

# CDFCI and Beyond: Tackling Large-Scale Eigenvalue Problems in Many-Body Quantum Mechanics with Coordinate Descent Methods

Yuejia Zhang

Fudan University

March 2026

# Background and Challenges

- In quantum mechanics, the Hamiltonian operator  $\hat{H}$  represents the total energy of the system, including the contributions of kinetic energy and potential energy.
- The evolution of the state of the system over time (according to the Schrödinger picture)

$$i\hbar \frac{\partial}{\partial t} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle. \quad (1)$$

- The spectrum of the Hamiltonian operator determines the energy level structure and dynamic behavior of the system

$$\hat{H}\Psi = E\Psi. \quad (2)$$

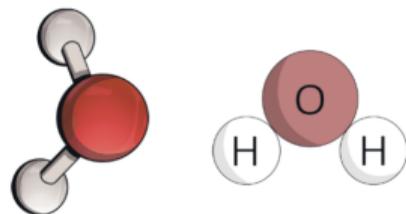
- The essential difficulty of the many-body problem: no analytical solution for  $n_{\text{elec}} > 1$ .  
The size of the problem grows exponentially with  $n_{\text{elec}}$ .

# Numerical Discretization of Schrödinger Equation

- Let's assume  $\Psi \in L^2(\mathbb{R}^{3n_{\text{elec}}}, \mathbb{C})$  for now. ( $\|\Psi\|_{L^2} = 1$ )
- **Finite Element Method** with naive grid  $10 \times 10 \times 10$ :  $10^{3n_{\text{elec}}}$  grid points.
- **Spectral Method** with (truncated) one-electron basis functions  $\{\varphi_p\}_{p=1}^{10}$ :

$$L^2(\mathbb{R}^{3n_{\text{elec}}}, \mathbb{C}) = \bigotimes_{i=1}^{n_{\text{elec}}} L^2(\mathbb{R}^3, \mathbb{C}),$$

$10^{n_{\text{elec}}}$  many-body basis functions.



# Famous Quote of Paul Dirac

“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and **the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.** It therefore becomes desirable that **approximate practical methods of applying quantum mechanics should be developed,** which can lead to an explanation of the main features of complex atomic systems without too much computation.”  
(P. A. M. Dirac, 1929.)



# System-Level Approximations

- Under Born–Oppenheimer approximation, the Hamiltonian operator with  $n_{\text{nuc}}$  nuclei and  $n_{\text{elec}}$  electrons is

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^{n_{\text{elec}}} \nabla_i^2 + \sum_{i=1}^{n_{\text{elec}}} V_{\text{ext}}(r_i; \{R_l\}_{l=1}^{n_{\text{nuc}}}) + \sum_{i < j}^{n_{\text{elec}}} \frac{1}{\|r_i - r_j\|}.$$

- **Hartree–Fock method:** mean-field approximation, neglecting electron correlation.
- **Density Functional Theory:** reformulate the problem in terms of electron density, with approximated exchange-correlation functionals.

$$\hat{H} \approx \sum_{i=1}^{n_{\text{elec}}} \left( -\frac{1}{2} \nabla_i^2 + V_{\text{ext}}(r_i; \{R_l\}_{l=1}^{n_{\text{nuc}}}) + V_{\text{approx}}(r_i) \right).$$

# Dealing with Electron Correlation — Full Configuration Interaction (FCI)

- Work in space  $\bigwedge^{n_{\text{elec}}} L^2(\mathbb{R}^3, \mathbb{C})$  to ensure anti-symmetry of the wavefunction (Pauli exclusion principle).

# Dealing with Electron Correlation — Full Configuration Interaction (FCI)

- Work in space  $\bigwedge^{n_{\text{elec}}} L^2(\mathbb{R}^3, \mathbb{C})$  to ensure anti-symmetry of the wavefunction (Pauli exclusion principle).
- From a truncated one-electron basis set  $\{\chi_p\}_{p=1}^{n_{\text{orb}}}$ , the basis functions of the many-body space are **Slater determinants** (or, **anti-symmetrized tensor products**):

$$|D_i\rangle = |\chi_{p_1}\chi_{p_2}\cdots\chi_{p_n}\rangle = \chi_{p_1} \wedge \chi_{p_2} \wedge \cdots \wedge \chi_{p_n},$$

with  $n = n_{\text{elec}}$  and  $1 \leq p_1 < p_2 < \cdots < p_n \leq n_{\text{orb}}$ .

# Dealing with Electron Correlation — Full Configuration Interaction (FCI)

- Work in space  $\bigwedge^{n_{\text{elec}}} L^2(\mathbb{R}^3, \mathbb{C})$  to ensure anti-symmetry of the wavefunction (Pauli exclusion principle).
- From a truncated one-electron basis set  $\{\chi_p\}_{p=1}^{n_{\text{orb}}}$ , the basis functions of the many-body space are **Slater determinants** (or, **anti-symmetrized tensor products**):

$$|D_i\rangle = |\chi_{p_1}\chi_{p_2}\cdots\chi_{p_n}\rangle = \chi_{p_1} \wedge \chi_{p_2} \wedge \cdots \wedge \chi_{p_n},$$

with  $n = n_{\text{elec}}$  and  $1 \leq p_1 < p_2 < \cdots < p_n \leq n_{\text{orb}}$ .

- FCI variational space dimension:  $N_{\text{FCI}} = \binom{n_{\text{orb}}}{n_{\text{elec}}}$ .

# Dealing with Electron Correlation — Full Configuration Interaction (FCI)

- Work in space  $\bigwedge^{n_{\text{elec}}} L^2(\mathbb{R}^3, \mathbb{C})$  to ensure anti-symmetry of the wavefunction (Pauli exclusion principle).
- From a truncated one-electron basis set  $\{\chi_p\}_{p=1}^{n_{\text{orb}}}$ , the basis functions of the many-body space are **Slater determinants** (or, **anti-symmetrized tensor products**):

$$|D_i\rangle = |\chi_{p_1}\chi_{p_2}\cdots\chi_{p_n}\rangle = \chi_{p_1} \wedge \chi_{p_2} \wedge \cdots \wedge \chi_{p_n},$$

with  $n = n_{\text{elec}}$  and  $1 \leq p_1 < p_2 < \cdots < p_n \leq n_{\text{orb}}$ .

- FCI variational space dimension:  $N_{\text{FCI}} = \binom{n_{\text{orb}}}{n_{\text{elec}}}$ .
- Schrödinger equation can be transformed to FCI eigenvalue problem

$$H\mathbf{c} = E_0\mathbf{c}, \quad H \in \mathbb{R}^{N_{\text{FCI}} \times N_{\text{FCI}}}, \quad \mathbf{c} \in \mathbb{R}^{N_{\text{FCI}}}.$$

# Exponential Growth of FCI Problem Size

Table: Different Molecule Systems and Storage cost

Molecule	Basis	Electrons	Spin-Orbitals	Dimension	Memory
H <sub>2</sub> O	cc-pVDZ	10	48	$\sim 10^8$	$\sim 1$ GB
N <sub>2</sub>	cc-pVDZ	14	56	$\sim 10^{11}$	$\sim 1$ TB
N <sub>2</sub>	cc-pVTZ	14	120	$\sim 10^{16}$	$\sim 100$ PB
Cr <sub>2</sub>	Ahlrichs	48	84	$\sim 10^{22}$	-

# Observations (Good news!)

- ① We know every entry of the Hamiltonian matrix  $H$  explicitly and the sparsity pattern *a priori*!
- ② Entry:  $H_{ij} = H_{ji} = \langle D_i | \hat{H} | D_j \rangle$ . For off-diagonals  $|D_i\rangle \neq |D_j\rangle$ ,
  - If  $|D_i\rangle = a_r^\dagger a_p |D_j\rangle$ ,  $H_{ij} = \langle r | \hat{H} | p \rangle + \sum_k \langle rk | |pk\rangle$ .
  - If  $|D_i\rangle = a_r^\dagger a_s^\dagger a_p a_q |D_j\rangle$ ,  $H_{ij} = \langle rs | |pq\rangle$ .
  - Otherwise,  $H_{ij} = 0$ .

Consequence:  $H$  has  $O(n_{\text{elec}}^2 n_{\text{orb}}^2)$  entries per row.
- ③ Eigenvector  $\mathbf{c}$  of a few low-lying eigenvalues usually sparse (in the sense of truncation).

# Observations (Good news!)

- ① We know every entry of the Hamiltonian matrix  $H$  explicitly and the sparsity pattern *a priori*!
- ② Entry:  $H_{ij} = H_{ji} = \langle D_i | \hat{H} | D_j \rangle$ . For off-diagonals  $|D_i\rangle \neq |D_j\rangle$ ,
  - If  $|D_i\rangle = a_r^\dagger a_p |D_j\rangle$ ,  $H_{ij} = \langle r | \hat{H} | p \rangle + \sum_k \langle rk | |pk \rangle$ .
  - If  $|D_i\rangle = a_r^\dagger a_s^\dagger a_p a_q |D_j\rangle$ ,  $H_{ij} = \langle rs | |pq \rangle$ .
  - Otherwise,  $H_{ij} = 0$ .

Consequence:  $H$  has  $O(n_{\text{elec}}^2 n_{\text{orb}}^2)$  entries per row.

- ③ Eigenvector  $\mathbf{c}$  of a few low-lying eigenvalues usually sparse (in the sense of truncation).

Idea: we can select an index set of “important” coefficients  $\mathcal{I}$  and compute the lowest eigenvalues of  $H_{\mathcal{I},\mathcal{I}}$ . (**variational methods**:  $\mathcal{I}$  is the variational space)

# Observations (Good news!)

- ① We know every entry of the Hamiltonian matrix  $H$  explicitly and the sparsity pattern *a priori*!
- ② Entry:  $H_{ij} = H_{ji} = \langle D_i | \hat{H} | D_j \rangle$ . For off-diagonals  $|D_i\rangle \neq |D_j\rangle$ ,
  - If  $|D_i\rangle = a_r^\dagger a_p |D_j\rangle$ ,  $H_{ij} = \langle r | \hat{H} | p \rangle + \sum_k \langle rk | |pk\rangle$ .
  - If  $|D_i\rangle = a_r^\dagger a_s^\dagger a_p a_q |D_j\rangle$ ,  $H_{ij} = \langle rs | |pq\rangle$ .
  - Otherwise,  $H_{ij} = 0$ .

Consequence:  $H$  has  $O(n_{\text{elec}}^2 n_{\text{orb}}^2)$  entries per row.

- ③ Eigenvector  $\mathbf{c}$  of a few low-lying eigenvalues usually sparse (in the sense of truncation).

Idea: we can select an index set of “important” coefficients  $\mathcal{I}$  and compute the lowest eigenvalues of  $H_{\mathcal{I},\mathcal{I}}$ . (**variational methods**:  $\mathcal{I}$  is the variational space)

Question: How to select  $\mathcal{I}$ ?

# Existing Methods

- **Selected CI:** Pick the important configurations based on perturbation theory estimates.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad H_0 = \begin{bmatrix} H_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \quad \mathbf{c}^{(0)} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{0} \end{bmatrix}.$$

$$\mathbf{c}^{(1)} = \begin{bmatrix} \mathbf{0} \\ -(D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0 \end{bmatrix}, \quad E^{(2)} = -\mathbf{v}_0^T H_{12} (D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0.$$

# Existing Methods

- **Selected CI:** Pick the important configurations based on perturbation theory estimates.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad H_0 = \begin{bmatrix} H_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \quad \mathbf{c}^{(0)} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{0} \end{bmatrix}.$$

$$\mathbf{c}^{(1)} = \begin{bmatrix} \mathbf{0} \\ -(D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0 \end{bmatrix}, \quad E^{(2)} = -\mathbf{v}_0^T H_{12} (D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0.$$

- 1 From selected configurations, generate new ones by single/double excitations.

# Existing Methods

- **Selected CI:** Pick the important configurations based on perturbation theory estimates.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad H_0 = \begin{bmatrix} H_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \quad \mathbf{c}^{(0)} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{0} \end{bmatrix}.$$

$$\mathbf{c}^{(1)} = \begin{bmatrix} \mathbf{0} \\ -(D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0 \end{bmatrix}, \quad E^{(2)} = -\mathbf{v}_0^T H_{12} (D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0.$$

- 1 From selected configurations, generate new ones by single/double excitations.
- 2 Sort by perturbation estimates, and select the top ones.

# Existing Methods

- **Selected CI:** Pick the important configurations based on perturbation theory estimates.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad H_0 = \begin{bmatrix} H_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \quad \mathbf{c}^{(0)} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{0} \end{bmatrix}.$$

$$\mathbf{c}^{(1)} = \begin{bmatrix} \mathbf{0} \\ -(D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0 \end{bmatrix}, \quad E^{(2)} = -\mathbf{v}_0^T H_{12} (D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0.$$

- 1 From selected configurations, generate new ones by single/double excitations.
- 2 Sort by perturbation estimates, and select the top ones.
- 3 Diagonalize the Hamiltonian in the selected space.

# Existing Methods

- **Selected CI:** Pick the important configurations based on perturbation theory estimates.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad H_0 = \begin{bmatrix} H_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \quad \mathbf{c}^{(0)} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{0} \end{bmatrix}.$$

$$\mathbf{c}^{(1)} = \begin{bmatrix} \mathbf{0} \\ -(D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0 \end{bmatrix}, \quad E^{(2)} = -\mathbf{v}_0^T H_{12} (D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0.$$

- 1 From selected configurations, generate new ones by single/double excitations.
- 2 Sort by perturbation estimates, and select the top ones.
- 3 Diagonalize the Hamiltonian in the selected space.
- 4 Repeat until convergence/out of memory.

# Existing Methods

- **Selected CI**: Pick the important configurations based on perturbation theory estimates.

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad H_0 = \begin{bmatrix} H_{11} & 0 \\ 0 & D_{22} \end{bmatrix}, \quad \mathbf{c}^{(0)} = \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{0} \end{bmatrix}.$$

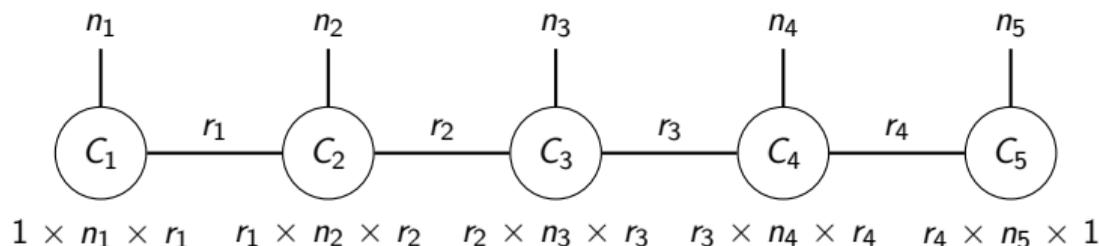
$$\mathbf{c}^{(1)} = \begin{bmatrix} \mathbf{0} \\ -(D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0 \end{bmatrix}, \quad E^{(2)} = -\mathbf{v}_0^T H_{12} (D_{22} - E_0 I)^{-1} H_{21} \mathbf{v}_0.$$

- 1 From selected configurations, generate new ones by single/double excitations.
  - 2 Sort by perturbation estimates, and select the top ones.
  - 3 Diagonalize the Hamiltonian in the selected space.
  - 4 Repeat until convergence/out of memory.
- ... and other FCI methods such as **DMRG** or **FCIQMC**.

## Other Approximations to Wavefunctions

- **DMRG**: Inspired by the success in one-dimensional quantum systems, uses a variational ansatz called Matrix Product State (MPS) to represent the wavefunction efficiently.

$$c(\mu_1, \dots, \mu_L) = \sum_{r_1, \dots, r_{L-1}} \prod_{k=1}^n C_k(r_{k-1}, \mu_k, r_k), \quad r_0 = r_L = 1.$$



- **FCIQMC**: A stochastic method that simulates the imaginary-time evolution of the wavefunction using a population of walkers (spawning, death/cloning, and annihilation).

$$\mathbf{c}(\tau + \Delta\tau) = \mathbf{c}(\tau) - \Delta\tau(H - SI)\mathbf{c}(\tau).$$

# Coordinate Descent FCI (CDFCI)<sup>1</sup>

**Coordinate gradient descent method + new way of formalizing the problem.**

Consider the unconstrained minimization problem

$$\min_{\mathbf{c} \in \mathbb{R}^{N_{\text{FCI}}}} f(\mathbf{c}) = \min_{\mathbf{c}} \|H + \mathbf{c}\mathbf{c}^T\|_F^2.$$

- Gradient  $\nabla f(\mathbf{c}) = 4H\mathbf{c} + 4(\mathbf{c}^T\mathbf{c})\mathbf{c}$ .
- Hessian  $\nabla^2 f(\mathbf{c}) = 4H + 8\mathbf{c}\mathbf{c}^T + 4(\mathbf{c}^T\mathbf{c})I$ .
- Stationary points: 0 and  $\pm\sqrt{-\lambda_k}\mathbf{v}_k$  for any  $\lambda_k < 0$ .
- If  $\lambda_1 < \lambda_2$  and  $\lambda_1 < 0$ , only two **local minimizers**  $\pm\sqrt{-\lambda_1}\mathbf{v}_1$ , which are also **global minimizers**, and strongly convex in a neighborhood!

<sup>1</sup>Z. Wang, Y. Li, J. Lu, J. Chem. Theory Comput., 2019.

# Why Coordinate Gradient Descent?

$$\mathbf{c} \leftarrow \mathbf{c} + \alpha_i \mathbf{e}_i, \quad i = 1, 2, \dots, N_{\text{FCI}}$$

- Suitable for large-scale problems with sparse solutions<sup>2</sup>.
- Linear convergence guaranteed for strongly convex problems with Lipschitz continuous gradients<sup>3</sup>.
- Selecting the coordinate with the largest gradient magnitude (Gauss–Southwell rule) can lead to faster convergence<sup>4</sup>.
- Automatically selects the most “important” coordinates into the variational space and ignores the unimportant ones throughout the iterations.

---

<sup>2</sup>S. Wright, *Mathematical Programming*, 2015.

<sup>3</sup>Y. Nesterov, *SIAM J. Optim.*, 2012.

<sup>4</sup>J. Nutini, M. Schmidt, I. H. Laradji, et al., *ICML*, 2015.

# CDFCI Framework

Initialize  $\mathbf{c}^{(0)}$ ,  $\mathbf{b}^{(0)} = H\mathbf{c}^{(0)}$ .

For iteration  $\ell = 1, 2, \dots$

- 1 Select coordinate  
 $i^{(\ell)} = \arg \max_i |\nabla_i f(\mathbf{c}^{(\ell-1)})|$ .
- 2 Find stepsize by exact line search  
 $\alpha^{(\ell)} = \arg \min_{\alpha} f(\mathbf{c}^{(\ell-1)} + \alpha \mathbf{e}_{i^{(\ell)}})$ .
- 3 Update  $\mathbf{c}^{(\ell)} = \mathbf{c}^{(\ell-1)} + \alpha^{(\ell)} \mathbf{e}_{i^{(\ell)}}$ ,  
 $\mathbf{b}^{(\ell)} = \mathbf{b}^{(\ell-1)} + \alpha^{(\ell)} H_{:,i^{(\ell)}}$ .

Remark: Gradient

$$\nabla f(\mathbf{c}) = 4H\mathbf{c} + 4\mathbf{c}^T \mathbf{c} \mathbf{c} = 4\mathbf{b} + 4\mathbf{c}^T \mathbf{c} \mathbf{c}.$$

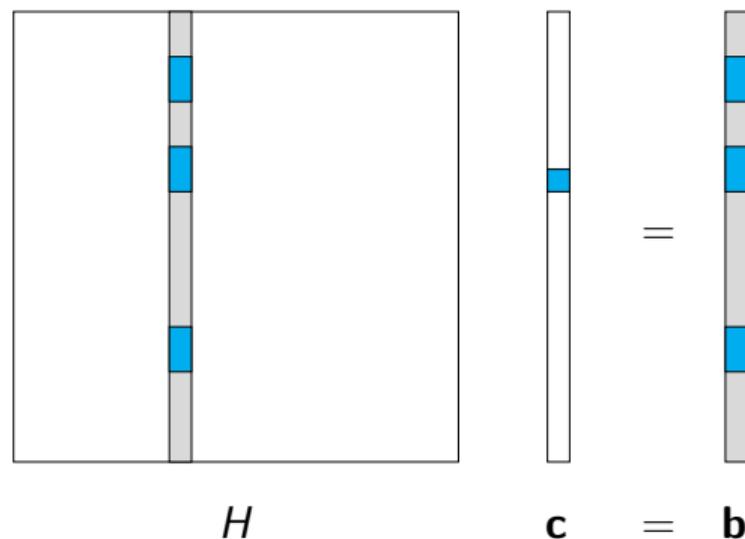


Figure: Update for one coordinate.

# CDFCI Framework - for Two Coordinates?

Initialize  $\mathbf{c}^{(0)}$ ,  $\mathbf{b}^{(0)} = H\mathbf{c}^{(0)}$ .

For iteration  $\ell = 1, 2, \dots$

- 1 Select coordinate

$$i^{(\ell)} = \arg \max_i |\nabla_i f(\mathbf{c}^{(\ell-1)})|,$$

$$j^{(\ell)} = \arg \max_{j \neq i^{(\ell)}} |\nabla_j f(\mathbf{c}^{(\ell-1)})|.$$

- 2 Find stepsize

$$\alpha^{(\ell)} = \arg \min_{\alpha} f(\mathbf{c}^{(\ell-1)} + \alpha \mathbf{e}_{i^{(\ell)}}),$$

$$\beta^{(\ell)} = \arg \min_{\beta} f(\mathbf{c}^{(\ell-1)} + \beta \mathbf{e}_{j^{(\ell)}}).$$

- 3 Update

$$\mathbf{c}^{(\ell)} = \mathbf{c}^{(\ell-1)} + \alpha^{(\ell)} \mathbf{e}_{i^{(\ell)}} + \beta^{(\ell)} \mathbf{e}_{j^{(\ell)}},$$

$$\mathbf{b}^{(\ell)} = \mathbf{b}^{(\ell-1)} + \alpha^{(\ell)} H_{:,i^{(\ell)}} + \beta^{(\ell)} H_{:,j^{(\ell)}}.$$

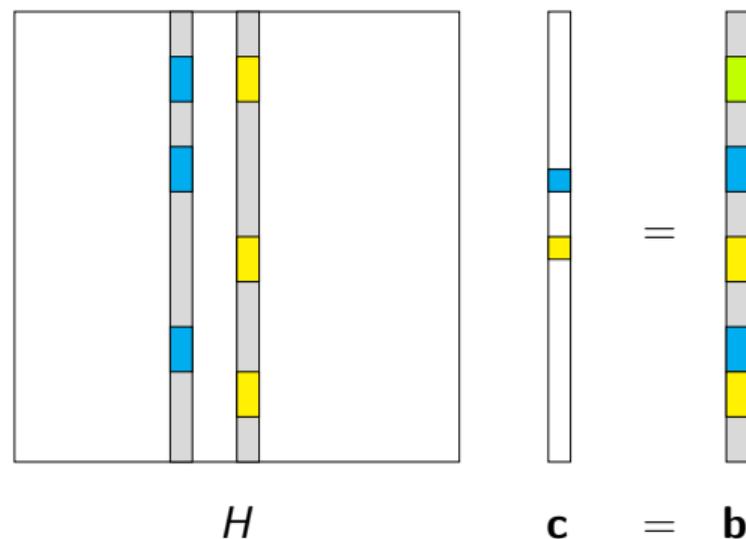


Figure: Update for two coordinates.

# CDFCI Framework - Exact Line Search?

Initialize  $\mathbf{c}^{(0)}$ ,  $\mathbf{b}^{(0)} = H\mathbf{c}^{(0)}$ .

For iteration  $\ell = 1, 2, \dots$

- 1 Select coordinate

$$i^{(\ell)} = \arg \max_i |\nabla_i f(\mathbf{c}^{(\ell-1)})|,$$

$$j^{(\ell)} = \arg \max_{j \neq i^{(\ell)}} |\nabla_j f(\mathbf{c}^{(\ell-1)})|.$$

- 2 Find stepsize  $\alpha^{(\ell)}, \beta^{(\ell)} =$

$$\arg \min_{\alpha, \beta} f(\mathbf{c}^{(\ell-1)} + \alpha \mathbf{e}_{i^{(\ell)}} + \beta \mathbf{e}_{j^{(\ell)}}).$$

- 3 Update

$$\mathbf{c}^{(\ell)} = \mathbf{c}^{(\ell-1)} + \alpha^{(\ell)} \mathbf{e}_{i^{(\ell)}} + \beta^{(\ell)} \mathbf{e}_{j^{(\ell)}},$$

$$\mathbf{b}^{(\ell)} = \mathbf{b}^{(\ell-1)} + \alpha^{(\ell)} H_{:,i^{(\ell)}} + \beta^{(\ell)} H_{:,j^{(\ell)}}.$$

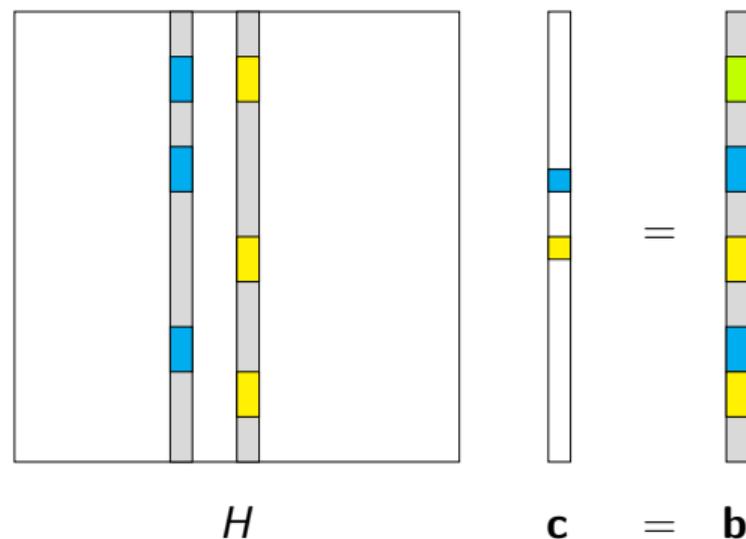


Figure: Update for two coordinates.

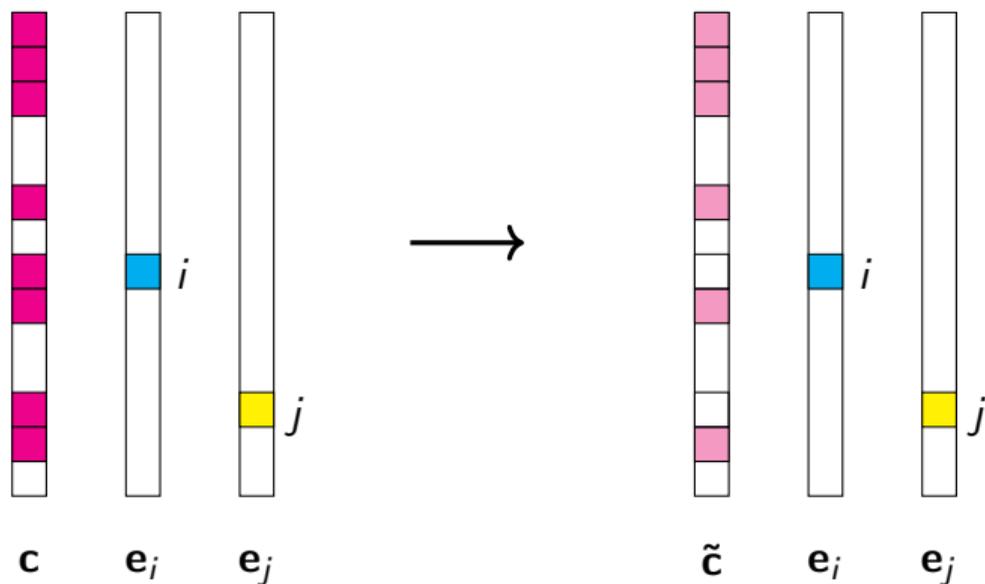
# Add a Scalar $\gamma$ for Exact Line Search

Modify the minimization problem from  $\min_{\alpha, \beta} f(\mathbf{c} + \alpha \mathbf{e}_i + \beta \mathbf{e}_j)$  to

$$\begin{aligned} \min_{\gamma, \alpha, \beta} f(\gamma \mathbf{c} + \alpha \mathbf{e}_i + \beta \mathbf{e}_j) &= f\left(\begin{bmatrix} \mathbf{c} & \mathbf{e}_i & \mathbf{e}_j \end{bmatrix} \begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix}\right) \\ &= \left\| H + \begin{bmatrix} \mathbf{c} & \mathbf{e}_i & \mathbf{e}_j \end{bmatrix} \begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix} \begin{bmatrix} \gamma & \alpha & \beta \end{bmatrix} \begin{bmatrix} \mathbf{c}^T \\ \mathbf{e}_i^T \\ \mathbf{e}_j^T \end{bmatrix} \right\|_F^2. \end{aligned}$$

# Matrix Orthogonalization

Construct  $[\tilde{\mathbf{c}} \ \mathbf{e}_i \ \mathbf{e}_j]$ , where  $\|\tilde{\mathbf{c}}\|_2 = 1$ ,  $(\tilde{\mathbf{c}}, \mathbf{e}_i) = 0$ ,  $(\tilde{\mathbf{c}}, \mathbf{e}_j) = 0$ .



# Add $\gamma$ and $\tilde{\mathbf{c}}$ for Exact Line Search

Modify the minimization problem from  $\min_{\alpha, \beta} f(\mathbf{c} + \alpha \mathbf{e}_i + \beta \mathbf{e}_j)$  to

$$\begin{aligned}
 \min_{\gamma, \alpha, \beta} f(\gamma \tilde{\mathbf{c}} + \alpha \mathbf{e}_i + \beta \mathbf{e}_j) &= f\left(\begin{bmatrix} \tilde{\mathbf{c}} & \mathbf{e}_i & \mathbf{e}_j \end{bmatrix} \begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix}\right) \\
 &= \left\| \begin{bmatrix} H + \begin{bmatrix} \tilde{\mathbf{c}} & \mathbf{e}_i & \mathbf{e}_j \end{bmatrix} \begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{c}}^T \\ \mathbf{e}_i^T \\ \mathbf{e}_j^T \end{bmatrix} \end{bmatrix} \right\|_F^2 \\
 &= \underbrace{\left\| \begin{bmatrix} \tilde{\mathbf{c}}^T \\ \mathbf{e}_i^T \\ \mathbf{e}_j^T \end{bmatrix} H \begin{bmatrix} \tilde{\mathbf{c}} & \mathbf{e}_i & \mathbf{e}_j \end{bmatrix} + \begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix} \begin{bmatrix} \gamma & \alpha & \beta \end{bmatrix} \right\|_F^2}_{\in \mathbb{R}^{3 \times 3}} + \text{constant}.
 \end{aligned}$$

# Block Coordinate Extension: Multi-Coordinate Descent FCI (mCDFCI)

- Select a set of coordinates  $I = \{i_1, \dots, i_k\}$ ,  $1 \leq i_j \leq N_{\text{FCI}}$  based on gradient  $\nabla f(\mathbf{c}) = 4H\mathbf{c} + 4\mathbf{c}^T\mathbf{c}\mathbf{c}$ .
- Denote  $\mathcal{E}_I = [\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k}] \in \mathbb{R}^{N_{\text{FCI}} \times k}$ .
- The update is given by

$$\mathbf{c} \leftarrow \gamma\mathbf{c} + \mathcal{E}_I\mathbf{a}.$$

- The values of  $\gamma$  and  $\mathbf{a}$  are given by the eigenvector of

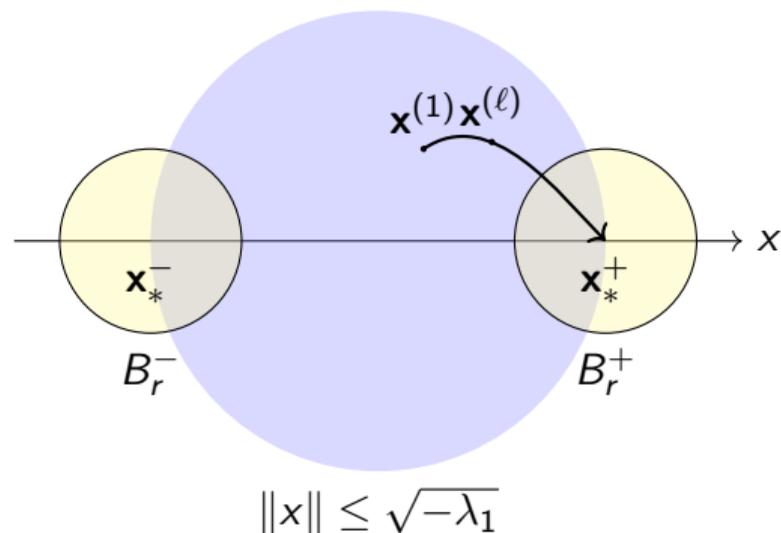
$$Q^T H Q := \begin{bmatrix} \tilde{\mathbf{c}}^T \\ \mathcal{E}_I^T \end{bmatrix} H \begin{bmatrix} \tilde{\mathbf{c}} & \mathcal{E}_I \end{bmatrix} \in \mathbb{R}^{(k+1) \times (k+1)}$$

corresponding to the minimal eigenvalue  $\lambda_{\min}$ , which is the current energy estimate.

**Subspace Methods with Rayleigh–Ritz projection!**

# Local Linear Convergence

- First-order methods almost surely avoid strict saddle points<sup>5</sup>.
- The iterates converge to the global minimizer  $\pm\sqrt{-\lambda_1}\mathbf{v}_1$  with a linear convergence rate in a neighborhood.
- The Rayleigh quotient estimator  $RQ(\mathbf{c}) = \frac{\mathbf{c}^\top H \mathbf{c}}{\mathbf{c}^\top \mathbf{c}}$  converges to the ground state energy  $E_0$ .



<sup>5</sup>J. D. Lee, I. Panageas, G. Piliouras, et al., *Mathematical Programming*, 2019.

# Compression Strategy

- Compression occurs when updating  $\mathbf{b} = H\mathbf{c}$ :

$$\mathbf{b} \leftarrow \gamma \mathbf{b} + H\mathcal{E}_I \mathbf{a},$$

$$b_i \leftarrow \gamma b_i + \sum_{j \in I} H_{ij} a_j.$$

- Update  $H_{ij} a_j$  is discarded if  $\mathbf{c}_i = 0$  and  $|H_{ij} a_j| < \tau$ .
- Not affecting the Rayleigh quotient estimator  $\text{RQ}(\mathbf{c}) = \frac{\mathbf{c}^T H \mathbf{c}}{\mathbf{c}^T \mathbf{c}} = \frac{\mathbf{c}^T \mathbf{b}}{\mathbf{c}^T \mathbf{c}}$ .
- Tolerance  $\tau$  balances between memory-cost and accuracy.

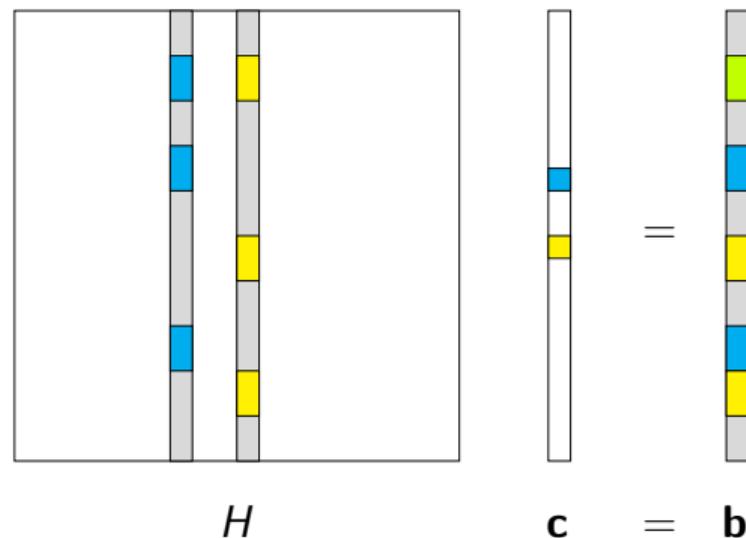


Figure: Update for two coordinates.

# Parallel Implementation Details

- Shared memory parallelism based on OpenMP.
- Sparse vectors  $\mathbf{c}$  and  $\mathbf{b} = H\mathbf{c}$  are stored in a concurrent hash table<sup>6</sup> that supports parallel read/write.
- Two levels of parallelism: row-wise and column-wise.

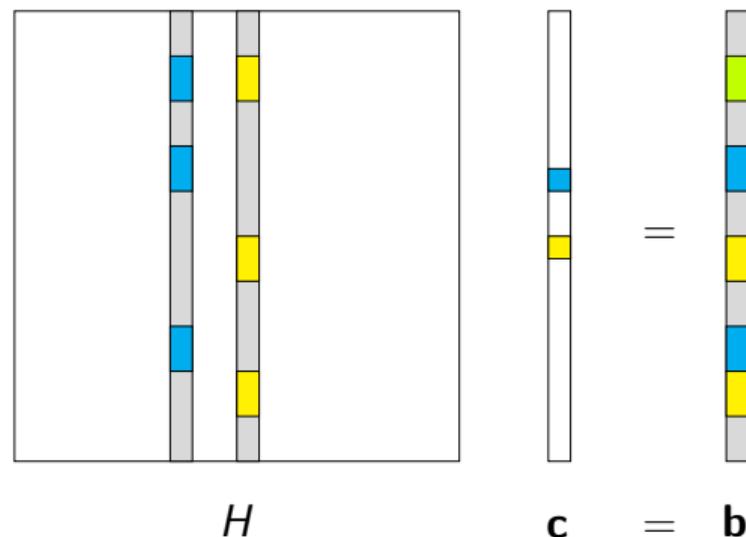
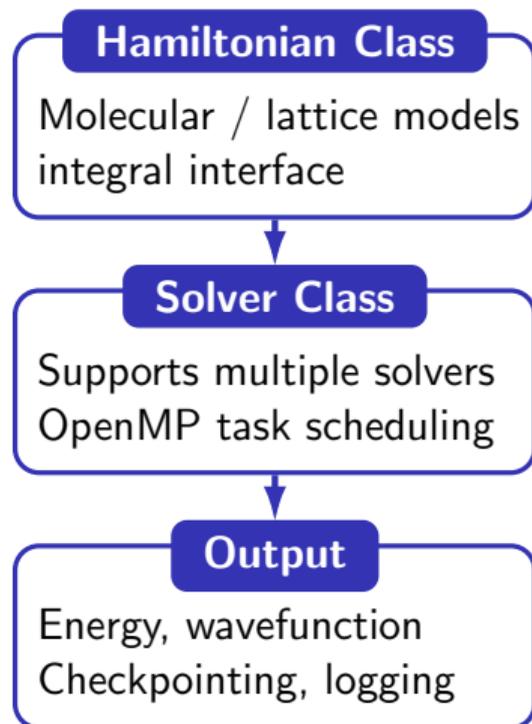


Figure: Update for two coordinates.

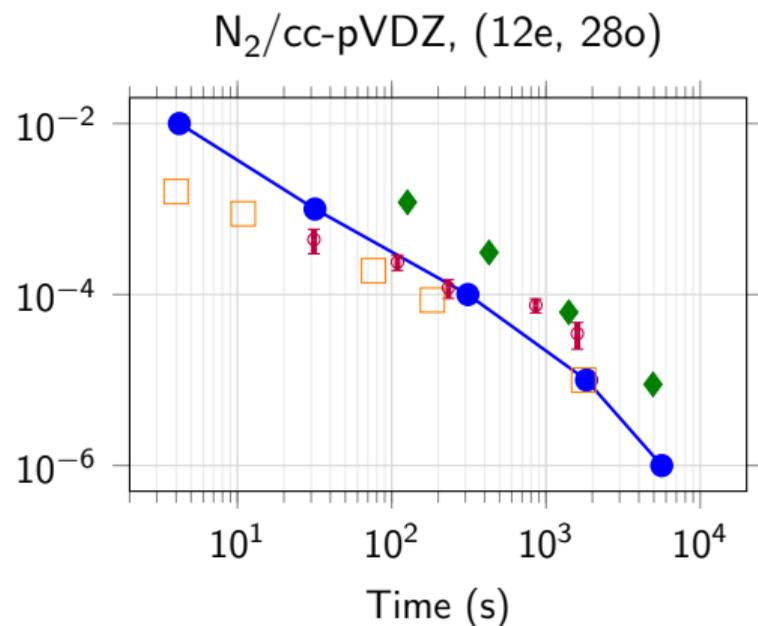
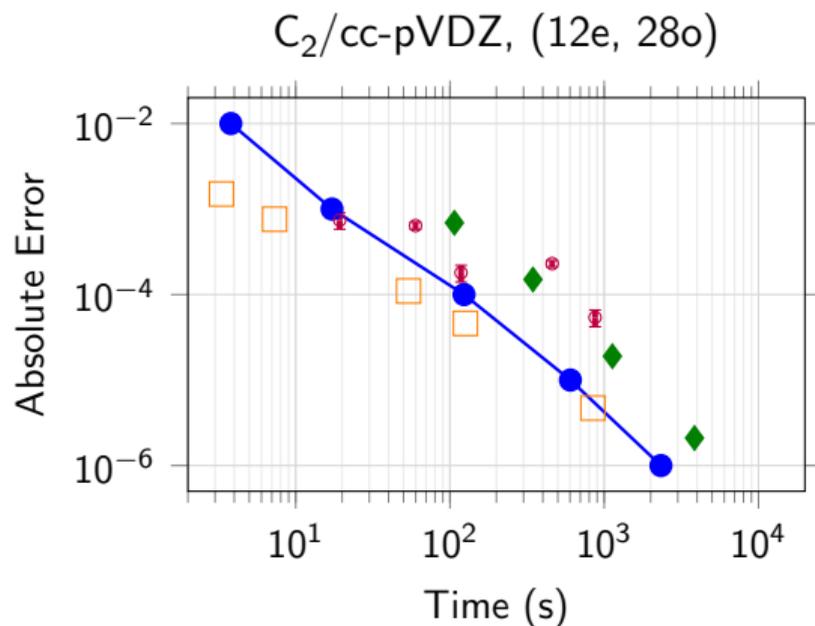
<sup>6</sup><https://github.com/efficient/libcuckoo>

# CDFCI Software Package

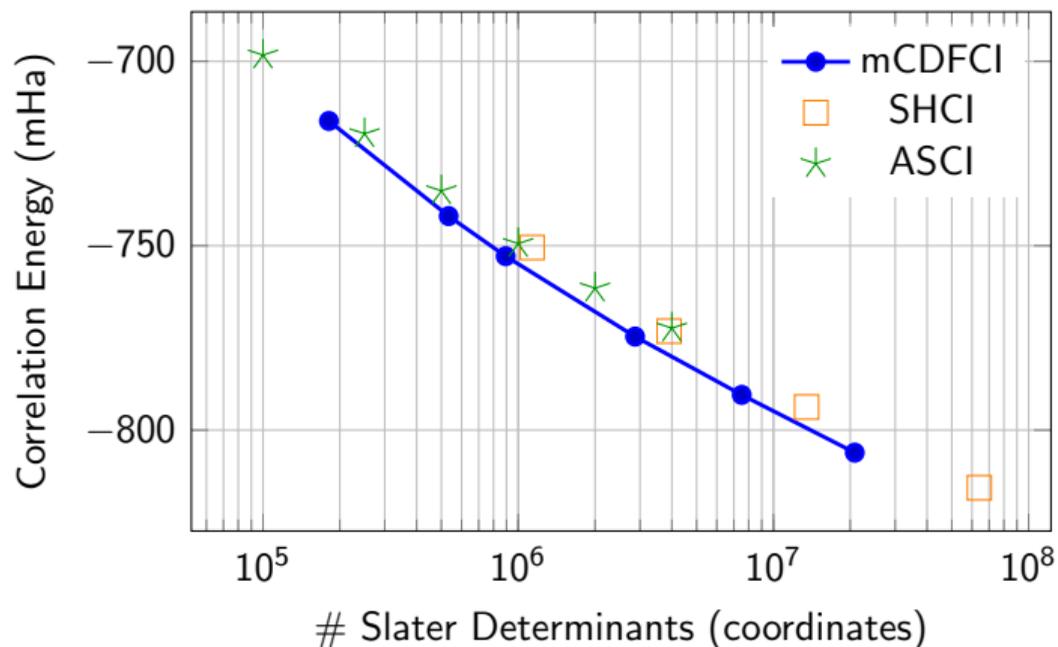
- Written in **C++17**, modular design with **OpenMP** parallelization
- Scalable to determinant spaces with **tens of millions of basis functions**
- Dynamic task scheduling and load balancing with a multi-solver framework
- Supports excited states calculations (xCDFCI) and single-particle orbitals compression (OptOrbFCI)
- **Open-source software**: reproducible code and reliable results



## CDFCI vs other FCI methods



# CDFCI vs selected CI methods



**Figure:** Correlation energy vs. number of Slater determinants (coordinates) for the benzene benchmark. CDFCI results are compared with reference data from SHCI and ASCI.

## mCDFCI vs CDFCI

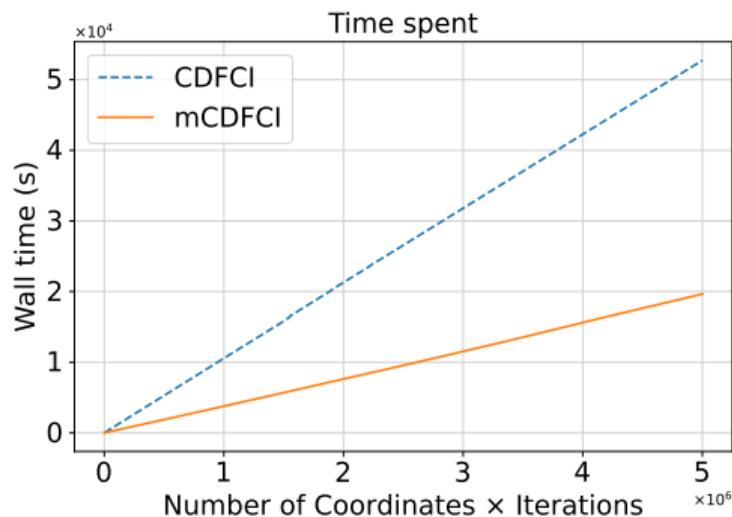
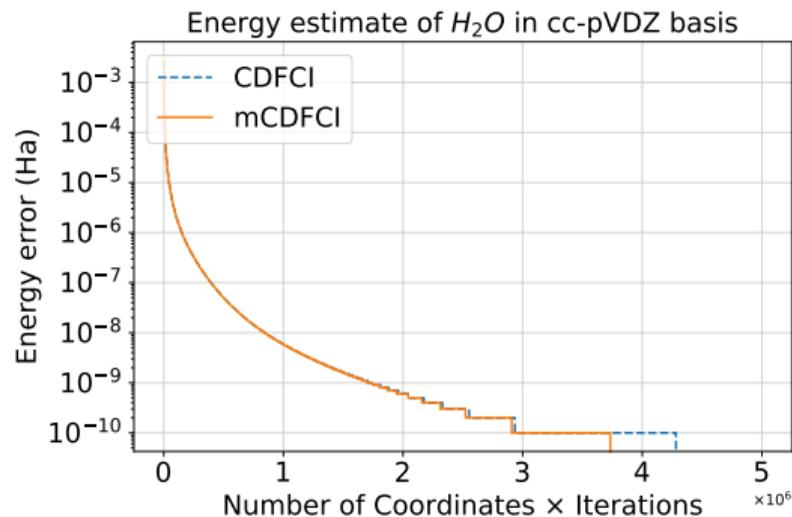
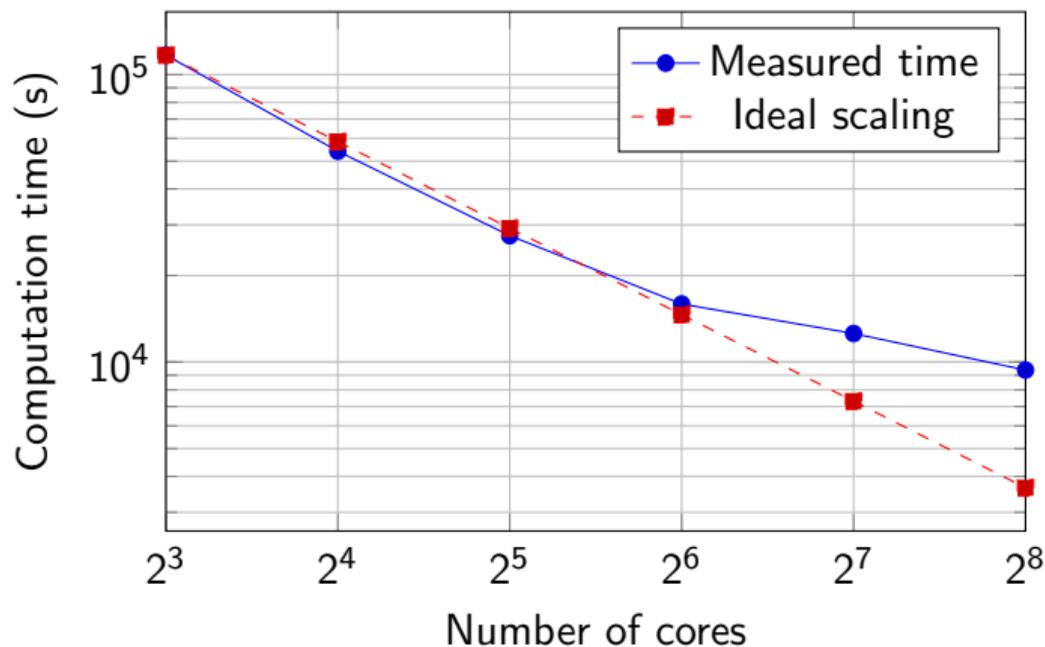


Figure: Speedup of mCDFCI compared with CDFCI, both in 64 threads and  $k = 64$ .

# Parallel Scaling



**Figure:** Strong scaling of CDFCI for the  $N_2/cc\text{-pVQZ}$  system. Experiments were conducted on a system equipped with two AMD EPYC 9754 processors (128 cores each) and 1.5 TB of memory.

# Open Questions

- Coordinate gradient descent only converges linearly — how can we achieve accelerations without destroying the sparsity structure?
- Possible directions: Nesterov's momentum, conjugate gradient, Krylov subspace structure, preconditioners, etc.
- How to implement the method in a distributed memory setting?
- How to extend the framework to bosonic systems and relativistic Hamiltonians?

# Thanks for Your Attention!

I'd love to hear any questions or suggestions!

This is joint work with Zhe Wang, Jianfeng Lu and Yingzhou Li.

Contact: [yuejiazhang21@m.fudan.edu.cn](mailto:yuejiazhang21@m.fudan.edu.cn)

Related work to this talk:

- **CDFCI Core Algorithm:** Z. Wang, Y. Li, J. Lu, JCTC, 15(6), 2019.
- **The Parallel Version:** Y. Zhang, W. Gao, Y. Li, JCTC, 21(5), 2025.
- **Other methods implemented in the software package:**
  - Y. Li, J. Lu, JCTC, 16(10), 2020.
  - Z. Wang, Z. Zhang, J. Lu, Y. Li, JCTC, 19(21), 2023.