Truly Scalable O(N) Approach for First-Principles Molecular Dynamics of Non-Metallic Systems

Jean-Luc Fattebert and Daniel Osei-Kuffuor
Center for Applied Scientific Computing

Lawrence Livermore National Laboratory

LLNL-PRES-673679

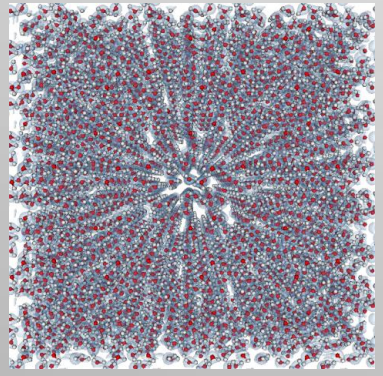
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

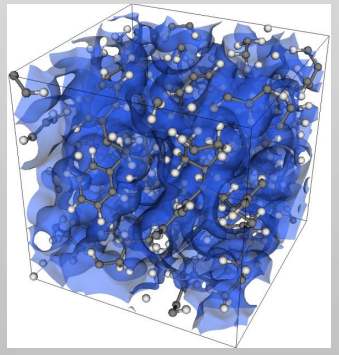
We want to take advantage of growing computational power to simulate larger and more realistic problems in material sciences

Sequoia, IBM BGQ, 1,572,864 cores

O(N) scalable algorithm

For insulators, semiconductors



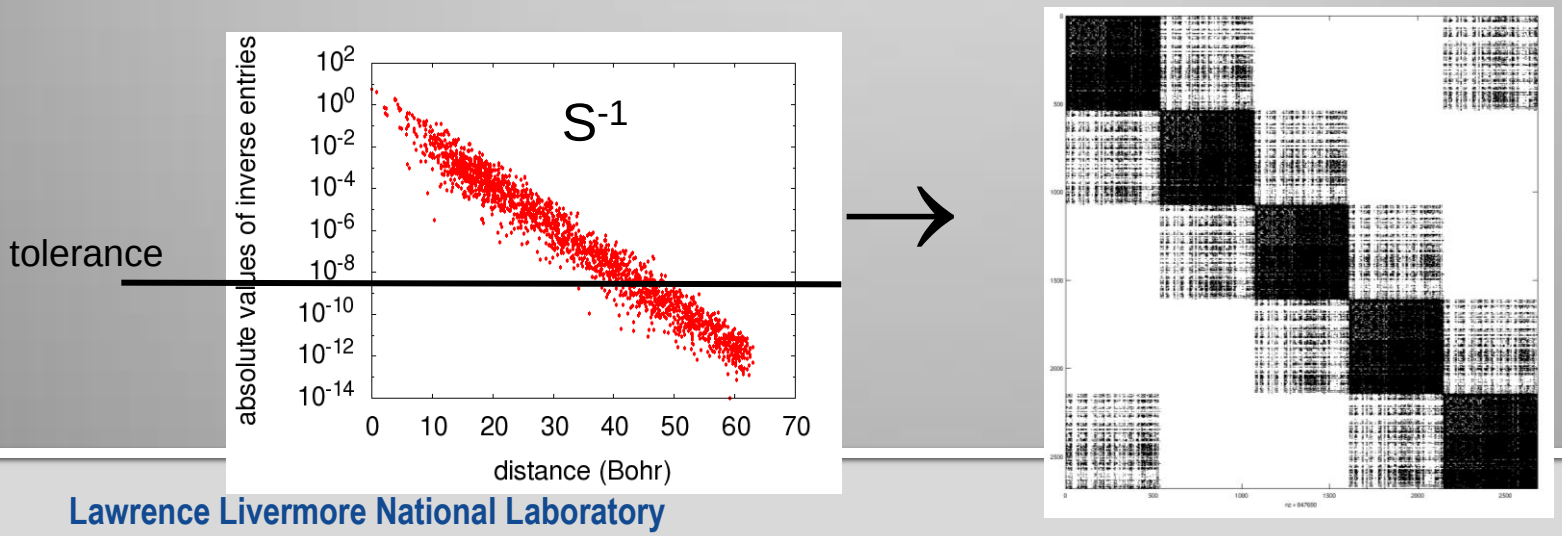


Large problems
> 5,000 atoms

- DFT with *Planewaves* pseudopotential accuracy (LDA, PBE)
- Fast time to solution
  - 1 step in minutes (not hours!!!) to be useful for MD

Developing an O(N) algorithm for FPMD implies truncations / approximations

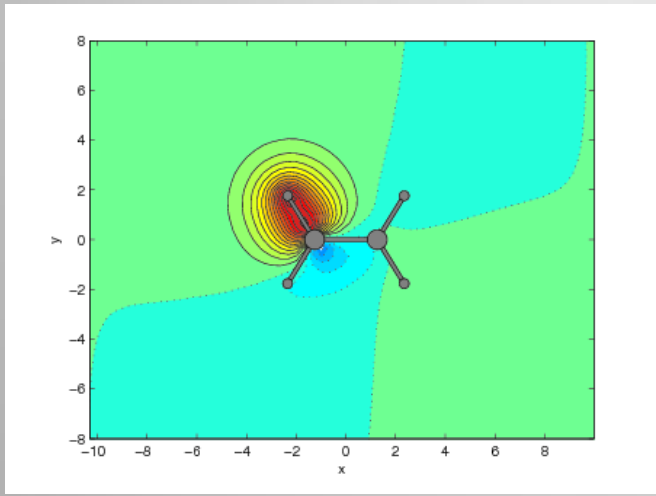
- Unlike classical physics models, in Quantum models the number of physical variables (electrons) grows with system size
  - $\rightarrow$ $O(N^2)$ degrees of freedom and $O(N^3)$ operations in DFT
- Reducing computational complexity to O(N) typically implies
  - Introduction of controllable approximations / truncate fast decaying terms
  - More complicated data structures – sparse vs. full matrices

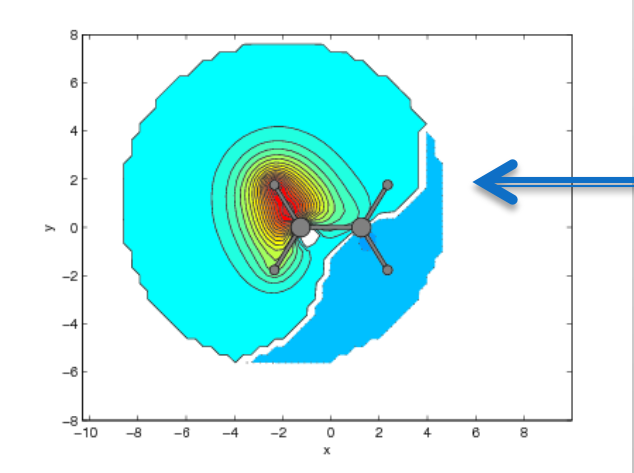


For systems with band-gap, one can find a representation of the electronic structure with localized functions

- Example: C_2H_4

Strictly localized, non-orthogonal,
Not centered on atoms (adaptive)





Auxilliary
“basis set”

- (orthogonal) Maximally Localized Wannier functions
  - Minimize the sum of the spread of all the functions [Marzari and Vanderbilt, PRB 1997]

$$\sum_{i=1}^{N} \left\langle \phi_i \left| \left(\hat{X} - \left\langle \phi_i \left| \hat{X} \right| \phi_i \right\rangle \right)^2 \right| \phi_i \right\rangle$$

Density Functional Theory: general formulation for non-orthogonal orbitals

- Energy minimization for general non-orthogonal orbitals
  [Galli and Parrinello, PRL 1992]

$$E_{KS}\left[\{\phi\}_{i=1}^{N}\right] = \sum_{i,j=1}^{N}\left(S^{-1}\right)_{ij}\int_{\Omega}\phi_i(r)\Delta\phi_j(r) + F[\rho] + \sum_{i,j=1}^{N}\left(S^{-1}\right)_{ij}\int_{\Omega}\phi_i(r)\left(V_{ext}\phi_j\right)(r)$$

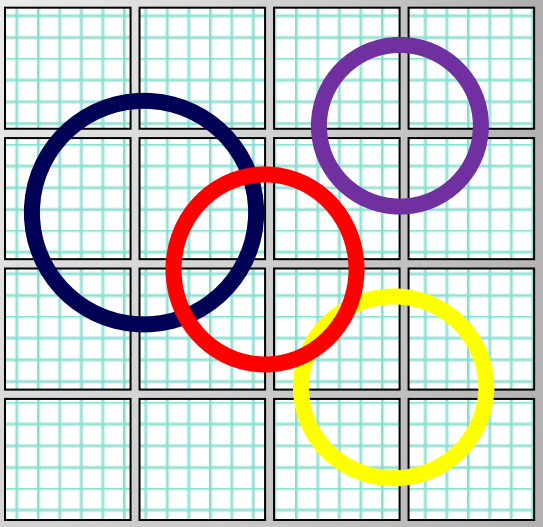
$$\rho(r) = \sum_{i,j=1}^{N}\left(S^{-1}\right)_{ij}\phi_i(r)\phi_j(r)$$

To take into account non-orthogonality

$$S_{ij} = \int_{\Omega}\phi_i(r)\phi_j(r)dr$$

- Assume finite gap $\varepsilon_N < \varepsilon_{N+1}$
- Assuming functions ϕ_i are linearly independent…
- No need for any eigenvalue computation!

DFT O(N) algorithm for localized functions



- Real-space (finite difference) discretization
- Norm-conserving pseudopotentials
- Parallel domain decomposition
- Confine functions to finite spherical regions
  - Each Φ_i lives on Finite Difference mesh, in a localization region of center $\mathbf{R}_i$ and radius Rc
  - O(1) d.o.f. for each orbital
- Iterative solver: direct minimization of energy functional
  - follow preconditioned steepest descent directions + block Anderson extrapolation scheme [JLF, J. Comp Phys 2010]
  - Truncate trial solution at each step [JLF and Bernholc, PRB 2000, JLF and F. Gygi, Comp Phys Comm 2004, PRB 2006]

There remains an O(N^3) operation…

$$E_{KS}\left[\{\phi\}_{i=1}^{N}\right] = \sum_{i,j=1}^{N} \left(S^{-1}\right)_{ij} \int_{\Omega} \phi_i(r) \Delta \phi_j(r) + F[\rho] + \sum_{i,j=1}^{N} \left(S^{-1}\right)_{ij} \int_{\Omega} \phi_i(r) \left(V_{ext} \phi_j\right)(r)$$

$$\rho(r) = \sum_{i,j=1}^{N} \left(S^{-1}\right)_{ij} \phi_i(r) \phi_j(r)$$

- Not even expensive, but requiring a lot of communications
  - O(N^3) solver becomes a bottleneck beyond 10,000 atoms and/or 10,000 MPI tasks
- Smaller size than in Tight-Binding models or LCAO methods
- “Global” coupling
- Need to calculate selected elements of the inverse of Gram matrix S
- We essentially need the elements S^{-1}_{ij} s.t. $S_{ij} \neq 0$

Gram (overlap) Matrix Properties

$$S_{ij} = \int_{\Omega} \phi_i(r)\phi_j(r)dr$$

$$S_{ij} = 0 \Rightarrow \left|\vec{c}_i - \vec{c}_j\right| > 2R_c$$

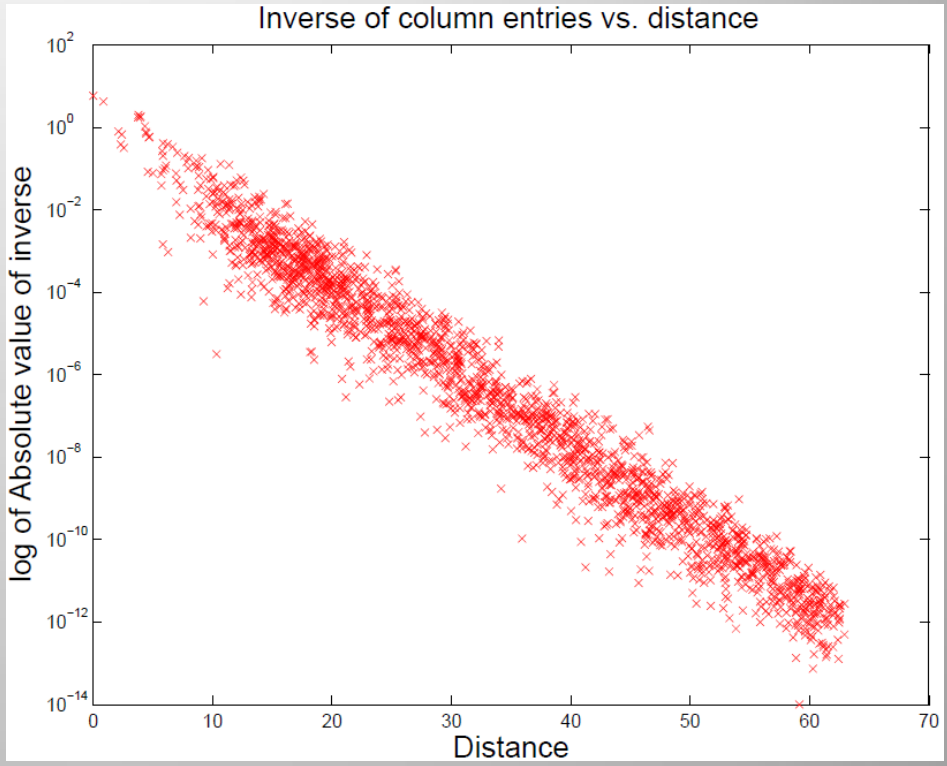
$$\vec{c}_i = \text{position of } \phi_i$$

$$R_c = \text{local function radius}$$

- S is sparse, Symmetric Positive Definite
- Condition number is independent of problem size!!
- Inverse
  - In principle full matrix…
  - …But off-diagonal elements decay exponentially fast [Demko et al., Math. Comp. 1984] [Benzi & Razouk, ETNA 2007]
  - Assumption: spectrum of S bounded away from 0, independently of N

We verify fast exponential decay of off-diagonal elements of the inverse of Gram matrix

- Polymers, 1888 atoms
- How to make efficient use of it on large parallel computers?



O(N) short-range calculation of selected elements of S^{-1}

- Based on the approximate inverse strategy

$$Solve: \arg\min_{M \in \Re^{N \times N}} \|SM - I\|_F \Rightarrow M \approx S^{-1}, I = \text{identity matrix}$$

- Sparsity pattern of *M* is predetermined by geometric distance

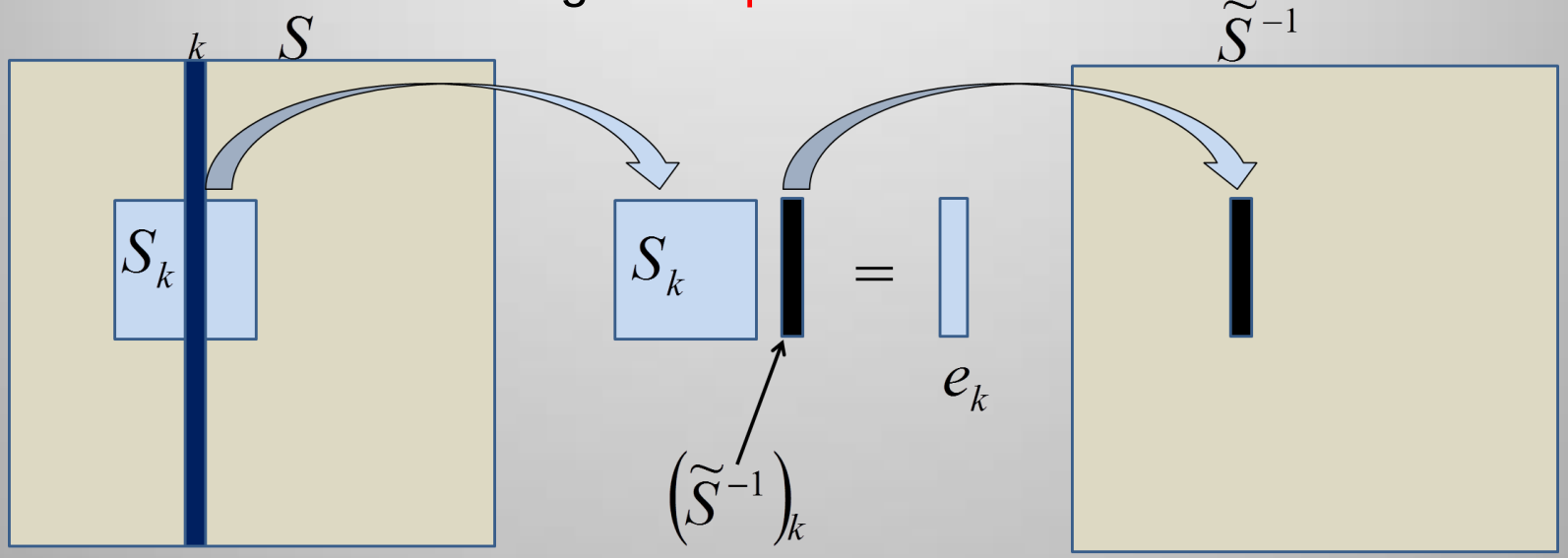
$$\forall \phi_j \,|\, \vec{c}_j \in \Omega_L, \text{define } \mathcal{J} = \left\{ k \,|\, \left| \vec{c}_j - \vec{c}_k \right| < R_s \right\}$$

$$\text{and set } M_{jk} \neq 0 \forall k \in \mathcal{J}, \text{for some distance } R_s$$

- R_s determines accuracy of selected elements of the inverse

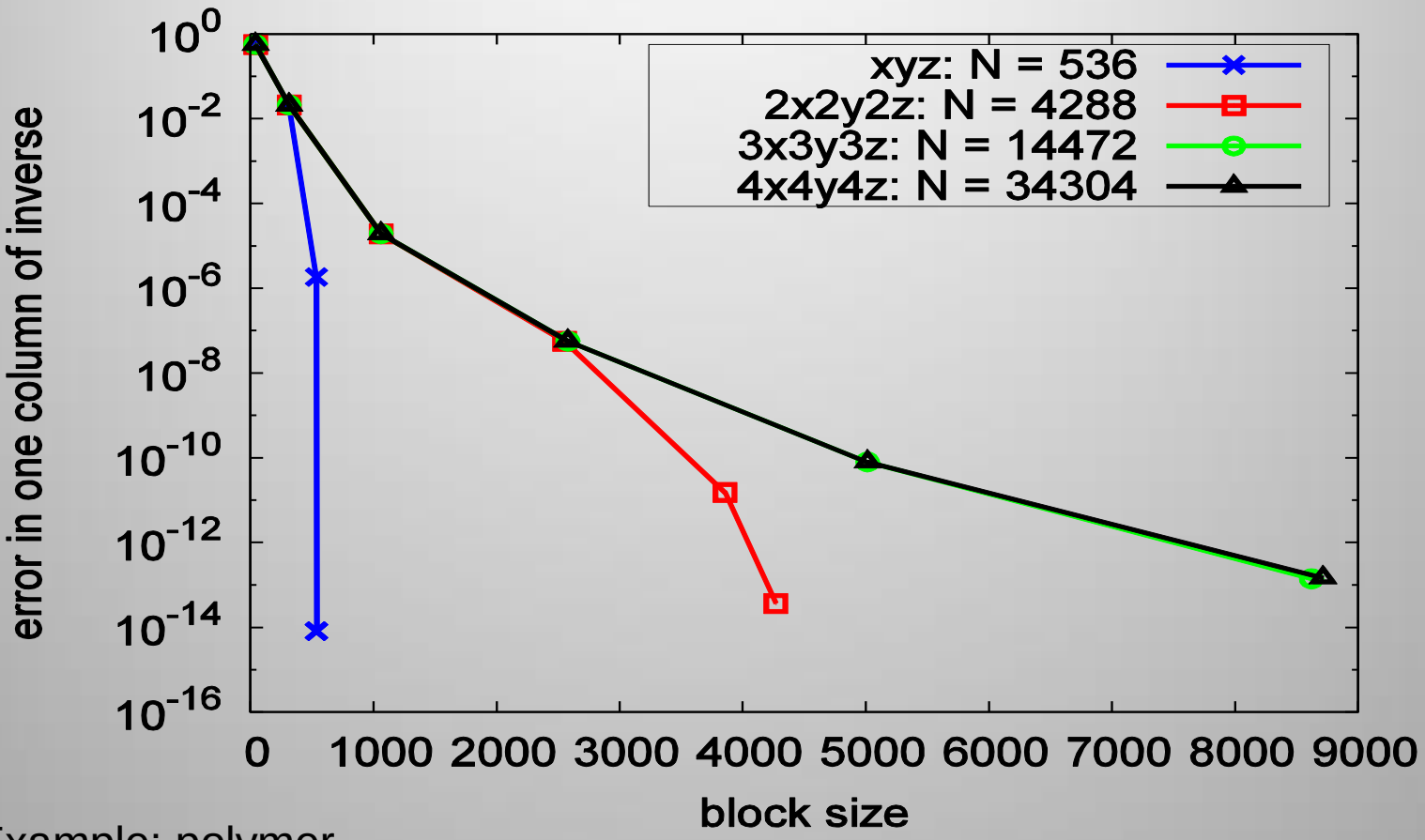
Computation of selected elements boils down to inverting local principal submatrix

- Include rows and columns of S corresponding to closest local functions (distance between centers)
- Solve for column k using ILU0–preconditioned GMRES



- Note: S not close to Identity matrix!!! (unlike in Tight-Binding or LCAO approach where no preconditioner is needed [Stechel et al. PRB 1994])

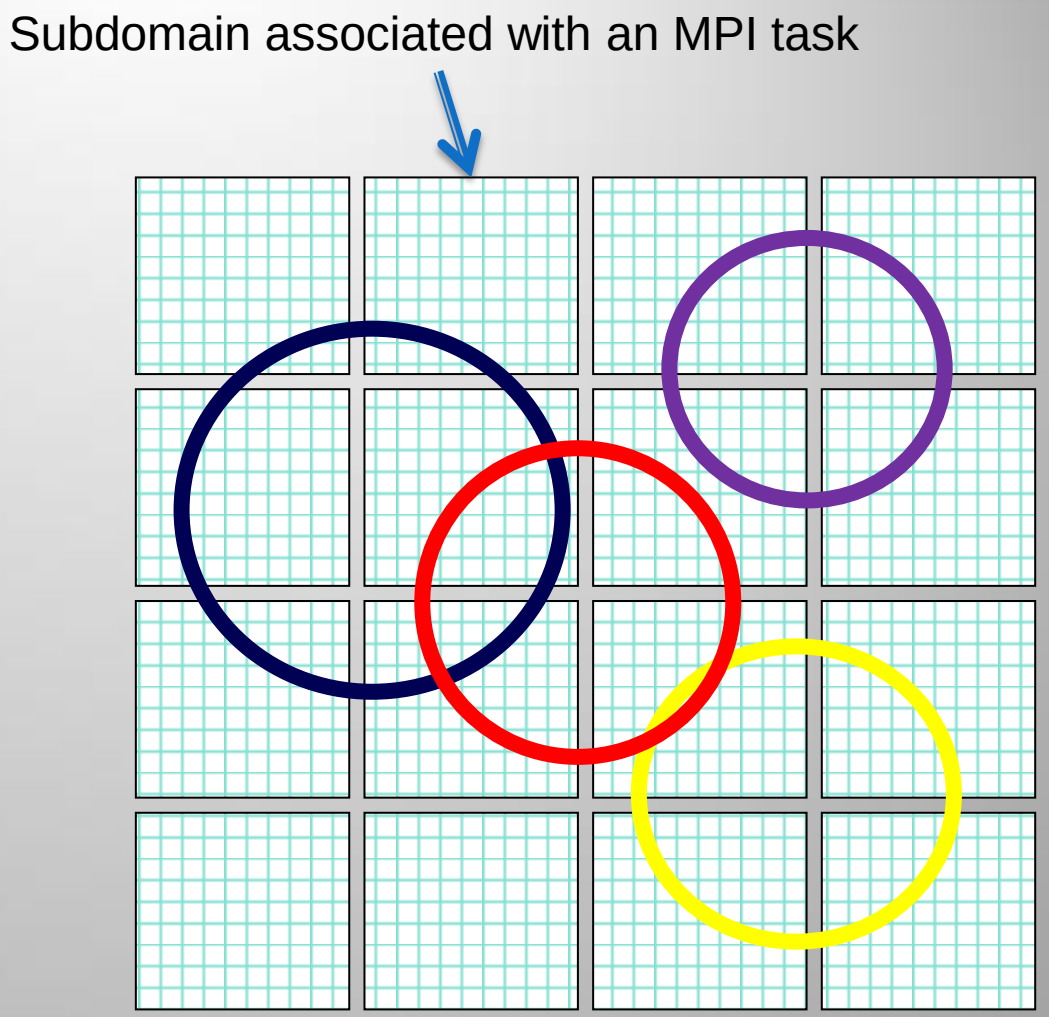
Error on approximate inverse decays fast with principal submatrix size



Example: polymer

Data layout

- Localized orbital are distributed across processors
- Each MPI task owns pieces of several functions
- Each MPI task computes partial contribution to the global matrices (overlap,…)



Exploiting sparsity poses challenges for fast parallel implementation compared to O(N^3) algorithms

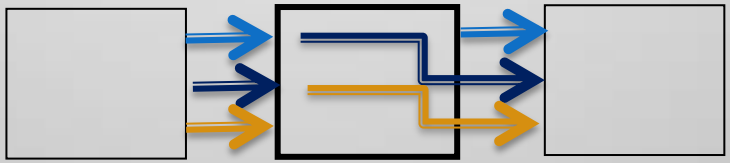
- Energy can be written as:

$$E_{ks} = Tr(S^{-1}H_{\phi}) + F(\rho), \text{where } H_{\phi} = \Phi^T H \Phi$$

- S^{-1} is approximated, sparse and has complete but distributed entries
- H_{ϕ} is sparse and distributed (incomplete entries)
- Each PE only needs entries corresponding to locally centered functions
  - Need to consolidate partial contributions of H_{ϕ}
- Efficient data communication and assembling algorithm is needed

Parallel data communication / matrix assembly is key to efficiency

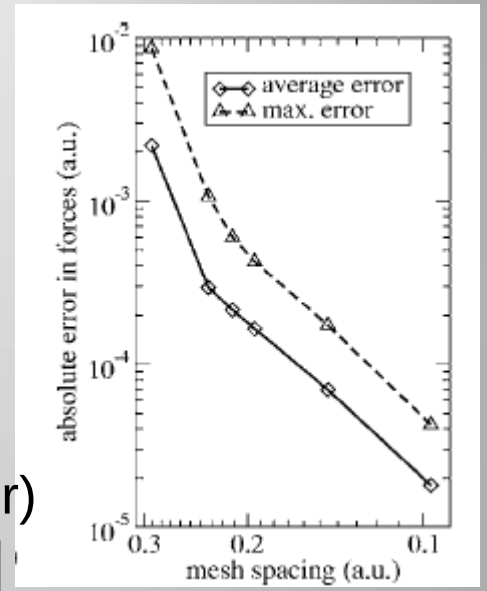
- Each parallel task compute partial contributions to some matrix elements
- Need to assemble local principal submatrix matrix
  - Sum up partial dot products computed on various processors
- We use a short range communication pattern where data is passed down to nearest neighbor only, one direction at a time, for as many steps as needed



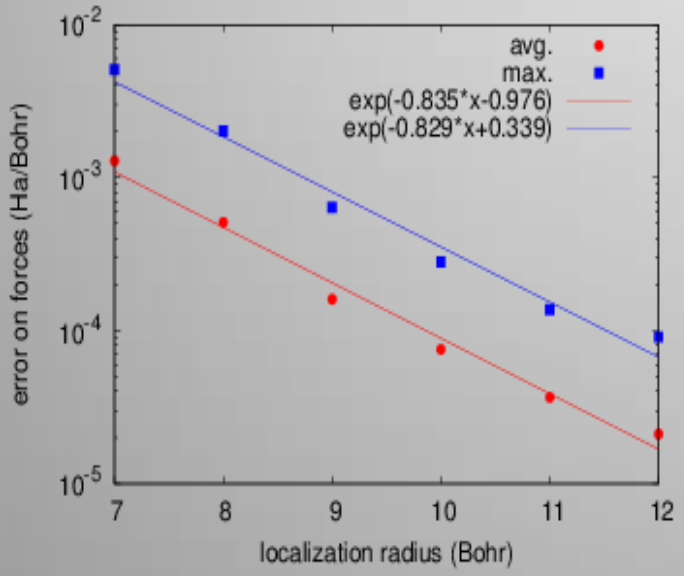
- Overlap communication and computation
  - Accumulate received data in sparse data structure while sending data for next step
- Need to scatter results to *adjacent* processors that need column j of S^{-1}

Three parameters to control error and achieve needed accuracy

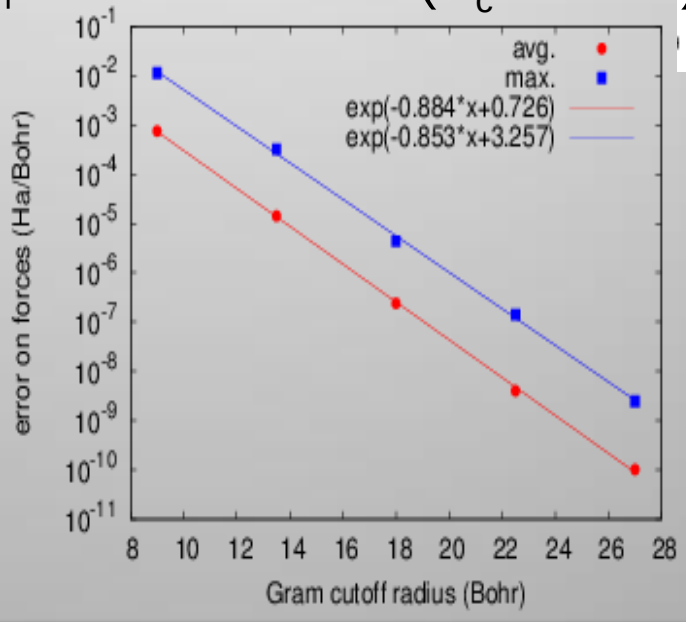
- Finite difference Mesh spacing
  - error $O(h^4)$
- 2 parameters to control O(N) truncation



Localization of functions ϕ

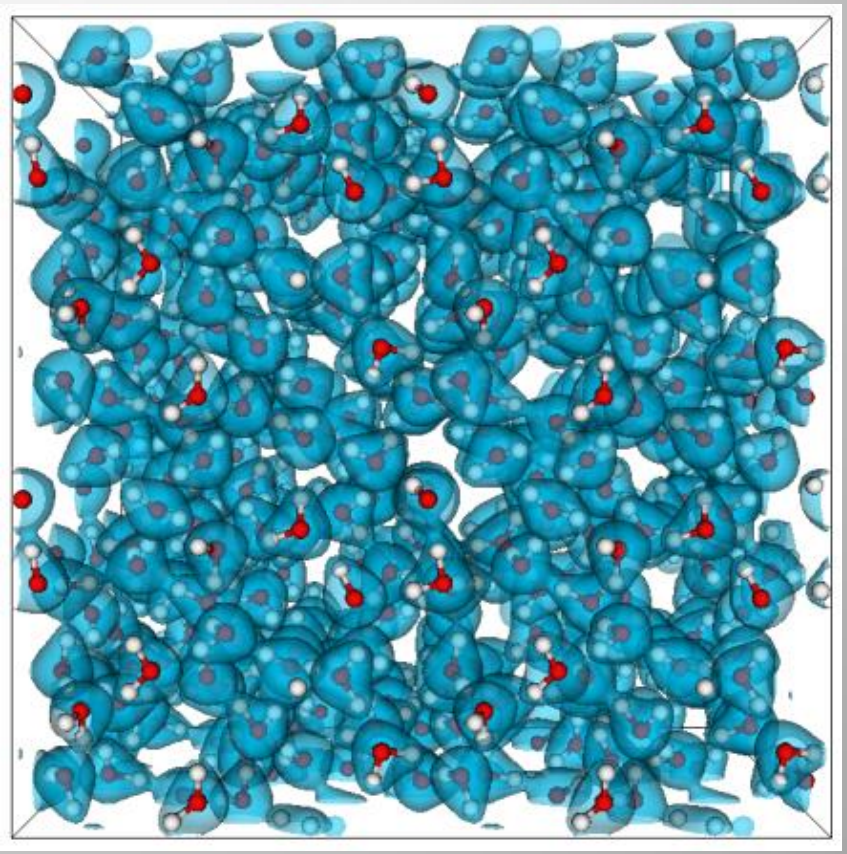


Cutoff for S^{-1} (R_c=9 Bohr)

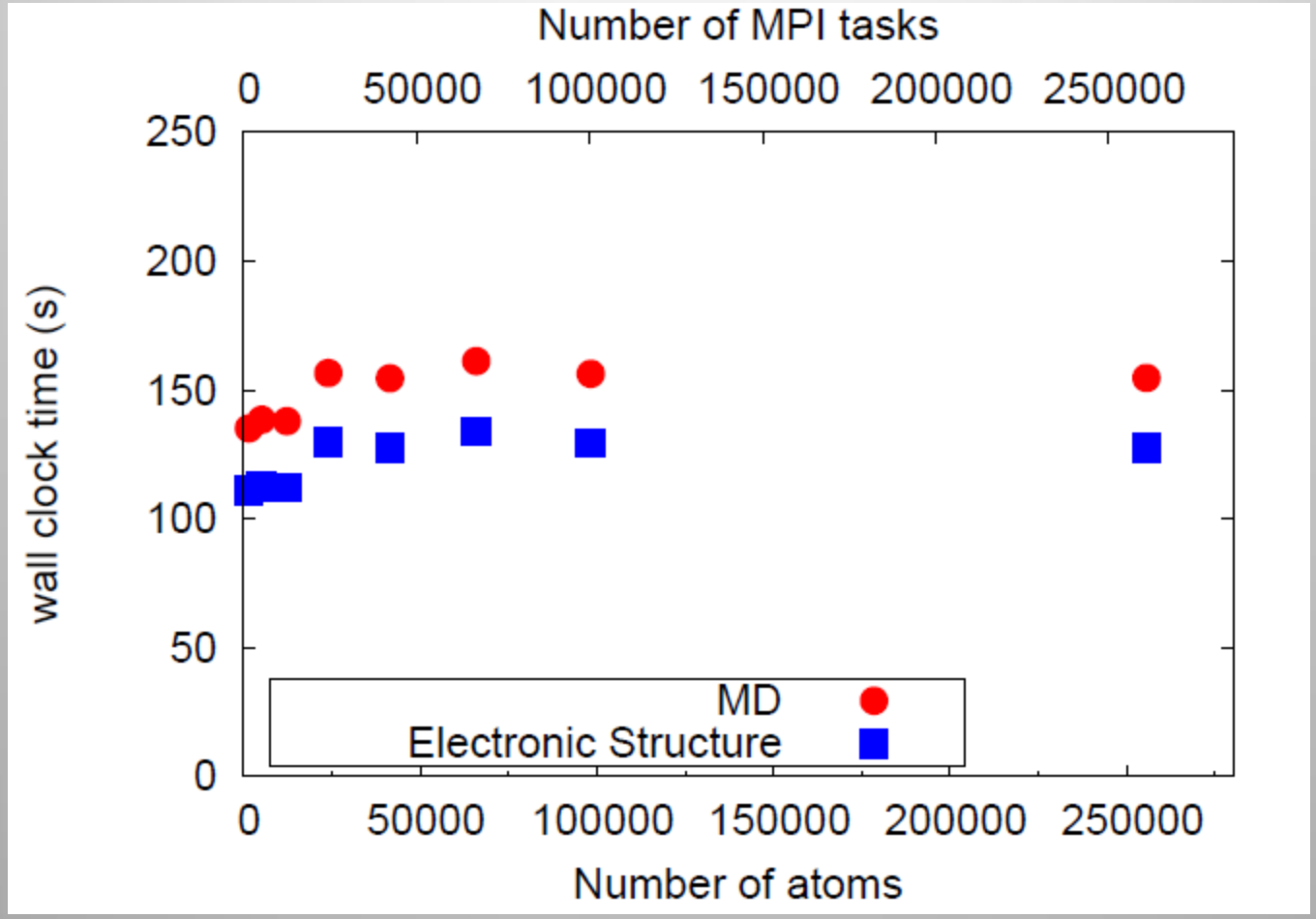


Weak scaling: Test application

- Liquid water
  - 1536 atoms (512 molecules)
  - 2048 orbitals
- Replicate
  - 2x2x2
  - 3x3x3
  - 4x4x4
  - …

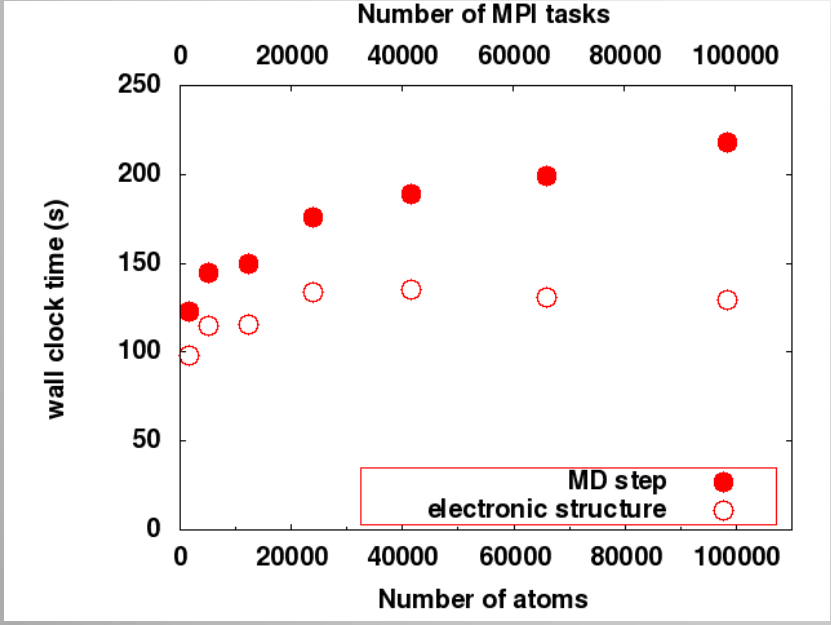


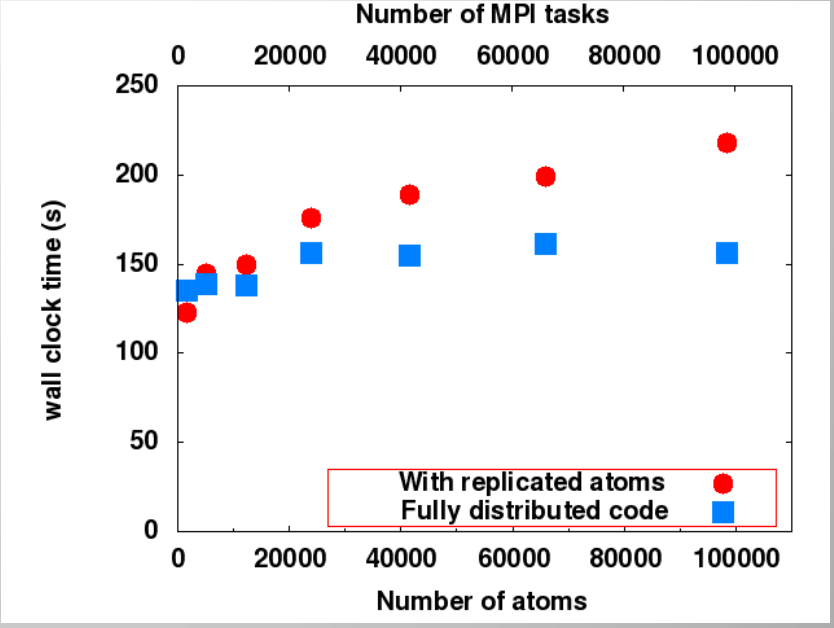
Numerical results show excellent weak scaling and fast time to solution (No limit to scaling – in principle)



IBM BGQ

At 100K MPI tasks and beyond, everything needs to be distributed!





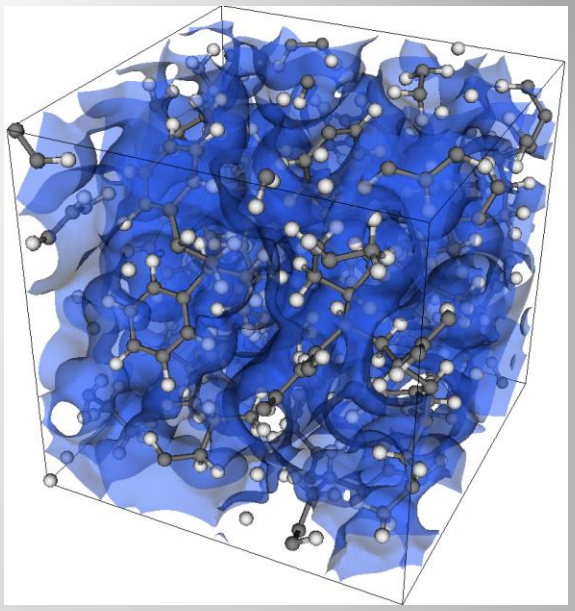
Distributing atomic positions

Liquid water on IBM/BGQ

Nearsightedness principle for parallel computation

- W. Kohn's nearsightedness principle [PRL 1996]
- Nearsightedness in computational algorithm leads to O(N) and parallel scaling beyond 100,000 MPI tasks
  - Practical accuracy achieved with short range communications / no global communications for insulators

Conclusion



- Research supported by LLNL LDRD program
- Recent Publications
  - D. Osei-Kuffuor and JLF, PRL 2014
  - D. Osei-Kuffuor and JLF, SIAM J. Sci. Comput. 2014
- Future
  - Speed-up time-to-solution (threading)
  - Applications
    - Distribution of ions in dilute solution: K + Cl in water
    - Biology
  - Extension to metals…