Case Study: DDCMP: Beyond Homogeneous Decomposition with ddcMD

Scaling Long-Range Forces on Massively Parallel Systems

Donald Frederick, LLNL
Presented at Supercomputing ‘11
Case Study: Outline

• Problem Description
• Computational Approach
• Changes for Scaling
The temporal evolution of plasmas depends on the complex interaction of collisional, radiative, and reactive processes.

**Photon**s

**Free and bound** electrons

**Light** ions $P, D, T, \text{He}^3, \text{He}^4$

**Partially or fully** ionized Hi-Z ions

**Radiation**
- Bremsstrahlung
- Compton
- Blackbody in LTE

**Coulomb Collisions**
- Classical
- Quantum

**Atomic Processes**
- Electron impact ionization
- Photo-ionization
- Radiative recombination
- 3-body recombination

**Light ion TN burn**
- $D + T \Rightarrow n + \alpha$
- $D + D \Rightarrow T + p$
- $D + D \Rightarrow \text{He}^3 + n$
- $T + T \Rightarrow \alpha + 2n$
Equations of motion for the particles are solved in detail

All particles (including electrons) are dynamic

Fully non-equilibrium

Approximations are isolated and understood

Opportunity for discovery science

Requires large numbers of particles

Charged particles interact via long-range forces
Two Methods to Achieve Fusion

Conventional ICF
- Fuel is compressed and heated until it auto-ignites
- Similar to a diesel engine
- Compression details are critical

Fast Ignition
- Fuel partially compressed then locally heated by high energy beam
- Like a spark plug in a gas engine
- Beam interactions are critical
Realistic Energy Flow Requires Micron Scale

Energy flow processes must be accurately understood in two limits:

Short range, fast scattering processes govern beam heating, e- conduction, diffusion

- $dr < 1 \, \text{Å}$
- $v \sim 100$’s Å/fs

Long-range charge interactions
- Uniform T => local charge neutrality
- T gradient => Thermoelectric field
- Strong, long-range field affects energy flow

**Scatter:** timestep $\sim 10^{-4}$ fs

**Transport:** hydrodynamic length scale $\sim \mu$m
Multi-Physics Effects of Beam Heating

- High energy beam causes non-uniform heating.
- Electrons conduct heat away from heated zone.
- System is no longer locally neutral.
- Induced electric field feeds back on thermal conductivity and density response.
- Smaller, shorter simulations do not capture thermoelectric complexity.

We have demonstrated it is possible to perform very large simulations with long-range forces and >80% parallel efficiency.

Electron potential
Multi-Physics Effects of Beam Heating

- High energy beam causes non-uniform heating
- Electrons conduct heat away from heated zone
- System is no longer locally neutral
- Induced electric field feeds back on thermal conduction and density response.
- Smaller, shorter simulations do not capture thermoelectric complexity

We have demonstrated it is possible to perform very large simulations with long-range forces and >80% parallel efficiency.
**Dawn (LLNL)**
Nodes: 36,864
Cores: 147,456
Peak: 501 TFlop/s
Memory: 147 Tbytes (4 GB/node)

**Jugene (Jülich)**
Nodes: 73,728
Cores: 294,912
Peak: 1.0 PFlop/s
Memory: 147 Tbytes (2 GB/node)
Interaction Range and Scalability for MD

Short-Range Forces

Long-Range Coulomb Forces

- Long-Range Force Challenges
- $O(N^2)$ Scaling
- All-to-all Communication
Particle-Particle Particle-Mesh (PPPM) Method

Divides problem into long-range and short-range parts

Screening width is an optimization parameter

$$U = \frac{1}{2} \sum_{ij} q_i q_j \sum_{n} e^{-\frac{qr}{2\eta L}} + \frac{1}{2} \iiint \frac{\rho(r)\rho(r')}{r - r'}$$

Short-range: Converges quickly in real space
Long-range: Converges quickly in Fourier space (3D FFT)
Homogeneous PPPM Implementation

- Pair interactions and FFTs are performed in serial
- All tasks have an equivalent slice of the problem and work synchronously
- Width of screening charge balances work between pair and FFT parts
- Scaling breaks down on large numbers of tasks due to all-to-all required for 3D FFTs
- Poor time-to-solution

Update Positions

$t = t_0$

Calculate Explicit Pair Interactions

$t = t_0$ + $\Delta t$

Distribute Charge to Grid

Solve Poisson Equation in Reciprocal Space using 3D FFTs

Interpolate Potential from Grid

Compute Forces

This problem appears to be un-scalable
Beyond Homogeneous Decomposition

Particle/mesh task split is an additional optimization parameter
A 3D FFT is a Series of 1D FFTs

Map 3D density $\rho(x,y,z)$ onto 2D process grid (nprow x npcol)

- npcol process columns
- nprow process rows

Alltoallv (over cols)
- x local
- y distributed over columns
- z distributed over rows

Alltoallv (over rows)
- x distributed over columns
- y local
- z distributed over rows

x distributed over columns
- y distributed over rows
- z local
3D FFTs have inherently poor weak scaling, no matter how well optimized. Excellent strong scaling of standalone FFT yields maximum flexibility in load balancing large-scale PPPM.
Single task 1D FFT kernel

Once parallel scaling is optimized, performance depends on efficient single task FFT kernel

benchmark: 1D FFT of randomized Gaussian, fwd + bwd 100 times (seconds)

<table>
<thead>
<tr>
<th>size</th>
<th>ESSL</th>
<th>FFTW 2.1.3</th>
<th>Spiral FFT *</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.00102</td>
<td>0.00142</td>
<td>0.00089</td>
</tr>
<tr>
<td>256</td>
<td>0.00169</td>
<td>0.00317</td>
<td>0.00173</td>
</tr>
<tr>
<td>512</td>
<td>0.00584</td>
<td>0.00699</td>
<td>0.00507</td>
</tr>
<tr>
<td>1024</td>
<td>0.01250</td>
<td>0.02067</td>
<td>0.01489</td>
</tr>
<tr>
<td>2048</td>
<td>0.03207</td>
<td>0.05367</td>
<td>0.04197</td>
</tr>
<tr>
<td>4096</td>
<td>0.06246</td>
<td>0.11930</td>
<td>0.09260</td>
</tr>
</tbody>
</table>

* http://www.spiral.net

On BG/P, ESSL is fastest for large sizes, but Spiral optimization is ongoing
Pair Kernel Optimization for Blue Gene/P Dual Floating Point Unit (Double Hummer)

- Particles assigned to cells and then sorted by cell
- Compute distance \( r_{ij} \) for all particle pairs between neighboring cells. Inner and outer loops unrolled by 4
- Filter separations \( r_{ij} < r_c \)
- \( \text{erfc}(\alpha r_{ij})/r_{ij} \) and derivative calculated as a series expansion using a custom vector function. Length 5 is optimal

\[
\begin{align*}
\mathbf{r}_{ij}^2 &= (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 \\
\frac{\mathbf{r}_{ij}^2}{2} &= \frac{r_i^2}{2} + \frac{r_j^2}{2} - x_i x_j - y_i y_j - z_i z_j
\end{align*}
\]

8 flops = 3 sub + 1 mult + 2 madd = 3 cycles

7 flops = 1 add + 3 msub = 2 cycles
Efficient Communication Between Task Groups

Particle tasks and mesh tasks must exchange data (twice) each time step. Asymmetric overhead is ok.

**Stage 1**
*Compute Density*

Density spreads beyond spatial domain of each particle task

**Stage 2**
*Gather/Reduce*

Tasks gather and reduce their part of the grid

**Stage 3**
*Pgather*

DMA allows particle tasks to overlap computation and communication
Animation of Data Motion

- Charge Density
- Gather/Reduce
- Pgather
- Pair Calculation
- FFT Calculation
- Pscatter
- Scatter/Spread
- Electric Field
We Demonstrate 324.8 TFlop/s on a 2.4-Billion Particle Simulation of a Dense Hydrogen Plasma.

- Achieve over 80% parallel efficiency across 278,528 CPU’s
- Fastest time-to-solution on Coulomb problem: 5-Billion particle-updates/sec (> 25X prior state-of-the-art)
- Sustained performance greater than 30% of peak on Jugene supercomputer
- Demonstrated highly scalable 3D FFT implementation
- Largest ever N-body simulation of beam-plasma interaction

**Heterogeneous decomposition enables the solution of today’s apparently un-scalable problems on tomorrow’s super-massively parallel computers**
Conclusion

• We have demonstrated that charged particle simulations can be scaled to hundreds of thousands of tasks using heterogeneous decomposition
  – Peak performance of 324.8 Tflop/s (>30% of peak) TFlop/s on Jugene despite complex communication patterns and short iteration cycles
  – Particle update rates 25x previous efforts make long time scales accessible

• A new simulation capability for plasma physics

• Heterogeneous decomposition will be a useful technique for next generation supercomputers
Acknowledgments

The ddcMD Team - Dave Richards – LLNL
J.N. Glosli, B. Chan, M.R. Dorr, E.W. Draeger, J.-L. Fattebert,
W.D. Krauss, T. Spelce, F.H. Streitz, M.P. Surh,
J.A. Gunnels (IBM)