

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Accelerating Atomic Orbital-based Electronic Structure Calculation via Pole Expansion plus Selected Inversion

Lin Lin¹, Mohan Chen¹, Chao Yang¹, Lixin He¹

^a*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*

^b*Key Laboratory of Quantum Information, CAS, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China*

Abstract

We describe how to apply the recently developed pole expansion plus selected inversion (PEpSI) technique to Kohn-Sham density function theory (DFT) electronic structure calculations that are based on atomic orbital discretization. We give analytic expressions for evaluating charge density, total energy, Helmholtz free energy and atomic forces without using the eigenvalues and eigenvectors of the Kohn-Sham Hamiltonian. We also show how to update the chemical potential without using Kohn-Sham eigenvalues. The advantage of using PEpSI is that it has a much lower computational complexity than that associated with the matrix diagonalization procedure. We demonstrate the performance gain by comparing the timing of PEpSI with that of diagonalization on insulating and metallic nanotubes. For these quasi-1D systems, the complexity of PEpSI is linear with respect to the number of atoms. This linear scaling can be observed in our computational experiments when the number of atoms in a nanotube is larger than a few hundreds. Both the wall clock time and the memory requirement of PEpSI is modest. This makes it even possible to perform Kohn-Sham DFT calculations for 10,000-atom nanotubes on a single processor. We also show that the use of PEpSI does not lead to loss of accuracy required in a practical DFT calculation.

Keywords: Kohn-Sham density functional theory, pole expansion, selected inversion, atomic orbital, nonorthogonal basis function, fast algorithm, large scale computation

Email addresses: linlin@lbl.gov (Lin Lin), mohan@mail.ustc.edu.cn (Mohan Chen), cyang@lbl.gov (Chao Yang), helx@ustc.edu.cn (Lixin He)

Preprint submitted to Elsevier

February 22, 2012

1. Introduction

Electronic structure calculations based on solving the Kohn-Sham equations play an important role in the analysis of electronic, structural and optical properties of molecules, solids and other nano structures. The efficiency of such a calculation depends largely on the computational cost associated with the evaluation of the electron charge density for a given potential within a self-consistent field (SCF) iteration. The most straightforward way to perform such an evaluation is to partially diagonalize the Kohn-Sham Hamiltonian by computing a set of eigenvectors corresponding to the algebraically smallest eigenvalues of the Hamiltonian. The complexity of this approach is $\mathcal{O}(N_e^3)$, where N_e is the number of electrons in the atomistic system of interest. As the number of atoms or electrons in the system increases, the cost of diagonalization becomes prohibitively expensive.

An alternative scheme for evaluating the charge density is to express the charge density as the diagonal of the Fermi-Dirac function evaluated at a fixed Kohn-Sham Hamiltonian. By approximating the Fermi-Dirac function through a pole expansion technique [1], we can reduce the problem of computing the charge density to that of computing the diagonal of the inverses of a number of shifted Kohn-Sham Hamiltonians. This approach was pursued by a number of researchers in the past. The cost of this approach depends on the number of poles required to expand the Fermi-Dirac function and the cost for computing the diagonal of the inverse of a shifted Kohn-Sham Hamiltonian.

The recent work by Lin et al. [1] provides an accurate and efficient pole-expansion scheme for approximating the Fermi-Dirac function. The number of poles required in this approach is proportional to $\log(\beta\Delta E)$, where β is proportional to the inverse of the temperature factor, and ΔE is the spectral width of the Kohn-Sham Hamiltonian. (i.e. the difference between the largest and the smallest eigenvalues). This pole count is significantly lower than the previous approaches [2, 3, 4, 5].

Furthermore, an efficient selected inversion algorithm for computing the inverse of the diagonal of a shifted Kohn-Sham Hamiltonian without computing the full inverse of the Hamiltonian has been developed [6, 7, 8]. The complexity of this algorithm is $\mathcal{O}(N_e)$ for quasi-1D systems such as nanorods,

nanotubes and nanowires, $\mathcal{O}(N_e^{3/2})$ for quasi-2D systems such as graphene and surfaces, and $\mathcal{O}(N_e^2)$ for 3D bulk systems. In exact arithmetic, the selected inversion algorithm gives the exact diagonal of the inverse, i.e., the algorithm does not rely on any type of localization or truncation scheme. For insulating systems, the use of localization and truncation can be combined with selected inversion to reduce the complexity of the algorithm further to $\mathcal{O}(N_e)$ even for general 3D systems.

In the previous work [7, 8] we used the pole expansion plus selected inversion (PEpSI) technique to solve the Kohn-Sham problem discretized by a finite difference scheme. However, it is worth pointing out that PEpSI is a general technique that is not limited to discretized problems obtained from finite difference. In particular, it can be readily applied to discretized Kohn-Sham problems obtained from any localized basis expansion technique. In this paper, we describe how PEpSI can be used to speed up solving a discretized Kohn-Sham problem obtained from an atomic orbital basis expansion. We show that electron charge density, total energy, Helmholtz free energy and atomic forces can all be efficiently calculated by using PEpSI.

We demonstrate the performance gain we can achieve by comparing PEpSI with the LAPACK diagonalization subroutine `dsygv` on two types of nanotubes. We show that by using the PEpSI technique, it is possible to perform electronic structure calculations accurately for a nanotube that contains 10,000 atoms on a single processor within a reasonable amount of time. The crossover point beyond which the computational complexity of PEpSI exhibits linear scaling with respect to the number of atoms is around a few hundred atoms.

This paper is organized as follows. In section 2, we show how the PEpSI technique previously developed [7, 8] can be extended to discretized Kohn-Sham problems obtained from an atomic orbital expansion scheme. In particular, we will show how charge density, total energy, free energy and force can be calculated in this formalism. We will also discuss how to update the chemical potential. In section 3, we report the performance of PEpSI on two quasi-1D test problems.

Throughout the paper, we use $\Im\mathbf{m}(A)$ to denote the imaginary part of a complex matrix A . A properly defined inner product between two functions f and g is sometimes denoted by $\langle f|g\rangle$. The diagonal of a matrix A is sometimes denoted by $\text{diag}(A)$.

2. Theory

The ground-state electron charge density $\rho(x)$ of an atomistic system can be obtained from the self-consistent solution to the Kohn-Sham equations

$$\hat{H}[\rho(x)]\psi_i(x) = \psi_i(x)\varepsilon_i, \quad (1)$$

where \hat{H} is the Kohn-Sham Hamiltonian that depends on $\rho(x)$, $\{\psi_i(x)\}$ are the Kohn-Sham orbitals that satisfy the orthonormality constraints

$$\int \psi_i^*(x)\psi_j(x)dx = \delta_{ij}, \quad (2)$$

and the eigenvalue ε_i is often known as the i th Kohn-Sham quasi-particle energy. Using the Kohn-Sham orbitals, we can define the charge density by

$$\rho(x) = \sum_i^{\infty} |\psi_i(x)|^2 f_i, \quad i = 1, 2, \dots, \infty, \quad (3)$$

with occupation numbers $0 \leq f_i \leq 2$, $i = 1, 2, \dots, \infty$. At finite temperature $T = 1/(k_B\beta)$, the occupation numbers in (3) can be chosen according to the Fermi-Dirac distribution function

$$f_i = f_\beta(\varepsilon_i - \mu) = \frac{2}{1 + e^{\beta(\varepsilon_i - \mu)}}, \quad (4)$$

and μ is the chemical potential chosen to ensure that

$$\int \rho(x)dx = N_e. \quad (5)$$

Note that $\rho(x)$ is simply the diagonal of the single particle density matrix defined by

$$\hat{\gamma}(x, x') = \sum_{i=1}^{\infty} \psi_i(x) f_\beta(\varepsilon_i - \mu) \psi_i^*(x'), \quad (6)$$

and the charge sum rule in (5) can be expressed alternatively by

$$\text{Tr}[\hat{\gamma}(x, x')] = N_e, \quad (7)$$

where Tr denotes the trace of an operator.

It follows from (1) and (6) that the electron density $\rho(x)$ is a fixed point of the Kohn-Sham map defined by

$$\rho(x) = \text{diag} \left(f_{\beta}(\hat{H}[\rho(x)] - \mu\delta(x, x')) \right), \quad (8)$$

where μ is chosen to satisfy (7). The most widely used algorithm for finding the solution to (7) and (8) is a Broyden type of quasi-Newton algorithm. In the physics literature, this is often referred to as the self-consistent field (SCF) iteration. The most time consuming part of this algorithm is the evaluation of $\rho(x) = \hat{\gamma}(x, x)$ in (8).

2.1. Basis expansion by nonorthogonal basis functions

An infinite-dimensional Kohn-Sham problem can be discretized in a number of ways (e.g., plane-wave expansion, finite difference, finite element etc.) In this paper, we focus on a discretization scheme in which a Kohn-Sham orbital ψ_i is expanded by a linear combination of a finite number of basis functions $\{\varphi_j\}$, i.e.,

$$\psi_i(x) = \sum_{j=1}^N \varphi_j(x) c_{ji}. \quad (9)$$

We should note that the total number of basis functions N is generally proportional to the number of electrons N_e or atoms in the system to be studied. These basis functions $\{\varphi_j\}$ can be constructed to have local nonzero support. But they may not necessarily be orthonormal to each other. Examples of these basis functions are wavelet basis functions [9, 10], adaptive local basis functions [11], Gaussian type orbitals [12, 13] and local atomic orbitals [14, 15, 16, 17, 18, 19]. In numerical examples presented in section 3, we use a set of nonorthogonal local atomic orbitals.

Substituting (9) into (1) yields a generalized eigenvalue problem

$$HC = SC\Xi, \quad (10)$$

where C is an $N \times N$ matrix with c_{ij} being its (i, j) th entry, Ξ is a diagonal matrix with ε_i on its diagonal, $S_{ij} = \langle \varphi_i | \varphi_j \rangle$, and $H_{ij} = \langle \varphi_i | \hat{H} | \varphi_j \rangle$. For orthogonal basis functions, the overlap matrix S is an identity matrix, and Eq. (10) reduces to a standard eigenvalue problem. When local atomic orbitals are used as the basis, S is generally not an identity matrix, but both H and S are sparse.

Without loss of generality, we assume the basis functions and the Kohn-Sham orbitals to be real in the following discussion. Let $\Psi = [\psi_1, \dots, \psi_N]$ and $\Phi = [\phi_1, \dots, \phi_N]$, Then Eq. (9) can be written in a compact form

$$\Psi = \Phi C. \quad (11)$$

Consequently, the single-particle density matrix (6) becomes

$$\begin{aligned} \hat{\gamma}(x, x') &= \Psi(x) f(\Xi - \mu) \Psi^T(x') \\ &= \Phi(x) C f_\beta(\Xi - \mu) C^T \Phi^T(x'). \end{aligned} \quad (12)$$

2.2. Pole expansion and selected inversion for nonorthogonal basis functions

The most straightforward way to evaluate $\hat{\gamma}(x, x')$ is to follow the right hand side of (12), which requires solving the generalized eigenvalue problem (10). The computational complexity of this approach is $\mathcal{O}(N^3)$. This approach becomes prohibitively expensive when the number of electrons or atoms in the system increases.

An alternative way to evaluate $\hat{\gamma}(x, x')$, which circumvents the cubic scaling of the diagonalization process, is to approximate $\hat{\gamma}(x, x')$ by a Fermi operator expansion (FOE) method [20]. In an FOE scheme, the function $f_\beta(\Xi - \mu)$ is approximated by a linear combination of a number of simpler functions, each of which can be evaluated directly without diagonalizing the matrix pencil (H, S) . A variety of FOE schemes have been developed. They include polynomial expansion [20], rational expansion [2, 1, 3], and a hybrid scheme in which both polynomials and rational functions are used [4, 21]. In all these schemes, the number of simple functions used in the expansion is asymptotically determined by $\beta\Delta E$, where $\Delta E = \max_{i=1}^N |\varepsilon_i - \mu|$ is the spectrum width for the discrete problem.

While most of the FOE schemes require as many as $\mathcal{O}(\beta\Delta E)$ or $\mathcal{O}(\sqrt{\beta\Delta E})$ terms of simple functions, the recently developed pole expansion [1] is particularly promising since it requires only $\mathcal{O}(\log \beta\Delta E)$ terms of simple rational functions. The pole expansion takes the form

$$f_\beta(\varepsilon - \mu) \approx \Im \sum_{l=1}^P \frac{\omega_l^p}{\varepsilon - (z_l + \mu)}, \quad (13)$$

where $z_l, \omega_l^p \in \mathbb{C}$ are complex shifts and weights respectively. We will refer to $\{z_l\}$ as *poles* in the following discussions. The construction of the pole expansion is based on the observation that the non-analytic part of the Fermi-Dirac

function lies only on the imaginary axis within $\left[\frac{i\pi}{\beta}, +i\infty\right] \cup \left[-i\infty, -\frac{i\pi}{\beta}\right]$. A dumbbell-shaped Cauchy contour (see Fig. 1) is carefully chosen and discretized to circle the eigenvalues $\{\varepsilon_i\}$ on the real axis, while avoiding the intersection with the non-analytic region. The pole expansion does not require a band gap between the occupied and unoccupied states. Therefore, it is applicable to both insulating and metallic systems. Furthermore, the construction of the pole expansion relies only on the analytical structure of the Fermi-Dirac function rather than its detailed shape. This is a key property that is crucial for constructing pole expansions for other functions, including the reduced free energy density matrix and the reduced energy density matrix which are discussed in section 2.3 for the purpose of computing Helmholtz free energy and atomic forces.

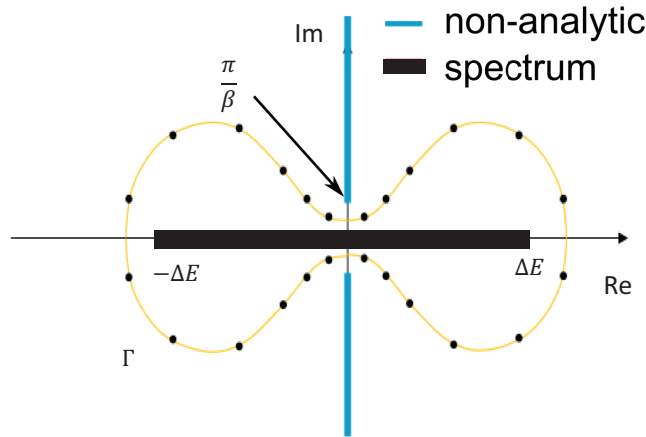


Figure 1: (color online) A schematic view of the placement of poles used in a pole expansion approximation of $f_\beta(z)$. The thick black line on the real axis indicates the range of $\varepsilon_i - \mu$, and the thin blue line on the imaginary axis indicates the non-analytic part of $f_\beta(z)$. The yellow dumbbell shaped contour is chosen to exclude the non-analytic part of the complex plane. Each block dot on the contour corresponds to a pole used in the pole expansion approximation.

Following the derivation in the appendix, we can use (13) to approximate

the single particle density matrix by

$$\begin{aligned}\hat{\gamma}(x, x') &\approx \Phi(x) \mathfrak{Im} \left(\sum_{l=1}^P \frac{\omega_l^\rho}{H - (z_l + \mu)S} \right) \Phi^T(x') \\ &\equiv \Phi(x) \gamma \Phi^T(x').\end{aligned}\tag{14}$$

In the above expression, γ is an $N \times N$ matrix. It is often referred to as the reduced single particle density matrix. Using Eq. (14), we can evaluate the electron density in the real space as the diagonal elements of $\hat{\gamma}$, i.e.,

$$\hat{\rho}(x) \approx \Phi(x) \gamma \Phi^T(x) = \sum_{ij} \gamma_{ij} \varphi_j(x) \varphi_i(x).\tag{15}$$

We assume that each basis function $\varphi_i(x)$ is compactly supported in the real space. In order to evaluate $\hat{\rho}(x)$ for any particular x , we only need γ_{ij} such that $\varphi_j(x) \varphi_i(x) \neq 0$, or $S_{ij} \neq 0$. This set of γ_{ij} 's is a subset of $\{\gamma_{ij} | H_{ij} \neq 0\}$. To obtain these *selected elements*, we need to compute the corresponding elements of $(H - (z_l + \mu)S)^{-1}$ for all z_l .

The recently developed selected inversion method [6, 7, 8] provides an efficient way of computing the selected elements of an inverse matrix. For a symmetric matrix of the form $A = H - zS$, the selected inversion algorithm first constructs an LDL^T factorization of A , where L is a block lower diagonal matrix called the Cholesky factor, and D is a block diagonal matrix. In the second step, the selected inversion algorithm computes all the elements A_{ij}^{-1} such that $L_{ij} \neq 0$. Since $L_{ij} \neq 0$ implies that $H_{ij} \neq 0$, all the selected elements of A^{-1} required in (15) are computed. As a result, the computational scaling of the selected inversion algorithm is only proportional to the number of nonzero elements in the Cholesky factor L . In particular, the selected inversion algorithm has a complexity of $\mathcal{O}(N)$ for quasi-1D systems, $\mathcal{O}(N^{1.5})$ for quasi-2D systems, and $\mathcal{O}(N^2)$ for 3D bulk systems. The selected inversion algorithm achieves universal improvement over the diagonalization method for systems of all dimensions. It should be noted that selected inversion algorithm is an *exact* method for computing selected elements of A^{-1} if exact arithmetic is to be employed, and in practice the only source of error is the roundoff error. Particularly, the selected inversion algorithm does not rely on any localization property of A^{-1} , though combined with localization property for insulating systems the computational cost can be further reduced which will be studied in future work.

2.3. Total energy, Helmholtz free energy and atomic force evaluation

In addition to reducing the computational complexity of charge density calculation in each SCF iteration, the PEPsi technique can also be used to compute the total energy, the Helmholtz free energy as well as the atomic forces efficiently without diagonalizing the Kohn-Sham Hamiltonian.

When Kohn-Sham orbitals $\{\psi_i\}$ are available, the total energy associated with an insulating system can be evaluated as

$$E_{\text{tot}} = \sum_{i=1}^{\infty} f_i \langle \psi_i, \hat{H} \psi_i \rangle - \frac{1}{2} \iint \frac{\hat{\rho}(x)\hat{\rho}(y)}{|x-y|} dx dy + E_{\text{xc}}[\hat{\rho}] - \int V_{\text{xc}}[\hat{\rho}](x)\hat{\rho}(x) dx. \quad (16)$$

An alternative expression for E_{tot} is

$$E_{\text{tot}} = \text{Tr}[\gamma H] - \frac{1}{2} \iint \frac{\hat{\rho}(x)\hat{\rho}(y)}{|x-y|} dx dy + E_{\text{xc}}[\hat{\rho}] - \int V_{\text{xc}}[\hat{\rho}](x)\hat{\rho}(x) dx, \quad (17)$$

where γ is the reduced density matrix defined in (6). Note that in this expression, the first term depends on the trace of the product of γ and H . The computation of this term requires only the (i, j) th entry of γ for (i, j) satisfying $H_{ij} \neq 0$. These entries are already available from the charge density calculation, thus using them for total energy evaluation does not introduce additional complexity. All other terms in the total energy expression depends on the electron density $\hat{\rho}(x)$, which we already know how to compute by the PEPsi technique. Here we assume LDA [22] or GGA [23, 24] exchange-correlation functional is used for the Kohn-Sham total energy expression.

For metallic systems, the Helmholtz free energy \mathcal{F}_{tot} is the quantity of interest because at finite temperature \mathcal{F}_{tot} takes into account both energy and entropy [25]. The Helmholtz free energy can be written as [26]

$$\mathcal{F}_{\text{tot}} = -2\beta^{-1} \text{Tr} \ln(1 + \exp(\beta(\mu - \Xi))) + \mu N_e - \frac{1}{2} \iint \frac{\hat{\rho}(x)\hat{\rho}(y)}{|x-y|} dx dy + E_{\text{xc}}[\hat{\rho}] - \int V_{\text{xc}}[\hat{\rho}](x)\hat{\rho}(x) dx. \quad (18)$$

It is straightforward to show that as $\beta \rightarrow \infty$, $\mathcal{F}_{\text{tot}} \rightarrow E_{\text{tot}}$. Again, in Eq. (18), only the first term requires special treatment. Note that the function

$$f_{\beta}^{\mathcal{F}}(\varepsilon - \mu) = -2\beta^{-1} \ln(1 + \exp(\beta(\mu - \varepsilon))) \quad (19)$$

is analytic everywhere in the complex plane, except for segments of the imaginary axis within $\left[\frac{i\pi}{\beta}, +i\infty\right] \cup \left[-i\infty, -\frac{i\pi}{\beta}\right]$. In this sense, $f_{\beta}^{\mathcal{F}}$ shares the same analytic structure as that of the Fermi-Dirac function f_{β} . The pole expansion technique can be applied with the same choice of poles $\{z_l\}$ but different weights, denoted by $\{\omega_l^{\mathcal{F}}\}$, *i.e.*

$$f_{\beta}^{\mathcal{F}}(\varepsilon - \mu) \approx \Im \sum_{l=1}^P \frac{\omega_l^{\mathcal{F}}}{\varepsilon - (z_l + \mu)}. \quad (20)$$

Following the derivation in the appendix, we can rewrite the Helmholtz free energy as

$$\begin{aligned} \mathcal{F}_{\text{tot}} = & \text{Tr}[\gamma^{\mathcal{F}} S] + \mu N_e - \frac{1}{2} \iint \frac{\hat{\rho}(x)\hat{\rho}(y)}{|x-y|} dx dy \\ & + E_{\text{xc}}[\hat{\rho}] - \int V_{\text{xc}}[\hat{\rho}]\hat{\rho}(x) dx, \end{aligned} \quad (21)$$

where the reduced free energy density matrix $\gamma^{\mathcal{F}}$ is given by

$$\gamma^{\mathcal{F}} \approx \Im \sum_{l=1}^P \frac{\omega_l^{\mathcal{F}}}{H - (z_l + \mu)S}. \quad (22)$$

Again, the selected elements of $[H - (z_l + \mu)S]^{-1}$ required for evaluation the first terms of (21) are already available from the charge density calculation. No additional computation is required to obtain these elements.

To perform geometric optimization or ab initio molecular dynamics, we need to compute atomic forces associated with different atoms. Atomic force is the derivative of the Helmholtz free energy with respect to the position of an atom. Following the derivation in the appendix, we can express the atomic force associated with the I -th atom as

$$F_I = -\frac{\partial \mathcal{F}}{\partial R_I} = -\text{Tr} \left[\gamma \frac{\partial H}{\partial R_I} \right] + \text{Tr} \left[\gamma^E \frac{\partial S}{\partial R_I} \right]. \quad (23)$$

where γ^E is the reduced energy density matrix defined by

$$\gamma^E = C \Xi f_\beta(\Xi - \mu) C^T. \quad (24)$$

It is clear that the function

$$f_\beta^E(\varepsilon - \mu) = \varepsilon f_\beta(\varepsilon - \mu) \quad (25)$$

shares the same analytic structure as that of the Fermi-Dirac function f_β . Thus, the reduced energy density matrix can be approximated by the same pole expansion used to approximate the reduced density matrix (14). In particular, there is no difference in the choice of poles z_l . But the weights of the expansion, which we denote by ω_l^E , for the reduced energy density matrix approximation, are different. To be specific, the reduced energy density matrix can be written as

$$\gamma^E \approx C \Im \sum_{l=1}^P \frac{\omega_l^E}{\Xi - (z_l + \mu)I} C^T = \sum_{l=1}^P \frac{\omega_l^E}{H - (z_l + \mu)S}. \quad (26)$$

Again the selected elements of γ^E required in (23) can be easily computed from the selected elements of $[H - (z_l + \mu)S]^{-1}$ which are available from the charge density calculation.

2.4. Chemical potential update

The true chemical potential μ required in the pole expansions (14), (21) and (26) is not known a priori. It must be solved iteratively as part of the solution to (7) and (8). For a fixed Hamiltonian H associated with a fixed charge density, it is easy to show that the left hand side (7), which can be expressed as,

$$N(\mu) = \text{Tr}[\hat{\gamma}] = \text{Tr}[\gamma \Phi^T \Phi] = \text{Tr}[\gamma S] \quad (27)$$

is a monotonic function with respect μ . Hence the root of (7) is unique. It can be obtained by either the Newton Raphson or the bisection method.

In an SCF iteration, ρ and μ are often updated in an alternating fashion. When the Kohn-Sham quasi-particle energies ϵ_i associated with a fixed charge density are available, both $N(\mu)$ and its derivative can be easily evaluated in the Newton's method. However, if $\hat{\gamma}$ is approximated via a pole expansion (14), a new expansion is needed whenever μ is updated. Furthermore, the derivative of $N(\mu)$ can be approximated by finite difference. In practice, one

or two Newton's iterations are sufficient to produce a reasonably accurate μ after the first SCF iteration. When μ^k is sufficiently close to the true chemical potential, the derivative of $N(\mu^k)$ can be approximated by

$$N'(\mu^k) \approx \frac{N(\mu^k) - N(\mu^{k-1})}{\mu^k - \mu^{k-1}}. \quad (28)$$

Thus each Newton's iteration requires only one more selected inversion calculation. This type of iterative strategy for updating the chemical potential has also been discussed in literature [27, 5].

3. Numerical results

In this section, we report the performance gain we achieved by applying the PEpSI technique to an existing electronic structure calculation code that uses local atomic orbital expansion to discretize the Kohn-Sham equations.

The test problems we used are two types of nanotubes. One is a boron nitride nanotube (BNNT) with chirality (8,0), which is an insulating system shown in Figure 2. The other is a carbon nanotube (CNT) with chirality (8,8) shown in Figure 3, which is a metallic system. According to the formula $d = \frac{\sqrt{3}a}{\pi} \sqrt{n^2 + mn + m^2}$, where a is the bond length and (n, m) is the chirality of nanotubes [28], the diameter for BNNT (8,0) is 12.09 Bohr and for CNT (8,8) is 20.50 Bohr. The longitudinal length of BNNT (8,0) with 256 atoms is roughly the same as CNT (8,8) with 512 atoms.

We performed our calculation at Γ point only. Because Brillouin zone sampling can be trivially parallelized, adding more k -points will not affect the performance of our calculation.

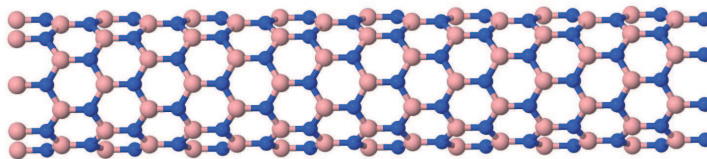


Figure 2: (color online) Boron nitride nanotube (8,0) with 256 atoms. The boron atoms are labeled as pink (light gray) balls while the nitrogen atoms are labeled as blue (dark gray) balls. The bond length between a pair of adjacent boron and nitride atoms is 1.45 Angstrom.

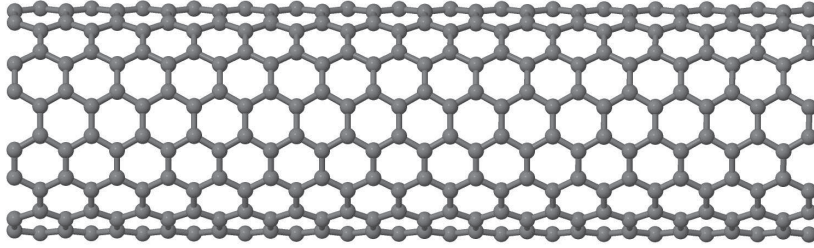


Figure 3: (color online) Carbon nanotube (8,8) with 512 atoms. The carbon atoms are labeled as gray balls. The bond length between a pair of adjacent carbon atoms is 1.42 Angstrom.

Our computational experiments were performed on the Hopper system at the National Energy Research Scientific Computing (NERSC) center. The performance results reported below were obtained from running the existing and modified codes on a single core of Hopper which is part of a node that consists of two twelve-core AMD 'MagnyCours' 2.1-GHz processors. Each Hopper node has 32 gigabytes (GB) DDR3 1333-MHz memory. Each core processor has 64 kilobytes (KB) L1 cache and 512KB L2 cache. It also has access to a 6 megabytes (MB) of L3 cache shared among 6 cores.

Although the existing code has been parallelized using MPI and ScaLAPACK, the parallelization of selected inversion is still work in progress. Hence, the performance study reported here is limited to single-processor runs. However, we expect that the new approach of using the PEpSI technique to compute the charge density, total energy, Helmholtz free energy and force will have a more favorable parallel scalability compared to diagonalizing the Kohn-Sham Hamiltonian by ScaLAPACK because it can take advantage of an additional level of parallelism introduced by the pole expansion. Due to the availability of such parallelism, the cost of the computational time of PEpSI is reported as the wall clock time for evaluating the selected elements of one single pole.

In addition to comparing the performance of the existing and new approaches in terms of wall clock time, we will also report the accuracy of our calculation and memory usage.

3.1. Atomic Orbitals and the Sparsity of H and S

The electronic structure calculation code we used for the performance study is based on a local atomic orbital expansion scheme [15, 16]. We will refer to this scheme as the CGH scheme below. In the CGH scheme, an atomic orbital $\phi_\mu(\mathbf{r})$ is expressed as the product of a radial wave function $f_{\mu,l}(r)$ and a spherical harmonic $Y_{lm}(\hat{r})$, where $\mu = \{\alpha, i, \zeta, l, m\}$, and α, i, ζ, l, m represent the atom type, the index of an atom, the multiplicity of the radial functions, the angular momentum and the magnetic quantum number respectively. The radial function $f_{\mu,l}(r)$ is constructed as a linear combination of spherical Bessel functions within a cutoff radius r_c , i.e.,

$$f_{\mu,l}(r) = \begin{cases} \sum_q c_{\mu q} j_l(qr), & r < r_c \\ 0 & r \geq r_c. \end{cases} \quad (29)$$

where $j_l(qr)$ is a spherical Bessel function with q chosen to satisfy $j_l(qr_c)=0$, and the coefficients $c_{\mu q} j_l(qr)$ are chosen to minimize a ‘‘spillage factor’’ [29, 30] associated with a reference system that consists of a set of (4 or 5) dimers. We refer readers to Ref. [15, 16] for the details on the construction of the CGH local atomic orbitals.

The cutoff radius r_c determines the sparsity of the reduced Kohn-Sham Hamiltonian H and the overlap matrix S . The smaller the radius, the sparser H and S are. The cutoff radius for the atomic orbitals is set to 8.0 Bohr for B and N atoms in BNNT, and 6.0 Bohr for C atoms in CNT, respectively.

Another parameter that affects the dimension of H and S is the multiplicity ζ of the radial function $f_{\mu,l}(r)$. The multiplicity determines the number of basis functions per atom. A higher multiplicity results in larger number of basis functions per atom, which in turn results in more rows and columns in H and S . In our experiments, we used both single- ζ (SZ) orbitals and double- ζ plus polar orbitals (DZP). The number of local atomic orbitals is 4 for SZ and 13 for DZP.

We measure the sparsity by the percentage of the nonzero elements in the matrix H denoted by

$$H_{\text{nnz}}\% = \frac{\text{nnz}(H)}{N^2(H)} \times 100. \quad (30)$$

Here $\text{nnz}(H)$ is the number of nonzero elements of H and $N(H)$ is the dimension of H respectively. Since the computational cost of the selected inversion method is determined by the sparsity of $L + L^T$ for the Cholesky factor L

of $H - zS$, we will also report the percentage of the nonzero elements in the matrix $L + L^T$ (denoted by $L_{\text{nnz}}\%$) below. To reduce the amount of non-zero fill-in of L , we use the nested dissection (ND) technique [31] to reorder the sparse matrix $H - zS$ before it is factored. Fig. 4 (a) depicts the sparsity pattern of the H matrix associated with a 5120-atom BNNT (8,0) obtained from SZ atomic orbitals after it is reordered by ND. The sparsity pattern of $L + L^T$ for the corresponding Cholesky factor L of the same problem is shown in Fig. 4 (b).

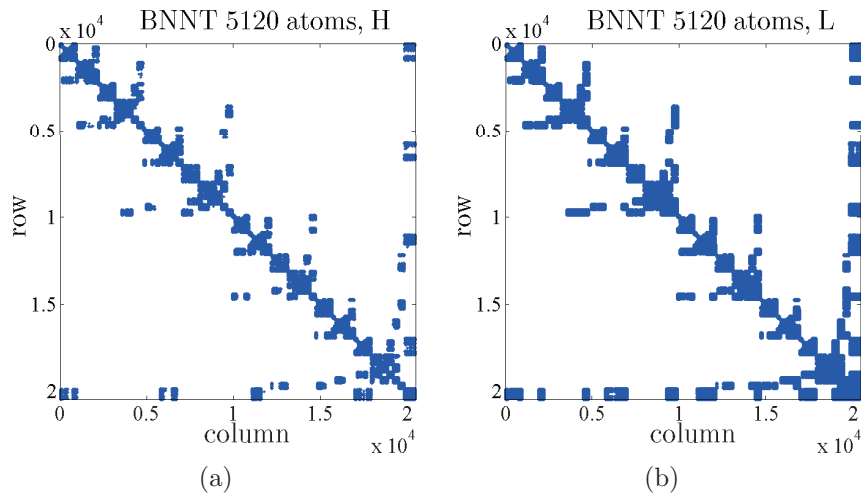


Figure 4: (color online) The sparsity pattern of H (a) and $L + L^T$ (b) for an 5120-atom BNNT (8,0) with SZ orbitals. Nested dissection reordering is used.

Table 1 shows the sparsity of Hamiltonian matrices associated with BNNT (8,0) and CNT (8,8) systems that consist of 64 to 10240 atoms. The Hamiltonians for these systems are constructed from SZ atomic orbitals. We report both the $H_{\text{nnz}}\%$ and $L_{\text{nnz}}\%$ values. We can clearly see from this table that H , and consequently L , are quite dense when the number of atoms in the nanotubes is relatively small (less than 512). This is due to fact that a large percentage of atoms in these small systems are within the r_c distance from each other. When the system size becomes larger (with more than 512 atoms), both $H_{\text{nnz}}\%$ and $L_{\text{nnz}}\%$ are inversely proportional to the system size. This is because for quasi-1D systems, the numerator in Eq. (30) scales linearly with respect to $N(H)$ for large $N(H)$. Hence, the resulting matrices

become increasingly sparse, thereby making the selected inversion method more favorable.

# Atoms		64	128	256	512	1024	1920	5120	10240
BNNT (8,0)	$H_{\text{nnz}}\%$	100.00	85.54	42.77	21.43	11.69	5.70	2.13	1.06
	$L_{\text{nnz}}\%$	100.00	99.48	77.94	46.13	25.07	13.70	5.26	2.64
CNT (8,8)	$H_{\text{nnz}}\%$	40.63	38.67	19.53	9.77	4.88	2.60	0.97	0.49
	$L_{\text{nnz}}\%$	69.92	68.45	68.70	54.38	31.75	17.54	7.42	3.79

Table 1: The percentage of nonzero elements $H_{\text{nnz}}\%$ and $L_{\text{nnz}}\%$ for BNNT (8,0) and CNT (8,8) of various sizes.

3.2. Performance comparison between diagonalization and selected inversion

We now compare the efficiency of selected inversion with that of diagonalization for computing the charge density in a single SCF iteration. In the existing code, the diagonalization of the matrix pencil (H, S) is performed by using the LAPACK subroutine `dsygv` when the code is run on a single processor. The selected inversion is performed by the `SelInv` software [7].

We use BNNT(8,0) and CNT(8,8) nanotubes of different lengths to study the scalability of the computation with respect to the number of atoms in the nanotube. The number of atoms in these tubes ranges from 64 to 10240.

Fig. 5 shows how the wall clock time used by `SelInv` compares with that used by `dsygv` for BNNT(8,0) of different sizes. When SZ atomic orbitals are used, `SelInv` takes almost the same amount of time as that used by `dsygv` for a BNNT with 64 atoms. When the number of atoms is larger than 64, `SelInv` is more efficient than `dsygv`. The cubic scaling of `dsygv` with respect to the number of atoms can be clearly seen from the slope of the blue loglog curve, which is approximately 3. The linear scaling of `SelInv`, which is indicated by the slope of the red curve, is evident when the number of atoms exceeds 200. For systems with less than 200 atoms, the wall clock time consumed by `SelInv` scales cubically with respect to the number of atoms also. This is due to the fact that the H and S matrices associated with these small systems are nearly dense. Similar observations can be made when the DZP atomic orbitals are used. In this case, `SelInv` is already more efficient than `dsygv` when the number of atoms is only 64. The linear scaling of `SelInv` can be observed when the number of atoms exceeds 128.

Fig. 6 shows the timing comparison between `SelInv` and `dsygv` for CNT (8,8) of different sizes. Because the cutoff radius for the carbon atom is chosen to be 6.0, which is smaller than that associated with the boron and nitrogen atoms, the H and S matrices associated with CNT (8,0) are sparser even when the number of atoms in the tube is relatively small. This explains why `SelInv` is already more efficient than `dsygv` already for a CNT with 64 atoms regardless whether SZ or DZP atomic orbitals are used. However, the linear scaling of `SelInv` timing with respect to the number of atoms does not show up until the number of atoms reaches 500. The increase in the crossover point is due to the fact that the sparsity of H is asymptotically determined by the number of atoms per unit length of the nanotube. Because the CNT (8,0) we use in our experiment has a large diameter, there are more atoms along the radial direction per unit length in CNT than that in BNNT. Consequently, it takes almost twice as many atoms for CNT to reach the same length along the longitudinal direction when compared to BNNT, as we can see from Fig. 2 and Fig. 3.

We should note here that it is possible to combine the PEPsi technique with a SZ atomic orbital based Kohn-Sham DFT solver to perform electron structure calculation on quasi-1D systems with more than 10,000 atoms. On the Hopper machine, the wall clock time used to perform a single selected inversion of the $H - zS$ matrix associated with a 5,120-atom BNNT(8,0) is 26.72 seconds. When the number of atoms increases to 10240, the wall clock time increases to 50.07 seconds. Similar performance is observed for CNT(8,8). It takes 47.59 seconds to perform a selected inversion for a 5120-atom CNT(8,8) tube, and 97.16 seconds for a 10240-atom tube.

3.3. Memory usage

We should also remark that the memory requirement for `SelInv` increases linearly with respect to the number of atoms when the nanotube reaches a certain size. For a nanotube that consists of 10240 atoms, the amount of memory required to store L and the selected elements of $[H - (z_l + \mu)S]^{-1}$ is 0.66 GB and 0.93 GB respectively. The relatively low memory requirement of `SelInv` for quasi-1D system suggests that the method may even be applicable to quasi-1D systems that contain more than 100,000 atoms on a single processor.

3.4. Accuracy

When selected inversion can be computed to high accuracy, which is often the case in practice, the only source of error introduced by the PEpSI technique comes from the limited number of terms in the pole expansion (14). The number of poles needed in (14) to achieve a desired level of accuracy in total energy (or Helmholtz free energy) and force is largely determined by the inverse temperature $\beta = 1/(k_B T)$ used in (4) and the spectrum width ΔE . Here we show that at room temperature $T = 300K$, the number of poles required to provide an accurate pole expansion approximation is modest even for a metallic system such as CNT(8,8). Table 2 shows that when diagonalization is replaced by PEpSI for a single Γ -point calculation, the errors in total energy and force decrease as the number of poles in (14) increases. When the number of poles reaches 80, the difference between the final total energies produced by the existing code and the modified code (which replaces diagonalization with PEpSI) is 3.6×10^{-7} eV. The difference in the mean absolute error (MAE) is 2×10^{-6} eV/Angstrom, which is quite small for all practical purposes.

# Poles	$E_{\text{PEpSI}} - E_{\text{ref}}$ (eV)	MAE Force (eV/Angstrom)
20	5.868351108	0.400431
40	0.007370583	0.001142
60	0.000110382	0.000026
80	0.000000360	0.000002

Table 2: The difference between the total energy and atomic force produced by the existing electronic structure code and modified version in which diagonalization is replaced by PEpSI. The difference in atomic force is measured in terms of the mean absolute error (MAE).

4. Conclusion

In this paper, we generalized the recently developed pole expansion and selected inversion technique (PEpSI) for solving finite dimensional Kohn-Sham equations obtained from an atomic orbital expansion. We gave expressions for evaluating electron density, total energy, Helmholtz free energy and atomic forces without using eigenvalues and eigenvectors of a Kohn-Sham Hamiltonian. These expressions are derived from an FOE approximation

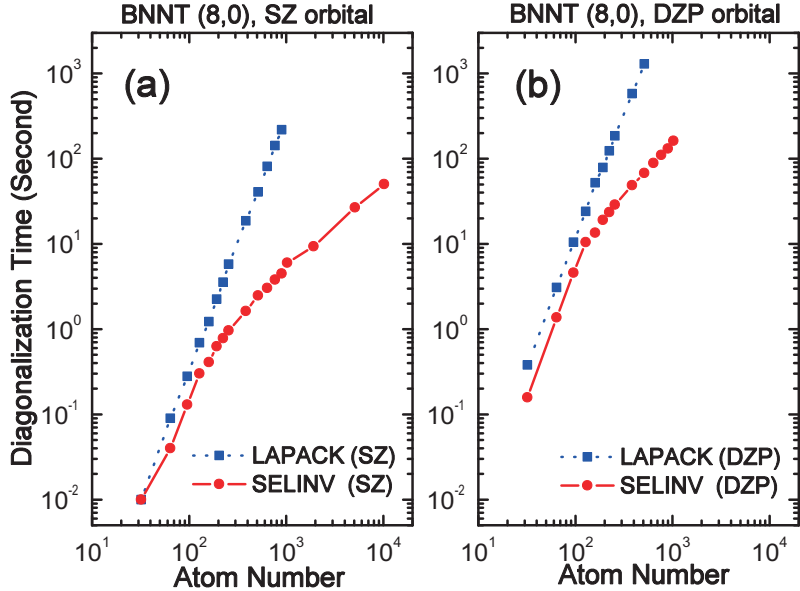


Figure 5: (color online) Comparisons of the wallclock time used by selected inversion (at one pole) required for PEPsi and by the LAPACK `dsygv` used to diagonalize a Kohn-Sham Hamiltonian associated with BNNT (8,0). The Hamiltonians are constructed from SZ orbitals (4 basis per atom) in (a) and DZP orbitals (13 basis per atom) in (b).

to the Fermi-Dirac function using an efficient and accurate pole expansion technique. They only use selected elements of the reduced density matrix, reduced energy density matrix and reduced free energy density matrix. These selected elements can be obtained from computing selected elements of the inverse of a shifted Kohn-Sham Hamiltonian through the selected inversion technique. The complexity of selected inversion is $\mathcal{O}(N_e)$ for quasi-1D systems such as nanorods, nanotubes and nanowires, $\mathcal{O}(N_e^{3/2})$ for quasi-2D systems such as graphene and surfaces, and $\mathcal{O}(N_e^2)$ for 3D bulk systems. It compares favorably to the complexity of diagonalization, which is $\mathcal{O}(N_e^3)$. We reported the performance gain we can achieve by comparing the efficiency of PEPsi with that of diagonalization on two types of nanotubes. The lin-

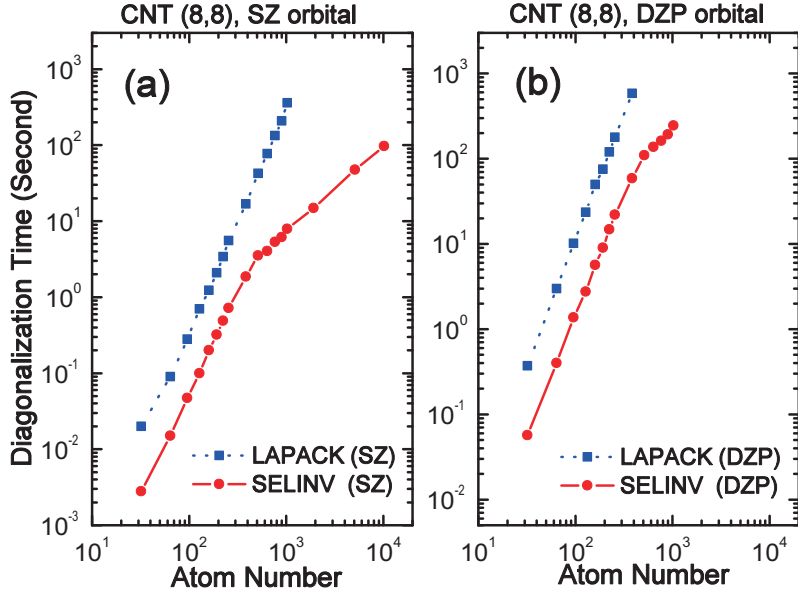


Figure 6: (color online) Comparisons of the wallclock time by selected inversion (at one pole) required for PEPsi and by the LAPACK `dsygv` used to diagonalize a Kohn-Sham Hamiltonian associated with CNT(8,8). The Hamiltonians are constructed from SZ orbitals (4 basis per atom) in (a) and DZP orbitals (13 basis per atom) in (b).

ear scaling of PEPsi with respect to the number of atoms is clear when the number of atoms in these quasi-1D systems is larger than a few hundreds. Even when these nanotubes contain fewer than a hundred atoms, PEPsi still appears to outperform a diagonalization based DFT calculation. For quasi-2D and quasi-3D systems, we expect the crossover point over which PEPsi exhibits $\mathcal{O}(N_e^{3/2})$ and $\mathcal{O}(N_e^2)$ scaling to be much larger. However, based on the experiments presented here, PEPsi may still be more efficient than diagonalization (before the crossover point is reached) as long as the Cholesky factors of the shifted Kohn-Sham Hamiltonian are not completely dense.

The computational experiments we presented above were performed on a single processor. For quasi-1D systems such as nanotubes, the use PEPsi

allows us to tackle problems that contain as many as 10,000 atoms. This cannot be done by using a diagonalization based approach. For quasi-2D and 3D systems, a parallel implementation of the PEpSI, which we are currently working on, is required to solve problems with that many atoms. We will report the performance for these large-scale calculations in a future publication.

Acknowledgment: This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under the U.S. Department of Energy contract number DE-AC02-05CH11231 (L. L. and C. Y.), and by the Chinese National Natural Science Funds for Distinguished Young Scholars (L. H.).

Appendix A.

Derivation of Eq. (14):

Ξ is a diagonal matrix, and the pole expansion (13) can be applied to each component of Ξ as

$$f_\beta(\Xi - \mu) \approx \Im \sum_{l=1}^P \frac{\omega_l^\rho}{\Xi - (z_l + \mu)I},$$

where I is an $N \times N$ identity matrix. Using Eq. (12), the single particle density matrix can be written as

$$\begin{aligned} \hat{\gamma}(x, x') &\approx \Phi(x) C \Im \sum_{l=1}^P \frac{\omega_l^\rho}{\Xi - z_l I} C^T \Phi^T(x') \\ &= \Phi(x) \Im \sum_{l=1}^P \frac{\omega_l^\rho}{C^{-T} \Xi C^{-1} - z_l C^{-T} C^{-1}} \Phi^T(x'). \end{aligned} \quad (\text{A.1})$$

Since the generalized eigenvalue problem (10) implies the identity

$$C^T H C = \Xi, \quad C^T S C = I, \quad (\text{A.2})$$

the single particle density matrix takes the form

$$\hat{\gamma}(x, x') \approx \Phi(x) \Im \sum_{l=1}^P \frac{\omega_l^\rho}{H - (z_l + \mu)S} \Phi^T(x') \quad (\text{A.3})$$

which is Eq. (14).

Derivation of Eq. (21):

The first term in the Helmholtz free energy functional is

$$\begin{aligned}\mathrm{Tr}[f_\beta^{\mathcal{F}}(\Xi - \mu)] &= \mathrm{Tr}[C f_\beta^{\mathcal{F}}(\Xi - \mu) C^T C^{-T} C^{-1}] \\ &\equiv \mathrm{Tr}[\gamma^{\mathcal{F}} S].\end{aligned}\tag{A.4}$$

The second equal sign in Eq. (A.4) defines the reduced free energy density matrix $\gamma^{\mathcal{F}}$, which can be evaluated using the pole expansion (20) as

$$\begin{aligned}\gamma^{\mathcal{F}} &\approx C \Im \sum_{l=1}^P \frac{\omega_l^{\mathcal{F}}}{\Xi - z_l I} C^T \\ &= \Im \sum_{l=1}^P \frac{\omega_l^{\mathcal{F}}}{C^{-T} H C^{-1} - z_l C^{-T} C^{-1}} \\ &= \Im \sum_{l=1}^P \frac{\omega_l^{\mathcal{F}}}{H - z_l S},\end{aligned}\tag{A.5}$$

which is Eq. (21).

Derivation of Eq. (23):

The atomic force is in general given by the derivative of the Helmholtz free energy with respect to the atomic positions. Using the representation of the Helmholtz free energy in Eq. (18), and the fact that

$$(f_\beta^{\mathcal{F}})'(z) = f_\beta(z), \quad N_e = \mathrm{Tr}[f_\beta(\Xi - \mu)],\tag{A.6}$$

it can be derived that

$$\begin{aligned}
F_I &= -\frac{\partial}{\partial R_I} \mathcal{F}_{\text{tot}} = -\frac{\partial}{\partial R_I} (\text{Tr}[f_\beta^{\mathcal{F}}(\Xi - \mu)] - \mu N_e) \\
&= -\text{Tr} \left[(f_\beta^{\mathcal{F}})'(\Xi - \mu) \left(\frac{\partial \Xi}{\partial R_I} - \frac{\partial \mu}{\partial R_I} \right) \right] + N_e \frac{\partial \mu}{\partial R_I} \\
&= -\text{Tr} \left[f_\beta(\Xi - \mu) \frac{\partial \Xi}{\partial R_I} \right] + \frac{\partial \mu}{\partial R_I} (N_e - \text{Tr}[f_\beta(\Xi - \mu)]) \\
&= -\text{Tr} \left[f_\beta(\Xi - \mu) C^T \frac{\partial H}{\partial R_I} C \right] - \text{Tr} \left[f_\beta(\Xi - \mu) \frac{\partial C^T}{\partial R_I} H C \right] \\
&\quad - \text{Tr} \left[f_\beta(\Xi - \mu) C^T H \frac{\partial C}{\partial R_I} \right] \\
&= -\text{Tr} \left[\gamma \frac{\partial H}{\partial R_I} \right] - \text{Tr} \left[f_\beta(\Xi - \mu) \frac{\partial C^T}{\partial R_I} H C \right] \\
&\quad - \text{Tr} \left[f_\beta(\Xi - \mu) C^T H \frac{\partial C}{\partial R_I} \right]
\end{aligned} \tag{A.7}$$

The second and the third terms in Eq. (A.7) come from the nonorthogonality of the basis functions and should be further simplified. We have

$$\begin{aligned}
&\text{Tr} \left[f_\beta(\Xi - \mu) \frac{\partial C^T}{\partial R_I} H C \right] + \text{Tr} \left[f_\beta(\Xi - \mu) C^T H \frac{\partial C}{\partial R_I} \right] \\
&= \text{Tr} \left[(C^{-T} C^{-1}) [C(C^T H C) f_\beta(\Xi - \mu) C^T] (C^{-T} C^{-1}) C \frac{\partial C^T}{\partial R_I} \right] \\
&\quad + \text{Tr} \left[C^{-T} C^{-1} [C f_\beta(\Xi - \mu) (C^T H C) C^T] C^{-T} C^{-1} \frac{\partial C}{\partial R_I} C^T \right] \\
&\equiv \text{Tr} \left[(C \Xi f_\beta(\Xi - \mu) C^T) \left(S C \frac{\partial C^T}{\partial R_I} S + S \frac{\partial C}{\partial R_I} C^T S \right) \right].
\end{aligned} \tag{A.8}$$

Define the reduced energy density matrix as in Eq. (24), and Eq. (A.8) can be simplified as

$$\begin{aligned}
&\text{Tr} \left[\gamma^E S \left(C \frac{\partial C^T}{\partial R_I} + \frac{\partial C}{\partial R_I} C^T \right) S \right] \\
&= \text{Tr} \left[\gamma^E S \frac{\partial S^{-1}}{\partial R_I} S \right] = -\text{Tr} \left[\gamma^E \frac{\partial S}{\partial R_I} \right]
\end{aligned} \tag{A.9}$$

Combining Eq. (A.9) and Eq. (A.7), we have

$$F_I = -\frac{\partial \mathcal{F}}{\partial R_I} = -\text{Tr} \left[\gamma \frac{\partial H}{\partial R_I} \right] + \text{Tr} \left[\gamma^E \frac{\partial S}{\partial R_I} \right]. \quad (\text{A.10})$$

which proves Eq. (23).

- [1] L. Lin, J. Lu, L. Ying, W. E, Pole-based approximation of the Fermi-Dirac function, *Chinese Ann. Math.* 30B (2009) 729.
- [2] S. Baroni, P. Giannozzi, Towards very large-scale electronic-structure calculations, *Europhys. Lett.* 17 (6) (1992) 547.
- [3] T. Ozaki, Continued fraction representation of the Fermi-Dirac function for large-scale electronic structure calculations, *Phys. Rev. B* 75 (2007) 035123.
- [4] M. Ceriotti, T. Kühne, M. Parrinello, An efficient and accurate decomposition of the Fermi operator., *J. Chem. Phys.* 129 (2008) 024707.
- [5] T. Ozaki, Efficient low-order scaling method for large-scale electronic structure calculations with localized basis functions, *Phys. Rev. B* 82 (2010) 075131.
- [6] L. Lin, J. Lu, L. Ying, R. Car, W. E, Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems, *Comm. Math. Sci.* 7 (2009) 755.
- [7] L. Lin, C. Yang, J. Meza, J. Lu, L. Ying, W. E, SelInv – An algorithm for selected inversion of a sparse symmetric matrix, *ACM. Trans. Math. Software* 37 (2010) 40.
- [8] L. Lin, C. Yang, J. Lu, L. Ying, W. E, A fast parallel algorithm for selected inversion of structured sparse matrices with application to 2D electronic structure calculations, *SIAM J. Sci. Comput.* 33 (2011) 1329.
- [9] T. Arias, Multiresolution analysis of electronic structure: semicardinal and wavelet bases, *Rev. Mod. Phys.* 71 (1999) 267.
- [10] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, R. Schneider, Daubechies wavelets as a basis set for density functional pseudopotential calculations, *J. Chem. Phys.* 129 (2008) 014109.

- [11] L. Lin, J. Lu, L. Ying, W. E, Adaptive local basis set for Kohn-Sham density functional theory in a discontinuous Galerkin framework I: Total energy calculation, *J. Comput. Phys.* 231 (2012) 2140.
- [12] M. Frisch, J. Pople, J. Binkley, Self-consistent molecular orbital methods 25. supplementary functions for gaussian basis sets, *J. Chem. Phys.* 80 (1984) 3265.
- [13] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, QUICKSTEP: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach, *Comput. Phys. Commun.* 167 (2005) 103.
- [14] J. Junquera, O. Paz, D. Sanchez-Portal, E. Artacho, Numerical atomic orbitals for linear-scaling calculations, *Phys. Rev. B* 64 (2001) 235111.
- [15] M. Chen, G. C. Guo, L. He, Systematically improvable optimized atomic basis sets for ab initio calculations, *J. Phys.: Condens. Matter* 22 (2010) 445501.
- [16] M. Chen, G. C. Guo, L. He, Electronic structure interpolation via atomic orbitals, *J. Phys.: Condens. Matter* 23 (2011) 325501.
- [17] S. D. Kenny, A. P. Horsfield, H. Fujitani, Transferable atomic-type orbital basis sets for solids, *Phys. Rev. B* 62 (2000) 4899.
- [18] T. Ozaki, Variationally optimized atomic orbitals for large-scale electronic structures, *Phys. Rev. B* 67 (2003) 155108.
- [19] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, Ab initio molecular simulations with numeric atom-centered orbitals, *Comput. Phys. Commun.* 180 (2009) 2175.
- [20] S. Goedecker, Integral representation of the fermi distribution and its applications in electronic-structure calculations, *Phys. Rev. B* 48 (1993) 17573.
- [21] L. Lin, J. Lu, R. Car, W. E, Multipole representation of the Fermi operator with application to the electronic structure analysis of metallic systems, *Phys. Rev. B* 79 (2009) 115133.

- [22] D. M. Ceperley, B. J. Alder, Ground state of the electron gas by a stochastic method, *Phys. Rev. Lett.* 45 (1980) 566.
- [23] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A* 38 (1988) 3098.
- [24] C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B* 37 (1988) 785.
- [25] N. Mermin, Thermal properties of the inhomogeneous electron gas, *Phys. Rev.* 137 (1965) A1441.
- [26] A. Alavi, J. Kohanoff, M. Parrinello, D. Frenkel, Ab initio molecular dynamics with excited electrons, *Phys. Rev. Lett.* 73 (1994) 2599.
- [27] S. Goedecker, Linear scaling electronic structure methods, *Rev. Mod. Phys.* 71 (1999) 1085.
- [28] J.-C. Charlier, X. Blase, S. Roche, Electronic and transport properties of nanotubes, *Rev. Mod. Phys.* 79 (2007) 677.
- [29] D. Sánchez-Portal, E. Artacho, J. M. Soler, Projection of plane-wave calculations into atomic orbitals, *Solid State Commun.* 95 (1995) 685.
- [30] D. Sánchez-Portal, E. Artacho, J. M. Soler, Analysis of atomic orbital basis sets from the projection of plane-wave results, *J. Phys.: Condens. Matter* 8 (1996) 3859.
- [31] A. George, Nested dissection of a regular finite element mesh, *SIAM J. Numer. Anal.* 10 (1973) 345.