

# Linearly Scaling Three Dimensional Fragment Method for Large Scale Electronic Structure Calculations

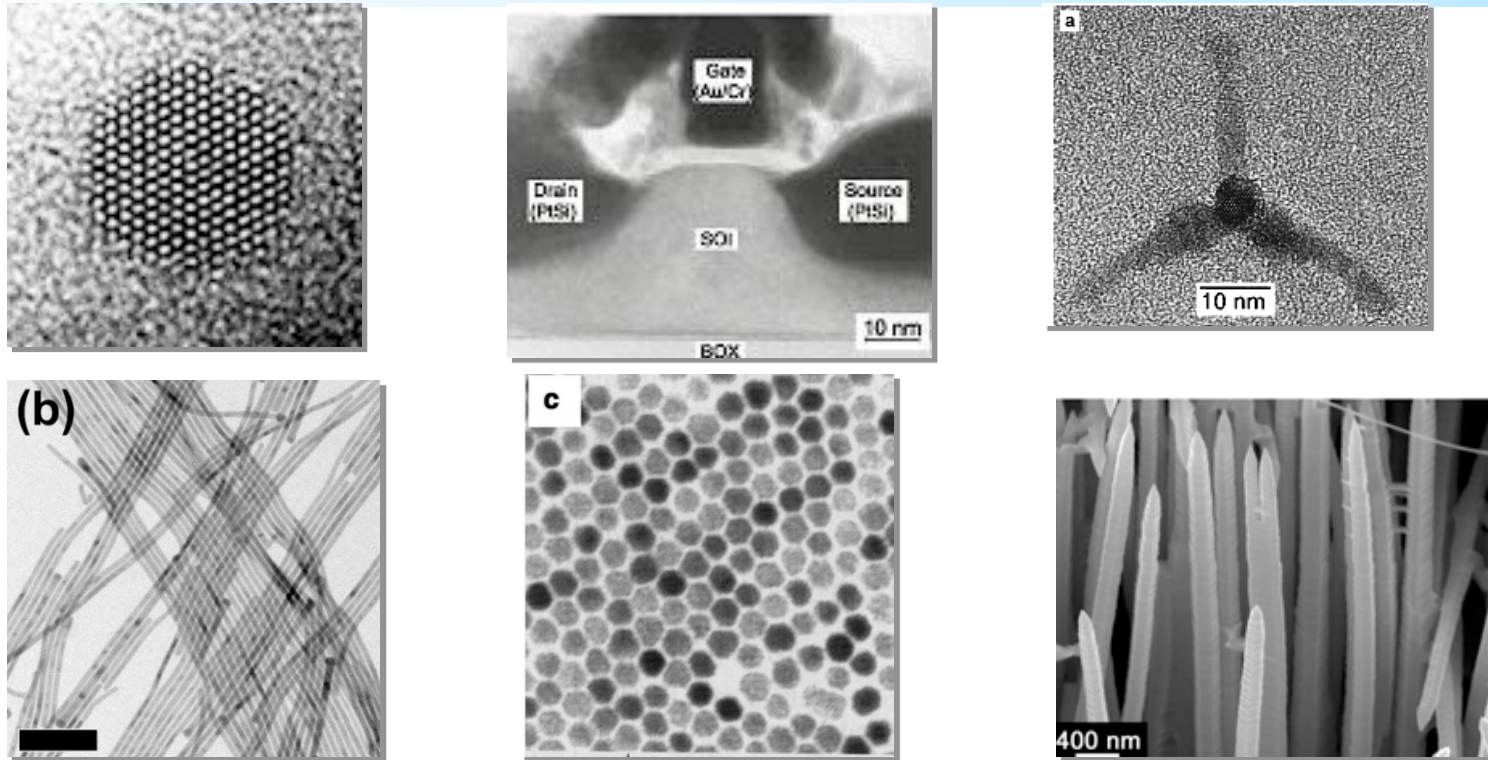
Lin-Wang Wang<sup>1,2</sup>, Byounghak Lee<sup>1</sup>, Zhengji Zhao<sup>2</sup>, Hongzhang Shan<sup>1,2</sup>, Juan Meza<sup>1</sup>, David Bailey<sup>1</sup>, Erich Strohmaier<sup>1,2</sup>

<sup>1</sup>)Computational Research Division

<sup>2</sup>)National Energy Research Scientific Computing Center (NERSC)  
Lawrence Berkeley National Laboratory

US Department of Energy, Office of Science  
Basic Energy Sciences and Advanced Scientific Computing Research

# Nanostructures have wide applications including: solar cells, biological tags, electronics devices



- ❖ Different electronic structures than bulk materials
- ❖ 1,000 ~ 100,000 atom systems are too large for direct  $O(N^3)$  *ab initio* calculations
- ❖  $O(N)$  computational methods are required
- ❖ Parallel supercomputers critical for the solution of these systems

# Why are quantum mechanical calculations so computationally expensive?

$$\left[-\frac{1}{2}\nabla^2 + V_{tot}(r)\right]\psi_i(r) = \varepsilon_i\psi_i(r)$$

- ❖ If the size of the system is  $N$ :
- ❖  $N$  coefficients to describe one wavefunction  $\psi_i(r)$
- ❖  $i = 1, \dots, M$  wavefunctions  $\psi_i(r)$ ,  $M$  is proportional to  $N$ .
- ❖ Orthogonalization:  $\int \psi_i(r) \psi_j^*(r) d^3r$ ,  $M^2$  wavefunction pairs, each with  $N$  coefficients:  $N \cdot M^2$ , i.e.  $N^3$  scaling.

**The repeated calculation of these orthogonal wavefunctions make the computation expensive,  $O(N^3)$ .**

**For large systems, an  $O(N)$  method is critical**

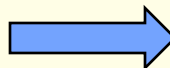
# Previous Work on Linear Scaling DFT methods

- ❖ Three main approaches:
  - Localized orbital method
  - Truncated density matrix method
  - Divide-and-conquer method
- ❖ Some current methods include:
  - Parallel SIESTA (atomic orbitals, not for large parallelization)
  - Many quantum chemistry codes (truncated D-matrix, Gaussian basis, not for large parallelization)
  - ONETEP (M. Payne, PW to local orbitals, then truncated D-matrix)
  - CONQUEST (D. Bowler, UCL, localized orbital)
- ❖ Most of these use localized orbital or truncated-D matrix
- ❖ None of them scales to tens of thousands of processors

# Linearly Scaling 3 Dimensional Fragment method (LS3DF)

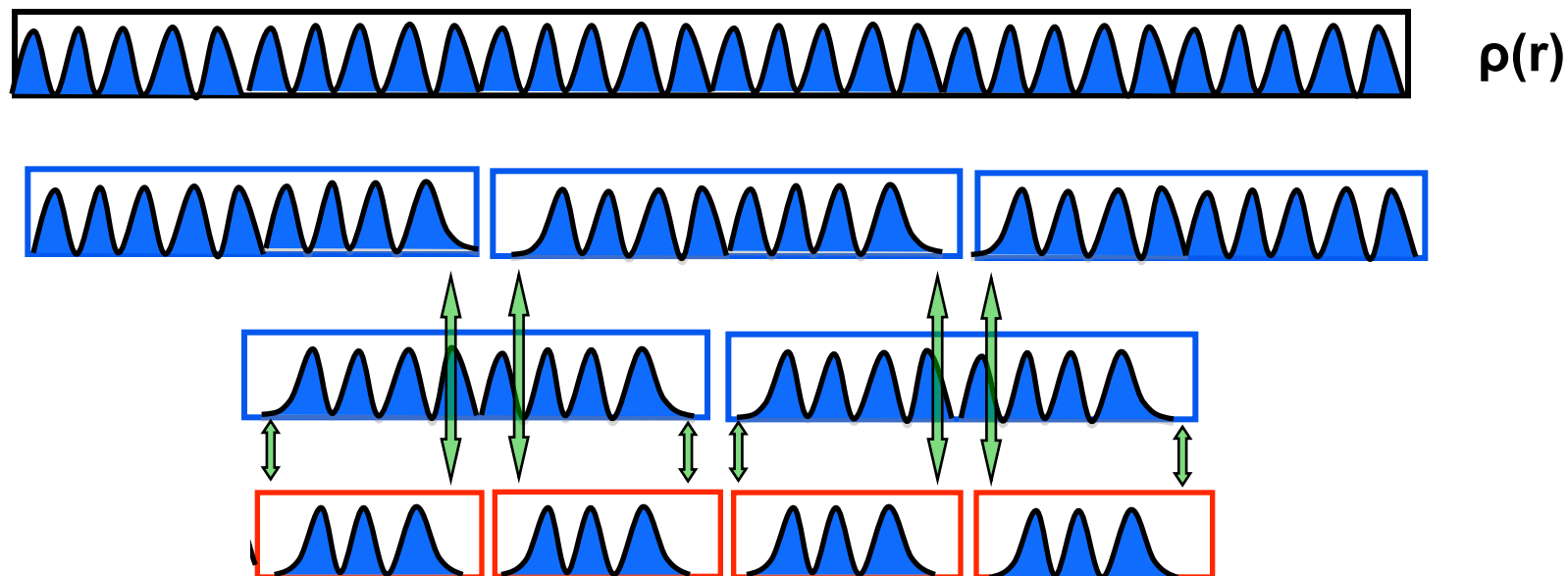
- ❖ A novel divide and conquer scheme with a new approach for patching the fragments together
- ❖ No spatial partition functions needed
- ❖ Uses overlapping positive and negative fragments
- ❖ New approach minimizes artificial boundary effects

**divide-and-conquer method**



**$O(N)$  scaling**  
**Massively parallelizable**

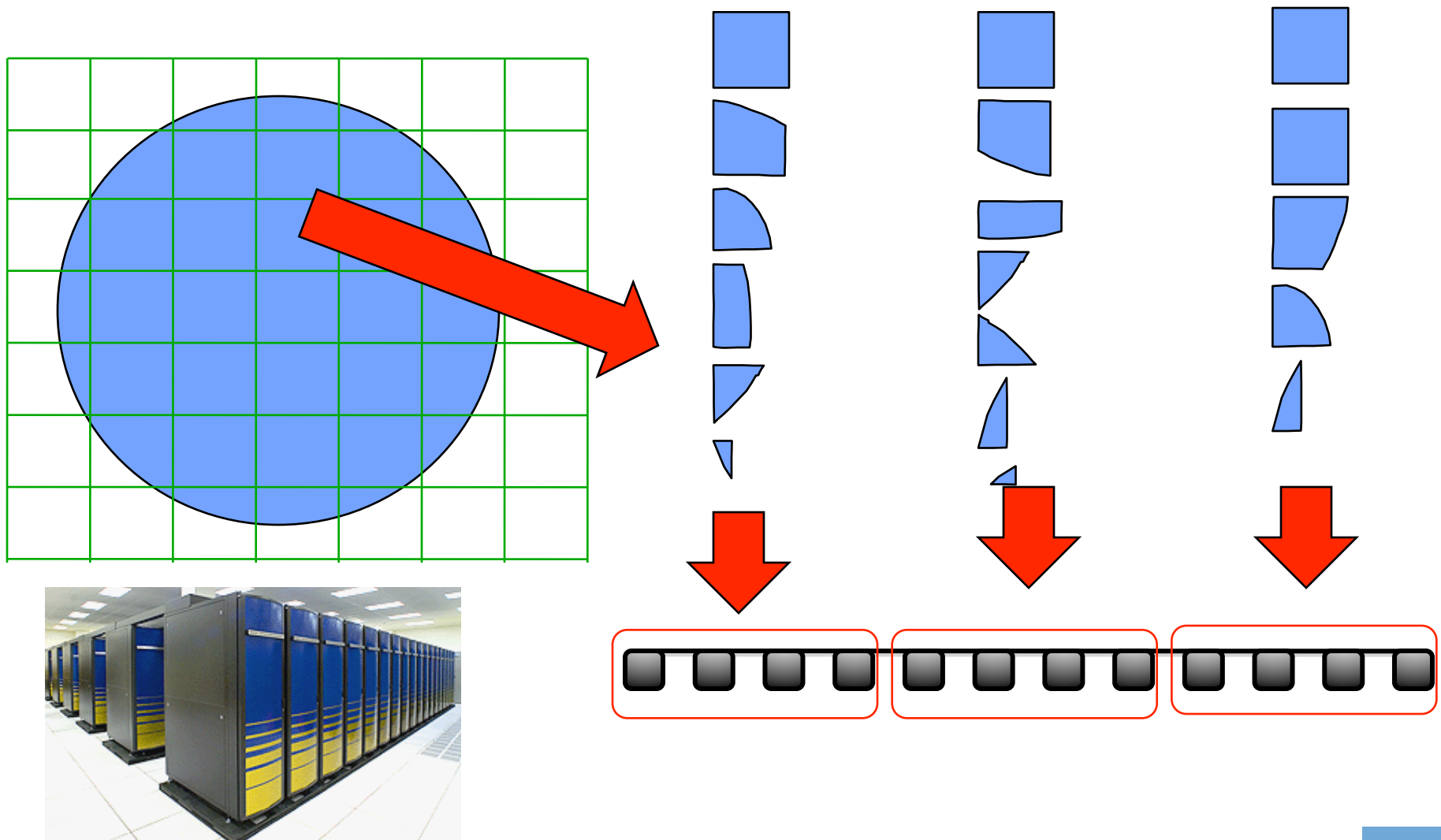
# LS3DF: 1D Example



$$\text{Total} = \sum_F \{ \boxed{\phantom{\rho(r)}}_F - \boxed{\phantom{\rho(r)}}_F \}$$

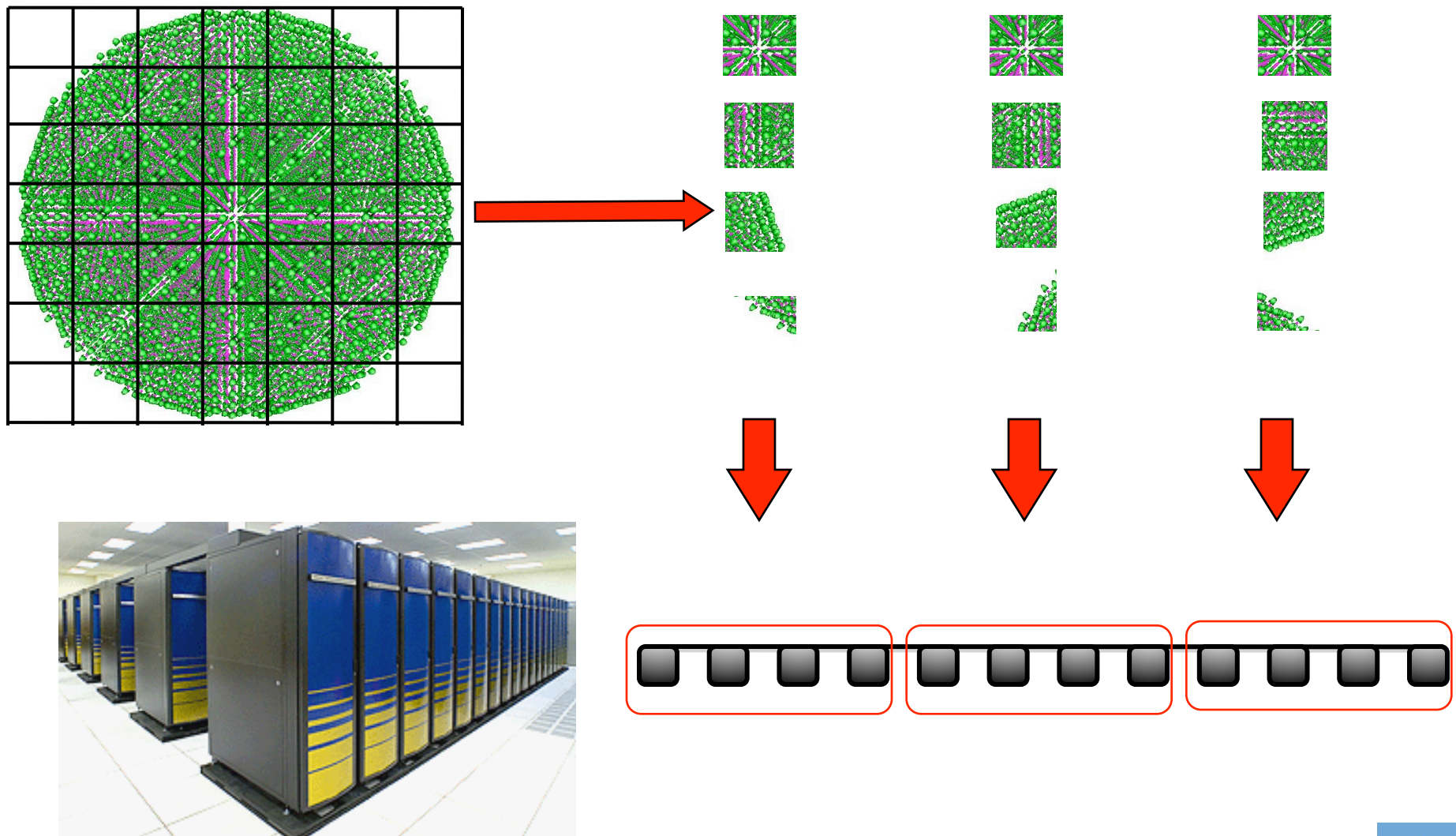
Phys. Rev. B 77, 165113 (2008); J. Phys: Cond. Matt. 20, 294203 (2008)

# Schematic for LS3DF calculation





# Schematic for LS3DF calculation





# Major components of LS3DF method

1. Generate fragment potentials  $V_F$
2. Solve for fragment wave functions
3. Compute total charge density
4. Solve global Poisson equation

Based on the plane wave PEtot code: <http://hpcrd.lbl.gov/~linwang/PEtot/PEtot.html>

# Overview of computational effort in LS3DF

- ❖ Most time consuming part of LS3DF calculation is for the fragment wavefunctions
  - Modified from the stand alone PEtot code
  - Uses planewave pseudopotential (like VASP, Qbox)
  - All-band algorithm takes advantage of BLAS3
- ❖ 2-level parallelization:
  - q-space (Fourier space)
  - band index ( $i$  in  $\psi_i(r)$ )
- ❖ PEtot efficiency > 50% for large systems (e.g, more than 500 atoms), 30-40% for our fragments.

PEtot code: <http://hpcrd.lbl.gov/~linwang/PEtot/PEtot.html>

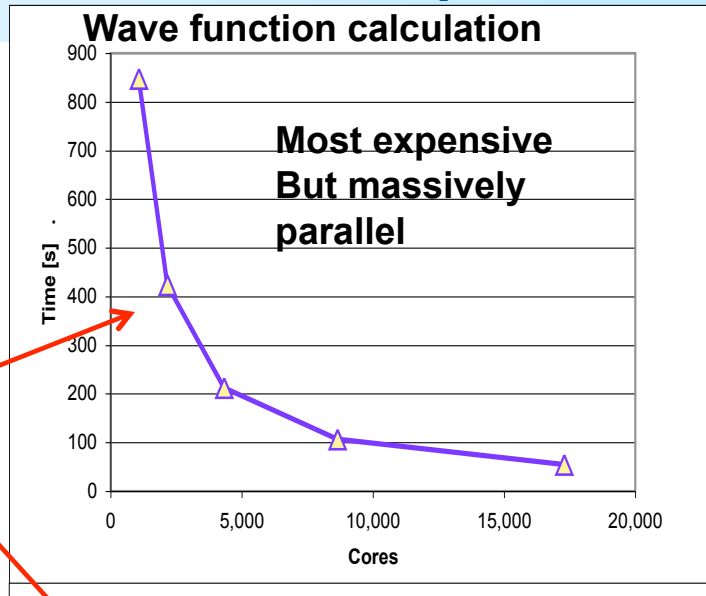
## Details on the LS3DF divide and conquer scheme

- ❖ Variational formalism, sound mathematics
- ❖ The division into fragments is done automatically, based on atom's spatial locations
- ❖ Typical large fragments (2x2x2) have ~100 atoms and the small fragments (1x1x1) have ~ 20 atoms
- ❖ Processors are divided into  $M$  groups, each with  $N_p$  processors.
  - $N_p$  is usually set to 16 – 128 cores
  - $M$  is between 100 and 10,000
- ❖ Each processor group is assigned  $N_f$  fragments, according to estimated computing times, load balance within 10%.
  - $N_f$  is typically between 8 and 100

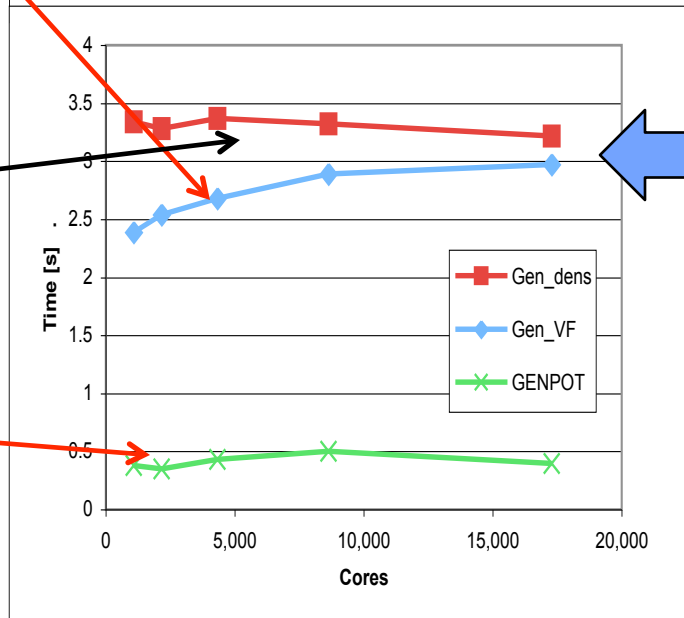
# The performance of LS3DF method (strong scaling, NERSC Franklin)

1. Generate fragment potentials  $V_F$
2. Solve for fragment wave functions
3. Compute total charge density
4. Solve global Poisson equation

Time (second)

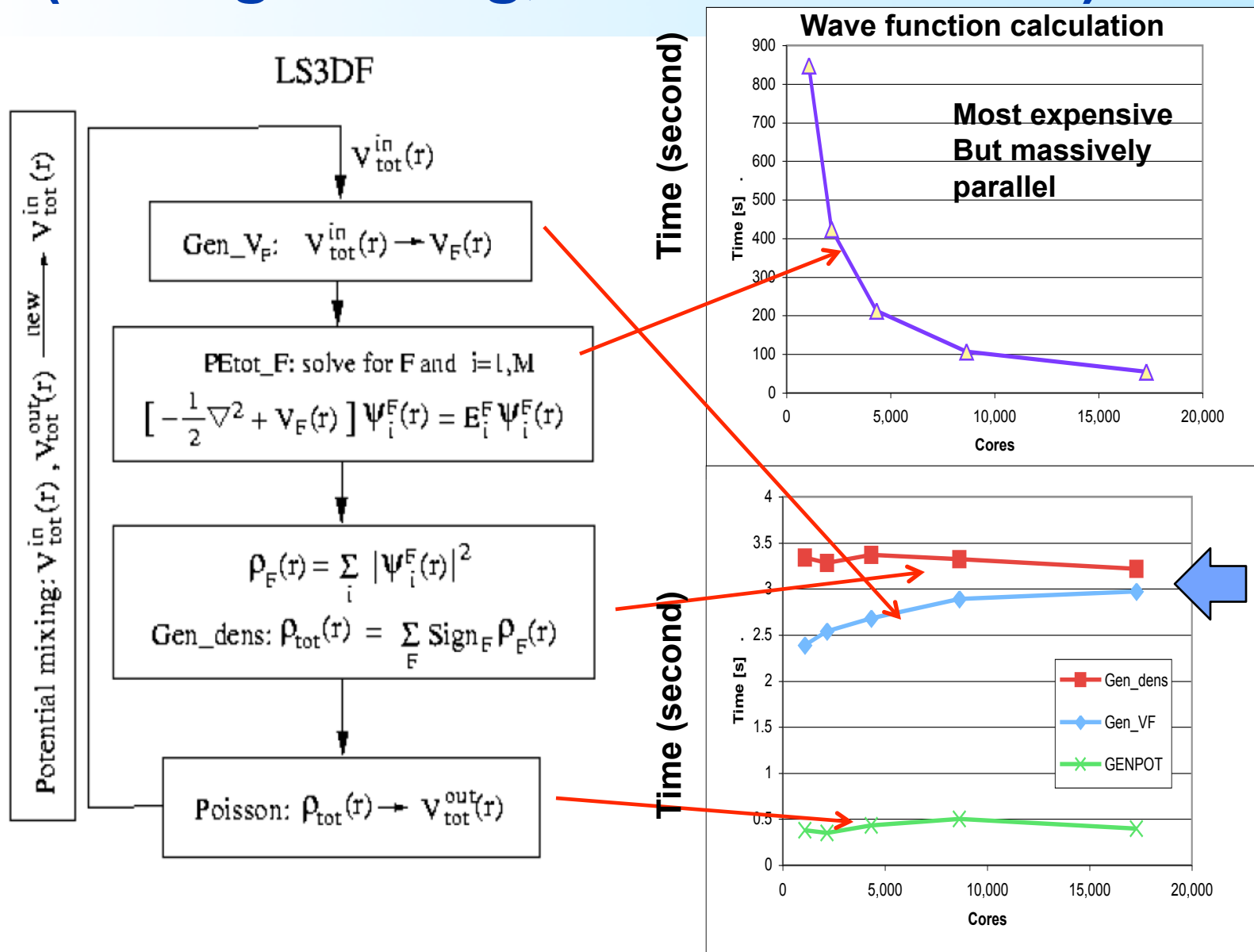


Time (second)

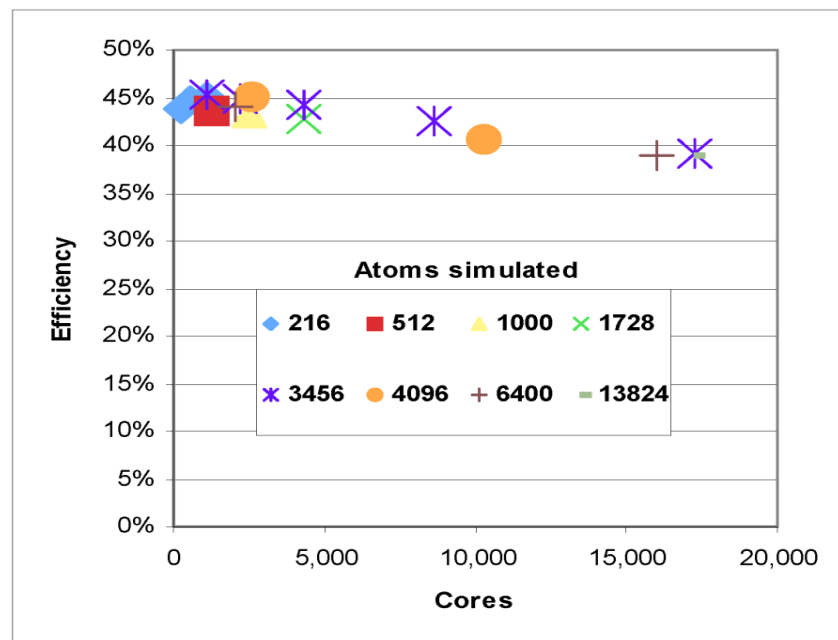
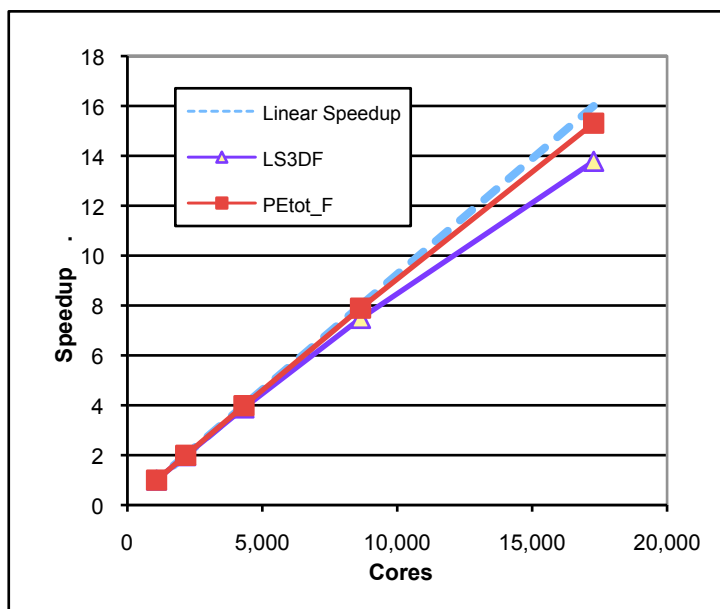


data movement

# The performance of LS3DF method (strong scaling, NERSC Franklin)



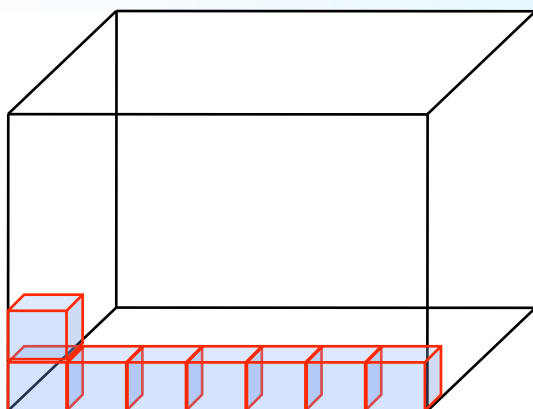
# NERSC Franklin results



- ❖ 3456 atom system, 17280 cores:
  - one min. per SCF iteration, one hour for a converged result
- ❖ 13824 atom system, 17280 cores,
  - 3-4 min. per SCF iteration, 3 hours for a converged result
- ❖ LS3DF is 400 times faster than PEtot on the 13824 atom system



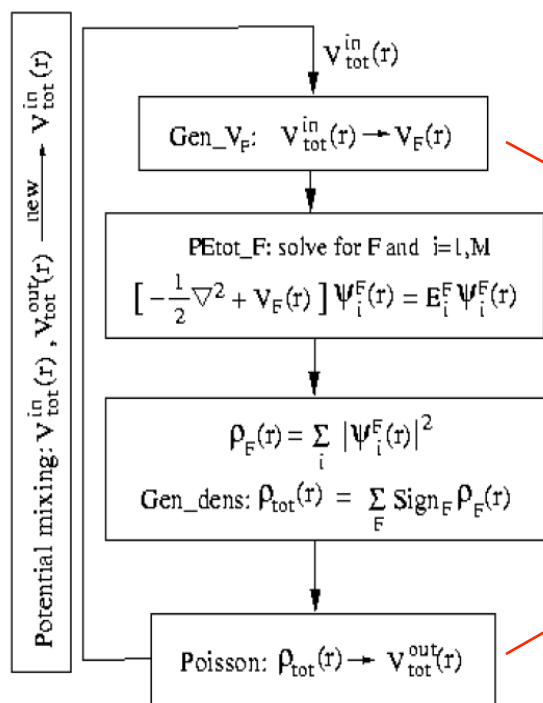
# Node mapping and performance on BlueGene/P



Map all the groups into identical compact cubes, for good intra-group FFT communication, and inter-group load balance.

Time: 50% inside group FFT  
50% inside group DGEMM

LS3DF

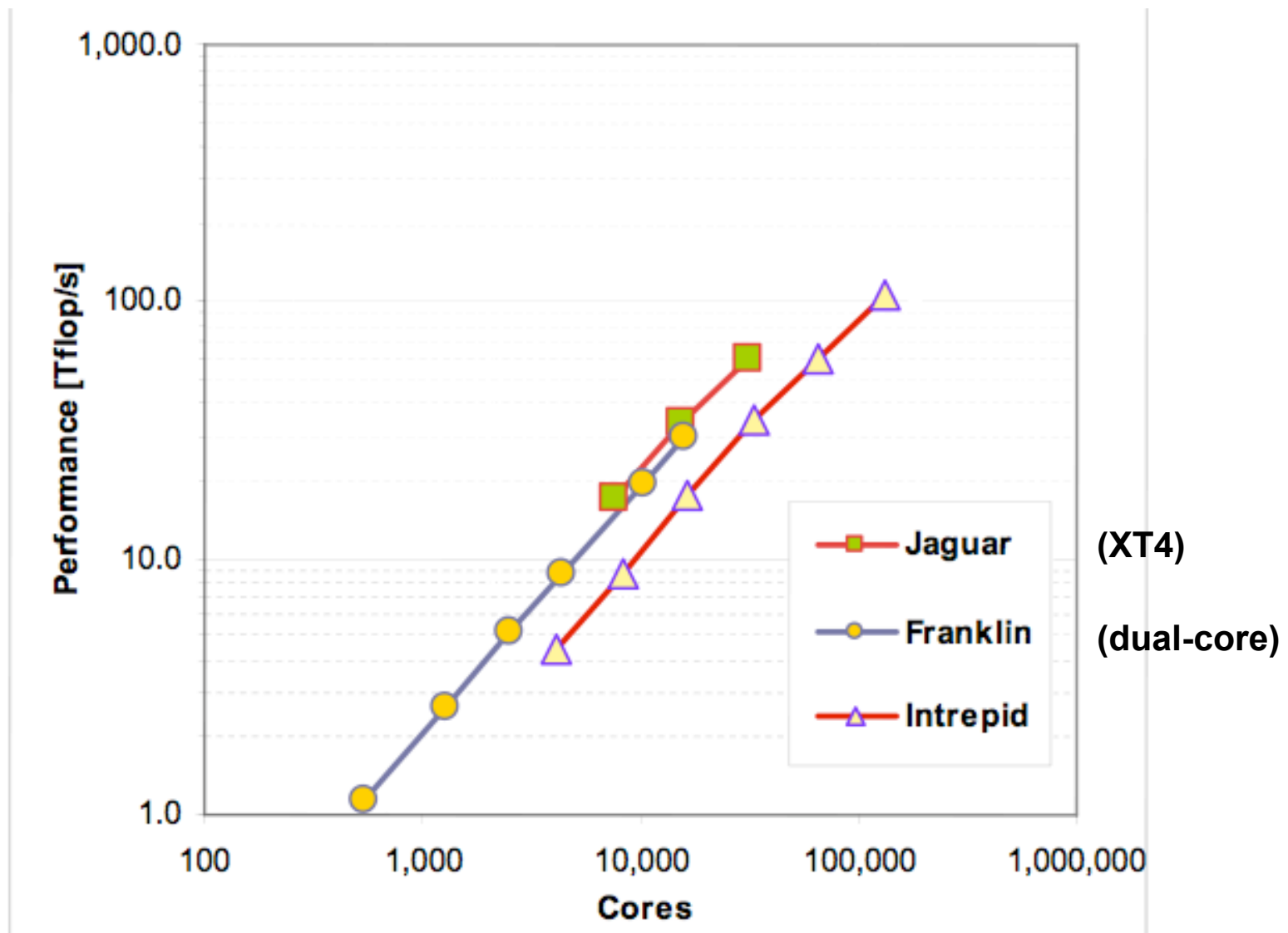


Times on diff. parts of the code (sec)

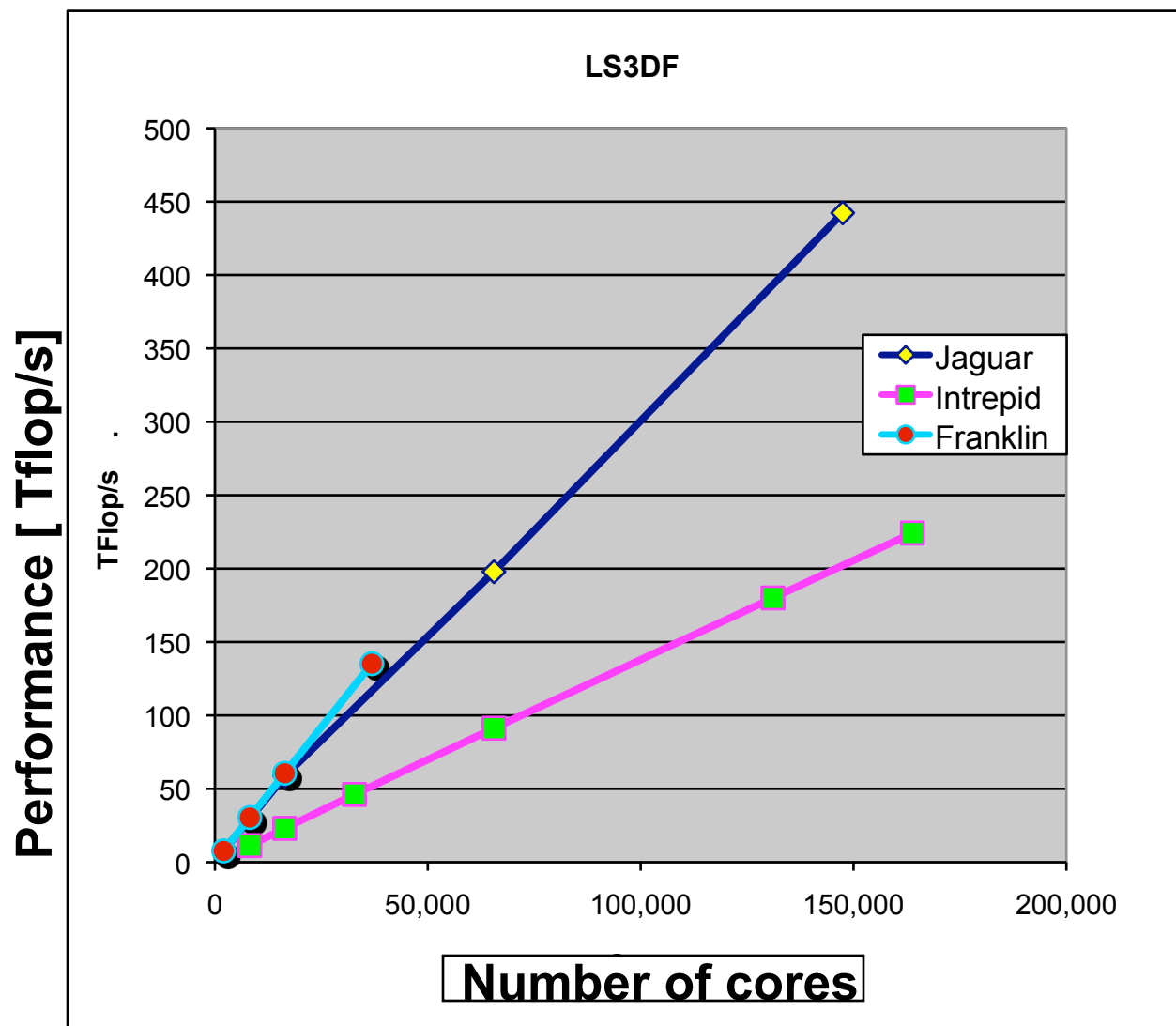
core	8,192	32,768	163,840
atom	512	2048	10,240
gen_VF	0.08	0.08	0.23
PEtot_F	69.30	68.81	69.87
gen_dens	0.08	0.14	0.37
Poisson	0.12	0.22	0.76

Perfect weak scaling

# Near perfect speedup across a wide variety of systems (weak scaling)



# ZnTeO alloy weak scaling calculations



Note: Ecut = 60Ryd with *d* states, up to 36864 atoms

# System Performance Summary



- ❖ 135 Tflops/s on 36,864 processors of the quad-core Cray XT4 Franklin at NERSC, 40% efficiency



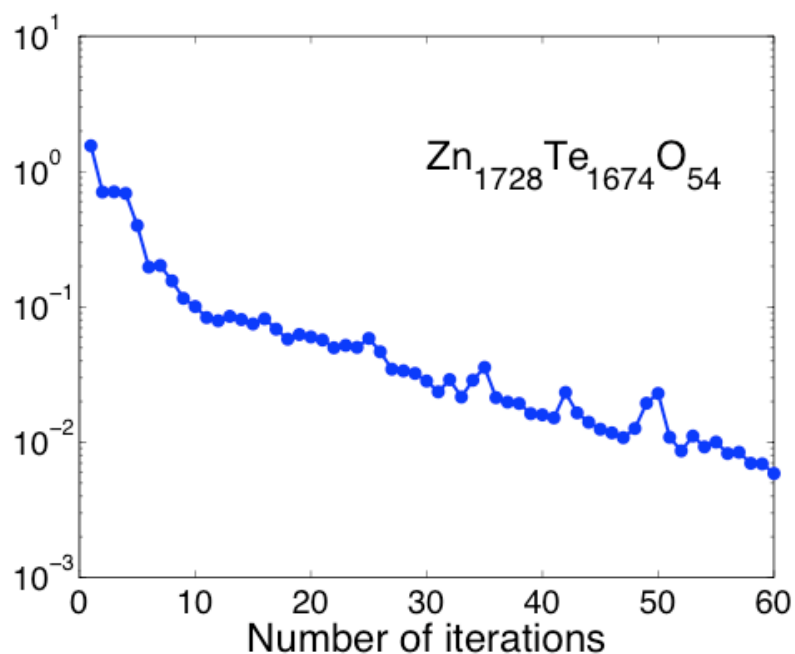
- ❖ 224 Tflops/s on 163,840 processors of the BlueGene/P Intrepid at ALCF, 40% efficiency



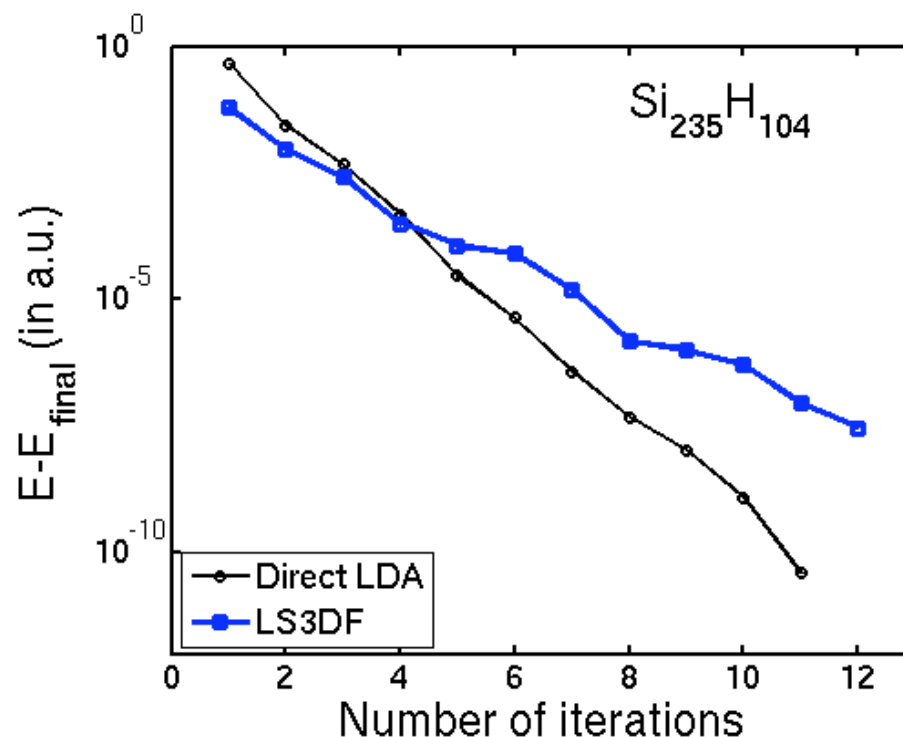
- ❖ 442 Tflops/s on 147,456 processors of the Cray XT5 Jaguar at NCCS, 33% efficiency

**For the largest physical system (36,000 atoms),  
LS3DF is 1000 times faster than direct DFT codes**

# Selfconsistent convergence of LS3DF



Measured by potential



Measured by total energy

- ❖ SCF convergence of LS3DF is similar to direct LDA method
- ❖ It doesn't have the SCF problem some other  $O(N)$  methods have

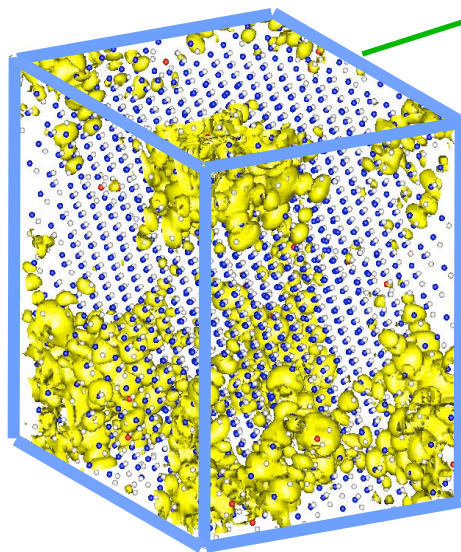
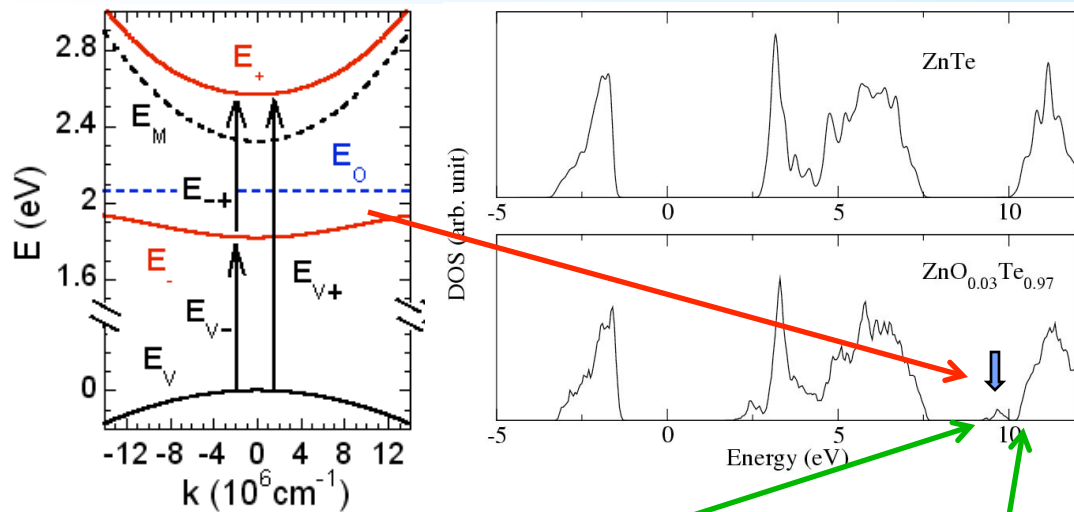
# LS3DF Accuracy is determined by fragment size

- ❖ A comparison to direct LDA calculation, with an 8 atom 1x1x1 fragment size division:
  - The total energy error: 3 MeV/atom  $\sim$  0.1 kcal/mol
  - Charge density difference: 0.2%
  - Better than other numerical uncertainties (e.g. PW cut off, pseudopotential)
- ❖ Atomic force difference:  $10^{-5}$  a.u.
  - Smaller than the typical stopping criterion for atomic relaxation
- ❖ Other properties:
  - The dipole moment error:  $1.3 \times 10^{-3}$  Debye/atom, 5%
  - Smaller than other numerical errors

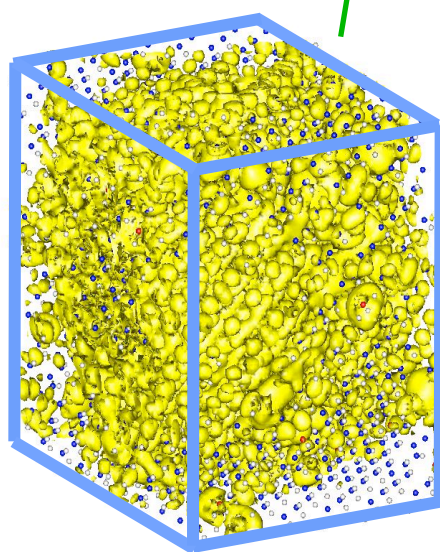
**For most practical purposes, LS3DF is the same as direct LDA**



# Can one use an intermediate state to improve solar cell efficiency?



Highest O induced state

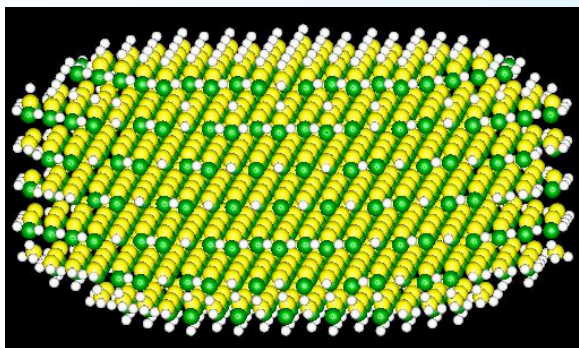


ZnTe bottom of cond. band state

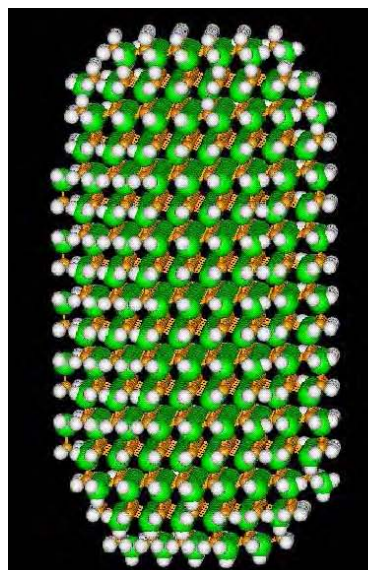
- ❖ Single band material theoretical PV efficiency is 30%
- ❖ With an intermediate state, the PV efficiency could be 60%
- ❖ One proposed material ZnTe:O
  - Is there really a gap?
  - Is it optically forbidden?
- ❖ LS3DF calculation for 3500 atom 3% O alloy [one hour on 17,000 cores of Franklin]
- ❖ Yes, there is a gap, and O induced states are very localized.

INCITE project, NERSC, NCCS.

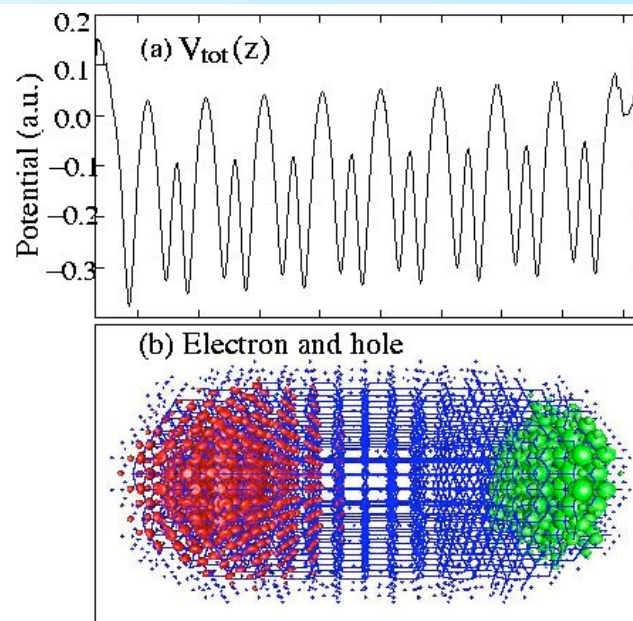
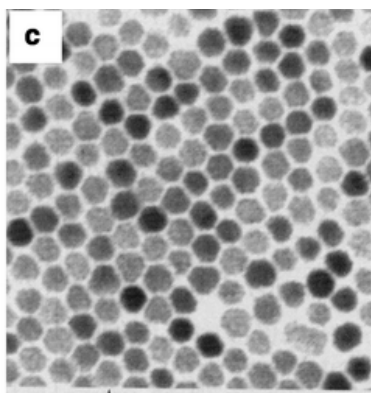
# LS3DF computations yield dipole moments of nanorods and the effects on electrons



**P = 30.3 Debye**



**P=73.3 Debye**



**Cd<sub>714</sub>Se<sub>724</sub>**  
**WZ**

- ❖ Equal volume nanorods can have different dipole moments
- ❖ The inequality comes from shape dependent self-screening
- ❖ Dipole moments depend on bulk and surface contributions
- ❖ Dipole moments can significantly change the electron and hole wave functions

INCITE project at NCCS and NERSC

C O M P U T A T I O N A L R E S E A R C H D I V I S I O N



# Summary and Conclusions

- ❖ LS3DF scales linearly to over 160,000 processors. It reached 440 Tflops/s. It runs on different platforms with little retuning
- ❖ The numerical results are the same as a direct DFT based on an  $O(N^3)$  algorithm, but at only  $O(N)$  computational cost
- ❖ LS3DF can be used to compute electronic structures for  $>10,000$  atom systems with total energy and forces in 1-2 hours. It can be 1000 times faster than  $O(N^3)$  direct DFT calculations.
- ❖ Enables us to yield new scientific results predicting the efficiency of proposed new solar cell materials

# Acknowledgements

- ❖ **National Energy Scientific Computing Center (NERSC)**
- ❖ **National Center for Computational Sciences (NCCS)**  
(Jeff Larkin at Cray Inc)
- ❖ **Argonne Leadership Computing Facility (ALCF)**  
(Katherine M Riley, William Scullin)
- ❖ **Innovative and Novel Computational Impact on Theory and Experiment (INCITE)**
- ❖ **SciDAC/PERI (Performance Engineering Research Institute)**
- ❖ **DOE/SC/Basic Energy Science (BES)**  
**DOE/SC/Advanced Scientific Computing Research (ASCR)**



# LS3DF Team



**Lin-Wang Wang**



**Zhengji Zhao**



**Byounghak Lee**



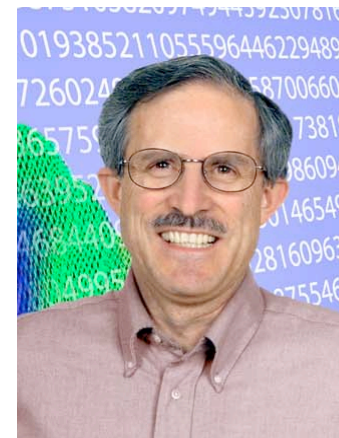
**HongZhang Shan**



**Juan Meza**



**Erich Strohmaier**



**David Bailey**