LeanCP: Terascale Car-Parrinello *ab initio* molecular dynamics using charm++

Application Team

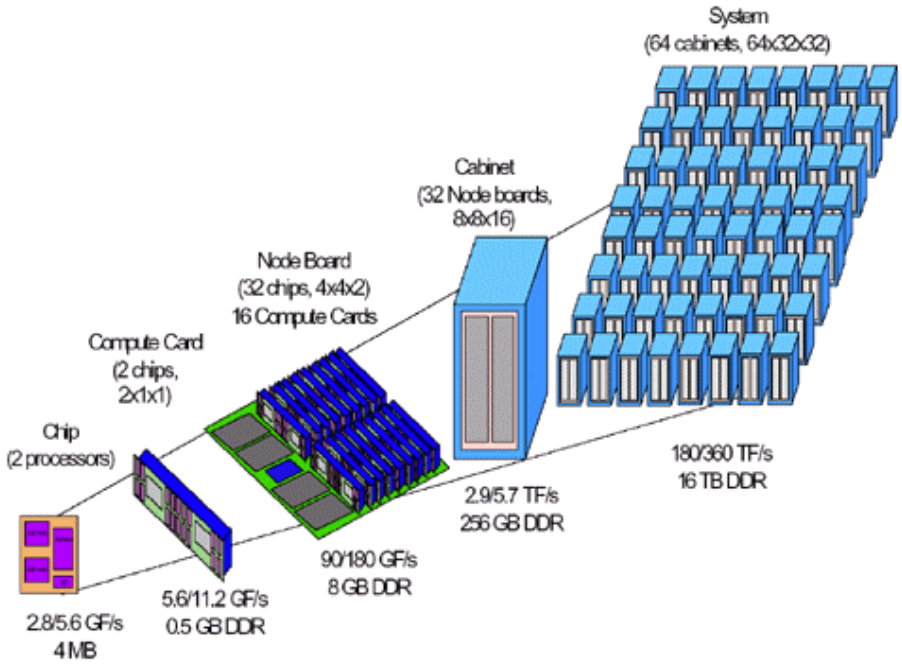
Glenn J. Martyna, Physical Sciences Division, IBM Research
Jason Crain, School of Physics, Edinburgh University
Susan Allison, School of Physics, Edinburgh University
Simon Bates, School of Physics, Edinburgh University
Bin Chen, Department of Chemistry, Louisiana State University
Troy Whitfield, IBM Research, Physical Sciences Division
Yves Mantz, IBM Research, Physical Sciences Division

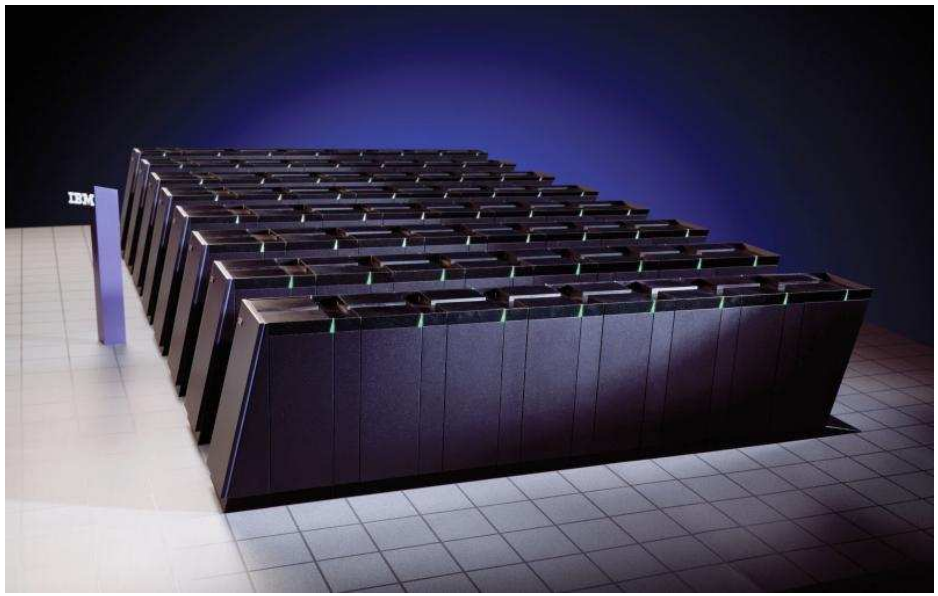
Methods/Software Development Team

Glenn J. Martyna, Physical Sciences Division, IBM Research
Mark E. Tuckerman, Department of Chemistry, NYU
Peter Minary, Computer Science/Bioinformatics, Stanford University.
Laxmikant Kale, Computer Science Department, UIUC
Ramkumar Vadali, Computer Science Department, UIUC
Sameer Kumar, Computer Science, IBM Research
Eric Bohm, Computer Science Department, UIUC
Abhinav Bhatele, Computer Science Department, UIUC

Funding : NSF, IBM Research

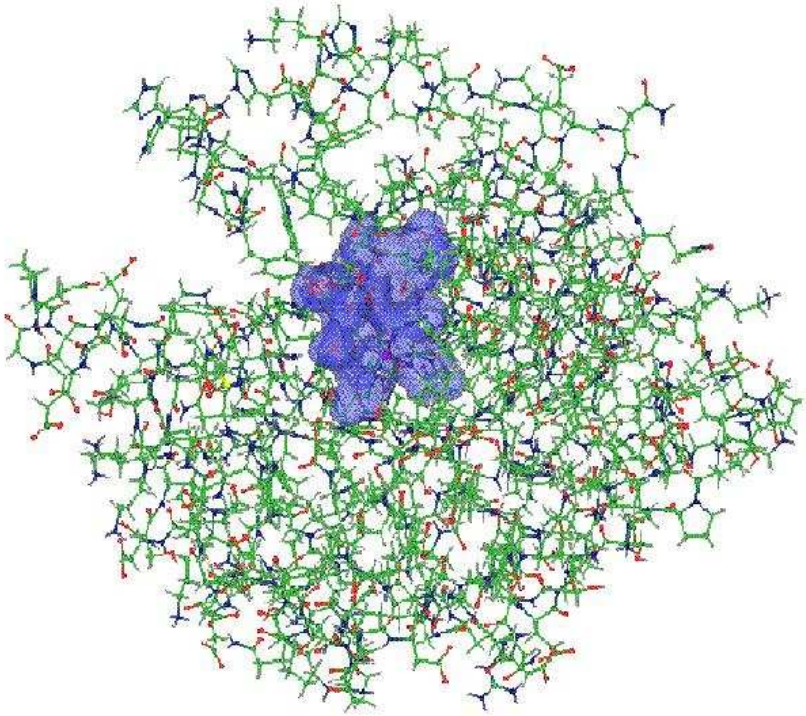
IBM’s Blue Gene/L network torus supercomputer

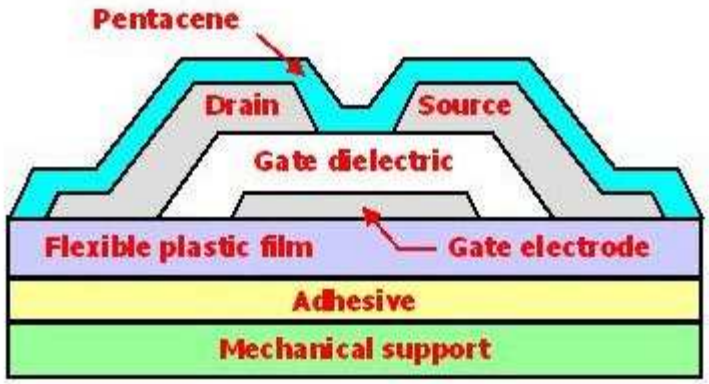


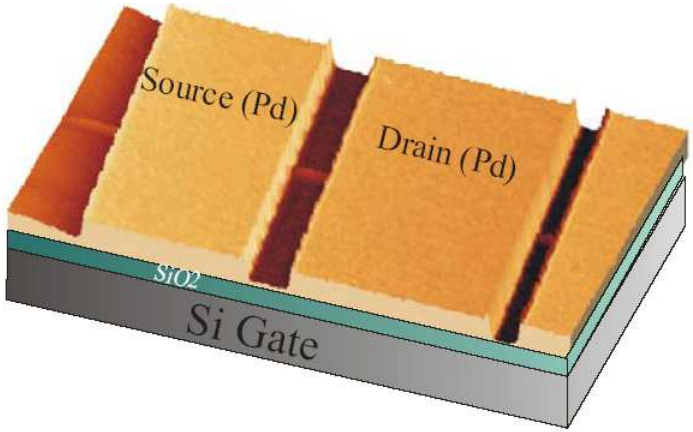


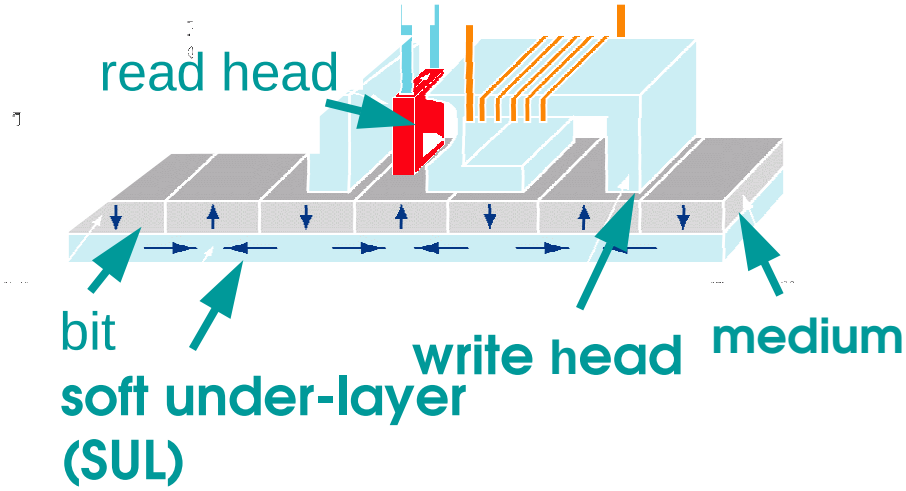
The worlds fastest supercomputer!

Goal : The accurate treatment of complex heterogeneous systems to gain physical insight.









Characteristics of current models

Empirical Models: Fixed charge,
non-polarizable,
pair dispersion.

Ab Initio Models: GGA-DFT,
Self interaction present,
Dispersion absent.

Problems with current models (empirical)

Dipole Polarizability : Including dipole polarizability changes solvation shells of ions and drives them to the surface.

Higher Polarizabilities: Quadrupolar and octapolar polarizabilities are NOT SMALL.

All Manybody Dispersion terms: Surface tensions and bulk properties determined using accurate pair potentials are incorrect. Surface tensions and bulk properties are both recovered using manybody dispersion and an accurate pair potential. An effective pair potential destroys surface properties but reproduces the bulk.

The force fields cannot treat chemical reactions:

Problems with current models (DFT)

- **Incorrect treatment of self-interaction/exchange**: Errors in electron affinity, band gaps …

- **Incorrect treatment of correlation**: Problematic treatment of spin states. The ground state of transition metals (Ti, V, Co) and spin splitting in Ni are in error. Ni oxide incorrectly predicted to be metallic when magnetic long-range order is absent.

- **Incorrect treatment of dispersion** : Both exchange and correlation contribute.

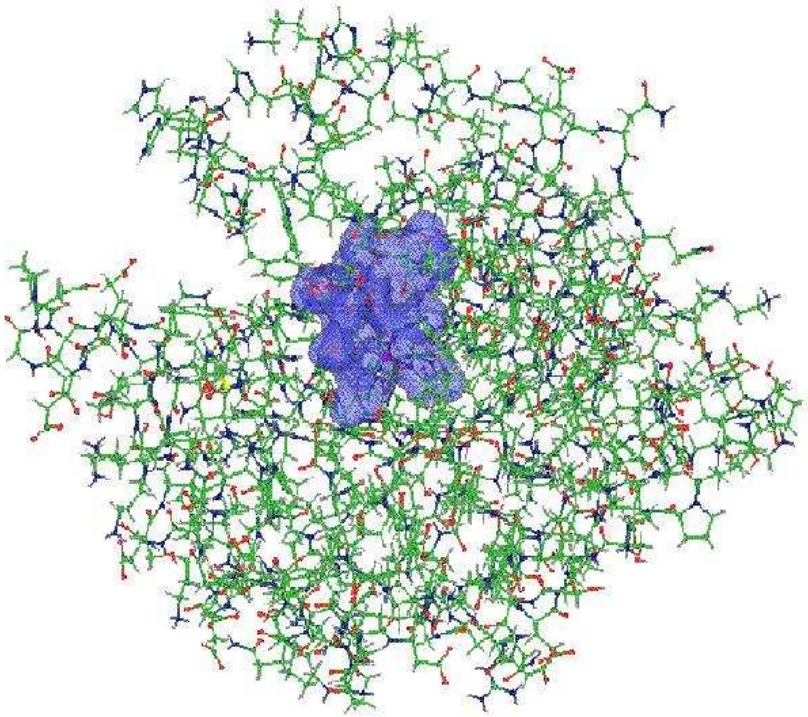
- **KS states are NOT physical objects** : The bands of the exact DFT are problematic. TDDFT with a frequency dependent functional (exact) is required to treat excitations even within the Born-Oppenheimer approximation.

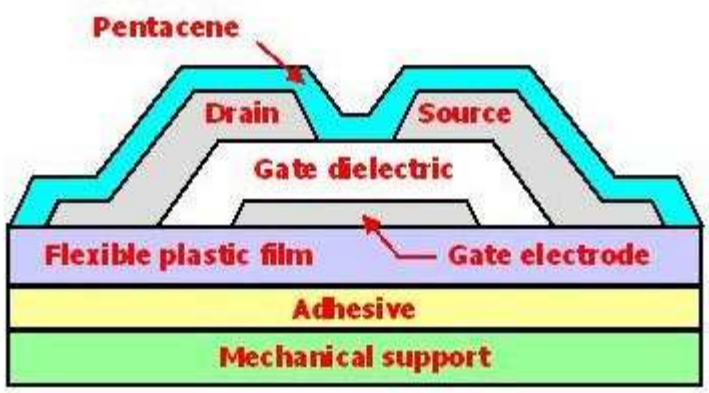
Conclusion : Current Models

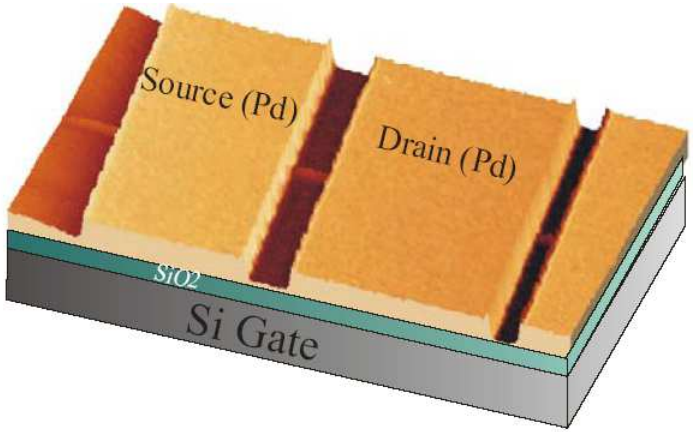
- Simulations are likely to provide semi-quantitative accuracy/agreement with experiment.
- Simulations are best used to obtain insight and examine physics .e.g. to promote understanding.

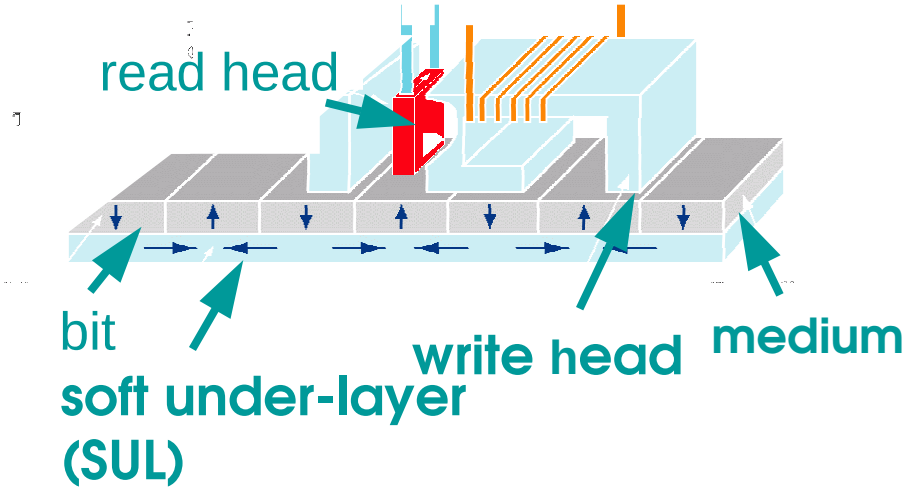
Nonetheless, in order to provide truthful solutions of the models, simulations must be performed to long time scales!

Goal : The accurate treatment of complex heterogeneous systems to gain physical insight.









Evolving the model systems in time:

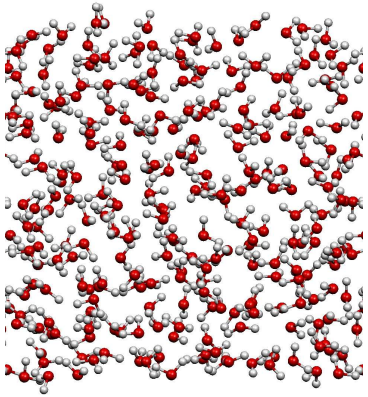
- **Classical Molecular Dynamics :** Solve Newton's equations or a modified set numerically to yield averages in alternative ensembles (NVT or NPT as opposed to NVE) on an empirical parameterized potential surface.
- **Path Integral Molecular Dynamics :** Solve a set of equations of motion numerically on an empirical potential surface that yields canonical averages of a classical ring polymer system isomorphic to a finite temperature quantum system.
- ***Ab Initio* Molecular Dynamics** : Solve Newton's equations or a modified set numerically to yield averages in alternative ensembles (NVT or NPT as opposed to NVE) on a potential surface obtained from an *ab initio* calculation.
- **Path Integral *ab initio* Molecular Dynamics** : Solve a set of equations of motion numerically on an *ab initio* potential surface to yield canonical averages of a classical ring polymer system isomorphic to a finite temperature quantum system.

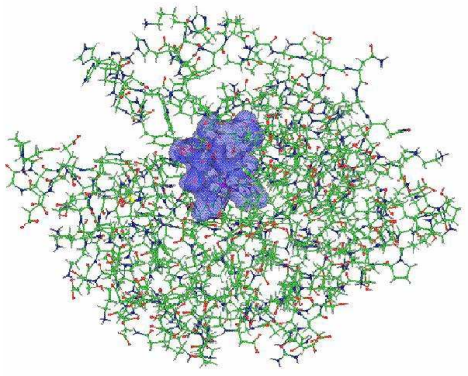
Reaching longer time scales : Recent efforts

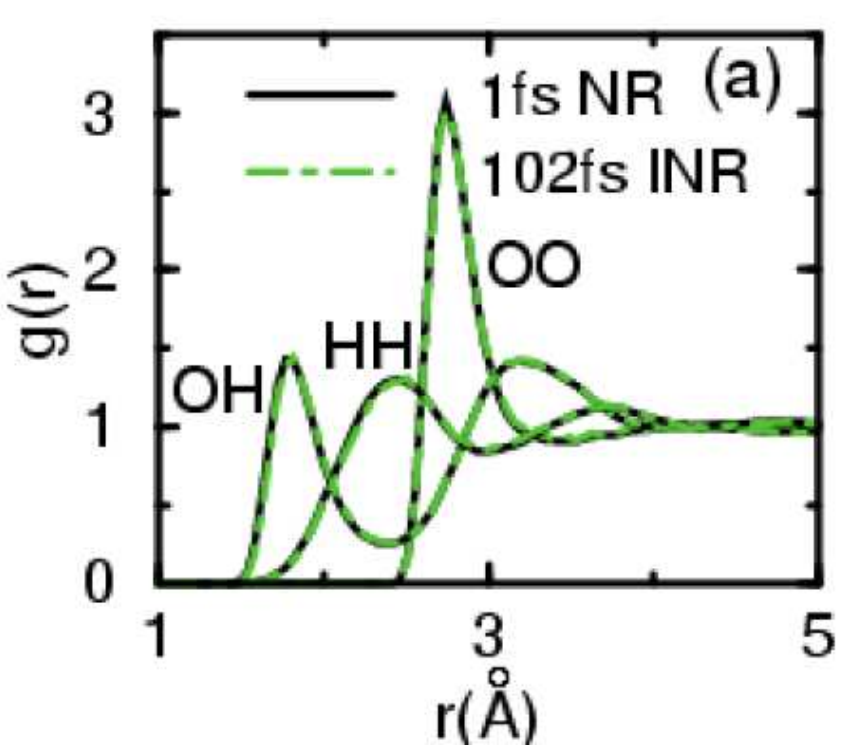
- Increase the **time step 100x from 1fs to 100fs** in numerical solutions of empirical model MD simulations.
- Reduce the **scaling** of **non-local pseudopotential** computations in plane wave based DFT with the number of atoms in the system, **N**, from $\mathbf{N^3}$ **to** $\mathbf{N^2}$.
- Increasing the **stability of Car-Parrinello** *ab initio* MD under **extreme conditions** for studying metals and chemical reactions.
- Naturally extend the plane-wave basis sets to treat clusters, surfaces and wires.

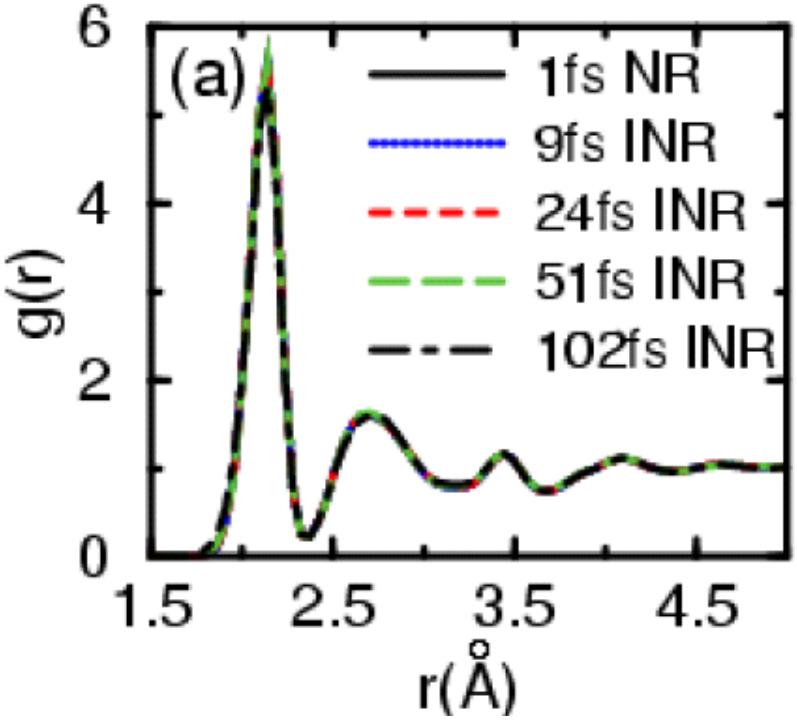
G.J. Martyna et al *Phys. Rev. Lett.* **93**, 150201 (2004).
Chem. Phys. Phys. Chem. **6**, 1827 (2005).
J. Chem. Phys. **118**, 2527 (2003).
J. Chem. Phys. **121**, 11949 (2004).

Improving Molecular Dynamics



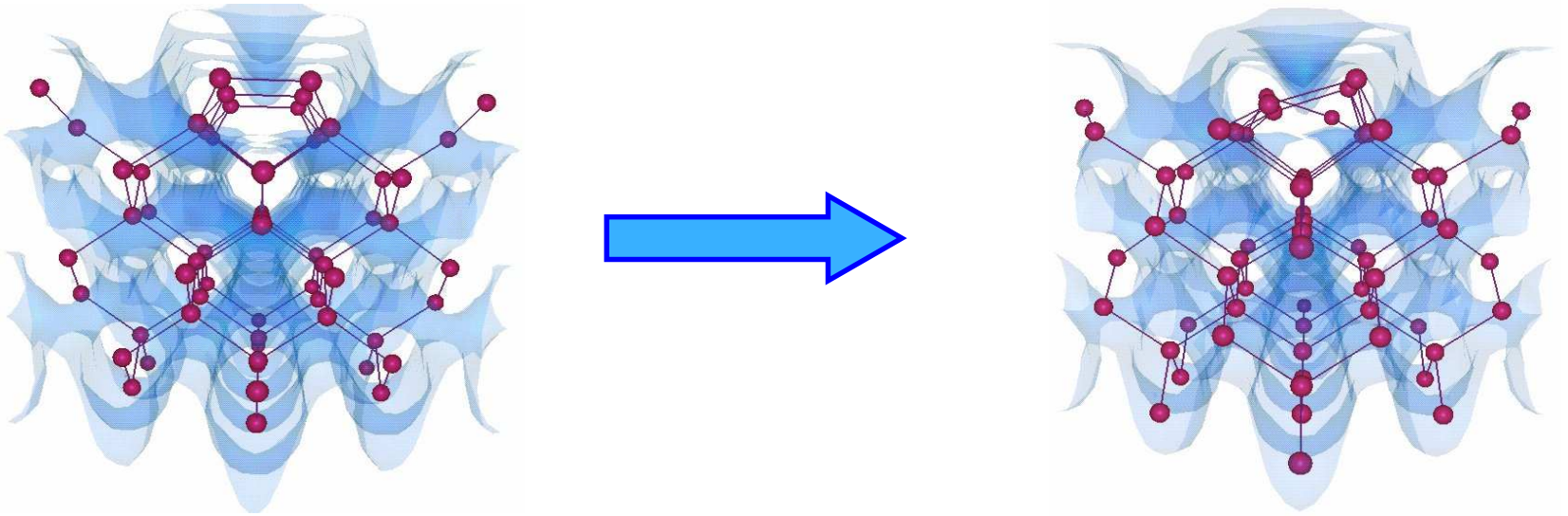


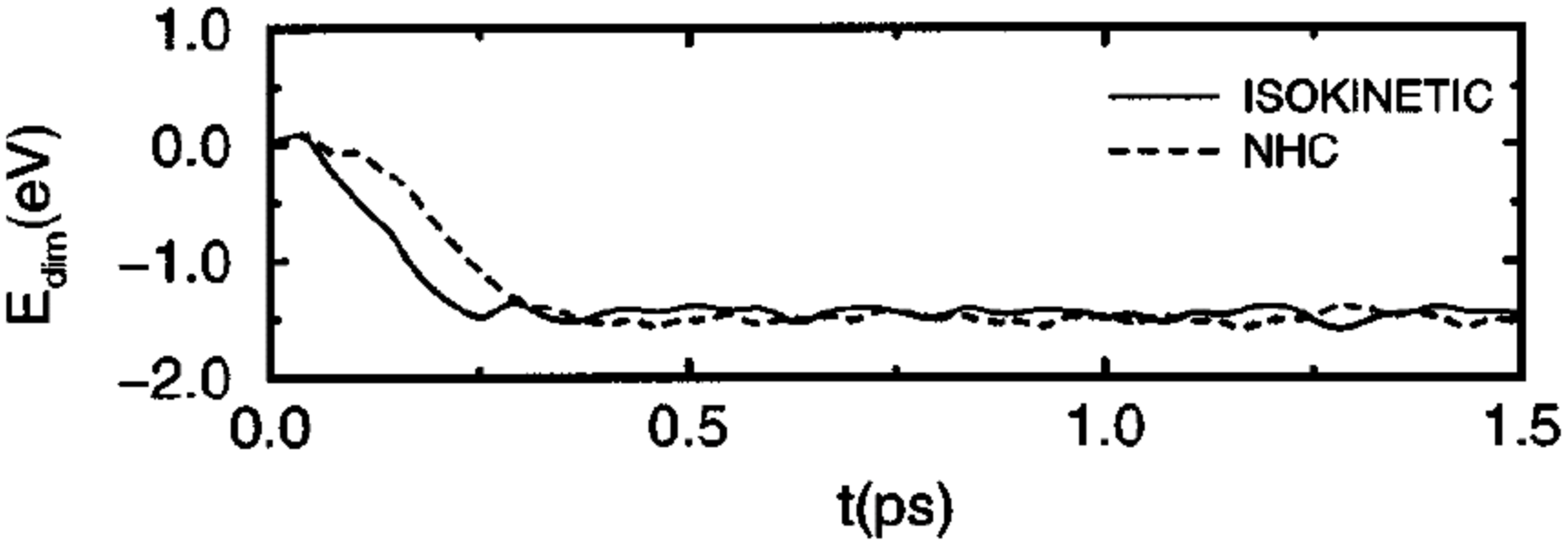




Based on the Statistical Theory of Non-Hamiltonian Systems of Martyna and Tuckerman (Euro. Phys. Lett. 2001).

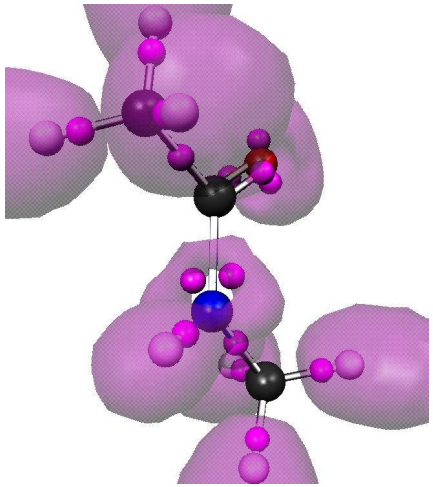
Improving ab initio MD :
2x1 Reconstruction of Si (100)



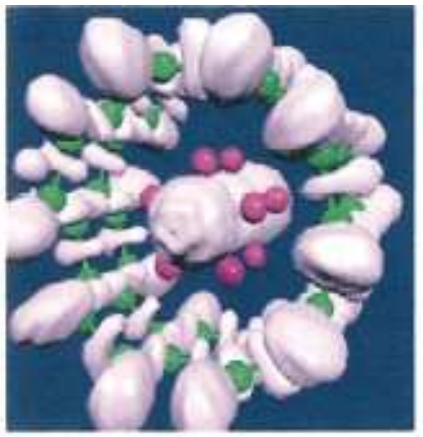


Unified Treatment of Long Range Forces: Point Particles and Continuous Charge Densities.

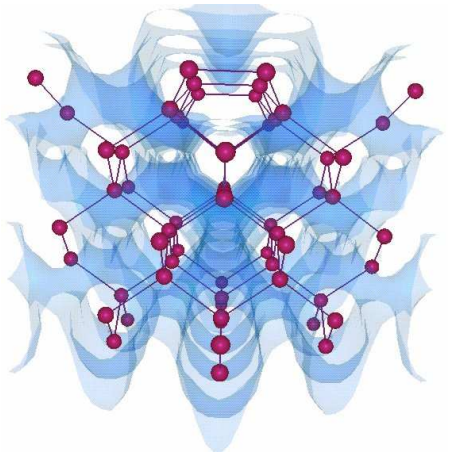
Clusters :



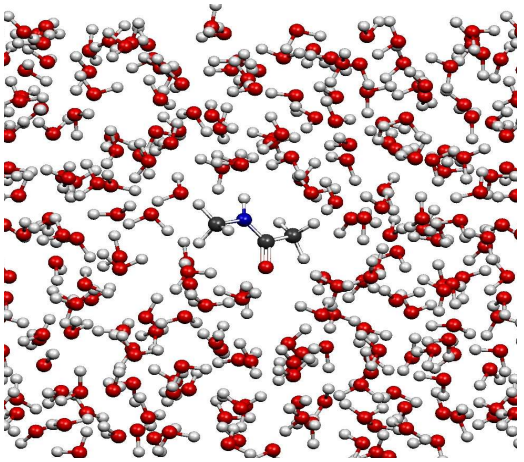
Wires:



Surfaces:



3D-Solids/Liquids:



Limitations of *ab initio* MD

(despite our efforts/improvements!)

- Limited to small systems (100-1000 atoms)*.
- Limited to short time dynamics and/or sampling times.
- Parallel scaling only achieved for

 # processors <= # electronic states

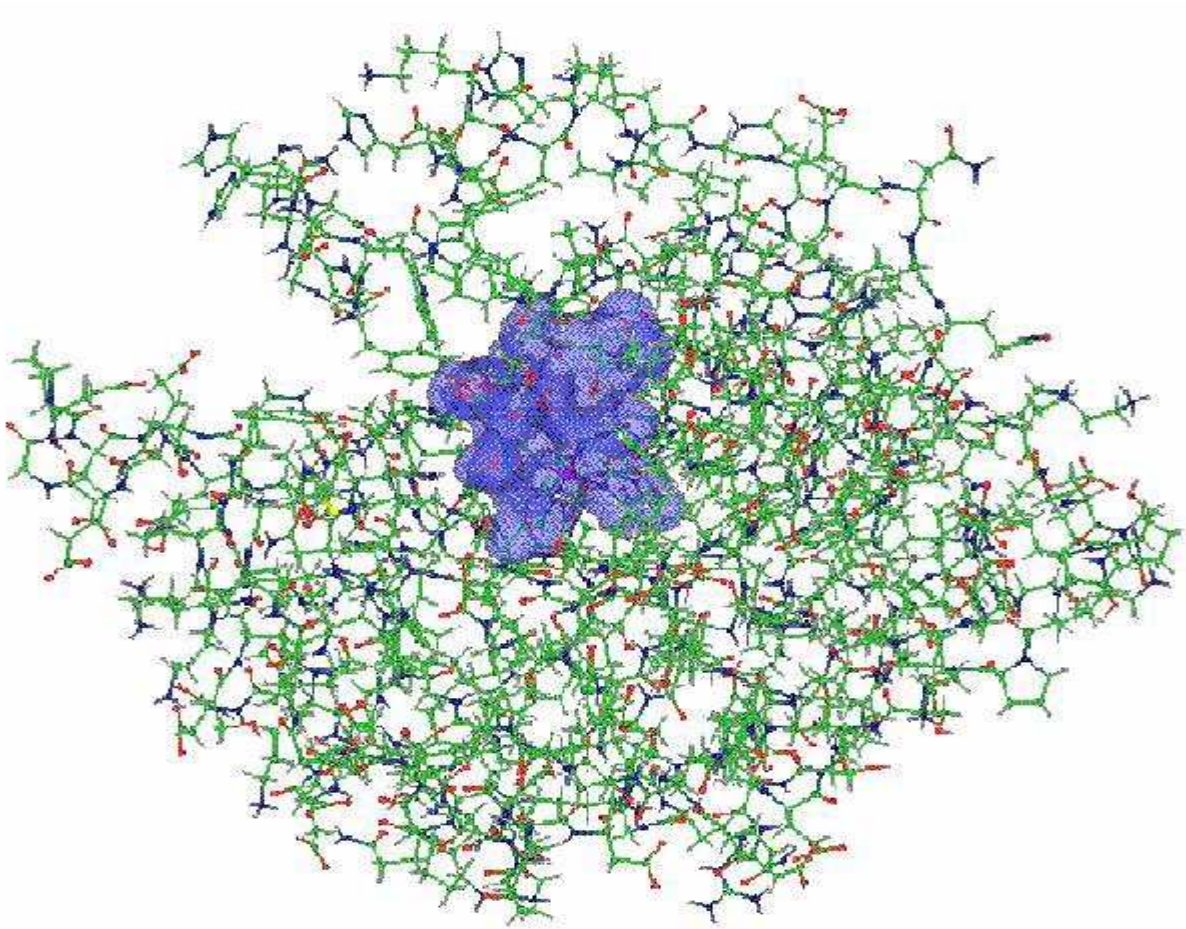
until recent efforts by ourselves and others.

*The methodology employed herein scales as $O(N^3)$ with system size due to the orthogonality constraint, only.

Solution: Fine grained Parallelization of CPAIMD.

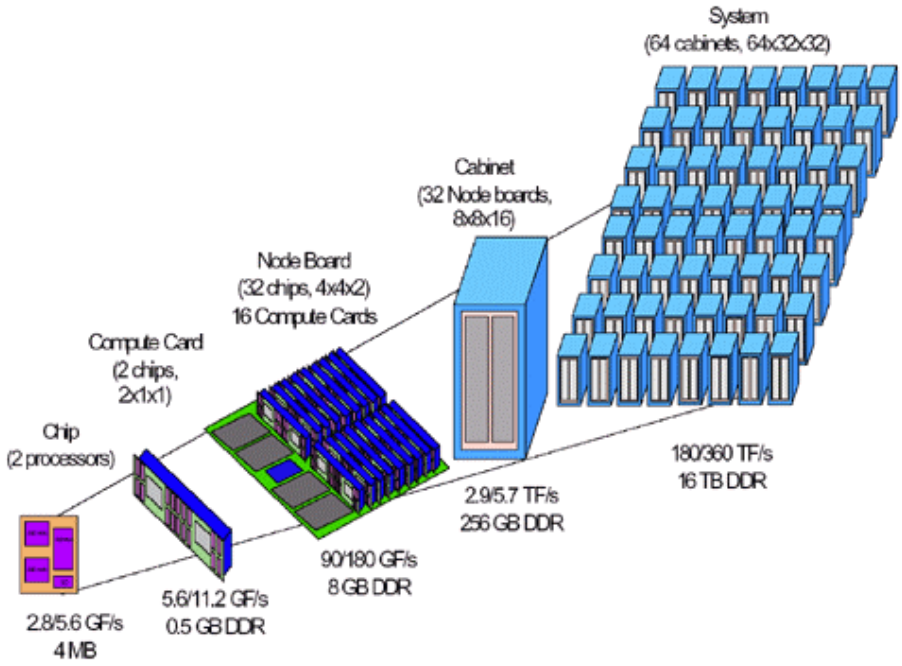
Scale small systems to 10^5 processors!!

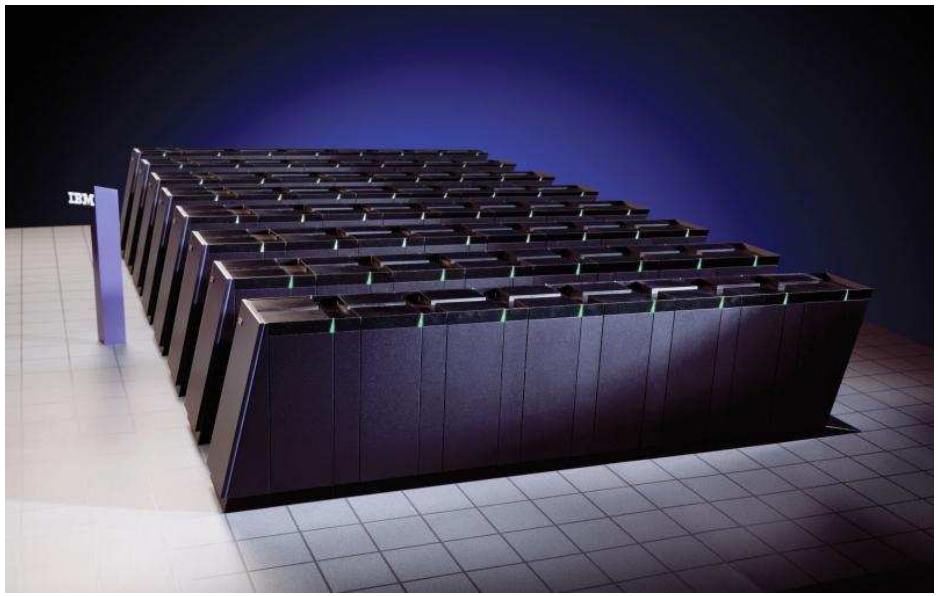
Study long time scale phenomena!!



(The charm++ QM/MM application is work in progress.)

IBM's Blue Gene/L network torus supercomputer





The worlds fastest supercomputer!

Its low power architecture requires fine grain parallel algorithms/software to achieve optimal performance.

Density Functional Theory : DFT

In the Kohn-Sham formulation of density functional theory, the electron density is expanded in a set of orbitals, $\{\psi_i(\mathbf{r})\}$,

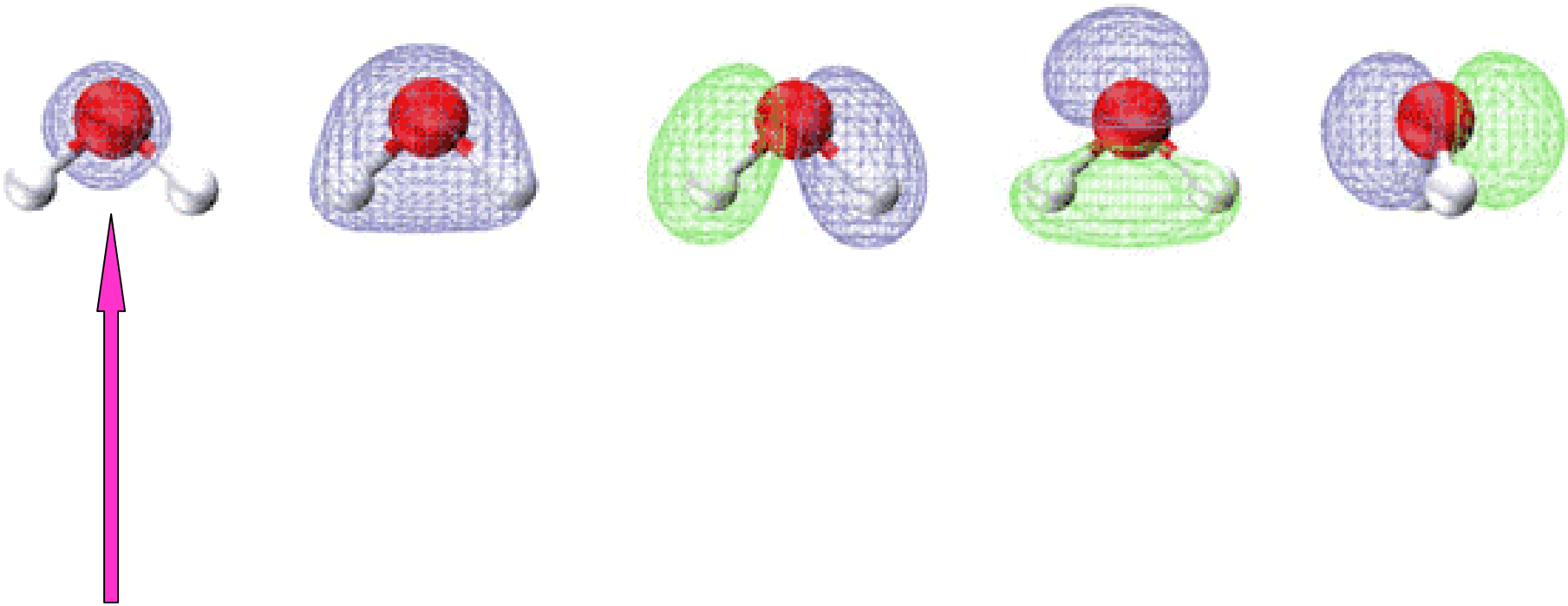
$$n(\mathbf{r}) = \sum_{i=1}^{n_{\mathrm{occ}}} f_i |\psi_i(\mathbf{r})|^2$$

subject to the orthogonality constraint, $(\langle\psi_j|\psi_k\rangle = \delta_{jk})$, where n_{occ} is the number of occupied orbitals and the f_i are the occupation numbers.

The energy functional is given by

$$E[n] = T_s[\{\psi_i\}] + E_{\mathrm{H}}[n] + E_{\mathrm{ext,loc}}[n] + E_{\mathrm{ext,non-loc}}[\{\psi_i\}] + E_{\mathrm{xc}}[n]$$

Electronic states/orbitals of water



Removed by introducing a non-local electron-ion interaction.

Plane Wave Basis Set:

In plane-wave based calculations at the Γ-point, the orbitals and, hence, the density are expanded

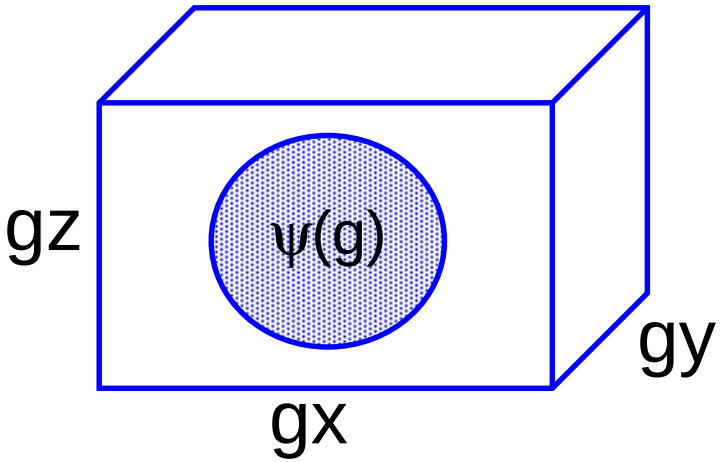
$$\psi_i(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{g}} \bar{\psi}_i(\mathbf{g}) \exp(i\mathbf{g}\cdot\mathbf{r})$$

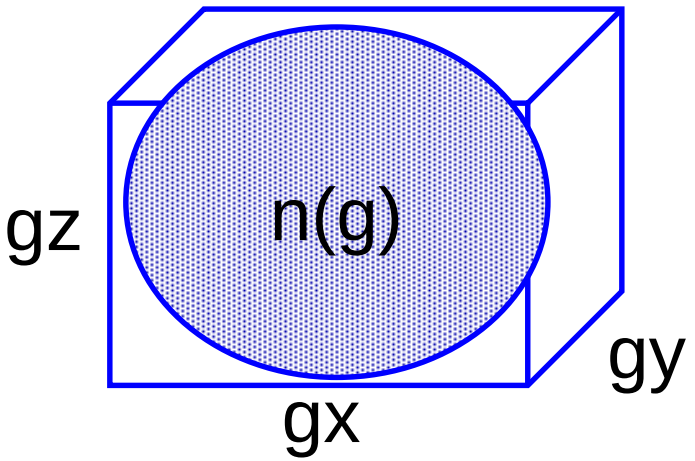
$$n(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{g}} \bar{n}(\mathbf{g}) \exp(i\mathbf{g}\cdot\mathbf{r})$$

Here, $\hat{\mathbf{g}}$ is the vector of integers $\{j,k,m\}$, $\mathbf{g} = 2\pi\hat{\mathbf{g}}\mathbf{h}^{-1}$, $V = \det \mathbf{h}$ is the volume, $\mathbf{h}$ is the matrix whose columns are the Cartesian components of the three vectors describing the parallelepiped enscribing the system and $\mathbf{r} = \mathbf{hs}$.

Plane Wave Basis Set:

Two Spherical cutoffs in G-space





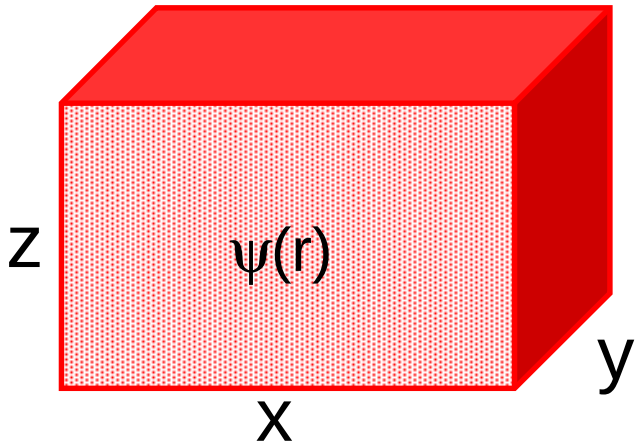
$\psi(g)$: radius g_{cut}

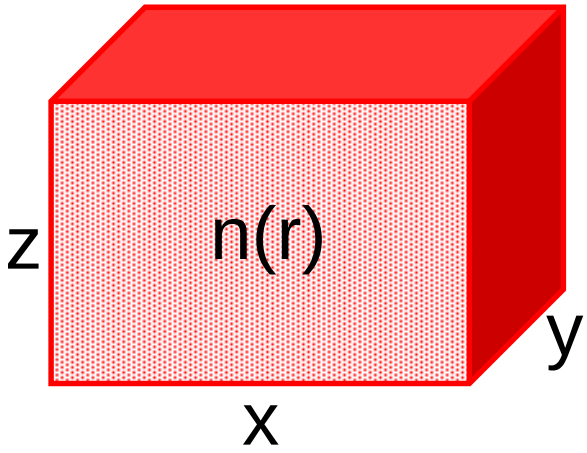
$n(g)$: radius $2g_{cut}$

g-space is a discrete regular grid due to finite size of sys

Plane Wave Basis Set:

The dense discrete real space mesh.





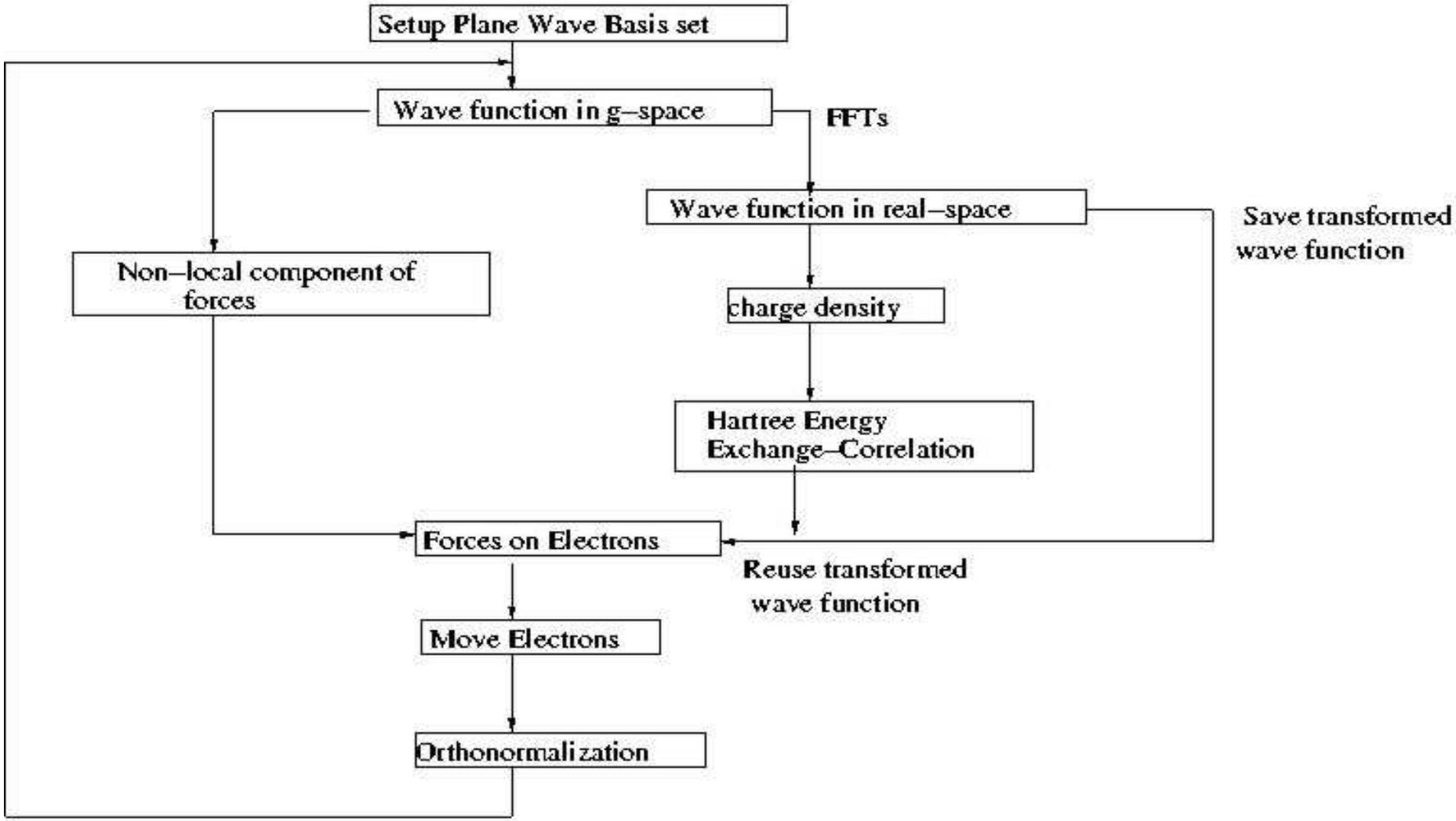
$\psi(r) = \text{3D-FFT}\{ \psi(g)\}$

$n(r) = \Sigma_k |\psi_k(r)|^2$

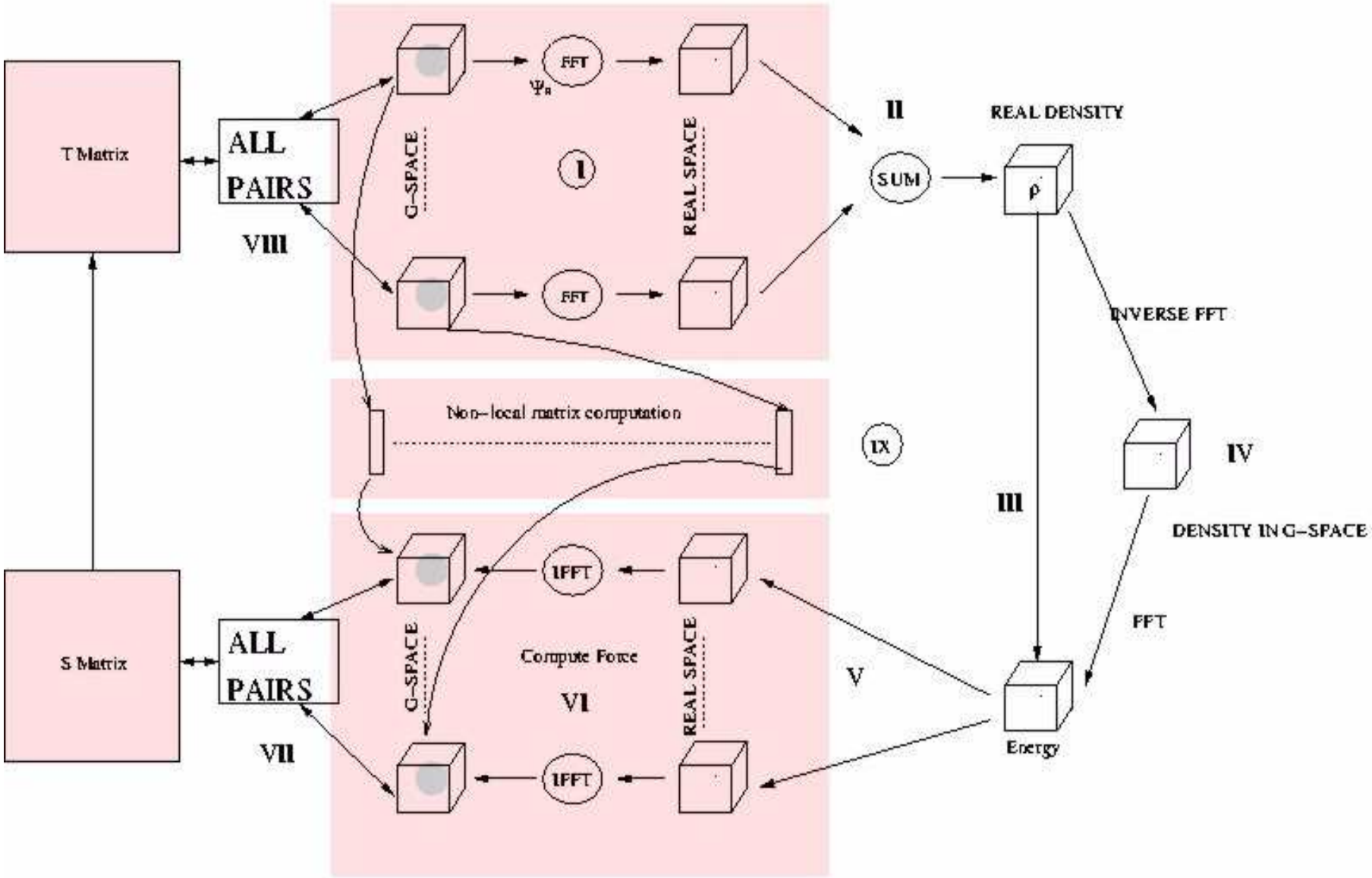
$n(g) = \text{3D-IFFT}\{n(r)\}$ exactly

Although r-space is discrete dense mesh, n(g) is generated exactly!

Simple Flow Chart : Scalar Ops



Flow Chart : Data Structures



Effective Parallel Strategy:

- The problem must be finely discretized.
- The discretizations must be deftly chosen to
  - Minimize the communication between processors.
  - Maximize the computational load on the processors.

NOTE , PROCESSOR AND DISCRETIZATION ARE

SEPARATE CONCEPTS!!!

Ineffective Parallel Strategy

- The discretization size is controlled by the number of physical processors.
- The size of data to be communicated at a given step is controlled by the number of physical processors.
- For the above paradigm :
  - Parallel scaling is limited to # processors=coarse grained parameter in the model.

THIS APPROACH IS TOO LIMITED TO ACHIEVE FINE GRAINED PARALLEL SCALING

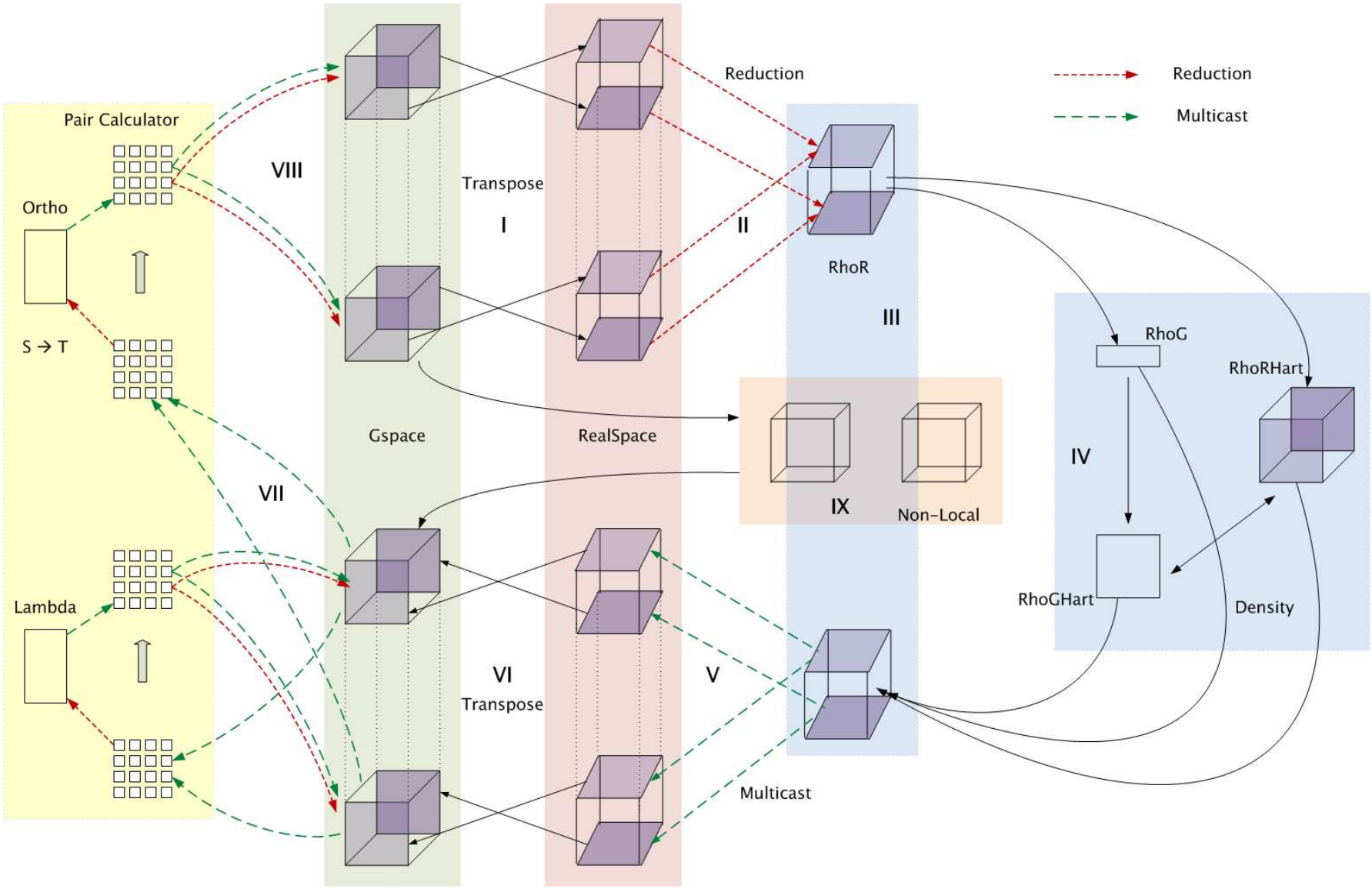
Virtualization and Charm++

- Discretize the problem into a large number of very fine grained parts.
- Each discretization is associated with some amount of computational work and communication.
- Each discretization is assigned to a light weight thread or a ``virtual processor'' (VP).
- VPs are rolled into and out of physical processors as physical processors become available.
- **The Charm++ middleware** provides the data structures and controls required to choreograph this complex dance.

Charm++ and CPAIMD

- The DFT based *ab initio* MD science modules or components are invoked by a component driver called Lean-CP written using Charm++.
- Lean-CP consists of arrays of VPs that control various aspects of the calculation.
- Anyone's ``plane wave GGA-DFT'' science modules can be plugged into Lean-CP.
- Lean-CP and the current science components, Open-Atom, will be released as Open-Source Code under the CPL.

Parallelization under charm++

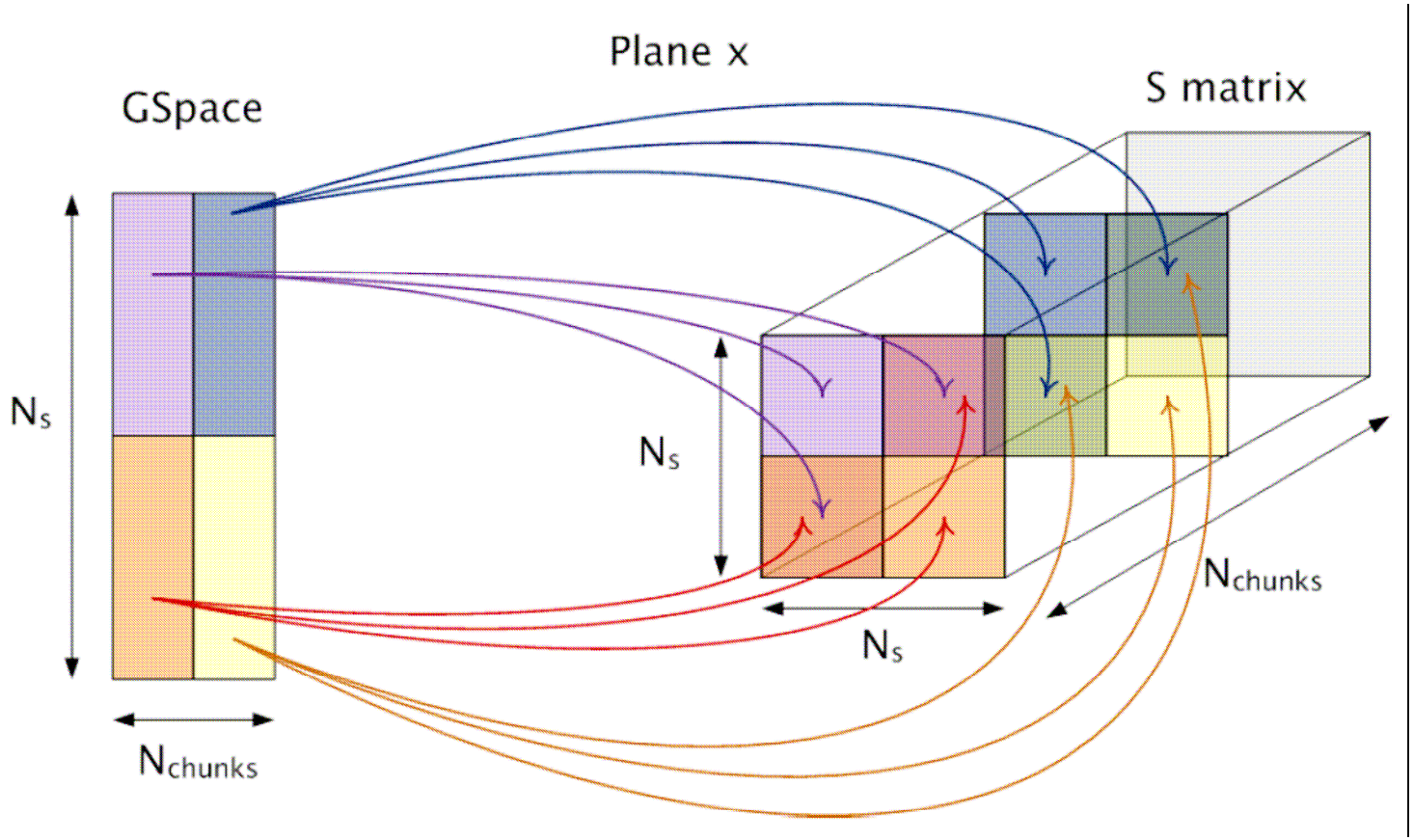


Challenges to scaling:

- Multiple concurrent 3D-FFTs to generate the states in real space require AllToAll communication patterns.
- Reduction of states (~N^2 data points) to the density (~N data points) in real space.
- Multicast of the KS potential computed from the density (~N points) back to the states in real space (~N copies).
- Applying the orthogonality constraint requires N^3 operations.
- Mapping the chare arrays/VPs to BG/L processors in a topologically aware fashion.

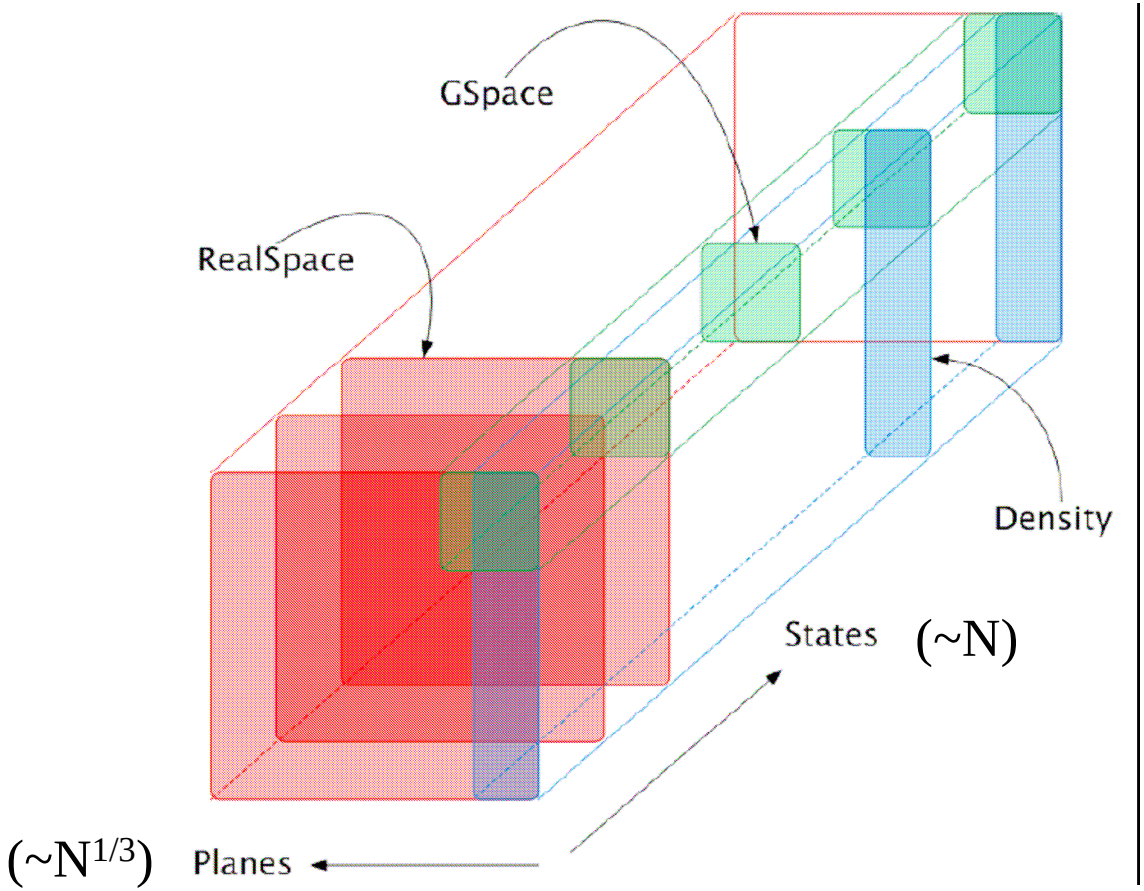
Bottlenecks due to non-local and local electron-ion interactions removed by the introduction of new methods!

The orthogonality computation



The application of the orthogonality constraint to the states requires matrix multiplications of size $(N_s \times N_g) \times (N_g \times N_s)$ to generate required overlap matrices and multiplications of size $(N_s \times N_s) \times (N_s \times N_g)$ to generate forces and states on the surface of constraint.

Topologically aware mapping for CPAIMD



- The states are confined to rectangular prisms cut from the torus to minimize 3D-FFT communication.
- The density placement is optimized to reduced its 3D-FFT communication and the multicast/reduction operations.

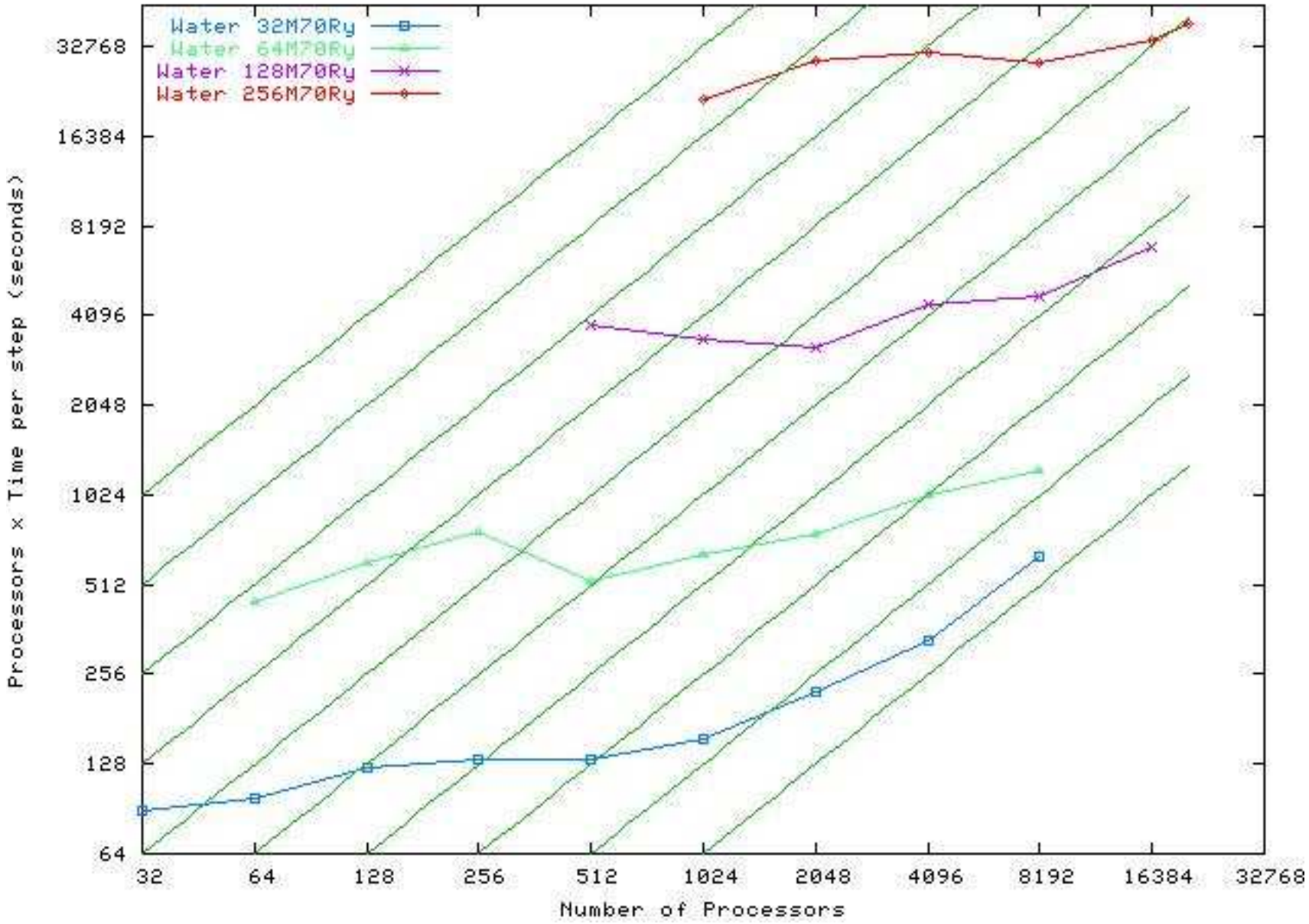
Parallel scaling of liquid water* as a function of system size on the Blue Gene/L installation at YKT:

CO Mode Native Layer with Optimizations											
Nodes	32	64	128	256	512	1024	2048	4096	8192	16384	20480
Processors	32	64	128	256	512	1024	2048	4096	8192	16384	20480
W8 Time s/step	0.22	0.10	0.082	0.071	0.046	0.026	0.020				
W16 Time s/step	0.73	0.40	0.23	0.15	0.106	0.061	0.041	0.035			
W32 Time s/step	2.71	1.52	0.95	0.44	0.26	0.15	0.11	0.081	0.063		
W64 Time s/step		6.72	3.77	1.88	0.87	0.51	0.31	0.21	0.15		
W128 Time s/step					7.4	3.31	1.57	1.09	0.581	0.425	
W256 Time s/step						21.1	14.3	7.64	3.54	2.09	1.90

*Liquid water has 4 states per molecule.

- Weak scaling is observed!
- Strong scaling on processor numbers up to ~60x the number of states!

Scaling Water on Blue Gene/L



Software : Summary and future work

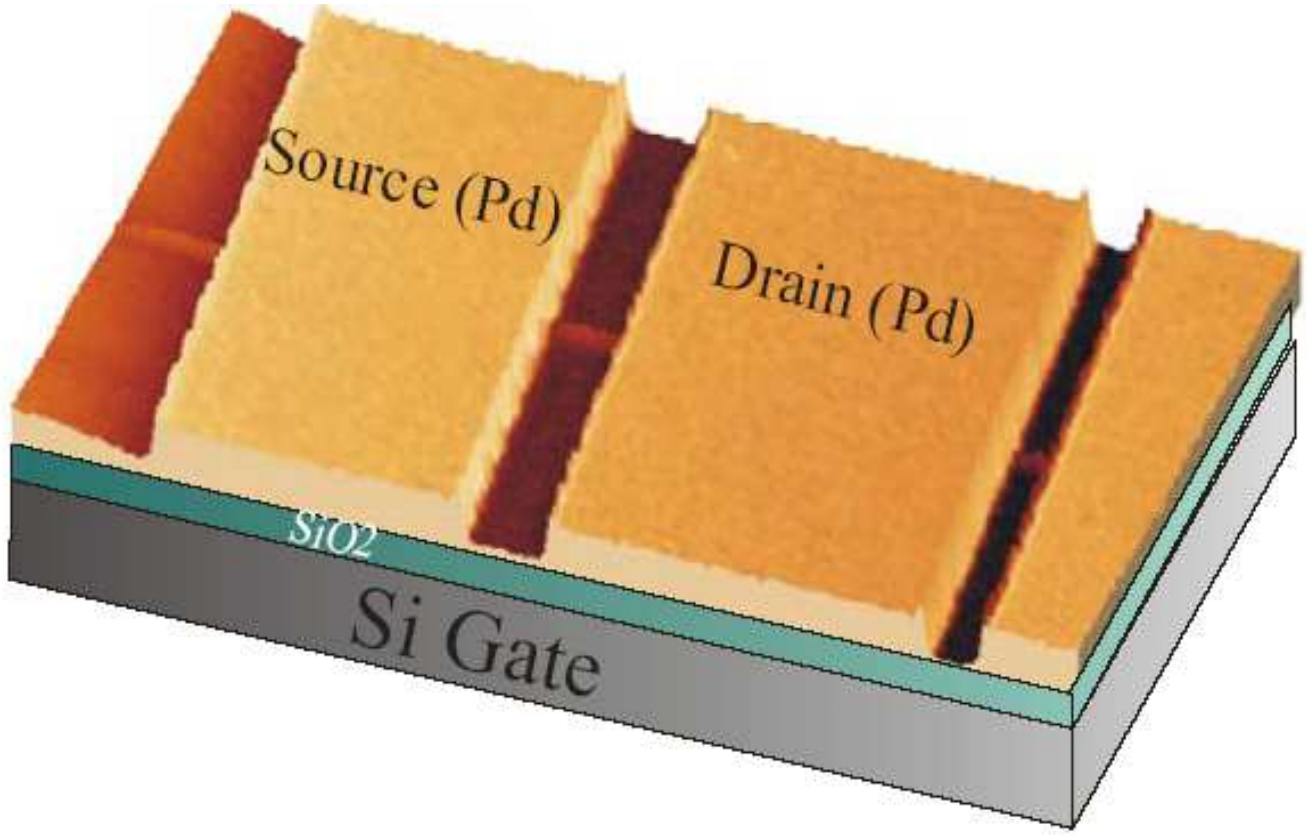
- Fine grained parallelization of the Car-Parrinello *ab initio* MD method demonstrated on thousands of processors :

  # processors >> # electronic states.

- Long time simulations of small systems are now possible on large massively parallel supercomputers.

- Future Work :
  - Utilize software to perform computations on BG/L.
  - Expand functionality of the Lean-CP driver.

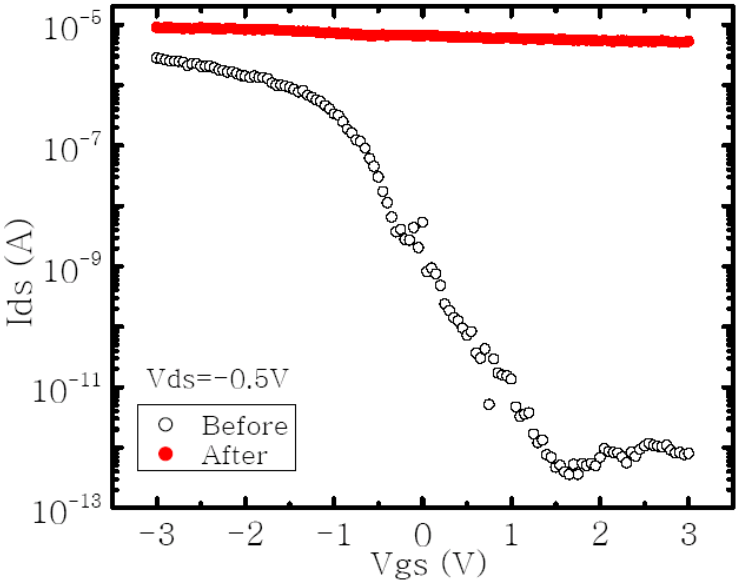
Carbon Nanotube Transistors?



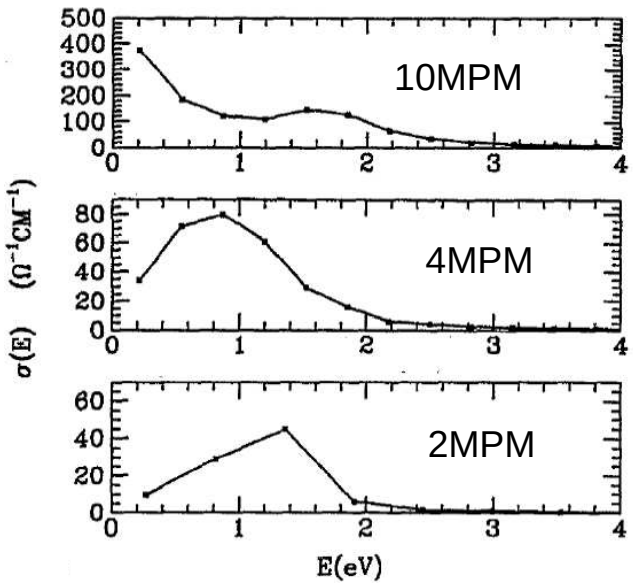
Ab initio MD studies of the Mott transition in SWCNT

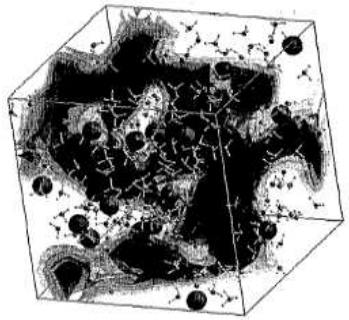
- The ambipolar behavior of SWCNT FETs is undesirable.
- Doping SWCNTs is difficult due to geometric constraints. Rydberg states are strongly bound, unlike 3D systems, necessitating specific chemical dopants.
- Excess doping of 3D semiconductors leads to a Mott transition. Is LSDA given its SIC to treat a Mott transition?
- In order to generate useful basic SWCNT science, a study of the doped SWCNT , in collaboration with D. Newns, P. Avouris and J. Chen

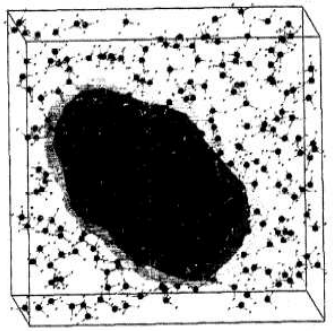
Mott Transition in SWCNT?
Chen et al (IEDM)



Metal Insulator Transition in Metal Ammonia:
Martyna et al (PRL)



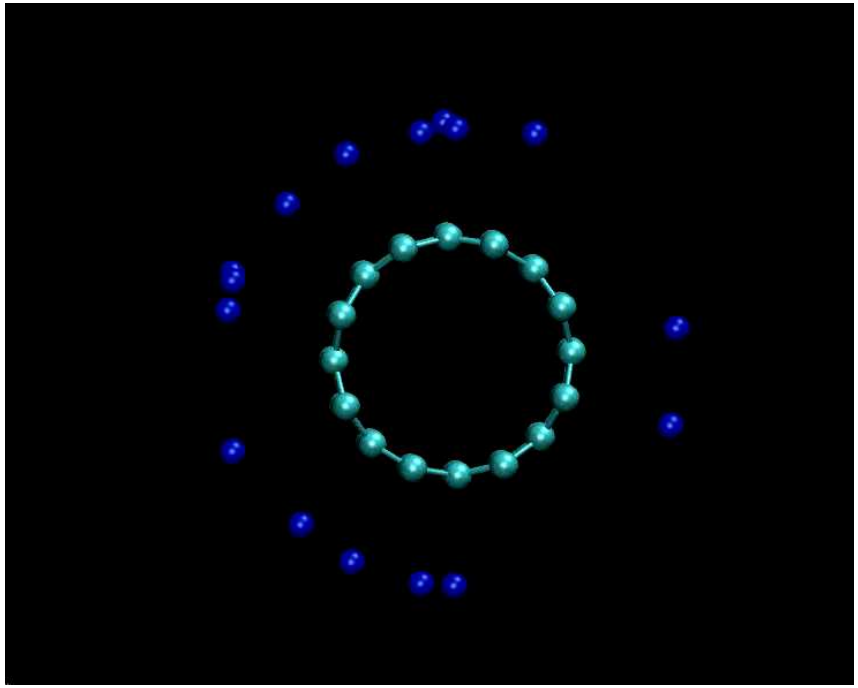


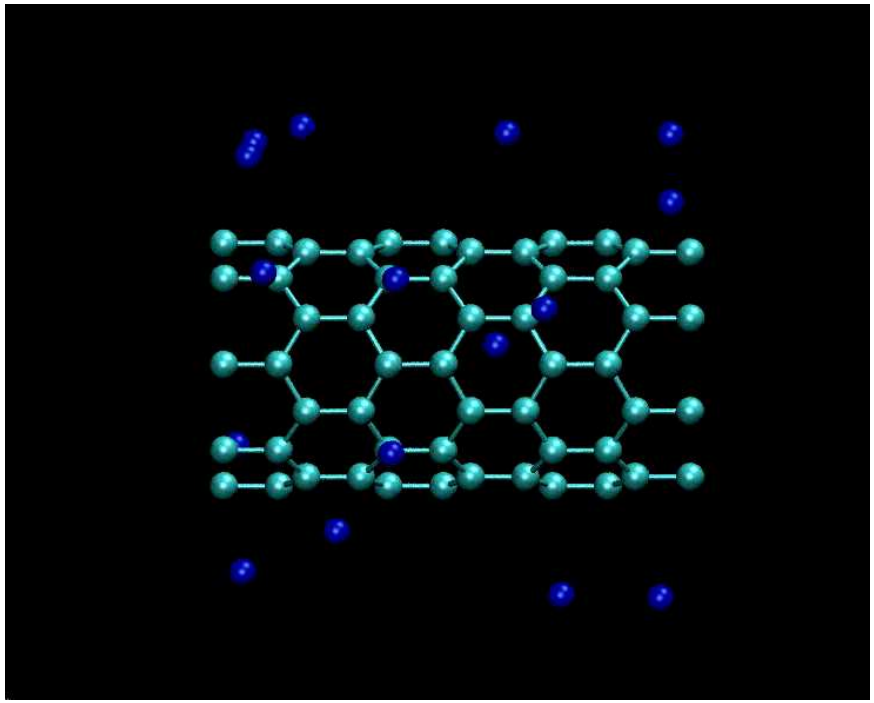


Sodium doped SWCNT

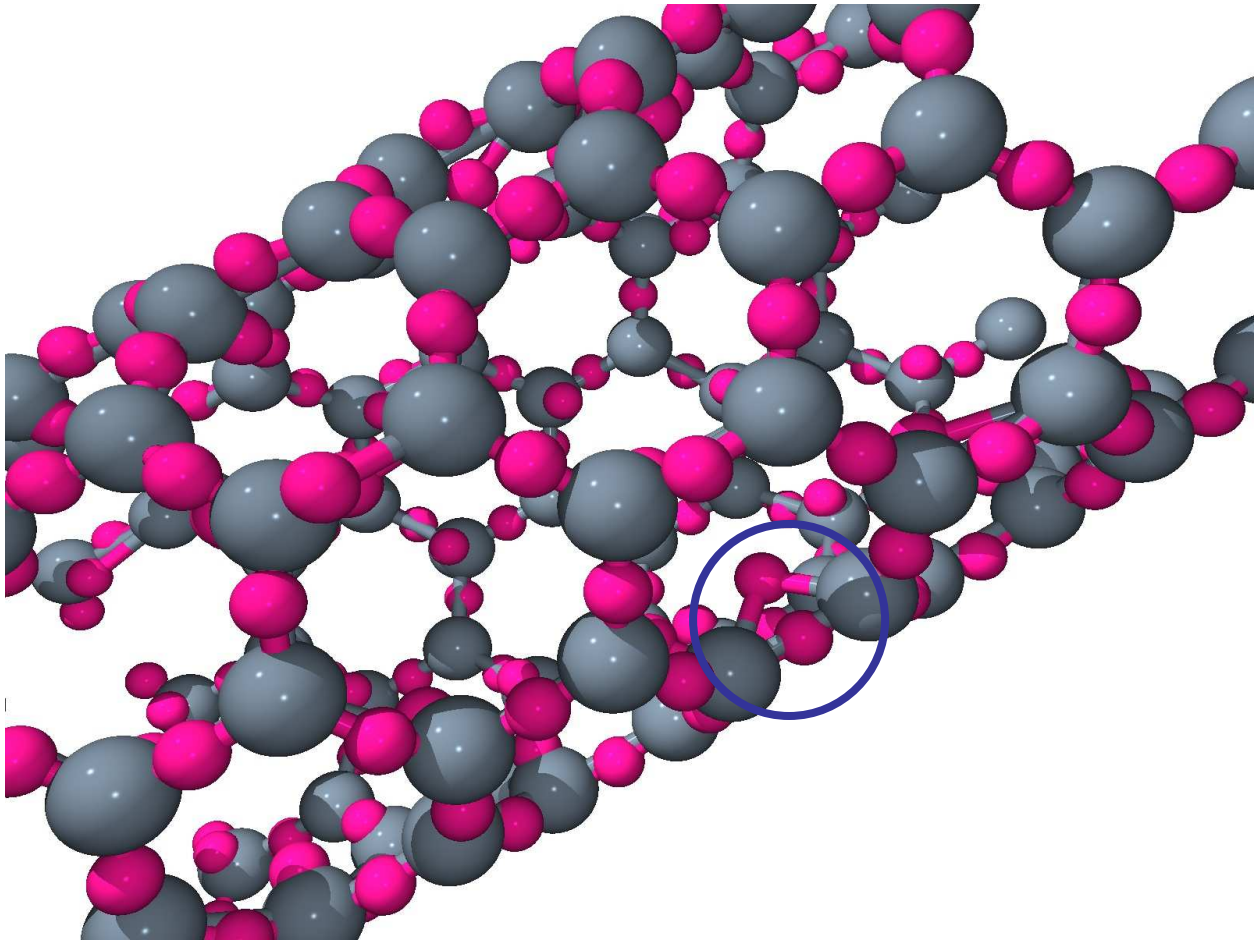
6 Carbon atoms per Sodium

T=400K





Maximally localized Wannier Function Analysis :

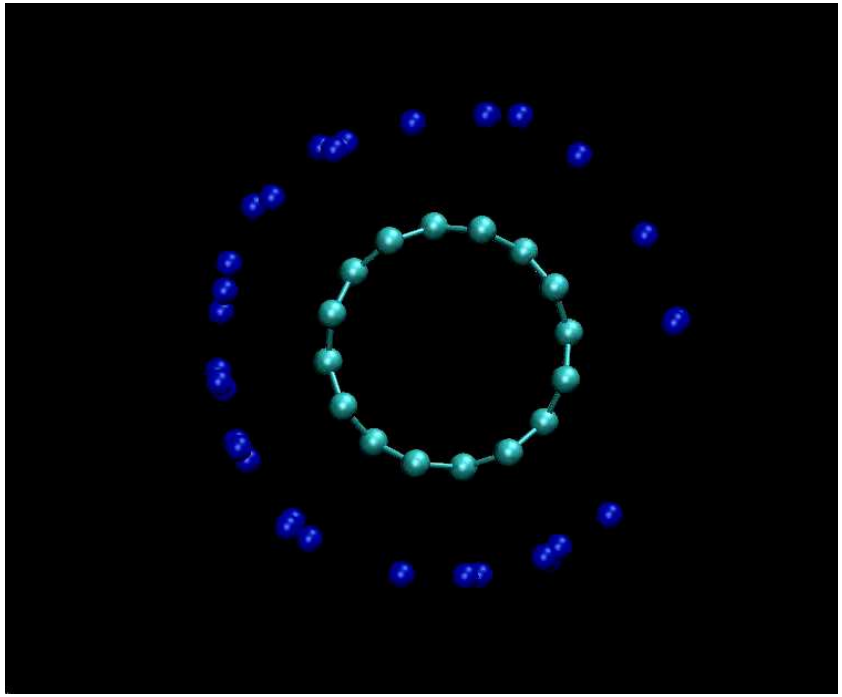


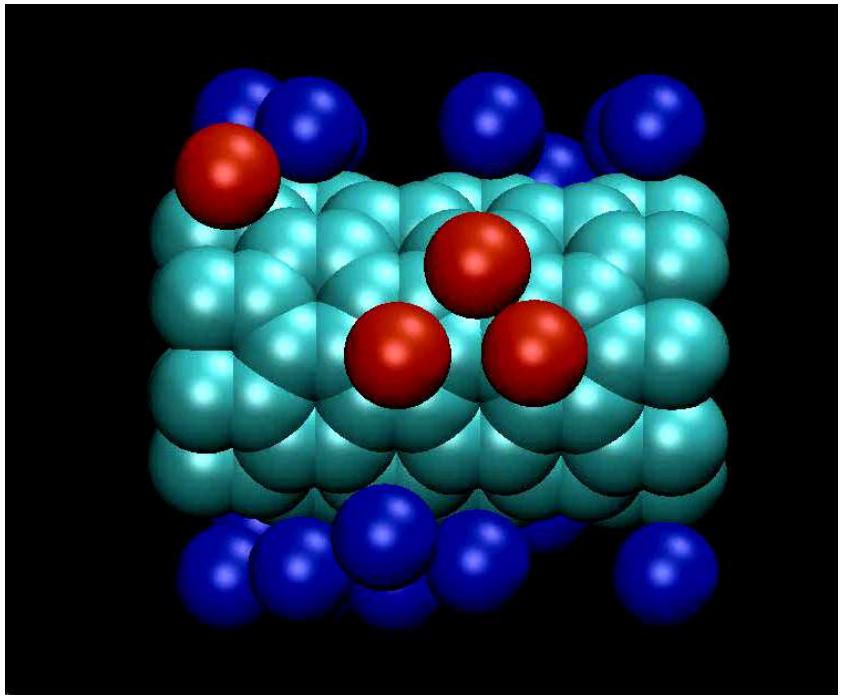
Charge Transfer from sodium atoms to the tube observed!

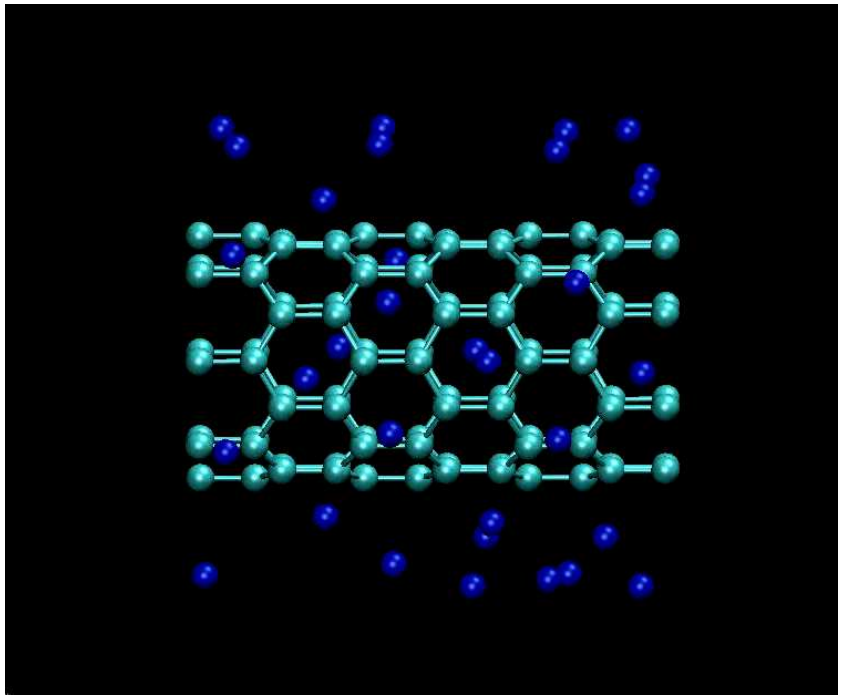
Sodium doped SWCNT

3 Carbon atoms per Sodium

T=400K

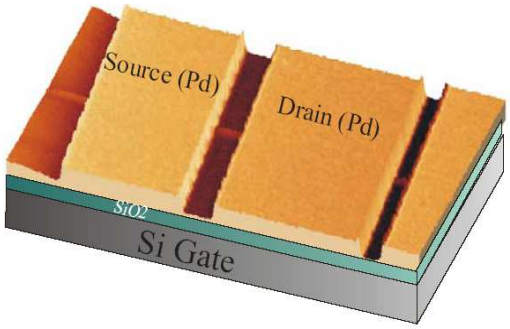


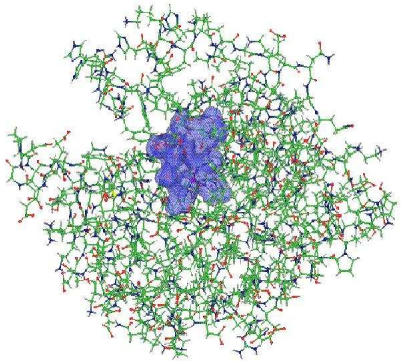




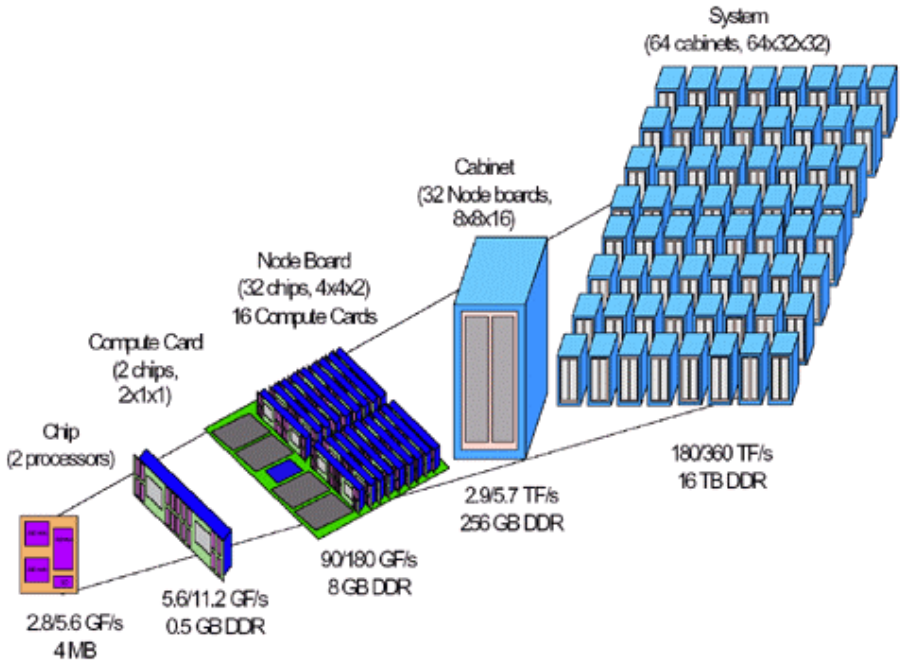
Conclusions

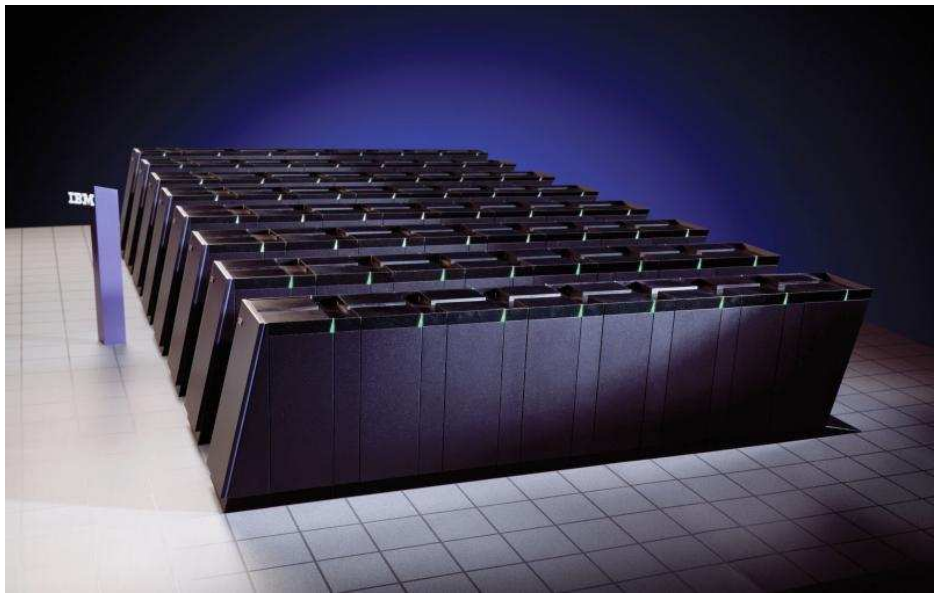
- Important physical insight can be gleaned from high quality, large scale computer simulation studies.
- The parallel algorithm development required necessitates cutting edge computer science.
- New methods must be developed hand-in-hand with new parallel paradigms.
- Using clever hardware with better methods and parallel algorithms shows great promise to impact science and technology.





IBM’s Blue Gene/L network torus supercomputer





Running on the worlds fastest supercomputer is fun!