World Wide Applications of the IBM Blue Gene Supercomputer

Dr. George Chiu
IEEE Fellow
IBM T.J. Watson Research Center
Yorktown Heights, NY
## Overall Efficiencies of BG Applications - Major Scientific Advances

1. **Qbox (DFT) LLNL:** 56.5%; 2006 Gordon-Bell Award 128Ki L, 64Ki P
   - CPMD IBM: 30% highest scaling 128Ki L
   - MGDC highest scaling 128Ki P
2. **ddcMD (Classical MD) LLNL:** 27.6%; 2005 Gordon-Bell Award 128Ki L
   - New ddcMD LLNL: 17.4%; 2007 Gordon-Bell Award 208Ki L
   - MDCASK LLNL, SPaSM LANL: highest scaling 128Ki L
   - LAMMPS SNL: highest scaling 128Ki L, PRXFF, GMD: highest scaling 128Ki L
   - Rosetta UW: highest scaling 20Ki L
   - AMBER 8Ki L
3. **Quantum Chromodynamics CPS:** 30%; 2006 GB Special Award 128Ki L, P
   - MILC, Chroma 128Ki P
4. **sPPM (CFD) LLNL:** 18%; highest scaling 128Ki L
   - Miranda, Raptor LLNL: highest scaling 128Ki L
   - DNS3D highest scaling 128Ki P
   - NEK5 (Thermal Hydraulics) ANL: 22%
   - HYPO4D, PLB (Lattice Boltzmann) highest scaling 128Ki P
5. **ParaDis (dislocation dynamics) LLNL:** highest scaling 128Ki L
6. **WRF (Weather) NCAR:** 10%; highest scaling 128Ki L
   - POP (Oceanography): highest scaling 16Ki L
   - HOMME (Climate) NCAR: 12%; highest scaling 32Ki L, 96Ki P
7. **GTC (Plasma Physics) PPPL:** 7%; highest scaling 128Ki L, 160Ki P
   - FLASH (Supernova Ia) highest scaling 128Ki L, 160Ki P
   - Cactus (General Relativity) highest scaling 32Ki L, 128Ki P
8. **DOCK5, DOCK6** highest scaling 128Ki P
9. **Argonne v18 Nuclear Potential, GFMC 16%** 2010 Bonner Prize 128Ki P
10. **“Cat” Brain** 2009 GB Special Award 144Ki P
FLASH Center Goals

- To simulate matter accretion onto the surfaces of compact stars, nuclear ignition of the accumulated (and possibly stellar) material, and the subsequent evolution of the star’s interior, surface, and exterior
  - Novae (on white dwarf surfaces)
  - Type Ia supernovae (in white dwarf interiors)
  - X-ray bursts (on neutron star surfaces)
FLASH Simulation Movie
FLASH on BG/L and BG/P
Katherine Riley, Andrew Siegel (U Chicago, ANL)

**Sod Discontinuity Scaling Test - Total Time**

- **QSC**: 256node x 4way HP Alpha, LANL
- **Seaborg**: IBM SP, 1.5GF/node NERSC
- **Jazz**: 350node, 2.4GHz Xeon, ANL
- **MCR**: 1152node, 2.4GHz Xeon, LLNL
- **Big BGL**: 16 Racks, coprocessor, 440
- **BG/P**: from 65,536 to 163,840 cores
Large scale vortex particle Direct Numerical Simulation (DNS) of aircraft wakes

In this work wake instabilities are studied through:

- A hybrid mesh-particle vortex method (1) built on top of the ppm library(2)
- An implementation on the massively parallel architecture IBM Blue Gene
- An effective scalability of the scheme that enables unprecedented billion particle simulations (up to 7 billions 16K nodes).

Philippe Chatelain*, Alessandro Curioni**, Michael Bergdorf*, Diego Rossinelli*, Wanda Andreoni**, Petros Koumoutsakos*
*Computational Science - ETH Zurich, Switzerland
**Computational Sciences - IBM Research Zurich, Switzerland
Gas Turbine Engine Simulations from CERFACS, France on Argonne BG/P

- 90% efficiency to 4096 cores out-of-box,
- 95% efficiency cache tweaking cache
AVBP Movie
HOMME and CCSM on Argonne BG/P

- Aqua planet experiment runs
  - Full physics, no land model
- Excellent performance to scale between L and P
  - P runs in VN (quad) mode
  - L runs in CO mode
- BG/P, at 0.50 degree, achieves an integration rate of over 20 Simulated Year Per Day
CCSM Development on Blue Gene/P

Science

• CCSM is a climate simulation code used by the DOE and NSF climate change experiments

• Ultra-high resolution atmosphere simulations on Intrepid
  CAM-HOMME atmospheric component
  1/8th degree (12.km avg grid) coupled with land model at 1/4th degree and ocean/ice
  Testing up to 56K cores, 0.5 simulated years per day with full I/O

• ½ degree finite-volume CAM with tropospheric chemistry and 399 tracers

• New Code Developments
  CAM performance doubled with threading and parallel decomposition of advection by species
  CCSM4 nearly complete with the carbon cycle, long-term integrations and decadal forecasts
  Parallel I/O using PIO and PNETCDF getting 700 MB/s

• 37 publications, presentations and posters

Figure. Value of an Impulse Boundary Propagator tracer 30 years after its introduction at the surface from the 1/10º POP simulation. The image plane descends from the surface in the north to the abyss in the south.

• Participated in the parallel I/O development
• Assisted with large-scale CAM-HOMME runs
• Ongoing work on OMP on BG/P
Weather Modeling - WRF

WRF: Weather Research and Forecasting Model
Used for both research and operational forecasting
Scaled to 128K cores on BG/P
POP: Parallel Ocean Program (NCAR)

- Simulates Global Eddies in World’s Oceans.
- Major component in climate models
- Goal is to “hindcast” 55 year ocean behavior from 1955 – present.
- Originally run at LLNL, but showed poor performance.
- After development effort from NCAR, scaling problems resolved, and tested during BGW day.
- On 16 Racks, 55 year hindcast can be completed in 8 days!

- One short term possibility – RFP for Tsunami Early Warning Systems. RFP Discussions beginning
  - Dr. J. Sairamesh (Ramesh),

2009/12/14
Classical MD – ddcMD
2005 Gordon Bell Prize Winner!!

- Scalable, general purpose code for performing classical molecular dynamics (MD) simulations using highly accurate MGPT potentials

- MGPT semi-empirical potentials, based on a rigorous expansion of many body terms in the total energy, are needed in to quantitatively investigate dynamic behavior of d-shell and f-shell metals.

524 million atom simulations on 64K nodes achieved 101.5 TF/s sustained. Superb strong and weak scaling for full machine - (“very impressive machine” says PI)

Visualization of important scientific findings already achieved on BG/L: Molten Ta at 5000K demonstrates solidification during isothermal compression to 250 GPa
Performance of ddcMD on BlueGene/L
Weak scaling: MGPT Uranium and Tantalum
Rapid Resolidification in Tantalum: Excellent scaling of ddcMD on BG/L

- Nucleation is initiated at multiple independent sites in each sample cell
- Growth of solid grains initiates independently, but soon leads to grain boundaries which span the simulation cell
- 101.5 sustained teraflop/s achieved.

Lawrence Livermore National Laboratory
Blue Gene/L Simulation Results Using ddcMD Code

The ddcMD team is currently using 131,072 CPUs of BG/L for unprecedented five hundred million atom MGPT simulations
Scaling ddcMD up to 131,072 CPUs

... but allows unprecedented scaling of size or time

- Weak scaling is virtually flat across the entire machine - enables simulation of tens of billions of atoms (roughly a cubic micron of material)
- Strong scaling shows speedup down to 8 atoms/CPU - enables simulations involving millions of steps (typically ns of simulated time)
Performance measured using hardware counters

Scaling behavior
- weak scaling almost linear
- strong scaling deviates due to small system size

Sustained Performance : 103.9 TFlop/s on 104 racks
Benchmark Performance : 115.1 TFlop/s
Update Rate : 20 billion atoms/s
I/O Rate : 14 GB/s → 2.7 TB in 3 min
Parity Recovery : 79 times → 2 CPU millennia saved
Simulating the Kelvin-Helmholtz Instability

- Simulating KH instabilities via quasi 2D MD
  - micron length scales in two dimensions with a nanometer length scale in the third dimension
  - equal parts molten copper and aluminum separated by a planar interface
  - a shear was applied with a velocity difference of 2000 m/s

- 9 billion atoms simulated for over 1.4 nanoseconds
  - this represents a total of 6.4 quadrillion \((10^{15})\) position and velocity updates requiring 36 quintillion \((10^{18})\) flops
- Currently simulating 62.5 billion atoms in a fully 3D geometry – this represents a cubic micron of material
Modeling hydrodynamics at the atomic scale

- BlueGene/L upgrade enables unprecedented 12.0 µm x 6.0 µm quasi-2D simulation of 9 billion atoms developing a Kelvin-Helmholtz instability
- Currently performing tour-de-force simulation of 1 cubic micron of sheared molten metal (62.5 Billion atoms) - first ever in 3D
- Atomistic simulation paves the way for creation of fully-resolved sub-zone model
Qbox: First Principles Molecular Dynamics
Francois Gygi UCD, Erik Draeger, Martin Schulz, Bronis de Supinski, LLNL
Franz Franchetti Carnegie mellon, John Gunnels, Vernon Austel, Jim Sexton, IBM

- Treats electrons quantum mechanically
- Treats nuclii classically
- Developed at LLNL
- BG Supported provided by IBM
- Simulated 1,000 Mo atoms with 12,000 electrons
- Achieves 207.3 Teraflops sustained.
  - (56.8% of peak).

Qbox simulation of the transition from a molecular solid (top) to a quantum liquid (bottom) that is expected to occur in hydrogen under high pressure.
## Qbox

<table>
<thead>
<tr>
<th></th>
<th># calls</th>
<th>avg. msg size (bytes)</th>
<th>time (s)</th>
<th>% MPI</th>
<th>% total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI_Bcast</td>
<td>4048</td>
<td>950681.5</td>
<td>401.0</td>
<td>70.2%</td>
<td>14.5%</td>
</tr>
<tr>
<td>MPI_Allreduce</td>
<td>54256</td>
<td>35446.1</td>
<td>103.6</td>
<td>18.1%</td>
<td>3.7%</td>
</tr>
<tr>
<td>MPI_Alltoallv</td>
<td>5635</td>
<td>399.1</td>
<td>41.8</td>
<td>7.3%</td>
<td>1.5%</td>
</tr>
<tr>
<td>MPI_Reduce</td>
<td>940</td>
<td>186007.1</td>
<td>15.2</td>
<td>2.7%</td>
<td>0.5%</td>
</tr>
<tr>
<td>MPI_Barrier</td>
<td>2352</td>
<td>0.0</td>
<td>8.8</td>
<td>1.5%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

![Graph showing performance metrics](image)
Computational Nuclear Structure

- Green’s Function Monte Carlo (GFMC)
  - Ab initio calculation of properties of light nuclei
  - Common benchmark for other methods
- Calculations of $^{12}\text{C}$ with complete Hamiltonian
  - INCITE time, 32K cores
  - Best converged ab initio calculations of $^{12}\text{C}$ ever
  - Key calculation to launch into study of effects of different terms in the nuclear interaction and compute excited states

ADLB
- SciDAC funded ANL-MCS developed multipurpose library for distributing load
- Required for $^{12}\text{C}$ simulation
- OpenMP and MPI programming model

Outstanding paper award
Quantum ChromoDynamics on BG/L

• June 11, 2004: QCD ran for 8 hours 40 min continuously at ~19% efficiency (24% flop rate) on one rack. This gives 1 Tflops sustained for that period (5.7Tflops * 0.19 = 1.06 Tflops) - a first for QCD.
  – A dynamical Wilson fermion full QCD with hybrid Monte Carlo evolution with an even/odd preconditioned matrix operator kernel on a 4x4x4x16 local lattice.
  – One rack of BG/L is 8x8x16x2 so the global lattice was 32x32x64x32. This is a large dynamical lattice.
  – The run did 50 evolutions in that period.
  – Used Virtual Node Mode, both FPU pipelines are used.

• November SC2006: QCD ran on the 64-rack Blue Gene/L at LLNL. We simulated QCD with a 70.5 Tflops (19.3%) sustained performance! 2006 Gordon Bell Special Prize winner
  – A dynamical Wilson fermion full QCD with hybrid Monte Carlo evolution with an even/odd preconditioned matrix operator kernel on a 4x4x4x16 local lattice.
  – The global lattice was 128x128x128x256x32. This is a large finite temperature dynamical lattice. It has been a QCD dream for the last 30 years.
  – Used Virtual Node Mode, both FPU pipelines are used.

• A Wilson fermion full QCD with a non-preconditioned matrix operator kernel achieved 29.23% with KEK code.
Gordon Bell Special Achievement Award 2006, P. Vranas et. al.
QCD Speedup on BG/L to 70.5 sustained Teraflops.

- Dirac Operator = 19.3%
- CG inverter = 18.7%
**Lattice QCD**

**Science**

- Addresses fundamental questions in high energy and nuclear physics, directly related to major experimental programs
- Determine parameters for Standard Model, including quark mass
- **MILC lattice generation for the lightest quark mass complete, moving onto HISQ quarks**

**Methods and Challenges**

- Rational Hybrid Monte Carlo
- For scalability and performance developed
  - QLA : 3x3 matrix linear algebra operations
  - QMP : low-level routines, partial MPI replacement
- Tuning complex algorithms after unexpected fall off in Monte Carlo acceptance rate

**Diagram**

- DWF Production time history
  - $32^3 \times 64 \times 16$, $m_c = 0.03$

- $m_l = 0.004$ QCDOC
- $m_l = 0.004$ BG/P (2K,4K)
- $m_l = 0.006$ QCDOC 4K (UKQCD)
- $m_l = 0.006$ ANL BG/P (2K,4K)
- $m_l = 0.008$ ANL BG/P 2K
- $48^3 \times 64 \times 16$, $m_l = 0.002$ ANL BG/P (4,8K)

**150x improvement**

Directed MILC code to do a series of successive local 1d-FFTs using FFTW and shuffling the data as necessary.
Lattice QCD

64^4 strong scaling
36.8% peak
Strong Scaling

- Test on 72 racks BG/P installation at supercomputer center Jülich (Gerhold, Herdioza, Urbach, K.J.)
- using tmHMC code (Urbach, K.J.)
Summary

● Emerging HPC landscape is extremely complex

● A time of extraordinary potential
  ● Game-changing capability is now available

● A time of significant challenge
  ● Just as HPC starts to have real scientific and industrial impact - it gets extraordinarily hard.

● A radical research and development approach required
  ● Multidisciplinary from domain to system design
  ● Collaboration across research teams within universities
  ● Collaboration between universities and industry