Case Case Study DDCMP: Beyond Homogeneous Decomposition with ddcMD

Scaling Long-Range Forces on Massively Parallel Systems

Donald Frederick, LLNL

Presented at Supercomputing '11

Lawrence Livermore National Laboratory, P. O. Box 808, Livermore, CA 94551

LLNL-PRES-508651

Case Study: Outline

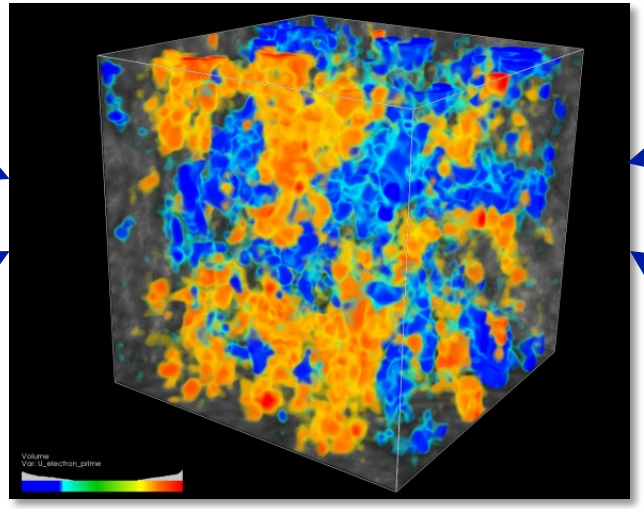
- Problem Description
- Computational Approach
- Changes for Scaling

Plasma Physics and Controlled Fusion

The temporal evolution of plasmas depends on the complex interaction of collisional, radiative, and reactive processes

Photons

Free and bound electrons

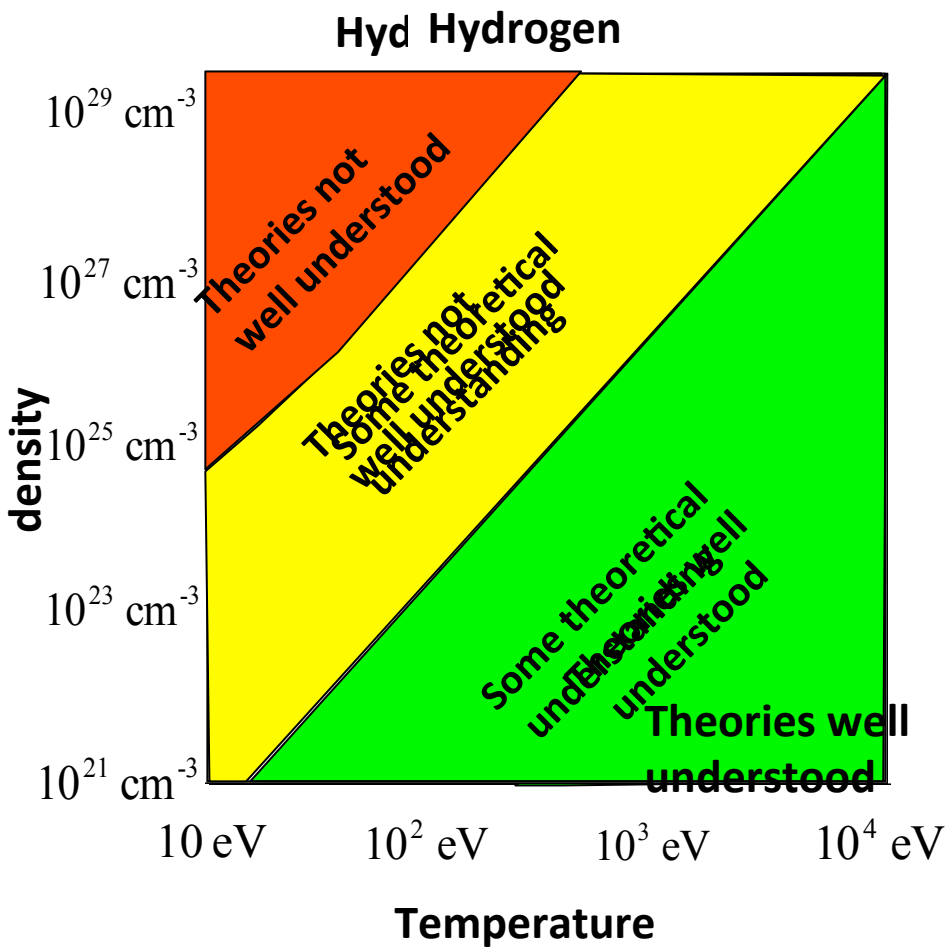


Partially or fully ionized Hi-Z ions

Light ions
P, D, T, He^3,He^4

Radiation	Coulomb Collisions	Atomic Processes	Light ion TN burn
Bremsstrahlung	Classical	Electron impact ionization	$D + T \Rightarrow n + \alpha$
Compton	Quantum	Photo-ionization	$D + D \Rightarrow T + p$
Blackbody in LTE		Radiative recombination	$D + D \Rightarrow {}^3He + n$
		3-body recombination	$T + T \Rightarrow \alpha + 2n$

Simulation is an Essential Tool

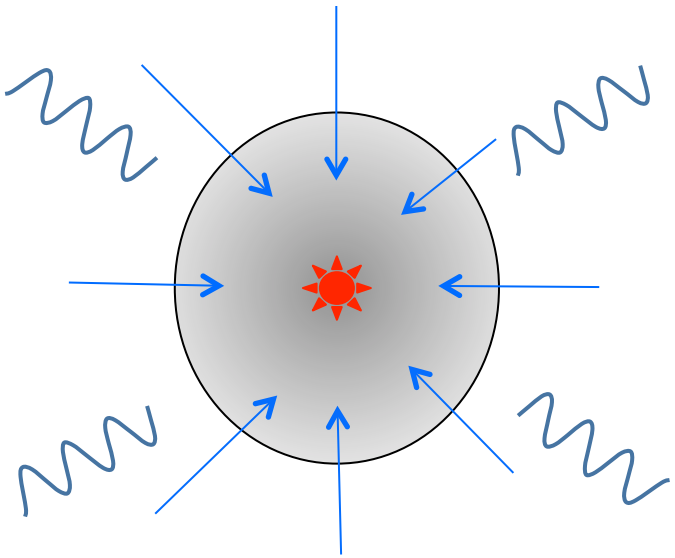


Pros/Cons for MD simulation of hot, dense, plasmas

- 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*

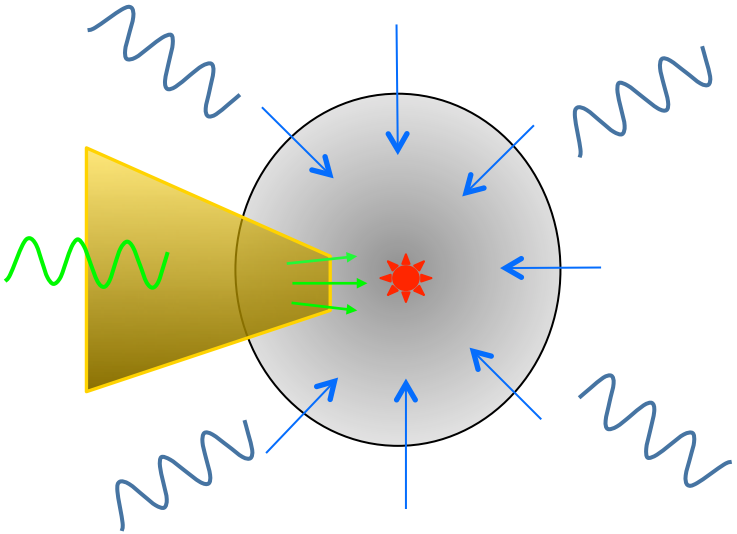
Two Methods to Achieve Fusion

Conventional ICF



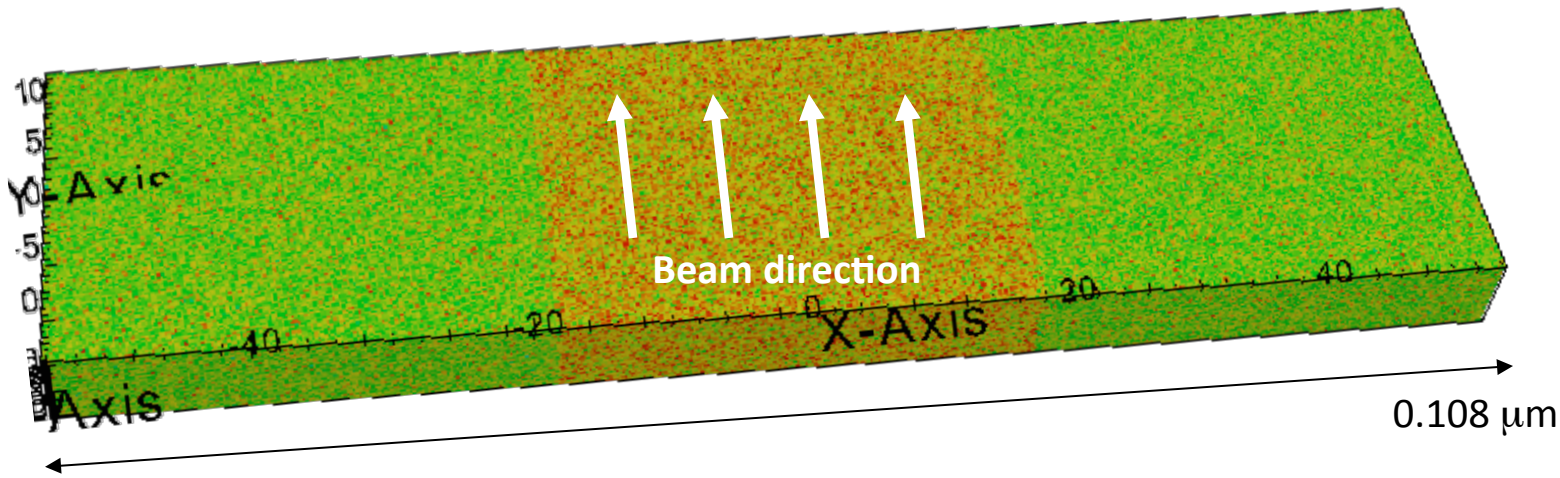
- Fuel is compressed and heated until it auto-ignites
- Similar to a diesel engine
- Compression details are critical

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 Å

v ~ 100's Å/fs

Scatter: timestep ~ 10^{-4} fs

Long-range charge interactions

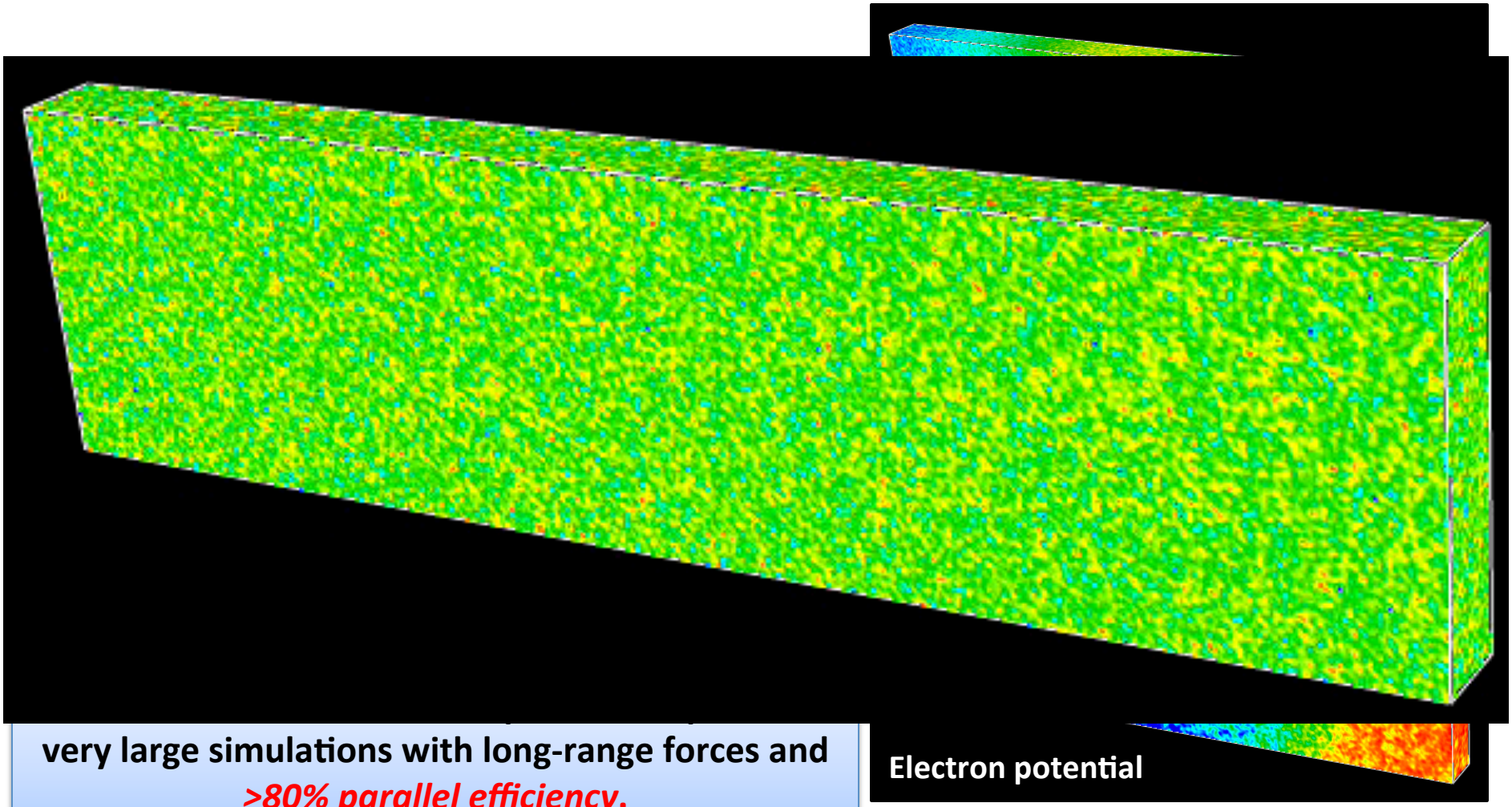
Uniform T => local charge neutrality

T gradient => Thermoelectric field

Strong, long-range field affects energy flow

Transport: hydrodynamic length scale ~μm

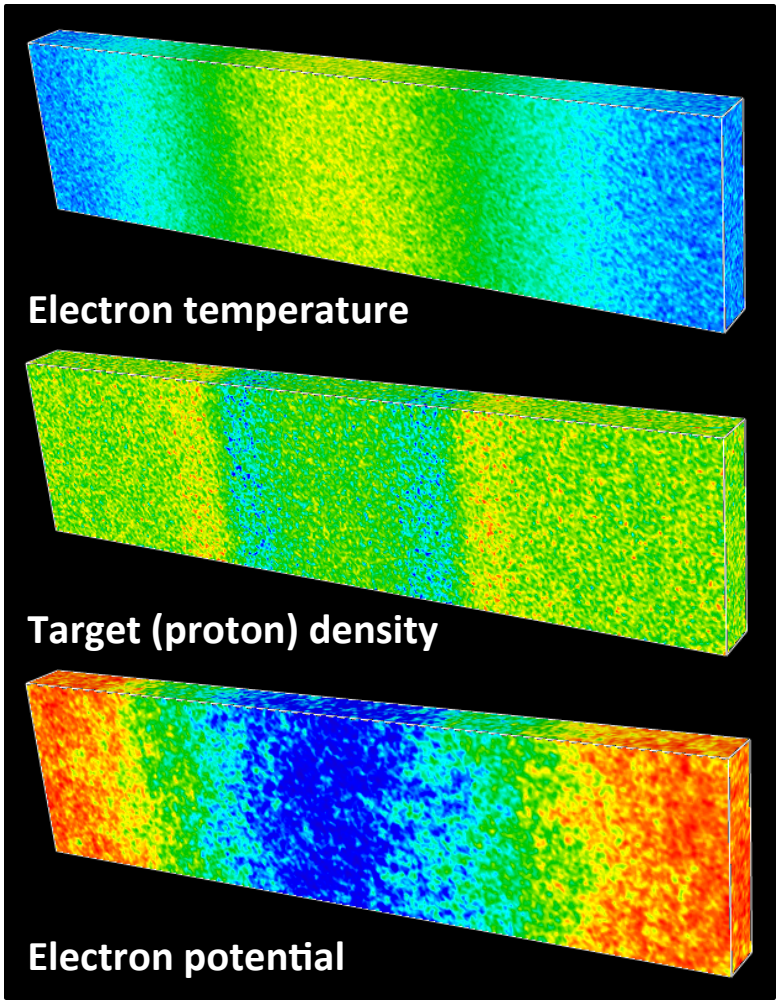
Multi-Physics Effects of Beam Heating



very large simulations with long-range forces and *>80% parallel efficiency*.

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 and Jugene

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

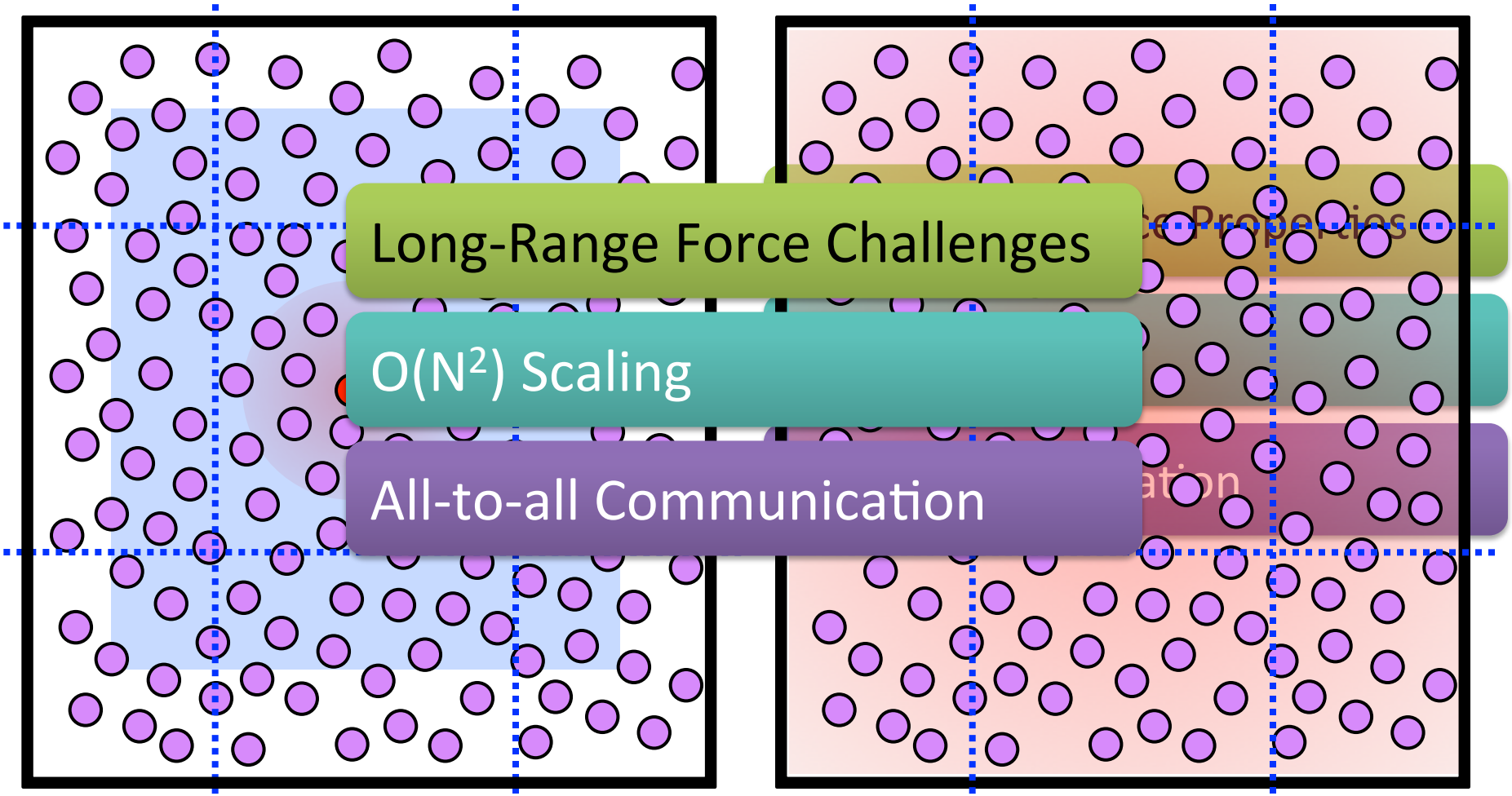
Peak:1.0 PFlop/s

Memory: 147 Tbytes (2 GB/node)

Interaction Range and Scalability for MD

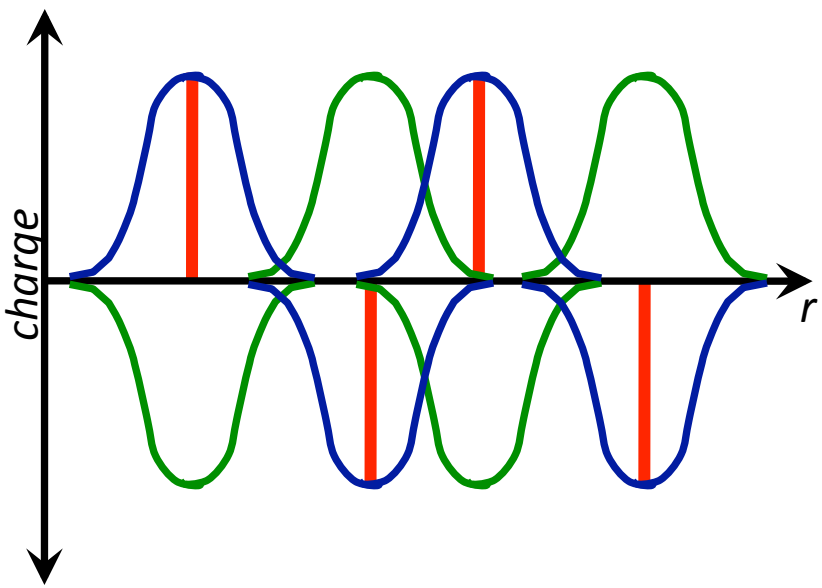
Short-Range Forces

Long-Range Coulomb Forces



Particle-Particle Particle-Mesh (PPPM) Method

Divides problem into long-range and short-range parts



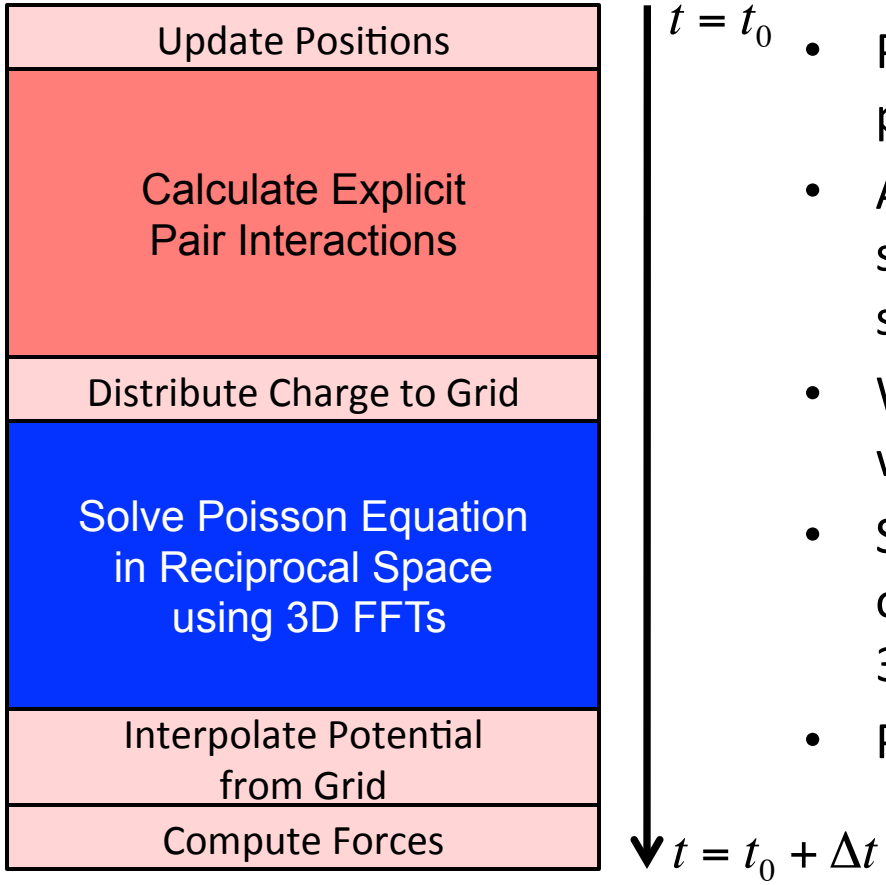
Short-range: Converges quickly in real space

$$U = \frac{1}{2}\sum_{ij}\frac{q_i q_j}{r_{ij}} = \sum_{ij}\sum_{n}\frac{q_i q_j}{r_{ij}}\operatorname{erfc}\left(\frac{r_{ij}}{\sqrt{2}}\right) + \frac{1}{2}\iint\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$$

Long-range: Converges quickly in Fourier space (3D FFT)

Screening width is a optimization parameter

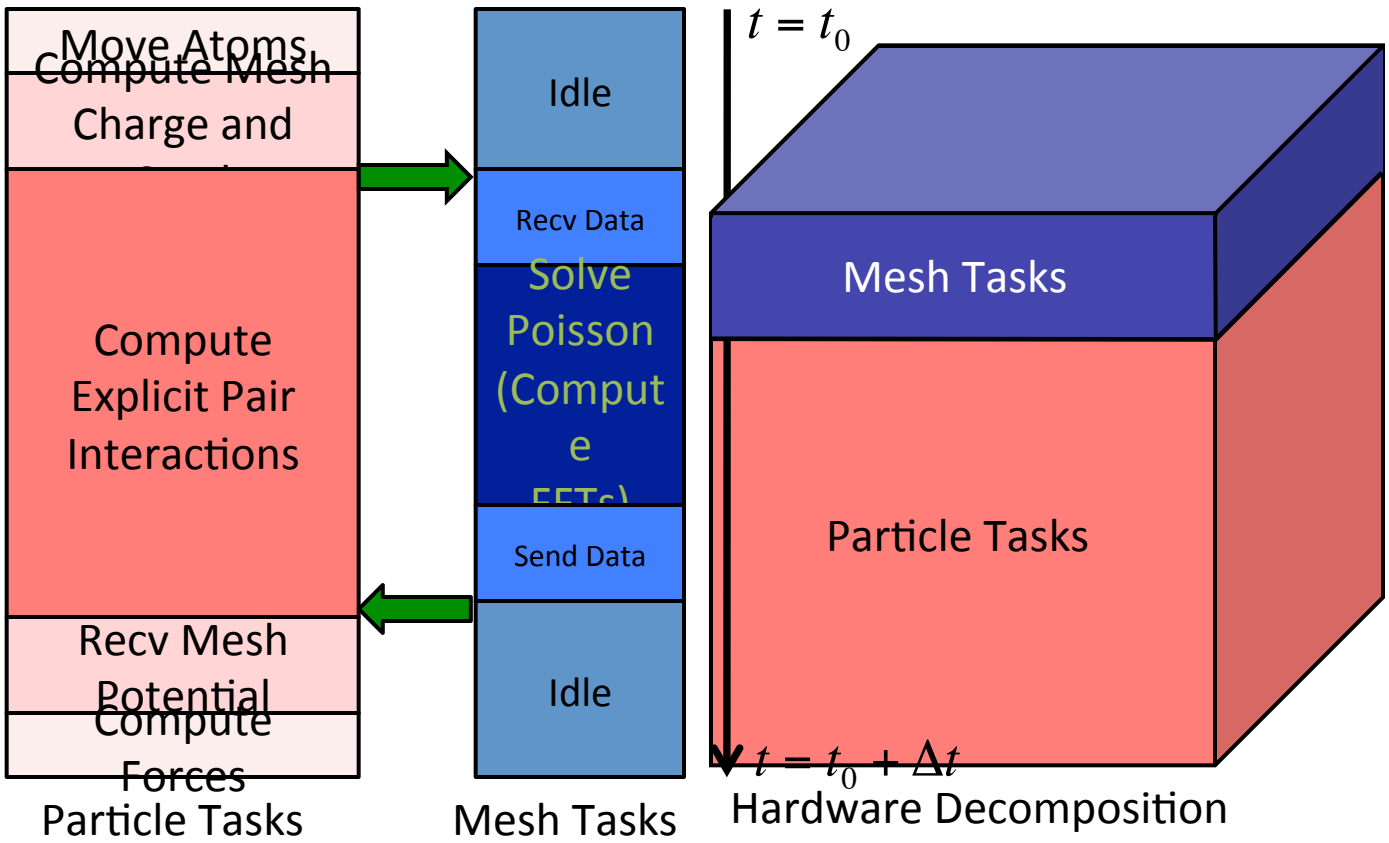
Homogeneous PPPM Implementation



- 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-allv required for 3D FFTs
- Poor time-to-solution

This problem appears to be un-scalable

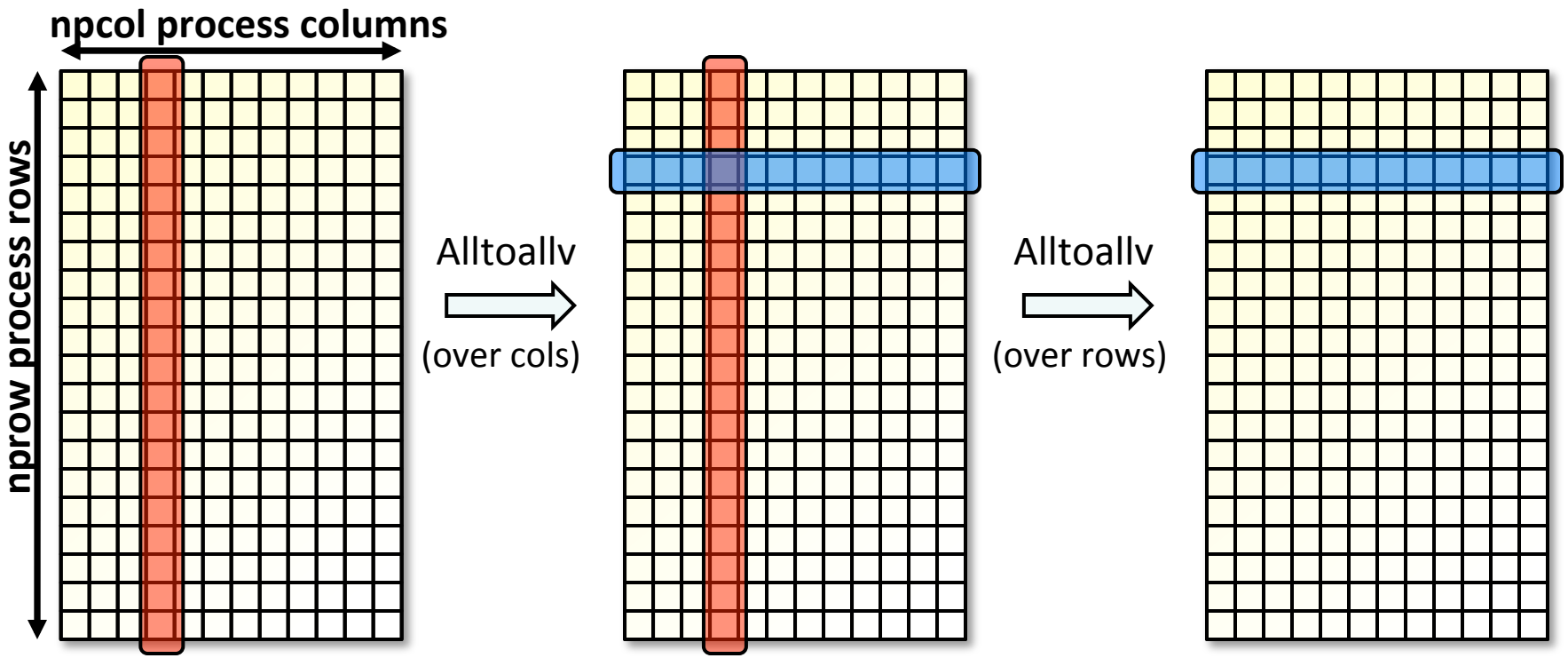
Beyond Homogeneous Decomposition



Particle/mesh task split is an additional optimization parameter

A 3D FFT is a Series of 1D FFTs

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

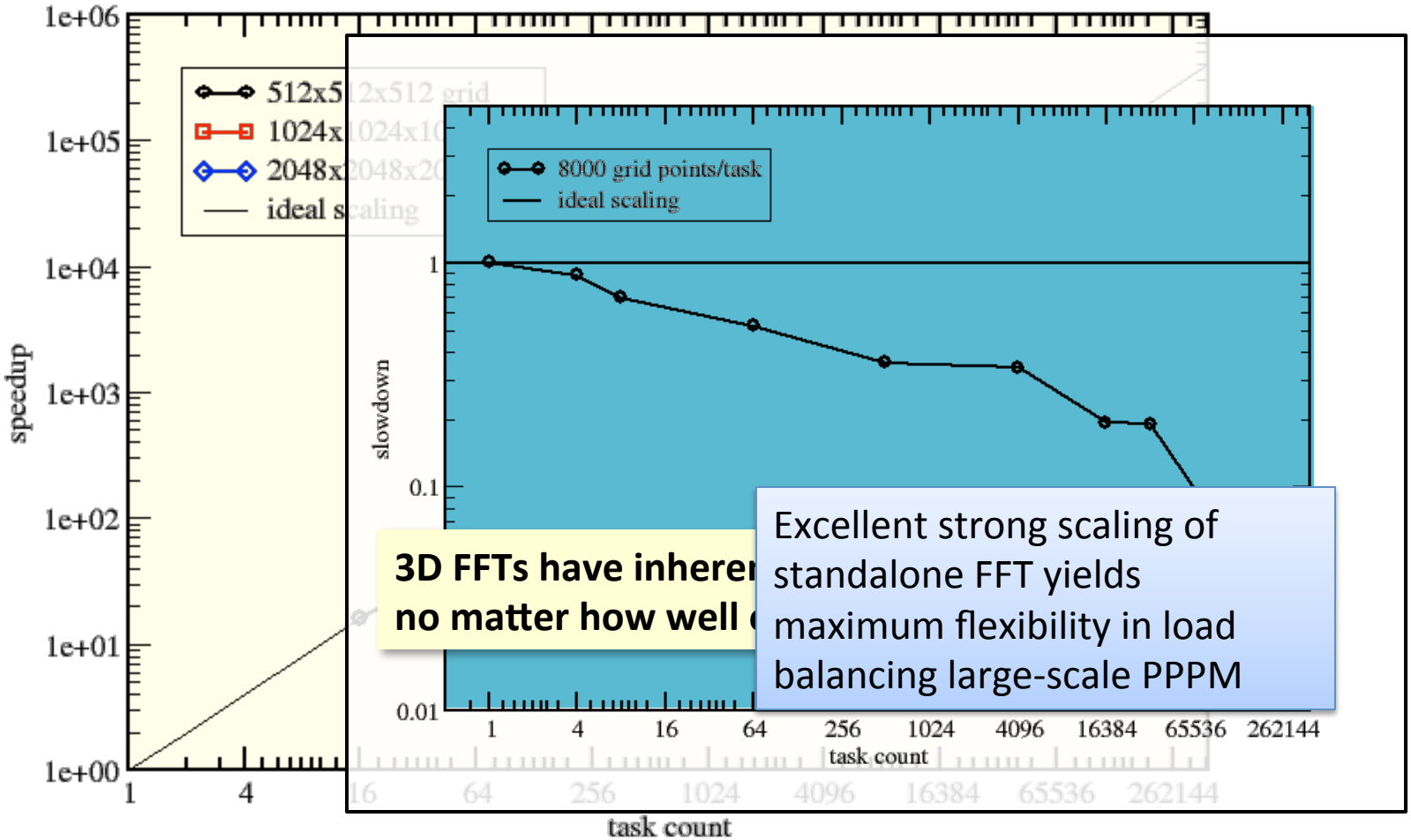


x local
y distributed over columns
z distributed over rows

x distributed over columns
y local
z distributed over rows

x distributed over columns
y distributed over rows
z local

3D FFT: Scaling Results



3D FFTs have inheren
no matter how well

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

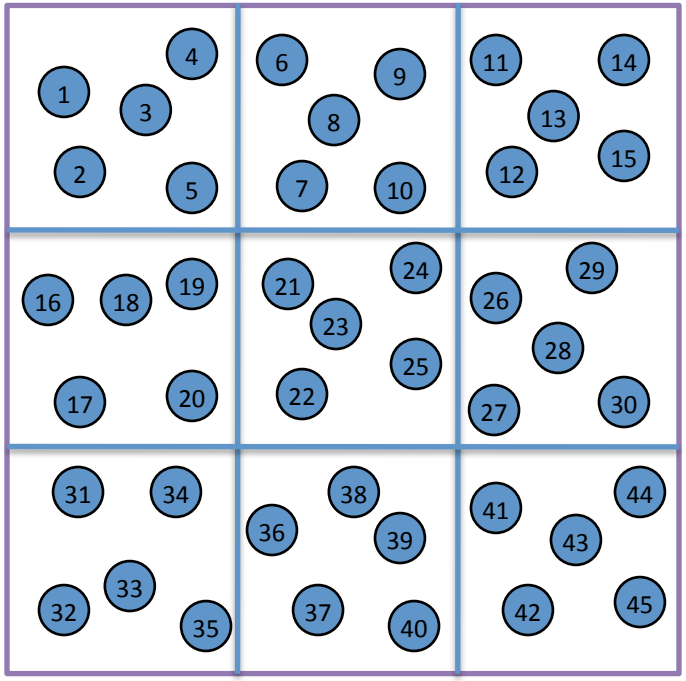
benchmark: 1D FFT of randomized Gaussian, fwd + bwd 100 times (seconds)

size	ESSL	FFTW 2.1.3	Spiral FFT *
128	0.00102	0.00142	0.00089
256	0.00169	0.00317	0.00173
512	0.00584	0.00699	0.00507
1024	0.01250	0.02067	0.01489
2048	0.03207	0.05367	0.04197
4096	0.06246	0.11930	0.09260

*** 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 $r_{ij}<r_c$
- $\mathrm{erfc}(\alpha r_{ij})/r_{ij}$ and derivative calculated as a series expansion using a custom vector function. Length 5 is optimal

$$r_{ij}^{\ 2} = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2$$

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

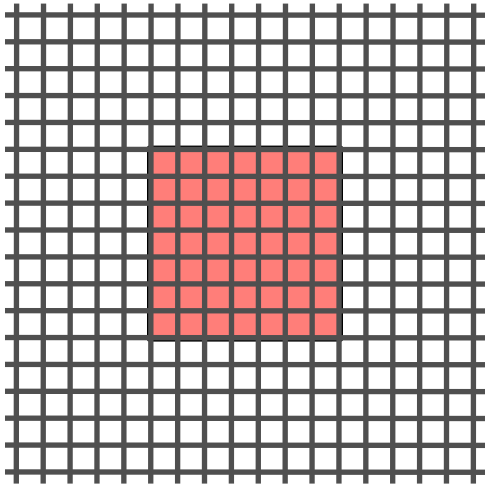
$$\frac{r_{ij}^{\ 2}}{2} = \frac{r_i^2}{2} + \frac{r_j^2}{2} - x_i x_j - y_i y_j - z_i z_j$$

7 flops =1 add+ 3 msub =2 cycles

Efficient Communication Between Task Groups

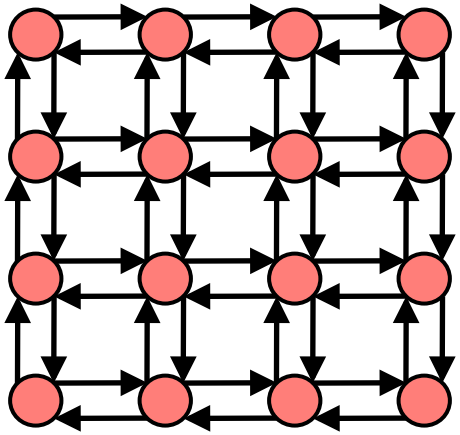
Particle tasks and mesh tasks must exchange data (twice) each time step. Asymmetric overhead is ok.

Stage 1
Compute Density



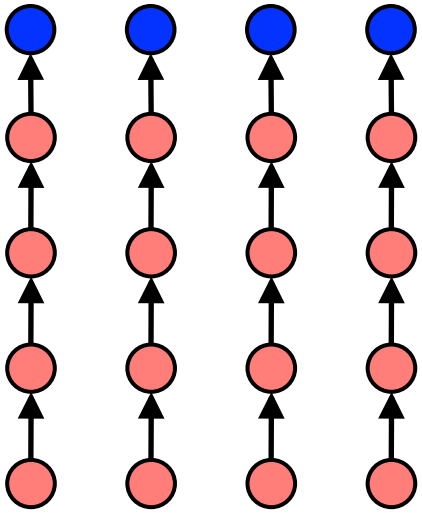
Density spreads beyond spatial domain of each particle task

Stage 2
Gather/Reduce



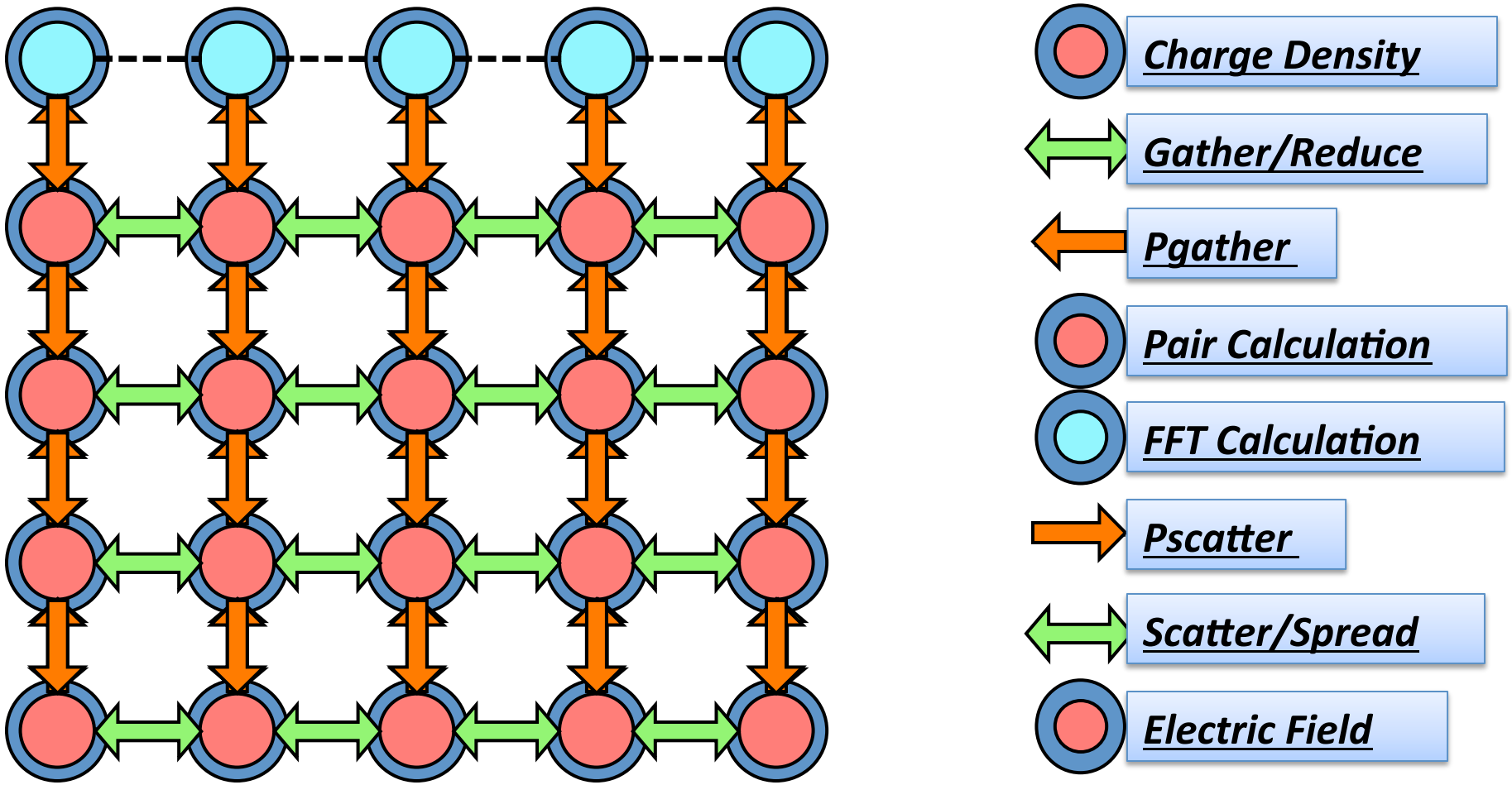
Tasks gather and reduce their part of the grid

Stage 3
Pgather

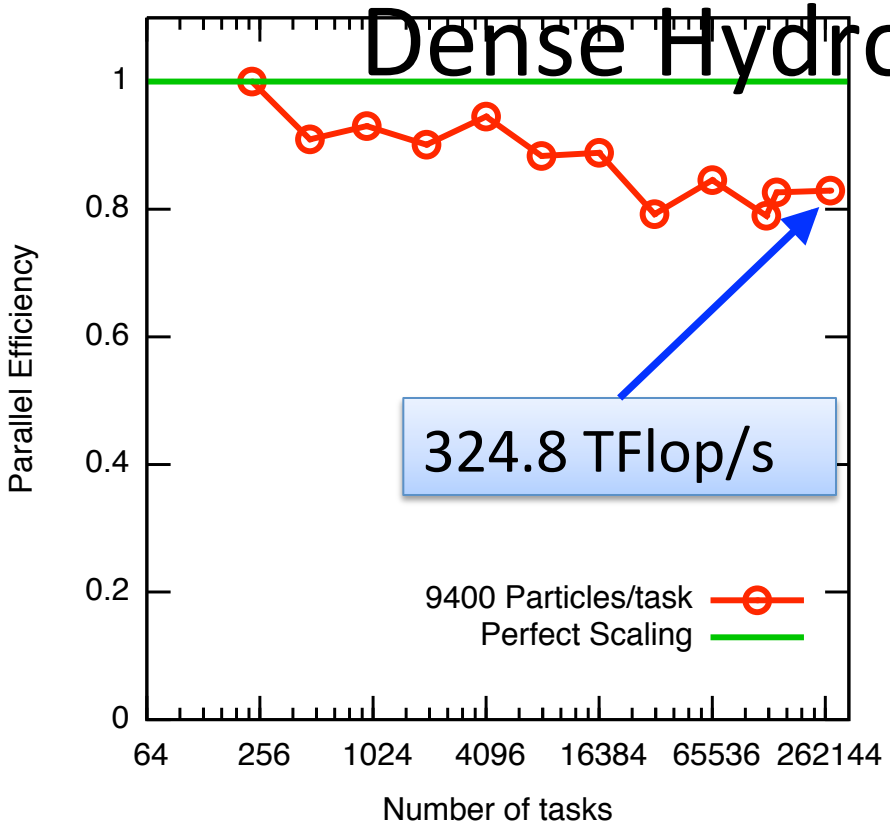


DMA allows particle tasks to overlap computation and communication

Animation of Data Motion



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)*