Chemistry Applications.
Development and Performance on the Bradford Xeon Cluster

Martyn F. Guest
CCLRC Daresbury Laboratory
m.f.guest@daresbury.ac.uk

Acknowledgements:
P. Sherwood, H.J. van Dam, DL
J. Nieplocha, E. Apra, PNNL

http://www.cse.clrc.ac.uk/disco/hw-perf.shtml
Outline

- Background - Distributed computing at DL
- Commodity-based and High-end Systems
  - Single-node performance & the Interconnect bottleneck
  - Prototype Commodity Systems; CS1 - CS20
  - High-end Systems: IBM SP/p690+, SGI Altix 3700, plus HP/Compaq Alpha Server SC and SGI Origin 3800
  - Performance Metrics
- Application performance
  - Molecular Simulation (DLPOLY, DLMULTI and CHARMM)
  - Electronic Structure - Global Arrays (GAs) ; Linear Algebra (PeIGS)
    - NWChem and GAMESS-UK
  - Materials Simulation (CRYSTAL, CPMD & CASTEP) and Computational Engineering (ANGUS, SBLI)
- Summary
### Capability and Capacity Computing

<table>
<thead>
<tr>
<th>Specification</th>
<th>Usage</th>
<th>Cost Units</th>
<th>CPU</th>
<th>Memory</th>
<th>I/O</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPP/ASCI</td>
<td>HPC community</td>
<td>10,000</td>
<td>2,000</td>
<td>2,000</td>
<td>200-300</td>
</tr>
<tr>
<td>SMP</td>
<td>Department</td>
<td>100</td>
<td>15</td>
<td>20</td>
<td>20-30</td>
</tr>
<tr>
<td>PC</td>
<td>Desktop</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**MPP/ASCI**
- 1536 CPUs, IBM p690+ (1 TB)
- HPC community: 10,000 CPUs, IBM p690+ (1 TB)

**SMP**
- 16-processor SGI Altix 350, Itanium 2 (8GB RAM)
- Department: 100 CPUs, 15 GB RAM

**PC**
- Pentium-4 Xeon / 3GHz (512 MByte, 30 GB)
- Desktop: 1 CPU, 1 GB RAM

**Commodity Systems (1.5 x N)**
- 3 year continual access to 32 CPUs:
  - High-end (£0.5 / CPU hour): £420,000
  - In-house Beowulf: £50,000
# High-end Commodity-based Systems

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Site</th>
<th>Nodes/Processors</th>
<th>Peak CPU (Gflop)</th>
<th>Interconnect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MCR</td>
<td>LLNL, US</td>
<td>1152/2304</td>
<td>11174 P4/2400</td>
<td>QsNet</td>
</tr>
<tr>
<td>2. Corporate</td>
<td>GX Tech. US</td>
<td>1792/3264</td>
<td>5091 Ultra2/450+</td>
<td>GigabitE.</td>
</tr>
<tr>
<td>3. RHIC</td>
<td>Brookhaven</td>
<td>1369/2738</td>
<td>4780 PIII/P4</td>
<td>GigabitE.</td>
</tr>
<tr>
<td>4. hpc</td>
<td>S. Calif. US</td>
<td>759/1524</td>
<td>4693 P4/1400</td>
<td>Myrinet 2k</td>
</tr>
<tr>
<td>5. Supermike</td>
<td>Louisana US</td>
<td>512/1024</td>
<td>3686 P4/1800</td>
<td>Myrinet 2k</td>
</tr>
<tr>
<td>8. HELICS</td>
<td>Heidelberg</td>
<td>256/512</td>
<td>1434 AMD MP/1400</td>
<td>Myrinet</td>
</tr>
<tr>
<td>9. Holk</td>
<td>Gdansk</td>
<td>128/256</td>
<td>1331 Itanium2/1300</td>
<td>GigabitE.</td>
</tr>
</tbody>
</table>
DisCo: Technical Progress in 2003-4

Hardware and Software Evaluation:

- **CPU**
  - IA32, x-86 and IA64 systems -
    - Intel Pentium 4 and Xeon Systems (3.06 GHz),
      AMD Opteron 246 & 248 (2.0 & 2.2 GHz)
  - Itanium2 (Intel Tiger 1, 1.2 and 1.5 GHz;
    HP systems, 900 MHz, 1 and 1.5 GHz;
    SGI Altix 3700 - 1.3 (3MB L3) & 1.5 GHz (6MB L3)

- **Networks**
  - Gigabit Ethernet options, cards, switches, channel-bonding, ...
  - SCI, Infiniband and Myrinet (P4/2400, P4/2666, P4/2800, Opteron 246 & 248 Clusters:
    OCF, Streamline, ClusterVision & Workstations UK), Quadrics

- **System Software**
  - message passing S/W (LAM MPI, LAM MPI-VIA, MPICH, VMI, SCAMPI), libraries
    (ATLAS, NASA, MKL, ACML, ScaLAPACK), compilers (Absoft, PGI, Intel’s ifc and
    efc, Pathscale, GNU/g77), tools (GA tools, PNNL)
  - resource management software (PBS, TORQUE, GridEngine, LSF etc.)

www.cse.clrc.ac.uk/Activity/DisCo
CCP1: Molecular Electronic Structure

- Working Group (29 Members from 17 Universities)
  - Chaired by Prof P.J. Knowles (University of Birmingham)
- Software
  - GAMESS-UK, CRYSTAL and ChemShell
- Study Weekends and Workshops
  - QM/MM methods (St Andrews, 1995)
  - Quantum Chemistry on MPP Computers (Cambridge, 1995)
  - Quantum Mechanics of Large systems (Daresbury, 1996)
  - Ab Initio Molecular Dynamics (Daresbury, 1998)
- Flagship Projects
- DL Staff Support
  - M. F. Guest, P. Sherwood, H.J.J. van Dam
Working with the Community I.

Chemistry Department, University of Bristol (Gabriel Balint Kurti)

- Provided initial benchmarking Suite incorporated in the procurement process.
- Exercised initial machine as part of acceptance tests (early user access), working with Bristol support staff (Ian Stewart)
  - Installed and validated wide variety of application codes as part of this process (NWChem, GAMESS-UK, DL_POLY, CHARMM, CPMD etc.)
  - Revealed shortcomings in initial Myrinet software
- Presented findings in visit to the University post-acceptance
- Provided application-code base to User Community (subject to appropriate licencing etc.)
- Potential “applications-driven” model for future support activities

Biochemistry Department, University of Essex (Chris Reynolds)
Exercised initial machine during acceptance tests (early user access), working with Bradford support staff (Victoria Pennington)

- Installed and validated wide variety of application codes as part of this process (NWChem, GAMESS-UK, DL_POLY, CHARMM, CPMD etc.)
- Revealed shortcomings in initial Myrinet / Redhat environment

- Presented findings in visit to the University post-acceptance
- Provided application-code base to User Community (subject to appropriate licencing etc.)
## Commodity Systems (CSx) Prototype / Evaluation Hardware

<table>
<thead>
<tr>
<th>Systems</th>
<th>Location</th>
<th>CPUs</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS1</td>
<td>Daresbury</td>
<td>32</td>
<td>PentiumIII / 450 MHz + FE (EPSRC)</td>
</tr>
<tr>
<td>CS2</td>
<td>Daresbury</td>
<td>64</td>
<td>24 X dual UP2000/EV67-667, QSNet Alpha/LINUX cluster, 8 X dual CS20/EV67-833 (“loki”)</td>
</tr>
<tr>
<td>CS3</td>
<td>RAL</td>
<td>16</td>
<td>Athlon K7 850MHz + myrinet</td>
</tr>
<tr>
<td>CS4</td>
<td>Sara</td>
<td>32</td>
<td>Athlon K7 1.2 GHz + FE</td>
</tr>
<tr>
<td>CS6</td>
<td>CLiC</td>
<td>528</td>
<td>PentiumIII / 800 MHz; fast ethernet (Chemnitzer Cluster)</td>
</tr>
<tr>
<td>CS7</td>
<td>Daresbury</td>
<td>64</td>
<td>AMD K7/1000 MP + SCALI/SCI (“ukcp”)</td>
</tr>
<tr>
<td>CS8</td>
<td>NCSA</td>
<td>320</td>
<td>160 dual IBM Itanium/800 + Myrinet 2k (“titan”)</td>
</tr>
<tr>
<td>CS9</td>
<td>Bristol</td>
<td>96</td>
<td>Pentium4 Xeon/2000 + Myrinet 2k (“dirac”)</td>
</tr>
</tbody>
</table>

### Prototype Systems

<table>
<thead>
<tr>
<th>Systems</th>
<th>Location</th>
<th>CPUs</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS0</td>
<td>Daresbury</td>
<td>10</td>
<td>10 CPUS, Pentium II/266</td>
</tr>
<tr>
<td>CS5</td>
<td>Daresbury</td>
<td>16</td>
<td>8 X dual Pentium III/933, SCALI</td>
</tr>
</tbody>
</table>
## Commodity Systems (CSx) II.

<table>
<thead>
<tr>
<th>Systems</th>
<th>Location</th>
<th>CPUs</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS10</td>
<td><em>Hull</em></td>
<td>64</td>
<td>Pentium4 Xeon/2667 + Myrinet 2k (“eagle”), Streamline/SCORE</td>
</tr>
<tr>
<td>CS11</td>
<td><em>Workstations</em></td>
<td>32</td>
<td>Pentium4 Xeon/2667 + GbitEther, ScaMPI</td>
</tr>
<tr>
<td>CS12</td>
<td><em>Essex</em></td>
<td>48</td>
<td>Pentium4 Xeon/2400 + GbitEther (“sstream1”), Streamline/SCORE</td>
</tr>
<tr>
<td>CS13</td>
<td><em>White Rose, Leeds</em></td>
<td>256</td>
<td>Pentium4 Xeon/2200-2400 + M2k (“snowdon”), Streamline/SCORE</td>
</tr>
<tr>
<td>CS14</td>
<td><em>NCSA</em></td>
<td>1024</td>
<td>Pentium III Xeon/1000 + M2k (“platinum”)</td>
</tr>
<tr>
<td>CS15</td>
<td><em>SDSC</em></td>
<td>128</td>
<td>Pentium III Xeon/ 800 + M2k (“meteor”)</td>
</tr>
<tr>
<td>CS16</td>
<td><em>SDSC</em></td>
<td>256</td>
<td>dual-Itanium2/1.3 GHz + M2k (“Teragrid”)</td>
</tr>
<tr>
<td>CS17</td>
<td><em>Daresbury</em></td>
<td>32</td>
<td>Pentium4 Xeon/2667 + GbitEther (“ccp1”), Streamline/SCORE</td>
</tr>
<tr>
<td>CS18</td>
<td><em>Bradford</em></td>
<td>78</td>
<td>Pentium4 Xeon/2800 + M2k/GbitE (“grendel”)</td>
</tr>
<tr>
<td>CS19</td>
<td><em>Daresbury</em></td>
<td>64</td>
<td>dual-Opteron/246 2.0 GHz nodes + Infiniband, Gbit and SCI (“scaliwag”)</td>
</tr>
<tr>
<td>CS20</td>
<td><em>RAL</em></td>
<td>256</td>
<td>dual-Opteron/248 2.2 GHz nodes + Myrinet (“scarf”)</td>
</tr>
</tbody>
</table>
High-End Systems Evaluated

- Cray T3E/1200E (… historical …)
  - 816 processor system at Manchester (CSAR), 600 Mz Alpha EV56 CPU, 256 MB
- IBM pseries 690 and pseries 690+ (Daresbury)
  - IBM p690 (8-way LPAR’d nodes, 1280 X 1.3 GHz CPUs with colony, HPCx)
  - IBM p690+ (32-way nodes, 1600 X 1.7 GHz CPUs with HPS, HPCx- Phase2)
- Compaq AlphaServer SC
  - 4-way ES40/667 A21264A (APAC) and 833 MHz SMP nodes (2 GB RAM);
  - TCS1 system at PSC (750 4-way ES45 nodes - 3,000 EV68 CPUs - 4 GB memory per node, 8MB L2 cache), Quadrics interconnect (5 usec latency, 250 MB/sec B/W)
- SGI Origin 3800
  - SARA (1000 CPUs) - Numalink with MIPS R14k/500 CPUs
- SGI Altix 3700
  - Linux Cluster - Numalink with Itanium 2 1.3 GHz CPUs, 3MB L3 cache
    - CSAR (“newton” 512 CPUs) and SARA (“aster” - 416 CPUs - 7 nodes)
    - ORNL (“ram” 256 CPUs with Itanium 2 1.5 GHz CPUs, 6MB L3 cache)
Applications Performance Overview

• Serial (SPEC, DL) & Communication Benchmarks
• Parallel Applications Performance

1. Computational Chemistry: Molecular Simulation & Electronic Structure
2. Computational Materials Science
3. Atomic & Molecular Physics
4. Computational Engineering
5. Environmental Modelling

• Capacity-based group solution
• Issues of Cost effectiveness
• On e.g. 128-256 CPU cluster, modal job size is ~ 32 CPUs
• Increasing trend to hierarchical clusters - Gbit network with HEC core (with e.g. myrinet)

Capability and Capacity Computing Commodity vs. Proprietary Solutions

Attempted to quantify delivered performance from the Commodity-based systems against current MPP (CSAR Cray T3E/1200E) and ASCI-style SMP-node platforms (e.g. SGI Origin 3800) i.e.

Performance Metric (% 32-node Cray T3E)

T (32-nodes Cray T3E/1200E) / T (32 CPUs) CSx

[ T_{32-node} T3E / T_{32-node} CS1 Pentium III/450 + FE ]

T_{32-node} T3E / T_{32-node} CS6 Pentium III/800 + FE

T_{32-node} T3E / T_{32-CPU} CS2 Alpha Linux Cluster + Quadrix

Performance Metrics: 2002

Performance Metric (% 32-node AlphaServer SC [PSC])

T (32-CPU AlphaServer SC ES45/1000) / T (32 CPUs) CSx

T_{32-CPU} AlphaServer ES45 / T_{32-CPU} CS9 Pentium 4 Xeon / 2000 + Myrinet 2k
Commodity Comparisons with High-end Systems

% of 32 CPUs of Cray T3E/1200E

Cluster CS6
Pentium3/800 + Fast Ethernet

ANGUS (144)
CASTEP
CHARMM
DLPOLY_2 (Macromolecular)
DLPOLY-2 (Ionic)
NWChem (DFT Jfit)
GAMESS-UK (HF Forces)
GAMESS-UK (MP2 gradient)
GAMESS-UK (DFT gradient)
GAMESS-UK (DFT-Jfit)
GAMESS-UK (DFT)
GAMESS-UK (SCF)

CS6 - 98% of Cray T3E

2000
Commodity Comparisons with High-end Systems

% of 32 CPUs of Compaq AlphaServer SC ES45/1000

Cluster CS9
Pentium4/2000 Xeon + Myrinet

CS9 - 66% of ES45/1000

- ANGUS (288)
- ANGUS (144)
- CPMD
- CASTEP
- CHARMM
- DLPOLOGY_2 (Macromolecular)
- DLPOLOGY-2 (Ionic)
- GAMESS-UK + CHARMM
- NWChem (DFT Jfit)
- GAMESS-UK (HF Forces)
- GAMESS-UK (MP2 gradient)
- GAMESS-UK (DFT gradient)
- GAMESS-UK (DFT-Jfit)
- GAMESS-UK (DFT)
- GAMESS-UK (SCF)

2002
### SPEC CPU 2000 - SPECfp2000

Values relative to IBM p-series 690/pwr4 1.3 GHz

<table>
<thead>
<tr>
<th>System Description</th>
<th>SPEC CPU 2000</th>
<th>SPECfp2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD A7N8X XP3200+/2.2 GHz</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>AMD SK8V Opteron150/2.4GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dell PW360 / 3.4GHz P4 Extreme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP RX4640 Itanium2/1500</td>
<td>108</td>
<td>133</td>
</tr>
<tr>
<td>HP RX4640 Itanium2/1300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP AlphaServer ES45 68/1250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP AlphaServer GS1280 7/1300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun Blade 2000 / 1200 MHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun Java Wrkst. (Opteron/2.4 GHz)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fujitsu Celsius M420 (3.4 GHz P4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGI Origin 3200/R14k-600</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>SGI Altix 3000 Itanium 2/1300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGI Altix 3000 Itanium 2/1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP 9000/C3750-875</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM eServer p5 570/1900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM p-series 690/1700</td>
<td>100</td>
<td>134</td>
</tr>
<tr>
<td>IBM p-series 690/1300</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Values relative to IBM p-series 690/pwr4 1.3 GHz*
The GAMESS-UK Serial Benchmark

Performance relative to the IBM p-series 690/pwr4 1.3 GHz

- AMD MP2400+ / 2000: 72
- Pentium 4 Xeon / 3066: 107
- SGI Altix3700 Itanium2/1300: 107
- HP RX4640 Itanium2/1300-H: 117
- HP RX2600 Itanium2/1500-L: 128
- HP RX5670 Itanium2/1500-H: 127
- Intel Tiger Itanium2/1500: 127
- AMD Opteron 848/ 2200: 133
- AMD Opteron 244/ 1800: 113
- Compaq Alpha ES45/1250: 79
- Compaq Marvel EV7 /1000: 73
- SUN Blade 2000 / 1056 Cu: 54
- SUN FireV880 / 900 Cu: 45
- SGI Origin3800/R14k-600: 40
- HP PA-9000/RP7410-875: 77
- IBM p-series 690/1700: 137
- IBM p-series 690/1300: 100

2.3 minutes
Interconnects and Networking

- Ethernet, Gbit etc.
- Myrinet, Quadrics and Dolphin SCI
  - 64 bit PCI implementations, with faster PCI-X options near market
- Infiniband PCI-X interconnect

Performance of current interconnects.

<table>
<thead>
<tr>
<th>Interconnect</th>
<th>Latency (µs)</th>
<th>Bandwidth (MB/s)</th>
<th>Switch size</th>
<th>Message size (kB)</th>
<th>1MB transmit (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadrics Elan3</td>
<td>5</td>
<td>325 (680)</td>
<td>128-2000</td>
<td>1.6</td>
<td>3.08</td>
</tr>
<tr>
<td>Dolphin SCI</td>
<td>5 (1.5 intranode)</td>
<td>326</td>
<td>N/A</td>
<td>1.6</td>
<td>3.08</td>
</tr>
<tr>
<td>Myrinet</td>
<td>~8 (6.3)</td>
<td>243 (500)</td>
<td>8-128</td>
<td>2</td>
<td>4.12</td>
</tr>
<tr>
<td>Mellanox Infiniband</td>
<td>7-10 (5.5)</td>
<td>800 (830)</td>
<td>8-96</td>
<td>5.6-8</td>
<td>1.26</td>
</tr>
<tr>
<td>Gbit. Ethernet</td>
<td>30-100</td>
<td>125</td>
<td>64</td>
<td>3.7-12.5</td>
<td>8.1</td>
</tr>
</tbody>
</table>
Collective Operations
(Time - usec) - as function of no.of CPUs

Communications among groups of processors in the cluster e.g.
- Distribute data to all nodes in the cluster (scatter, all-to-all)
- Collect data from all nodes (gather)
- Collect information from all nodes to determine an overall condition e.g. minima or maxima (reduce).
Communications Benchmark
PMB: Pallas MPI Benchmark Suite (V2.2)

High-end Systems
- Cray T3E/1200E
- SGI Origin3800/R14k-500 (Teras)
- IBM SP/WH2-375 ("trailblazer" switch and SP/NH2-375 (single-plane colony)
- IBM p690 (8-way LPAR, HPCx - colony “Phase 1”)
- IBM p690+ (32-way SMP, HPCx - HPS “”Phase2”)
- Compaq AlphaServer SC ES45/1GHz -TCS1
- SGI Altix3700/Itanium2-1.3GHz (Newton) and 1.5 GHz (Ram)

Commodity-based Systems
- CS2 Alpha Linux Cluster dual UP2000/667
- CS8 Itanium/800 + Myrinet 2k (NCSA)
- CS9 dual P4/2000 Xeon + Myrinet 2k
- CS10 P4/2666 + Myrinet 2k (Streamline, Hull U.)
- CS12 P4/2400 + GbitEther (Streamline, Essex U.)
- CS13 P4/2200-2400 + Myrinet 2k (Streamline, Leeds U.)
- CS14 PIII/1000 + Myrinet (VMI, NCSA)
- CS16 Itanium2 1300 + Myrinet2k (SDSC)
- CS18 P4/2800 + Myrinet 2k, GBitE (Clustervision, Bradford U.)
- CS19 Opteron 246/2000 + Infiniband, SCI, GbitEther (OCF, Daresbury).
- CS20 Opteron 248/2200 + Myrinet, (Streamline, RAL).
PingPong Performance

Bandwidth (MByte/sec) vs. Message Length (Bytes)

- Cray T3E/1200E
- CS18 Pentium 4/2800 + Myrinet 2k
- CS15 Itanium2/1300 + Myrinet 2k
- CS10 Pentium 4/2666 + Myrinet 2k
- CS14 Platinum/PIII 1000 + Myrinet
- CS19 Opteron 246/2.0 + SCI
- CS19 Opteron 246/2.0 + IB
- CS12 Pentium 4/2400 + GbitEther
- CS18 Pentium 4/2800 + GbitEther

Commodity Systems
MPI_allreduce Performance

- Cray T3E/1200E
- CS18 Pentium 4/2800 + Myrinet 2k
- CS15 Itanium2/1300 + Myrinet 2k
- CS10 Pentium 4/2666 + Myrinet 2k
- CS14 Platinum/PIII 1000 + Myrinet
- CS19 Opteron 246/2.0 + SCI
- CS19 Opteron 246/2.0 + IB
- CS12 Pentium 4/2400 + GbitEther
- CS18 Pentium 4/2800 + GbitEther

PMB Benchmark (Pallas)

16 CPUs

Commodity Systems

Message Length (Bytes)
Application Codes

Application-driven performance comparisons between Commodity-based systems and both current MPP and ASCI-style SMP-node platforms:

- **Computational Chemistry**
  - **Molecular Simulation**
    - DL_POLY, DLMULTI - parallel MD codes with many applications
    - CHARMM - macromolecular MD and energy minimisation
  - **Molecular Electronic Structure**
    - GAMESS-UK and NWChem - Ab initio Electronic structure codes
- **Materials**
  - CPMD, CASTEP - Car-Parrinello plane wave codes
- **Engineering**
  - ANGUS, SBLI - regular-grid domain decomposition engineering codes

Performance Metric (% 32-node high-end system)
Performance Metrics: 2004

Attempt to quantify delivered performance from current Commodity-based systems against high-end SMP-node platforms: SGI Altix 3700/Itanium2 1.3 GHz (CSAR) and IBM p690+ (HPCx) i.e.

Performance Metric (% 32-node SGI Altix 3700/Itanium 2 1.3 GHz)

\[
\frac{T_{32-CPU \ SGI \ Altix \ 3700}}{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}} \text{ / } \frac{T_{32-CPU \ CS20 \ Opteron}}{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}}
\]

\[
\frac{T_{32-CPU \ SGI \ Altix \ 3700}}{T_{32-CPU \ CS20 \ Opteron \ 248}} \text{ / } \frac{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}}{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}} + \text{ Myrinet 2k}
\]

\[
\frac{T_{32-CPU \ SGI \ Altix \ 3700}}{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}} \text{ / } \frac{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}}{T_{32-CPU \ CS18 \ Pentium \ 4 \ Xeon}} + \text{ Gbit Ethernet}
\]
Molecular Simulation

Molecular Dynamics Codes:
DL_POLY, DLMULTI and CHARMM
DL_POLY: A Parallel Molecular Dynamics Simulation Package

- First major MD code for parallel platforms
  - Developed as CCP5 parallel MD code by W. Smith and T.R. Forester
  - UK + International user community
- 830 licences issued since 1994
- 10 industrial licences since 2000.
- Areas of application:
  - liquids, solutions, spectroscopy, ionic solids, molecular crystals, polymers, glasses, membranes, proteins, metals, solid and liquid interfaces, catalysis, clathrates, liquid crystals, biopolymers, polymer electrolytes.
Boundary Conditions
- None (e.g. isolated macromolecules)
- Cubic periodic boundaries
- Orthorhombic periodic boundaries
- Parallelepiped periodic boundaries
- Truncated octahedral periodic boundaries
- Rhombic dodecahedral periodic boundaries
- Slabs (i.e. x,y periodic, z nonperiodic)

Target Systems
- Atomic systems & mixtures (Ne, Ar, etc.)
- Ionic melts & crystals (NaCl, KCl etc.)
- Polarisable ionics (ZSM-5, MgO etc.)
- Molecular liquids & solids (CCl₄, Bz etc.)
- Molecular ionics (KNO₃, NH₄Cl, H₂O etc.)
- Synthetic polymers ([PhCHCH₂]n etc.)
- Biopolymers and macromolecules
- Polymer electrolytes, Membranes,
- Aqueous solutions, Metals

MD Algorithms/Ensembles
- Verlet leapfrog, Verlet leapfrog + RD-SHAKE
- Rigid units with FIQA and RD-SHAKE
- Linked rigid units with QSHAKE
- Constant T (Berendsen) with Verlet leapfrog and with RD-SHAKE
- Constant T (Evans) with Verlet leapfrog and with RD-SHAKE
- Constant T (Hoover) with Verlet leapfrog

DL_POLY: A Parallel MD Simulation Package
DL_POLY Parallel Benchmarks (Cray T3E/1200)

V2: Replicated Data

4. NaCl; Ewald, 27,000 ions
5. NaK-disilicate glass; 8,640 atoms, MTS+ Ewald
8. MgO microcrystal: 5,416 atoms

9. Model membrane/Valinomycin (MTS, 18,886)
7. Gramicidin in water (SHAKE, 12,390)
6. K/valinomycin in water (SHAKE, AMBER, 3,838)
1. Metallic Al (19,652 atoms, Sutton Chen)
3. Transferrin in Water (neutral groups + SHAKE, 27,593)
2. Peptide in water (neutral groups + SHAKE, 3993).
DL_POLY V2: Bench 4 - Commodity-based Systems

NaCl; 27,000 ions, Ewald, 75 time steps, Cutoff=24Å

Measured Time (seconds)

<table>
<thead>
<tr>
<th>System Configuration</th>
<th>Measured Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS12 P4/2400 + GbitE</td>
<td>115</td>
</tr>
<tr>
<td>CS10 P4/2666 + Myrinet</td>
<td>110</td>
</tr>
<tr>
<td>CS18 P4/2800 + GbitE</td>
<td>106</td>
</tr>
<tr>
<td>CS18 P4/2800 + Myrinet</td>
<td>103</td>
</tr>
<tr>
<td>CS19 Opteron246/2.0 + IB</td>
<td>61</td>
</tr>
<tr>
<td>CS19 Opteron246/2.0 + SCI</td>
<td>60</td>
</tr>
<tr>
<td>CS20 Opteron246/2.2 + Myrinet</td>
<td>50</td>
</tr>
<tr>
<td>CS16 Itanium2/1300 + Myrinet</td>
<td>40</td>
</tr>
<tr>
<td>SGI Altix 3700/Itanium2 1300</td>
<td>36</td>
</tr>
</tbody>
</table>

Number of CPUs

- 16: 61, 60, 50, 40, 36
- 32: 65, 61, 57, 30, 29, 27, 22, 18

T3E128 = 94

30%, 68%
DL_POLY V2: Bench 5 - Commodity-based Systems

NaK-disilicate glass; 8,640 atoms, MTS + Ewald: 270 time steps

Measured Time (seconds)

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Number of CPUs</th>
<th>Measured Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS12 P4/2400 + GbitE</td>
<td>16</td>
<td>74</td>
</tr>
<tr>
<td>CS10 P4/2666 + Myrinet</td>
<td>16</td>
<td>67</td>
</tr>
<tr>
<td>CS18 P4/2800 + GbitE</td>
<td>16</td>
<td>56</td>
</tr>
<tr>
<td>CS18 P4/2800 + Myrinet</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td>CS19 Opteron246/2.0 + SCI</td>
<td>32</td>
<td>44</td>
</tr>
<tr>
<td>CS19 Opteron246/2.0 + IB</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td>CS20 Opteron246/2.2 + Myrinet</td>
<td>16</td>
<td>44</td>
</tr>
<tr>
<td>CS16 Itanium2/1300 + Myrinet</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td>SGI Altix 3700/Itanium2 1300</td>
<td>16</td>
<td>25</td>
</tr>
</tbody>
</table>

T3E_{128} = 75

51%, 94%
DL_POLY V2: Bench 7 - Commodity-based Systems

Gramicidin in water; rigid bonds and SHAKE, 12,390 atoms, 500 time steps

Number of CPUs

Measured Time (seconds)

T3E_{128} = 166

31%, 78%

CS12 P4/2400 + GbitE
CS10 P4/2666 + Myrinet
CS18 P4/2800 + GbitE
CS18 P4/2800 + Myrinet
CS19 Opteron246/2.0 + SCI
CS19 Opteron246/2.0 + IB
CS20 Opteron246/2.2 + Myrinet
SGI Altix 3700/Itanium2 1300
Migration from Replicated to Distributed data: DL_POLY-3 : Domain Decomposition

- Distribute atoms, forces across the nodes
  - More memory efficient, can address much larger cases (10^5-10^7)
- Shake and short-ranges forces require only neighbour communication
  - Communications scale linearly with number of nodes
- Coulombic energy remains global
  - Strategy depends on problem and machine characteristics
  - Adopt Smooth Particle Mesh Ewald scheme
    - Includes Fourier transform smoothed charge density (reciprocal space grid typically 64x64x64 - 128x128x128)

W. Smith and T.R. Forester

http://www.cse.clrc.ac.uk/msi/software/DL_POLY/
Migration from Replicated to Distributed data

DL_POLY-3: Coulomb Energy Evaluation

- Conventional routines (e.g. fftw) assume plane or column distributions
- A global transpose of the data is required to complete the 3D FFT and additional costs are incurred re-organising the data from the natural block domain decomposition.
- An alternative FFT algorithm has been designed to reduce communication costs.
  - the 3D FFT are performed as a series of 1D FFTs, each involving communications only between blocks in a given column
  - More data is transferred, but in far fewer messages
  - Rather than all-to-all, the communications are column-wise only
DL_POLY3 Coulomb Energy Evaluation

Commodity-based Systems

Measured Time (seconds)

Number of CPUs

IPI, University of Bradford  36  26th August 2004
DL_POLY3 Macromolecular Simulations

Commodity-based Systems

Measured Time (seconds)

<table>
<thead>
<tr>
<th>Commodity-based System</th>
<th>Measured Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS12 Pentium 4/2400 + Gbit Ethernet</td>
<td>739</td>
</tr>
<tr>
<td>CS18 P4/2800 + GbitE</td>
<td>716</td>
</tr>
<tr>
<td>CS18 P4/2800 + Myrinet</td>
<td>675</td>
</tr>
<tr>
<td>CS10 Pentium 4/2666 + Myrinet 2k</td>
<td>480</td>
</tr>
<tr>
<td>CS19 Opteron246/2.0 + SCI</td>
<td>435</td>
</tr>
<tr>
<td>CS19 Opteron246/2.0 + IB</td>
<td>410</td>
</tr>
<tr>
<td>CS20 Opteron248/2.2 + Myrinet</td>
<td>378</td>
</tr>
<tr>
<td>SGI Altix 3700/Itanium2 1300</td>
<td>398</td>
</tr>
</tbody>
</table>

Gramicidin in water; rigid bonds + SHAKE: 792,960 ions, 50 time steps

Measuring 52%, 88%
CHARMM

- CHARMM (Chemistry at HARvard Macromolecular Mechanics) is a general purpose molecular mechanics, molecular dynamics and vibrational analysis package for modelling and simulation of the structure and behaviour of macromolecular systems (proteins, nucleic acids, lipids etc.)
- Supports energy minimisation and MD approaches using a classical parameterised force field.
- Parallel Benchmark - MD Calculation of Carboxy Myoglobin (MbCO) with 3830 Water Molecules.
- QM/MM model for study of reacting species
  - incorporate the QM energy as part of the system into the force field
  - coupling between GAMESS-UK (QM) and CHARMM.
Parallel CHARMM Benchmark

Benchmark MD Calculation of Carboxy Myoglobin (MbCO) with 3830 Water Molecules: 14026 atoms, 1000 steps (1 ps), 12-14 Å shift.

Measured Time (seconds)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>104</td>
<td>83</td>
<td>72</td>
<td>64</td>
<td>64</td>
<td>61</td>
<td>62</td>
<td>73</td>
</tr>
<tr>
<td>32</td>
<td>114</td>
<td>89</td>
<td>72</td>
<td>64</td>
<td>64</td>
<td>61</td>
<td>62</td>
<td>73</td>
</tr>
<tr>
<td>64</td>
<td>89</td>
<td>89</td>
<td>72</td>
<td>64</td>
<td>64</td>
<td>61</td>
<td>62</td>
<td>73</td>
</tr>
</tbody>
</table>

T$E_{16} = 665$, T$E_{128} = 106$
Molecular Electronic Structure

Ab initio Electronic structure Codes: NWChem and GAMESS-UK
Pictorial representation of the iterative SCF process in (i) a sequential process, and (ii) a distributed data parallel process: MOAO represents the molecular orbitals, P the density matrix and F the Fock or Hamiltonian matrix.
Exploiting HPC: The PNNL Collaboration

- Long standing collaboration with HPCC group within EMSL
- Tools
  - Global arrays:
    - portable distributed data tool:
      - Physically distributed data
  - NWChem
    - Highly efficient and portable MPP computational chemistry package
    - Distributed Data - Scalable with respect to chemical system size as well as MPP hardware size
  - PeIGS:
    - parallel eigensolver,
    - guaranteed orthogonality of eigenvectors
  - Wide range of parallel functionality essential for HPCx
- Extensible Architecture
  - Object-oriented design
    - abstraction, data hiding, handles, APIs
  - Parallel programming model
    - non-uniform memory access, global arrays
  - Infrastructure
    - GA, Parallel I/O, RTDB, MA, ...
Parallel Linear Algebra

Symmetric Eigensolver Routines

- ScALAPACK
  - drivers for solving standard and generalized dense symmetric or dense Hermitian Eigenproblems.
  - PDSYEV (QR Method) (Scalapack 1.5)
  - PDSYEVX (Bisection & Inverse Iteration) (Scalapack 1.5)
  - PDSYEVD (D&C Method) (Scalapack 1.7)
- PeIGS (PNNL)
  - General symmetric and standard symmetric eigenproblems
    - PDSPEV (Bisection & Inverse Iteration)
- BFG (I. Bush)
  - Block Jacobi Method on full dense symmetric matrix (+ Hermitian)
- Plapack
  - QR method
  - MRRR ‘Multiple Relatively Robust Representations’
GAMESS-UK

GAMESS-UK is the general purpose ab initio molecular electronic structure program for performing SCF-, MCSCF- and DFT-gradient calculations, together with a variety of techniques for post Hartree Fock calculations.

- Derived from the original GAMESS code, obtained from Michel Dupuis in 1981 (then at the NRCC), and has been extensively modified and enhanced over the past two decades.
- This work has included contributions from numerous authors†, and has been conducted largely at the CCLRC Daresbury Laboratory, under the auspices of the UK’s Collaborative Computational Project No. 1 (CCP1). Other major sources that have assisted in the on-going development and support of the program include various academic funding agencies in the Netherlands, and ICI plc.

Additional information on the code may be found from links at:
http://www.dl.ac.uk/CFS

GAMESS-UK features 1.

- Hartree Fock:
  - Segmented/ GC + spherical harmonic basis sets
  - SCF-Energies and Gradients: conventional, in-core, direct
  - SCF-Frequencies: numerical and analytic 2nd derivatives
  - Restricted, unrestricted open shell SCF and GVB.

- Density Functional Theory
  - Energies + gradients, conventional and direct including Dunlap fit
  - B3LYP, BLYP, BP86, B97, HCTH, B97-1, FT97 & LDA functionals
  - Numerical 2nd derivatives (analytic implementation in testing)

- Electron Correlation:
  - MP2 energies, gradients and frequencies, Multi-reference MP2, MP3 Energies
  - MCSCF and CASSCF Energies, gradients and numerical 2nd derivatives
  - MR-DCI Energies, properties and transition moments (semi-direct module)
  - CCSD and CCSD(T) Energies
  - RPA (direct) and MCLR excitation energies / oscillator strengths, RPA gradients
  - Full-CI Energies
  - Green’s functions calculations of IPs.
  - Valence bond (Turtle)
GAMESS-UK features 2.

- **Molecular Properties:**
  - Mulliken and Lowdin population analysis, Electrostatic Potential-Derived Charges
  - Distributed Multipole Analysis, Morokuma Analysis, Multipole Moments
  - Natural Bond Orbital (NBO) + Bader Analysis
  - IR and Raman Intensities, Polarizabilities & Hyperpolarizabilities
  - Solvation and Embedding Effects (DRF)
  - Relativistic Effects (ZORA)

- **Pseudopotentials:**
  - Local and non-local ECPs.

- **Visualisation:** tools include CCP1 GUI
- **Hybrid QM/MM (ChemShell + CHARMM QM/MM)**
- **Semi-empirical**: MNDO, AM1, and PM3 hamiltonians

- **Parallel Capabilities:**
  - MPP and SMP implementations (GA tools)
  - SCF/DFT energies, gradients, frequencies
  - MP2 energies and gradients
  - Direct RPA
Parallel Implementation of GAMESS-UK

- Early implementation based on message passing
- Subsequent activities under HEC Facilities Agreement with support from European projects
  - IMMP (1994-1997, part of EUROPORT)
    - Partners: Guest, Sherwood (Daresbury) - GAMESS-UK, Baerends (Amsterdam) - ADF, Clark (Erlangen) - VAMP
    - Focus on MPP systems (e.g. T3E)
    - Mapping of disk files into global memory (uses GAs)
    - First MPP MP2 algorithm
      - GA storage of transformed integrals
  - QUASI (1998-2001)
    - Application of QM/MM methods in Industry
    - Led by Daresbury, Partners: Catlow (RI), Thiel (MPI), BASF, ICI, Hydro
    - Focus on commodity systems, cost-effective computing in industry
      - demonstrated using Linux alpha commodity cluster at Daresbury.
Parallel Implementations of GAMESS-UK

- Extensive use of Global Array (GA) Tools and Parallel Linear Algebra from NWChem Project (EMSL)
- SCF and DFT energies and gradients
  - Replicated data, but …
  - GA Tools for caching of I/O for restart and checkpoint files
  - Storage of 3-centre 2-e integrals in DFT Jfit
  - Linear Algebra (via PeIGs, DIIS/MMOs, Inversion of 2c-2e matrix)
- SCF and DFT second derivatives
  - Distribution of $<\text{vvoo}>$ and $<\text{vovo}>$ integrals via GAs
- MP2 gradients
  - Distribution of $<\text{vvoo}>$ and $<\text{vovo}>$ integrals via GAs
- Efficiency dependent on optimised GA implementation
  - e.g. SCALI, GM vs. PM Myrinet port (CS7, CS12 and CS14)
Parallel Linux Implementations of GAMESS-UK

- Issues arising from the use of Global Array (GA) Tools
- Efficiency dependent on optimised GA implementation
  - e.g. SCALI, GM vs. PM Myrinet port (CS7, CS12 and CS14)
- SHMMAX
  - The Linux kernel has traditionally fairly small limit for the shared memory segment size (SHMMAX). In kernels 2.2.x it is 32MB on Intel
  - There are two ways to increase this limit:
    - rebuild the kernel after changing SHMMAX in `/usr/src/linux/include/asm-i386/shmparam.h`, for example, setting SHMMAX as 0x8000000 (128MB)
    - a system admin can increase the limit without rebuilding the kernel, e.g.
      - `echo "134217728" >/proc/sys/kernel/shmmax`
- More info on this subject:
  - and on the Linux SMP discussion list.

On “grendel” - `/proc/sys/kernel/shmmax` : 33554432
GAMES-UK ΔSCF Performance
IBM SP/p690, High-end and Commodity-based Systems

Elapsed Time (seconds)

Impact of Serial Linear Algebra:
\[ T_{CS20-Opt248}^{(16)} = 1306 \text{ [231]} \]
\[ T_{CS20-Opt248}^{(32)} = 1270 \text{ [160]} \]
GAMESS-UK. DFT B3LYP Performance
IBM SP/p690+, High-end and Commodity-based Systems

Elapsed Time (seconds)

Basis: 6-31G

Cyclosporin, 1000 GTOs
DFT HCTH on Valinomycin. Impact of Coulomb Fitting

Basis: DZV_A2 (Dgauss)
A1_DFT Fit: 882/3012

Measured Time (seconds)

Number of CPUs

$J_{\text{EXPLICIT}}$ 100%, 136%

$J_{\text{FIT}}$ 55%, 74%

IPI, University of Bradford 52 26th August 2004
DFT HCTH Performance

Elapsed Time (seconds)

- CS18 P4/2800 + GbitE
- CS20 Opteron 248/2.2 + Myrinet
- SGI Origin 3800/R14k-500
- IBM p690
- AlphaServer ES45/1000
- SGI Altix 3700 Itanium2/1300
- IBM p690+ // HPS (SP9)

Valinomycin, 1620 GTOs (DZVP2_A2 (Dgauss))

Commodity-based Systems

IPI, University of Bradford 53 26th August 2004
DFT BLYP Gradient:
High-end and Commodity-based Systems

Geometry optimisation of polymerisation catalyst
Cl(C_3H_5O)_2Pd[(P(CMe_3)_2)_2C_6H_4]
MP2 Gradient Algorithms

Serial

- Conventional
  - Integrals written to disk
  - Read back, transformed, written out, resorted etc.
  - Heavy I/O demands

- Direct/Semi-direct (Frisch, Head-Gordon & Pople, Hasse and Ahlrichs)
  - Replace all/some I/O with batched integral recomputation

Parallel

- Poor I/O-to-compute performance of MPPs
  - Direct approach

- Current MPPs have large global memories

- Store subset of MO integrals
  - Reduce number of integral recomputations
  - Increase communication overhead

- Subset includes VOVO, VVOO, VOOO,
  - VVVO-class too large to store
  - Compute VVVO-terms in separate step

IPI, University of Bradford 55 26th August 2004
Performance of MP2 Gradient Module
High-end and Commodity-based Systems

Elaborated Time (seconds)

- Mn(CO)$_5$H - MP2 geometry optimisation
  BASIS: TZVP + f (217 GTOs)

Latency sensitive - optimised GAs

45%, 69%
SCF and DFT Analytic 2nd Derivatives

**SCF**
- Solve DFT equations

**Derivative Fock**
- Compute 1st derivatives of DFT matrices

**CPHF**
- Solve systems of linear equations for 1st derivatives of DFT wavefunction

**2nd Derivative**
- Compute 2nd derivatives of DFT functionals

**Hessian**
- Sum partial derivatives

- Distribution of $\langle vvoo \rangle$ and $\langle vovo \rangle$ integrals via GAs
- All DFT contributions calculated on the fly
- Extensive use of screening
- Evaluation in AO or MO basis available
- CPHF arrays replicated but all systems of equations can be solved separately to save memory
- DFT terms in CPHF equations can be calculated at lower accuracy
SCF Analytic 2nd Derivatives Performance
IBM SP/p690+, High-end and Commodity-based Systems

Elapsed Time (seconds)

(C₆H₄(CF₃))₂: Basis 6-31G (196 GTO)

Terms from MO 2e-integrals in GA storage (CPHF & pert. Fock matrices); Calculation dominated by CPHF
DFT Analytic 2nd Derivatives Performance
Commodity-based Systems - HCTH functional

(C₆H₄(CF₃))₂: Basis 6-31G (196 GTO)

Terms from MO 2e-integrals in GA storage (CPHF & pert. Fock matrices);
Calculation dominated by CPHF:
Coupled QM/MM Calculations

- To extend application of *ab-initio* techniques to complex systems
  - Treat reacting core by Quantum Mechanics
    - High accuracy
    - High cost
  - Model environment by Molecular Mechanics
    - Parameterised force field
    - Classical electrostatics
  - QM-MM Junction
    - Link atoms (e.g. H)
    - Mechanical coupling
    - Polarisation of QM region
GAMESS-UK Interface with CHARMM

- Implemented in collaboration with Bernie Brooks, Eric Billings, (NIH, Bethesda Maryland)
- Functionality:
  - Similar to existing ab-initio interfaces; CHARMM side follows coupling to GAMESS(US) (Milan Hodoscek)
  - Support for Gaussian delocalised point charges implemented in GAMESS-UK, based on 2- and 3- centre integral and derivative integral drivers from the DFT module.
- Availability:
  - CHARMM-capable code incorporated into GAMESS-UK Version 6.2.
  - CHARMM (implemented in c26b2 onwards) requires independent licencing from Martin Karplus.
- Ported to a wide variety of systems including MPPs
  - Origin (Green), Alphaserver SC (PSC), Beowulfs
QM/MM Applications


Triosephosphate isomerase (TIM)
- Central reaction in glycolysis, catalytic interconversion of DHAP to GAP
- Demonstration case within QUASI (Partners UZH, and BASF)

- QM region 35 atoms (DFT BLYP)
  - include residues with possible proton donor/acceptor roles
  - GAMESS-UK, MNDO, TURBOMOLE
- MM region (4,180 atoms + 2 link)
  - CHARMM force-field, implemented in CHARMM, DL_POLY

Measured Time (seconds)

- IBM SP/p690 = 134 secs
- CS9 P4/2000 + Myrinet 2k
- SGI Origin3800/R14k-500
- AlphaServer SC ES45/1000
- IBM SP/p690

High-End Computational Chemistry
The NWChem Software

- **Capabilities (Direct, Semi-direct and conventional):**
  - RHF, UHF, ROHF using up to 10,000 basis functions; analytic 1st and 2nd derivatives.
  - DFT with a wide variety of local and non-local XC potentials, using up to 10,000 basis functions; analytic 1st and 2nd derivatives.
  - CASSCF; analytic 1st and numerical 2nd derivatives.
  - Semi-direct and RI-based MP2 calculations for RHF and UHF wave functions using up to 3,000 basis functions; analytic 1st derivatives and numerical 2nd derivatives.
  - Coupled cluster, CCSD and CCSD(T) using up to 3,000 basis functions; numerical 1st and 2nd derivatives of the CC energy.
  - Classical molecular dynamics and free energy simulations with the forces obtainable from a variety of sources.
Case Studies - Zeolite Fragments

- **Si$_8$O$_7$H$_{18}$**
  - 347/832
  - DFT Calculations with Coulomb Fitting
  - Basis (Godbout et al.)
    - DZVP - O, Si
    - DZVP2 - H
  - Fitting Basis:
    - DGAUSS-A1 - O, Si
    - DGAUSS-A2 - H

- **Si$_8$O$_{25}$H$_{18}$**
  - 617/1444

- **Si$_{26}$O$_{37}$H$_{36}$**
  - 1199/2818

- **Si$_{28}$O$_{67}$H$_{30}$**
  - 1687/3928

• NWChem & GAMESS-UK
  
Both codes use auxiliary fitting basis for coulomb energy, with 3 centre 2 electron integrals held in core.
DFT Coulomb Fit - NWChem 4.6

**Measured Time (seconds)**

- $\text{Si}_{26}\text{O}_{37}\text{H}_{36}$
- 1199/2818

<table>
<thead>
<tr>
<th>Number of CPUs</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS18 P4/2800 + GbitEther</td>
<td>1595</td>
<td>1279</td>
</tr>
<tr>
<td>CS18 P4/2800 + Myrnet</td>
<td>1314</td>
<td>866</td>
</tr>
<tr>
<td>CS20 Opteron 248/2.2 + Myrnet</td>
<td>741</td>
<td>741</td>
</tr>
<tr>
<td>AlphaServer SC ES45/1000</td>
<td>1193</td>
<td>785</td>
</tr>
<tr>
<td>SGI Altix 3700/1300</td>
<td>761</td>
<td>481</td>
</tr>
<tr>
<td>IBM p690+ HPS - SP9</td>
<td>404</td>
<td>404</td>
</tr>
</tbody>
</table>

39%, 106%
DFT Coulomb Fit - NWChem 4.6

Si$_{28}$O$_{67}$H$_{30}$  1687/3928

Measured Time (seconds)

- CS18 P4/2800 + GbitEther
- CS18 P4/2800 + Myrinet
- CS20 Opteron 248/2.2 + Myrinet
- AlphaServer SC ES45/1000
- SGI Altix 3700/1300
- IBM p690+ HPS - SP9

Number of CPUs

<table>
<thead>
<tr>
<th>Number of CPUs</th>
<th>32</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS18 P4/2800 + GbitEther</td>
<td>2055</td>
<td>1532</td>
</tr>
<tr>
<td>CS18 P4/2800 + Myrinet</td>
<td>1519</td>
<td>972</td>
</tr>
<tr>
<td>CS20 Opteron 248/2.2 + Myrinet</td>
<td>1519</td>
<td>829</td>
</tr>
<tr>
<td>AlphaServer SC ES45/1000</td>
<td>603</td>
<td>558</td>
</tr>
<tr>
<td>SGI Altix 3700/1300</td>
<td>1015</td>
<td>553</td>
</tr>
<tr>
<td>IBM p690+ HPS - SP9</td>
<td>876</td>
<td>-</td>
</tr>
</tbody>
</table>

46%, 109%
Commodity Comparisons with High-end Systems

% of 32 CPUs of SGI Altix 3700 / 1300

Cluster CS18
Pentium4/2800 Xeon + Gbit Ethernet

CS18 - 49% of Altix
Commodity Comparisons with High-end Systems

% of 32 CPUs of SGI Altix 3700 / 1300

Cluster CS20
AMD Opteron 248/2.2 + Myrinet

CS20 - 100% of Altix
Summary

- Commodity-based and High-end Systems
  - Commodity Systems under evaluation; CS1 – CS20
  - High-end systems from SGI, IBM and HP/Compaq
  - Serial CPU and Communications Performance; Evaluation Metrics

- Chemistry Application performance
  - Molecular Simulation (replicated & distributed data)
    - DL_POLY, DLMULTI and CHARMM
  - Electronic Structure
    - NWChem and GAMESS-UK

- Notable Performance of Opteron clusters with Myrinet:
  - CS20 Opteron248/2200 cluster with Myrinet delivers on average 100% of SGI Altix 3700/1300 (“newton”).
  - Cost effectiveness of the Bradford IPI Xeon / 2600 Cluster (CS18) with Gbit Ethernet - delivers 49% of the SGI Altix 3700/1300.
#!/bin/bash
#PBS -N GAMESS-UK
#PBS -j oe
#PBS -l nodes=16:ppn=2
#PBS -l walltime=3:00:00
# Application name:
APP=/home/martyn/GAMESS-UK/bin/gamess-uk
FLAGS=''
# Work directory
RUNDIR=/home/martyn/GAMESS-UK/examples/cluster
# Module environment to be loaded
MODULE_ENV=default-ethernet
cd $RUNDIR
mkdir $PBS_JOBID
cd $PBS_JOBID
cp ../par_41.in datain
# Load modules
if [ -x /etc/custom/profile.custom ]; then
  . /etc/custom/profile.custom
fi
module purge
module load $MODULE_ENV
module load mpiexec/0.75
export OMP_NUM_THREADS=1
env
mpiexec -verbose -comm mpich-p4 -mpich-p4-no-shmem $APP $FLAGS

csub -q eth qsub.clustervision.ethernet

diring to examine running jobs - 
qpeek 4317 | tail -50

csub.clustervision.ethernet

qsub.clustervision.ethernet
Shared Memory Segments on Grendel

rsh node01 ./shmem_check

#!/bin/sh
echo `hostname` ===============

ipcs -m
ids=`ipcs -m | awk '{ print $2 }'`
for id in $ids;do
    if [ "$id" != "Shared" -a "$id" != "shmid" ]; then
        echo "$id"
        #ipcrm shm $id
    fi
done
exit 0
Useful Scripts on Grendel

Show all Active processes

```
#!/bin/csh
set nodes="node01 node02 node03 node04 node05 node06 node07 node08 node09 node10 node11 node12 node13 node14 node15 node16 node17 node18 node19 node20 node21 node22 node23 node24 node25 node26 node27 node28 node29 node30 node31 node32 node33 node34 node35 node36 node37 node38 node39"
foreach node  ($nodes )
    echo $node
    rsh $node ps -ef
end
```

or ... rsh $node ps -flu martyn

Cancel all Users processes

```
#!/bin/csh
set nodes=" ... Node List above ..."
foreach node  ($nodes )
    set list=`rsh $node ps -uh | grep -v tcsh | awk '{print $2}'`
    echo "$node $list"
    rsh $node kill -9 $list
end
```