Petascale Computing Research Challenges - A Manycore Perspective

Stephen Pawlowski
Intel Senior Fellow
GM, Architecture & Planning
CTO, Digital Enterprise Group

IEEE 13th International Symposium on High-Performance Computer Architecture
Hyatt Regency Phoenix - Phoenix, Arizona - February 10-14, 2007
Yesterday, Today and Tomorrow in HPC

ENIAC
20 Numbers in Main Memory

1946

CDC 6600 - First successful Supercomputer 9MFlops

1965 - 1977

~2008 Beyond

Climate

Astrophysics

Cell-base Community Simulation

Petascale Platforms

ASCI Red
(word fastest on top500 till 2000)
First Teraflop Computer,
9298 Intel Pentium® II Xeon Processors

1997 - 2006

Intel ENDEAVOR
464 Intel® Xeon® Processors 5100 series, 6.85 Teraflop MP Linpack, #68 on top500

2006

Yesterday’s Supercomputing is Today’s Personal Computing
PF on all Top500 reached '04

PF on a single system at ~2009 ..

It takes merely 8 years to move from #1 to being off the list!

1 TF barrier to entry to Top500 in 2005 ..
Many believe that global warming will produce more extremes weather (drought/flooding).

Current models are too coarse for predicting climate change at the national level.

To predict regional climate change:

- Community climate model resolution goal is 10 km
- Currently can simulate 50 days/day on Red Storm at 10 km using NCAR/SNL SEAM
- Typical climate simulation is for 100 yrs.

To Simulate 100 Year Climate:
1.6 PFlops in 6 Month of Computing; 40 PFlops per Week
Highly Multi-threaded Apps and Primitives

- Rigid body game physics
- Fluid simulation
- Portfolio management
- Text mining
- Signal / image processing primitives
- Derivative pricing suite
- Stochastic optimization suite
- Partitioning structure collision tests
- Dense and Sparse matrix primitives
Embarrassingly Parallel Applications

- **Computer Vision**: Body tracking, 4 DV cameras
- **Ray Tracing**: Beetle car scene, 1 mega-pixel
- **Physical Simulation**: Computational Fluid Dynamics (CFD), 150x100x100, 30 fps
- **Financial Analytics**: Asset Liability Management, 8 time-steps, 6 assets, 10 branches, 1 sec

**Research Challenge**: Develop parallel software applications, languages, tools, OS, Firmware
Scalability with Cores

Source: Intel Labs
Exploding Demand for Data Processing

Example: HPC in Medical Imaging

- 3-D renderings of the images
- Computer aided diagnostic algorithms
- Fusions of images from different modalities
  - MRI, CT, PET, and SPECT
- Real-time applications are appearing

Full Body CT - 256 slice/10,000 images: a 20GB file
Processor Performance

Reaching Petascale with ~3,000 Processors in 2010

Source: Intel
A Sample Many Core System

65nm, 4 Cores
1V, 3GHz
10mm die, 5mm each core
Core Logic: 6MT, Cache: 44MT
Total transistors: 200M

45nm, 8 Cores
1V, 3GHz
3.5mm each core
Total: 400MT

32nm, 16 Cores
1V, 3GHz
2.5mm each core
Total: 800MT

22nm, 32 Cores
1V, 3GHz
1.8mm each core
Total: 1.6BT

16nm, 64 Cores
1V, 3GHz
1.3mm each core
Total: 3.2BT

Research Challenge:
Asymmetric vs. symmetric, Homogenous vs. heterogeneous
What kind of applications will benefit?

Note: the above pictures don’t represent any current or future Intel products
Intra-chip Interconnect
Bus for Future Many Core Chip?

Issues:
Slow
Shared, limited scalability?

Benefits:
Power?
Simpler cache coherency

Traditional Bus is Not a Good Interconnect Option
Intra-chip Interconnect Options to Evaluate
Bandwidth, Link Bandwidth and Power

Topology Effect on Bandwidth

Energy Iso-Bandwidth

Interconnect Area Iso-Bandwidth
Performance within the Power Envelope

**Rule of thumb**

<table>
<thead>
<tr>
<th>Voltage</th>
<th>Frequency</th>
<th>Power</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1%</td>
<td>3%</td>
<td>0.66%</td>
</tr>
</tbody>
</table>

![Diagram showing Core and Cache](image)

- **Voltage** = 1
- **Freq** = 1
- **Power** = 1
- **Perf** = 1

- **Voltage** = -20%
- **Freq** = -20%
- **Power** = 1
- **Perf** = ~1.7
**How Do We Feed the Machine?**

**RMS Workload - Bandwidth and Computation Requirements**

<table>
<thead>
<tr>
<th></th>
<th>Memory Bandwidth (GB/s)</th>
<th>Computation (GFlops/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ray Tracing:</td>
<td>150 GB/sec</td>
<td>0.002 B/Flop</td>
</tr>
<tr>
<td>Computer Vision:</td>
<td></td>
<td>0.45 B/Flop</td>
</tr>
<tr>
<td>Physical Sim:</td>
<td></td>
<td>0.17 B/Flop</td>
</tr>
<tr>
<td>Financial Analytics:</td>
<td></td>
<td>6.5 B/Flop</td>
</tr>
</tbody>
</table>

**Memory Bandwidth and Processor Performance Need to Keep Pace**

Source: Intel Labs
Memory Bandwidth Vision: 3D Die Stacking

- Power and IO signals go through DRAM to CPU
- Thin DRAM die
- Through DRAM vias

DRAM, Voltage Regulators, and High Voltage I/O All on the 3D integrated die
Memory Performance for Balanced Computing

Bytes Per FLOP

Source: Intel

Byte : Flop Ratio has been Consistent and Steady
Silicon Photonics Future I/O Vision

HPC and Data Center Fabrics

Chip-to-Chip Interconnects

Backplane and Display Interconnects

Chemical Analysis

Medical Lasers

Research Challenge: Intra-chip and Inter-chip I/O Architecture and Topology Options
System Power/Cooling Efficiency

Silicon:
- Moore’s law
- Strained silicon
- Transistor leakage control techniques
- Clock gating

Processor:
- Policy-based power allocation
- Multi-threaded cores

System Power Delivery:
- Fine grain power management
- Ultra fine grain power management

Facilities:
- Air cooling and liquid cooling options
- Vertical integration of cooling solutions

Research Challenge:
- OS, VMM, manageability software
- take control of system and facility level power management
Reliable Systems With Unreliable Components

Architectural Techniques
- **Micro Solutions**
  - Parity
  - SECDED ECC
  - π bit

- **Macro Solutions**
  - Lockstepping
  - Redundant multithreading (RMT)
  - Redundant multi-core CPU

Circuit Techniques
- **Device Param Tuning**
- **Rad-hard Cell Creation**

Process Techniques
- **State-of-Art Processes**

Research Challenge:
From software (Apps, OS, VMM, etc.) to hardware
a reliable Petascale HPC system needs management top down

Detect, Correct, Log and Signal the errors
What can we expect!

Reaching Petascale and Beyond with Energy Efficiency

Source: Dr. Steve Chen, "The Growing HPC Momentum in China", June 30th, 2006, Dresden, Germany