Fast, scalable and accurate finite-element based *ab initio* calculations using mixed precision computing: 46 PFLOPS simulation of a metallic dislocation system.

Sambit Das[†*], Phani Motamarri[†*], Vikram Gavini[†‡], Bruno Turcksin[§], Ying Wai Li[¶‖], Brent Leback[**]
[†]*Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA*
[‡]*Department of Materials Science & Engineering, University of Michigan, Ann Arbor, MI 48109, USA*
[§]*Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States*
[¶]*National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States*
[‖]*Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[**]*Nvidia Corporation, Santa Clara, CA 95050, USA*
[*] Sambit Das and Phani Motamarri contributed equally to this work.

***Abstract*—Accurate large-scale first principles calculations based on density functional theory (DFT) in metallic systems are prohibitively expensive due to the asymptotic cubic scaling computational complexity with number of electrons. Using algorithmic advances in employing finite-element discretization for DFT (DFT-FE) in conjunction with efficient computational methodologies and mixed precision strategies, we delay the onset of this cubic scaling by significantly reducing the computational prefactor while increasing the arithmetic intensity and lowering the data movement costs. This has enabled fast, accurate and massively parallel DFT calculations on large-scale metallic systems on both many-core and heterogeneous architectures, with time-to-solution being an order of magnitude faster than state-of-the-art plane-wave DFT codes. We demonstrate an unprecedented sustained performance of 46 PFLOPS (27.8% peak FP64 performance) on a dislocation system in Magnesium containing 105,080 electrons using 3,800 GPU nodes of Summit supercomputer, which is the highest performance to-date among DFT codes.**

I. Justification for ACM Gordon Bell Prize

Our spectral finite-element based DFT code with FP32-FP64 computing demonstrates 46 PFLOPS (27.8% FP64 peak) sustained performance on OLCF's[1] Summit, which is 3.1× higher than the current DFT record. We demonstrate strong scaling speedup of 14.5× – 26.8× on ALCF's[2] Theta, with time-to-solution 9× faster than state-of-the-art codes.

II. Performance Attributes

Category of achievement	peak performance, scalability, time-to-solution
Type of method used	N/A
Results reported on the basis of	whole application including I/O
Precision reported	mixed precision
System scale	results measured on full-scale system
Measurement mechanism	timers and FLOP count

[1]Oak Ridge Leadership Computing Facility
[2]Argonne Leadership Computing Facility

III. Overview: Accurate DFT Calculations in Metallic Systems

First principles calculations, based on quantum mechanics, have been immensely successful in predicting a wide variety of materials properties. In particular, Kohn-Sham density functional theory calculations (DFT) have provided numerous key insights into materials behavior (mechanical, chemical, electronic & optical properties), and consume a substantial fraction of world's computational resources today. The Kohn-Sham approach to DFT [1] reduces the exponential computational complexity (in number of electrons) of solving the many-electron Schrödinger equation (SE) to cubic computational complexity, and this important advance was awarded the *Nobel Prize* in Chemistry in 1998 [2]. This advance was accomplished by reducing the SE to an equivalent problem of non-interacting single electrons in an effective mean field governed by the electron-density—the probability density of finding an electron at a spatial point $\mathbf{x}$. Mathematically, the computation of ground-state electronic structure in Kohn-Sham DFT [3] involves a self consistent field (SCF) iteration solving the following non-linear eigenvalue problem (Kohn-Sham equations):

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\rho)\right)\psi_i = \epsilon_i\psi_i, \quad \rho(\mathbf{x}) = \sum_i 2f_i(\epsilon_i)|\psi_i(\mathbf{x})|^2 .$$

In the above, $\{\psi_i\}$ and $\{\epsilon_i\}$ are the single-electron Kohn-Sham orbitals (wavefunctions) and corresponding energy levels, respectively; V_{eff} is the effective single-electron potential; ρ is the electron-density; f_i denotes the orbital occupancy of the ith state, which is evaluated as a function of ϵ_i (typically, using the Fermi-Dirac distribution). As DFT is concerned with evaluating the ground-state properties, $f_i = 0$ for the unoccupied states, and, thus, the solution to the above equation entails solving the eigenvalue problem corresponding to the lowest N states (corresponding to non-trivial orbital occupancy). The resulting computational complexity is cubic-scaling with number of electrons (N_e).

The stringent accuracy requirements in DFT ($\sim \mathcal{O}(10^{-5})$ relative error in energy) needed to compute meaningful materials properties, in conjunction with the cubic-scaling computational complexity, demand significant computational resources for accurate DFT calculations, and, thus, routinely limit materials systems to at most a few thousands of electrons. Numerous efforts have been undertaken over the past two decades to tackle this challenge, which can broadly be categorized into: (i) reduced-order scaling techniques [4]–[7]; (ii) reduced-order basis techniques [8], [9]. The former relies on avoiding the direct computation of Kohn-Sham wavefunctions, and designing algorithms to directly compute the electron-density. While this has been successful in reducing the computational complexity to be close to linear scaling for materials systems that have a band-gap, these techniques are either not applicable, or are not robust and accurate for metallic systems. On the other hand, approaches relying on the use of a reduced-order basis, while efficient, do not offer systematic convergence and the desired accuracy, especially for metallic systems. Thus, despite significant method development efforts into alternate approaches, the conventional approach of solving the Kohn-Sham equations with a plane-wave discretization [10]–[12], despite all its limitations, has remained the method of choice for metallic systems.

In this work, we report a significant advance in the state-of-the-art for accurate DFT calculations in metallic systems via the development of DFT−FE [13], which is a result of algorithmic advances combined with implementation innovations, that has enabled fast, scalable and accurate large-scale DFT calculations on metallic systems as large as 100,000 electrons. This has been made possible by: (i) the development of efficient and accurate spatially adaptive discretization strategies using higher-order finite-element discretization; (ii) developing efficient and scalable algorithms in conjunction with mixed-precision strategies for the solution of Kohn-Sham equations; (iii) implementation innovations, both on many-core and hybrid architectures, that significantly reduce the data movement costs and increase arithmetic intensity. As will be demonstrated, these developments have resulted in DFT−FE providing a time-to-solution that is 9× faster than the state-of-the-art for large-scale metallic systems at similar accuracy, and a sustained performance of 46 PFLOPS that is 3.1× greater than that of any previously reported DFT code.

The reported advance in this work has wide ranging implications on a number of fields, including applied physics, chemistry, materials science, and metallurgy. There are many critical scientific and technological problems that can be tackled by availing the predictive capability of DFT calculations for large-scale metallic systems, made possible by DFT−FE. These include, to name a few: (i) studying catalytic properties of nanoparticles [14], whose sizes are beyond those that are currently accessible, to accelerate research in catalysis; (ii) designing efficient solid-state electrolytes with high ionic conductivity and interfacial stability [15], which require large-scale and long time-scale molecular dynamics (MD) simulations; (iii) predicting the properties of high entropy alloys [16] to develop next generation materials with unprecedented properties; (iv) design of light-weight structural alloys [17]. In this submission, we focus on demonstrating the critical capability advance made possible by DFT−FE that has the potential to provide a breakthrough in the design of light-weight structural alloys. It is expected that every 10% reduction in the weight of a vehicle will result in a 6-8% increase in the fuel efficiency, which has significant implications to economic savings and reducing the carbon footprint. Magnesium (Mg) being the lightest structural metal is an ideal candidate [18], but a technological solution has thus far been elusive due to the lack of ductility of Mg (needed to form parts). Dislocations are line defects, whose energetics control the ductility in a crystalline material. In particular, in Mg, the energetics and stability of a particular type of dislocation, namely the pyramidal dislocation, has been identified as critical to improving ductility in Mg [17], [19]. Understanding its energetics require highly accurate DFT calculations on large-scale systems involving many thousands of atoms, which have been out of reach thus far. We demonstrate here that DFT−FE is capable of tackling this outstanding challenge by demonstrating calculations on pyramidal dislocations in Mg, with system-sizes reaching ∼10,000 atoms (∼100,000 electrons) with the desired high-accuracy.

IV. Current State of the Art

Given the large importance of DFT calculations, numerous efforts have been undertaken over the past three decades to develop accurate and computationally efficient approaches for solving the Kohn-Sham DFT problem. These include efficient discretization schemes for the Kohn-Sham problem, development of efficient and reduced-order scaling algorithms for solving the Kohn-Sham equations, as well as efforts to improve the parallel scalability of the developed codes. Among the discretization schemes, the plane-wave (PW) basis remains to date the method of choice, especially for metallic systems, due to the systematic convergence and computational efficiency afforded by the spectral convergence of the PW basis. This has resulted in the wide adoption of the codes such as VASP [12] (commercial) and QUANTUM ESPRESSO [11] (QE) (open source) by the electronic structure community. Despite the popularity of the PW based DFT codes, the limitations of a Fourier space discretization are widely known—limited scalability on parallel computing architectures, and inefficiency in treating non-periodic systems. Significant efforts have been undertaken to improve the parallel scalability of PW based codes, such as the development of QBox [20] for *ab initio* MD simulations, which was employed by the Gordon Bell prize winning team in 2006. As will be demonstrated in this paper, despite the decades long effort into developing scalable and efficient PW based DFT codes, DFT−FE significantly outperforms both QBox and QE (state-of-the-art PW codes) by 5× in time-to-solution on a medium-sized benchmark metallic system containing 8,630 electrons, solved to a similar accuracy level. Furthermore, DFT−FE has been demonstrated to outperform QE up to 16×, even on periodic systems where

PW discretization is most efficient, for systems containing ∼30,000 electrons [13].

In order to tackle some of the limitations of the PW basis, numerous real-space techniques have been developed, of which reduced order basis functions constructed from Gaussian type orbitals [8], [9] have been the most popular. Notably, the CP2K [8] package is a massively parallel DFT code that uses such a reduced order basis. The Gordon Bell finalists in 2015 [21] demonstrated a sustained performance of 15 DP-PFLOPS using CP2K, the highest reported among DFT codes thus far. While such reduced-order basis sets can provide computational efficiency, they lack systematic convergence for generic materials systems, and, hence, are not suitable for accurate DFT calculations involving metallic systems. Furthermore, as will be demonstrated in this paper, DFT-FE attained sustained performance of 46 PFLOPS, 3.1× higher than [21].

The lack of systematic convergence using Gaussian type orbitals has led to recent efforts in developing systematically improvable, efficient and scalable real-space DFT techniques based on finite-difference [22], [23] and finite-element discretization [24], [25]. The 2011 Gordon Bell prize winning team [26] used a finite-difference technique to perform DFT calculations on a semi-conducting system achieving a sustained performance of 3.08 DP-PFLOPS on the K supercomputer, a previous generation machine. However, the simulations demonstrated were not commensurate with chemical accuracy. The 2016 Gordon Bell prize finalist [27] also employed finite-difference discretization in conjunction with a linear scaling technique to conduct large-scale first principles molecular dynamics simulations. However, this linear scaling technique is only suited for materials systems with a band-gap, and thus is not applicable for metallic systems. Reduced order scaling DFT calculations on metallic systems at chemical accuracy have recently been demonstrated using the ONETEP code [28], employing psinc basis functions and a Fermi-operator expansion type approach [29]. We note that such an approach has a very high computational prefactor for metallic systems. Furthermore, parallel scalability has not been demonstrated, thereby resulting in a large time to solution—up to 3 hours per SCF iteration on system sizes involving ∼100,000 electrons. In comparison, as will be demonstrated, DFT-FE has achieved a per SCF time of 2.4 minutes for similar system sizes.

Overall, the current state-of-the-art DFT methods for large-scale metallic systems are either limited by scalability, accuracy, or by the computational efficiency they afford. Thus, DFT calculations on metallic systems have been severely limited, and is the key bottleneck in tackling many scientific and technologically important questions across a range of fields. The development of DFT-FE that combines various strategies, algorithmic advances and implementation innovations, as discussed subsequently in Sec. V, provides an important advance for fast and accurate large-scale DFT calculations on metallic systems, and a tool to tackle a range of open questions that have not been possible heretofore.

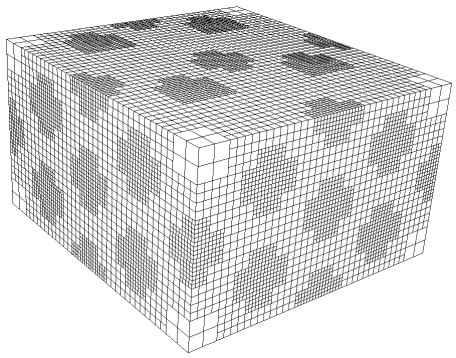

Fig. 1: Adaptive finite-element mesh with refinement around atomic positions for a periodic HCP Mg supercell with a vacancy at the corner of the simulation domain.

V. Innovations Realized

We employ the finite-element (FE) discretization of the Kohn-Sham equations to avail the systematic convergence afforded by FE basis, and to exploit its significant parallel scaling advantage in comparison to widely used basis sets like plane-waves (PW). In particular, FE basis is localized with a compact support on the FE cells containing/sharing a FE node. Thus, only the FE nodes on processor boundaries need to be communicated, which has a significantly smaller communication cost in comparison to the all-to-all communication required in global basis sets like PW.

A. Algorithmic advances in using finite-elements for DFT

Two major challenges prevented the FE basis from being widely adopted in *ab initio* calculations: (i) naively implemented FE discretization has a significant degrees of freedom (DoFs)[3] disadvantage, in comparison to PW, at chemical accuracy; (ii) FE discretization results in a generalized Hermitian eigenvalue problem (GHEP)[4], $\mathbf{H}\hat{\boldsymbol{\Psi}}_i = \epsilon_i^h \mathbf{M}\hat{\boldsymbol{\Psi}}_i$, which is more challenging to solve in comparison to a standard Hermitian eigenvalue problem (SHEP). We overcame the first challenge by employing an *error-analysis informed adaptive higher-order FE discretization* (polynomial order ≥ 4) [13] that significantly reduced the DoFs needed to achieve chemical accuracy. In particular, the spatial adaptivity is realized via an *a priori* mesh adaption scheme (cf. Fig. 1), which guides the mesh refinement based on a local FE cell error indicator obtained from the error analysis of the Kohn-Sham problem [13]. In DFT-FE, the adaptive mesh generation infrastructure is based on the p4est library [30] via the deal.II finite-element library [31]. We overcame the second challenge by employing spectral finite-elements with nodal points coincident with Gauss-Lobatto-Legendre (GLL) points [25], which in conjunction with GLL quadrature for numerical integration[5] renders $\mathbf{M}$ diagonal. Thus, we can trivially transform the GHEP to SHEP, given by: $\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_i = \epsilon_i^h \widetilde{\boldsymbol{\Psi}}_i$, where $\widetilde{\boldsymbol{\Psi}}_i = \mathbf{M}^{1/2}\hat{\boldsymbol{\Psi}}_i$ and $\widetilde{\mathbf{H}} = \mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{-1/2}$.

[3]DoFs denotes # basis functions.

[4]$\mathbf{H}$ and $\mathbf{M}$ denote the discrete sparse Hamiltonian and overlap matrices of size $M \times M$, where M denotes # FE basis. $\hat{\boldsymbol{\Psi}}_i$ denotes the discrete vector corresponding to the i^{th} wavefunction.

[5]Accuracy and sufficiency of GLL quadrature on $\mathbf{M}$ is established in [25].

The resulting discrete Kohn-Sham nonlinear eigenvalue problem is solved using a self consistent field (SCF) iteration (cf. Algorithm 1) by employing the Chebyshev filtered subspace iteration procedure (ChFSI) [25], [32] in each SCF iteration. The most dominant computational step here is ChFSI, which entails constructing a subspace that is rich in the eigenspace of the occupied single-electron states and solving the SHEP in this subspace. ChFSI has been demonstrated to be more computationally efficient ($10\times$ – $20\times$) [25] and scalable in comparison to traditional approaches computing eigenpairs explicitly. The ChFSI procedure implemented in `DFT-FE` (cf. Algorithm 2) involves the following three steps: Chebyshev filtering (CF), orthonormalization (CholGS), and the Rayleigh-Ritz procedure (RR). Below we discuss the innovations focused on improving arithmetic intensity and reducing data movement costs of CF, CholGS and RR steps.

Algorithm 1 Self Consistent Field (SCF) iteration in `DFT-FE`

M: # FE basis; N: # eigenstates; $(.)^h$: FE discretized field

1: Start with an initial guess for input electron-density $\rho^h_{\text{in}}(\mathbf{x})$, and an initial guess for $\widetilde{\mathbf{\Psi}}$.

2: [EP] Get effective potential, $V^h_{\text{eff}}(\rho^h_{\text{in}}(\mathbf{x}), \mathbf{R})$, by solving a discrete Poisson equation ($\mathcal{O}(M\log(M))$). Compute FE cell level Hamiltonian matrices of $\widetilde{\mathbf{H}}$.

3: Perform ChFSI procedure: $[\widetilde{\mathbf{\Psi}}^{\mathbf{R}}, \mathbf{D}] = \text{ChFSI}\left(\widetilde{\mathbf{\Psi}}, \widetilde{\mathbf{H}}\right)$ (call Algorithm 2).

4: [DC] Compute new output electron density, $\rho^h_{\text{out}}(\mathbf{x})$, using $\widetilde{\mathbf{\Psi}}^{\mathbf{R}}$ and $\mathbf{D}$. ($\mathcal{O}(MN)$)

5: If $\left\|\rho^h_{\text{out}}(\mathbf{x}) - \rho^h_{\text{in}}(\mathbf{x})\right\| \leq$ tolerance, *stop*; Else, compute new $\rho^h_{\text{in}}(\mathbf{x})$ using a mixing scheme, and go to step 2.

Algorithm 2 ChFSI procedure: $[\widetilde{\mathbf{\Psi}}^{\mathbf{R}}, \mathbf{D}] = \text{ChFSI}\left(\widetilde{\mathbf{\Psi}}, \widetilde{\mathbf{H}}\right)$

$\widetilde{\mathbf{\Psi}}$ is a $M \times N$ matrix

1: [CF] Chebyshev polynomial filtering of $\widetilde{\mathbf{\Psi}}$ (cf. Sec. V-C). ($\mathcal{O}(MN)$)

2: [CholGS] Orthonormalize the Chebyshev filtered basis $\widetilde{\mathbf{\Psi}}$:

a: [CholGS-S] Compute overlap matrix, $\mathbf{S} = \widetilde{\mathbf{\Psi}}^{\dagger}\widetilde{\mathbf{\Psi}}$. ($\mathcal{O}(MN^2)$)

b: [CholGS-CI] Perform Cholesky factorization of the overlap matrix, $\mathbf{S} = \mathbf{L}\mathbf{L}^{\dagger}$, and compute $\mathbf{L}^{-1}$. ($\mathcal{O}(N^3)$)

c: [CholGS-O] Construct orthonormal basis: $\widetilde{\mathbf{\Psi}}^{\mathbf{o}} = \widetilde{\mathbf{\Psi}}\mathbf{L}^{-1^{\dagger}}$. ($\mathcal{O}(MN^2)$)

3: [RR] Perform the Rayleigh-Ritz procedure:

a: [RR-P] Compute projected Hamiltonian: $\hat{\mathbf{H}} = \widetilde{\mathbf{\Psi}}^{\mathbf{o}^{\dagger}}\widetilde{\mathbf{H}}\widetilde{\mathbf{\Psi}}^{\mathbf{o}}$. ($\mathcal{O}(MN^2)$)

b: [RR-D] Diagonalization of $\hat{\mathbf{H}}$: $\hat{\mathbf{H}}\mathbf{Q} = \mathbf{Q}\mathbf{D}$. ($\mathcal{O}(N^3)$)

c: [RR-SR] Subspace rotation step: $\widetilde{\mathbf{\Psi}}^{\mathbf{R}} = \widetilde{\mathbf{\Psi}}^{\mathbf{o}}\mathbf{Q}$. ($\mathcal{O}(MN^2)$)

B. General GPU acceleration strategy

To achieve maximum performance on GPUs, we have ported all computationally intensive steps in the ChFSI procedure to GPUs. Further, the algorithm is implemented such that we minimize CPU-GPU data transfers[6], which can be a rate limiting step relative to the high arithmetic performance on the GPU. In ChFSI procedure, the operations corresponding to CF, CholGS-S, CholGS-O, RR-P and RR-SR are ported to GPUs using CUDA kernels, and cuBLAS library for the `Xgemm` operations. The CholGS-CI and RR-D steps are performed on CPUs in parallel on a subset of the MPI ranks using the ELPA library [33]. These operations could not be performed on GPUs as the $\mathcal{O}(N^2)$ memory of the overlap ($\mathbf{S}$) and projected Hamiltonian ($\hat{\mathbf{H}}$) matrices would be too large to store in serial[7] on a single GPU for very large scale problems considered in this work ($N \sim 60,000$). Further, these operations performed on CPUs do not constitute a bottleneck, even at large system sizes[8]. The aforementioned memory overflow issue is also present in the compute intensive CholGS-S, CholGS-O, RR-P RR-SR steps, but we circumvent them by using a blocked approach [13]. To elaborate the blocked approach in the context of CholGS-S, the sub-matrices of $\mathbf{S}$, corresponding to a block of wavefunctions, are computed one at a time by assembling the local contributions computed on the GPUs, and subsequently copied to the ELPA parallelized $\mathbf{S}$ on the CPUs. An important aspect of our implementation is that, by porting the aforementioned steps in ChFSI procedure to GPUs in conjunction with GPU porting of the electron-density computation (DC), we *completely eliminate* the otherwise required large data transfer of $\widetilde{\mathbf{\Psi}}$ between CPU and GPU during the SCF procedure.

C. Reduction of memory access & communication costs in CF

Chebyshev polynomial filtering procedure (CF) constitutes the action of a degree m (typically 30-50) Chebyshev polynomial filter on $\widetilde{\mathbf{\Psi}}$, computed using a recursive iteration which involves m $\widetilde{\mathbf{H}}\mathbf{X}$ evaluations. Each $\widetilde{\mathbf{H}}\mathbf{X}$ evaluation is performed using a blocked approach to reduce peak memory. To this end, blocks of B_f wavefunction vectors, denoted by $\mathbf{X_b}$, are filtered sequentially. We significantly reduce the memory access costs in the core computational kernel $\widetilde{\mathbf{H}}\mathbf{X_b}$ by employing FE cell level dense matrix operations as shown in Fig. 2, instead of global sparse matrix approaches. In particular, we recast the $\mathbf{Y_b} = \widetilde{\mathbf{H}}\mathbf{X_b}$ kernel as

$$\mathbf{Y_b} = \text{ASEMB}\left\{\widetilde{\mathbf{H}}^{c_i}\mathbf{X_b}^{c_i}\right\}, \tag{1}$$

where c_i denotes the i^{th} FE cell, $\widetilde{\mathbf{H}}^{c_i}$ and $\mathbf{X_b}^{c_i}$ are matrices with sizes $M_{\text{cell}} \times M_{\text{cell}}$ and $M_{\text{cell}} \times B_f$ respectively, with M_{cell}[9] denoting the number of DoFs in a FE cell. ASEMB denotes the assembly operation of vector contributions from all FE cells. The above FE cell level matrix operations involving many small dense matrix-matrix multiplications are performed simultaneously for all cells on GPUs using cuBLAS's `XgemmStridedBatched` routine. Furthermore,

[6]Denotes data transfer in both directions.

[7]Currently there is no GPU alternative to ScaLAPACK, but upcoming libraries like SLATE (https://icl.utk.edu/slate/) will address this gap.

[8]As reported in Table IV, these operations constitute 10.3% of the SCF wall-time for a 61,640 electrons system.

[9]We employ a polynomial degree of 4, which corresponds to $M_{\text{cell}} = 125$.

as shown in Fig. 2, we structure the memory layout of $\mathbf{X_b}$ to ensure coalesced memory access across GPU threads, thereby significantly reducing memory access costs in extracting $\mathbf{X_b^{c_i}}$ from $\mathbf{X_b}$ and in the assembly operation into $\mathbf{Y_b}$.

We minimize communication latencies and overheads in $\widetilde{\mathbf{H}}\mathbf{X_b}$ by exploiting the fact that all wavefunction vectors have identical MPI point-to-point communication pattern across the FE domain decomposition partition boundaries. This allows us to perform the MPI communication for all wavefunction vectors in $\mathbf{X_b}$ simultaneously, which incurs minimal network latency compared to communicating the wavefunction vectors one by one. The above optimization further benefits from the memory layout of $\mathbf{X_b}$ (see Fig. 2) when copying the wavefunctions data to and from the MPI buffer. Furthermore, we use FP32 for this MPI communication and it has been observed to retain FP64 accuracy in ground-state solutions while reducing the communication cost by a factor of 2. However, we still observed significant overheads on GPUs resulting from invoking independent CUDA kernel launches[10] needed for copying the wavefunctions data of each DoF on the FE partition boundary to and from the MPI buffer. We overcame this bottleneck by caching the FE partition boundary DoF indices in an array at the beginning of the program, and, subsequently, during the communication steps in $\widetilde{\mathbf{H}}\mathbf{X_b}$, we are able to handle all the boundary DoF indices simultaneously inside the CUDA kernel using the cached index array.

These implementation innovations lead to a high overall throughput for CF, as demonstrated in Fig. 3. Notably, we achieved 20.5% of the FP64 peak using block size[11] $B_f = 200$ on a single Tesla V100 GPU of Summit. We note that efforts to efficiently port the $\widetilde{\mathbf{H}}\mathbf{X}$ kernel to GPUs have been attempted in the context of finite-difference (FD) discretization [34]. However, a much lower throughput of 7.6% of FP64 peak was realized using a single Tesla K20X GPU. The significantly higher throughput demonstrated in this work is attributed to the aforementioned optimizations realized by utilizing the FE cell structure, which are not accessible in a FD discretization.

D. Mixed precision computation in CholGS and RR

Mixed precision computing strategies are being increasingly used in many scientific computing areas to accelerate performance on modern computing architectures due to significant reductions in computational cost, data movement costs and memory size. However, there has only been a limited exploration of mixed precision ideas in DFT solvers [35]. Furthermore, mixed precision algorithms have not been explored in the context of the ChFSI procedure, where the subspace spanned by the Chebyshev filtered vectors $\widetilde{\mathbf{\Psi}}$ adaptively approaches the eigensubspace corresponding to the lowest N states of $\widetilde{\mathbf{H}}$. In `DFT-FE` [13], we have exploited this behaviour of the ChFSI procedure to develop and validate mixed precision strategies for both CholGS (cf. Algorithm 3)

[10]Each CUDA kernel call has a launch overhead which can become a significant cost if a large number of CUDA kernel launches are performed.

[11]For smaller block sizes, increased memory access costs and other overheads reduce the throughput.

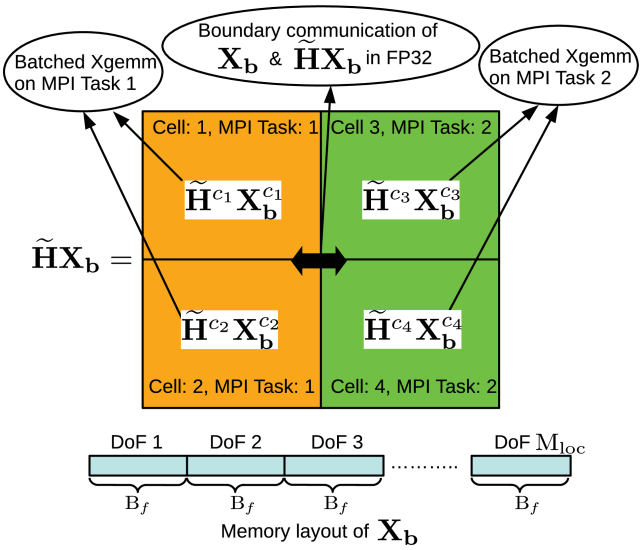

Fig. 2: Schematic of $\widetilde{\mathbf{H}}\mathbf{X_b}$ computation over four FE cells distributed over two MPI tasks using batched `Xgemm` operations. Memory layout of $\mathbf{X_b}$ where wavefunction values are stored contiguously for each degree of freedom provides coalesced memory access across GPU threads. M_{loc} denotes number of DoFs owned locally by a MPI task.

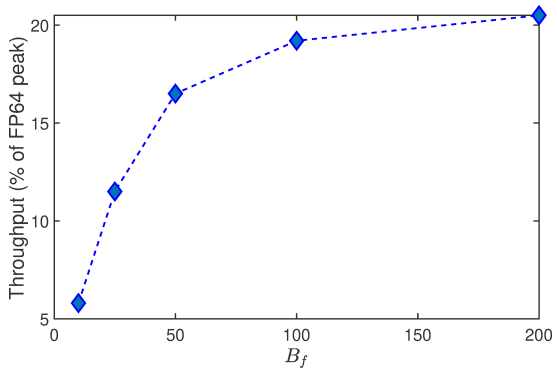

Fig. 3: Chebyshev filtering (CF) throughput on a single Tesla V100 GPU of Summit using 3 MPI tasks (via Multi Process Service) for various block sizes (B_f). FP64 peak of Summit's Tesla V100: 7.3 TFLOPS (cf. Sec. VI-C). Case study: Mg super cell with mono-vacancy containing 310 electrons. FE Mesh DoFs: 254,097.

and RR (cf. Algorithm 4) steps. This renders a majority ($\sim 85\%$) of the total FLOP count in the ChFSI procedure being performed in FP32 for the large-scale performance demonstrations in Sec. VII, while the remaining are performed in FP64. The above mixed precision algorithms are designed such that the *contribution of the FP32 computations to the electron-density tend to zero* as the SCF approaches convergence, and thereby the ground-state solution (energy and forces) retains FP64 accuracy. Further, we are able to achieve robust SCF convergence, with no change in the number of SCF iterations [13]. In Table I, using Summit GPU nodes, we demonstrate mixed-precision performance improvements of $\sim 2\times$ in the $\mathcal{O}(MN^2)$ scaling steps of CholGS and RR. We note that the use of FP32 in network communication (all-to-all communication of N^2 entries via `MPI_Allreduce`)

is crucial to realizing this enhanced performance. We remark that similar $2\times$ mixed-precision speedups have been realized in DFT−FE on NERSC's[12] Cori-KNL CPU nodes [13].

Overall, as demonstrated in Table II, all the innovations discussed above have led to *substantial GPU acceleration*, with overall SCF iteration (including all steps) speedups of $19.6\times$ on Summit nodes.

Algorithm 3 Mixed precision algorithm for CholGS.

DP{} denotes FP64 computing, and SP{} denotes FP32 computing.

1: [CholGS-S-MP] Compute overlap matrix in mixed precision:

$$\mathbf{S} = \mathrm{DP}\left\{\mathbf{S_d}\right\} + \mathrm{SP}\left\{\mathbf{S_{od}}\right\}$$

$\mathbf{S_d}$ and $\mathbf{S_{od}}$ are matrices containing the diagonal and off-diagonal entries of $\mathbf{S}$, respectively.

2: Perform CholGS-CI in double precision.

3: [CholGS-O-MP] Construct orthonormal basis in mixed precision:

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

$\mathbf{L}_{\mathbf{d}}^{-\mathbf{1}^{\dagger}}$ and $\mathbf{L}_{\mathbf{od}}^{-\mathbf{1}^{\dagger}}$ are matrices containing the diagonal and off-diagonal entries of $\mathbf{L}^{-\mathbf{1}^{\dagger}}$ respectively.

Algorithm 4 Mixed precision algorithm for RR.

DP{} denotes FP64 computing, and SP{} denotes FP32 computing.

1: [RR-P-MP] Compute projected Hamiltonian in mixed precision:

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

Sub-matrices $\hat{\mathbf{H}}_{\mathbf{oc-oc}}$, $\hat{\mathbf{H}}_{\mathbf{oc-fr}}$, $\hat{\mathbf{H}}_{\mathbf{fr-oc}}$, and $\hat{\mathbf{H}}_{\mathbf{fr-fr}}$ have sizes $N_{\mathrm{oc}} \times N_{\mathrm{oc}}$, $N_{\mathrm{oc}} \times N_{\mathrm{fr}}$, $N_{\mathrm{fr}} \times N_{\mathrm{oc}}$, and $N_{\mathrm{fr}} \times N_{\mathrm{fr}}$, respectively, with $N_{\mathrm{oc}}+N_{\mathrm{fr}} = N$. N_{oc} denotes the number of eigenstates with orbital occupancy $f_i = 1$, and N_{fr} denotes remaining fractionally occupied or unoccupied eigenstates. N_{fr} is $(10-15)\%$ of N for metallic systems. We note that, as the SCF approaches convergence, the error in electron-density introduced due to FP32 computation of $\hat{\mathbf{H}}_{\mathbf{oc-oc}}$ tends to zero [13].

2: Perform RR-D in double precision.

3: [RR-SR-MP] Perform subspace rotation step in mixed precision:

$$\widetilde{\mathbf{\Psi}}^{\mathbf{R}} = \mathrm{DP}\left[\widetilde{\mathbf{\Psi}}^{\mathbf{o}}\mathbf{Q_d}\right] + \mathrm{SP}\left[\widetilde{\mathbf{\Psi}}^{\mathbf{o}}\mathbf{Q_{od}}\right]$$

$\mathbf{Q_d}$ and $\mathbf{Q_{od}}$ are matrices containing the diagonal and off-diagonal entries of $\mathbf{Q}$ respectively.

VI. How performance was measured.

A. Systems and Environment

All simulations reported in this work using hybrid CPU-GPU architecture were executed on the Summit supercomputer, whereas the simulations using CPUs were conducted on the Theta and Cori supercomputers.

[12]National Energy Research Scientific Computing Center

TABLE I: Performance improvement due to mixed precision computation in CholGS-S and RR-P. Similar speedups are achieved for CholGS-O and RR-SR. Case study: 61,640 electrons Mg dislocation system using 1,300 Summit nodes (FP64 peak: 56.65 PFLOPS).

Step	Wall-time (sec)	Flop count (PFLOP)	PFLOPS (% of FP64 peak)
CholGS-S	13.3	224.1	16.8 (29.7%)
RR-P	16.2	228.7	14.1 (24.9%)

(a) Double precision.

Step	Wall-time (sec)	Flop count (PFLOP)	PFLOPS (% of FP64 peak)
CholGS-S-MP	7.5	224.1	29.9 (52.8%)
RR-P-MP	9.5	228.7	24.1 (42.5%)

(b) Mixed precision.

TABLE II: GPU speedup of single SCF iteration step with respect to CPU on Summit nodes. Case study: 18,480 electrons Mg dislocation system using 140 nodes. CPU simulation used 40 MPI tasks per node, with each task bound to 1 CPU core (total 42 cores in each node). CPU linear algebra performed using IBM ESSL. GPU simulation used 18 MPI tasks across 6 GPUs (via MPS) on each node.

Step	Wall-time CPU (sec)	Wall-time GPU (sec)	Speedup
Single SCF Total	844.8	43.1	19.6×

Summit is currently the fastest supercomputer in the world, with 200.79 PFLOPS FP64 peak. Summit comprises of 4,608 IBM Power System AC922 nodes with two IBM POWER9 processors (42 physical cores) and six NVIDIA Volta V100 GPUs in each node. Each node contains 512 GB of DDR4 memory for use by the POWER9 processors and 16 GB of HBM2 for each V100 GPU. Summit nodes are connected to a dual-rail EDR InfiniBand network providing a node injection bandwidth of 23 GB/s. On Summit, we have compiled DFT−FE using NVIDIA CUDA/10.1.105, GCC/6.4.0, IBM Spectrum-MPI/10.3.0.0, and IBM ESSL/6.1.0.

Theta comprises of 4,392 compute nodes with 11.69 PFLOPS peak. Each compute node has: single 1.3 GHz Intel Xeon Phi 7230 SKU chip with 64 cores, 16 GB MCDRAM and 192 GB DDR4 memory. Theta uses Dragonfly topology interconnect with bisection bandwidth of 7.2 TB/sec. Cori, the second CPU based supercomputer we have used, contains 9,688 compute nodes with 29.5 PFLOPS peak. Each compute node has: single-socket 1.4 GHz Intel Xeon Phi 7250 processor with 68 cores, 16 GB MCDRAM and 96 GB DDR4 memory. Cori uses a Cray Aries with Dragonfly topology for inter-node communication with 45.0 TB/s global peak bidirectional bandwidth. On Theta and Cori, we have compiled DFT-FE using Intel/18 compiler, Cray-MPICH, and Intel MKL.

B. What Applications were used to Measure Performance

In a recent work on the effect of alloying elements on ductility in Magnesium (Mg) [17], it was shown that small

energy difference between pyramidal I and II screw dislocations can be tuned to significantly improve the ductility of Mg. In order to use DFT calculations to guide the alloy design, it is imperative to compute the energy difference between pyramidal I and II screw dislocations (ΔE_{I-II}) to an accuracy that is better than 10^{-4} Ha/Å [17]. Achieving this requires employing a systematically convergent basis, as well as the capability to handle large system sizes involving many thousands of atoms to compute the energetics of dislocations accurately. Such large-scale and accurate DFT calculations on metallic systems are currently not feasible using state-of-the-art DFT codes. Thus, in this work, we choose an isolated pyramidal II screw dislocation in Mg, with system sizes ranging up to 105,080 electrons, as the primary application problem on which we demonstrate the performance of DFT-FE in Sec. VII. We consider three different system sizes: pyrIIScrewA with 18,480 electrons (1,848 atoms), pyrIIScrewB with 61,640 electrons (6,164 atoms), and pyrIIScrewC with 105,080 electrons (10,508 atoms). Additionally, we have also considered pyramidal I and II screw dislocations on smaller computational domains with less than 1,000 atoms, where we conducted a convergence study with respect to FE discretization to determine the discretization parameters to achieve the targeted accuracy[13] of better than 10^{-4} Ha/Å in ΔE_{I-II}. These discretization parameters have been used for pyrIIScrewA, pyrIIScrewB and pyrIIScrewC simulations.

Further, we consider other metallic benchmark systems involving HCP Mg periodic supercells with a vacancy, ranging from 2,550 to 39,990 electrons, to assess the scalability and time to solution afforded by DFT-FE in comparison to state-of-the-art DFT codes. The discretization parameters in DFT-FE, QE, and QBox for the above benchmark systems are chosen to be commensurate with chemical accuracy (discretization errors of $\sim 10^{-4}$ Ha and $\sim 10^{-4}$ Ha/Bohr in energy per atom and ionic forces, respectively).

C. Measurement Methodology

Performance of DFT-FE is measured by using the large-scale metallic systems discussed in Sec. VI-B. Comparison of DFT-FE is carried out with state-of-the-art PW DFT codes like QBox v1.66.2 [10] and QUANTUM ESPRESSO v6.3 [11] (QE) using some of the aforementioned materials systems. In the comparison of DFT-FE with QE, we use the stable single SCF iteration time[14] as a metric. We note that, although different solution methodologies are employed in DFT-FE (ChFSI) and QE (Davidson solver), we use the same electron-density mixing schemes and other appropriate parameters in both these codes that results in a similar number of SCF iterations. We have verified this on smaller benchmark systems by allowing the ground-state energy to converge to 10^{-8} Ha/atom between successive SCF iterations. Thus, the stable single SCF iteration time between DFT-FE and QE serves as an equivalent relative performance metric, and avoids the use of significant computational resources for large system sizes given the limited parallel scaling of QE. In the case of QBox, we conduct a comparative study by considering the full ground-state solution time[15], including initialization costs, for one representative benchmark materials system to achieve the same ground-state energy convergence criterion as above.

For simulations using DFT-FE, we used 3 MPI tasks per GPU (using Multi-Process Service) or equivalently 18 MPI tasks per node on Summit, 32 MPI tasks per node and 4 OpenMP threads per MPI task on Theta, and 32 MPI tasks per node and 2 OpenMP threads per MPI task on Cori. QE simulations were run on Cori using 16 MPI tasks per node and 4 OpenMP threads per MPI task for smaller problem sizes ($< 10,000$ electrons), and 8 MPI tasks per node and 4 OpenMP threads per MPI task for larger problem sizes where more memory per MPI task is required. QBox simulations were run on Theta with 32 MPI tasks per node and 4 OpenMP threads per MPI task.

Time measurements for the various computational steps and the total run-times in DFT-FE were performed using MPI_Wtime for CPU only simulations, and a combination of MPI_Wtime and cudaDeviceSynchronize for hybrid CPU-GPU simulations on Summit. FLOP counts were measured for a single SCF iteration step in the hybrid CPU-GPU simulations using nvprof with the DFT-FE solver mode set to double precision[16], while timings were measured with the solver mode set to mixed precision. We note that FLOP count measurements using nvprof are very slow (many hours in wall time) when measured for all the MPI tasks in large scale runs (> 500 Summit nodes). To circumvent this, we measured the FLOP counts at two different MPI tasks, and used the average FLOP count per MPI task multiplied by total number of MPI tasks to obtain the total FLOP count. The total FLOP count obtained in this manner is very close ($\sim 3\%$ error[17]) to explicitly measuring and adding FLOP counts for all MPI tasks, as the load balancing in DFT-FE ensures almost equal number of FE DoFs in each MPI task. Finally, we remark that the theoretical peak FP64 FLOPS, for a given number of Summit nodes, are obtained based on scaling with respect to the reported 200.79 DP-PFLOPS[18] for 4,608 nodes.

VII. Performance Results

In this section, we demonstrate the parallel scaling performance, time-to-solution, and sustained performance of DFT-FE on large-scale metallic systems involving accurate pseudopotential DFT calculations reaching 105,800 electrons. GGA [36] exchange correlation functional of the PBE type [37] and norm-conserving ONCV [38] pseudopotentials from

[13]Discretization parameters can also be chosen to achieve tighter accuracy of 5×10^{-5} Ha/Å or better in ΔE_{I-II}.

[14]It takes a few initial SCF iterations for the times to become stable in both DFT-FE and QE.

[15]We did not achieve a stable single SCF time in the case of QBox.

[16]FLOP counts have been measured in uniform precision as per ACM Gordon Bell guidelines.

[17]We verified this on a small system with 254,097 DoFs and using 6 MPI tasks, where we find the difference in the FLOP count between the two approaches is 3.3%.

[18]https://www.top500.org/list/2019/06/

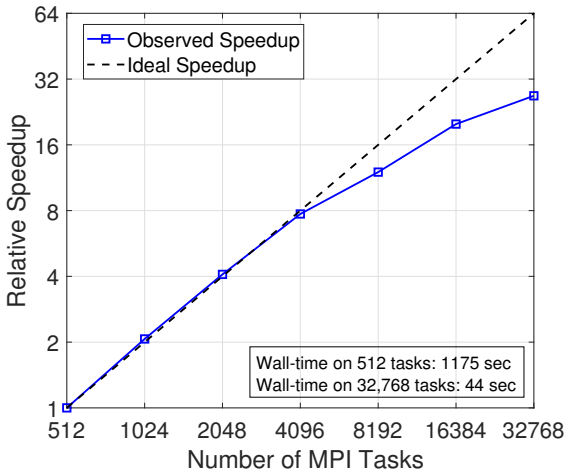

Fig. 4: Strong scaling of wall-time per SCF iteration on Theta using DFT-FE. Case study: Mg HCP supercell with 8,630 electrons (863 atoms) discretized using 23.36 million DoFs (FE basis) using FE polynomial degree 4.

the SG15 database [39] are employed in all the DFT simulations presented here.

A. Scalability & Time-to-Solution

We first examine the strong parallel scalability of DFT-FE on HCP Mg supercell with a vacancy containing 8,630 electrons (863 atoms), a representative medium-sized metallic system employing periodic boundary conditions. The scaling study conducted on Theta is shown in Fig. 4, which employed domain decomposition parallelism till 8,192 MPI tasks (corresponding to 2,852 DoFs per MPI task). To extend the scalability beyond 8,192 MPI tasks, parallelism across wavefunctions [13] is employed in conjunction with domain decomposition. A relative speedup of 26.7× is obtained reducing the wall-time per SCF iteration from 1,175 sec on 512 MPI tasks to around 44 sec on 32,768 MPI tasks. To further highlight the implications of the scaling performance of DFT-FE, we conducted a comparative study with QBox [10], state-of-the-art PW code and a previous Gordon Bell prize winner [20]. To this end, we first obtained the minimum wall-time/SCF iteration by conducting a scaling study on QBox, which was attained at 8,192 MPI tasks on 256 Theta nodes. Notably, DFT-FE outperforms QBox by 4.6×, when solved to chemical accuracy, as shown in Table III. This is a result of the low communication costs afforded by locality of FE basis, and the algorithmic and implementation innovations discussed in Sec. V. To further corroborate the computational efficiency afforded by DFT-FE in comparison to state-of-the-art PW codes, we conducted a comparative study of minimum wall-times per SCF iteration obtained using DFT-FE and QE on increasing system sizes of Mg HCP supercells with a vacancy. As shown in Fig. 5, DFT-FE significantly outperforms QE, and, notably, attains 9× speedup for the system containing 20,470 electrons.

TABLE III: Comparison of total run time of DFT-FE ground-state calculation with QBox on Theta. Case study: Mg supercell with a vacancy (8,630 electrons).

Code	Wall-time (sec)	MPI tasks	Energy/atom (Ha)
Qbox	6480	8192	-54.32655
DFT-FE	1403	32768	-54.32650

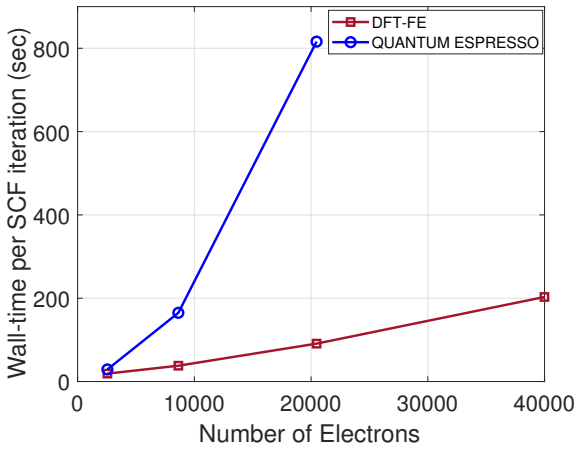

Fig. 5: Comparison of minimum wall-time per SCF iteration between DFT-FE and QE on Cori KNL nodes. Case study: Mg supercells with increasing system size.

We next consider our primary application problem, the pyramidal II screw dislocation in Mg. The strong scaling study conducted on pyrIIScrewA containing 18,480 electrons (1,848 Mg atoms) is shown in Fig. 6 for Theta and Fig. 7 for Summit. We obtain 82% efficiency on Theta at 16,384 MPI tasks with 3,385 DoFs per task. This corresponds to a relative speedup of 14.5×, thus, reducing the wall-time per SCF iteration from 1,511 sec on 2,048 MPI tasks to 104 sec on 65,536 MPI tasks[19]. On Summit GPU nodes, we obtain 96% efficiency at 280 nodes (5,040 MPI tasks) with 11,000 DoFs per task. The wall-time per SCF iteration reduced from 97.6 sec on 1,260 MPI tasks to around 13.99 sec on 20,160 MPI tasks. We remark that DFT-FE's scaling range on Summit is smaller compared to that obtained using Theta due to the high GPU compute speedups of DFT-FE relative to CPUs (cf. Table II for GPU speedups obtained on the same system).

Subsequently, we examine the weak scaling performance of DFT-FE on Summit nodes. DFT-FE simulations have been set up such that the memory associated with the wavefunctions per MPI task ($\mathcal{O}(M_{\text{loc}}N)$), which constitutes the dominant memory footprint, remains a constant across all benchmark systems. In DFT-FE, the computational complexity of CF scales as $\mathcal{O}(M_{\text{loc}}N)$ per MPI task, while the scaling of CholGS and RR is $\mathcal{O}(M_{\text{loc}}N^2)$. The departure from ideal weak scaling efficiency in Fig. 8 is expected because of the $\mathcal{O}(M_{\text{loc}}N^2)$ complexity becoming dominant with increase in the number of electrons.

[19]This run used 2,048 nodes, which corresponds to half the total number of Theta nodes. Scalability of DFT-FE has also been demonstrated on 192,000 MPI tasks using ∼60% of Cori-KNL nodes [13].

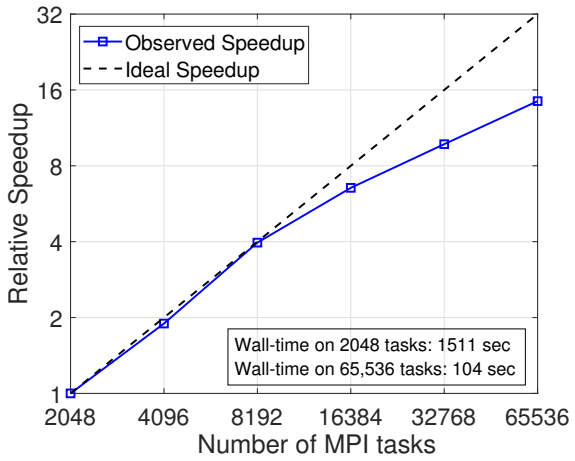

Fig. 6: Strong scaling of wall-time per SCF iteration on Theta using DFT-FE. Case study: Mg screw dislocation system (pyrIIScrewA) with 18,480 electrons (1,848 atoms) discretized with 55.11 million DoFs.

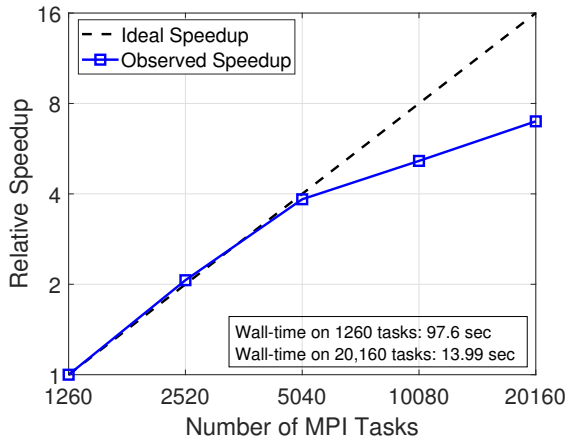

Fig. 7: Strong scaling of wall-time per SCF iteration on Summit GPU nodes using DFT-FE. Case study: Mg dislocation system with 18,480 electrons (1,848 atoms). Each GPU is associated with 3 MPI tasks via MPS.

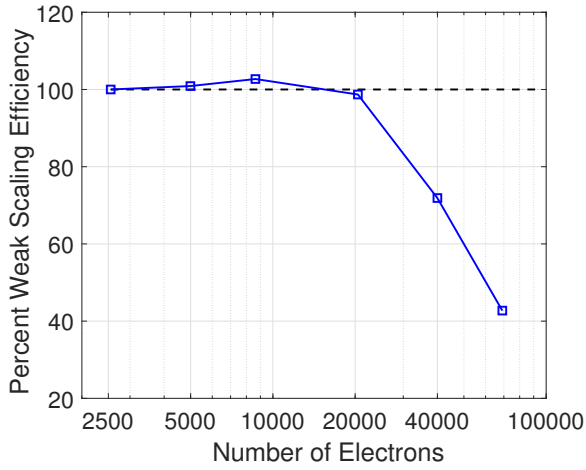

Fig. 8: Weak scaling efficiency of wall-time per SCF iteration on Summit (MPI tasks used are 54, 180, 576, 3294, 12744 and 38250). Case study: Mg supercells with increasing system size.

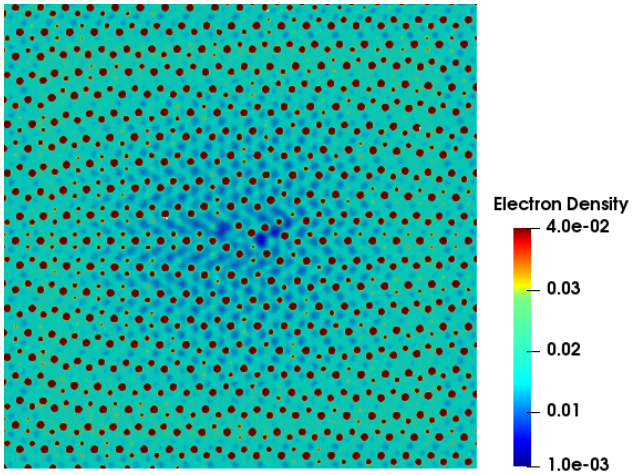

Fig. 9: Electron density contour of pyramidal II screw dislocation system in Mg with 61,640 electrons (6,164 Mg atoms).

B. Large-scale dislocation systems performance: Time-to-Solution & Sustained Performance

We now demonstrate the performance of DFT-FE on large-scale Mg dislocation systems: pyrIIScrewB with 61,640 electrons (6,164 Mg atoms), and pyrIIScrewC with 105,080 electrons (10,508 Mg atoms) using Summit GPU nodes. First, in Table IV, we report the time-to-solution and performance of the pyrIIScrewB system, which has been discretized with 179.03 million DoFs. This simulation achieved 16.7 PFLOPS (29.5% efficiency) sustained performance in the ground-state calculation involving 56 SCF iterations[20], and almost similar sustained performance of 14.7 PFLOPS (26.0% efficiency) over the entire run time of the program including initialization costs[21]. In Table IV, we also measured the performance of the various computational steps in a SCF iteration, demonstrating the significantly high efficiencies achieved in the range of 42.5% – 119.5% for the $\mathcal{O}(MN^2)$ scaling mixed-precision compute steps, which constitute 82.4% of the total FLOP count. We note that the FLOP counts of CholGS-S-MP, CholGS-O-MP and RR-P-MP are smaller compared to RR-SR-MP, which is attributed to the optimizations we have performed by exploiting the Hermiticity and triangular nature of the matrices involved [13].

Finally, in Table V, we report the wall-time and sustained performance of a single SCF iteration step[22] of the pyrIIScrewC system, the largest system size considered in this work with 105,080 electrons (10,508 Mg atoms) and discretized with 304.92 million DoFs (17.69 trillion wavefunction values).

[20]Multiple calls to the ChFSI procedure (cf. Algorithm 2) are performed in the initial few SCFs. Hence total FLOP count is calculated by multiplying the single SCF iteration step FLOP count with total number of ChFSI calls.

[21]The initialization costs will be reduced in half after a known issue in deal.II library (https://github.com/dealii/dealii/issues/7053) is resolved in the short-term. Our current temporary fix involves an expensive serial operation.

[22]Due to our limited resource allocation on Summit, we could not perform a full ground-state calculation on the pyrIIScrewC system (105,080 electrons). However as demonstrated on the pyrIIScrewB system (61,640 electrons) in Table IV the performance of a single SCF iteration step is very close (within 12%) to that of the full run-time.

TABLE IV: Time-to-solution and performance of pyrIIScrewB (61,640 electrons) system. Simulation performed using 1,300 Summit nodes (FP64 peak: 56.65 PFLOPS).

Procedure	Wall-time (sec)	FLOP count (PFLOP)	PFLOPS (% of FP64 peak)
Initialization	981	-	-
Ground-state	7377	123174	16.7 (29.5%)
Total	**8358**	**123174**	**14.7 (26.0%)**

(a) Total program run time and sustained performance. Breakdown of run time into initialization and ground-state calculation costs.

Step	Wall-time (sec)	FLOP count (PFLOP)	PFLOPS (% of FP64 peak)
CF	36.3	212.6	5.9 (10.3%)
CholGS-S-MP	7.5	224.1	29.9 (52.8%)
CholGS-CI	1.6	-	-
CholGS-O-MP	5.6	228.9	40.9 (72.2%)
RR-P-MP	9.5	228.7	24.1 (42.5%)
RR-D	7.0	-	-
RR-SR-MP	6.6	446.6	67.7 (119.5%)
DC	0.9	27.7	30.8 (54.3%)
EP	4.6	-	-
Others	3.7	-	-
Single SCF Total	**83.3**	**1368.6**	**16.4 (29.0%)**

(b) Breakdown of performance of single SCF iteration. FLOP count for operations on CPU (CholGS-CI, RR-D and EP), which constitute a minor portion of the total FLOP count are not measured, though their wall-times are considered in the single SCF total time.

TABLE V: Wall-time and sustained performance of a single SCF iteration step of pyrIIScrewC (105,080 electrons) system. Simulation performed using 3,800 Summit nodes (FP64 peak: 165.58 PFLOPS).

Step	Wall-time (sec)	FLOP count (PFLOP)	PFLOPS (% of FP64 peak)
Single SCF	142.7	6,563.7	46 (27.8%)

(a) Performance of single SCF iteration.

Step	Wall-time (sec)
CF	6.1
Chol-GS	36.9
RR	54.0
DC+EP+Others	5.7

(b) Breakdown of single SCF iteration wall-time into major steps.

We note that the SCF wall-time in Table V is obtained by taking the average over 10 calls to the ChFSI procedure (cf. Algorithm 2), thus demonstrating sustainability of the performance. This simulation achieves two significant landmarks. First, the single SCF wall-time of 142.7 sec demonstrates that fast large-scale and chemically accurate Kohn-Sham DFT simulations of metallic systems reaching $\sim 100,000$ electrons are now possible. Second, we achieve an unprecedented sustained performance of 46 PFLOPS (27.8% efficiency) utilizing 3,800 nodes out of total 4,608 nodes on Summit, which is 3.1× higher than the current DFT record of 15 DP-PFLOPS [21].

VIII. Implications

Applications: The recent disruptive advancements in computing architectures require a paradigm shift in the design and implementation of computational algorithms to exploit these developments and enable accurate DFT calculations at larger length- and time-scales than possible heretofore. This work constitutes an important step in this direction. In particular, the advances reported here, make possible, for the first time, accurate DFT calculations on metallic systems with tens of thousands of atoms computationally feasible and practical. This opens the possibility of tackling a wide range of scientifically and technologically important problems that have been out of reach. For instance, accurate *ab initio* studies on dislocation energetics now becomes a reality, which can guide and accelerate the discovery of new light weight structural alloys. To elaborate, the electronic ground-state of a dislocation system containing $\sim$5,000 – 10,000 atoms ($\sim$50,000 – 100,000 electrons) can be computed in $\sim$ 2 – 4 hrs, which is unprecedented—current state-of-the-art codes will require a few days to a week. Other application areas that benefit from the ability to handle large-scale metallic systems in an efficient manner, to name a few, include: design and discovery of new catalytic materials, studies on high entropy alloys, novel energy storage materials, organometallic complexes in biomolecular electronics. Furthermore, the significant improvement in the time-to-solution will also enable large-scale *ab initio* MD simulations for longer time-scales than possible thus far, which will be instrumental in studying transport (electronic and ionic), reaction kinetics, kinetics of phase transitions etc., over a wide range of materials systems.

Future architectures: The upcoming exascale machines are expected to have significantly more compute capability per node, in addition to having more compute nodes, compared to current pre-exascale machines. However, the data movement and memory bandwidths are not expected to increase significantly. In this regard, the innovations realized in `DFT-FE`, which significantly improved arithmetic intensity and reduced data movement costs, make it well-placed to leverage future exascale machines. Further, `DFT-FE` will also be able to take advantage of larger number of compute nodes on exascale machines to extend the parallel scalability, through the use of parallelization across wavefunctions (band parallelization)[23], to enable simulations of even larger metallic systems. Implementation of enriched finite-element basis [40] in `DFT-FE`, which has the potential to substantially reduce DoFs for pseudopotential DFT calculations, while maintaining the same accuracy and parallel scaling, can further expand the accessible system sizes to millions of electrons on exascale machines. At such extreme sizes, building on the developments reported here, further algorithmic advancements such as development of accurate "divide and conquer" approaches for generic materi-

[23]Band parallelization was not activated on Summit for the $\sim 100,000$ electrons system as we exhausted almost all the compute nodes (3,800 out of 4,608) for performing domain decomposition parallelization.

als systems have the potential to push DFT−FE's performance into the exaFLOPS range.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support from DOE-BES (DE-SC0008637) and Toyota Research Institute. This work used resources of OLCF (DE-AC05-00OR22725), ALCF (DE-AC02-06CH11357), and NERSC (DE-AC02-05CH11231). V.G. also gratefully acknowledges support from AFOSR and ARO that supported some algorithmic developments, and B.T. acknowledges support of LDRD program of ORNL.

REFERENCES

[1] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects,Phys. Rev. 140(4A) (1965) A1133.

[2] https://www.nobelprize.org/prizes/chemistry/1998/summary/

[3] R. M. Martin, Electronic structure: basic theory and practical methods, Cambridge university press (2004).

[4] S. Goedecker, Linear scaling electronic structure methods, Rev. Mod. Phys. 71 (1999) 1085–1123.

[5] D. R. Bowler, T. Miyazaki, $\mathcal{O}(\mathcal{N})$ methods in electronic structure calculations, Rep. Prog. Phys. 75(3) (2012) 036503.

[6] L.-W. Wang, Z. Zhao, J. Meza, Linear-scaling three-dimensional fragment method for large-scale electronic structure calculations, Phys. Rev. B 77 (2008) 165113.

[7] M. Eisenbach, C.-G. Zhou, D. M. Nicholson, G. Brown, J. Larkin, and T. C. Schulthess, A scalable method for ab initio computation of free energies in nanoscale systems, In Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis (2009).

[8] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele, cp2k: atomistic simulations of condensed matter systems, Wiley Interdiscip. Rev.: Comput. Mol. Sci. 4(1) (2014) 15–25.

[9] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, W. de Jong, NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations, Comput. Phys. Commun. 181(9) (2010) p.1477.

[10] F. Gygi, Architecture of Qbox: A scalable first-principles molecular dynamics code, IBM J. Res. Dev. 52 (2008) 137–144.

[11] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car,C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. D. Corso, S. de Giron-coli, P. Delugas, R. A. D. Jr, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer,U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Kkbenli,M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. O. de-laRoza, L. Paulatto, S. Ponc, D. Rocca, R. Sabatini, B. Santra,M. Schlipf, A. P. Seitsonen,A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, S. Baroni, Advanced capabilities for materials modelling with QUANTUM ESPRESSO, J. Phys. Condens. Matter. 29(46) (2017) 465901.

[12] G. Kresse, and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Physical review B, 54(16) (1996) 11169.

[13] P. Motamarri, S. Das, S. Rudraraju, K. Ghosh, D. Davydov, V. Gavini, DFT-FE: A massively parallel adaptive finite-element code for large-scale density functional theory calculations, arXiv:1903.10959.

[14] L. Liu, A. Corma, Metal catalysts for heterogeneous catalysis: From single atoms to nanoclusters and nanoparticles, Chem. Rev. 118(10) (2018) 4981–5079.

[15] J. Janek, and W.G. Zeier, A solid future for battery development, Nat. Energy, 500(400) (2016) 30.

[16] O. El-Atwani, N. Li, M. Li, A. Devaraj, J. K. S. Baldwin, M. M. Schneider, D. Sobieraj, J. S. Wróbel, D. Nguyen-Manh, S. A. Maloy, E. Martinez, Outstanding radiation resistance of tungsten-based high-entropy alloys, Sci. Adv. 5(3) (2019) eaav2002.

[17] Z. Wu, R. Ahmad, B. Yin, S. Sandlöbes, and W.A. Curtin, Mechanistic origin and prediction of enhanced ductility in magnesium alloys. Sci. 359(6374) (2018) 447–452.

[18] T. M. Pollock, Weight loss with magnesium alloys, Sci. 328(5981) (2010) 986–987.

[19] Z. Wu, and W.A. Curtin, The origins of high hardening and low ductility in magnesium. Nat. 526(7571) (2015) p.62.

[20] F. Gygi, E. W. Draeger, M. Schulz, B. R. de Supinski, J. A. Gunnels, V. Austel, J. C. Sexton, F. Franchetti, S. Kral, C. W. Ueberhuber, and J. Lorenz, Large-scale electronic structure calculations of high-Z metals on the BlueGene/L platform. In Proceedings of the 2006 ACM/IEEE conference on Supercomputing (2006).

[21] M. Calderara, S. Brück, A. Pedersen, M. H. Bani-Hashemian, J. VandeVondele, and M. Luisier, Pushing back the limit of ab-initio quantum transport simulations on hybrid supercomputers. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (2015).

[22] J. R. Chelikowsky, N. Troullier, and Y. Saad, Finite-difference-pseudopotential method: Electronic structure calculations without a basis, Phys. Rev. Lett. 72 1240 (1994).

[23] L. Kronik, A. Makmal, M. L. Tiago, M. M. G. Alemany, M. Jain, X. Huang, Y. Saad, J. R. Chelikowsky, PARSEC – the pseudopotential algorithm for real-space electronic structure calculations: recent advances and novel applications to nano-structures, Phys. Status Solidi B 243(5) (2006) 1063–1079.

[24] J. E. Pask, P. A. Sterne, Finite element methods in ab initio electronic structure calculations, Modell. Simul. Mater. Sci. Eng. 13 (3) (2005).

[25] P. Motamarri, M. Nowak, K. Leiter, J. Knap, V. Gavini, Higher-order adaptive finite-element methods for Kohn-Sham density functional theory, J. Comput. Phys. 253 (2013) 308–343.

[26] Y. Hasegawa, J.-I. Iwata, M. Tsuji, D. Takahashi, A. Oshiyama, K. Minami, T. Boku, F. Shoji, A. Uno, M. Kurokawa, H. Inoue, I. Miyoshi, and M. Yokokawa, First-principles calculations of electron states of a silicon nanowire with 100,000 atoms on the K computer, In Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis (2011).

[27] J.-L. Fattebert, D. Osei-Kuffuor, E. W. Draeger, T. Ogitsu, and W. D. Krauss, Modeling dilute solutions using first-principles molecular dynamics: computing more than a million atoms with over a million cores. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (2016).

[28] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, M. C. Payne, Introducing ONETEP: Linear-scaling density functional simulations on parallel computers, J. Chem. Phys. 122(8) (2005) 084119.

[29] J. Aarons, C.-K. Skylaris, Electronic annealing Fermi operator expansion for DFT calculations on metallic systems, J. Chem. Phys. 148 (2018) 074107.

[30] C. Burstedde, L. C. Wilcox, O. Ghattas, p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees, SIAM J. Sci. Comput 33(3) (2011) 1103–1133.

[31] G. Alzetta, D. Arndt, W. Bangerth, V. Boddu, B. Brands, D. Davydov, R. Gassm oller,T. Heister, L. Heltai, K. Kormann, M. Kronbichler, M. Maier, J.-P. Pelteret, B. Turcksin, D. Wells, The deal.II Library, Version 9.0, J. Numer. Math. 26(4) (2018) 173–183.

[32] Y. Zhou, Y. Saad, M. L. Tiago, J. R. Chelikowsky, Self-consistent-field calculations using Chebyshev-filtered subspace iteration, J. Comput. Phys. 219(1) (2006) 172–184.

[33] A. Marek, V. Blum, R. Johanni, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H.-J.Bungartz, H. Lederer, The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science, J. Phys.: Condens. Matter 26(21) (2014) 213201.

[34] X. Andrade, and A. Aspuru-Guzik, Real-space density functional theory on graphical processing units: computational approach and comparison to Gaussian basis set methods. J. Chem. Theory Comput. 9(10) (2013) 4360–4373.

[35] E. Tsuchida, Y.-K. Choe, Iterative diagonalization of symmetric matrices in mixed precision and its application to electronic structure calculations, Comput. Phys. Commun. 183(4) (2012) 980–985.

[36] D. C. Langreth, M. J. Mehl, Beyond the local-density approximation in calculations of ground-state electronic properties, Phys. Rev. B 28 (1983) 18091834.

[37] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868.

[38] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, Phys. Rev. B 88(1995) 239906.

[39] M. Schlipf, F. Gygi, Optimization algorithm for the generation of ONCV pseudopotentials, Comput. Phys. Commun. 196 (2015) 36–44.

[40] B. Kanungo, V. Gavini, Large-scale all-electron density functional theory calculations using an enriched finite-element basis, Phys. Rev. B 95 (2017) 035112.