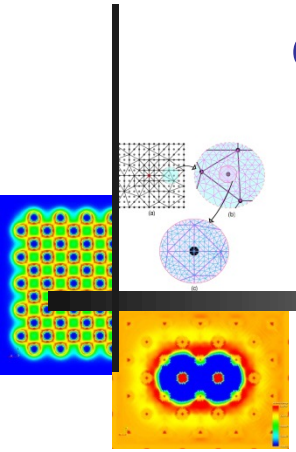
Large-scale DFT simulations commensurate with quantum accuracy: our experience on the exascale Frontier machine

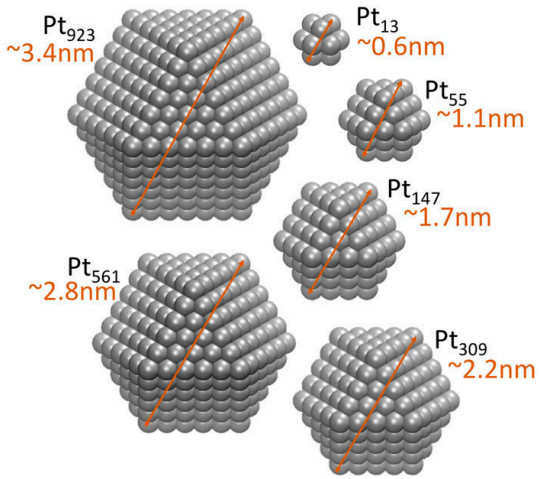


Speakers: Sambit Das (U. Mich.); Bikash Kanungo (U. Mich.);
David Rogers (ORNL);

Joint work with: *Vikram Gavini (U. Mich.); Vishal Subramanian (U. Mich.); Gourab Panigrahi (IISc); Phani Motamarri (IISc); Paul Zimmerman (U. Mich)*

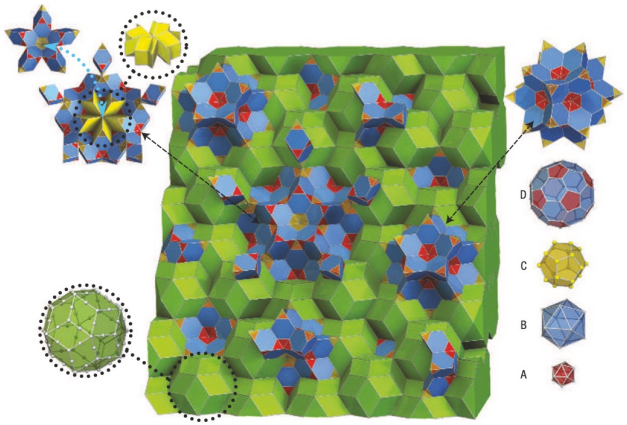
Acknowledgements: DoE-BES, TRI, AFOSR, NERSC, OLCF

Need for accurate large-scale electronic structure calculations



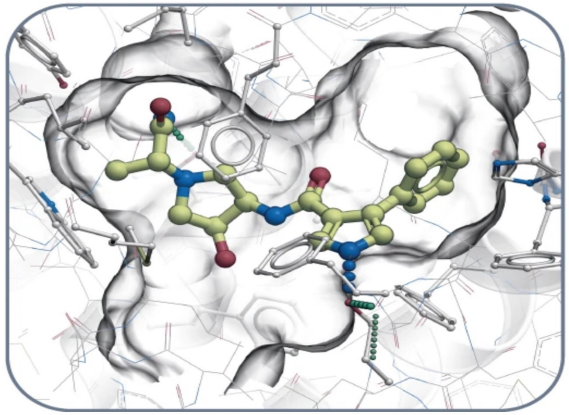
Aarons et al. JCP (2016)

Chemical properties of nanoparticles



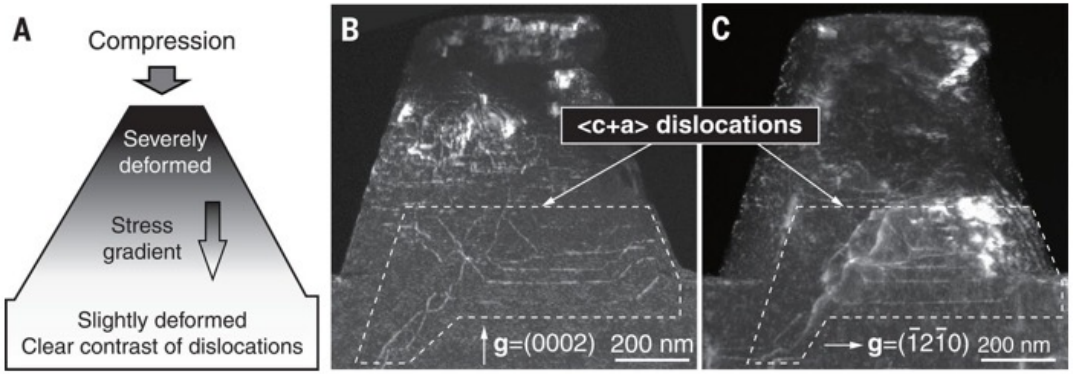
icosahedral-QC YbCd: Takakura et al. Nat. Mat. (2007)

Quasicrystals



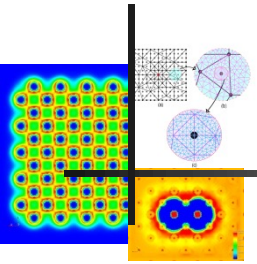
Sadybekov et al. Nature (2023)

Drug design



<c+a> dislocations in submicrometer-size magnesium samples: Liu et al. Sci. (2019)

Defects in Materials



Quantum Mechanics

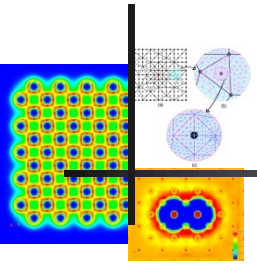
- Schrödinger equation - $H\psi = E\psi$

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{A=1}^{M}\sum_{B=1,B>A}^{M}\frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}$$

$$\psi = \psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N, \mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_M)$$

- Computational complexity - $\psi \in \mathbf{R}^{3N}$!!

- Low rank approximation of many-body wavefunction
  - Configuration interactions: O(10) e-; Accuracy (GS) ~ 1 mHa/atom (Chemical/Quantum Accuracy)
  - Coupled Cluster; MultiConfiguration
  - Quantum Monte Carlo: O(10^3) e-; Accuracy (GS) ~ 5 mHa/atom



Density-functional theory – Kohn-Sham approach

- Ground-state properties are a functional of electron-density !! (Kohn & Sham, 1964-65)

- **Kohn-Sham eigenvalue problem**:

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}[\rho;\mathbf{R}] \right)\psi_i = \epsilon_i \psi_i$$

$$\Downarrow$$

$$V_{\text{eff}}[\rho;\mathbf{R}] = V_{\text{ext}}(\mathbf{R}) + V_H(\rho) + V_{xc}(\rho)$$

$$\Uparrow$$

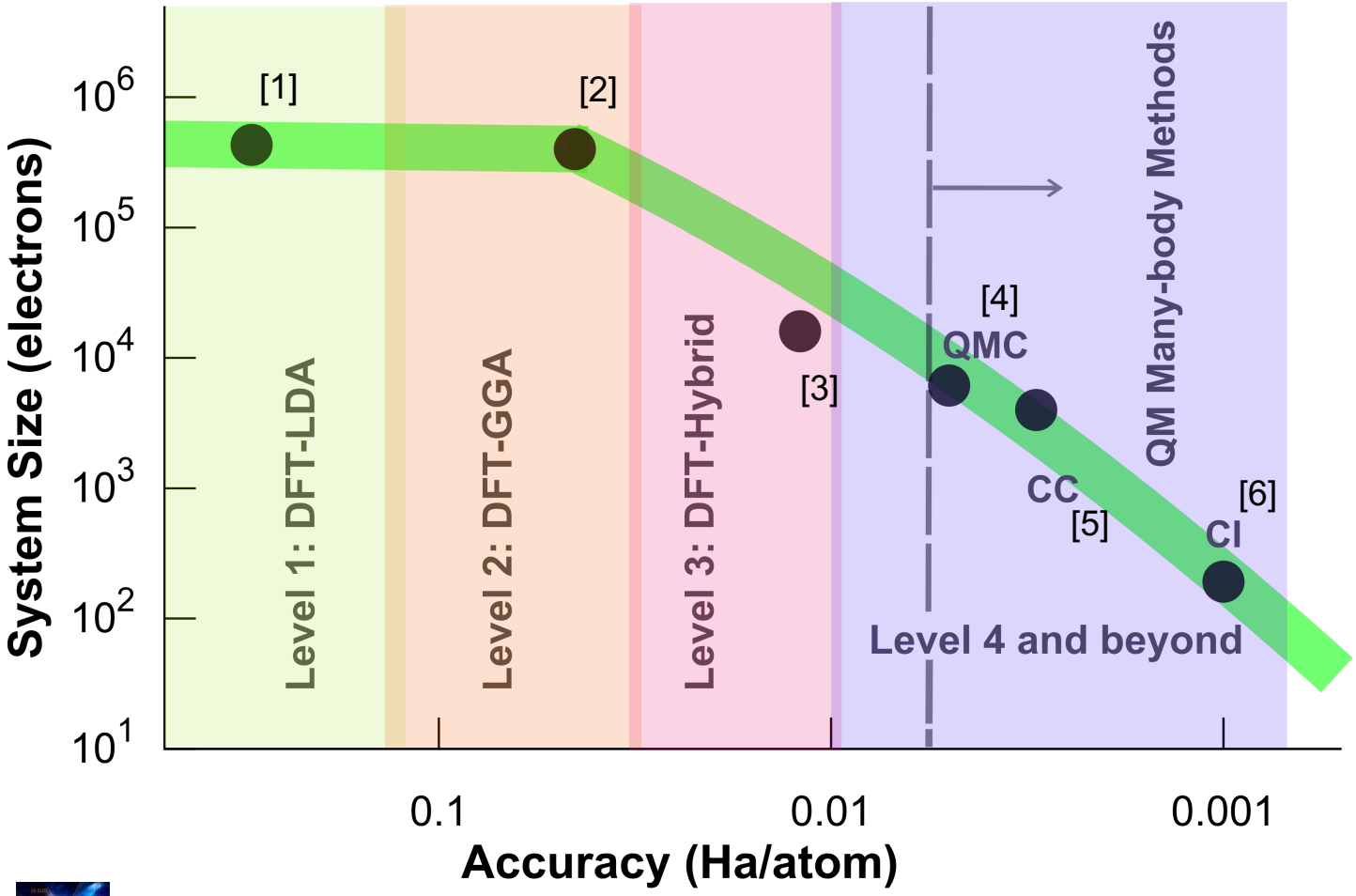
$$\rho(\mathbf{r}) = 2\sum_i f_i |\psi_i(\mathbf{r})|^2$$

Self Consistent Field (SCF) iteration
(Kohn-Sham map)

- $V_{xc} = \dfrac{\delta E_{xc}(\rho)}{\delta \rho}$; $V_{xc}(\rho)\,, E_{xc}(\rho)$ ⟶ Universal functionals

- Computational Complexity O(N^3)
  - O(10^4) e-; Accuracy (GS) ~ 10-100 mHa/atom

Accuracy/Length-scale trade off



[1] Hasegawa et al., SC11 (2011).

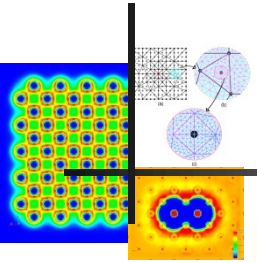
[2] Dogan et al., Phys. Rev. Mat. (2023).

[3] Hu et al. J. Chem Theory Comput. (2017)

[4]*https://ecpannualmeeting.com/assets/overview/posters/*

[5] Nagy, J. Chem Theory Comput. (2019)

[6] Rask & Zimmerman, J. Phys. Chem. A (2021)



Overview of the approach

QM Many Body Physics

$$\hat{\mathcal{H}}_{\mathrm{QMB}}\boldsymbol{\Psi}_{\mathrm{QMB}} = E_0\boldsymbol{\Psi}_{\mathrm{QMB}}$$

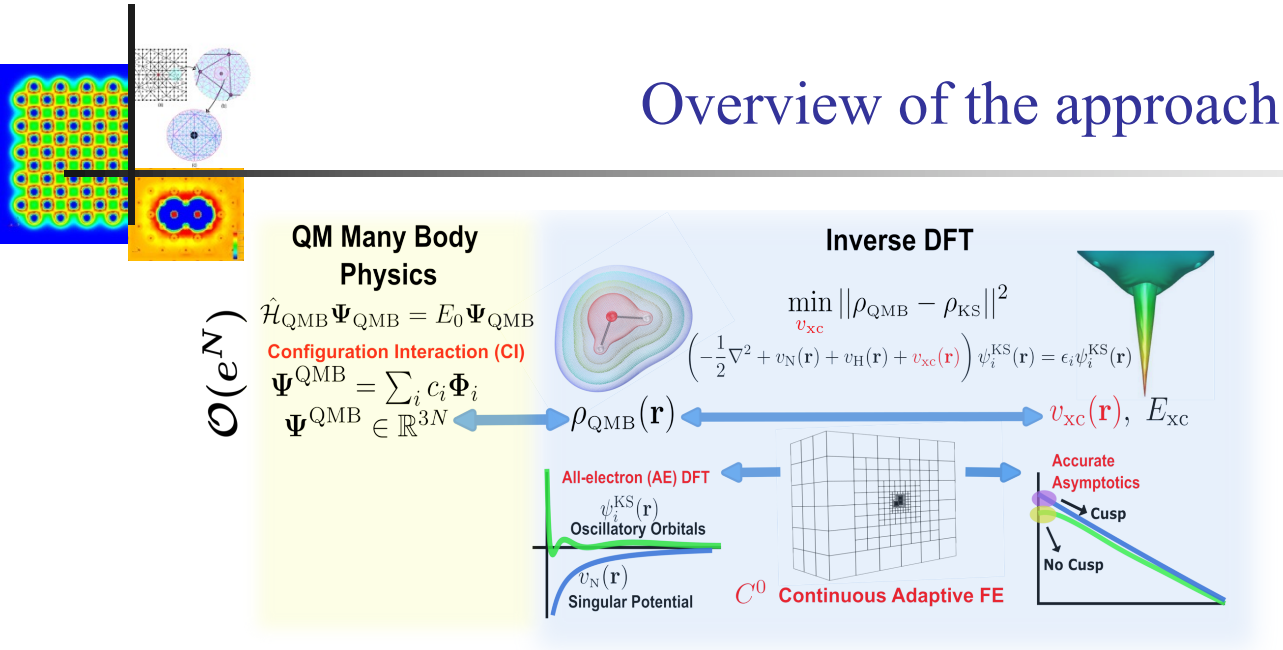
$\mathcal{O}(e^N)$

Configuration Interaction (CI)

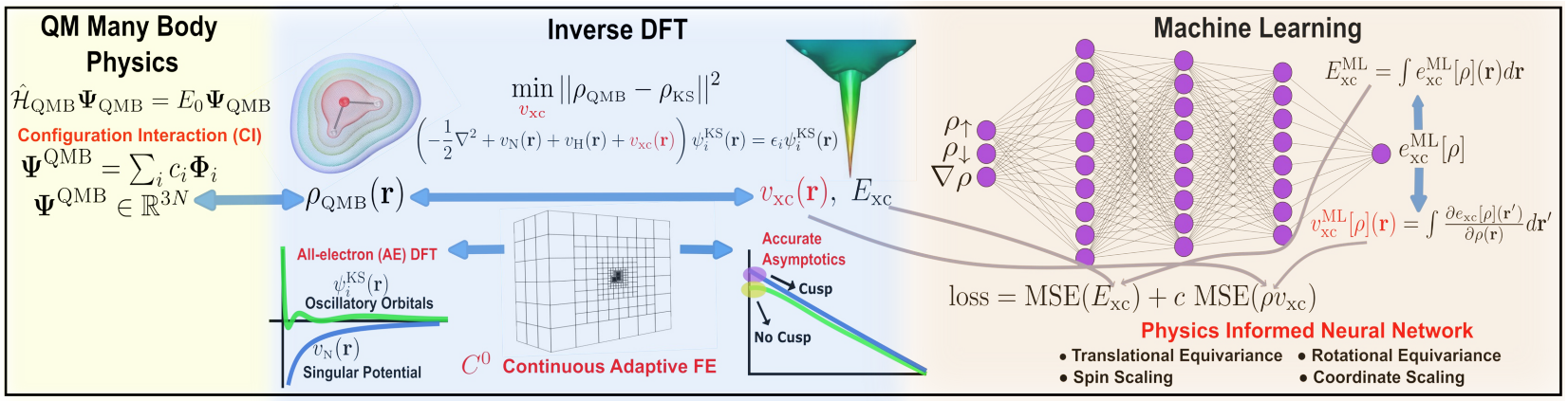
$$\boldsymbol{\Psi}^{\mathrm{QMB}} = \sum_i c_i \boldsymbol{\Phi}_i$$

$$\boldsymbol{\Psi}^{\mathrm{QMB}} \in \mathbb{R}^{3N}$$

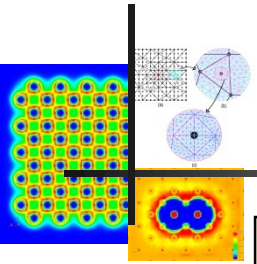
Overview of the approach



Overview of the approach



Overview of the approach



$\mathcal{O}(e^N)$

QM Many Body Physics

$\hat{\mathcal{H}}_{\rm QMB}\Psi_{\rm QMB} = E_0\Psi_{\rm QMB}$

Configuration Interaction (CI)

$\Psi^{\rm QMB} = \sum_i c_i \Phi_i$

$\Psi^{\rm QMB} \in \mathbb{R}^{3N}$

Inverse DFT

$\min_{v_{\rm xc}} ||\rho_{\rm QMB} - \rho_{\rm KS}||^2$

$\left(-\frac{1}{2}\nabla^2 + v_{\rm N}(\mathbf{r}) + v_{\rm H}(\mathbf{r}) + v_{\rm xc}(\mathbf{r})\right)\psi_i^{\rm KS}(\mathbf{r}) = \epsilon_i\psi_i^{\rm KS}(\mathbf{r})$

$\rho_{\rm QMB}(\mathbf{r})$ ⟷ $v_{\rm xc}(\mathbf{r}),\ E_{\rm xc}$

All-electron (AE) DFT

$\psi_i^{\rm KS}(\mathbf{r})$ Oscillatory Orbitals

$v_{\rm N}(\mathbf{r})$ Singular Potential

C^0 Continuous Adaptive FE

Accurate Asymptotics

Cusp

No Cusp

Machine Learning

$\rho\uparrow$, $\rho\downarrow$, $\nabla\rho$

$E_{\rm xc}^{\rm ML} = \int e_{\rm xc}^{\rm ML}[\rho](\mathbf{r})d\mathbf{r}$

$e_{\rm xc}^{\rm ML}[\rho]$

$v_{\rm xc}^{\rm ML}[\rho](\mathbf{r}) = \int \frac{\partial e_{\rm xc}[\rho](\mathbf{r}')}{\partial\rho(\mathbf{r})}d\mathbf{r}'$

$\text{loss} = \text{MSE}(E_{\rm xc}) + c\ \text{MSE}(\rho v_{\rm xc})$

Physics Informed Neural Network

- Translational Equivariance
- Rotational Equivariance
- Spin Scaling
- Coordinate Scaling

MLXC

$\mathcal{O}(N^3)$

Large-scale DFT

Kohn-Sham Non-linear Eigenvalue Problem

$\left(-\frac{1}{2}\nabla^2 + v_{\rm N}(\mathbf{r}) + v_{\rm H}[\rho](\mathbf{r}) + v_{\rm xc}^{\rm ML}[\rho](\mathbf{r})\right)\psi_i^{\rm KS} = \epsilon_i\psi_i^{\rm KS}$

$\rho(\mathbf{r}) = \sum_i^N f_i|\psi_i^{\rm KS}(\mathbf{r})|^2 \qquad (N \sim \mathcal{O}(10^5))$

Chebyshev Accelerated Eigensolver

Wanted Spectrum

Unwanted Spectrum

$\bar{\mathbf{H}} = c_1\tilde{\mathbf{H}} + c_2$

-1.0 1.0

Chebyshev Filter

$\tilde{\Psi}_F = T_m(\bar{\mathbf{H}})\tilde{\Psi}$

$T_m(\bar{\mathbf{H}})\mathbf{X} = [2\bar{\mathbf{H}}T_{m-1}(\bar{\mathbf{H}}) - T_{m-2}(\bar{\mathbf{H}})]\mathbf{X}$

- Magnifies wanted spectrum (left of -1) and supresses unwanted spectrum (between -1 and 1)

- Project discrete eigenvalue problem into the filtered subspace and solve a much smaller eigenvalue problem (Rayleigh-Ritz step)

Overview of the approach

$\mathcal{O}(e^N)$

MLXC

$\mathcal{O}(N^3)$

QM Many Body Physics

$\hat{\mathcal{H}}_{\mathrm{QMB}}\Psi_{\mathrm{QMB}} = E_0\Psi_{\mathrm{QMB}}$

Configuration Interaction (CI)

$\Psi^{\mathrm{QMB}} = \sum_i c_i \Phi_i$

$\Psi^{\mathrm{QMB}} \in \mathbb{R}^{3N}$

Inverse DFT

$\min_{v_{\mathrm{xc}}} ||\rho_{\mathrm{QMB}} - \rho_{\mathrm{KS}}||^2$

$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{N}}(\mathbf{r}) + v_{\mathrm{H}}(\mathbf{r}) + v_{\mathrm{xc}}(\mathbf{r})\right)\psi_i^{\mathrm{KS}}(\mathbf{r}) = \epsilon_i\psi_i^{\mathrm{KS}}(\mathbf{r})$

$\rho_{\mathrm{QMB}}(\mathbf{r})$ ⟷ $v_{\mathrm{xc}}(\mathbf{r})$, E_{xc}

All-electron (AE) DFT

$\psi_i^{\mathrm{KS}}(\mathbf{r})$ Oscillatory Orbitals

$v_{\mathrm{N}}(\mathbf{r})$ Singular Potential

C^0 Continuous Adaptive FE

Accurate Asymptotics

Cusp

No Cusp

Machine Learning

$E_{\mathrm{xc}}^{\mathrm{ML}} = \int e_{\mathrm{xc}}^{\mathrm{ML}}[\rho](\mathbf{r})d\mathbf{r}$

$e_{\mathrm{xc}}^{\mathrm{ML}}[\rho]$

$v_{\mathrm{xc}}^{\mathrm{ML}}[\rho](\mathbf{r}) = \int \frac{\partial e_{\mathrm{xc}}[\rho](\mathbf{r}')}{\partial\rho(\mathbf{r})}d\mathbf{r}'$

$\rho\uparrow$ $\rho\downarrow$ $\nabla\rho$

$\mathrm{loss} = \mathrm{MSE}(E_{\mathrm{xc}}) + c\ \mathrm{MSE}(\rho v_{\mathrm{xc}})$

Physics Informed Neural Network

- Translational Equivariance
- Rotational Equivariance
- Spin Scaling
- Coordinate Scaling

Large-scale DFT

Kohn-Sham Non-linear Eigenvalue Problem

$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{N}}(\mathbf{r}) + v_{\mathrm{H}}[\rho](\mathbf{r}) + v_{\mathrm{xc}}^{\mathrm{ML}}[\rho](\mathbf{r})\right)\psi_i^{\mathrm{KS}} = \epsilon_i\psi_i^{\mathrm{KS}}$

$\rho(\mathbf{r}) = \sum_i^N f_i|\psi_i^{\mathrm{KS}}(\mathbf{r})|^2 \qquad (N \sim \mathcal{O}(10^5))$

Chebyshev Accelerated Eigensolver

Wanted Spectrum | Unwanted Spectrum

$\bar{\mathbf{H}} = c_1\tilde{\mathbf{H}} + c_2$

-1.0 1.0

Chebyshev Filter

$\tilde{\Psi}_F = T_m(\bar{\mathbf{H}})\tilde{\Psi}$

$T_m(\bar{\mathbf{H}})\mathbf{X} = [2\bar{\mathbf{H}}T_{m-1}(\bar{\mathbf{H}}) - T_{m-2}(\bar{\mathbf{H}})]\mathbf{X}$

- Magnifies wanted spectrum (left of -1) and supresses unwanted spectrum (between -1 and 1)
- Project discrete eigenvalue problem into the filtered subspace and solve a much smaller eigenvalue problem (Rayleigh-Ritz step)

Exascale Computational Framework

Accelerated Discrete Hamiltonian Multiplication through FE Cell Level Dense Linear Algebra

$\mathbf{Y} = \mathbf{H}\mathbf{X} \equiv \mathrm{Assembly}_{\mathrm{FE}}\{\mathbf{H}_{c_i}\mathbf{X}_{c_i}\}$

$\mathbf{X} \to M \times N$ Dense Matrix

$\mathbf{H}_{c_i} \to m_c \times m_c$ Dense matrices

$\mathbf{X}_{c_i} \to m_c \times N$ Dense matrices

$m_c \sim 5^3$–9^3 (Higher order FE)

$\mathbf{X}_p$ i $i+B_f$ (M_{loc}, N)

FE Cell c_i

Global to cell extraction

(m_c, m_c) $\mathbf{H}_{c_1}\mathbf{X}_{c_1}^b$ (m_c, N) $\mathbf{H}_{c_2}\mathbf{X}_{c_2}^b$ $\mathbf{H}_{c_n}\mathbf{X}_{c_n}^b$

Strided Batched GEMM over all cells

Cell to global assembly

$\mathbf{Y}_p$ i $i+B_f$ (M_{loc}, N)

MPI partition boundary assembly in FP32

Mixed Precision Compute and Communication Algorithms for $\mathcal{O}(MN^2)$ Rayleigh-Ritz steps

FP32 Arithmetic

FP64 Arithmetic

$\mathbf{X}_p$ or $\tilde{\mathbf{H}}\mathbf{X}_p$ i $i+B_v$

XGEMM

$\mathbf{X}_p^\dagger$ (M_{loc}, N)

$\mathbf{X}^\dagger\mathbf{X}$ or $\mathbf{X}^\dagger\tilde{\mathbf{H}}\mathbf{X}$

i $i+B_v$ (N, M_{loc})

i $i+B_v$ (N, N)

GPU Streams

NCCL/RCCL GPUDirect AllReduce (+)

Distribute to Block-Cyclic Parallel Matrix

Non-Blocking MPI Point-Point

Overlapping of Compute and Communication

Compute For Block k

Data movement/communication for Block k

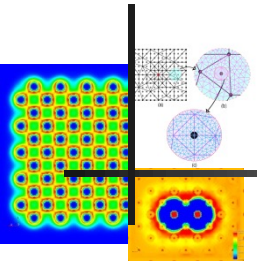
Compute For Block (k+1)

Data movement/communication for Block (k+1)

Run Time

Optimal Memory bandwidth strategies

- Coalesced memory access
- Fused GPU kernels



Inverse DFT

Inverse DFT

Remained an open problem for 25 years. Numerically and conceptually challenging!

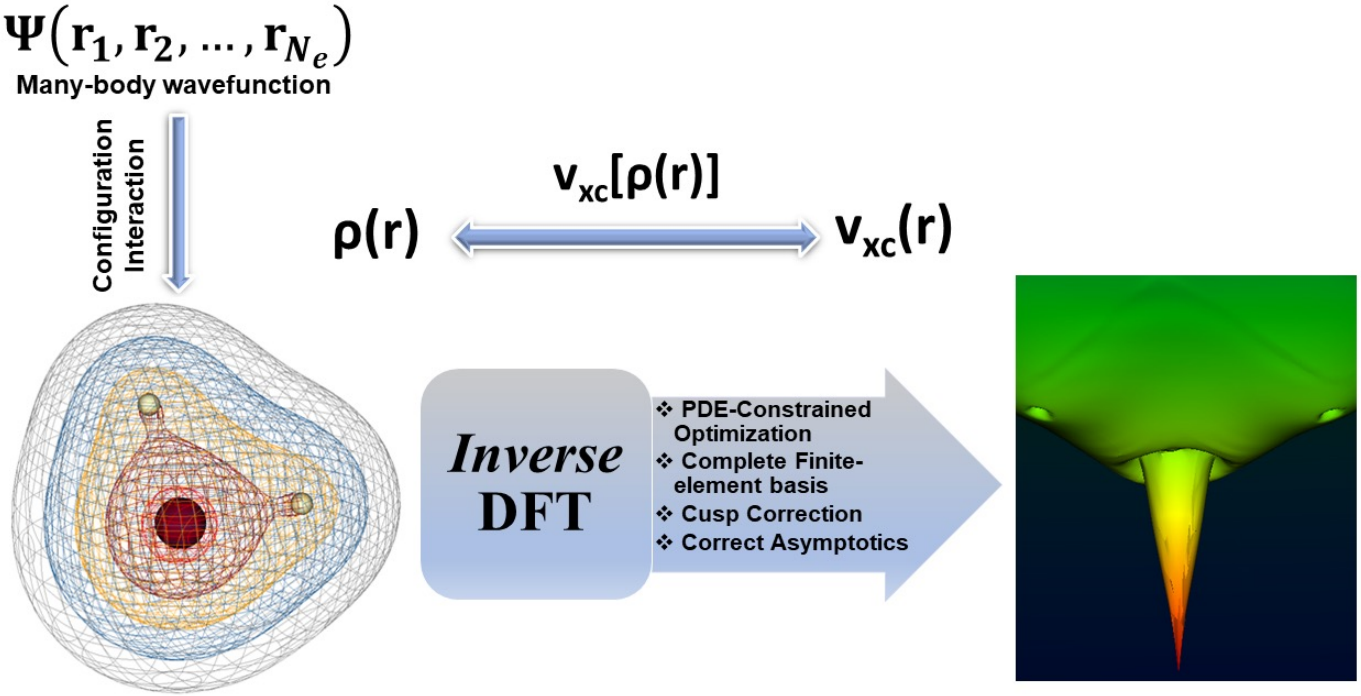
Proposed accurate solution to the inverse DFT problem.

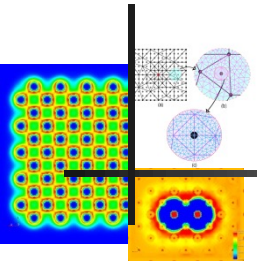
Demonstrated on molecular systems that are both weakly and strongly correlated.

Kanungo et. al., *Nature Communications* **10**, 4497 (2019)

Kanungo et. al., *JPCL* **12**, 12012 (2021)

Kanungo et. al., *JPCL* **14**, 10039 (2023)





Inverse DFT – Key Ideas

- PDE constrained optimization

$$\arg\min_{v_{\mathrm{xc}}(\mathbf{r})} \int w(\mathbf{r}) \left(\rho_{data}(\mathbf{r}) - \rho(\mathbf{r})\right)^2 d\mathbf{r}$$

subject to

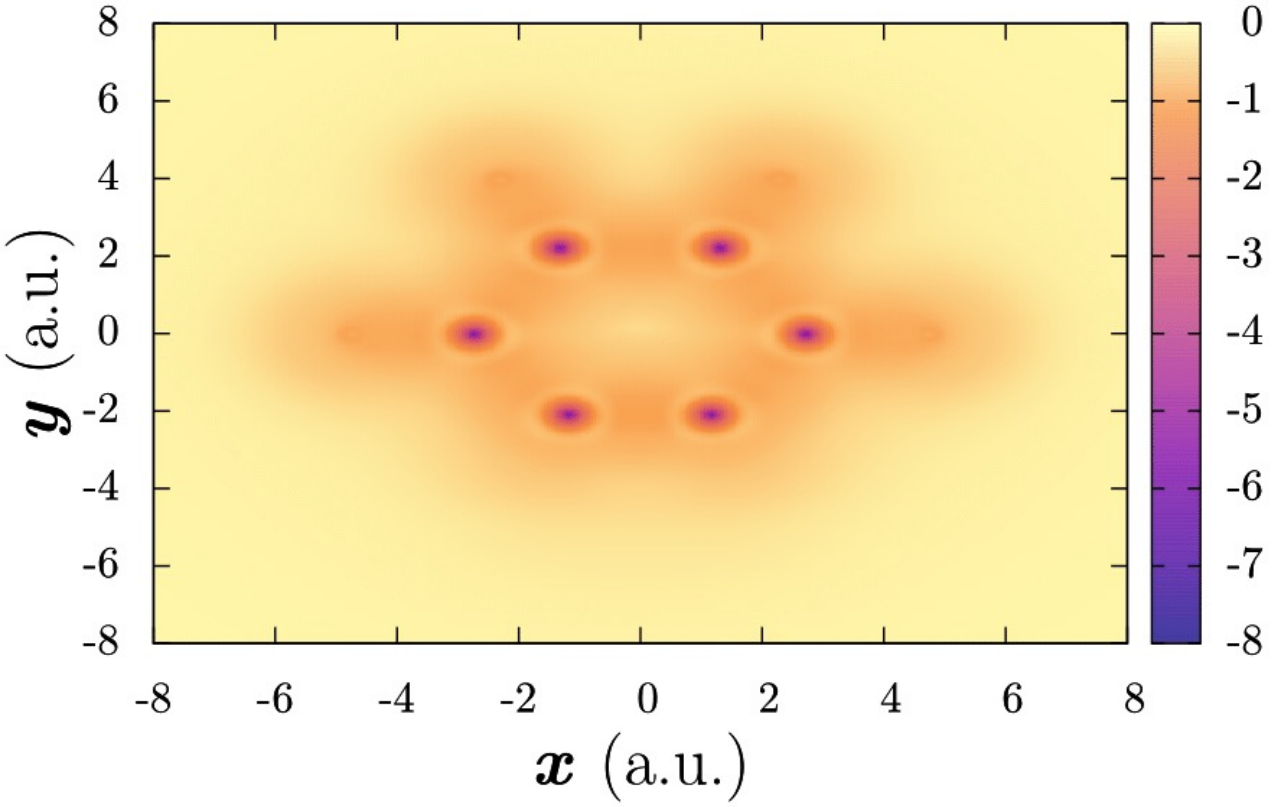
$$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{ext}}(\mathbf{r}) + v_{\mathrm{H}}(\mathbf{r}) + v_{\mathrm{xc}}(\mathbf{r})\right)\psi_i = \epsilon_i \psi_i\,, \qquad \int |\psi_i(\mathbf{r})|^2\, d\mathbf{r} = 1\,.$$

- Higher-order FE basis for discretization – ensures completeness
- Cusp correction: $\Delta\rho(\mathbf{r}) = \rho_{FE}^{DFT}(\mathbf{r}) - \rho_{G}^{DFT}(\mathbf{r})$
- Far-field asymptotics: Start with a guess for $V_{xc}(\mathbf{r})$ with correct far-field asymptotics and use homogeneous Dirichlet boundary conditions on the adjoint fields.

Inverse DFT using highly accurate CI densities

Materials system C_6H_4 (*ortho*-Benzyne) – strongly correlated system

Exact V_{xc} (data from full CI calculation)

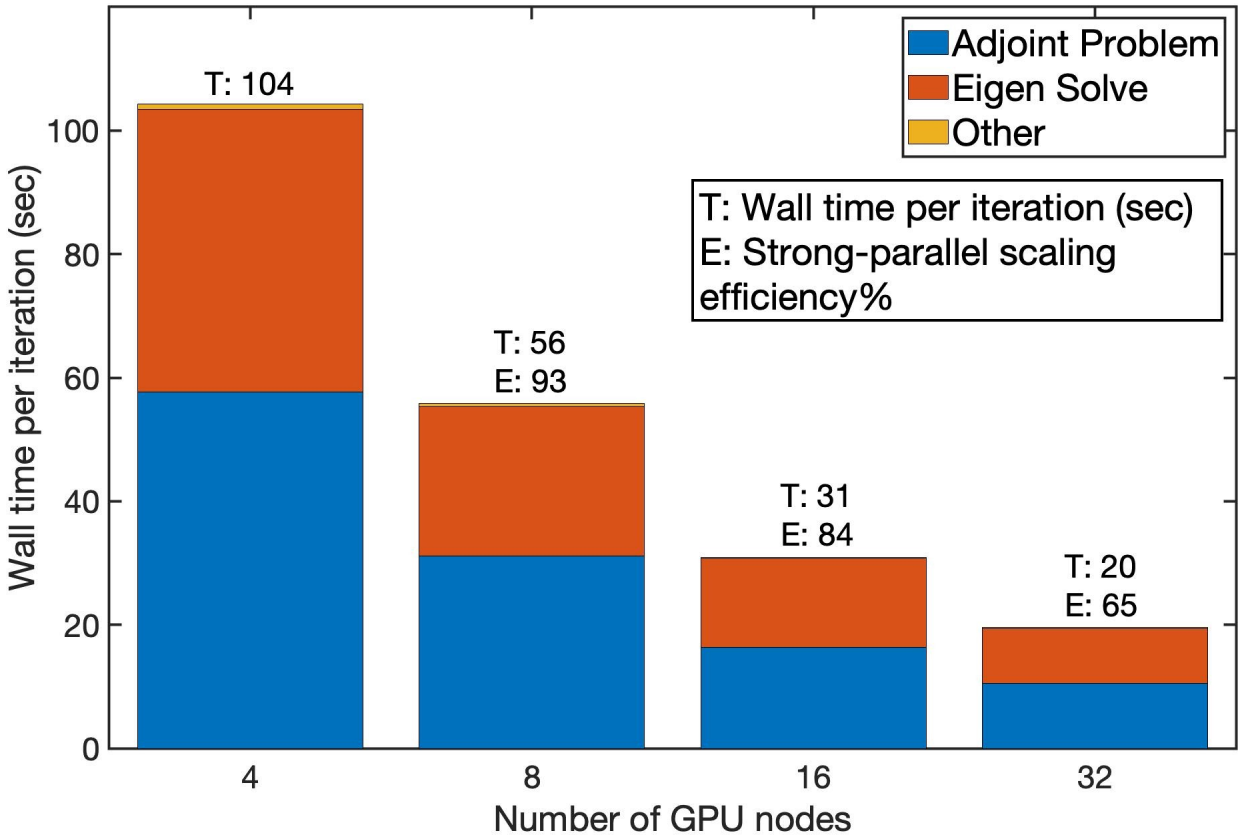


Verification of Koopmans' theorem:

HOMO eigenvalue = -0.354 Ha; $-I_p$= -0.355 Ha

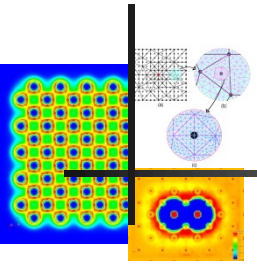
Performance of Inverse DFT implementation

System: Orthobenzyne C_6H_4 (strongly correlated); Perlmutter

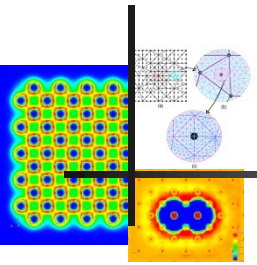


- CPU-GPU speed up: 18x (node-hours basis)
- Wall-time of ~20 sec per BFGS step; Full inverse calculation wall time is 3.5hours (preconditioner optimizations)
- The same system took us ~7 days of wall time previously; ~50x improvement!
- Fast evaluation of 'exact XC potentials' now possible to train DFT functionals.

Inverse DFT CPU-GPU open-source package to be released this year



ML-XC



Learning XC functional — ML-XC

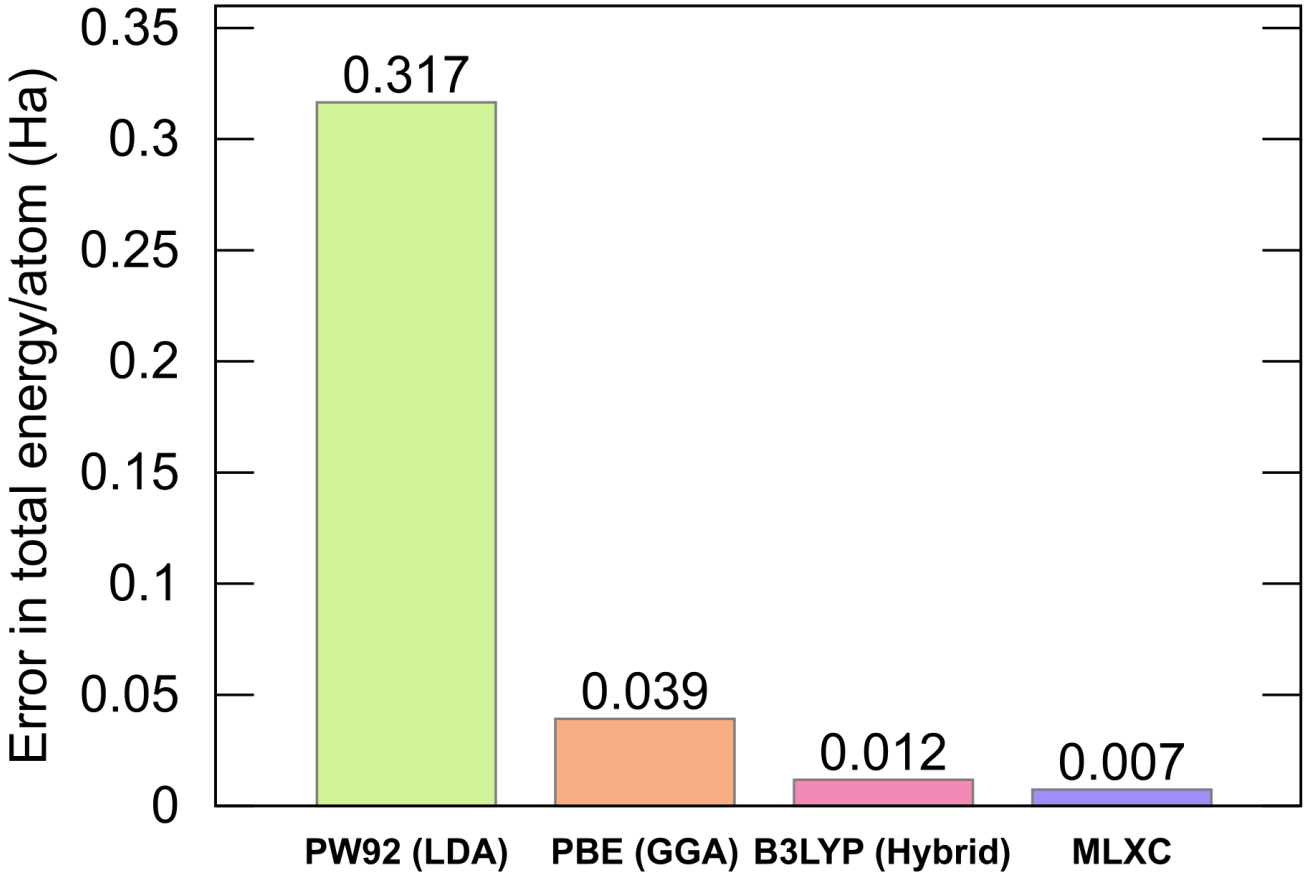
- Express $E_{\rm xc}[\rho] = \int e_{\rm xc}[\rho](\mathbf{r})d\mathbf{r}$ where $e_{\rm xc}[\rho](\mathbf{r})d\mathbf{r}$: energy density
- $e_{xc}^{ML} = \rho^{4/3}(\mathbf{r})\phi(\mathbf{r})F^{DNN}(\rho,\xi,s)$
  - $\xi(\mathbf{r}) = \frac{\rho_\uparrow(\mathbf{r}) - \rho_\downarrow(\mathbf{r})}{\rho(\mathbf{r})}$ $\phi(\mathbf{r}) = \frac{1}{2}\left((1+\xi(\mathbf{r})^{4/3} + (1-\xi(\mathbf{r}))^{4/3}\right)$ $s(\mathbf{r}) = \frac{(3\pi^2)^{1/3}|\nabla\rho(\mathbf{r})|}{2\rho^{4/3}(\mathbf{r})}$
  - Satisfied translational and rotational equivariance; Enforces known coordinate- and spin-scaling relations

$$\mathcal{L}: \sum_I \int \left(v_{\rm xc}^{\rm exact}[\rho_I](\mathbf{r}) - v_{\rm xc}^{\rm ML}[\rho_I](\mathbf{r})\right)^2 d\mathbf{r} + c\sum_I (E_{\rm xc}^{\rm exact,I} - E_{\rm xc}^{\rm ML,I})^2$$

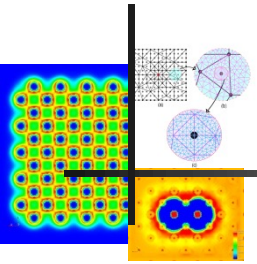
$$v_{\rm xc}^{\rm ML}(\mathbf{r}) = \frac{\delta E_{\rm xc}^{\rm ML}[\rho]}{\delta\rho(\mathbf{r})} = \int \frac{\delta e_{\rm xc}^{\rm ML}(\mathbf{r}')}{\delta\rho(\mathbf{r})}d\mathbf{r}'$$

- Learning is on $e_{\rm xc}[\rho](\mathbf{r})$, but optimization is on $v_{xc}(\mathbf{r})$

Accuracy of ML-XC



- Trained on 'exact XC potentials' from *invDFT* and XC energy of 3 atoms (Li, N, Ne) and 2 molecules (H_2 , LiH).
- The MAE in energy per atom from G2 thermochemistry data.



ML-XC: The Way Ahead

- Expand the training data to solids
  - Accurate groundstate densities from QMC (QMCPACK code)
  - Extension of inverse DFT to solids

- More sophisticated models
  - Laplacian meta-GGA:

$$e_{\mathrm{XC}}(\mathbf{r}) = e_{\mathrm{XC}}^{\mathrm{ML}}[\rho, |\nabla\rho|, \nabla^2\rho](\mathbf{r})$$

  - Kinetic-energy meta-GGA:

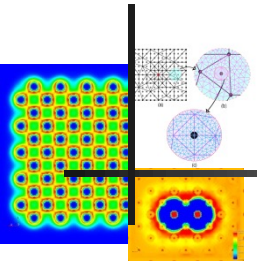
$$e_{\mathrm{XC}}(\mathbf{r}) = e_{\mathrm{XC}}^{\mathrm{ML}}[\rho, |\nabla\rho|, \tau](\mathbf{r}) \qquad \tau(\mathbf{r}) = \frac{1}{2}\sum_i |\nabla\psi_i(\mathbf{r})|^2$$ (Kinetic energy density)

  - Fractional derivatives based non-local models

$$e_{\mathrm{XC}}(\mathbf{r}) = e_{\mathrm{XC}}^{\mathrm{ML}}[\rho, |\nabla|\rho, D^{\alpha_1}\rho, D^{\alpha_2}\rho, \ldots](\mathbf{r})$$

$$D^{\alpha}\rho(\mathbf{r}) = C(\alpha)\int \frac{\rho(\mathbf{r}) - \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^{3+2\alpha}} d\mathbf{r}'$$ (Fractional Laplacian)

    - Naturally encapsulates non-local features
    - Modulation of range of non-locality and decay-rate via different α's

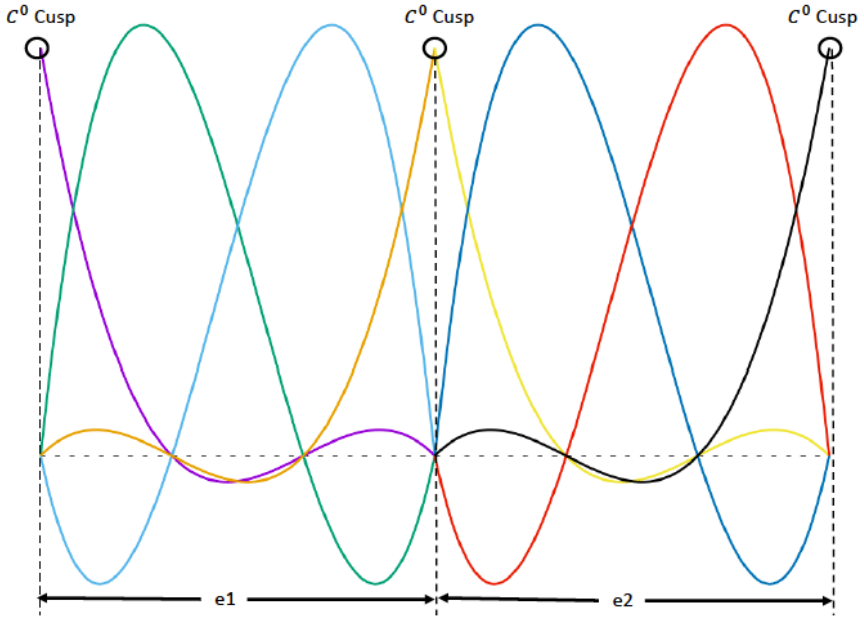


Large-scale DFT calculations

DFT – Finite Element discretization

Use finite-element basis for computing –

$$\psi_i^h(\mathbf{r}) = \sum_k \psi_{ik} N_k(\mathbf{r}) \quad i = 1, 2, \ldots, N$$



<u>Features of FE basis</u>

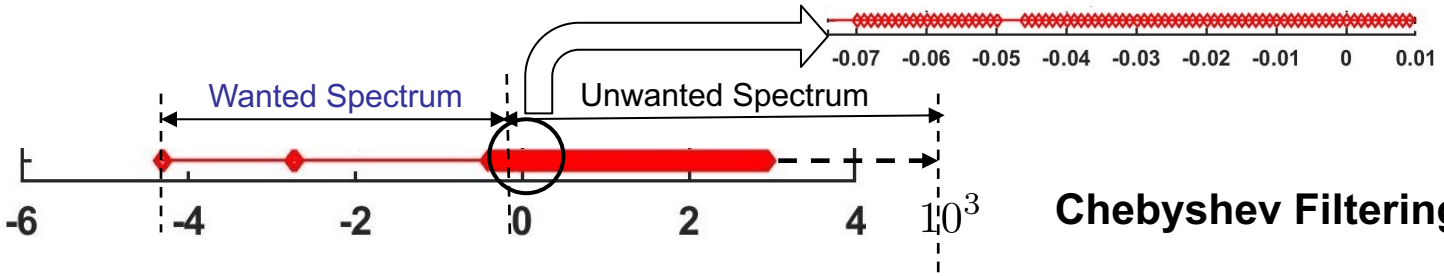
- Systematic convergence
  - Element size
  - Polynomial order
- Adaptive refinement
- Potential for excellent parallel scalability

- Higher-order finite-element discretization; p=6-8 for computational efficiency
- Spectral finite-elements
  - Better conditioning of basis
  - In conjunction with Gauss Lobatto Legendre quadrature results in a standard eigenvalue problem

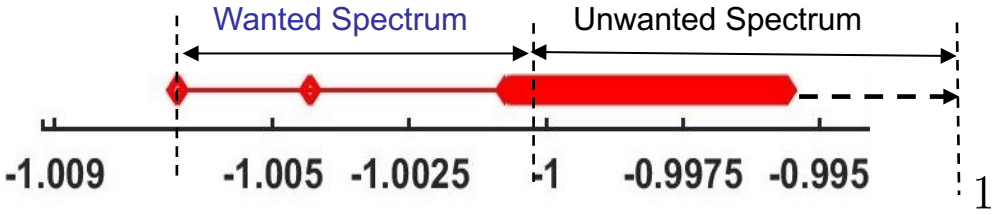
Eigen-space computation: Chebyshev acceleration

(Zhou et al. J. Comput. Phys. 219 (2006); Motamarri et al. J. Comp. Phys. 253, 308-343 (2013))

Kohn-Sham eigenvalue problem: $\widetilde{\mathbf{H}}\widetilde{\psi}_k = \epsilon_k \widetilde{\psi}_k$ for $k = 1,2,\ldots N$ $(N \sim N_e)$

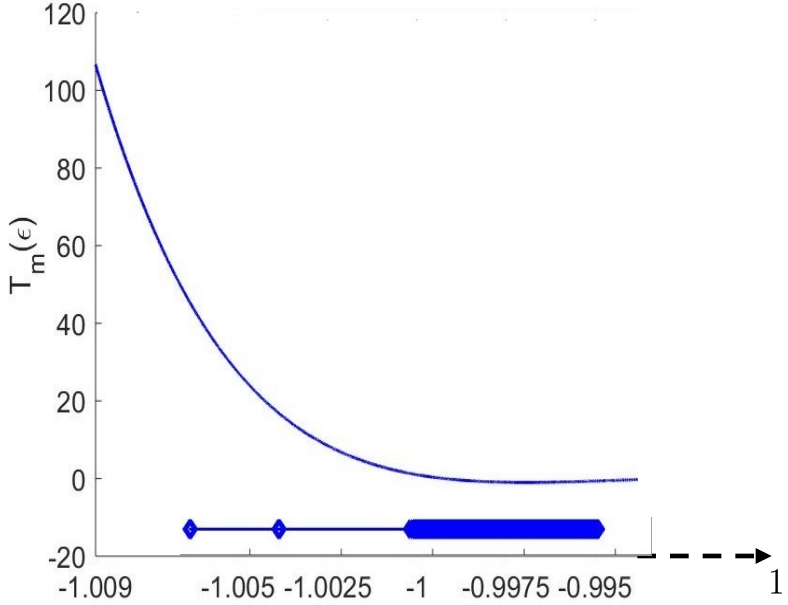


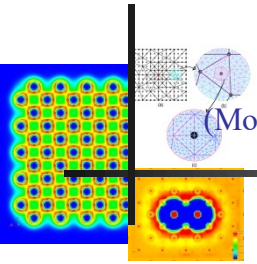
$$\bar{\mathbf{H}} = c_1 \widetilde{\mathbf{H}} + c_2$$



Chebyshev Filtering: $T_m(\bar{\mathbf{H}})\widetilde{\mathbf{\Psi}} = \widetilde{\mathbf{\Psi}}_F$

$$T_m(\bar{\mathbf{H}})\mathbf{X} = [2\bar{\mathbf{H}}T_{m-1}(\bar{\mathbf{H}}) - T_{m-2}(\bar{\mathbf{H}})]\mathbf{X}$$





Numerical algorithm

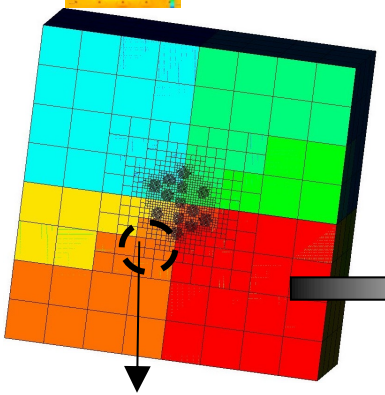
(Motamarri and Das et al. Comput. Phys. Commun (2020), Das and Motamarri et al. Comput. Phys. Commun (2022))

1. Start with initial guess for electron density $\rho_{in}^{h}(\mathbf{r}) = \rho_0(\mathbf{r})$ and the initial wavefunctions
2. Compute the discrete Hamiltonian $\tilde{\mathbf{H}}$ using the input electron density ρ_{in}^{h}
3. **CF:** Chebyshev filtering: $\widetilde{\mathbf{\Psi}}_F = T_m(\bar{\mathbf{H}})\widetilde{\mathbf{\Psi}}$
4. **Orthonormalize** CF basis: $\widetilde{\mathbf{\Psi}}_F \rightarrow \widetilde{\mathbf{\Psi}}_F^o$
5. **Rayleigh-Ritz procedure**:
   - Compute projected Hamiltonian: $\hat{\mathbf{H}} = \widetilde{\mathbf{\Psi}}_F^{o\dagger}\widetilde{\mathbf{H}}\widetilde{\mathbf{\Psi}}_F^{o}$
   - Diagonalize $\hat{\mathbf{H}}$: $\hat{\mathbf{H}}\mathbf{Q} = \mathbf{Q}\mathbf{D}$
   - Subspace rotation: $\widetilde{\mathbf{\Psi}}^{\mathbf{R}} = \widetilde{\mathbf{\Psi}}_{\mathbf{F}}^{o}\mathbf{Q}$
6. Compute electron density $\rho_{out}^{h}(\mathbf{r}) = 2\sum_{i=1}^{N} f(\epsilon_i^h, \mu)|\psi_i^h(\mathbf{r})|^2$
7. If $||\rho_{out}^{h}(\mathbf{r}) - \rho_{in}^{h}(\mathbf{r})|| < tol$, EXIT; else, compute new ρ_{in}^{h} using a mixing scheme and go to (2).

Chebyshev Filtering

$$\mathbf{Y} = \mathbf{H}\,\mathbf{X}$$

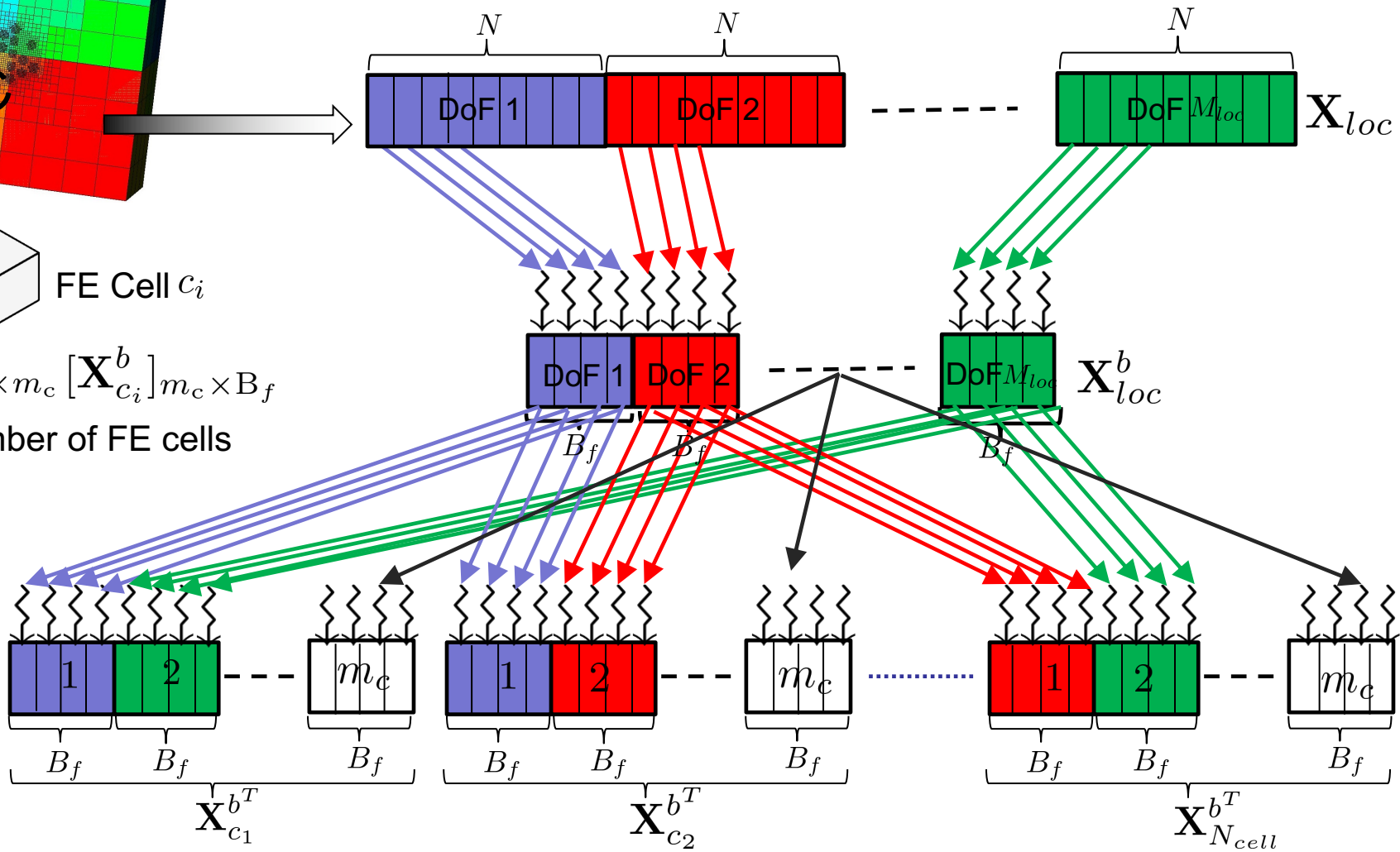
$\mathbf{H} \rightarrow$ Sparse Matrix $(M \times M)$
$\mathbf{X} \rightarrow$ Dense Matrix $(M \times N)$
$\mathbf{Y} \rightarrow$ Dense Matrix $(M \times N)$



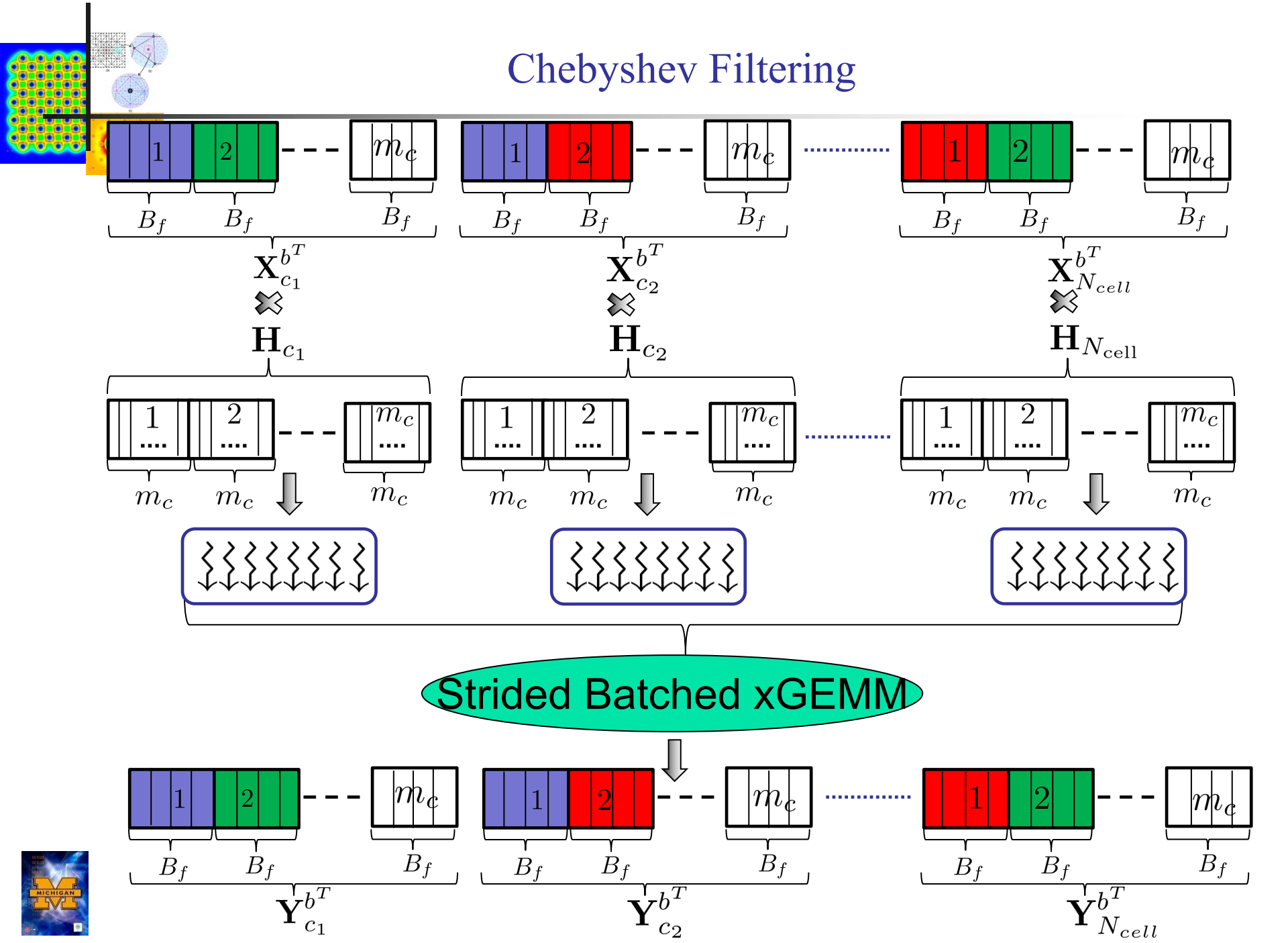
FE Cell c_i

$[\mathbf{H}_{c_i}]_{m_c \times m_c}\,[\mathbf{X}^b_{c_i}]_{m_c \times B_f}$

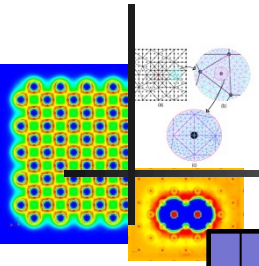
N_{cell} : Number of FE cells

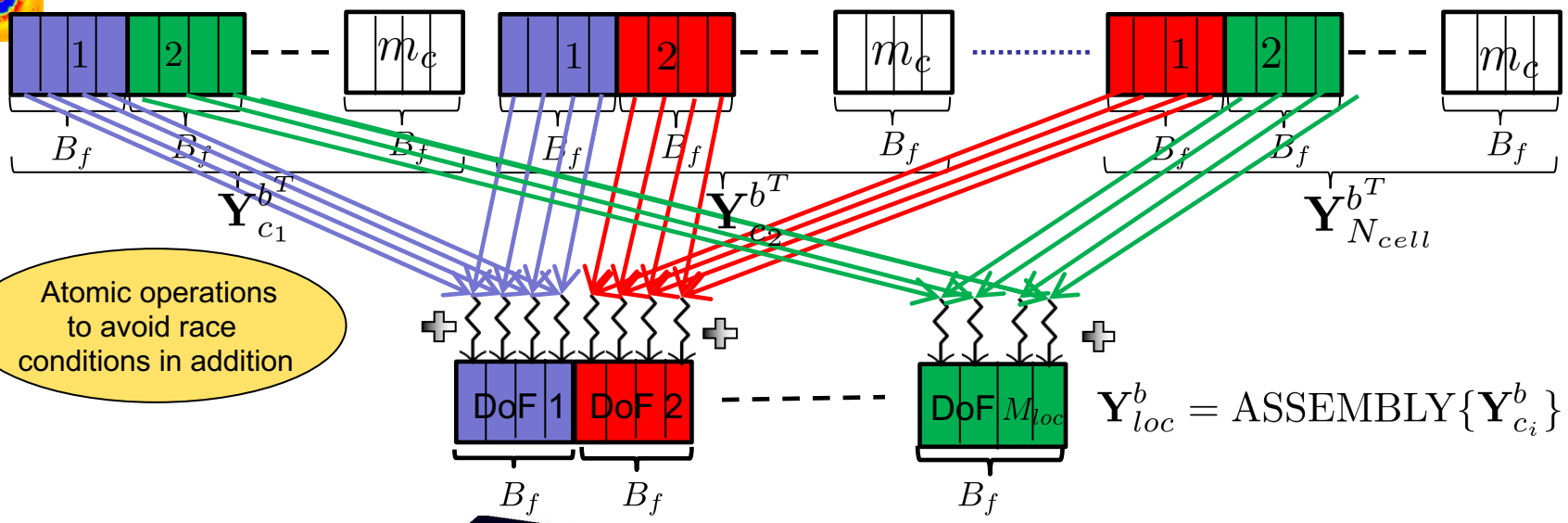


Chebyshev Filtering



Chebyshev Filtering

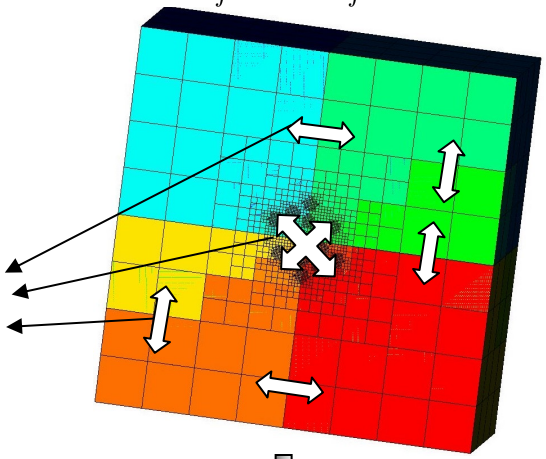




Atomic operations to avoid race conditions in addition

$$\mathbf{Y}^b_{loc} = \text{ASSEMBLY}\{\mathbf{Y}^b_{c_i}\}$$

$$\mathbf{Y}^b = \text{ASSEMBLY}\{\mathbf{Y}^b_{loc}\}$$



Assembly across processor boundaries: Communication in FP32

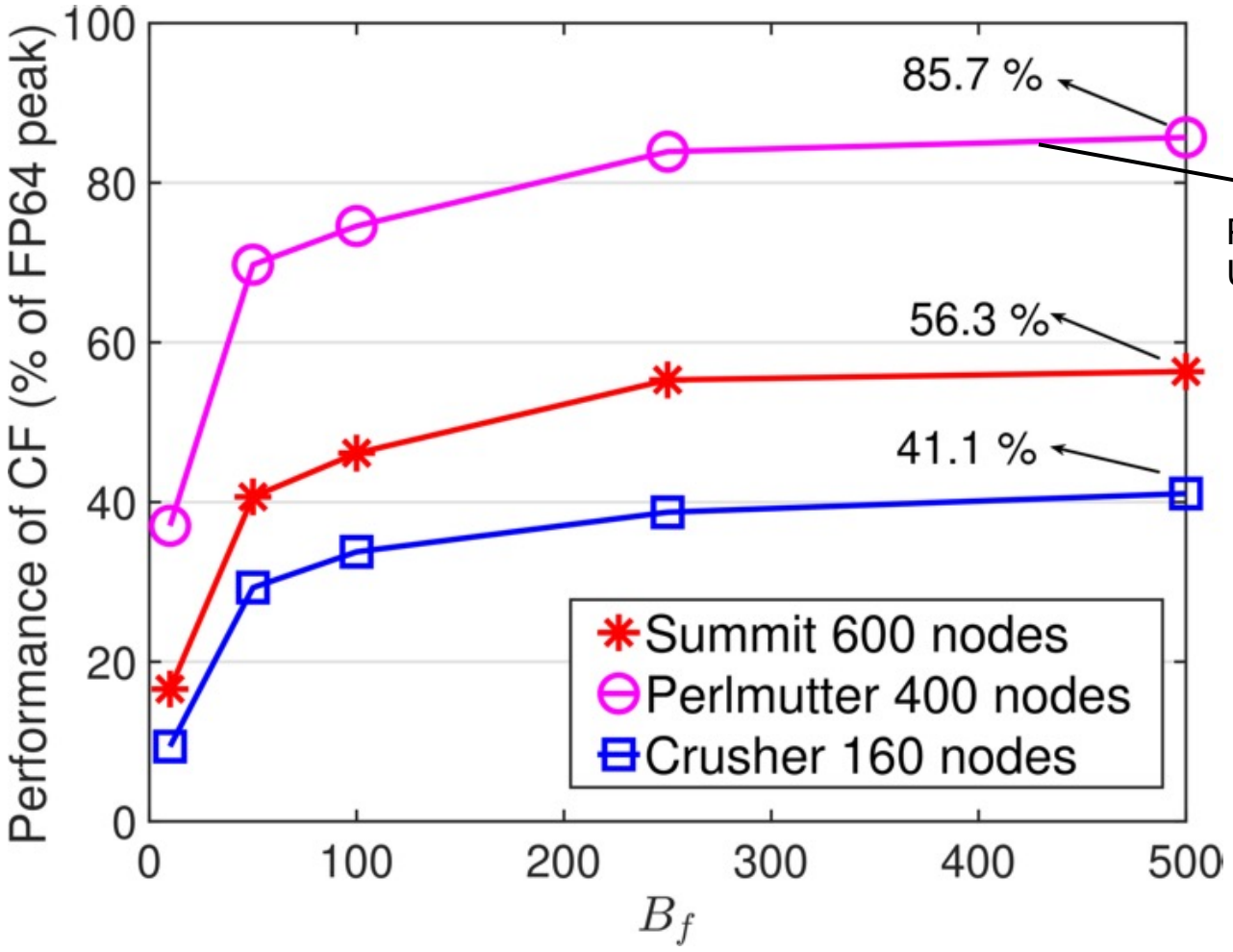
Blocked approach allows overlapping compute of one block with communication of another block

$$\mathbf{Y}^b = T_m(\mathbf{H})\mathbf{X}^b = [2\,\mathbf{H}\,T_{m-1}(\mathbf{H}) - T_{m-2}(\mathbf{H})]\mathbf{X}^b$$

Repeat for $b = 1 \cdots \frac{N}{B_f}$

Performance of Chebyshev filtering

Case study: Mg-Y dislocation system; 6,016 atoms (~12,000 e-)



Perlmutter:
Using FP64 tensor cores

Orthogonalization: Cholesky Gram-Schmidt

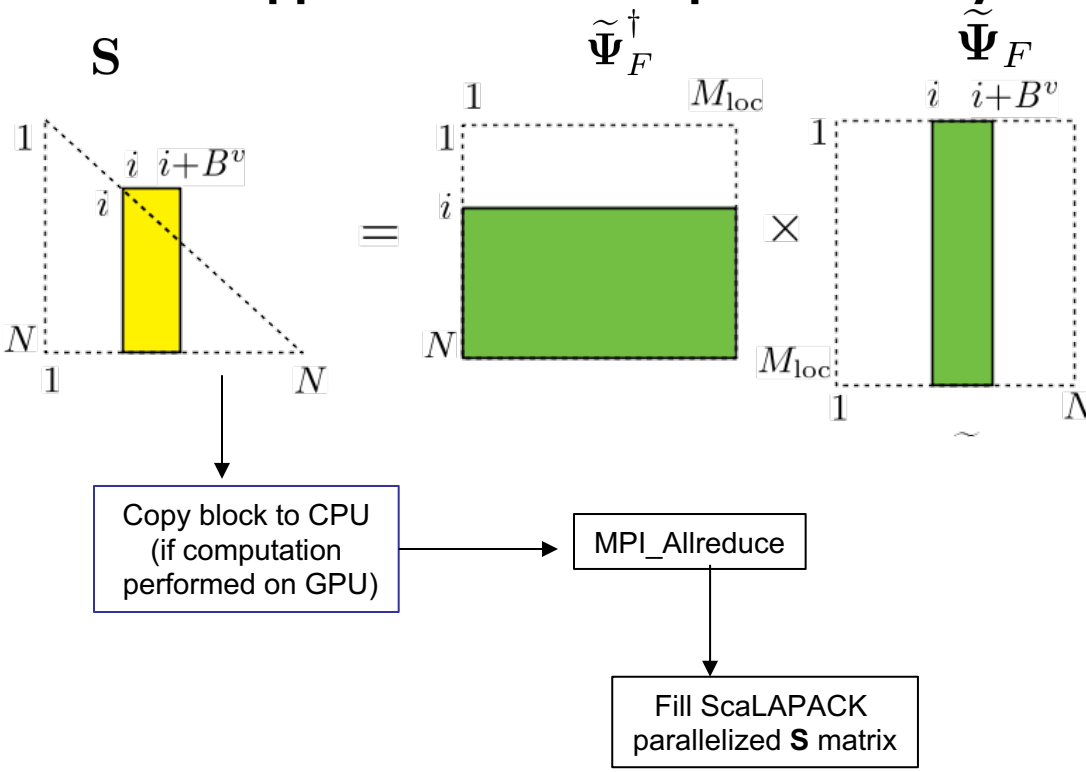
- Cholesky factorization of the overlap matrix: $\mathbf{S} = \widetilde{\mathbf{\Psi}}_F^{\dagger}\widetilde{\mathbf{\Psi}}_F = \mathbf{L}\mathbf{L}^{\dagger}.\ \mathcal{O}(MN^2)$
- Orthonormal basis construction: $\widetilde{\mathbf{\Psi}}_F^{\mathrm{o}} = \widetilde{\mathbf{\Psi}}_F \mathbf{L}^{-\mathbf{1}^{\dagger}}.\ \mathcal{O}(MN^2)$

Mixed precision computation for Chol-GS

1. $\mathbf{S} = \mathrm{DP}\left\{\mathbf{S_d}\right\} + \mathrm{SP}\left\{\mathbf{S_{od}}\right\}$
2. $\mathbf{S} = \mathbf{L}\mathbf{L}^{\dagger}$ in double precision.
3. Orthonormal basis construction:

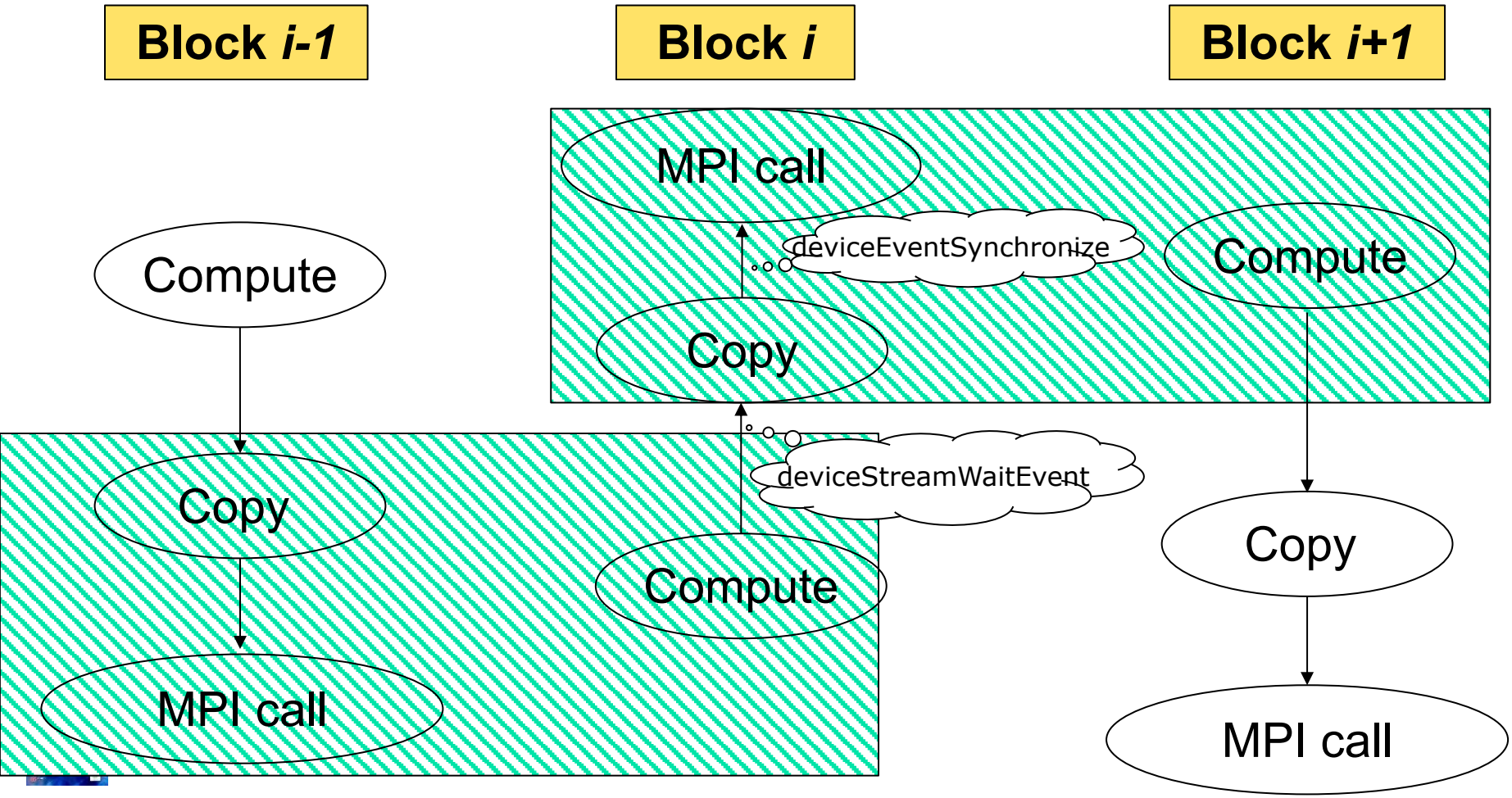
$$\widetilde{\mathbf{\Psi}}_F^{\mathrm{o}} = \mathrm{DP}\left\{\widetilde{\mathbf{\Psi}}_F \mathbf{L}_{\mathbf{d}}^{-\mathbf{1}^{\dagger}}\right\} + \mathrm{SP}\left\{\widetilde{\mathbf{\Psi}}_F \mathbf{L}^{-\mathbf{1}^{\dagger}}_{\mathbf{od}}\right\}$$

Blocked approach to reduce peak memory



Overlapping compute and data movement on GPUs

Execute copy and MPI calls of current block asynchronously with compute of the successive block.



Rayleigh-Ritz procedure

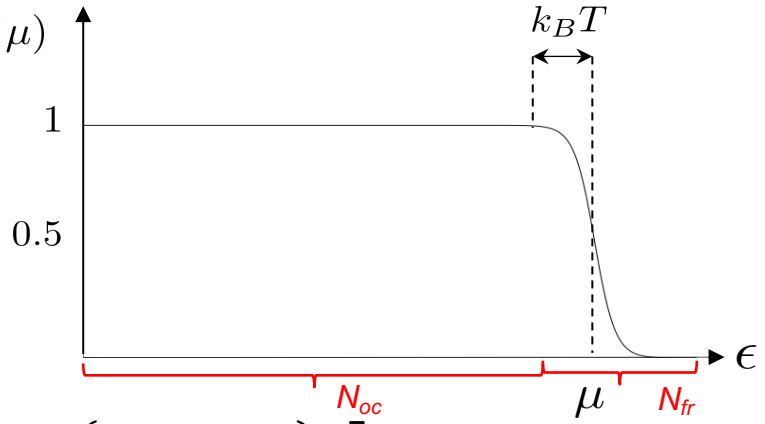
- Compute projected Hamiltonian: $\hat{\mathbf{H}} = \widetilde{\mathbf{\Psi}}_F^{\mathrm{o}\dagger} \widetilde{\mathbf{H}} \widetilde{\mathbf{\Psi}}_F^{\mathrm{o}}$. $\mathcal{O}(MN^2)$
- Diagonalization of $\hat{\mathbf{H}}$: $\hat{\mathbf{H}}\mathbf{Q} = \mathbf{Q}\mathbf{D}$. $\mathcal{O}(N^3)$
- Subspace rotation step: $\widetilde{\mathbf{\Psi}}^{\mathbf{R}} = \widetilde{\mathbf{\Psi}}_{\mathbf{F}}^{\mathrm{o}} \mathbf{Q}$. $\mathcal{O}(MN^2)$

Mixed precision computation for RR

Compute projected Hamiltonian:

$$\rho_{\mathrm{out}}^h(\mathbf{x}) = 2\sum_{i=1}^{N} f(\epsilon_i^h, \mu)|\psi_i^h(\mathbf{x})|^2$$

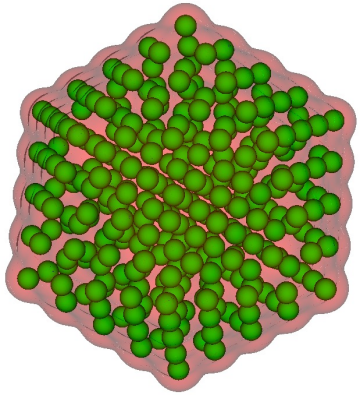
$$\widetilde{\mathbf{\Psi}}_F^{\mathrm{o}} = \left[\widetilde{\mathbf{\Psi}}_{\mathrm{oc}}^{\mathrm{o}} \; \widetilde{\mathbf{\Psi}}_{\mathrm{fr}}^{\mathrm{o}}\right]$$



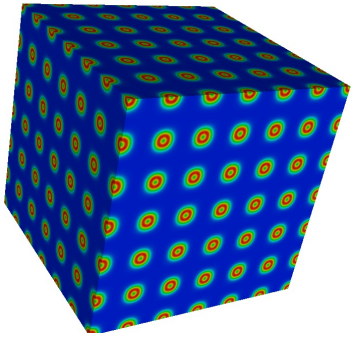
$$\left[\begin{array}{c|c} \hat{\mathbf{H}}_{\mathbf{oc-oc}} & \hat{\mathbf{H}}_{\mathbf{oc-fr}} \\ \hline \hat{\mathbf{H}}_{\mathbf{fr-oc}} & \hat{\mathbf{H}}_{\mathbf{fr-fr}} \end{array}\right] = \left[\begin{array}{c|c} \mathrm{SP}\left\{\widetilde{\mathbf{\Psi}}_{\mathrm{oc}}^{\mathrm{o}\dagger} \widetilde{\mathbf{H}} \widetilde{\mathbf{\Psi}}_{\mathrm{oc}}^{\mathrm{o}}\right\} & \mathrm{SP}\left\{\widetilde{\mathbf{\Psi}}_{\mathrm{oc}}^{\mathrm{o}\dagger} \widetilde{\mathbf{H}} \widetilde{\mathbf{\Psi}}_{\mathrm{fr}}^{\mathrm{o}}\right\} \\ \hline \mathrm{SP}\left\{\widetilde{\mathbf{\Psi}}_{\mathrm{fr}}^{\mathrm{o}\dagger} \widetilde{\mathbf{H}} \widetilde{\mathbf{\Psi}}_{\mathrm{oc}}^{\mathrm{o}}\right\} & \mathrm{DP}\left\{\widetilde{\mathbf{\Psi}}_{\mathrm{fr}}^{\mathrm{o}\dagger} \widetilde{\mathbf{H}} \widetilde{\mathbf{\Psi}}_{\mathrm{fr}}^{\mathrm{o}}\right\} \end{array}\right]$$

Subspace rotation step: $\widetilde{\mathbf{\Psi}}^{\mathrm{R}} = \mathrm{DP}\left[\widetilde{\mathbf{\Psi}}_F^{\mathrm{o}} \mathbf{Q}_{\mathbf{d}}\right] + \mathrm{SP}\left[\widetilde{\mathbf{\Psi}}_F^{\mathrm{o}} \mathbf{Q}_{\mathbf{od}}\right]$

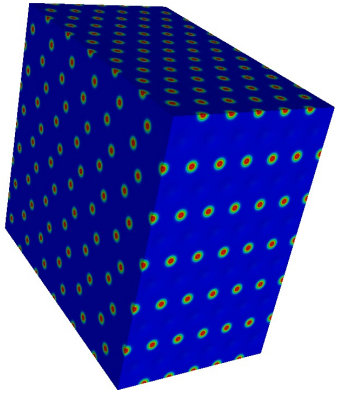
Accuracy and robustness of mixed precision computations



I. Cu nanoparticle
5871 electrons



II. Mo periodic
supercell w/ vacancy
6034 electrons

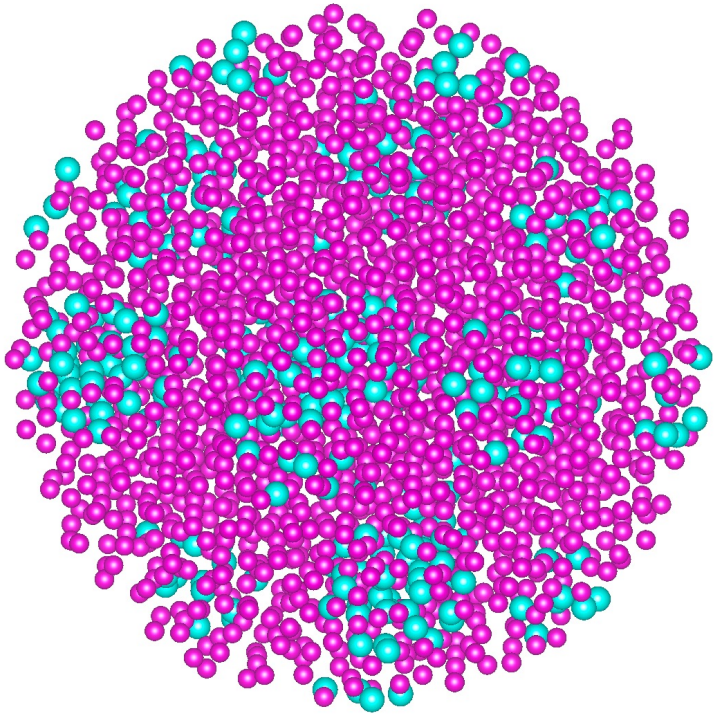


III. Mg periodic
supercell w/ vacancy
8630 electrons

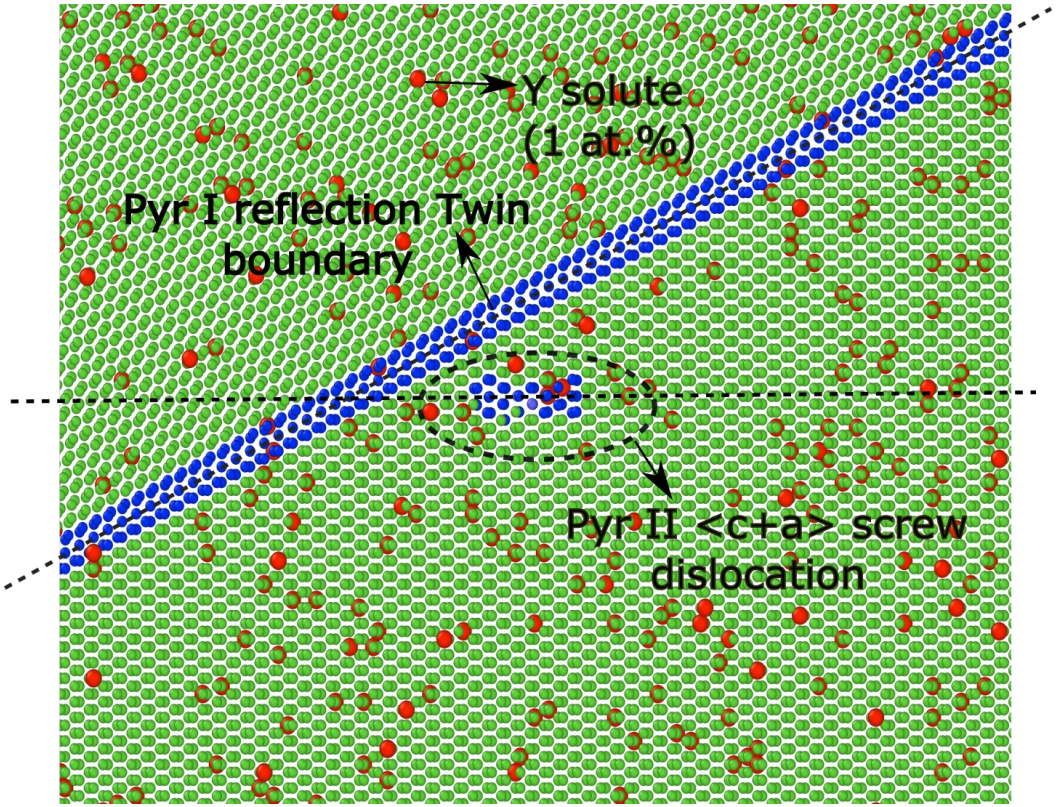
Energy error (Ha/atom)	5×10^{-12}	7×10^{-12}	3×10^{-12}
Max force error (Ha/Bohr)	3×10^{-6}	7×10^{-7}	2×10^{-6}
Total SCFs (DP, MP)	(46,46)	(30,30)	(18,18)

Benchmark Systems

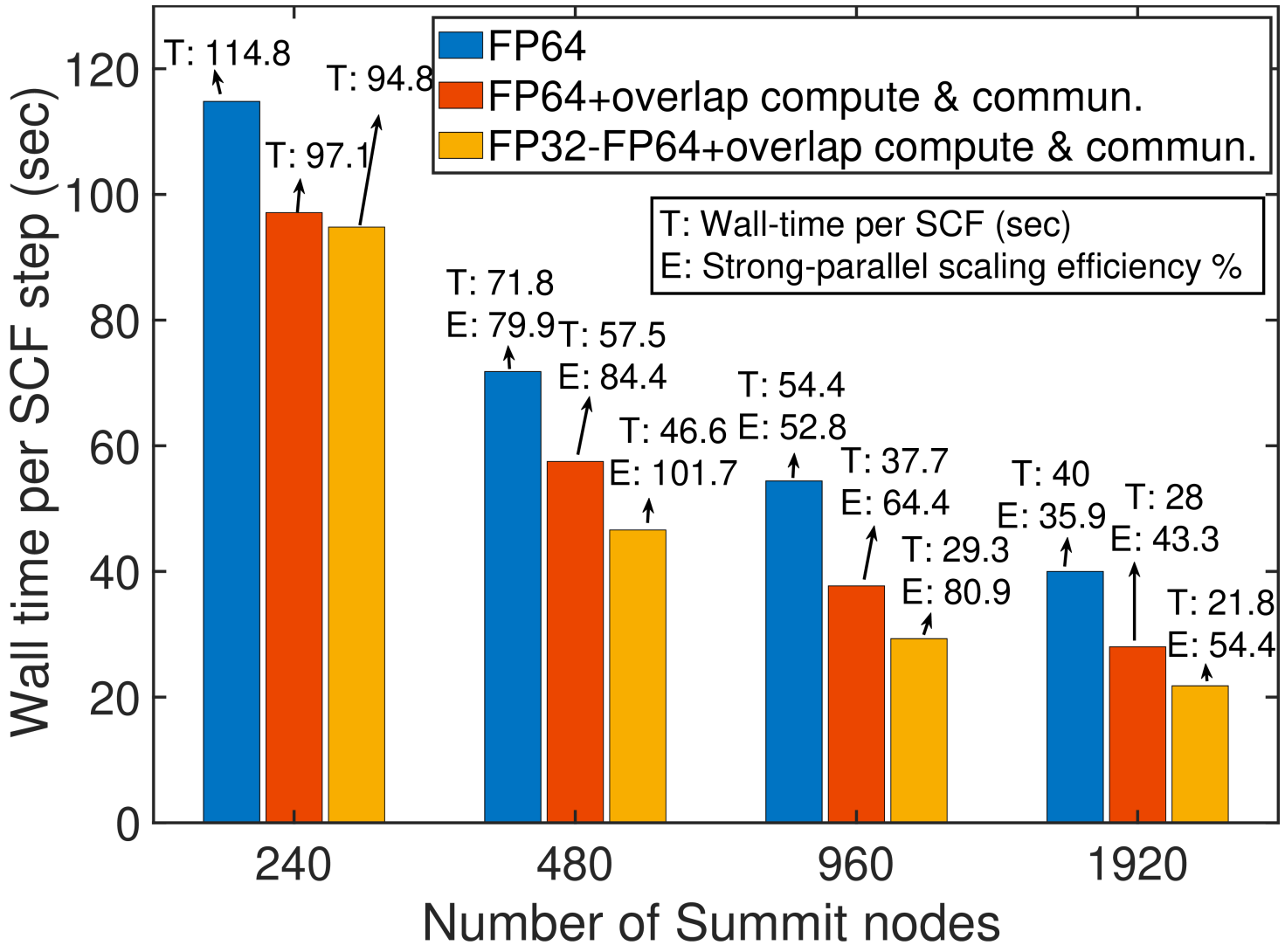
1. *YbCd quasicrystal*: 1,940 atoms, 40,040 e- ; Full ground state calculation



2. *Mg-Y alloy, <c+a> dislocation twin boundary interaction*: (74,164 atoms, 154,781 e-)x4 k-pts



DFT-FE-ML-XC : Yb-Cd quasicrystal (~40,000 e-)



Strong-scaling: Yb-Cd quasicrystal (~40,000 e-)

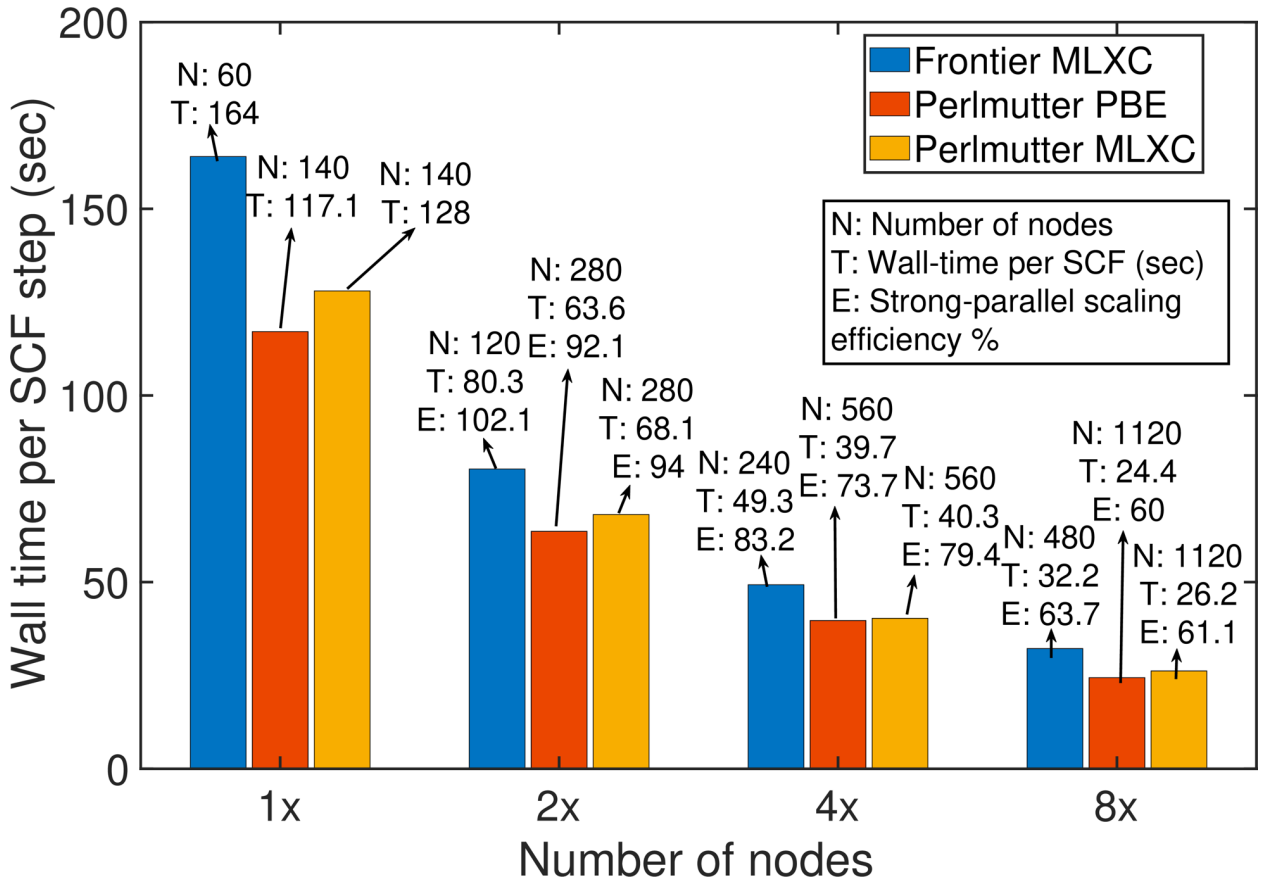


Table: Time to solution for ground-state (1,120 Perlmutter nodes)

Initialization (sec)	Total SCF (sec)	Total run (sec)
69	2023 (34 SCFs)	2092

Full ground-state in ~30 mins

Large-scale Interacting Extended Defects: Performance (Frontier)

(Das, Kanungo and Subramanian et al. SC23)

TwinDislocMgY (A): 36,344 atoms (75,667 e^-) x 4-kpoints

- 2400 Frontier nodes (FP64 peak: 458.9 PFLOPS)

TwinDislocMgY (B): 74,164 atoms (154,781) x 4-kpoints

- 1.7 billion DoF
- 605 trillion wavefunction values
- 8000 Frontier nodes (FP64 peak: 1529.6 PFLOPS)

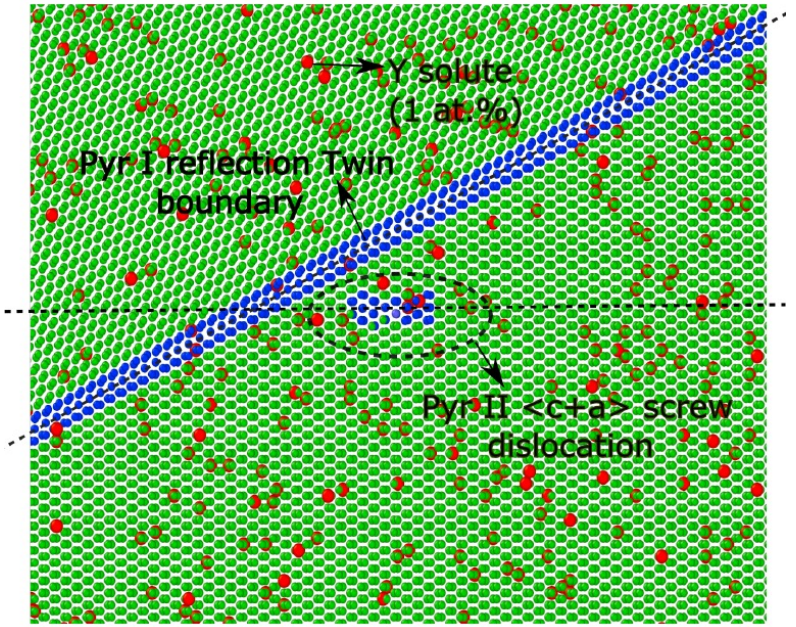
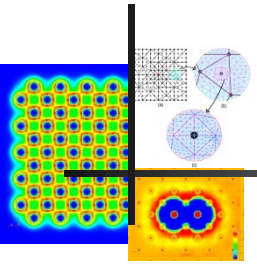


Table: Program run time and sustained performance per SCF iteration.

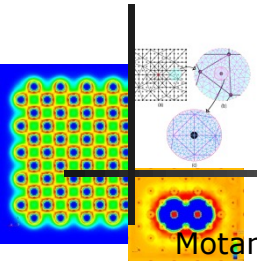
System	Wall-time (sec)	FLOP count (PFLOP)	Performance (PFLOPS)	% of FP64 peak
TwinDislocMgY (A)	223	50,457	226.3	49.3%
TwinDislocMgY (B)	513.7	338,863	659.7	43.1%

Unprecedented sustained performance for ab-initio GS calculation;
10x higher than previous best (64 PFLOPS; SC22)



Performance breakdown

Step	Wall-time (sec)	FLOP count (PFLOP)	Performance (PFLOPS)	% of FP64 peak
CF	135.4	57,810	427	27.9%
CholGS-S-MP	79.3	54,430	686.4	44.9%
CholGS-CI	8.8	-	-	-
CholGS-O-MP	49.6	54,430	1097.4	71.7%
RR-P-MP	66.7	61,036	915.1	59.8%
RR-D	22.3	-	-	-
RR-SR-MP	93.5	108,858	1164.3	76.1%
DC	4.3	2303	535.5	35%
Others	53.8	-	-	-
Single SCF Total	**513.7**	**338,863**	**659.7**	**43.1%**



DFT-FE open source code

https://github.com/dftfeDevelopers/dftfe;

Motamarri and Das et al. Comput. Phys. Commun (2020), Das and Motamarri et al. (Comput. Phys. Commun (2022))

- Boundary conditions: Periodic, non-periodic, semi-periodic
- Physics:
  - Norm conserving pseudopotentials (TM; ONCV) and All-electron calculations (classical FE)
  - Density and density gradient based XC functionals, including spin-polarization
- Calculations:
  - Ground-state energy
  - Geometry (ionic and cell) relaxation (Motamarri & Gavini, Phys. Rev. B 97, 165132 (2018), Das and Motamarri et al. Comput. Phys. Commun (2022))
  - Ab-initio MD (NVE, NVT), Nudged elastic band
- Scaling: Tested on Frontier, Summit, Perlmutter, Cori, Stampede2
- Current development branch with GPU acceleration on both NVIDIA and AMD GPUs:
  - ~20x speedups in comparison to CPUs on a node-to-node basis
  - 660 PFLOPS of sustained performance achieved on Frontier for a ~155,000 electrons (x4 k-pts) metallic system; ~43% efficiency
  - Nominated 2019 and 2023 ACM Gordon Bell prize.
- Upcoming capabilities:
  - Meta-GGA and non-local MLXC functionals
  - Non-collinear magnetism + spin-orbit coupling
  - Softer PAW pseudopotentials

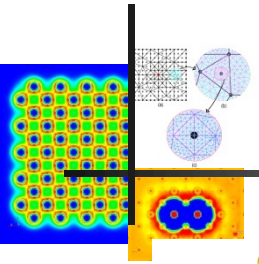
Application: Technological challenge of low ductility in Mg

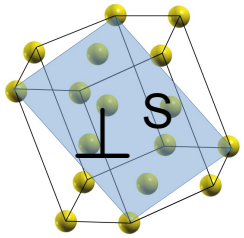
- Magnesium is the lightest structural metal with high strength to weight ratio
  - 75% lighter than Steel and 30% lighter than Aluminum
- Every 10% reduction in the weight of a vehicle will result in 6-8% increase in fuel efficiency.
  - Important implications to fuel efficiency and reducing carbon footprint
- Low ductility key issue in the manufacturability of structural components. Main limitation in the adoptability of Mg and Mg alloys in automotive and aerospace sectors. (T.M. Pollock, *Science* **328**, 986-987 (2010))

Courtesy: https://www.audi-technology-portal.de/en/body
Current state of art: Hybrid Steel and Aluminum construction

S. Sandlöbes et al. Scientific Reports 7, 10458 (2017).

Ductility enhancement of Magnesium using solutes

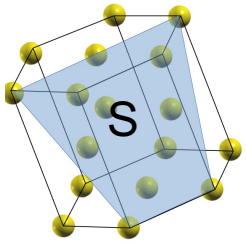




<c+a> edge and screw Pyramidal II (glissile)

(cross-slip of screw segment)

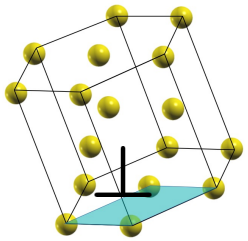
ΔG_{XS} ~ 1 eV



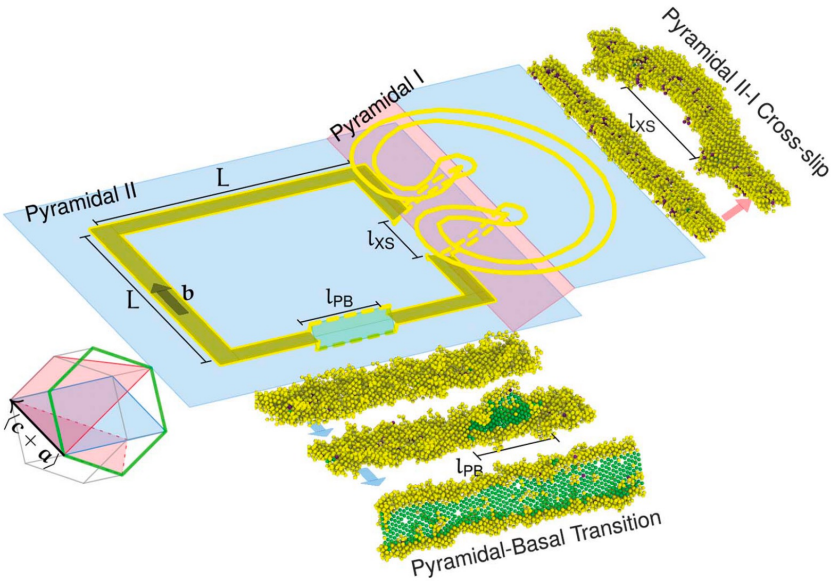
<c+a> screw Pyramidal I (glissile)

ΔG_{PB} ~ 0.5 eV

(edge segment)



Basal dissociated <c+a> edge (sessile)



(From Wu et al. Sci. 2018)

- In pure Magnesium: $\Delta GXS > \Delta GPB$

$v_0 L/l_{PB} \exp(-\Delta GPB/kT) > v_0 L/l_{XS} \exp(-\Delta G_{XS}/kT)$

L: length of screw segment

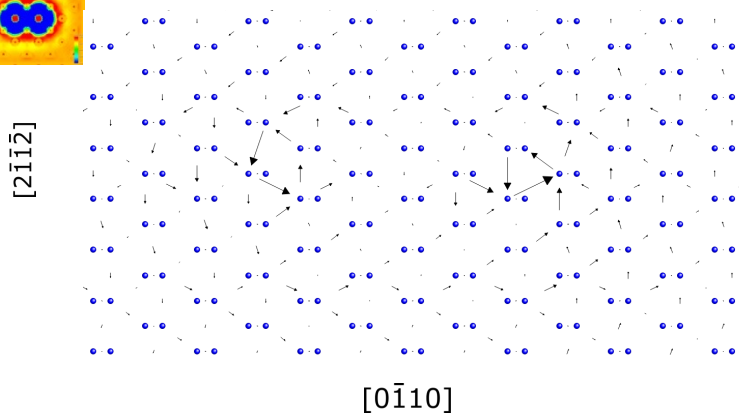
l_{XS}: critical cross-slip nucleation length (> 10 nm)

l_{PB}: critical pyramidal to basal transformation nucleation length (~2 nm)

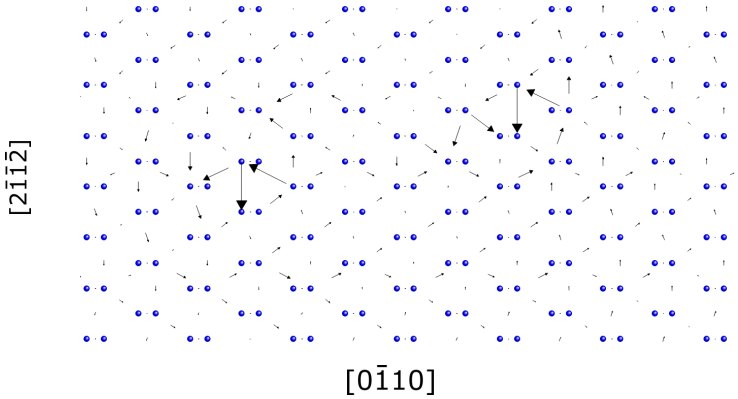
High ductility can be achieved by dilute solute additions to enable ΔGXS (c) < ΔGPB

$$v_0 L/l_{XS} \exp(-\Delta G_{XS}(c)/kT) >> v_0 L/l_{PB} \exp(-\Delta GPB/kT)$$

Pyramidal dislocation core energetics in Mg

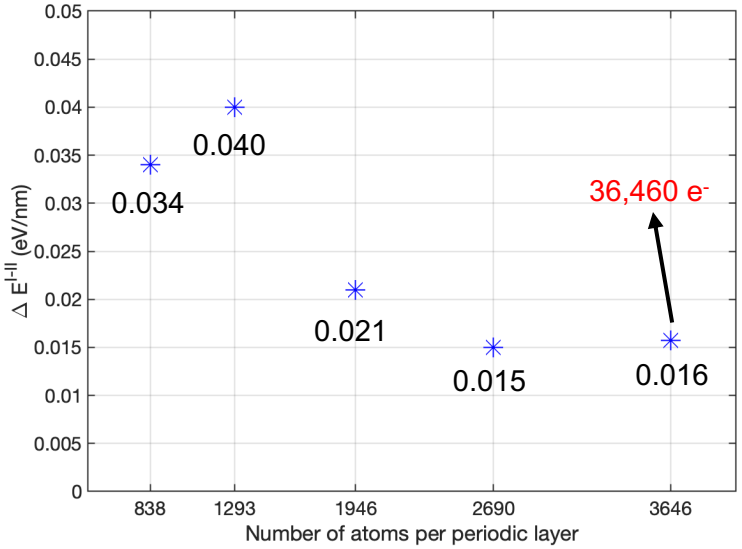


Pyramidal II dissociated <c+a> screw DD plot



Pyramidal I dissociated <c+a> screw DD plot

- Discretization accuracy for all calculations (<0.1 mHa/atom energy and <0.1 mHa/Bohr in forces)
- Ionic forces relaxed to under 0.2 mHa/Bohr
- 12 k-points along the dislocation line



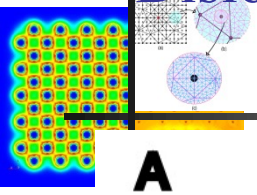
1-2 directions along [0$\bar{1}$10] – [2$\bar{1}\bar{1}$2]

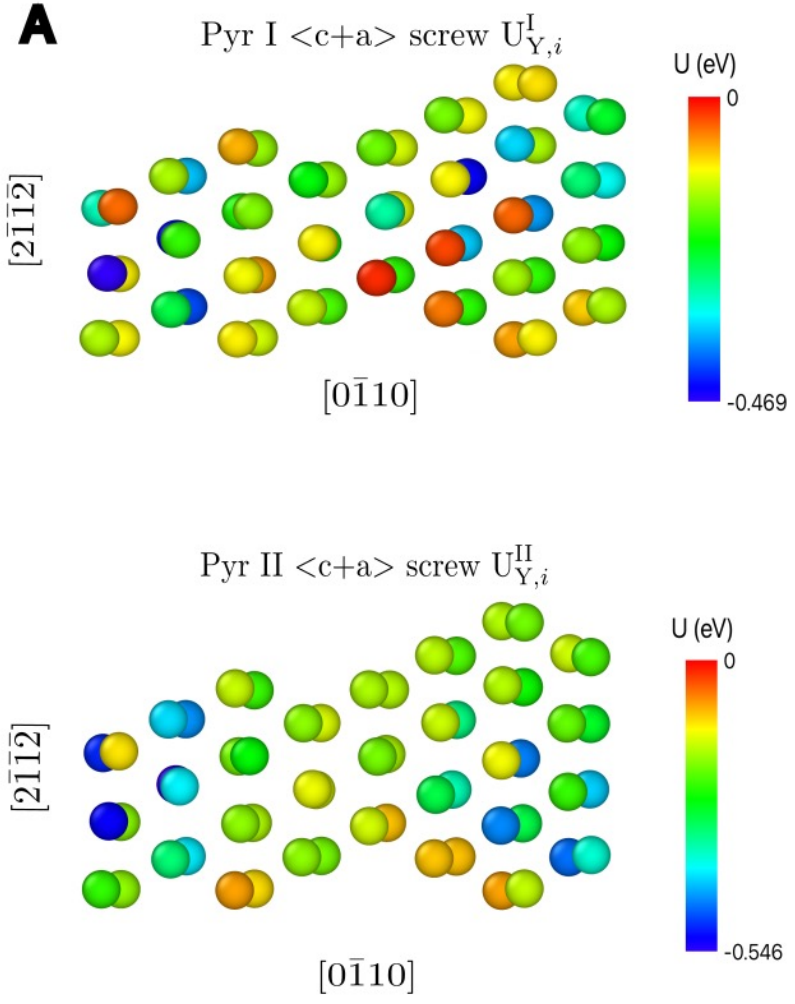
Normal to Pyr II plane

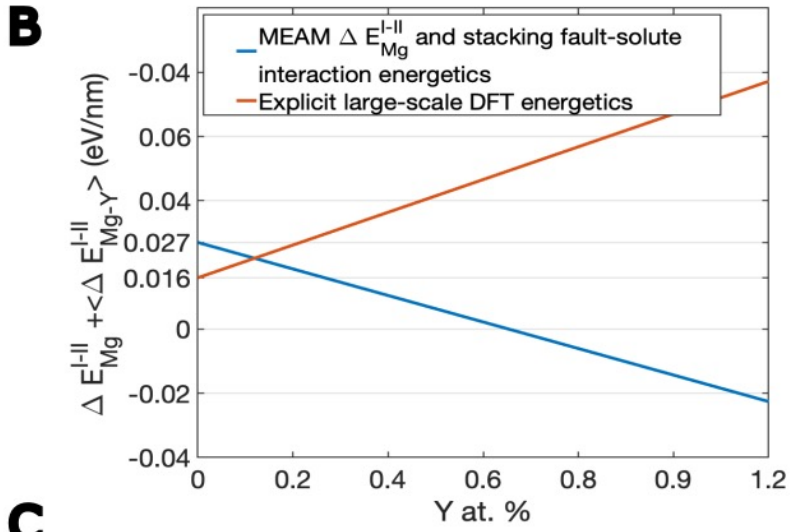
Uniaxial non-glide Strains	$\Delta E^{I-II}(\epsilon)$ Slope (eV/nm)
ε_{11}	-1.450
ε_{22}	-3.507
ε_{33}	0.301

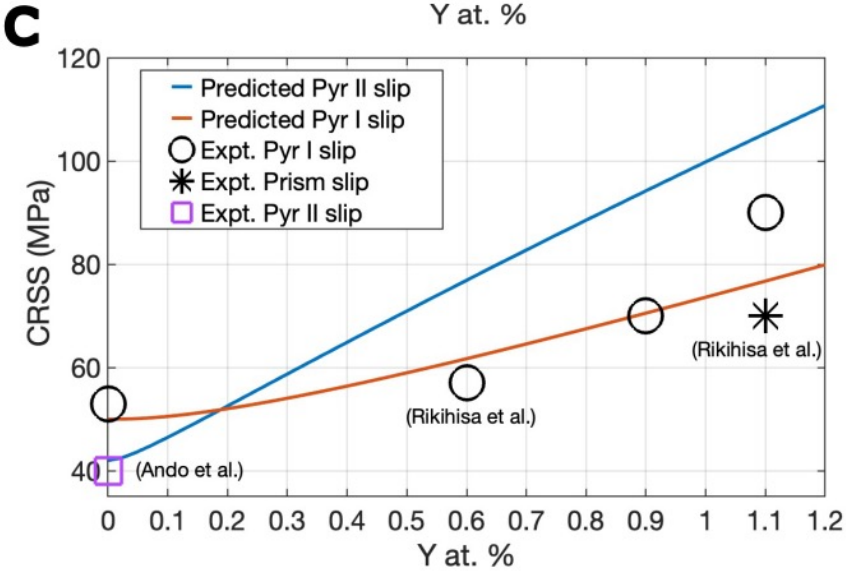
- **Cell size effects in ΔE^{I-II} till ~3000 atoms (30,000 electrons). Converged ΔE^{I-II} =16 meV/nm**

Dislocation-solute interaction energetics for Mg-Y alloy and solute strengthening predictions







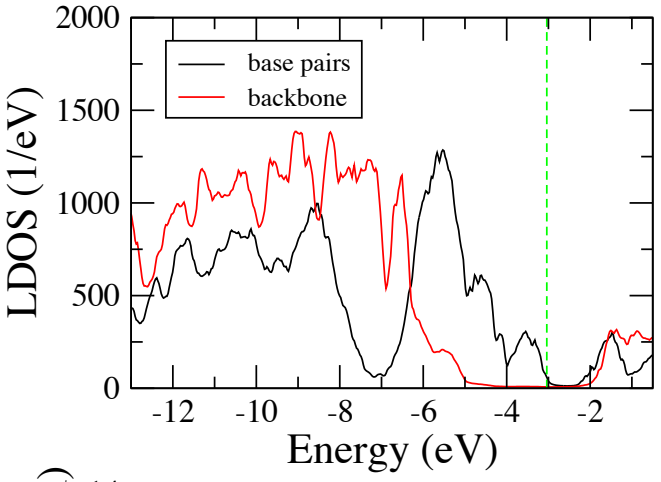


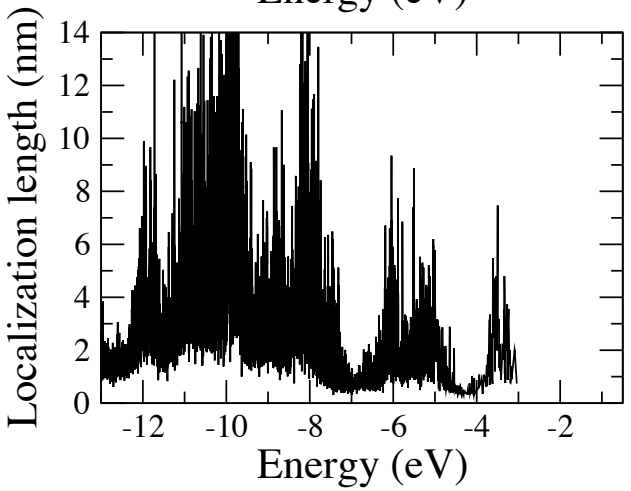
- **Calculations on Zn and Ca solutes ongoing**

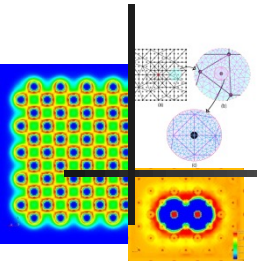
Other Application Studies using DFT-FE

- Understanding electron transport in DNA molecules (*Nature Nanotechnology 15 836 (2020)*)
  - Large-scale simulations involving 100 basis pairs (~6,200 atoms) simulating experiments
  - Provided new insights into the role of backbone in electron charge transport

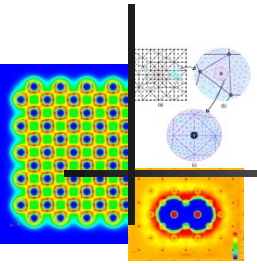
- Spin-spin interactions in defects in solids (*npj Computational Materials, 50 (2021); Phys. Rev. Mat. 3 043801 (2019)*)
  - Computed spin Hamiltonian parameters that describe electron-electron and electron-nuclear spin interactions
  - Systematically convergent calculations with all-electron accuracy, possible for the first time
  - Use mixed pseudopotential and all-electron calculations leveraging the flexibility of the DFT-FE framework





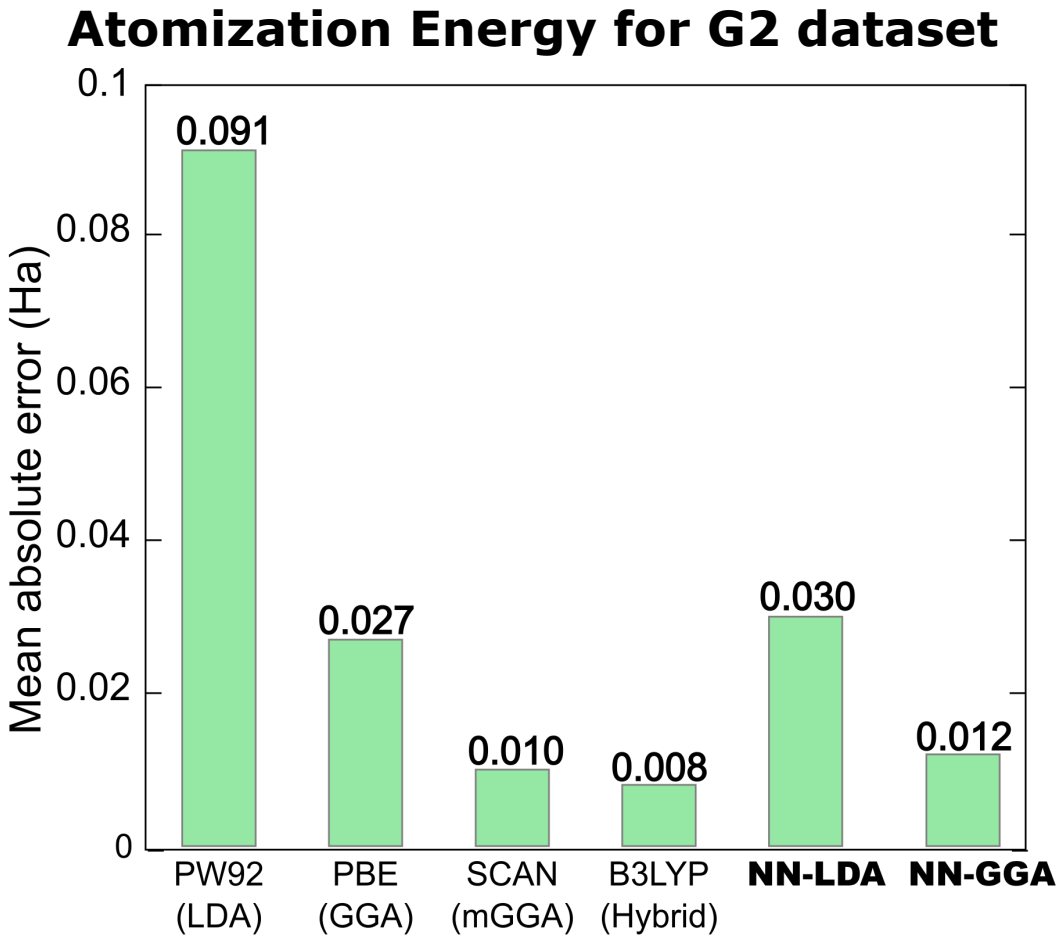


THANK YOU!



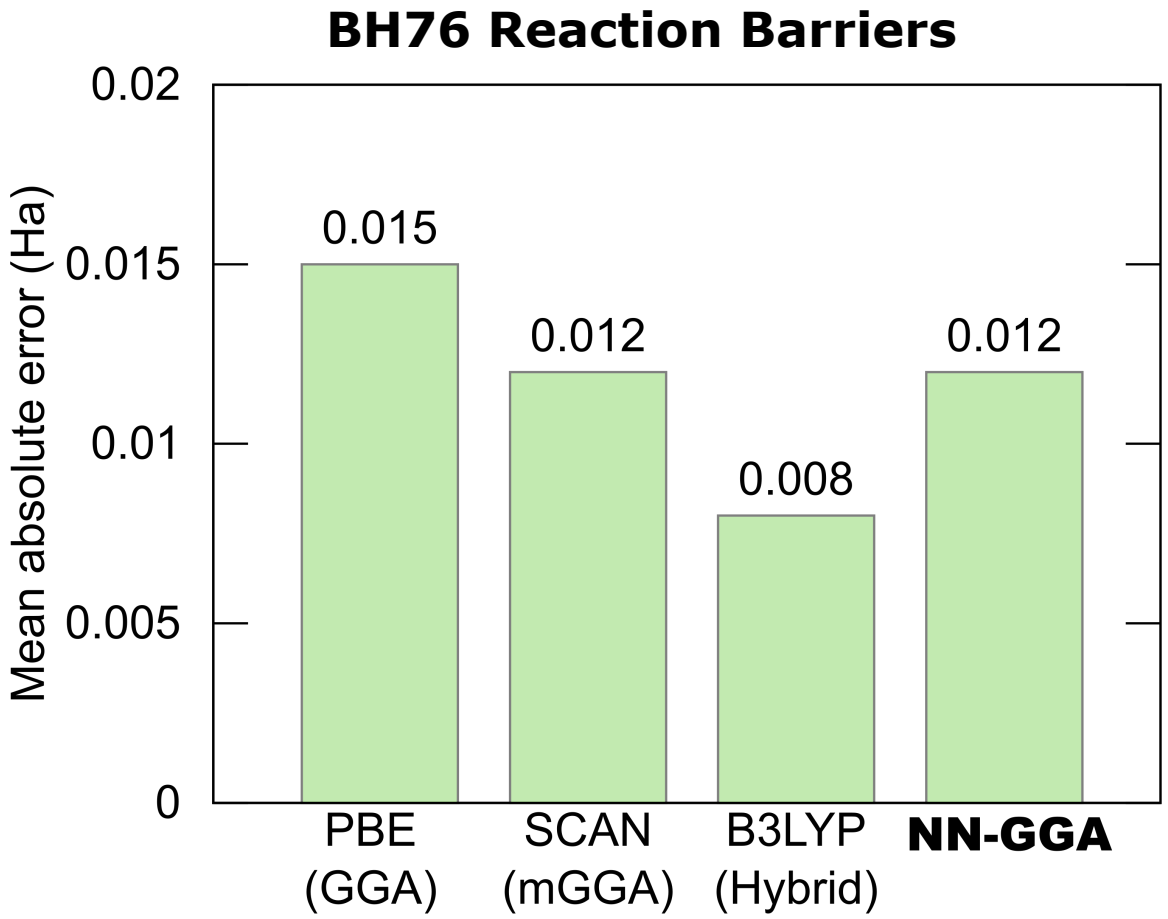
Backup Slides

ML-XC



- NN-GGA achieves close to SCAN accuracy
- ML-XC at a lower rung achieves higher-rung accuracy!

ML-XC



- NN-GGA achieves SCAN accuracy
- ML-XC at a lower rung achieves higher-rung accuracy!