Adaptive local basis for Kohn-Sham density functional theory in Discontinuous Galerkin framework

Lin Lin

Program in Applied and Computational Mathematics, Princeton University

Joint work with *Weinan E, Jianfeng Lu and Lexing Ying*.

Density Functional Theory: Fundamentals and Applications in Condensed Matter Physics, Banff, 2011

Outline

Kohn-Sham Density Functional Theory

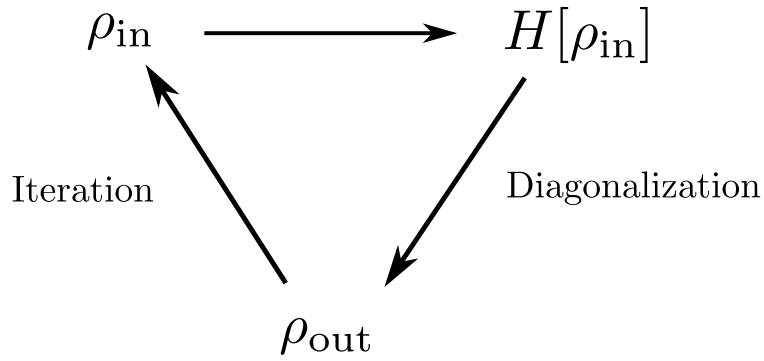
$$H[\rho]\psi_i(x) = \left(-\frac{1}{2}\Delta + V_{\text{ion}} + \int dx' \, \frac{\rho(x')}{|x - x'|} + V_{\text{xc}}[\rho]\right)\psi_i(x) = \epsilon_i \psi_i(x),$$

$$\rho(x) = 2\sum_{i=1}^{N/2} |\psi_i(x)|^2, \quad \int \psi_i \psi_j = \delta_{ij}.$$

- Single particle formalism with in principle exact accuracy.
- N is the number of electrons (2 comes from spin). V_{ion} is the ionic potential. V_{xc} is the exchange-correlation potential.
- Nonlinear eigenvalue problem.

Standard flowchart

Self-consistent field iteration:



- Diagonalization: Lowest $N/2$ eigenvalues ε_i and eigenfunctions ψ_i, $\rho_{\text{out}}(x) = 2\sum_{i=1}^{N/2}|\psi_i(x)|^2$.
- Most time consuming step.
- Focus of this talk: $H[\rho_{\text{in}}] \to \rho_{\text{out}}$.

Computational cost of KSDFT

Typical cubic computational scaling:

Size	Time
10	1 sec
100	17 min
1000	11.6 day

$\mathcal{O}(nN^3)$ cost:

- diagonalization cost: N^3.
- discretization cost: the number of basis functions per atom (n), n is typically $500 \sim 5000$.

Importance of discretization cost

- Diagonalization: $\mathcal{O}(nN^3)$ for all systems.
- Selected inversion: $\mathcal{O}(N)$ for quasi-1D system, $\mathcal{O}(N^{1.5})$ for quasi-2D system, and $\mathcal{O}(N^2)$ for 3D bulk system.

$$\rho = \text{diag}\ \frac{2}{1 + e^{\beta(H-\mu)}} \approx \text{diag} \sum_{i=1}^{P} \frac{\omega_i}{z_i I - H}$$

Importance of discretization cost

- Diagonalization: $\mathcal{O}(nN^3)$ for all systems.
- Selected inversion: $\mathcal{O}(n^3N)$ for quasi-1D system, $\mathcal{O}(n^3N^{1.5})$ for quasi-2D system, and $\mathcal{O}(n^3N^2)$ for 3D bulk system.

$$\rho = \text{diag}\ \frac{2}{1+e^{\beta(H-\mu)}} \approx \text{diag}\ \sum_{i=1}^{P} \frac{\omega_i}{z_i I - H}$$

Importance of discretization cost

- Diagonalization: $\mathcal{O}(nN^3)$ for all systems.
- Selected inversion: $\mathcal{O}(n^3N)$ for quasi-1D system, $\mathcal{O}(n^3N^{1.5})$ for quasi-2D system, and $\mathcal{O}(n^3N^2)$ for 3D bulk system.

$$\rho = \text{diag}\,\frac{2}{1+e^{\beta(H-\mu)}} \approx \text{diag}\sum_{i=1}^{P}\frac{\omega_i}{z_i I - H}$$

n	1D Crossover $N \sim n$	2D Crossover $N \sim n^{1.33}$	3D Crossover $N \sim n^2$
20	20	54	400
50	50	184	2,500
500	500	3,887	250,000
1000	1000	9,772	1,000,000

Importance of discretization cost

- Diagonalization: $\mathcal{O}(nN^3)$ for all systems.
- Selected inversion: $\mathcal{O}(n^3N)$ for quasi-1D system, $\mathcal{O}(n^3N^{1.5})$ for quasi-2D system, and $\mathcal{O}(n^3N^2)$ for 3D bulk system.

$$\rho = \mathrm{diag}\,\frac{2}{1+e^{\beta(H-\mu)}} \approx \mathrm{diag}\sum_{i=1}^{P}\frac{\omega_i}{z_i I - H}$$

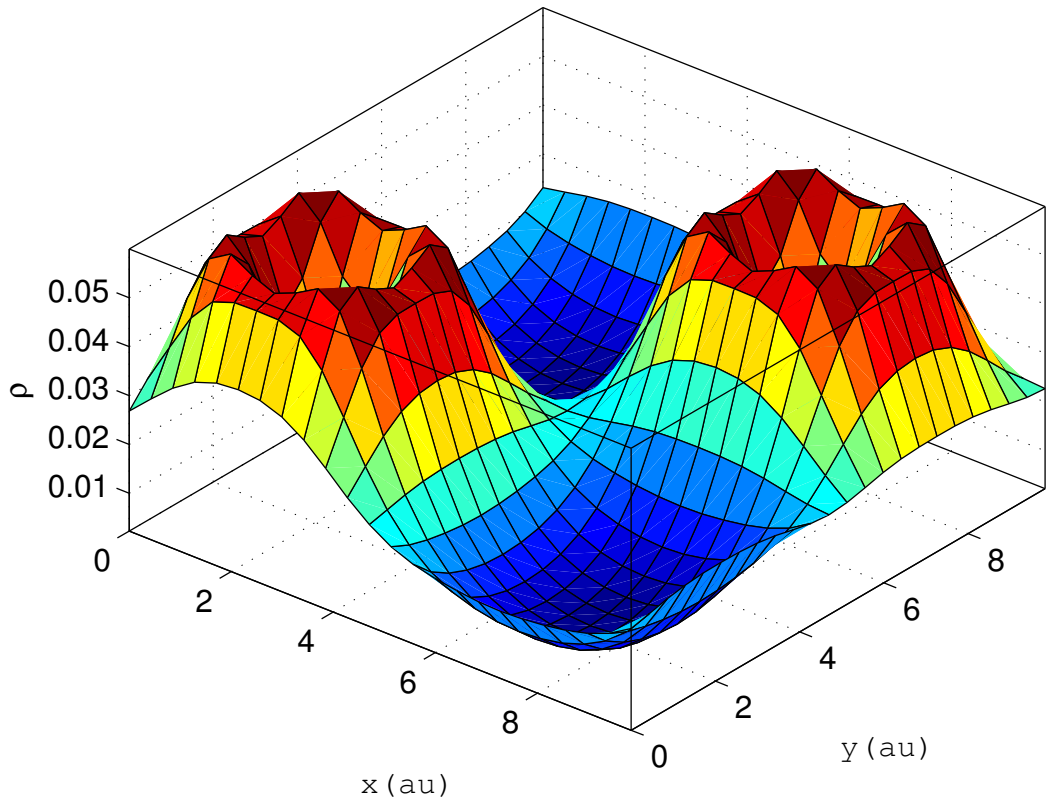
n	1D Crossover $N \sim n$	2D Crossover $N \sim n^{1.33}$	3D Crossover $N \sim n^2$
20	20	54	400
50	50	184	2,500
500	500	3,887	250,000
1000	1000	9,772	1,000,000

- Goal today: Reduce the number of basis functions per atom.

Outline

Building atomic physics into basis functions I

- Rapid but “inert” oscillation near the nucleus.
- Smooth oscillation in the interstitial bonding region.



Building atomic physics into basis functions II

Atomic orbital as basis functions:

- Gaussian type orbitals (GTO)
- Numerical atomic orbital (NAO) [Blum et al. 2009]

Mixed (enriched) basis functions:

- Augmented plane-wave (APW) [Slater, 1937]
- Linear augmented-plane-wave (LAPW) [Andersen, 1975]
- Projector augmented-wave (PAW) [Blöchl, 1994]
- Enriched finite element [Sukumar and Pask, 2009]

Difficulty of atomic orbitals

Lack of environmental effect

Difficulty of atomic orbitals

Lack of environmental effect

Atomic orbital as basis functions:

- Fine tunning of the parameters for different chemical elements, exchange-correlation functional and even different environment.
- Overcomplete and incomplete basis sets.

Difficulty of atomic orbitals

Lack of environmental effect

Atomic orbital as basis functions:

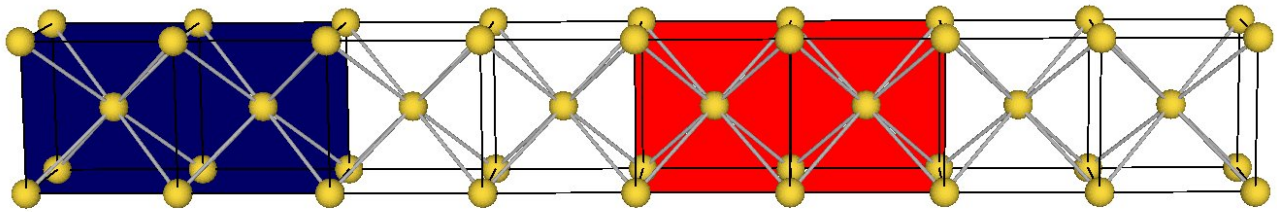
- Fine tunning of the parameters for different chemical elements, exchange-correlation functional and even different environment.
- Overcomplete and incomplete basis sets.

Mixed (enriched) basis functions:

- Fine tunning.
- Relatively large number of basis functions per atom.

Building environmental effect into basis functions

Construct the local basis functions on the fly by solving a small part of the system.



Questions:

- How to obtain basis function?
- How to patch basis functions together?

Our answer:

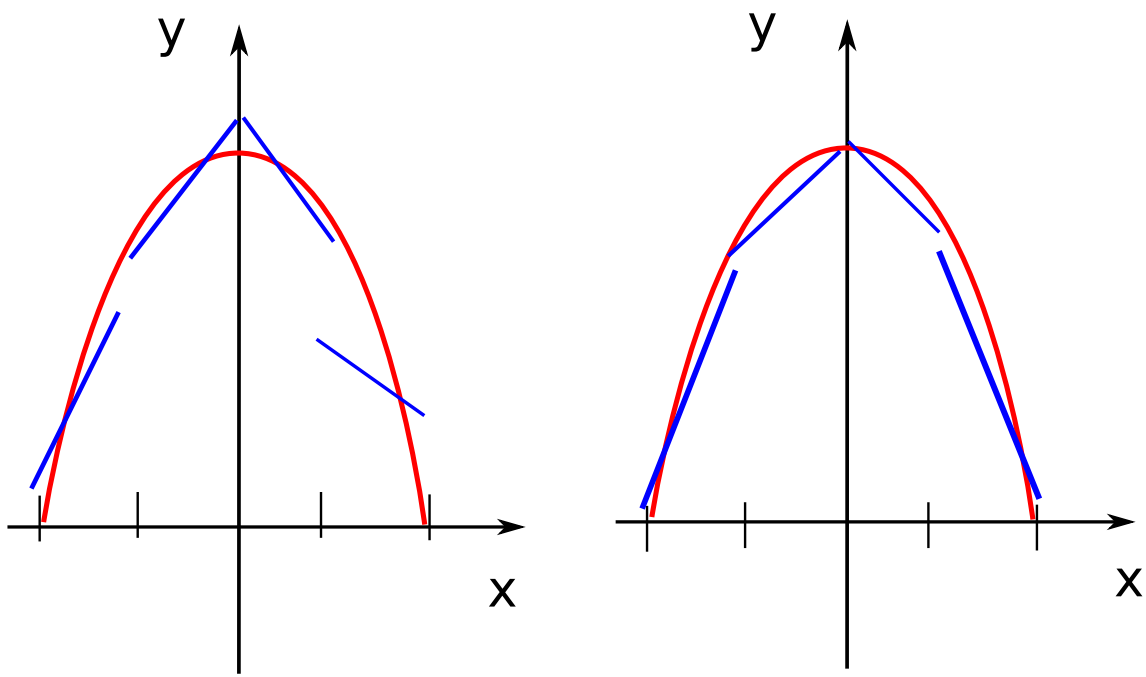
- Local solve to obtain discontinuous basis function.
- Discontinuous Galerkin framework to patch basis functions together.

Outline

Introduction

Discontinuous Galerkin:

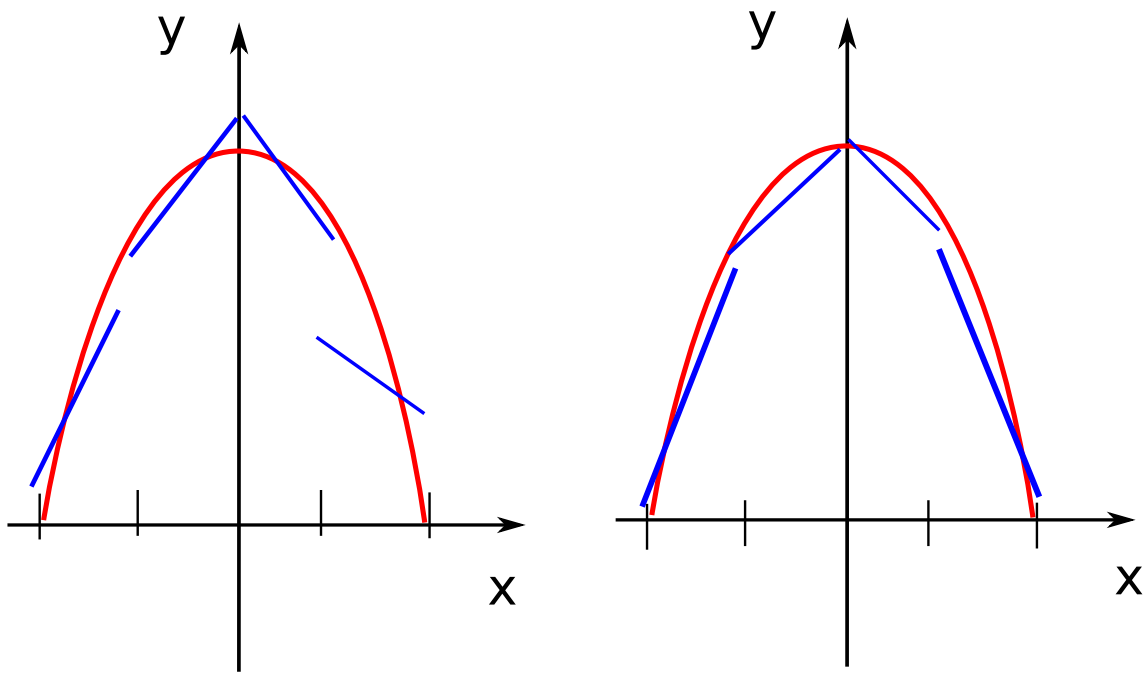
- Finite element method with discontinuous basis functions.
- Penalty on the inter-element jump to enforce inter-element continuity.



Introduction

Discontinuous Galerkin:

- Finite element method with discontinuous basis functions.
- Penalty on the inter-element jump to enforce inter-element continuity.



APW, LAPW etc. already discontinuous.

History more than four decades

Early work on using penalty to impose boundary condition:

- [Lions, 1968], [Aubin, 1970],[Babuška, 1973], [Nitsche, 1971]

Interior penalty for elliptic and parabolic problems:

- [Babuška and Zlámal, 1973] Penalty to impose C^1 continuity for fourth-order problem.
- [Douglas and Dupont, 1976]: elliptic and parabolic.
- [Arnold, 1982] interior penalty formulation.
- [Arnold, Brezzi, Cockburn et al 2002] unified analysis.

Hyperbolic equation:

- [Cockburn and Shu, 1991], [Cockburn and Shu, 1998] Runge-Kutta DG.
- [Cockburn, Karniakdakis and Shu, 2000]: Review

Setup

Let $\mathcal{T}$ be a collection of uniform rectangular partitions of Ω

$$\mathcal{T} = \{E_1, E_2, \cdots, E_M\},$$

and $\mathcal{S}$ be the collection of surfaces corresponds to $\mathcal{T}$.
Inner products

$$\begin{aligned}
\langle v, w \rangle_E &= \int_E v^*(x) w(x) \, \mathrm{d}x, \\
\langle \boldsymbol{v}, \boldsymbol{w} \rangle_S &= \int_S \boldsymbol{v}^*(x) \cdot \boldsymbol{w}(x) \, \mathrm{d}s(x), \\
\langle v, w \rangle_{\mathcal{T}} &= \sum_{i=1}^{M} \langle v, w \rangle_{E_i}, \\
\langle \boldsymbol{v}, \boldsymbol{w} \rangle_{\mathcal{S}} &= \sum_{S \in \mathcal{S}} \langle \boldsymbol{v}, \boldsymbol{w} \rangle_S .
\end{aligned}$$

Kohn-Sham energy functional

The variational problem

$$E_{\text{eff}}(\{\psi_i\}) = \frac{1}{2}\sum_{i=1}^{N}\int |\nabla\psi_i(x)|^2\,\mathrm{d}x + \int V_{\text{eff}}(x)\rho(x)\,\mathrm{d}x + \sum_{\ell}\gamma_\ell\sum_{i=1}^{N}|\langle b_\ell,\psi_i\rangle$$

with orthonormality constraints $\langle\psi_i,\psi_j\rangle = \delta_{ij}$. The effective one-body potential V_{eff}

$$V_{\text{eff}}[\rho](x) = V_{\text{ext}}(x) + \int \frac{\rho(y)}{|x-y|}\,\mathrm{d}y + \epsilon'_{\text{xc}}[\rho(x)].$$

DG energy functional: interior penalty formulation

$$E_{\mathrm{DG}}(\{\psi_i\}) = \frac{1}{2}\sum_{i=1}^{N}\langle\nabla\psi_i,\nabla\psi_i\rangle_{\mathcal{T}}+\langle V_{\mathrm{eff}},\rho\rangle_{\mathcal{T}}+\sum_{\ell}\gamma_\ell\sum_{i=1}^{N}|\langle b_\ell,\psi_i\rangle_{\mathcal{T}}|^2$$
$$-\sum_{i=1}^{N}\langle\{\{\nabla\psi_i\}\},[\![\psi_i]\!]\rangle_{\mathcal{S}}+\frac{\alpha}{h}\sum_{i=1}^{N}\langle[\![\psi_i]\!],[\![\psi_i]\!]\rangle_{\mathcal{S}}.$$

$\{\{\cdot\}\}$ and $[\![\cdot]\!]$ are average and jump operators across surfaces.

$$\{\{q\}\} = \tfrac{1}{2}(q_1+q_2) \quad \text{on } S.$$
$$[\![u]\!] = u_1 n_1 + u_2 n_2 \quad \text{on } S.$$

- S: shared by elements K_1 and K_2.
- n_1 and n_2 as unit normal vectors on S pointing exterior to K_1 and K_2.

Galerkin I

For each element E_k choose a set of basis functions $\{\varphi_{k,j}\}_{j=1}^{J_k}$, and extend $\varphi_{k,j}$ to the whole computational domain Ω by setting it to be 0 on the complement of E_k.

$$\mathcal{V} = \text{span}\{\varphi_{k,j},\ E_k \in \mathcal{T},\ j = 1, \cdots, J_k\}.$$

Minimize for $\{\psi_i\} \subset \mathcal{V}$ in the approximation space $\mathcal{V}$ leads to the eigenvalue problem for any $v \in \mathcal{V}$

$$\frac{1}{2}\langle \nabla v, \nabla \psi_i \rangle_{\mathcal{T}} - \frac{1}{2}\left\langle [\![v]\!], \{\{\nabla \psi_i\}\} \right\rangle_{\mathcal{S}} - \frac{1}{2}\left\langle \{\{\nabla v\}\}, [\![\psi_i]\!] \right\rangle_{\mathcal{S}} + \frac{\alpha}{h}\left\langle [\![v]\!], [\![\psi_i]\!] \right\rangle_{\mathcal{S}} + \langle v, V_{\text{eff}}\psi_i \rangle_{\mathcal{T}} + \sum_{\ell} \gamma_\ell \langle v, b_\ell \rangle_{\mathcal{T}} \langle b_\ell, \psi_i \rangle_{\mathcal{T}} = \lambda_i \langle v, \psi_i \rangle_{\mathcal{T}}.$$

Galerkin II

Setting $v = \varphi_{k',j'}$ and denoting

$$\psi_i = \sum_{k \in \mathcal{T}} \sum_{j=1}^{J_k} c_{i;k,j} \varphi_{k,j},$$

gives the following linear system

$$\sum_{k,j} \Bigg(\frac{1}{2} \left\langle \nabla \varphi_{k',j'}, \nabla \varphi_{k,j} \right\rangle_{\mathcal{T}} - \frac{1}{2} \left\langle \left[\!\left[\varphi_{k',j'} \right]\!\right], \left\{\!\left\{ \nabla \varphi_{k,j} \right\}\!\right\} \right\rangle_{\mathcal{S}}$$
$$- \frac{1}{2} \left\langle \left\{\!\left\{ \nabla \varphi_{k',j'} \right\}\!\right\}, \left[\!\left[\varphi_{k,j} \right]\!\right] \right\rangle_{\mathcal{S}} + \frac{\alpha}{h} \left\langle \left[\!\left[\varphi_{k',j'} \right]\!\right], \left[\!\left[\varphi_{k,j} \right]\!\right] \right\rangle_{\mathcal{S}} + \left\langle \varphi_{k',j'}, V_{\text{eff}} \varphi_{k,j} \right\rangle_{\mathcal{T}}$$
$$+ \sum_{\ell} \gamma_\ell \left\langle \varphi_{k',j'}, b_\ell \right\rangle_{\mathcal{T}} \left\langle b_\ell, \varphi_{k,j} \right\rangle_{\mathcal{T}} \Bigg) c_{i;k,j} = \lambda_i \sum_{k,j} \left\langle \varphi_{k',j'}, \varphi_{k,j} \right\rangle c_{i;k,j}.$$

Eigenvalue problem

$$Ac_i = \lambda_i B c_i.$$

We can actually choose $B = I$. Then we have standard eigenvalue problem

$$Ac_i = \lambda_i c_i.$$

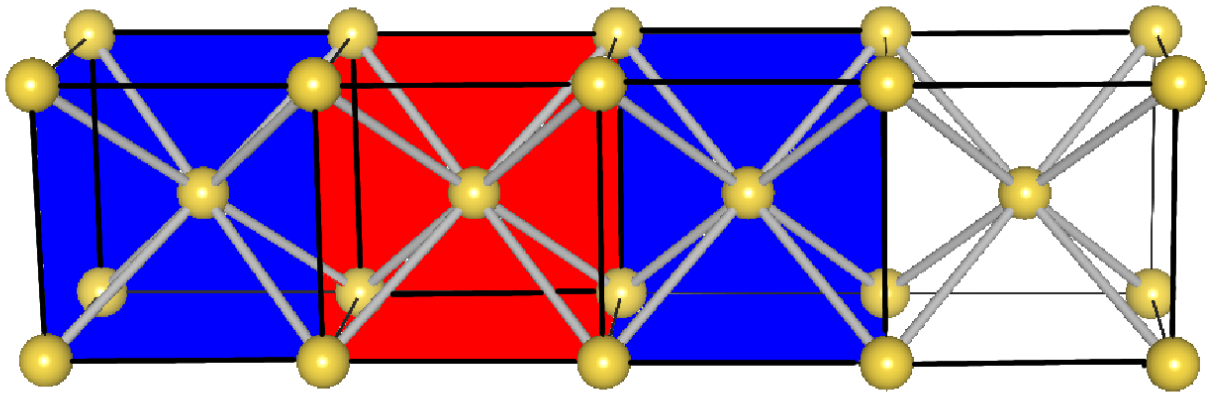
The electron density is given by

$$\widetilde{\rho} = \sum_{i=1}^{N} \sum_{k} (\sum_{j=1}^{J_k} c_{i;k,j} \varphi_{k,j})^2.$$

Outline

Setup of local problem

- Buffer region associated with E_k: $Q_k \supset E_k$.
- Restrict the effective Hamiltonian on Q_k by assuming the periodic boundary condition on ∂Q_k and obtain H_{eff,Q_k}.
- Take the first several eigenfunctions of H_{eff,Q_k} called $\{\varphi_{k,j}\}$, $j = 1, \cdots, J_k$ and restrict them on E_k.



Red: E_k; Red+Blue: Q_k

Standard eigenvalue problem

Recall that

$$Ac_i = \lambda_i B c_i, \quad B = \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_M \end{pmatrix}, \quad \{B_k\}_{jj'} = \left\langle \varphi_{k',j'}, \varphi_{k,j} \right\rangle.$$

SVD filtering to obtain an orthonormal basis set and to avoid numerical degeneracy:

1. For each k, form matrix $M_k = (\varphi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,J_k})$;
2. Calculate SVD decomposition $UDV^* = M_k$;
3. For a threshold δ, find $\widetilde{J}_k$ such that $|\lambda_{k,\widetilde{J}_k}| > \delta$ and $|\lambda_{k,\widetilde{J}_k+1}| < \delta$. Take $\widetilde{\varphi}_{k,j}$ be the j-th column of U.
4. Set $J_k \leftarrow \widetilde{J}_k$ and $\varphi_{k,j} \leftarrow \widetilde{\varphi}_{k,j}$ for $j = 1, \cdots, \widetilde{J}_k$.

Overall algorithm

1. Set $n = 0$, take elements partition $\mathcal{T}$ and an initial guess of density ρ_0;
2. Form the effective potential $V_{\text{eff}}[\rho_n]$ and the effective Hamiltonian $H_{\text{eff}}[\rho_n]$;
3. For each element $E_k \in \mathcal{T}$, solve the restricted Hamiltonian on buffer region Q_k, and obtain $\{\varphi_{k,j}\}$, $j = 1, \cdots, J_k$ using SVD filtering.
4. Solve standard eigenvalue problem $Ac_i = \lambda_i c_i$. to obtain $c_{i;k,j}$ and get the density $\widetilde{\rho}$.
5. Mixing step: Determine ρ_{n+1} from ρ_n and $\widetilde{\rho}$;
6. If $\|\rho_n - \widetilde{\rho}\| \leq \delta$, stop; otherwise, go to step (2) with $n \leftarrow n + 1$.

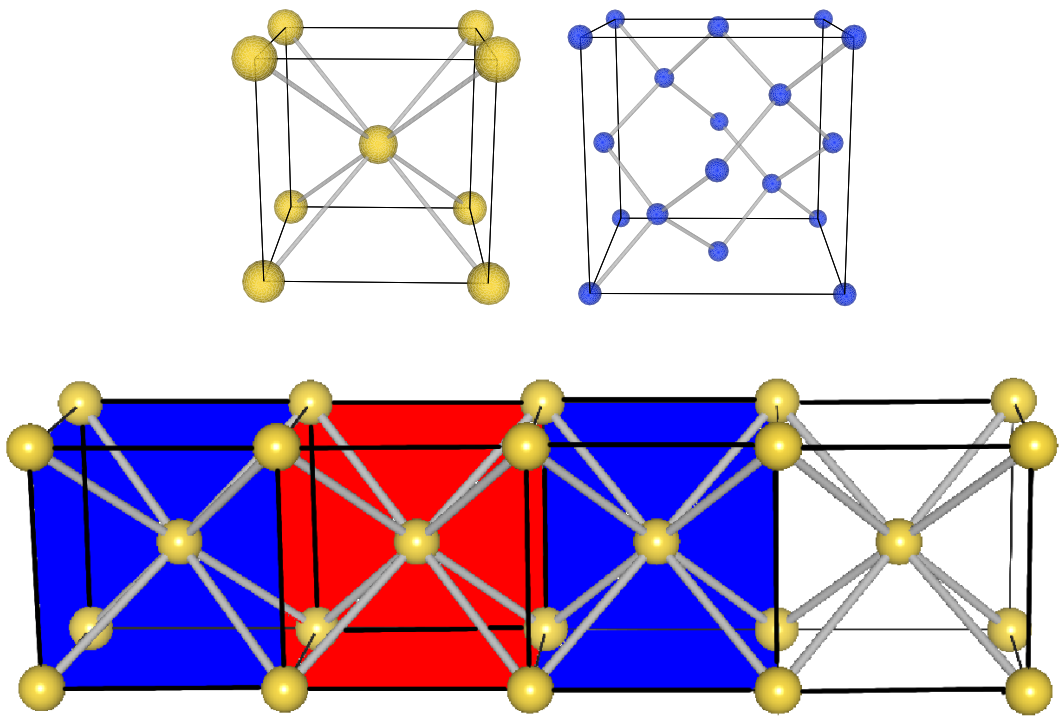
Advantage of the current framework

- Automatic basis set reduction.
- Capture both atomic and environmental effect automatically.
- Complete basis set.
- Flexible framework due to discontinuous character: can be combined with polynomials or other existing basis functions.

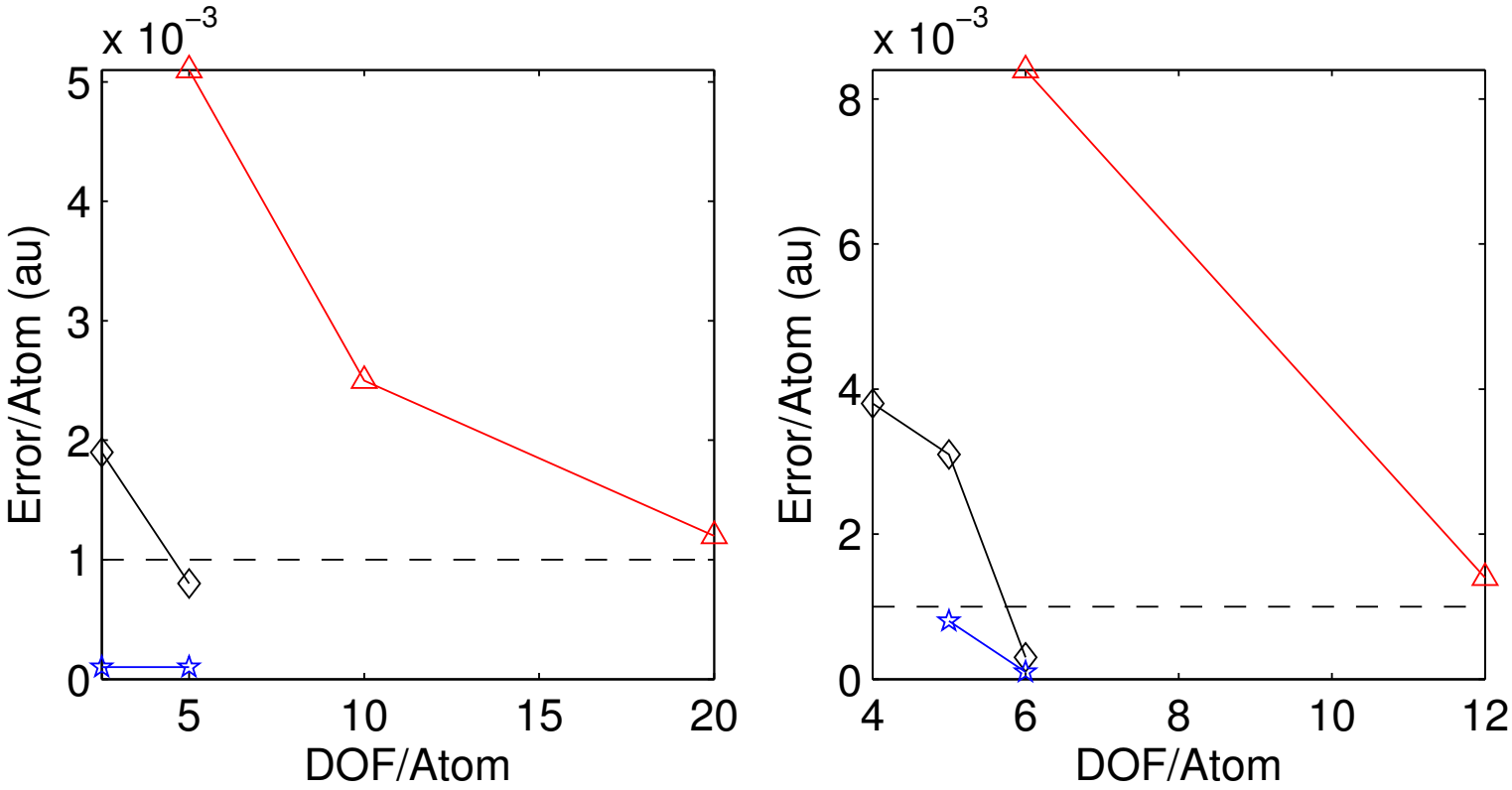
Outline

Setup

- Target accuracy (chemical accuracy): total energy error 10^{-3} au/atom.
- LDA. Real space local and non-local pseudopotential [Shaw, 1968].
- Buffer extra size.
- Sodium (Na) and Silicon (Si).

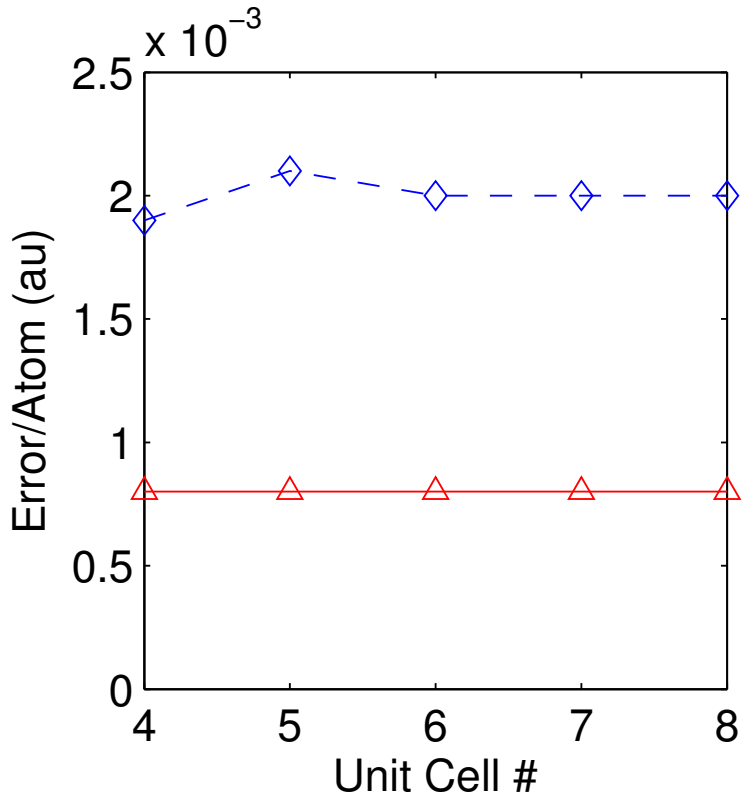


Quasi-1D: periodic case



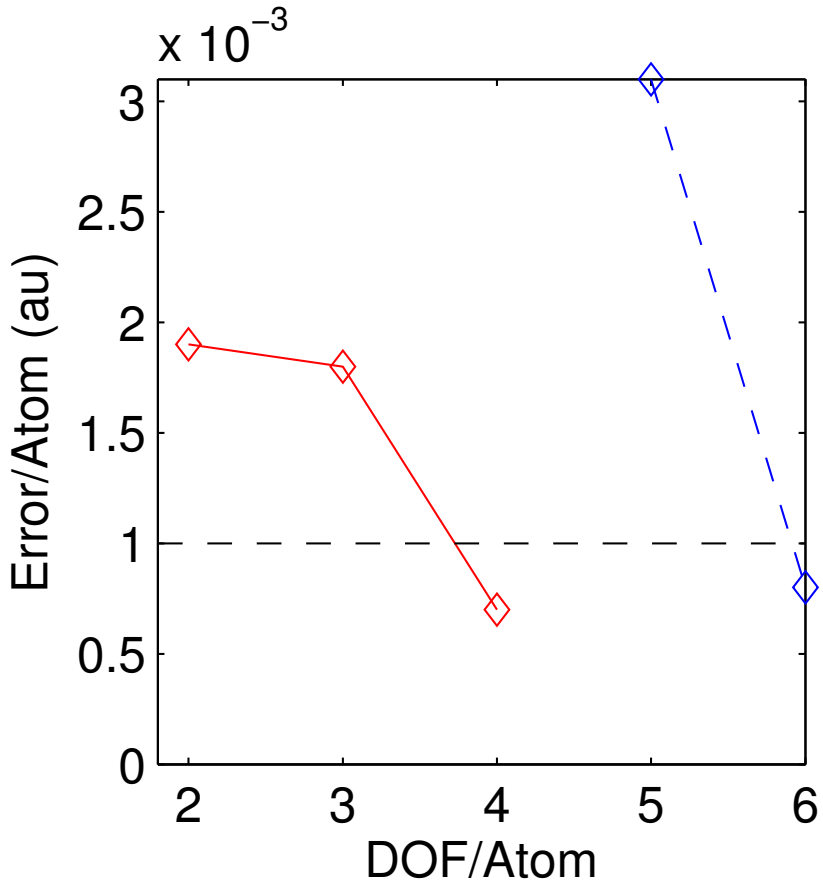
Left: Na with 4 unit cells. Buffer extra size 0.25 (red triangle with solid line), 0.50 (black diamond with solid line), and 0.75 (blue star with solid line). Right: Si with 4 unit cells. The legend is the same as in (a). The black dashed horizontal line refers to the target accuracy 10^{-3} au per atom.

Quasi-1D: Long supercell



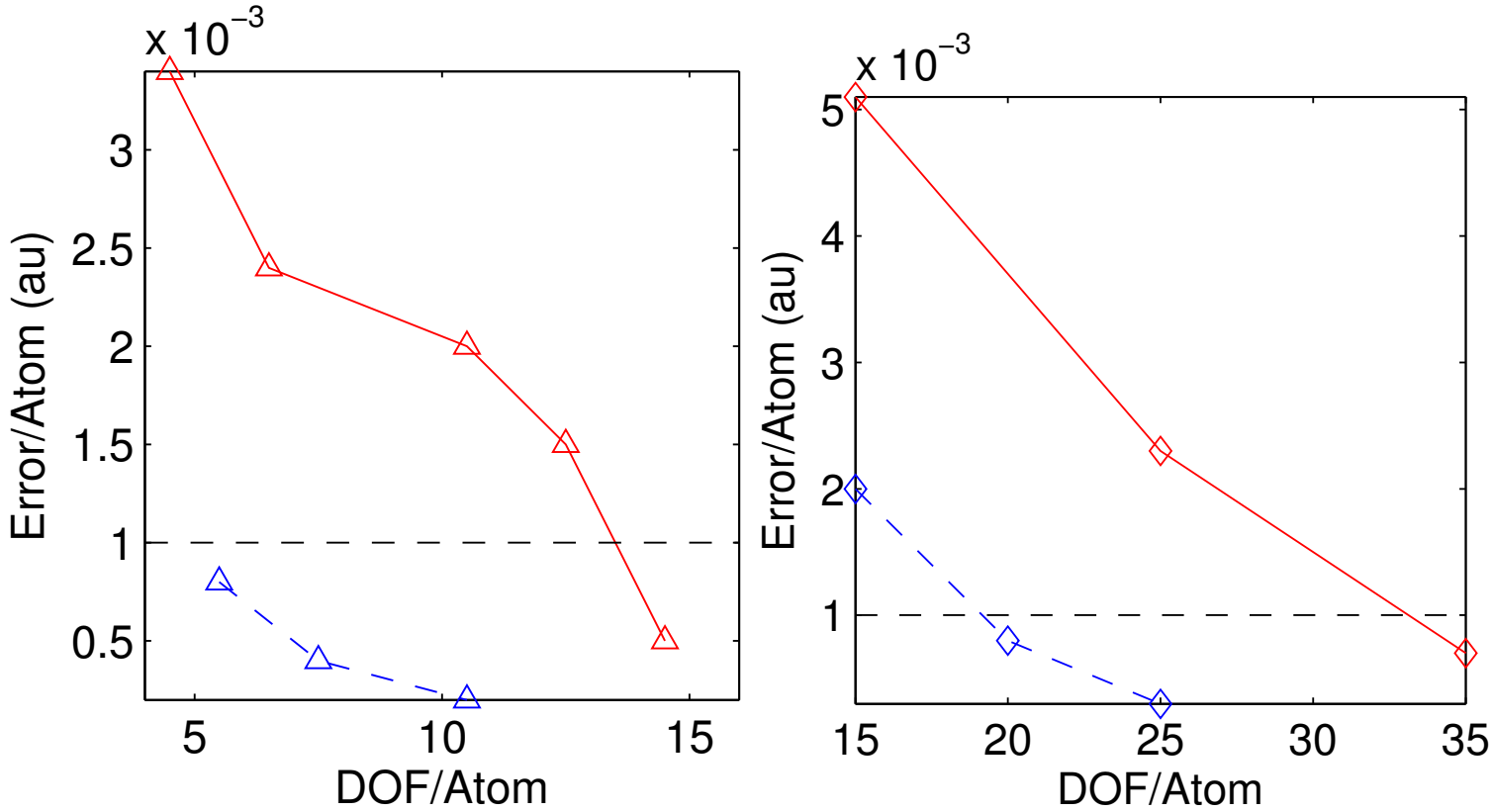
Na system with increasing supercell dimension along z-direction. The buffer extra size is 0.50. 3 enrichment basis function per atom and 5 enrichment basis function per atom are represented by blue diamond with dashed line, and red triangle with solid line, respectively.

Quasi-1D: random perturbation



Disordered Na (red diamond with solid line) and Si (blue diamond with dashed line), with buffer extra size 0.50.

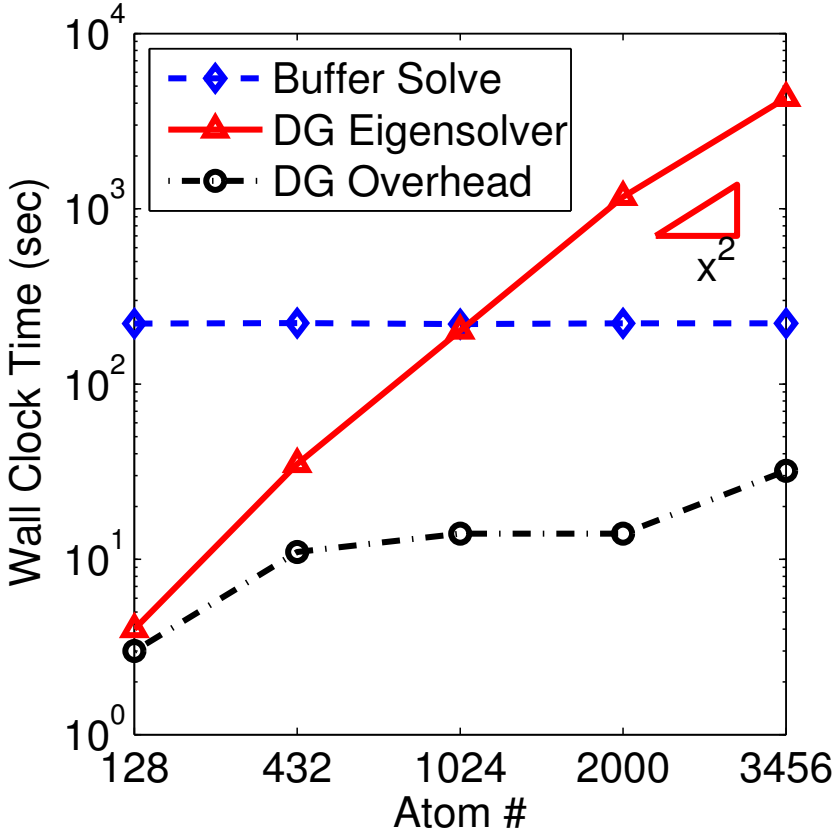
Quasi-2D and 3D Bulk system



Left: Quasi-2D Na system with buffer extra sizes 0.50 (red triangle with solid line), and buffer extra size 1.00 (blue triangle with dashed line). Right: Bulk 3D Na system with buffer extra sizes 0.50 (red diamond with solid line), and buffer extra size 1.00 (blue diamond with dashed line).

Computational efficiency

- Treat the matrix A as dense matrix.
- Parallelized implementation of buffer solve.
- Parallelized implementation of DG matrix assembly.
- Use ScaLAPACK subroutine pdsyevd as parallel eigenvalue solver.



Computational time per processor comparison:

Atom#	Proc#	Global time	DG time
128	64	35 s	4 s
432	216	248 s	35 s

Outline

Future work

Methodology:

- Improvement over periodic boundary condition in the buffer.
- All electron calculation.
- Molecular dynamics.

Code development:

- Discontinuous Galerkin framework with existing atomic orbital type basis functions.
- Combine with Fermi Operator Expansion methods: $\mathcal{O}(n^3 N)$ for quasi-1D system, $\mathcal{O}(n^3 N^{1.5})$ for quasi-2D system, and $\mathcal{O}(n^3 N^2)$ for 3D bulk system.
- Pseudopotentials. GGA.

Application:

- Long molecules, nanotube and nanowire (quasi-1D)
- Large scale surface problem. (quasi-2D)

Conclusion

Adaptive:

- Tunning-free basis set reduction with small number of enrichment functions.
- Capture both atomic and environmental effect automatically.

Local:

- Discontinuous nature.
- Important for the block sparsity of the assembled DG matrix.

Flexible:

- Flexible framework due to discontinuous character.

Fast:

- Computationally more efficient due to reduced basis set.

Thanks for your attention!