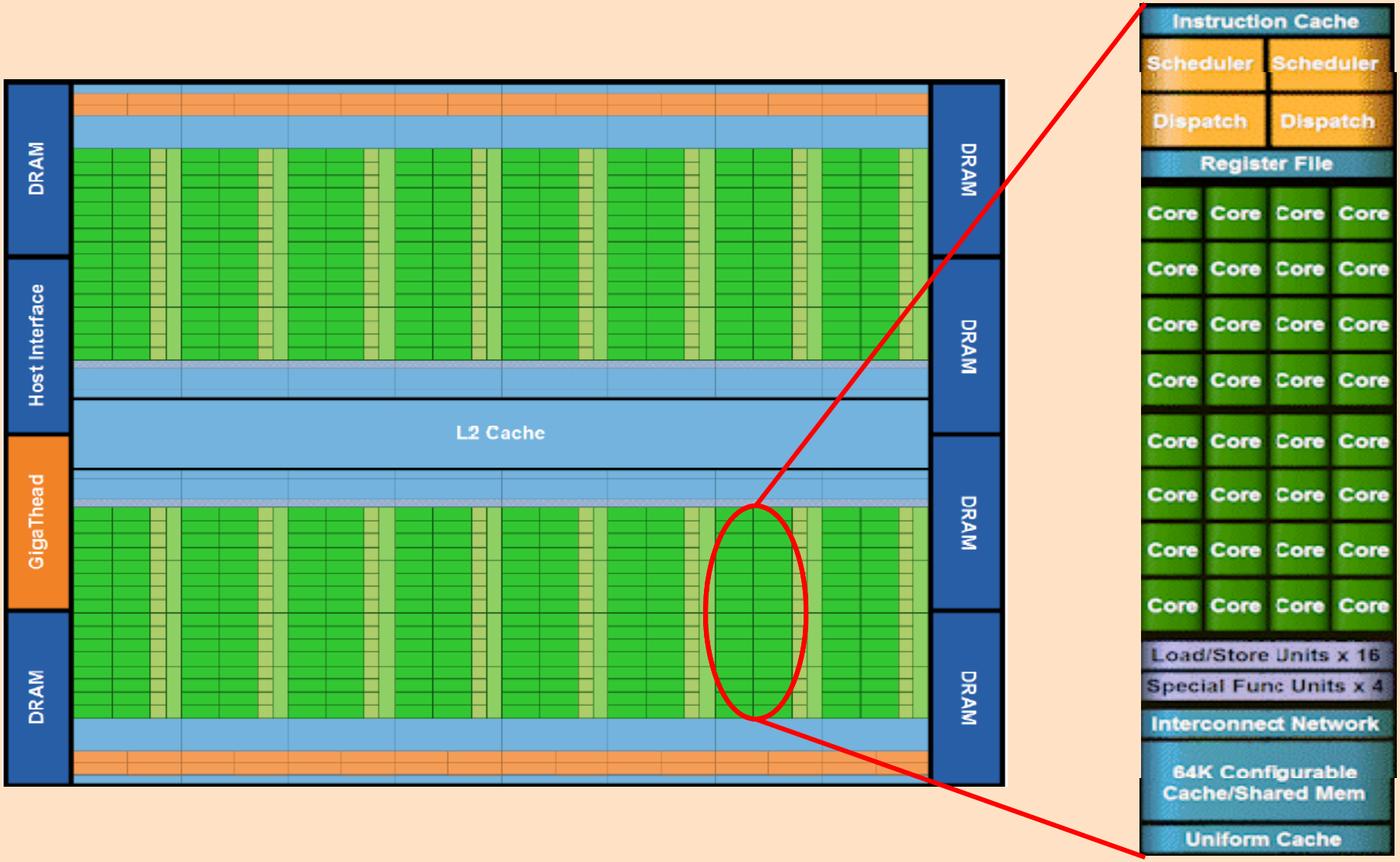
- Multithreaded Architecture,
- Software Distributed Shared Memory Architecture
- Cluster computing
  - Off-loading computation to Communication procs., improving commun. performance thro. compression

- **HPC using Accelerator-based Architectures**
- **Programming Models, Languages, and compilers**

Accelerators: Hype or Reality?

Rank	Site	Manufacturer	Computer	Country	Cores	Rmax [Pflops]	Power [MW]
1	Oak Ridge National Labs	Cray	Titan Cray XK7, Opteron 6274 (2.2GHz) + NVIDIA Kepler K-20	USA	560,640	17.59	8.20
2	Lawrence Livermore Labs	IBM	Sequoia – BlueGene/Q	USA	1,572,864	16.32	7.89
3	RIKEN Advanced Institute for Computational Science	Fujitsu	K Computer SPARC64 VIIIfx 2.0GHz, Tofu Interconnect	Japan	705,024	10.51	12.66
4	DOE/SC/ANL	IBM	BlueGene/Q Power BQC 16C/1.6 GHz	USA	786,432	8.16	3.90
5	Juelich, Germany	IBM	BlueGene/Q Power BQC 16C/1.6 GHz	Ger.	393,216	4.14	1.90
6	Leibniz, Germany	IBM	SuperMUC - iDataPlex Xeon E5-2680 (2.7GHz)	Ger.	147,456	2.89	3.42
7	Texas Advanced Computing Centre	Dell	Stampede Xeon E5-2680 (2.7GHz) + Intel Xeon Phi	USA	204,900	2.66	
8	National SuperComputer Center in Tianjin	NUDT	Tianhe-1A NUDT TH MPP, Xeon 6C, NVidia, FT-1000 8C	China	186,368	2.566	4.04

Accelerator – Fermi S2050



Handling the Multi-Core Challenge

- Shared and Distributed Memory Programming Languages
  - OpenMP
  - MPI
- Other Parallel Languages (partitioned global address space languages)
  - X10, UPC, Chapel, …
- Emergence of Programming Languages for GPU
  - CUDA
  - OpenCL

GPU Programming: Good News

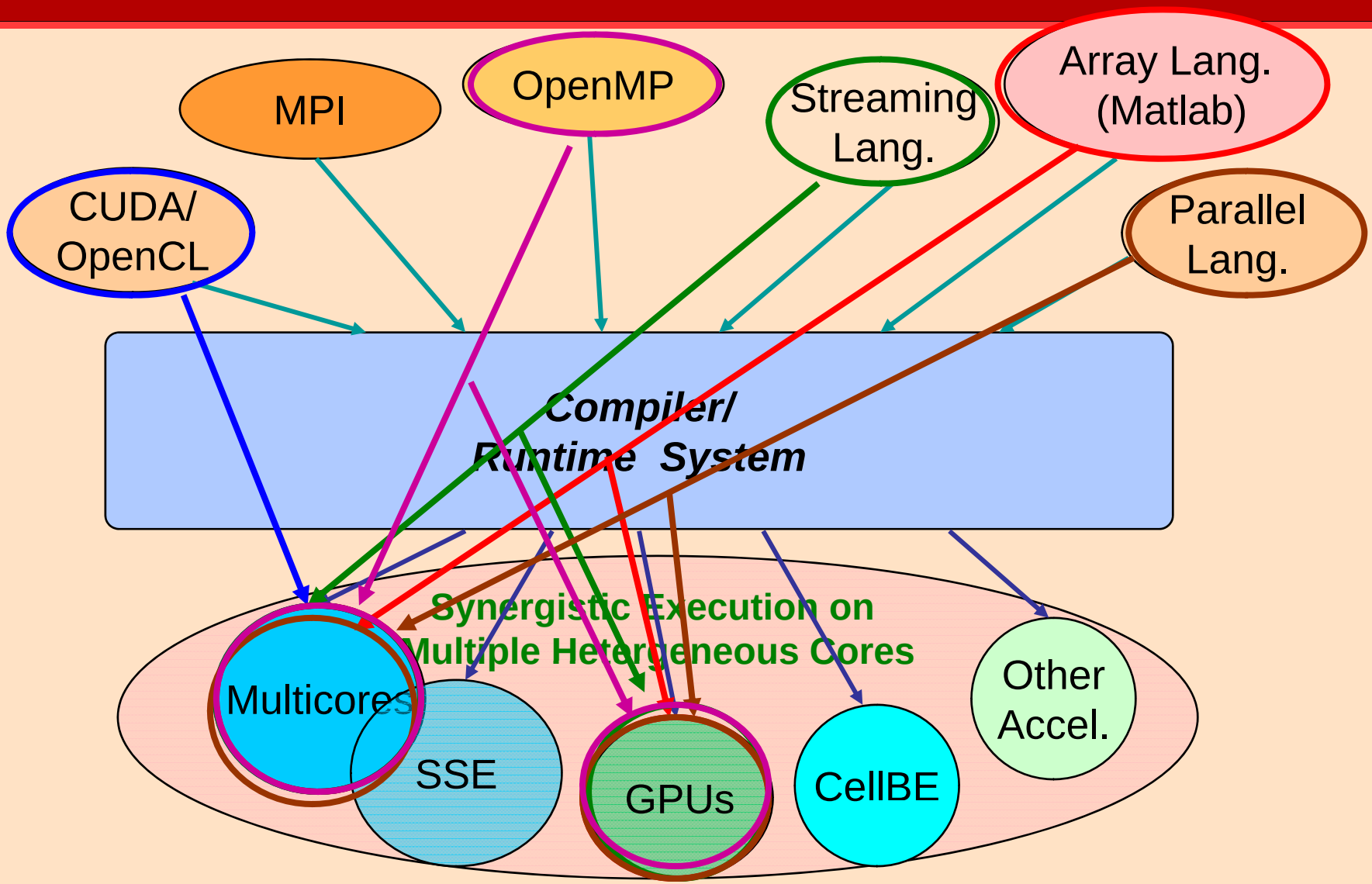
- Emergence of Programming Languages for GPU
  - CUDA
  - OpenCL – Open Standards
- Growing collection of code base
  - CUDAzone
  - Packages supporting GPUs by ISV
- Impressive performance
  - Yes!
- What about Programmer Productivity?

GPU Programming: Boon or Bane

- Challenges in GPU programming
  - Managing parallelism across SMs and SPMD cores
  - Transfer of data between CPU and GPU
  - Managing CPU-GPU memory bandwidth efficiently
  - Efficient use of different level of memory (GPU memory, Shared Memory, Constant and Texture Memory, …
  - Efficient buffer layout scheme to ensure all accesses to GPU memory are coalesced.
  - Identifying appropriate execution configuration for efficient execution
  - Synchronization across multiple SMs

What Parallelism(s) to Exploit?

Our Approach



Other GPU-Related Works

- FluidiCL
  - Cooperative Execution of OpenCL Programs on Multiple Heterogeneous Devices
  - Achieve this automatically (without programmer involvement) and achieve higher performance
- GPU Concurrency [ASPLOS 2013]
  - Allowing multiple concurrent kernels execution in GPU
  - Avoiding many serialization that currently happen in GPUs/CUDA runtime
- Runtime Identification of Parallelism in Do-Across Loops using GPUs [CGO 2013]

Computer Systems Research in IISc

- Embedded Systems & Reconfigurable SoC
- Computer Architecture
- Compiler Analysis & Optimizations
- High Performance Computing
- Grid and Cloud Computing
- Storage Systems

Embedded Reconfigurable SoCs

- Methodologies for mapping applications and algorithms to embedded systems and dynamically reconfigurable SoC platforms
- Compiling applications to Architecture
- REDEFINE: Application Synthesis on General Purpose ASICs
- Network on Chip
- Streaming applications, Security (crytpography), Face Detection, …

Prof. S.K.Nandy
CAD Lab., SERC, IISc
(nandy@serc.iisc.ernet.in)

Computer Architecture

- Performance evaluation methodologies, applying design of experiments for processor architecture using statistical learning techniques
- Performance and energy efficient structures for superscalar and multi-core processors
- Novel Cache architecture design for General Purpose for single and multi-cores
- Performance-centric and power-aware memory system (DRAM, Newer Memory Technology, …)

Prof. R. Govindarajan and Matthew Jacob
HPC and Architecture Lab., SERC, IISc
{govind,mjt}@serc.iisc.ernet.in)

Compiler Analysis & Optimization

- Performance guarantees (WCET) in embedded systems
- Pointer analysis
- Compiler analysis for software testing, …
- Program analysis and verification
- Compiler analysis and optimization for multi-core and GPU systems using Polyhedral models

Prof. Srikant, Raghavan, Aditya Kanade Govindarajan, and Uday Kumar
CSA, IISc

High Performance Computing

- Programming Models, Languages, and Compiling techniques for Heterogeneous Accelerator (GPU)-based architectures
- Compiling programs written in Matlab, X10, OpenMP, StreamIT, and other languages for CPU-GPU synergistic execution
- Compile-time and Runtime methods

Prof. R. Govindarajan and Matthew Jacob
HPC and Architecture Lab., SERC, IISc
{govind,mjt}@serc.iisc.ernet.in)

Grid Computing

- Application-oriented Grid middleware
- Performance modeling and prediction of Grid applications
- Scheduling, Rescheduling, and Meta-scheduling
- Check-pointing and Recovery
- HPC Applications on GPUs

Prof. Sathish Vadhiyar
Grid Appln. Res. Lab., SERC, IISc
(vss@serc.iisc.ernet.in)

Virtualization & Cloud Computing

- System virtualization for QoS properties like performance, security, availability, fault tolerance, etc.
- Cloud computing with focus on SaaS
- QoS features of different cloud architectures
- End-to-end analysis of I/O virtualization architectures

Dr. J. Lakshmi
SERC, IISc
(jlakshmi@serc.iisc.ernet.in)

Large-scale storage

- Long term secure archival storage
- Low power highly available and secure petascale storage
- Integrated nanoscale designs for processing memory and storage
- Storage structures for efficient privacy-aware search
- Access control/Information flow/Privacy Infrastructure for large scale storage systems
- Instrumentation for monitoring performance and detecting anomalies

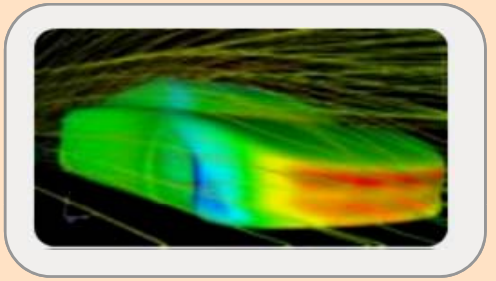
Prof. K. Gopinath
CSA, IISc
(gopi@csa.iisc.ernet.in)

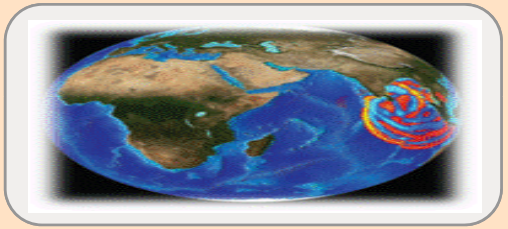
Computer Systems Research at Other Institutions

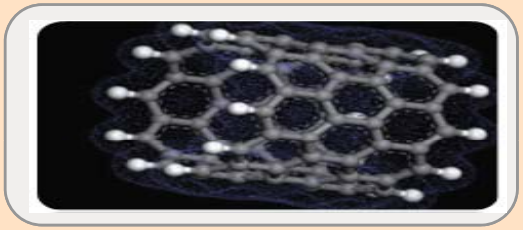
- IIT Kanpur
  - Architecture Group
  - Compiler groups
- IIT Delhi
  - VLSI and Appln. Specific Architectures group
- IIT Bombay
  - Compiler Analysis and GCC group
- IIT Chennai
  - VLSI Group

HPC Applications

- HPC Applications
  - Drug design
  - Computational Fluid Dynamics
  - Computational Chemistry
  - Molecular Dynamics
  - Weather and Climate modeling
- Use of Application packages
  - Gaussian, NAMD, Gromacs, Accelerys, Fluent/Ansys, …
- Development of home-grown code
  - CFD, Physics, weather modeling, …







Strengths - Summary

- Strong research group on computer systems covering all aspects
- Focused development groups
- Diverse HPC applications research group
- Growing interest/expertise in HPC Area

Weaknesses

- Limited Hardware system development activities
- Development of large open source software yet to mature
- Develop into large research groups (critical mass)
- Need to have Large Collaborative projects

Opportunities

- Good synergy and many possible collaboration with HiPEAC
- New thrust/initiative in high performance computing
- Large investment from Indian Govt. on HPC
- Setting up Multiple, tiered HPC centres
- Setting up research labs on
  - System software
  - HPC applications
  - Other areas?

Opportunities

- EU-India Funding
- Already many EU-India Collaboration and synergy!
- EU Computing models (PRACE) – plenty of lessons to learn

Thank You !!