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

Amartya Banerjee^{1,2}

Joint work with L. Lin^{1,3}, C. Yang¹, W. Hu¹, J. Pask⁴ and P. Suryanarayana⁵ ¹Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA ²Mechanical, Aerospace & Biomedical Engineering, University of Tennessee, Knoxville, TN, USA ³Department of Mathematics, University of California, Berkeley, CA, USA ⁴Physics Division, Lawrence Livermore National Laboratory, Livermore, CA, USA ⁵College of Engineering, Georgia Institute of Technology, GA, USA

MolSSI Workshop and ELSI Conference

Richmond, Virginia.

August 17, 2018





Introduction

 Energy storage devices : Lithium lon batteries





CHARGING

DISCHARGING

• In practice reliability / safety is a major issue









Anode-electrolyte interface at the atomic scales

- Often related to stability of anodeelectrolyte interface
- Need for accurate "bottom-up approach" to simulating these interfaces.

Images courtesy of internet resources

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

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 :

$$\begin{split} &-\frac{1}{2}\nabla^2\psi_i(\mathbf{x}) + \hat{V}_{eff}\psi_i(\mathbf{x}) = \varepsilon_i\psi_i(\mathbf{x}), \\ &\hat{V}_{eff} = V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc}, \\ &V_I^\ell = \sum_a V_{I,a}(\mathbf{x}), \\ &\hat{V}_I^{n\ell}\psi_i = \sum_a \int d\mathbf{x}' V_{I,a}^{n\ell}(\mathbf{x},\mathbf{x}')\psi_i(\mathbf{x}'), \\ &V_H = -\int d\mathbf{x}' \frac{\rho_e(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}, \\ &V_{xc} = V_{xc}(\mathbf{x};\rho_e), \\ &\rho_e = -\sum_i f_i\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}), \end{split}$$

SCF Iterations



Kohn-Sham Density Functional Theory

• The problem at hand :

$$egin{aligned} &-rac{1}{2}
abla^2\psi_i(\mathbf{x})+\hat{V}_{eff}\psi_i(\mathbf{x})=arepsilon_i\psi_i(\mathbf{x}),\ \hat{V}_{eff}=V_I^\ell+\hat{V}_I^{n\ell}+V_H+V_{xc},\ V_I^\ell&=\sum_a V_{I,a}(\mathbf{x}),\ \hat{V}_I^{n\ell}\psi_i&=\sum_a \int d\mathbf{x}'V_{I,a}^{n\ell}(\mathbf{x},\mathbf{x}')\psi_i(\mathbf{x}'),\ V_H&=-\int d\mathbf{x}'rac{
ho_e(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|},\ V_{xc}&=V_{xc}(\mathbf{x};
ho_e),\
ho_e&=-\sum_i f_i\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}), \end{aligned}$$

SCF Iterations



- Linear Problem: Solve the Schrödinger equation in an efficient and scalable way.
- Non-linear Problem: Achieve self-consistency in a robust fashion.

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 :

$$\begin{split} -\frac{1}{2} \nabla^2 \psi_i(\mathbf{x}) + \hat{V}_{eff} \psi_i(\mathbf{x}) &= \varepsilon_i \psi_i(\mathbf{x}), \\ \hat{V}_{eff} &= V_I^{\ell} + \hat{V}_I^{n\ell} + V_H + V_{xc}, \\ V_I^{\ell} &= \sum_a V_{I,a}(\mathbf{x}), \\ \hat{V}_I^{n\ell} \psi_i &= \sum_a \int d\mathbf{x}' V_{I,a}^{n\ell}(\mathbf{x}, \mathbf{x}') \psi_i(\mathbf{x}'), \\ V_H &= -\int d\mathbf{x}' \frac{\rho_e(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}, \\ V_{xc} &= V_{xc}(\mathbf{x}; \rho_e), \\ \rho_e &= -\sum_i f_i \psi_i^*(\mathbf{x}) \psi_i(\mathbf{x}), \end{split}$$

SCF Iterations



Part I: CheFSI within DGDFT

The problem at hand :

SCF Iterations



$$\begin{split} -\frac{1}{2}\nabla^2\psi_i(\mathbf{x}) + \hat{V}_{eff}\psi_i(\mathbf{x}) &= \varepsilon_i\psi_i(\mathbf{x}),\\ \hat{V}_{eff} &= V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc},\\ V_I^\ell &= \sum_a V_{I,a}(\mathbf{x}),\\ \hat{V}_I^{n\ell}\psi_i &= \sum_a \int d\mathbf{x}' V_{I,a}^{n\ell}(\mathbf{x},\mathbf{x}')\psi_i(\mathbf{x}'),\\ V_H &= -\int d\mathbf{x}' \frac{\rho_e(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|},\\ V_{xc} &= V_{xc}(\mathbf{x};\rho_e),\\ \rho_e &= -\sum_i f_i\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}), \end{split}$$

• 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.

[1] Lin, Lu, Ying, E, JCP 2012; [2] Zhang, Lin, Hu, Yang, Pask, JCP 2017

Discontinuous Galerkin Density Functional Theory (cont.)

Adaptive Local Basis Functions.



Kohn-Sham Hamiltonian becomes :

$$H_{DG}(k',j';k,j) = \frac{1}{2} \langle \nabla u_{k',j'}, \nabla u_{k,j} \rangle_{\mathcal{T}} + \alpha \langle [[u_{k',j'}]], [[u_{k,j}]] \rangle_{\mathcal{S}} - \frac{1}{2} \langle [[u_{k',j'}]], \{\{\nabla u_{k,j}\}\} \rangle_{\mathcal{S}} - \frac{1}{2} \langle \{\{\nabla u_{k',j'}\}\}, [[u_{k,j}]] \rangle_{\mathcal{S}} + \langle u_{k',j'}, V_{\text{eff}} u_{k,j} \rangle_{\mathcal{T}} + \sum_{\ell} \gamma_{\ell} \langle u_{k',j'}, b_{\ell} \rangle_{\mathcal{T}} \langle b_{\ell}, u_{k,j} \rangle_{\mathcal{T}}$$

= elements $\mathcal{S} = \text{element surfaces}$

 $u_{k,i} = j$ th basis function in kth element

 $\{\{\cdot\}\}\$ and $[[\cdot]] =$ average and jump operators across surfaces

- Discretized Kohn-Sham equations : $H_{DG}c_i = \varepsilon_i c_i$ Wavefunctions, density : $\psi_i = \sum_{E_k \in \mathcal{T}} \sum_{j=1}^{J_k} c_{i;k,j} u_{k,j}$ $\rho = \sum_{E_k \in \mathcal{T}} \sum_{i=1}^{N} |\sum_{j=1}^{J_k} c_{i;k,j} u_{k,j}|^2$
- Energy :

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

[1] Lin, Lu, Ying, E, JCP 2012; [2] Zhang, Lin, Hu, Yang, Pask, JCP 2017

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

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 : $(Y^THY)Q = \mu Q$.
- Subspace rotation : Set Y = YQ.
- · Accelerate this by polynomial filtering.

Chebyshev Polynomial Filtering: Compute pm(Hscaled)Y

- p_m() is the Chebyshev polynomial of degree m.
- H_{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).

	I	M = Num	ber of DC	d elements	s
element	1	P_e+1	2P _e +1		(M-1)P _e +1
s per DG	2	P _e +2	2P _e +2		(M-1)P _e +2
orocessor	3	P _e +3	2P _e +3		(M-1)P _e +3
mber of ₁				·	÷
$P_e = Nu$	P _e	2 <i>P</i> _e	3P _e		MP _e

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 : $Y_{new} = (V_{new})^T V_{old} Y_{old}$.
- Here, Y : wave-function expansion coefficients, V : basis functions.



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

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).





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

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.



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

Results: Strong Scaling Efficiency



• Li3D_{2*2*2} : 2544 atoms, ~ 4400 bands

- Graphene2D₆₄₈₀ : 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 n²/m.
- Li3D : Li3D_{1*1*1}, Li3D_{1*1*2}, Li3D_{1*2*2}
- Graphene2D : Graphene2D₁₈₀, Graphene2D₇₂₀, Graphene2D₂₈₈₀



- 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

System	No. of Atoms	No. of bands	Wall Time (s) ScaLAPACK	Wall Time (s) PEXSI	Wall Time (s) CheFSI
Li3D _{3*3*3} 13,824 procs	8586	~15,000	3323	3784	170
Graphene 2D ₁₁₅₂₀ 13,824 procs	11520	> 23000	2473	426	105

• CheFSI is up to an order of magnitude faster.

Results : Wall time benchmarks on ~ 55,000 cores

System	No. of Atoms	No. of bands	Wall Time (s) CheFSI
Li3D _{3*3*3} 13,824 procs	8586	~15,000	90
Graphene 2D ₁₁₅₂₀ 13,824 procs	11520	> 23000	75

- 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 :

$$\begin{split} -\frac{1}{2}\nabla^2\psi_i(\mathbf{x}) + \hat{V}_{eff}\psi_i(\mathbf{x}) &= \varepsilon_i\psi_i(\mathbf{x}),\\ \hat{V}_{eff} &= V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc},\\ V_I^\ell &= \sum_a V_{I,a}(\mathbf{x}),\\ \hat{V}_I^{n\ell}\psi_i &= \sum_a \int d\mathbf{x}' V_{I,a}^{n\ell}(\mathbf{x},\mathbf{x}')\psi_i(\mathbf{x}'),\\ V_H &= -\int d\mathbf{x}' \frac{\rho_e(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|},\\ V_{xc} &= V_{xc}(\mathbf{x};\rho_e),\\ \rho_e &= -\sum_i f_i\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}), \end{split}$$

SCF Iterations



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

Algorithm 1 CheFSI cycle

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

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

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

Algorithm 1 CheFSI cycle

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

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

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

Focus on the Rayleigh-Ritz (subspace diagonalization) + subspace rotation steps in the DG-CheFSI cycle

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



• 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

- 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 :

$$\tilde{D} = \tilde{I} - C * C^T \qquad \begin{array}{c} \tilde{D}, \tilde{I} : N_s * N_s \\ C : N_s * N_t \end{array}$$

- The matrix C is related to the Nt 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.
- DGDFT : Full density matrix can be constructed from projected density matrix in an efficient and scalable manner.

Image courtesy of internet resources

 Similar ideas : "Hole Density Matrix" (Mazziotti, 2002, 2003; Bowler et. al, 2016), "Partial Rayleigh-Ritz Method" (Guo et. al, 2016), "Fermi Operator Splitting" (Gavini et. al, 2017)

- 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.

Image courtesy of internet resources

Results : Choice of Systems

System	Туре	No. of atoms in unit cell	Types of atoms (elements)	No. of electrons in unit cell	ALBs per atom in DGDFT	Representative image of unit cell	
Electrolyte3D	Bulk, Insulating	318	C,H,F,Li,O	1104	40		-
SiDiamond3D	Bulk, Semiconducting	8	Si	16	40		 Wide variety of materials systems tested out : disordered bulk insulator, 2D-semi metal, semiconductor, bulk
Graphene2D	2D, Semimetallic	180	С	360	20		 crystalline metals, etc Unit cells of these materials replicated along X-Y-Z axes to obtain larger system
LiBCC3D	Bulk, Metallic	16	Li	48	35		sizes.
CuFCC3D	Bulk, Metallic	4	Cu	44	50		

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

Results : SCF Convergence



Results : Energy and Force Accuracy

System	CS Strateg for	y with LOBPCG top states	CS Strategy with CheFSI for top states (CS2CF strategy)		
	Energy per atom difference (Ha) Max. force component difference (Ha/Bohr)		Energy per atom difference (Ha)	Max. force component difference (Ha/Bohr)	
Electrolyte3D $_{1\times 2\times 2}$	5×10^{-5}	$2 imes 10^{-4}$	5×10^{-5}	2×10^{-4}	
SiDiamond $3D_{5 \times 5 \times 5}$	4×10^{-6}	4×10^{-6} 4×10^{-5}		1×10^{-5}	
$\text{Graphene2D}_{1\times 2\times 2}$	7×10^{-6}	8×10^{-5}	$9 imes 10^{-6}$	8×10^{-5}	
$CuFCC3D_{5 \times 5 \times 5}$	8×10^{-6}	$9 imes 10^{-5}$	4×10^{-6}	$7 imes 10^{-5}$	
$LiBCC3D_{4\times4\times4}$	$7 imes 10^{-6}$	$9 imes 10^{-5}$	5×10^{-6}	8×10^{-5}	

 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.)



(a) Electrolyte3D_{3×3×3} system (8,586 atoms, 29,808 electrons, left two plots) and SiDiamond3D_{10×10×10} system (8,000 atoms, 32,000 electrons, right two plots). 1,728 MPI processes used in all cases.





(b) Graphene2D_{8×8} system (11,520 atoms, 23,040 electrons). Right two figures use fewer extra and top states. 2,304 MPI processes used in all cases.



- 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

System	No. of atoms (No. of electrons)	Total computational cores (cores used in subspace prob.) [s]	ALB generation [s]	Hamiltonian update [s]	CS2CF strategy (subspace prob. time) [s]	Total SCF wall time via CS2CF strategy [s]
$\rm Electrolyte3D_{3\times3\times3}$	8,586 (29,808)	34,560 $(3,456)$	12	4	34 (19)	50
SiDiamond3D _{10×10×10}	8,000 (32,000)	34,560 $(3,456)$	9	2	40 (24)	51
$\rm Graphene2D_{8\times 8}$	11,520 (23,040)	27,648 $(4,608)$	4	2	35 (27)	41
$CuFCC3D_{10\times10\times10}$	4,000 (44,000)	$30,000\ (3,000)$	20	9	75 (46)	104
LiBCC3D _{12×12×12}	27,648 (82,944)	38,880 (12,960)	22	13	180 (165)	215

Results :SCF iteration wall times and Ab initial MD

• SCF iteration wall times with DGDFT-CS2CF and ELPA

System	No. of atoms (No. of electrons)	Total computational cores (cores used in subspace prob.) [s]	CS2CF strategy (subspace prob. time) [s]	Total SCF wall time via CS2CF strategy [s]	Direct diagonalization via ELPA [s]	Total SCF wall time via ELPA [s]
${\rm Electrolyte3D}_{3\times3\times3}$	8,586 (29,808)	$34,560\ (3,456)$	34 (19)	50	647	663
$SiDiamond3D_{10\times10\times10}$	8,000 (32,000)	34,560(3,456)	40 (24)	51	648	659
$\rm Graphene2D_{8\times 8}$	11,520 (23,040)	27,648 (4,608)	35 (27)	41	262	268
$CuFCC3D_{10\times10\times10}$	4,000 (44,000)	30,000 $(3,000)$	75 (46)	104	199	228
$LiBCC3D_{12 \times 12 \times 12}$	27,648 (82,944)	$38,880\ (12,960)$	180 (165)	215	5844	5879

Results :SCF iteration wall times and Ab initial MD

• SCF iteration wall times with DGDFT-CS2CF and ELPA

System	No. of atoms (No. of electrons)	Total computational cores (cores used in subspace prob.) [s]	CS2CF strategy ; (subspace prob. time [s]	Total SCF wall time I via CS2CF strategy [s]	irect diagonalizatio via ELPA [s]	Total SCF wall time via ELPA [s]
${\rm Electrolyte3D}_{3\times3\times3}$	8,586 (29,808)	34,560(3,456)	34 (19)	50	647	663
$SiDiamond3D_{10\times10\times10}$	8,000 (32,000)	34,560(3,456)	40 (24)	51	648	659
${\rm Graphene2D}_{8\times 8}$	11,520 (23,040)	27,648 (4,608)	35 (27)	41	262	268
$CuFCC3D_{10\times10\times10}$	4,000 (44,000)	30,000 $(3,000)$	75 (46)	104	199	228
LiBCC3D _{12×12×12}	27,648 (82,944)	38,880 (12,960)	180~(165)	215	5844	5879
				·		

Results :SCF iteration wall times and Ab initial MD

• SCF iteration wall times with DGDFT-CS2CF and ELPA

System	No. of atoms (No. of electrons)	Total computational cores (cores used in subspace prob.) [s]	CS2CF strategy ; (subspace prob. time [s]	Total SCF wall time via CS2CF strategy [s]	I irect diagonalizatio via ELPA [s]	Total SCF wall time via ELPA [s]
${\rm Electrolyte3D}_{3\times3\times3}$	8,586 (29,808)	34,560(3,456)	34 (19)	50	647	663
$SiDiamond3D_{10\times10\times10}$	8,000 (32,000)	34,560(3,456)	40 (24)	51	648	659
${\rm Graphene2D}_{8\times 8}$	11,520 (23,040)	27,648 (4,608)	35 (27)	41	262	268
$CuFCC3D_{10\times10\times10}$	4,000 (44,000)	30,000 $(3,000)$	75~(46)	104	199	228
LiBCC3D _{12×12×12}	27,648 (82,944)	38,880 (12,960)	180~(165)	215	5844	5879

• 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 ?

Image courtesy of internet resources

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 :

SCF Iterations



$$\begin{split} &-\frac{1}{2}\nabla^2\psi_i(\mathbf{x}) + \hat{V}_{eff}\psi_i(\mathbf{x}) = \varepsilon_i\psi_i(\mathbf{x}), \\ &\hat{V}_{eff} = V_I^\ell + \hat{V}_I^{n\ell} + V_H + V_{xc}, \\ &V_I^\ell = \sum_a V_{I,a}(\mathbf{x}), \\ &\hat{V}_I^{n\ell}\psi_i = \sum_a \int d\mathbf{x}' V_{I,a}^{n\ell}(\mathbf{x}, \mathbf{x}')\psi_i(\mathbf{x}'), \\ &V_H = -\int d\mathbf{x}' \frac{\rho_e(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}, \\ &V_{xc} = V_{xc}(\mathbf{x}; \rho_e), \\ &\rho_e = -\sum_i f_i\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}), \end{split}$$

Part III : Periodic Pulay Mixing

The problem at hand :

SCF Iterations



$egin{aligned} &-rac{1}{2} abla^2\psi_i(\mathbf{x})+\hat{V}_{eff}\psi_i(\mathbf{x})=arepsilon_i\psi_i(\mathbf{x}),\ \hat{V}_{eff}=V_I^\ell+\hat{V}_I^{n\ell}+V_H+V_{xc},\ V_I^\ell&=\sum_a V_{I,a}(\mathbf{x}),\ \hat{V}_I^{n\ell}\psi_i=\sum_a \int d\mathbf{x}' V_{I,a}^{n\ell}(\mathbf{x},\mathbf{x}')\psi_i(\mathbf{x}'),\ V_H&=-\int d\mathbf{x}'rac{ ho_e(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|},\ V_{xc}=V_{xc}(\mathbf{x}; ho_e),\ ho_e&=-\sum_i f_i\psi_i^*(\mathbf{x})\psi_i(\mathbf{x}), \end{aligned}$

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

Amartya Banerjee : Pushing the Envelope of Large Scale First Principles Simulations of Non-Insulating Systems MoISSI 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 : $\rho = g(\rho)$
 - $\boldsymbol{\rho} \in \mathbb{R}^{N \times 1}$

$$\mathbf{g}: \mathbb{R}^{N \times 1} \to \mathbb{R}^{N \times 1}$$

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

- $f(\rho) = g(\rho) \rho$
 - 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

$$\boldsymbol{\rho}_{i+1} = \boldsymbol{\rho}_i + \mathbf{C}_i \mathbf{f}_i$$

- $\mathbf{f}_i = \mathbf{f}(\boldsymbol{\rho}_i) \quad \mathbf{C}_i = \alpha \mathbf{I}$
- 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

$$\boldsymbol{\rho}_{i+1} = \boldsymbol{\rho}_i + \mathbf{C}_i \mathbf{f}_i$$
$$\mathbf{f}_i = \mathbf{f}(\boldsymbol{\rho}_i) \qquad \mathbf{C}_i = \alpha \mathbf{I} - (\mathbf{R}_i + \alpha \mathbf{F}_i)(\mathbf{F}_i^T \mathbf{F}_i)^{-1} \mathbf{F}_i^T$$
$$\mathbf{R}_i = [\Delta \boldsymbol{\rho}_{i-n+1}, \Delta \boldsymbol{\rho}_{i-n+2}, \dots, \Delta \boldsymbol{\rho}_i] \in \mathbb{R}^{N \times n},$$
$$\mathbf{F}_i = [\Delta \mathbf{f}_{i-n+1}, \Delta \mathbf{f}_{i-n+2}, \dots, \Delta \mathbf{f}_i] \in \mathbb{R}^{N \times n},$$
$$\Delta \boldsymbol{\rho}_i = \boldsymbol{\rho}_i - \boldsymbol{\rho}_{i-1} \qquad \Delta \mathbf{f}_i = \mathbf{f}_i - \mathbf{f}_{i-1},$$

- 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.
- Has lead to highly efficient and scalable linear solver -Alternating-Anderson Accelerated Jacobi (AAJ) method (Pratapa, Suryanarayana, Pask : J. Comp. Phys., vol. 306, 2016).
- Application of AAJ to SCF iterations : Periodic Pulay Mixing

Periodic Pulay Mixing (cont.)

• Extremely easy to integrate into existing electronic structure codes.

Algorithm 1. Periodic Pulay method.

Input: ρ_0 , α , n, k, and tolrepeat i = 0, 1, 2... $\begin{vmatrix} \mathbf{f}_i = \mathbf{g}(\rho_i) - \rho_i \\ \mathbf{if} \ (i+1)/k \in \mathbb{N} \text{ then} \\ \mid \rho_{i+1} = \rho_i + \alpha \mathbf{f}_i - (\mathbf{R}_i + \alpha \mathbf{F}_i)(\mathbf{F}_i^T \mathbf{F}_i)^{-1} \mathbf{F}_i^T \mathbf{f}_i \\ \mathbf{else} \\ \mid \rho_{i+1} = \rho_i + \alpha \mathbf{f}_i \\ \mathbf{until} \| \mathbf{f}_i \| < tol;$ Output: ρ_i

Implemented in DG-DFT, ClusterES and SIESTA

Results

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

System	Description
Benzene	Benzene molecule (12 atoms)
32-H ₂ O	32 molecules of water (96 atoms)
Pd-bulk	3 × 3 × 3 FCC bulk Palladium (108 atoms) Γ-point calculation
Pt ₁₃	Octahedral platinum cluster (13 atoms)
Fe ₃ -noncollinear	Iron cluster with non-collinear spin (3 atoms)
Fe-bulk	Bulk BCC iron (1 atom). Brillouin zone integration with 63 k-points
SiC-slab	Silicon carbide slab with hydrogen saturated surface (78 atoms)
ptcda-Au	Molecules of 3, 4, 9, 10 perylenetetracarboxylic dianhydride adsorbed onto gold (111) surface (204 atoms)

• Convergence statistics : fixed mixing parameters, varying histories, periods

System	Pulay	Pulay				Periodic Pulay			
	μ	σ	Max	Min	μ	σ	Max	Min	
Benzene	12	3	14	9	11	2	13	9	
32-H ₂ O	19	3	23	16	17	2	21	16	
Pd-bulk	60	40	125	24	35	20	85	22	
Pt ₁₃	98	33	157	74	76	16	113	62	
Fe ₃ -noncollinear	112	38	173	70	77	24	100	65	
Fe-bulk	72	31	122	43	29	5	36	23	
SiC-slab	51	10	64	41	31	3	37	26	
ptcda-Au	89	11	103	72	75	8	97	65	

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

Results (cont.)

Convergence behavior



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

System	Pulay					Periodic Pulay			
	μ	σ	Max	Min	μ	σ	Max	Min	
Pd-bulk SiC-slab	56 26	18 4	77 35	27 22	36 22	10 2	52 26	19 19	

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

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.

Temperature (K)	Pulay	Periodic Pulay	
		k=2	k=3
100	78	25	36
50	164	75	37
25	а	39	64

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.
- Periodic Pulay method for robust and efficient convergence acceleration of selfconsistent field iterations: AS Banerjee, P Suryanarayana, JE Pask; Chemical Physics Letters 647, 31-35

Acknowledgments

- NERSC computing facilities
- Funding Agencies : US Department of Energy BES, ASCR, SciDAC programs.

Thank You !

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