Accelerating self-consistent field iterations in Kohn-Sham density functional theory using a low-rank approximation of the dielectric matrix

Sambit Das ●

Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA

Vikram Gavini ●*

Department of Mechanical Engineering and Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA

We present an efficient preconditioning technique for accelerating the fixed-point iteration in real-space Kohn-Sham density functional theory (DFT) calculations. The preconditioner uses a low-rank approximation of the dielectric matrix (LRDM) based on Gâteaux derivatives of the residual of fixed-point iteration along appropriately chosen direction functions. We develop a computationally efficient method to evaluate these Gâteaux derivatives in conjunction with the Chebyshev filtered subspace iteration procedure, an approach widely used in large-scale Kohn-Sham DFT calculations. Further, we propose a variant of LRDM preconditioner based on adaptive accumulation of low-rank approximations from previous self-consistent field iterations, and also extend the LRDM preconditioner to spin-polarized Kohn-Sham DFT calculations. We demonstrate the robustness and efficiency of the LRDM preconditioner against other widely used preconditioners on a range of benchmark systems with sizes ranging from ∼100 to 1100 atoms (∼500–20 000 electrons). The benchmark systems include various combinations of metal-insulating-semiconducting heterogeneous material systems, nanoparticles with localized *d* orbitals near the Fermi energy, nanofilm with metal dopants, and magnetic systems. In all benchmark systems, the LRDM preconditioner converges robustly within 20–30 iterations. In contrast, other widely used preconditioners show slow convergence in many cases, as well as divergence of the fixed-point iteration in some cases. Finally, we demonstrate the computational efficiency afforded by the LRDM method, with up to 3.4-fold reduction in computational cost for the total ground-state calculation compared to other preconditioners.

I. INTRODUCTION

Electronic-structure calculations based on Kohn-Sham density functional theory (KS-DFT) [1,2] provide an excellent balance between accuracy and computational efficiency by reducing the many-body Schrödinger problem of interacting electrons into an equivalent problem of noninteracting electrons in an effective mean field that is governed by the electron density. This has led to KS-DFT being one of the most widely used electronic-structure methods for the predictive modeling of materials and for gaining qualitative and quantitative insights into various material properties. The significant increase in computational resources over the last decade, including the advent of hybrid CPU-GPU architectures, has also played an important role in the wide adoption of KS-DFT. Furthermore, the simultaneous development of efficient and scalable numerical schemes in conjunction with systematically convergent real-space discretizations [3–8], including reduced-order scaling approaches (cf., e.g., [9–12]), have advanced the ability to conduct fast and accurate DFT calculations using large-scale computing platforms. As a result, applications using KS-DFT are increasingly targeting larger as well as more complex heterogeneous material

systems [13–21]. However, existing numerical methods to solve the Kohn-Sham equations suffer from instabilities for heterogeneous systems, with the convergence worsening for larger system sizes. To elaborate, the ground-state solution in KS-DFT is often computed via the solution of the nonlinear Kohn-Sham eigenvalue problem, that is posed as a fixed-point iteration—commonly referred to as the self-consistent field (SCF) iteration—in the electron density, written as $\rho = F[V_{\text{eff}}[\rho]]$, where ρ denotes the electron density and $V_{\text{eff}}[\rho]$ is the Kohn-Sham effective mean-field potential. As each step in the fixed-point iteration involves the computation of Kohn-Sham eigenstates that scales cubically with the number of electrons (N_e), the slow convergence is a serious computational bottleneck for large-scale DFT calculations. As will be discussed below, existing state-of-the-art methods for accelerating the Kohn-Sham fixed-point iteration are either not suitable to generic heterogeneous material systems or incur significant computational overhead.

The origin of the instabilities in the Kohn-Sham SCF iteration is due to the large condition number of the Jacobian operator corresponding to the residual of the fixed-point iteration. The Jacobian operator, denoted by $J = \frac{\delta}{\delta\rho}(F[V_{\text{eff}}[\rho]] - \rho)$, is related to the physical dielectric operator of the material system, $J = -\epsilon^\dagger$. Methods such as the Ho-Ihm-Joannopoulos (HIJ) method [22] and the extrapolar preconditioner [23] have been proposed to directly compute ϵ that allow efficient

*vikramg@umich.edu

convergence of the SCF iteration. However, the associated quartic-scaling computational cost of computing ϵ limits their application to small system sizes of ~ 100 atoms [23]. The quartic-scaling cost arises from the computation of the static susceptibility matrix (χ_0), a portion of ϵ, using the Alder-Wiser expression [24,25] that involves a double summation over all occupied and many unoccupied eigenstates for each matrix element of χ_0. Thus, DFT codes have primarily relied on cheaper quasi-Newton techniques with an approximation of J as a preconditioner to accelerate the SCF iteration. Broadly, two groups of such methods have found wide usage in the DFT community, one using direct numerical approximation of J from the history of previous SCF iterations and the other based on physically motivated approximations of ϵ. Concerning the former group of methods, the most widely used techniques are those of Anderson [26], Pulay [27], and Broyden [28] and direct inversion of the iterative subspace (DIIS) [29]. These techniques can be generically considered as multisecant approximations of J or J^{-1} [30], where the preconditioner is constructed to optimally satisfy the secant approximation of J at multiple steps, simultaneously, based on a history of ρ and $F[V_{\text{eff}}[\rho]]$ from previous SCF iterations. Further, such schemes have been shown to behave like Krylov subspace methods with Q-superlinear convergence near the ground-state solution [31,32]. However, as shown in previous studies [33,34], for large metallic systems and heterogeneous systems with large condition numbers of J, multisecant approaches demonstrate slow and system-size-dependent convergence. These issues can be further compounded by potential strong nonlinearities in $F[V_{\text{eff}}[\rho]]$ in heterogeneous systems, resulting in divergence of the SCF iteration as will be demonstrated in this work. Given the limitations of the multisecant methods, several physically motivated approximations of ϵ have been proposed, and they are typically combined with the aforementioned multisecant methods of Anderson, Pulay, or Broyden. The Kerker preconditioner [35] is one such widely used approximation, which is based on the Thomas-Fermi screening theory of homogeneous electron gas. Although the Kerker preconditioner captures the long wavelength divergent eigenvalues of ϵ in bulk metallic systems, it is not suitable for semiconducting and insulating systems as it does not model the screening behavior in these systems. In order to better capture the screening effects in semiconducting and insulating systems, the Resta [36] and truncated-Kerker [37] preconditioners have been proposed based on material-specific parametrizations related to the static dielectric constant. However, these preconditioners are still not suitable for heterogeneous material systems with spatially varying screening behavior. We refer to [38] for a more in-depth review of the numerical convergence aspects of the above preconditioners, and their systematic comparison on robustness and efficiency measures assessed on a test suite of benchmarks that encapsulates various sources of ill conditioning of the Kohn-Sham SCF iteration.

In order to address the aforementioned challenges posed by heterogeneous systems, preconditioners such as the Thomas-Fermi–von Weizsacker (TFW) preconditioner [39], elliptic preconditioner [33], and local density of states (LDOS)-based preconditioner [34] have recently been developed. The TFW preconditioner approximates the χ_0 portion of ϵ, relying on the equivalence between χ_0 and the inverse of the double functional derivative of the noninteracting kinetic energy functional ($T_s[\rho]$) and using the TFW approximation for T_s. However, the TFW functional approximation cannot accurately capture the complex dielectric response in general heterogeneous systems due to the semilocal nature of the functional, and thus limits its suitability to simpler metal-vacuum systems. In the case of elliptic and LDOS preconditioners, a key limitation is that they consider only long-range eigenmodes of χ_0 in the construction of the preconditioner. This prevents the elliptic and LDOS preconditioners from appropriately accounting for the eigenmodes of χ_0 with large eigenvalues related to localized states near the Fermi energy [34] that would require resolving eigenmodes with atomic-scale wavelengths. Such localized states near the Fermi energy can occur for metallic elements with d and f valence orbitals. Further, strong nonlinearities in $F[V_{\text{eff}}[\rho]]$ are not accounted for in these preconditioners. In addition to the above-discussed preconditioners for the Kohn-Sham fixed-point iteration map, methods to directly minimize the Kohn-Sham finite-temperature free energy functional over the Kohn-Sham orbitals and fractional occupancies have also been developed, for example, the ensemble DFT method [40]. Due to its improved global convergence, in general, over fixed-point-iteration-based methods, ensemble DFT has been demonstrated to converge robustly for challenging SCF problems. However, the computational cost associated with direct minimization can be substantially larger than fixed-point iteration methods [38].

In this work, we present a robust and computationally efficient preconditioning approach for the Kohn-Sham SCF iteration based on a low-rank approximation of J, or equivalently ϵ. This approach, which we refer to as the low-rank dielectric matrix (LRDM) preconditioning approach, constructs an approximation of J based on a sum of rank-1 tensor products between direction functions corresponding to an approximate Krylov subspace of J and the Gâteaux derivative of the residual of the fixed-point iteration, $F[V_{\text{eff}}[\rho]] - \rho$, along the direction functions. We note that this preconditioning strategy was proposed in a recent work by Niklasson [41], in the context of self-consistent charge density functional tight-binding theory employing a reduced-order atom-centered basis. In this work, we build upon this idea to apply the LRDM method to Kohn-Sham DFT calculations employing systematically convergent higher-order finite-element basis sets [42]. Further, we demonstrate the robustness and efficiency of LRDM using extensive large-scale heterogeneous benchmark systems, and comparing against widely used preconditioners. The primary challenge here is the efficient computation of the the Gâteaux derivatives of $F[V_{\text{eff}}[\rho]]$ along the direction functions—referred to as the first-order density response functions—that constitutes the most computationally expensive portion of the LRDM method. Further, developing strategies to enable modest values of rank for a wide range of heterogeneous systems is critical to minimize the overheads associated with the method and boost the practical computational efficiency. Regarding the first aspect, one of the key contributions of the present work is to develop a computationally efficient method to approximate the first-order density response functions in real space, and within

the context of Chebyshev filtered subspace iteration (ChFSI) procedure [3]. We note that ChFSI is a computationally efficient and scalable eigensolver that progressively approximates the eigensubspace corresponding to the occupied Kohn-Sham eigenstates instead of performing an exact diagonalization in each SCF iteration. ChFSI provides significant computational gains over other iterative eigensolvers for solution of the Kohn-Sham nonlinear eigenvalue problem [3,42]. As a result, many real-space DFT codes, based on finite-element basis [6,8] or finite-difference discretization [7,43,44], currently employ the ChFSI procedure. In the proposed work, the evaluation of the first-order density response functions in conjunction with the ChFSI procedure involves two steps. First, we compute the first-order density-matrix response in the approximate eigensubspace of dimension $\sim N_e$, obtained in each iteration of the ChFSI procedure. Subsequently, we transform the density-matrix response to the space spanned by the finite-element basis, and obtain the density response functions from the diagonal of the density-matrix response. This results in an $\mathcal{O}(MN_e^2)$ scaling method (M denoting the size of the finite-element basis) with a small computational prefactor, as will be demonstrated in this work. The second important aspect of the present work targets reduction of the average rank by developing an accumulated variant of LRDM, which adaptively accumulates the Jacobian approximation from previous SCF iterations. Our proposed numerical approach for accumulation entails taking new direction functions that are orthogonal to the ones from the previous SCF iterations in conjunction with an adaptive strategy that either continues or clears the accumulation based on numerical metrics informing the linearity of the residual function with respect to the electron density, and the low-rank approximation error. Furthermore, in the present work, we extend the formulation of the LRDM preconditioner to collinear spin-polarized KS-DFT calculations.

We demonstrate the robustness and efficiency of the LRDM preconditioner, and compare it against Anderson, Kerker, and TFW preconditioners on a comprehensive set of benchmark heterogeneous material systems ranging up to $\sim$1100 atoms ($\sim$20 000 electrons), including spin-polarized magnetic systems. In all the benchmark systems LRDM demonstrates robust and system-size-independent convergence. In contrast, other preconditioners studied in this work do not converge, in some cases, for larger-scale heterogeneous benchmark systems. We also demonstrate the advantages of the proposed accumulated LRDM variant in reducing the average rank. Finally, we demonstrate the computational efficiency of the LRDM preconditioner against Anderson and/or Kerker preconditioners, where we observe up to a 3.4-fold reduction in the computational times for full ground-state calculations on the benchmark problems.

The remainder of the paper is organized as follows. Section II starts with a brief theoretical background on the convergence aspects of the Kohn-Sham SCF iteration. Subsequently, in Sec. II B we present the formulation of the LRDM preconditioner and its proposed accumulated variant. In Sec. II C, we develop the computational method for evaluation of the density response functions in real-space Kohn-Sham DFT calculations within the ChFSI procedure. Finally, in Sec. II D we extend the formulation to spin-polarized

calculations. Section III presents computational results on heterogeneous benchmark systems comparing the SCF convergence of the LRDM preconditioner against other widely used preconditioners. Section III also demonstrates the computational efficiency of the LRDM preconditioner on hybrid CPU-GPU architectures. We finally conclude with an outlook in Sec. IV.

II. FORMULATION

We first introduce the mathematical notation that will be used in the subsequent sections. We assume the electron density, Kohn-Sham wavefunctions, and other electronic fields appearing in the formulation to belong to an appropriate function space, $\Upsilon(\Omega)$. In the case of nonperiodic calculations, Ω corresponds to a large enough domain containing the compact support of the electronic fields, and, in periodic calculations, it corresponds to a periodic domain. We denote the inner product between two functions as $\langle g_1, g_2 \rangle = \int_\Omega g_1^*(\mathbf{x}) g_2(\mathbf{x}) d\mathbf{x}$, and $\|\cdot\|$ to be the norm induced from this inner product. In what follows, we denote the action of an infinite-dimensional bounded linear operator $A : \Upsilon \to \Upsilon$ on g as $A\,g := \int_\Omega A(\mathbf{x}, \mathbf{x}')\, g(\mathbf{x}') d\mathbf{x}'$, where $A(\mathbf{x}, \mathbf{x}')$ and $g(\mathbf{x})$ denote the real-space representations of A and g, respectively. Finally, we denote the action of a nonlinear operator F on g as $F[g] : \Upsilon \to \Upsilon$.

A. Background on Kohn-Sham SCF convergence

We begin by considering the Kohn-Sham fixed-point iteration. For a materials system with N_{at} nuclei and N_e electrons, the spin-restricted ground-state properties in Kohn-Sham density functional theory are given by solving the $N/2$ lowest eigenstates of the following nonlinear eigenvalue problem [1]:

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}[\rho, \boldsymbol{R}]\right)\psi_k = \varepsilon_k\,\psi_k,$$

$$2\sum_k f(\epsilon_k, \mu) = N_e, \quad f(\epsilon, \mu) = \frac{1}{1 + \exp\left(\frac{\epsilon - \mu}{k_B T}\right)},$$

$$\rho(\mathbf{x}) = 2\sum_k f(\varepsilon_k, \mu)|\psi_k(\mathbf{x})|^2,$$

$$V_{\text{eff}}[\rho, \boldsymbol{R}] = V_{\text{xc}}[\rho] + V_{\text{H}}[\rho] + V_{\text{ext}}(\boldsymbol{R}),$$

where $\boldsymbol{R} = \{\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_{N_{at}}\}$ denotes the positions of the N_{at} nuclei, ρ denotes the electron density, $V_{\text{eff}}[\rho, \boldsymbol{R}]$ denotes the effective single-electron Kohn-Sham potential, ε_k and ψ_k denote the eigenstates of the Kohn-Sham Hamiltonian (henceforth denoted by $\mathcal{H}[\rho]$), and $f(\epsilon, \mu)$ denotes the Fermi-Dirac distribution with μ being the Fermi energy or the chemical potential. $V_{\text{eff}}[\rho, \boldsymbol{R}]$ is composed of the exchange-correlation potential ($V_{\text{xc}}[\rho]$) accounting for the many-body quantum mechanical interactions, the Hartree electrostatic potential corresponding to the electron density ($V_{\text{H}}[\rho]$), and the external electrostatic potential from the nuclei ($V_{ext}(\boldsymbol{R})$). The above nonlinear eigenvalue problem can be viewed as the fixed-point iteration $\rho = F[V_{\text{eff}}[\rho]]$, which is commonly referred to as the SCF iteration. Within this SCF procedure, the evaluation of $F[V_{\text{eff}}[\rho]]$ entails solving a linear eigenvalue problem. Solution of this fixed-point problem is equivalent
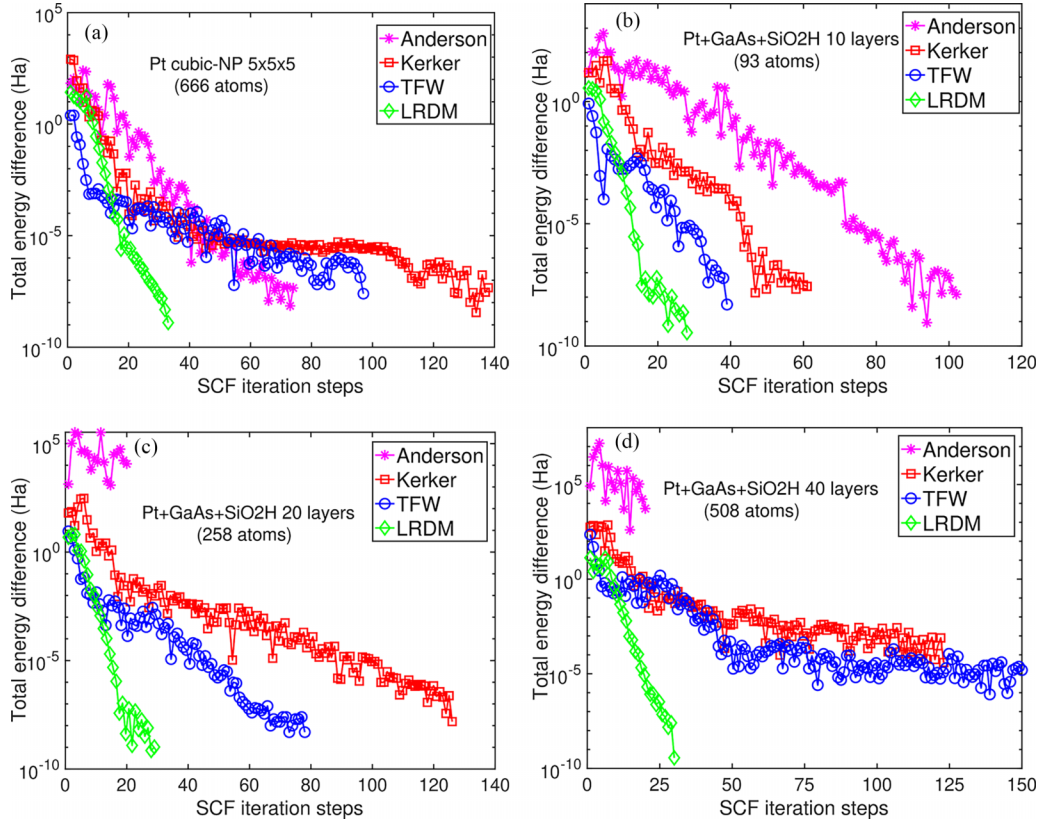
FIG. 1. Convergence of the difference in total energy between successive Kohn-Sham SCF iterations for the various methods—Anderson mixing, Kerker preconditioner (Thomas-Fermi screening approximation of ϵ), TFW preconditioner (Thomas-Fermi–von Weizsacker approximation of ϵ), and the LRDM preconditioner. Absolute values of the total energy difference are plotted against the SCF iteration number (starting from 1).

to finding the root of the residual in the electron density, $R[\rho] = F[V_{\text{eff}}[\rho]] - \rho = 0$. Further, in the neighborhood of the ground state, fast quadratic convergence in the residual can be achieved by using the Newton method:

$$\rho^{(n+1)} = \rho^{(n)} - J^{-1}(F[V_{\text{eff}}[\rho^{(n)}]] - \rho^{(n)}), \qquad (1)$$

where n denotes the nth SCF iteration, and J denotes the Jacobian corresponding to $R[\rho]$ given by

$$J = \frac{\delta F[V_{\text{eff}}]}{\delta V_{\text{eff}}} \frac{\delta V_{\text{eff}}[\rho]}{\delta \rho}\bigg|_{\rho = \rho^{(n)}} - I := \chi_0 K - I = -\epsilon^{\dagger},$$

$$K = \frac{\delta V_{\text{eff}}[\rho]}{\delta \rho}\bigg|_{\rho = \rho^{(n)}} = \frac{\delta V_{\text{xc}}[\rho]}{\delta \rho}\bigg|_{\rho = \rho^{(n)}} + \frac{\delta V_{\text{H}}[\rho]}{\delta \rho}\bigg|_{\rho = \rho^{(n)}}$$

$$:= K_{\text{xc}} + K_{\text{c}}. \qquad (2)$$

In the above, χ_0 is commonly referred to as the independent particle susceptibility operator, ϵ is the dielectric operator, K_{c} is the Coulomb kernel, and K_{xc} is the exchange-correlation kernel, which is usually a small contribution to ϵ compared to K_{c} (random-phase approximation). However, as the exact evaluation of ϵ, in particular the evaluation of χ_0, is computationally expensive, most widely implemented SCF acceleration strategies rely on a quasi-Newton iteration,

$$\rho^{(n+1)} = \rho^{(n)} - \alpha P(F[V_{\text{eff}}[\rho^{(n)}]] - \rho^{(n)}), \qquad (3)$$

where $\alpha \in (0, 1]$ is the damping parameter, and P is a linear operator that approximates the inverse Jacobian of the residual, thereby acting as a preconditioner for the quasi-Newton step. In the neighborhood of the ground-state solution $\rho^{(*)}$, it can easily be shown that the necessary condition for convergence is

$$s(I - \alpha P_* J_*) < 1, \qquad (4)$$

where $s(A)$ denotes the spectral radius of a diagonalizable operator A, J_* is the Jacobian operator evaluated at $\rho^{(*)}$, and P_* is the approximation to the inverse of J_*. The role of α in satisfying the above convergence condition can be understood as follows. Considering an appropriately constructed $P \approx J^{-1}$, such that $P_* J_*$ has a positive real eigenspectrum,[1] we can choose $0 < \alpha \leqslant 1$ to satisfy the convergence condition in Eq. (4). Further, an optimal value of α can be chosen to minimize s or equivalently maximize the linear convergence rate. The corresponding minimal value of s is given by

$$s_{opt} = \frac{\bar{\kappa}(P_* J_*) - 1}{\bar{\kappa}(P_* J_*) + 1}, \qquad (5)$$

---

[1]For the simplest case of $P = -I$, it can be shown that $-J_*$ has a positive real eigenspectrum using the generally assumed random-phase approximation [33,34].
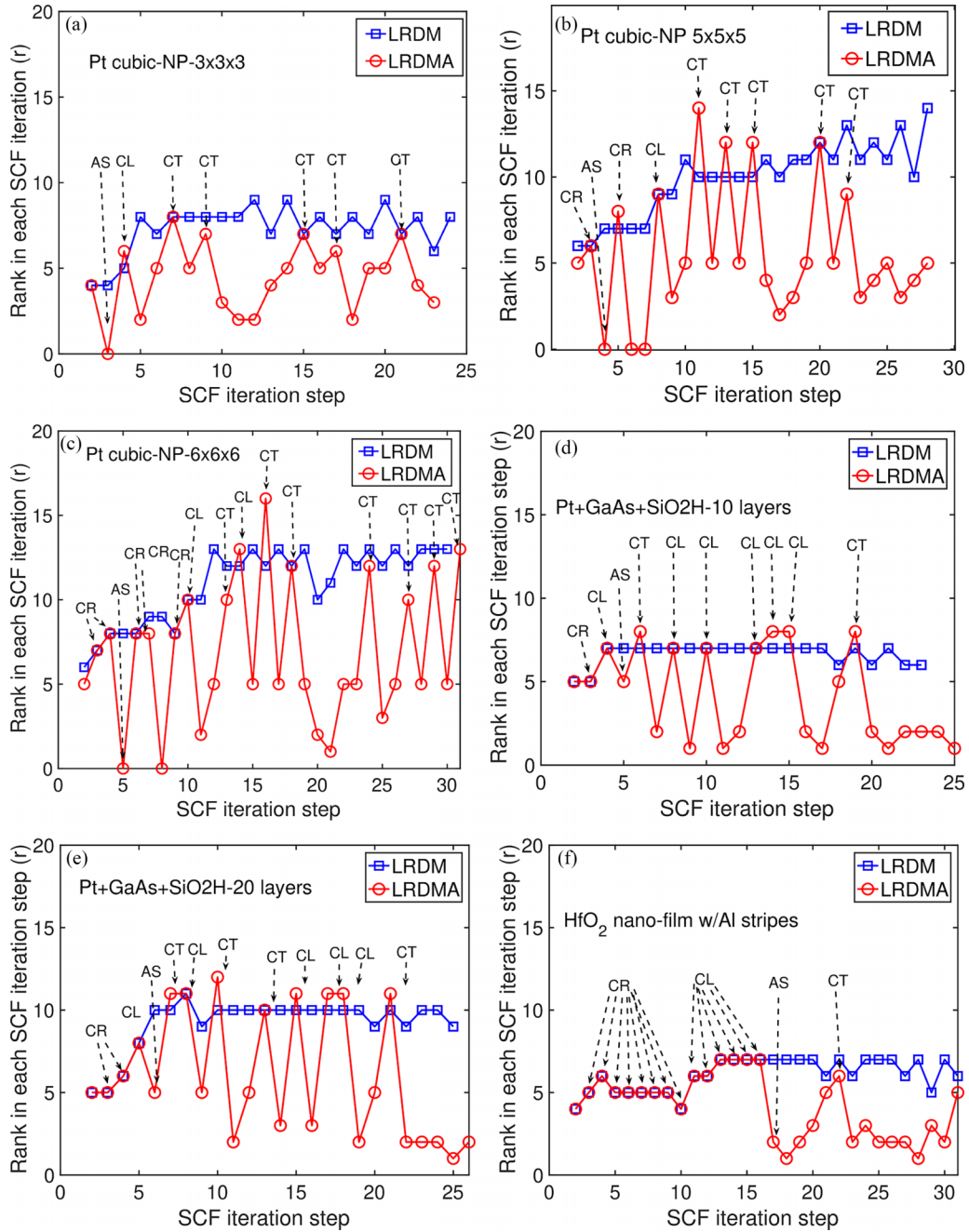
FIG. 2. LRDM and LRDMA ranks in each SCF iteration step for the various heterogeneous benchmark systems. The plot for LRDMA indicates the additional rank-1 updates in each SCF iteration. Further, in the case of LRDMA, the label AS denotes the first accumulated SCF step, and labels CR, CL, and CT denote clearing of accumulation from previous SCF iteration steps due to nonsatisfaction of density residual norm criteria, linearity-indicator criteria, and low-rank approximation error criteria, respectively.

where $\bar{\kappa}(A)$ denotes the condition number of a diagonalizable operator A, given by the ratio of the largest to the smallest eigenvalue magnitude of A. Thus, the slow convergence in the Kohn-Sham SCF iteration arises from the large condition number of J, or, equivalently, the dielectric operator ϵ. Practically, an optimal value of α is difficult to estimate due to the high computational cost of obtaining $\bar{\kappa}(J)$. Further, the lack of a suitable preconditioner P will necessitate the use of small values of α resulting in slow convergence, or for larger values of α may result in divergence of the fixed-point iteration.

Large condition numbers of ϵ arise primarily from the following scenarios encountered in large heterogeneous material systems. The first and the most common one is the divergence of the Coulomb kernel $K_c(|\mathbf{q}|) \propto \frac{1}{|\mathbf{q}|^2}$ as $|\mathbf{q}| \to 0$ in the Fourier space. In metals, it can be shown from the Alder-Wiser expression of χ_0 in Fourier space that $\chi_0(|\mathbf{q}|)$ converges to a finite value as $|\mathbf{q}| \to 0$, overall resulting in divergence of $\chi_0 K_c(|\mathbf{q}|)$ as $|\mathbf{q}| \to 0$. The divergence manifests as the well-known charge-sloshing behavior observed in large metallic materials systems, with the long-wavelength eigenvalue of χ_0 scaling as

L^2, L being the extent of the metallic region. However, in the case of insulators and semiconductors, the long-wavelength divergence of $\chi_0 K_c$ is suppressed as $\chi_0(|\mathbf{q}|) \propto |\mathbf{q}|^2$ as $|\mathbf{q}| \to 0$. We refer to [34] for a detailed discussion on the derivation of the long-wavelength-limit screening behavior. The second important source of instabilities is from the localized d and f orbitals with large density of states near the Fermi energy, which lead to large eigenvalues of χ_0. We refer to [31] for an insightful discussion on such instabilities using a model of a $3d$ impurity in jellium. Importantly, both long-ranged and short-ranged wavelengths can be potentially associated with eigenmodes corresponding to the large eigenvalues of χ_0 [34]. However, current preconditioners primarily approximate the long-ranged modes. Overall, given that a material system can involve any combination of the above scenarios in different spatial regions, a robust and generic preconditioning strategy that appropriately and efficiently accounts for the above sources of large eigenvalues of ϵ has remained a challenge. In this work, we use a low-rank approximation of ϵ to efficiently precondition the SCF iteration that provides robust convergence for complex and large-scale heterogeneous material systems.

B. Approximate Krylov subspace-based low-rank approximation of J

We now present the low-rank approximation of J, or equivalently $-\epsilon^\dagger$, in a continuous real-space setting, and with an adaptive determination of the rank based on an error indicator. The continuous real-space setting is useful to employ the proposed method in conjunction with the ChFSI procedure, as will be discussed in Sec. II C. Furthermore, we propose an adaptively accumulated variant that reuses the low-rank approximation from previous SCF iterations. As will be discussed below, our adaptive accumulation strategy relies on an additional numerical indicator of the linearity of $R[\rho]$ with respect to the electron density. Additionally, we extend the formulation to the spin-polarized case, as will be presented in Sec. II D. We remark that the idea of using a low-rank approximation of the dielectric matrix or related quantities has been explored in electronic-structure calculations, with a recent work employing this idea [41] in the context of extended Lagrangian Born-Oppenheimer molecular dynamics. In [41], which employed self-consistent charge density functional tight-binding theory using a reduced-order basis, the use of a low-rank approximation of J to accelerate the SCF iteration was also suggested. However, it was only demonstrated on a single insulating system of a reactive nitromethane mixture system containing ~50 atoms. The present work applies the low-rank approximation approach to Kohn-Sham DFT using a systematically convergent basis set and demonstrates the robustness of the approach on a wide range of medium- to large-scale benchmark systems that include combinations of metal-insulating-semiconducting heterogeneous systems and magnetic systems. Testing on large-scale systems with metallic regions is critical as long-range charge-sloshing effects during SCF convergence primarily manifest in such systems.

In the LRDM approach, we construct a rank-r approximation J_r^{lr} of J in each SCF iteration (indexed by n), based on generalized directional (Gâteaux) derivatives, t_i, of the residual $R[\rho^{(n)}]$ along orthonormal direction functions u_i:

$$J(\mathbf{x}, \mathbf{x}') \approx J_r^{lr}(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^r t_i(\mathbf{x}) u_i(\mathbf{x}'),$$

$$\text{where } t_i = \left. \frac{\partial R[\rho^{(n)} + \lambda u_i]}{\partial \lambda} \right|_{\lambda=0} \text{ with } \langle u_i, u_j \rangle = \delta_{ij}. \quad (6)$$

In the above, the direction functions u_i are related to the functions in the Krylov subspace of J as will be discussed subsequently. We also note that t_is are related to the density response functions corresponding to the direction functions u_is. We remark that the above expansion converges to J as $r \to \infty$, based on analogy to canonical decomposition of tensors of arbitrary order as a sum of tensor products of rank-1 components [45]. We seek an approximate solution, $\Delta\bar{\rho}$, of the equation $J\Delta\rho = R[\rho^{(n)}]$ that provides the update in the quasi-Newton step $\rho^{(n+1)} = \rho^{(n)} - \alpha\Delta\bar{\rho}$. Since J is not explicitly known, we construct the Krylov subspace progressively as follows as we build the low-rank approximation. We start by choosing the first normalized direction function as $u_1 = R[\rho^{(n)}]/\|R[\rho^{(n)}]\|$, and compute its corresponding Gâteaux derivative: $t_1 = \frac{\partial R[\rho^{(n)} + \lambda u_1]}{\partial \lambda}|_{\lambda=0}$. Subsequently, we choose the remaining orthonormal direction functions u_i $(1 < i \leqslant r)$ using the following iterative procedure based on Gram-Schmidt orthonormalization:

$$u_i = t_{i-1},$$
$$u_i = u_i - \sum_{k=1}^{i-1} \langle u_i, u_k \rangle u_k,$$
$$u_i = u_i / \|u_i\|. \quad (7)$$

Next, we solve $J_r^{lr} \Delta\bar{\rho} = R[\rho^{(n)}]$, whose solution is given by

$$\Delta\bar{\rho} = P_r^{lr} R[\rho^{(n)}] = \sum_{i,j=1}^r u_i S_{ij}^{-1} \langle t_j, R[\rho^{(n)}] \rangle,$$

$$\text{where } S_{ij} = \langle t_i, t_j \rangle, \quad (8)$$

and P_r^{lr} denotes the pseudoinverse of J_r^{lr}.

Overall, in the above low-rank formulation, we have two important parameters, r and α, that significantly control the convergence behavior of the fixed-point iteration. We now discuss strategies for choosing these parameters to achieve robust convergence without case-by-case manual tuning. First, considering r, we note that $J_r^{lr} P_r^{lr} = \sum_{i,j=1}^r t_i S_{ij}^{-1} t_j = I_r^t$ represents the rank-r resolution of the identity operator in the nonorthogonal t_i basis. I_r^t tends to I as $r \to \infty$. This aspect can be used to design an adaptive metric for deciding the rank r. Thereby, we use following error metric based on the relative error of $(I_r^t - I)$ applied to the current residual:

$$s_{rel} = \left\| \sum_{i,j=1}^r t_i S_{ij}^{-1} \langle t_j, R[\rho^{(n)}] \rangle - R[\rho^{(n)}] \right\| / \|R[\rho^{(n)}]\|, \quad (9)$$

where $s_{rel} \to 0$ as $r \to \infty$. In each SCF iteration step, we increase r until s_{rel} goes below a set tolerance, s_{tol}. As will be demonstrated in Sec. III, we find that a value of $s_{tol} \sim 0.3$ is sufficient to achieve accelerated and system-size-independent

convergence for all the heterogeneous benchmark systems considered in this work.

Next, turning our attention to the damping parameter α, if J_r^{lr} closely approximates J, one could use $\alpha = 1$ to obtain a quadratically convergent Newton step [$s_{opt} = 0$ from Eq. (5)]. However, there are a few practical issues with such a strategy when applied to the SCF iterations in KS-DFT. First, the rank r required to obtain a very close approximation to J^{-1} could be very high in a complex heterogeneous system, thus significantly increasing the computational overhead of the preconditioner. Second, the Newton step assumes the starting point $\rho^{(1)}$ to lie within the linear approximation zone of $R[\rho]$ about the solution $\rho^{(*)}$. However, this may not be true for a typical starting initial guess, $\rho^{(1)}$, obtained as a superposition of single atomic electron densities. Particularly, in condensed matter systems, $\rho^{(1)}$ can be quite far from $\rho^{(*)}$. Hence, in this work, we set a small value of $\alpha = 0.1$ for the initial steps until $\|R[\rho^{(n)}]\|$ goes below a threshold of $\mathcal{O}(1)$, below which we increase α to 0.5 to achieve accelerated convergence. This choice is determined based on our numerical experiments and, as will be demonstrated in Sec. III, we obtain robust convergence for all the benchmark systems using this choice of α.

Accumulated low-rank approximation

We now propose an accumulated variant of LRDM, referred to as LRDMA, that accumulates the low-rank approximation of J from previous SCF iteration steps. This could potentially reduce the average rank r of the LRDM during the SCF convergence, thereby reducing the computational overhead. The primary consideration here is that the density response functions of $F[V_{\mathrm{eff}}[\rho]]$ ($\frac{\partial}{\partial\lambda}(F[V_{\mathrm{eff}}[\rho + \lambda u_i]])|_{\lambda=0} = t_i + u_i$) from the previous SCF iteration steps can provide a good approximation to the density response functions of the present iteration, especially if the electron densities from the previous iterations are close to the current iteration. Such a condition is expected to exist close to the solution of the fixed-point iteration $\rho^{(*)}$. However, even when ρ is not necessarily close to $\rho^{(*)}$, such an accumulation can benefit when the densities in the preceding iterations are close to the density in the current iteration and within the linear approximation zone, i.e., the region of the function space where the linear term in the Taylor series is dominant. To this end, we develop a numerical indicator that can guide the adaptive accumulation procedure.

We consider the Taylor series expansion of $R[\rho]$ about $\rho^{(n-1)}$ [density in the $(n-1)$th SCF iteration] to the linear order. The linear approximation of $R[\rho^{(n)}]$ is given by

$$R^{\mathrm{lin}}[\rho^{(n)}] = R[\rho^{(n-1)}] + J(\rho^{(n)} - \rho^{(n-1)})$$
$$= R[\rho^{(n-1)}] - \alpha J \Delta\bar{\rho}$$
$$= R[\rho^{(n-1)}] - \alpha J P_r^{\mathrm{lr}} R[\rho^{(n-1)}],$$

where J is the Jacobian at $\rho = \rho^{(n-1)}$. Although J is unknown, using Eqs. (6) and (8) we can express $R^{\mathrm{lin}}[\rho^{(n)}]$ as

$$R^{\mathrm{lin}}[\rho^{(n)}] = R[\rho^{(n-1)}] - \alpha \sum_{i,j=1}^{r} t_i S_{ij}^{-1} t_j R[\rho^{(n-1)}]$$
$$= (I - \alpha I_r^t) R[\rho^{(n-1)}]. \quad (10)$$

Using the above, we design the following linearity indicator based on the ratio of norms of the predicted and the actual residuals at the nth step:

$$\beta = |\,\|R^{\mathrm{lin}}[\rho^{(n)}]\| / \|R[\rho^{(n)}]\| - 1|. \quad (11)$$

We note that β will be close to zero in the linear approximation zone, and β correlates with the strength of the nonlinear terms in the Taylor series expansion of $R[\rho]$.

The proposed LRDMA approach uses β, s_{rel}, and $\|R[\rho^{(n)}]\|$ to determine whether to use and further accumulate on the low-rank approximation from the previous SCF iteration, or to clear any accumulation and construct the low-rank approximation solely from the current SCF iteration. In order to determine if the density response functions from previous SCF iterations are useful in the current iteration, we first rely on the linearity indicator β. If $\beta > \beta_{\mathrm{tol}}$ then there is sufficient nonlinearity which suggests that using density response functions from previous iterations can have substantial errors and may not provide a good approximation to J for the current iteration. Thus, we clear any rank accumulation from previous iteration (labeled as CL in Fig. 2 to denote clearing based on linearity indicator), and construct the low-rank approximation of J afresh using the direction functions and density response functions from the current SCF iteration. Based on numerical experiments, we find $\beta_{\mathrm{tol}} \sim 0.1$ is a good choice. If $\beta < \beta_{\mathrm{tol}}$, then we check s_{rel}. If $s_{\mathrm{rel}} < s_{\mathrm{tol}}$, then the available low-rank approximation of the Jacobian from the previous SCF iteration is also a good approximation for the current iteration and no further update to the Jacobian is needed. On the other hand, if $s_{\mathrm{rel}} > s_{\mathrm{tol}}$ we will need to decide whether it is beneficial to continue the accumulation of the rank-1 updates using the density response functions from the current iteration, or to purge or clear and construct the low-rank approximation afresh in the current iteration. We note that if the current iterate is far from the solution $\rho^{(*)}$, then the direction response functions computed from previous iterations may not be useful in the current iteration and may result in a very large rank if the accumulation continues. Thus, it is beneficial to clear the accumulation and construct the approximation to the Jacobian in the current iteration. To this end, we use the $\|R[\rho^{(n)}]\|$ as a proxy for how far or close the iterate is from $\rho^{(*)}$. If $\|R[\rho^{(n)}]\| > 1.0$, then we clear the accumulation (labeled as CR in Fig. 2 to denote clearing based on residual indicator). If a clearing of the accumulation based on the linearity indicator and the residual indicator is not triggered, we use the low-rank approximation from the previous SCF iteration ($J_{r,(n-1)}^{\mathrm{lr}}$) and continue the accumulation using rank-1 updates from the current iteration to improve the Jacobian approximation as

$$J_{(r+r_n),(n)}^{\mathrm{lr}}(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{r_n} t_i^{(n)}(\mathbf{x}) u_i^{(n)}(\mathbf{x}') + J_{r,(n-1)}^{\mathrm{lr}}(\mathbf{x}, \mathbf{x}'). \quad (12)$$

In the above, r_n is the additional rank added in the nth SCF iteration and the direction functions $u_i^{(n)}$ are chosen from the Krylov subspace of J at the current nth SCF iteration, but orthogonalized with respect to all the other direction functions. We note that $u_1^{(n)}$ is chosen to be the normalized orthogonal complement of $R[\rho^{(n)}]$ to the previous history of direction functions. We continue the accumulation until either $s_{\mathrm{rel}} < s_{\mathrm{tol}}$

or $r_n = r_{\text{iter}}^{\max}$, with $r_{\text{iter}}^{\max} = 5$. If the tolerance condition on s_{rel} is not reached, we also trigger an accumulation clearing event for the next SCF step (labeled as CT in Fig. 2). We investigate the efficiency and robustness of the proposed LRDMA method in comparison to the LRDM method in Sec. III C, where the numerical results demonstrate a reduction in the average rank across most benchmark systems resulting in improved computational efficiency.

C. First-order density response computation in Chebyshev filtered subspace

We now develop an efficient numerical methodology for the computation of the first-order density response functions ($\frac{\partial}{\partial\lambda}(F[[V_{\text{eff}}[\rho^{(n)} + \lambda u_i]])|_{\lambda=0}$), i.e., the Gâteaux derivatives along the direction functions, in Kohn-Sham DFT. We focus on Kohn-Sham DFT calculations using systematically convergent complete basis sets, such as the finite-element basis, where the nonlinear Kohn-Sham eigenvalue problem [cf. Eq. (1)] is discretized with M nonorthogonal real-space basis functions, with $M \gg N$. The main aspect of our numerical method is to approximately compute the first-order density response in a finite-dimensional subspace with a dimension that is much smaller compared to M, and further take advantage of the efficient canonical density-matrix perturbation approach [46] that avoids explicit computation of first-order perturbations in the wavefunctions.

To begin, we first briefly discuss the ChFSI procedure in real-space Kohn-Sham DFT [3] to compute the electron density, and subsequently detail our approach to compute the first-order density response. We choose our real-space basis functions to be higher-order spectral finite element (FEs) $\{l_a\}_{(1 \leqslant a \leqslant M)}$ that are strictly local piecewise polynomial basis functions. Since the spectral FE basis is nonorthogonal, resulting in a generalized Hermitian eigenvalue problem (GHEP), the Löwdin orthogonalized FE basis functions (denoted as $\{p_a\}_{(1 \leqslant a \leqslant M)}$) are used to obtain a standard Hermitian eigenvalue problem (SHEP) [42]:

$$\widetilde{\mathbf{H}}[\rho]\tilde{\boldsymbol{\psi}}_k = \varepsilon_k^h \tilde{\boldsymbol{\psi}}_k, \quad k = 1, 2, \ldots, N \quad \text{with} \quad N > \frac{N_e}{2}, \quad (13)$$

where $\widetilde{\mathbf{H}}[\rho]$ is the discrete Kohn-Sham Hamiltonian in the $\{p_a\}$ basis, and $\tilde{\boldsymbol{\psi}}_k$ are the expansion coefficients of the single-electron Kohn-Sham eigenfunctions in the $\{p_a\}$ basis. In order to solve the nonlinear SHEP, we use the ChFSI procedure, which exploits the fact that we are only interested in the occupied eigensubspace that is a very small portion at the lower end of the spectrum of $\widetilde{\mathbf{H}}$. Instead of solving Eq. (13) exactly in every SCF step, the ChFSI procedure progressively approximates the occupied eigensubspace. In particular, this involves applying a Chebyshev polynomial filter of degree c, $T_c(\bar{\mathbf{H}})$, to a trial subspace $\mathbf{X}$ with dimension N. Here, $\bar{\mathbf{H}}$ denotes a scaled and shifted Hamiltonian constructed from $\widetilde{\mathbf{H}}$ such that the wanted (occupied) spectrum is mapped to $(-\infty, -1)$ and the unwanted spectrum is mapped to $[-1, 1]$. As $T_c(y)$ monotonically and rapidly increases for decreasing values of $y < -1$, the action of $T_c(\bar{\mathbf{H}})$ on $\mathbf{X}$ results in a filtered subspace $\widetilde{\boldsymbol{\Psi}}_{\mathbf{f}}$ that is a close approximation to the eigensubspace corresponding to the occupied spectrum of $\widetilde{\mathbf{H}}[\rho]$.

Subsequently, $\widetilde{\boldsymbol{\Psi}}_{\mathbf{f}}$ is orthonormalized to obtain $\widetilde{\boldsymbol{\Psi}}_{\mathbf{o}}$, and the discrete SHEP is projected onto the subspace spanned by $\widetilde{\boldsymbol{\Psi}}_{\mathbf{o}}$ to solve the eigendecomposition of the projected Hamiltonian:

$$\hat{\mathbf{H}}\mathbf{Q} = \mathbf{Q}\mathbf{D}, \quad \text{where} \quad \hat{\mathbf{H}} = \widetilde{\boldsymbol{\Psi}}_{\mathbf{o}}^{\dagger}\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{o}}, \quad (14)$$

where $\mathbf{Q}$ is the matrix comprised of the eigenvectors and $\mathbf{D}$ is a diagonal matrix with the corresponding eigenvalues, represented in the Chebyshev filtered subspace. The corresponding eigenvectors in the Löwdin orthogonalized FE basis are obtained from the transformation $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}} = \widetilde{\boldsymbol{\Psi}}_{\mathbf{o}}\mathbf{Q}$. Finally, the output electron density ($F[V_{\text{eff}}[\rho]](\mathbf{x}) = \rho_{\text{out}}(\mathbf{x})$) in each SCF iteration is computed as

$$\rho_{\text{out}}(\mathbf{x}) = 2\,(\boldsymbol{n}^e(\mathbf{x}))^T[\mathbf{M}^{-1/2}\,\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\,f(\mathbf{D}, \mu)\,\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}\,\mathbf{M}^{-1/2^{\dagger}}]\boldsymbol{n}^e(\mathbf{x}), \quad (15)$$

where $\boldsymbol{n}^e(\mathbf{x}) = [l_1^e(\mathbf{x})\,l_2^e(\mathbf{x})\cdots l_{M_{cell}}^e(\mathbf{x})]^T$ denotes the FE basis functions associated with the given finite-element cell (M_{cell} denotes the number of nodes in the cell), and $\mathbf{M}$ is the positive-definite and symmetric finite-element overlap matrix ($\mathbf{M}_{ab} = \langle l_a, l_b\rangle$) in the original nonorthogonal FE basis. We refer to [6,8,42] for more details on the algorithmic aspects of employing ChFSI in Kohn-Sham DFT calculations using the FE basis.

The electron-density expression in Eq. (15) can also equivalently be obtained from $\boldsymbol{\Gamma}_{\text{proj}} = f(\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}, \mu)$, which denotes the density matrix corresponding to the projection of $\widetilde{\mathbf{H}}$ onto $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$. The resulting electron-density expression is given by

$$\rho_{\text{out}}(\mathbf{x}) = 2\,(\boldsymbol{n}^e(\mathbf{x}))^T[\mathbf{M}^{-1/2}\,\boldsymbol{\Gamma}_{\text{proj}}\,\mathbf{M}^{-1/2^{\dagger}}]\boldsymbol{n}^e(\mathbf{x}). \quad (16)$$

The equivalence between the two expressions for the electron density can be seen from

$$\boldsymbol{\Gamma}_{\text{proj}} = f(\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}, \mu)$$

$$= \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}f(\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}, \mu)\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger} = \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}f(\mathbf{D}, \mu)\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}, \quad (17)$$

where the second equality follows from the power-series representation of the analytic function $f(\varepsilon, \mu)$ combined with the spectral decomposition of the Hermitian matrix $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^{\dagger}$ [11]. In the above, we note that $\boldsymbol{\Gamma}_{\text{proj}}$ systematically approaches the density matrix corresponding to the discrete FE Hamiltonian, $\boldsymbol{\Gamma} = f(\widetilde{\mathbf{H}}[\rho], \mu)$, as the SCF approaches convergence. This follows from the ChFSI procedure, where $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$ progressively approaches the occupied eigensubspace of $\widetilde{\mathbf{H}}[\rho]$ as the SCF approaches convergence, i.e., $\widetilde{\mathbf{H}}[\rho] \rightarrow \widetilde{\mathbf{H}}[\rho^{(*)}]$.

We now consider a perturbation to the input electron density ρ, in a given SCF iteration step, along the direction function u_i expressed as $\rho(\lambda) = \rho + \lambda u_i$. We seek to compute the response in ρ_{out} to first order, corresponding to the perturbations. The first-order response of ρ_{out} can be obtained from the explicit first-order response of all the occupied eigenfunctions and eigenvalues. While the first-order response in the eigenvalues can be obtained without significant computational overhead by using the first-order perturbation theory involving only the first-order response of the Hamiltonian [47], the first-order response in eigenfunctions requires a solution of the

Sternheimer equations that are linear systems of equations of dimension M, one for each eigenfunction perturbation. The solution to Sternheimer equations involves computation of the perturbations in a large subspace of the full finite-dimensional space that is orthogonal to the occupied eigensubspace [48]. This results in a computational cost that is $\mathcal{O}(MN_e^2)$, and typically with high prefactors. Thus, in systematically convergent real-space basis sets, where $M \gg N$, solution of Sternheimer equations can result in substantial computational overheads. In this work, we adopt a different approach. We consider the first-order perturbation response of the FE discretized Hamiltonian,

$$\widetilde{\mathbf{H}}[\rho(\lambda)] = \widetilde{\mathbf{H}}[\rho] + \lambda \widetilde{\mathbf{H}}', \quad \widetilde{\mathbf{H}}' = \left.\frac{\partial \widetilde{\mathbf{H}}[\rho + \lambda u_i]}{\partial \lambda}\right|_{\lambda=0}, \quad (18)$$

and approximately compute the first-order density response corresponding to the orthogonal projection of $\widetilde{\mathbf{H}}[\rho(\lambda)]$ onto an unperturbed subspace of small dimension ($\ll M$), which we choose to be $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$. Utilizing the density-matrix-based representation of ρ_{out} in Eq. (16), we obtain the approximate density response as

$$\frac{\partial}{\partial \lambda}(F[V_{\text{eff}}[\rho + \lambda u_i]])|_{\lambda=0}$$
$$\approx 2\,(\boldsymbol{n}^e(\mathbf{x}))^T\big[\mathbf{M}^{-1/2}\,\boldsymbol{\Gamma}^{(1)}_{\text{proj}}\,\mathbf{M}^{-1/2^\dagger}\big]\boldsymbol{n}^e(\mathbf{x}). \quad (19)$$

In the above, $\boldsymbol{\Gamma}^{(1)}_{\text{proj}}$ is the approximate first-order density-matrix response computed in the unperturbed subspace $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$. Using Eq. (17), we arrive at

$$\boldsymbol{\Gamma}^{(1)}_{\text{proj}} = \frac{\partial}{\partial \lambda}(f(\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^\dagger(\widetilde{\mathbf{H}} + \lambda\widetilde{\mathbf{H}}')\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^\dagger, \mu + \lambda\mu'))|_{\lambda=0}$$
$$= \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\frac{\partial}{\partial \lambda}(f(\breve{\mathbf{H}} + \lambda\breve{\mathbf{H}}', \mu + \lambda\mu'))|_{\lambda=0}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^\dagger, \quad (20)$$

where $\breve{\mathbf{H}} = \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^\dagger\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}} = \mathbf{D}$, $\breve{\mathbf{H}}' = \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^\dagger\widetilde{\mathbf{H}}'\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$. We remark that there are two controllable approximations in the computation of $\boldsymbol{\Gamma}^{(1)}_{\text{proj}}$. The first is from the consideration of an unperturbed subspace $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$ of dimension $N \ll M$. Since $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$ is an orthonormal basis, the associated approximation error can be systematically decreased by increasing the dimension N of the subspace $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$, albeit with an increase in the computational cost. Practically, based on our numerical experiments, we find that the additional buffer states already used in the ChFSI procedure [3,42] provide a large enough subspace size to obtain a sufficiently accurate $\boldsymbol{\Gamma}^{(1)}_{\text{proj}}$ for the purpose of SCF preconditioning. The second approximation arises from the nature of the ChFSI procedure itself—we compute the density-matrix response $\boldsymbol{\Gamma}^{(1)}_{\text{proj}}$ about the Chebyshev-filtered density matrix $\boldsymbol{\Gamma}_{\text{proj}}$ that closely approximates $\boldsymbol{\Gamma}$ in each SCF iteration step, with the approximation error systematically decreasing with SCF convergence. Thus, $\boldsymbol{\Gamma}^{(1)}_{\text{proj}}$ can be considered to progressively approximate the true density-matrix response $\boldsymbol{\Gamma}^{(1)} = \frac{\partial}{\partial \lambda}(f(\widetilde{\mathbf{H}}[\rho(\lambda)], \mu(\lambda)))|_{\lambda=0}$, within the constraint of the small unperturbed subspace $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$. Again, practically, we find that the typical choices for the Chebyshev polynomial degree c and the eigenpair residual norm tolerance for the occupied

states, $\|\widetilde{\mathbf{H}}\widetilde{\Psi}_{\mathbf{r},i} - \varepsilon_i^h\widetilde{\Psi}_{\mathbf{r},i}\|$, are sufficient for the purpose of preconditioning.

We now briefly discuss the approach for an efficient computation of the Gâteaux derivative of the subspace projected density matrix appearing in Eq. (20),

$$\breve{\boldsymbol{\Gamma}}^{(1)} = \frac{\partial}{\partial \lambda}(f(\breve{\mathbf{H}} + \lambda\breve{\mathbf{H}}', \mu + \lambda\mu'))|_{\lambda=0}, \quad (21)$$

without performing additional eigendecomposition calculations that will result in significant overhead for the proposed low-rank preconditioning approach. Since $\breve{\boldsymbol{\Gamma}}^{(1)}$ can be interpreted as the first-order perturbation response of $f(\breve{\mathbf{H}}[\rho(\lambda)], \mu(\lambda))$, we employ canonical density-matrix perturbation theory [46], which is based on perturbation of the recursive Fermi operator expansion of $f(\breve{\mathbf{H}}[\rho(\lambda)], \mu(\lambda))$:

$$\frac{\partial}{\partial \lambda}(f(\breve{\mathbf{H}}[\rho(\lambda)], \mu(\lambda)))|_{\lambda=0}$$
$$\approx \frac{\partial}{\partial \lambda}(\mathcal{P}_T(\mathcal{P}_{T-1}(\cdots\mathcal{P}_0(\breve{\mathbf{H}}[\rho(\lambda)], \mu(\lambda))\dots)))|_{\lambda=0}, \quad (22)$$

where T denotes the degree of the recursive expansion and

$$\boldsymbol{X}_0 = \mathcal{P}_0(\breve{\mathbf{H}}[\rho(\lambda)]) = 0.5\boldsymbol{I} - 2^{-(T+2)}\frac{\breve{\mathbf{H}}[\rho(\lambda)] - \mu(\lambda)\boldsymbol{I}}{k_B T},$$

$$\boldsymbol{X}_n = \mathcal{P}_n(\boldsymbol{X}_{n-1}) = \frac{\boldsymbol{X}_{n-1}^2}{\boldsymbol{X}_{n-1}^2 + (\boldsymbol{I} - \boldsymbol{X}_{n-1})^2}. \quad (23)$$

We note that μ' in Eq. (21) is obtained from the requirement of traceless density-matrix response, $\text{Tr}[\breve{\boldsymbol{\Gamma}}^{(1)}] = 0$, for ensuring the conservation of number of electrons. The Padé polynomial functions $\mathcal{P}_n(\boldsymbol{X}_{n-1})$ used in the above recursive expansion enable rapid convergence in the density-matrix first-order response calculations [41,46], with $T = 8$–10 found to be sufficient in numerical experiments. This aspect is also supported by our numerical studies in this work, where we find the SCF convergence for the various benchmark systems to be insensitive to values of T beyond 10. Importantly, the above recursive expansion further exploits the fact that $\breve{\mathbf{H}}[\rho(\lambda = 0)]$ is a diagonal matrix, thereby rendering the algorithm to only consist of computationally cheap matrix-vector multiplications and diagonal matrix inversion operations [41]. Our numerical benchmarks indeed show that evaluation of Eq. (22) is a negligible cost compared to the $\mathcal{O}(MN_e^2)$ scaling steps involving the computation of $\breve{\mathbf{H}}' = \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}^\dagger\widetilde{\mathbf{H}}'\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}$ and $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}}\breve{\boldsymbol{\Gamma}}^{(1)}$, required for evaluating the density response in Eq. (19). However, since the above operations are performed only once for each direction response computation, they incur a much lower computational prefactor compared to the iterative solution of Sternheimer equations requiring $\mathcal{O}(MN_e^2)$ computations in each iteration step. Further, due to the involvement of dense matrix-matrix multiplications, these operations can be performed very efficiently on hybrid CPU-GPU architectures as will be demonstrated in Sec. III D.

Finally, we mention that the following changes are required for extending the density response approach to periodic systems which entails integration over the first Brillouin zone (BZ). The first consideration is the choice of the unperturbed subspace for the response contribution corresponding to each k point, $\mathbf{k}$, in the BZ. We choose this unperturbed

subspace to be the approximate occupied eigensubspace of the **k**-dependent FE discretized Hamiltonian, $\widetilde{\mathbf{H}}_{\mathbf{k}}[\rho(\lambda = 0)]$, obtained from the ChFSI procedure. Second, the constraint on conservation of number of electrons is extended to BZ sampling:

$$\oint_{BZ} \int_{\Omega} 2\, \boldsymbol{n}^{e^T}(\mathbf{x}) \big[\mathbf{M}^{-1/2}\, \boldsymbol{\Gamma}^{(1)}_{\mathrm{proj},\mathbf{k}}\, \mathbf{M}^{-1/2^{\dagger}}\big] \boldsymbol{n}^e(\mathbf{x})\, d\mathbf{x}\, d\mathbf{k} = 0, \tag{24}$$

where the value μ' is now determined using the above constraint.

We remark that this diagonalization free first-order perturbation response methodology can, in principle, be applied to other finite-temperature smearing schemes, such as Methfessel-Paxton smearing [49] and cold smearing [50], in conjunction with appropriate expansions [51]. Methfessel-Paxton and cold-smearing schemes allow for usage of relatively higher smearing temperatures compared to Fermi-Dirac smearing due to their lower finite-temperature errors in ground-state internal energy and forces, and hence are popularly used in various DFT codes to reduce charge-sloshing effects. We will demonstrate in Sec. III the robust performance of the developed LRDM preconditioner for bulk and heterogeneous metallic systems with first BZ sampling, while employing low to modest Fermi-Dirac smearing temperatures of $T = 10$ K and $T = 500$ K. Thus, the requirement of higher smearing temperatures to accelerate SCF convergence can be avoided.

D. Extension of low-rank approximation of J to spin-polarized Kohn-Sham DFT

We develop an extension of the LRDM preconditioner discussed in Sec. II B to collinear spin-polarized Kohn-Sham DFT calculations. To this end, we define the residual in the spin-polarized case to be given by

$$R[(\rho_{\uparrow}, \rho_{\downarrow})] = F[V_{\mathrm{eff}}[(\rho_{\uparrow}, \rho_{\downarrow})]] - (\rho_{\uparrow}, \rho_{\downarrow}), \tag{25}$$

where $(\rho_{\uparrow}, \rho_{\downarrow}) \in \Upsilon \times \Upsilon$ denotes an ordered pair of spin-up and spin-down density functions belonging to the Cartesian product of the function space Υ with itself. $F[V_{\mathrm{eff}}[(\rho_{\uparrow}, \rho_{\downarrow})]] : \Upsilon \times \Upsilon \to \Upsilon \times \Upsilon$ represents the Kohn-Sham input-to-output spin-polarized density map in each SCF iteration. The inner product between two ordered pairs of functions $(f_1, f_2) \in \Upsilon \times \Upsilon$ and $(g_1, g_2) \in \Upsilon \times \Upsilon$ is defined as

$$\langle (f_1, f_2), (g_1, g_2) \rangle = \langle f_1, g_1 \rangle + \langle f_2, g_2 \rangle. \tag{26}$$

The associated norm induced by the inner product for an ordered pair (f_1, f_2) is given by $\|(f_1, f_2)\| = \sqrt{\langle (f_1, f_2), (f_1, f_2) \rangle}$.

We solve for $R[(\rho_{\uparrow}, \rho_{\downarrow})] = 0_{\Upsilon \times \Upsilon}$ using a damped quasi-Newton iteration scheme:

$$(\rho_{\uparrow}^{(n+1)}, \rho_{\downarrow}^{(n+1)}) = (\rho_{\uparrow}^n, \rho_{\downarrow}^n) - \alpha\, P\, R[(\rho_{\uparrow}^n, \rho_{\downarrow}^n)], \tag{27}$$

$$P \approx J^{-1} \quad \text{with} \quad J = \frac{\partial R[(\rho_{\uparrow}, \rho_{\downarrow})]}{\partial (\rho_{\uparrow}, \rho_{\downarrow})}\bigg|_{(\rho_{\uparrow}, \rho_{\downarrow}) = (\rho_{\uparrow}^n, \rho_{\downarrow}^n)}, \tag{28}$$

where $\alpha \in (0, 1]$, and $P : \Upsilon \times \Upsilon \to \Upsilon \times \Upsilon$ approximates the inverse Jacobian of the residual. Analogous to the spin-restricted case, the Jacobian in the spin-polarized case also

admits a canonical decomposition into a sum of tensor products of rank-1 components. In particular, we consider the following rank-r approximation form:

$$J_r^{\mathrm{lr}}(\mathbf{x}, \mathbf{x}', \sigma, \sigma') = \sum_{i=1}^{r} (t_{\uparrow}, t_{\downarrow})_{\sigma}^i(\mathbf{x})(u_{\uparrow}, u_{\downarrow})_{\sigma'}^i(\mathbf{x}'),$$

$$\text{where} \quad (t_{\uparrow}, t_{\downarrow})^i = \frac{\partial}{\partial \lambda}(R[(\rho_{\uparrow}^n, \rho_{\downarrow}^n) + \lambda(u_{\uparrow}, u_{\downarrow})^i])|_{\lambda=0}. \tag{29}$$

In the above, $(t_{\uparrow}, t_{\downarrow})^i \in \Upsilon \times \Upsilon$ represents the generalized directional derivative of $R[(\rho_{\uparrow}, \rho_{\downarrow})]$ along an ordered pair of orthonormal direction functions, $(u_{\uparrow}, u_{\downarrow})^i \in \Upsilon \times \Upsilon$. Further, $(\cdot, \cdot)_{\sigma}$ denotes the choice of one of the functions $\sigma = \uparrow$ or $\sigma = \downarrow$ from the ordered pair. The orthonormality condition is given by $\langle (u_{\uparrow}, u_{\downarrow})^i, (u_{\uparrow}, u_{\downarrow})^j \rangle = \delta_{ij}$. Following Sec. II B, for efficient convergence of the preconditioner P_r^{lr} with respect to r, the ordered pairs $(u_{\uparrow}, u_{\downarrow})^i$ are obtained from an approximate r-rank Krylov subspace of J, where we first choose $(u_{\uparrow}, u_{\downarrow})^1 = R[(\rho_{\uparrow}^n, \rho_{\downarrow}^n)]/\|R[(\rho_{\uparrow}^n, \rho_{\downarrow}^n)]\|$ and subsequently use the following iterative procedure for $1 < i \leqslant r$:

$$(u_{\uparrow}, u_{\downarrow})^i = (t_{\uparrow}, t_{\downarrow})^{i-1},$$

$$(u_{\uparrow}, u_{\downarrow})^i = (u_{\uparrow}, u_{\downarrow})^i - \sum_{k=1}^{i-1} \langle (u_{\uparrow}, u_{\downarrow})^i, (u_{\uparrow}, u_{\downarrow})^k \rangle (u_{\uparrow}, u_{\downarrow})^k,$$

$$(u_{\uparrow}, u_{\downarrow})^i = (u_{\uparrow}, u_{\downarrow})^i / \|(u_{\uparrow}, u_{\downarrow})^i\|. \tag{30}$$

The spin-polarized extension of the remaining aspects of the low-rank formulation, including the accumulated variant, follow along similar lines as discussed in Sec. II B.

Finally, we mention the additional considerations in extending the Chebyshev filtered subspace projected density response computation discussed in Sec. II C to the spin-polarized case. Adopting a similar approach as the spin-restricted case, we approximately evaluate the density-matrix response corresponding to the projection of $\widetilde{\mathbf{H}}_{\sigma}[(\rho_{\uparrow}, \rho_{\downarrow})(\lambda)]$ onto $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}, \sigma}$ as

$$\boldsymbol{\Gamma}^{(1)}_{\mathrm{proj},\sigma} = \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}, \sigma} \frac{\partial}{\partial \lambda}(f(\breve{\mathbf{H}}_{\sigma} + \lambda \breve{\mathbf{H}}_{\sigma}', \mu + \lambda \mu'))|_{\lambda=0} \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}, \sigma}^{\dagger}, \tag{31}$$

where

$$\breve{\mathbf{H}}_{\sigma}' = \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}, \sigma}^{\dagger} \frac{\partial}{\partial \lambda}(\widetilde{\mathbf{H}}_{\sigma}[(\rho_{\uparrow}, \rho_{\downarrow}) + \lambda(u_{\uparrow}, u_{\downarrow})^i])|_{\lambda=0} \widetilde{\boldsymbol{\Psi}}_{\mathbf{r}, \sigma} \tag{32}$$

is the first-order response of $\widetilde{\mathbf{H}}_{\sigma}[(\rho_{\uparrow}, \rho_{\downarrow})(\lambda)]$ along $(u_{\uparrow}, u_{\downarrow})^i$ projected onto $\widetilde{\boldsymbol{\Psi}}_{\mathbf{r}, \sigma}$. Subsequently, the desired directional derivative of $F[V_{\mathrm{eff}}[(\rho_{\uparrow}, \rho_{\downarrow})]]$ along $(u_{\uparrow}, u_{\downarrow})^i$ is obtained as

$$\frac{\partial}{\partial \lambda}(F[V_{\mathrm{eff}}[(\rho_{\uparrow}, \rho_{\downarrow}) + \lambda(u_{\uparrow}, u_{\downarrow})^i]]_{\sigma})|_{\lambda=0}$$

$$\approx (\boldsymbol{n}^e(\mathbf{x}))^T \big[\mathbf{M}^{-1/2}\, \boldsymbol{\Gamma}^{(1)}_{\mathrm{proj},\sigma}\, \mathbf{M}^{-1/2^{\dagger}}\big] \boldsymbol{n}^e(\mathbf{x}). \tag{33}$$

Finally, we remark that the above numerical method for the first-order density response calculations and the proposed spin-polarized LRDM formulation are adaptable to other discretizations (plane waves, finite difference, wavelets) and eigensolver strategies other than the ChFSI procedure [Davidson, residual minimization method (RMM)-DIIS] implemented in various Kohn-Sham DFT codes. For instance, using the Davidson iterative eigensolver implemented in the

QUANTUM ESPRESSO software (plane-wave basis), the smaller subspace required for the efficient first-order density response calculations can be constructed from the eigenpairs computed in the Davidson method. Other aspects of the LRDM method including the use of the canonical density-matrix perturbation approach are agnostic to the choice of the discretization and eigensolver strategy.

III. RESULTS AND DISCUSSION

In this section, we demonstrate the robustness and computational efficiency of the LRDM preconditioner for Kohn-Sham DFT calculations. We compare the performance against three other widely used preconditioners, namely, Anderson mixing [26], Anderson mixing with Kerker preconditioner [35], and Broyden mixing with TFW preconditioner [39]. We consider various heterogeneous benchmark systems with system sizes up to $\sim$1100 atoms ($\sim$20 000 electrons) and study the convergence of LRDM preconditioner and the accumulated variant, LRDMA. Furthermore, we also demonstrate the robustness of the proposed extension of the LRDM preconditioner to spin-polarized DFT calculations. Finally, we comment on the computational cost of the LRDM preconditioner, and compare against Anderson and Kerker preconditioners on hybrid CPU-GPU architectures.

A. General calculation details

In all the DFT calculations reported in this work, we use the Perdew-Burke-Ernzerhof exchange-correlation functional [52] and ONCV [53] pseudopotentials from the PseudoDojo [54] database. Unless otherwise specified, we use Fermi-Dirac smearing with $T = 500$ K in all our simulations. Additionally for periodic benchmark systems, we use shifted Monkhorst-Pack k-point grids to sample the first Brillouin zone with the minimum k-point spacing chosen to be $\sim$0.3 Å^{-1}.

We have implemented the LRDM preconditioner in the DFT-FE software [6,8,42], a recently developed open-source code for massively parallel large-scale real-space Kohn-Sham DFT calculations based on a finite-element discretization. DFT-FE already has implementations of Anderson mixing and Anderson with Kerker preconditioning. Thus, in this work, simulations using Anderson and Kerker preconditioners are performed using DFT-FE, while simulations using the TFW preconditioner are performed using the implementation available in QUANTUM ESPRESSO software (QE) [55,56] using the same ONCV pseudopotential input files. All simulations using DFT-FE for benchmarking the LRDM, Anderson, and Kerker preconditioners are performed till 10^{-5} stopping tolerance in the L_2 norm of electron-density residual ($\|R[\rho]\|$). The simulations using QE for the TFW preconditioner employ 5×10^{-8} Ha as stopping tolerance in the total energy between consecutive SCF iteration steps. We remark that these two stopping criteria are approximately equivalent for the range of system sizes considered in this work, as will be demonstrated in the convergence studies reported in Sec. III B. Further, we note that the finite-element and plane-wave discretization parameters for DFT-FE and QE are chosen such that we obtain $\sim$10 meV/atom accuracy in the ground-state

energies. We additionally remark that the different eigensolver implementations—the ChFSI procedure in DFT-FE and the Davidson iteration diagonalization in QE—can lead to differences in the SCF convergence behavior for the Anderson multisecant method and its coupling with Kerker. We refer to Table I in the Supplemental Material [57] for a numerical comparison of the multisecant preconditioner performance, both with and without the Kerker preconditioner, between DFT-FE and QE on a subset of the benchmark systems studied below. The benchmark results demonstrate similar trends between DFT-FE and QE implementations.

Next, we mention the choice of the damping parameter, α, and other preconditioner-specific choices used in the benchmark studies. First, in the case of the LRDM preconditioner, we choose $s_{\mathrm{tol}} = 0.2$–0.3 for adaptively setting the rank r in each SCF iteration step [cf. Eq. (9)]. Further, we choose the damping parameter $\alpha = 0.1$ till $\|R[\rho]\|$ goes below 2.0, and subsequently switch to $\alpha = 0.5$. Additionally, for the accumulated variant of the LRDM preconditioner proposed in Sec. II B (LRDMA), we use the tolerance on the linearity indicator to be $\beta_{\mathrm{tol}} = 0.1$, the maximum additional rank that can be accumulated in an SCF iteration step to be $r_{\mathrm{iter}}^{\mathrm{max}} = 5$, and further allow accumulation only when $\|R[\rho]\| < 1.0$. The above parameters are determined based on numerical experiments, and provide a robust convergence for a wide range of benchmark material systems with increasing complexity and system sizes considered in this work, without any system specific tuning in the parameter values. In the SCF convergence studies to be discussed below, we also use the LRDM preconditioner computational framework to approximately compute the condition number of J, $\kappa(J)$, at the converged ground-state solution $\rho^{(*)}$, to assess the difficulty of convergence of the SCF iteration for each material system. Specifically, for estimating $\kappa(J)$, we use a stringent low-rank approximation tolerance of $s_{\mathrm{tol}} = 5 \times 10^{-4}$ along with employing a relatively larger Chebyshev filtered eigensubspace ($\sim$25% buffer over $N_e/2$) to obtain a closer approximation J_r^{lr} to J, and finally perform a power iteration on J_r^{lr} and P_r^{lr} to obtain estimates of the highest and lowest eigenvalues. Second, in the case of Anderson mixing, we use values of α between 0.015 and 0.05 and a mixing history range of 20–50. The relatively small values of α for Anderson mixing are required due to the large condition numbers of J for the heterogeneous material systems considered in this work. Third, in the case of Anderson mixing with Kerker preconditioner, we use $\alpha = 0.5$, mixing history of 20, and Thomas-Fermi screening wave-vector value of 0.8 bohr^{-1}. In the case of TFW preconditioner, the parameters are set to be $\alpha = 0.5$–0.7 and a mixing history of 15. The above parameter choices for Kerker and TFW preconditioners are determined based on numerical experiments to be broadly optimal for all the benchmark systems considered in this work. Finally, we remark that for the LRDM, Anderson, and Kerker preconditioners implemented in DFT-FE, we perform multiple sweeps of the ChFSI procedure in each SCF iteration till the residual norm of the eigenpair closest to the Fermi energy is below a tolerance of 2×10^{-3}.

All the numerical simulations with computational times reported in this work were executed using the using the Phase 1 GPU-accelerated nodes of the NERSC Perlmutter supercomputer, with each node containing four NVIDIA A100

TABLE I. Comparison of the performance of LRDM against other preconditioners. All DFT calculations are spin restricted, and use k-point sampling of the Brillouin zone for semiperiodic systems. A stopping tolerance of 10^{-5} in the L_2 norm of the electron-density residual ($\|R[\rho]\|$) is used. Based on numerical experiments, this is approximately equivalent to a stopping tolerance of 5×10^{-8} Ha in the total energy difference between consecutive SCF iteration steps. DNC denotes that calculation did not converge within 200–250 SCF iterations.

System	N_{at}	$\kappa(J)$	Anderson	Kerker	TFW	LRDM (r_{avg})
Pt cubic NP, $3 \times 3 \times 3$	172	170	53	182	78	24 (7.3)
Pt cubic NP, $5 \times 5 \times 5$	666	390	73	138	97	28 (10.0)
Pt cubic NP, $6 \times 6 \times 6$	1099	557	90	DNC		30 (11.0)
SiO$_2$H–20 layers	98	1.7	16	58	37	22 (1)
SiO$_2$–20 layers	90	382	87	75	34	25 (6.1)
Li$_{10}$GeP$_2$S$_{12}$ layers	400	133	DNC	69	85	23 (4.9)
Pt+SiO$_2$H–10 layers	93	552	112	68	35	23 (6.5)
Pt+GaAs+SiO$_2$H–10 layers	133	454	92	61	39	23 (6.9)
Au+GaAs+SiO$_2$H–10 layers	133	222	93	30	26	22 (4.9)
Al+GaAs+SiO$_2$H–10 layers	133	199	55	26	23	23 (4.8)
Pt+GaAs+SiO$_2$H–20 layers	258	1662	DNC	126	78	25 (9.5)
Au+GaAs+SiO$_2$H–20 layers	258	888	DNC	46	39	24 (6.6)
Al+GaAs+SiO$_2$H–20 layers	258	708	DNC	40	46	31 (7.8)
Pt+GaAs+SiO$_2$H–40 layers	508	9841	DNC	248	DNC	28 (13.8)
Au+GaAs+SiO$_2$H–40 layers	508	1294	DNC	74	DNC	29 (11.0)
Al+GaAs+SiO$_2$H–40 layers	508	2051	DNC	76	112	30 (9.7)
HfO$_2$ nanofilm with Al stripes	864	535	125	85	155	31 (6.2)

Tensor Core GPUs and a single AMD Milan CPU (64 physical cores). Some of the simulations reported in this work were also executed on the OLCF Summit and XSEDE Stampede2 supercomputers. Summit comprises 4608 IBM Power System AC922 nodes with two IBM POWER9 processors (42 physical cores) and six NVIDIA Volta V100 GPUs in each node.

B. Comparison of SCF convergence

We first study the SCF convergence for the metal-vacuum-type material systems with increasing system sizes. As benchmark systems, we consider nonperiodic fcc Pt nanoparticles (NPs) of increasing sizes, with the cube faces along the {100} crystallographic planes and a vacuum layer of ~15 Å around the nanoparticle. We note that the localized $5d$ orbitals of Pt near the Fermi energy can lead to large eigenvalues of χ_0 that make these calculations challenging. Table I reports the approximate value of $\kappa(J)$ for three different sizes—$3 \times 3 \times 3$, $5 \times 5 \times 5$, and $6 \times 6 \times 6$ containing 172–1099 atoms (3096–19 782 electrons)—and compares the SCF convergence of Anderson mixing, Kerker, TFW, and LRDM preconditioners. We observe that the convergence of Anderson mixing deteriorates with system size, requiring up to 90 SCF iteration steps for the Pt cubic-NP $6 \times 6 \times 6$ system. This corresponds to the approximately fourfold increase in $\kappa(J)$ as would be excepted from the L^2 scaling of the long-wavelength eigenmode from the Coulomb kernel in the metallic region (cf. Sec. II A), where L denotes the extent of the metallic region. We remark that the convergence of Anderson mixing did not improve upon increasing the mixing history from 20 to 50. The Kerker preconditioner performs worse than Anderson for this benchmark system, requiring around 150 SCF iterations or more. This can be attributed to the unsuitability of Thomas-Fermi screening for metal-vacuum systems. Likewise, the TFW preconditioner, which approximates a heterogeneous dielectric function, also requires a large number

of SCF iterations of ~100. We note that although the orbital-free Thomas-Fermi–von Weizsacker model can qualitatively capture the metal-vacuum transition region [39], it is not suited for systems with large eigenvalues of χ_0 due to the localized states near the Fermi energy. Finally, considering the LRDM preconditioner, it achieves accelerated convergence within 24–30 SCF iteration steps for all the system sizes considered, up to three times fewer iteration steps compared to Anderson, the next best performing method for these benchmark systems. We note that even for the largest system size of 1099 atoms (19 782 electrons), r_{avg} in LRDM is a modest value of 11, with further reduction in the rank achieved by using LRDMA as will be demonstrated in Sec. III C. We also refer to Fig. 1(a), where we observe a noticeable plateauing of the energy difference in successive SCF iterations beyond the initial 20–40 SCF iteration steps for Kerker and TFW preconditioners. In contrast, LRDM demonstrates close to an exponential convergence of energy differences.

Next, we consider semiconductor-vacuum benchmark systems. We consider three semiperiodic benchmark systems: (i) 20 (110) plane monolayers of SiO$_2$ with the silica surfaces passivated by hydrogen, (ii) the same silica layer system but without the surface passivation, and (iii) a disordered Li$_{10}$GeP$_2$S$_{12}$ (LGPS superionic conductor) slab. Benchmark system (i) is obtained from a previous work related to the development and testing of a LDOS-based preconditioner [34]. Benchmark system (iii) is constructed based on the atomic coordinates and occupation factors obtained from x-ray structure analysis of LGPS crystal [58], and further a vacuum layer is considered normal to the slab surfaces.[2] First, analyzing the SCF convergence of silica layers with surface passivation, we observe that both the Anderson mixing and

---

[2]This corresponds to a model system for investigations of surface effects in LGPS.

LRDM preconditioner converge, efficiently taking $\sim$20 iterations, whereas Kerker and TFW preconditioners demonstrate a slow convergence despite a low $\kappa(J) \sim 2$ for this system. The slower convergence is attributed to the well-known inability of Kerker and TFW preconditioners to qualitatively approximate the dielectric function of semiconductors, which can, in turn, even deteriorate the net conditioning of the preconditioned quasi-Newton problem [cf. Eq. (4)]. In case of the unpassivated silica layers, the convergence of Anderson and Kerker deteriorates due to high $\kappa(J) \sim 400$ caused by the dangling bonds, whereas the LRDM preconditioner requires a similar number of SCF iterations as the passivated case. Subsequently, analyzing the convergence for the LGPS system, which has a much larger value of $\kappa(J)$ compared to the silica layers, we observe that the Anderson mixing did not converge. Both the Kerker and TFW preconditioners also demonstrate slower convergence for LGPS compared to the silica system, taking 3 to 3.7 times more SCF iteration steps compared to LRDM (23 iterations).

Next, we consider metal-semiconductor-insulator-vacuum heterogeneous benchmark systems. In particular, we consider semiperiodic layered systems of the type (fcc metal)+GaAs+SiO$_2$H–N_l layers, with N_l denoting the number of monolayers in each of the metal and semiconductor regions. The benchmark systems are adapted from previous work [34,59], wherein they are constructed by first arranging N_l (100) plane monolayers of the fcc metallic crystal, followed by N_l (110) plane monolayers of GaAs and SiO$_2$. The metal and silica surfaces are exposed to the vacuum. Further, we note that the silica surfaces are passivated by hydrogen and all the interfaces are made to be coherent by applying appropriate affine deformations to the respective regions. We choose three sets of system sizes, $N_l = 10$, 20, and 40 containing 133–508 atoms (up to $\sim$6100 electrons), and three different metallic elements, Al, Au, and Pt. We note that for $N_l = 40$, we use 20 monolayers for the SiO$_2$ region as the benchmark systems obtained from [34,59] are only available till $N_l = 20$. Going to $N_l = 40$ from $N_l = 20$, we increased the layers of only the metallic and semiconducting layers to maintain the similar hydrogen passivated silica surfaces as $N_l = 20$. Examining the SCF convergence of these benchmark systems from Table I and Figs. 1(b)–1(d) (for Pt), we observe that except in the case of Pt/Al/Au+GaAs+SiO$_2$H–10 layers, Anderson mixing did not converge despite using a small $\alpha = 0.015$. We anticipate this to be a result of the large values of $\kappa(J)$ that increase with number of layers, reaching up to $\sim$10 000 for Pt+GaAs+SiO$_2$H–40 layers, in combination with the strong nonlinear effects in such highly heterogeneous systems, as will be numerically demonstrated in Sec. III C. We also remark that the values of $\kappa(J)$ are higher for Pt+GaAs+ SiO$_2$H–N_l compared to Au+GaAs+SiO$_2$H–N_l due to the partially filled $5d$ subshell of Pt in comparison to the fully filled $5d$ subshell in Au. Analyzing the SCF convergence of Kerker and TFW, we observe both approaches deteriorating with increasing N_l as neither the Thomas-Fermi homogeneous screening nor the Thomas-Fermi–von Weizsacker model can qualitatively capture the transition of the dielectric function between metal, semiconductor, insulator, and vacuum regions. In contrast, the LRDM preconditioner performs remarkably well for these highly heterogeneous

and large-scale benchmark systems, converging within $\sim$20–30 SCF iteration steps for all the nine different benchmark systems considered here. Notably, for the Pt+GaAs+SiO$_2$H–40-layer system, except for the LRDM preconditioner, all the other preconditioners exhibit significantly deteriorated convergence requiring $\sim$250 SCF iterations or more. The relatively higher rank $r_{avg} = 13.8$ required by the LRDM preconditioner for the Pt+GaAs+SiO$_2$H–40-layer system indicates the complex polarization response of the dielectric matrix for such systems.

Finally, we consider a large-scale semiconductor-metal-vacuum semiperiodic benchmark system, HfO$_2$ nanofilm with Al stripes, used to investigate the origins of ferroelectricity in Al-doped HfO$_2$ nanofilms [21]. This benchmark system is constructed as a 3-nm-thick tetragonal-phase HfO$_2$ (001) nanofilm with vacuum on both sides, and 18 Al dopant atoms distributed along two separate (001) layers or stripes in the nanofilm. Overall, this system contains 864 atoms (6750 electrons). The results from the SCF convergence study for this system in Table I show slow convergence for Anderson, TFW, and Kerker. In contrast, LRDM converges within 31 SCF iteration steps. The deteriorated performance for Anderson, TFW, and Kerker, which use the previous SCF history, could be due to nonlinear effects. This is corroborated by our findings in Sec. III C below, where we numerically demonstrate strong nonlinear effects for the HfO$_2$-Al benchmark system.

1. SCF convergence comparison for magnetic systems

We now test the performance of LRDM for magnetic benchmark systems using the spin-polarized extension of LRDM proposed in Sec. II D. We consider two benchmark systems: (i) a ferromagnetic bcc Fe $3 \times 3 \times 3$ periodic supercell with a monovacancy, and (ii) a ferromagnetic Pt$_3$Ni cubic $2 \times 2 \times 2$ nanoparticle with a vacuum layer of $\sim$15 Å around the nanoparticle. We remark that Pt$_3$Ni nanoparticles have important applications as catalysts for the oxygen reduction reaction in hydrogen fuel cells [60]. Table II reports the SCF convergence comparison of spin-polarized LRDM against other preconditioners for these benchmark systems. In the case of the close-to-homogeneous bcc Fe$_{vac}$ $3 \times 3 \times 3$ system, we observe efficient convergence for Kerker, TFW, and LRDM preconditioners requiring $\sim$20 SCF iteration steps, compared to Anderson which requires about two times more iterations. Next, considering the Pt$_3$Ni cubic NP, we observe accelerated convergence only with the LRDM preconditioner requiring 25 SCF iteration steps, while Anderson, Kerker, and TFW require more than four times the iterations. The relatively higher r_{avg} of around 15, compared to the spin-restricted Pt metallic nanoparticle benchmarks discussed previously, indicates the complex nature of the dielectric matrix due to the additional spin-density response of $R[(\rho_\uparrow, \rho_\downarrow)]$. Additionally, we note that the net magnetization at convergence obtained from LRDM is within $0.1\mu_B$ of the net magnetization obtained from other preconditioners.

2. SCF convergence comparison at lower smearing temperatures

We now test the performance of LRDM for a lower Fermi-Dirac smearing temperature $T = 10$ K compared to the 500 K used in the earlier benchmarks. Lower smearing temperature

TABLE II. Comparison of LRDM performance against other preconditioners for spin-polarized DFT benchmark calculations with nonzero net magnetization. A stopping tolerance of 10^{-5} in the L_2 norm of the electron-density residual ($\|R[(\rho_\uparrow, \rho_\downarrow)]\|$) is used. Based on numerical experiments, this is approximately equivalent to a stopping tolerance of 5×10^{-8} Ha in the total energy difference between consecutive SCF iteration steps. DNC denotes that the calculation did not converge within 200–250 SCF iteration steps. M_{tot} denotes the net magnetization.

System	N_{at}	$\kappa(J)$	Anderson	Kerker	TFW	LRDM (r_{avg})	M_{tot} (μ_B)
bcc Fe$_{vac}$, $3 \times 3 \times 3$	53	64	41	18	16	24 (5.0)	125.8
Pt$_3$Ni cubic NP, $2 \times 2 \times 2$	63	128	102	DNC	108	25 (14.8)	43.9

leads to larger total density of states at the Fermi energy for metallic systems, which can increase $\kappa(J)$ as $\chi_0(|\mathbf{q}|)$ is proportional to the total density of states at the Fermi energy as $|\mathbf{q}| \to 0$. This can also be interpreted more physically as sharpening of the occupancies near the Fermi energy leading to pronounced occupancy sloshing across degenerate or close-to-degenerate states near the Fermi energy. We reconsider three benchmark systems with metallic character: (i) a fcc $3 \times 3 \times 3$ Pt NP, (ii) a Pt+GaAs+SiO$_2$H layered metal-semiconductor-insulator-vacuum heterogeneous system, and (iii) a ferromagnetic bcc Fe $3 \times 3 \times 3$ periodic supercell with a monovacancy. Table III reports the lower temperature SCF convergence comparison of LRDM against other preconditioners for these benchmark systems. Comparing the $T = 10$ K results to $T = 500$ K results for benchmark systems (i) and (ii) (cf. Table I), we first observe that $\kappa(J)$ has increased by 1.5- to 2 times. Overall this results in slower convergence for Anderson, Kerker, and TFW preconditioners, whereas the LRDM preconditioner converges in the same number of iterations as with $T = 500$ K, although with a slightly higher average rank due to increase in $\kappa(J)$. Subsequently, considering the ferromagnetic bcc Fe monovacancy system, we observe that both LRDM and Kerker preconditioners demonstrate robust convergence similar to the $T = 500$ K benchmark results in Table II, whereas the Anderson and TFW convergence is significantly deteriorated. We note that to obtain robust convergence for LRDM in the bcc Fe system, the damping parameter α was reduced from 0.5 to 0.4.

C. Results on r_{avg} reduction using LRDMA

In the SCF convergence studies discussed in Sec. III B for the LRDM preconditioner, we find that r_{avg} can be high for systems with large $\kappa(J)$. In Sec. II B, we proposed an improvement to LRDM that can potentially reduce r_{avg} by accumulating the low-rank approximation of J_r^{lr} from previous SCF steps combined with an adaptive algorithm that controls the accumulation. We now demonstrate and analyze the r_{avg} reduction achieved by the accumulated variant (LRDMA)

over LRDM for a subset of the difficult heterogeneous benchmark systems discussed in Sec. III B, namely, the Pt cubic nanoparticles, Pt+GaAs+SiO$_2$H–N_l-layer benchmark systems, and the HfO$_2$ nanofilm with Al stripes benchmark system. Table IV reports the total SCF iterations and r_{avg} required by LRDM and LRDMA approaches. We further analyze the rank accumulation in Fig. 2, which shows the adaptively determined rank in each SCF iteration step for the LRDM and LRDMA approaches, and also marks the accumulation clearing events with labels CL, CR, and CT related to nonsatisfaction of linearity-indicator criteria, residual norm criteria, and low-rank approximation error criteria, respectively (cf. Sec. II B).

In the case of Pt cubic NPs, we observe an appreciable $\sim$1.8-fold reduction in r_{avg} for LRDMA compared to LRDM, while the number of SCF iterations are the same except for the largest system, Pt cubic-NP $6 \times 6 \times 6$ taking an additional SCF step in LRDMA. Next, in the case of the more heterogeneous Pt+GaAs+SiO$_2$H–N_l-layer system and HfO$_2$-Al system, we observe around a 1.6-fold reduction in r_{avg} with a slight increase in the number of SCF iterations by $\sim$10% for the most difficult Pt+GaAs+SiO$_2$H–40 system. Although LRDMA performs robustly for the highly heterogeneous systems, the relatively lower advantage as compared to the Pt cubic NPs is attributed to the stronger nonlinear effects in the more heterogeneous layered material systems when the SCF iteration is not close to convergence. The above reasoning is supported by Fig. 2, which shows that in the case of Pt cubic NPs only a small number of CL-labeled clearing events occur across all the LRDMA preconditioned SCF iterations. On the other hand, Pt+GaAs+SiO$_2$H–N_l and HfO$_2$-Al systems have a much larger number of CL clearing events when the SCF iteration is not close to convergence. Specifically, we observe successive CL-labeled events in these systems, that correspond to nonlinear regimes in $R[\rho]$ when considering the Taylor series expansion about ρ at the current iterate. Relying on CT events only in such regimes leads to suboptimal accumulation, which we have observed from our numerical experiments. Overall, the above studies demonstrate that

TABLE III. Comparison of the performance of LRDM against other preconditioners at low Fermi-Dirac smearing temperature of $T = 10$ K. DNC denotes the calculation did not converge within 200–250 SCF iteration steps.

System	N_{at}	$\kappa(J)$	Anderson	Kerker	TFW	LRDM (r_{avg})
Pt cubic NP, $3 \times 3 \times 3$	172	244	64	DNC	149	24 (8.9)
Pt+GaAs+SiO$_2$H–10 layers	133	875	127	72	54	23 (6.9)
bcc Fe$_{vac}$, $3 \times 3 \times 3$	53	69	92	20	DNC	33 (5.7)

TABLE IV. Comparison of performance of LRDM and LRDMA preconditioners.

System	LRDM (r_{avg})	LRDMA (r_{avg})
Pt cubic NP, $3 \times 3 \times 3$	24 (7.3)	23 (3.7)
Pt cubic NP, $5 \times 5 \times 5$	28 (10.0)	28 (5.5)
Pt cubic NP, $6 \times 6 \times 6$	30 (11.0)	31 (6.8)
Pt+GaAs+SiO2H–10 layers	23 (6.9)	25 (4.1)
Pt+GaAs+SiO2H–20 layers	25 (9.5)	26 (6.0)
Pt+GaAs+SiO2H–40 layers	28 (13.8)	32 (8.8)
HfO$_2$ nanofilm with Al stripes	31 (6.2)	31 (4.2)

the LRDMA approach, in general, is an improvement over LRDM, with the r_{avg} reduction dependent on the strength of the nonlinearities of $F[V_{eff}[\rho]]$ in the material system.

D. Computational cost of LRDM and its system size scaling

We now demonstrate the computational efficiency of the LRDM preconditioner and the accumulated variant (LRDMA) against Anderson and Kerker preconditioners on a subset of the benchmark systems considered in Sec. III B, with system sizes ranging from ~3000 to 20 000 electrons. The benchmark calculations are performed using implementations of Anderson, Kerker, LRDM, and LRDMA preconditioners in the DFT-FE code. All simulations are performed using multiple GPU nodes of the NERSC Perlmutter supercomputer, with comparisons conducted between simulations using the same number of nodes. Further, we note that the key computational steps in the LRDM preconditioner, in particular the first-order density response computation in the Chebyshev filtered subspace (cf. Sec. II C), have been efficiently ported to GPUs. On the other hand, the electron-density mixing operations required by Anderson and Anderson with Kerker preconditioners are performed on CPUs, as their overheads are less than 5% compared to the ChFSI eigensolver cost in each SCF iteration in DFT-FE. We also remark that the ChFSI procedure in DFT-FE was recently GPU ported [8], and achieves significant CPU-GPU speedups of ~15- to 20-fold on OLCF Summit. We first consider the Pt cubic nanoparticle benchmark systems with system sizes ranging from 172 to 1099 atoms (3096–19 782 electrons). Table V reports the computational cost in GPU Node-hours for Anderson, LRDM, and LRDMA. We present both the computational cost of the full SCF convergence as well as the average per SCF iteration cost. We observe that the total SCF computational cost speedups of LRDM and LRDMA with respect to Anderson are ~1.4-fold

and ~1.6-fold, respectively, for the larger system sizes. The relatively higher speedup for LRDMA is due to the previously demonstrated lower r_{avg} for LRDMA compared to LRDM for these benchmark systems. Furthermore, the per SCF computational cost ratio of LRDM to Anderson is quite modest, with an average value across the three system sizes of 2.0- and 1.6-fold for LRDM and LRDMA, respectively, where the computational benefit of LRDMA over LRDM is further evident. The above computational cost ratio also demonstrates a weak system size dependence, which is related to the $\mathcal{O}(MN_e^2)$ scaling of LRDM. From Table V, it can be numerically verified that the additional per SCF computational cost for LRDM and LRDMA compared to Anderson scales close to cubically with respect to the number of electrons when the system size is increased from Pt cubic NP $5 \times 5 \times 5$ (666 atoms) to Pt cubic NP $6 \times 6 \times 6$ (1099 atoms). Thus, for larger system sizes, given the asymptotic cubic scaling of KS-DFT and the weak dependence of r_{avg} on the system size (cf. Table I), the computational cost ratio can be expected to approach a constant factor for very large system sizes.

Next, we investigate the computational cost of LRDM and LRDMA for the more heterogeneous benchmark systems— Pt+GaAs+SiO$_2$H (metal-semiconductor-insulator-vacuum) layers with increasing sizes and a large-scale HfO$_2$ nanofilm with Al stripes (metal-semiconductor-vacuum) system. For these material systems we compare against the Kerker preconditioner, which is relatively more robust for such systems compared to Anderson. Table VI reports the results for this study, where we observe LRDM to achieve 1.6- to 3.4-fold speedups in the total SCF cost compared to Kerker for the above benchmark systems. The per SCF computational cost ratio of LRDM to Kerker is 1.5- to 2.5-fold, with up to a ~1.2-fold reduction observed for the per SCF computational cost ratio while using LRDMA. We remark that LRDMA does not provide an improvement in the total SCF cost reduction relative to LRDM for the Pt+GaAs+SiO$_2$H–40-layer system due to the slight increase in the number of SCFs (cf. Sec. III C) that negates the reduction in the per SCF cost.

IV. CONCLUSIONS

We present a robust and efficient approach to accelerate the self-consistent field iteration in Kohn-Sham DFT calculations. In particular, we use a low-rank approximation of the dielectric matrix (LRDM)—constructed as a sum of rank-1 tensor products—as the preconditioner for solving the self-consistent fixed-point iteration. We note that the low-rank approximation is adaptive and systematically convergent, which lends

TABLE V. Comparison of computational efficiency between Anderson, LRDM, and LRDMA approaches as implemented in the DFT-FE code for benchmark systems comprising Pt cubic nanoparticles of various sizes. Computational cost is reported in terms of NERSC Perlmutter GPU Node-hours for the total SCF solve and the average cost per SCF iteration. The first SCF iteration, which involves multiple passes of Chebyshev filtering, is excluded in measuring the average per SCF iteration cost. The total SCF cost includes all SCF iteration steps.

System	Anderson total SCF	LRDM total SCF	LRDMA total SCF	Anderson per SCF	LRDM per SCF	LRDMA per SCF	LRDM-Anderson ratio per SCF	LRDMA-Anderson ratio per SCF
Pt cubic NP, $3 \times 3 \times 3$	0.78	0.71	0.56	0.013	0.026	0.021	2.0	1.62
Pt cubic NP, $5 \times 5 \times 5$	13.14	9.18	8.04	0.157	0.284	0.239	1.81	1.52
Pt cubic NP, $6 \times 6 \times 6$	56.94	40.85	36.0	0.564	1.174	1.0	2.08	1.77

TABLE VI. Comparison of computational efficiency between Kerker, LRDM, and LRDMA approaches as implemented in the DFT-FE code for heterogeneous metal-insulator-semiconductor benchmark systems of various sizes. Computational cost is reported in terms of NERSC Perlmutter GPU Node-hours for the total SCF solve and the average cost per SCF iteration. The first SCF iteration, which involves multiple passes of Chebyshev filtering, is excluded in measuring the average per SCF iteration cost. The total SCF cost includes all SCF iteration steps.

System	Kerker total SCF	LRDM total SCF	LRDMA total SCF	Kerker per SCF	LRDM per SCF	LRDMA per SCF	LRDM-Kerker ratio per SCF	LRDMA-Kerker ratio per SCF
Pt+GaAs+SiO2H–10 layers	3.06	1.86	1.74	0.047	0.072	0.061	1.53	1.30
Pt+GaAs+SiO2H–20 layers	22.31	8.66	7.43	0.166	0.290	0.263	1.75	1.58
Pt+GaAs+SiO2H–40 layers	213.30	63.5	64.58	0.837	2.112	1.937	2.52	2.31
HfO$_2$ nanofilm with Al stripes	6.30	4.04	3.50	0.069	0.122	0.104	1.77	1.51

to the robustness of the approach for tackling generic heterogeneous material systems, including systems with large condition number and strong nonlinearities in the fixed-point iteration. A key contribution of the present effort is the development of an efficient numerical method for computing the first-order density response in real space that is critical to the efficiency of the LRDM procedure. To this end, we compute the density response in the Chebyshev filtered subspace iteration (ChFSI) procedure, which adaptively approximates the occupied eigensubspace in each SCF iteration step. The smaller dimension of the Chebyshev filtered subspace, in conjunction with the use of canonical density-matrix perturbation theory, provides a computationally efficient approach for using the LRDM preconditioner. In order to further improve the computational efficiency of the LRDM method, we developed an accumulated variant of the formulation, which adaptively accumulates the low-rank dielectric matrix approximation from the previous SCF iterations using indicators based on the extent of linearity of the fixed-point iteration map and the residual norm. Additionally, we extended the LRDM preconditioner to collinear spin-polarized KS-DFT, using a generalization of the low-rank formulation to spin densities and their corresponding response functions. We note that this generalization is also extensible to noncollinear spin-polarized KS-DFT, which is a topic for future studies. We additionally remark that the proposed numerical method for the LRDM procedure, although developed in the context of a finite-element basis and ChFSI eigensolver, is adaptable to other Kohn-Sham DFT implementations that use different discretizations (e.g., plane waves, finite difference, and wavelets) and eigensolver strategies (e.g., Davidson and RMM-DIIS).

We investigated the robustness and efficiency of the LRDM approach on a comprehensive set of heterogeneous nonperiodic and periodic benchmark systems. To this end, we chose material systems such as metallic and bimetallic nanoparticles, layered materials with various combinations of metal-semiconductor-insulator systems, semiconducting nanofilms with metal dopants, and magnetic systems with spin polarization. The benchmark systems range from small

to large system sizes comprising ∼100–1100 atoms (∼500–20 000 electrons). In all the benchmark systems we compared the SCF convergence of LRDM against three other widely used preconditioners—Anderson mixing [26], Anderson mixing with Kerker preconditioner [35], and Broyden mixing with Thomas-Fermi–von Weizsacker preconditioner [39]. Our studies have demonstrated that LRDM achieves robust convergence within 20–30 SCF iterations, for all the benchmark systems considered here, with a weak dependence observed on the system size. In comparison, Anderson, Kerker, and Broyden showed slow convergence for systems with high condition numbers, requiring three to five times more SCF iterations compared to LRDM. The accumulated variant (LRDMA) showed promise in reducing the average (over the SCF iterations) adaptive rank r_{avg} by 1.5- to 1.8-fold across various benchmark systems without affecting the robustness of the SCF iteration. We also found the LRDM or LRDMA approach to outperform Anderson and Kerker preconditioners in computational cost for full ground-state calculations by ∼1.4- to 3.4-fold for Pt cubic nanoparticles and layered heterogeneous benchmark systems.

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

[2] W. Kohn, Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms, Phys. Rev. Lett. **76**, 3168 (1996).

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

[4] J.-I. Iwata, D. Takahashi, A. Oshiyama, T. Boku, K. Shiraishi, S. Okada, and K. Yabana, A massively-parallel electronic-structure calculations based on real-space density functional theory, J. Comput. Phys. **229**, 2339 (2010).

[5] L. Genovese, B. Videau, M. Ospici, T. Deutsch, S. Goedecker, and J.-F. Méhaut, Daubechies wavelets for high performance electronic structure calculations: The BigDFT project, C. R. Mec. **339**, 149 (2011).

[6] P. Motamarri, S. Das, S. Rudraraju, K. Ghosh, D. Davydov, and V. Gavini, DFT-FE—A massively parallel adaptive finite-element code for large-scale density functional theory calculations, Comput. Phys. Commun. **246**, 106853 (2020).

[7] Q. Xu, A. Sharma, B. Comer, H. Huang, E. Chow, A. J. Medford, J. E. Pask, and P. Suryanarayana, SPARC: Simulation package for ab-initio real-space calculations, SoftwareX **15**, 100709 (2021).

[8] S. Das, P. Motamarri, V. Subramanian, D. M. Rogers, and V. Gavini, DFT-FE 1.0: A massively parallel hybrid CPU-GPU density functional theory code using finite-element discretization, Comput. Phys. Commun. **280**, 108473 (2022).

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

[10] L. Lin, M. Chen, C. Yang, and L. He, Accelerating atomic orbital-based electronic structure calculation via pole expansion and selected inversion, J. Phys.: Condens. Matter **25**, 295501 (2013).

[11] P. Motamarri and V. Gavini, Subquadratic-scaling subspace projection method for large-scale Kohn-Sham density functional theory calculations using spectral finite-element discretization, Phys. Rev. B **90**, 115127 (2014).

[12] C.-C. Lin, P. Motamarri, and V. Gavini, Tensor-structured algorithm for reduced-order scaling large-scale Kohn-Sham density functional theory calculations, npj Comput. Mater. **7**, 50 (2021).

[13] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard, Density functional theory in surface chemistry and catalysis, Proc. Natl. Acad. Sci. USA **108**, 937 (2011).

[14] A. Fernando, K. D. M. Weerawardene, N. V. Karimova, and C. M. Aikens, Quantum mechanical studies of large metal, metal oxide, and metal chalcogenide nanoparticles and clusters, Chem. Rev. **115**, 6112 (2015).

[15] D. J. Cole and N. D. Hine, Applications of large-scale density functional theory in biology, J. Phys.: Condens. Matter **28**, 393001 (2016).

[16] T. A. Pham, Y. Ping, and G. Galli, Modelling heterogeneous interfaces for solar water splitting, Nat. Mater. **16**, 401 (2017).

[17] B. Das and S. Mahapatra, An atom-to-circuit modeling approach to all-2D metal–insulator–semiconductor field-effect transistors, npj 2D Mater. Appl. **2**, 28 (2018).

[18] S. Das, P. Motamarri, V. Gavini, B. Turcksin, Y. W. Li, and B. Leback, Fast, scalable and accurate finite-element based ab initio calculations using mixed precision computing: 46 PFLOPS simulation of a metallic dislocation system, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (Association for Computing Machinery, New York, 2019), pp. 1–11.

[19] R. Zhuravel, H. Huang, G. Polycarpou, S. Polydorides, P. Motamarri, L. Katrivas, D. Rotem, J. Sperling, L. A. Zotti, A. B. Kotlyar, J. C. Cuevas, V. Gavini, S. S. Skourtis, and D. Porath, Backbone charge transport in double-stranded DNA, Nat. Nanotechnol. **15**, 836 (2020).

[20] K. Ghosh, H. Ma, M. Onizhuk, V. Gavini, and G. Galli, Spin-spin interactions in defects in solids from mixed all-electron and pseudopotential first-principles calculations, npj Comput. Mater. **7**, 123 (2021).

[21] L. Yao, S. Das, X. Liu, K. Wu, Y. Cheng, V. Gavini, and B. Xiao, Modulating the microscopic lattice distortions through the Al-rich layers for boosting the ferroelectricity in Al: HfO_2 nanofilms, J. Phys. D **55**, 455501 (2022).

[22] K.-M. Ho, J. Ihm, and J. D. Joannopoulos, Dielectric matrix scheme for fast convergence in self-consistent electronic-structure calculations, Phys. Rev. B **25**, 4260 (1982).

[23] P.-M. Anglade and X. Gonze, Preconditioning of self-consistent-field cycles in density-functional theory: The extrapolar method, Phys. Rev. B **78**, 045126 (2008).

[24] S. L. Adler, Quantum theory of the dielectric constant in real solids, Phys. Rev. **126**, 413 (1962).

[25] N. Wiser, Dielectric constant with local field effects included, Phys. Rev. **129**, 62 (1963).

[26] D. G. Anderson, Iterative procedures for nonlinear integral equations, J. Assoc. Comput. Mach. **12**, 547 (1965).

[27] P. Pulay, Convergence acceleration of iterative sequences. The case of SCF iteration, Chem. Phys. Lett. **73**, 393 (1980).

[28] D. D. Johnson, Modified Broyden's method for accelerating convergence in self-consistent calculations, Phys. Rev. B **38**, 12807 (1988).

[29] K. N. Kudin, G. E. Scuseria, and E. Cancès, A black-box self-consistent field convergence algorithm: One step closer, J. Chem. Phys. **116**, 8255 (2002).

[30] H.-r. Fang and Y. Saad, Two classes of multisecant methods for nonlinear acceleration, Numer. Linear Algebra Appl. **16**, 197 (2009).

[31] P. H. Dederichs and R. Zeller, Self-consistency iterations in electronic-structure calculations, Phys. Rev. B **28**, 5462 (1983).

[32] Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, Philadelphia, 2003).

[33] L. Lin and C. Yang, Elliptic preconditioner for accelerating the self-consistent field iteration in Kohn-Sham density functional theory, SIAM J. Sci. Comput. **35**, S277 (2013).

[34] M. F. Herbst and A. Levitt, Black-box inhomogeneous preconditioning for self-consistent field iterations in density functional theory, J. Phys.: Condens. Matter **33**, 085503 (2021).

[35] G. P. Kerker, Efficient iteration scheme for self-consistent pseudopotential calculations, Phys. Rev. B **23**, 3082 (1981).

[36] R. Resta, Thomas-Fermi dielectric screening in semiconductors, Phys. Rev. B **16**, 2717 (1977).

[37] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[38] N. D. Woods, M. Payne, and P. Hasnip, Computing the self-consistent field in Kohn-Sham density functional theory, J. Phys.: Condens. Matter **31**, 453001 (2019).

[39] D. Raczkowski, A. Canning, and L. W. Wang, Thomas-Fermi charge mixing for obtaining self-consistency in density functional calculations, Phys. Rev. B **64**, 121101(R) (2001).

[40] N. Marzari, D. Vanderbilt, and M. C. Payne, Ensemble Density-Functional Theory for *Ab Initio* Molecular Dynamics of Metals and Finite-Temperature Insulators, Phys. Rev. Lett. **79**, 1337 (1997).

[41] A. M. N. Niklasson, Extended Lagrangian Born-Oppenheimer molecular dynamics using a Krylov subspace approximation, J. Chem. Phys. **152**, 104103 (2020).

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

[43] X. Andrade, D. Strubbe, U. De Giovannini, A. H. Larsen, M. J. T. Oliveira, J. Alberdi-Rodriguez, A. Varas, I. Theophilou, N. Helbig, M. J. Verstraete, L. Stella, F. Nogueira, A. Aspuru-Guzik, A. Castro, M. A. L. Marques, and A. Rubio, Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems, Phys. Chem. Chem. Phys. **17**, 31371 (2015).

[44] V. Michaud-Rioux, L. Zhang, and H. Guo, Rescu: A real space electronic structure method, J. Comput. Phys. **307**, 593 (2016).

[45] F. L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, J. Math. Phys. **6**, 164 (1927).

[46] A. M. N. Niklasson, M. J. Cawkwell, E. H. Rubensson, and E. Rudberg, Canonical density matrix perturbation theory, Phys. Rev. E **92**, 063301 (2015).

[47] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, UK, 2004).

[48] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, Rev. Mod. Phys. **73**, 515 (2001).

[49] M. Methfessel and A. T. Paxton, High-precision sampling for Brillouin-zone integration in metals, Phys. Rev. B **40**, 3616 (1989).

[50] N. Marzari, D. Vanderbilt, A. De Vita, and M. C. Payne, Thermal Contraction and Disordering of the Al(110) Surface, Phys. Rev. Lett. **82**, 3296 (1999).

[51] W. Liang, C. Saravanan, Y. Shao, R. Baer, A. T. Bell, and M. Head-Gordon, Improved Fermi operator expansion methods for fast electronic structure calculations, J. Chem. Phys. **119**, 4117 (2003).

[52] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[53] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, Phys. Rev. B **88**, 085117 (2013).

[54] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table, Comput. Phys. Commun. **226**, 39 (2018).

[55] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos *et al.*, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, J. Phys.: Condens. Matter **21**, 395502 (2009).

[56] 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 Gironcoli, P. Delugas, R. A. DiStasio, Jr., A. Ferretti, A. Floris, G. Fratesi, G. Fugallo *et al.*, Advanced capabilities for materials modelling with QUANTUM ESPRESSO, J. Phys.: Condens. Matter **29**, 465901 (2017).

[57] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.107.125133 for numerical comparison of multi-secant preconditioner performance as implemented in the DFT-FE code and Quantum Espresso code.

[58] A. Kuhn, J. Köhler, and B. V. Lotsch, Single-crystal x-ray structure analysis of the superionic conductor $Li_{10}GeP_2S_{12}$, Phys. Chem. Chem. Phys. **15**, 11620 (2013).

[59] M. F. Herbst, Black-box inhomogeneous preconditioning for self-consistent field iterations in density functional theory, https://github.com/mfherbst/supporting-ldos-preconditioning (2020).

[60] X. Shen, C. Zhang, S. Zhang, S. Dai, G. Zhang, M. Ge, Y. Pan, S. M. Sharkey, G. W. Graham, A. Hunt *et al.*, Deconvolution of octahedral Pt_3Ni nanoparticle growth pathway from in situ characterizations, Nat. Commun. **9**, 4485 (2018).