The Basics

- GPUs are a CPU co-processor
- Developed for the video gaming industry
- Millions of GPUs are sold every year
- GPUs come standard with many desktop and laptop systems

GPUs Explained

Mark Govett
NOAA Earth System Research Laboratory

CPU – GPU Comparison at a Glance

<table>
<thead>
<tr>
<th>CHIP TYPE</th>
<th>CPU Nahalem</th>
<th>GPU NVIDIA Tesla</th>
<th>GPU NVIDIA Fermi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cores</td>
<td>4</td>
<td>240</td>
<td>512</td>
</tr>
<tr>
<td>Parallelism</td>
<td>Medium Grain</td>
<td>Fine Grain</td>
<td>Fine Grain</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Precision</td>
<td>47 GFlops</td>
<td>933 GFlops</td>
<td>1040 GFlops</td>
</tr>
<tr>
<td>Double Precision</td>
<td>60 GFlops</td>
<td>500 GFlops</td>
<td>1040 GFlops</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>130W</td>
<td>150W</td>
<td>220W</td>
</tr>
<tr>
<td>Transistors</td>
<td>730 million</td>
<td>1.4 billion</td>
<td>3.0 billion</td>
</tr>
</tbody>
</table>

Why are GPUs So Fast?

- Design maximizes computational efficiency
- Chip space is dominated by processing units
- High number of compute cores
- Cores are simple, lightweight, low-power

Getting to Operational PetaFlop Computing

Operational Computing at NCEP

Where We Are Today

~180 TeraFlops (2 systems)
5000 IBM Power 6
68th, 69th fastest on Top500 (Nov2009)

Power
0.5 MegaWatts

Reliability
Use two 90 TeraFlop Systems
99.9 reliability requirement

Research Computing: DOE Jaguar

State of the Art in CPU Computing
Innovative building design, cooling, power efficiency
2.3 PetaFlops
250,000 AMD cores
284 cabinets of computing

Power
7-10 MegaWatts
(sufficient for 8-10,000 homes)

Reliability
MTBF: estimated in hours

Cost
Facilities ($73M), System (~ $100M)
Annual Power (~ $4M)

GPU Cluster Computing

Alternative Fermi System
1.0 PetaFlops
1000 NVIDIA Fermi GPUs
500 Intel CPU Nodes
10 cabinets of computing

Power
0.5 MegaWatts

Reliability
MTBF: Estimated in weeks

Cost
System (~ $5M)
Annual Power (~ $250K)

Alternative Computing Technologies

ATI Radeon GPU (2010)
5.0 TeraFlops Performance
Graphics card only
No HPC language support

Intel SandyBridge CPU (2011)
8 cores + "GPU" extensions