

SQDFT: Spectral Quadrature method for large-scale parallel $\mathcal{O}(N)$ Kohn-Sham calculations at high temperature

Phanish Suryanarayana^{*,a}, Phanisri P. Pratapa^a, Abhiraj Sharma^a, John E. Pask^b

^aCollege of Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

^bPhysics Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Abstract

We present SQDFT: a large-scale parallel implementation of the Spectral Quadrature (SQ) method for $\mathcal{O}(N)$ Kohn-Sham Density Functional Theory (DFT) calculations at high temperature. Specifically, we develop an efficient and scalable finite-difference implementation of the infinite-cell Clenshaw-Curtis SQ approach, in which results for the infinite crystal are obtained by expressing quantities of interest as bilinear forms or sums of bilinear forms, that are then approximated by spatially localized Clenshaw-Curtis quadrature rules. We demonstrate the accuracy of SQDFT by showing systematic convergence of energies and atomic forces with respect to SQ parameters to reference diagonalization results, and convergence with discretization to established planewave results, for both metallic and insulating systems. We further demonstrate that SQDFT achieves excellent strong and weak parallel scaling on computer systems consisting of tens of thousands of processors, with near perfect $\mathcal{O}(N)$ scaling with system size and wall times as low as a few seconds per self-consistent field iteration. Finally, we verify the accuracy of SQDFT in large-scale quantum molecular dynamics simulations of aluminum at high temperature.

Key words: Electronic structure, Linear scaling, Metallic systems, High temperature, Quantum molecular dynamics, High performance computing, Parallel computing

1. Introduction

Kohn-Sham Density Functional Theory (DFT) [1, 2] is a powerful tool for predicting and understanding a wide range of materials properties, from the first principles of quantum mechanics, with no empirical or adjustable parameters. The tremendous popularity of DFT is a consequence of its high accuracy to cost ratio relative to other such ab initio theories. However, the solution of the Schrödinger type eigenproblem for the Kohn-Sham orbitals remains a challenging task. In particular, since the orbitals need to be orthogonal and increase in number linearly with the number of atoms N , the overall computational complexity of DFT calculations scales as $\mathcal{O}(N^3)$ and the memory requirement scales as $\mathcal{O}(N^2)$ (see, e.g., [3, 4, 5]). The orthogonality constraint on the orbitals also results in global communications between processors in parallel computing, which limits parallel scalability. The need for high performance parallel computing is especially crucial

*Corresponding Author (phanish.suryanarayana@ce.gatech.edu)

for quantum molecular dynamics (QMD) calculations [6, 7], wherein tens or hundreds of thousands of Kohn-Sham solutions can be required to complete a single simulation.

In order to overcome the critical $\mathcal{O}(N^3)$ scaling bottleneck, much research in the past two decades has been devoted to the development of $\mathcal{O}(N)$ solution strategies (see, e.g., [3, 4, 5] and references therein). Rather than calculate the orthonormal Kohn-Sham orbitals, these techniques directly determine the electron density, energy, and atomic forces in $\mathcal{O}(N)$ operations by exploiting the decay of the density matrix [8, 9, 10, 11, 12].¹ These efforts have yielded significant advances, culminating in mature implementations of a number of approaches [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. However, significant challenges remain. In particular, the accuracy and stability of $\mathcal{O}(N)$ methods remain ongoing concerns due to the need for additional computational parameters, subtleties in determining sufficient numbers and/or centers of localized orbitals, limitations of underlying basis sets, and calculation of accurate atomic forces, as required for structural relaxation and molecular dynamics simulations [4, 26]. In addition, efficient large-scale parallelization poses a significant challenge due to complex communications patterns and load balancing issues. Finally, and perhaps most importantly, the assumption of a band gap in the electronic structure makes existing methods inapplicable to metallic systems [4, 5].²

High-temperature DFT calculations present additional challenges [27, 28]. Such calculations have a number of applications, including the study of warm dense matter and dense plasmas, as occur in laser experiments, and the interiors of giant planets and stars [27, 28, 29, 30, 31, 32]. Particular challenges include the need for a significantly larger number of orbitals to be computed, as the number of partially occupied states increases, and need for more diffuse orbitals, as higher-energy states become less localized. Consequently, $\mathcal{O}(N^3)$ methods as well as local-orbital based $\mathcal{O}(N)$ methods have very large prefactors, which makes high-temperature QMD calculations for even small systems intractable. Recent work to address these challenges includes orbital-free molecular dynamics (OFMD) [33] wherein the standard Kohn-Sham kinetic energy is replaced by an approximation in terms of the density, extended first principles molecular dynamics (ext-FPMD) [34] wherein higher-energy states are approximated as planewaves rather than computed explicitly, and finite-temperature potential functional theory (PFT) [35] wherein an orbital-free free energy approximation is constructed through a coupling-constant formalism. While OFMD can miss electronic shell structure effects [36], ext-FPMD and PFT have been shown to capture such effects in initial applications.

The recently developed Spectral Quadrature (SQ) method for $\mathcal{O}(N)$ Kohn-Sham calculations [37, 38] addresses both scaling with number of atoms and scaling with temperature, while retaining systematic convergence to standard $\mathcal{O}(N^3)$ results for metals and insulators alike. In this approach, all quantities of interest are expressed as bilinear forms or sums of bilinear forms, which are then approximated by Clenshaw-Curtis quadrature rules that remain spatially localized by exploiting the locality of electronic interactions in real-space [39], i.e, the exponential decay of the

¹The real-space density matrix has exponential decay for insulating systems as well as metallic systems at finite temperature [8, 12].

²This is also the case for insulating systems at sufficiently high temperature which results in conduction bands becoming partially occupied.

density matrix in real-space for insulators as well as metals at finite temperature. In conjunction with local reformulation of the electrostatics, this technique enables the $\mathcal{O}(N)$ evaluation of the electronic density, energy, and atomic forces. The computational cost of SQ decreases rapidly with increasing temperature due to the enhanced locality of the electronic interactions and the increased smoothness of the Fermi-Dirac function. Further, it is well suited to scalable high-performance parallel computing since a majority of the communication is localized to nearby processors, whose pattern remains fixed throughout the simulation. The SQ approach also permits infinite-crystal calculations without recourse to Brillouin zone integration or large supercells, a technique referred to as the infinite-cell method [38].

In this paper, we present SQDFT: a parallel implementation of the SQ method for $\mathcal{O}(N)$ Kohn-Sham DFT calculations at high temperature.³ Specifically, we develop a finite-difference implementation of the infinite-cell variant of the Clenshaw-Curtis SQ method that can efficiently scale on large-scale parallel computers. We verify the accuracy of SQDFT by showing systematic convergence of energies and atomic forces to reference diagonalization results, and convergence with discretization to established planewave results, for both metallic and insulating systems. We further show that SQDFT achieves excellent strong and weak parallel scaling on computer systems consisting of tens of thousands of cores, with near perfect $\mathcal{O}(N)$ scaling with system size and wall times as low as a few seconds per self-consistent field (SCF) iteration. Finally, we verify the accuracy of SQDFT in large-scale quantum molecular dynamics simulations of aluminum at high temperature.

The remainder of this paper is organized as follows. In Section 2, we review the $\mathcal{O}(N)$ density matrix formulation of DFT. Next, we discuss the formulation and implementation of SQDFT in Section 3 and study its accuracy, efficiency, and scaling in Section 4. Finally, we provide concluding remarks in Section 5.

2. $\mathcal{O}(N)$ Density Functional Theory

Consider a cuboidal domain Ω containing N atoms, the unit cell of an infinite crystal. Let the nuclei be positioned at $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N\}$ and let there be a total of N_e valence electrons. Neglecting spin and Brillouin zone integration, the nonlinear eigenproblem for the electronic ground state in Kohn-Sham Density Functional Theory (DFT) can be written as [38, 40]

$$\mathcal{D} = g(\mathcal{H}, \mu, \sigma) = \left(1 + \exp\left(\frac{\mathcal{H} - \mu\mathcal{I}}{\sigma}\right) \right)^{-1}, \quad (1)$$

$$\mathcal{H} = -\frac{1}{2}\nabla^2 + V_{xc} + \phi + \mathcal{V}_{nl}, \quad (2)$$

where \mathcal{D} is the density matrix; g is the Fermi-Dirac function; μ is the Fermi level, which is determined by solving for the constraint on the total number of electrons, i.e., $2\text{Tr}(\mathcal{D}) = N_e$; $\sigma = k_B T$

³Though we focus on high-temperature calculations in this work, SQDFT is also capable of performing $\mathcal{O}(N)$ DFT calculations at ambient temperature, with a larger prefactor (Appendix A).

is the smearing, where k_B is Boltzmann's constant and T is the electronic temperature⁴; \mathcal{H} is the Hamiltonian; V_{xc} is the exchange-correlation potential; \mathcal{V}_{nl} is the nonlocal pseudopotential; and ϕ is the electrostatic potential, the solution to the Poisson equation [41, 42, 43, 44]

$$-\frac{1}{4\pi}\nabla^2\phi(\mathbf{x}, \mathbf{R}) = \rho_{\mathcal{D}}(\mathbf{x}) + b(\mathbf{x}, \mathbf{R}) \quad (3)$$

subject to periodic boundary conditions. Above,

$$\rho_{\mathcal{D}}(\mathbf{x}) = 2\mathcal{D}(\mathbf{x}, \mathbf{x}) \quad (4)$$

is the electron density and $b = \sum_I b_I$ is the total pseudocharge density of the nuclei [41], where b_I is the pseudocharge density of the I^{th} nucleus and the index I extends over all atoms in \mathbb{R}^3 .

Once the electronic ground state has been determined, the free energy can be written as [38]⁵

$$\begin{aligned} \mathcal{F}(\mathbf{R}) &= 2\text{Tr}(\mathcal{D}\mathcal{H}) + E_{xc}(\rho_{\mathcal{D}}) - \int_{\Omega} V_{xc}(\rho_{\mathcal{D}}(\mathbf{x}))\rho_{\mathcal{D}}(\mathbf{x}) \, d\mathbf{x} + \frac{1}{2} \int_{\Omega} (b(\mathbf{x}, \mathbf{R}) - \rho_{\mathcal{D}}(\mathbf{x}))\phi(\mathbf{x}, \mathbf{R}) \, d\mathbf{x} \\ &+ \frac{1}{2} \int_{\Omega} (\tilde{b}(\mathbf{x}, \mathbf{R}) + b(\mathbf{x}, \mathbf{R})) V_c(\mathbf{x}, \mathbf{R}) \, d\mathbf{x} - \frac{1}{2} \sum_I \int_{\Omega} \tilde{b}_I(\mathbf{x}, \mathbf{R}_I) \tilde{V}_I(\mathbf{x}, \mathbf{R}_I) \, d\mathbf{x} \\ &+ 2\sigma\text{Tr}(\mathcal{D} \log \mathcal{D} + (\mathcal{I} - \mathcal{D}) \log(\mathcal{I} - \mathcal{D})) , \end{aligned} \quad (5)$$

where E_{xc} is the exchange-correlation energy; $\tilde{b} = \sum_I \tilde{b}_I$ is the total reference pseudocharge density, where \tilde{b}_I is the reference pseudocharge density of the I^{th} nucleus and the summation index I runs over all atoms in \mathbb{R}^3 ; $V_c = \sum_I V_{c,I} = \sum_I (\tilde{V}_I - V_I)$, where \tilde{V}_I and V_I are the potentials generated by \tilde{b}_I and b_I , respectively; and \mathcal{I} is the identity operator. The first term is the band structure energy (E_{band}) and the last term is the energy associated with the electronic entropy (E_{ent}).

In order to update the positions of the ions during the Born-Oppenheimer quantum molecular dynamics (QMD) simulations, the Hellmann-Feynman force on the I^{th} nucleus can be written as [38]⁶

$$\begin{aligned} \mathbf{f}_I &= \sum_{I'} \int_{\Omega} \nabla b_{I'}(\mathbf{x}, \mathbf{R}_{I'}) \phi(\mathbf{x}, \mathbf{R}) \, d\mathbf{x} + \frac{1}{2} \sum_{I'} \int_{\Omega} \left[(\tilde{b}(\mathbf{x}, \mathbf{R}) + b(\mathbf{x}, \mathbf{R})) \nabla V_{c,I'} \right. \\ &\left. + (\nabla \tilde{b}_{I'}(\mathbf{x}, \mathbf{R}_{I'}) + \nabla b_{I'}(\mathbf{x}, \mathbf{R}_{I'})) V_c(\mathbf{x}, \mathbf{R}) \right] \, d\mathbf{x} - 4 \sum_{I'} \text{Tr}(\mathcal{V}_{nl}^{I'} \nabla \mathcal{D}) , \end{aligned} \quad (6)$$

⁴The electronic temperature/smearing is typically set to be equal to the ionic temperature in QMD simulations, particularly for those performed at high temperature.

⁵The repulsive energy correction for overlapping pseudocharges [45, 46] has been explicitly included in the free energy expression. This term plays a particularly important role in high-temperature simulations since the ions get significantly closer compared to ambient temperature.

⁶For the reasons mentioned in footnote 5, the force corresponding to the repulsive energy correction for overlapping pseudocharges [45, 46] has been explicitly included in the atomic force expression. Note that this term has been significantly simplified from its original form by utilizing the relation $\int \nabla f(|\mathbf{x} - \mathbf{R}_I|)g(|\mathbf{x} - \mathbf{R}_I|) \, d\mathbf{x} = 0$ for spherically symmetric functions f and g , which in the present case are b_I , $b_{I'}$, V_I , and $V_{I'}$.

where the summation index I' runs over the I^{th} atom and its periodic images, and $\mathcal{V}_{nl,I'}$ is the nonlocal pseudopotential associated with the I^{th} atom. The first two terms together constitute the local component of the force (\mathbf{f}_I^l) and the last term is the nonlocal component of the force (\mathbf{f}_I^{nl}).

Within a real-space representation, the density matrix \mathcal{D} has exponential decay for insulators as well as metallic systems at finite electronic temperature/smearing [8, 12]. This decay in the density matrix is exploited by linear-scaling methods through truncation within the calculations to enable $\mathcal{O}(N)$ computation of the electron density, energy, and atomic forces. In doing so, it has been observed that there is exponential convergence in the energy and forces with the size of the truncation region for insulating [47, 48] as well as metallic systems at finite temperature [49]. Note that in the above description for DFT, we have employed a local reformulation of the electrostatics to enable $\mathcal{O}(N)$ scaling for the complete Kohn-Sham problem.

3. Formulation and implementation of SQDFT

SQDFT is a large-scale parallel implementation of the Clenshaw-Curtis Spectral Quadrature (SQ) method⁷ [37, 38] for $\mathcal{O}(N)$ Kohn-Sham density functional calculations. In this approach, all quantities of interest (in discrete form) are expressed as bilinear forms or sums of bilinear forms, that are then approximated by spatially localized quadrature rules. The method is identically applicable to insulating and metallic systems and well suited to massively parallel computation. Furthermore, the SQ method becomes more efficient as temperature is increased, since electronic interactions become more localized and the representation of the Fermi-Dirac function becomes more compact [49].

We employ the *infinite-cell* version of the Clenshaw-Curtis SQ method, wherein the results corresponding to the infinite crystal are obtained without recourse to Brillouin zone integration or large supercells [38]. Specifically, rather than employ Bloch boundary conditions for the orbitals on Ω , zero-Dirichlet boundary conditions are prescribed at infinity, and the relevant components of the density matrix for spatial points within Ω are calculated by utilizing the potential within the truncation region surrounding that point [38, 49]. Periodic boundary conditions are retained for the electrostatic potential. Indeed, the infinite-cell SQ approach is equivalent to the standard Γ -point SQ calculation when the size of the truncation region is smaller than the size of the domain, a situation common in large-scale DFT simulations, particularly those at high temperature.

We utilize a high-order finite-difference discretization in order to exploit the locality of electronic interactions in real space, enable systematic convergence, and facilitate large-scale parallel implementation. We solve the fixed-point problem in Eq. 1 using the self consistent field (SCF) method [51]⁸, whose convergence is accelerated using the Periodic Pulay mixing scheme [52], a technique that significantly outperforms the well-established Anderson/Pulay mixing [53, 54]. We solve the Poisson problem in Eq. 3 using the Alternating Anderson-Richardson (AAR) method [55, 56], an approach that outperforms the conjugate gradient method [57] in the context of large-scale parallel computations [56]. We perform NVE (microcanonical) simulations using

⁷The Clenshaw-Curtis variant of SQ is chosen here because it is more efficient compared to its Gauss counterpart [50], particularly in the computation of the nonlocal component of the forces [38].

⁸In SQDFT, we perform a fixed-point iteration with respect to the effective potential: $V_{\text{eff}} = V_{xc} + \phi$.

the leapfrog method and NVT (canonical) simulations using the Verlet algorithm with the Nose-Hoover thermostat. At each MD step, we extrapolate the electron density using information from the previous two steps [58] to reduce SCF iterations. We parallelize the calculations using domain decomposition, with the communication between processors handled via the Message Passing Interface (MPI) [59].

In Fig. 1, we outline the key steps in a QMD simulation. We employ the Clenshaw-Curtis SQ method for calculating the electron density in each SCF iteration as well as for computing the energy and atomic forces (nonlocal component) once the electronic ground state has been determined. In the sections below, we discuss the calculation of the electron density, energy, and atomic forces in SQDFT. For additional details on the formulation and implementation of the pseudocharges, AAJ/AAR method, and Periodic Pulay mixing scheme, we refer the reader to the relevant previous works [60, 55, 52, 56].

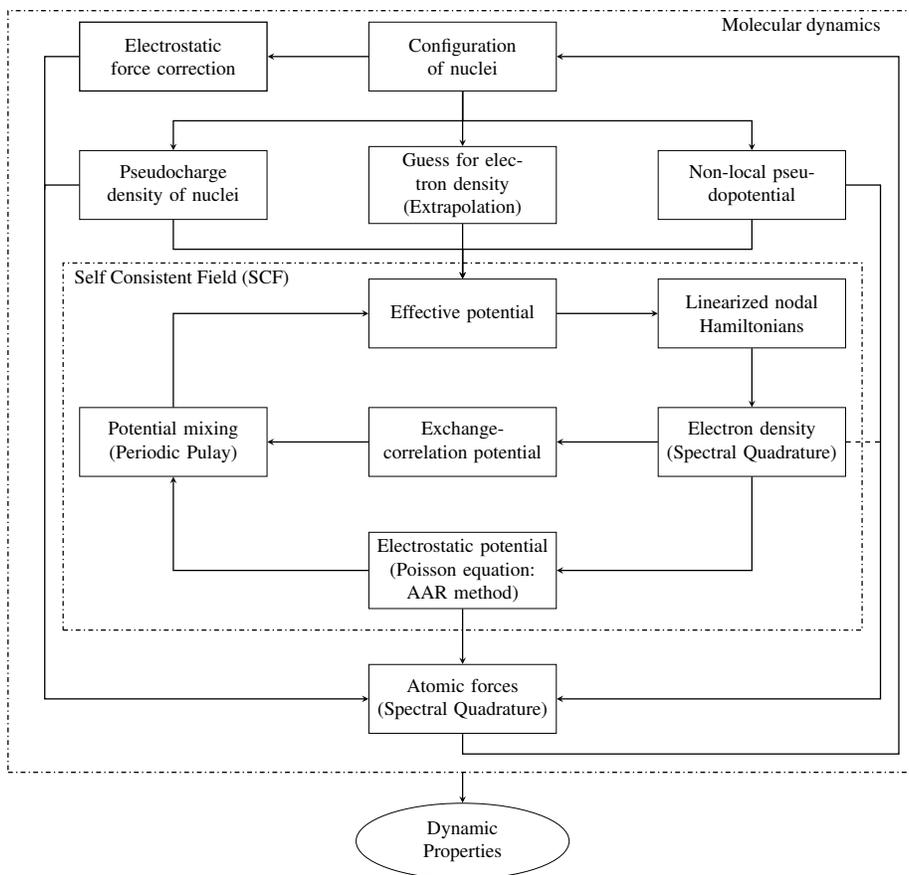


Figure 1: Outline of quantum molecular dynamics (QMD) simulation.

3.1. Finite-difference discretization

We consider a cubical domain Ω and discretize it with a uniform grid of spacing h , resulting in $N_d = n^3$ finite-difference nodes, the collection of which is referred to as K_Ω . We parallelize

the calculations by employing domain decomposition, i.e., we partition the domain into cubes of equal size such that $\Omega = \bigcup_{p=1}^{N_P} \Omega_p$, where N_P is the total number of processors and Ω_p denotes the domain local to the p^{th} processor, each of which contains $N_d^p = N_d/N_P$ finite-difference nodes. We refer to the collection of finite-difference nodes belonging to the p^{th} processor as K_Ω^p , where $K_\Omega = \bigcup_{p=1}^{N_P} K_\Omega^p$ with $K_\Omega^p \cap K_\Omega^q = \emptyset$ if $p \neq q$.

We approximate the Laplacian arising in the Hamiltonian and the local electrostatic reformulation using the central finite-difference approximation:

$$\begin{aligned} \nabla_h^2 f|^{(i,j,k)} &\approx \sum_{d=0}^{n_o} w_d \left(f^{(i+d,j,k)} + f^{(i-d,j,k)} + f^{(i,j+d,k)} + f^{(i,j-d,k)} + f^{(i,j,k+d)} + f^{(i,j,k-d)} \right), \\ w_0 &= -\frac{1}{h^2} \sum_{q=1}^{n_o} \frac{1}{q^2}, \\ w_d &= \frac{2(-1)^{d+1}}{h^2 d^2} \frac{(n_o!)^2}{(n_o-d)!(n_o+d)!}, \quad d = 1, 2, \dots, n_o, \end{aligned} \quad (7)$$

where $f^{(i,j,k)}$ denotes the value of the function f at the node indexed by (i, j, k) and $2n_o$ is the order of the approximation. Similarly, we approximate the gradient operator arising in the atomic forces using central finite-differences:

$$\begin{aligned} \nabla_h f|^{(i,j,k)} &\approx \sum_{d=1}^{n_o} \tilde{w}_d \left((f^{(i+d,j,k)} - f^{(i-d,j,k)}) \hat{\mathbf{e}}_1 + (f^{(i,j+d,k)} - f^{(i,j-d,k)}) \hat{\mathbf{e}}_2 + (f^{(i,j,k+d)} - f^{(i,j,k-d)}) \hat{\mathbf{e}}_3 \right), \\ \tilde{w}_d &= \frac{(-1)^{d+1}}{hd} \frac{(n_o!)^2}{(n_o-d)!(n_o+d)!}, \quad d = 1, 2, \dots, n_o, \end{aligned} \quad (8)$$

where $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, and $\hat{\mathbf{e}}_3$ are the unit vectors along the edges of Ω . We enforce periodic boundary conditions by mapping any index that does not correspond to a node in the finite-difference grid to its periodic image within Ω . We approximate the spatial integrals arising in the Hamiltonian (projectors of the nonlocal pseudopotential), energy, and atomic forces using the trapezoidal rule:

$$\int_{\Omega} f(\mathbf{x}) \, d\mathbf{x} \approx h^3 \sum_{i,j,k=1}^n f^{(i,j,k)}. \quad (9)$$

Even though this is a low order quadrature scheme, it has been chosen to ensure that the discrete free energy is consistent with the discrete Kohn-Sham equations, i.e., the calculated electronic ground state corresponds to the minimum of the free energy within the finite-difference discretization [61, 60].

In accordance with the nearsightedness principle [39], we define the *region of influence* for any finite-difference node as the cube of side $2R_{\text{cut}}$ centered at that node. A cube is chosen rather than a sphere due to its simplicity and efficiency within the finite-difference implementation. The

parameter R_{cut} corresponds to the distance beyond which electronic interactions are ignored, i.e., theoretically R_{cut} corresponds to the truncation radius for the density matrix. Within the above described finite-difference framework, we define the nodal Hamiltonian $\mathbf{H}_q \in \mathbb{R}^{N_s \times N_s}$ for any node $q \in K_\Omega$ as the restriction of the Hamiltonian to its region of influence [38], where $N_s = (2R_{\text{cut}}/h + 1)^3$ is the number of finite-difference nodes within the region of influence. Similarly, $\mathbf{w}_q \in \mathbb{R}^{N_s \times 1}$, $\nabla_{h,q} \in \mathbb{R}^{N_s \times N_s}$, and $\mathbf{V}_{nl,q}^I \in \mathbb{R}^{N_s \times N_s}$ represent the restriction of the standard basis vector, gradient matrix, and nonlocal pseudopotential matrix of the I^{th} atom to the region of influence, respectively. It is important to note that \mathbf{H}_q , $\nabla_{h,q}$, and $\mathbf{V}_{nl,q}^I$ are not explicitly determined/stored in SQDFT, rather their multiplication with a vector is directly computed in a matrix-free way.

3.2. Electron density

In each iteration of the SCF method, the electron density (Eq. 4) needs to be computed at all the finite-difference nodes. In SQDFT, the electron density at the $q \in K_\Omega^p$ node in the p^{th} processor is calculated using the relations [38]

$$\rho_q = \frac{2}{h^3} \sum_{j=0}^{n_{pl}} c_q^j \rho_q^j, \quad (10)$$

$$c_q^j = \frac{2}{\pi} \int_{-1}^1 \frac{g(r, \hat{\mu}_q, \hat{\sigma}_q) T_j(r)}{\sqrt{1-r^2}} dr, \quad j = 0, 1, \dots, n_{pl}, \quad (11)$$

$$\rho_q^j = \begin{cases} \mathbf{w}_q^T \mathbf{t}_q^j, & j = 0, 1, \dots, \frac{n_{pl}}{2} \\ \rho_q^1 - 2 \left(\mathbf{t}_q^{\frac{j+1}{2}} \right)^T \mathbf{t}_q^{\frac{j-1}{2}}, & j = \frac{n_{pl}}{2} + 1, \frac{n_{pl}}{2} + 3, \dots, n_{pl} - 1 \\ \rho_q^0 - 2 \left(\mathbf{t}_q^{\frac{j}{2}} \right)^T \mathbf{t}_q^{\frac{j}{2}}, & j = \frac{n_{pl}}{2} + 2, \frac{n_{pl}}{2} + 4, \dots, n_{pl} \end{cases} \quad (12)$$

where n_{pl} is the order of the Clenshaw-Curtis quadrature, chosen here to be a multiple of four for simplicity; T_j denotes the Chebyshev polynomial of degree j ; c_q^j is the coefficient of T_j in the polynomial expansion of the Fermi-Dirac function, with the value of c_q^0 half of that given in the expression; $\hat{\mu}_q = (\mu - \chi_q)/\zeta_q$ is the scaled and shifted Fermi energy, where $\chi_q = (\lambda_q^{\text{max}} + \lambda_q^{\text{min}})/2$ and $\zeta_q = (\lambda_q^{\text{max}} - \lambda_q^{\text{min}})/2$, with λ_q^{max} and λ_q^{min} denoting the maximum and minimum eigenvalues of \mathbf{H}_q , respectively; $\hat{\sigma}_q = \sigma/\zeta_q$ is the scaled smearing; and $\mathbf{t}_q^j = T_j(\hat{\mathbf{H}}_q) \mathbf{w}_q \in \mathbb{R}^{N_s \times 1}$, which is determined using the following iteration, obtained as a consequence of the three term recurrence relation of Chebyshev polynomials:

$$\begin{aligned} \mathbf{t}_q^{i+1} &= 2\hat{\mathbf{H}}_q \mathbf{t}_q^i - \mathbf{t}_q^{i-1}, \quad i = 1, 2, \dots, \frac{n_{pl}}{2} \\ \mathbf{t}_q^1 &= \hat{\mathbf{H}}_q \mathbf{w}_q, \quad \mathbf{t}_q^0 = \mathbf{w}_q, \end{aligned} \quad (13)$$

where $\hat{\mathbf{H}}_q = (\mathbf{H}_q - \chi_q \mathbf{I})/\zeta_q$ is the scaled and shifted nodal Hamiltonian whose spectrum lies in the interval $[-1, 1]$. Note that the iteration in Eq. 13 proceeds only up to $n_{pl}/2$ rather than n_{pl} , since we have employed the product property of Chebyshev polynomials⁹ to directly calculate ρ_q^j for

⁹ $2T_j(r)T_k(r) = T_{j+k}(r) + T_{|j-k|}(r)$. Note that in previous work where the expression for the electron density has been derived [38], this property has not been utilized, and therefore the iteration in Eq. 13 proceeds up to n_{pl} . The

$j = n_{pl}/2 + 1, n_{pl}/2 + 2, \dots, n_{pl}$, as described by Eq. 12. Since the values of ρ_q^j are independent of the Fermi level μ , they are first computed and stored. Next, μ is determined by satisfying the constraint on the total number of electrons:

$$2 \sum_{p=1}^{N_P} \sum_{q \in K_p} \sum_{j=0}^{n_{pl}} c_q^j \rho_q^j = N_e, \quad (14)$$

where the values of c_q^j are given by Eq. 11. Finally, for the c_q^j corresponding to the Fermi level μ , the electron density is calculated via Eq. 10.

Since the simulation domain Ω is cubical and we have employed a uniform finite-difference grid with uniform domain decomposition, the layout of effective potential values V_{eff} required from neighboring processors as part of the nodal Hamiltonians for grid points in each processor is identical. To accomplish this, we utilize the MPI command `MPI_Ineighbor_alltoallv` [62] to communicate the required values of V_{eff} (i.e., those within the region of influence for grid points in each processor) between processors. Doing so reduces the number of MPI related calls that would otherwise be required in every matrix-vector multiplication. After the communication is complete, for every finite-difference node $q \in K_\Omega^p$, we first calculate λ_q^{\max} and λ_q^{\min} —maximum and minimum eigenvalues of \mathbf{H}_q , respectively—using the Lanczos method [63, 64]. Next, we perform the recursive iteration in Eq. 13, the results of which are used to calculate ρ_q^j as given in Eq. 12. The matrix-vector multiplications required as part of the Lanczos method and the recursive iteration in Eq. 13 are performed in a matrix-free manner. The Chebyshev coefficients c_q^j are calculated using the discrete orthogonality of the Chebyshev polynomials [65]. The Fermi level is determined using Brent’s method [66], with Newton-Raphson’s method becoming the preferred choice as the temperature is increased. While doing so, the global communication between processors is handled by using the `MPI_Allreduce` command.

During the electron density calculation, the memory costs are dominated by the storage of three vectors during the recursive iteration in Eq. 13 (t_q^{i+1} , t_q^i , and t_q^{i-1}) and the storage of ρ_q^j for the calculation of the electron density in Eq. 10. Therefore, the memory costs per processor scale as $\mathcal{O}(3N_s + n_{pl}N_d^p)$. The computational costs are dominated by the the matrix-vector products in the Lanczos iteration and the recursive iteration in Eq. 13. Therefore, the computational cost per processor scales as $\mathcal{O}(n_{\text{lancz}}N_sN_d^p + \frac{1}{2}n_{pl}N_sN_d^p)$, where n_{lancz} is the number of Lanczos iterations required for determining λ_q^{\max} and λ_q^{\min} . Since N_s , n_{pl} , and n_{lancz} remain independent of system size, the overall memory and computational costs scale linearly with the number of finite-difference nodes in the domain Ω , and therefore $\mathcal{O}(N)$ with respect to the number of atoms.

3.3. Free energy

Once the electronic ground state has been determined, i.e., the SCF iteration has converged, the free energy (Eq. 5) needs to be calculated.

present approach reduces the cost for the computation of ρ_q^j by a factor of two.

Band structure energy. The band structure energy in the Clenshaw-Curtis SQ method [38] takes the following form in parallel computations

$$E_{band} = 2 \sum_{p=1}^{N_P} \sum_{q \in K_p} \sum_{j=0}^{n_{pl}} (\chi_q c_q^j + \zeta_q d_q^j) \rho_q^j, \quad (15)$$

$$d_q^j = \frac{2}{\pi} \int_{-1}^1 \frac{r g(r, \hat{\mu}_q, \hat{\sigma}_q) T_j(r)}{\sqrt{1-r^2}} dr, \quad (16)$$

where d_q^j is the coefficient of T_j in the polynomial expansion of the band structure energy function (i.e., $rg(r)$), with the value of d_q^0 half of that given in the expression. In addition, c_q^j and ρ_q^j are as given in Eqs. 11 and 12, respectively.

Electronic entropy energy. The electronic entropy energy in the Clenshaw-Curtis SQ approach [38] takes the following form in parallel computations

$$S = 2\sigma \sum_{p=1}^{N_P} \sum_{q \in K_p} \sum_{j=0}^{n_{pl}} e_q^j \rho_q^j, \quad (17)$$

$$e_q^j = \frac{2}{\pi} \int_{-1}^1 \frac{g(r, \hat{\mu}_q, \hat{\sigma}_q) \log g(r, \hat{\mu}_q, \hat{\sigma}_q) + (1 - g(r, \hat{\mu}_q, \hat{\sigma}_q)) \log(1 - g(r, \hat{\mu}_q, \hat{\sigma}_q)) T_j(r)}{\sqrt{1-r^2}} dr, \quad (18)$$

where e_q^j is the coefficient of T_j in the polynomial expansion of the electronic entropy energy function (i.e., $g(r) \log g(r) + (1 - g(r)) \log(1 - g(r))$), with the value of e_q^0 half of that given in the expression. Again, ρ_q^j is as given in Eq. 12.

Free energy. The free energy of the system in SQDFT is computed as

$$\begin{aligned} \mathcal{F}(\mathbf{R}) = & h^3 \sum_{p=1}^{N_P} \sum_{q \in K_p} \left(\frac{2}{h^3} \sum_{j=0}^{n_{pl}} (\chi_q c_q^j + \zeta_q d_q^j) \rho_q^j + \epsilon_{xc}(\rho_q) \rho_q - V_{xc}(\rho_p) \rho_q + \frac{1}{2} (b_q - \rho_q) \phi_q \right. \\ & \left. + \frac{1}{2} (\tilde{b}_q + b_q) V_{c,q} - \frac{1}{2} \sum_{I \in D_p^b} \tilde{b}_{I,q} \tilde{V}_{I,q} + \frac{2\sigma}{h^3} \sum_{j=0}^{n_{pl}} e_q^j \rho_q^j \right), \quad (19) \end{aligned}$$

where the spatial integrals in Eq. 5 have been approximated using the trapezoidal rule in Eq. 9 and D_p^b is the set of all atoms (considering all atoms in \mathbb{R}^3) whose pseudocharges have overlap with the processor domain Ω_p . We note that the exchange correlation energy E_{xc} has been modeled using the Local Density Approximation (LDA) [2], wherein $\epsilon_{xc}(\rho)$ is the sum of the exchange and correlation energy per particle of a uniform electron gas of density ρ .

During the free energy calculation, the values of c_q^j and ρ_q^j determined as part of the electron density computation in the last SCF iteration are directly utilized. The Chebyshev coefficients d_q^j and e_q^j are determined using the discrete orthogonality of the Chebyshev polynomials [65]. One `MPI_Allreduce` command is utilized to simultaneously compute all the components of the energy. The computational cost per processor scales as $\mathcal{O}(n_{pl} N_a^p)$, which translates to an overall scaling of $\mathcal{O}(N)$ with respect to the number of atoms.

3.4. Atomic forces

In order to update the positions of the atoms during the course of the QMD simulation, the Hellmann-Feynman forces on the nuclei (Eq. 6) need to be computed.

Local component. The local component of the force has the following discrete form in parallel computations

$$\mathbf{f}_I^l = h^3 \sum_{p=1}^{N_P} \sum_{I' \in D_{p,I}^b} \sum_{q \in K_p} \left(\nabla_h b_{I'}|_q \phi_q + \frac{1}{2} (\tilde{b}_q + b_q) \nabla_h V_{c,I'}|_q + \frac{1}{2} (\nabla_h \tilde{b}_{I'}|_q + \nabla_h b_{I'}|_q) V_{c,q} \right) \quad (20)$$

where the integrals in Eq. 6 have been approximated using the trapezoidal rule (Eqn. 9) and $D_{p,I}^b$ is the set of the I^{th} atom and its images whose pseudocharges have overlap with the processor domain Ω_p .

Nonlocal component. The nonlocal component of the force in the SQ approach [38] takes the following form in parallel computations

$$\mathbf{f}_I^{nl} = -4 \sum_{p=1}^{N_P} \sum_{I' \in D_{p,I}^c} \sum_{q \in K_p} \mathbf{w}_q^T \mathbf{V}_{nl,q}^{I'} \nabla_{h,q} \left(\sum_{j=0}^{n_{pl}} c_q^j \mathbf{t}_q^j \right) \quad (21)$$

where $D_{p,I}^c$ is the set of the I^{th} atom and its images whose nonlocal projectors have overlap with the processor domain Ω_p , c_q^j is the coefficient of T_j in the polynomial expansion of the Fermi-Dirac function (Eq. 11); and \mathbf{t}_q^j is determined using the recurrence relation:

$$\begin{aligned} \mathbf{t}_q^{i+1} &= 2\hat{\mathbf{H}}_q \mathbf{t}_q^i - \mathbf{t}_q^{i-1}, \quad i = 1, 2, \dots, n_{pl} \\ \mathbf{t}_q^1 &= \hat{\mathbf{H}}_q \mathbf{w}_q, \quad \mathbf{t}_q^0 = \mathbf{w}_q, \end{aligned} \quad (22)$$

Note that unlike the iteration in Eq. 13 which proceeds up to $n_{pl}/2$, the above iteration proceeds up to n_{pl} since some of the off-diagonal components of the density matrix are needed for the calculation of the nonlocal force in Eq. 21, i.e., ρ_q^j are not sufficient, rather \mathbf{t}_q^j are required.

Total atomic force. The atomic force in SQDFT is then calculated as

$$\begin{aligned} \mathbf{f}_I &= h^3 \sum_{p=1}^{N_P} \sum_{I' \in D_{p,I}^b} \sum_{q \in K_p} \left(\nabla_h b_{I'}|_q \phi_q + \frac{1}{2} (\tilde{b}_q + b_q) \nabla_h V_{c,I'}|_q + \frac{1}{2} (\nabla_h \tilde{b}_{I'}|_q + \nabla_h b_{I'}|_q) V_{c,q} \right) \\ &- 4 \sum_{p=1}^{N_P} \sum_{I' \in D_{p,I}^c} \sum_{q \in K_p} \mathbf{w}_q^T \mathbf{V}_{nl,q}^{I'} \nabla_{h,q} \left(\sum_{j=0}^{n_{pl}} c_q^j \mathbf{t}_q^j \right). \end{aligned} \quad (23)$$

During the computation of the atomic forces, the values of c_q^j that were determined as part of the electron density calculation during the last SCF iteration are directly utilized. As mentioned previously, the nodal gradient matrix $\nabla_{h,q}$ is not generated/stored explicitly, rather its product

with \mathbf{t}_q^j is calculated in matrix-free fashion. The total atomic force on all atoms is simultaneously computed using a single `MPI_Allreduce`. Since the storage per processor scales as $\mathcal{O}(3N_s)$ and the computational effort per processor scales as $\mathcal{O}(n_{pl}N_sN_d^p) + \mathcal{O}(3N_s)^{10}$, the overall storage and computational cost scales as $\mathcal{O}(N)$ with respect to the number of atoms.

Relation to classical Fermi Operator Expansion (FOE). The Clenshaw-Curtis SQ method bears some resemblance to the classical Fermi Operator Expansion (FOE) [67, 68] in that both techniques use Chebyshev polynomials as the underlying basis for expanding the Fermi-Dirac function of a matrix. However, the key underlying difference is that the matrix in FOE corresponds to the Hamiltonian, whereas in SQ it corresponds to the nodal Hamiltonian. Therefore, truncation is automatically included within the SQ method and the key operation is reduced to local sparse matrix-vector products, as opposed to the global sparse matrix-matrix products in the FOE method. This makes SQ more efficient since (i) the Fermi level calculation does not require an outer loop, and (ii) inter-processor communication is needed just once per SCF iteration, unlike the FOE where it is required for every matrix-matrix multiplication. Finally, SQ also requires significantly less storage compared to FOE, making it especially well suited for modern high performance computing platforms.

4. Results and discussion

In this section, we demonstrate the accuracy, efficiency, and scaling of SQDFT in Kohn-Sham Density Functional Theory (DFT) calculations at high temperature. In all simulations, we employ a twelfth-order finite-difference discretization, norm-conserving Troullier-Martins pseudopotentials [69], and the Local Density Approximation (LDA) [2] with the Perdew-Wang parametrization [70] of the correlation energy calculated by Ceperley-Alder [71]. To ensure accuracy of the standard 3s3p pseudopotentials employed, temperatures were limited to $T \lesssim 80000$ K, where 2p states can be treated as fully occupied.¹¹ To demonstrate applicability to metals and insulators alike, we consider two systems: (i) aluminum, a prototypical metal, and (ii) lithium hydride, a prototypical insulator. We compare the results obtained by SQDFT to benchmarks obtained by the finite-difference code SPARC [61, 60] and planewave code ABINIT [72], both of which solve the Kohn-Sham problem via diagonalization.

4.1. Accuracy and convergence

We first study the accuracy of SQDFT, i.e., we verify the convergence of computed energies and atomic forces with respect to key SQ parameters (i.e., quadrature order n_{pl} and truncation radius R_{cut}) as well as spatial discretization (i.e., mesh-size h). As representative systems, we consider a 32-atom cell of aluminum at the equilibrium lattice constant of 7.78 Bohr and a 64-atom cell of lithium hydride at the equilibrium lattice constant of 7.37 Bohr, with all atoms randomly

¹⁰This corresponds to the calculation of the nonlocal component of the force, which is the dominant cost in the atomic force calculation in SQDFT. Note that $\mathcal{O}(3N_s)$ arises due to the three matrix-vector products arising in the multiplication with the finite-difference gradient.

¹¹Calculations with deeper 2s2p3s3p pseudopotentials show 2p occupation of 5.9991 at 80000 K.

displaced by up to 15% of the equilibrium interatomic distance. We select the electronic temperature/smearing to be $\sigma = 4$ eV.

First, we verify the convergence of SQDFT energies and forces with respect to n_{pl} and R_{cut} in Fig. 2, with the reference diagonalization answers obtained by SPARC at the same mesh-size and a $4 \times 4 \times 4$ Monkhorst-Pack grid for Brillouin zone integration.¹² We choose mesh-sizes of $h = 0.7780$ and $h = 0.5264$ Bohr for the aluminum and lithium hydride systems, respectively. It is clear that SQDFT obtains exponential convergence in the energy and atomic forces with respect to both parameters, in agreement with previous studies [37, 38, 49]. In particular, $\{n_{pl}, R_{cut}\} \sim \{28, 6\}$ and $\{n_{pl}, R_{cut}\} \sim \{40, 6\}$ are sufficient to obtain chemical accuracy in both the energy and forces for the aluminum and lithium hydride systems, respectively. Importantly, these values further reduce as the smearing/temperature is increased [38], which makes SQDFT particularly attractive for high-temperature simulations. Note that neither the energy nor the atomic forces are variational with respect to n_{pl} and R_{cut} , hence the non-monotonic convergence in Fig. 2.

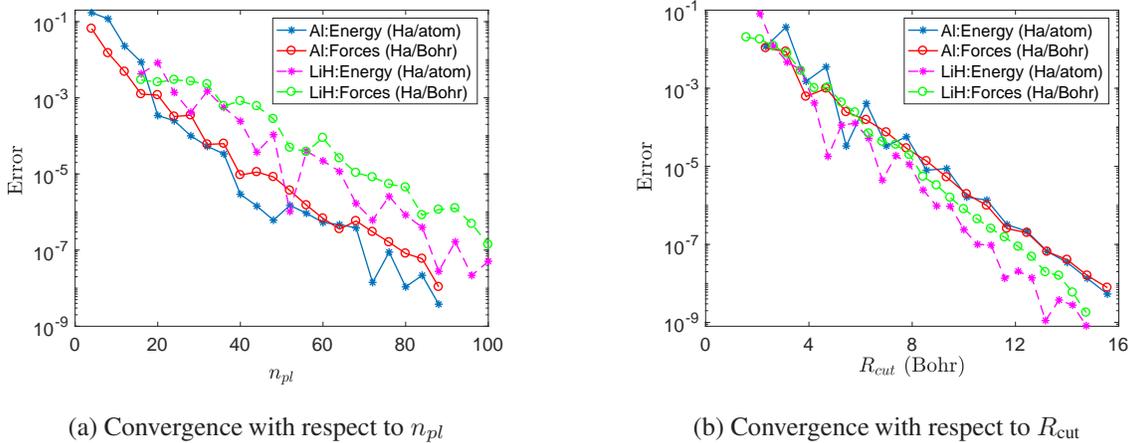


Figure 2: Convergence of energy and forces in SQDFT with respect to quadrature order n_{pl} and truncation radius R_{cut} for aluminum and lithium hydride systems. The error in energy is the magnitude of the difference and the error in forces is the maximum difference in any component, with diagonalization result obtained by finite-difference code SPARC at the same mesh-size as reference.

Next, we verify the convergence of SQDFT energies and forces with mesh-size h to those computed by the established planewave code ABINIT. To do so, we utilize $n_{pl} = 160$ and $R_{cut} = 10$ Bohr in SQDFT, which are sufficient to put the associated errors well below the mesh errors of interest (see Fig. 2). In ABINIT, we employ a planewave cutoff of 50 Ha and a $4 \times 4 \times 4$ Monkhorst-Pack grid for Brillouin zone integration, which results in energy and forces that are converged to within 10^{-6} Ha/atom and 10^{-6} Ha/Bohr, respectively. As shown in Fig. 3, both the energy and atomic forces in SQDFT converge rapidly and systematically, with chemical accuracy readily obtained. Notably, we see that energies and forces converge at comparable rates, without the need for additional measures such as double-grid [73] or high-order integration [74]. Therefore,

¹²Unlike standard codes, SQDFT can obtain the infinite crystal result without recourse to Brillouin zone integration.

accurate forces are easily obtained, as needed for structural relaxations and molecular dynamics simulations.

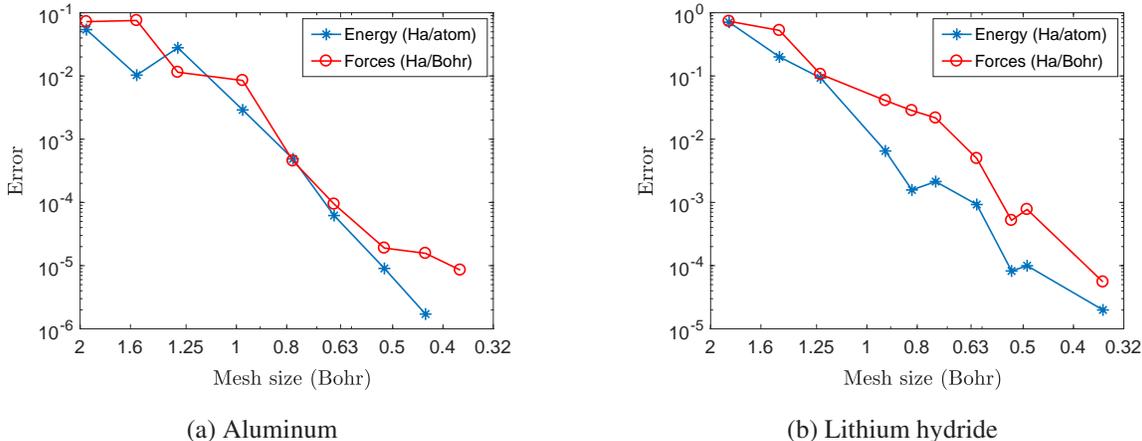


Figure 3: Convergence of energy and forces in SQDFT with respect to mesh-size h for aluminum and lithium hydride systems. The error in energy is the magnitude of the difference and error in forces is the maximum difference in any component, with results obtained by planewave code ABINIT as reference.

4.2. Scaling and performance

We now study the scaling and performance of SQDFT on large-scale parallel computers with up to tens of thousands of processors. Specifically, we investigate the strong and weak scaling of SQDFT for aluminum and lithium hydride systems on the *Sierra* and *Quartz* supercomputers at the Lawrence Livermore National Laboratory (LLNL) [75]. In all calculations, we employ a smearing of $\sigma = 4$ eV and utilize (i) $h = 0.7780$ Bohr and $\{n_{pl}, R_{cut}\} = \{28, 6.224 \text{ Bohr}\}$ for the aluminum systems, and (ii) $h = 0.5264$ Bohr and $\{n_{pl}, R_{cut}\} = \{40, 6.387 \text{ Bohr}\}$ for the lithium hydride systems. These parameters are sufficient to obtain chemical accuracy of 0.001 Ha/atom and 0.001 Ha/Bohr in the energy and atomic forces, respectively, as demonstrated in the previous section.

First, we perform a strong scaling study for a 2048-atom aluminum system on *Sierra* and a 1728-atom lithium hydride system on *Quartz*, with atoms randomly displaced in both systems. For aluminum, the number of processors on *Sierra* is varied from 64 to 8000. For lithium hydride, the number of processors on *Quartz* is varied from 125 to 27000. The wall times per SCF iteration so obtained are presented in Fig. 4a. Relative to the smallest number of processors, on the largest number of processors SQDFT achieves 97% parallel efficiency for aluminum on *Sierra* and 95% for lithium hydride on *Quartz*. It is clear that SQDFT demonstrates excellent strong scaling. Notably, the wall time per SCF iteration for systems containing ~ 2000 atoms can be reduced to less than 5 seconds.

Next, we perform a weak scaling study for aluminum and lithium hydride. For aluminum on *Sierra*, we increase the system size from 32 to 6912 atoms, while increasing the number of processors from 64 to 13824, maintaining two processors per atom for all systems. For lithium hydride on *Quartz*, we increase the system size from 8 to 10648 atoms, while increasing the

number of processors from 27 to 35937, maintaining ~ 3.4 processors per atom for all systems. The systems are generated by replicating 4-atom and 8-atom unit cells of aluminum and lithium hydride, respectively, with one atom in each unit cell randomly perturbed. We present the results so obtained in Fig. 4b. We find the scaling with system size for aluminum on Sierra to be $\mathcal{O}(N^{1.00})$ and the scaling for lithium hydride on Quartz to be $\mathcal{O}(N^{1.01})$. It is clear that SQDFT demonstrates excellent weak scaling, with near perfect $\mathcal{O}(N)$ scaling with respect to system size in practical calculations.

Overall, the excellent strong and weak scaling of SQDFT up to tens of thousands of processors makes it possible to perform high-temperature Kohn-Sham molecular dynamics simulations at large length and time scales, as we show below.

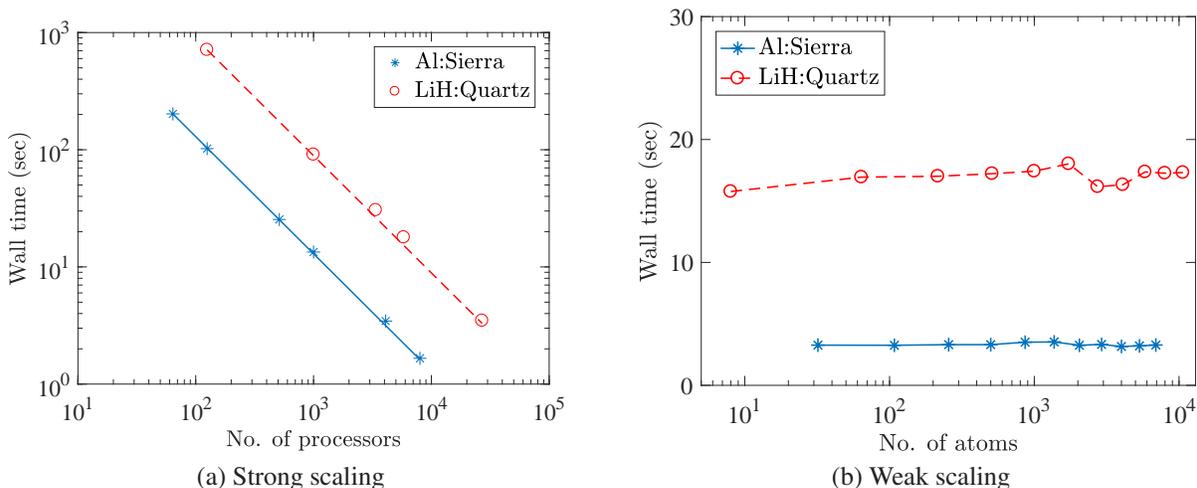


Figure 4: Parallel scaling of a single SCF iteration in SQDFT on Sierra and Quartz supercomputers at the Lawrence Livermore National Laboratory (LLNL) [75]. In the strong scaling plot, the straight lines represent ideal scaling. All parameters have been chosen so as to achieve chemical accuracy of 0.001 Ha/atom and 0.001 Ha/Bohr in energy and atomic forces, respectively.

4.3. High-temperature quantum molecular dynamics

We now consider the accuracy and efficiency of SQDFT in high-temperature Born-Oppenheimer quantum molecular dynamics (QMD) simulations. Such simulations are a cornerstone of modern warm dense matter theory [27], providing equation-of-state and shock-compression predictions of unprecedented accuracy, up to temperatures of ~ 100 eV and pressures of 100s of Mbar; see, e.g., [36, 76, 77, 78, 79, 80]. At temperatures above ~ 100 eV, conventional Kohn-Sham methods become prohibitively expensive, and so alternative methods such as OFMD [33] and, more recently, path integral Monte Carlo (PIMC) [80] have been employed to reach temperatures of 1000s of eV and higher. With sufficiently deep potentials, however, the new SQ methodology makes possible Kohn-Sham MD at temperatures of 1000s of eV as well.

As a representative example, we choose an 864-atom aluminum system and perform a 0.15

ps NVE QMD simulation with time step of 0.1 fs.¹³ We use a mesh-size $h = 0.7780$ Bohr, quadrature order $n_{pl} = 28$, truncation radius $R_{cut} = 6.224$ Bohr, initial ionic temperature $T = 116045$ K, initial atomic positions close to perfect FCC crystal with one atom in each 4-atom unit cell randomly displaced by the same amount, and initial velocities randomly assigned based on the Maxwell-Boltzmann distribution. We integrate the equations of motion using the Leapfrog method [81]. The values of n_{pl} and R_{cut} have been chosen so as to put the associated errors close to an order of magnitude lower than the discretization error, which is ~ 0.001 Ha/atom and ~ 0.001 Ha/Bohr in the energy and forces, respectively. At each MD step, we set the electronic temperature equal to the ionic temperature, e.g, $\sigma = 10$ eV at the start of the simulation.

We perform the simulation on 3375 processors on Quartz to obtain a wall clock time of ~ 30 seconds per QMD step¹⁴. In Fig. 5, we plot the variation of the total energy and temperature of the system over the course of the simulation. We observe that the temperature settles after ~ 30 fs, subsequent to which the mean and standard deviation of the total energy are -3.3451 and 4.7×10^{-4} Ha/atom, respectively. In addition, the drift in total energy as obtained from a linear fit is $\sim 2.7 \times 10^{-4}$ Ha/atom-ps. SQDFT thus shows excellent energy conservation, consistent with the accurate atomic forces obtained. We also plot the calculated radial distribution function in Fig. 6, which is in agreement with previous studies [32].

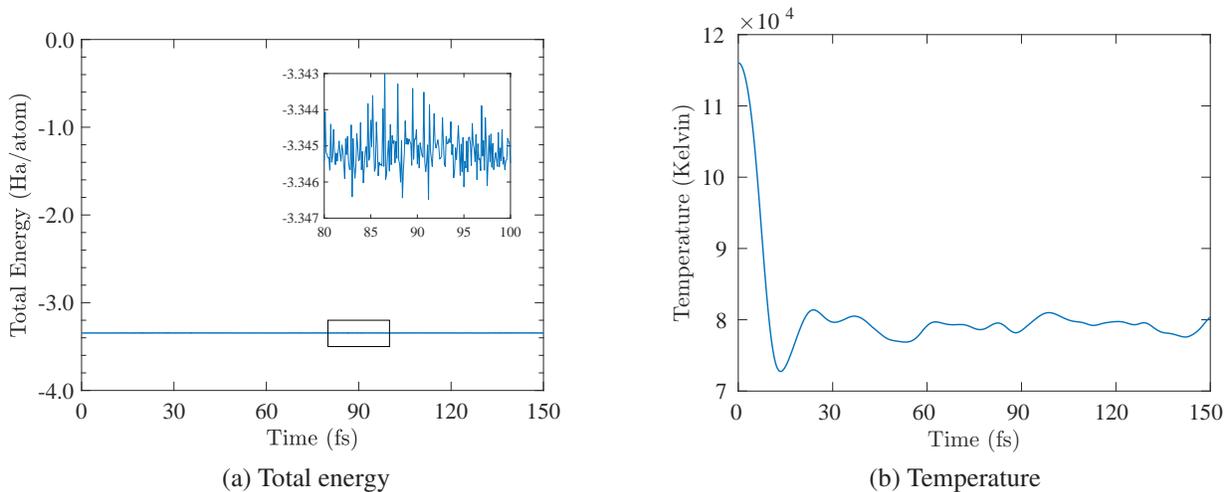


Figure 5: Variation of total energy and temperature during 864-atom aluminum NVE QMD simulation.

5. Concluding remarks

We presented SQDFT: a large-scale parallel implementation of the Spectral Quadrature (SQ) method for $\mathcal{O}(N)$ Kohn-Sham density functional theory calculations at high temperature. Specif-

¹³Due to the high velocities of the ions in high-temperature simulations, the time step has been chosen significantly smaller than in ambient calculations [34].

¹⁴Since the computational cost of SQDFT reduces with temperature, the wall time will further reduce as the temperature is increased.

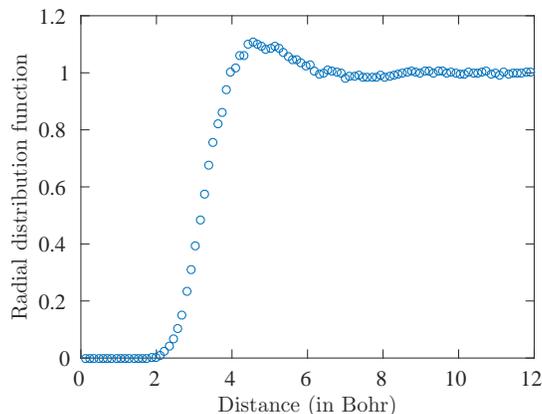


Figure 6: Radial distribution function for 864-atom aluminum NVE QMD simulation.

ically, we developed an efficient and scalable finite-difference implementation of the infinite-cell Clenshaw-Curtis SQ approach, in which results for the infinite crystal are obtained by expressing quantities of interest as bilinear forms or sums of bilinear forms, that are then approximated by spatially localized Clenshaw-Curtis quadrature rules. We demonstrated the accuracy of SQDFT by showing systematic convergence of energies and atomic forces with respect to quadrature order and truncation radius to reference diagonalization results, and convergence with mesh spacing to established planewave results, for both metallic and insulating systems. In all cases, chemical accuracy was readily obtained. We demonstrated excellent strong and weak parallel scaling on computer systems consisting of tens of thousands of processors, with near perfect $\mathcal{O}(N)$ scaling with system size, and wall clock times as low as a few seconds per SCF iteration for insulating and metallic systems of ~ 2000 atoms. Finally, we verified the accuracy and efficiency of SQDFT in large-scale quantum molecular dynamics (QMD) simulations at high temperature, demonstrating excellent energy conservation and QMD step times of ~ 30 seconds for an 864-atom aluminum system at ~ 80000 K.

In the present work, we have focused on high-temperature Kohn-Sham DFT calculations. However, the SQ method is applicable at lower temperatures as well, with larger prefactor, as we show in Appendix A. A possible approach to reduce this prefactor is to generate a localized orthonormal reduced basis (e.g., [82, 83]), subsequent to which the SQ method is applied to the finite-difference Hamiltonian projected into this basis. This is indeed a promising path to $\mathcal{O}(N)$ DFT calculations of metals and insulators at ambient conditions which the authors are pursuing presently.

Acknowledgements

This work was supported in part by the National Science Foundation (Grant number 1333500), and performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Early support from the Exascale Co-design Center for Materials in Extreme Environments supported by Office of Science Advanced Scientific Computing Research Program, and subsequent support from the Laboratory Directed Research and Development program at the Lawrence Livermore National Laboratory is gratefully

acknowledged.

Appendix

A. Ambient temperature Kohn-Sham calculations

Though the focus of the present work has been Kohn-Sham calculations at high temperature, SQDFT can also be utilized at ambient temperature, albeit with a larger prefactor. This increase in prefactor may, however, be mitigated by the excellent parallel scaling of SQDFT on large-scale parallel computers. In order to demonstrate this, we consider an 864-atom randomly perturbed aluminum system with smearing $\sigma = 0.27$ eV, as typical in calculations of metallic systems at ambient conditions in order to facilitate self-consistent convergence [84, 72]. We utilize $h = 0.7780$ Bohr and $\{n_{pl}, R_{cut}\} = \{320, 18.672 \text{ Bohr}\}$, which are sufficient to obtain chemical accuracy of 0.001 Ha/atom and 0.001 Ha/Bohr in the energy and atomic forces, respectively. We perform the calculations on Quartz, where the number of processors is varied from 1000 to 27000, the results of which are presented in Fig. 7. Relative to 1000 processors, the efficiency of SQDFT on 8000 processors is larger than 98%, but on 27000 processors, the efficiency drops to 51%. The reduced efficiency at this temperature at the largest processor counts arises due to the increased communications required for the larger nodal Hamiltonians ($R_{cut} = 18.672$ Bohr) relative to the computational work per processor. However, SQDFT is still able to achieve wall times of less than a minute per SCF iteration, which demonstrates its ability to perform large-scale Kohn-Sham quantum molecular dynamics simulations (QMD) even at ambient temperature, given sufficient number of processors.

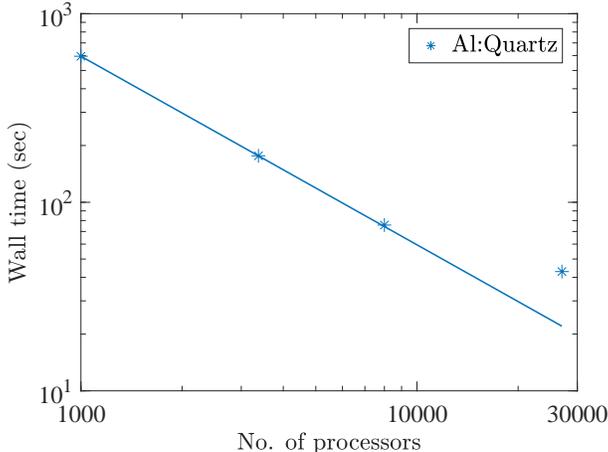


Figure 7: Strong scaling of single SCF iteration in SQDFT for 864-atom aluminum system at $\sigma = 0.27$ eV on Quartz supercomputer at the Lawrence Livermore National Laboratory (LLNL) [75]. The straight line represents ideal scaling. All parameters have been chosen so as to achieve chemical accuracy of 0.001 Ha/atom and 0.001 Ha/Bohr in energy and atomic forces, respectively.

References

- [1] P. Hohenberg, W. Kohn, *Physical Review* 136 (1964) B864–B871.
- [2] W. Kohn, L. J. Sham, *Physical Review* 140 (1965) A1133–A1138.
- [3] S. Goedecker, *Rev. Mod. Phys.* 71 (1999) 1085–1123.
- [4] D. R. Bowler, T. Miyazaki, *Reports on Progress in Physics* 75 (2012) 036503.
- [5] J. Aarons, M. Sarwar, D. Thompsett, C.-K. Skylaris, *The Journal of Chemical Physics* 145 (2016) 220901.
- [6] D. Marx, J. Hutter, *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, 2009.
- [7] G. Kresse, J. Hafner, *Physical Review B* 47 (1993) 558.
- [8] S. Goedecker, *Physical Review B* 58 (1998) 3501.
- [9] S. Ismail-Beigi, T. Arias, *Physical review letters* 82 (1999) 2127.
- [10] X. Zhang, D. Drabold, *Physical Review B* 63 (2001) 233109.
- [11] S. Taraskin, P. Fry, X. Zhang, D. Drabold, S. Elliott, *Physical Review B* 66 (2002) 233101.
- [12] M. Benzi, P. Boito, N. Razouk, *SIAM Review* 55 (2013) 3–64.
- [13] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, D. Sanchez-Portal, *J. Phys.: Condes. Matter* 14 (2002) 2745–2779.
- [14] SIESTA: www.icmab.es/siesta, accessed 2017-07-05.
- [15] M. J. Gillan, D. R. Bowler, A. S. Torralba, T. Miyazaki, *Comput. Phys. Commun.* 177 (2007) 14–18.
- [16] Conquest: www.order-n.org, accessed 2017-07-05.
- [17] C. K. Skylaris, P. D. Haynes, A. A. Mostofi, M. C. Payne, *J. Chem. Phys.* 122 (2005).
- [18] ONETEP: www.onetep.org, accessed 2017-07-05.
- [19] E. Tsuchida, *J. Phys. Soc. Jpn.* 76 (2007).
- [20] D. Osei-Kuffuor, J.-L. Fattebert, *Phys. Rev. Lett.* 112 (2014).
- [21] S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch, S. Goedecker, *J. Chem. Phys.* 140 (2014).
- [22] BigDFT: bigdft.org, accessed 2017-07-05.

- [23] OpenMX: www.openmx-square.org, accessed 2017-07-05.
- [24] N. Bock, M. Challacombe, C. K. Gan, G. Henkelman, K. Nemeth, A. M. N. Niklasson, A. Odell, E. Schwegler, C. J. Tymczak, V. Weber, FREEON, 2014. Los Alamos National Laboratory (LA-CC 01-2; LA-CC-04-086), Copyright University of California.
- [25] FreeON: www.openhub.net/p/freeon, accessed 2017-07-05.
- [26] A. Ruiz-Serrano, N. D. M. Hine, C.-K. Skylaris, *J. Chem. Phys.* 136 (2012).
- [27] F. Graziani, M. P. Desjarlais, R. Redmer, S. B. Trickey (Eds.), *Frontiers and Challenges in Warm Dense Matter, Lecture Notes in Computational Science and Engineering*, Springer, 2014.
- [28] F. R. Graziani, V. S. Batista, L. X. Benedict, J. I. Castor, H. Chen, S. N. Chen, C. A. Fichtl, J. N. Glosli, P. E. Grabowski, A. T. Graf, S. P. Hau-Riege, A. U. Hazi, S. A. Khairallah, L. Krauss, A. B. Langdon, R. A. London, A. Markmann, M. S. Murillo, D. F. Richards, H. A. Scott, R. Shepherd, L. G. Stanton, F. H. Streitz, M. P. Surh, J. C. Weisheit, H. D. Whitley, *High Energy Density Physics* 8 (2012) 105–131.
- [29] P. Renaudin, C. Blancard, J. Clérouin, G. Faussurier, P. Noiret, V. Recoules, *Physical review letters* 91 (2003) 075002.
- [30] M. Dharma-Wardana, *Physical Review E* 73 (2006) 036401.
- [31] R. Ernstorfer, M. Harb, C. T. Hebeisen, G. Sciaini, T. Dartigalongue, R. D. Miller, *Science* 323 (2009) 1033–1037.
- [32] T. White, S. Richardson, B. Crowley, L. Pattison, J. Harris, G. Gregori, *Physical review letters* 111 (2013) 175002.
- [33] F. Lambert, J. Clérouin, G. Zerah, *Phys. Rev. E* 73 (2006) 016403.
- [34] S. Zhang, H. Wang, W. Kang, P. Zhang, X. He, *Physics of Plasmas* 23 (2016) 042707.
- [35] A. Cangi, A. Pribram-Jones, *Physical Review B* 92 (2015) 161113.
- [36] S. X. Hu, B. Militzer, L. A. Collins, K. P. Driver, J. D. Kress, *Phys. Rev. B* 94 (2016).
- [37] P. Suryanarayana, *Chemical Physics Letters* 584 (2013) 182–187.
- [38] P. P. Pratapa, P. Suryanarayana, J. E. Pask, *Computer Physics Communications* 200 (2016) 96–107.
- [39] E. Prodan, W. Kohn, *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005) 11635–11638.
- [40] A. Anantharaman, E. Cancès, *Annales de l’Institut Henri Poincaré (C) Non Linear Analysis* 26 (2009) 2425–2455.

- [41] J. E. Pask, P. A. Sterne, *Phys. Rev. B* 71 (2005) 113101.
- [42] V. Gavini, J. Knap, K. Bhattacharya, M. Ortiz, *Journal of the Mechanics and Physics of Solids* 55 (2007) 669 – 696.
- [43] P. Suryanarayana, V. Gavini, T. Blesgen, K. Bhattacharya, M. Ortiz, *Journal of the Mechanics and Physics of Solids* 58 (2010) 256 – 280.
- [44] P. Suryanarayana, K. Bhattacharya, M. Ortiz, *Journal of Computational Physics* 230 (2011) 5226 – 5238.
- [45] P. Suryanarayana, D. Phanish, *Journal of Computational Physics* 275 (2014) 524 – 538.
- [46] S. Ghosh, P. Suryanarayana, *Journal of Computational Physics* 307 (2016) 634–652.
- [47] D. Bowler, T. Miyazaki, M. Gillan, *Journal of Physics: Condensed Matter* 14 (2002) 2781.
- [48] C.-K. Skylaris, P. D. Haynes, *The Journal of chemical physics* 127 (2007) 164712.
- [49] P. Suryanarayana, *Chemical Physics Letters* 679 (2017) 146–151.
- [50] P. Suryanarayana, K. Bhattacharya, M. Ortiz, *Journal of the Mechanics and Physics of Solids* 61 (2013) 38 – 60.
- [51] L. Lin, C. Yang, *SIAM Journal on Scientific Computing* 35 (2013) S277–S298.
- [52] A. S. Banerjee, P. Suryanarayana, J. E. Pask, *Chemical Physics Letters* 647 (2016) 31–35.
- [53] D. G. Anderson, *Journal of the ACM (JACM)* 12 (1965) 547–560.
- [54] P. Pulay, *Chemical Physics Letters* 73 (1980) 393–398.
- [55] P. P. Pratapa, P. Suryanarayana, J. E. Pask, *Journal of Computational Physics* 306 (2016) 43–54.
- [56] P. Suryanarayana, P. P. Pratapa, J. E. Pask, arXiv preprint arXiv:1606.08740 (2016).
- [57] J. R. Shewchuk, *An introduction to the conjugate gradient method without the agonizing pain*, 1994.
- [58] D. Alfè, *Computer Physics Communications* 118 (1999) 31–33.
- [59] W. Gropp, E. Lusk, A. Skjellum, *Using MPI: portable parallel programming with the message-passing interface, volume 1*, MIT press, 1999.
- [60] S. Ghosh, P. Suryanarayana, *Computer Physics Communications* 216 (2017) 109 – 125.
- [61] S. Ghosh, P. Suryanarayana, *Computer Physics Communications* 212 (2017) 189 – 204.

- [62] W. Gropp, T. Hoefler, R. Thakur, E. Lusk, Using advanced MPI: Modern features of the message-passing interface, MIT Press, 2014.
- [63] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, United States Governm. Press Office Los Angeles, CA, 1950.
- [64] P. P. Pratapa, P. Suryanarayana, *Mechanics Research Communications* 78 (2016) 27–33.
- [65] A. Gil, J. Segura, N. M. Temme, Numerical methods for special functions, SIAM, 2007.
- [66] W. H. Press, Numerical recipes 3rd edition: The art of scientific computing, Cambridge university press, 2007.
- [67] S. Goedecker, M. Teter, *Physical Review B* 51 (1995) 9455.
- [68] S. Goedecker, L. Colombo, *Physical Review Letters* 73 (1994) 122.
- [69] N. Troullier, J. L. Martins, *Physical Review B* 43 (1991) 1993–2006.
- [70] J. P. Perdew, Y. Wang, *Physical Review B* 45 (1992) 13244.
- [71] D. M. Ceperley, B. J. Alder, *Phys. Rev. Lett.* 45 (1980) 566–569.
- [72] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J. Y. Raty, D. C. Allan, *Computational Materials Science* 25 (2002) 478–492(15).
- [73] T. Ono, K. Hirose, *Phys. Rev. Lett.* 82 (1999) 5016–5019.
- [74] N. S. Bobbitt, G. Schofield, C. Lena, J. R. Chelikowsky, *Phys. Chem. Chem. Phys.* (2015). DOI: 10.1039/c5cp02561c.
- [75] Lawrence Livermore National Laboratory (LLNL) high performance computing systems: <https://computation.llnl.gov/computing/machine-catalog>, accessed 2017-06-27.
- [76] S. Zhang, K. P. Driver, F. Soubiran, B. Militzer, *High Energy Density Physics* 21 (2016) 16–19.
- [77] S. Zhang, K. P. Driver, F. Soubiran, B. Militzer, *J. Chem. Phys.* 146 (2017).
- [78] A. Shamp, E. Zurek, T. Ogitsu, D. E. Fratanduono, S. Hamel, *Phys. Rev. B* 95 (2017).
- [79] K. P. Driver, F. Soubiran, S. Zhang, B. Militzer, *High Energy Density Physics* 23 (2017) 81–89.
- [80] S. Zhang, K. P. Driver, F. Soubiran, B. Militzer, *Phys. Rev. E* 96 (2017).
- [81] D. C. Rapaport, The art of molecular dynamics simulation, Cambridge university press, 2004.

- [82] L. Lin, J. Lu, L. Ying, E. Weinan, *Journal of Computational Physics* 231 (2012) 2140–2154.
- [83] G. Zhang, L. Lin, W. Hu, C. Yang, J. E. Pask, *Journal of Computational Physics* 335 (2017) 426–443.
- [84] G. Kresse, J. Furthmüller, *Physical Review B* 54 (1996) 11169–11186.