Grassmann manifold, gauge, and quantum chemistry

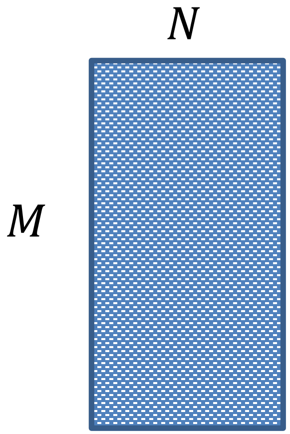
Lin Lin

Department of Mathematics, UC Berkeley;
Lawrence Berkeley National Laboratory

Flatiron Institute, April 2019

Stiefel manifold

- $\mathbb{C}^{M\times N}$ $(M > N)$



- Stiefel manifold (1935): N orthogonal vectors in $\mathbb{C}^M$

$$St_{M,N} = \{X \in \mathbb{C}^{M\times N} | X^*X = I_N\}$$

Grassmann manifold

- Grassmann manifold (1848): set of N dimensional subspace in $\mathbb{C}^M$

$$Gr_{M,N} = St_{M,N}/\mathcal{U}_N$$

$\mathcal{U}_N$: N dimensional unitary group (i.e. the set of unitary matrices in $\mathbb{C}^{N\times N}$)

- Any point in $St_{M,N}$ can be viewed as a representation of a point in $Gr_{M,N}$

Example (in optimization)

$$\min_X F(X)$$
$$s.t. X^*X = I_N$$

- $F(X) = F(XU), U \in \mathcal{U}_N$. invariant to the choice of basis ⇒ Optimization on a Grassmann manifold.

- The representation U is called the gauge in quantum physics / chemistry. Gauge invariant.

Example (in optimization)

$$\min_{X} F(X)$$
$$s.t. X^{*}X = I_N$$

- Otherwise, if $F(X)$ is gauge dependent ⇒ Optimization on a Stiefel manifold.

- [Edelman, Arias, Smith, SIMAX 1998] many works following

This talk: Quantum chemistry

- Time-dependent density functional theory (TDDFT)
- Kohn-Sham density functional theory (KSDFT)
- Localization.
- Gauge choice is the key
- Used in community software packages: Quantum ESPRESSO, Wannier90, Octopus, PWDFT etc

Time-dependent density functional theory

Parallel transport gauge

Time-dependent density functional theory

[Runge-Gross, PRL 1984]
(spin omitted, N_e: number of electrons)

$$i\partial_t \psi_i(x,t) = \left[H\big(t, P(t)\big)\psi_i\right](x,t), \qquad i = 1, \dots, N_e$$

$$P(x,x',t) = \sum_{i=1}^{N_e} \psi_i(x,t)\psi_i^*(x',t)$$

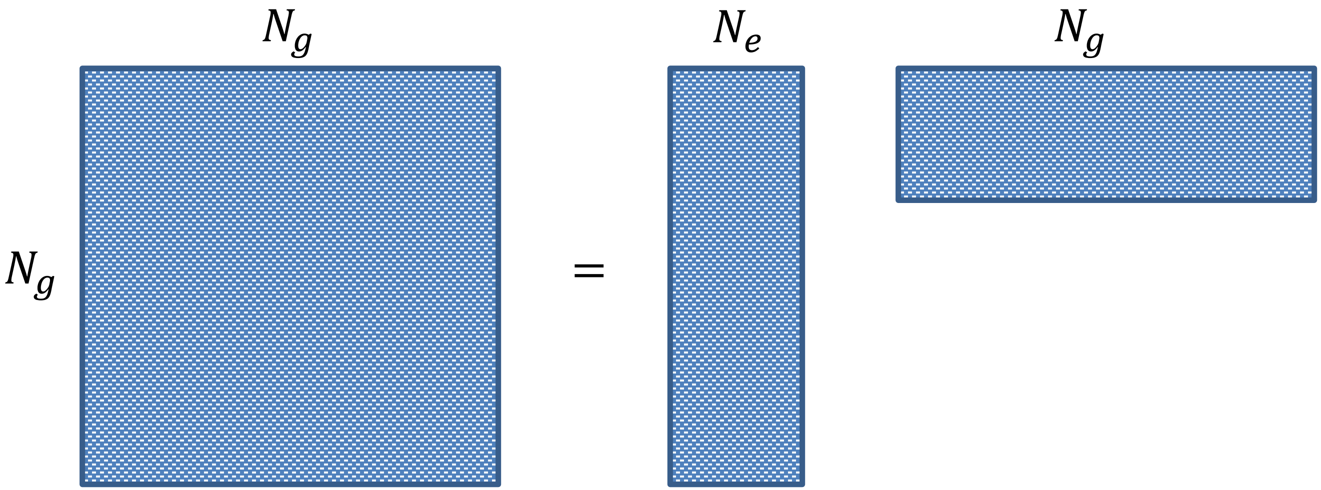
Hamiltonian

$$H\big(t, P(t)\big) = -\frac{1}{2}\Delta + V_{ext}(t) + V_{Hxc}[P(t)]$$

Density matrix

- Key quantity throughout the talk
- $\Psi(\mathrm{t}) = \left[\psi_1(t), \ldots, \psi_{N_e}(t)\right] \in \mathbb{C}^{N_g \times N_e}$.
  N_g: number of discretized grid points to represent each ψ_i

$$P(t) = \Psi(t)\left(\Psi(t)\right)^*$$



Gauge invariance

- Unitary rotation

$$\Phi = \Psi U, \qquad U^* U = I$$

$$P = \Phi\Phi^* = \Psi(UU^*)\Psi^* = \Psi\Psi^*$$

- Propagation of the density matrix, von Neumann equation (quantum Liouville equation)

$$i\partial_t P(t) = [H, P](t) = H\big(t, P(t)\big)P(t) - P(t)H\big(t, P(t)\big)$$

Gauge choice affects the time step

Extreme case (boring Hamiltonian)

$$H(t) \equiv H_0$$

Initial condition

$$H_0 \psi_i(0) = \psi_i(0)\varepsilon_i$$

Exact solution

$$\psi_i(t) = \psi_i(0)e^{-i\varepsilon_i t}$$ → Small time step

$$P(t) = P(0)$$ → (Formally) arbitrarily large time step

P v.s. Ψ

- Propagation of $P(t)$ requires storing $N_g \times N_g$ matrix. Often impractical for large basis set $(N_g = 10^3 \sim 10^7)$

- Propagation of $\Psi(t)$ requires storing $N_g \times N_e$ matrix, but is non-optimal due to gauge choice $(N_e = 10 \sim 10^4)$

- P is only a rank N_e matrix. Not enough information to justify storing as dense matrix.

- Common theme in quantum chemistry.

Propagation on the Grassmann manifold

- The Schrödinger gauge can be seen as

$$\Phi(t) = \Psi(t)U(t), \; U(t) \equiv I$$

- Optimal choice of $U(t)$ at each time step, to yield slowest possible dynamics ⇒ Optimization on the Grassmann manifold

- Related work: dynamical low rank decomposition [Koch-Lubich 2007; Lubich-Oseledets, 2014]

Optimization of gauge choice

- Gauge transformed wavefunctions

$$\Phi(t) = \Psi(t)U(t), \qquad U^*(t)U(t) = I$$

- Goal:

$$\min_{U(t)} ||\partial_t \Phi(t)||_F^2$$

- $P(t)\Phi(\mathrm{t}) = \Phi(t)$, optimization at each time step

Optimization of gauge choice

- Orthogonal decomposition

$$\|\dot{\Phi}\|_F^2 = \|P\dot{\Phi}\|_F^2 + \|(I-P)\dot{\Phi}\|_F^2$$

- After some derivation

$$\|(I-P)\dot{\Phi}\|_F^2 = \mathrm{Tr}\left[\dot{P}^2 P\right]$$

→ Gauge invariant!

- Optimizer (parallel transport condition)

$$P\dot{\Phi} = 0$$

- Also induce a parallel transport propagator

$$i\partial_t \mathcal{T} = [i\partial_t P, P]\mathcal{T}, \qquad \mathcal{T}(0) = I, \qquad \Phi(t) = \mathcal{T}(t)\Psi_0$$

Parallel transport (PT) dynamics

- Schrödinger dynamics

$$i\partial_t \Psi = H\Psi, \qquad \Psi(0) = \Psi_0$$

- Parallel transport dynamics

$$i\partial_t \Phi = H\Phi - \Phi(\Phi^* H\Phi), \qquad \Phi(0) = \Psi_0$$

  - Gauge implicitly defined by the dynamics
  - Only one additional mixing term needed, simple to implement
  - Optimal gauge among all possible choices
  - Mixing term $O(N^3)$ cost

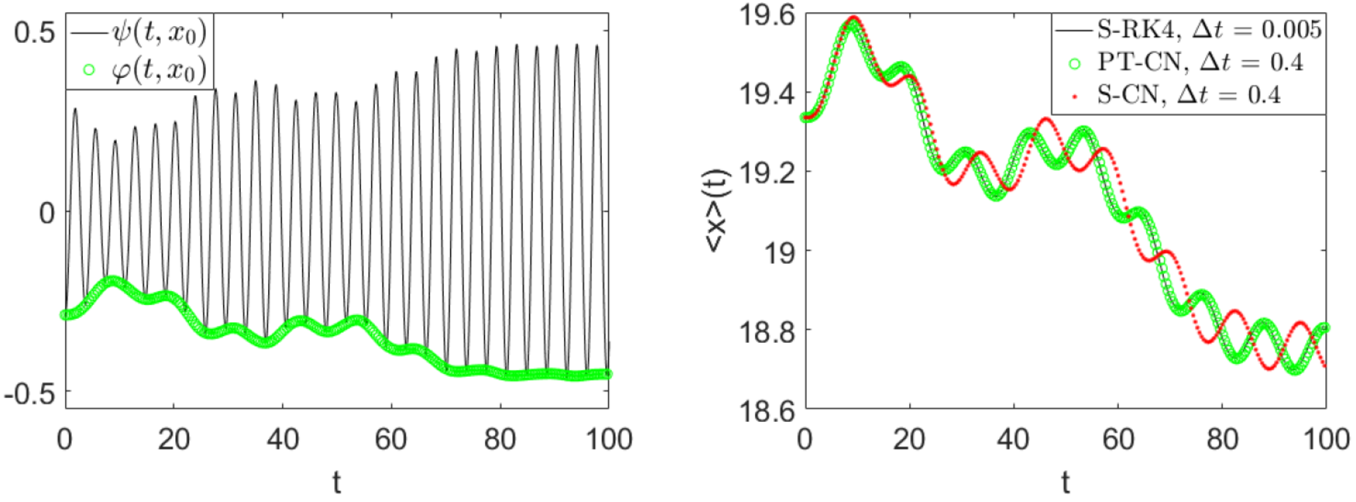
[An, L., arXiv:1804.02095]
[Jia, An, Wang, L., J. Chem. Theory Comput., 2018]

How PT dynamics works

- Boring Hamiltonian: $H(t) \equiv H_0,\ H_0\Psi_0 = \Psi_0\Lambda_0,\ \Phi_0 = \Psi_0$
- $H\Phi - \Phi(\Phi^* H\Phi) = 0 \Rightarrow\ i\partial_t\Phi = 0$

  Same behavior as density matrix

- 1D time-dependent Schrödinger:



(a) Wavefunctions

(b) Dipole moment

Implicit time integrator

- PT dynamics can be discretized with any integrator

- Smooth dynamics ⇒ Larger time

- Implicit integrator: Allow $||H\Delta t|| \gg 1$

Implicit time integrator

- Example: PT-CN (Crank-Nicolson scheme, a.k.a. trapezoidal rule)

$$\Phi_{n+1} + \imath\frac{\Delta t}{2}\{H_{n+1}\Phi_{n+1} - \Phi_{n+1}(\Phi^*_{n+1}H_{n+1}\Phi_{n+1})\}$$
$$=\Phi_n - \imath\frac{\Delta t}{2}\{H_n\Phi_n - \Phi_n(\Phi^*_n H_n\Phi_n)\}$$

- Φ_{n+1} needs to be solved self-consistently. Preconditioner. Mixing.

- Natural to combine with Ehrenfest dynamics

- Many others are possible

Analysis

- Near-adiabatic regime, singularly perturbed ($\varepsilon \ll 1$)

$$
\begin{aligned}
i\varepsilon\partial_t\psi &= H(t)\psi \\
i\varepsilon\partial_t\varphi &= H(t)\varphi - \varphi(\varphi^* H(t)\varphi)
\end{aligned}
$$

Theorem [An-L., 2018] symplectic integrator of order k, up to $T \sim O(1)$,

$$
\text{error} \sim \begin{cases} \dfrac{\Delta t^k}{\varepsilon^{k+1}}, & \text{Schrödinger} \\ \dfrac{\Delta t^k}{\varepsilon^k}, & \text{parallel transport} \end{cases}
$$

[An, L., arXiv:1804.02095]

Proof sketch

Key: Refined adiabatic theorem

Lemma There exists decomposition

$$\varphi(t) = \varphi_A(t) + \varepsilon\varphi_R(t)$$

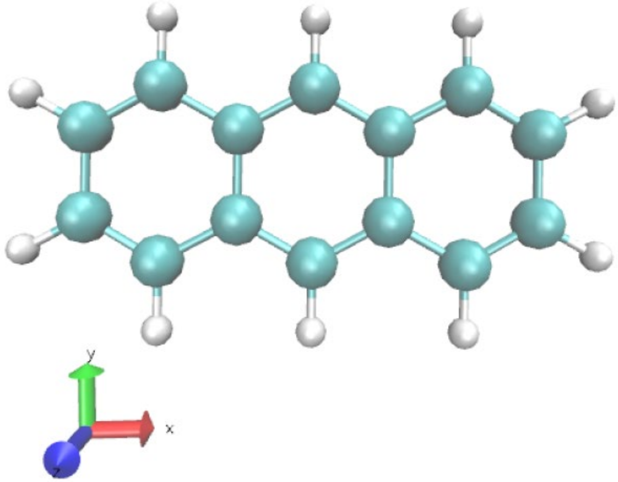
up to $T \sim O(1)$. $\varphi_A(t)$ is eigenstate of $H(t)$ and is ε-independent (including phase factor). $\|\varphi_R(t)\|_2$ is bounded ε-independent.

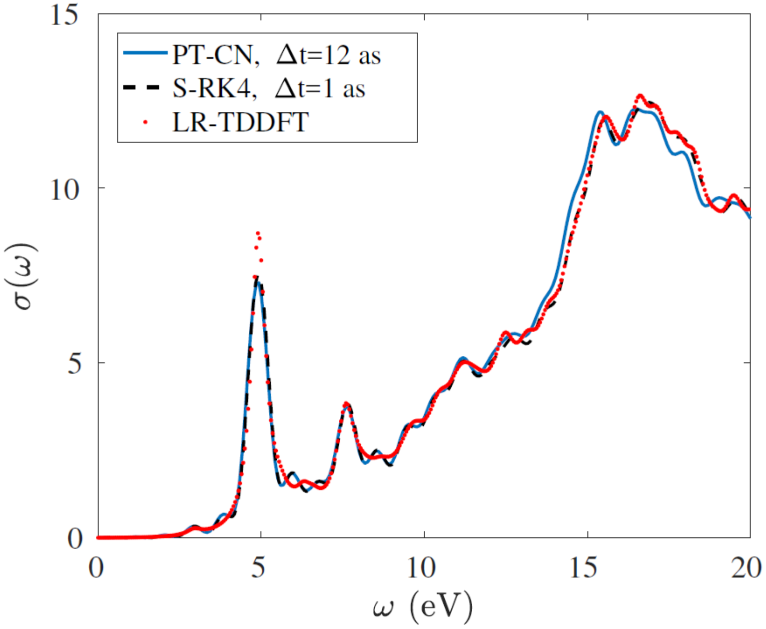
Implementation in electronic structure software packages

- DGDFT / PWDFT
- Octopus (most widely used package for TDDFT, in progress)

Absorption spectrum

- Anthracene ($\mathrm{C_{14}H_{10}}$)
- Compare RT-TDDFT (PWDFT) and linear response TDDFT (Quantum ESPRESSO).
- PT-CN 8 times faster than S-RK4

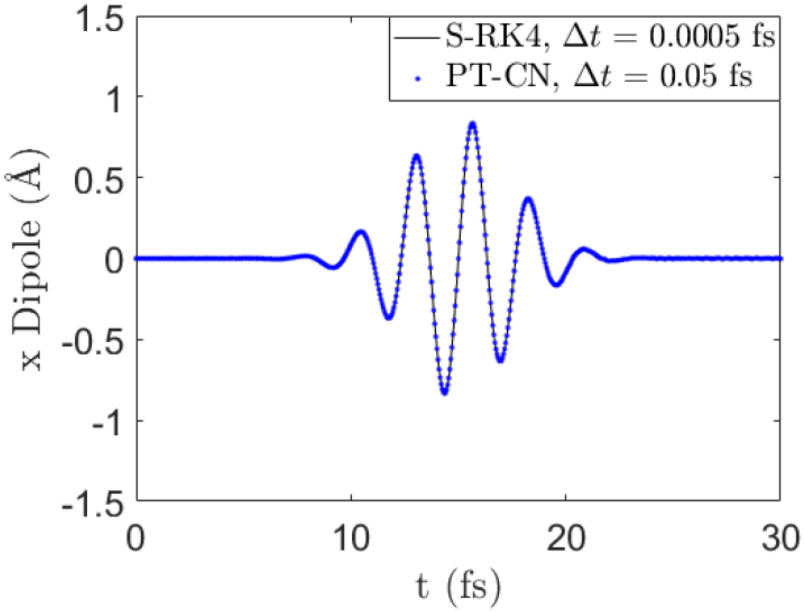


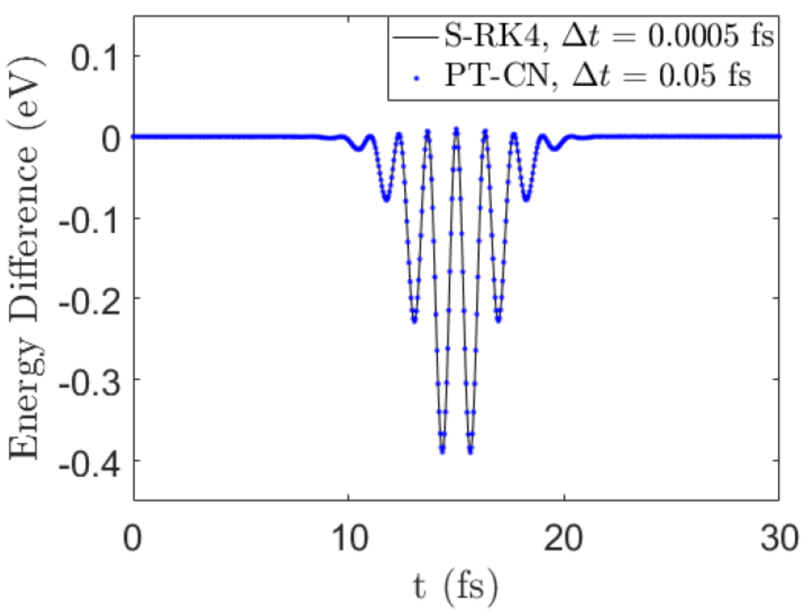


[Jia, An, Wang, L., J. Chem. Theory Comput., 2018]

Ultrafast dynamics

- Benzene
- Slow laser (800 nm wavelength)
- PT-CN 32 times faster than S-RK4

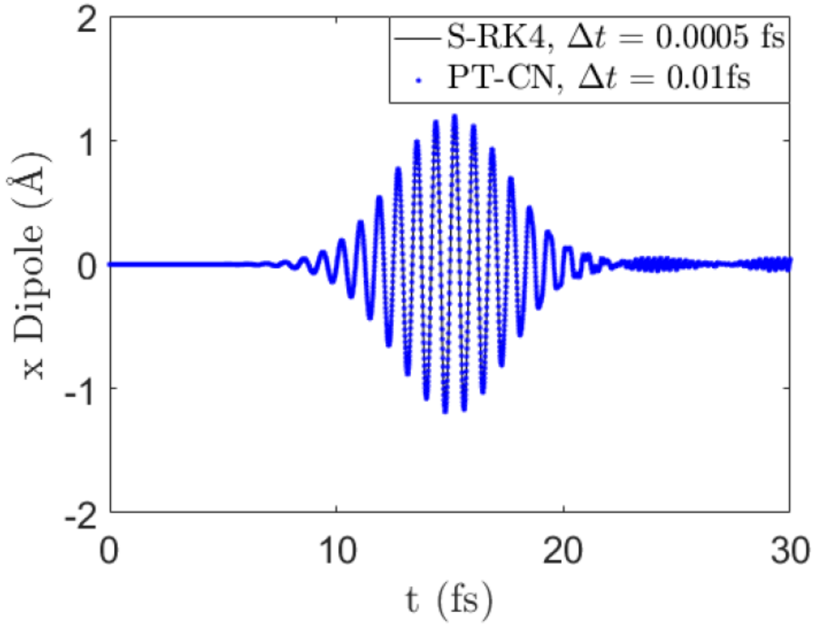


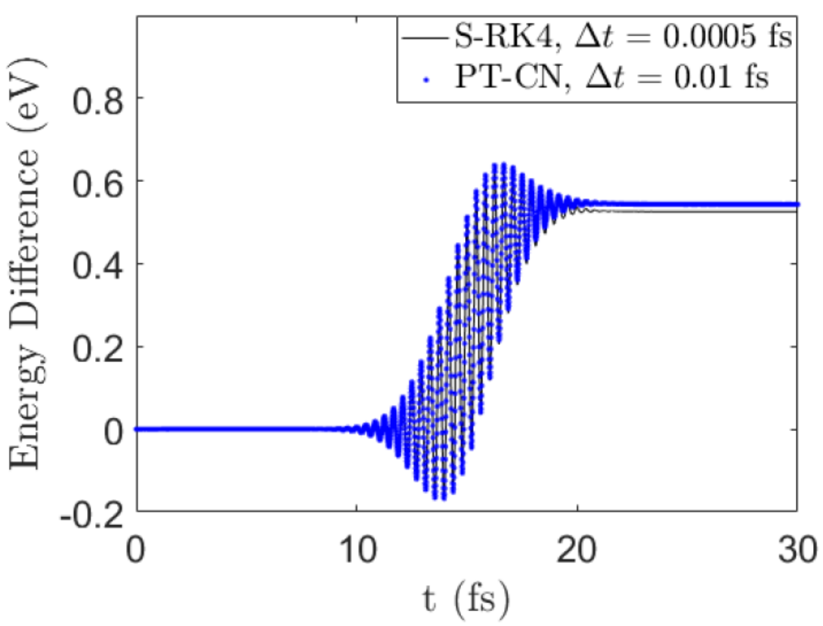


[Jia, An, Wang, L., J. Chem. Theory Comput., 2018]

Ultrafast dynamics

- Fast laser (250 nm wavelength)
- PT-CN 10 times faster than S-RK4

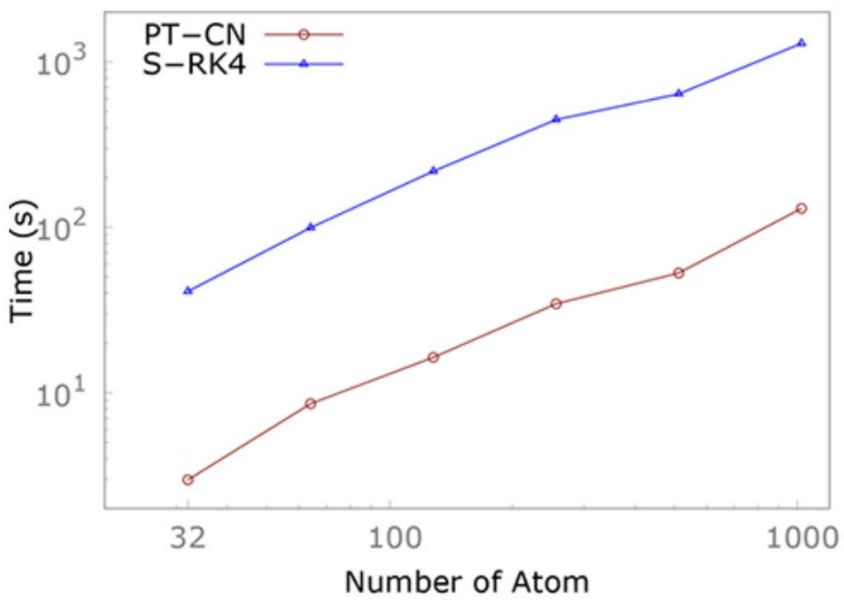




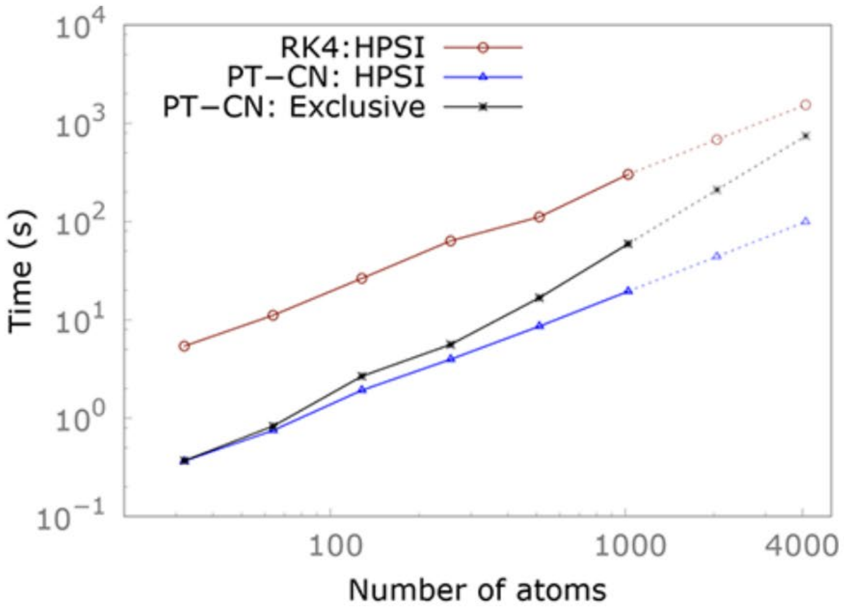
[Jia, An, Wang, L., J. Chem. Theory Comput., 2018]

Efficiency

- Silicon from 32 to 1000 atoms, ~10 times faster
- Parallel implementation up to 2000 cores.



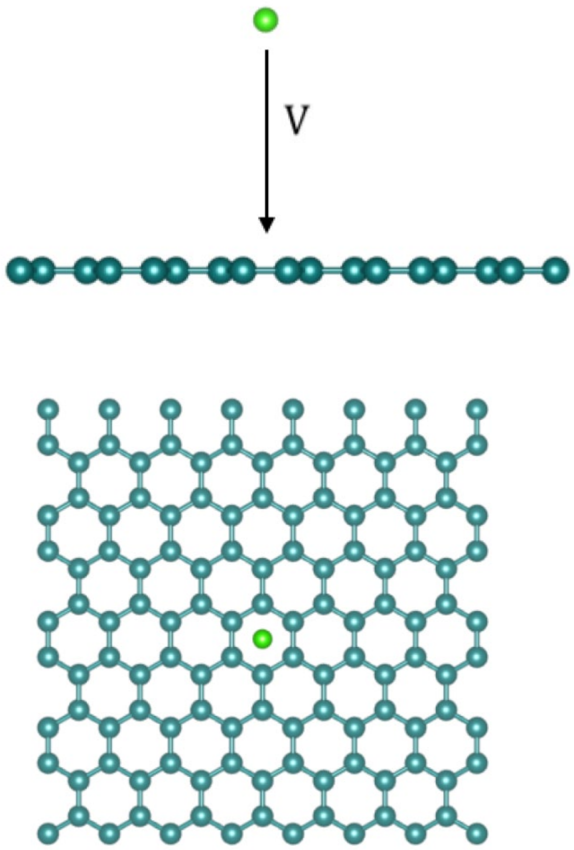
(a) Total wall clock time

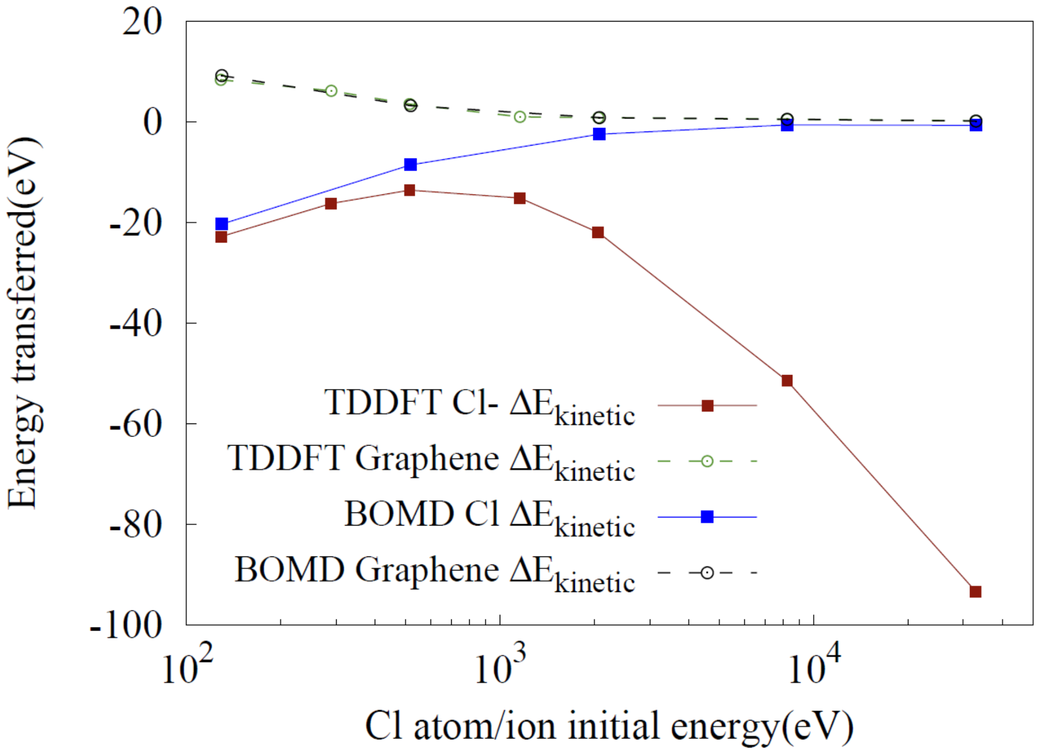


(b) Scalable components

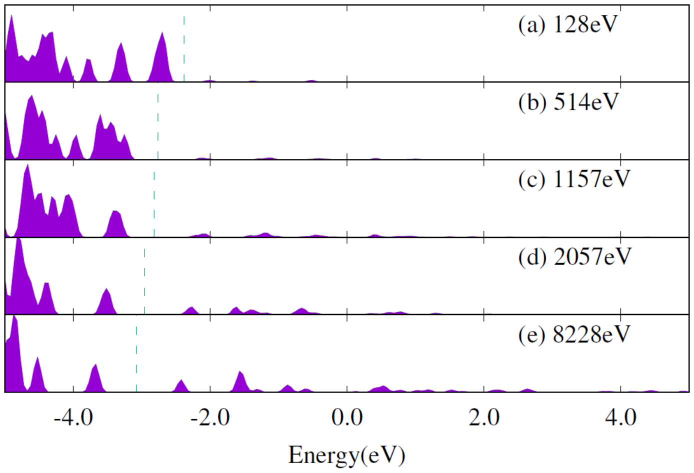
Ion collision

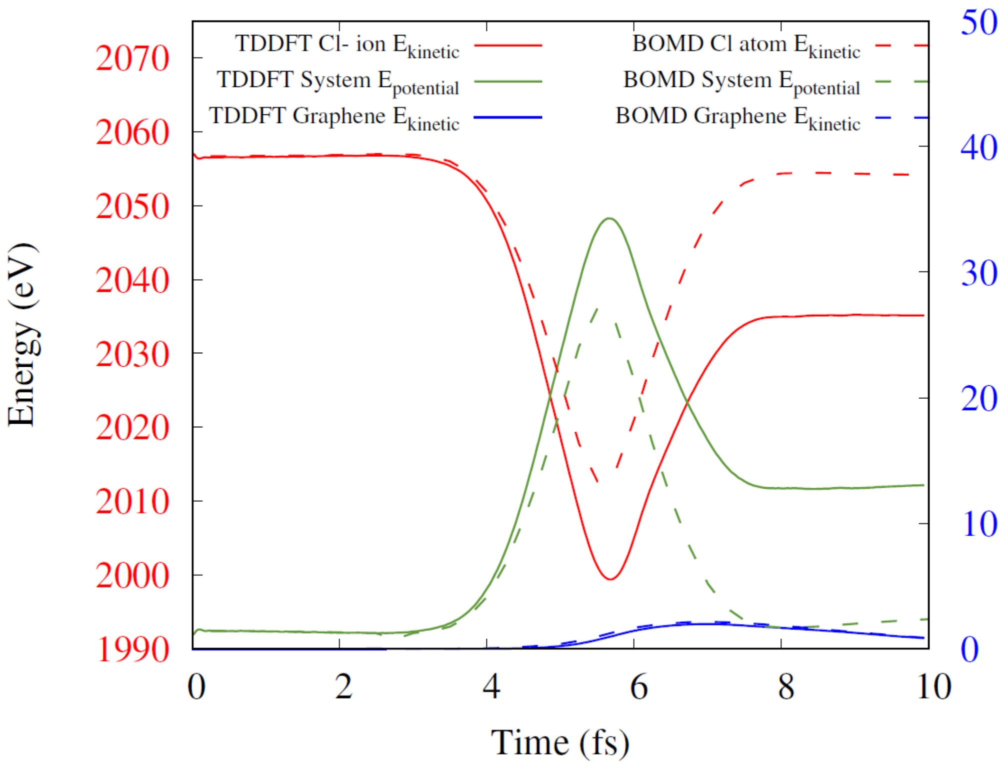
- Cl^- ion through graphene nanoflake (113 atoms)





Ion collision

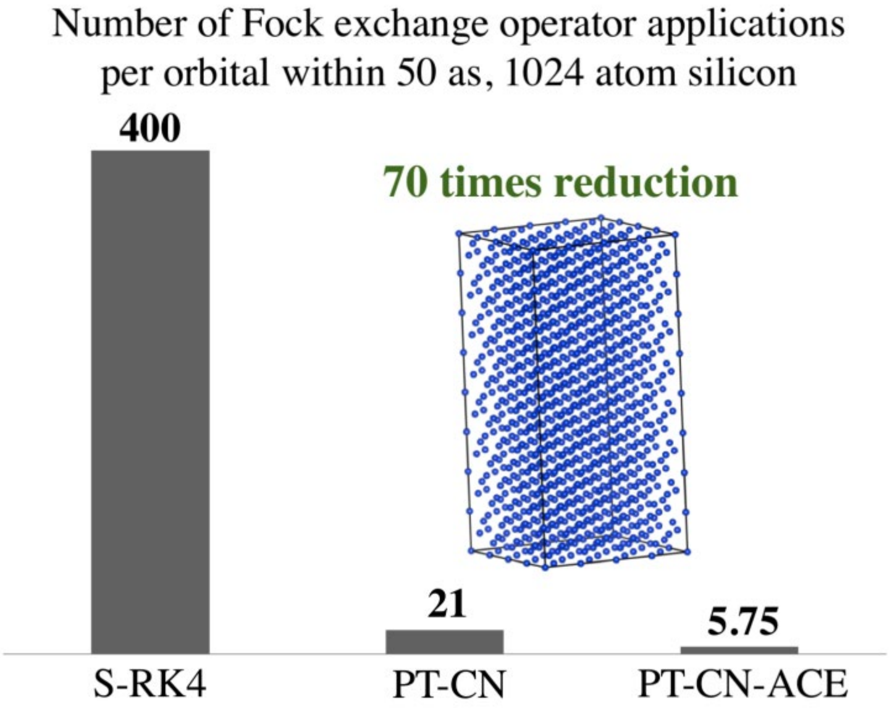




S-RK4 Δt: 0.5 as, 78 hours, 228 cores
PT-CN Δt: 50 as, 5.2 hours, 228 cores

Hybrid functional RT-TDDFT

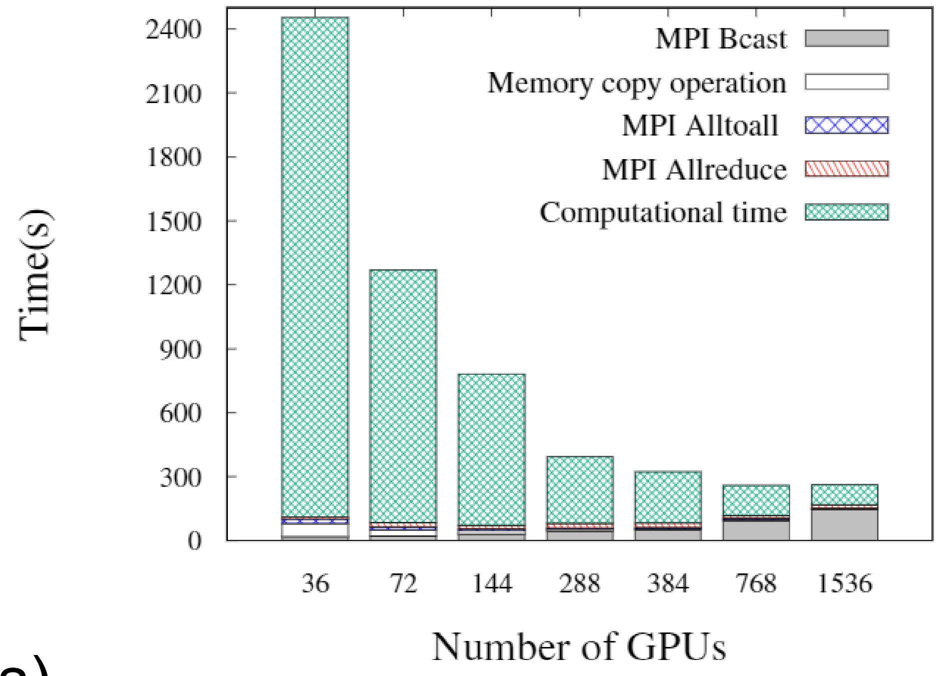
- For the first time, practical hybrid functional RT-TDDFT calculations with a large basis set



[Jia, L., Comput. Phys. Commun. 2019]

GPU acceleration on Summit

- No. 1 supercomputer in the Top500 list in November, 2018
- Silicon with 1536 atoms
- Scale to 768 GPUs
- 34 x times faster than CPU with 3072 cores
- 1.5 hr per femtosecond (fs)



[Jia, Wang, L., submitted]

Conclusion

- Parallel transport dynamics: optimal gauge choice

- Cost $\sim$ Schrödinger dynamics (wavefunction)
  Time step $\sim$ von Neumann dynamics (density matrix)

- Implicit time integrators demonstrated to be effective for TDDFT for the first time (c.f. Pueyo et al JCTC 2018). Significantly improved efficiency.

Electron localization

Selected columns of density matrix

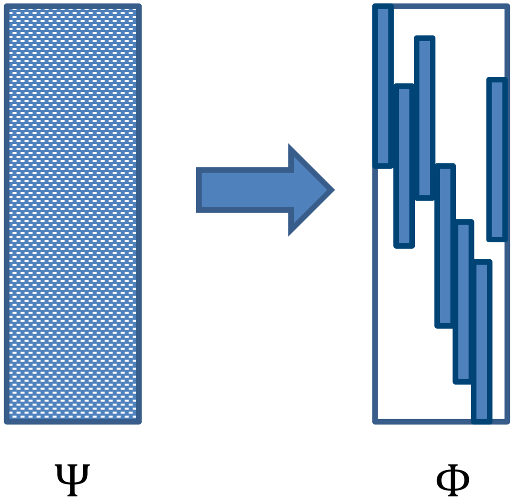
Localization problem

- Ψ: Kohn-Sham orbitals
  dense, unitary matrix of size $N_g \times N_e$ $(N_g \gg N_e)$

- ε-sparse representation

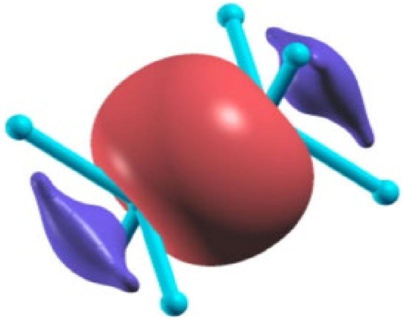
$$\|\Phi - \Psi\mathrm{U}\|_F \leq \varepsilon$$

  - Each column of Φ is sparse.
  - U is an $N_e \times N_e$ unitary matrix.

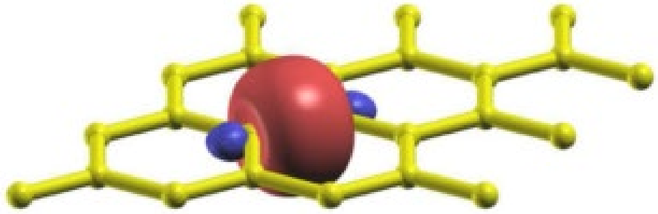


Wannier functions

- Maximally localized Wannier function (MLWF) [Marzari-Vanderbilt, Phys. Rev. B 1997].



Silicon



Graphene

- Reason for the existence of MLWF for insulating systems [Kohn, PR 1959] [Nenciu, CMP 1983] [Panati, AHP 2007], [Brouder et al, PRL 2007] [Benzi-Boito-Razouk, SIAM Rev. 2013] etc

Application of Wannier functions

- Analysis of chemical bonding
- Band-structure interpolation
- Basis functions for DFT calculations (representing occupied orbitals ψ_i)
- Basis functions for excited state calculations (representing Hadamard product of orbitals $\psi_i \odot \psi_j$)
- Strongly correlated systems (DFT+U)
- Phonon calculations
- etc

[Marzari et al. Rev. Mod. Phys. 2012]

Maximally localized Wannier functions

- Geometric intuition: Minimization of “spread” or second moment.

$$\min_{\substack{\Phi=\Psi U,\\ U^*U=I}} \Omega[\Phi]$$

$$\Omega[\Phi] = \sum_{j=1}^{N_e} \int \left|\phi_j(x)\right|^2 x^2 \, dx - \left(\int \left|\phi_j(x)\right|^2 x \, dx \right)^2$$

- U: gauge degrees of freedom

Maximally localized Wannier functions

Robustness

- Initialization: Nonlinear optimization and possible to get stuck at local minima.
- Entangled band: Localization in the absence of band gap.

Both need to be addressed for high throughput computation.

Density matrix perspective

Ψ is unitary, then

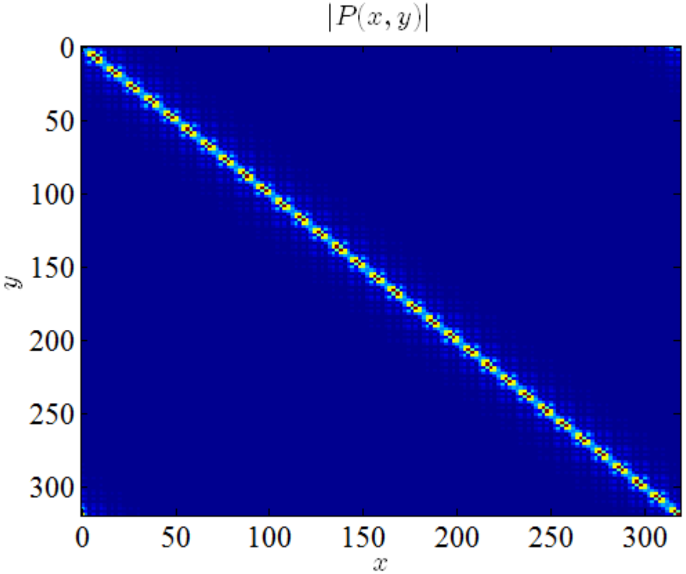
$$P = \Psi\Psi^*$$

is a projection operator, and is gauge invariant.

$$P = \Psi\Psi^* = \Phi(U^*U)\Phi^* = \Phi\Phi^*$$

is close to a sparse matrix.

- Can one construct sparse representation directly from the density matrix?



Algorithm: Selected columns of the density matrix (SCDM)

Code (MATLAB. Psi: matrix of size $N_g \times N_e$)

```
[U,R,perm] = qr(Psi', 0);
```

← Pivoted QR

```
Phi = Psi * U;
```

← GEMM

- Very easy to code and to parallelize!
- Deterministic, no initial guess.
- perm encodes selected columns of the density matrix

[Damle-L.-Ying, JCTC, 2015]

k-point

- Strategy: find columns using one “anchor” k-point (such as Γ), and then apply to all k-points

$$P(\mathbf{k}) = \sum_{\varepsilon_{b,\mathbf{k}} \in \mathcal{I}} |\psi_{b,\mathbf{k}}\rangle\langle\psi_{b,\mathbf{k}}| = \sum_{\varepsilon_{b,\mathbf{k}} \in \mathcal{I}} |\widetilde{\psi}_{b,\mathbf{k}}\rangle\langle\widetilde{\psi}_{b,\mathbf{k}}|$$

$$(\Xi^*(\mathbf{k})\Xi(\mathbf{k}))_{b,b'} = \sum_{b''=1}^{N_b} \psi_{b'',\mathbf{k}}(\mathbf{r}_b)\psi^*_{b'',\mathbf{k}}(\mathbf{r}_{b'}) = P(\mathbf{r}_b, \mathbf{r}_{b'}; \mathbf{k})$$

$$U(\mathbf{k}) = \Xi(\mathbf{k})\left[\Xi^*(\mathbf{k})\Xi(\mathbf{k})\right]^{-\frac{1}{2}}$$

SCDM: A flexible framework

- Molecular or periodic systems
- Isolated or entangled bands
- Variational formulation

[Damle-L., MMS 2018]
[Damle-L.-Levitt, MMS 2019]

SCDM in software packages

- MATLAB/Julia code

  https://github.com/asdamle/SCDM

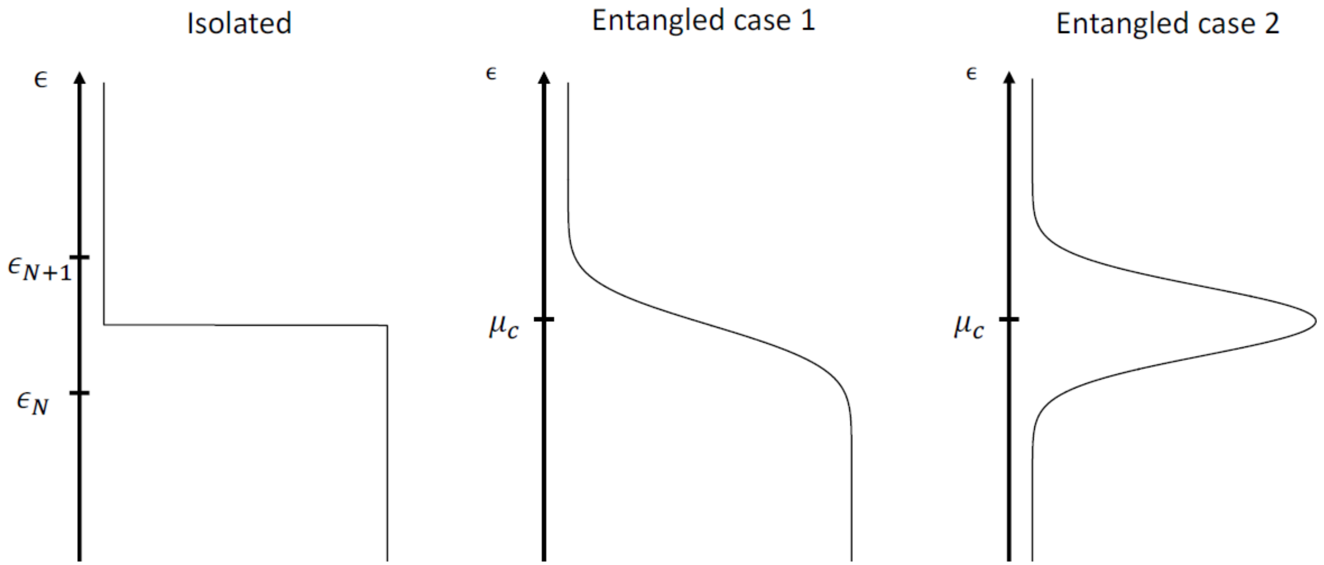
  https://github.com/antoine-levitt/wannier

- Quantum ESPRESSO [I. Carnimeo, S. Baroni, P. Giannozzi, arXiv: 1801.09263]

- Wannier90 [V. Vitale et al] (starting from v3.0)

  http://www.wannier.org/

Entangled bands



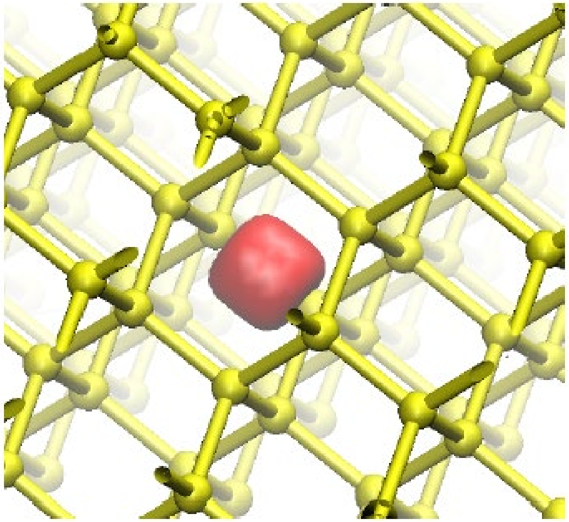
Entangled case 1 (metal, valence + conduction):

$$f(\varepsilon) = \frac{1}{2}\mathrm{erfc}\left(\frac{\varepsilon-\mu_c}{\sigma}\right) = \frac{1}{\sqrt{\pi\sigma^2}}\int_{\varepsilon}^{\infty}\exp\left(-\frac{(t-\mu_c)^2}{\sigma^2}\right)\,\mathrm{d}t.$$

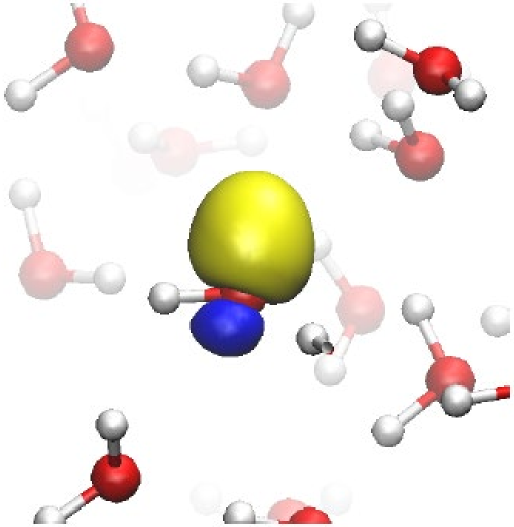
Entangled case 2 (near Fermi energy):

$$f(\varepsilon) = \exp\left(-\frac{(\varepsilon-\mu_c)^2}{\sigma^2}\right)$$

Examples of SCDM orbitals (Γ-point)

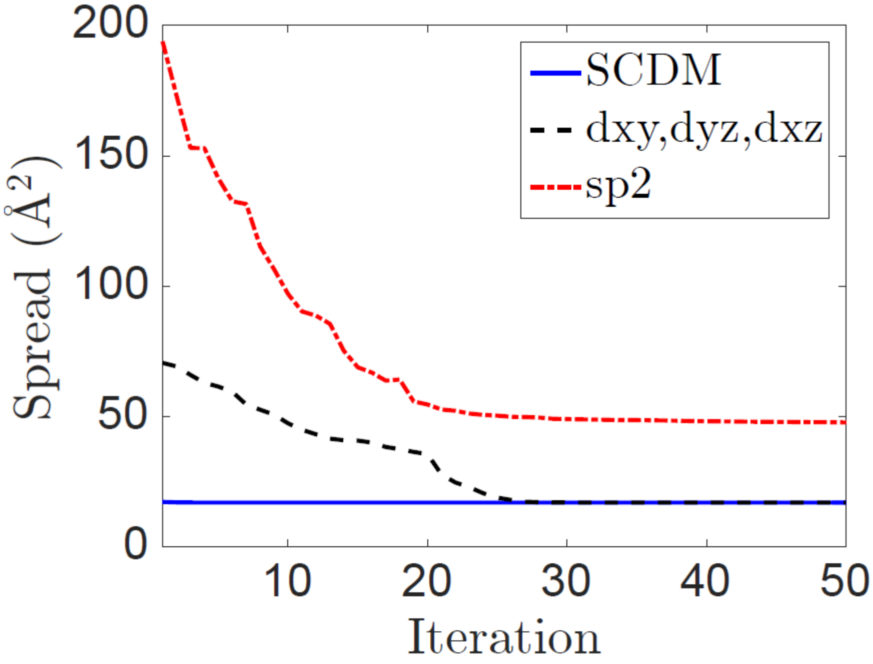


Silicon

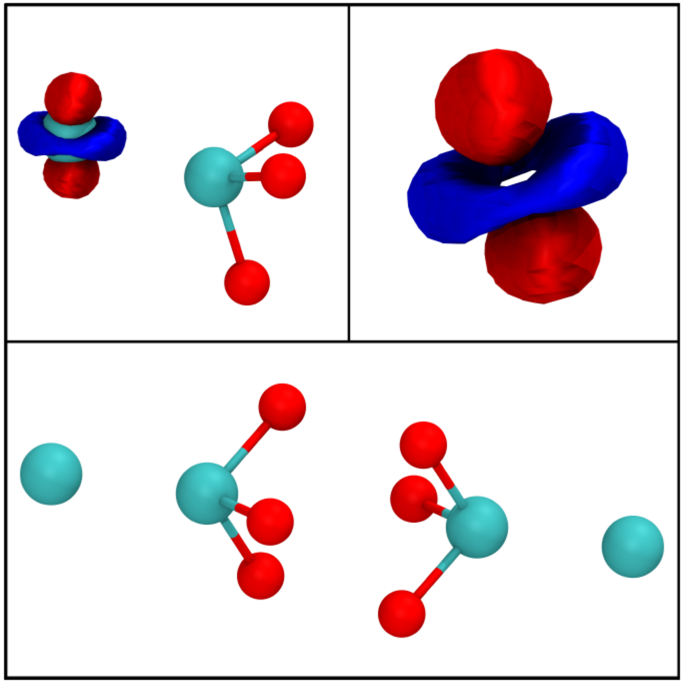


Water

Examples of SCDM orbitals (k-point)



(a)



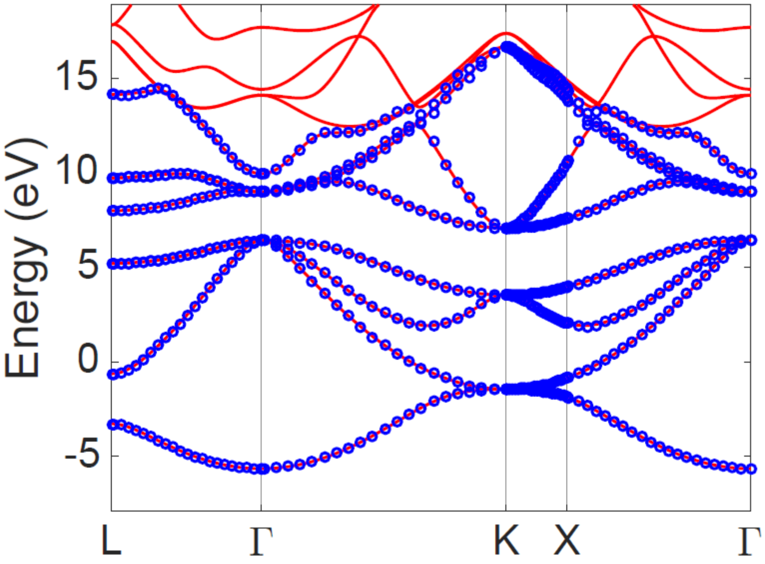
(b)

Cr2O3. k-point grid $6 \times 6 \times 6$

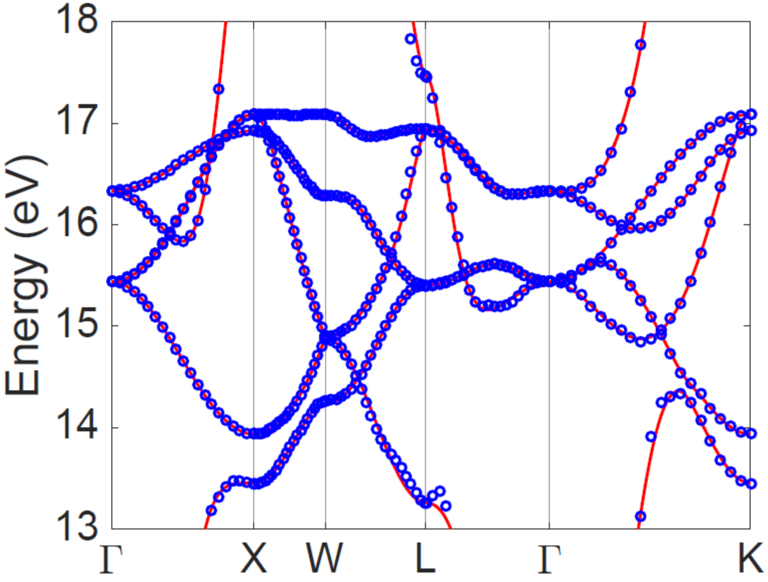
Initial spread from SCDM: 17.22 Å^2

MLWF converged spread: 16.98 Å^2

Example: band interpolation



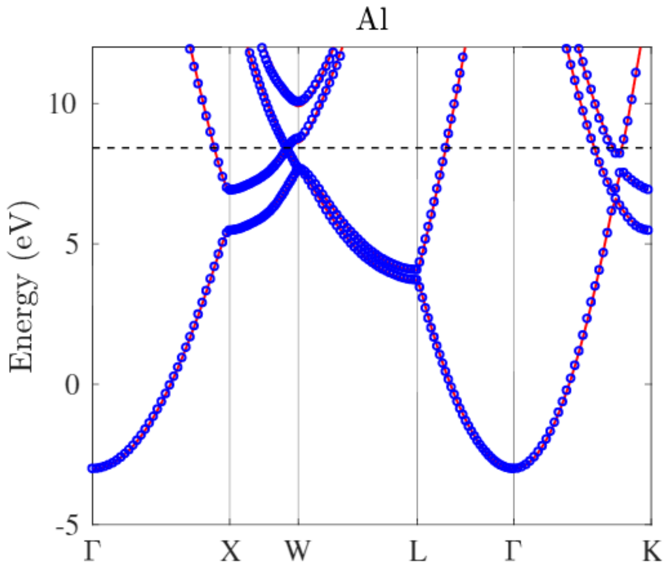
Si



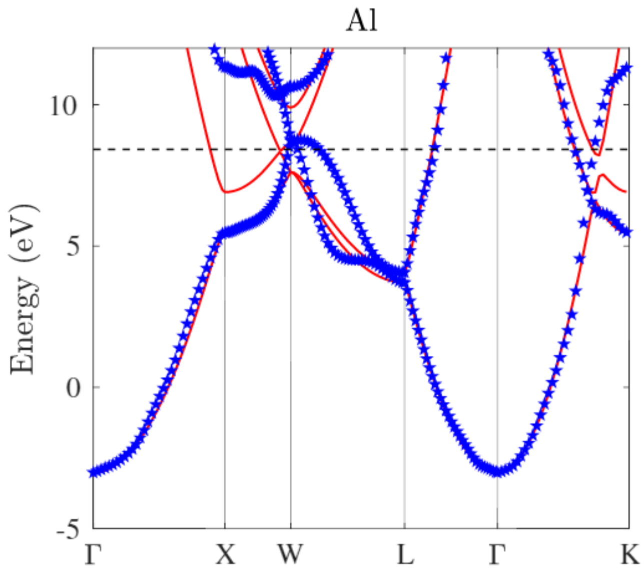
Cu

Band structure interpolation: Al

10x10x10 k-points, 6 bands ⇒ 4 bands (no disentanglement)



SCDM spread:
18.38 Å^2

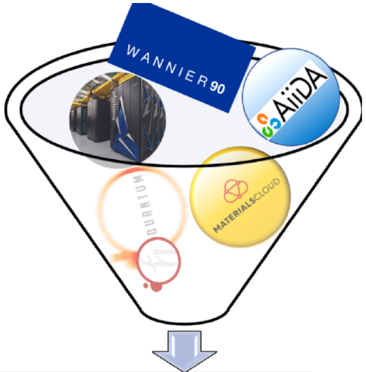


Wannier: optimized spread:
12.42 Å^2

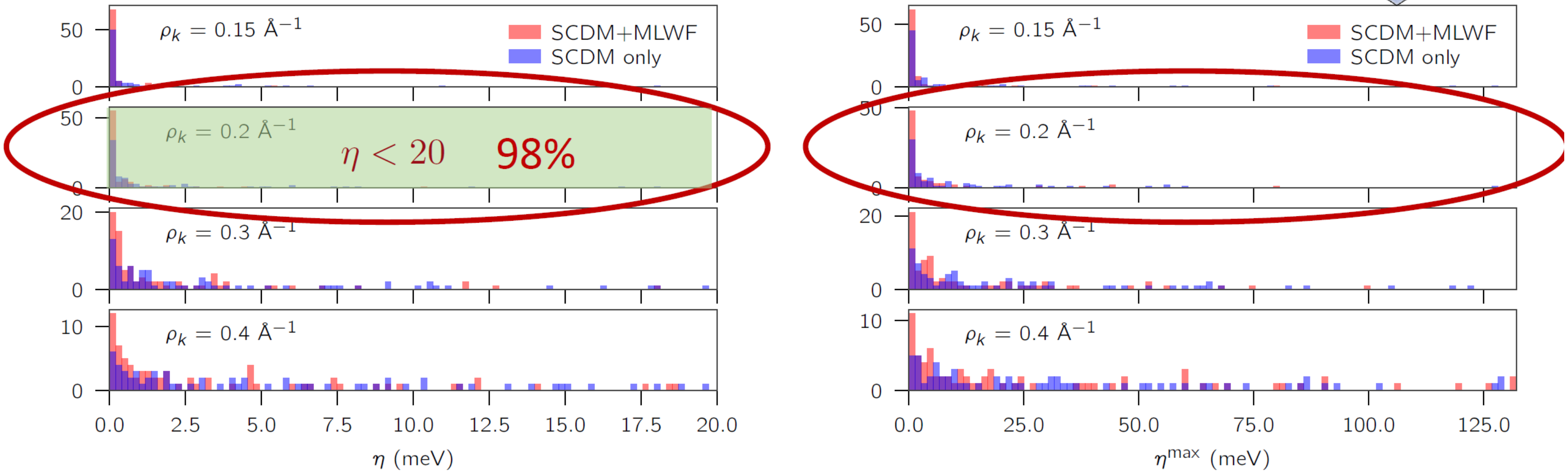
Smaller spread ↛ Better interpolation

Implementation in Wannier 90

Validation --- Results for insulators



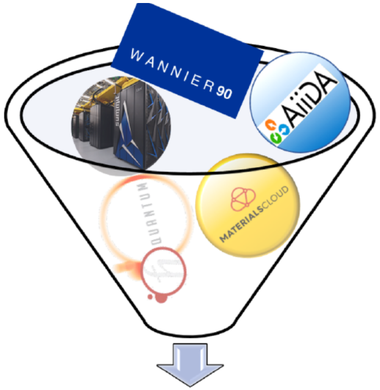
- **81 insulating** 3D materials spanning the chemical and structural space
- PARAMETER-FREE



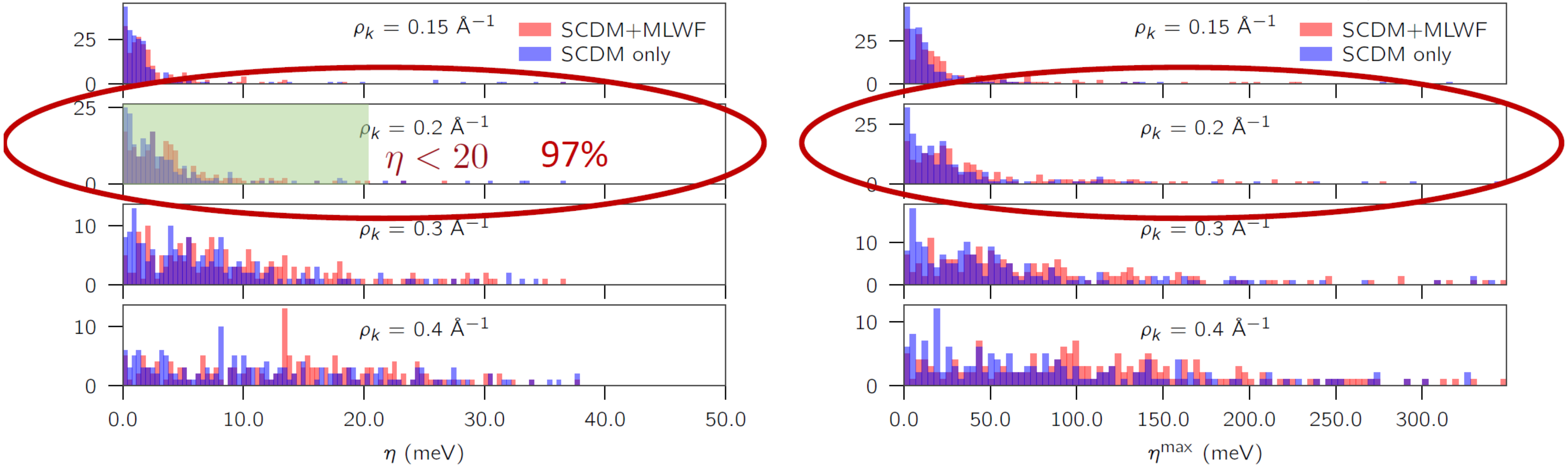
Credit: Dr. Valerio Vitale (APS March meeting 2019)

Implementation in Wannier 90

High-throughput --- Results for metals



- **Valence and conduction** bands of a set of **200** 3D materials spanning the chemical and structural space
- TWO PARAMETERS (μ,σ)



Credit: Dr. Valerio Vitale (APS March meeting 2019)

Conclusion

- Extract information from the gauge-invariant density matrix

- Selected columns of the density matrix (SCDM) is a simple, robust and deterministic method to find localized orbitals. Alternative method to the Maximally Localized Wannier functions.

- Wannier localization can be robustly initialized with SCDM (already in Wannier90). High-throughput materials simulation

Kohn-Sham density functional theory

Projected commutator DIIS

Kohn-Sham density functional theory

$$H[\rho]\psi_i(x) = \left(-\frac{1}{2}\Delta + V_{ext}(x) + V_{Hxc}[P](x)\right)\psi_i(x) = \varepsilon_i\psi_i(x)$$

$$P(x,x',t) = \sum_{i=1}^{N_e} \psi_i(x,t)\psi_i^*(x',t),$$

$$\int dx\, \psi_i^*(x)\psi_j(x) = \delta_{ij}, \qquad \varepsilon_1 \le \varepsilon_2 \le \cdots \le \varepsilon_{N_e}$$

- In principle exact many body ground state energy.
- Most widely used electronic structure theory for condensed matter systems and molecules
- [Hohenberg-Kohn,1964], [Kohn-Sham, 1965], Nobel Prize in Chemistry, 1998

Self consistent field / Fixed point problem

$$x = f(x)$$

Kohn-Sham density functional theory

$$P = \mathcal{F}[P]$$

More precisely

$$H[\mathrm{P}]\psi_i = \psi_i \lambda_i, \qquad P = \sum_i \psi_i \psi_i^*$$

SCF: Fixed point problem

- Fixed point iteration $x_{k+1} = f(x_k)$. Diverges quickly.

- Simple mixing. $x_{k+1} = \alpha f(x_k) + (1 - \alpha)x_k$. Often works but can be very slow

- Broyden-type methods: method of choice for practical calculations.

- Direct Inversion of the iterative subspace (DIIS) [Pulay 1980; 1982] (another equivalent version is Anderson acceleration [Anderson 1965])

SCF: Quantum chemist’s viewpoint

- commutator-DIIS (c-DIIS). Most widely used method in quantum chemistry software packages

- Residual

$$R[P] = [H[P], P] = H[P]P - PH[P]$$

- Necessary condition for SCF convergence

$$R[P] = 0$$

[Pulay CPL 1982; Kudin-Scuseria-Cances JCP 2002; Kudin-Scuseria M2AN 2007]
[T. Rohwedder, R. Schneider, JMC 2011]

C-DIIS

- Extrapolate the Fock matrix

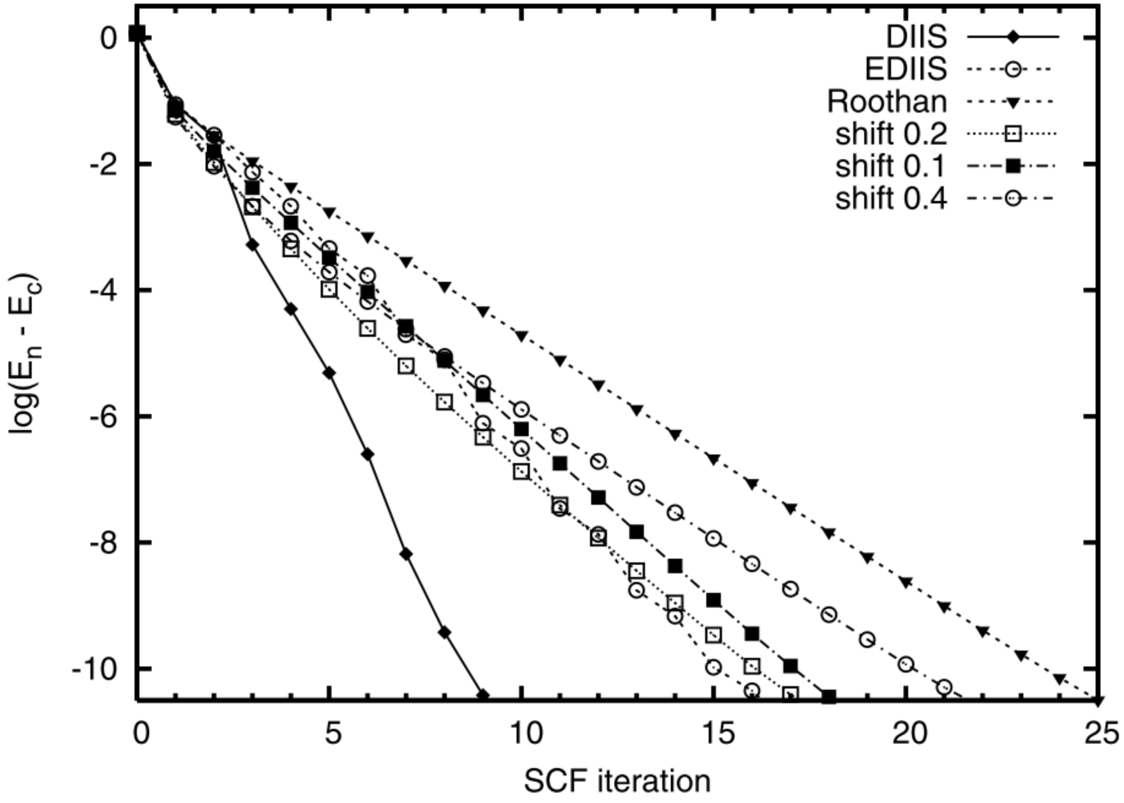
$$H^{(k+1)} = \sum_{j=k-\ell}^{k} \alpha_j H^{(j)}$$

- Obtain coefficients by minimizing the residual

$$\min_{\{\alpha_j\}} \left\| \sum_{j=k-\ell}^{k} \alpha_j R^{(j)} \right\|_F^2, \qquad \text{s.t.} \quad \sum_{j=k-\ell}^{k} \alpha_j = 1.$$

- Equivalent to a version of Broyden’s method

Convergence of C-DIIS



Comparison of SCF convergence patterns for (CH3)2CHOH with the Huckel guess at the RHF/6-31G(d) level of theory. [Kudin-Scuseria M2AN 2007]

C-DIIS is not suitable for large basis set

- Large basis: planewaves, finite elements, wavelets etc

$$P = \Psi\Psi^* \in \mathbb{C}^{N_g \times N_g}, \qquad \Psi \in \mathbb{C}^{N_g \times N_e}$$

- $N_g \sim 1000 N_e$. Cannot store H, R, P

Projected commutator DIIS

- Project to a set of gauge fixing orbitals $\Phi_{ref} \in \mathbb{C}^{N_g \times N_e}$ with full column rank

$$\Phi = P\Phi_{ref}$$

- Use Φ as the mixing variable (gauge-invariant)

[Hu-L.-Yang, JCTC, 2017]

Projection

- $\Phi \Leftrightarrow P$, if Φ has full column rank

$$P = \Phi(\Phi^*\Phi)^{-1}\Phi^*$$

- Use projected commutator

$$R_\Phi = H[P]P\Phi_{ref} - PH[P]\Phi_{ref}$$

- Enables C-DIIS for large basis set for the first time

[Hu-L.-Yang, JCTC, 2017]

PC-DIIS

- Minimize

$$\min_{\sum_j \alpha_j = 1} \left\| \sum_{j=k-l}^{k} \alpha_j R_\Phi^{(j)} \right\|_F$$

- Update

$$\widetilde{\Phi} = \sum_{j=k-l}^{k} \alpha_j \Phi_k$$

- Compute

$$\Phi_{k+1} = \widetilde{\Phi}\left(\widetilde{\Phi}^* \widetilde{\Phi}\right)^{-1}\left(\widetilde{\Phi}^* \Phi_{ref}\right)$$

PC-DIIS

Projected commutator

$$R_\Phi = H[P]P\Phi_{ref} - PH[P]\Phi_{ref}$$

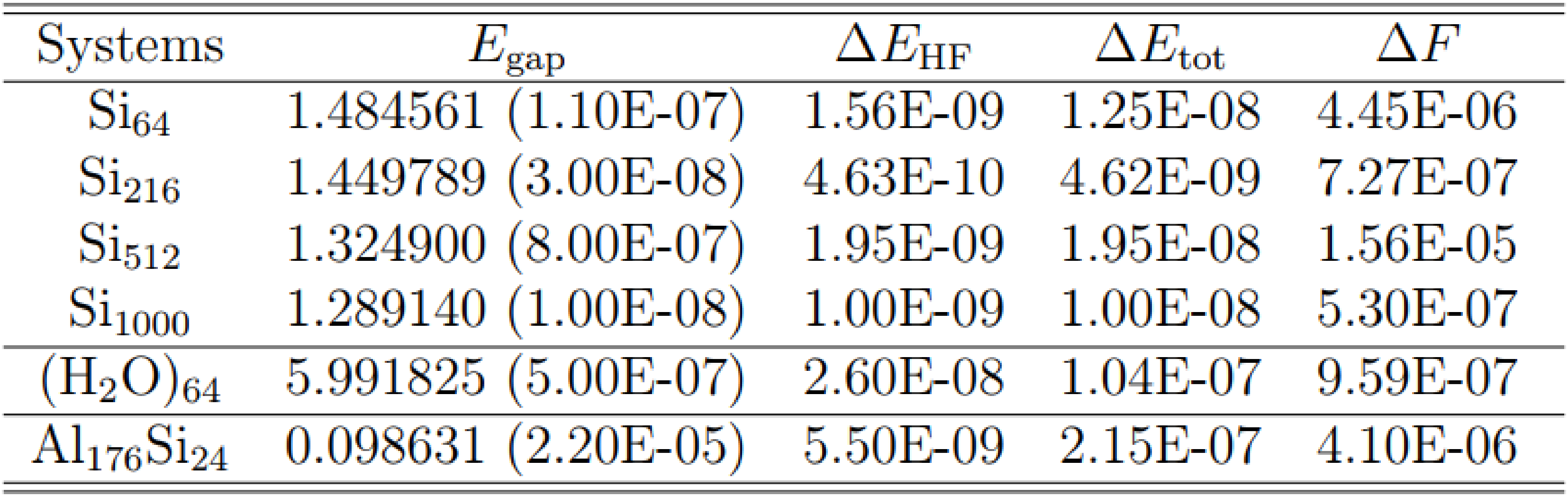
- If $\Phi_{ref} = \Psi$

$$R_\Phi = H[P]\Psi - \Psi\Lambda$$

⇒ Residual of eigenvalue problem.

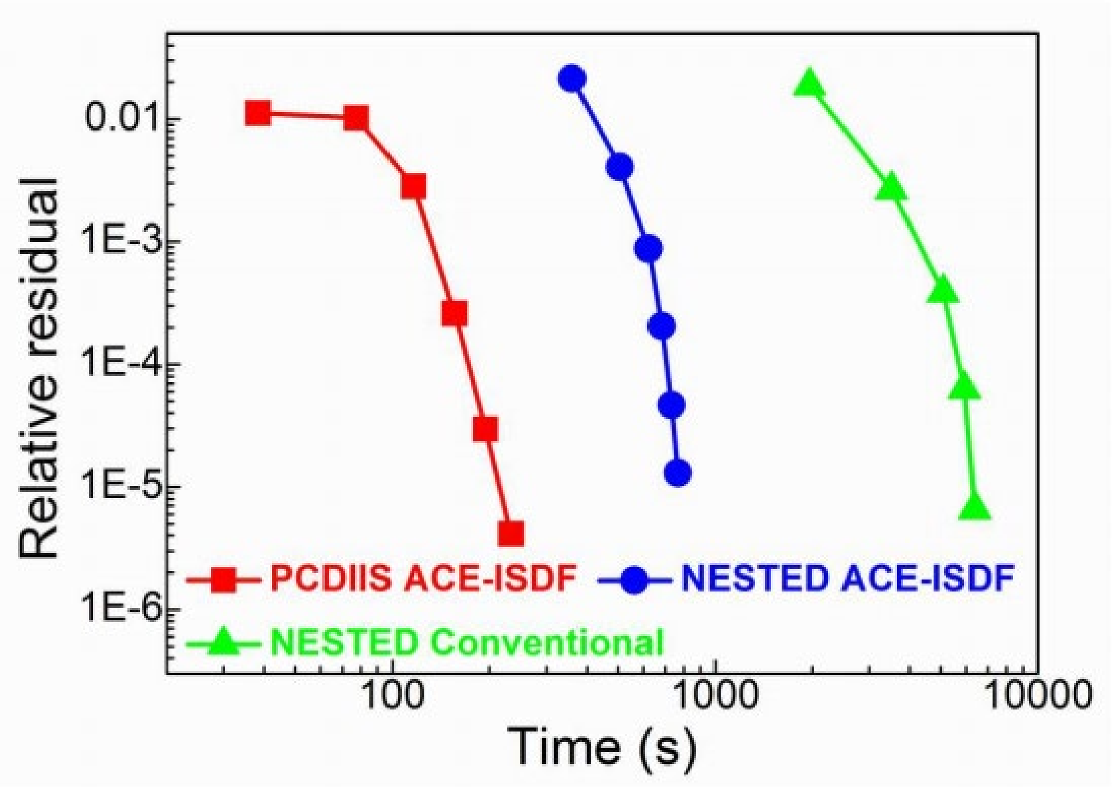
- All operations are limited to matrices of size $N_g \times N_e$

Accuracy



Systems	E_{gap}	ΔE_{HF}	ΔE_{tot}	ΔF
Si_{64}	1.484561 (1.10E-07)	1.56E-09	1.25E-08	4.45E-06
Si_{216}	1.449789 (3.00E-08)	4.63E-10	4.62E-09	7.27E-07
Si_{512}	1.324900 (8.00E-07)	1.95E-09	1.95E-08	1.56E-05
Si_{1000}	1.289140 (1.00E-08)	1.00E-09	1.00E-08	5.30E-07
$(\text{H}_2\text{O})_{64}$	5.991825 (5.00E-07)	2.60E-08	1.04E-07	9.59E-07
$\text{Al}_{176}\text{Si}_{24}$	0.098631 (2.20E-05)	5.50E-09	2.15E-07	4.10E-06

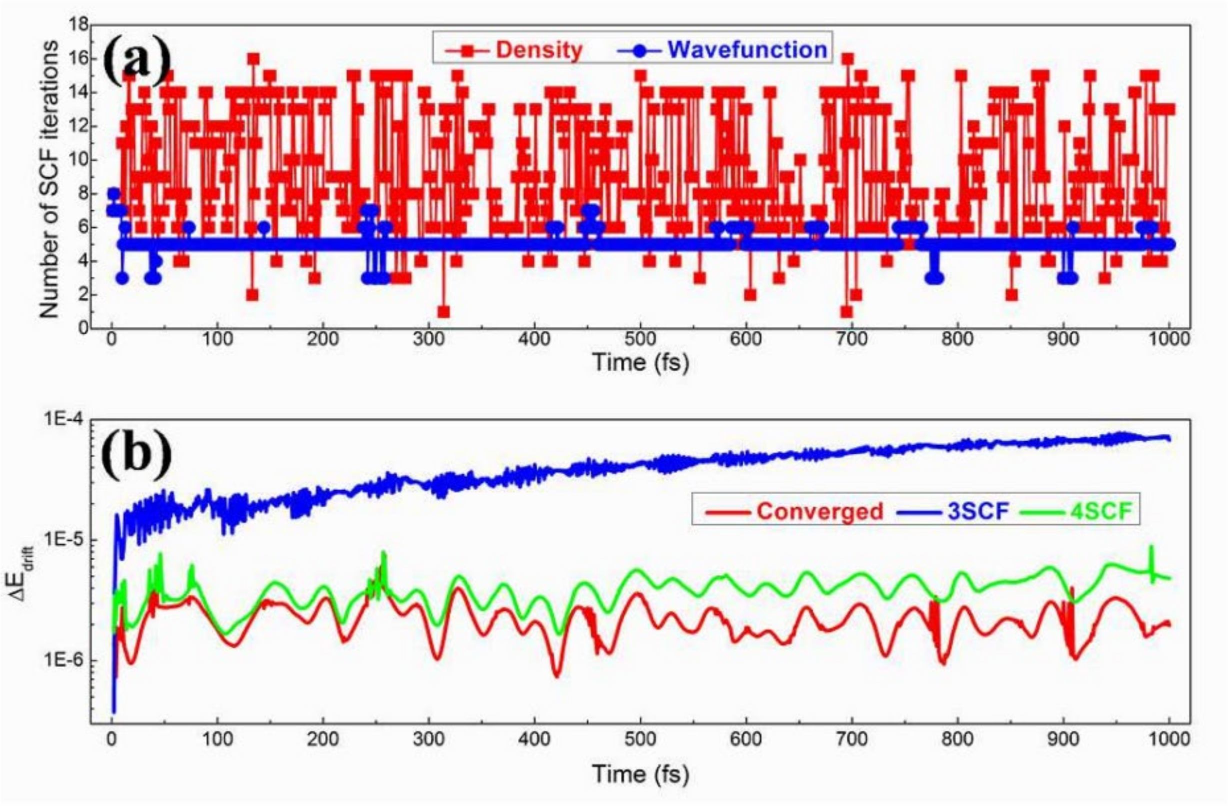
Efficiency



Si 1000 atom on 2000 cores

Ab initio molecular dynamics

- Wavefunction extrapolation with fixed gauge



Conclusion

- Projected C-DIIS enables C-DIIS for large basis set by a gauge fixing procedure.
- Recent usage in quantum embedding theory (joint work with Garnet Chan)
- Remark: Convergence in the nonlinear setup for DIIS is very much missing

Acknowledgement

- Dong An (Berkeley)
- Anil Damle (NSF postdoc, now Assistant Professor at Cornell)
- Wei Hu (LBNL, now Professor at USTC)
- Weile Jia (Berkeley)

- Lin-Wang Wang (LBNL)
- Chao Yang (LBNL)
- Lexing Ying (Stanford)

DOE Base Math, CAMERA, SciDAC, Early Career
NSF CAREER.
NERSC and BRC for Computing Resources

Thank you for your attention!