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

Chemical properties of nanoparticles

Quasicrystals

Drug design

Defects in Materials

Aarons et al. JCP (2016)

Sadybekov et al. Nature (2023)

Quantum Mechanics

- Schrödinger equation: \[ H \psi = E \psi \]

\[
H = -\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 - \frac{1}{2} \sum_{A=1}^{M} \frac{1}{M_A} \nabla_A^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{|r_i - R_A|} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_i - r_j|} + \sum_{A=1}^{M} \sum_{B=1}^{M} \frac{Z_A Z_B}{|R_A - R_B|}
\]

\[
\psi = \psi(x_1, x_2, \ldots, x_N, R_1, R_2, \ldots, R_M)
\]

- Computational complexity: \( \psi \in \mathbb{R}^{3N} \) !!

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

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

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

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

- \( V_{xc} = \frac{\delta E_{xc}(\rho)}{\delta \rho} \); \( V_{xc}(\rho), E_{xc}(\rho) \) → Universal functionals

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

- Level 1: DFT-LDA
- Level 2: DFT-GGA
- Level 3: DFT-Hybrid
- QM Many-body Methods

Overview of the approach

QM Many Body Physics

\[ \hat{H}_{QMB} \Psi_{QMB} = E_0 \Psi_{QMB} \]

Configuration Interaction (CI)

\[ \Psi_{QMB} = \sum_i c_i \Phi_i \]

\[ \Psi_{QMB} \in \mathbb{R}^{3N} \]
Overview of the approach

QM Many Body Physics

\[ \hat{H}_{QMB} \Psi_{QMB} = E_0 \Psi_{QMB} \]

Configuration Interaction (CI)

\[ \Psi_{QMB} = \sum_i c_i \Phi_i \quad \in \mathbb{R}^{3N} \]

\[ \rho_{QMB}(r) \]

Inverse DFT

\[ \min_{\rho_{KS}} \left\| \rho_{QMB} - \rho_{KS} \right\|^2 \]

\[ \left( \frac{-1}{2} \nabla^2 + v_N(r) + v_H(r) + v_{xc}(r) \right) \psi_{KS}(r) = c_i \phi_i^{KS}(r) \]

All-electron (AE) DFT

\[ \psi_{KS}(r) \]

Continuous Adaptive FE

\[ C^0 \]

Oscillatory Orbitals

\[ v_N(r) \]

Singular Potential

Accurate Asymptotics

Cusp

No Cusp
Overview of the approach

QM Many Body Physics
\[ \hat{H}_{QMB} \Psi_{QMB} = E_0 \Psi_{QMB} \]
Configuration Interaction (CI)
\[ \Psi_{QMB} = \sum_i c_i \Phi_i \quad \in \mathbb{R}^{3N} \]

Inverse DFT
\[ \min_{\psi_{xc}} \left\| \rho_{QMB} - \rho_{KS} \right\|^2 \]
\[ \left( -\frac{1}{2} \nabla^2 + v_N(r) + v_H(r) + v_{xc}(r) \right) \psi_{KS}^i(r) = c_i \psi_{KS}^i(r) \]
\[ \rho_{xc}(r), \quad E_{xc} \]

Machine Learning
\[ E_{xc}^{ML} = \int \psi_{xc}^{ML}(\rho)(r) dr \]
\[ \rho_{xc}[\rho] = \int \frac{\partial E_{xc}[\rho(r)]}{\partial \rho(r)} dr' \]

Physics Informed Neural Network
- Translational Equivariance
- Rotational Equivariance
- Spin Scaling
- Coordinate Scaling

\[ \mathcal{O}(eN) \]

All-electron (AE) DFT
\[ \psi_{KS}(r) \]
Oscillatory Orbitals
\[ v_N(r) \]
Singular Potential

Continuous Adaptive FE

\[ C^0 \]
Overview of the approach

**QM Many Body Physics**

\( \hat{H}_{QBMB} \Psi_{QBMB} = E_0 \Psi_{QBMB} \)

Configuration Interaction (CI)

\( \Psi_{QBMB} = \sum_{i} c_i \Phi_i \)

\( \rho_{QBMB}(r) \)

All-electron (AE) DFT

Oscillatory Orbitals

\( V_N(r) \) Singular Potential

Inverse DFT

\[ \min_{\rho_{QBMB}} ||\rho_{QBMB} - \rho_{KS}||^2 \]

\[ \left( \frac{1}{2} \nabla^2 + V_N(r) + V_H(r) + V_{xc}(r) \right) \psi_{KS}(r) = \epsilon \psi_{KS}(r) \]

Accurate Asymptotics

Cusp

Continuous Adaptive FE

**Machine Learning**

\( E_{xc}^{ML} = \int \rho_{xc}^{ML}[\rho](r) dr \)

\( \rho_{xc}^{ML}[\rho] = \int \frac{\partial V_{xc}[\rho](r')}{\partial \rho(r)} dr' \)

\( \rho_{xc}^{ML}[\rho] \)

\( C_0 \) Physics Informed Neural Network

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

**Large-scale DFT**

Kohn-Sham Non-linear Eigenvalue Problem

\[ \left( \frac{1}{2} \nabla^2 + V_N(r) + V_H[\rho](r) + V_{xc}[\rho](r) \right) \psi_{KS}(r) = \epsilon \psi_{KS}(r) \]

\( \rho(r) = \sum_{i} f_i |\psi_{KS}(r)|^2 \quad (N \sim O(10^3)) \)

**Chebyshev Accelerated Eigensolver**

Wanted Spectrum

Unwanted Spectrum

\( \hat{H} = c_1 \hat{H} + c_2 \)

Chebyshev Filter

\( \Psi_F = T_m(H) \Psi \)

\( T_m(H)X = [2HT_{m+1}(H) - T_{m-1}(H)]X \)

- 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

QM Many Body Physics

\[ \hat{H}_{\text{QMB}} \Psi_{\text{QMB}} = E_0 \Psi_{\text{QMB}} \]

Configuration Interaction (CI)

\[ \Psi_{\text{QMB}} = \sum_{i} c_i \Phi_i \]

Inverse DFT

\[ \min_{\rho_{\text{QMB}}} \| \rho_{\text{QMB}} - \rho_{\text{KS}} \|^2 \]

\[ \left( \frac{1}{2} \nabla^2 + v_N(r) + v_H(r) + v_{xc}(r) \right) \psi_{\text{KS}}(r) = \varepsilon \psi_{\text{KS}}(r) \]

Machine Learning

\[ E_{\text{ML}}^\text{ML} = \int_{E_{\text{xc}}^\text{ML}} [\rho](r) dr \]

\[ \rho_{\text{ML}}(r) = \int \frac{\partial v_{\text{xc}}}(\rho(r')) dr' \]

Physics Informed Neural Network

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

Large-scale DFT

Kohn-Sham Non-linear Eigenvalue Problem

\[ \left( -\frac{1}{2} \nabla^2 + v_N(r) + v_H[\rho](r) + v_{\text{ML}}(\rho)(r) \right) \psi_{\text{KS}}(r) = \varepsilon \psi_{\text{KS}}(r) \]

\[ \rho(r) = \sum_i f_i \psi_i^2(r) \quad (N \sim O(10^3)) \]

Continuous Adaptive FE

Exascale Computational Framework

Accelerated Discrete Hamiltonian Multiplication through FE Cell Level Dense Linear Algebra

\[ Y = HX \equiv \text{Assembly}_{\text{FE}} \{ \Phi_i, X_{\Phi_i} \} \]

\[ X \rightarrow M \times N \text{ Dense Matrix} \]

\[ H \rightarrow m_c \times m_c \text{ Dense matrices} \]

\[ m_c \approx 5^2 \text{ to } 3^3 \text{ (Higher order FE)} \]

Chebyshev Accelerated Eigensolver

\[ \hat{H} = c_1 \hat{H} + c_2 \]

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

Chebyshev Filter

\[ \Psi_F = T_m(H) \Psi \]

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

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

Rolled Strided Batched GEMM over all cells

- Non-blocking MPI Point-to-Point

Optimal Memory bandwidth strategies

- Coalesced memory access
- Fused GPU kernels

Overlapping of Compute and Communication

- Data movement/communication for Block \(k\)
- Compute for Block \((k+1)\)
- Data movement/communication for Block \((k+1)\)

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

- FP32 Arithmetic
- FP64 Arithmetic
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.

\[
\Psi(r_1, r_2, \ldots, r_{N_e})
\]

Many-body wavefunction

\[
\rho(r) \leftrightarrow v_{xc}[\rho(r)] \rightarrow v_{xc}(r)
\]


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

Kanungo et. al., *JPCL* **14**, 10039 (2023)
Inverse DFT – Key Ideas

- PDE constrained optimization

\[
\arg \min_{\rho_{\text{data}}(\mathbf{r})} \int w(\mathbf{r}) (\rho_{\text{data}}(\mathbf{r}) - \rho(\mathbf{r}))^2 \, d\mathbf{r}
\]

subject to

\[
\left(-\frac{1}{2} \nabla^2 + v_{\text{ext}}(\mathbf{r}) + v_H(\mathbf{r}) + v_{\text{xc}}(\mathbf{r})\right) \psi_i = \epsilon_i \psi_i, \quad \int |\psi_i(\mathbf{r})|^2 \, d\mathbf{r} = 1.
\]

- Higher-order FE basis for discretization – ensures completeness

- Cusp correction:

  \[
  \Delta \rho(\mathbf{r}) = \rho_{\text{FE}}^{\text{DFT}}(\mathbf{r}) - \rho_{\text{G}}^{\text{DFT}}(\mathbf{r})
  \]

- Far-field asymptotics: Start with a guess for \( V_{\text{xc}}(\mathbf{r}) \) with correct far-field asymptotics and use homogeneous Dirichlet boundary conditions on the adjoint fields.
Inverse DFT using highly accurate CI densities

Materials system C₆H₄ (ortho-Benzene) – 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 3.5 hours (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 \( E_{xc}[\rho] = \int e_{xc}[\rho](r)dr \) where \( e_{xc}[\rho](r)dr \): energy density

- \( e_{xc}^{ML} = \rho^{4/3}(r)\phi(r)F^{DNN}(\rho, \xi, s) \)

  - \( \xi(r) = \frac{\rho_{\uparrow}(r) - \rho_{\downarrow}(r)}{\rho(r)} \)
  - \( \phi(r) = \frac{1}{2} \left((1 + \xi(r))^{4/3} + (1 - \xi(r))^{4/3}\right) \)
  - \( s(r) = \frac{(3\pi^2)^{1/3}|\nabla \rho(r)|}{2\rho^{4/3}(r)} \)

- Satisfied translational and rotational equivariance; Enforces known coordinate- and spin-scaling relations

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

- \( v_{xc}^{ML}(r) = \frac{\delta E_{xc}^{ML}[\rho]}{\delta \rho(r)} = \int \frac{\delta e_{xc}^{ML}(r')}{\delta \rho(r)}dr' \)

- Learning is on \( e_{xc}[\rho](r) \), but optimization is on \( v_{xc}(r) \)
Accurate ML-XC

- Trained on ‘exact XC potentials’ from invDFT and XC energy of 3 atoms (Li, N, Ne) and 2 molecules (H₂, LiH).
- The MAE in energy per atom from G2 thermochemistry data.

![Bar chart showing error in total energy per atom for different methods: PW92 (LDA), PBE (GGA), B3LYP (Hybrid), MLXC. The PW92 (LDA) has an error of 0.317 Ha, PBE (GGA) 0.039 Ha, B3LYP (Hybrid) 0.012 Ha, and MLXC 0.007 Ha.](image)
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:
    \[ e_{Xc}(r) = e_{Xc}^{ML}[\rho, |\nabla\rho|, \nabla^2\rho](r) \]
  - Kinetic-energy meta-GGA:
    \[ e_{Xc}(r) = e_{Xc}^{ML}[\rho, |\nabla\rho|, \tau](r) \]
    \[ \tau(r) = \frac{1}{2} \sum_i |\nabla\psi_i(r)|^2 \] (Kinetic energy density)
  - Fractional derivatives based non-local models
    \[ e_{Xc}(r) = e_{Xc}^{ML}[\rho, |\nabla|\rho, D^{\alpha_1}\rho, D^{\alpha_2}\rho, \ldots](r) \]
    \[ D^\alpha \rho(r) = C(\alpha) \int \frac{\rho(r') - \rho(r')}{|r-r'|^{3+2\alpha}} dr' \] (Fractional Laplacian)
    - Naturally encapsulates non-local features
    - Modulation of range of non-locality and decay-rate via different \(\alpha\)'s
Large-scale DFT calculations
DFT – Finite Element discretization

Use finite-element basis for computing –

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

Features of FE basis

- 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**


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

\[ \widetilde{H} = c_1 \tilde{H} + c_2 \]

**Chebyshev Filtering:** \( T_m(\tilde{H}) \widetilde{\Psi} = \widetilde{\Psi}_F \)

\[ T_m(\tilde{H})X = [2\tilde{H}T_{m-1}(\tilde{H}) - T_{m-2}(\tilde{H})]X \]
Numerical algorithm

1. Start with initial guess for electron density $\rho_{in}^h(r) = \rho_0(r)$ and the initial wavefunctions

2. Compute the discrete Hamiltonian $\tilde{H}$ using the input electron density $\rho_{in}^h$

3. **CF:** Chebyshev filtering: $\tilde{\Psi}_F = T_m(\tilde{H})\tilde{\Psi}$

4. **Orthonormalize** CF basis: $\tilde{\Psi}_F \rightarrow \tilde{\Psi}_F^o$

5. **Rayleigh-Ritz procedure:**
   - Compute projected Hamiltonian: $\hat{H} = \tilde{\Psi}_F^o \tilde{H} \tilde{\Psi}_F^o$
   - Diagonalize $\hat{H}$: $\hat{H}Q = QD$
   - Subspace rotation: $\tilde{\Psi}^R = \tilde{\Psi}_F^o Q$

6. Compute electron density $\rho_{out}^h(r) = 2 \sum_{i=1}^{N} f(e_i^h, \mu) |\psi_i^h(r)|^2$

7. If $||\rho_{out}^h(r) - \rho_{in}^h(r)|| < tol$, EXIT; else, compute new $\rho_{in}^h$ using a mixing scheme and go to (2).
Chebyshev Filtering

\[ Y = H X \]

\[ H \rightarrow \text{Sparse Matrix} \ (M \times M) \]
\[ X \rightarrow \text{Dense Matrix} \ (M \times N) \]
\[ Y \rightarrow \text{Dense Matrix} \ (M \times N) \]

\[ \begin{bmatrix} H_{c_i} \end{bmatrix}_{m_c \times m_c} \begin{bmatrix} X_{b,c_i}^b \end{bmatrix}_{m_c \times B_f} \]

\[ N_{cell} : \text{Number of FE cells} \]

FE Cell \( c_i \)
Atomic operations to avoid race conditions in addition

\[ Y^b = \text{ASSEMBLY}\{Y^b_{loc}\} \]

Assembly across processor boundaries: Communication in FP32

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

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

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-)

![Graph showing performance of Chebyshev filtering](image)

- **85.7%** using Perlmutter with FP64 tensor cores
- **56.3%** using Summit with 600 nodes
- **41.1%** using Crusher with 160 nodes

Legend:
- *Summit 600 nodes*
- *Perlmutter 400 nodes*
- *Crusher 160 nodes*
Orthogonalization: Cholesky Gram-Schmidt

- Cholesky factorization of the overlap matrix: $S = \tilde{\Psi}_F^\dagger \tilde{\Psi}_F = LL^\dagger. O(MN^2)$

- Orthonormal basis construction: $\tilde{\Psi}_F^\circ = \tilde{\Psi}_F L^{-1}\dagger. O(MN^2)$

**Mixed precision computation for Chol-GS**

1. $S = DP \{S_d\} + SP \{S_{od}\}$

2. $S = LL^\dagger$ in double precision.

3. Orthonormal basis construction:

   $\tilde{\Psi}_F^\circ = DP \{\tilde{\Psi}_F L_d^{-1}\dagger\} + SP \{\tilde{\Psi}_F L_{od}^{-1}\dagger\}$

**Blocked approach to reduce peak memory**

- $\tilde{\Psi}_F^\dagger \tilde{\Psi}_F = M_{loc}$
- Copy block to CPU (if computation performed on GPU)
- MPI_Allreduce
- Fill ScaLAPACK parallelized $S$ matrix
Overlapping compute and data movement on GPUs

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

Block \(i-1\)
- Compute
  - Copy
    - MPI call

Block \(i\)
- MPI call
  - deviceEventSynchronize
- Compute
  - Copy
    - deviceStreamWaitEvent

Block \(i+1\)
- Compute
  - Copy
    - MPI call
Rayleigh-Ritz procedure

- Compute projected Hamiltonian: \( \hat{H} = \tilde{\Psi}^o_F \bar{H} \tilde{\Psi}^o_F. \mathcal{O}(MN^2) \)
- Diagonalization of \( \hat{H} \): \( \hat{H} Q = Q \Lambda. \mathcal{O}(N^3) \)
- Subspace rotation step: \( \tilde{\Psi}^R = \tilde{\Psi}^o_F Q. \mathcal{O}(MN^2) \)

Mixed precision computation for RR

Compute projected Hamiltonian: \( f(\epsilon, \mu) \)

\[
\rho^h_{\text{out}}(x) = 2 \sum_{i=1}^{N} f(\epsilon^h_i, \mu) |\psi_i^h(x)|^2
\]

\[
\tilde{\Psi}^o_F = \begin{bmatrix} \tilde{\Psi}^o_{\text{oc}} & \tilde{\Psi}^o_{\text{fr}} \end{bmatrix}
\]

\[
\begin{bmatrix} \hat{H}_{\text{oc-oc}} & \hat{H}_{\text{oc-fr}} \\ \hat{H}_{\text{fr-oc}} & \hat{H}_{\text{fr-fr}} \end{bmatrix} = \begin{bmatrix} \text{SP} \left\{ \tilde{\Psi}^o_{\text{oc}} \bar{H} \tilde{\Psi}^o_{\text{oc}} \right\} & \text{SP} \left\{ \tilde{\Psi}^o_{\text{oc}} \bar{H} \tilde{\Psi}^o_{\text{fr}} \right\} \\ \text{SP} \left\{ \tilde{\Psi}^o_{\text{fr}} \bar{H} \tilde{\Psi}^o_{\text{oc}} \right\} & \text{DP} \left\{ \tilde{\Psi}^o_{\text{fr}} \bar{H} \tilde{\Psi}^o_{\text{fr}} \right\} \end{bmatrix}
\]

Subspace rotation step: \( \tilde{\Psi}^R = \text{DP} \left[ \tilde{\Psi}^o_F Q_d \right] + \text{SP} \left[ \tilde{\Psi}^o_F Q_{od} \right] \)
### Accuracy and robustness of mixed precision computations

<table>
<thead>
<tr>
<th></th>
<th>I. Cu nanoparticle</th>
<th>II. Mo periodic supercell w/ vacancy</th>
<th>III. Mg periodic supercell w/ vacancy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Energy error</strong></td>
<td>5 x 10^{-12}</td>
<td>7 x 10^{-12}</td>
<td>3 x 10^{-12}</td>
</tr>
<tr>
<td><strong>Max force error</strong></td>
<td>3 x 10^{-6}</td>
<td>7 x 10^{-7}</td>
<td>2 x 10^{-6}</td>
</tr>
<tr>
<td><strong>Total SCFs</strong></td>
<td>(46,46)</td>
<td>(30,30)</td>
<td>(18,18)</td>
</tr>
</tbody>
</table>
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-)

The graph shows the wall time per SCF step for different numbers of Summit nodes. The x-axis represents the number of nodes, ranging from 240 to 1920. The y-axis represents the wall time per SCF step in seconds.

- **FP64**
  - Number of nodes: 240, Wall time: T: 114.8 sec
  - Number of nodes: 480, Wall time: T: 94.8 sec
  - Number of nodes: 960, Wall time: T: 71.8 sec
  - Number of nodes: 1920, Wall time: T: 40 sec

- **FP64+overlap compute & commun.**
  - Number of nodes: 240, Wall time: T: 97.1 sec
  - Number of nodes: 480, Wall time: E: 79.9, T: 57.5 sec
  - Number of nodes: 960, Wall time: E: 84.4, T: 54.4 sec
  - Number of nodes: 1920, Wall time: E: 43.3, T: 28 sec

- **FP32-FP64+overlap compute & commun.**
  - Number of nodes: 240, Wall time: 101.7 sec
  - Number of nodes: 480, Wall time: E: 64.4, T: 29.3 sec
  - Number of nodes: 960, Wall time: E: 80.9, T: 21.8 sec
  - Number of nodes: 1920, Wall time: E: 54.4, T: 21.8 sec

The graph also shows the strong-parallel scaling efficiency (E) as a percentage.
Strong-scaling: Yb-Cd quasicrystal (~40,000 e-)

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

<table>
<thead>
<tr>
<th>Initialization (sec)</th>
<th>Total SCF (sec)</th>
<th>Total run (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>2023 (34 SCFs)</td>
<td>2092</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>System</th>
<th>Wall-time (sec)</th>
<th>FLOP count (PFLOP)</th>
<th>Performance (PFLOPS)</th>
<th>% of FP64 peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>TwinDislocMgY (A)</td>
<td>223</td>
<td>50,457</td>
<td>226.3</td>
<td>49.3%</td>
</tr>
<tr>
<td>TwinDislocMgY (B)</td>
<td>513.7</td>
<td>338,863</td>
<td>659.7</td>
<td>43.1%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Step</th>
<th>Wall-time (sec)</th>
<th>FLOP count (PFLOP)</th>
<th>Performance (PFLOPS)</th>
<th>% of FP64 peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>135.4</td>
<td>57,810</td>
<td>427</td>
<td>27.9%</td>
</tr>
<tr>
<td>CholGS-S-MP</td>
<td>79.3</td>
<td>54,430</td>
<td>686.4</td>
<td>44.9%</td>
</tr>
<tr>
<td>CholGS-CI</td>
<td>8.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CholGS-O-MP</td>
<td>49.6</td>
<td>54,430</td>
<td>1097.4</td>
<td>71.7%</td>
</tr>
<tr>
<td>RR-P-MP</td>
<td>66.7</td>
<td>61,036</td>
<td>915.1</td>
<td>59.8%</td>
</tr>
<tr>
<td>RR-D</td>
<td>22.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RR-SR-MP</td>
<td>93.5</td>
<td>108,858</td>
<td>1164.3</td>
<td>76.1%</td>
</tr>
<tr>
<td>DC</td>
<td>4.3</td>
<td>2303</td>
<td>535.5</td>
<td>35%</td>
</tr>
<tr>
<td>Others</td>
<td>53.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Single SCF Total</strong></td>
<td><strong>513.7</strong></td>
<td><strong>338,863</strong></td>
<td><strong>659.7</strong></td>
<td><strong>43.1%</strong></td>
</tr>
</tbody>
</table>
DFT-FE open source code
https://github.com/dftfeDevelopers/dftfe;

- 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
  - 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
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

Brittle Mg (pure)

Ductility enhancement of Magnesium using solutes

In pure Magnesium: $\Delta G_{XS} > \Delta G_{PB}$

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

$L$: 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 $\Delta G_{XS}(c) < \Delta G_{PB}$

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

(From Wu et al. Sci. 2018)
Pyramidal dislocation core energetics in Mg

- 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

<table>
<thead>
<tr>
<th>Uniaxial non-glide Strains</th>
<th>$\Delta E_{I-II}(\epsilon)$ Slope (eV/nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_{11}$</td>
<td>-1.450</td>
</tr>
<tr>
<td>$\varepsilon_{22}$</td>
<td>-3.507</td>
</tr>
<tr>
<td>$\varepsilon_{33}$</td>
<td>0.301</td>
</tr>
</tbody>
</table>

- Cell size effects in $\Delta E_{I-II}$ till ~3000 atoms (30,000 electrons). Converged $\Delta E_{I-II} = 16$ 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

Atomization Energy for G2 dataset

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

BH76 Reaction Barriers

Mean absolute error (Ha)

- PBE (GGA) 0.015
- SCAN (mGGA) 0.012
- B3LYP (Hybrid) 0.008
- NN-GGA 0.012

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