

Sampling-free computation of finite temperature material properties in isochoric and isobaric ensembles using the mean-field anharmonic bond model

Raynol Dsouza ^{1,*}, Marvin Poul ¹, Liam Huber ², Thomas D. Swinburne ^{3,†} and Jörg Neugebauer ¹

¹Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, 40327, Germany

²Grey Haven Solutions, Victoria, British Columbia, V8R 4C6, Canada

³Aix-Marseille Université, CNRS, CINaM UMR 7325, Campus de Luminy, Marseille 13288, France



(Received 15 November 2023; revised 18 January 2024; accepted 5 February 2024; published 20 February 2024)

The recently introduced mean-field anharmonic bond model has shown remarkable accuracy in predicting finite temperature free energies for certain potential models of fcc crystals without thermodynamic sampling. In this work, we extend the model to treat modern machine learning potentials in both isochoric and isobaric ensembles while preserving existing vibrational correlations and ensuring thermodynamic self-consistency. Testing against molecular dynamics simulations of bulk fcc Al and Cu, we find free energies with an accuracy of a few meV/atom up to the melting temperature under typical operating pressures, with similar accuracy for thermal expansion. Our sampling-free estimation is universally superior to the quasiharmonic approximation for less than ten percent of the computational cost and many orders of magnitude more efficient than thermodynamic integration. We discuss applications of the method in modern computational materials science workflows.

DOI: [10.1103/PhysRevB.109.064108](https://doi.org/10.1103/PhysRevB.109.064108)

I. INTRODUCTION

Accurately determining temperature-dependent material properties, such as free energy and thermal expansion, is essential for gaining insights into and predicting material behavior. Unlike at $T = 0$ K, where a single minimized configuration is sufficient, computing these properties at finite temperatures demands statistical averaging over a vast ensemble of distinct configurations [1,2]. Sampling configurations on this scale becomes impractical when using density functional theory (DFT) based [3] methods, which are highly precise but resource intensive. Consequently, well-established approximations like the harmonic approximation and quasiharmonic approximation (QHA), which allow for the analytical evaluation of thermodynamic properties, are routinely employed [4,5].

While the QHA has proved effective in addressing anharmonicity due to volume expansion, several studies have shown that it falls short of accurately capturing the full extent of anharmonic effects [6–11]. Accounting for anharmonicity beyond the QHA involves utilizing sampling techniques such as thermodynamic integration (TI) [12,13], which require a suitable reference system, typically taken to be the QHA. The numerical efficiency of TI depends on the similarity between the potential energy surfaces of the reference and target

systems within the region of configuration space where thermodynamic sampling occurs [14]. When the surfaces diverge, it demands extensive sampling to ensure accurate results.

To construct a reference system that better approximates the target system than the QHA, Glensk *et al.* [8] proposed a “local anharmonic” (LA) approximation, which assumes that contributions to the internal energy of an atom beyond the harmonic approximation are dominated by its nearest neighbors. For a fcc system, it was deduced that contributions from just the first nearest neighbors are sufficient to describe anharmonicity. This deduction was based on the observation that the density of bonds between an atom and its first nearest neighbors, as obtained from *ab initio* molecular dynamics (MD) simulations, is symmetric exclusively along the transversal directions (perpendicular to the bond) but asymmetric along the longitudinal direction (along the bond). Meanwhile, the bond densities of the second and third nearest neighbors were found to be nearly symmetric along all three directions.

Following this observation, a simple first nearest neighbor LA bonding potential was parameterized by sampling the potential energy surface around an atom along the longitudinal and transversal directions. Passing trajectories to this bonding potential yielded forces and energies that closely matched those obtained from *ab initio* MD, suggesting it is a better reference for TI than the QHA. Additionally, the LA-predicted free energies were in close agreement with those obtained from *ab initio* TI, typically differing by only a few meV/atom across various fcc materials, significantly speeding up TI convergence. However, the requirement to perform thermodynamic sampling still resulted in a much higher computational cost when compared to the QHA.

To eliminate the need for thermodynamic sampling, Swinburne *et al.* [15] introduced the mean-field anharmonic bond (MAB) model [16] for fcc crystals. In the MAB model,

*dsouza@mpie.de

†thomas.swinburne@cnrs.fr

the first nearest neighbor bond density at any given temperature is estimated analytically as a function of the bonding potential by using bond symmetries and bond-pair correlations and enforcing thermodynamic self-consistency. The anharmonic free energies predicted by this model were also found to be within a few meV/atom of those obtained from TI, requiring mere seconds of computation time. The MAB model offers the same level of simplicity as the harmonic models while demonstrating the ability to predict anharmonic free energies with a comparable computational efficiency. However, in [15], the model was tested only on directionally separable first nearest neighbor bonding potentials and was restricted to the NVT ensemble.

This work extends the MAB model to utilize not just analytical potential models as bonding potentials but also potential models of any functional form using the concept of LA bonding potentials (introduced by Glensk *et al.* [8]) in both the NVT and NPT ensembles of bulk fcc crystals.

We first revisit the core principles of the LA approximation and provide a detailed insight into how a mean-field approximation of the bond density simplifies the evaluation of thermodynamic properties. By incorporating physically motivated offsets to the analytical pressure and internal energy of the model from a $T = 0$ K energy-volume calculation, we extend its applicability to the NPT ensemble at negligible additional computational expense. We also make refinements to the estimated bond density based on correlations observed in the effective potentials of the bond densities obtained from MD trajectories. The improved model is assessed by employing a highly anharmonic analytical first nearest neighbor Morse potential and LA bonding potentials parameterized from complex machine learning potential models for Al and Cu [17]. We aim to demonstrate the numerical efficiency of the improved MAB model over harmonic models in accurately estimating finite temperature thermodynamic properties (internal energy, thermal expansion, and free energy) of fcc crystals without the need for thermodynamic sampling in both NVT and NPT ensembles.

II. METHODOLOGY

A. Bond lattice coordinates

Consider a crystal of N atoms in d dimensions with a volume $V_0(1 + \epsilon)^d$, where ϵ is an isotropic strain, such that $\epsilon = 0$ corresponds to zero strain. Each atom i at position $\{\mathbf{R}_i\}_{i=1}^{i=N}$ has m nearest neighbors, with the index of its l th neighbor, $l \in [1, m]$, given by $l(i)$. We define a ‘‘bond vector’’ $\mathbf{b}_i^{(l)} \in \mathbb{R}^d$ as the separation between atoms i and $l(i)$, namely,

$$\mathbf{b}_i^{(l)} \equiv \mathbf{R}_{l(i)} - \mathbf{R}_i. \quad (1)$$

Without any loss of generality, we are free to represent the atomic positions by the set $\{\mathbf{b}\}$ of all bond vectors, but as they clearly have mdN degrees of freedom, physical bond configurations must lie on some dN -dimensional surface $\mathcal{C}(\{\mathbf{b}\}) = 0$. Previous work [15] showed that this surface is defined by the compatibility constraint

$$\mathbf{b}_i^{(l)} - \mathbf{b}_{k(i)}^{(l)} = \mathbf{b}_i^{(k)} - \mathbf{b}_{l(i)}^{(k)}, \quad (2)$$

such that neighboring bonds do not move independently of each other. At this stage, we have simply defined an overcomplete set of coordinates for the atomic system under study.

B. Local anharmonic approximation

The LA approximation [8] assumes that nearest neighbor (or ‘‘local’’) interactions dominate anharmonic contributions to the internal energy of a target system. The LA model thus connects the m nearest neighbors of each atom and has a total energy

$$U_{\text{LA}}(\{\mathbf{b}\}) = \frac{1}{2} \sum_{i=1}^N \sum_{l=1}^m V_l(\mathbf{b}_i^{(l)}), \quad (3)$$

with V_l being the ‘‘bonding potential’’ for the bond vector $\mathbf{b}_i^{(l)}$. The factor of $1/2$ accounts for each pair of atoms sharing a bond. The expected anharmonic internal energy per atom at some temperature T and strain ϵ then reads

$$\begin{aligned} \mathcal{U}_{\text{LA}}^{\text{ah}}(T, \epsilon) &= \langle U_{\text{LA}}(\{\mathbf{b}\})/N \rangle_{T, \epsilon} - \frac{3}{2} k_B T \\ &= \frac{1}{2} \sum_l^m \int V_l(\mathbf{b}) \rho_l(\mathbf{b}, \epsilon, T) d\mathbf{b} - \frac{3}{2} k_B T, \end{aligned} \quad (4)$$

where $\rho_l(\mathbf{b}, \epsilon, T)$ is the bond density of the l th bond at T and ϵ , i.e., the $d = 3$ dimensional equilibrium distribution of $\mathbf{b}_i^{(l)}$, which for unaries is identical for all $i \in [1, N]$. $\mathcal{U}_{\text{LA}}^{\text{ah}}$ can be used to calculate the per atom anharmonic free energy

$$\mathcal{F}_{\text{LA}}^{\text{ah}}(T, \epsilon) \equiv T \int_0^T \mathcal{U}_{\text{LA}}^{\text{ah}}(T', \epsilon) / (T')^2 dT'. \quad (5)$$

The model also has a virial temperature

$$T_{\text{vir}}(T, \epsilon) = \frac{1}{6k_B} \sum_l^m \int (\mathbf{b} - \mathbf{a}_l^\epsilon) \cdot \nabla V_l(\mathbf{b}) \rho_l(\mathbf{b}, \epsilon, T) d\mathbf{b} \quad (6)$$

and a pressure of

$$P(T, \epsilon) = -\frac{P_0(\epsilon)}{2} \sum_l^m \int \mathbf{a}_l^\epsilon \cdot \nabla V_l(\mathbf{b}) \rho_l(\mathbf{b}, \epsilon, T) d\mathbf{b}, \quad (7)$$

where $P_0(\epsilon) = N/[3V_0(1 + \epsilon)^3]$ and \mathbf{a}_l^ϵ is the associated equilibrium lattice vector. We discuss how these expressions provide self-consistency requirements for the mean-field model in Sec. II F.

C. LA approximation for bulk fcc crystals

For a relaxed, bulk fcc crystal, all N atoms are equivalent in terms of their local environments, each with m nearest neighbors. If we assume that only the $m = 12$ first nearest neighbors of an atom contribute to the anharmonicity in the system, in the absence of lattice vibrations, $\mathbf{b}^{(l)} \equiv \mathbf{a}_l(1 + \epsilon)$, i.e., the associated equilibrium lattice vector with the appropriate strain. Any two equilibrium lattice vectors \mathbf{a}_l and \mathbf{a}_k are related to each other by a symmetry operation represented as a matrix $\mathbf{G}_{kl} \in \mathbb{R}^{d \times d}$, such that $\mathbf{a}_l = \mathbf{G}_{kl} \mathbf{a}_k$. This reduces the bonding potentials and bond densities for l to

$$V_l(\mathbf{b}^{(l)}) \equiv V_1(\mathbf{G}_{1l} \mathbf{b}^{(l)}), \quad \rho_l(\mathbf{b}^{(l)}) \equiv \rho_1(\mathbf{G}_{1l} \mathbf{b}^{(l)}), \quad (8)$$

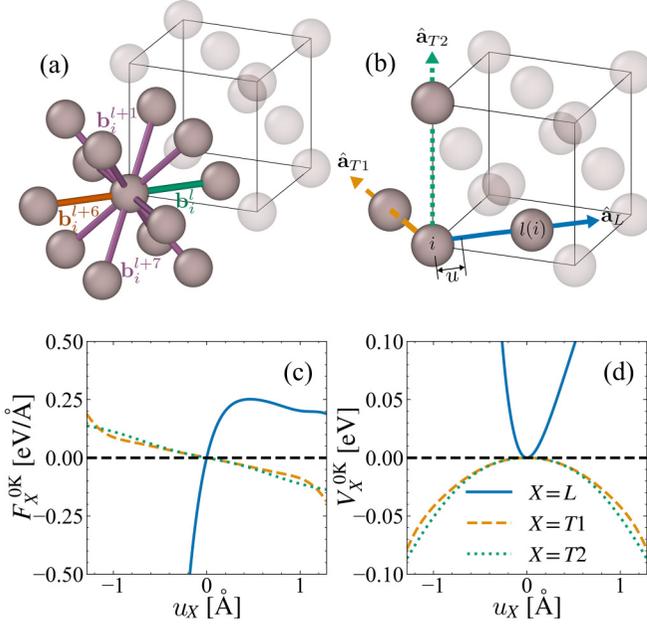


FIG. 1. (a) “Cage” of first nearest neighbors of an atom in an fcc crystal, with the reference bond in green, its antiparallel bond in orange, and the remaining bonds in violet. (b) Displacement directions of atom i for the parametrization of the LA bonding potential from the Cu-ACE potential model with the resultant (c) forces and (d) potential along each direction (blue solid line along $\hat{\mathbf{a}}_L$, dashed yellow line along $\hat{\mathbf{a}}_{T1}$, and dotted green line along $\hat{\mathbf{a}}_{T2}$).

respectively. The per atom anharmonic internal energy as a function of T and ϵ can then be rewritten for bulk fcc crystals as

$$\mathcal{U}_{\text{fcc}}^{\text{ah}}(T, \epsilon) \equiv 6 \int V_1(\mathbf{b}) \rho_1(\mathbf{b}, T, \epsilon) d\mathbf{b} - \frac{3}{2} k_B T. \quad (9)$$

D. LA bonding potential

The LA approximation requires the definition of a single bonding potential V_1 , which must depend exclusively on the nearest neighbor bond vectors. A simple first nearest neighbor bonding potential can be parameterized from any potential model from the relaxed $T = 0$ K structure by sampling the potential energy surface around each inequivalent atom along the longitudinal ($\hat{\mathbf{a}}_L$) and two transversal ($\hat{\mathbf{a}}_{T1}$ and $\hat{\mathbf{a}}_{T2}$) directions [8].

As all atoms in a bulk fcc crystal are equivalent, one of the atoms i is displaced along each of the three principal directions ($\hat{\mathbf{a}}_L$, $\hat{\mathbf{a}}_{T1}$, and $\hat{\mathbf{a}}_{T2}$), as shown in Fig. 1(b), and the resultant force $\mathbf{F}_{l(i)}$ on its neighbor $l(i)$ is recorded. The force [Fig. 1(c)] and potential [Fig. 1(d)] along the principal directions are then

$$F_X(u) = \mathbf{F}_{l(i)}(u\hat{\mathbf{a}}_X) \cdot \hat{\mathbf{a}}_X', \quad V_X(u) = - \int_0^u F_X(u') du', \quad (10)$$

respectively, where u is the displacement along the direction $X = L, T1, T2$. The vector $\hat{\mathbf{a}}_X'$ corresponds to the fact that for displacements along the transversal directions, $\hat{\mathbf{a}}_{l(i)}$ [the separation between atoms i and $l(i)$], $\hat{\mathbf{a}}_{T1}$, and $\hat{\mathbf{a}}_{T2}$ are no longer orthogonal, as $\hat{\mathbf{a}}_L$ is replaced by $\hat{\mathbf{a}}_{il(i)}$. The orthonormal

basis can be recovered by applying a Gram-Schmidt process to the vectors spanned by $\hat{\mathbf{a}}_{il(i)}$, giving $\hat{\mathbf{a}}'_{T1}$ and $\hat{\mathbf{a}}'_{T2}$. For displacements along the transversal directions, $\mathbf{F}_{l(i)}$ also has a component along $\hat{\mathbf{a}}_{il(i)}$, which makes the force along each transversal direction

$$F_{Tx}(u_{Tx}) = \mathbf{F}_{l(i)} - F_{il(i)}(u_{il(i)}), \quad (11)$$

where $x = 1, 2$, with $u_{il(i)} = |\mathbf{b}|$, $u_{T1} = \mathbf{b} \cdot \hat{\mathbf{a}}'_{T1}$, and $u_{T2} = \mathbf{b} \cdot \hat{\mathbf{a}}'_{T2}$. The resulting bonding potential is then

$$V_1^{\text{LA}} = V_L(u_{il(i)}) + V_{T1}(u_{T1}) + V_{T2}(u_{T2}). \quad (12)$$

Due to symmetry relations of the fcc crystal, it is sufficient to generate only two sets of displacements (one along $\hat{\mathbf{a}}_L$ and another along $\hat{\mathbf{a}}_{T2}$, as elaborated further in Sec. II H to fully parametrize an LA bonding potential.

E. Mean-field anharmonic bond model

To remove the computational burden of thermodynamic sampling, Swinburne *et al.* [15] recently introduced the MAB model, which aims to estimate thermodynamic properties of the LA model directly from properties of the bonding potential. The target of the MAB model is to estimate a single bond density $\rho_{\text{MAB}}(\mathbf{b}, \epsilon, T)$, or, more specifically, its Boltzmann inverse, the “effective” bonding potential

$$\tilde{V}(\mathbf{b}, \epsilon, T) = -k_B T \ln |\rho_{\text{MAB}}(\mathbf{b}, \epsilon, T)|, \quad (13)$$

such that $\rho_{\text{MAB}}(\mathbf{b}, \epsilon, T) \approx \rho_1(\mathbf{b}, \epsilon, T)$. If achieved, this allows efficient evaluation of $\mathcal{U}_{\text{fcc}}^{\text{ah}}$ in NVT and NPT ensembles through rapid three-dimensional quadrature using Eq. (9).

The mean-field approximation is applied to the compatibility constraint in Eq. (2), simplifying the treatment of bond-pair correlations. The first approximation simply confines each atom within a “cage” of static nearest neighbors, as illustrated in Fig. 1(a), implying

$$\mathbf{b}^{(l)} - \mathbf{b}^{(k)} = \langle \mathbf{b}_i^{(l)} \rangle - \langle \mathbf{b}_i^{(k)} \rangle = \mathbf{a}_l^\epsilon - \mathbf{a}_k^\epsilon. \quad (14)$$

Under this approximation, when one of the neighbor bonds changes, it becomes straightforward to ascertain how the bonds that connect the atom to its other neighbors are affected. Using the relations defined in Eq. (14), the effective potential for the mean-field model of the fcc crystal is written as

$$\tilde{V}_{\text{mf}}(\mathbf{b}, \epsilon) = \frac{1}{2} \sum_{l=1}^{12} V_l[\mathbf{G}_{1l}(\mathbf{b} - \mathbf{a}_l^\epsilon) + \mathbf{a}_l^\epsilon]. \quad (15)$$

However, the approximation in Eq. (14) was found to be too restrictive; in particular, any reference bond l [green bond in Fig. 1(a)] had a strong correlation to its antiparallel bond (orange bond), with index $l + 6$. The mean-field model is modified to reflect this, giving

$$\mathbf{b}^{(l)} - \mathbf{b}^{(k)} = \langle \mathbf{b}_i^{(l)} \rangle - \langle \mathbf{b}_i^{(k)} \rangle = \mathbf{a}_l^\epsilon - \mathbf{a}_k^\epsilon, \quad k \neq l + 6, \quad (16)$$

which effectively extends the cage along the equilibrium bond direction, forming an infinite “chain” of bonds. This one-dimensional system can be solved analytically, giving the *correlated* mean-field (mfc) effective potential

$$\tilde{V}_{\text{mfc}}(\mathbf{b}, \epsilon, T) = \frac{1}{2} \{ V_1(\mathbf{b}) - V_1[\mathbf{G}_{17}(\mathbf{b} - \mathbf{a}_1^\epsilon) + \mathbf{a}_1^\epsilon] \} + \tilde{V}_{\text{mf}}(\mathbf{b}, \epsilon) + \lambda(\epsilon, T) \hat{\mathbf{a}} \cdot \mathbf{b}, \quad (17)$$

where $\lambda(\epsilon, T)$ is a scalar Lagrange multiplier that conserves the volume of the system, such that $\langle \mathbf{b} \rangle = \mathbf{a}^\epsilon$ at temperature T and strain ϵ . This is achieved by finding the unique root of $f(\lambda) = \langle \mathbf{b}; \lambda(\epsilon, T) = \lambda \rangle \cdot \mathbf{a}^\epsilon - \|\mathbf{a}^\epsilon\|^2$, which typically requires only milliseconds of CPU effort.

F. Thermodynamic self-consistency

The expressions for the virial temperature [Eq. (6)] and pressure [Eq. (7)] introduced in Sec. II B provide valuable self-consistent criteria for the mean-field model in both NVT and NPT ensembles. The MAB model already has one self-consistent property: the Lagrange multiplier $\lambda(\epsilon, T)$ introduced in Eq. (17), which ensures self-consistency for the volume such that $\langle \mathbf{b} \rangle = \mathbf{a}^\epsilon$.

In a sufficiently equilibrated MD simulation, where many-body interactions are properly accounted for, $T_{\text{vir}} = T$. However, in earlier work [15], it was found that $T_{\text{vir}} < T$ for the bond lattice model. This is because a self-consistent temperature definition is not guaranteed in a mean-field treatment unless all the nearest neighbor bonds and their respective correlations with each other are considered in the model. As our MAB model accounts for only the first nearest neighbor bonds and their correlations, we enforce a self-consistent temperature renormalization condition

$$T_{\text{vir}}(T_{\text{eff}}(T), \epsilon) \equiv T, \quad (18)$$

giving a renormalized bond density

$$\rho_{\text{MAB}}(\mathbf{b}, \epsilon, T) \rightarrow \rho_{\text{MAB}}(\mathbf{b}, \epsilon, T_{\text{eff}}(T)). \quad (19)$$

Similar to the Lagrange multiplier, $T_{\text{eff}}(T)$ is determined by finding the unique root of $f(T_{\text{eff}}) = T_{\text{vir}}(T_{\text{eff}}, \epsilon) - T$. This raises the cost of the MAB model to around 0.02–0.2 CPU seconds per temperature and strain.

For the NPT ensemble, the model must satisfy the same temperature condition [Eq. (18)] as for NVT , but now the strain is also free to vary such that

$$P(T, \epsilon) = P. \quad (20)$$

This can typically be determined in $O(10)$ iterations, giving a final cost of around 0.2–2 CPU seconds per temperature and pressure.

The original MAB model was tested on directionally separable first nearest neighbor bonding potentials in the NVT ensemble. In this work, we target much more complex, many-body potentials. Targeting meV/atom accuracy for these more complex systems necessitates two additional refinements to the MAB model, both of which incur no significant additional computational effort.

G. Refinement 1: Energy-volume dependence

As $T \rightarrow 0$, the bond density ρ_1 converges to a δ function centered around $\mathbf{b} = \mathbf{a}^\epsilon = (1 + \epsilon)\mathbf{a}_0$, with the internal energy per atom reducing to

$$\lim_{T \rightarrow 0} \mathcal{U}_{\text{fcc}}(T, \epsilon) = 6V_1[(1 + \epsilon)\mathbf{a}_0]. \quad (21)$$

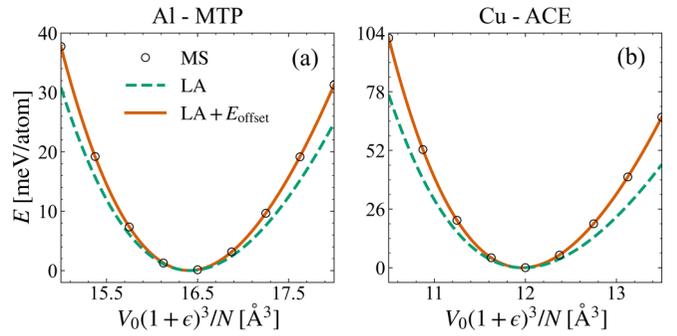


FIG. 2. Energy-volume curves of the (a) Al-MTP and (b) Cu-ACE potential models from static calculations (black circles, labeled MS) and the LA bonding potential, without (dashed green line) and with (solid orange line) the offset term.

We can therefore calculate an energy-volume curve, giving a zero temperature pressure

$$P_0(\epsilon) = -\frac{dU}{dV} = \frac{-1}{3V_0(1 + \epsilon)^2} \frac{dU}{d\epsilon}, \quad (22)$$

using the identity $dV = 3V_0(1 + \epsilon)^2 d\epsilon$.

When applying an isotropic strain to an atomic system at $T = 0$ K, the volume V of the system (and hence the equilibrium bond \mathbf{a}^ϵ) changes proportionally to the applied strain. For a simple first nearest neighbor Morse bonding potential, $E_{\text{Morse}}(V) = E_{\text{LA}}(V)$, as there are no other interactions contributing to E_{Morse} . For molecular static (MS) calculations with many-body atomic interactions, $E_{\text{MS}}(V) \neq E_{\text{LA}}(V)$ in general, as seen in Fig. 2, due to the first nearest neighbor assumption in E_{LA} . In order to account for this inequality, an additional correction term to the internal energy of the system, $E_{\text{offset}}(\epsilon) = E_{\text{MS}}(\epsilon) - E_{\text{LA}}(\epsilon)$, is necessary. Equation (9) thus becomes

$$\mathcal{U}_{\text{fcc}}^{\text{ah}}(T, \epsilon) \rightarrow \mathcal{U}_{\text{fcc}}^{\text{ah}}(T, \epsilon) + \frac{E_{\text{offset}}(\epsilon)}{N}. \quad (23)$$

As the pressure of a strained system becomes nonzero when $\frac{dU}{dV} \neq 0$, the correction term is also reflected in the pressure evaluated by the model, modifying Eq. (7) to

$$P(T, \epsilon) \rightarrow P(T, \epsilon) - \frac{N}{3V_0(1 + \epsilon)^2} \frac{d}{d\epsilon} E_{\text{offset}}(\epsilon). \quad (24)$$

Evaluation of $E_{\text{MS}}(\epsilon)$ using an interatomic potential is inexpensive and is a necessary step in obtaining $T = 0$ K structures from DFT methods. As such, this step does not add any additional computational overhead. However, we find it essential for accurate predictions in the NPT ensemble.

H. Refinement 2: Transversal effective potential

Our final model refinement concerns the transversal components of the bond density, defined by the lattice directions $(\hat{\mathbf{a}}_L, \hat{\mathbf{a}}_{T1}, \hat{\mathbf{a}}_{T2})$. Figure 3(a) shows contour lines for $\rho_{\text{MAB}}^{\text{mfc}}$ superimposed on contours of ρ_{MD} extracted from finite temperature MD. “Projections” of their effective potentials along the lattice directions are shown in Fig. 3(b), with the longitudinal projection defined in Eq. (25) below.

Figure 3(b) clearly shows that the MAB model accurately predicts the longitudinal projection, and both the model and

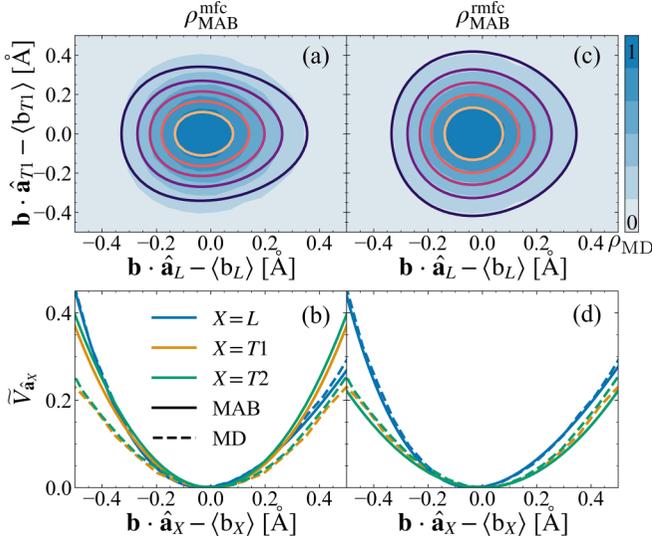


FIG. 3. Contours of the first nearest neighbor single bond density for a first nearest neighbor Morse potential from the (a) original MAB model and the (c) refined model, superimposed on the colored MD single bond density (intensifying color indicates increasing density). The one-dimensional projections of their effective potentials along $\hat{\mathbf{a}}_L$ (blue), $\hat{\mathbf{a}}_{T1}$ (yellow), and $\hat{\mathbf{a}}_{T2}$ (green) are shown in (b) and (d), respectively. The solid lines indicate the model projections, while the dashed lines indicate the MD projections.

MD give symmetric transversal projections. However, while the MD transversal projections are essentially harmonic and closely follow the attractive portion of the longitudinal projection, the MAB transversal projections are much stiffer and more closely follow the repulsive portion of the longitudinal projection. This behavior was observed in all test systems studied; it is expected, as the system will typically make many-body rearrangements (not captured in the MAB model) to avoid high-energy configurations.

Rather than build new constraints into the MAB model to incorporate these observations, which would raise the computational cost, we instead employ a simple *ansatz* that gives highly accurate transversal projections for negligible additional effort. We first project the MAB effective potential [Eq. (17)] along the longitudinal direction

$$\tilde{V}_{\text{mfc}}^L(\mathbf{b}) \equiv \tilde{V}_{\text{mfc}}(\mathbf{b} \cdot \hat{\mathbf{a}}_L). \quad (25)$$

We then fit a harmonic potential $\kappa_T \mathbf{b}^2/2$ to $\tilde{V}_{\text{mfc}}^L(|\mathbf{b}|)$ for the transverse displacements, retaining $\tilde{V}_{\text{mfc}}^L(\mathbf{b})$ for longitudinal displacements. This yields a “refined” effective potential

$$\tilde{V}_{\text{mfc}}(\mathbf{b}) = \tilde{V}_{\text{mfc}}^L(\mathbf{b} \cdot \hat{\mathbf{a}}_L) + \frac{\kappa_T}{2} (\mathbf{b} \cdot \hat{\mathbf{a}}_{T1})^2 + \frac{\kappa_T}{2} (\mathbf{b} \cdot \hat{\mathbf{a}}_{T2})^2. \quad (26)$$

As evident from Figs. 3(c) and 3(d), the resulting bond density $\rho_{\text{MAB}}^{\text{rmfc}}$ gives much better agreement with ρ_{MD} than $\rho_{\text{MAB}}^{\text{mfc}}$. We note that while Eq. (26) is separable, this cannot be leveraged to accelerate quadrature as the LA bonding potential [Eq. (12)] is not separable in general.

I. Computational details

The workflow for the MAB model is implemented within the PYIRON [18,19] integrated development environment. The model was tested for fcc structures with three different potential models: a highly anharmonic first nearest neighbor Morse potential (with parameters $D_e = 0.143$, $\alpha = 2.163$, and $r_e = 2.856$), a publicly available atomic cluster expansion (ACE) [20] potential for Cu [17], and a newly fitted general-purpose moment tensor potential (MTP) [21,22] for Al. The Al-MTP was fitted according to the methodology outlined in [23] and is available on request; a detailed description of the potential properties will be the subject of a separate publication.

Reference MD data for each potential model was generated using the LAMMPS simulator [24] within PYIRON using a $4 \times 4 \times 4$ supercell for 15 temperatures ranging from 5 up to 1000 K for the Morse potential and up to their respective experimental melting temperatures for Al-MTP (933.5 K) and Cu-ACE (1357.8 K). Ten independent runs of 100 000 steps each (with 1000 snapshots) were carried out for each temperature in both the *NVT* and *NPT* ensembles to reduce statistical errors. Reference QHA data were generated using the PHONOPY [25] package within PYIRON from nine strained structures for the same 15 temperatures. The MD anharmonic free energies were derived from the thermodynamic integration of the anharmonic internal energies over the square of the inverse temperature using Eq. (5). The energy-volume curves necessary for obtaining the offsets in Eqs. (23) and (24) were also evaluated from these strained structures.

The LA bonding potentials for Al-MTP and Cu-ACE were parameterized from the $T = 0$ K relaxed structure by performing static calculations at 15 displacements towards the first nearest neighbor (along $\hat{\mathbf{a}}_L$) and 15 more towards the second nearest neighbor (along $\hat{\mathbf{a}}_{T2}$) using the methodology described in Sec. IID. A $100 \times 50 \times 50$ ($L \times T1 \times T2$) mesh grid of bond vectors $\{\mathbf{b}\}$ was used for the MAB model. The thermodynamic self-consistency optimizations in the *NVT* and *NPT* ensembles were performed according to the formalism described in Sec. IIF using the Levenberg-Marquardt algorithm [26] in SCIPY.OPTIMIZE.ROOT [27].

III. RESULTS AND DISCUSSION

A. Accuracy of the LA bonding potential

We first compare the pointwise accuracy of per atom forces and energies predicted by the LA bonding potential and the QHA against MD reference data for Al-MTP and Cu-ACE. For a given configuration of a system with N atoms, we determine the average energy per atom U and the $3N$ atomic forces $F_{i\alpha}$, where $i \in [1, N]$ and $\alpha \in [x, y, z]$.

While force comparisons are straightforward, for energy evaluations, our primary focus lies in the deviation of the average energy from its harmonic expectation of $1.5k_B T$. However, as discussed in Sec. IIF, the LA model requires a temperature renormalization of T_{vir} . This renormalization is necessary as the LA model typically underestimates T by approximately 1%–7%, depending on the system and temperature. To account for this adjustment, we use the following

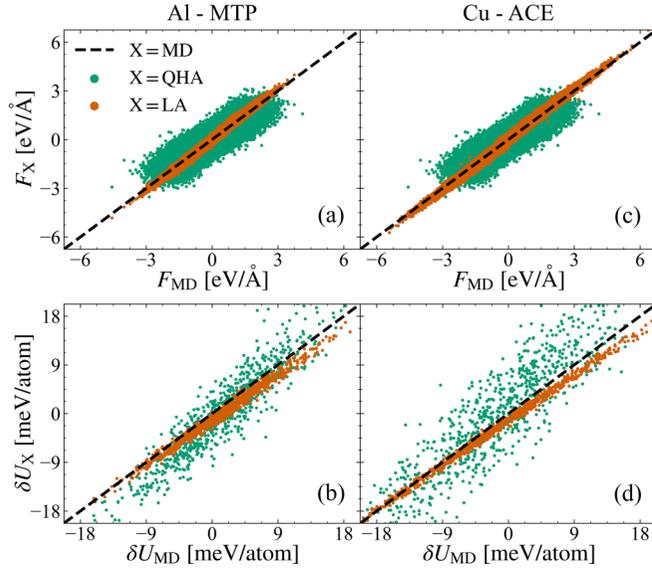


FIG. 4. Comparison of the pointwise forces and average internal energy per atom fluctuations predicted by the LA bonding potential for the (a) and (b) Al-MTP and (c) and (d) Cu-ACE potential models. $\delta U_{\text{MD,LA,QH}}$ are defined in (27), (28), and (29) respectively. We plot the most anharmonic case, the highest simulation temperatures ($T = 933.5$ K for Al and $T = 1357.8$ K for Cu), in the NVT ensemble at the $T = 0$ K volume.

definition for the per atom energy deviations, which do not affect force evaluations:

$$\delta U_{\text{MD}} = U_{\text{MD}} - U_0 - 1.5k_{\text{B}}T_{\text{MD}}, \quad (27)$$

$$\delta U_{\text{LA}} = U_{\text{LA}} - 1.5k_{\text{B}}T_{\text{LA}}, \quad (28)$$

$$\delta U_{\text{QH}} = U_{\text{QH}} - U_0 - 1.5k_{\text{B}}T_{\text{QH}}, \quad (29)$$

where T_{X} is the virial temperature calculated by each method. By construction, $T_{\text{QH}} = T_{\text{MD}} = T$. The comparison is illustrated in Fig. 4 for a target temperature of $T = 933.5$ K for Al-MTP and $T = 1357.8$ K for Cu-ACE in the NVT ensemble. The system volume was chosen to be the $T = 0$ K simulation volume, as it represents the most challenging test case, where second neighbor interactions are strongest.

Similar to the findings of Glensk *et al.* [8], our results show excellent agreement between the forces ($F_{\text{LA}}^{\text{rmse}} \approx 0.09$ eV/Å) and renormalized energies ($\delta U_{\text{LA}}^{\text{rmse}} \approx 1.5$ meV/atom) predicted by the LA bonding potential and the MD reference for both the machine learning potentials. The deviations in the QHA forces (energies) are the effect of fitting a linear (quadratic) function to small displacements (≈ 0.1 Å) from the equilibrium position of the atoms. In contrast, the displacements in the LA model are large (> 1 Å). While third- or fourth-order polynomials could effectively fit the transversal bonding forces and potentials [28], applying the same to the longitudinal force and potential leads to significant underfitting and fails to capture its anharmonic curvature as illustrated in Figs. 1(c) and 1(d). Although various functional forms that accurately fit the bonding force (potential) data points could be considered, we choose a cubic spline [29] function for its versatility.

B. Error propagation

One of the primary advantages of an analytical, sampling-free prediction of material properties is the elimination of statistical errors, which must be highly converged when targeting meV/atom accuracies. Even with over a million sampled configurations at each temperature, our MD data still exhibit noticeable uncertainty. In particular, the integral expression for $\mathcal{F}^{\text{ah}}(T, \epsilon)$ [Eq. (5)] accumulates sampling errors in $\mathcal{U}^{\text{ah}}(T, \epsilon)$, weighted by the square of the inverse temperature. This means that small errors in the internal energy at low temperatures are amplified, while errors at high temperatures are suppressed. The effect of this error propagation can be seen in the shaded regions and occasional discontinuities of the black circles in Figs. 5 and 6.

The reduction of statistical errors can be achieved only through more extensive thermodynamic sampling, which, while enhancing accuracy, reduces computational efficiency, making approaches like the MAB model an attractive alternative. We will return to considerations of computational efficiency in Sec. III E.

C. NVT ensemble

The predicted anharmonic internal energy, free energy, and system pressure in the NVT ensemble are shown in Fig. 5 for the three potential models. In addition to the machine learning potentials, we choose a first nearest neighbor Morse potential with relatively high anharmonicity. As there are no contributions from atoms beyond the first nearest neighbors, this choice establishes an ideal test system for the MAB model.

For the many-body potentials Al-MTP and Cu-ACE, the MAB model gives an excellent prediction of the small anharmonic internal energies up to around 65% of the melting temperature [Figs. 5(d) and 5(e)]. We attribute the underestimation at high temperatures to the first nearest neighbor approximation inherent in the MAB approach. At high temperatures and fixed volume, the high (> 4 GPa) system pressures cause second nearest neighbors to approach the central atom closely enough to modify the predicted bond density, leading to an upward shift in anharmonic internal energies. This attribution of error is supported by the absence of such an upward shift in the MD-predicted internal energies for the first nearest neighbor Morse potential [Fig. 5(a)].

Additionally, the model demonstrates very good predictive capability for pressure across all temperatures [see Figs. 5(c), 5(f) and 5(i)]. The systematic overestimation of the pressure for Cu-ACE with temperature could also be the result of a larger influence of second nearest neighbors. However, we find that despite the mismatch in pressure, the MAB model still predicts the correct thermal expansion in the NPT ensemble, as can be seen in Fig. 6(i).

We emphasize that the QHA, represented by the dashed green lines, does not account for anharmonic contributions or pressure at constant volume.

D. NPT ensemble

Figure 6 summarizes the predicted thermodynamic properties for the NPT ensemble for the three potential models at $P = 0$. As in the NVT ensemble, the influence of second

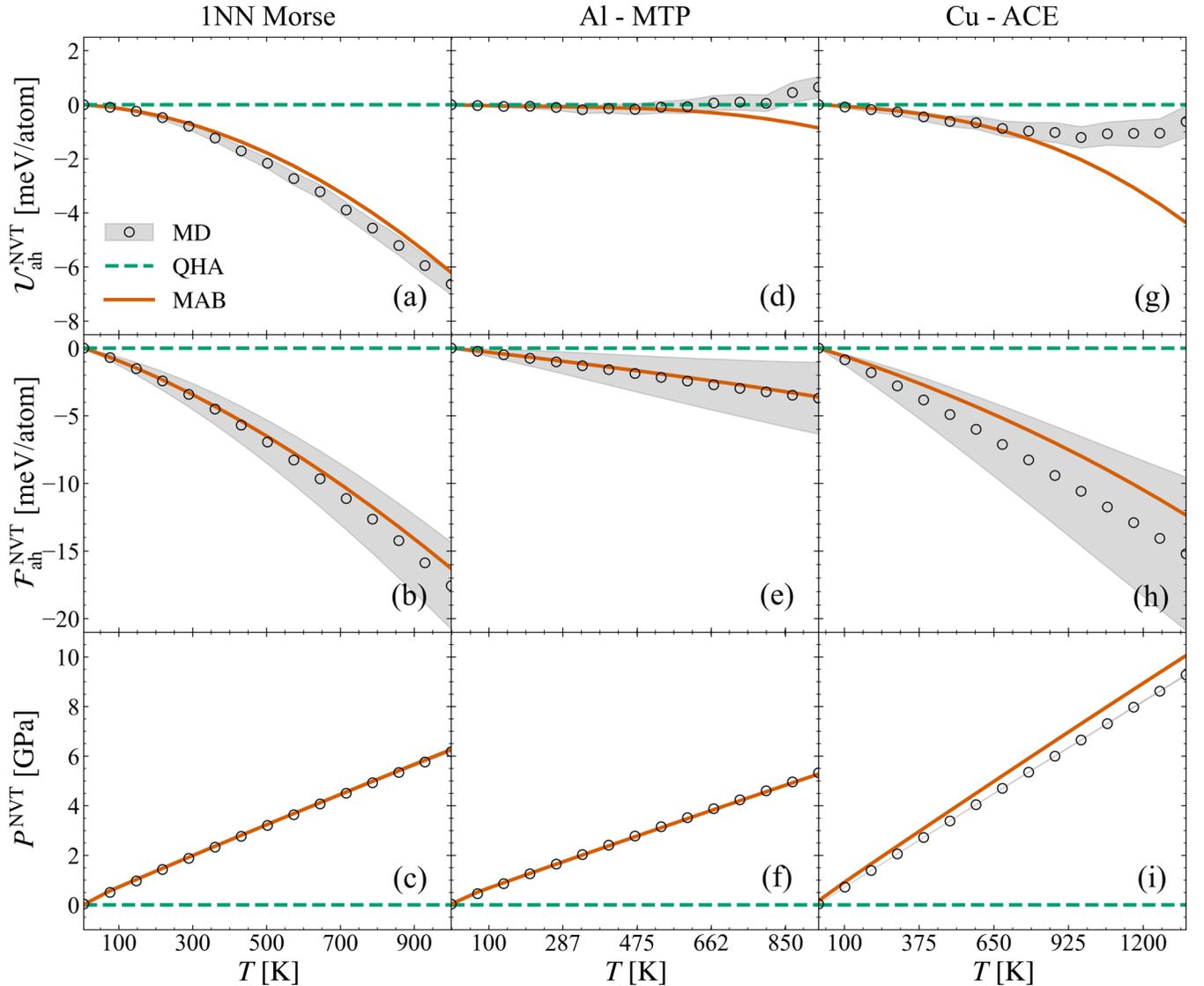


FIG. 5. Finite temperature per atom (a), (d), and (g) anharmonic internal energy, (b), (e), and (h) free energy, and (c), (f), and (i) system pressure predicted by MD (black circles), QHA (dashed green line), and the refined MAB model (solid orange line) for the NVT ensemble. Each column corresponds to properties evaluated from a different potential model: (a)–(c) the first nearest neighbor Morse, (d)–(f) the Al-MTP, and (g)–(i) the Cu-ACE potential models.

nearest neighbor interactions at high temperatures leads to a slight underestimation of the anharmonic internal energy, although the onset of these effects is delayed due to thermal expansion. However, it is noteworthy that even with this underestimation, the MAB model outperforms the QHA, as seen in the insets of Fig. 6.

The model-predicted anharmonic free energies [Figs. 6(b), 6(e) and 6(h)] are <7 meV/atom for Al-MTP and Cu-ACE, significantly better than the QHA. High temperature deviations in the anharmonic internal energy have limited impact on the anharmonic free energies as they are scaled by a factor of $1/T^2$. Thus, our primary quantity of interest, NPT anharmonic free energies, shows better agreement with brute force MD than the widely used QHA, despite being nearly 2 orders of magnitude more computationally efficient. This constitutes our main contribution.

In addition, we find that the predicted thermal expansion from the MAB model agrees very well with MD [Figs. 6(c), 6(f) and 6(i)]. Both the anharmonic internal energy and thermal expansion at any temperature are properties derived from the very same bond density, highlighting the thermodynamic self-consistency of the model.

E. Computational efficiency

Similar to the QHA, a key feature of the MAB model is its ability to estimate thermodynamic properties analytically without the need for thermodynamic sampling. As demonstrated in the preceding sections, the MAB model consistently outperforms the QHA in terms of numerical accuracy for the considered potential models. We now proceed to compare the computational efficiency of the model. The comprehensive

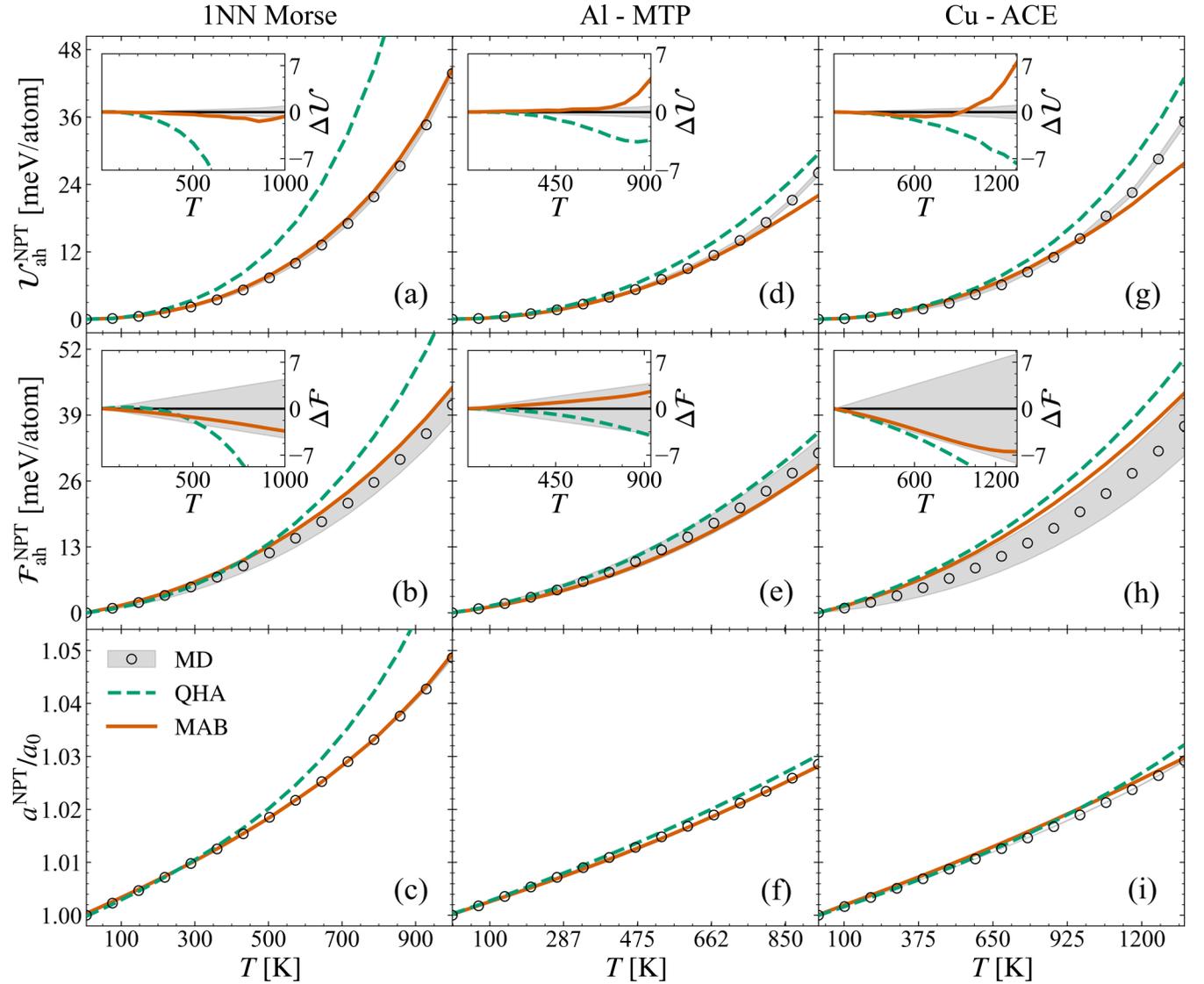


FIG. 6. Finite temperature per atom (a), (d), and (g) anharmonic internal energy, (b), (e), and (h) free energy, and (c), (f), and (i) lattice constant normalized by the $T = 0$ K lattice constant predicted by MD (black circles), QHA (dashed green line), and the refined MAB model (solid orange line) for the NPT ensemble. Each column corresponds to properties evaluated from a different potential model: (a)–(c) the first nearest neighbor Morse, (d)–(f) the Al-MTP, and (g)–(i) the Cu-ACE potential models. The difference between the MD-predicted and MAB-model-predicted properties are shown in the insets.

computational cost associated with evaluating the properties of an NPT ensemble using MD, QHA, and the MAB model is detailed in Table I for each of the potential models. Table I

also provides information on the number of explicit force calls and the speedup factors of the model with respect to MD and the QHA.

TABLE I. Total computational cost, with the number of explicit force calls of different approaches for the potential models studied, for the NPT ensemble. Speedup factors of the model (including the cost of LA bonding potential parametrization) over MD and QHA are also shown.

Method	Time (min/core)			No. force calls
	1NN Morse	Al-MTP	Cu-ACE	
MD	12.5	7248	2226	10^6
QHA	27.1	37.5	29.6	9+9
LA+MAB	0+0.09	1.93+0.43	1.75+0.37	(15+15)+9
Speedup factor				
MD/MAB	139	3071	1050	
QHA/MAB	300	16	14	

It is important to note that obtaining reliable MD reference data required a substantial number of force evaluations (up to a million time steps). The difference in the MD computation times between the first nearest neighbor Morse potential and the complex machine learning potentials represents the wide range of time per force evaluations of different interatomic potentials. In addition to force evaluations at the reference and displaced structures ($9 + 9$), the QHA costs include the cost of diagonalization of the dynamical matrix for each of the strained structures. Most of the cost for the MAB model comes from the LA parametrization of the bonding potential (15 displacements along $\hat{\mathbf{a}}_L$ and 15 more along $\hat{\mathbf{a}}_{T1}$), along with nine static calculations to obtain the energy-volume dependence. From the speedup factors in Table I, it is evident that the MAB model is significantly more efficient than MD and over an order of magnitude faster than the QHA. For the first nearest neighbor Morse potential, the cost of LA parametrization is zero, as it can be used directly as a bonding potential. It is also worth noting that for force evaluations involving DFT methods, the speedup factor between the model and the QHA may decrease, as the number of force calls will dominate the computational cost of the method. However, given that the MAB model offers superior numerical accuracy compared to the QHA, this could be considered a reasonable trade-off.

IV. CONCLUSION AND PERSPECTIVES

In the present work, we extended the original mean-field anharmonic bond model to allow the sampling-free evaluation of anharmonic thermodynamic properties of bulk fcc crystals in both NVT and NPT ensembles, irrespective of the potential model. The improved model is able to predict anharmonic free energies with an accuracy of a few meV/atom when compared to the MD-derived quantity from machine learning force fields, with a similar accuracy for the thermal expansion and pressure. The model shows better numerical accuracy than the QHA, being able to predict anharmonic properties for around an order of magnitude less computational cost.

While the first nearest neighbor approximation used by the MAB model is accurate for many-body potentials down to a few meV/atom, the systematic underestimation of the anharmonic internal energy above 65% of the melting temperature

motivates the inclusion of second nearest neighbors in future developments. In addition to making the model more accurate at high temperatures, this extension will also aid application to bcc and hcp crystal systems, which show a larger degree of anharmonicity, possibly due to a larger influence of the second nearest neighbors [11].

The MAB model is classical; i.e., it cannot directly account for anharmonicity due to nuclear quantum effects, which are prevalent at temperatures below the Debye temperature of the material. Such effects can be indirectly approximated using approaches like the temperature remapping approximation [30]. The application of the MAB model to different crystal systems, when coupled with such an approach, would allow the rapid prediction of quantum-accurate phase diagrams.

Another possible area of application of such a model would be to systems with defects such as vacancies, alloys, stacking faults, and free surfaces. Thermodynamic data from the model would be useful in the quick construction of defect phase diagrams [31]. However, it is important to note that the number of mean-field approximations required would scale with each additional inequivalent atom in the system. For systems with simple defects like monovacancies and dilute alloys, we expect efficiency gains similar to those observed for bulk systems.

ACKNOWLEDGMENTS

R.D., M.P., L.H., and J.N. gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Projects No. 405621160 and No. 409476157 (SFB1394). R.D. would also like to acknowledge the Max-Planck-Gesellschaft (MPG) and the Centre National de la Recherche Scientifique (CNRS) for their research support provided via the SALTO exchange program. T.D.S. gratefully acknowledges support from the ANR grant MeMoPas (Grant No. ANR-19-CE46-0006-1), IDRIS allocations A0090910965 and A0120913455, and Euratom Grant No. 633053. T.D.S. and J.N. gratefully acknowledge the Institute of Pure and Applied Mathematics at the University of California, Los Angeles, for providing a stimulating environment during the long program “New Mathematics for the Exascale: Applications to Materials Science.”

-
- [1] O. Hellman, I. A. Abrikosov, and S. I. Simak, *Phys. Rev. B* **84**, 180301(R) (2011).
- [2] O. Hellman and I. A. Abrikosov, *Phys. Rev. B* **88**, 144301 (2013).
- [3] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [4] B. Fultz, *Prog. Mater. Sci.* **55**, 247 (2010).
- [5] A. van de Walle and G. Ceder, *Rev. Mod. Phys.* **74**, 11 (2002).
- [6] B. Grabowski, L. Ismer, T. Hickel, and J. Neugebauer, *Phys. Rev. B* **79**, 134106 (2009).
- [7] A. Glensk, B. Grabowski, T. Hickel, and J. Neugebauer, *Phys. Rev. X* **4**, 011018 (2014).
- [8] A. Glensk, B. Grabowski, T. Hickel, and J. Neugebauer, *Phys. Rev. Lett.* **114**, 195901 (2015).
- [9] X. Zhang, B. Grabowski, F. Körmann, C. Freysoldt, and J. Neugebauer, *Phys. Rev. B* **95**, 165126 (2017).
- [10] B. Grabowski, Y. Ikeda, P. Srinivasan, F. Körmann, C. Freysoldt, A. I. Duff, A. Shapeev, and J. Neugebauer, *npj Comput. Mater.* **5**, 80 (2019).
- [11] P. Srinivasan, A. Shapeev, J. Neugebauer, F. Körmann, and B. Grabowski, *Phys. Rev. B* **107**, 014301 (2023).
- [12] D. Frenkel and A. J. C. Ladd, *J. Chem. Phys.* **81**, 3188 (1984).
- [13] M. de Koning, A. Antonelli, and S. Yip, *Phys. Rev. Lett.* **83**, 3973 (1999).
- [14] D. Frenkel and B. Smit, in *Understanding Molecular Simulation*, 2nd ed. (Academic, San Diego, 2002), pp. 241–267.

- [15] T. D. Swinburne, J. Janssen, M. Todorova, G. Simpson, P. Plechac, M. Luskin, and J. Neugebauer, *Phys. Rev. B* **102**, 100101(R) (2020).
- [16] Previous work labeled the computational scheme BLaSa: Bond lattice sampling and analytical approximation.
- [17] Y. Lysogorskiy, C. v. d. Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, and R. Drautz, *npj Comput. Mater.* **7**, 97 (2021).
- [18] J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, and J. Neugebauer, *Comput. Mater. Sci.* **163**, 24 (2019).
- [19] PYIRON, <https://github.com/pyiron>.
- [20] R. Drautz, *Phys. Rev. B* **99**, 014104 (2019).
- [21] A. V. Shapeev, *Multiscale Model. Simul.* **14**, 1153 (2016).
- [22] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, *Mach. Learn.: Sci. Technol.* **2**, 025002 (2020).
- [23] M. Poul, L. Huber, E. Bitzek, and J. Neugebauer, *Phys. Rev. B* **107**, 104103 (2023).
- [24] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *Comput. Phys. Commun.* **271**, 108171 (2022).
- [25] A. Togo, L. Chaput, T. Tadano, and I. Tanaka, *J. Phys.: Condens. Matter* **35**, 353001 (2023).
- [26] J. J. Moré, The Levenberg-Marquardt algorithm: Implementation and theory, in *Numerical Analysis*, edited by G. A. Watson (Springer, Berlin, Heidelberg, 1978), pp. 105–116.
- [27] P. Virtanen *et al.*, *Nat. Methods* **17**, 261 (2020).
- [28] S.-Y. Yue, X. Zhang, G. Qin, S. R. Phillpot, and M. Hu, *Phys. Rev. B* **95**, 195203 (2017).
- [29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1986).
- [30] R. Dsouza, L. Huber, B. Grabowski, and J. Neugebauer, *Phys. Rev. B* **105**, 184111 (2022).
- [31] S. Korte-Kerzel, T. Hickel, L. Huber, D. Raabe, S. Sandlöbes-Haut, M. Todorova, and J. Neugebauer, *Int. Mater. Rev.* **67**, 89 (2022).