Subquadratic-scaling subspace projection method for large-scale Kohn-Sham density functional theory calculations using spectral finite-element discretization

Phani Motamarri and Vikram Gavini

A subquadratic-scaling subspace projection method for large-scale Kohn-Sham density functional theory calculations using spectral finite-element discretization

Phani Motamarri and Vikram Gavini

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

We present a subspace projection technique to conduct large-scale Kohn-Sham density functional theory calculations using higher-order spectral finite-element discretization. The proposed method treats both metallic and insulating materials in a single framework, and is applicable to both pseudopotential as well as all-electron calculations. The key ideas involved in the development of this method include: (i) employing a higher-order spectral finite-element basis that is amenable to mesh adaption; (ii) using a Chebyshev filter to construct a subspace which is an approximation to the occupied eigenspace in a given self-consistent field iteration; (iii) using a localization procedure to construct a non-orthogonal localized basis spanning the Chebyshev filtered subspace; (iv) using a Fermi-operator expansion in terms of the subspace-projected Hamiltonian represented in the non-orthogonal localized basis to compute relevant quantities like the density matrix, electron density and band energy. We demonstrate the accuracy and efficiency of the proposed approach on benchmark systems involving pseudopotential calculations on aluminum nano-clusters up to 3430 atoms and on alkane chains up to 7052 atoms, as well as all-electron calculations on silicon nano-clusters up to 3920 electrons. The benchmark studies revealed that accuracies commensurate with chemical accuracy can be obtained with the proposed method, and a subquadratic-scaling with system size was observed for the range of materials systems studied. In particular, for the alkane chains—representing an insulating material—close to linear-scaling is observed, whereas, for aluminum nano-clusters—representing a metallic material—the scaling is observed to be $\mathcal{O}(N^{1.46})$. For all-electron calculations on silicon nano-clusters, the scaling with the number of electrons is computed to be $\mathcal{O}(N^{1.75})$. In all the benchmark systems, significant computational savings have been realized with the proposed approach, with $\sim 10-$fold speedups observed for the largest systems with respect to reference calculations.

I. INTRODUCTION

Past few decades have seen an increasingly important role played by electronic structure calculations based on density functional theory in the investigation of materials properties. The Kohn-Sham approach to density functional theory (DFT)[1,2], in particular, provides a computationally tractable approach to conduct quantum-mechanically informed calculations on ground-state materials properties. The Kohn-Sham approach reduces the many-body problem of interacting electrons into an equivalent problem of non-interacting electrons in an effective mean field that is governed by the electron density. While this formulation has no approximations and is exact in principle for ground-state properties, the quantum mechanical interactions between electrons manifest in the form of an unknown exchange-correlation term in DFT. Various models have been proposed for the exchange-correlation interactions[3], and these have been shown to predict a wide range of materials properties across various materials systems. Though the Kohn-Sham approach greatly reduces the computational complexity of the many-body Schrödinger problem, large-scale electronic structure calculations with DFT are still computationally very demanding, and accurate numerical methods that improve the computational efficiency of DFT calculations are desirable.

The widely used numerical implementations of DFT employ a plane-waves basis[4–6], which provides an efficient computation of the electrostatic interactions arising in DFT naturally through Fourier transforms. Further, the plane-wave basis provides variational convergence in the ground-state energy with exponential convergence rates. However, the plane-wave basis also suffers from some notable disadvantages. In particular, the simulations are restricted to periodic boundary conditions that are not well suited for material systems containing extended defects (for e.g. dislocations), as well as isolated materials systems such as molecules and nano-clusters. Further, the plane-wave basis does not offer adaptive spatial resolution which can be inefficient in the treatment of certain types of calculations—for e. g. all-electron calculations, isolated systems—where higher basis resolution is required in some spatial regions and a coarser resolution suffices elsewhere. Moreover, the plane-wave basis functions are extended in real space, which significantly affects the scalability of computations on parallel computing platforms. On the other hand, atomic-orbital-type basis functions[7–12], which constitute other widely employed basis sets in electronic structure calculations, are well suited for isolated systems, but cannot easily handle other boundary conditions. Further, these basis sets do not offer systematic convergence for all materials systems, and due to the non-locality of the basis functions the parallel scalability is significantly affected. Thus, there has been an increasing focus on systematically improvable and scalable real-space techniques for electronic structure calculations over the past decade[13–21].

Among real-space techniques, the finite-element basis[22]—a piecewise polynomial basis—presents some key

advantages for electronic structure calculations. In particular, the finite-element basis naturally allows for arbitrary boundary conditions, which enables the consideration of isolated, semi-periodic, as well as, periodic materials systems under a single framework. Further, the locality of the basis provides good scalability on parallel computing platforms. Further, the finite-element basis is amenable to adaptive spatial resolution which can effectively be exploited for efficient solution of all-electron DFT calculations[23–25] as well as the development of coarse-graining techniques that seamlessly bridge electronic structure calculations with continuum[26,27]. There have been significant efforts in the recent past to develop real-space electronic structure calculations based on a finite-element discretization[14,23–25,28–43].

Although the finite-element basis offers some unique advantages for electronic structure calculations, initial studies[23,28,38], which employed linear finite-element basis functions, suggested that they require a large number of linear basis functions—of the order of 100,000 basis functions per atom—to achieve chemical accuracy in electronic structure calculations. This compares poorly with plane-wave basis or atomic-orbital-type basis. A recent investigation[25] has indicated that the use of adaptive higher-order spectral finite-elements can significantly improve the computational efficiency of real-space electronic structure calculations. In particular, staggering computational savings—of the order of $1000-$fold—relative to linear finite-elements for both all-electron and local pseudopotential calculations have been obtained by using higher-order finite-element discretizations. Further, for accuracies commensurate with chemical accuracy, it was demonstrated that the computational efficiency afforded by higher-order finite-element discretizations is competing with plane-wave discretization for non-periodic calculations, and is comparable to the Gaussian basis for all-electron calculations to within an order of magnitude. Moreover, the parallel scalability of the finite-element basis was demonstrated on up to 200 processors, where over 90% parallel efficiency was observed. Using modest computational resources, local pseudopotential calculations had been demonstrated on up to 1688 atoms and all-electron calculations had been demonstrated on up to 600 electrons.

However, we note that the traditional self-consistent approach to solving the discretized nonlinear Kohn-Sham eigenvalue problem involves the diagonalization of the Hamiltonian to obtain orthonormal eigenvectors, the computational complexity of which typically scales as $\mathcal{O}(M\,N^2)$ where M denotes the number of basis functions and N denotes the system size (number of atoms or number of electrons). This cost becomes prohibitively expensive, approaching cubic-scaling, as the system size becomes larger. To this end, numerous efforts have focused on either reducing the prefactor[44,45] associated with the computational cost of the Kohn-Sham DFT calculations or reducing the computational complexity to have improved-scaling behavior for DFT calculations.

The latter methods usually exploit locality in the wavefunctions[46] directly or indirectly, and can be broadly categorized[47] into two types: one which calculate the single-electron density matrix, and another which work with its representation in terms of localized Wannier functions. The divide and conquer method[48–50], Fermi-operator expansion[51–54], density-matrix minimization[55,56] approach belong to the former category, whereas, the Fermi-operator projection method[57,58] and the orbital minimization approach[59–62] belong to the latter category. A comprehensive review of these methods has been provided by Göedecker[47], and more recently by Bowler and Miyazaki[63]. These methods, which rely on the locality of the Wannier functions or the exponential decay of the density matrix in real-space, have been demonstrated to work well for insulating systems, exhibiting linear-scaling with system size. However, for metallic systems, due to the slower decay of the density matrix, the computational complexity of these approaches can deviate significantly, in practice, from linear-scaling. Further, we note that, some of the developed techniques[55,57,58,62] assume the existence of a band-gap, thus restricting these techniques solely to insulating systems. The Fermi-operator expansion method[47,51,54], which is equally applicable to both insulating and metallic systems, computes the finite-temperature density-matrix through a Chebyshev polynomial approximation of the Fermi distribution function (also referred to as Fermi function) of the Kohn-Sham Hamiltonian. The accuracy of such an expansion depends on the smearing parameter ($\sigma = k_B T$) in the Fermi distribution and the width of the eigenspectrum (ΔE) of the discretized Hamiltonian. In fact, the number of polynomial terms required to achieve a prescribed accuracy[54] is $\mathcal{O}(\frac{\Delta E}{\sigma})$. Numerous recent efforts[65–69] have focused towards developing alternate approximations to the Fermi function, or approximations to its spectral representation[70]. A majority of these methods aim to reduce the number of terms used in the expansion to approximate the Fermi function. However, a major drawback with these methods is that they are not efficient for local real-space basis functions like finite-elements, where, typically, more refined discretizations are needed. In a recent study[25], it was observed that the width of the eigenspectrum of the finite-element discretized Hamiltonian (using higher-order finite-elements) is $\mathcal{O}(10^3)\ Ha$ for pseudopotential calculations, and $\mathcal{O}(10^5)\ Ha$ for all-electron calculations. In this article, building on our previous work[25], we propose a reduced-scaling subspace projection technique in the framework of spectral finite-element discretization. To this end, we borrow localization ideas from Garcia et al.[62] and develop a subspace iteration technique that treats both metallic and insulating systems on a similar footing. Further, besides pseudopotential calculations, the proposed technique is also applicable to all-electron calculations, as demonstrated in our benchmark studies. The main ideas used in our approach are: (i) employ Chebyshev filtered subspace iteration to compute the occupied eigenspace; (ii) em-

ploy a localization procedure to generate non-orthogonal localized wavefunctions spanning the Chebyshev filtered subspace; (iii) use adaptive tolerances to truncate the wavefunctions, with looser tolerances being employed in initial self-consisted field (SCF) iterations and progressively tightening as the SCF iteration approaches convergence, and (iv) employ Fermi-operator expansion in terms of the projected Hamiltonian expressed in the non-orthogonal localized basis to compute the density matrix, electron density and band energy.

We first present an abstract mathematical framework in which the projection of the Hamiltonian into a sub-space, corresponding to the occupied eiegenspace, and the associated density matrix are expressed in a non-orthogonal basis spanning the subspace. We then derive expressions for the computation of electron density, con-straint on the number of electrons and the band energy in terms of the projected Hamiltonian, which are subse-quently used to formulate the subspace projection tech-nique within the framework of finite-element discretiza-tion. To this end, the Kohn-Sham Hamiltonian and the corresponding wave-functions are represented in the Löwdin orthonormalized finite-element basis constructed using spectral finite-elements in conjunction with Gauss-Lobatto-Legendre quadrature rules. The SCF iteration begins with an initial subspace spanned by the localized single-atom wavefunctions, and a Chebyshev filter is ap-plied on this subspace to compute an approximation to the occupied eigenspace. Next, we employ a localization procedure to construct non-orthogonal localized wave-functions spanning the Chebyshev filtered subspace. The localized wavefunctions are then truncated using a trun-cation tolerance, below which the localized wavefunctions are set to zero. We note that the proposed approach of providing a compact support for the wavefunctions by using a truncation tolerance on the wavefunctions differs from commonly employed approach of using truncation radius, and presents a more efficient approach as the lo-calized wavefunctions may not necessarily be spherically symmetric about the localization center. The Kohn-Sham Hamiltonian is then projected into the localized basis, and a Fermi-operator expansion in terms of the projected Hamiltonian is employed to compute the finite-temperature density matrix and the electron density. If the truncated wavefunctions are sufficiently sparse, the computational cost of each of the above steps is shown to scale linearly with number of atoms.

The proposed approach is implemented in a par-allel computing framework, and the performance of the algorithm is investigated on representative bench-mark atomic systems involving metallic aluminum nano-clusters (pseudopotential calculations), insulating alkane chains (pseudopotential calculations), and semi-conducting silicon nano-clusters (all-electron calcula-tions). The scaling behavior is assessed on these materi-als systems with varying system sizes up to 3430 atoms in the case of aluminum nano-clusters, up to 7052 atoms in the case of alkane chains, and up to 3920 electrons in

the case of all-electron silicon nano-clusters. The scal-ing of the computational time per SCF iteration with the system size is computed to be $\mathcal{O}(N^{1.46})$ for alu-minum nano-clusters, $\mathcal{O}(N^{1.18})$ for the alkane chains, and $\mathcal{O}(N^{1.75})$ for the all-electron silicon nano-clusters. One factor contributing to the deviation from linearity is the use of adaptive tolerances—using looser tolerances in the initial SCF iterations and progressively tighten-ing the tolerances as the SCF approaches convergence—which partially sacrifices the scaling for good accuracy in the ground-state energies. We note that the computed ground-state energies using the proposed approach are within 5 meV per atom for pseudopotential calculations and are within 5 mHa per atom for all-electron calcu-lations with respect to the reference ground-state ener-gies. Further, in our implementation, we switch from us-ing sparse data-structures to dense data-structures when the density fraction of the localized wavefunctions ex-ceeds 2%, as the computational cost of using parallel sparse data-structures is observed to exceed that of dense data-structures beyond this point. This is another fac-tor contributing to the deviation from linear-scaling as discussed in the Appendix. Our results suggest that sig-nificant computational savings can be realized using the proposed approach, where $\sim 10-$fold speedups are ob-tained with respect to reference benchmark calculations for the largest systems.

The remainder of the paper is organized as follows. Section II describes the real-space formulation of the Kohn-Sham DFT problem, followed by the presentation of the mathematical formalism in Section III. Section IV describes the various steps involved in the subspace pro-jection technique within the framework of spectral finite-element discretization. Section V presents the numerical study on three representative materials systems demon-strating the accuracy, computational efficiency and scal-ing of our approach. We finally conclude with a summary and outlook in Section VI.

II. KOHN-SHAM DENSITY FUNCTIONAL THEORY

The variational problem of evaluating the ground-state properties in density functional theory for a materials system consisting of N_a nuclei is equivalent to solving the following non-linear Kohn-Sham eigenvalue problem[2]:

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\rho, \mathbf{R})\right)\psi_i = \epsilon_i\psi_i, \quad i = 1, 2, \cdots \quad (1)$$

where ϵ_i and ψ_i denote the eigenvalues and correspond-ing eigenfunctions (also referred to as the canonical wave-functions) of the Hamiltonian, respectively. We denote by $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \cdots \mathbf{R}_{N_a}\}$ the collection of all nuclear positions in the materials system. In the present work, we restrict ourselves to a non-periodic setting and present the formulation for this case. However, we note that the

ideas presented in the work can easily be generalized to periodic or semi-periodic materials systems. Further, for the sake of simplicity, we present the formulation for a spin independent Hamiltonian, and note that the extension to spin-dependent Hamiltonians[71] follows along similar lines. The electron density in terms of the canonical wavefunctions is given by

$$\rho(\mathbf{x}) = 2 \sum_i f(\epsilon_i, \mu) |\psi_i(\mathbf{x})|^2 , \qquad (2)$$

where $f(\epsilon_i, \mu)$ is the orbital occupancy function, whose range lies in the interval $[0, 1]$, and μ represents the Fermi-energy. We note that the factor 2 in the above equation represents the case of a spin independent system, where each orbital is occupied by two electrons. In ground-state calculations, the orbital occupancy function $f(\epsilon, \mu)$ is given by the Heaviside function

$$f(\epsilon, \mu) = \begin{cases} 1 & \text{if } \epsilon < \mu, \\ 0 & \text{otherwise .} \end{cases} \qquad (3)$$

However, it is common in density functional theory calculations to represent f by the Fermi distribution[4,47] given by

$$f(\epsilon, \mu) = \frac{1}{1 + \exp\left(\frac{\epsilon - \mu}{\sigma}\right)} , \qquad (4)$$

where σ is a smearing parameter. We note that as $\sigma \to 0$, the Fermi distribution tends to the Heaviside function. Such a smearing of the orbital occupancy function avoids numerical instabilities that may arise in the solution of the non-linear Kohn-Sham eigenvalue problem, especially in materials systems where there are a large number of eigenstates around the Fermi energy. Often, the numerical value of σ is chosen to be $\sigma = k_B T$, where k_B denotes the Boltzmann constant and T denotes the finite temperature used for smearing the orbital occupancy function. The Fermi-energy μ is computed from the constraint on the total number of electrons in the system (N_e) given by

$$\int \rho(\mathbf{x}) \, d\mathbf{x} = 2 \sum_i f(\epsilon_i, \mu) = N_e . \qquad (5)$$

Henceforth, we denote by f the finite-temperature Fermi distribution given in equation (4).

The effective single-electron potential, $V_{\text{eff}}(\rho, \mathbf{R})$, in the Hamiltonian in equation (1) is given by

$$V_{\text{eff}}(\rho, \mathbf{R}) = V_{\text{xc}}(\rho) + V_{\text{H}}(\rho) + V_{\text{ext}}(\mathbf{R}) . \qquad (6)$$

In the above, $V_{\text{xc}}(\rho)$ denotes the exchange-correlation potential that accounts for quantum-mechanical interactions between electrons, and is given by the first varia-

tional derivative of the exchange-correlation energy E_{xc} :

$$V_{\text{xc}}(\rho) = \frac{\delta E_{\text{xc}}}{\delta \rho} . \qquad (7)$$

In this work, we adopt the local-density approximation (LDA)[71] for the exchange-correlation functional. Other approximations such as the local spin density approximation (LSDA)[71] and generalized gradient approximation (GGA)[72,73] can be incorporated into the present formulation in a straightforward manner. In the case of LDA[74,75], the exchange-correlation energy is given by

$$E_{\text{xc}}(\rho) = \int \varepsilon_{\text{xc}}(\rho) \rho(\mathbf{x}) \, d\mathbf{x} , \qquad (8)$$

where $\varepsilon_{\text{xc}}(\rho) = \varepsilon_x(\rho) + \varepsilon_c(\rho)$, and

$$\varepsilon_x(\rho) = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \rho^{1/3}(\mathbf{x}) , \qquad (9)$$

$$\varepsilon_c(\rho) = \begin{cases} \frac{\gamma}{(1 + \beta_1 \sqrt{(r_s)} + \beta_2 r_s)} & r_s \geq 1, \\ A \log r_s + B + C \, r_s \log r_s + D \, r_s & r_s < 1, \end{cases} \qquad (10)$$

and $r_s = (3/4\pi\rho)^{1/3}$. Specifically, we use the Ceperley and Alder constants[75] in equation (10). The remainder of the effective single-electron potential accounts for the electrostatic interactions. In particular, $V_{\text{H}}(\rho)$ (Hartree potential) denotes the classical electrostatic potential corresponding to the electron density distribution and is given by

$$V_{\text{H}}(\rho) = \int \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \, d\mathbf{x}' . \qquad (11)$$

$V_{\text{ext}}(\mathbf{x}, \mathbf{R})$ denotes the external electrostatic potential corresponding to the nuclear charges, and is given by

$$V_{\text{ext}}(\mathbf{x}, \mathbf{R}) = -\sum_{I=1}^{N_a} \frac{Z_I}{|\mathbf{x} - \mathbf{R}_I|} , \qquad (12)$$

with Z_I denoting the atomic number of the Ith nucleus in the given materials system. The tightly bound core electrons close to the nucleus of an atom may not influence the chemical bonding, and may not play a significant role in governing many material properties. Hence, it is a common practice to adopt the pseudopotential approach, where only the wavefunctions for the valence electrons are computed. The pseudopotential, which provides the effective electrostatic potential of the nucleus and core electrons, is often defined by an operator

$$\mathcal{V}_{\text{PS}} = \mathcal{V}_{\text{loc}} + \mathcal{V}_{\text{nl}} , \qquad (13)$$

where $\mathcal{V}_{\text{loc}}$ is the local part of the pseudopotential operator and $\mathcal{V}_{\text{nl}}$ is the non-local part of the operator. In this work, we use the norm-conserving Troullier-

Martins pseudopotential[76] in the Kleinman-Bylander form[77]. The action of these operators on a Kohn-Sham wavefunction in the real-space is given by

$$V_{\text{loc}}(\mathbf{x}, \mathbf{R})\psi(\mathbf{x}) = \sum_{J=1}^{N_a} V_{\text{loc}}^J(\mathbf{x} - \mathbf{R}_J)\psi(\mathbf{x}), \qquad (14)$$

$$V_{\text{nl}}(\mathbf{x}, \mathbf{R})\psi(\mathbf{x}) = \sum_{J=1}^{N_a} \sum_{lm} C_{lm}^J V_{lm}^J \phi_{lm}^J(\mathbf{x} - \mathbf{R}_J)\Delta V_l^J(\mathbf{x} - \mathbf{R}_J), \qquad (15)$$

where

$$\Delta V_l^J(\mathbf{x} - \mathbf{R}_J) = V_l^J(\mathbf{x} - \mathbf{R}_J) - V_{\text{loc}}^J(\mathbf{x} - \mathbf{R}_J),$$

$$C_{lm}^J = \int \phi_{lm}^J(\mathbf{x} - \mathbf{R}_J)\Delta V_l^J(\mathbf{x} - \mathbf{R}_J)\psi(\mathbf{x})\, d\mathbf{x},$$

$$\frac{1}{V_{lm}^J} = \int \phi_{lm}^J(\mathbf{x} - \mathbf{R}_J)\Delta V_l^J(\mathbf{x} - \mathbf{R}_J)\phi_{lm}^J(\mathbf{x} - \mathbf{R}_J)\, d\mathbf{x}.$$

In the above, $V_l^J(\mathbf{x} - \mathbf{R}_J)$ denotes the pseudopotential component of the atom J corresponding to the azimuthal quantum number l, $V_{\text{loc}}^J(\mathbf{x} - \mathbf{R}_J)$ is the corresponding local potential, and $\phi_{lm}^J(\mathbf{x} - \mathbf{R}_J)$ is the corresponding single-atom pseudo-wavefunction with azimuthal quantum number l and magnetic quantum number m. We note that the computation of the Hartree potential (V_{H}), the external potential (V_{ext}) in the case of all-electron calculations, or the local part of pseudopotential (V_{loc}), is extended in real-space. However, noting that the kernel corresponding to the extended interactions is the Green's function of the Laplace operator, these quantities can be efficiently computed by taking recourse to the solution of a Poisson problem. We refer to Suryanarayana et al.[38] and Motamarri et al.[25] for details of the local reformulation of the extended electrostatic interactions.

Finally, for given positions of nuclei, the system of equations corresponding to the Kohn-Sham eigenvalue problem are:

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{xc}}(\rho) + V_{\text{H}}(\rho) + V_{\text{ext}}(\mathbf{R})\right)\psi_i = \epsilon_i\psi_i, \quad (16a)$$

$$2\sum_i f(\epsilon_i, \mu) = N_e, \qquad (16b)$$

$$\rho(\mathbf{x}) = 2\sum_i f(\epsilon_i, \mu)|\psi_i(\mathbf{x})|^2. \qquad (16c)$$

In the case of all-electron calculations, V_{ext} denotes the Coulomb-singular potential corresponding to all nuclei, and further, all the wavefunctions, including those of the core electrons, are computed. In the case of pseudopotential calculations, $V_{\text{ext}} = V_{\text{loc}} + V_{\text{nl}}$ and only the wavefunctions corresponding to the valence electrons are computed. Furthermore, the formulation in (16) represents a nonlinear eigenvalue problem which has to be solved self-consistently. Upon self-consistently solving (16), the ground-state energy of the system is given by

$$E_{\text{tot}} = E_{\text{band}} + \int (\varepsilon_{\text{xc}}(\rho) - V_{\text{xc}}(\rho))\, \rho\, d\mathbf{x}$$
$$- \frac{1}{2}\int \rho V_{\text{H}}(\rho)\, d\mathbf{x} + E_{\text{ZZ}}, \qquad (17)$$

where E_{band}, denoting the band energy, is given by

$$E_{\text{band}} = 2\sum_i f(\epsilon_i, \mu)\epsilon_i, \qquad (18)$$

and E_{ZZ}, denoting the nuclear-nuclear repulsive energy, is given by

$$E_{\text{ZZ}} = \sum_{\substack{I,J=1 \\ I \neq J}}^{N_a} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \qquad (19)$$

In the above, Z_I denotes the valence charge of the Ith nucleus in the case of pseudopotential calculations, and denotes the atomic number in the case of all-electron calculation.

III. MATHEMATICAL FORMULATION

In this section, we discuss some mathematical preliminaries and present the expression for the projection of the Hamiltonian operator into a subspace spanned by a non-orthogonal basis. Subsequently, we derive the density matrix corresponding to the projected Hamiltonian, and present the expressions for the computation of electron density, constraint on the number of electrons and the band energy in terms of the projected Hamiltonian. This constitutes the mathematical formulation for the subspace projection technique within the framework of spectral finite-element discretization, described in the next section.

Let $\mathcal{H}$ denote the Hermitian operator representing the Hamiltonian of interest defined on the infinite dimensional Hilbert space $\mathbb{H}$. We note that $\mathbb{H}$ is a space of functions equipped with the inner product $\langle .|.\rangle$ and a norm $\| . \|$ derived from the inner product. Let $\mathbb{V}_h^M \subset \mathbb{H}$ be the finite-dimensional subspace of $\mathbb{H}$ with dimension M, and $\{|q_i\rangle\}$ denote an orthonormal basis for $\mathbb{V}_h^M$. In the present case, the representation of $|q_i\rangle$ in the real space, $\langle \mathbf{x}|\, q_i\rangle = q_i(\mathbf{x})$, denotes the Löwdin orthonormalized finite-element basis employed in this study, as discussed subsequently in section IV. We define the projection operator into the subspace $\mathbb{V}_h^M$ to be $\mathcal{P}^q : \mathbb{H} \rightarrow \mathbb{V}_h^M$ given by

$$\mathcal{P}^q = \sum_{i=1}^M |q_i\rangle \langle q_i|. \qquad (20)$$

The projection of the Hamiltonian into $\mathbb{V}_h^M$ is given by $\mathcal{P}^q\mathcal{H} : \mathbb{V}_h^M \rightarrow \mathbb{V}_h^M$, or equivalently $\mathcal{P}^q\mathcal{H}\mathcal{P}^q : \mathbb{H} \rightarrow \mathbb{V}_h^M$.

We denote the operator corresponding to the projected Hamiltonian to be $\tilde{\mathcal{H}} \equiv \mathcal{P}^q \mathcal{H} \mathcal{P}^q$, and the matrix corresponding to $\tilde{\mathcal{H}}$ expressed in $\{|q_i\rangle\}$ basis by $\tilde{\mathbf{H}}$ with the matrix element given by $\tilde{\mathrm{H}}_{ij} = \left\langle q_i \middle| \tilde{\mathcal{H}} \middle| q_j \right\rangle$.

A. Projection of the Hamiltonian into a non-orthogonal basis

We consider a subspace $\mathbb{V}^N \subset \mathbb{V}_h^M$, which approximates the occupied eigenspace of $\tilde{\mathcal{H}}$ that can be computed, for instance, using a Chebyshev filtering approach[25,44] (discussed subsequently in section IV). Let $\{|\phi_\alpha\rangle\}$ represent a non-orthonormal basis which spans $\mathbb{V}^N$. We denote by $\mathcal{P}^\phi : \mathbb{V}_h^M \to \mathbb{V}^N$ the projection operator into the space $\mathbb{V}^N$, and is given by

$$\mathcal{P}^\phi = \sum_{\alpha,\beta=1}^N |\phi_\alpha\rangle S_{\alpha\beta}^{-1} \langle\phi_\beta| , \qquad (21)$$

where $\mathbf{S}$ denotes the overlap matrix with matrix elements $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$. We denote the projection of $\tilde{\mathcal{H}}$ into $\mathbb{V}^N$ by $\mathcal{H}^\phi$, and is given by $\mathcal{H}^\phi \equiv \mathcal{P}^\phi \tilde{\mathcal{H}} \mathcal{P}^\phi : \mathbb{V}_h^M \to \mathbb{V}^N$. Denoting the matrix corresponding to $\mathcal{H}^\phi$ expressed in $\{|q_i\rangle\}$ basis to be $\mathbf{H}^q$ and in $\{|\phi_\alpha\rangle\}$ basis to be $\mathbf{H}^\phi$, we derive the expressions for the corresponding matrix elements H_{ij}^q and $\mathrm{H}_{\alpha\beta}^\phi$ using equations (20) and (21) as follows:

$$\mathrm{H}_{ij}^q = \left\langle q_i \middle| \mathcal{H}^\phi \middle| q_j \right\rangle = \left\langle q_i \middle| \mathcal{P}^\phi \tilde{\mathcal{H}} \mathcal{P}^\phi \middle| q_j \right\rangle$$

$$= \sum_{\substack{\alpha,\beta=1 \\ \gamma,\delta=1}}^N \langle q_i| \phi_\alpha\rangle S_{\alpha\beta}^{-1} \left\langle \phi_\beta \middle| \tilde{\mathcal{H}} \middle| \phi_\gamma \right\rangle S_{\gamma\delta}^{-1} \langle \phi_\delta| q_j\rangle . \quad (22)$$

Since $\mathcal{P}^q$ is idempotent ($\mathcal{P}^q \mathcal{P}^q = \mathcal{P}^q$), we note that $\tilde{\mathcal{H}} = \mathcal{P}^q \tilde{\mathcal{H}} \mathcal{P}^q$. Hence, equation (22) can be written as

$$\mathrm{H}_{ij}^q = \sum_{\substack{\alpha,\beta=1 \\ \gamma,\delta=1}}^N \langle q_i| \phi_\alpha\rangle S_{\alpha\beta}^{-1} \left\langle \phi_\beta \middle| \mathcal{P}^q \tilde{\mathcal{H}} \mathcal{P}^q \middle| \phi_\gamma \right\rangle S_{\gamma\delta}^{-1} \langle \phi_\delta| q_j\rangle$$

$$= \sum_{\substack{\alpha,\beta=1 \\ \gamma,\delta=1}}^N \sum_{k,l=1}^M \phi_{i\alpha} S_{\alpha\beta}^{-1} \phi_{k\beta}^* \tilde{\mathrm{H}}_{kl} \phi_{l\gamma} S_{\gamma\delta}^{-1} \phi_{j\delta}^* , \quad (23)$$

where $\phi_{i\alpha} = \langle q_i| \phi_\alpha\rangle$ and $\phi_{k\beta}^*$ denotes the complex conjugate of $\phi_{k\beta}$. The above equation can be conveniently recast in terms of matrices as:

$$\mathbf{H}^q = \mathbf{\Phi}\mathbf{\Phi}^+ \tilde{\mathbf{H}}\mathbf{\Phi}\mathbf{\Phi}^+ \qquad (24)$$

where $\mathbf{\Phi}$ denotes a matrix whose column vectors are the components of $|\phi_\alpha\rangle$ in $|q_i\rangle$ basis, and $\mathbf{\Phi}^+$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{\Phi}$ given by

$$\mathbf{\Phi}^+ = \mathbf{S}^{-1}\mathbf{\Phi}^\dagger \qquad (25)$$

with $\mathbf{\Phi}^\dagger$ denoting the conjugate transpose of the matrix $\mathbf{\Phi}$. We now derive the expression for the matrix element corresponding to the operator $\mathcal{H}^\phi$ expressed in the non-orthonormal basis $\{|\phi_\alpha\rangle\}$:

$$\mathrm{H}_{\alpha\beta}^\phi = \sum_{\gamma=1}^N S_{\alpha\gamma}^{-1} \langle\phi_\gamma| \mathcal{H}^\phi |\phi_\beta\rangle = \sum_{\gamma=1}^N S_{\alpha\gamma}^{-1} \left\langle\phi_\gamma \middle| \mathcal{P}^\phi \tilde{\mathcal{H}} \mathcal{P}^\phi \middle| \phi_\beta\right\rangle$$

$$= \sum_{\gamma=1}^N S_{\alpha\gamma}^{-1} \left\langle\phi_\gamma \middle| \tilde{\mathcal{H}} \middle| \phi_\beta\right\rangle = \sum_{\gamma=1}^N S_{\alpha\gamma}^{-1} \left\langle\phi_\gamma \middle| \mathcal{P}^q \tilde{\mathcal{H}} \mathcal{P}^q \middle| \phi_\beta\right\rangle$$

$$= \sum_{\gamma=1}^N \sum_{k,l=1}^M S_{\alpha\gamma}^{-1} \langle\phi_\gamma| q_k\rangle \left\langle q_k \middle| \tilde{\mathcal{H}} \middle| q_l\right\rangle \langle q_l| \phi_\beta\rangle$$

$$= \sum_{\gamma=1}^N \sum_{k,l=1}^M S_{\alpha\gamma}^{-1} \phi_{k\gamma}^* \tilde{\mathrm{H}}_{kl} \phi_{l\beta} . \qquad (26)$$

Using matrices, the above equation (26) can be conveniently expressed as

$$\mathbf{H}^\phi = \mathbf{\Phi}^+ \tilde{\mathbf{H}}\mathbf{\Phi} . \qquad (27)$$

We also note the following relation between $\mathbf{H}^\phi$ and $\mathbf{H}^q$ using equations (24) and (27):

$$\mathbf{H}^q = \mathbf{\Phi}\mathbf{H}^\phi\mathbf{\Phi}^+ . \qquad (28)$$

B. Density Matrix

We now consider the single particle density operator (Γ) corresponding to $\mathcal{H}^\phi$ given by

$$\Gamma = \sum_{i=1}^N f(\epsilon_i^\phi) \left|\psi_i^\phi\right\rangle \left\langle\psi_i^\phi\right|$$

$$= f(\mathcal{H}^\phi) , \qquad (29)$$

where ϵ_i^ϕ and $\left|\psi_i^\phi\right\rangle$ denote the eigenvalues and the corresponding eigenvectors of $\mathcal{H}^\phi$, and we note that the relation $\Gamma = f(\mathcal{H}^\phi)$ follows from the spectral decomposition of the Hermitian operator $\mathcal{H}^\phi$. Denoting the matrix representation of Γ in $\{|q_i\rangle\}$ basis by $\mathbf{\Gamma}$, we have the following relation between the matrices $\mathbf{\Gamma}$ and $\mathbf{H}^q$ from equation (29):

$$\mathbf{\Gamma} = f(\mathbf{H}^q) . \qquad (30)$$

We now derive the expression for $\mathbf{\Gamma}$ in terms of the matrices $\mathbf{H}^\phi$ and $\mathbf{\Phi}$. To this end, we note that the function $f(\epsilon)$ represents the Fermi distribution (equation (4)), which is an analytic function. Hence, this admits a power series representation given by

$$f(\mathbf{H}^q) = \sum_{k=0}^\infty a_k \left(\mathbf{H}^q\right)^k . \qquad (31)$$

Using equation (24) and the relation $\mathbf{\Phi}^+\mathbf{\Phi} = \mathbf{I}_N$, where $\mathbf{I}_N$ denotes an identity matrix of dimension N, we have

$$
\begin{aligned}
\mathbf{\Gamma} &= \sum_{k=0}^{\infty} a_k \left(\mathbf{H}^q\right)^k = \sum_{k=0}^{\infty} a_k \left(\mathbf{\Phi}\mathbf{\Phi}^+\tilde{\mathbf{H}}\mathbf{\Phi}\mathbf{\Phi}^+\right)^k \\
&= \sum_{k=0}^{\infty} a_k \, \mathbf{\Phi}(\mathbf{\Phi}^+\tilde{\mathbf{H}}\mathbf{\Phi})^k\mathbf{\Phi}^+ = \mathbf{\Phi} \left(\sum_{k=0}^{\infty} a_k(\mathbf{H}^\phi)^k\right) \mathbf{\Phi}^+ \\
&= \mathbf{\Phi}f(\mathbf{H}^\phi)\mathbf{\Phi}^+ \, .
\end{aligned}
\tag{32}
$$

We note that the electron density $\rho(\mathbf{x})$ is related to the diagonal of the density operator Γ expressed in the real space, and is given by

$$
\rho(\mathbf{x}) = 2 \, \langle\mathbf{x}| \, \Gamma \, |\mathbf{x}\rangle = 2 \sum_{i,j=1}^{M} \Gamma_{ij} \, q_i(\mathbf{x}) \, q_j(\mathbf{x}) \,,
\tag{33}
$$

where Γ_{ij} denote the matrix elements of $\mathbf{\Gamma}$. We note that factor 2 in the above equation represents the case of a spin independent system, where each orbital is occupied by two electrons. We now derive an expression for the constraint on the total number of electrons as given in equation (5). The total number of electrons in the materials system is given by

$$
\begin{aligned}
\int \rho(\mathbf{x}) \, d\mathbf{x} &= 2 \int \sum_{i,j=1}^{M} \Gamma_{ij} \, q_i(\mathbf{x}) \, q_j(\mathbf{x}) \, d\mathbf{x} = 2 \sum_{i=1}^{M} \Gamma_{ii} \\
&= 2 \operatorname{tr}\left(\mathbf{\Gamma}\right) = 2 \operatorname{tr}\left(\mathbf{\Phi}f(\mathbf{H}^\phi)\mathbf{\Phi}^+\right) \\
&= 2 \operatorname{tr}\left(f(\mathbf{H}^\phi)\mathbf{\Phi}^+\mathbf{\Phi}\right) = 2 \operatorname{tr}\left(f(\mathbf{H}^\phi)\right) \, . \quad (34)
\end{aligned}
$$

Hence, the constraint on the total number of electrons in equation (5) can be reformulated as

$$
2 \operatorname{tr}\left(f(\mathbf{H}^\phi)\right) = N_e \,.
\tag{35}
$$

Finally, the band-energy (E_b), which is required in the calculation of the ground-state energy, is given by

$$
\begin{aligned}
E_b &= 2 \operatorname{tr}\left(\mathbf{\Gamma}\,\mathbf{H}^q\right) = 2 \operatorname{tr}\left(\mathbf{\Phi}f(\mathbf{H}^\phi)\mathbf{\Phi}^+\mathbf{\Phi}\mathbf{H}^\phi\mathbf{\Phi}^+\right) \\
&= 2 \operatorname{tr}\left(\mathbf{\Phi}f(\mathbf{H}^\phi)\mathbf{H}^\phi\mathbf{\Phi}^+\right) = 2 \operatorname{tr}\left(\mathbf{\Phi}^+\mathbf{\Phi}f(\mathbf{H}^\phi)\mathbf{H}^\phi\right) \\
&= 2 \operatorname{tr}\left(f(\mathbf{H}^\phi)\mathbf{H}^\phi\right) \, .
\end{aligned}
\tag{36}
$$

IV. SUBSPACE-PROJECTION ALGORITHM USING FINITE-ELEMENT BASIS

In this section, we introduce the finite-element discretization of the Kohn Sham eigenvalue problem along the lines of our prior work[25], and subsequently present the subspace projection algorithm used to reduce the computational complexity involved in the solution of the Kohn-Sham problem.

A. Discrete Kohn-Sham eigenvalue problem

As denoted in the previous section, let $\mathbb{V}_h^M$ represent the finite-dimensional subspace with dimension M. In particular, we consider the finite-dimensional space spanned by the finite-element basis, which is a piecewise polynomial basis generated from a finite-element discretization[22] whose characteristic mesh-size is denoted by h. The representation of the various fields in the Kohn-Sham eigenvalue problem (16)—the wavefunctions and the electrostatic potential—in the finite-element basis is given by

$$
\psi_i^h(\mathbf{x}) = \sum_{j=1}^{M} N_j^h(\mathbf{x})\psi_i^j \,,
\tag{37}
$$

$$
\phi^h(\mathbf{x}) = \sum_{j=1}^{M} N_j^h(\mathbf{x})\phi^j \,,
\tag{38}
$$

where $N_j^h : 1 \leq j \leq M$ denotes the finite-element basis spanning $\mathbb{V}_h^M$. We note that ψ_i^h and ϕ^h denote the finite-element discretized fields, with ψ_i^j and ϕ^j denoting the coefficients in the expansion of the ith discretized wavefunction and electrostatic potential, respectively. Further, from the Kronecker delta property[22] of the finite-element basis, ψ_i^j and ϕ^j also correspond to the nodal values of the respective fields at the jth node on the finite-element mesh. We note that the finite-element basis is a non-orthogonal basis, and, thus, the discretization of the Kohn-Sham eigenvalue problem (16) results in a generalized eigenvalue problem given by

$$
\mathbf{H}\hat{\mathbf{\Psi}}_i = \epsilon_i^h \mathbf{M}\hat{\mathbf{\Psi}}_i \,,
\tag{39}
$$

where $\mathbf{H}$ denotes the discrete Hamiltonian matrix with matrix elements H_{jk}, $\mathbf{M}$ denotes the overlap matrix (or commonly referred to as the mass matrix in finite-element literature) with matrix elements M_{jk}, and ϵ_i^h denotes the ith eigenvalue corresponding to the discrete eigenvector $\hat{\mathbf{\Psi}}_i$. The expression for the discrete Hamiltonian matrix H_{jk} in a non-periodic setting with $\mathbb{H} = H_0^1(\Omega)$ (space of functions with compact support on Ω) is given by

$$
H_{jk} = H_{jk}^{\text{loc}} + H_{jk}^{\text{nl}} \,,
\tag{40}
$$

where

$$
\begin{aligned}
\mathrm{H}_{jk}^{\text{loc}} = \quad &\frac{1}{2} \int \nabla N_j(\mathbf{x}).\nabla N_k(\mathbf{x}) \, d\mathbf{x} \\
&+ \int V_{\text{eff,loc}}^h(\mathbf{x}, \mathbf{R}) N_j(\mathbf{x}) N_k(\mathbf{x}) \, d\mathbf{x} \, .
\end{aligned}
\tag{41}
$$

In the above, $V_{\text{eff,loc}}^h$ denotes the local part of the effective single-electron potential computed in the finite-element basis (the discretized effective single-electron potential). In the case of pseudopotential calculations, $V_{\text{eff,loc}}^h = V_{\text{xc}}^h + V_{\text{H}}^h + V_{\text{loc}}^h$, where V_{loc}^h denotes the discretized local

part of the pseudopotential (cf. equation (14)). In the case of all-electron calculations, $V_{\text{eff,loc}}^h = V_{\text{xc}}^h + V_{\text{H}}^h + V_{\text{ext}}^h$ with V_{ext}^h denoting the discretized all-electron external Coulomb potential from all nuclei. As noted in section II, V_{H}^h and V_{ext}^h are computed as solutions of the Poisson's equation in the finite-element basis. In the case of all-electron calculations, the term H_{jk}^{nl} in (40) is zero, while in the case of pseudopotential calculations it is given by

$$H_{jk}^{nl} = \sum_{J=1}^{N_a} \sum_{lm} C_{lm,j}^J V_{lm}^J C_{lm,k}^J \,, \qquad (42)$$

where

$$C_{lm,j}^J = \int \phi_{lm}^J(\mathbf{x} - \mathbf{R}_J) \Delta V_l^J(\mathbf{x} - \mathbf{R}_J) N_j(\mathbf{x}) \, d\mathbf{x} \,. \quad (43)$$

Finally, the matrix elements of the overlap matrix $\mathbf{M}$ are given by $\mathrm{M}_{jk} = \int N_j(\mathbf{x}) N_k(\mathbf{x}) \, d\mathbf{x}$. We note that the matrices $\mathbf{H}^{\text{loc}}$ and $\mathbf{M}$ are sparse as the finite-element basis functions are local in real space and have a compact support (a finite region where the function is non-zero and zero elsewhere). Further, the vectors $C_{lm,j}^J$ in $\mathbf{H}^{nl}$ are also sparse since both $\phi_{lm}^J(\mathbf{x} - \mathbf{R}_J)$ and $\Delta V_l^J(\mathbf{x} - \mathbf{R}_J)$ have a compact support, thus rendering a sparse structure to the discrete Hamiltonian $\mathbf{H}$.

It is now desirable to transform the generalized eigenvalue problem (39) into a standard eigenvalue problem for which a wide range of efficient solution procedures are available. Since the matrix $\mathbf{M}$ is positive definite symmetric, there exists a unique positive definite symmetric square root of $\mathbf{M}$, and is denoted by $\mathbf{M}^{1/2}$. Hence, the following holds true:

$$\mathbf{H}\hat{\boldsymbol{\Psi}}_i = \epsilon_i^h \mathbf{M}\hat{\boldsymbol{\Psi}}_i$$
$$\Rightarrow \quad \mathbf{H}\hat{\boldsymbol{\Psi}}_i = \epsilon_i^h \mathbf{M}^{1/2}\mathbf{M}^{1/2}\hat{\boldsymbol{\Psi}}_i$$
$$\Rightarrow \quad \tilde{\mathbf{H}}\tilde{\boldsymbol{\Psi}}_i = \epsilon_i^h \tilde{\boldsymbol{\Psi}}_i \,, \qquad (44)$$

where

$$\tilde{\boldsymbol{\Psi}}_i = \mathbf{M}^{1/2}\hat{\boldsymbol{\Psi}}_i$$
$$\tilde{\mathbf{H}} = \mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{-1/2} \,.$$

We note that $\tilde{\mathbf{H}}$ is a Hermitian matrix, and (44) represents a standard Hermitian eigenvalue problem. The actual eigenvectors are recovered by the transformation $\hat{\boldsymbol{\Psi}}_i = \mathbf{M}^{-1/2}\tilde{\boldsymbol{\Psi}}_i$. We remark that $\tilde{\boldsymbol{\Psi}}_i$ is a vector containing the expansion coefficients of the discretized eigenfunction $\psi_i^h(\mathbf{x})$ expressed in Löwdin orthonormalized finite-element basis $q_j(\mathbf{x}) : 1 \leq j \leq M$ spanning the finite-element space. We note the following relation between $q_j(\mathbf{x})$ and $N_j(\mathbf{x})$:

$$q_j(\mathbf{x}) = \sum_{k=1}^{M} M_{jk}^{-1/2} N_k(\mathbf{x}) \,. \qquad (45)$$

Furthermore, we note that the transformation to a standard eigenvalue problem (44) is computationally advantageous if the matrix $\mathbf{M}^{-1/2}$ can be evaluated with modest computational cost and the matrix $\tilde{\mathbf{H}}$ has the same sparsity structure as the matrix $\mathbf{H}$. This is immediately possible by using a spectral finite-element basis in conjunction with the use of Gauss-Lobatto-Legendre (GLL) quadrature for the evaluation of integrals in the overlap matrix, that renders the overlap matrix diagonal[25]. This has been employed in Motamarri et al.[25] to develop an efficient and scalable approach using adaptive higher-order spectral finite-element discretization of the Kohn-Sham DFT problem. Materials systems as large as 1688 atom aluminum nano-clusters have been simulated, and the parallel scalability of the algorithms has been demonstrated up to 200 processors[25]. In Motamarri et al.[25], the self-consistent field (SCF) iteration consists of employing a Chebyshev filtering approach[44] to compute the occupied eigenspace of the spectral finite-element discretized Hamiltonian $\tilde{\mathbf{H}}$. The computational complexity of the Chebyshev filtering scales as $\mathcal{O}(M\,N)$, where N denotes the number of eigenstates of interest. Upon computing the approximate eigenspace, the Chebyshev filtered vectors spanning the eigenspace are orthonormalized, and the projection of the Hamiltonian into this orthonormal basis is computed. The computational complexity of this orthonormalization and projection scales as $\mathcal{O}(M\,N^2)$. Finally, the projected Hamiltonian is diagonalized to compute the eigenvalues and eigenvectors, which, in turn, are used in the computation of the electron density. The computational complexity of this step scales as $\mathcal{O}(N^3)$. Hence, the approach developed in Motamarri et al.[25], subsequently referred to as the Chebyshev filtered subspace iteration for the finite-element basis (ChFSI-FE), scales as $\mathcal{O}(N^3)$ asymptotically (as $M \propto N$), thus limiting accessible systems to a few thousand atoms. In order to address this significant limitation, we build on our prior work[25] to develop a subspace projection technique that reduces the computational complexity of solving the Kohn-Sham problem. The proposed approach treats both insulating and metallic systems under a single framework, and is applicable for both pseudopotential and all-electron calculations. The key ideas involved in the method for a single self-consistent field (SCF) iteration are discussed below.

B. Chebyshev filtered subspace iteration

Chebyshev filtered subspace iteration (ChFSI)[78] belongs to the class of subspace iteration techniques which are generalizations of the power method applied to a subspace. As the ground-state electron density, and subsequently the ground-state energy, depends solely on the occupied eigenspace—the vector space spanned by the eigenfunctions corresponding to the occupied states—the ChFSI technique exploits the fast growth property of Chebyshev polynomial in $(-\infty, -1)$ to magnify the relevant spectrum and thereby providing an efficient approach for the solution of the Kohn-Sham eigenvalue

problem. We refer to Zhou et al.[44] and Motamarri et al.[25] for the application of this technique to electronic structure calculations in the context of finite-difference and finite-element discretizations, respectively. We also refer to Garcia et al.[62] for a linear-scaling subspace iteration technique based on Chebyshev filtering for insulating systems. In the present work, we also adopt the Chebyshev filtering procedure to find the occupied eigenspace at any given SCF iteration. We start with an initial subspace $\overline{\mathbb{V}}^N$ of dimension N ($N > N_e/2$) spanned by the set of localized wavefunctions $\{\psi_1(\mathbf{x}), \psi_2(\mathbf{x}), \cdots, \psi_N(\mathbf{x})\}$ obtained from the previous SCF iteration (cf. section IV C for details on the construction of the localized wavefunctions). We note that, here, and subsequently, all the electronic fields (wavefunctions and electron density) denote finite-element discretized fields, and we have dropped the superscript h for notational simplicity. We denote by $\mathbf{X}$ the matrix whose column vectors are the coefficients of expansion of these localized wavefunctions in the Löwdin orthonormalized spectral finite-element basis ($q_j : 1 \leq j \leq M$). The Chebyshev filtered subspace iteration then proceeds as follows:

a. Construction of shifted and scaled Hamiltonian: The discretized Hamiltonian $\tilde{\mathbf{H}}$ is scaled and shifted to construct $\bar{\mathbf{H}}$ such that the unwanted (unoccupied) spectrum of $\tilde{\mathbf{H}}$ is mapped to $[-1, 1]$ and the wanted (occupied) spectrum into $(-\infty, -1)$. Hence

$$\bar{\mathbf{H}} = \frac{1}{e}(\tilde{\mathbf{H}} - c\boldsymbol{I}) \text{ where } e = \frac{b-a}{2} \quad c = \frac{a+b}{2}. \quad (46)$$

Here a and b denote the upper bound of the wanted and unwanted spectrum of $\tilde{\mathbf{H}}$, respectively. The upper bound b is obtained inexpensively using a very small number of Lanczos iterations[79] whose computational complexity scales as $\mathcal{O}(M)$. The upper bound of the wanted spectrum is chosen as the largest Rayleigh quotient of $\tilde{\mathbf{H}}$ in the occupied eigenspace computed in the previous SCF iteration.

b. Construction of Chebyshev filter: In a given SCF iteration, the action of a Chebyshev filter on $\mathbf{X}$ is given by

$$\mathbf{Y} = T_m(\bar{\mathbf{H}})\mathbf{X}, \quad (47)$$

where $\mathbf{Y}$ denotes the matrix whose column vectors are the coefficients of expansion of the Chebyshev filtered wavefunctions $\{\psi_1^f(\mathbf{x}), \psi_2^f(\mathbf{x}), \cdots, \psi_N^f(\mathbf{x})\}$ expressed in the Löwdin orthonormalized finite-element basis. We note that the subspace $\mathbb{V}^N$ spanned by these Chebyshev filtered wavefunctions is a close approximation to the occupied eigenspace. In the above, the filter $T_m(\bar{\mathbf{H}})$ is constructed using a Chebyshev polynomial $T_m(x)$ of degree m, and the action of the filter on $\mathbf{X}$ can be recursively computed[80] as

$$T_m(\bar{\mathbf{H}})\mathbf{X} = \left[2\bar{\mathbf{H}} T_{m-1}(\bar{\mathbf{H}}) - T_{m-2}(\bar{\mathbf{H}}) \right] \mathbf{X}, \quad (48)$$

with $T_0(\bar{\mathbf{H}})\mathbf{X} = \mathbf{X}$ and $T_1(\bar{\mathbf{H}})\mathbf{X} = \bar{\mathbf{H}}\mathbf{X}$. It is evident from equation (48) that the application of Chebyshev filter on $\mathbf{X}$ involves matrix-vector multiplications between the discretized Hamiltonian $\tilde{\mathbf{H}}$ and the vectors obtained during the course of recursive iteration (48). Further, we note that the finite-element basis functions $N_j : 1 \leq j \leq M$ are local. Hence, the discretized Hamiltonian $\tilde{\mathbf{H}}$ expressed in the Löwdin orthonormalized spectral finite-element basis $q_j : 1 \leq j \leq M$ is sparse, and has the exact same sparsity as $\mathbf{H}$ if the matrix $\mathbf{M}$ is evaluated using the GLL quadrature. Thus, if the vectors obtained during the process of recursive iteration of Chebyshev filtering procedure are sparse, the computational complexity of the relevant matrix-vector multiplications scales as $\mathcal{O}(M)$.

In practice, we exploit the locality of finite-element basis to construct the elemental-matrices corresponding to $\mathbf{H}^{\text{loc}}$, i.e. the individual contributions from each finite-element to $\mathbf{H}^{\text{loc}}$, without explicitly assembling the global matrix. We note that building these elemental-matrices scales as $\mathcal{O}(M)$. In the case of pseudopotential calculations, we compute $C_{lm,j}^J$ at the finite-element level only for those elements in the compact support of ΔV_l^J. Further, the vectors obtained during the course of the recursive iteration (48) are truncated using a truncation tolerance δ_c, below which the values are set to zero. The matrix-vector multiplications in the Chebyshev filtering procedure are performed at the finite-element level only if the vector has a non-zero value in the finite-element considered. This leads to the computation of only the non-trivial elemental matrix-vector products, thereby rendering the computational complexity of the Chebyshev filtering procedure to $\mathcal{O}(M)$. The resulting element-level vectors are then assembled to form the global vectors $\mathbf{Y}$. We choose δ_c to be in the range of 10^{-4} to 10^{-14} in our subsequent numerical simulations (cf. section V) with looser tolerances being employed in the intial SCF iterations, where the solution is far away from the ground-state solution, and, adaptively employing tighter tolerances in later iterations as the SCF approaches convergence.

We next discuss the localization procedure employed to construct the localized wavefunctions spanning the Chebyshev filtered subspace.

C. Localization and truncation

Various procedures have been employed in the literature to achieve linear-scaling (cf.[47,63] for a comprehensive review) for pseudopotential calculations, and the use of localized functions spanning the occupied eigenspace (or an approximation to this space) has been one of the ideas exploited in developing linear-scaling algorithms for materials systems with a band-gap. In this regard, we remark that Wannier functions[81–84] have played an important role. In particular, the maximally localized Wannier functions[85] have been effectively used as an orthogonal localized basis for Kohn-Sham DFT calculations, specifically in periodic systems. However, techniques employing non-orthogonal localized functions have also been pro-

posed[59,60,86,87], which have better localization properties than orthogonal functions. A discussion on general localization properties of bases spanning the eigenspace is provided in E et al.[88] and Ozolins et al.[89]. In the present work, we adopt the technique proposed in Garcia-Cervera et al.[62] to construct a localized basis for the subspace $\mathbb{V}^N$ spanned by Chebyshev filtered wavefunctions $\{\psi_1^f(\mathbf{x}), \psi_2^f(\mathbf{x}), \cdots, \psi_N^f(\mathbf{x})\}$. Here, the localized basis is obtained as

$$\arg \min_{\psi \in \mathbb{V}^N, ||\psi||=1} \int w(\mathbf{x}) |\psi(\mathbf{x})|^2 \, d\mathbf{x}, \qquad (49)$$

where $w(\mathbf{x}) \geq 0$ is chosen to be a smooth weighting function $|\mathbf{x} - \boldsymbol{b}_I|^{2p}$, with p being a positive integer and $\boldsymbol{b}_I$ denoting a localization center. Such a choice of $w(\mathbf{x})$ minimizes the spread of the wavefunctions from a localization center, similar in spirit to the construction of maximally localized Wannier functions[85]. In the present work, we choose p to be 1 and $\boldsymbol{b}_I$ to be an atom center $\mathbf{R}_I$. Let n_I denote the number of localized functions we desire to compute at every atom center $\mathbf{R}_I$. Also, letting

$$\psi(\mathbf{x}) = \sum_i \alpha_i \psi_i^f(\mathbf{x}) \ \in \ \mathbb{V}^N, \qquad (50)$$

the minimization problem (49) is equivalent to solving the following generalized eigenvalue problem for the smallest n_I eigenvalues

$$\mathbf{W}^I \boldsymbol{\alpha} = \lambda \mathbf{S} \boldsymbol{\alpha}, \qquad (51)$$

where

$$W_{ij}^I = \int |\mathbf{x} - \mathbf{R}_I|^2 \, \psi_i^f(\mathbf{x}) \, \psi_j^f(\mathbf{x}) \, d\mathbf{x} \quad i, j = 1 \cdots N \tag{52a}$$

$$S_{ij} = \int \psi_i^f(\mathbf{x}) \, \psi_j^f(\mathbf{x}) \, d\mathbf{x}. \tag{52b}$$

In the present work, we choose n_I to be equal to the number of occupied single-atom orbitals corresponding to the Ith atom. If $\sum_I n_I < N$, then we randomly pick some atoms to compute additional localized functions. We note that we can rewrite $\mathbf{W}^I$ in (52a) using matrix notation to be

$$\mathbf{W}^I = \mathbf{L}^T \mathbf{K}^I \mathbf{L}, \qquad (53)$$

where columns of the matrix $\mathbf{L}$ correspond to the nodal values of the wavefunctions $\{\psi_1^f(\mathbf{x}), \psi_2^f(\mathbf{x}), \cdots, \psi_N^f(\mathbf{x})\}$ obtained after Chebyshev filtering, and

$$K_{ij}^I = \int |\mathbf{x} - \mathbf{R}_I|^2 \, N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x}. \qquad (54)$$

We note that the computational complexity of computing the matrix $\mathbf{W}^I$ for an atom I scales as $\mathcal{O}(MN^2)$ as the wavefunctions $\{\psi_1^f, \psi_2^f, \cdots, \psi_N^f\}$ obtained after Cheby-

shev filtering need not be sparse. Hence the total computational complexity for N_a atoms ($N_a \propto N$) in a given materials system scales as $\mathcal{O}(MN^3)$. We propose the following procedure in order to reduce the computational complexity to $\mathcal{O}(M)$. We first expand $\mathbf{K}^I$ in terms of atom $I = 0$ as follows:

$$
\begin{aligned}
K_{ij}^I &= \int |\mathbf{x} - \mathbf{R}_0 + \mathbf{R}_0 - \mathbf{R}_I|^2 \, N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x} \\
&= \int \Big[|\mathbf{x} - \mathbf{R}_0|^2 \, N_i(\mathbf{x}) N_j(\mathbf{x}) + |\mathbf{R}_0 - \mathbf{R}_I|^2 \, N_i(\mathbf{x}) N_j(\mathbf{x}) \\
&\quad + 2 \left(\mathbf{x} - \mathbf{R}_0 \right) . \left(\mathbf{R}_0 - \mathbf{R}_I \right) N_i(\mathbf{x}) N_j(\mathbf{x}) \Big] \, d\mathbf{x} \\
&= K_{ij}^0 + |\mathbf{R}_0 - \mathbf{R}_I|^2 M_{ij} + 2 \left(R_{0x} - R_{Ix} \right) B_{ij}^x \\
&\quad + 2 \left(R_{0y} - R_{Iy} \right) B_{ij}^y + 2 \left(R_{0z} - R_{Iz} \right) B_{ij}^z
\end{aligned}
$$

where

$$K_{ij}^0 = \int |\mathbf{x} - \mathbf{R}_0|^2 \, N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x}, \qquad (55\text{a})$$

$$M_{ij} = \int N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x}, \qquad (55\text{b})$$

$$B_{ij}^x = \int (x - R_{0x}) \, N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x}, \qquad (55\text{c})$$

$$B_{ij}^y = \int (y - R_{0y}) \, N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x}, \qquad (55\text{d})$$

$$B_{ij}^z = \int (z - R_{0z}) \, N_i(\mathbf{x}) N_j(\mathbf{x}) \, d\mathbf{x}. \qquad (55\text{e})$$

We use Gauss-Lobatto-Legendre quadrature rules to evaluate each of the above integrals which renders the matrices in (55) diagonal. Further, the matrix $\mathbf{W}^I$ for any atom I is constructed as a linear combination of five matrices, and is given by

$$
\begin{aligned}
\mathbf{W}^I = {}& \mathbf{L}^T \mathbf{K}^0 \mathbf{L} + |\mathbf{R}_0 - \mathbf{R}_I|^2 \mathbf{L}^T \mathbf{M} \mathbf{L} + \\
& 2 \left(R_{0x} - R_{Ix} \right) \mathbf{L}^T \mathbf{B}^x \mathbf{L} + 2 \left(R_{0y} - R_{Iy} \right) \mathbf{L}^T \mathbf{B}^y \mathbf{L} + \\
& 2 \left(R_{0z} - R_{Iz} \right) \mathbf{L}^T \mathbf{B}^z \mathbf{L}.
\end{aligned}
\qquad (56)
$$

We note that the five matrices $\mathbf{L}^T \mathbf{K}^0 \mathbf{L}$, $\mathbf{L}^T \mathbf{M} \mathbf{L}$, $\mathbf{L}^T \mathbf{B}^x \mathbf{L}$, $\mathbf{L}^T \mathbf{B}^y \mathbf{L}$ and $\mathbf{L}^T \mathbf{B}^z \mathbf{L}$ are independent of I, and can be computed *a priori*. Further, in order to reduce the computational complexity of computing these five matrices, we introduce a truncation tolerance $\delta_w > \delta_c$ ($\delta_w \sim 10^{-4} - 10^{-8}$) to truncate the Chebyshev filtered wavefunctions used in the construction of matrices $\mathbf{L}$ and $\mathbf{S}$. We note that this truncation in the Chebyshev filtered wavefunctions is introduced only in the construction of $\mathbf{L}$ and $\mathbf{S}$, and not in other operations involving the Chebyshev filtered wavefunctions, in particular, the linear combination in (50). Introducing the truncation δ_w renders a sparse structure to both $\mathbf{L}$ and $\mathbf{S}$, and thus the computation of $\mathbf{W}^I$ for all the atoms $I = 1 \cdots N_a$ scales as $\mathcal{O}(M)$. We note that the use of GLL quadrature rules in the evaluation of matrix elements in (55), as well as, the use of truncation tolerance δ_w in the construction

of $\mathbf{L}$ and $\mathbf{S}$ introduces approximation errors in the solution of the eigenvectors $\boldsymbol{\alpha}$ in the generalized eigenvalue problem (51). However, we note that these approximations errors do not alter the space $\mathbb{V}^N$ spanned by the localized wavefunctions (cf. equation (50)), as the vector space remains invariant under any linear combination.

Using the eigenvectors $\boldsymbol{\alpha}$ from the solution of the eigenvalue problem in (51) for each atom I, and the linear combination in (50), the non-orthogonal localized wavefunctions are computed and are denoted by $\{\phi_1^L(\mathbf{x}), \phi_2^L(\mathbf{x}), \cdots, \phi_N^L(\mathbf{x})\}$ which span $\mathbb{V}^N$. In order to provide a compact support for these non-orthogonal localized wavefunctions, we introduce a truncation tolerance δ_l, where the nodal values of these functions that are below this tolerance are set to zero. We note that, upon truncating these wavefunctions, the space spanned by these functions is only an approximation to $\mathbb{V}^N$, where the approximation error in the subspace depends on the choice of δ_l. As in the case of δ_c, the truncation tolerance introduced during the Chebyshev filtering, δ_l is also chosen adaptively with looser tolerances in the initial SCF iterations and progressively employing tighter tolerances as the SCF approaches convergence (see section V for details). We denote by $\boldsymbol{\Phi}_L$ the matrix whose column vectors are the expansion coefficients of these compactly-supported non-orthogonal localized wavefunctions expressed in the Löwdin orthonormalized finite-element basis. We note that the locality of the non-orthogonal wavefunctions $\phi_i^L(\mathbf{x}) : 1 \leq i \leq N$ renders sparsity to the matrix $\boldsymbol{\Phi}_L$.

D. Subspace projection in the non-orthogonal basis

We now discuss the steps involved in the subspace projection of the Hamiltonian into the non-orthogonal localized basis represented by $\boldsymbol{\Phi}_L$.

a. Computation of overlap matrix: The overlap matrix $\mathbf{S}$ resulting from the non-orthogonal localized wavefunctions is given by

$$\mathbf{S} = \boldsymbol{\Phi}_L^T \boldsymbol{\Phi}_L. \tag{57}$$

We note that the computational complexity of evaluating $\mathbf{S}$ scales as $\mathcal{O}(N)$ if $\boldsymbol{\Phi}_L$ is a sparse matrix.

b. Computation of projected Hamiltonian: We recall from the discussion in section III that the projection of the Hamiltonian into a non-orthogonal localized basis is given by

$$\mathbf{H}^\phi = \mathbf{S}^{-1} \boldsymbol{\Phi}_L^T \tilde{\mathbf{H}} \boldsymbol{\Phi}_L. \tag{58}$$

The above equation involves the inverse of the overlap matrix $\mathbf{S}^{-1}$, which can be evaluated using scaled third-order Newton-Schulz iteration[62,90]. If $\mathbf{S}$ and $\mathbf{S}^{-1}$ are exponentially localized, $\mathbf{S}^{-1}$ can be computed in $\mathcal{O}(N)$ complexity[91]. Since the discretized Hamiltonian $\tilde{\mathbf{H}}$ is sparse, and if the matrix $\boldsymbol{\Phi}_L$ is sparse with a bandwidth independent of N, $\mathbf{H}^\phi$ can be computed in $\mathcal{O}(N)$ com-

plexity.

E. Electron density computation

The Fermi-operator expansion techniques[47,63] have been widely adopted to avoid explicit diagonalization of the discretized Hamiltonian in order to compute the electron density. One of the widely used Fermi-operator expansion technique[52,53], which works for both insulating as well as metallic systems, approximates the Fermi distribution (cf. equation (4)) by means of a Chebyshev polynomial expansion. The accuracy of such an expansion depends on the smearing parameter in the Fermi distribution, σ, and the width of the eigenspectrum (spectral width) of discretized Hamiltonian denoted by ΔE. In particular, the degree of polynomial required to achieve a desired accuracy in the approximation[54] of the Fermi distribution is $\mathcal{O}(\frac{\Delta E}{\sigma})$, and is $\mathcal{O}(\sqrt{\Delta E})$ for a given σ and the occupied spectrum[64]. We also note that numerous recent efforts have focused on developing alternate approximations to the Fermi distribution[65–69] with reduced computational complexity.

One of the major challenges in employing the Fermi-operator expansion technique on a finite-element discretized Hamiltonian is the large spectral width of the Hamiltonian—$\mathcal{O}(10^3)$ Ha for pseudopotential calculations and $\mathcal{O}(10^5)$ Ha for all-electron calculations[25]—that deteriorates the accuracy of the Fermi-operator expansion. To circumvent this, we employ the Fermi-operator expansion in terms of the subspace projected Hamiltonian $\mathbf{H}^\phi$ whose spectral width is commensurate with that of the occupied eigenspectrum. We recall from equation (33) that the electron density in terms of the density matrix is given by

$$\rho(\mathbf{x}) = 2 \sum_{i,j=1}^{M} \Gamma_{ij} q_i(\mathbf{x}) q_j(\mathbf{x}). \tag{59}$$

Using equations (45) and (32) in the above equation, we have

$$\begin{aligned} \rho(\mathbf{x}) &= 2\,\boldsymbol{N}^T(\mathbf{x})\,\mathbf{M}^{-1/2}\,\boldsymbol{\Phi}_L\,f(\mathbf{H}^\phi)\,\boldsymbol{\Phi}_L^\dagger\,\mathbf{M}^{-1/2}\boldsymbol{N}(\mathbf{x}) \\ &= 2\,\boldsymbol{N}^T(\mathbf{x})\,\mathbf{M}^{-1/2}\,\boldsymbol{\Phi}_L\,f(\mathbf{H}^\phi)\,\mathbf{S}^{-1}\,\boldsymbol{\Phi}_L^T\mathbf{M}^{-1/2}\boldsymbol{N}(\mathbf{x}), \end{aligned} \tag{60}$$

where $\boldsymbol{N}^T(\mathbf{x}) = [N_1(\mathbf{x}), N_2(\mathbf{x}), N_3(\mathbf{x}), \cdots, N_M(\mathbf{x})]$ and

$$f(\mathbf{H}^\phi) = \frac{1}{1 + \exp\left(\frac{\mathbf{H}^\phi - \mu}{\sigma}\right)} \tag{61}$$

with μ being the chemical potential and $\sigma = k_B T$. We note that as the self-consistent field iteration converges, $f(\mathbf{H}^\phi)$ represents the finite-temperature density matrix expressed in the non-orthogonal localized basis. We remark that, for a jellium approximation (a simplified representation for a metallic system), it was shown that the finite-temperature density operator exhibits an ex-

ponential decay in real-space[92,93]. However, this remains an open question beyond the jellium approximation for metallic systems. For an insulating system with a band-gap, the density operator, even at zero temperature, has an exponential decay in real-space[47]. Assuming that the density operator decays in real-space, and recalling that $\mathbf{\Phi}_L$ represents a localized basis, we note that $f(\mathbf{H}^\phi)$ will have a sparse structure with the extent of sparsity depending on the decay properties of the density operator.

We use Chebyshev polynomial expansion to approximate $f(\mathbf{H}^\phi)$, and compute the electron density. To this end, we begin by scaling and shifting $\mathbf{H}^\phi$ to obtain $\mathbf{H}_s^\phi$ such that its spectrum lies in $[-1,1]$ and then $f(\mathbf{H}^\phi)$ is approximated using a finite number of Chebyshev polynomials[54] as

$$f(\mathbf{H}^\phi) = \sum_{n=0}^{R} a_n(\sigma_s, \mu_s) T_n(\mathbf{H}_s^\phi), \qquad (62)$$

where

$$\mathbf{H}_s^\phi = \frac{\mathbf{H}^\phi - \bar{\epsilon}}{\Delta\epsilon} \quad ; \quad \sigma_s = \frac{\Delta\epsilon}{\sigma} \quad ; \quad \mu_s = \frac{\mu - \bar{\epsilon}}{\Delta\epsilon}, \qquad (63)$$

$$\Delta\epsilon = \frac{\epsilon_{max} - \epsilon_{min}}{2} \quad ; \quad \bar{\epsilon} = \frac{\epsilon_{max} + \epsilon_{min}}{2}, \qquad (64)$$

and

$$a_n(\sigma_s, \mu_s) = \frac{2 - \delta_{n\,0}}{\pi} \int_{-1}^{1} \frac{T_n(x)}{\sqrt{1-x^2}} \frac{1}{1 + e^{\sigma_s(x - \mu_s)}} dx, \qquad (65)$$

where δ_{ij} denotes the Kronecker delta. In the above, ϵ_{max} and ϵ_{min} denote the upper and lower bounds for the spectrum of $\mathbf{H}^\phi$. Estimates for ϵ_{max} and ϵ_{min} are computed using the Krylov-Schur method[94]. As $\mathbf{H}^\phi$ is the projection of the Hamiltonian into a localized basis, $\mathbf{H}^\phi$ is a sparse matrix, and these estimates for the spectral width can be computed in $\mathcal{O}(N)$ complexity. We also remark that if $\mathbf{H}^\phi$ is sufficiently sparse, $f(\mathbf{H}^\phi)$ can be computed in $\mathcal{O}(N)$ complexity[54]. Further, we note that the degree R of the Chebyshev expansion in equation (62) is proportional to the spectral width $\Delta E = \epsilon_{max} - \epsilon_{min}$ of $\mathbf{H}^\phi$. As discussed earlier, since $\mathbf{H}^\phi$ is the projected Hamiltonian in the space containing the occupied eigenstates and only a few unoccupied eigenstates, the spectral width of $\mathbf{H}^\phi$ is $\mathcal{O}(1\,Ha)$ for pseudopotential calculations and $\mathcal{O}(10\,Ha)$ for all-electron calculations for low atomic numbers. Thus, the Fermi-operator expansion can be computed efficiently and accurately using a Chebyshev polynomial expansion of $\mathcal{O}(100)$ for pseudopotential calculations and $\mathcal{O}(1000)$ for all-electron calculations for moderate temperatures ($\sim 500K$) used in the smearing parameter.

The Fermi-energy (μ), which is required in the computation of the Fermi-operator expansion for $f(\mathbf{H}^\phi)$ and the electron density, is evaluated using the constraint in (35)

$$2\,\mathrm{tr}\left(f(\mathbf{H}^\phi)\right) = N_e, \qquad (66)$$

where N_e is the number of electrons in the given system. The above equation is solved using the Newton-Raphson method[95] and an initial guess to the Newton-Raphson method is computed using the bisection algorithm[95]. It is evident from equation (62) that the dependence of the expansion on the Fermi-energy is only in the coefficients of the expansion. Exploiting this fact, the Fermi-energy can be efficiently computed using methods described in Baer et al.[54], which scales as $\mathcal{O}(N)$. To this end, the mth column of $f(\mathbf{H}^\phi)$ is obtained by the application of the expansion in (62) on a unit column vector $\boldsymbol{v}^m$ containing all zeros except at the mth position. Hence,

$$\left[f(\mathbf{H}^\phi)\right]_m = \sum_{n=0}^{R} a_n(\sigma_s, \mu_s) \boldsymbol{v}_n^m, \qquad (67)$$

where $\boldsymbol{v}_n^m = T_n(\mathbf{H}_s^\phi)\boldsymbol{v}^m$ denotes the mth column of $T_n(\mathbf{H}_s^\phi)$. We note that $\boldsymbol{v}_n^m$ can be computed efficiently using the Chebyshev polynomial recursion given by

$$v_0^m = v^m$$
$$v_1^m = \mathbf{H}^\phi v^m$$
$$v_{n+1}^m = 2\,\mathbf{H}^\phi v_n^m - v_{n-1}^m.$$

As the vectors $\boldsymbol{v}_n^m$ can be precomputed and stored, the evaluation of $\mathrm{tr}\left(f(\mathbf{H}^\phi)\right)$ for every trial Fermi-energy results in a trivial computational cost.

Upon evaluating the Fermi energy, the band energy (E_b) can also be expressed in terms of $\mathbf{H}^\phi$ as

$$E_b = 2\,\mathrm{tr}\left(f(\mathbf{H}^\phi)\mathbf{H}^\phi\right). \qquad (68)$$

Finally, we note that the computational complexity of computing the electron density at a quadrature point $\mathbf{x}$ is independent of the system size. This can be seen by rewriting the expression (60) as

$$\rho(\mathbf{x}) = 2\,\mathbf{r}^T(\mathbf{x})\,f(\mathbf{H}^\phi)\,\mathbf{S}^{-1}\,\mathbf{r}(\mathbf{x}), \qquad (69)$$

where $\mathbf{r}(\mathbf{x}) = \mathbf{\Phi}_L^T \mathbf{M}^{-1/2} \boldsymbol{N}(\mathbf{x})$ is a sparse vector if $\mathbf{\Phi}_L$ is sparse. As $f(\mathbf{H}^\phi)\mathbf{S}^{-1}$ can be pre-computed *a priori*, the computation of $\rho(\mathbf{x})$ is independent of the system size. Since the total number of quadrature points scales linearly with system size, computation of electron density for a given material system is $\mathcal{O}(M)$ complexity.

V. RESULTS AND DISCUSSION

In this section, we investigate the accuracy, performance, and scaling of the proposed subspace projec-

tion technique. As benchmark systems, we consider non-periodic three dimensional systems involving metallic, insulating and semi-conducting materials systems. The benchmark metallic systems chosen for this study include aluminum nano-clusters of varying sizes, containing $3 \times 3 \times 3$ (172 atoms), $5 \times 5 \times 5$ (666 atoms), $7 \times 7 \times 7$ (1688 atoms) and $9 \times 9 \times 9$ (3430 atoms) face-centered-cubic (fcc) unit-cells. The benchmark insulating systems chosen for this study include alkane chains of varying lengths, containing 101, 302, 902, 2702 and 7058 atoms. Silicon nano-clusters containing $1 \times 1 \times 1$ (252 electrons), $2 \times 1 \times 1$ (434 electrons), $2 \times 2 \times 1$ (756 electrons), $2 \times 2 \times 2$ (1330 electrons) and $3 \times 3 \times 3$ (3920 electrons) diamond-cubic (dia) unit-cells are chosen for the benchmark semi-conducting materials systems. Norm-conserving Troullier-Martins pseudopotentials[76] have been employed in the case of aluminum nano-clusters and alkane chains, whereas all-electron calculations have been performed in the case of silicon nano-clusters. In all our simulations, we use the n-stage Anderson[96] mixing scheme on the electron density in self-consistent field iteration of the Kohn-Sham problem. Further, all the numerical simulations reported in this section are conducted using a parallel implementation of the code based on MPI, and are executed on a parallel computing cluster with the following specifications: dual-socket eight-core Intel Core Sandybridge CPU nodes with 16 processors (cores) per node, 64 GB memory per node, and 40 Gbps Infiniband networking between all nodes for fast MPI communications.

As discussed in section IV, the linear-scaling of the proposed subspace projection technique relies on the locality of the finite-element basis ($q_i : 1 \leq i \leq M$) as well as the localized wavefunctions ($\phi_j^L : 1 \leq j \leq N$), and subsequently, the sparsity of the various matrices involved in our formulation ($\tilde{\mathbf{H}}, \mathbf{H}^\phi, \mathbf{\Phi}_L, \mathbf{S}, \mathbf{W}^I$). We recall that the compact support of localized wavefunctions and the sparsity in various matrices is achieved by introducing truncation tolerances δ_c (in Chebyshev filtering) and δ_l (for localized wavefunctions). Further, as mentioned in section 4, and elaborated subsequently, the truncation tolerances are chosen adaptively with looser tolerances being employed in the initial SCF iterations and progressively tightening these during the course of the SCF convergence. As demonstrated in our benchmark studies, such an adaptive tolerance provides significant computational efficiency while retaining the accuracy of the solution to the Kohn-Sham DFT problem. We note that the sparsity of various matrices is governed by the eigenspectrum of discrete Hamiltonian as well as the values of the truncation tolerances employed, which changes with each SCF iteration. In the present implementation of the subspace projection technique, we employ efficient parallel data-structures provided by PETSc package[97], where necessary, to represent various matrices and perform arithmetic operations between them. We observe in our simulations that the operations involving sparse data-structures provided by PETSc are efficient only when the

fraction of non-zero entries (density fraction) in the matrix is $< 1 - 2\%$. Beyond this density-fraction, the overhead cost of using a sparse data-structure is prohibitively expensive, and it is more efficient to use dense data-structures. To maximize the computational efficiency, in the present work, we employ sparse data-structures only when the density fraction is $< 2\%$, beyond which we switch to dense data-structures. This is the main source of deviation from linear-scaling, in practice, in the present implementation.

In order to assess the accuracy, performance and scaling of the proposed approach, we use as reference the recently developed Chebyshev filtered subspace iteration for the finite-element basis[25] (ChFSI-FE). The ChFSI-FE involves the projection of the Hamiltonian into an orthogonal basis spanning the Chebyshev filtered space, and an explicit computation of the eigenvalues and eigenvectors of the projected Hamiltonian to estimate the electron density. It was demonstrated[25] that ChFSI-FE technique with the use of higher-order finite-element discretization presents a computationally efficient real-space approach for Kohn-Sham DFT calculations, which can handle both pseudopotential and all-electron calculations. We also note that the accuracy of ChFSI-FE was ascertained, on benchmark problems of varying sizes, using ABINIT software[6] for pseudopotential calculations and GAUSSIAN software[98] for all-electron calculations. In the present work, we use ChFSI-FE as our reference to assess the accuracy and performance of the proposed approach.

A. Aluminum nano-clusters: Pseudopotential study

We consider aluminum nano-clusters formed from fcc unit cells with lattice spacing of 7.45 $a.u.$. The norm-conserving Troullier-Martins pseudopotential in the Kleinman-Bylander form[76,77] is employed in these simulations. We consider the $3s$ and $3p$ components to compute the projector, while the $3d$ component is chosen to be the local component of the pseudopotential. The pseudopotentials are generated using the fhi98pp[99] software and the default cut-off radii are used for $3s$, $3p$ and $3d$ components, which are 1.8 $a.u.$, 2.0 $a.u.$ and 2.15 $a.u.$, respectively. In order to choose finite-element meshes which provide a discretization error of less than 5 meV per atom for the various nano-clusters considered in the present study, we first obtain the converged ground-state energy for the aluminum cluster containing $3 \times 3 \times 3$ fcc unit cells which comprises of 172 atoms with 516 electrons. To this end, we use a sequence of increasingly refined fourth-order spectral finite-element (HEX125SPECT) meshes on a cubic simulation domain of side 400 $a.u.$ employing Dirichlet boundary conditions. Here, and subsequently, we use the *a priori* mesh adaption techniques developed in Motamarri et. al[25] to construct the finite-element meshes, and refer to this prior work for details. For these sequence of meshes,

the ground-state energy is computed using ChFSI-FE with a Chebyshev filter of degree 20 and employing a Fermi smearing parameter of 0.00158 Ha (T=500K). The computed discrete ground-state energies (E_h) for these meshes are tabulated in Table I, where h denotes a measure of the finite-element mesh-size. The extrapolation procedure proposed in Motamarri et. al[25] allows us to estimate the ground-state energy in the limit as $h \to 0$, denoted by E_0. Using this extrapolation procedure, we computed the ground-state energy per atom to be $E_0 = -56.6966935 \ eV$. In order to ascertain the accuracy of this extrapolated ground-state energy, we conducted a plane wave simulation with ABINIT using a cell-size of 80 $a.u.$ and an energy cut-off of 30 Ha with one k-point (the most refined calculation possible within our memory limitations). In the case of ABINIT, the ground-state energy per atom was found to be $-56.6966719 \ eV$, which only differs from E_0 by 0.02 meV.

TABLE I: Convergence of the finite-element discretization (HEX125SPECT element) for a $3 \times 3 \times 3$ fcc aluminum cluster.

Degrees of freedom (DoF)	Energy per atom (eV)	Relative error $\left\lvert \frac{E_h - E_0}{E_0} \right\rvert$
$222,553$	-55.5622017	2.0×10^{-2}
$1,760,305$	-56.6677104	5.1×10^{-4}
$14,003,809$	-56.6966331	1.06×10^{-6}

We next choose a finite-element mesh with fifth-order spectral finite-elements (HEX216SPECT) for the $3 \times 3 \times 3$ aluminum nano-cluster, where the mesh-size is chosen such that the discretization error, measured with respect to E_0, is less than 5 meV per atom. Using the same characteristic mesh-size, which is expected to result in a similar discretization error, we construct the finite-element meshes for the varying sizes of aluminum clusters containing $5 \times 5 \times 5$ (666 atoms), $7 \times 7 \times 7$ (1688 atoms) and $9 \times 9 \times 9$ (3430 atoms) fcc unit-cells. Using a Chebyshev filter of degree 20 and a Fermi smearing parameter of 0.00158 Ha (T=500K), we compute the reference ground-state energies using ChFSI-FE. These are tabulated in Table II.

We next compute the ground-state energies using the proposed subspace projection algorithm (cf. section IV) with identical meshes and parameters (Chebyshev filter degree, Fermi smearing parameter) employed in our reference calculations. The polynomial degree R used in the Chebyshev expansion of the Fermi-function of the projected Hamiltonian (62) is chosen to be 400. We recall that we use truncation tolerances (δ_c, δ_l) to achieve compact support for the non-orthogonal localized wavefunctions ($\mathbf{\Phi}_L$) and sparsity in the various matrices involved in our formulation. In the initial SCF iterations we use looser truncation tolerances, and employ tighter

tolerances as the self-consistent iteration proceeds towards convergence. In particular, we choose $\delta_c = \delta_l$ and vary the truncation tolerance as a function of the relative change in the ground-state energy between two successive SCF iterations (δE_r). The specific choice of the truncation tolerance employed in the present study is

$$\delta_c = \begin{cases} C_1 & \text{if } \delta E_r \geq 1, \\ C_1 (\delta E_r)^p & \text{if } 10^{-1.5} \leq \delta E_r < 1, \\ C_2 (\delta E_r)^q & \text{otherwise}. \end{cases} \quad (70)$$

Here, we use a truncation tolerance of $C_1 = 10^{-4}$ when δE_r is greater than 1, then use a power law (cf. equation (70)), subsequently, with $p = \frac{2}{3}$ and $q = \frac{3}{2}$. In order to ensure continuity in δ_c, C_2 is chosen to be 0.00178. We note that the choice of this truncation tolerance is arbitrary, however, this form of the truncation tolerance has provided robust convergence of the SCF for all benchmark calculations, where the number of SCF iterations using the subspace projection approach is only marginally ($\sim 4 - 5$ iterations) greater than those using ChFSI-FE. In all our simulations, using the proposed method as well as ChFSI-FE, the SCF is terminated when $\delta E_r < 10^{-7}$ for three successive iterations.

TABLE II: Ground-state energies per atom (eV) for the various sizes of aluminum nano-clusters computed using the proposed subspace projection algorithm technique and ChFSI-FE[25].

Cluster	DoF	Proposed Method	ChFSI-FE
3x3x3	$1,107,471$	-56.694963	-56.6949697
5x5x5	$4,363,621$	-56.876491	-56.876518
7x7x7	$11,085,371$	-56.959021	-56.9623511
9x9x9	$22,520,721$	-57.010587	-57.0145334

The ground-state energies computed using the proposed subspace projection algorithm for the range of nano-clusters considered in this work are tabulated in Table II. The above results show that the proposed subspace projection technique provides good accuracies in the ground-state energies, where the computed energies are within 5 meV per atom of the reference energies computed using ChFSI-FE. The computational times for the full SCF convergence for the range of nano-clusters are plotted in figure 1 for the proposed method, as well as, the reference calculations using ChFSI-FE. In this plot, the computational time denoted on the Y-axis is the total computational CPU time in hours (CPU time = Number of cores $\times$ wall-clock time in hours). These results show that the proposed approach is computationally efficient, compared to ChFSI-FE, for system sizes beyond 172 atoms. Further, we note that the subspace projection technique provides a factor of ~ 8.5 speedup for the
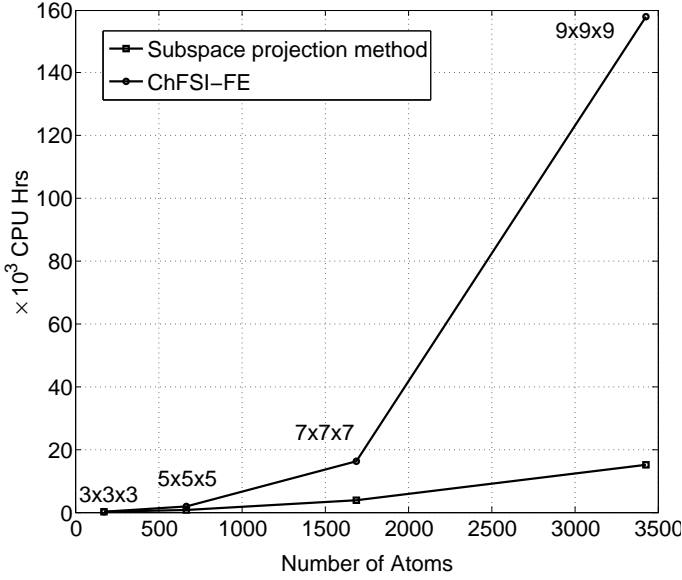
FIG. 1: Total computational times for the proposed method and ChFSI-FE. Case study: Aluminum nano-clusters.
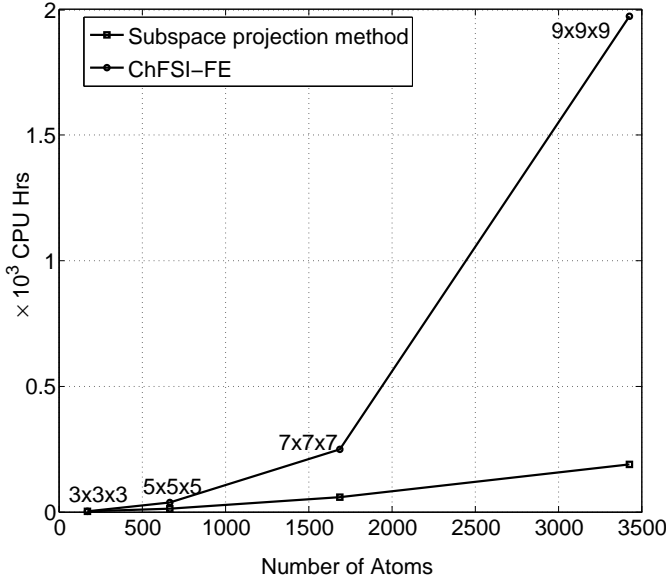

FIG. 2: Average computational times per SCF iteration for the proposed method and ChFSI-FE. Case study: Aluminum nano-clusters.
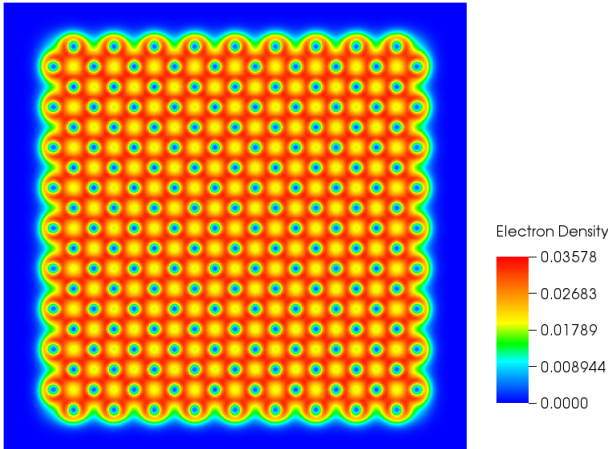

FIG. 3: Electron density contours on the mid-plane of $9 \times 9 \times 9$ fcc aluminum nano-cluster.

nano-cluster containing 3430 atoms. Using these computational times, we estimated the scaling of the proposed approach and our reference calculations using ChFSI-FE as a function of system size (number of atoms). The scaling for the proposed approach is found to be approximately $\mathcal{O}(N^{1.66})$, while the scaling in the case of ChFSI-FE is $\mathcal{O}(N^{2.37})$. The deviation from linear-scaling of the proposed method, in practice, is primarily due to two factors. Firstly, we observe that the number of SCF iterations increase with increasing system size. For instance, the number of SCF iterations increase from 45 SCF iterations for the $3 \times 3 \times 3$ nano-cluster to 80 iterations for the $9 \times 9 \times 9$ nano-cluster. This issue of increasing SCF iterations with system size can potentially be mitigated using improved mixing schemes[100]. Figure 2 shows the computational time per SCF iteration, and the scaling with system size is found to be $\mathcal{O}(N^{1.46})$ for the proposed subspace projection method and $\mathcal{O}(N^{2.17})$ for ChFSI-FE. The second factor which results in the deviation of the proposed subspace projection approach from being linear-scaling is due to the use of dense data-structures when the density fraction is above 2%. Although the use of dense data-structure affects the overall scaling, it is still computationally efficient in comparison to using sparse data-structures for density fractions greater than 2%. We refer to the Appendix for the scaling of various components of the subspace projection algorithm and a discussion. Figure 3 shows the electron density contours on the mid-plane of $9 \times 9 \times 9$ nano-cluster computed using the proposed subspace projection technique.

B. Alkane chains: Pseudopotential study

We next consider three-dimensional alkane chains with individual repeating units of CH_2 with C-C and C-H bond lengths to be 2.91018 a.u. and 2.0598 a.u., respectively. The H-C-H and C-C-C bond angles are taken to be 109.47^0. Norm-conserving Troullier-Martins pseudopotential in the Kleinman-Bylander form[76,77] has been employed in these simulations. In the case of carbon, we consider the 2s component to compute the projector in the Kleinman-Bylander form, while the 2p component is chosen to be the local component of the pseudopotential. We consider the local pseudopotential for hydrogen corresponding to the 1s component. The pseudopotentials are generated using the software package fhi98PP[99] using the default cut-off radii, which is 1.5 a.u. for both 2s and 2p components of carbon and 1.3 a.u. for the 1s component of hydrogen. In order to choose finite-element meshes with discretization errors less than 5 meV per atom for various alkane chains, we first consider the case of $C_{33}H_{68}$ containing 101 atoms with 200 valence electrons and obtain the converged value of the ground-state energy. To this end, as in the case of aluminum nano-clusters (cf. section V A), we use a sequence of increasingly refined fourth-order spectral finite-elements (HEX125SPECT) on a cuboidal simulation do-
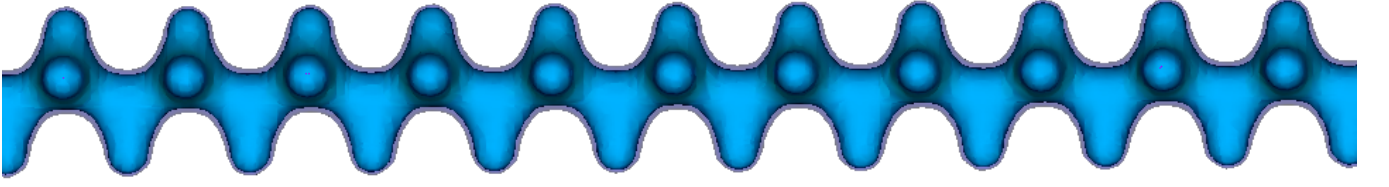
FIG. 4: Electron density isocontours of $C_{900}H_{1802}$.

main of dimensions 100 $a.u. \times$ 100 $a.u. \times$ 220 $a.u..$ For these sequence of meshes, the ground-state energy E_h is computed using ChFSI-FE with a Chebyshev filter of degree 35 and employing a Fermi smearing parameter of 0.00158 Ha (T=500K), and these results are tabulated in Table III. Using the extrapolation procedure[25], we find the reference ground-state energy per atom to be $E_0 = -61.44173873$ eV. Further, we also compared our reference ground-state energy with ABINIT using a cuboidal simulation of size 40 $a.u. \times$ 40 $a.u. \times$ 160 $a.u.$ and an energy cut-off of 30 Ha with one k-point (the most refined calculation possible within our memory limitations). In the case of ABINIT, the ground-state energy per atom was found to -61.44219366 eV, which differs from E_0 by ~ 0.4 meV.

TABLE III: Convergence of the finite-element discretization for $C_{33}H_{68}$ using HEX125SPECT element.

Deg. of freedom	Energy per atom (eV)	Relative error
$391,893$	-61.31623538	2.0×10^{-3}
$3,096,585$	-61.43711471	7.5×10^{-5}
$24,621,969$	-61.44173469	5.01×10^{-8}

We next choose a finite-element mesh with fifth-order spectral finite-elements (HEX216SPECT) for $C_{33}H_{68}$, where the mesh-size is chosen such that the discretization error, measured with respect to E_0, is less than 5 meV per atom. Using the same characteristic mesh-size which is expected to result in a similar discretization error, we construct finite-element meshes for varying lengths of alkane chains, namely $C_{100}H_{202}$ (302 atoms), $C_{300}H_{602}$ (902 atoms), $C_{900}H_{1802}$ (2702 atoms) and $C_{2350}H_{4702}$ (7052 atoms). Using a Chebyshev filter of degree 35 and a Fermi smearing parameter of 0.00158 Ha (T=500K), we compute the reference ground-state energies using ChFSI-FE for these systems, and are tabulated in Table IV.

The subspace projection algorithm is then used to compute the ground-state energies using identical meshes and parameters (Chebyshev filter degree, Fermi smearing pa-

rameter) employed in our reference calculations. The polynomial degree R used in the Chebyshev expansion of the Fermi function of the projected Hamiltonian (62) is chosen to be 400. Just as in the case of aluminum nano-clusters, we use adaptive truncation tolerances (δ_c, δ_l). In particular, we choose $\delta_c = \delta_l$ and use the specific choice of truncation tolerance given in equation (70) with identical values of C_1, p and q as employed in the case of our previous benchmark calculations involving aluminum nano-clusters. In all our simulations, using the proposed method as well as ChFSI-FE, the SCF is terminated when $\delta E_r < 10^{-7}$ for three successive iterations.

TABLE IV: Comparison of ground-state energies for various alkane chains.

Alkane Chain	DoF	Proposed Method	ChFSI-FE
C_{33} H_{68}	$870,656$	-61.438671	-61.438680
C_{100} H_{202}	$2,491,616$	-62.041530	-62.041532
C_{300} H_{602}	$7,354,496$	-62.240148	-62.240277
C_{900} H_{1802}	$21,943,138$	-62.303101	-62.303608

The ground-state energies computed using the proposed subspace projection algorithm for different lengths of alkane chains considered in this work are tabulated in Table IV. These results indicate that the proposed method provides good accuracies in the ground-state energies, where the computed energies are within 1 meV per atom of the reference energies computed using ChFSI-FE. The computational times for the full SCF convergence for varying lengths of alkane chains are plotted in figure 5 for the proposed method, as well as, the reference calculations using ChFSI-FE. These results indicate that the proposed approach is computationally efficient, compared to ChFSI-FE, for system sizes beyond 101 atoms and provides a factor of ~ 8 speedup for the alkane chain containing 2702 atoms. Using these computational times, the estimated scaling for the proposed approach is found to be approximately $\mathcal{O}(N^{1.33})$, while scaling in the case of ChFSI-FE is $\mathcal{O}(N^{2.13})$. The av-

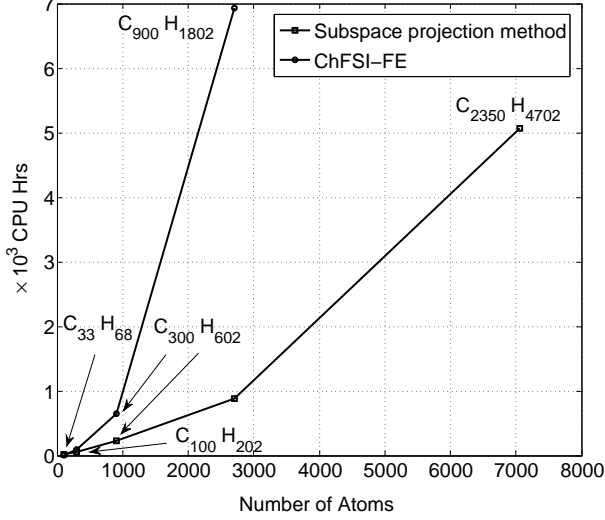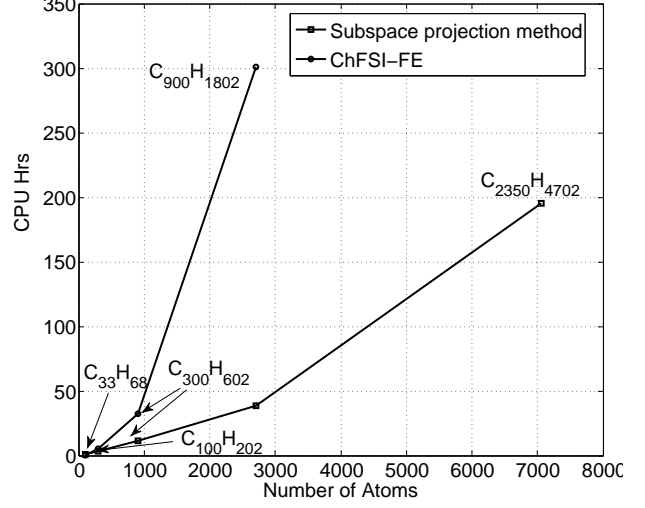FIG. 5: Total computational times for the proposed method and ChFSI-FE. Case study: Alkane chains.

FIG. 6: Average computational times per SCF iteration for the proposed method and ChFSI-FE. Case study: Alkane chains.

erage computational time per SCF iteration is shown in figure 6, and the scaling with system size is found to be $\mathcal{O}(N^{1.18})$ for the proposed subspace projection method and $\mathcal{O}(N^{1.98})$ for ChFSI-FE. We refer to the Appendix for the scaling of various components of the subspace projection algorithm and a discussion. We note that the proposed subspace projection technique exhibits better scaling behavior for the alkane chains, which is an insulating system, in comparison to the aluminum nano-clusters, which is a metallic system. This is due to the better localization of the wavefunctions in the insulating system in comparison to the metallic system (cf. Appendix). Figure 4 shows the isocontours of electron density of the alkane chain containing 2702 atoms.

C. Silicon nano-clusters: All-electron study

We consider silicon nano-clusters comprising of diamond-cubic unit cells with a lattice constant of 10.26 $a.u.$ and conduct all-electron calculations to test the performance of the subspace projection method. In order to choose finite-element meshes which provide a discretization error of less than 5 mHa per atom for various nano-clusters, we first obtain a converged value of the ground state energy by conducting a very refined simulation using the GAUSSIAN package[98] on silicon nano-cluster containing $1 \times 1 \times 1$ unit-cell which comprises of 18 atoms with 252 electrons. To this end, we employ the polarization consistent DFT basis sets (pc-n) and introduce them as an external basis set in GAUSSIAN pack-

age. Using the most refined pc-4 basis set, the computed ground-state energy per atom (E_0) is -288.3179669 Ha.

We next chose a finite-element mesh with fifth-order spectral finite-elements (HEX216SPECT) for the $1 \times 1 \times 1$ silicon nano-cluster, where the discretization error, measured with respect to E_0, is less than 3 mHa. Using the same characteristic mesh-size, we construct the finite-element meshes for the varying sizes of silicon nano-clusters containing $1 \times 1 \times 1$ (252 electrons), $2 \times 1 \times 1$ (434 electrons), $2 \times 2 \times 1$ (756 electrons), $2 \times 2 \times 2$ (1330 electrons) and $3 \times 3 \times 3$ (3920 electrons) diamond-cubic unit cells. The finite-element mesh is locally refined near the nuclei since all-electron calculations involve highly oscillatory wavefunctions near the nuclei. The order of the Chebyshev filter thus required in these simulations is around 1000 to effectively filter out the unwanted spectrum. Using a Chebyshev filter of degree 1000, and a Fermi smearing parameter of 0.00475 Ha (T=1500K), we compute the reference ground-state energies using ChFSI-FE which are tabulated in Table V.

The subspace projection algorithm is then used to compute the ground-state energies with identical meshes and parameters employed in our reference calculations. The polynomial degree R used in the Chebyshev expansion of the Fermi-operator of the projected Hamiltonian (62) is chosen to be 1250 since the width of the occupied spectrum is larger than that of a pseudopotential calculation. Just as in the case of the previous benchmark calculations, we use adaptive truncation tolerances (δ_c, δ_l) with $\delta_l = 10^4 \delta_c$ and vary δ_c as a function of relative change
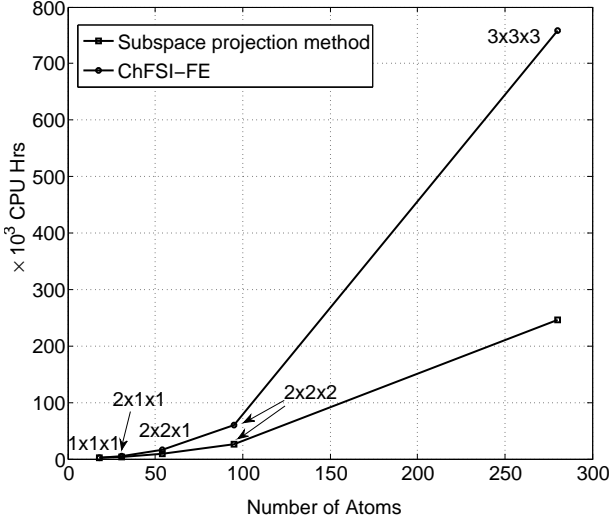
FIG. 7: Total computational times for the proposed method
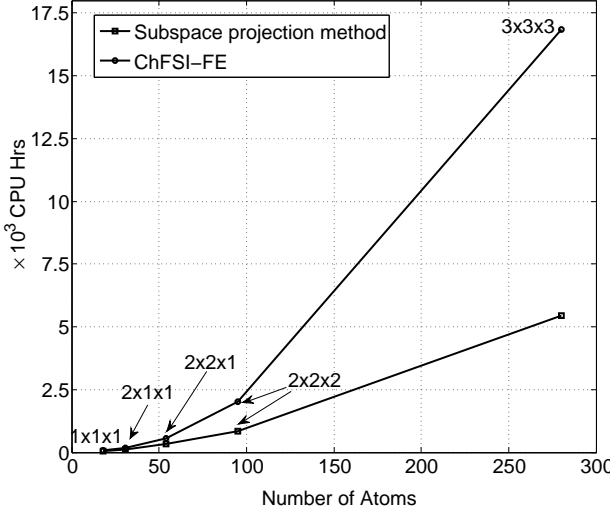and ChFSI–FE. Case study: Silicon nano-clusters.

FIG. 8: Average computational times per SCF iteration for
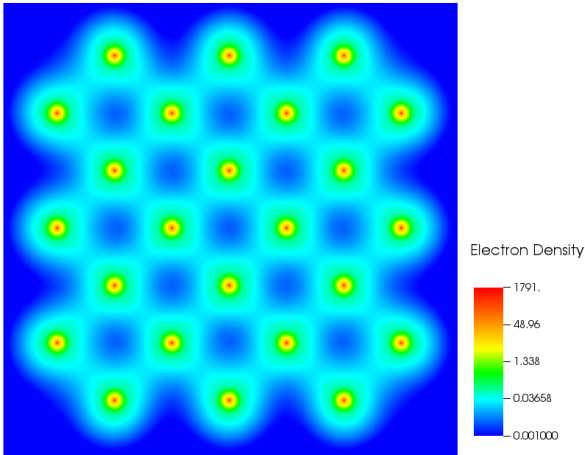the proposed method and ChFSI–FE. Case study: Silicon
nano-clusters.

FIG. 9: Electron density contours on the mid-plane of
$3 \times 3 \times 3$ silicon nano-cluster.

in the ground-state energy between two successive itera-
tions (δE_r). We use the specific choice of truncation tol-
erance given in equation (70) with $C_1 = 10^{-11}$, $p = 2/3$
and $q = 3/2$. We note that much tighter tolerances have
been used for δ_c in comparison to pseudopotential calcu-
lations in order to control the accumulation of truncation
errors in applying a very high degree Chebyshev filter. In
all our simulations, using the proposed method as well as
ChFSI-FE, the SCF is terminated when $\delta E_r < 10^{-6}$ for
three successive iterations.

TABLE V: Comparison of ground-state energies for various
sizes of silicon clusters.

Cluster	DoF	Proposed Method	ChFSI-FE
1x1x1	$5,136,901$	-288.32016	-288.32046
2x1x1	$9,676,481$	-288.33380	-288.33411
2x2x1	$16,358,791$	-288.34715	-288.34791
2x2x2	$27,208,731$	-288.35985	-288.36140
3x3x3	$79,226,681$	-288.37132	-288.37341

The ground-state energies computed using the pro-
posed subspace projection algorithm for varying sizes of
silicon nano-clusters considered in this work are tabu-
lated in Table V. These results indicate that the pro-
posed method provides good accuracies in the ground-
state energies, where the computed energies are within
$3 \, mHa$ per atom of the reference energies computed us-
ing ChFSI-FE. The computational times for the full SCF
convergence for varying sizes of silicon nano-clusters are
plotted in figure 7 for the proposed method, as well as,
the reference calculations using ChFSI-FE. These results
indicate that the proposed approach is computationally
efficient, compared to ChFSI-FE, for system sizes beyond
18 atoms and provides a factor of ~ 3 speedup for the 3
$\times$ 3 $\times$ 3 nano-cluster containing 280 atoms. Using these
computational times, the estimated scaling for the pro-
posed approach is found to be approximately $\mathcal{O}(N^{1.85})$,
while scaling in the case of ChFSI-FE is $\mathcal{O}(N^{2.21})$. The
average computational time per SCF iteration is shown in
figure 8, and the scaling is computed to be $\mathcal{O}(N^{1.75})$ for
the proposed subspace projection method and $\mathcal{O}(N^{2.11})$
for ChFSI-FE. The deterioration in the scaling for all-
electron calculations in comparison to pseudopotential
calculations is due to tighter truncation tolerances em-
ployed in order to control the accumulation of the trun-
cation errors during the application of a very high degree
Chebyshev filter. Nevertheless, the proposed approach
provides significant savings, which will increase with in-
creasing system size. We refer to the Appendix for the
scaling of various components of the subspace projection
algorithm in the case of all-electron calculations and a
discussion. Figure 9 shows the electron density contours
on the mid-plane of $3 \times 3 \times 3$ nano-cluster computed using
the proposed subspace projection technique.

VI. SUMMARY

In the present study, we formulated a subspace projection technique in the framework of higher-order spectral finite-element discretization of the Kohn-Sham DFT problem in order to reduce the computational complexity involved in traditional solution approaches that compute the canonical orthonormal Kohn-Sham eigenfunctions. The proposed approach provides a single unified framework to handle both insulating and metallic materials system. Further, both pseudopotenial as well as all-electron calculations can be conducted using the proposed methodology.

The development of the proposed approach involved bringing together four main ideas. Firstly, we employed a higher-order spectral finite-element basis for the discretization of the Kohn-Sham DFT problem. The adaptive nature of the finite-element basis is crucial for efficiently handling all-electron DFT calculations. Secondly, we employed the Chebyshev filtering approach[44] to directly compute an approximation of the occupied eigenspace in each SCF iteration. In this Chebyshev filtering step, we effectively exploited the finite-element structure to conduct matrix-vector products, associated with the action of the Chebyshev filter on a space of localized trial wavefunctions from previous SCF iteration, in linear-scaling complexity. We subsequently employed the localization procedure proposed by Garcia et al.[62] to compute atom-centered non-orthogonal localized basis (localized wavefunctions) spanning the Chebyshev filtered subspace. We employed an adaptive tolerance, where looser tolerances are used in initial iteration and progressively become tighter as the SCF approaches convergence, to provide a compact support for the localized wavefunctions. The use of adaptive tolerance provides strict control on the accuracy of the calculation, which is reflected in our benchmark calculations. Finally, we computed the projection of the Hamiltonian into the non-orthogonal localized basis, and used the Fermi-operator expansion[54] to compute the relevant quantities, including the finite-temperature density matrix, electron density and the band energy. We note that as the Fermi-operator expansion is computed in the projected subspace, the spectral width of the projected Hamiltonian is bounded—$\mathcal{O}(1\,Ha)$ for pseudopotential calculations and $\mathcal{O}(10\,Ha)$ for all-electron calculations—and can efficiently be computed for both pseudopotential and all-electron calculations. We demonstrated from complexity estimates that, for well-localized wavefunctions with a compact support, all operations in the proposed algorithm are linear-scaling in complexity.

The accuracy and performance of the proposed method was investigated on three different materials system: (i) a series of aluminum nano-clusters up to 3430 atoms representing a metallic system; (ii) a series of alkane chains up to 7052 atoms representing an insulating system; (iii) a series of silicon nano-clusters up to 3920 electrons representing a semiconducting system. Pseudopotential calculations were conducted on aluminum nano-clusters and alkane chains, whereas all-electron calculations were performed on silicon nano-clusters. In all the cases, the proposed method provided ground-state energies that are in excellent agreement with reference calculations, with accuracies commensurate with chemical accuracy. From these benchmark calculations, the computational complexity of the proposed approach was computed to be $\mathcal{O}(N^{1.46})$ for aluminum nano-clusters, $\mathcal{O}(N^{1.18})$ for the alkane chains, and $\mathcal{O}(N^{1.75})$ for the all-electron silicon nano-clusters. The deviation from linear-scaling, in practice, is due to the use of adaptive tolerances with tighter tolerances in the later SCF iterations in order to ensure strict control on the accuracy of the calculations. This affects the scaling due to reduced sparsity in the localized wavefunctions. We further note that using the proposed approach $\sim 10-$fold speedups were obtained with respect to reference benchmark calculations for the largest systems.

The present work demonstrates a methodology to conduct large scale electronic structure calculation using spectral finite-element discretizations at reduced scaling and presents an important direction for electronic structure calculations employing the finite-element basis. The computational efficiency as well as scaling in the case of all-electron calculations can be further improved by using a finite-element basis enriched with single-atom wavefunctions, or, alternately, the partitions-of-unity finite-element approach[101], and is currently being investigated. Also, the subspace projection method, which exploits the locality of the electronic structure, constitutes an important step in developing seamless coarse-graining techniques for Kohn-Sham density functional theory, in the similar spirit of quasi-continuum reduction techniques for electronic structure calculations[26,70].

for providing the computing resources through the Flux computing platform.

Appendix A: Scaling performance of individual components of the subspace projection technique

1. Case study: Pseudopotential calculations

We report scaling of individual components of the proposed subspace projection algorithm with system size for the benchmark calculations involving aluminum nano-clusters and alkane chains reported in section V. The average CPU-times per SCF iteration of the various components involved in the proposed technique (cf. section IV)—namely: a) Chebyshev filtered subspace iteration (ChFSI) b) Localization procedure (Loc) c) Subspace projection in the non-orthogonal basis (SubProj) and d) Electron-density computation (ElecDen)—have been recorded and are plotted against number of atoms.

a. Aluminum nano-clusters: Figure 10 shows the average computational CPU-times per SCF iteration for individual components of the subspace projection technique in the case of aluminum nano-clusters. These results indicate a computational complexity of $\mathcal{O}(N^{1.34})$ for the Chebyshev filtered subspace iteration, $\mathcal{O}(N^{1.91})$ for the localization procedure, $\mathcal{O}(N^{1.93})$ for the subspace projection and $\mathcal{O}(N^{1.94})$ for the computation of electron-density. The deviation of the scaling behaviour from linearity for the individual components of the algorithm is primarily due to the use of an adaptive tolerance (cf. equation (70)) to truncate the localized wavefunctions, which is important to ensure the accuracy of the subspace projection technique. We recall that looser truncation tolerances are employed in the initial SCF iterations with progressive tightening of the tolerances during the course of SCF convergence. We note that the Chebyshev filtering procedure has the dominant computational cost, and also exhibits better scaling in comparison to the other components of the subspace projection algorithm. The better scaling behaviour of Chebyshev filtering procedure can be attributed to the matrix-vector multiplications performed at the finite-element level only if the relevant wave-functions have a non-zero value in the finite-element considered (cf. Section IV B). This allows us to naturally exploit the sparsity of the wavefunctions at the finite-element level even with tight tolerances. The higher scaling observed for the other components of the subspace projection algorithm is due to the use of dense data-structures of PETSc when the density fraction of the relevant matrices is above 2%, as sparse PETSc data-structures have been observed to be efficient only when the density fraction is $< 1 - 2\%$. Figure 11 shows the average computational CPU-times per SCF iteration when the sparse data-structures are active. The results indicate a computational complexity of $\mathcal{O}(N^{1.20})$ for the localization procedure, $\mathcal{O}(N^{1.32})$ for the subspace
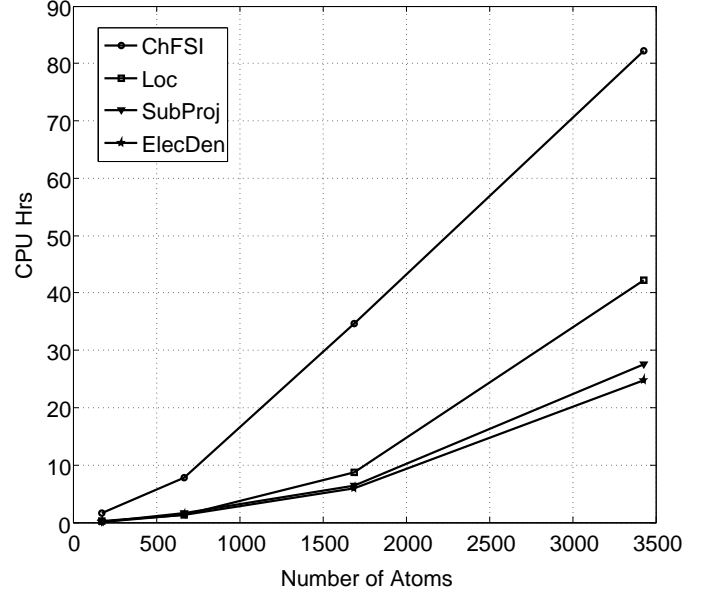

FIG. 10: Average computational times per SCF iteration for individual components of the proposed method. Case Study: Aluminum nano-clusters.
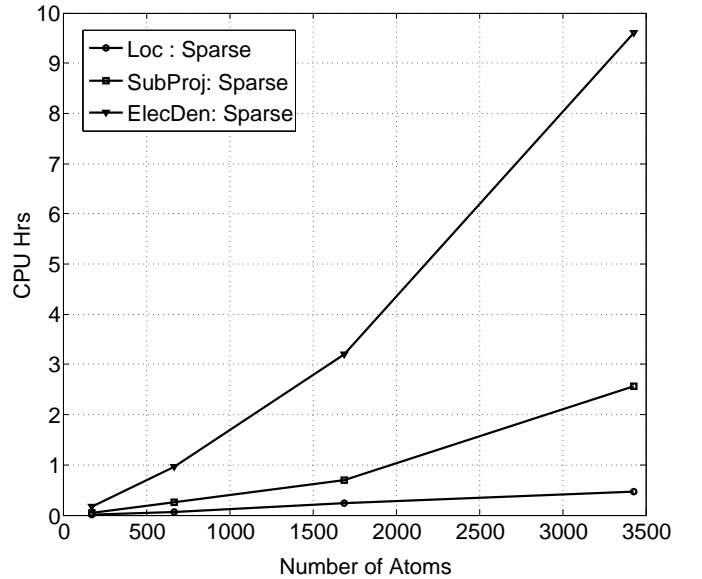

FIG. 11: Average computational times per SCF iteration for individual components when sparse data-structures are active. Case Study: Aluminum nano-clusters.

projection and $\mathcal{O}(N^{1.34})$ for the computation of electron-density. We remark that the scaling exponents of these individual components are closer to linearity when the sparse data-structures are active, but still deviate from linearity. We attribute this deviation from linearity to the delocalized nature of the wavefunctions for a metallic system. This delocalized nature of wavefunctions results in a higher density fraction (lesser sparsity) of truncated wavefunctions ($\mathbf{\Phi}_L$) for any given truncation tolerance in comparison to the insulating alkane chains (see discussion below on alkane chains). We also note that, as expected, the scaling of the individual components of the subspace projection technique (Loc, SubProj, ElecDen) is close to cubic-scaling ($\sim \mathcal{O}(N^{2.8})$) when dense data-structures are employed. Figure 14 shows the variation of density fraction of $\mathbf{\Phi}_L$ with SCF iteration in the case of aluminum nano-cluster (7x7x7 cluster with 1688 atoms) highlighting the SCF iterations in which sparse data-structures are active. Figure 15 shows the variation of relative error in ground-state energy with SCF iteration for the same benchmark problem.

b. Alkane Chains: Figure 12 shows the average computational CPU-times per SCF iteration for individual components of the subspace projection technique in the case of alkane chains. These results indicate a computational complexity of $\mathcal{O}(N^{1.15})$ for the Chebyshev filtered subspace iteration, $\mathcal{O}(N^{1.80})$ for the localization procedure, $\mathcal{O}(N^{1.85})$ for the subspace projection and $\mathcal{O}(N^{1.91})$ for the computation of electron-density. As in the case of aluminum clusters, the Chebyshev filtering procedure comprises the dominant computational cost of the subspace projection algorithm, and is almost linear-scaling for this system. The improved scaling of the Chebyshev filtering procedure in comparison to aluminum nano-clusters can be attributed to better localization of wavefunctions in the case of alkane chains, an insulating material system. The higher scaling of other components of the subspace projection algorithm is once again due to the use of dense data-structures of PETSc when density fraction of the relevant matrices is above 2%, while sparse data-structures are employed only when the density fraction is < 2%. Figure 13 shows the average computational CPU-times per SCF iteration when sparse data-structures are active. The results indicate a computational complexity of $\mathcal{O}(N^{1.13})$ for the localization procedure, $\mathcal{O}(N^{1.25})$ for the subspace projection and $\mathcal{O}(N^{1.29})$ for the computation of electron-density. We remark that these scaling exponents are smaller in comparison to those of aluminum nano-cluster (metallic nature) due to the localized nature of the wavefunctions, thus resulting in smaller density fractions (greater sparsity) of the truncated wavefunctions ($\mathbf{\Phi}_L$). Figure 14 shows the variation of density fraction with SCF iteration in the case of $C_{900}H_{1802}$ (2702 atoms) highlighting the SCF iterations in which sparse data-structures are active. Figure 15 shows the variation of relative error in ground-state energy with SCF iterations for the same problem.
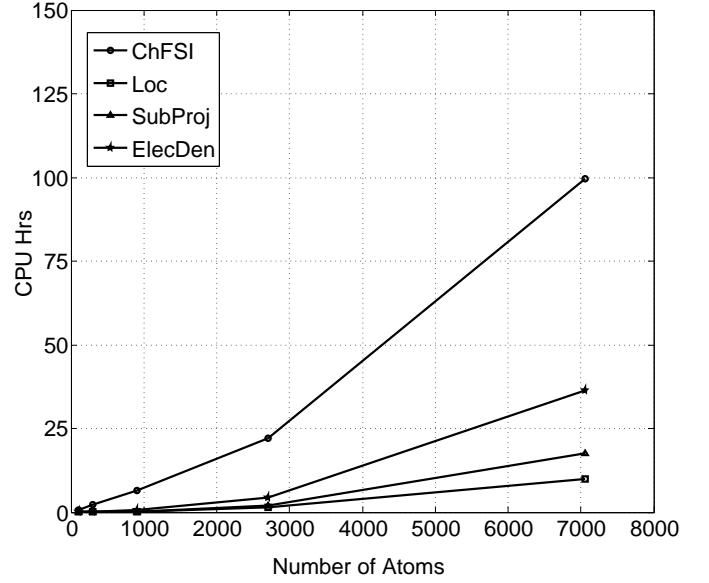

FIG. 12: Average computational times per SCF iteration for individual components of the proposed method. Case Study: Alkane chain.
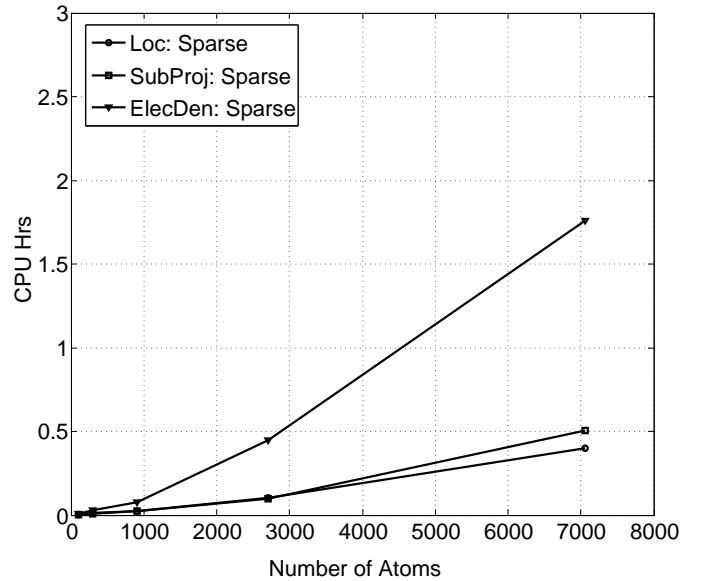

FIG. 13: Average computational times per SCF iteration for individual components when sparse data-structures are active. Case Study: Alkane chain.
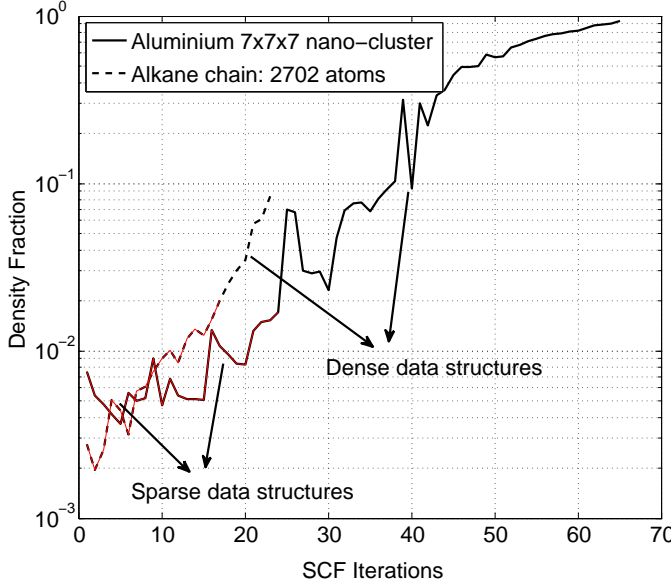
FIG. 14: Variation of density fraction with SCF iteration showing where the sparse data-structures (brown line) and dense data-structures are used. Case Study: Aluminum 7x7x7 nano-cluster and Alkane chain $C_{900}H_{1802}$.
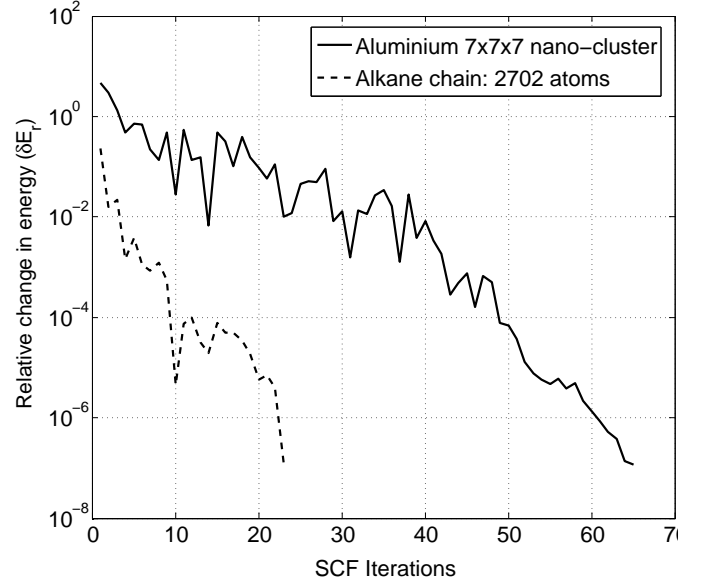

FIG. 15: Relative change in ground-state energy with SCF iteration. Case Study:Aluminum 7x7x7 nano-cluster and Alkane chain $C_{900}H_{1802}$.

2. Case study: All-electron calculations

Silicon nano-clusters: Figure 16 shows the average computational CPU-times per SCF iteration for individual components of the subspace projection technique in the case of all-electron calculations performed on silicon nano-clusters. These results indicate a computational complexity of $\mathcal{O}(N^{1.74})$ for the Chebyshev filtered subspace iteration, $\mathcal{O}(N^{2.72})$ for the localization procedure, $\mathcal{O}(N^{2.55})$ for the subspace projection and $\mathcal{O}(N^{2.51})$ for the computation of electron density. In comparison to the pseudopotential calculations, the scaling of the individual components deviate significantly from linearity. The main reason for this significant deviation is due to the tighter adaptive tolerances employed in all-electron calculations. These tighter tolerances were necessary to avoid error accumulations during the Chebyshev filtering procedure as a very high order Chebyshev filter is needed in all-electron calculations to filter the large unwanted spectrum. For these tighter tolerances, the density fractions for the various relevant matrices were greater than 2% even during the initial SCF iterations, which explains the observed close to cubic-scaling of the localization procedure, subspace projection, and computation of electron density. However, we note that the significantly dominant cost for all-electron calculations is the Chebyshev
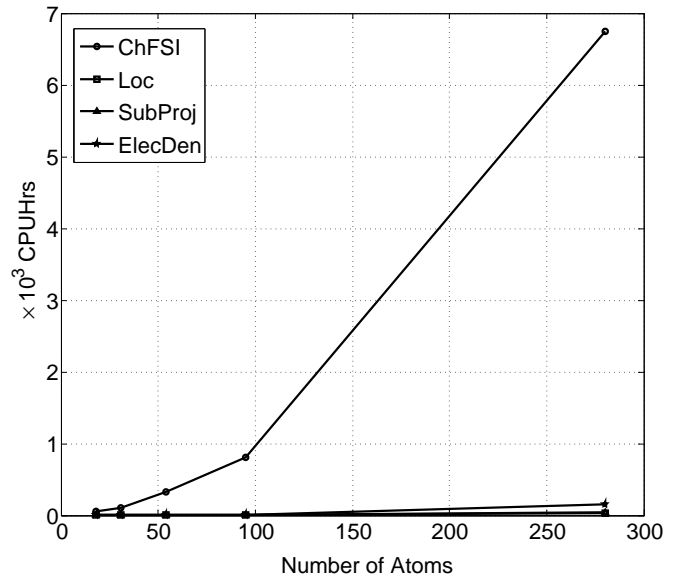

FIG. 16: Average computational times per SCF iteration for individual components of the proposed method. Case Study: Silicon nano-clusters.

filtering step, which naturally exploits the sparsity in truncated wavefunctions, even for high density fractions, by computing the matrix-vector products at the finite-element level only if the relevant wave-functions have a non-zero value in the finite-element being considered (cf. Section IV B). Thus, the overall scaling of the proposed technique for all-electron calculations is determined by the scaling of the Chebyshev filtering step, even for modestly sized systems. While, the performance of the proposed approach for all-electron calculations is not as good as the performance for pseudopotential calculations, the proposed approach does offer significant computational savings ($\sim$ 3 fold in comparison to ChFSI-FE for the

silicon nano-cluster containing $\sim$ 4000 electrons). Further, we note that the need of a very high degree Chebyshev filter for all-electron calculations can be mitigated by employing an enriched finite-element basis, where the finite-element basis functions are enriched by numerically computed single-atom wavefunctions, and this can potentially lead to better overall scaling of the method for all-electron calculations.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] G. E. Scuseria and V. N. Staroverov in *Progress in the development of exchange-correlation functionals*, edited by C. E. Dykstra, G. Frenking, K. S. Kim and G. E. Scuseria, Theory and Applications of Computational Chemistry: The First Forty Years (Elsevier B.V., Amsterdam, 2005).
[4] G. Kresse, J. Furthmüller, Phys. Rev. B **54** 11169 (1996).
[5] M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark and M. C. Payne,, J. Phys. Cond. Matt. **14** 2717 (2002).
[6] X. Gonze, J. -M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. -M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph. Ghosez, J. -Y. Raty and D. C. Allan, Comp. Mat. Sci. **25** 478 (2002).
[7] W. J. Hehre, R. F. Stewart and J. A. Pople, J. Chem. Phys. **51** 2657 (1969).
[8] F. Jensen, *Introduction to Computational Chemistry* (Wiley Publishers, New York, 1999).
[9] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko,S. V. Levchenko, D. P. O'Neill, R. A. Distasio Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M., Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, Phys. Chem. Chem. Phys. **8** 3172 (2006).
[10] H. -J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, WIREs Comput. Mol. Sci. **2** 242 (2012).
[11] J. M. Soler, E. Artacho, J. D. Gale, A. Garca, J. Junquera, P. Ordejn and D. Snchez-Portal, J. Phys.: Condens. Matter **14** 2745 (2002).
[12] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, Comput. Phys. Commun. **180** 2175 (2009).
[13] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72** 1240 (1994).
[14] N. A. Modine, G. Zumbach and E. Kaxiras, Phys. Rev. B **55** 10289 (1997).
[15] T.L. Beck, Rev. Mod. Phys. **72** 1041 (2000).
[16] A. Castro, H. Appel, M. Oliveira, C. A. Rozzi, X. Andrade, F. Lorenzen, M. A. L. Marques, E. K. U. Gross and A. Rubio, Phys. Status Solidi B **243** 2465 (2006).
[17] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman and R. Schneider, J. Chem. Phys. **129** 014109 (2008).
[18] D. Bowler, R. Choudhury, M. Gillan and T. Miyazaki, Phys. Status. Solidi. B **243** 989 (2006).
[19] C. -K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, J. Chem. Phys. **122** 084119 (2005).
[20] K. Varga, Z. Zhang, S. T. Pantelides, Phys. Rev. Lett. **93** 176403 (2004).
[21] A. S. Banerjee, R. S. Elliot, R. D. James, http://arxiv-web3.library.cornell.edu/pdf/1404.3773v1.pdf.
[22] S.C. Brenner, L.R. Scott, *The mathematical Theory of Finite-element Methods* (Springer, New York, 2002).
[23] E. J. Bylaska, M. Host and J. H. Weare., J. Chem. Theory. Comput. **5** 937 (2009).
[24] L. Lehtovaara, V. Havu and M. Puska, J. Chem. Phys. **131** 054103 (2009).
[25] P. Motamarri, M. Nowak, K. Leiter, J. Knap and V. Gavini, J. Comput. Phys. **253**, 308 (2013).
[26] V. Gavini, K. Bhattacharya, M. Ortiz, J. Mech. Phys. Solids **55**, 697 (2007).
[27] B. G. Radhakrishnan, V. Gavini, Phys. Rev. B **82**, 094117 (2010).
[28] B. Hermansson, D. Yevick, Phys. Rev. B **33**, 7241 (1986).
[29] S. R. White, J. W. Wilkins and M. P. Teter, Phys. Rev. B **39** 5819 (1989).
[30] E. Tsuchida and M. Tsukada, Phys. Rev. B **52** 5573 (1995).
[31] E. Tsuchida and M. Tsukada, Phys. Rev. B **54** 7602 (1996).
[32] E. Tsuchida and M. Tsukada, J. Phys. Soc. Jpn. **67** 3844 (1998).

[33] J. E. Pask, B. M. Klein, C. Y. Fong and P. A. Sterne, Phys. Rev. B **59** 12352 (1999).

[34] P. F. Batcho, Phys. Rev. E **61** 7169 (2000).

[35] J. E. Pask, B. M. Klein, P. A. Sterne and C. Y. Fong, Comp. Phys. Comm. **135** 1 (2001).

[36] J. E. Pask and P. A. Sterne, Mod. Sim. Mat. Sci. Eng. **13** R71 (2005).

[37] D. Zhang, L. Shen, A. Zhou and X. Gong, Phys. Lett. A **372** 5071 (2008).

[38] P. Suryanarayana, V. Gavini, T. Blesgen, K. Bhattacharya and M. Ortiz, J. Mech. Phys. Solids **58**, 256 (2010).

[39] L. Lin, J. Lu and W. E, J. Comp. Phys. **231** 2140 (2012).

[40] J. Fang, X. Gao and A. Zhou, J. Comp. Phys. **231** 3166 (2012).

[41] G. Bao, G. Hu and D. Liu, J. Comp. Phys. **231** 4967 (2012).

[42] A. Masud, R. Kannan, Comput. Methods Appl. Mech. Engrg. **241** 112 (2012).

[43] P. Motamarri, M. Iyer, J. Knap and V. Gavini, J. Comput. Phys. **231**, 6596 (2012).

[44] Y.Zhou, Y.Saad, M.L.Tiago and J.R.Chelikowsky, Phys. Rev. E **74**, 066704 (2006).

[45] G. Schofield, J. R. Chelikowsky and Y. Saad, Comp. Phys. Comm **183**, 497 (2012).

[46] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[47] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[48] W. Yang, J. Mol. Struc. Theochem. **255**, 461 (1992).

[49] T. Ozaki, Phys. Rev. B **74**, 245101 (2006).

[50] M. Barrault, E. Cancés, W. W. Hager and C. L. Bris, J. Comput. Phys. **222**, 86 (2007).

[51] S. Goedecker, Phys. Rev. B **48**, 17573 (1993).

[52] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994)

[53] S. Goedecker and M. Teter, Phys. Rev. B **51**, 9455 (1995).

[54] R. Baer, M. Head-Gordon, J. Chem. Phys. **107**, 10003 (1997).

[55] X. P. Li, R. W. Nunes and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).

[56] P. D. Haynes, C. K. Skylaris, A. A. Mostofi and M. C. Payne, Phys. Stat. Sol.(b) **243**, 2489 (2006).

[57] O. Sankey, D. Drabold and A. Gibson, Phys. Rev. B **50**, 1376 (1994).

[58] U. Stephan and D. Drabold, Phys. Rev. B **57**, 6391 (1998).

[59] F. Mauri, G. Galli and R. Car, Phys. Rev. B **47**, 9973 (1993).

[60] J. Kim, F. Mauri and G. Galli, Phys. Rev. B **52**, 1640 (1995).

[61] W. Gao, W. E, Discret. Contin. Dyn. Sys. A **23**, 249 (2009).

[62] C. J. Garcia-Cervera, J. Lu, Y. Xuan and W. E, Phys. Rev. B **79**, 115110 (2009).

[63] D. R. Bowler and T. Miyazaki, Rep. Prog. Phys. **75**, 036503 (2012).

[64] P. Suryanarayana, Chem. Phys. Lett. **584**, 182 (2013).

[65] T. Ozaki, Phys. Rev. B **75**, 035123 (2007).

[66] M. Ceriotti, T. Kuhne and M. Parrinello, J. Chem. Phys. **129**, 024707 (2008).

[67] L. Lin, J. Lu, L. Ying and W. E, Chin. Ann. Math. **30B**, 729 (2009).

[68] T. Ozaki, Phys. Rev. B **82**, 075131 (2010).

[69] L. Lin, M. Chen, C. Yang and L. He, J. Phys:Condens. Matter **25**, 295501 (2013).

[70] P. Suryanarayana, K. Bhattacharya, M. Ortiz, J. Mech. Phys. Solids **61**, 38 (2013).

[71] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2011).

[72] D. C. Langreth and M. J. Mehl, Phys. Rev. B **28**, 1809 (1983).

[73] J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[74] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45** 566 (1980).

[75] J. P. Perdew and A. Zunger, Phys. Rev. B **23** 5048 (1981).

[76] N. Troullier and J. L. Martin, Phys. Rev. B **43**, 1993(1991).

[77] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[78] Y. Saad, *Numerical Methods for Large Eigenvalue Problems* (John Wiley, New York, 1992).

[79] C. Lanczos, J. Res. Nat. Bureau Standards, Sec. B **45**, 255 (1950).

[80] T.J.Rivlin, *Chebyshev Polynomials:From Approximation Theory to Algebra and Number Theory*(John Wiley, New York, 1973).

[81] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

[82] W. Kohn, Phys. Rev. **115**, 809 (1959).

[83] W. Kohn, Phys. Rev. B **7**, 4388 (1973).

[84] W. Kohn, Chem. Phys. Lett. **208**, 167 (1993).

[85] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[86] P. Ordejon, D. A. Drabold, M. P. Grumbach and R. M. Martin, Phys. Rev. B **48**, 14646 (1993).

[87] P. W. Anderson, Phys. Rev. Lett. **21**, 13 (1968).

[88] W. E, L. Tiejun and L. Jianfeng, Proc. Nat. Acad. Sci. **107**, 1273 (2010).

[89] V. Ozolins, R. Lai, R. Calflisch and S. Osher, Proc. Nat. Acad. Sci. **110**, 18368 (2013).

[90] B. Jansik, S. Host, P. Jorgensen, J. Olsen and T. Helgaker, J. Chem. Phys **126**, 124104(2007).

[91] E. Rubenson and P. Salek, J. Comput. Chem. **26**, 1628 (2005).

[92] S. Goedecker, Phys. Rev. B **58**, 3501 (1998).

[93] S. Ismail-Beigi and T. Arias, Phys. Rev. Lett. **82**, 2127 (1999).

[94] G.W.Stewart, SIAM J. Matrix Anal. Appl. **23**, 601614 (2001).

[95] K. E. Atkinson, *An Introduction to Numerical Analysis*(John Wiley & Sons, Inc, 1989).

[96] D. G. Anderson, J. Assoc. Comput. Mach. **12**, 547 (1965)

[97] S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, H. Zhang, *PETSc 3.4 Users Manual*, Argonne National Laboratory, 2013.

[98] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo,

J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision A.1, Gaussian, Inc., Walling-ford CT, 2009.

[99] M. Fuchs and M. Scheffler, Comp. Phys. Comm. **119**, 67 (1999).

[100] L. Lin and C. Yang, SIAM J. Sci. Comput. **35**, S277 (2013).

[101] N. Sukumar, J. E. Pask, Int. J. Numer. Meth. Engng. **77** 1121 (2009).