Performance comparisons with HECToR

Alan Gray (EPCC) / Mike Ashworth (STFC)

HPCx Terascaling Team
Acknowledgements

**EPCC**
- Joachim Hein
- Fiona Reid
- Michele Weiland
- Tom Edwards

**STFC**
- Roderick Johnstone
- Ian Bush
- Ilian Todorov
- Andy Sunderland

- CSCS and ORNL for access to systems
Overview

- System Introduction
- Synthetic Benchmark Results
- Application Benchmark Results
- Conclusions
Systems for comparison

• HPCx (Phase 3): 160 IBM e-Server p575 nodes
  - SMP cluster, 16 Power5 1.5GHz cores per node
  - 32 GB of RAM per node (2GB per core)
  - IBM Federation interconnect
  - 12.9 TFLOP/s Linpack, No 101 on top500

• HECToR (Phase 1): Cray XT4
  - MPP, 5664 nodes, 2 Opteron 2.8GHz cores per node
  - 6 GB of RAM per node (3 GB per core)
  - Cray Seastar2 torus network
  - 54.6 TFLOP/s Linpack, No 17 on top500

• Also included in some plots are:
  - HECToR Test and Development system (TDS)
    • Cray XT4, 64 nodes: 2.6GHz dual core, 4GB RAM/node
  - ORNL Cray XT4, 2.6GHz dual core, 4GB RAM/node
  - CSCE Cray XT3, 880 cores, 2.6 GHz
  - IBM BlueGene/L
<table>
<thead>
<tr>
<th></th>
<th>HPCx</th>
<th>HECToR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chip</td>
<td>IBM Power5 (dual core)</td>
<td>AMD Opteron (dual core)</td>
</tr>
<tr>
<td>Clock</td>
<td>1.5 GHz</td>
<td>2.8 GHz</td>
</tr>
<tr>
<td>FPUs</td>
<td>2 FMA</td>
<td>1 M, 1 A</td>
</tr>
<tr>
<td>Peak Perf/core</td>
<td>6.0 GFlop/s</td>
<td>5.6 GFlop/s</td>
</tr>
<tr>
<td>cores</td>
<td>2560</td>
<td>11328</td>
</tr>
<tr>
<td>Peak Perf</td>
<td>15.4 TFLOP/s</td>
<td>63.4 TFLOP/s</td>
</tr>
<tr>
<td>Linpack</td>
<td>12.9 TFLOP/s</td>
<td>54.6 TFLOP/s</td>
</tr>
</tbody>
</table>
Synthetic Benchmarks

• Memory Bandwidth
  - Streams

• MPI Bandwidth
  - Intel MPI Benchmarks
    • PingPing
    • AlltoAll
Memory bandwidth - Streams

Streams Triad Benchmark
System Comparison

Bandwidth (load+store) (MB/s)

Array Size (Bytes)

- TDS: 2 cores per node
- TDS: 1 core per node
- HECToR: 2 cores per node
- HECToR: 1 core per node
- HPCx: 16 cores per node
- HPCx: 8 cores per node
- BG/L: 2 cores per node
- BG/L: 1 core per node
Memory bandwidth - Streams

• Can clearly see caches
• HECToR better at L1, slightly better on main memory
  - HPCx has advantage for intermediate array sizes.
• Underpopulating nodes (1 core per chip) gives improvements on both systems
  - memory bandwidth cannot sustain 2 cores per chip
  - HECToR worse than HPCx, especially on main memory
  - Of course, 1 core/chip means double the resource for same no. tasks
• TDS has lower clock rate than HECToR, but has higher bandwidth from main memory!
  - 4=2+2 GB RAM on TDS is symmetric, interleaving possible
  - 6=4+2 GB RAM on HECToR only allows partial interleaving
MPI bandwidth - PingPing

Intel MPI Multi Ping Ping Benchmark
System Comparison

- HECToR: 2 cores per node
- HPCx: 16 cores per node
MPI bandwidth - AlltoAll

HPCx Annual Seminar
26th November 2007

AlltoAll Benchmark
System Comparison

- HECToR: 2 cores per node
- HPCx: 16 cores per node
- BlueGene/L: 2 cores per node
Applications

- Fluid Dynamics
  - PDNS3D
- Fusion
  - Centori
  - GS2
- Ocean Modelling
  - POLCOMS
- Molecular Dynamics
  - DL_POLY
  - NAMD
  - AMBER
  - LAMMPS
Fluid Dynamics: PDNS3D (PCHAN)

- Finite difference code for Turbulent Flow
  - shock/boundary layer interaction (SBLI)
- Simulates the flow of fluids to study turbulence
- T3 benchmark - Involves a 360x360x360 grid
- Developed by Neil Sandham, University of Southampton
PDNS3D - compilation optimization

PDNS3D T3 Benchmark
HECToR PGI Compilation Flag Comparison, 64 cores

Time (s)

-00
-01
-02
-03
-04
-fast
-fast -03
-fast -04
-fast -03 -Mipa=fast,inline
-fast -04 -Mipa=fast,inline

HPCx Annual Seminar
26th November 2007
PDNS3D - system comparison

PDNS3D T3 Benchmark System Comparison

- HECToR
- HPCx Phase 3

Time * Cores (s) vs Cores
• Underpopulating nodes gives huge improvement (in terms of performance/core) on HECToR, slight improvement on HPCx
• TDS outperforms HECToR
• c.f. streams
PDNS3D – Optimised version

- New optimised version less sensitive to memory bandwidth
PDNS3D – Optimised version

HPCx Annual Seminar 26th November 2007
Fusion

- **Centori**
  - simulates the fluid flow inside a tokamak reactor developed by UKAEA Fusion in collaboration with EPCC

- **GS2**
  - Gyrokinetic simulations of low-frequency turbulence in tokamak developed by Bill Dorland et al.

*ITER tokamak reactor (www.iter.org)*
Centori, 128x128x128 problem
System Comparison

- HECToR
- HPCx
- TDS

Time * Cores (s)

Cores

1 10 100 1000 10000
GS2 NEV02 benchmark System Comparison

- HECToR
- HPCx
- HECToR, 1 core per node
- TDS
- TDS, 1 core per node

Time * Cores (s)

Cores

HECToR, 1 core per node
Ocean Modelling: POLCOMS

• Proudman Oceanographic Laboratory Coastal Ocean Modelling System (POLCOMS)
  - Simulation of the marine environment
  - Applications include coastal engineering, offshore industries, fisheries management, marine pollution monitoring, weather forecasting and climate research
  - Uses 3-dimensional hydrodynamic model
Molecular dynamics

- **DL_POLY**
  - general purpose molecular dynamics package which can be used to simulate systems with very large numbers of atoms

- **LAMMPS**
  - Classical Molecular Dynamics - can simulate wide range of materials

- **NAMD**
  - classical molecular dynamics code designed for high-performance simulation of large biomolecular systems

- **AMBER**
  - General purpose biomolecular simulation package

Protein Dihydrofolate Reductase
DL_POLY - system comparison

DL_POLY 3.08 - GRAMICIDIN A WITH WATER SOLVATING
(792960 atoms)
System Comparison

HPCx Annual Seminar
26th November 2007
DL_POLY - system comparison

DL_POLY 3.08 - GRAMICIDIN A WITH WATER SOLVATING
(792960 atoms)

HECToR

HECToR
- 1 core per node

Cores

0 50 100 150 200 250 300 350

timestep time * Cores (s)
LAMMPS

LAMMPS Rhodopsin benchmark, 4096000 atoms
System Comparison

- HPCx Phase 3
- HPCx Phase 3 SMT
- HECToR

Loop Time * Cores (s) vs. Cores

HPCx Annual Seminar
26th November 2007
LAMMPS Rhodopsin benchmark, 4096000 atoms
HECToR

Loop Time * Cores (s)

10 100 1000 10000

Cores

HECToR
HECToR 1 core per node
HECToR TDS
HECToR TDS 1 core per node
NAMD Benchmarks
System Comparison

- HECToR: F1ATP
- HPCx: F1ATP
- HECToR: AP0-A1
- TDS: APO-A1
- HPCx: AP0-A1
- HECToR: AP0-A1, 1 c/n
AMBER PMEMD Factor IX Benchmark (90906 atoms)

System Comparison

- HECToR
- TDS
- HPCx
- HECToR, 1 c/n
- HPCx: 8 c/n
- BG

Time * Cores (s)

1 10 100 1000

Cores

26th November 2007
Conclusions

- On a core by core basis, not much difference to HPCx in terms of application performance
  - But HECToR has many more cores and a more scalable interconnect
- Scaling slightly better at high core counts on HECToR
- Memory bandwidth cannot sustain fully populated nodes for both systems
  - general problem for HPC systems these days
  - This is seen in performance of memory bandwidth sensitive applications
  - Problem is worse for HECToR, especially with current non-symmetric memory setup.
- To gain benefit from the increased performance of HECToR, your app has to use more cores and therefore be more scalable.