An Overview of Current and Future Accelerator Architectures

ILLINOIS INSTITUTE OF TECHNOLOGY

DataSys
Data-Intensive Distributed Systems Laboratory

Scott J. Krieder
Department of Computer Science
Illinois Institute of Technology
skrieder@iit.edu

Ioan Raicu
Department of Computer Science
Illinois Institute of Technology
iraicu@cs.iit.edu

Abstract

Accelerator technologies are now quite common in Supercomputers, Clusters, Grids, and personal desktops. This work will provide an overview of the current technologies that are available today, and examine future accelerator technologies. This work examines the 3 major competitors in the Accelerator market; nVIDIA, Intel, and AMD.

Why Accelerators?

Accelerators allow the machine to offload work from the CPU to the Accelerator. The Accelerator then completes the computation and returns the solution back to the host CPU.
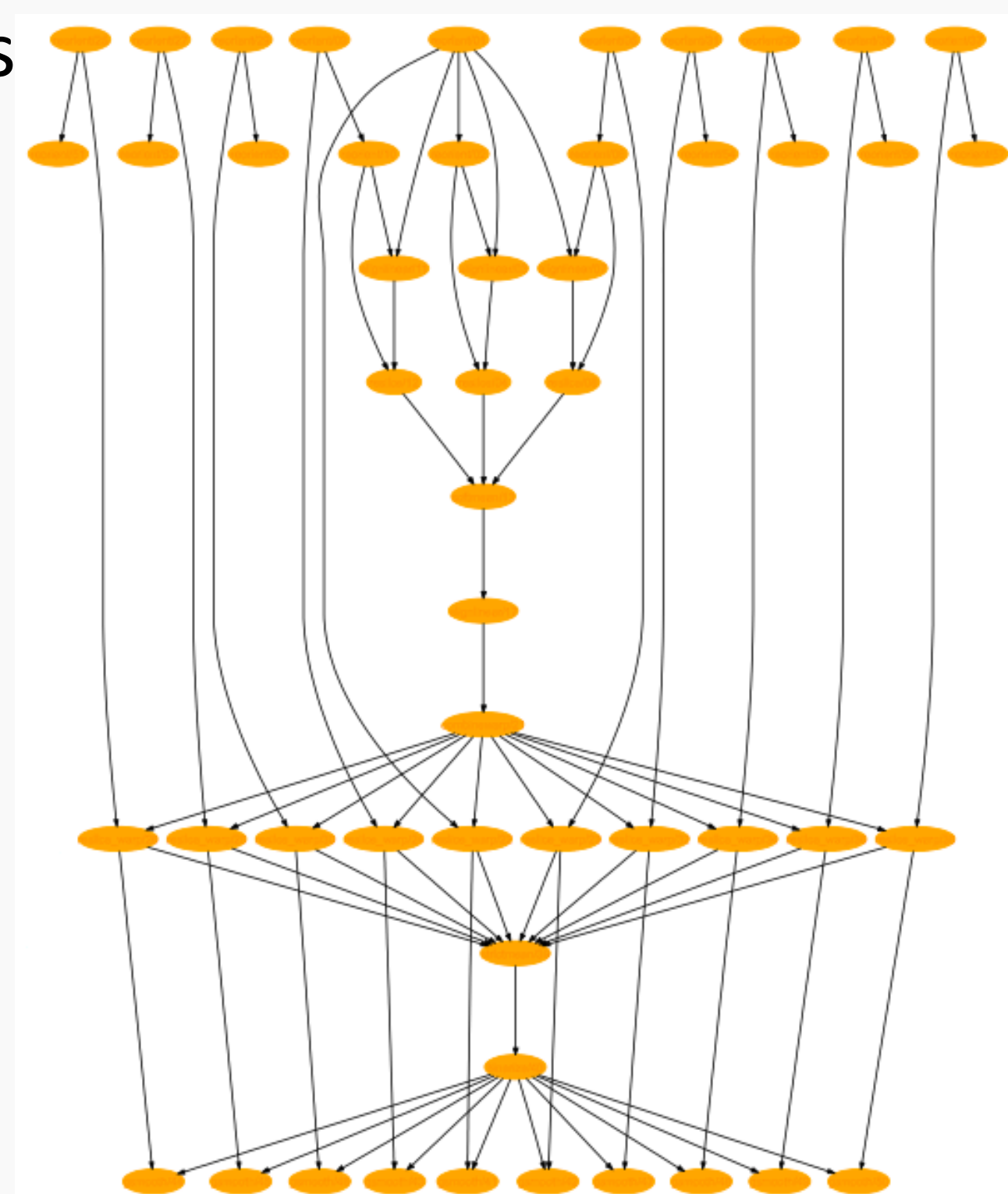
Motivation
- Many Task Computing
- Workflow Systems
- Many Small Scientists
- Programmability
- Bridge gap between Clusters and GPU compute
- Swift + Accelerators

Advantages
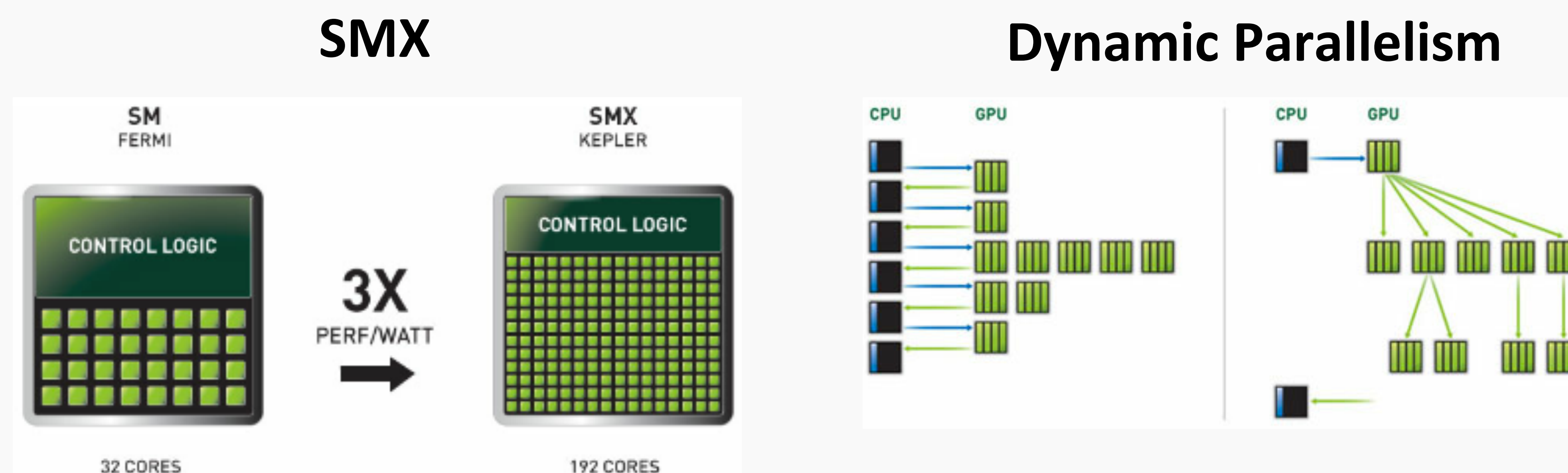- Many cores
- Power efficient

Disadvantages
- Data Transfers
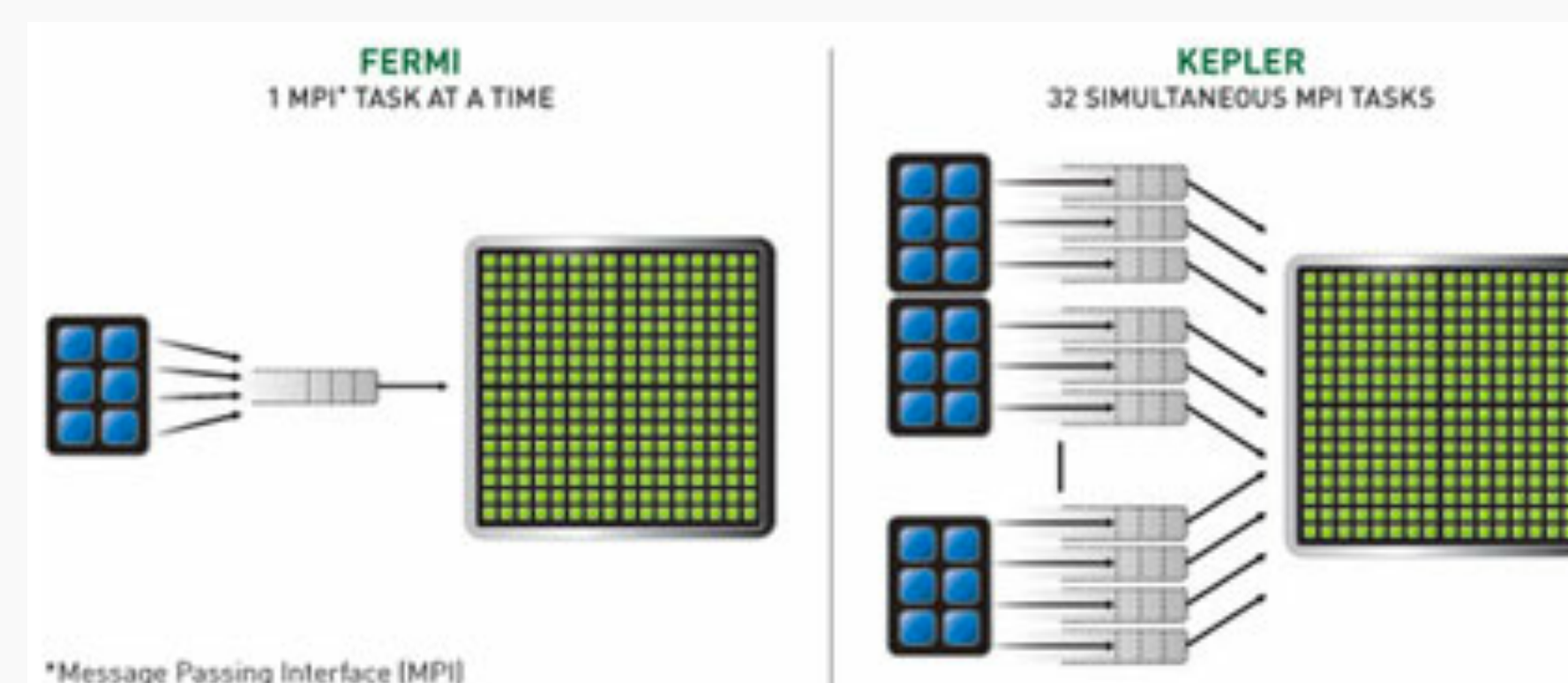- Communication
- Programmability

References

NVIDIA

Kepler Architecture

SMX

SM FERMI
CONTROL LOGIC
32 CORES

3X PERF/WATT

SMX KEPLER
CONTROL LOGIC
192 CORES

Dynamic Parallelism

CPU GPU   CPU GPU

Hyper Q

FERMI
1 MPI* TASK AT A TIME

KEPLER
32 SIMULTANEOUS MPI TASKS

*Message Passing Interface (MPI)

CUDA 5.0
Advanced Libraries
GPU to GPU via. P2P
C, C++, Fortran

Products
GeForce = Desktop Graphics
Quadro = Professional Graphics
Tesla = Server Edition

Intel

Many Integrate Cores (MIC)

Architecture
- 22 Nanometer Chip
- 50 Intel Cores
- X86

Programmability
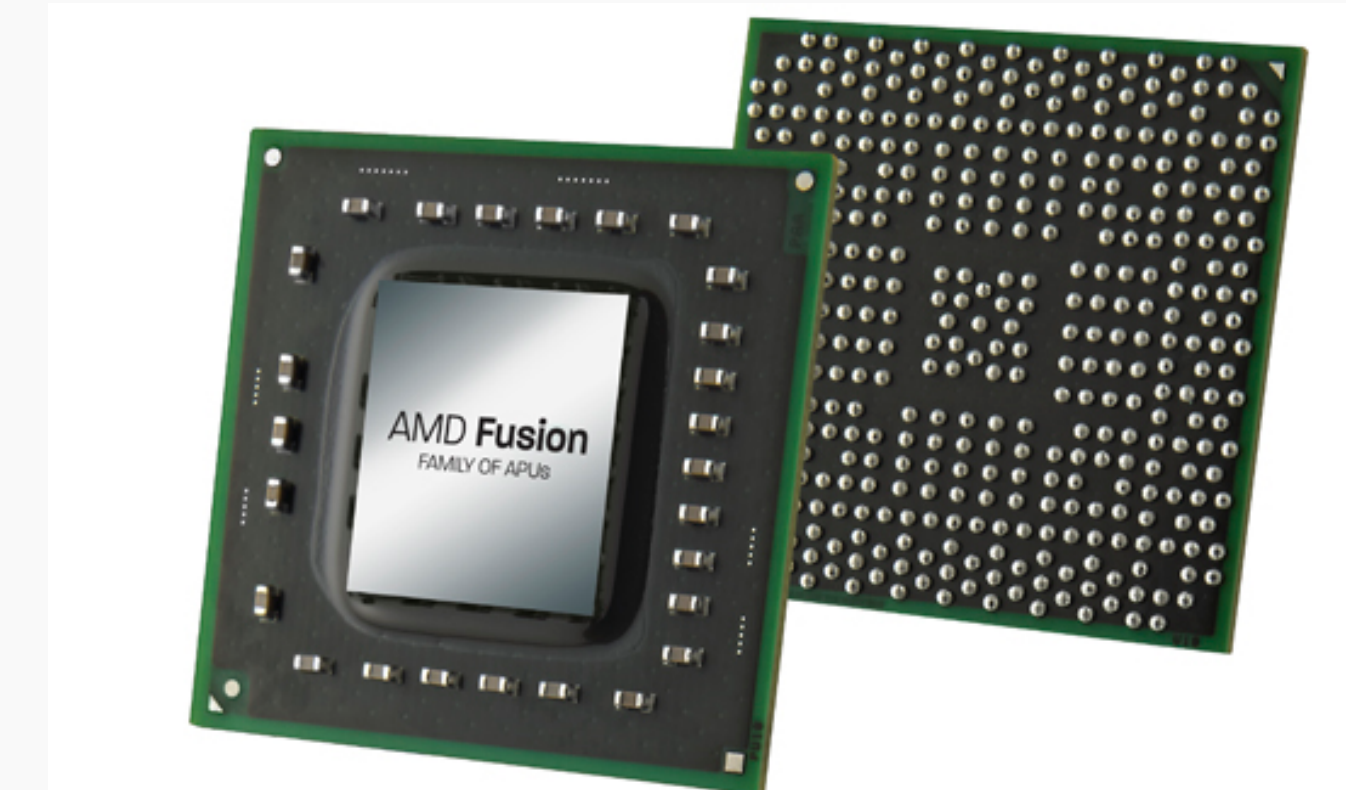- C Programming
- C++ Programing
- Fortran Programming

Products
- **Knights Ferry**
- Research Product
- **Knights Corner**
- Future Product
- Stampede at TACC

AMD

Programmability
- OpenCL
- JavaCL
- Fortran
- C++
- OpenCL .NET
- Pythoncl

Architectures
- GPU 40nm Firestream
- APU 32nm Fusion

Conclusions

Device	Pro	Con
NVIDIA	Maturity Raw Perf.	Programming
Intel	Programming	Availability
AMD	Openness	Adoption

Future Work

- GPU Virtualization
- Workflows with Accelerators

nVIDIA - http://www.nvidia.com/object/nvidia-kepler.html
Intel – http://tinyurl.com/dxhjs3g
AMD - http://developer.amd.com/zones/OpenCLZone/Pages/default.aspx