IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Accelerating atomic orbital-based electronic structure calculation via pole expansion and selected inversion

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2013 J. Phys.: Condens. Matter 25 295501

(http://iopscience.iop.org/0953-8984/25/29/295501)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 92.151.153.153
The article was downloaded on 27/06/2013 at 08:27

Please note that terms and conditions apply.

J. Phys.: Condens. Matter **25** (2013) 295501 (14pp) doi:10.1088/0953-8984/25/29/295501

Accelerating atomic orbital-based electronic structure calculation via pole expansion and selected inversion

Lin Lin[1], **Mohan Chen**[2], **Chao Yang**[1] **and Lixin He**[2]

[1] Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[2] Key Laboratory of Quantum Information, CAS, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China

Received 31 January 2013, in final form 5 June 2013
Published 26 June 2013
Online at stacks.iop.org/JPhysCM/25/295501

Abstract

We describe how to apply the recently developed pole expansion and selected inversion (PEXSI) technique to Kohn–Sham density function theory (DFT) electronic structure calculations that are based on atomic orbital discretization. We give analytic expressions for evaluating the charge density, the total energy, the Helmholtz free energy and the atomic forces (including both the Hellmann–Feynman force and the Pulay force) 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 PEXSI is that it has a computational complexity much lower than that associated with the matrix diagonalization procedure. We demonstrate the performance gain by comparing the timing of PEXSI with that of diagonalization on insulating and metallic nanotubes. For these quasi-1D systems, the complexity of PEXSI 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 PEXSI are modest. This even makes it possible to perform Kohn–Sham DFT calculations for 10 000-atom nanotubes with a sequential implementation of the selected inversion algorithm. We also perform an accurate geometry optimization calculation on a truncated (8, 0) boron nitride nanotube system containing 1024 atoms. Numerical results indicate that the use of PEXSI does not lead to loss of the accuracy required in a practical DFT calculation.

(Some figures may appear in colour only in the online journal)

1. Introduction

Electronic structure calculations based on solving the Kohn–Sham density functional theory (KSDFT) play an important role in the analysis of electronic, structural and optical properties of molecules, solids and other nanostructures. 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.

Linear scaling algorithms (or $\mathcal{O}(N_e)$ scaling methods; see for example [1–6], and review articles [7, 8]) are attractive alternatives for solving KSDFT. The traditional linear scaling methods use the nearsightedness principle, which asserts that the density perturbation induced by a local change in the external potential decays exponentially away from where the perturbation is applied. Consequently, the off-diagonal elements of the density matrix decay exponentially away from the diagonal [9, 10]. Strictly speaking, the nearsightedness

© 2013 IOP Publishing Ltd Printed in the UK & the USA

property is valid for insulating systems but not for metallic systems.

In order to design a fast algorithm that is accurate for both insulating and metallic systems, we use an equivalent formulation of KSDFT, in which the charge density is evaluated 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 [11], 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* [11] 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, and ΔE is the spectral width of the Kohn–Sham Hamiltonian (i.e. the difference between the largest and the smallest eigenvalues). This number of expansion terms, or the pole count, here is significantly lower than those given in the previous approaches [12–16]. The favorable scaling of the pole expansion allows us to treat both insulating and metallic systems efficiently at room temperature or even lower temperature.

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 [17–19]. The idea of using the inverse of a shifted Hamiltonian operator (Green's function) for reducing the complexity of Kohn–Sham density functional theory has also been pursued in other recent works [16, 20]. In the selected inversion method, 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 [18, 19], we used the pole expansion and selected inversion (PEXSI) technique to solve the Kohn–Sham problem discretized by a finite difference scheme. However, it is worth pointing out that PEXSI 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 PEXSI can be used to speed up the solution of 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 PEXSI.

We demonstrate the performance gain that we can achieve by comparing PEXSI with the LAPACK diagonalization subroutine `dsygv` on nanotubes of two types. We show that by using the PEXSI technique, it is possible to perform electronic structure calculations accurately for a nanotube that contains 10 000 atoms with a sequential implementation of the selected inversion algorithm within a reasonable amount of time. This is not possible with the sequential LAPACK subroutine. For this example, PEXSI exhibits linear scaling when the system size exceeds a few hundred atoms.

This paper is organized as follows. In section 2, we show how the PEXSI technique previously developed [11, 17–19] can be extended to solve 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 PEXSI on two quasi-1D test problems.

Throughout the paper, we use $\mathfrak{Im}(A)$ 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 $\mathrm{diag}(A)$. We use $\hat{H}(x,x')$ to denote the Hamiltonian operator, and H, S to denote the discretized Hamiltonian matrix and the corresponding overlap matrix obtained from a basis set Φ. Similarly $\hat{\gamma}(x,x')$ denotes the single-particle density matrix operator, and the corresponding electron density is denoted by $\hat{\rho}(x)$. The matrix Γ denotes the single-particle density matrix represented under a basis set Φ. It will be used to define the electron density $\hat{\rho}$ and the total energy E_{tot}. In a finite temperature *ab initio* molecular dynamics simulation, we also need the Helmholtz free energy $\mathcal{F}_{\mathrm{tot}}$, and the atomic forces on the nuclei $\{F_I\}$. To compute these quantities without using Kohn–Sham eigenvalues and Kohn–Sham orbitals, we need the free energy density matrix $\Gamma^{\mathcal{F}}$ and the energy density matrix Γ^{E}. In PEXSI, these matrices are approximated by finite P-term pole expansions, denoted by Γ_P, $\Gamma_P^{\mathcal{F}}$, Γ_P^{E} respectively. However, to simplify notation, we will drop the subscript P and simply use $\Gamma, \Gamma^{\mathcal{F}}, \Gamma^{E}$ to denote the approximated matrices unless otherwise noted.

2. Theory

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

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

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

$$\int \psi_i^*(x)\psi_j(x)\,\mathrm{d}x = \delta_{ij}, \tag{2}$$

and the eigenvalue ε_i is often known as the ith Kohn–Sham energy level. Using the Kohn–Sham orbitals, we can define the charge density by

$$\hat{\rho}(x) = \sum_i^{\infty} |\psi_i(x)|^2 f_i, \qquad i = 1, 2, \ldots, \infty, \tag{3}$$

with occupation numbers $0 \leq f_i \leq 2$, $i = 1, 2, \ldots \infty$. The occupation numbers in (3) can be chosen according to the Fermi–Dirac distribution function

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

where μ is the chemical potential chosen to ensure that

$$\int \hat{\rho}(x)\,\mathrm{d}x = N_\mathrm{e}, \tag{5}$$

and β is the inverse of the temperature, i.e., $\beta = 1/(k_\mathrm{B}T)$ with k_B being the Boltzmann constant.

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

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

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

$$\mathrm{Tr}[\hat{\gamma}(x, x')] = N_\mathrm{e}, \tag{7}$$

where Tr denotes the trace of an operator.

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

$$\hat{\rho}(x) = \mathrm{diag}(f_\beta(\hat{H}[\hat{\rho}(x)] - \mu\delta(x, x'))), \tag{8}$$

where μ is chosen to satisfy (7). The algorithm most widely used 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 $\hat{\rho}(x) = \hat{\gamma}(x, x)$ in (8).

2.1. Basis expansion in 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 methods etc). In this paper, we focus on a discretization scheme in which a Kohn–Sham orbital ψ_i is expanded as a linear combination of a finite number of basis functions $\{\varphi_j\}$, i.e.,

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

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 include Gaussian-type orbitals [21, 22] and local atomic orbitals [23–28], adaptive curvilinear coordinates [29], optimized nonorthogonal orbitals [1–3] and adaptive local basis functions [30]. 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

$$HC = SC\Xi, \tag{10}$$

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 equation (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 $\Psi = [\psi_1, \ldots, \psi_N]$ and $\Phi = [\varphi_1, \ldots, \varphi_N]$; then equation (9) can be written in a compact form:

$$\Psi = \Phi C. \tag{11}$$

Consequently, the single-particle density matrix (6) becomes [2]

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

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 [13]. 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 [13], rational expansion [11, 12, 14], and a hybrid scheme in which both polynomials and rational functions are used [15, 31]. 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. An upper bound of ΔE can be obtained inexpensively via a very small number of Lanczos steps [32].

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 [11] is particularly promising since it requires only $\mathcal{O}(\log\beta\Delta E)$ terms of simple rational functions. The favorable scaling of the pole expansion allows us to treat both insulating and metallic systems efficiently at room or even lower temperature. The pole expansion has the analytic expression

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

where

$$w_l^\rho = \frac{4K\sqrt{mM}}{\pi k P} \frac{\mathrm{cn}(t_l)\mathrm{dn}(t_l)}{z_l(k^{-1}-\mathrm{sn}(t_l))^2} f_\beta(z_l), \tag{14}$$

with $m = \frac{\pi^2}{\beta^2}, M = \Delta E^2 + \frac{\pi^2}{\beta^2}, k = \frac{\sqrt{M/m}-1}{\sqrt{M/m}+1}$. The functions cn, dn, sn are Jacobi elliptic functions, and $K, \{z_l\}, \{t_l\}$ are chosen carefully and computed from analytic expressions. We refer the readers to [11] for more detailed explanations. In the following discussions, we will also refer to $\{z_l\}$ as the complex shifts or *poles*, and refer to $\{\omega_l^\rho\}$ as the complex weights. The complex shifts and weights are determined only by β, ΔE and the number of poles P. All quantities in the pole expansion are known explicitly and their calculation takes a negligible amount of time. The construction of pole expansion is based on the observation that the non-analytic part of the Fermi–Dirac function lies only on the imaginary axis within $\left[\frac{i\pi}{\beta}, +i\infty\right] \bigcup \left[-i\infty, -\frac{i\pi}{\beta}\right]$. A dumbbell-shaped Cauchy contour (see figure 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 free energy density matrix and the energy density matrix which are discussed in section 2.3 for the purpose of computing Helmholtz free energy and atomic forces (including both the Hellmann–Feynman force and the Pulay force). In such cases, one only needs to substitute f_β in the weight function in equation (14) by the corresponding function that shares the same analytic structure as the Fermi–Dirac function f_β.

Following the derivation in the appendix, we can use (13) to approximate the single-particle density matrix $\hat{\gamma}$ by its P-term pole expansion, denoted by $\hat{\gamma}_P$ as

$$\begin{aligned} \hat{\gamma}_P(x,x') &= \Phi(x)\mathfrak{Im}\left(\sum_{l=1}^{P} \frac{\omega_l^\rho}{H-(z_l+\mu)S}\right)\Phi^T(x') \\ &\equiv \Phi(x)\Gamma\Phi^T(x'). \end{aligned} \tag{15}$$

In the above expression, Γ is an $N \times N$ matrix represented in terms of the atomic orbitals Φ. To simplify our notation, we will drop the subscript P from the P-term pole expansion approximation of the single-particle density matrix $\hat{\gamma}$ unless otherwise noted. A similar treatment will be performed for the electron density $\hat{\rho}$, the total energy E_{tot}, the Helmholtz free energy $\mathcal{F}_{\mathrm{tot}}$, and the atomic force on the Ith nuclei F_I. Using equation (15), we can evaluate the electron density in the real space as the diagonal elements of $\hat{\gamma}$, i.e.,

$$\hat{\rho}(x) = \Phi(x)\Gamma\Phi^T(x) = \sum_{ij} \Gamma_{ij}\varphi_j(x)\varphi_i(x). \tag{16}$$

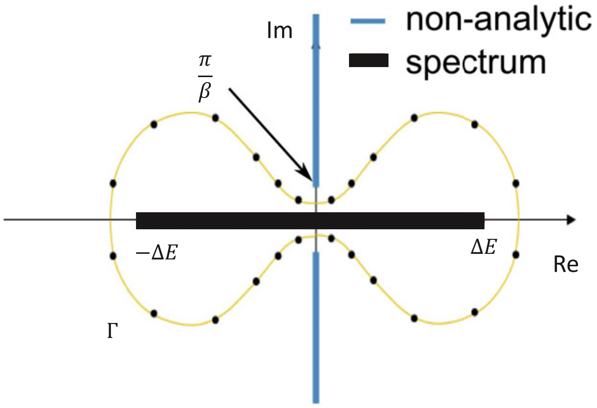

Figure 1. A schematic view of the placement of poles used in a pole expansion approximation of $f_\beta(z)$. The thick black line on the real axis indicates the range of ε_i–μ, and the thin blue line on the imaginary axis indicates the non-analytic part of $f_\beta(z)$. The yellow dumbbell-shaped contour 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.

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} 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 [17–19] 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 (16) 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 the 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. In particular, the selected inversion algorithm does not rely on any localization property of A^{-1}. However, it can be combined with localization properties of insulating systems to further reduce the computational cost. We will pursue this approach in future work. We also remark that the PEXSI technique can be applied whenever H and S are sparse matrices. However, since the selected inversion method relies on an LDL^T factorization of $H - zS$, the preconstant of the selected inversion method asymptotically scales cubically with respect to the number of basis functions per atom. The numbers of basis functions or degrees of freedom per atom associated with the finite difference method [33] and the finite element method [34] are usually much larger than those associated with methods based on contracted basis functions such as local atomic orbitals. Therefore the finite difference method and the finite element method do not benefit as much from the PEXSI technique as methods that are based on local atomic orbitals.

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

In addition to reducing the computational complexity of the charge density calculation in each SCF iteration, the PEXSI technique can also be used to compute the total energy, the Helmholtz free energy and the atomic forces (including both the Hellmann–Feynman force and the Pulay force) efficiently without diagonalizing the Kohn–Sham Hamiltonian.

It is well known that equations (1)–(5) can be derived as the first-order necessary condition for minimizing the Mermin free energy [35–39]

$$\mathcal{F}_{\mathrm{tot}}[\{\psi_i\},\{f_i\}] = E_{\mathrm{tot}}[\{\psi_i\},\{f_i\}] - TS[\{f_i\}], \tag{17}$$

under the constraints (2) and $\sum_{i=1}^{\infty} f_i = N_{\mathrm{e}}$, where

$$E_{\mathrm{tot}}[\{\psi_i\},\{f_i\}] = \sum_{i=1}^{\infty} f_i\varepsilon_i - \frac{1}{2}\iint \frac{\hat\rho(x)\hat\rho(y)}{|x-y|}\,\mathrm{d}x\,\mathrm{d}y + E_{\mathrm{xc}}[\hat\rho] - \int V_{\mathrm{xc}}[\hat\rho](x)\hat\rho(x)\,\mathrm{d}x \tag{18}$$

is called the internal energy or the total energy, and

$$S[\{f_i\}] = -2k_{\mathrm{B}}\sum_{i=1}^{\infty}(\tilde f_i \log \tilde f_i + (1-\tilde f_i)\log(1-\tilde f_i)) \tag{19}$$

is the entropy due to fractional occupation where $\tilde f_i = f_i/2$ is used so that $0 \le \tilde f_i \le 1$. The chemical potential μ in (4) is simply the Lagrange multiplier associated with the occupation number constraint $\sum_{i=1}^{\infty} f_i = N_{\mathrm{e}}$.

Furthermore, it is the derivative of the Mermin free energy (rather than the total energy) with respect to the atomic positions that gives rise to the correct force in *ab initio* molecular dynamics simulation [36–39].

The evaluation of the Mermin free energy functional $\mathcal{F}_{\mathrm{tot}}$ requires explicit knowledge of the Kohn–Sham eigenvalues $\{\varepsilon_i\}$ which are not available in the PEXSI scheme. However, it has been shown in [40] that the Mermin free energy can be equivalently computed in the form of the following Helmholtz free energy, which does not contain the Kohn–Sham eigenvalues explicitly:

$$\mathcal{F}_{\mathrm{tot}} = -2\beta^{-1}\mathrm{Tr}\ln(1+\exp(\beta(\mu-\Xi))) + \mu N_{\mathrm{e}} - \frac{1}{2}\iint \frac{\hat\rho(x)\hat\rho(y)}{|x-y|}\,\mathrm{d}x\,\mathrm{d}y + E_{\mathrm{xc}}[\hat\rho] - \int V_{\mathrm{xc}}[\hat\rho](x)\hat\rho(x)\,\mathrm{d}x. \tag{20}$$

Here we assume that the LDA [41] or GGA [42, 43] exchange–correlation functional is used for the Kohn–Sham total energy expression. In section 2.2 we have shown that the electron density $\hat\rho(x)$ can be computed in the PEXSI scheme. Therefore in equation (20), only the first term requires extra treatment. Note that the function

$$f_\beta^{\mathcal{F}}(\varepsilon-\mu) = -2\beta^{-1}\ln(1+\exp(\beta(\mu-\varepsilon))) \tag{21}$$

is different from the Fermi–Dirac function f_β in equation (4). In fact $f_\beta^{\mathcal{F}}$ is directly related to f_β as

$$(f_\beta^{\mathcal{F}})'(z) = f_\beta(z). \tag{22}$$

Nonetheless $f_\beta^{\mathcal{F}}(z)$ is analytic everywhere in the complex plane, except for segments of the imaginary axis within $\left[\frac{\mathrm{i}\pi}{\beta}, +\mathrm{i}\infty\right] \bigcup \left[-\mathrm{i}\infty, -\frac{\mathrm{i}\pi}{\beta}\right]$. In this sense, $f_\beta^{\mathcal{F}}$ has the same analytic structure as 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.,

$$f_\beta^{\mathcal{F}}(\varepsilon-\mu) \approx \mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\mathcal{F}}}{\varepsilon-(z_l+\mu)}. \tag{23}$$

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

$$\mathcal{F}_{\mathrm{tot}} = \mathrm{Tr}[\Gamma^{\mathcal{F}}S] + \mu N_{\mathrm{e}} - \frac{1}{2}\iint \frac{\hat\rho(x)\hat\rho(y)}{|x-y|}\,\mathrm{d}x\,\mathrm{d}y + E_{\mathrm{xc}}[\hat\rho] - \int V_{\mathrm{xc}}[\hat\rho]\hat\rho(x)\,\mathrm{d}x, \tag{24}$$

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

$$\Gamma^{\mathcal{F}} = \mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\mathcal{F}}}{H-(z_l+\mu)S}. \tag{25}$$

Note that in the expression (24), the first term depends on the trace of the product of $\Gamma^{\mathcal{F}}$ and S. The computation of this term requires only the (i,j)th entry of $\Gamma^{\mathcal{F}}$ for (i,j) satisfying $S_{ij} \neq 0$ or $H_{ij} \neq 0$. Since the poles $\{z_l\}$ are the same as those used for computing the electron density, the selected elements of $\Gamma^{\mathcal{F}}$ correspond to the same selected elements of $(H-(z_l+\mu)S)^{-1}$ as are used for the charge density calculation. Thus using them for computing $\mathcal{F}_{\mathrm{tot}}$ does not introduce additional complexity.

It is worth mentioning that the above formulation can be simplified for insulating systems with a relatively large band gap (even at zero temperature). In such cases, f_i can be chosen to be 2 for occupied states and 0 for unoccupied states. Then the entropy term S vanishes and $\mathcal{F}_{\mathrm{tot}} = E_{\mathrm{tot}}$. Furthermore, like for the Helmholtz free energy, we can give an alternative expression for E_{tot}, as follows:

$$E_{\mathrm{tot}} = \mathrm{Tr}[\Gamma H] - \frac{1}{2}\iint \frac{\hat\rho(x)\hat\rho(y)}{|x-y|}\,\mathrm{d}x\,\mathrm{d}y + E_{\mathrm{xc}}[\hat\rho] - \int V_{\mathrm{xc}}[\hat\rho](x)\hat\rho(x)\,\mathrm{d}x, \tag{26}$$

where Γ is the 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.

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 free energy with respect to the position of an atom. For a nonorthogonal atomic basis set, the force calculation is not trivial, and standard methods have been established in [36] for calculating the force. The calculation includes both the Hellmann–Feynman force and the Pulay force [44], where the Pulay force is induced by the change of basis functions with respect to atomic positions. Following the derivation in the appendix, we can express the atomic force associated with the Ith atom in a compact way as

$$F_I = -\frac{\partial \mathcal{F}_{\mathrm{tot}}}{\partial R_I} = -\mathrm{Tr}\left[\Gamma \frac{\partial H}{\partial R_I}\right] + \mathrm{Tr}\left[\Gamma^E \frac{\partial S}{\partial R_I}\right], \tag{27}$$

where Γ^E is the energy density matrix defined by

$$\Gamma^E = C \Xi f_\beta(\Xi - \mu) C^T. \tag{28}$$

We remark that equation (27) itself is not new. We re-derive this formula in the appendix using linear algebra notation to make the work more accessible to readers not familiar with this subject. The concept of the energy density matrix has been used before [36, 45], and the last term in equation (27) is also referred to as the 'orthogonalization force' in the appendix of [36], which takes into account the fact that eigenfunctions must be orthogonalized after atomic positions change.

To illustrate more clearly that both the Hellmann–Feynman force and the Pulay force are taken into account correctly, let us look into the first term in equation (27),

$$\frac{\partial H_{ij}}{\partial R_I} = \left\langle \frac{\partial \varphi_i}{\partial R_I}, \hat{H}\varphi_j \right\rangle + \left\langle \varphi_i, \frac{\partial \hat{H}}{\partial R_I}\varphi_j \right\rangle + \left\langle \varphi_i, \hat{H}\frac{\partial \varphi_j}{\partial R_I} \right\rangle. \tag{29}$$

The terms $\frac{\partial \varphi_i}{\partial R_I}$ are automatically included to reflect the change of the atom-centered basis functions with respect to atomic positions, which gives rise to the Pulay force. From a computational point of view, the terms in equation (29) that are related to the kinetic and non-local pseudopotential parts can be solved by efficient two-center integrals techniques, while the terms related to local potential parts can be solved on a real space uniform grid. The Hartree potential and the exchange correlation potential are involved in the first term and the third term on the right-hand side of equation (29), but make no contribution to the second term on the right-hand side of equation (29). Once all the terms in equation (29) are evaluated, one only needs to multiply them with density matrix Γ, which is obtained directly from the PEXSI method.

In order to compute the energy density matrix in equation (28), and therefore the orthogonalization force without using the Kohn–Sham eigenvalues $\{\varepsilon_i\}$ and Kohn–Sham orbitals $\{\psi_i\}$, it is sufficient to note that the function

$$f^E_\beta(\varepsilon - \mu) = \varepsilon f_\beta(\varepsilon - \mu) \tag{30}$$

has the same analytic structure as the Fermi–Dirac function f_β. Thus, the energy density matrix can be approximated by the same pole expansion used to approximate the density matrix (15). In particular, there is no difference in the choice of poles z_l. But the weights of the expansion, which we denote by ω^E_l, for the energy density matrix approximation are different. To be specific, the energy density matrix can be written using the pole expansion as

$$\Gamma^E = C \Im\mathrm{m} \sum_{l=1}^{P} \frac{\omega^E_l}{\Xi - (z_l + \mu)I} C^T = \sum_{l=1}^{P} \frac{\omega^E_l}{H - (z_l + \mu)S}. \tag{31}$$

Again the selected elements of Γ^E required in (27) 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 (15), (24) and (31) 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 of (7), which can be expressed as

$$N(\mu) = \mathrm{Tr}[\hat{\gamma}] = \mathrm{Tr}[\Gamma \Phi^T \Phi] = \mathrm{Tr}[\Gamma S], \tag{32}$$

is a non-decreasing function with respect to μ. Hence the root of (7) can be obtained by either Newton's method or the bisection method. Other strategies for updating the chemical potential have also been discussed in more detail in the literature [7, 16].

In an SCF iteration, $\hat{\rho}$ and μ are often updated in an alternating fashion. When the Kohn–Sham energies ε_i associated with a fixed charge density are available, both $N(\mu)$ and its derivative can be easily evaluated in Newton's method. However, if $\hat{\gamma}$ is approximated via a pole expansion (15), a new expansion is needed whenever μ is updated. In Newton's method, the derivative of $N(\mu)$ can be approximated by the finite difference. When μ^k is sufficiently close to the true chemical potential, the derivative of $N(\mu^k)$ can be approximated by

$$N'(\mu^k) \approx \frac{N(\mu^k) - N(\mu^{k-1})}{\mu^k - \mu^{k-1}}. \tag{33}$$

We remark that although Newton's method converges rapidly near the correct chemical potential as can be seen from the numerical results in section 3, it may not always be robust and may give very large correction when the derivative (33) is small. In such a case a damped Newton's method or the bisection method can be used instead to ensure the convergence of the chemical potential iteration. It remains challenging to update the chemical potential both efficiently and robustly for all systems with a wide range of initial guesses, especially in the presence of gap states, and dispersive bands which require global Fermi level finding across multiple k-points. We will develop efficient and robust schemes to overcome this difficulty in our future work.

2.5. The flowchart of PEXSI

In algorithm 1 we summarize the main steps of the PEXSI technique for accelerating atomic orbital-based electronic structure calculation with the SCF iteration. We see that PEXSI replaces the diagonalization procedure in solving KSDFT, and obtains the electron density, the total energy, the Helmholtz free energy and the atomic force accurately without computing eigenvalues and eigenfunctions of the Hamiltonian operator.

3. Numerical results

In this section, we report the performance achieved by applying the PEXSI 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 nanotubes of two types. 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 the nanotubes [46], 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 that of CNT(8, 8) with 512 atoms.

```
Algorithm 1: Flowchart of the PEXSI technique.
Input:  Atomic position $\{R_I\}$. Basis set $\Phi$. A subroutine to construct matrices $H, S$ and matrices
        $\left\{\frac{\partial H}{\partial R_I}\right\}, \left\{\frac{\partial S}{\partial R_I}\right\}$ given any electron density $\hat{\rho}$.
Output: Converged electron density $\hat{\rho}$. Total energy $E_{\mathrm{tot}}$.
        Helmholtz free energy $\mathcal{F}_{\mathrm{tot}}$. Atomic forces $\{F_I\}$.
        Chemical potential $\mu$.

while $\hat{\rho}$ has not converged do
    Update $\hat{\rho}$ via charge mixing schemes for the SCF iteration.
    Construct matrices $H, S$ using the updated electron density $\hat{\rho}$.
    while $\mu$ has not converged do
        Update the chemical potential $\mu$.
        for each pole $l = 1, \ldots, P$ do
            Compute the selected elements of each Green's function $\frac{1}{H-(z_l+\mu)}$ using selected inversion.
        end for
        Compute $\Gamma$ via equation (15), and compute the number of electrons $N(\mu)$ via equation (32).
    end while
end while
Compute the free energy density matrix $\Gamma^{\mathcal{F}}$ via equation (25), and the energy density matrix $\Gamma^{E}$ via
    equation (31) using the selected elements of the same set of Green's functions for computing $\Gamma$.
Compute the converged electron density $\hat{\rho}$ via equation (16), the total energy $E_{\mathrm{tot}}$ via equation (26), the
    Helmholtz free energy $\mathcal{F}_{\mathrm{tot}}$ via equation (24), and the atomic forces $\{F_I\}$ via equation (27).
```

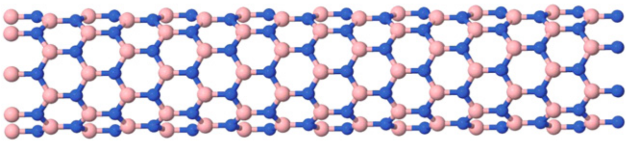

Figure 2. Boron nitride nanotube (8, 0) with 256 atoms. The boron atoms are labeled as pink (light gray) balls while the nitrogen atoms are labeled as blue (dark gray) balls. The bond length between a pair of adjacent boron and nitride atoms is 1.45 Å.

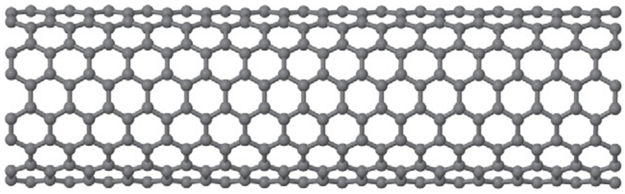

Figure 3. Carbon nanotube (8, 8) with 512 atoms. The carbon atoms are labeled as gray balls. The bond length between a pair of adjacent carbon atoms is 1.42 Å.

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

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 GB (gigabytes) DDR3 1333 MHz memory. Each core processor has 64 kB (kilobytes) L1 cache and 512 kB L2 cache. It also has access to a 6 MB (megabytes) of L3 cache shared among six 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 PEXSI 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 using 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 PEXSI 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 that we used for the performance study is based on a local atomic orbital expansion scheme [24, 25]. We will refer to this scheme as the CGH scheme below. In the CGH scheme, an atomic

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

# Atoms		64	128	256	512	1024	1920	5120	10 240
BNNT(8, 0)	$H_{\mathrm{nnz}}\%$	100.00	85.54	42.77	21.43	11.69	5.70	2.13	1.06
	$L_{\mathrm{nnz}}\%$	100.00	99.48	77.94	46.13	25.07	13.70	5.26	2.64
CNT(8, 8)	$H_{\mathrm{nnz}}\%$	40.63	38.67	19.53	9.77	4.88	2.60	0.97	0.49
	$L_{\mathrm{nnz}}\%$	69.92	68.45	68.70	54.38	31.75	17.54	7.42	3.79

orbital $\varphi_\mu(\mathbf{r})$ is expressed as the product of a radial wavefunction $f_{\mu,l}(r)$ 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.,

$$f_{\mu,l}(r) = \begin{cases} \sum_q c_{\mu q} j_l(qr), & r < r_c \\ 0 & r \geq r_c. \end{cases} \qquad (34)$$

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' [47, 48] associated with a reference system that consists of a set of (four or five) dimers. We refer readers to [24, 25] for the details on the construction of the CGH local atomic orbitals.

The cutoff radius r_c determines the sparsity of the 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. The reason why we choose a larger cutoff radius for B, N atoms is that the spillage factor for the B and N atoms is larger than that for the C atoms if 6.0 Bohr cutoff is used for all atoms, which affects the accuracy of the atomic orbitals. In general, the cutoff radius of most atomic orbitals can be chosen below 10 Bohr.

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

$$H_{\mathrm{nnz}}\% = \frac{\mathrm{nnz}(H)}{N^2(H)} \times 100. \qquad (35)$$

Here $\mathrm{nnz}(H)$ is the number of nonzero elements of H and $N(H)$ is the dimension of H. 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 $L_{\mathrm{nnz}}\%$) below. To reduce the amount of nonzero fill-in of L, we use the nested dissection (ND) technique [49] to reorder the sparse matrix $H - zS$ before it is factored. Figure 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 figure 4(b).

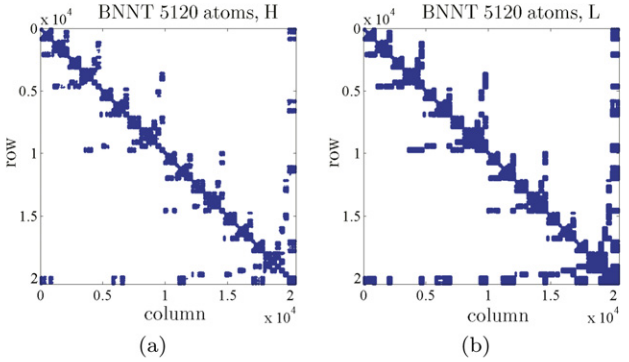

Figure 4. The sparsity pattern of H (a) and $L + L^T$ (b) for a 5120-atom BNNT(8, 0) with SZ orbitals. Nested dissection reordering is used.

Table 1 shows the sparsity of Hamiltonian matrices associated with BNNT(8, 0) and CNT(8, 8) systems that consist of 64–10 240 atoms. The Hamiltonians for these systems are constructed from SZ atomic orbitals. We report both the $H_{\mathrm{nnz}}\%$ and $L_{\mathrm{nnz}}\%$ 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 the fact that a large percentage of the 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 $H_{\mathrm{nnz}}\%$ and $L_{\mathrm{nnz}}\%$ are inversely proportional to the system size. This is because for quasi-1D systems, the numerator in equation (35) 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.

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 [18].

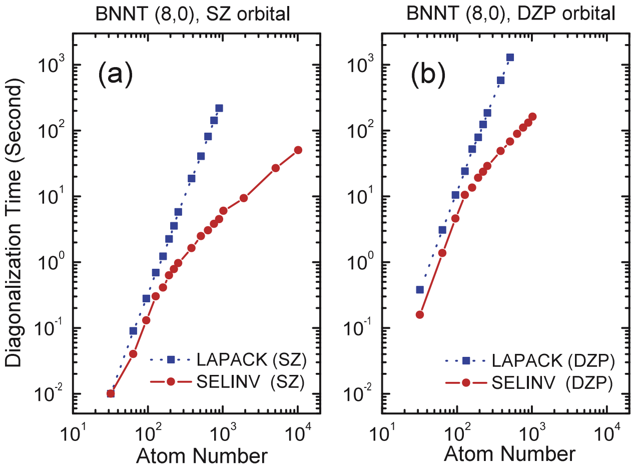

Figure 5. Comparisons of the wall clock time used by selected inversion (at one pole) required for PEXSI and by the LAPACK `dsygv` used to diagonalize a Kohn–Sham Hamiltonian associated with BNNT(8, 0). The Hamiltonians are constructed from SZ orbitals (four basis functions per atom) in (a) and DZP orbitals (thirteen basis functions per atom) in (b).

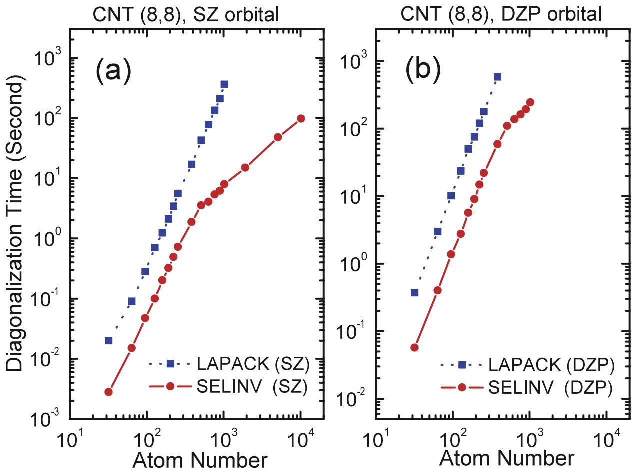

Figure 6. Comparisons of the wall clock time by selected inversion (at one pole) required for PEXSI and by the LAPACK `dsygv` used to diagonalize a Kohn–Sham Hamiltonian associated with CNT(8, 8). The Hamiltonians are constructed from SZ orbitals (four basis functions per atom) in (a) and DZP orbitals (thirteen basis functions per atom) in (b).

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 10 240.

Figure 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 `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 log–log 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.

Figure 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` for a CNT with 64 atoms regardless of 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) that 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 as atoms for CNT to reach the same length along the longitudinal direction when compared to BNNT, as we can see from figures 2 and 3.

We should note here that it is possible to combine the PEXSI 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 5120-atom BNNT(8, 0) is 26.72 s. When the number of atoms increases to 10 240, the wall clock time increases to 50.07 s. Similar performance is observed for CNT(8, 8). It takes 47.59 s to perform a selected inversion for a 5120-atom CNT(8, 8) tube, and 97.16 s for a 10 240-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 10 240 atoms, the amounts of memory required to store L and the selected elements of $[H - (z_l + \mu)S]^{-1}$ are 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

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

# Poles	$E_{\mathrm{PEXSI}} - E_{\mathrm{ref}}$ (eV)	MAE force (eV Å^{-1})
20	5.868 351 108	0.400 431
40	0.007 370 583	0.001 142
60	0.000 110 382	0.000 026
80	0.000 000 360	0.000 002

introduced by the PEXSI technique comes from the limited number of terms in the pole expansion (15). The number of poles needed in (15) to achieve a desired level of accuracy in total energy (or free energy) and force is largely determined by the inverse temperature $\beta = 1/(k_{\mathrm{B}}T)$ used in (4) and the spectrum width ΔE. Here we show that at room temperature $T = 300$ K, 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 PEXSI for a single Γ point calculation, the errors in total energy and force decrease as the number of poles in (15) increases. The force difference is measured between the force calculated with the PEXSI scheme using equation (27), and that calculated by the LAPACK diagonalization subroutine `dsygv` using standard methods [36] previously implemented in the CGH atomic orbital scheme [24, 25]. 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 PEXSI) is 3.6×10^{-7} eV. The difference in the mean absolute error (MAE) is 2×10^{-6} eV Å^{-1}, which is quite small for all practical purposes.

The number of chemical potential iterations, as well as the error of the number of electrons at different SCF steps for a metallic CNT(8, 8) system with 1024 atoms using the SZ basis set are reported in figure 7. The chemical potential is relaxed until the error associated with the total electron number (4096 electrons in this system) is within a given tolerance τ. The average number of chemical potential iterations is 2.01 for the low accuracy case ($\tau = 10^{-1}$), and 5.21 for the high accuracy case ($\tau = 10^{-8}$), respectively. Notice that in both cases, the number of chemical potential iterations is 1–2 when the SCF gets close to convergence. Similar behavior is also observed in the geometry optimization example in section 3.6 for which the change of chemical potential in consecutive steps is small. We further remark that the chemical potential does not need to be performed very accurately in the first few SCF steps. So the tolerance τ can be chosen dynamically with respect to the accuracy of the current SCF step, in order to further reduce the number of chemical potential iterations in the case of high accuracy calculation. We note that `SelInv` is a direct method for computing selected elements of the Green's function accurately. When low accuracy is allowed, it is possible to reduce the computational cost of this method further by discarding elements in the Cholesky factor with small magnitude. This approach will be pursued in our future work.

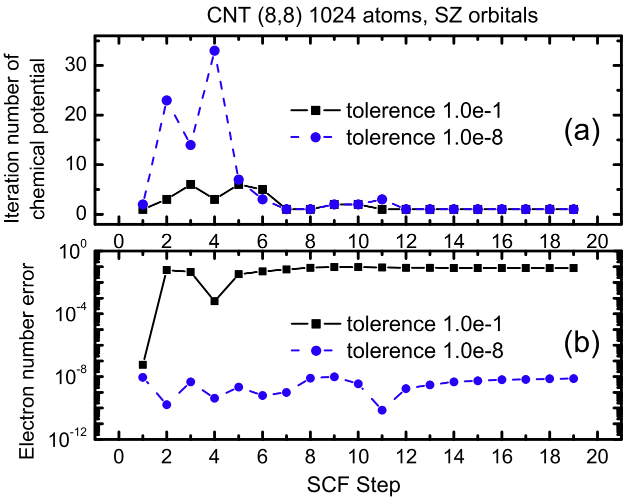

Figure 7. The numbers of chemical potential iteration steps (a), and the error associated with the number of electrons (b) at different SCF iterations for CNT(8, 8) with 1024 atoms using the SZ basis set. The chemical potential is relaxed until the error of the total number of electrons (4096 electrons in this system) is within 10^{-1} (blue dashed lines with dots) and within 10^{-8} (black solid lines with squares).

3.5. Overall performance

With a sequential machine, the total wall clock time consumed by each PEXSI-based SCF iteration is $t_{\mathrm{selinv}} \times P \times k_{\mu}$, where t_{selinv} is the time required to perform one selected inversion, P is the number of poles used in the pole expansion (13) and k_{μ} is the average number of chemical potential iterations. In practice, $P = 80$ is often more than sufficient to yield an accurate approximation in (13) as we can see from table 2. The average k_{μ} can be 1–2 especially in geometry optimization and molecular dynamics. If we take $P = 80$ and $k_{\mu} = 2$, the total wall clock time of a PEXSI-based SCF iteration is compared with a LAPACK diagonalization-based SCF iteration for BNNT and CNT of various sizes in figures 8 and 9, respectively. Since the LAPACK diagonalization routine cannot perform as large a calculation as PEXSI due to memory constraints, we extrapolate the wall clock time of the LAPACK diagonalization routine in figures 8 and 9, and we find that the number of atoms beyond which the sequential PEXSI method outperforms the diagonalization method is 1650 atoms for BNNT(8, 0) discretized by SZ orbitals, and 1800 atoms for BNNT(8, 0) discretized by DZP orbitals. Similarly, the crossover for the sequential PEXSI method to outperform the diagonalization method is 1750 atoms for CNT(8, 8) discretized by SZ orbitals, and 1700 atoms for CNT(8, 8) discretized by DZP orbitals.

However, when a large number of processors are available, the advantage of PEXSI becomes apparent. Because each term in (13) can be evaluated independently, we achieve an automatic P-fold speedup whereas the speedup that can be achieved by a parallel diagonalization procedure implemented in, for example, the ScaLAPACK software package, is often limited. Furthermore, each selected inversion can be parallelized, and our current work, which we will publish in

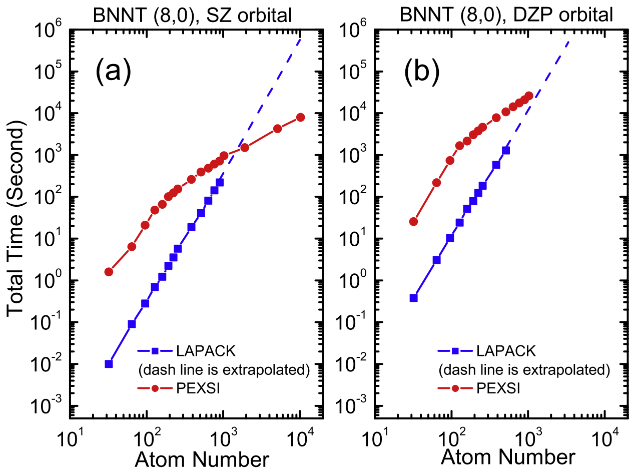

Figure 8. Comparisons of the total wall clock time used to perform a PEXSI-based SCF iteration (using eighty poles and two iterations of chemical potential) and to perform a LAPACK `dsygv` diagonalization-based SCF iteration for BNNT(8, 0) configured with different numbers of atoms. The Hamiltonians are constructed from SZ orbitals (four basis functions per atom) in (a) and DZP orbitals (thirteen basis functions per atom) in (b).

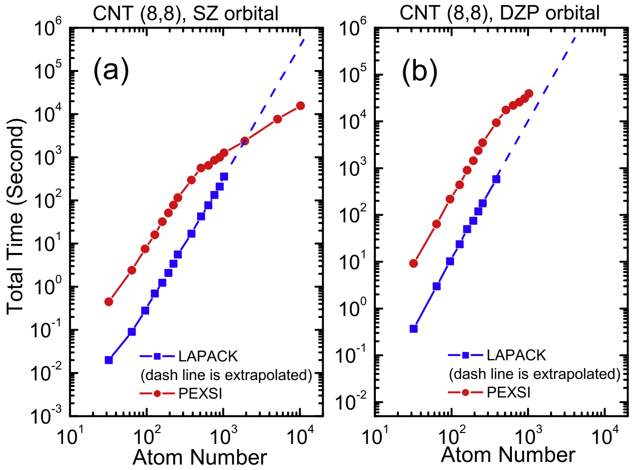

Figure 9. Comparisons of the total wall clock time used to perform a PEXSI-based SCF iteration (using eight poles and two iterations of chemical potential) and a LAPACK `dsygv` diagonalization-based SCF iteration for CNT(8, 8). The Hamiltonians are constructed from SZ orbitals (four basis functions per atom) in (a) and DZP orbitals (thirteen basis functions per atom) in (b).

a separate publication, indicates that excellent speedup can be achieved for this calculation on hundreds of processors. As a result, the PEXSI-based SCF iteration can easily scale to tens of thousands of processors, whereas it is difficult to make ScaLAPACK diagonalization procedures work efficiently on that many processors.

3.6. Geometry optimization

The PEXSI scheme with atomic orbitals can also be used for accurate geometry relaxation of large-scale atomic systems. We use a truncated boron nitride nanotube (8, 0) with 1024 atoms, shown in figure 10, as an example to illustrate the efficiency of PEXSI in this type of calculation. The nanotube contains 504 boron atoms (B) and 504 nitride atoms (N). Each end of the nanotube is passivated by eight hydrogen atoms (H). We used DZP orbitals for all three atomic elements. The cutoff radius for B and N is set to 8.0 Bohr. The cutoff radius for H is set to 6.0 Bohr. We used 96 poles in the pole expansion for both energy and force calculations.

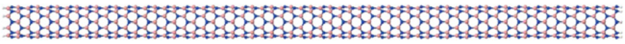

Figure 10. A truncated boron nitride nanotube (8, 0) with 1024 atoms, among which 504 boron atoms are labeled as pink (light gray) balls, 504 nitride atoms are labeled as blue (dark gray) balls, and 16 hydrogen atoms are labeled as small white balls. The hydrogen atoms are used to passivate both ends of the nanotube.

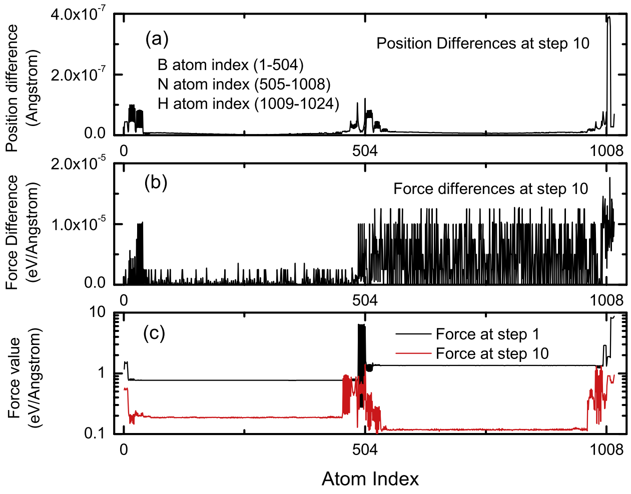

Figure 11. The differences of the atomic positions (a) and forces (b) obtained from separate simulations using the PEXSI method and the diagonalization method, starting from the same initial condition. The result is obtained at the tenth geometry optimization step for the boron nitride nanotube (8, 0) system with 1024 atoms. The absolute values of the forces at the first and the tenth geometry optimization steps are also presented (c). The tolerance for the error of the total number of electrons is chosen to be 10^{-8}.

Convergence is reached after 105 steps of ionic relaxation steps are taken in the BFGS method. The maximum atomic force associated with the converged structure is less than 0.04 eV Å^{-1}. To demonstrate the accuracy of the PEXSI method, we compare the differences of the atomic positions and forces obtained from separate geometry optimization simulations using the PEXSI method and the diagonalization method, starting from the same initial condition. Figure 11 shows that at the tenth geometry optimization step, the maximum difference of the atomic positions among all 1024 atoms is less than 5×10^{-7} Å (figure 11(a)), and the maximum difference of the forces is less than 2×10^{-5} eV Å^{-1} (figure 11(b)). Figure 11(c) shows that at the tenth geometry optimization step, the absolute value of the force is still as large as 0.1–1 eV Å^{-1}, and the relative error of the forces obtained from the PEXSI method is around 0.01%. This result shows that the PEXSI scheme is accurate for evaluating the forces for this system.

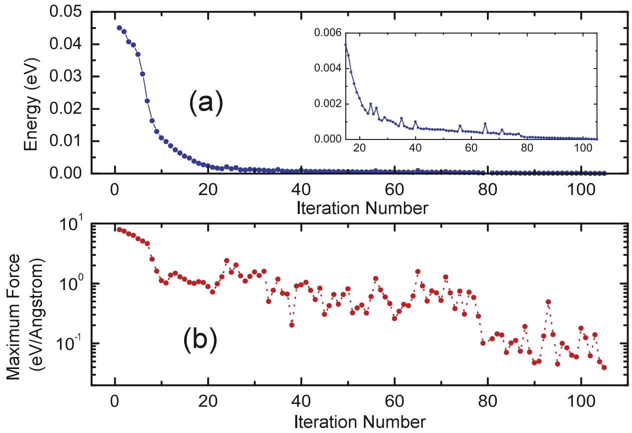

Figure 12. The energy per atom (a) and the maximum force (b) for each geometry optimization iteration step. The criterion for the convergence of the force is set to 0.04 eV Å^{-1}. The energy per atom at the last iteration step is set to zero.

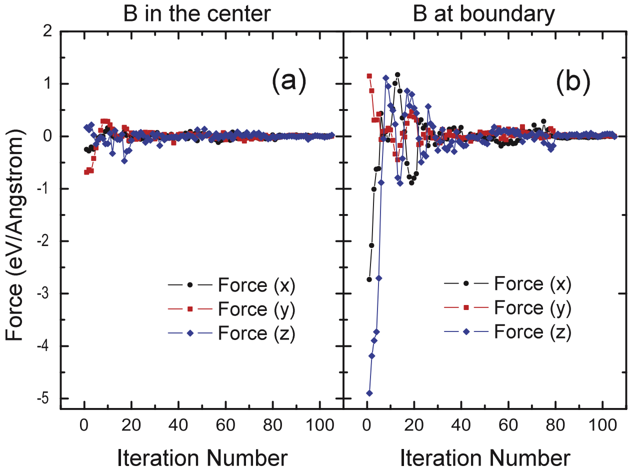

Figure 13. The force (x, y, z directions) acting on the boron atoms near the center of the nanotube (a) and near the boundary of the nanotube (b).

The convergence history of the energy per atom and the convergence history of the maximum force with respect to the iteration number in the geometry optimization procedure are plotted in figures 12(a) and (b), respectively. In figure 12(a), the energy per atom at the last iteration step is set to zero. The energy per atom converges rapidly from 0.05 to 0.005 eV during the first 16 steps. Correspondingly, in figure 12(b), the maximum force converges rapidly during the first few steps. This is mainly because the initial positions of the hydrogen and boron atoms near the end of the nanotube are not far from the equilibrium value. After the hydrogen and boron atoms at the boundary are relaxed to more reasonable positions, the maximum force begins to decrease slowly but with some oscillations. In order to illustrate more clearly the origin of the oscillation, we show the forces of boron atoms in figure 13. Figures 13(a) and (b) show the forces of the boron atoms near the center of the nanotube and near the boundary of the nanotube, respectively. We find that the forces acting on the boron atoms near the center of the nanotube are much smaller than those near the boundary. This is mainly due to the fact that the atomic configuration near the center of the nanotube is close to the bulk configuration. The magnitude of the force acting on the atoms near the boundary is much larger, and is more difficult to converge in the numerical optimization.

4. Conclusion

In this paper, we generalized the recently developed pole expansion and selected inversion technique (PEXSI) for solving finite-dimensional Kohn–Sham equations obtained from an atomic orbital expansion. We gave expressions for evaluating the electron density, the total energy, the Helmholtz free energy and the atomic forces (including both the Hellmann–Feynman force and the Pulay force) without using eigenvalues and eigenvectors of a Kohn–Sham Hamiltonian. These expressions are derived from an FOE approximation to the Fermi–Dirac function using an efficient and accurate pole expansion technique. The favorable $\log(\beta\Delta E)$ scaling of the pole expansion allows us to treat both insulating and metallic systems efficiently at room temperature or even lower temperature. The pole expansion only uses selected elements of the density matrix, energy density matrix and free energy density matrix. These selected elements can be obtained from computing the selected elements of the inverse of a shifted Kohn–Sham Hamiltonian through the selected inversion technique. The complexity of the 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 achieved by comparing the efficiency of PEXSI with that of diagonalization on nanotubes of two types. The linear scaling behavior of PEXSI 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. For quasi-2D and quasi-3D systems, we expect the crossover point over which PEXSI exhibits $\mathcal{O}(N_e^{3/2})$ and $\mathcal{O}(N_e^2)$ scaling to be much larger. However, on the basis of the experiments presented here, it seems that PEXSI 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 that we presented above were performed with a sequential implementation of the selected inversion algorithm. For quasi-1D systems such as nanotubes, the use of PEXSI allows us to tackle problems that contain as many as 10 000 atoms. This cannot be done by using a diagonalization-based approach. We further demonstrate the applicability of the PEXSI scheme by performing the geometry optimization of a truncated boron nitride nanotube with 1024 atoms. For quasi-2D and 3D systems, a parallel implementation of the PEXSI, 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.

Acknowledgments

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under the US Department of Energy contract number DE-AC02-05CH11231 and by the Scientific Discovery through Advanced Computing (SciDAC) program funded by the US Department of Energy, Office of Science, Advanced Scientific Computing Research and Basic Energy Sciences (LL and CY), and by the Chinese National Natural Science Funds for Distinguished Young Scholars (LH).

Appendix

Derivation of equation (15):

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

$$f_\beta(\Xi-\mu) \approx \mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\rho}}{\Xi-(z_l+\mu)I}, \tag{36}$$

where I is an $N \times N$ identity matrix. Using equation (12), the approximation of the single-particle density matrix using P terms of the pole expansion (still denoted by $\hat{\gamma}$ to simplify the notation) can be written as

$$\begin{aligned}\hat{\gamma}(x,x') &= \Phi(x)C\mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\rho}}{\Xi-z_l I}C^T\Phi^T(x')\\ &= \Phi(x)\mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\rho}}{C^{-T}\Xi C^{-1}-z_l C^{-T}C^{-1}}\Phi^T(x').\end{aligned} \tag{37}$$

Since the generalized eigenvalue problem (10) implies the identity

$$C^T H C = \Xi, \qquad C^T S C = I, \tag{38}$$

the single-particle density matrix takes the form

$$\hat{\gamma}(x,x') = \Phi(x)\mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\rho}}{H-(z_l+\mu)S}\Phi^T(x') \tag{39}$$

which is equation (15).

Derivation of equation (24):

The first term in the Helmholtz free energy functional is

$$\begin{aligned}\mathrm{Tr}[f_\beta^{\mathcal{F}}(\Xi-\mu)] &= \mathrm{Tr}[Cf_\beta^{\mathcal{F}}(\Xi-\mu)C^T C^{-T}C^{-1}]\\ &\equiv \mathrm{Tr}[\Gamma^{\mathcal{F}}S].\end{aligned} \tag{40}$$

The second equal sign in equation (40) defines the free energy density matrix $\Gamma^{\mathcal{F}}$, which can be evaluated using the pole expansion (23) as

$$\begin{aligned}\Gamma^{\mathcal{F}} &= C\mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\mathcal{F}}}{\Xi-z_l I}C^T\\ &= \mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\mathcal{F}}}{C^{-T}HC^{-1}-z_l C^{-T}C^{-1}}\\ &= \mathfrak{Im}\sum_{l=1}^{P}\frac{\omega_l^{\mathcal{F}}}{H-z_l S},\end{aligned} \tag{41}$$

which is equation (24).

Derivation of equation (27):

The atomic force is in general given by the derivative of the Helmholtz free energy $\mathcal{F}_{\mathrm{tot}}$ with respect to the atomic positions. Since the free energy is minimized with respect to $\{\psi_i\}$, $\{f_i\}$ at each atomic configuration $\{R_I\}$, all the terms in $\mathcal{F}_{\mathrm{tot}}$ that do not explicitly depend on R_I will not contribute to the atomic force F_I. In particular, the double-counting terms $-\frac{1}{2}\iint\frac{\hat{\rho}(x)\hat{\rho}(y)}{|x-y|}\,\mathrm{d}x\,\mathrm{d}y + E_{\mathrm{xc}}[\hat{\rho}] - \int V_{\mathrm{xc}}[\hat{\rho}](x)\hat{\rho}(x)\,\mathrm{d}x$ do not contribute to the atomic force. Therefore

$$F_I = -\frac{\mathrm{d}}{\mathrm{d}R_I}\mathcal{F}_{\mathrm{tot}} = -\frac{\partial}{\partial R_I}\mathcal{F}_{\mathrm{tot}}. \tag{42}$$

Using the representation of the Helmholtz free energy in equation (20), and the fact that

$$(f_\beta^{\mathcal{F}})'(z) = f_\beta(z), \qquad N_{\mathrm{e}} = \mathrm{Tr}[f_\beta(\Xi-\mu)], \tag{43}$$

it can be derived that

$$\begin{aligned}F_I &= -\frac{\partial}{\partial R_I}\mathcal{F}_{\mathrm{tot}} = -\frac{\partial}{\partial R_I}(\mathrm{Tr}[f_\beta^{\mathcal{F}}(\Xi-\mu)]+\mu N_{\mathrm{e}})\\ &= -\mathrm{Tr}\left[(f_\beta^{\mathcal{F}})'(\Xi-\mu)\left(\frac{\partial\Xi}{\partial R_I}-\frac{\partial\mu}{\partial R_I}\right)\right]-N_{\mathrm{e}}\frac{\partial\mu}{\partial R_I}\\ &= -\mathrm{Tr}\left[f_\beta(\Xi-\mu)\frac{\partial\Xi}{\partial R_I}\right]-\frac{\partial\mu}{\partial R_I}(N_{\mathrm{e}}-\mathrm{Tr}[f_\beta(\Xi-\mu)])\\ &= -\mathrm{Tr}\left[f_\beta(\Xi-\mu)C^T\frac{\partial H}{\partial R_I}C\right]-\mathrm{Tr}\left[f_\beta(\Xi-\mu)\frac{\partial C^T}{\partial R_I}HC\right]\\ &\quad -\mathrm{Tr}\left[f_\beta(\Xi-\mu)C^T H\frac{\partial C}{\partial R_I}\right]\\ &= -\mathrm{Tr}\left[\Gamma\frac{\partial H}{\partial R_I}\right]-\mathrm{Tr}\left[f_\beta(\Xi-\mu)\frac{\partial C^T}{\partial R_I}HC\right]\\ &\quad -\mathrm{Tr}\left[f_\beta(\Xi-\mu)C^T H\frac{\partial C}{\partial R_I}\right].\end{aligned} \tag{44}$$

The second and the third terms in equation (44) come from the nonorthogonality of the basis functions and should be further simplified. We have

$$\begin{aligned}&\mathrm{Tr}\left[f_\beta(\Xi-\mu)\frac{\partial C^T}{\partial R_I}HC\right]+\mathrm{Tr}\left[f_\beta(\Xi-\mu)C^T H\frac{\partial C}{\partial R_I}\right]\\ &\quad = \mathrm{Tr}\left[(C^{-T}C^{-1})[C(C^T HC)f_\beta(\Xi-\mu)C^T]\right.\\ &\quad\times\left.(C^{-T}C^{-1})C\frac{\partial C^T}{\partial R_I}\right]+\mathrm{Tr}\left[C^{-T}C^{-1}[Cf_\beta(\Xi-\mu)\right.\\ &\quad\times\left.(C^T HC)C^T]C^{-T}C^{-1}\frac{\partial C}{\partial R_I}C^T\right]\\ &\quad\equiv\mathrm{Tr}\left[(C\Xi f_\beta(\Xi-\mu)C^T)\left(SC\frac{\partial C^T}{\partial R_I}S+S\frac{\partial C}{\partial R_I}C^T S\right)\right].\end{aligned} \tag{45}$$

Define the energy density matrix as in equation (28), and equation (45) can be simplified as

$$\begin{aligned}&\mathrm{Tr}\left[\Gamma^E S\left(C\frac{\partial C^T}{\partial R_I}+\frac{\partial C}{\partial R_I}C^T\right)S\right]\\ &\quad = \mathrm{Tr}\left[\Gamma^E S\frac{\partial S^{-1}}{\partial R_I}S\right] = -\mathrm{Tr}\left[\Gamma^E\frac{\partial S}{\partial R_I}\right].\end{aligned} \tag{46}$$

Combining equations (46) and (44), we have

$$F_I = -\frac{\partial \mathcal{F}}{\partial R_I} = -\mathrm{Tr}\left[\Gamma \frac{\partial H}{\partial R_I}\right] + \mathrm{Tr}\left[\Gamma^E \frac{\partial S}{\partial R_I}\right]. \qquad (47)$$

which proves equation (27).

References

[1] Bowler D R, Miyazaki T and Gillan M J 2002 *J. Phys.: Condens. Matter* **14** 2781

[2] Fattebert J L and Bernholc J 2000 *Phys. Rev.* B **62** 1713

[3] Hine N D, Haynes P D, Mostofi A A, Skylaris C K and Payne M C 2009 *Comput. Phys. Commun.* **180** 1041

[4] Yang W 1991 *Phys. Rev. Lett.* **66** 1438

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

[6] McWeeny R 1960 *Rev. Mod. Phys.* **32** 335

[7] Goedecker S 1999 *Rev. Mod. Phys.* **71** 1085

[8] Bowler D R and Miyazaki T 2012 *Rep. Prog. Phys.* **75** 036503

[9] Kohn W 1996 *Phys. Rev. Lett.* **76** 3168

[10] Prodan E and Kohn W 2005 *Proc. Natl Acad. Sci.* **102** 11635

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

[12] Baroni S and Giannozzi P 1992 *Europhys. Lett.* **17** 547

[13] Goedecker S 1993 *Phys. Rev.* B **48** 17573

[14] Ozaki T 2007 *Phys. Rev.* B **75** 035123

[15] Ceriotti M, Kühne T and Parrinello M 2008 *J. Chem. Phys.* **129** 024707

[16] Ozaki T 2010 *Phys. Rev.* B **82** 075131

[17] Lin L, Lu J, Ying L, Car R and E W 2009 *Commun. Math. Sci.* **7** 755

[18] Lin L, Yang C, Meza J, Lu J, Ying L and E W 2011 *ACM Trans. Math. Softw.* **37** 40

[19] Lin L, Yang C, Lu J, Ying L and E W 2011 *SIAM J. Sci. Comput.* **33** 1329

[20] Varga K 2010 *Phys. Rev.* B **81** 045109

[21] Frisch M, Pople J and Binkley J 1984 *J. Chem. Phys.* **80** 3265

[22] VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T and Hutter J 2005 *Comput. Phys. Commun.* **167** 103

[23] Junquera J, Paz O, Sanchez-Portal D and Artacho E 2001 *Phys. Rev.* B **64** 235111

[24] Chen M, Guo G C and He L 2010 *J. Phys.: Condens. Matter* **22** 445501

[25] Chen M, Guo G C and He L 2011 *J. Phys.: Condens. Matter* **23** 325501

[26] Kenny S D, Horsfield A P and Fujitani H 2000 *Phys. Rev.* B **62** 4899

[27] Ozaki T 2003 *Phys. Rev.* B **67** 155108

[28] Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K and Scheffler M 2009 *Comput. Phys. Commun.* **180** 2175

[29] Tsuchida E and Tsukada M 1998 *J. Phys. Soc. Japan* **67** 3844

[30] Lin L, Lu J, Ying L and E W 2012 *J. Comput. Phys.* **231** 2140

[31] Lin L, Lu J, Car R and E W 2009 *Phys. Rev.* B **79** 115133

[32] Lanczos C 1950 *J. Res. Natl Bur. Stand.* **45** 255

[33] Chelikowsky J, Troullier N and Saad Y 1994 *Phys. Rev. Lett.* **72** 1240

[34] Tsuchida E and Tsukada M 1995 *Phys. Rev.* B **52** 5573

[35] Mermin N 1965 *Phys. Rev.* **137** A1441

[36] Soler J M, Artacho E, Gale J D, García A, Junquera J, Ordejón P and Sánchez-Portal D 2002 *J. Phys.: Condens. Matter* **14** 2745

[37] Kresse G and Furthmüller J 1996 *Phys. Rev.* B **54** 11169

[38] Weinert M and Davenport J W 1992 *Phys. Rev.* B **45** 13709

[39] Wentzcovitch R M, Martins J L and Allen P B 1992 *Phys. Rev.* B **45** 11372

[40] Alavi A, Kohanoff J, Parrinello M and Frenkel D 1994 *Phys. Rev. Lett.* **73** 2599

[41] Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566

[42] Becke A D 1988 *Phys. Rev.* A **38** 3098

[43] Lee C, Yang W and Parr R G 1988 *Phys. Rev.* B **37** 785

[44] Pulay P 1969 *Mol. Phys.* **17** 197

[45] Sankey O F and Niklewski D J 1989 *Phys. Rev.* B **40** 3979

[46] Charlier J-C, Blase X and Roche S 2007 *Rev. Mod. Phys.* **79** 677

[47] Sánchez-Portal D, Artacho E and Soler J M 1995 *Solid State Commun.* **95** 685

[48] Sánchez-Portal D, Artacho E and Soler J M 1996 *J. Phys.: Condens. Matter* **8** 3859

[49] George A 1973 *SIAM J. Numer. Anal.* **10** 345