

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Adaptive local basis set for Kohn–Sham density functional theory in a discontinuous Galerkin framework II: Force, vibration, and molecular dynamics calculations

### Permalink

<https://escholarship.org/uc/item/0ch4r4rv>

### Authors

Zhang, Gaigong

Lin, Lin

Hu, Wei

et al.

### Publication Date

2017-04-01

### DOI

10.1016/j.jcp.2016.12.052

Peer reviewed

# Adaptive local basis set for Kohn-Sham density functional theory in a discontinuous Galerkin framework II: Force, vibration, and molecular dynamics calculations

Gaigong Zhang<sup>b</sup>, Lin Lin<sup>a,b</sup>, Wei Hu<sup>b</sup>, Chao Yang<sup>b</sup>, John E. Pask<sup>c</sup>

<sup>a</sup>*Department of Mathematics, University of California, Berkeley, Berkeley, CA 94720.*

<sup>b</sup>*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.*

<sup>c</sup>*Physics Division, Lawrence Livermore National Laboratory, Livermore, CA 94550.*

---

## Abstract

Recently, we have proposed the adaptive local basis set for electronic structure calculations based on Kohn-Sham density functional theory in a pseudopotential framework. The adaptive local basis set is efficient and systematically improvable for total energy calculations. In this paper, we present the calculation of atomic forces, which can be used for a range of applications such as geometry optimization and molecular dynamics simulation. We demonstrate that, under mild assumptions, the computation of atomic forces can scale nearly linearly with the number of atoms in the system using the adaptive local basis set. We quantify the accuracy of the Hellmann-Feynman forces for a range of physical systems, benchmarked against converged planewave calculations, and find that the adaptive local basis set is efficient for both force and energy calculations, requiring at most a few tens of basis functions per atom to attain accuracies required in practice. Since the adaptive local basis set has implicit dependence on atomic positions, Pulay forces are in general nonzero. However, we find that the Pulay force is numerically small and systematically decreasing with increasing basis completeness, so that the Hellmann-Feynman force is sufficient for basis sizes of a few tens of basis functions per atom. We verify the accuracy of the computed forces in static calculations of quasi-1D and 3D disordered Si systems, vibration calculation of a quasi-1D Si system, and molecular dynamics calculations of H<sub>2</sub> and liquid Al-Si alloy systems, where we show systematic convergence to benchmark planewave results and results from the literature.

*Keywords:* Electronic structure, Kohn-Sham density functional theory, Discontinuous Galerkin, Adaptive local basis set, Hellmann-Feynman force, Pulay force, Molecular dynamics

---

*Email addresses:* [linlin@math.berkeley.edu](mailto:linlin@math.berkeley.edu) (Lin Lin), [whu@lbl.gov](mailto:whu@lbl.gov) (Wei Hu), [cyang@lbl.gov](mailto:cyang@lbl.gov) (Chao Yang), [pask1@llnl.gov](mailto:pask1@llnl.gov) (John E. Pask)

## 1. Introduction

Kohn-Sham density functional theory (KSDFT) [1, 2] is the most widely used electronic structure model for molecules and condensed matter systems. Kohn-Sham density functional theory gives rise to a nonlinear eigenvalue problem, which is commonly solved using the self-consistent field (SCF) iteration method [3]. At each SCF step, a linear eigenvalue problem with a fixed Kohn-Sham Hamiltonian defined by a fixed electron density  $\rho$  is solved. The solution to this linear eigenvalue problem is used to update the electron density and Kohn-Sham Hamiltonian in the SCF iteration. This is the most computationally expensive part of the SCF iteration. Although the asymptotic complexity of the computation with respect to the number of atoms depends on the algorithm used to solve the algebraic eigenvalue problem, the prefactor, which is related to the number of basis functions per atom, is characterized by how the problem is discretized. Methods such as the planewave method [4], finite difference method [5], and finite element method [6–10] exhibit systematic convergence with respect to the number of basis functions per atom, but can require a large number of basis functions per atom, from hundreds to thousands or more. The number of degrees of freedom can be reduced by incorporating atomic orbital physics into the basis [11–18]. Compared to methods such as the planewave method, however, it is more difficult to improve the quality of such atomic-orbital bases in a systematic fashion. The improvement can rely heavily on the practitioner’s experience with the underlying chemical system.

In a recent publication [19], we presented a new basis to discretize the Kohn-Sham Hamiltonian, called the adaptive local basis (ALB). The basic idea is to partition the global domain into a number of subdomains (called elements), and solve the Kohn-Sham problem locally around each element to generate the basis functions in each element. The basis so constructed is discontinuous across element boundaries. Therefore, we use the discontinuous Galerkin (DG) method [20] to construct a finite dimensional Kohn-Sham Hamiltonian in the discontinuous representation. The DG approach for solving the Kohn-Sham equations is also explored recently with enriched polynomial basis functions [21]. Recently, the adaptive local basis has been implemented in the DGDFT massively parallel electronic structure code [22]. By virtue of the adaptive local basis, the DGDFT solution converges systematically to the exact Kohn-Sham solution as the number of basis functions is increased, and the error can be measured by a posteriori error estimators [23]. We remark that the idea of generating localized basis functions on the fly has also been explored in other electronic structure software packages such as ONETEP [24] and BigDFT [25], where localized basis functions are continuous and are improved through an optimization procedure. Filter diagonalization [26] is another approach for contracting basis functions and has been applied to contract Gaussian type functions. Filter

diagonalization requires choosing trial functions, although the choice of trial functions may not be straightforward for an initial set of fine basis functions such as planewaves, finite elements or wavelets. Compared to ONETEP and other approaches mentioned above, one main advantage of the ALBs is that the basis functions are strictly local and form an orthonormal basis set. Therefore the reduced eigenvalue problem is well conditioned. The DG Hamiltonian matrix has a regular structure and resembles that obtained from finite difference discretization. We find that this facilitates load balancing in the massively parallel context, and allows the DGDFT code to scale beyond 100,000 processors [22, 27].

In KSDFT, many quantities of interest can be obtained from the total energy and atomic forces. We have previously demonstrated the effectiveness of the adaptive local basis [19] and a variant, the element orbitals [28], for computing the total energy of systems such as disordered bulk Na and Si, graphene with defects, and edge reconstruction of large scale armchair phosphorene nanoribbon systems [29]. In order to compute the atomic forces, which are given by the derivatives of the total energy with respect to atomic positions, the Hellmann-Feynman theorem [30, 31] is typically employed, and the resulting force is called the Hellmann-Feynman force. Since the adaptive local basis set depends implicitly on the atomic positions, however, the atomic force and Hellmann-Feynman force are in general not the same, and their difference, the Pulay force [32], reflects the effect of the atomic-position dependence of the basis. Although we have recently demonstrated that the Pulay force can be eliminated systematically through an additional optimization procedure [33], the procedure can be costly, especially in three-dimensional simulations. Therefore, it is of interest to determine the extent to which the Pulay force is reduced without such additional optimization as the size of the adaptive local basis is increased, and to determine the size of basis required for accurate quantum mechanical forces in practice.

Here, we describe the details to compute Hellmann-Feynman forces in the adaptive local basis, which can be evaluated with near linear scaling cost with respect to the number of atoms, provided the density matrix represented in the adaptive local basis is obtained. We quantify the accuracy of the Hellmann-Feynman forces for a range of systems compared to converged planewave calculations, and find that the adaptive local basis set is accurate and efficient for both energy and force calculations, achieving accuracies required in practice with a few tens of basis functions per atom. We quantify the Pulay force for two test systems, including a quasi-1D disordered Si system and a 3D disordered Si system. We find that the magnitude of the Pulay force is readily reduced to  $10^{-4}$  au with a few tens of basis functions per atom. To demonstrate that such accuracy is sufficient in practice, we compute the vibrational frequencies for the quasi-1D disordered Si system using a frozen phonon approach, and find that the vibrational frequencies agree well with those obtained from converged planewave calculations using ABINIT [34]. We further validate the accuracy of the computed forces through molecular dynamics simulations, and vibrational calculations of  $H_2$  molecules and pair-correlation functions of a liquid Al-Si alloy,

and we find excellent agreement with independent results in literature.

The remainder of this paper is organized as follows. Section 2 introduces the discontinuous Galerkin framework for Kohn-Sham density functional theory and the construction of the adaptive local basis functions. Section 3 discusses the computation of the Hellmann-Feynman force. We report numerical results in Section 4, followed by discussion and conclusions in Section 5.

## 2. Discontinuous Galerkin framework for Kohn-Sham density functional theory

### 2.1. Kohn-Sham density functional theory

We consider a system consisting of  $N_A$  nuclei and  $N$  electrons. In the Born-Oppenheimer approximation, for each set of nuclear positions  $\{R_I\}_{I=1}^{N_A}$ , the electrons are relaxed to their ground state. The ground state total energy is denoted by  $\mathcal{E}_{\text{tot}}(\{R_I\}_{I=1}^{N_A})$ , and can be computed in Kohn-Sham density functional theory [1, 2] according to

$$\mathcal{E}_{\text{tot}}(\{R_I\}_{I=1}^{N_A}) = \min_{\{\psi_i\}_{i=1}^N} \mathcal{E}_{\text{KS}}(\{\psi_i\}_{i=1}^N; \{R_I\}_{I=1}^{N_A}). \quad (1)$$

For simplicity, we assume all quantities are real, and neglect spin degeneracy, as well as temperature effects leading to fractional occupation. We also omit the range of indices  $I, i$  unless otherwise specified.  $\mathcal{E}_{\text{KS}}$  is the Kohn-Sham energy functional and is given by

$$\begin{aligned} & \mathcal{E}_{\text{KS}}(\{\psi_i\}; \{R_I\}) \\ &= \frac{1}{2} \sum_{i=1}^N \int |\nabla \psi_i(x)|^2 dx + \int V_{\text{loc}}(x; \{R_I\}) \rho(x) dx \\ &+ \sum_{i=1}^N \int \psi_i(x) V_{\text{nl}}(x, y; \{R_I\}) \psi_i(y) dx dy + \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x-y|} dx dy \\ &+ E_{\text{xc}}[\rho] + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|R_I - R_J|}. \end{aligned} \quad (2)$$

We include the  $\{R_I\}$  dependence explicitly to facilitate the derivation of atomic forces. Here,

$$\rho(x) = \sum_{i=1}^N |\psi_i(x)|^2 \quad (3)$$

is the electron density. The eigenfunctions (also called Kohn-Sham orbitals)  $\{\psi_i\}$  satisfy the orthonormality constraints

$$\int \psi_i(x) \psi_j(x) dx = \delta_{ij}. \quad (4)$$

In (2), we use a norm-conserving pseudopotential [3]. **Here**

$$V_{\text{loc}}(x; \{R_I\}) = \sum_{I=1}^{N_A} V_{\text{loc},I}(x - R_I) \quad (5)$$

is the local part of the pseudopotential. Each term  $V_{\text{loc},I}(x - R_I)$  is centered on the  $I$ -th atom, and decays asymptotically as  $-\frac{Z_I}{|x - R_I|}$  for large  $|x - R_I|$ , where  $Z_I$  is the charge of the  $I$ -th nucleus. The nonlocal part of the pseudopotential takes the Kleinman-Bylander form [35]

$$V_{\text{nl}}(x, y; \{R_I\}) = \sum_{I=1}^{N_A} \sum_{\ell=1}^{L_I} \gamma_{I,\ell} b_{I,\ell}(x - R_I) b_{I,\ell}(y - R_I). \quad (6)$$

For each atom  $I$ , there are  $L_I$  functions  $\{b_{I,\ell}\}$  called projectors of the nonlocal pseudopotential. Each  $b_{I,\ell}$  is centered at  $R_I$  and is supported locally in real space around  $R_I$ .  $\gamma_{I,\ell}$  is a real scalar.  $E_{\text{xc}}$  is the exchange-correlation energy. Here, we assume local or semi-local exchange-correlation functionals are used. The last term in Eq. (2) is the ion-ion Coulomb interaction energy. We note that for extended systems, modeled as infinite periodic structures, both the local-pseudopotential and ion-ion terms require special treatment in order to avoid divergences due to the long-range  $1/r$  nature of the Coulomb interaction. We provide corresponding expressions for this case in the appendix A.

When the atomic positions  $\{R_I\}$  are fixed, we may simplify the notation and drop the  $\{R_I\}$ -dependence in  $V_{\text{loc}}$  and  $V_{\text{nl}}$ . The Kohn-Sham equation is the Euler-Lagrange equation associated with (2):

$$H_{\text{eff}}[\rho]\psi_i = \left( -\frac{1}{2}\Delta + V_{\text{eff}}[\rho] + V_{\text{nl}} \right) \psi_i = E_i \psi_i. \quad (7)$$

Here the effective single-particle potential  $V_{\text{eff}}$  is defined as

$$V_{\text{eff}}[\rho](x) = V_{\text{loc}}(x) + V_H(x) + V_{\text{xc}}[\rho](x), \quad (8)$$

in which the Coulomb potential is given by

$$V_H(x) = \int \frac{\rho(y)}{|x - y|} dy. \quad (9)$$

$V_{\text{xc}}[\rho](x) = \frac{\delta E_{\text{xc}}}{\delta \rho}(x)$  is the exchange-correlation potential. Note that Eq. (7) is a nonlinear eigenvalue problem, as  $V_{\text{eff}}$  depends on  $\rho$ , which is in turn determined by  $\{\psi_i\}$ . The electron density is self-consistent if both (3) and (7) are satisfied. After obtaining the self-consistent electron density, the total energy of the system can be expressed using the eigenvalues  $\{E_i\}$  and density  $\rho$  as [3]

$$\begin{aligned} \mathcal{E}_{\text{tot}} = & \sum_{i=1}^N E_i - \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x - y|} dx dy + E_{\text{xc}}[\rho] - \int V_{\text{xc}}[\rho](x)\rho(x) dx \\ & + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|R_I - R_J|}. \end{aligned} \quad (10)$$

In the self-consistent field iteration for solving the Kohn-Sham equations, the total computational time is usually dominated by the following step: Given an input electron density and associated effective potential  $V_{\text{eff}}(x)$ , we find the output electron density  $\tilde{\rho}(x)$  from

$$\tilde{\rho}(x) = \sum_{i=1}^N |\psi_i(x)|^2, \quad (11)$$

where  $\{\psi_i\}$  are the lowest  $N$  eigenfunctions of  $H_{\text{eff}}$  in Eq. (7). The  $\{\psi_i\}$  then minimize the quadratic energy functional

$$\begin{aligned} \mathcal{E}_{\text{eff}}(\{\psi_i\}) = & \frac{1}{2} \sum_{i=1}^N \int |\nabla \psi_i(x)|^2 dx + \sum_{i=1}^N \int V_{\text{eff}}(x) |\psi_i(x)|^2 dx \\ & + \sum_{i=1}^N \int \psi_i(x) V_{\text{nl}}(x, y) \psi_i(y) dx dy, \end{aligned} \quad (12)$$

with  $\{\psi_i\}$  being orthonormal. Note that the ion-ion interaction is a constant depending only on  $\{R_I\}$ , and is dropped in Eq. (12).

## 2.2. Adaptive local basis and discontinuous Galerkin framework

In [19], the adaptive local basis functions in a discontinuous Galerkin (DG) framework have been proposed to reduce the computational time for solving the equations of KSDFT. The DG method relaxes the continuity constraint on basis functions, and provides flexibility in choosing the basis set for efficient discretization. Among the different formalisms in the DG framework, we use the interior penalty method [36, 37], which naturally generalizes the variational principle (12).

We denote by  $\Omega$  the computational domain with periodic boundary conditions, which corresponds to  $\Gamma$  point sampling in the Brillouin zone [3]. The domain  $\Omega$  is also referred to as the global domain in the following. More general Bloch boundary conditions may be accommodated as well. Let  $\mathcal{T}$  be a collection of quasi-uniform rectangular partitions of  $\Omega$ ,

$$\mathcal{T} = \{E_1, E_2, \dots, E_M\}, \quad (13)$$

and  $\mathcal{S}$  be the collection of surfaces that correspond to  $\mathcal{T}$ . Each  $E_k$  is called an element of  $\Omega$ . For a typical choice of partitions used in practice, the elements are chosen to be of the same size.

We define the following inner products:

$$\langle v, w \rangle_E = \int_E v(x)w(x) dx, \quad \langle \mathbf{v}, \mathbf{w} \rangle_S = \int_S \mathbf{v}(x) \cdot \mathbf{w}(x) ds(x), \quad (14)$$

$$\langle v, w \rangle_{\mathcal{T}} = \sum_{i=1}^M \langle v, w \rangle_{E_i}, \quad \langle \mathbf{v}, \mathbf{w} \rangle_{\mathcal{S}} = \sum_{S \in \mathcal{S}} \langle \mathbf{v}, \mathbf{w} \rangle_S. \quad (15)$$

In the interior penalty method, the energy functional corresponding to (12) is given by

$$\begin{aligned}
E_{\text{DG}}(\{\psi_i\}) = & \frac{1}{2} \sum_{i=1}^N \langle \nabla \psi_i, \nabla \psi_i \rangle_{\mathcal{T}} + \langle V_{\text{eff}}, \rho \rangle_{\mathcal{T}} + \sum_{I=1}^{N_A} \sum_{\ell=1}^{L_I} \gamma_{I,\ell} \sum_{i=1}^N |\langle b_{I,\ell}(\cdot - R_I), \psi_i \rangle_{\mathcal{T}}|^2 \\
& - \sum_{i=1}^N \langle \{\{\nabla \psi_i\}\}, [\psi_i] \rangle_{\mathcal{S}} + \alpha \sum_{i=1}^N \langle [\psi_i], [\psi_i] \rangle_{\mathcal{S}}.
\end{aligned} \tag{16}$$

Here,  $\{\{ \cdot \}\}$  and  $[\![ \cdot ]\!]$  are the average and the jump operators across surfaces, defined as follows. Because of the periodic boundary condition, each surface  $S \in \mathcal{S}$  is an interior surface in the sense that  $S$  is shared by elements  $K_1$  and  $K_2$ . Denote by  $n_1$  and  $n_2$  the unit normal vectors on  $S$  pointing exterior to  $K_1$  and  $K_2$ , respectively. With  $u_i = u|_{\partial K_i}$ ,  $i = 1, 2$ , we set

$$[\![ u ]\!] = u_1 n_1 + u_2 n_2 \quad \text{on } S \in \mathcal{S}. \tag{17}$$

For vector-valued function  $q$ , we define

$$\{\{ q \}\} = \frac{1}{2}(q_1 + q_2) \quad \text{on } S \in \mathcal{S}, \tag{18}$$

where  $q_i = q|_{\partial K_i}$ . The second to last term in Eq. (16) comes from integration by parts of the Laplacian operator, which cures the ill-defined operation of applying the Laplacian operator to discontinuous functions in order to define the kinetic energy. The last term in Eq. (16) is a penalty term which penalizes the jumps of functions across element surfaces to guarantee stability [38], and the constant  $\alpha$  is a positive penalty parameter. We have demonstrated that the adjustable penalty parameter  $\alpha$  is mainly used to ensure the stability of the numerical scheme, and has relatively little effect of the accuracy of the scheme when it takes a large range of values [19, 22].

Assume that we have chosen for each element  $E_k$  a set of basis functions  $\{\varphi_{k,j}\}_{j=1}^{J_k}$ , where  $J_k$  is the number of basis functions in  $E_k$ . We extend each  $\varphi_{k,j}$  to the whole computational domain  $\Omega$  by setting it to zero on the complement set of  $E_k$ . Define the function space  $\mathcal{V}$  as

$$\mathcal{V} = \text{span}\{\varphi_{k,j}, k = 1, \dots, M; j = 1, \dots, J_k\}. \tag{19}$$

The local basis functions  $\{\varphi_{k,j}\}_{j=1}^{J_k}$  which we use to discretize the Kohn-Sham problem are constructed as follows. For each  $E_k \in \mathcal{T}$ , we introduce an associated *extended element*  $Q_k \supset E_k$ , with  $Q_k \setminus E_k$  a buffer region surrounding  $E_k$ . We define  $V_{\text{eff}}^{Q_k} = V_{\text{eff}}|_{Q_k}$  to be the restriction of the effective potential at the current SCF step to  $Q_k$ , and  $V_{\text{nl}}^{Q_k} = V_{\text{nl}}|_{Q_k}$  to be the restriction of the nonlocal potential to  $Q_k$ . We solve the local eigenvalue problem on each extended element

$$\left( -\frac{1}{2}\Delta + V_{\text{eff}}^{Q_k} + V_{\text{nl}}^{Q_k} \right) \tilde{\varphi}_{k,j} = \lambda_{k,j} \tilde{\varphi}_{k,j}. \tag{20}$$



The lowest  $J_k$  eigenvalues  $\{\lambda_{k,j}\}_{j=1}^{J_k}$  and corresponding orthonormal eigenfunctions  $\{\tilde{\varphi}_{k,j}\}_{j=1}^{J_k}$  are computed. We then restrict  $\{\tilde{\varphi}_{k,j}\}_{j=1}^{J_k}$  from  $Q_k$  to  $E_k$ . The truncated functions are not necessarily orthonormal. Therefore, we apply a singular value decomposition (SVD) to obtain  $\{\varphi_{k,j}\}_{j=1}^{J_k}$ . The SVD procedure can ensure the orthonormality of the basis functions inside each element, as well as eliminating the linearly dependent and nearly linearly dependent functions in the basis set. We then extend each  $\varphi_{k,j}$  to the global domain by setting it to zero outside of  $E_k$ , so that it is in general discontinuous across the boundary of  $E_k$ . As a result, the overlap matrix corresponding to the adaptive local basis set is an identity matrix.

There are a number of possible ways to set the boundary conditions for the local problem (20). In practice, we use periodic boundary conditions for all eigenfunctions  $\{\tilde{\varphi}_{k,j}\}_{j=1}^{J_k}$  in  $Q_k$ . In some sense, the details of boundary condition do not affect the accuracy of the adaptive local basis set much as the buffer size increases. This permits the use of highly efficient Fourier based solution methods for the local problem. **Note that while the eigenfunctions are required to satisfy periodic boundary conditions, the potential is obtained from the restriction of the global Kohn-Sham potential to the extended element.** The size of each extended element should be chosen to balance between the effectiveness of the basis functions and the computational cost for obtaining them. For a typical choice used in practice, the elements are chosen to be of the same size, and each element contains on average a few atoms. The partition does not need to be updated when the atomic configuration is changed, as in the case of structure optimization and molecular dynamics.

After obtaining the basis functions, we minimize (16) for  $\{\psi_i\} \subset \mathcal{V}$ , i.e.,

$$\psi_i(x) = \sum_{k=1}^M \sum_{j=1}^{J_k} c_{i;k,j} \varphi_{k,j}(x). \quad (21)$$

The output electron density is then computed as

$$\tilde{\rho}(x) = \sum_{i=1}^N \sum_{k=1}^M \left| \sum_{j=1}^{J_k} c_{i;k,j} \varphi_{k,j}(x) \right|^2. \quad (22)$$

Note that the computation of the electron density can be performed locally in each element, since for each  $x \in \Omega$  there is a unique  $k \equiv k(x)$  such that  $\varphi_{k,j}(x) \neq 0$ . We refer readers to [19] for details of solving the minimization problem (16) as an eigenvalue problem in the DG formulation.

### 3. Calculation of atomic forces

Once the SCF iteration reaches convergence to yield converged electron density  $\rho(x)$  and Kohn-Sham orbitals  $\{\psi_i\}$ , the force on the  $I$ -th atom can be computed as the negative derivative of the total energy with respect to the atomic

position  $R_I$ :

$$F_I = -\frac{\partial \mathcal{E}_{\text{tot}}(\{R_I\})}{\partial R_I}. \quad (23)$$

The required derivative can be computed directly, e.g., via finite differences. However, even for first order accuracy, the number of energy evaluations for a system containing  $N_A$  atoms is  $3N_A + 1$ , i.e., the Kohn-Sham equations must be solved  $3N_A + 1$  times independently. This approach becomes prohibitively expensive as the system size increases. The cost of the force calculation is greatly reduced via the Hellmann-Feynman theorem, which states that, at self-consistency, the partial derivative  $\frac{\partial}{\partial R_I}$  only needs to be applied to terms in Eq. (2) which depend *explicitly* on the atomic position  $R_I$ . The Hellmann-Feynman (HF) force is then given by

$$\begin{aligned} F_I^{\text{HF}} = & -\int \frac{\partial V_{\text{loc}}(x; \{R_I\})}{\partial R_I} \rho(x) dx - \sum_{i=1}^N \int \psi_i(x) \frac{\partial V_{\text{nl}}(x, y; \{R_I\})}{\partial R_I} \psi_i(y) dx dy \\ & + \sum_{J \neq I} \frac{Z_I Z_J}{|R_I - R_J|^3} (R_I - R_J). \end{aligned} \quad (24)$$

Note that Eq. (5) gives

$$\frac{\partial V_{\text{loc}}(x; \{R_I\})}{\partial R_I} = \frac{\partial V_{\text{loc},I}}{\partial R_I}(x - R_I) = -\nabla_x V_{\text{loc},I}(x - R_I),$$

and similarly Eq. (6) gives

$$\begin{aligned} & \frac{\partial V_{\text{nl}}(x, y; \{R_I\})}{\partial R_I} \\ = & \sum_{\ell=1}^{L_I} \gamma_{I,\ell} \left( \frac{\partial b_{I,\ell}}{\partial R_I}(x - R_I) b_{I,\ell}(y - R_I) + b_{I,\ell}(x - R_I) \frac{\partial b_{I,\ell}}{\partial R_I}(y - R_I) \right) \\ = & -\sum_{\ell=1}^{L_I} \gamma_{I,\ell} (\nabla_x b_{I,\ell}(x - R_I) b_{I,\ell}(y - R_I) + b_{I,\ell}(x - R_I) \nabla_y b_{I,\ell}(y - R_I)). \end{aligned}$$

Then the Hellmann-Feynman force in Eq. (24) can be written as

$$\begin{aligned} F_I^{\text{HF}} = & \int \nabla_x V_{\text{loc},I}(x - R_I) \rho(x) dx \\ & + 2 \sum_{i=1}^N \sum_{\ell=1}^{L_I} \gamma_{I,\ell} \left( \int \psi_i(x) \nabla_x b_{I,\ell}(x - R_I) dx \right) \left( \int \psi_i(y) b_{I,\ell}(y - R_I) dy \right) \\ & + \sum_{J \neq I} \frac{Z_I Z_J}{|R_I - R_J|^3} (R_I - R_J). \end{aligned} \quad (25)$$

From the computational cost point of view, if we denote by  $N_g$  the number of grid points to discretize quantities such as  $\rho(x)$  in the global domain, then the cost of computing each integral in the form  $\int \nabla_x V_{\text{loc},I}(x - R_I) \rho(x) dx$  is  $\mathcal{O}(N_g)$ , since  $V_{\text{loc},I}(x - R_I)$  is a delocalized quantity in the global domain. On the other hand, each nonlocal projector  $b_{I,\ell}(x - R_I)$  is localized around  $R_I$ , and the cost of evaluating the integral  $(\int \psi_i(x) \nabla_x b_{I,\ell}(x - R_I) dx)$  or  $(\int \psi_i(y) b_{I,\ell}(y - R_I) dy)$  is a constant  $N_l$  independent of the global number of grid points  $N_g$ . The computation of the last term  $\sum_{J \neq I} \frac{Z_I Z_J}{|R_I - R_J|^3} (R_I - R_J)$  involves only scalar operations, and its cost is usually negligibly small in electronic structure calculations.  $N_g$  and  $N_A$  are proportional to the number of electrons  $N$ . Hence, neglecting constant terms independent of  $N$ , we have that the computational cost of the Hellmann-Feynman force on each atom is  $\mathcal{O}(N_g + N L_I N_l) \sim \mathcal{O}(N)$ , so that the cost for all atoms is  $\mathcal{O}(N^2)$ .

We now demonstrate that with auxiliary quantities, the cost of computing the Hellmann-Feynman forces on all atoms can be reduced to  $\mathcal{O}(N)$  without loss of accuracy in the DG formulation. The strategy is different for the local pseudopotential term and nonlocal pseudopotential term, respectively.

For the local term, let us rewrite  $V_{\text{loc},I}(x - R_I)$  as

$$V_{\text{loc},I}(x - R_I) = - \int \frac{\rho_{\text{loc},I}(y - R_I)}{|x - y|} dy. \quad (26)$$

The term  $\rho_{\text{loc},I}$  is called the ionic pseudocharge density, or smeared ionic density corresponding to the local pseudopotential [39, 40], and the minus sign in Eq. (26) reflects the opposite sign of electronic and ionic charge. While  $V_{\text{loc},I}$  is delocalized in the global domain due to the long-range Coulomb interaction, the corresponding pseudocharge  $\rho_{\text{loc},I}$  is localized around the nuclear position  $R_I$ , similar to the projector  $b_{I,\ell}$  of the nonlocal pseudopotential. Then,

$$\begin{aligned} \int \nabla_x V_{\text{loc},I}(x - R_I) \rho(x) dx &= \int \nabla_x \left( - \int \frac{\rho_{\text{loc},I}(y - R_I)}{|x - y|} dy \right) \rho(x) dx \\ &= \int \rho_{\text{loc},I}(y - R_I) \nabla_y \left( \frac{1}{|x - y|} \right) \rho(x) dx dy \\ &= \int \rho_{\text{loc},I}(y - R_I) \nabla_y \left( \int \frac{\rho(x)}{|x - y|} dx \right) dy \\ &= \int \rho_{\text{loc},I}(y - R_I) \nabla_y V_H(y) dy. \end{aligned} \quad (27)$$

Here,  $V_H$  is the Coulomb potential defined in Eq. (9). Note that  $V_H$  and its gradient only needs to be evaluated once for all atoms. Due to the locality of the ionic pseudocharge, the cost of numerical integration using the last equality in Eq. (27) is independent of the global number of grid points  $N_g$  and number of electrons  $N$ . So the computation of the local pseudopotential part of the Hellmann-Feynman force for all atoms scales as  $\mathcal{O}(N)$ .

For the nonlocal part of the pseudopotential, the cost arises from the fact that all  $N$  Kohn-Sham orbitals need to participate in the evaluation of the force

for each atom  $I$ . The cost can be reduced by an alternative formulation using the density matrix, defined as

$$P(x, y) = \sum_{i=1}^N \psi_i(x) \psi_i(y). \quad (28)$$

Recalling the expansion of the Kohn-Sham orbitals in Eq. (21) in terms of the adaptive local basis functions, the density matrix can also be expanded as

$$P(x, y) = \sum_{i=1}^N \sum_{k=1}^M \sum_{j=1}^{J_k} \sum_{k'=1}^M \sum_{j'=1}^{J_{k'}} \varphi_{k,j}(x) \varphi_{k',j'}(y) c_{i;k,j} c_{i;k',j'}.$$

Omitting the range of summation for  $k, j, k', j'$ , the nonlocal part of the force becomes

$$\begin{aligned} & 2 \sum_{i=1}^N \sum_{\ell=1}^{L_I} \left( \int \psi_i(x) \nabla_x b_{I,\ell}(x - R_I) dx \right) \left( \int \psi_i(y) b_{I,\ell}(y - R_I) dy \right) \\ &= 2 \sum_{\ell=1}^{L_I} \int \nabla_x b_{I,\ell}(x - R_I) P(x, y) b_{I,\ell}(y - R_I) dx dy \\ &= 2 \sum_{\ell=1}^{L_I} \sum_{k,j,k',j'} (V_{\text{nl},I,\ell})_{k,j;k',j'} P_{k,j;k',j'} \equiv 2 \sum_{\ell=1}^{L_I} \text{Tr}[V_{\text{nl},I,\ell} P]. \end{aligned} \quad (29)$$

Here,

$$(V_{\text{nl},I,\ell})_{k,j;k',j'} = \left( \int \varphi_{k,j}(x) \nabla_x b_{I,\ell}(x - R_I) dx \right) \left( \int \varphi_{k',j'}(y) b_{I,\ell}(y - R_I) dy \right)$$

is the nonlocal pseudopotential matrix element for the  $\ell$ -th nonlocal pseudopotential projector of the  $I$ -th atom, represented in the adaptive local basis set, and

$$P_{k,j;k',j'} = \sum_{i=1}^N c_{i;k,j} c_{i;k',j'}$$

is the corresponding density matrix element represented in the adaptive local basis set. Since the basis functions and projectors are localized, the cost of evaluating each matrix element of  $V_{\text{nl},I,\ell}$  is independent of  $N$ , and the number of nonzeros in  $V_{\text{nl},I,\ell}$  is independent of  $N$  as well. Thus the cost of evaluating the trace operation in Eq. (29) for each atom  $I$  is independent of  $N$ . Therefore, if the density matrix represented in the adaptive local basis set is computed, the computational cost of the nonlocal pseudopotential component of the force for all atoms scales as  $\mathcal{O}(N)$  as well. We remark that diagonalization-free numerical methods for computing the density matrix represented in the adaptive local basis set are available. They include linear scaling methods [41, 42] for insulating systems, and the recently developed pole expansion and selected inversion

(PEXSI) method [43–45] for both insulating and metallic systems. The DG Hamiltonian matrix corresponding to the adaptive local basis set has a block stencil like sparsity structure and the overlap matrix is an identity matrix, and is therefore well suited for such diagonalization free methods.

In sum, the Hellmann-Feynman force can be compactly written and computed in the DG formulation as

$$F_I^{\text{HF}} = \int \rho_{\text{loc},I}(x - R_I) \nabla_x V_H(x) dx + 2 \sum_{\ell=1}^{L_I} \text{Tr}[V_{\text{nl},I,\ell} P] + \sum_{J \neq I} \frac{Z_I Z_J}{|R_I - R_J|^3} (R_I - R_J). \quad (30)$$

The computational cost to obtain the forces on all atoms is  $\mathcal{O}(N)$ , provided that the density matrix represented in the adaptive local basis set is computed, neglecting the small cost of the ion-ion term.

In general, the force  $F_I$  in Eq. (23) and Hellmann-Feynman force  $F_I^{\text{HF}}$  in Eq. (30) are not the same when the basis used in discretization depends on  $R_I$ . But they can be the same if one of the following three conditions is satisfied

1. The basis set is complete.
2. The basis set is not complete, but is independent of the atomic positions. This is the case, e.g., for plane-wave basis functions.
3. The basis set is not complete, but the basis set can be embedded in a larger subspace which is independent of atomic positions, and the basis set achieves the minimal energy among all choices of basis sets within the larger subspace. This is the case, e.g., for the recently proposed optimized local basis (OLB) functions [33], which is a variant of the ALB in the DG framework.

If none of the conditions above is satisfied, as e.g., for Gaussian type orbitals (GTO) or atomic orbitals (AO), an additional term called the Pulay force [32] (denoted by  $F_I^{\text{Pulay}}$ ) is needed so that

$$F_I = F_I^{\text{HF}} + F_I^{\text{Pulay}}. \quad (31)$$

Since the Pulay force arises from the  $\{R_I\}$ -dependence of the basis, it is present for the adaptive local basis also. However, due to the local optimality of the construction in the vicinity of each element, the Pulay force is small and readily reduced to chemical accuracy, as we show below.

#### 4. Results and discussion

In order to ascertain the accuracy and convergence of the obtained Hellmann-Feynman forces in the adaptive local basis, we consider a range of test cases, including quasi-1D and 3D, metallic and insulating. We first consider static configurations, where we compare directly to converged plane-wave calculations

to determine the absolute accuracy and convergence of the computed forces and size of the Pulay component as the number of ALBs is increased. We then employ the forces so obtained to compute vibrational frequencies using a frozen phonon approach and pair correlation functions from *ab initio* molecular dynamics simulations, comparing to independent calculations and previous work.

The ALB and DG based calculation is carried out using our new massively parallel code called DGDFE [22]. Along each dimension ( $x, y, z$ ), the size of the extended element is fixed to be 3 times the size of the element, unless there is only 1 element along this dimension. Since periodic boundary conditions are assumed for the extended elements, we use the planewave basis set to expand the ALBs in the extended element. It follows from the standard planewave convention that the number of grid points in the extended element is determined by the kinetic energy cutoff denoted by  $E_{\text{cut}}$ . The relationship between  $E_{\text{cut}}$  and the number of uniform grid points along the  $i$ th direction ( $N_i$ ), where  $i \in \{x, y, z\}$ , can be written as

$$N_i = \frac{\sqrt{2E_{\text{cut}}L_i}}{\pi}. \quad (32)$$

where  $L_i$  is the dimension of the extended element along the  $i$ th direction. Since the wavefunctions are generally smoother than the electron density and potential, a second set of uniform grids is employed to accurately represent the density and potential with a higher kinetic energy cutoff  $E_{\text{cut}}^{\text{den}}$ . In all of our calculations, we set  $E_{\text{cut}}^{\text{den}} = 4E_{\text{cut}}$ . In addition to the wavefunction and density grids, we generate another grid which we call the Legendre-Gauss-Lobatto (LGL) grid on each element to perform accurate numerical quadrature. To be consistent in notation, the number of LGL grid points along each dimension is also defined in terms of a kinetic energy cutoff denoted by  $E_{\text{cut}}^{\text{LGL}}$ , **where for accuracy consistent with  $E_{\text{cut}}$ , we set  $E_{\text{cut}}^{\text{LGL}} = 4E_{\text{cut}}$ .** We note that  $E_{\text{cut}}^{\text{LGL}}$  does not carry any physical meaning for LGL grids, but is merely used as a convenient notation for fixing the number of grid points from an equation analogous to Eq. (32). More details of the implementation in DGDFE can be found in [19, 22].

For all calculations here, we use the local density approximation (LDA) [46, 47] for exchange and correlation, and Hartwigsen-Goedecker-Hutter (HG) [48] pseudopotentials to model the ions.

We use the ABINIT planewave DFT code [34] as reference to assess the accuracy of our calculations. The same exchange-correlation functionals and pseudopotentials are employed in ABINIT and DGDFE so that results can be compared directly, and errors assessed rigorously.

All calculations were carried out on the Hopper and Edison systems at the National Energy Research Scientific Computing Center (NERSC). There are 24 processors on each computational node on both Hopper and Edison, with 32 and 64 gigabyte (GB) of memory per node, respectively.

#### 4.1. Static calculations

We first examine the accuracy of the Hellmann-Feynman force for quasi-1D and 3D disordered Si systems.

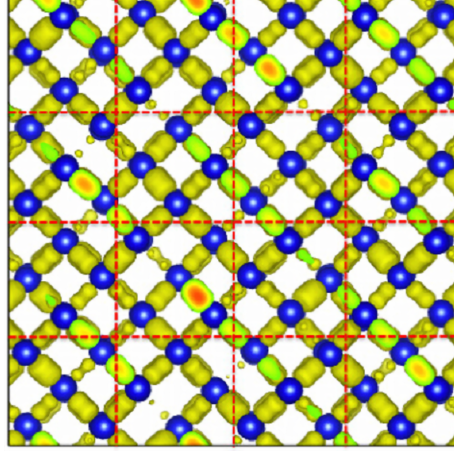


Figure 1: (Color online) 3D-bulk disordered Si system with 216 atoms is partitioned (red dashed lines) into 4 elements along each direction, 64 elements in total, viewed along the  $z$  direction. Blue balls represent Si atoms, yellow surfaces represent the charge density at a given isosurface, and green/orange surfaces represent cross sections of charge density at boundaries. The disorder of the structure is manifested in the differences between bonds.

For the quasi-1D Si system, we replicate the 8-atom unit cell with diamond structure along the  $z$  direction 4 times, forming a  $1 \times 1 \times 4$  structure with 32 atoms in the supercell. The  $1 \times 1 \times 4$  Si system is partitioned into 6 elements along the  $z$  direction resulting in about 5 atoms per element. For the 3D bulk Si system, we replicate the unit cell along  $x$ ,  $y$ , and  $z$  directions 3 times, to form a  $3 \times 3 \times 3$  structure with 216 atoms in the supercell. The  $3 \times 3 \times 3$  structure is partitioned into 4 elements along each direction. **This gives 64 elements in total, and the size of each element along each direction is 7.65 Bohr.** Each element contains around 3 atoms. The partition of the  $3 \times 3 \times 3$  Si structure is shown in Fig. 1. The lattice constant of the Si unit cell is 10.2 au. The positions of atoms in both Si systems are slightly perturbed by applying a random displacement uniformly distributed within  $[-0.2, 0.2]$  au.

To assess accuracy and convergence, we define the force error to be

$$\max_I |\Delta F_I|, \quad (33)$$

where  $I$  is the atom index,

$$\Delta F_I = F_I^{\text{HF}} - F_I^{\text{ABINIT}},$$

with  $F_I^{\text{HF}}$  the Hellmann-Feynman force computed by DGDFE and  $F_I^{\text{ABINIT}}$  the fully converged force computed by ABINIT. The magnitude of the force error, denoted by  $|\Delta F_I|$ , is then given by

$$|\Delta F_I| = \sqrt{\Delta F_{I,x}^2 + \Delta F_{I,y}^2 + \Delta F_{I,z}^2},$$

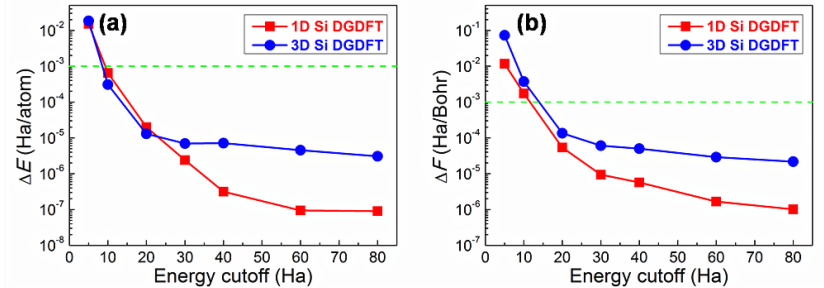


Figure 2: (Color online) Convergence of DGDFT total energy and forces for quasi-1D and 3D Si systems to reference plane-wave results with increasing kinetic energy cutoff  $E_{\text{cut}}$  (Ha). (a) Total energy error per atom  $\Delta E$  (Ha/atom). (b) Maximum atomic force error  $\Delta F$  (Ha/Bohr). The dashed green line corresponds to chemical accuracy. The number of ALB functions per atom is 45.0 and 71.1 for quasi-1D and 3D Si systems, respectively, and the penalty parameter  $\alpha = 40$ . When  $E_{\text{cut}}$  is sufficiently large, the error is well below the target accuracy (green dashed line).

where  $\Delta F_{I,x}$ ,  $\Delta F_{I,y}$ , and  $\Delta F_{I,z}$  represent the  $x$ ,  $y$ , and  $z$  components of  $\Delta F_I$ , respectively.

The main contributors to the force error defined by (33) are: 1) numerical quadrature error that results from an insufficient number of grid points required to perform numerical integration of several quantities such as ALBs, density, and potential; 2) an insufficient number of ALBs required to accurately represent the Kohn-Sham wavefunctions. We shall refer the latter as basis set error. Note that the basis set error contributes directly to the Pulay force through the incompleteness of the basis.

In order to examine the Pulay force and compare it with the error (33), we first reduce the numerical quadrature error by increasing the kinetic energy cutoff  $E_{\text{cut}}$ . Specifically, we start from a large number of ALBs, and increase the  $E_{\text{cut}}$  in DGDFT gradually until the difference between the total energy obtained from DGDFT and that obtained from a fully converged ABINIT calculation is negligible, at least two orders of magnitude below chemical accuracy. Once a sufficient  $E_{\text{cut}}$  is determined, we use this energy cutoff in subsequent calculations to examine how the force error changes with respect to the number of basis functions per element.

#### 4.1.1. Convergence with respect to energy cutoff

We first check the quadrature errors of the total energy and atomic forces with respect to the kinetic energy cutoff. Figure 2 shows the convergence of the DGDFT total energy and forces for quasi-1D and 3D Si systems to reference plane-wave results with increasing kinetic energy cutoff  $E_{\text{cut}}$ .

To achieve convergence of the total energy in ABINIT, we increase  $E_{\text{cut}}$  until the change of the computed total energy is below  $10^{-8}$  Ha per atom. We find that the smallest  $E_{\text{cut}}$  that achieves this level of convergence is 80 Ha. We then take  $E_{\text{cut}} = 100$  Ha results as reference for subsequent error computations.



In the DGDFT calculations, we use a large number of basis functions (45.0 and 71.1 ALBs per atom respectively for the quasi-1D and bulk 3D Si systems) to ensure that the basis set error is small. When  $E_{\text{cut}}$  is increased to 60 Ha in DGDFT, the difference between the DGDFT total energy and converged ABINIT total energy is below  $10^{-5}$  Ha per atom. Figure 2 indicates that we may set  $E_{\text{cut}}$  to 60 Ha in DGDFT to make quadrature errors negligible in subsequent calculations.

#### 4.1.2. Convergence with respect to number of ALB functions

We next check the basis set error of the total energy and atomic forces in DGDFT. Figure 3 shows the convergence of the total energy and atomic forces for the quasi-1D and 3D Si systems with increasing number of ALB functions per atom. We set the energy cutoff to 60 Ha and the penalty parameter  $\alpha$  to 40. First, we see that chemical accuracy is obtained with less than 10 basis functions per atom for the quasi-1D system and a few tens of basis functions per atom for the 3D system. Furthermore, for the quasi-1D Si system, we find that the total energy error can be as small as  $2.78 \times 10^{-8}$  Ha/atom and the maximum error of the atomic forces can be as small as  $8.47 \times 10^{-7}$  Ha/Bohr when 22.5 ALB functions per atom are used. For the bulk 3D Si system, the total energy error can be as small as  $2.10 \times 10^{-6}$  Ha/atom and the maximum error in atomic forces can be as small as  $1.70 \times 10^{-5}$  Ha/Bohr when 71.11 ALB functions per atom are used.

We note that the errors for the 3D system are generally larger than for the quasi-1D system for comparable number of ALB functions per atom. This is due to the fact that the ALB functions are discontinuous at element boundaries and the boundary surface to volume ratio in 3D systems is larger than in 1D systems. Indeed, when we increase the size of the 3D elements to 10.2 Bohr along each direction with 54 basis functions per atom, the the total energy error is reduced to  $9.49 \times 10^{-9}$  Ha/atom and the maximum error in atomic forces is reduced to  $5.12 \times 10^{-6}$  Ha/Bohr.

#### 4.1.3. Penalty parameter

The penalty parameter  $\alpha$  in Eq. (16) is used to penalize potential large jumps of Kohn-Sham orbitals between neighboring elements. If  $\alpha$  is too small, the resulting numerical error can be large. This leads to significantly lower total energy and is easily detected. On the other hand, as the penalty parameter is increased beyond a certain point, energies and forces remain stable and so the method is insensitive to the choice [19, 22].

Figure 4 shows how the errors in the total energy and atomic forces change with respect to  $\alpha$  for the quasi-1D and 3D Si systems. We set the energy cutoff to 100 Ha and choose 45 ALBs per atom in the quasi-1D case and 71.1 ALBs per atom in the 3D case to obtain highly accurate results. For the quasi-1D system, when  $\alpha$  is increased from 4 to 400, the error in total energy remains around  $2.78 \times 10^{-8}$  Ha/atom, while the maximum error in atomic forces decreases slightly from  $9.41 \times 10^{-7}$  to  $8.47 \times 10^{-7}$  Ha/Bohr. For the bulk 3D Si system, the error in total energy changes from  $2.31 \times 10^{-6}$  Ha/atom to  $1.06 \times 10^{-6}$  Ha/atom

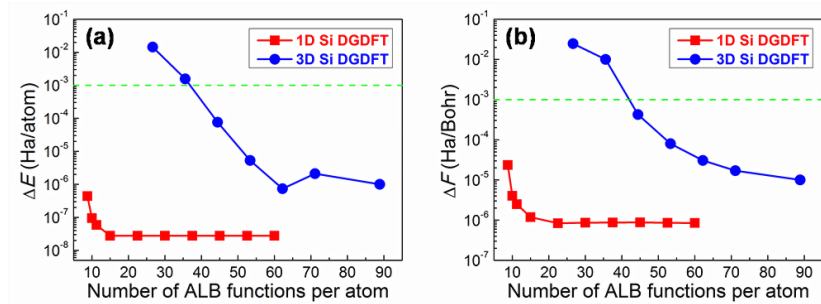


Figure 3: (Color online) Convergence of DGDFT total energy and forces for quasi-1D and 3D Si systems to reference planewave results with increasing number of ALB functions per atom. (a) Total energy error per atom  $\Delta E$  (Ha/atom). (b) Maximum atomic force error  $\Delta F$  (Ha/Bohr). The dashed green line corresponds to chemical accuracy. Energy cutoff  $E_{\text{cut}} = 60$  Ha and penalty parameter  $\alpha = 40$ . When the number of basis functions per atom is sufficiently large, the error is well below the target accuracy (green dashed line).

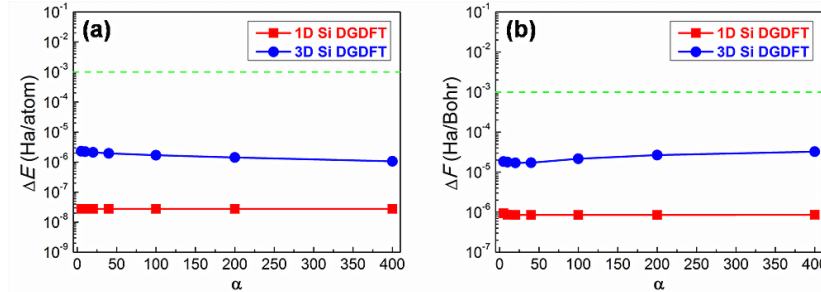


Figure 4: (Color online) Total energy error per atom  $\Delta E$  (Ha/atom) and maximum atomic force error  $\Delta F$  (Ha/Bohr) for quasi-1D and 3D Si systems as a function of penalty parameter  $\alpha$  in DGDFT calculations. The dashed green line corresponds to chemical accuracy. Kinetic energy cutoff  $E_{\text{cut}} = 100$  Ha and ALB functions per atom is 45.0 and 71.1 for quasi-1D and 3D Si systems, respectively. For all values of  $\alpha$ , the error is well below the target accuracy (green dashed line).

when  $\alpha$  is increased from 4 to 400, while the maximum error in atomic forces changes from  $1.72 \times 10^{-5}$  Ha/Bohr to  $3.24 \times 10^{-5}$  Ha/Bohr. In both cases, we see that the error remains stable over a broad range of penalty parameter values, so that the method is generally insensitive to the choice.

#### 4.1.4. Pulay force

In order to compute the Pulay force  $F_I^{\text{Pulay}}$  defined by Eq. (31), we use a second order finite difference method with grid spacing 0.05 au to compute the required derivatives to obtain the force  $F_I$  (Eq. (23)), and subtract the Hellmann-Feynman force  $F_I^{\text{HF}}$  (Eq. (30)). For all examples, we report the Pulay force  $F_I^{\text{Pulay}}$  for the atom  $I$  with the largest absolute force error in magnitude.

The computed DGDFT force error (33) and Pulay force (31) for the quasi-1D Si system are shown in Table 1 for several choices of #ALB/atom. We find

Table 1: DGDFT force error ( $\Delta F$ ) for quasi-1D Si system and Pulay force ( $F^{\text{Pulay}}$ ) with respect to the number of adaptive local basis functions per atom. #ALB/atom indicates the number of adaptive local basis functions per atom. Forces are in units of Ha/Bohr.

#ALB/atom	$\Delta F_x$	$\Delta F_y$	$\Delta F_z$	$F_x^{\text{Pulay}}$	$F_y^{\text{Pulay}}$	$F_z^{\text{Pulay}}$
9.0	$1.26 \times 10^{-4}$	$7.38 \times 10^{-5}$	$-6.55 \times 10^{-4}$	$2.45 \times 10^{-4}$	$-6.13 \times 10^{-5}$	$-6.61 \times 10^{-4}$
15.8	$-1.18 \times 10^{-5}$	$3.00 \times 10^{-6}$	$8.68 \times 10^{-5}$	$-1.29 \times 10^{-5}$	$-1.21 \times 10^{-5}$	$5.08 \times 10^{-5}$
25.5	$2.92 \times 10^{-6}$	$-4.27 \times 10^{-7}$	$-3.60 \times 10^{-6}$	$1.82 \times 10^{-6}$	$4.46 \times 10^{-6}$	$4.00 \times 10^{-7}$

Table 2: DGDFT force error for 3D bulk Si system ( $\Delta F$ ) and Pulay force ( $F^{\text{Pulay}}$ ) with respect to the number of adaptive local basis functions per atom. #ALB/atom indicates the number of adaptive local basis functions per atom. Force errors are in units of Ha/Bohr.

#ALB/atom	$\Delta F_x$	$\Delta F_y$	$\Delta F_z$	$F_x^{\text{Pulay}}$	$F_y^{\text{Pulay}}$	$F_z^{\text{Pulay}}$
35.6	$1.78 \times 10^{-4}$	$-1.14 \times 10^{-4}$	$4.94 \times 10^{-3}$	$3.63 \times 10^{-3}$	$4.79 \times 10^{-3}$	$5.40 \times 10^{-3}$
47.4	$-9.38 \times 10^{-5}$	$-6.00 \times 10^{-5}$	$-6.68 \times 10^{-5}$	$6.12 \times 10^{-5}$	$8.20 \times 10^{-5}$	$-1.53 \times 10^{-4}$
71.1	$1.50 \times 10^{-5}$	$-5.58 \times 10^{-6}$	$-1.32 \times 10^{-5}$	$3.38 \times 10^{-5}$	$1.62 \times 10^{-5}$	$-1.83 \times 10^{-5}$

that the DGDFT force error decreases rapidly as the number of ALBs per atom increases. The force error is on the order of  $10^{-5}$  au or below when approximately 16 ALBs per atom are used. If we increase the number of ALBs further, the force error can decrease to  $10^{-6}$  au or smaller. We also see from Table 1 that the computed Pulay forces are on the same order as the total force error. This observation suggests that, after reducing the numerical quadrature error, the error in forces is comparable in magnitude to the Pulay contribution, which decreases rapidly with the number of adaptive local basis functions used. Therefore, DGDFT force error can be readily reduced to a level typically required in molecular dynamics and geometry optimization applications, around  $10^{-3}$  au with  $\sim 10$  basis functions/atom.

The same trend is observed for the 3D bulk Si system, as shown in Table 2. For the 3D bulk system, using around 71 ALBs per atom allows us to reduce the force error to  $10^{-4}$  Ha/Bohr or below. The latter system requires a few times more basis functions per atom due to the 3D element partition, consistent with previous findings [19] for total energies.

#### 4.1.5. Translation and boundary effects

In electronic structure calculations of periodic systems, the total energy and atomic forces should remain the same when all atoms are translated along a given direction. However, DGDFT employs a fixed element partition, and therefore can give slightly different energies and forces for atomic configurations that differ by translation only. We call this the translation effect. Furthermore, DGDFT uses discontinuous basis functions, hence one might expect differing errors when atoms are near boundaries versus far away. We call this the boundary effect. Below we study these effects by translating the disordered 3D Si<sub>216</sub> system relative to the fixed element partition along X, Y, and Z directions. In each of the shifted configurations, there are a number of atoms on or close to element boundaries. Figure 5 shows two such examples, shifting the system in the positive and negative directions along the X axis. We report the corre-

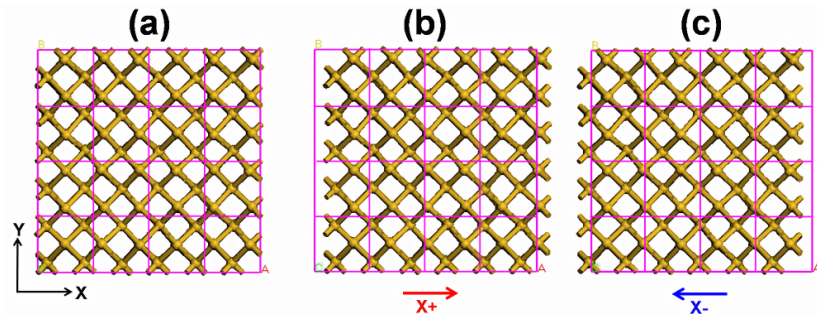


Figure 5: (Color online) Three dimensional Si<sub>216</sub> systems translated along the X direction about 0.018 Bohr. (a) Original structure of Si<sub>216</sub>. (b) Structure translated along positive X direction (Si<sub>216</sub>X+). (c) Structure translated along negative X direction (Si<sub>216</sub>X-).

Table 3: Errors in total energy  $\Delta E$  (Ha/atom) and atomic forces  $\Delta F$  (Ha/Bohr) for bulk 3D Si systems translated along X, Y, and Z directions. The energy cutoff is 60 Ha and the number of ALB functions per atom is 71.1.

3D Si systems	$\Delta E$	$\Delta F$
Si <sub>216</sub>	$4.56 \times 10^{-6}$	$2.92 \times 10^{-5}$
Si <sub>216</sub> X+	$8.60 \times 10^{-6}$	$6.47 \times 10^{-5}$
Si <sub>216</sub> X-	$9.08 \times 10^{-6}$	$6.31 \times 10^{-5}$
Si <sub>216</sub> Y+	$7.28 \times 10^{-6}$	$7.05 \times 10^{-5}$
Si <sub>216</sub> Y-	$1.08 \times 10^{-5}$	$5.39 \times 10^{-5}$
Si <sub>216</sub> Z+	$8.80 \times 10^{-6}$	$6.51 \times 10^{-5}$
Si <sub>216</sub> Z-	$1.13 \times 10^{-5}$	$5.58 \times 10^{-5}$

sponding errors in the total energy and atomic forces in Table 3. We set the energy cutoff to 60 Ha and the number of ALB functions per atom to 71.1. We see that translation and boundary effects remain small for all configurations, as expected given the construction of ALBs for each configuration.

#### 4.1.6. Linear scaling cost of force computation

Here we demonstrate that the computation of atomic forces in DGDFE scales linearly with respect to system size. Figure 6 shows the wall time for computing the atomic forces with respect to the number of silicon atoms in a series of bulk 3D Si systems. We set the energy cutoff to 40 Ha and use 64 ALB functions per atom in the calculations. For all calculations, we use 4096 computational cores, on which we obtain wall times for computing forces of 0.35s, 1.21s, 2.20s, and 5.22s for Si<sub>64</sub>, Si<sub>512</sub>, Si<sub>1728</sub>, and Si<sub>4096</sub>, respectively. Hence, the cost associated with atomic force computation in DGDFE is a small fraction in the total, and increases linearly with the number of atoms in the system.

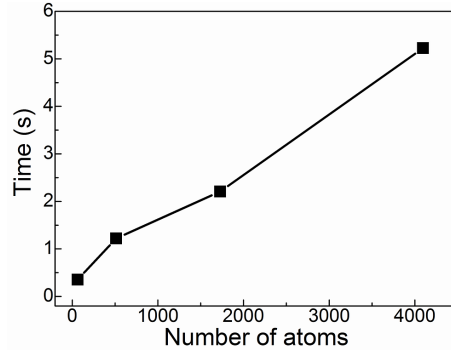


Figure 6: Wall time for computing atomic forces for a series of bulk Si systems (Si<sub>64</sub>, Si<sub>512</sub>, Si<sub>1728</sub> and Si<sub>4096</sub>) using 4096 computational cores. The energy cutoff is 40 Ha with 64 ALB functions per atom.

#### 4.1.7. Vibrational frequencies

Vibrational frequencies are important observables for characterizing material properties at finite temperature. For a system at equilibrium atomic configuration, the dynamical matrix is defined as

$$D_{I;J} = \frac{1}{\sqrt{M_I M_J}} \frac{\partial^2 \mathcal{E}_{\text{tot}}(\{R_K^*\})}{\partial R_I \partial R_J},$$

which is a square matrix of dimension  $3N_A$ . Here  $\{R_K^*\}$  is the equilibrium atomic position for which all atomic forces vanish. The square root of its eigenvalues gives the vibrational frequencies. The computation of vibrational frequencies requires highly accurate force calculations since the second order derivatives of the energy are needed.

We compute the vibrational frequencies of the quasi-1D Si system using the frozen phonon method. The second derivatives of the total energy are computed by applying a central difference formula to the Hellmann-Feynman forces computed by DGDFT as follows. **Starting from the perfect crystal configuration  $\{R_K^*\}$ , we have**

$$\begin{aligned} \frac{\partial^2 \mathcal{E}_{\text{tot}}(\{R_K^*\})}{\partial R_I \partial R_J} &\approx \frac{1}{2h} \left( \frac{\partial \mathcal{E}_{\text{tot}}(\{R_K^* + \delta R_K^{(I)}\})}{\partial R_J} - \frac{\partial \mathcal{E}_{\text{tot}}(\{R_K^* - \delta R_K^{(I)}\})}{\partial R_J} \right) \\ &= \frac{1}{2h} \left( F_J(\{R_K^* - \delta R_K^{(I)}\}) - F_J(\{R_K^* + \delta R_K^{(I)}\}) \right). \end{aligned}$$

Here the change of the atomic configuration is denoted by  $\delta R^{(I)}$ , which is a vector of length  $3N_A$ . The vector  $\delta R^{(I)}$  has only one nonzero entry corresponding to the  $X$ ,  $Y$ , or  $Z$  direction of the position of the  $I$ -th atom  $R_I$ , and the value of this entry is denoted by  $h$ . In all calculations we choose  $h = 0.01$  Bohr. The first order derivative of the energy is then directly computed as the negative of the atomic force. Hence each phonon calculation for the quasi-1D Si system

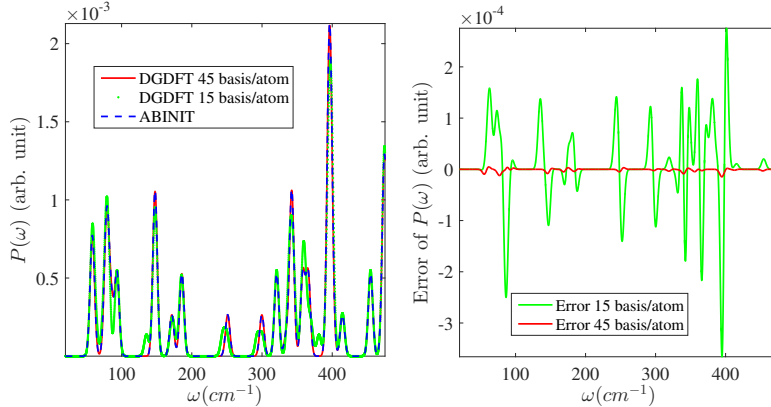


Figure 7: (Color online) Vibrational spectrum for quasi-1D Si obtained from DGDFT using 45 basis functions per atom and 15 basis functions per atom, and ABINIT. The error is measured against the result from ABINIT.

requires  $6 \times 32 = 192$  independent DGDFT calculations for the shifted configurations, and no symmetry is explicitly used. We note that the purpose of the present calculation is to demonstrate the applicability of ALBs for the computation of such second derivative information, rather than to produce a physically meaningful vibrational spectrum, which would require a considerably larger supercell and/or  $k$ -point sampling to obtain. At the equilibrium configuration, all eigenvalues of the dynamical matrix are non-negative, and the vibrational frequencies are the square roots of the eigenvalues of  $D$ , i.e.,

$$Du_k = \omega_k^2 u_k. \quad (34)$$

The vibrational spectrum is the spectral density of the vibrational frequencies smeared by a Gaussian distribution, which is defined as

$$P(\omega) = \sum_k \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(\omega-\omega_k)^2}{2\sigma^2}}. \quad (35)$$

with smearing parameter  $\sigma = 10\text{cm}^{-1}$ .

To compute vibrational frequencies, we set the kinetic energy cutoff to 60 Ha in both DGDFT and ABINIT, and use 15 and 45 ALBs per atom in DGDFT. As shown in Fig. 7, we find that the vibrational frequencies obtained from DGDFT converge to those from ABINIT as the number of ALBs is increased, with excellent agreement at 45 ALBs/atom, as expected given the agreement of DGDFT and ABINIT forces shown in Section 4.1.2.

#### 4.2. Molecular dynamics

We have shown for randomly displaced configurations that the DGDFT force error can be as small as  $10^{-4}$  Ha/Bohr with a moderate number of adaptive local

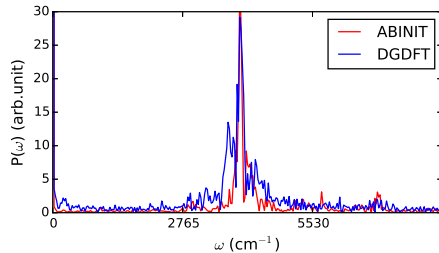


Figure 8: (Color online) Vibrational frequency of  $\text{H}_2$  molecule obtained from DGDFT (blue solid line) and ABINIT (red dashed line) at 300 K.

basis functions per atom. In this section, we demonstrate that this level of accuracy, as for standard planewave methods, yields converged *ab initio* molecular dynamics simulations as well.

We consider two systems: one molecular and the other condensed matter. The first system consists of four  $\text{H}_2$  molecules **initially** evenly spaced in a box of dimension  $10 \times 10 \times 40$  au. The other system is a liquid Al-Si alloy. When performing molecular dynamics simulations in DGDFT, we partition the four  $\text{H}_2$  molecules into 4 elements of equal size along the  $z$  direction, and partition the Al-Si alloy into  $4 \times 4 \times 4$  elements of equal size. These two systems constitute a quasi-1D structure and 3D structure, respectively.

#### 4.2.1. $\text{H}_2$

The four  $\text{H}_2$  molecules are simulated in the constant-temperature-constant-volume (NVT) ensemble at 300 K, using a Nose-Hoover thermostat [49, 50]. The integration time step is 25 au ( $\sim 0.6$  fs), and the thermostat masses  $Q_s$  are set to 20 au for both DGDFT and ABINIT. In DGDFT, we set the kinetic energy cutoff to 160 Ha. With 10 adaptive local basis functions per atom, the force computed by DGDFT is accurate to  $10^{-6}$  au in the initial configuration. We perform 2000 MD simulation steps in both ABINIT and DGDFT, which corresponds to 1.2 ps of simulation time.

The vibrational spectrum is given by the Fourier transform of the bond length of one  $\text{H}_2$  molecule along the MD trajectory. Fig. 8 (blue solid line) shows the vibrational spectrum calculated using DGDFT for the first  $\text{H}_2$  molecule. The general profile of the calculated spectrum agrees well with that obtained from ABINIT (red dashed line). The characteristic vibrational frequencies estimated by DGDFT and ABINIT differ by only a few  $\text{cm}^{-1}$ . Although we obtain good quantitative agreement from the 1.2 ps trajectories, longer trajectories may be required to eliminate the multiple peaks around the characteristic frequency.

#### 4.2.2. Liquid Al-Si alloy

Simulations of liquid Al-Si alloy  $\text{Al}_{0.88}\text{Si}_{0.12}$  were performed in a 200-atom cell with lattice constant 15.68 Å in the canonical ensemble at target temperature 973K, as in Ref. [51]. In the initial configuration, Al and Si atoms are

randomly placed in the cubic cell. The system is then relaxed at 4000 K, well above the target temperature, so that any biases in the initial configuration are removed. We then cool down the system from 4000 K to 973 K gradually, at the rate of 0.471 K/fs, and continue the simulation in the canonical ensemble at 973 K for more than 2.0 ps to equilibrate. The simulation continues under these conditions for another 7.0 ps, and data is collected during this period to compute the properties of the Al-Si alloy. We use a Nose-Hoover chain thermostat. In DGDFT, we set the integration time step for the equation of motion to 100 au ( $\sim 2.42$  fs), the thermostat mass to 10000 au, the energy cutoff to 10 Ha, and employ a basis of 51.2 adaptive local basis functions per atom on average. The Nose-Hoover thermostat has a conserved energy, which can be used to check the consistency of energies and forces. We find a drift of the conserved energy of 2.6 meV/ps/atom, consistent with the high accuracy of forces.

We next compare the statistical properties of the liquid Al-Si alloy obtained from DGDFT with those obtained from SIESTA [39], and results presented in Ref. [51], obtained from the PARSEC electronic structure code. In particular, we examine the pair correlation function, a statistical quantity widely used to characterize liquid structure and coordination. For alloy systems, the total pair correlation is given by:

$$g^{\text{tot}}(r) = \frac{1}{b^2}(x_i^2 b_i^2 g_{ii}(r) + 2x_i x_j b_i b_j g_{ij}(r) + x_j^2 b_j^2 g_{jj}(r)) \quad (36)$$

according to the Faber-Ziman formalism [52]. Here, the indices  $i$  and  $j$  indicate different types of atoms,  $x_i$ ,  $x_j$  are the corresponding molar fractions,  $b_i$ ,  $b_j$  are neutron scattering lengths or x-ray form factors,  $g_{ij}(r)$  are the partial pair correlation functions, and  $b = x_i b_i + x_j b_j$ . For Al and Si, we use the neutron scattering length ratio  $b_{Al}/b_{Si} = 0.8318$ .

In Fig. 9, we show the total pair correlation functions  $g(r)$  computed using DGDFT and SIESTA, and results from Ref. [51] (labeled NVE). The curve produced by DGDFT matches well with the other two results, especially near the first peak. The liquid structure and coordination are thus well described by all three codes. Slight differences among the three are likely due to differences in pseudopotentials, basis sets, and ensembles used.

In Ref. [51], the functions  $G(r) = r(g(r) - 1)$  and  $G_{ij}(r) = r(g_{ij}(r) - 1)$  are considered. We plot the corresponding quantities in Fig. 10. The  $G(r)$ ,  $G_{Al-Al}(r)$ , and  $G_{Al-Si}(r)$  obtained by all three codes agree well. The agreement of the curves, especially up to the first minimum, indicates that DGDFT describes the Al-Si and Al-Al bond lengths and coordinations well. In addition to the differences in basis sets used in three methods, the number of atoms used in the DGDFT and SIESTA simulations is different from that used in the PARSEC simulations. The result in Ref. [51] in Fig. 10 used a 500-atom Al-Si system, while 200 atoms were used in the DGDFT and SIESTA simulations. The difference in number of atoms can also contribute to the slight differences in computed pair correlation functions at larger  $r$ . Unlike the other partial pair correlation functions, the differences in  $G_{Si-Si}(r)$  among the three methods is more noticeable. As noted in Ref. [51], the uncertainties are primarily a result



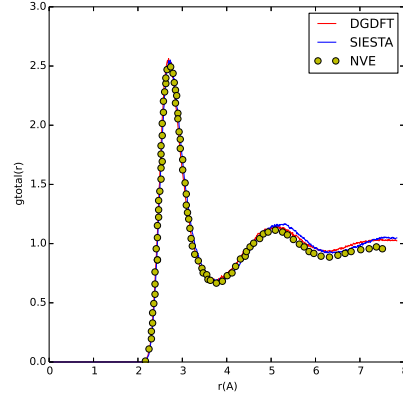


Figure 9: (Color online) Comparison of total pair correlation function of liquid  $Al_{0.88}Si_{0.12}$  calculated at 973 K from 200-atom simulations: DGDFT (red solid line), SIESTA (blue solid line), and previous work [51] (yellow dots).

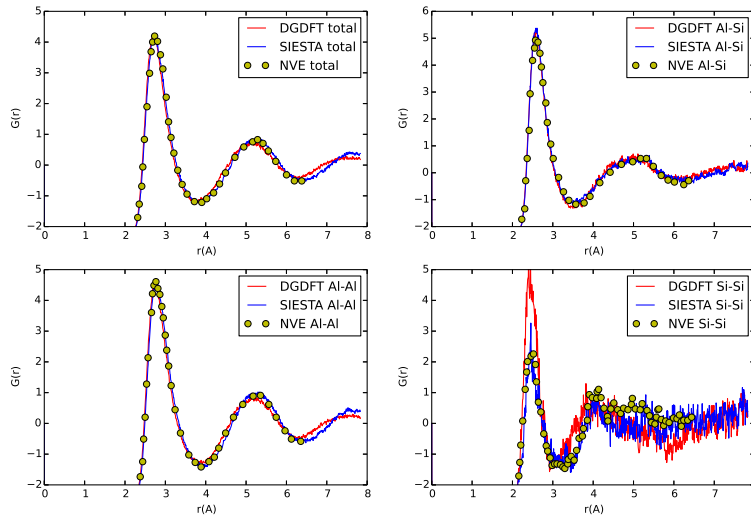


Figure 10: (Color online)  $G(r)$  derived from partial and total pair correlation functions for liquid  $Al_{0.88}Si_{0.12}$  calculated at 973K from 200-atom simulations using DGDFT (red solid line) and SIESTA (blue solid line), and from 500-atom simulations in previous work [51] (yellow dots).

of insufficient statistics for the Si, which occurs in much smaller number than Al in the present alloy.

## 5. Conclusion

In this work, we detail the calculation of atomic forces in Kohn-Sham density functional theory using an adaptive local basis. We demonstrate that, under mild assumptions, the computation of atomic forces can scale nearly linearly with the number of atoms in the system using the adaptive local basis set. The method is implemented in the recently developed massively parallel DGDFT electronic structure code. We quantify the accuracy of the Hellmann-Feynman forces for a range of physical systems, benchmarked against converged plane-wave calculations, and find that the adaptive local basis set is efficient for both force and energy calculations, requiring at most a few tens of basis functions per atom to attain accuracies required in practice. Since the adaptive local basis set has implicit dependence on atomic positions, Pulay forces are in general nonzero. However, by virtue of the adaptive local construction, we find that the Pulay force is small and systematically decreasing with increasing basis completeness; so that the Hellmann-Feynman force is sufficient for basis sizes of a few tens of basis functions per atom. We verified the accuracy of computed Hellmann-Feynman forces in static calculations of quasi-1D and 3D Si systems, finding convergence of forces to  $10^{-4}$  Ha/Bohr with at most a few tens of basis functions per atom in all cases. We further verified the accuracy of the computed forces in frozen phonon calculations of the vibrational spectrum of quasi-1D Si, and molecular dynamics simulations of H<sub>2</sub> molecules and liquid Al-Si alloy, finding in all cases excellent agreement with benchmark plane-wave results and literature.

Whereas the Hellmann-Feynman force is sufficient for adaptive local bases of a few tens of basis functions per atom or more, if still smaller bases are desired, as for example in density-matrix based  $\mathcal{O}(N)$  calculations, then Pulay forces will become significant. In such case, optimized local basis functions (OLB) [33] may be a better candidate than adaptive local basis functions (ALB) since the Pulay force can be reduced systematically and substantially. However, the optimized local basis functions introduce other difficulties in implementation, particularly for 3D systems. We plan to investigate this in the future.

## Acknowledgments

This work was supported by the Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research and Basic Energy Sciences (G. Z., L. L., W. H., C. Y. and J. P.), and the Center for Applied Mathematics for Energy Research Applications (CAMERA), which is a partnership between Basic Energy Sciences (BES) and Advanced Scientific Computing Research (ASRC) at the U.S. Department of Energy (L. L. and C. Y.). This work was performed, in part, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. We thank the

National Energy Research Scientific Computing (NERSC) center for providing the computational resources.

## Appendix A. Appendix: Treatment of electrostatic interactions in extended systems

In this appendix, we discuss how extended periodic systems can be treated, and provide corresponding expressions for the total energy and atomic forces. The most widely used method for treating periodic systems relies on the use of Ewald summation for long-range interactions, which appear both in the local part of the pseudopotential and in the ion-ion interaction. An alternative approach is to use the pseudocharge formulation as given in Eq. (26) to replace long-range ionic potentials by corresponding short-range charge densities, eliminating the need for Ewald summation altogether [39, 40]. The advantage of such a formulation is that the electrons and ions are treated on the same footing, in a single Poisson solution, permitting both electron-ion and ion-ion interactions to be evaluated at once in  $\mathcal{O}(N)$  operations.

The total ionic pseudocharge in a unit cell is the sum of pseudocharges from all atoms:

$$\rho_{\text{loc}}(x) = \sum_{I=1}^{N_A} \rho_{\text{loc},I}(x - R_I), \quad (\text{A.1})$$

where  $N_A$  includes all atoms with nonvanishing pseudocharge  $\rho_{\text{loc},I}$  in the unit cell. The Kohn-Sham energy functional can then be written as

$$\begin{aligned} & \mathcal{E}_{\text{KS}}(\{\psi_i\}; \{R_I\}) \\ &= \frac{1}{2} \sum_{i=1}^N \int |\nabla \psi_i(x)|^2 dx + \sum_{i=1}^N \int \psi_i(x) V_{\text{nl}}(x, y; \{R_I\}) \psi_i(y) dx dy \\ & \quad + \frac{1}{2} \int (\rho(x) - \rho_{\text{loc}}(x)) V_C(x) dx + E_{\text{xc}}[\rho] - E_s. \end{aligned} \quad (\text{A.2})$$

Since  $\int \rho(x) - \rho_{\text{loc}}(x) dx = 0$  in the unit cell, the total Coulomb potential  $V_C$  due to ionic and electronic charge densities can be obtained by solving the Poisson equation

$$-\Delta V_C(x) = 4\pi(\rho(x) - \rho_{\text{loc}}(x)) \quad (\text{A.3})$$

subject to periodic boundary conditions on the unit cell.  $E_s$  is the self-interaction energy of pseudocharges  $\rho_{\text{loc},I}$  in the unit cell, a constant independent of atomic positions [40]. Having computed the total Coulomb potential  $V_C$ , the Hellmann-Feynman force is then given by

$$F_I^{\text{HF}} = \int \rho_{\text{loc},I}(x - R_I) \nabla_x V_C(x) dy + 2 \sum_{\ell=1}^{L_I} \text{Tr}[V_{\text{nl},I,\ell} P]. \quad (\text{A.4})$$

Compared to Eq. (30), the force due to the ion-ion interaction is taken into account by the total Coulomb potential.

- [1] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, *Phys. Rev.* 136 (1964) B864–B871.
- [2] W. Kohn, L. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* 140 (1965) A1133–A1138.
- [3] R. Martin, *Electronic structure: basic theory and practical methods*, Cambridge Univ. Press, Cambridge, 2004.
- [4] M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, J. D. Joannopoulos, Iterative minimization techniques for *ab initio* total energy calculation: molecular dynamics and conjugate gradients, *Rev. Mod. Phys.* 64 (1992) 1045–1097.
- [5] J. Chelikowsky, N. Troullier, Y. Saad, Finite-difference-pseudopotential method: Electronic structure calculations without a basis, *Phys. Rev. Lett.* 72 (1994) 1240–1243.
- [6] E. Tsuchida, M. Tsukada, Electronic-structure calculations based on the finite-element method, *Phys. Rev. B* 52 (1995) 5573–5578.
- [7] J. Pask, B. Klein, C. Fong, P. Sterne, Real-space local polynomial basis for solid-state electronic-structure calculations: A finite-element approach, *Phys. Rev. B* 59 (1999) 12352–12358.
- [8] J. Pask, P. Sterne, Finite element methods in ab initio electronic structure calculations, *Modelling Simul. Mater. Sci. Eng.* 13 (2005) R71.
- [9] H. Chen, X. Dai, X. Gong, L. He, A. Zhou, Adaptive finite element approximations for Kohn–Sham models, *Multiscale Model. Simul.* 12 (2014) 1828–1869.
- [10] G. Bao, G. Hu, D. Liu, Numerical solution of the Kohn-Sham equation by finite element methods with an adaptive mesh redistribution technique, *J. Sci. Comput.* 55 (2) (2013) 372–391.
- [11] F. Averill, D. Ellis, An efficient numerical multicenter basis set for molecular orbital calculations: application to FeCl<sub>4</sub>, *J. Chem. Phys.* 59 (1973) 6412–6418.
- [12] B. Delley, D. Ellis, Efficient and accurate expansion methods for molecules in local density models, *J. Chem. Phys.* 76 (1982) 1949–1960.
- [13] H. Eschrig, *Optimized LCAO Method and the Electronic Structure of Extended Systems*, Akademie Verlag and Springer, Berlin, 1988.
- [14] K. Koepnik, H. Eschrig, Full-potential nonorthogonal local-orbital minimum-basis band-structure scheme, *Phys. Rev. B* 59 (1999) 1743–1757.
- [15] S. Kenny, A. Horsfield, H. Fujitani, Transferable atomic-type orbital basis sets for solids, *Phys. Rev. B* 62 (2000) 4899–4905.

- [16] J. Junquera, O. Paz, D. Sanchez-Portal, E. Artacho, Numerical atomic orbitals for linear-scaling calculations, *Phys. Rev. B* 64 (2001) 235111.
- [17] T. Ozaki, Variationally optimized atomic orbitals for large-scale electronic structures, *Phys. Rev. B* 67 (2003) 155108.
- [18] 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) 2175–2196.
- [19] 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) 2140–2154.
- [20] B. Cockburn, G. Karniadakis, C.-W. Shu, Discontinuous Galerkin methods: Theory, computation and applications, Vol. 11 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 2000.
- [21] T. Lu, W. Cai, J. Xin, Y. Guo, Linear scaling discontinuous Galerkin density matrix minimization method with local orbital enriched finite element basis: 1-d lattice model system, *Commun. Comput. Phys.* 14 (2013) 276–300.
- [22] W. Hu, L. Lin, C. Yang, DGDFT: A massively parallel method for large scale density functional theory calculations, *J. Chem. Phys.* 143 (2015) 124110.
- [23] J. Kaye, L. Lin, C. Yang, A posteriori error estimator for adaptive local basis functions to solve Kohn-Sham density functional theory, *Commun. Math. Sci.* 13 (2015) 1741.
- [24] C. Skylaris, P. Haynes, A. Mostofi, M. Payne, Introducing ONETEP: Linear-scaling density functional simulations on parallel computers, *J. Chem. Phys.* 122 (2005) 084119.
- [25] S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, S. Goedecker, Daubechies wavelets for linear scaling density functional theory, *J. Chem. Phys.* 140 (2014) 204110.
- [26] M. J. Rayson, P. R. Briddon, Highly efficient method for Kohn-Sham density functional calculations of 500–10 000 atom systems, *Phys. Rev. B* 80 (2009) 205104.
- [27] A. S. Banerjee, L. Lin, W. Hu, C. Yang, J. E. Pask, Chebyshev polynomial filtered subspace iteration in the Discontinuous Galerkin method for large-scale electronic structure calculations, arXiv:1606.03416.
- [28] L. Lin, L. Ying, Element orbitals for Kohn-Sham density functional theory, *Phys. Rev. B* 85 (2012) 235144.

- [29] W. Hu, L. Lin, C. Yang, Edge reconstruction in armchair phosphorene nanoribbons revealed by discontinuous Galerkin density functional theory, *Phys. Chem. Chem. Phys.* 17 (2015) 31397.
- [30] H. Hellmann, *Einführung in die Quantumchemie*, Franz Deuticke, Leipzig, 1937.
- [31] R. Feynman, Forces in molecules, *Phys. Rev.* 56 (1939) 340.
- [32] P. Pulay, Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules I. Theory, *Mol. Phys.* 17 (1969) 197–204.
- [33] L. Lin, J. Lu, L. Ying, W. E, Optimized local basis function for Kohn-Sham density functional theory, *J. Chem. Phys.* 231 (2012) 4515.
- [34] X. Gonze, B. Amadon, P. Anglade, J. M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Cote, et al., Abinit: First-principles approach to material and nanosystem properties, *Comput. Phys. Commun.* 180 (2009) 2582–2615.
- [35] L. Kleinman, D. Bylander, Efficacious form for model pseudopotentials, *Phys. Rev. Lett.* 48 (1982) 1425–1428.
- [36] I. Babuška, M. Zlámal, Nonconforming elements in the finite element method with penalty, *SIAM J. Numer. Anal.* 10 (1973) 863 – 875.
- [37] D. N. Arnold, An interior penalty finite element method with discontinuous elements, *SIAM J. Numer. Anal.* 19 (1982) 742 – 760.
- [38] D. N. Arnold, F. Brezzi, B. Cockburn, L. D. Marini, Unified analysis of discontinuous Galerkin methods for elliptic problems, *SIAM J. Numer. Anal.* 39 (2002) 1749.
- [39] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, D. Sánchez-Portal, The SIESTA method for ab initio order-N materials simulation, *J. Phys.: Condens. Matter* 14 (2002) 2745–2779.
- [40] J. Pask, P. Sterne, Real-space formulation of the electrostatic potential and total energy of solids, *Phys. Rev. B* 71 (2005) 113101.
- [41] S. Goedecker, Linear scaling electronic structure methods, *Rev. Mod. Phys.* 71 (1999) 1085.
- [42] D. R. Bowler, T. Miyazaki, O(N) methods in electronic structure calculations, *Rep. Prog. Phys.* 75 (2012) 036503.
- [43] L. Lin, A. García, G. Huhs, C. Yang, SIESTA-PEXSI: Massively parallel method for efficient and accurate *ab initio* materials simulation without matrix diagonalization, *J. Phys.: Condens. Matter* 26 (2014) 305503.

- [44] L. Lin, M. Chen, C. Yang, L. He, Accelerating atomic orbital-based electronic structure calculation via pole expansion and selected inversion, *J. Phys.: Condens. Matter* 25 (2013) 295501.
- [45] 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, *Commun. Math. Sci.* 7 (2009) 755–777.
- [46] D. Ceperley, B. Alder, Ground state of the electron gas by a stochastic method, *Phys. Rev. Lett.* 45 (1980) 566–569.
- [47] J. Perdew, A. Zunger, Self-interaction correction to density-functional approximations for many-electron systems, *Phys. Rev. B* 23 (1981) 5048–5079.
- [48] C. Hartwigsen, S. Goedecker, J. Hutter, Relativistic separable dual-space gaussian pseudopotentials from h to rn, *Phys. Rev. B* 58 (1998) 3641.
- [49] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, *J. Chem. Phys.* 81 (1984) 511–519.
- [50] W. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* 31 (1985) 1695–1697.
- [51] K. H. Khoo, T. L. Chan, M. Kim, J. R. Chelikowsky, Ab initio molecular dynamics simulations of molten  $\text{Al}_{1-x}\text{Si}_x$  alloys, *Phy. Rev. B* 84 (2011) 214203.
- [52] T. E. Faber, J. M. Ziman, A theory of the electrical properties of liquid metals: III. the resistivity of binary alloys, *Phil. Mag.* 11 (1965) 153–173.