Pushing the Envelope of Large Scale First Principles Simulations of Non-Insulating Systems

Amartya Banerjee\textsuperscript{1,2}

Joint work with L. Lin\textsuperscript{1,3}, C. Yang\textsuperscript{1}, W. Hu\textsuperscript{1}, J. Pask\textsuperscript{4} and P. Suryanarayana\textsuperscript{5}

\textsuperscript{1}Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
\textsuperscript{2}Mechanical, Aerospace & Biomedical Engineering, University of Tennessee, Knoxville, TN, USA
\textsuperscript{3}Department of Mathematics, University of California, Berkeley, CA, USA
\textsuperscript{4}Physics Division, Lawrence Livermore National Laboratory, Livermore, CA, USA
\textsuperscript{5}College of Engineering, Georgia Institute of Technology, GA, USA

MolSSI Workshop and ELSI Conference

Richmond, Virginia.

August 17, 2018
Introduction

- Energy storage devices: Lithium Ion batteries
- In practice reliability/safety is a major issue

Images courtesy of internet resources
Introduction (cont.)

- Another example: Mechanics of Grain Boundaries - Tilt angle vs. GB Energy plot

- First principles calculations based on DFT: Workhorse of computational materials science — Many steps of Quantum Molecular Dynamics!
- Fundamental need: Computational ability to attack large systems efficiently and accurately.
- Focus on non-insulating (metallic / semiconducting) systems.

Images courtesy of B. Runnels et al and internet resources
Introduction (cont.)

- Another example: Mechanics of Grain Boundaries - Tilt angle vs. GB Energy plot

- First principles calculations based on DFT: Workhorse of computational materials science — Many steps of Quantum Molecular Dynamics!

- Fundamental need: Computational ability to attack large systems efficiently and accurately.

- Focus on non-insulating (metallic / semiconducting) systems.

Images courtesy of B. Runnels et. al and internet resources
Kohn-Sham Density Functional Theory

- The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_{\text{eff}} \psi_i(x) = \varepsilon_i \psi_i(x),\]

\[\hat{V}_{\text{eff}} = V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc},\]

\[V_I^\ell = \sum_a V_{I,a}(x),\]

\[\hat{V}_I^{n\ell} \psi_i = \sum_a \int d\mathbf{x}' V_I^{n\ell}(\mathbf{x}, \mathbf{x}') \psi_i(\mathbf{x}'),\]

\[V_H = -\int d\mathbf{x}' \frac{\rho_e(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|},\]

\[V_{xc} = V_{xc}(\mathbf{x}; \rho_e),\]

\[\rho_e = -\sum_i f_i \psi_i^* (\mathbf{x}) \psi_i (\mathbf{x}).\]

**SCF Iterations**

Start (guess): \(\rho_{in}\)

\[\hat{V}_{\text{eff}} [\rho_{in}]\]

\[\{\psi_i, \varepsilon_i\}_{i=1}^m\]

\[\rho_{out}\]

Repeat: new \(\rho_{in}\)

no \(\rho_{out} = \rho_{in}\)?

yes

Done:

\[\{\psi_i, \varepsilon_i\}_{i=1}^m \Rightarrow \text{materials properties}\]
Kohn-Sham Density Functional Theory

- The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_{\text{eff}} \psi_i(x) = \varepsilon_i \psi_i(x),\]
\[\hat{V}_{\text{eff}} = V_I^\ell + \hat{V}_I^{\text{n} \ell} + V_H + V_{xc},\]
\[V_I^\ell = \sum_a V_{I,a}(x),\]
\[\hat{V}_I^{\text{n} \ell} \psi_i = \sum_a \int dx' V_I^{\text{n} \ell}(x, x') \psi_i(x'),\]
\[V_H = -\int dx' \frac{\rho_e(x')}{|x - x'|},\]
\[V_{xc} = V_{xc}(x; \rho_e),\]
\[\rho_e = -\sum_i f_i \psi_i^*(x) \psi_i(x),\]

SCF Iterations

Start (guess): \( \rho_{m} \)

Repeat:

\[\hat{V}_{\text{eff}}[\rho_{m}] \]
\[\{\psi_i, \varepsilon_i\}_{i=1}^{m} \]
\[\rho_{out} \]

new \( \rho_{m} \)

no (mix)

\( \rho_{out} = \rho_{m} ? \)

yes

Done:

\[\{\psi_i, \varepsilon_i\}_{i=1}^{m} \Rightarrow \text{materials properties} \]

- Linear Problem: Solve the Schrödinger equation in an efficient and scalable way.
Outline

• Introduction

• **Part I**: Chebyshev Filtered Subspace Iterations (CheFSI) Within the Framework of Discontinuous Galerkin Density Functional Theory for Large Scale Abinitio Simulations

• Part II : Complementary Subspace Strategy for confronting the Rayleigh-Ritz bottle neck

• Part III : Periodic Pulay method for robust and efficient convergence acceleration of self-consistent field iterations

• Broad guiding principles: Maintain generality - as many systems as possible (metals, insulators…)

• Reduce pre-constants.
Part I: CheFSI within DGDFT

- The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_{\text{eff}} \psi_i(x) = \varepsilon_i \psi_i(x),\]

\[\hat{V}_{\text{eff}} = V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc},\]

\[V_I^\ell = \sum_a V_{I,a}(x),\]

\[\hat{V}_I^{n\ell} \psi_i = \sum_a \int dx' V_{I,a}^{n\ell}(x, x') \psi_i(x'),\]

\[V_H = - \int dx' \frac{\rho_e(x')}{|x - x'|},\]

\[V_{xc} = V_{xc}(x; \rho_e),\]

\[\rho_e = - \sum_i f_i \bar{\psi}_i^*(x) \psi_i(x),\]

SCF Iterations

Start (guess): \(\rho_{in}\)

Repeat:

new \(\rho_{in}\) \(\rightarrow\) \(\hat{V}_{\text{eff}}[\rho_{in}]\)

\(\{\psi_i, \varepsilon_i\}_{i=1}^m\)

\(\rho_{out}\)

\(\text{no (mix)}\)

\(\text{yes}\)

done:

\(\{\psi_i, \varepsilon_i\}_{i=1}^m \Rightarrow \text{materials properties}\)
Part I: CheFSI within DGDFT

• The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_\text{eff} \psi_i(x) = \varepsilon_i \psi_i(x),\]

\[\hat{V}_\text{eff} = V_I^\ell + \hat{V}_{I}^{nl} + V_H + V_{xc},\]

\[V_I^\ell = \sum_a V_{I,a}(x),\]

\[\hat{V}_{I}^{nl} \psi_i = \sum_a \int d\mathbf{x}' V_{I,a}^{nl}(x, x') \psi_i(x'),\]

\[V_H = -\int d\mathbf{x}' \frac{\rho_e(x')}{|x - x'|},\]

\[V_{xc} = V_{xc}(x; \rho_e),\]

\[\rho_e = -\sum_i f_i \psi_i^*(x) \psi_i(x),\]

SCF Iterations

Start (guess): \[\rho_{\text{in}}\]

Repeat:

\[\hat{V}_\text{eff}[\rho_{\text{in}}]\]

\[\{\psi_i, \varepsilon_i\}_{i=1}^m\]

\[\rho_{\text{out}}\]

new \[\rho_{\text{in}}\] no (mix)

\[\rho_{\text{out}} = \rho_{\text{in}}?\]

yes

Done:

\[\{\psi_i, \varepsilon_i\}_{i=1}^m \Rightarrow \text{materials properties}\]

• Linear Problem: Solve the Schrödinger equation in an efficient and scalable way.
Confronting the bottlenecks

• Observation 1: Irrespective of the choice of solution method, having fewer degrees of freedom per atom is highly desirable.

• Typically results in a lower pre-factor for any algorithm of choice.

• Plane-waves, Finite Differences, Finite Elements:
  • Systematically improvable, highly accurate solutions.
  • A few hundred to thousands of DOFs per atom for chemical accuracy.

• Atom centered basis functions (Gaussians, Numerical Orbitals,…)
  • Accuracy control is more involved: requires expertise, awareness of BSSE …
  • 10 - 15 DOFs per atom are often enough for chemical accuracy.

• Adaptive Local Basis (ALB) set.
  • Systematically improvable, accurate solutions.
  • A few tens of DOFs per atom enough for chemical accuracy.
Confronting the bottlenecks (cont.)

• Observation 2: The specific choice of algorithm matters

  • Traditional (cubic scaling) vs. lower complexity algorithms
    (Fermi Operator Expansion, Spectral Quadrature, Purification,
    Pole-Expansion and Selected Inversion, …)

    • Range of applicability needs to be as broad as possible: Different materials
      systems (2D - 3D, metal-insulator, …), wide range of temperatures, …

  • Traditional (cubic scaling) algorithms (Direct / Iterative
    diagonalization: ScaLAPACK, Variations of Preconditioned Conjugate
    Gradients, Lanczos, Polynomial filtering, spectrum slicing, …)

    • Value of pre-constant matters in practical simulations.
    • Parallel scalability is important for large calculations.

• Combine Observations 1 and 2
Observation 1: Discontinuous Galerkin Density Functional Theory

- Adaptive Local Basis functions for an interior penalty Discontinuous Galerkin formulation of Kohn Sham DFT.

- DG formalism allows solution of Kohn-Sham equations with discontinuous basis set: Number of desirable properties simultaneously.
- Few 10s of DOFs per atom.
- Systematically improvable.
- Strictly local: identically zero outside prescribed subdomain, zero overlap across subdomains.
- Orthonormal basis: standard, relatively well conditioned eigenvalue problem.

- How?

  - Partition domain into sub-domain (elements).
  - Solve Kohn-Sham equations in each (extended) element.
  - Basis is union of local Kohn-Sham solutions: Incorporates local materials physics.
  - Each of these steps can be done efficiently with good parallel scaling.

Discontinuous Galerkin Density Functional Theory (cont.)

- Adaptive Local Basis Functions.

• Kohn-Sham Hamiltonian becomes:

\[ H_{DG}(k', j'; k, j) =\frac{1}{2} \left< \nabla u_{k', j'}, \nabla u_{k, j} \right>_T + \alpha \left< [u_{k', j'}], [u_{k, j}] \right>_S - \frac{1}{2} \left< [u_{k', j'}], \left\{ \nabla u_{k, j} \right\} \right>_S \]

\[ - \frac{1}{2} \left< \left\{ \nabla u_{k', j'} \right\}, [u_{k, j}] \right>_S + \left< u_{k', j'}, V_{eff} u_{k, j} \right>_T + \sum_{\ell} \gamma_{\ell} \left< u_{k', j'}, b_{\ell} \right>_T \left< b_{\ell}, u_{k, j} \right>_T \]

\[ \mathcal{T} = \text{elements} \]
\[ \mathcal{S} = \text{element surfaces} \]
\[ u_{k, j} = j\text{th basis function in } k\text{th element} \]
\[ \left\{ \cdot \right\} \text{ and } [\cdot] = \text{average and jump operators across surfaces} \]

• Discretized Kohn-Sham equations:

\[ H_{DG} c_i = \varepsilon_i c_i \]

• Wavefunctions, density:

\[ \psi_i = \sum_{E_k \in \mathcal{T}} \sum_{j=1}^{J_k} c_{i; k, j} u_{k, j} \quad \rho = \sum_{E_k \in \mathcal{T}} \sum_{i=1}^{N} \sum_{j=1}^{J_k} |c_{i; k, j} u_{k, j}|^2 \]

• Energy:

\[ E_{\text{tot}} = \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} \iint \frac{\rho(x) \rho(y)}{|x - y|} \, dx \, dy + \int \epsilon_{xc}[\rho(x)] \, dx - \int \epsilon'_{xc}[\rho(x)] \rho(x) \, dx \]

Observation 2: Choice of Algorithmic Technique

• Focus on traditional (cubic scaling) techniques - but choose technique with lowest pre-constant: Ensures full range of applicability.

• Focus on iterative strategies - direct methods suffer from scaling issues, larger pre-constants.

• But many iterative strategies require knowledge of good preconditioners: not easily achievable!

• Chebyshev Polynomial Filtered Subspace Iterations:

  • Low pre-constant if efficient Hamiltonian matrix times block of wave-function vectors product is available.

  • Is known to far outperform other competing methods (observed in several codes PARSEC, ClusterES, …).

  • Works well in a molecular dynamics / geometry optimization setting: refinement of good guess of wave functions from previous step.

  • Basic idea goes back to the RITZIT code by Rutishauser (1970), application to DFT by Zhou, Chelikowsky and Saad (2000s).
Observation 2: Chebyshev Polynomial Filtered Subspace Iterations

- **Subspace iteration**
  - Orthonormalize a guess block of vectors $Y$.
  - Solve the Rayleigh-Ritz problem on a projected (occupied) subspace: $(Y^T H Y) Q = \mu Q$.
  - Subspace rotation: Set $Y = YQ$.
  - Accelerate this by polynomial filtering.

- **Chebyshev Polynomial Filtering: Compute $p_m(H_{\text{scaled}})Y$**
  - $p_m(\cdot)$ is the Chebyshev polynomial of degree $m$.
  - $H_{\text{scaled}}$ is a scaled version of the Hamiltonian whose occupied part of the spectrum lies below -1 and unoccupied part lies in [-1,1].
  - $Y$ is the block of eigenvectors.
  - Requires Hamiltonian times block of vectors product.

- **DFT setting: Do this on every SCF step.**
  - Successively improves wavefunction guess with progress of SCF.
Implementation of CheFSI in DG-DFT

• The DG Hamiltonian
  - Sparse.
  - Has a block structure.
  - Relatively low spectral width: High filter orders are not necessary.

• Hamiltonian times vector product: 2 level parallelization.
  - 1st level: Domain decomposition based on DG partition - nearest neighbor communication.
  - 2nd level: Band parallelization.

• Subspace iteration parallelization
  - Subspace problem size < 25,000 * 25,000
  - Dense problem - ScALAPACK solution.
  - Parallelized using 1 row of process grid (= # of DG elements).

\[ M = \text{Number of DG elements} \]

\[
\begin{array}{cccc}
1 & P_e + 1 & 2P_e + 1 & \cdots \\
2 & P_e + 2 & 2P_e + 2 & \cdots \\
3 & P_e + 3 & 2P_e + 3 & \cdots \\
\vdots & \vdots & \vdots & \vdots \\
P_e & 2P_e & 3P_e & \cdots \\
\end{array}
\]

\[
P_e = \text{Number of processors per DG element}
\]

Amartya Banerjee: Pushing the Envelope of Large Scale First Principles Simulations of Non-Insulating Systems
MolSSI Workshop and ELSI Conference, Richmond, VA; August 2018
Implementation of CheFSI in DG-DFT (cont.)

- The nuances of an on-the-fly basis.
  - The basis functions in DG-DFT change on every SCF step.
  - Wave-function guess from previous SCF step has to be expressed using the basis functions of the current SCF step.
  - Compute: \( Y_{\text{new}} = (V_{\text{new}})^T V_{\text{old}} Y_{\text{old}} \).
  - Here, \( Y \): wave-function expansion coefficients, \( V \): basis functions.
Results

• Comparison of CheFSI with 2 alternatives in DG-DFT
  • CheFSI : Chebyshev Polynomial Filtered Subspace Iterations.
  • ScaLAPACK : Direct diagonalization of DG Hamiltonian.
  • PEXSI : Pole Expansion and Selected Inversion : Sparse direct method with lower asymptotic computational complexity and very favorable parallel scaling properties.

• Two representative systems :
  • Graphene2D : Two dimensional graphene layer (unit cell : 180 atoms).
  • Li3D : Bulk Lithium Ion battery material with atoms of Li, C, F, O, P, H. (unit cell: 318 atoms)
  • Unit cells repeated to obtain larger system sizes (up to ~ 8600 atoms in 3D, ~12500 atoms in 2D).
Results: SCF Convergence

- Parameters carefully tuned for similar SCF convergence
  - ScaLAPACK: reference
  - CheFSI: Adjust filter order
  - PEXSI: Adjust number of poles, etc.
  - Time to solution per SCF step compared.
Results: Strong Scaling Efficiency

- **Li3D_{2^2*2}**: 2544 atoms, ~ 4400 bands
- **Graphene2D_{6480}**: 6480 atoms, ~ 13,000 bands

- Overall, CheFSI has a parallel scaling efficiency of about 40 - 50 % for up to about 12,500 processors.
- Scaling performance better than ScaLAPACK but not as good as PEXSI.
- Beyond ~ 8000 bands, the relatively poorly scaling subspace problem starts to dominate.
- The filtering routine continues to perform very well (~ 80 % for 3D system, ~ 90 % for 2D system). Similar to PEXSI or better.
Results: Weak Scaling Efficiency

- Anticipate CheFSI time will quadruple on doubling system size.
- Increase system size n fold, increase number of processors by m fold, graph should have a slope of $n^2/m$.
- Li3D: Li3D$^1$$^1$$^1$, Li3D$^1$$^1$$^2$, Li3D$^1$$^2$$^2$
- Graphene2D: Graphene2D$_{180}$, Graphene2D$_{720}$, Graphene2D$_{2880}$

![Graph of Wall Time vs Number of Computational Cores]

- CheFSI performance consistent with quadratic scaling with system size as long as subspace iteration cost does not dominate.
- Weak scaling efficiency is ~ 70 %, Filtering step scaling efficiency is ~ 90 %
## Results: Wall time benchmarks

<table>
<thead>
<tr>
<th>System</th>
<th>No. of Atoms</th>
<th>No. of bands</th>
<th>Wall Time (s) ScaLAPACK</th>
<th>Wall Time (s) PEXSI</th>
<th>Wall Time (s) CheFSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li3D(_{3 \times 3 \times 3}) 13,824 procs</td>
<td>8586</td>
<td>~15,000</td>
<td>3323</td>
<td>3784</td>
<td>170</td>
</tr>
<tr>
<td>Graphene 2D(_{11520}) 13,824 procs</td>
<td>11520</td>
<td>&gt; 23000</td>
<td>2473</td>
<td>426</td>
<td>105</td>
</tr>
</tbody>
</table>

- CheFSI is up to an order of magnitude faster.
Results: Wall time benchmarks on ~ 55,000 cores

<table>
<thead>
<tr>
<th>System</th>
<th>No. of Atoms</th>
<th>No. of bands</th>
<th>Wall Time (s) CheFSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li3D$_{3\times3\times3}$ 13,824 procs</td>
<td>8586</td>
<td>~15,000</td>
<td>90</td>
</tr>
<tr>
<td>Graphene 2D$_{11520}$ 13,824 procs</td>
<td>11520</td>
<td>&gt; 23000</td>
<td>75</td>
</tr>
</tbody>
</table>

- 50 - 60 % time spent on subspace problem
- More recently (softer pseudopotentials, optimization of parameters): Li3D system ~ 70 s, Graphene2D ~ 60 s
Summary: Part I (CheFSI within DG-DFT)

- Chebyshev Polynomial Filtered Subspace Iterations seem to work extremely well within the discontinuous formulation of DFT.
- Simulation wall times can be up to an order of magnitude shorter than competing methods.
- Allows complex materials systems with many thousands of atoms to be routinely attacked: Wavefunction and density extrapolation for more effective use of CheFSI for MD simulations.
- Overall performance and strong parallel scalability of CheFSI is held down by the wall time and poor scalability of the subspace problem solution: Remedies in Part II!
Summary : Part I (CheFSI within DG-DFT)

• Chebyshev Polynomial Filtered Subspace Iterations seem to work extremely well within the discontinuous formulation of DFT.

• Simulation wall times can be up to an order of magnitude shorter than competing methods.

• Allows complex materials systems with many thousands of atoms to be routinely attacked: Wavefunction and density extrapolation for more effective use of CheFSI for MD simulations.

• Overall performance and strong parallel scalability of CheFSI is held down by the wall time and poor scalability of the subspace problem solution : Remedies in Part II !
Outline

• Introduction

• Part I : Chebyshev Filtered Subspace Iterations (CheFSI) Within the Framework of Discontinuous Galerkin Density Functional Theory for Large Scale Abinitio Simulations

• Part II : Complementary Subspace Strategy for confronting the Rayleigh-Ritz bottle neck

• Part III : Periodic Pulay method for robust and efficient convergence acceleration of self-consistent field iterations

• Broad guiding principles : Maintain generality - as many systems as possible (metals, insulators…)

• Reduce pre-constants.
Part II: Complementary Subspace Strategy

- The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_{\text{eff}} \psi_i(x) = \varepsilon_i \psi_i(x),\]

\[\hat{V}_{\text{eff}} = V_I^L + \hat{V}^{nL}_I + V_H + V_{xc},\]

\[V_I^L = \sum_a V_{I,a}(x),\]

\[\hat{V}^{nL}_I \psi_i = \sum_a \int dx' V^{nL}_{I,a}(x, x') \psi_i(x'),\]

\[V_H = -\int dx' \frac{\rho_e(x')}{|x - x'|},\]

\[V_{xc} = V_{xc}(x; \rho_e),\]

\[\rho_e = -\sum_i f_i \psi_i^*(x) \psi_i(x),\]

- Linear Problem: Solve the Schrödinger equation in an efficient and scalable way.

SCF Iterations

Start (guess): \(\rho_{in}\)

Repeat:

\[\hat{V}_{\text{eff}}[\rho_{in}]\]

\[\{\psi_i, \varepsilon_i\}_{i=1}^m\]

\[\rho_{out}\]

\[\text{new } \rho_{in}\]

\[\text{no (mix)}\]

\[\rho_{out} = \rho_{in}\?\]

\[\text{yes}\]

Done:

\[\{\psi_i, \varepsilon_i\}_{i=1}^m \Rightarrow \text{materials properties}\]
Algorithm 1 CheFSI cycle

Input: Matrix $H^{DG}$, starting vector block $X$, filter order $m$

1. Compute lower bound $b_{\text{low}}$ using previous Ritz values and the upper bound $b_{\text{up}}$ using a few steps of the Lanczos algorithm.

2. Perform Chebyshev polynomial filtering, i.e., compute $\tilde{Y} = p_m(H^{DG})X$ with $[b_{\text{low}}, b_{\text{up}}]$ mapped to $[-1, 1]$.

3. Orthonormalize columns of $\tilde{Y}$: Set $S = \tilde{Y}^T \tilde{Y}$, compute $U^T U = S$, and solve $\tilde{Y} U = \tilde{Y}$.

4. Rayleigh-Ritz step: Compute the projected subspace matrix $\tilde{H} = \tilde{Y}^T H^{DG} \tilde{Y}$ and solve the eigenproblem $\tilde{H} Q = Q D$.

5. Perform a subspace rotation step $X_{\text{new}} = \tilde{Y} Q$.

Output: Vector block $X_{\text{new}}$ (approximate eigenvectors) and Ritz values $D$ (approximate eigenvalues).
Complementary Subspace Strategy (cont.)

**Algorithm 1** CheFSI cycle

**Input:** Matrix $H^{DG}$, starting vector block $X$, filter order $m$

1. Compute lower bound $b_{\text{low}}$ using previous Ritz values and the upper bound $b_{\text{up}}$ using a few steps of the Lanczos algorithm.

2. Perform Chebyshev polynomial filtering, i.e., compute $\tilde{Y} = p_m(H^{DG})X$ with $[b_{\text{low}}, b_{\text{up}}]$ mapped to $[-1, 1]$.

3. Orthonormalize columns of $\tilde{Y}$: Set $S = \tilde{Y}^T\tilde{Y}$, compute $U^TU = S$, and solve $\tilde{Y}U = \tilde{Y}$.

4. **Rayleigh-Ritz step:** Compute the projected subspace matrix $\hat{H} = \tilde{Y}^TH^{DG}\tilde{Y}$ and solve the eigenproblem $\hat{H}Q = QD$.

5. Perform a subspace rotation step $X_{\text{new}} = \hat{Y}Q$.

**Output:** Vector block $X_{\text{new}}$ (approximate eigenvectors) and Ritz values $D$ (approximate eigenvalues).

- Focus on the Rayleigh-Ritz (subspace diagonalization) + subspace rotation steps in the DG-CheFSI cycle
Complementary Subspace Strategy (cont.)

- Large benchmark problems: 50 - 60% time spent on subspace problem

```
Algorithm 1 CheFSI cycle

Input: Matrix $H^{DG}$, starting vector block $X$, filter order $m$

1. Compute lower bound $b_{\text{low}}$ using previous Ritz values and the upper bound $b_{\text{up}}$ using a few steps of the Lanczos algorithm.
2. Perform Chebyshev polynomial filtering, i.e., compute $\hat{Y} = p_m(H^{DG})X$ with $[b_{\text{low}}, b_{\text{up}}]$ mapped to $[-1, 1]$.
3. Orthonormalize columns of $\hat{Y}$: Set $S = Y^T \hat{Y}$, compute $U^T U = S$, and solve $YU = \hat{Y}$.
4. Rayleigh-Ritz step: Compute the projected subspace matrix $H = Y^T H^{DG} Y$ and solve the eigenproblem $HQ = QD$.
5. Perform a subspace rotation step $X_{\text{new}} = \hat{Y} Q$.

Output: Vector block $X_{\text{new}}$ (approximate eigenvectors) and Ritz values $D$ (approximate eigenvalues).
```

- Highly Scalable
- Can be made scalable (parallel dense linear algebra: SUMMA, GEMM 2.5D)
- Not intrinsically scalable,
- Consumes significant fraction of subspace problem time

- Need new strategy for the Rayleigh-Ritz and Subspace Rotation steps
Complementary Subspace Strategy within DG-CheFSI

- Electronic occupation function (Fermi-Dirac):
  
  ![Fermi-Dirac distribution for several temperatures](Image courtesy of internet resources)

  - At reasonable electronic temperatures, only a small fraction of states are partially occupied.
  - Can work with just these few states!

- Allows projected density matrix to be written as:
  
  \[ \tilde{D} = \tilde{I} - C \cdot C^T \]
  
  \[ \tilde{D}, \tilde{I}: N_s \times N_s \]
  
  \[ C : N_s \times N_t \]

- The matrix C is related to the \( N_t \) top few (~5 - 10% for metals at room temperature) eigenstates of the projected Hamiltonian matrix.

- Fermi and occupations level need to be solved for in a suitable way.

- DGDFD : Full density matrix can be constructed from projected density matrix in an efficient and scalable manner.
Complementary Subspace Strategy (cont.)

Complementary Subspace Strategy (cont.)

• Complementary subspace strategy: “approximation to Density Matrix”
  
  • How to compute the matrix C? Compute top states of projected Hamiltonian via iterative solvers: LOBPCG (unpreconditioned) or 2nd level of CheFSI!
  
  • Adjustable parameters: Accuracy of top states, how many top states.

• “Two level CheFSI scheme”: CS2CF
  
  • First or top level CheFSI evaluates subspace of occupied states of full Hamiltonian
  
  • Second or inner level CheFSI evaluates fractionally occupied states of projected Hamiltonian (small spectral width, very low filter order)
  
  • Parallelizes easily: Most operations are parallel dense GEMMs (ScaLAPACK)

• Not necessarily of lower computational complexity, but lower pre-factor and can lead to significantly lower computational wall times.
**Results : Choice of Systems**

<table>
<thead>
<tr>
<th>System</th>
<th>Type</th>
<th>No. of atoms in unit cell</th>
<th>Types of atoms (elements)</th>
<th>No. of electrons in unit cell</th>
<th>ALBs per atom in DGDFD</th>
<th>Representative image of unit cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrolyte3D</td>
<td>Bulk, Insulating</td>
<td>318</td>
<td>C,H,F,Li,O</td>
<td>1104</td>
<td>40</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>SiDiamond3D</td>
<td>Bulk, Semiconducting</td>
<td>8</td>
<td>Si</td>
<td>16</td>
<td>40</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>Graphene2D</td>
<td>2D, Semimetallic</td>
<td>180</td>
<td>C</td>
<td>360</td>
<td>20</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>LiBCC3D</td>
<td>Bulk, Metallic</td>
<td>16</td>
<td>Li</td>
<td>48</td>
<td>35</td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>CuFCC3D</td>
<td>Bulk, Metallic</td>
<td>4</td>
<td>Cu</td>
<td>44</td>
<td>50</td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- Wide variety of materials systems tested out: disordered bulk insulator, 2D-semi metal, semiconductor, bulk crystalline metals, etc…
- Unit cells of these materials replicated along X-Y-Z axes to obtain larger system sizes.
Results: SCF Convergence

- SCF converges at similar rates as original CheFSI
## Results: Energy and Force Accuracy

<table>
<thead>
<tr>
<th>System</th>
<th>CS Strategy with LOBPCG for top states</th>
<th>CS Strategy with CheFSI for top states (CS2CF strategy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Energy per atom difference (Ha)</td>
<td>Max. force component difference (Ha/Bohr)</td>
</tr>
<tr>
<td>Electrolyte3D(_{1 \times 2 \times 2})</td>
<td>(5 \times 10^{-5})</td>
<td>(2 \times 10^{-4})</td>
</tr>
<tr>
<td>SiDiamond3D(_{5 \times 5 \times 5})</td>
<td>(4 \times 10^{-6})</td>
<td>(4 \times 10^{-5})</td>
</tr>
<tr>
<td>Graphene2D(_{1 \times 2 \times 2})</td>
<td>(7 \times 10^{-6})</td>
<td>(8 \times 10^{-5})</td>
</tr>
<tr>
<td>CuFCC3D(_{5 \times 5 \times 5})</td>
<td>(8 \times 10^{-6})</td>
<td>(9 \times 10^{-5})</td>
</tr>
<tr>
<td>LiBCC3D(_{4 \times 4 \times 4})</td>
<td>(7 \times 10^{-6})</td>
<td>(9 \times 10^{-5})</td>
</tr>
</tbody>
</table>

- Energies and forces compare to 2e-4 a.u. or better accuracy w.r.t reference standard CheFSI results (< discretization error).
Results: Computational Efficiency

- Largest non-insulating systems: Subspace diagonalization and rotation step times brought down by factors of ~ 3.7 - 7.8
- Insulating system: Subspace diagonalization and rotation step time brought down by factor of 60
Results: Computational Efficiency (cont.)

- Overall subspace problem construction and solution wall time is brought down by a factor of ~1.7 - 2.2 for all examples.

- Gains can be made more significant by use of alternate smearing schemes (Gaussian / Methfessel-Paxton, etc.)
Results: SCF iteration wall times (cont.)

- SCF iteration wall times with DGDFT-CS2CF

<table>
<thead>
<tr>
<th>System</th>
<th>No. of atoms (No. of electrons)</th>
<th>Total computational cores (cores used in subspace prob.)</th>
<th>ALB generation [s]</th>
<th>Hamiltonian update [s]</th>
<th>CS2CF strategy (subspace prob. time) [s]</th>
<th>Total SCF wall time via CS2CF strategy [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrolyte3D$_{3\times3\times3}$</td>
<td>8,586 (29,808)</td>
<td>34,560 (3,456)</td>
<td>12</td>
<td>4</td>
<td>34 (19)</td>
<td>50</td>
</tr>
<tr>
<td>SiDiamond3D$_{10\times10\times10}$</td>
<td>8,000 (32,000)</td>
<td>34,560 (3,456)</td>
<td>9</td>
<td>2</td>
<td>40 (24)</td>
<td>51</td>
</tr>
<tr>
<td>Graphene2D$_{8\times8}$</td>
<td>11,520 (23,040)</td>
<td>27,648 (4,608)</td>
<td>4</td>
<td>2</td>
<td>35 (27)</td>
<td>41</td>
</tr>
<tr>
<td>CuFCC3D$_{10\times10\times10}$</td>
<td>4,000 (44,000)</td>
<td>30,000 (3,000)</td>
<td>20</td>
<td>9</td>
<td>75 (46)</td>
<td>104</td>
</tr>
<tr>
<td>LiBCC3D$_{12\times12\times12}$</td>
<td>27,648 (82,944)</td>
<td>38,880 (12,960)</td>
<td>22</td>
<td>13</td>
<td>180 (165)</td>
<td>215</td>
</tr>
</tbody>
</table>
### Results: SCF iteration wall times and Ab initial MD

- SCF iteration wall times with DGDFT-CS2CF and ELPA

<table>
<thead>
<tr>
<th>System</th>
<th>No. of atoms (No. of electrons)</th>
<th>Total computational cores (cores used in subspace prob.)</th>
<th>CS2CF strategy (subspace prob. time)</th>
<th>Total SCF wall time via CS2CF strategy</th>
<th>Direct diagonalization via ELPA</th>
<th>Total SCF wall time via ELPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrolyte3D$_{3\times3\times3}$</td>
<td>8,586 (29,808)</td>
<td>34,560 (3,456)</td>
<td>34 (19)</td>
<td>50</td>
<td>647</td>
<td>663</td>
</tr>
<tr>
<td>SiDiamond3D$_{10\times10\times10}$</td>
<td>8,000 (32,000)</td>
<td>34,560 (3,456)</td>
<td>40 (24)</td>
<td>51</td>
<td>648</td>
<td>659</td>
</tr>
<tr>
<td>Graphene2D$_{8\times8}$</td>
<td>11,520 (23,040)</td>
<td>27,648 (4,608)</td>
<td>35 (27)</td>
<td>41</td>
<td>262</td>
<td>268</td>
</tr>
<tr>
<td>CuFCC3D$_{10\times10\times10}$</td>
<td>4,000 (44,000)</td>
<td>30,000 (3,000)</td>
<td>75 (46)</td>
<td>104</td>
<td>199</td>
<td>228</td>
</tr>
<tr>
<td>LiBCC3D$_{12\times12\times12}$</td>
<td>27,648 (82,944)</td>
<td>38,880 (12,960)</td>
<td>180 (165)</td>
<td>215</td>
<td>5844</td>
<td>5879</td>
</tr>
</tbody>
</table>
Results: SCF iteration wall times and Ab initial MD

- SCF iteration wall times with DGDFT-CS2CF and ELPA

<table>
<thead>
<tr>
<th>System</th>
<th>No. of atoms (No. of electrons)</th>
<th>Total computational cores (cores used in subspace prob.)</th>
<th>CS2CF strategy (subspace prob. time) [s]</th>
<th>Total SCF wall time via CS2CF strategy [s]</th>
<th>Direct diagonalization via ELPA [s]</th>
<th>Total SCF wall time via ELPA [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrolyte3D&lt;sub&gt;3x3x3&lt;/sub&gt;</td>
<td>8,586 (29,808)</td>
<td>34,560 (3,456)</td>
<td>34 (19)</td>
<td>50</td>
<td>647</td>
<td>663</td>
</tr>
<tr>
<td>SiDiamond3D&lt;sub&gt;10x10x10&lt;/sub&gt;</td>
<td>8,000 (32,000)</td>
<td>34,560 (3,456)</td>
<td>40 (24)</td>
<td>51</td>
<td>648</td>
<td>659</td>
</tr>
<tr>
<td>Graphene2D&lt;sub&gt;8x8&lt;/sub&gt;</td>
<td>11,520 (23,040)</td>
<td>27,648 (4,608)</td>
<td>35 (27)</td>
<td>41</td>
<td>262</td>
<td>268</td>
</tr>
<tr>
<td>CuFCC3D&lt;sub&gt;10x10x10&lt;/sub&gt;</td>
<td>4,000 (44,000)</td>
<td>30,000 (3,000)</td>
<td>75 (46)</td>
<td>104</td>
<td>199</td>
<td>228</td>
</tr>
<tr>
<td>LiBCC3D&lt;sub&gt;12x12x12&lt;/sub&gt;</td>
<td>27,648 (82,944)</td>
<td>38,880 (12,960)</td>
<td>180 (165)</td>
<td>215</td>
<td>5844</td>
<td>5879</td>
</tr>
</tbody>
</table>
Results: SCF iteration wall times and Ab initio MD

- SCF iteration wall times with DGDFT-CS2CF and ELPA

<table>
<thead>
<tr>
<th>System</th>
<th>No. of atoms (No. of electrons)</th>
<th>Total computational cores (cores used in subspace prob.)</th>
<th>CS2CF strategy (subspace prob. time) [s]</th>
<th>Total SCF wall time via CS2CF strategy [s]</th>
<th>Direct diagonalization via ELPA [s]</th>
<th>Total SCF wall time via ELPA [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrolyte3D_{3\times 3}</td>
<td>8,586 (29,808)</td>
<td>34,560 (3,456)</td>
<td>31 (19)</td>
<td>50</td>
<td>647</td>
<td>663</td>
</tr>
<tr>
<td>SiDiamond3D_{10\times 10\times 10}</td>
<td>8,000 (32,000)</td>
<td>34,560 (3,456)</td>
<td>40 (24)</td>
<td>51</td>
<td>648</td>
<td>659</td>
</tr>
<tr>
<td>Graphene2D_{8\times 8}</td>
<td>11,520 (23,040)</td>
<td>27,648 (4,608)</td>
<td>35 (27)</td>
<td>41</td>
<td>262</td>
<td>268</td>
</tr>
<tr>
<td>CuFCC3D_{10\times 10\times 10}</td>
<td>4,000 (44,000)</td>
<td>30,000 (3,000)</td>
<td>75 (46)</td>
<td>104</td>
<td>199</td>
<td>228</td>
</tr>
<tr>
<td>LiBCC3D_{12\times 12\times 12}</td>
<td>27,648 (82,944)</td>
<td>38,880 (12,960)</td>
<td>180 (165)</td>
<td>215</td>
<td>5844</td>
<td>5879</td>
</tr>
</tbody>
</table>

- Abinitio MD simulation example: 10 * 10 * 10 Diamond Si system 8000 atoms; 32,000 electrons.
  - 300 K simulation, 2.5 fs time step, NVE
  - 34,560 MPI tasks, 12 * 12 * 12 DGDFT partition
  - 1 ps simulation takes ~ 28 hours (51 s / SCF step)
  - Drift remains small (<= 1e-5 Ha / atom picosecond)
Summary & Future Directions: Part II (CS2CF within DG-DFT)

- Complementary subspace strategy to confront Rayleigh-Ritz bottleneck (particularly: metallic systems).

- Two level CheFSI strategy (CS2CF) implemented in DG-DFT.

- Abinitio MD / geometry optimization of complex materials systems with thousands of atoms can be routinely carried out on moderate resources.

- Applications in various materials systems.

- Next bottleneck to confront: orthonormalization! Ways to reduce prefactor?

- Spectrum-Slicing strategies?
Outline

• Introduction

• Part I : Chebyshev Filtered Subspace Iterations (CheFSI) Within the Framework of Discontinuous Galerkin Density Functional Theory for Large Scale Abinitio Simulations

• Part II : Complementary Subspace Strategy for confronting the Rayleigh-Ritz bottle neck

• **Part III** : Periodic Pulay method for robust and efficient convergence acceleration of self-consistent field iterations

  • Broad guiding principles : Maintain generality - as many systems as possible (metals, insulators…)

  • Reduce pre-constants.
Part III : Periodic Pulay Mixing

- The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_\text{eff} \psi_i(x) = \varepsilon_i \psi_i(x),\]

\[\hat{V}_\text{eff} = V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc},\]

\[V_I^\ell = \sum_a V_{I,a}(x),\]

\[\hat{V}_I^{n\ell} \psi_i = \sum_a \int dx' V_{I,a}^{n\ell}(x, x') \psi_i(x'),\]

\[V_H = - \int dx' \frac{\rho_e(x')}{|x - x'|},\]

\[V_{xc} = V_{xc}(x; \rho_e),\]

\[\rho_e = - \sum_i f_i \psi_i^\ast(x) \psi_i(x),\]

**SCF Iterations**

Start (guess): \(\rho_{in}\)

\[\hat{V}_\text{eff}[\rho_{in}]\]

\[{(\psi_i, \varepsilon_i)_{i=1}^m}\]

\[\rho_{out}\]

Repeat:

new \(\rho_{in}\)

\[\rho_{out} = \rho_{in}\?\]

no \(\to \) \(\rho_{in}\)

yes \(\to \) \{\psi_i, \varepsilon_i\}_{i=1}^m \Rightarrow \text{materials properties}\]
Part III : Periodic Pulay Mixing

• The problem at hand:

\[-\frac{1}{2} \nabla^2 \psi_i(x) + \hat{V}_{\text{eff}} \psi_i(x) = \varepsilon_i \psi_i(x),\]

\[\hat{V}_{\text{eff}} = V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc},\]

\[V_I^\ell = \sum_a V_{I,a}(x),\]

\[\hat{V}_I^{n\ell} \psi_i = \sum_a \int dx' V_{I,a}^{n\ell}(x, x') \psi_i(x'),\]

\[V_H = -\int dx' \frac{\rho_e(x')}{|x - x'|},\]

\[V_{xc} = V_{xc}(x; \rho_e),\]

\[\rho_e = -\sum_i f_i^s \psi_i^s(x) \psi_i(x),\]

• Non-linear Problem: Robust attainment of self-consistency.

SCF Iterations

Start (guess): \(\rho_{in}\)

\[\hat{V}_{\text{eff}}[\rho_{in}]\]

\[\{\psi_i, \varepsilon_i\}_{i=1}^m\]

\[\rho_{out}\]

Repeat:

\[\text{new } \rho_{in}\]

\[\text{no mix} \]

Yes

Done:

\[\{\psi_i, \varepsilon_i\}_{i=1}^m \Rightarrow \text{materials properties}\]

Amartya Banerjee : Pushing the Envelope of Large Scale First Principles Simulations of Non-Insulating Systems
MolSSI Workshop and ELSI Conference, Richmond, VA ; August 2018
Mixing Schemes within DFT

- Required for convergence acceleration.
- Total simulation wall time strongly influenced (1st - order effect).
- Many mixing schemes in the literature.

- Basic mathematical setup:

  \[ \rho = g(\rho) \]

  \( \rho \in \mathbb{R}^{N \times 1} \)
  \( g : \mathbb{R}^{N \times 1} \rightarrow \mathbb{R}^{N \times 1} \)
  \( f(\rho) = g(\rho) - \rho \)

  - Given electron density
  - Evaluation of new electron density from given density
  - The residual

- Convergence in the neighborhood of a solution dictated by properties of Jacobian evaluated at solution.
Mixing Schemes within DFT (cont.)

- **Linear mixing**: Under relaxed fixed point iteration
  \[ \rho_{i+1} = \rho_i + C_i f_i \]
  \[ f_i = f(\rho_i) \quad C_i = \alpha I \]
  - Guaranteed to converge for small enough damping.
  - Extremely slow and rarely used in practice.
  - In practice (most widely used): **Pulay mixing (DIIS)**
Mixing Schemes within DFT (cont.)

- **Pulay mixing (DIIS):** A quasi-Newton Broyden type method

\[
\rho_{i+1} = \rho_i + C_i f_i
\]

\[
f_i = f(\rho_i) \quad C_i = \alpha I - (R_i + \alpha F_i)(F_i^T F_i)^{-1} F_i^T
\]

\[
R_i = [\Delta \rho_{i-n+1}, \Delta \rho_{i-n+2}, \ldots, \Delta \rho_i] \in \mathbb{R}^{N \times n},
\]

\[
F_i = [\Delta f_{i-n+1}, \Delta f_{i-n+2}, \ldots, \Delta f_i] \in \mathbb{R}^{N \times n},
\]

\[
\Delta \rho_i = \rho_i - \rho_{i-1} \quad \Delta f_i = f_i - f_{i-1}
\]

- Approximation of inverse Jacobian
- Can stagnate or perform poorly in complex materials
Periodic Pulay Mixing

- Combine strategies: Pulay and Linear mixing
- Heuristic: Combination of strategies likely to help escape metastable wells.
- Wavelength analysis: Linear mixing (or damping) eliminates high frequency components of residual most effectively. Anderson extrapolation damps lower frequency components.
- Application of AAJ to SCF iterations: **Periodic Pulay Mixing**
Periodic Pulay Mixing (cont.)

• Extremely easy to integrate into existing electronic structure codes.

\[\textbf{Algorithm 1.} \quad \text{Periodic Pulay method.}\]

\textbf{Input:} \(\rho_0, \alpha, n, k, \text{ and } tol\)
\[\text{repeat } i = 0, 1, 2 \ldots \]
\[f_i = g(\rho_i) - \rho_i\]
\[\text{if } (i + 1)/k \in \mathbb{N} \text{ then}\]
\[\rho_{i+1} = \rho_i + \alpha f_i - (R_i + \alpha F_i)(F_i^T F_i)^{-1} F_i^T f_i\]
\[\text{else}\]
\[\rho_{i+1} = \rho_i + \alpha f_i\]
\[\text{until } \|f_i\| < tol;\]
\[\textbf{Output: } \rho_i\]

• Implemented in DG-DFT, ClusterES and SIESTA
Results

- Variety of materials systems chosen: insulators, conductors, large inhomogeneities, different geometries, magnetic systems, etc.

Results include:

1. Variety of materials systems chosen: insulators, conductors, large inhomogeneities, different geometries, magnetic systems, etc.

2. Convergence statistics: fixed mixing parameters, varying histories, periods

---

<table>
<thead>
<tr>
<th>Materials test cases.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>Benzene molecule (12 atoms)</td>
</tr>
<tr>
<td>32-H$_2$O</td>
<td>32 molecules of water (96 atoms)</td>
</tr>
<tr>
<td>Pd-bulk</td>
<td>$3 \times 3 \times 3$ FCC bulk Palladium (108 atoms) $\Gamma$-point calculation</td>
</tr>
<tr>
<td>Pt$_{13}$</td>
<td>Octahedral platinum cluster (13 atoms)</td>
</tr>
<tr>
<td>Fe$_3$-noncollinear</td>
<td>Iron cluster with non-collinear spin (3 atoms)</td>
</tr>
<tr>
<td>Fe-bulk</td>
<td>Bulk BCC iron (1 atom), Brillouin zone integration with 63 k-points</td>
</tr>
<tr>
<td>SiC-slab</td>
<td>Silicon carbide slab with hydrogen saturated surface (78 atoms)</td>
</tr>
<tr>
<td>ptdca-Au</td>
<td>Molecules of 3, 4, 9, 10 perylenetetracarboxylic dianhydride adsorbed onto gold (111) surface (204 atoms)</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>System</th>
<th>Pulay</th>
<th>Periodic Pulay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>Benzene</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>32-H$_2$O</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>Pd-bulk</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Pt$_{13}$</td>
<td>98</td>
<td>33</td>
</tr>
<tr>
<td>Fe$_3$-noncollinear</td>
<td>112</td>
<td>38</td>
</tr>
<tr>
<td>Fe-bulk</td>
<td>72</td>
<td>31</td>
</tr>
<tr>
<td>SiC-slab</td>
<td>51</td>
<td>10</td>
</tr>
<tr>
<td>ptdca-Au</td>
<td>89</td>
<td>11</td>
</tr>
</tbody>
</table>
Results (cont.)

- Convergence behavior

![Graphs showing convergence behavior](image)

Pd Bulk

SiC slab

- Convergence statistics for different mixing parameters (fixed histories, etc.)

<table>
<thead>
<tr>
<th>System</th>
<th>Pulay</th>
<th>Periodic Pulay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>Pd-bulk</td>
<td>56</td>
<td>18</td>
</tr>
<tr>
<td>SiC-slab</td>
<td>26</td>
<td>4</td>
</tr>
</tbody>
</table>
Summary: Part III (Periodic Play Mixing)

- Periodic Pulay Mixing combines linear mixing with periodic Anderson / Pulay extrapolation.
- Performance gains of up to 2 - 3 in most systems.
- Convergence achieved even in many difficult systems.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>Pulay</th>
<th>Periodic Pulay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k = 2$</td>
</tr>
<tr>
<td>100</td>
<td>78</td>
<td>25</td>
</tr>
<tr>
<td>50</td>
<td>164</td>
<td>75</td>
</tr>
<tr>
<td>25</td>
<td>a</td>
<td>39</td>
</tr>
</tbody>
</table>

Bulk Iron at low temperatures

- Rigorous mathematical understanding is now emerging (M. Benzi et. al)
Associated Publications

- Chebyshev polynomial filtered subspace iteration in the discontinuous Galerkin method for large-scale electronic structure calculations: AS Banerjee, L Lin, W Hu, C Yang, JE Pask; The Journal of Chemical Physics 145 (15), 154101

- Two-Level Chebyshev Filter Based Complementary Subspace Method: Pushing the Envelope of Large-Scale Electronic Structure Calculations: AS Banerjee, L Lin, P Suryanarayana, C Yang, JE Pask; Journal of Chemical Theory and Computation; Volume 14 (6), Pages 2930 - 2946.

Acknowledgments

- NERSC computing facilites
Thank You!