Self-consistency error correction for accurate machine learning potentials from variational Monte Carlo

Giacomo Tenti,[*,†] Kousuke Nakano,[‡,¶] and Michele Casula[§]

†*International School for Advanced Studies (SISSA), Via Bonomea 265, 34136 Trieste, Italy*

‡*Center for Basic Research on Materials, National Institute for Materials Science (NIMS), 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*

¶*Center for Emergent Matter Science (CEMS), RIKEN, 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan.*

§*Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie (IMPMC), Sorbonne Université, CNRS UMR 7590, MNHN, 4 Place Jussieu, 75252 Paris, France*

* E-mail: gtenti@sissa.it

Abstract

Variational Monte Carlo (VMC) can be used to train accurate machine learning interatomic potentials (MLIPs), enabling molecular dynamics (MD) simulations of complex materials on time scales and for system sizes previously unattainable. VMC training sets are often based on partially optimized wave functions (WFs) to circumvent expensive energy optimizations of the whole set of WF parameters. However, frozen variational parameters lead to VMC forces and pressures not consistent with the underlying potential energy surface, a bias called the self-consistency error (SCE). Here, we demonstrate how the SCE can spoil the accuracy of MLIPs trained on these data, taking high-pressure hydrogen as test case. We then apply a recently introduced SCE correction [*Phys. Rev. B* 109, 205151 (2024)] to generate unbiased

VMC training sets based on a Jastrow-correlated single determinant WF with frozen Kohn-Sham orbitals. The MLIPs generated within this framework are significantly improved and can approach in quality those trained on datasets built with fully optimized WFs. Our conclusions are further supported by MD simulations, which show how MLIPs trained on SCE-corrected datasets systematically yield more reliable physical observables. Our framework opens the possibility of constructing extended high-quality training sets with VMC.

1 Introduction

Since their introduction,[1,2] machine learning interatomic potentials (MLIPs) have revolutionized the field of molecular dynamics (MD).[3–12] These models can learn the potential energy surface (PES) of a given system starting from a representative set of configurations, called the training set. Each element of this set is labeled by its physical properties (e.g., energy, ionic forces, pressure, etc.), usually computed using *ab initio* electronic structure methods. After it is trained, an MLIP is capable of reproducing the observables of the system on new configurations with the accuracy of the original method but with a computational cost of several orders of magnitude smaller.[2] This has opened the possibility of running MD simulations at *ab initio* accuracy with sizes and lengths that were previously out of reach.[13] By far, the most common method used for generating MLIPs datasets is density functional theory (DFT),[14] which can easily produce $\sim 10^4$-10^5 training points with modest computational resources, usually sufficient to train data-heavy machine learning frameworks.[7,15] For several systems, however, DFT fails to capture the strong effect of electronic correlations or may provide results that strongly depend on the choice of the exchange-correlation functional. Quantum chemistry methods, such as the density matrix renormalization group (DMRG),[16] and the complete active space self-consistent field (CASSCF),[17,18] provide a much better description of correlation, but their applicability to periodic extended systems is limited by their large computational cost. Because of this, even if significant progress has been made recently, also thanks to the implementation of new machine learning schemes,[19] the range of applicability of these approaches remains limited to a few tens of atoms and a few *ps*-long trajectories.

Quantum Monte Carlo (QMC) methods are a series of stochastic algorithms that provide state-of-the-art results for a large class of systems, from molecules to solids,[20,21] while being less computationally demanding. Despite this, QMC approaches still have a cost about 10-100 times larger than DFT. In the context of MLIPs, this implies that the number of training configurations that can be generated within a given amount of resources is generally smaller than the one attainable with mean-field methods. Recently proposed solutions to this problem make use of data-efficient schemes in combination with hierarchical machine learning (or Δ-learning[22]) to construct QMC-based MLIPs.[23,24] However, within the most popular QMC approach, i.e., variational Monte Carlo (VMC), another issue prevents a straightforward application of MLIPs, presented in what follows.

The efficiency of VMC is largely due to the complexity of the wave function (WF) employed in the calculations to describe the electronic part. A possible choice for the WF Ψ is

$$\Psi = \exp(J) \times \Phi_{\mathrm{AS}},$$

where $\exp(J)$ is a bosonic term (i.e., symmetric under particle exchange) called the Jastrow factor, and Φ_{AS} is an antisymmetric (AS) part. Both terms contain variational parameters, whose number can lead to flexible WFs. However, there is a practical difficulty in optimizing a large number of variables in a stochastic setting, particularly those belonging to the AS part. Thus, only the Jastrow parameters are usually optimized, and $\Phi_{\mathrm{AS}} \equiv \Phi_{\mathrm{SD}}$ is often taken as a single Slater determinant (SD) obtained from a DFT calculation.[21] This gives rise to the frozen-orbital Jastrow Slater determinant (JSD) wave function, an appealing candidate for VMC-based machine learning applications, since it provides an excellent tradeoff between accuracy and computational cost. Indeed, the Jastrow optimization is much less demanding than the full optimization also involving the determinant. Nevertheless, the use of the frozen Slater determinant introduces a bias in the VMC forces and pressures computed with this WF. This bias has been dubbed as the self-consistency error ,[25,26] and we are going to show that it is particularly relevant in the context of VMC-based MLIPs. Even though machine learning potentials are, by definition, consistent, a biased training set can

nonetheless spoil the final accuracy of the model. A solution would be to use only the energies for the training; this, however, usually increases by orders of magnitude the number of configurations necessary to reach a given accuracy[27,28].

In Ref. 26, a suitable correction was developed to remove the self-consistency error, adding a relatively small computational overhead to the calculation. The aim of this work is to show the actual impact of the self-consistency error on an MLIP in a working case scenario, and to demonstrate how its correction can be applied to generate unbiased VMC datasets with a JSD WF for machine-learning applications. As a test case, we consider high-pressure hydrogen, a widely studied system with applications ranging from planet modeling[29] to inertial confinement fusion,[30–32] and often used to benchmark numerical methods.[33] It is also a system where DFT usually fails, when compared with the experimental data,[34] making it an ideal testbed for our purposes. In particular, we focus on density and temperatures relevant for the Hugoniot curve, *i.e.*, the set of possible states reachable with a shock wave.[35] The deuterium Hugoniot was recently computed in Ref. 24 using MLIPs trained on accurate QMC data. There, the self-consistency error was avoided by relaxing the full set of variational parameters in the WF, *i.e.*, including both the Jastrow factor and the antisymmetric part. The resulting optimization improves the consistency of the forces and pressures and also changes the underlying PES, at a computational cost about 10 times larger than the standard frozen-orbital JSD optimization.

Here, we explore a different path, by showing that a very similar accuracy can be obtained by training models on a frozen-orbital JSD dataset supplemented by the self-consistency error correction. In particular, we compute VMC forces and pressures on the same configurations using a JSD WF, with and without corrections, and use them to train the different models. The performances of these MLIPs are compared with the one of an MLIP trained on the fully optimized dataset of Ref. 24, which we take as the reference. In this way, we show the relevance of the self-consistency error in the MLIP generation and the need for its correction, by studying the impact of both the correction and the full optimization of the WF on thermodynamic equilibrium quantities such as pressure and radial distribution functions. This work demonstrates the importance of obtaining unbiased

QMC datasets for MLIPs at the VMC level using cheap-to-optimize JSD-type WFs, including also ionic forces and stress components.

2 Methods

We briefly present the main methods used in this work. In Sec. 2.1 we introduce general aspects of MLIPs, focusing on one specific framework, *i.e.,* the kernel ridge regression, and on the loss function used for model training. In Sec.2.2, we review the forces and pressure correction developed in Ref. 26 to solve the self-consistency error in VMC and describe its implementation in our WF representation.

2.1 Generalities of MLIPs

The basic goal of any MLIPs is to predict the energy of a given N-atom configuration described by its ionic positions, i.e., $E = E(\mathbf{R}_1, \cdots, \mathbf{R}_N)$. We now suppose that E can be expressed as a sum of atomic contributions,[2] each depending on the relative coordinates of all the other atoms with respect to the i-th one:

$$E = \sum_{i=1}^{N} e(\mathscr{R}_i) \qquad \text{where} \quad \mathscr{R}_i = (\mathbf{R}_{i1}, \ldots, \mathbf{R}_{iN}) \quad \text{with} \quad \mathbf{R}_{ij} \equiv \mathbf{R}_j - \mathbf{R}_i. \tag{1}$$

Here, we refer to $\mathscr{R}_i$ as the local environment around the i-th atom and to $e(\mathscr{R}_i)$ as the local atomic energy, and suppose the minimum image convention holds when periodic boundary conditions are applied. Often, the function $e(\mathscr{R}_i)$ in Eq.(1) is considered local, i.e., only dependent on the atoms closer than a certain cutoff radius r_c. Different types of MLIPs may be distinguished by their internal representation of the local atomic energy. Many techniques have been successfully applied over the years, including deep neural networks,[2,4,11] Gaussian approximation potentials,[36,37] kernel ridge regression,[23,38] graph neural networks,[39,40] and more. Independently of the specific

architecture, $e(\mathscr{R}_i)$ usually depends on a set of model parameters $\mathbf{p}$ which have to be optimized:

$$e(\mathscr{R}_i) \equiv e(\mathscr{R}_i;\mathbf{p}). \tag{2}$$

Most MLIPs implement a functional form that already satisfies some of the symmetries required by physical constraints, such as permutational and rotational invariance. To determine the model parameters $\mathbf{p}$, we consider a set of N_{train} configurations, *i.e.*, the training set, for which we compute the desired physical quantities using a target numerical method. In particular, we indicate with $\mathscr{R}_i^\mu \equiv (\mathbf{R}_1^\mu - \mathbf{R}_i^\mu, \ldots, \mathbf{R}_{N_\mu}^\mu - \mathbf{R}_i^\mu)$ the i-th atomic environment belonging to the μ-th configuration of the set, with $\mu = 1, \ldots, N_{\text{train}}$, having N_μ atoms and volume V_μ (for simplicity we will consider the supercell to be cubic). Moreover let E_{ref}^μ, $\mathbf{F}_{j,\text{ref}}^\mu$ and P_{ref}^μ be the values of the total energy, the force acting on the j-th atom, and the (isotropic) virial pressure of the configuration. The model parameters $\mathbf{p}$ are then optimized by minimizing a loss function such as

$$\begin{aligned}
\mathscr{L}(\mathbf{p}) = {} & W_E \frac{1}{N_{\text{train}}} \sum_{\mu=1}^{N_{\text{train}}} \left[\frac{1}{N_\mu} \sum_{i=1}^{N_\mu} e\left(\mathscr{R}_i^\mu;\mathbf{p}\right) - \frac{1}{N_\mu} E_{\text{ref}}^\mu \right]^2 \\
& + W_F \frac{1}{N_{\text{train}}} \sum_{\mu=1}^{N_{\text{train}}} \frac{1}{3N_\mu} \sum_{j=1}^{N_\mu} \sum_{\alpha=x,y,z} \left[-\frac{\partial}{\partial R_{j,\alpha}^\mu} \sum_{i=1}^{N_\mu} e\left(\mathscr{R}_i^\mu;\mathbf{p}\right) - F_{j,\alpha,\text{ref}}^\mu \right]^2 \\
& + W_P \frac{1}{N_{\text{train}}} \sum_{\mu=1}^{N_{\text{train}}} \left[-\frac{\partial}{\partial V_\mu} \sum_{i=1}^{N_\mu} e\left(\mathscr{R}_i^\mu;\mathbf{p}\right) - P_{\text{ref}}^\mu \right]^2 .
\end{aligned} \tag{3}$$

The three terms in Eq.(3) are the mean squared error (MSE) of the energy per atom, ionic forces, and virial isotropic pressure, respectively, and W_E, W_F, and W_P are tunable weights multiplying the different MSEs. Other functional forms of the loss function are possible, including additional physical quantities (e.g., charges, all 6 independent components of the stress, etc). In neural networks, $\mathscr{L}(\mathbf{p})$ can be minimized using stochastic optimization methods that compute the gradients using only a small batch of training configurations. In this paper, we primarily use MLIPs based on the kernel ridge regression (KRR) method. Within KRR, the local atomic energy

is expressed as

$$e(\mathscr{R};\mathbf{p}) = \sum_{\nu=1}^{N_{\text{env}}} p_\nu K(\mathscr{R},\mathscr{R}_\nu) \tag{4}$$

where $\mathscr{R}_\nu$, $\nu = 1,\ldots,N_{\text{env}}$ belong to a subset of all the local environments in the training set, and $K(\mathscr{R},\mathscr{R}_\nu)$ is the normalized kernel between $\mathscr{R}$ and $\mathscr{R}_\nu$. Following Refs 23 and 24, we adopt a kernel based on a variant of the Smooth Overlap of Atomic Positions (SOAP) descriptor[41,42](see App. A.1), and we select the N_{env} local environments using the farthest point sampling method,[23] according to the distance introduced by the kernel K. Notice how plugging Eq.(4) into the loss function (3) and taking the derivatives with respect to the variational coefficients turns the minimization problem into a linear system, which is then solved by the conjugate gradient method. We conclude this Section by stressing the fact that, by definition, any MLIPs will give energy derivatives (e.g., forces and virial stress components) that are consistent with their underlying potential energy surface. Any effect of potential inconsistency in the dataset, such as the ones studied in this work, will manifest in the final performance of the models and its dependence on the relative training weights W_E, W_F, and W_P. This is demonstrated in Sec.3.2, where we compare the quality of MLIPs trained on different datasets as a function of W_E, W_F, and W_P.

2.2 Self-consistency error in VMC and its correction

Within VMC, the force acting on the i-th atom is computed using

$$\mathbf{F}_i^{\text{VMC}} = -\left\langle \frac{\partial}{\partial \mathbf{R}_i} E_L \right\rangle - 2\left\langle (E_L - E)\frac{\partial}{\partial \mathbf{R}_i}\log\Psi \right\rangle, \tag{5}$$

where $\langle\cdots\rangle$ indicates the quantum expectation value on the wave function Ψ evaluated stochastically, $E_L \equiv \hat{H}\Psi/\Psi$ the local energy with $\hat{H}$ being the Hamiltonian of the system and $E = \langle E_L\rangle$. To obtain an efficient and statistically meaningful value of the force in Eq. (5) with finite variance, techniques such as the zero-variance zero-bias principle,[43] the space warp transformation[44] and reweighting[43,45–49] are applied. $\mathbf{F}_i^{\text{VMC}}$ is unbiased only when all the variational parameters of the WF are at their variational minimum. As mentioned earlier, this assumption is not true for partially

optimized WFs, such as the frozen-orbital JSD, one of the most popular choices. In this case, an extra term

$$\mathbf{F}_i^{\mathrm{corr}} = -\sum_{k=1}^{N_{SD}} \frac{\partial E}{\partial \alpha_k^{\mathrm{SD}}} \frac{d\alpha_k^{\mathrm{SD}}}{d\mathbf{R}_i} \tag{6}$$

has to be considered, containing the derivatives of the N_{SD} variational coefficients α_k^{SD} included in the Slater determinant.

Discarding the term in Eq.(6) causes an inconsistency between the forces and the underlying PES computed with a JSD WF, meaning that $\mathbf{F}_i^{\mathrm{VMC}} \neq -\frac{dE}{d\mathbf{R}_i}$. This is called the self-consistency error (SCE).[25,26] An analogous discussion applies to other energy derivatives, such as the virial pressure, that contain a similar bias and for which we can derive similar corrections.

In our applications, following the implementation of the TURBORVB package[50] (here used for all the QMC calculations), we consider a representation of the SD part in terms of the antisymmetrized geminal power (AGP)

$$\Phi_{\mathrm{AGP}} = \mathscr{A}\left(g(x_1,x_2)\cdots g(x_{N-1},x_N)\right), \tag{7}$$

where $x_i = (\mathbf{r}_i, \sigma_i)$ indicates the collective spatial and spin coordinates of the i-th electron, with $i = 1,\ldots,N$, $\mathscr{A}$ is the antisymmetrizer operator and g is a pairing function. In Eq. (7) we also suppose N to be even for simplicity. In the spin unpolarized case, the pairing function can be expressed as the product of a spatial part f and a spin singlet, namely

$$g(x_i,x_j) = f(\mathbf{r}_i,\mathbf{r}_j)\frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}}. \tag{8}$$

In TURBORVB the spatial part of the pairing function is written in terms of a localized basis of L atomic orbitals (AOs) ϕ_k

$$f(\mathbf{r}_i,\mathbf{r}_j) = \sum_{k=1}^{L}\sum_{k'=1}^{L} \lambda_{k,k'}\phi_k^{\theta}(\mathbf{r}_i)\phi_{k'}^{\theta'}(\mathbf{r}_j). \tag{9}$$

In our case, the AOs are taken as Gaussian type orbitals (GTOs). For periodic systems, such as the ones considered in this work, the localized basis is appropriately generalized to satisfy periodic or

twisted boundary conditions,[51] depending on the phase θ (θ') for spin-up (spin-down) electrons. Here, $\theta = -\theta'$, such that the global phase factor is canceled, making the function f invariant under a global translation. Fixing this gauge leads to a numerically stable evaluation of the $d\lambda_{k,k'}/d\mathbf{R}$ derivatives implied by Eq. 6.

An alternative expression for the function f can also be obtained using molecular orbitals (MOs) $\Phi_n(\mathbf{r}) = \sum_k c_{n,k}\phi_k(\mathbf{r})$. In particular, when the number M of MOs is equal to $N/2$, the resulting AGP is exactly equivalent to a single SD, *i.e.,* $\Phi_{\mathrm{AGP}} \equiv \Phi_{\mathrm{SD}}$. Within TURBORVB, the MOs (and the corresponding $\lambda_{k,k'}$ matrix) are initialized by a local density approximation[52] (LDA) DFT calculation. This is performed in a basis consisting of the aforementioned AOs with an additional one-body Jastrow factor to automatically take into account Kato cusp conditions.[50,53] Because of this, the electronic integrals entering into the Kohn-Sham equations are evaluated on a real space grid with a suitably chosen lattice space. Finally, we also mention that the basis is regularized by removing any eigenvector of the overlap matrix $S_{k,k'} = \langle\phi_k|\phi_{k'}\rangle$ with a corresponding eigenvalue (s_i) satisfying the condition $s_i/s_{\max} < \varepsilon_{\mathrm{cut}}$,[54] where $s_{\max}$ is the maximum eigenvalue and $\varepsilon_{\mathrm{cut}}$ is set to 10^{-7} in this study. This regularization is very general and is also able to automatically cure the problem of basis-set redundancy coming from delocalized periodic Gaussians.

Finally, the expression in Eq. (6) is stochastically evaluated in VMC according to

$$\mathbf{F}_i^{\mathrm{corr}} = -2\,\mathrm{Re}\left\{\left\langle E_L \sum_{k=1}^{L}\sum_{k'=1}^{L}\left[\left(\mathscr{O}_{k,k'} - \overline{\mathscr{O}}_{k,k'}\right)\frac{d\lambda_{k,k'}}{d\mathbf{R}_i}\right]\right\rangle\right\}, \tag{10}$$

where $\mathscr{O}_{k,k'} = \frac{\partial \log\Psi}{\partial\lambda_{k,k'}}$ and $\overline{\mathscr{O}}_{k,k'} = \langle\mathscr{O}_{k,k'}\rangle$. In the previous expression, the logarithmic derivatives in $\mathscr{O}_{k,k'}$ are efficiently computed using the adjoint algorithmic differentiation,[47] while the derivatives of the parameters, $\frac{d\lambda_{k,k'}}{d\mathbf{R}_i}$, are evaluated within DFT-LDA using the finite differences method (FDM), as described in Ref. 26. For the virial isotropic pressure, an analogous derivation yields the following expression for the correction term:

$$P^{\mathrm{corr}} = -2\,\mathrm{Re}\left\{\left\langle E_L \sum_{k=1}^{L}\sum_{k'=1}^{L}\left[\left(\mathscr{O}_{k,k'} - \overline{\mathscr{O}}_{k,k'}\right)\frac{d\lambda_{k,k'}}{dV}\right]\right\rangle\right\}. \tag{11}$$

The above Equation can be easily extended to the non-isotropic case.

3 Results

Henceforth, we analyze how the SCE affects the training of MLIPs and discuss how the correction described in Sec. 2.2 can effectively be implemented to improve the accuracy of the resulting models. With this aim, we consider a dataset made of pristine hydrogen configurations for which we computed VMC energy, forces, and pressure using different types of WF and with/without bias correction. The dataset is the same as the one used in Ref. 24 and comprises 558 configurations of 128 atoms each. These configurations were extracted from DFT-driven MD simulations at temperatures between 4000 K and 20000 K and Wigner-Seitz radii r_s between 1.80 and 2.12. This range of thermodynamic conditions corresponds to the expected location of the Hugoniot curve of deuterium, defined as the set of states that can be reached by producing a shock wave in a sample. For these configurations, we fully optimized the wave function, going beyond the frozen-orbital JSD ansatz. As previously mentioned, we take this dataset as the reference one. In Sec. 3.1, we study the SCE for the same set of configurations as the reference dataset, and show the effect of both forces and pressure corrections. In Sec. 3.2 we directly compare the accuracy of MLIPs trained on a biased dataset (i.e. affected by the SCE) and a corrected one, respectively. We also compare both models with the reference dataset used in Ref. 24. Finally, in Sec. 3.3, we analyze the results of molecular dynamics simulations driven by the different MLIPs.

3.1 Correcting the SCE bias in the Hugoniot dataset

To assess the magnitude of the SCE in the Hugoniot dataset, we use a JSD WF with a basis set of [4s2p1d] and [2s2p1d] GTOs for the antisymmetric part and Jastrow factor, respectively. The same primitive basis set was used in Ref. 24. We initialize the SD by running an LDA-DFT calculation with twisted boundary conditions at the Baldereschi point, i.e, at $\mathbf{k} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ in crystal units. The Jastrow part is optimized using the stochastic reconfiguration method.[55]

We assess the consistency of the JSD forces and pressures by calculating the forces and pressure on a few selected configurations with two separate approaches: (i) evaluating Eq.(5) (along with the analogous expression for pressure) and (ii) using the FDM by fitting the PES of the system and evaluating the derivative numerically. As previously discussed, the SCE manifests itself as a difference between these two values. The results are shown in Fig. 1 for one of the force components and the virial pressure.

We observe a significant bias for both quantities, with a discrepancy as large as $\sim 10\%$ of their values. This clearly highlights the importance of curing the SCE for the biased dataset. For this reason, we applied the corrections of Eqs. (10) and (11) to forces and pressures, respectively. For forces, we computed the numerical derivatives of the parameters $\frac{d\lambda_{k,k'}}{d\mathbf{R}}$ with the FDM from DFT-LDA, the theory used to obtain $\lambda_{k,k'}$, with a displacement $\Delta R = \pm 0.003$ Å. The derivatives of $\lambda_{k,k'}$ with respect to the volume, entering in the pressure correction, were obtained with the FDM using relative volume variations of $\pm 0.03\%$. In Fig. 1 we observe that the corrected forces and pressure are perfectly compatible with their values directly computed through the fit of the VMC PES. Fig. 2 shows the values of both the biased and corrected force components (pressures) versus the forces (pressures) estimated from the PES, for several configurations. Notice how the pressure correction acts almost like a rigid shift of ~ 1 GPa. The root mean squared error (RMSE) of biased and corrected quantities is also shown, further demonstrating the effectiveness of the correction in removing the SCE.

Sometimes, the added correction significantly increases the error bar σ_F of the corrected forces. We then disregard the corrections that lead to forces with $3\sigma_F > 0.015$ Ha/Bohr. For pressures, the threshold value of 1.5×10^{-5} is applied to $3\sigma_p$. See App. A.2 for further details. In principle, the FDM used to estimate the parameter derivatives $\frac{d\lambda_{k,k'}}{d\mathbf{R}}$ can be replaced with more robust and faster approaches (*e.g.*, using linear response theory[56]). We leave this possibility for future work.

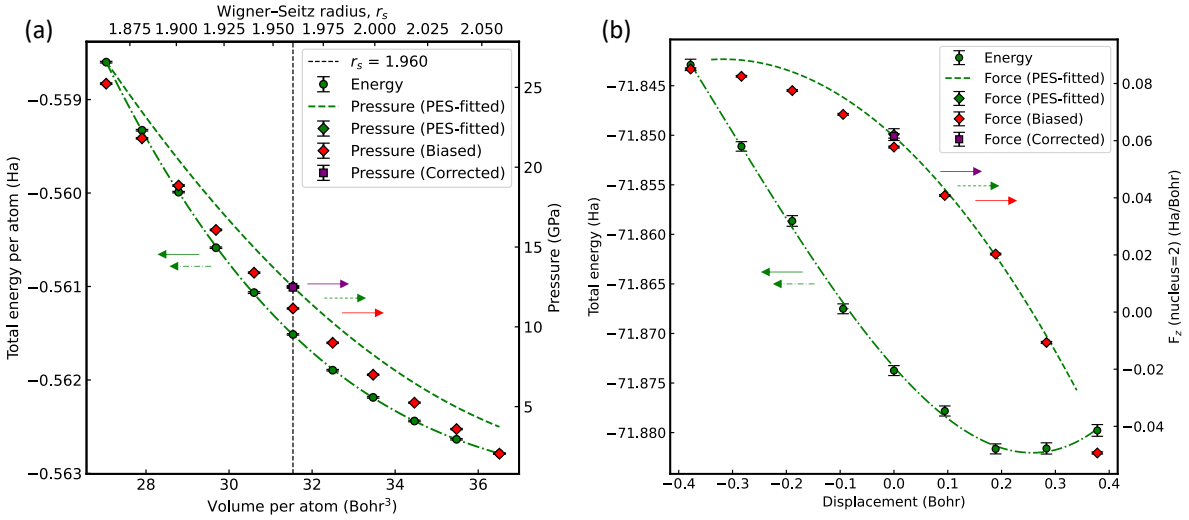

Figure 1: (a) Comparison among the biased pressure evaluated with an expression equivalent to Eq. (5) (red diamonds), the corrected pressure obtained by applying Eq. (11) (violet square), and the pressure obtained by fitting the PES (dashed green line and green diamond) for a selected configuration in the Hugoniot dataset.[24] The PES of the system is also shown (green dots and dash-dotted line). (b) Comparison among the biased force evaluated with Eq. (5) (red diamonds), the corrected force obtained by applying Eq. (10) (violet square), and the force calculated by fitting the PES (dashed green line and green diamond) for the z-component of the second hydrogen atom belonging to the same configuration as in (a). The PES of the system along the atomic displacement is also shown (green dots and dash-dotted line).

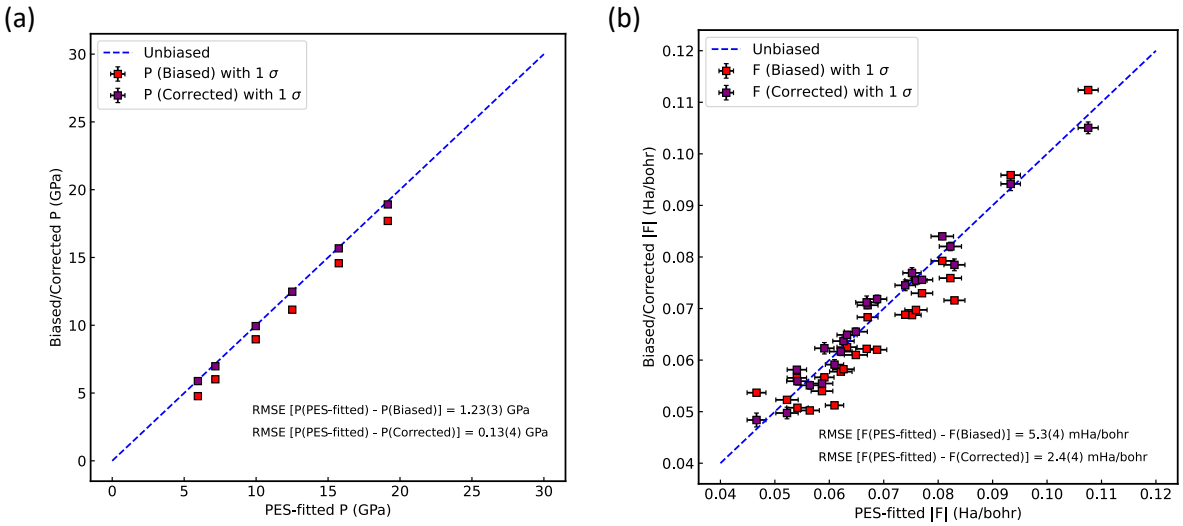

Figure 2: (a) Value of the biased (red markers) and corrected (violet markers) pressure for 6 different 128-atom configurations as a function of the numerical pressure estimated from the derivative of the PES with respect to the volume. (b) Values of the biased (red markers) and corrected (violet markers) force components as a function of the numerical force estimated by the PES for one of the configurations belonging to the Hugoniot dataset.[24] The dashed line is a reference for perfect consistency.

3.2 Comparison between VMC datasets

To illustrate how using an SCE-corrected JSD dataset affects the MLIPs generation, we compare the performances of MLIPs trained on three different datasets: (i) the aforementioned "biased dataset" obtained with a JSD WF and with forces and pressure computed with the standard expressions (Eq. 5), (ii) a "corrected dataset" where the corrections of Eqs. (10) and (11) are applied as described in Sec. 2.2 and (iii) a "reference dataset", *i.e.*, the one introduced in Ref. 24. In particular, the latter was obtained from an improved Jastrow-correlated AGP WF, where the Jastrow and antisymmetric part were both optimized, thanks to a combination of geminal embedded orbitals[57] to reduce the number of variational parameters and a restricted optimization based on the locality of the AGP $\lambda_{k,k'}$ matrix. The $\lambda_{k,k'}$ coefficients were optimized for those basis set elements k and k' centered on atoms closer than a cutoff distance of 4 Bohr radii.[24] No significant lowering of the variational energy was found beyond this threshold, suggesting that the dependence of the energy on the remaining parameters can be disregarded. Thanks to this approach, the SCE on this

dataset was effectively removed and we thus decided to take the models trained on it as the reference. Notice how the explicit optimization of the whole WF, as the one performed in Ref.,[24] is a procedure that is about an order of magnitude more expensive than the partial optimization of the frozen-orbital JSD WF, and requires advanced minimization techniques.[58,59]

For all the MLIPs, a KRR model is trained on the difference between VMC quantities and DFT ones, the latter using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional,[60] following the Δ-learning approach.[22–24] In this way, a much higher accuracy is achieved with relatively small datasets such as the ones used here. The values of the hyperparameters chosen for these models are reported in App. A.1. Among the 558 configurations of the datasets, we select 58 configurations for testing. We investigate the effect of varying the training weights in the loss function (Eq. (3)) on the RMSE computed for energy, forces, and pressure of the test set for each dataset.

Since our goal here is to compare MLIPs trained on different datasets (*i.e.*, different training values for energy E, forces $\mathbf{F}$ and pressure P), the RMSE may not be the best metric to assess the relative accuracy of one model with respect to the others. For this reason, we define the relative improvement of the model for each system property ($X = E, \mathbf{F}, P$) as:

$$\Delta_X = \frac{\mathrm{RMSE}(X_{\mathrm{test}} - X_{\mathrm{dummy}}) - \mathrm{RMSE}(X_{\mathrm{test}} - X_{\mathrm{pred}})}{\mathrm{RMSE}(X_{\mathrm{test}} - X_{\mathrm{dummy}})}, \tag{12}$$

where X_{pred} are the model predictions, X_{test} the test set values, $E_{\mathrm{dummy}} = \frac{1}{N_{test}} \sum_{i \in test} E_i$, $\mathbf{F}_{\mathrm{dummy}} = 0$, and $P_{\mathrm{dummy}} = 0$. In other words, Δ_X measures the relative improvement of the model prediction on the quantity X with respect to a "dummy" model, corresponding to a perfectly flat PES equal to the average energy on the test set for all configurations. Notice how $\Delta_X = 1$ when $\mathrm{RMSE}(X_{\mathrm{test}} - X_{\mathrm{pred}}) = 0$, indicating that the model has learned exactly the quantity X, while a value $\Delta_X \leq 0$ means that the model has an error equal to, or larger than, a flat model. In the present work, the test quantities refer to the difference between the VMC and the DFT-PBE baseline, while E_{dummy} is the average difference on the test set.

Figs. 3 and 4 show the results obtained for Δ_E, Δ_F, and Δ_P using MLIPs trained with different

values of the training weights W_E, W_F, and W_P. We first analyze what happens by progressively increasing the weight on the forces for the different models without explicitly training on pressures (*i.e.*, $W_P = 0$). The force bias in the biased dataset is evident from the lower Δ_F of the corresponding model compared to that of both the corrected and reference datasets (see Fig. 3). Notice how, for small values of W_F/W_E, Δ_F is negative in the biased case, which in the Δ-learning framework means that the correction on forces is detrimental. As the weight on forces is increased, all models reach a plateau in Δ_F, while simultaneously losing some accuracy on the other quantities. For the energy, this loss of accuracy is significantly larger in the models trained on the biased dataset, further demonstrating the effect of the SCE.

We observe very similar performances between the corrected and reference models. For pressure, the models including the SCE correction show the best performances for almost all the values of W_F. Since the loss does not include the pressure in this case, this is another indication of consistency. In Fig. 4 we plot the results obtained by varying the weight on pressure, for two values of W_F/W_E, namely $W_F/W_E = 3/128$ and $W_F/W_E = 15/128$. Also in this case, a better accuracy is generally achieved for the models trained on the corrected and reference datasets.

Finally, we mention how these conclusions exclusively depend on the quality of the datasets themself, *i.e.*, their internal consistency, and not on the MLIPs specific architecture. To demonstrate this feature, in App. A.3, we discuss the relative improvement Δ_X for models obtained with the MACE[40] framework, which show the same behavior as the one found for the KRR MLIPs.

3.3 MD results comparison

To conclude our analysis, we run a set of MD simulations using MLIPs trained on the three different datasets. For each model, we select the relative training weights W_F/W_E and W_P/W_E to have a good tradeoff between the RMSE values on energy, forces, and pressure. In particular, for the MLIPs trained on the reference and corrected weight, we choose the last value of W_F/W_E before the significant decrease in Δ_E (*i.e.*, the fourth smallest value in Fig. 3), in order to increase as much as possible the accuracy on forces without spoiling too much the energy. A similar reasoning is

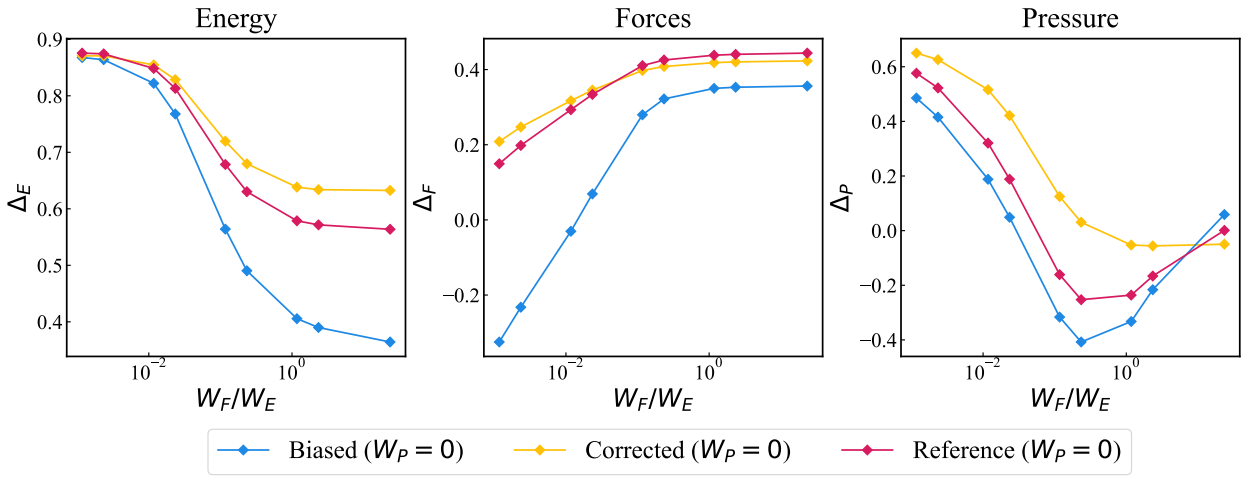

Figure 3: Relative RMSE variation Δ_X (Eq. 12) for energy, forces, and pressure, as a function of the force/energy weight ratio in the loss function of Eq. 3 (expressed in atomic units) for MLIPs trained on the biased, corrected and reference datasets, respectively. The weight on the pressure is set to zero for all models.

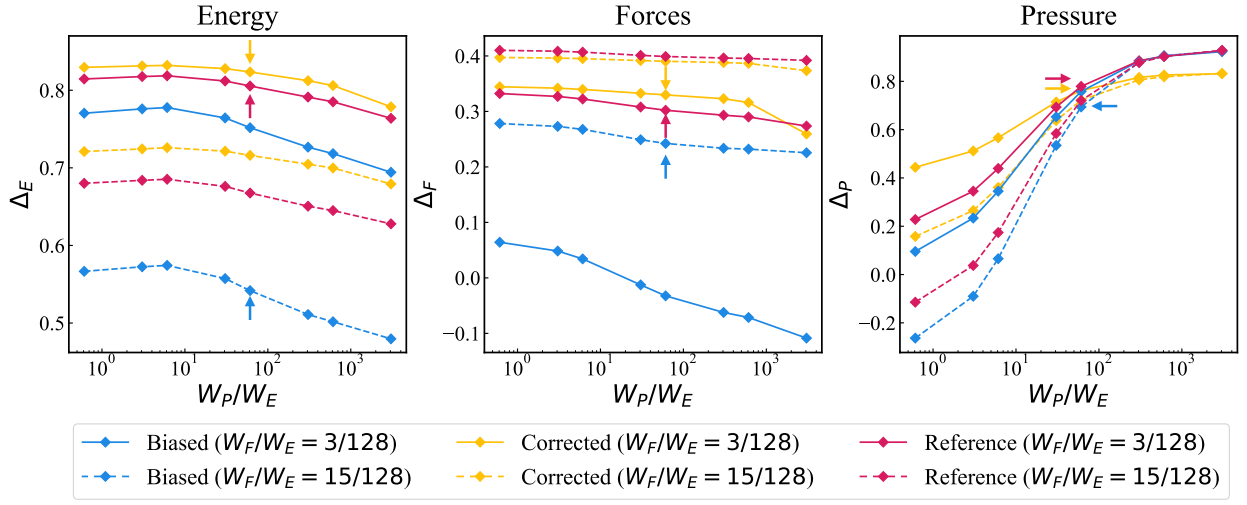

Figure 4: Δ_X (Eq. 12) for energy, forces, and pressure, as a function of the pressure/energy weight ratio in the loss function of Eq. 3 used for training (expressed in atomic units), for the different models. Solid lines correspond to $W_E = 1$ and $W_F = 3/128$, while dashed lines correspond to $W_E = 1$ and $W_F = 15/128$. The arrows highlight the "best" MLIPs for each of the three datasets, which are used for the MD simulations described in Sec. 3.3.

used to determine the optimal value of W_P/W_E (see Fig. 4). For the MLIP trained on the biased dataset, we choose a larger value of W_F/W_E, which is necessary to reach a reasonable accuracy on forces. This comes at the price of sacrificing the accuracy on the energy, a direct consequence of the SCE. The weights used for each MLIP are summarized in Tab. 1. While the values of the best relative training weights W_F/W_E and W_P/W_E are model dependent, we highlight that this procedure is generally applicable to any other system.

Table 1: Training weights for the MLIPs used in the MD simulations, for each dataset. For both the forces and pressure weights, we report the relative value with respect to W_E. All values are in atomic units.

Dataset	W_F/W_E	W_P/W_E
Biased	$15/128$	$10^6/128^2$
Corrected	$3/128$	$10^6/128^2$
Reference	$3/128$	$10^6/128^2$

We run MD simulations at three different densities and temperature conditions close to the position of the deuterium Hugoniot curve given by the VMC-MLIP of Ref. 24. In particular, we considered thermodynamic conditions that span both the molecular liquid and the atomic one. Because of the Δ-learning approach adopted, the resulting efficiency of these calculations is the same as a DFT Born-Oppeheimer MD simulation. Indeed, the energy, ionic forces, and pressure of the system are calculated at each step with DFT, and the corresponding corrections are then added using the MLIP. Notice how the DFT baseline can be, in principle, replaced by a faster potential (*e.g.*, using an MLIP trained on DFT data[61]), improving the overall efficiency. Further details regarding the computational setup of the MD simulations are given in App. A.4.

A comparison of the equilibrium pressure during the dynamics for three different temperatures is reported in Tab. 2. Notice that, in principle, the reference dataset and the corrected one can give different results. Indeed, the optimization of the antisymmetric part of the WF not only improves the consistency of forces and pressure, but can also modify the PES of the system. Nevertheless, if compared with the biased model, the MLIP trained on the corrected data gives results that are consistently much closer to the reference ones. This suggests that the main discrepancy between

the biased and reference models comes from the SCE, and that the SCE correction improves considerably the physical description of the system. We observe a maximum discrepancy of ~ 2 GPa at $r_s = 2.02$ and $T = 6000$ K between the biased and corrected results, while the corrected and reference ones are in statistical accordance within a joint error bar of 0.2 GPa at the same thermodynamic conditions. Remarkably, this state is very close to the Hugoniot compressibility peak, *i.e.*, the maximum in the density reached by the shocked state along the Hugoniot curve. This thermodynamic point is very sensitive to changes in the underlying equation of state, and a pressure shift as the one observed here between the corrected and biased models has a significant impact in the final position of the Hugoniot.[62] Similar conclusions can be reached by looking at the radial distribution function $g(r)$ (Fig. 5). Analogously to the pressure analysis, the $g(r)$ of the corrected model is compatible with the one given by the reference one. Also for this quantity the maximum discrepancy between the corrected and biased models is observed close to the compressibility peak at $T = 6000$ K, which corresponds to the onset of the the molecular-to-atomic transition along the Hugoniot curve. These results further highlight how SCE forces and pressure corrections significantly improve the physical outcome of the resulting MLIPs.

Table 2: Average pressure from MD simulations obtained with different MLIPs trained on the biased, corrected, and reference datasets, respectively. All models employ the Δ-learning technique with a DFT-PBE baseline.

	P_{bias} (GPa)	P_{corr} (GPa)	P_{ref} (GPa)
$r_s = 2.16$, $T = 4000$ K	17.1(1)	17.4(1)	17.4(1)
$r_s = 2.02$, $T = 6000$ K	26.8(1)	28.5(1)	28.7(1)
$r_s = 1.92$, $T = 8000$ K	38.3(1)	39.8(1)	39.9(1)

4 Conclusions

Machine learning models have become a key tool to reach system sizes and simulation times not achievable with first principle methods, while showing a similar accuracy. In particular, MLIPs trained on data generated by advanced quantum-mechanical algorithms, such as quantum Monte

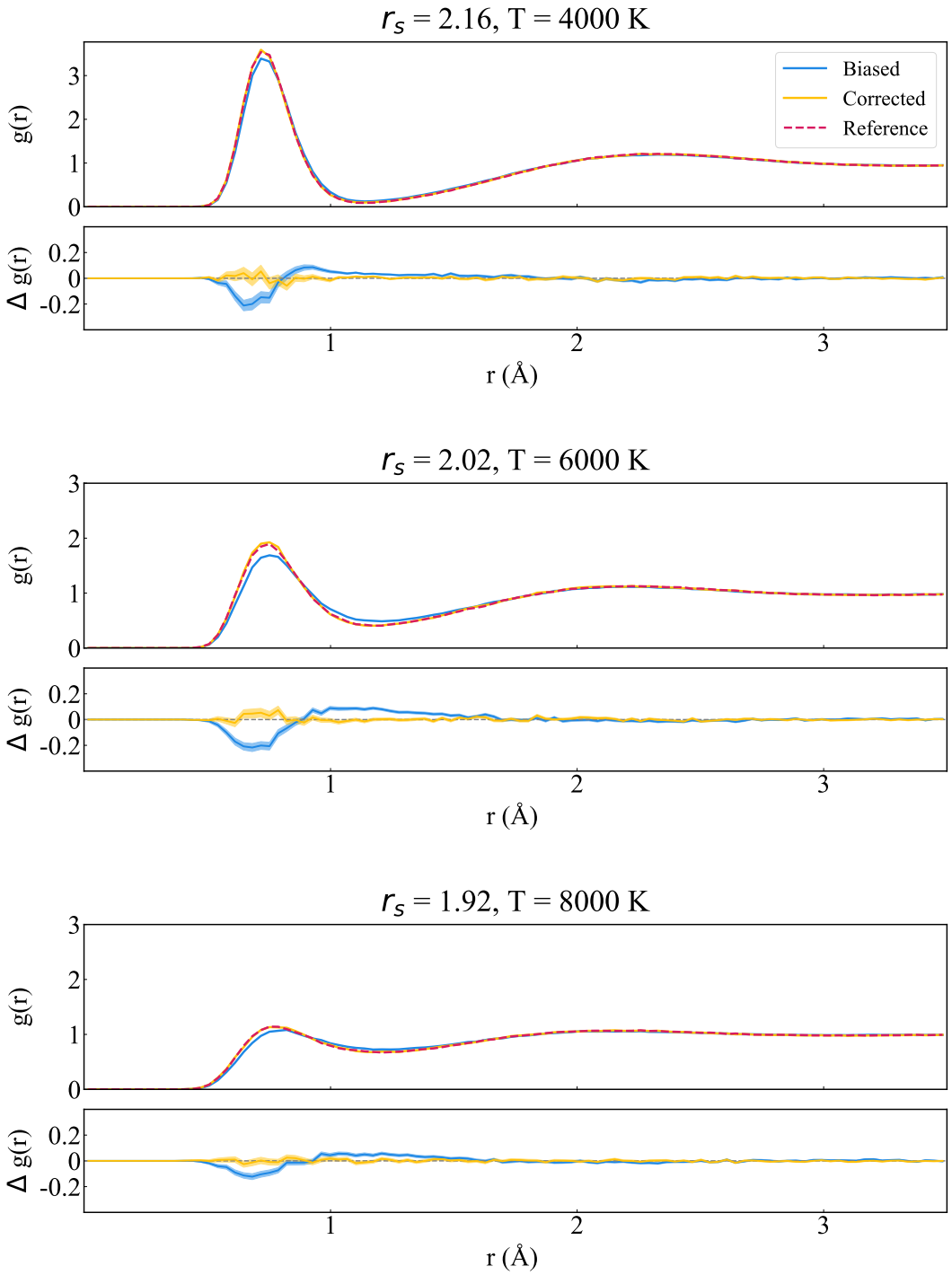

Figure 5: Radial distribution function $g(r)$ for densities and temperatures close to the Hugoniot curve, obtained with MLIPs trained on different datasets. The difference $\Delta g(r)$ with respect to the model trained on the reference dataset is also shown for both the "biased" and "corrected" models. The shaded area indicates the statistical uncertainty in $\Delta g(r)$.

Carlo (QMC), are essential to improve the description and our physical understanding of phenomena for which standard mean-field methods fail. In this work, we provided evidence on how VMC, in combination with a cheap-to-optimize JSD WF and a force/pressure correction, can be used to construct reliable QMC-based training sets. Our results can be summarized in the following three points.

First, we clearly demonstrated how the SCE affecting derivative quantities (*i.e.*, forces and pressures) computed with a partially optimized WF, such as the frozen-orbital JSD, can heavily spoil the accuracy of MLIPs trained on those data. This is manifested by a large dependence of the models' accuracy on the relative weights in the loss function.

Second, we showed how an easy-to-evaluate bias correction, previously developed in Ref. 26, can be applied to remove the SCE from the training set, and improve the accuracy of the resulting MLIPs.

Finally, we directly compared equilibrium thermodynamic quantities yielded by MD simulations driven by different MLIPs for the deuterium Hugoniot curve, taken as test system. We observed how the main discrepancy between MLIPs obtained from frozen-orbital and fully optimized WF datasets does not necessarily come from a different PES description of the dataset but from the intrinsic SCE affecting forces and pressures computed with partially optimized WFs. Indeed, in our working case provided by high-pressure hydrogen, the MLIPs trained on the corrected dataset give results in statistical agreement with those yielded by a reference model trained on VMC data with a fully optimized WF.

Our results reveal the importance of appropriately curing forces and pressures when they are computed with partially optimized WF in QMC and subsequently used to train MLIPs. The MLIPs quality degradation due to the SCE plaguing these VMC datasets is an effect that has been overlooked in previous applications. This could be even more relevant for heavier elements, since the SCE usually grows with the number of valence electrons.[25] Moreover, the SCE correction scheme is general and applicable to systems more complex than hydrogen, because it is based on cheaper DFT calculations. We have also shown that some numerical instabilities coming from the finite

difference approximation used to evaluate the SCE correction can be cured by an appropriate cut-off in the corresponding error bars. The development of a more efficient scheme for evaluating the SCE correction based on linear response theory is currently in progress. The availability of an SCE correction framework used in combination with other approaches, such as Δ-learning, to further increase the precision of the final models, opens the way to the efficient generation of accurate MLIP based on QMC electronic structure calculations.

Acknowledgments

K.N. is grateful for computational resources from the Numerical Materials Simulator at National Institute for Materials Science (NIMS), and from Research Institute for Information Technology at Kyushu University under the category of General Projects on the supercomputer GENKAI. M.C. acknowledges GENCI for providing computational resources on the CEA-TGCC Irene supercomputer cluster under project number A0170906493 and the TGCC special session. M. C. is grateful to EuroHPC for the computational grant EHPC-EXT-2024E01-064 allocated on Leonardo booster partition.

K.N. acknowledges financial support from MEXT Leading Initiative for Excellent Young Researchers (Grant No. JPMXS0320220025), from Murata Science and Education Foundation (Grant No. M24AN006), and from Japan Science and Technology Agency (JST), PRESTO (Grant No. JPMJPR24J9). M. C. thanks the European High Performance Computing Joint Undertaking (JU) for the partial support through the "EU-Japan Alliance in HPC" HANAMI project (Hpc AlliaNce for Applications and supercoMputing Innovation: the Europe - Japan collaboration).

Associated Content

Data Availability Statement

The *ab initio* QMC package used in this work, TurboRVB, is available from its GitHub repository `https://github.com/sissaschool/turborvb`.

All data necessary for reproducing the results reported in this work is available in the following GitHub repository `https://github.com/giacomotenti/unbiased_qmc_ML`

A Appendix

A.1 Kernel regression models parameters

In this Appendix we provide further details on the hyperparameters and specific kernels used for the models described in Sec. 3.2. The KRR models use a kernel as the one described in Ref. 23. Each environment is described by a density function written as a sum of Gaussian functions (with spread ε) multiplied by a localized function f_c depending on a cutoff radius r_c, *i.e.,*

$$\rho\left(\mathbf{r},\mathscr{R}_i\right) \propto \sum_{\left|\mathbf{R}_{ij}\right|\leq r_c} f_c\left(\left|\mathbf{R}_{ij}\right|\right) \exp\left(-\frac{\left|\mathbf{r}-\mathbf{R}_{ij}\right|^2}{\varepsilon}\right). \tag{13}$$

A similarity kernel between $\mathscr{R}_i$ and $\mathscr{R}_j$ is then defined by averaging the overlap between the two densities over a discrete set of N_{sym} symmetry operations $\hat{U}_k$:

$$\mathscr{K}\left(\mathscr{R}_i,\mathscr{R}_j\right) = \frac{1}{N_{\text{sym}}} \sum_{k=1}^{N_{\text{sym}}} \left[\int d^3\mathbf{r}\rho(\mathbf{r};\mathscr{R}_i)\rho(\mathbf{r};\hat{U}_k\mathscr{R}_j)\right]^n. \tag{14}$$

Finally, the normalized kernel is written as

$$K\left(\mathscr{R}_i,\mathscr{R}_j\right) = \left[\frac{\mathscr{K}\left(\mathscr{R}_i,\mathscr{R}_j\right)}{\mathscr{K}\left(\mathscr{R}_i,\mathscr{R}_i\right)\mathscr{K}\left(\mathscr{R}_j,\mathscr{R}_j\right)}\right]^\eta. \tag{15}$$

In the application presented here, we chose $r_c = 4$ Bohr and $\varepsilon = 1.5$ Bohr^2 in Eq. (13). In Eq. (14) we averaged over the cubic symmetry group (here appropriate since we are using cubic cells) and took $n = 2$. In Eq. (15) a value of $\eta = 2$ was employed.

For the local energy (Eq. (4)) we also added an additional pair-wise term as done in Ref.,[24] using cubic spline functions[63] defined on 10 equally spaced grid points between 0.3 Bohr and 5.5 Bohr.

Finally, the furthest point sampling method was used to select $N_{\text{env}} = 6000$ local environments among the training configurations.

A.2 Cutoff in the correction error

Figure 6 shows the error bars of the corrected forces in the dataset. We observed that, for certain force components and pressures ($\sim 10\%$ of all components), the correction is occasionally accompanied by large error bars. We identified the cause of this phenomenon in our current workflow, which estimates the parameter derivatives $\frac{d\lambda_{k,k'}}{d\mathbf{R}i}$ and $\frac{d\lambda_{k,k'}}{dV}$ in Eq.10 and Eq.11 using the FDM by performing multiple DFT calculations at displaced coordinates.[26] On the one hand, when employing localized basis sets in DFT, a cutoff is usually applied to reduce basis redundancy by eliminating elements corresponding to overlap matrix eigenvalues below a certain threshold, as written in Sec. 2.2. Without this cutoff, the forces and pressures computed according to Eq. 5 exhibit large error bars, as reported in Ref. 64. On the other hand, applying this cutoff causes the basis set to depend on atomic displacements in the FDM approach, as the quality of the basis set changes when ions move. This displacement-induced dependence occasionally deteriorates the error cancellation in Eqs. 10 and 11, resulting in larger error bars. Another source of large error bars is the finite smearing parameter used for molecular orbital (MO) occupations. Since hydrogen exhibits metallic behavior depending on its density, a smearing technique is necessary to achieve stable and smooth convergence for all configurations. However, employing smearing methods can introduce discontinuities when evaluating parameter derivatives $\frac{d\lambda_{k,k'}}{d\mathbf{R}i}$ and $\frac{d\lambda_{k,k'}}{dV}$ by

FDM. Additionally, the discrete real-space grid size used in DFT calculations contributes to error deterioration. Within the FDM approach, we need to perform $3N$ DFT calculations to compute the correction terms. To keep their computational cost low for large N, we chose a real-space grid size of 0.15^3 Bohr^3. We underline that the finite-difference step in ionic positions and cell volume for FDM, the smearing parameter, and the real-space grid have been chosen to provide converged forces and pressures, such that any possible bias is comparable with their total stochastic error bars. Yet, the occurrence of large error bars in the SCE correction could not be avoided for some particular ionic configurations by playing solely with these parameters. These factors indicate that the FDM method proposed in Ref. 26 can be improved for metallic systems in the future, although it is already robust for insulating systems with band gaps, as we have verified for c-BN, SiC, Si, and diamond.

To manage these erratic components, in this study, we retained the biased force values in the corrected data set whenever three standard deviations exceeded a given threshold, set here to 0.015 Ha/Bohr. As shown in Fig. 6, this threshold corresponds to a minimum in the standard deviation distribution for the corrected forces, indicating the onset of departure from normality. A similar strategy was employed for pressure, using a threshold of 1.5×10^{-5} a.u. We confirmed that forces and pressures with error bars below the threshold are unbiased, consistent with their corresponding potential energy surfaces (PESs), as illustrated in Figs.1 and 2. This choice was further validated by analyzing the behavior of Δ_E, Δ_F, and Δ_P (see Eq.(12)) as a function of the relative weight W_F/W_E in the loss function, using models trained on datasets corrected with different thresholds for σ_F. The results, depicted in Fig. 7, confirm that the threshold of 0.015 Ha/Bohr yields the best overall performance. Indeed, although Δ_E and Δ_P both steadily increase with increasing threshold values (for fixed W_F/W_E), the accuracy of forces deteriorates significantly when corrections with large error bars (corresponding to the largest threshold, i.e. $3\sigma_F \leq 0.04$ Ha/Bohr) are included.

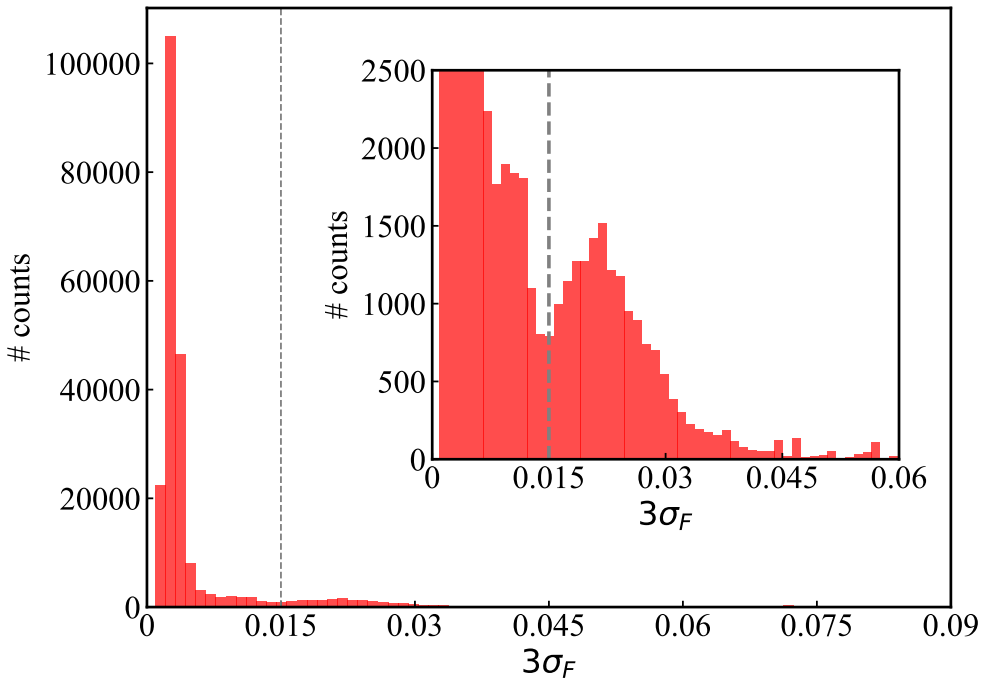

Figure 6: Distribution of three standard deviations for the corrected forces. The inset zooms-in its tail, to highlight the outliers. The vertical line indicates the 3σ threshold chosen in our application, *i.e.*, 0.015 Ha/Bohr.

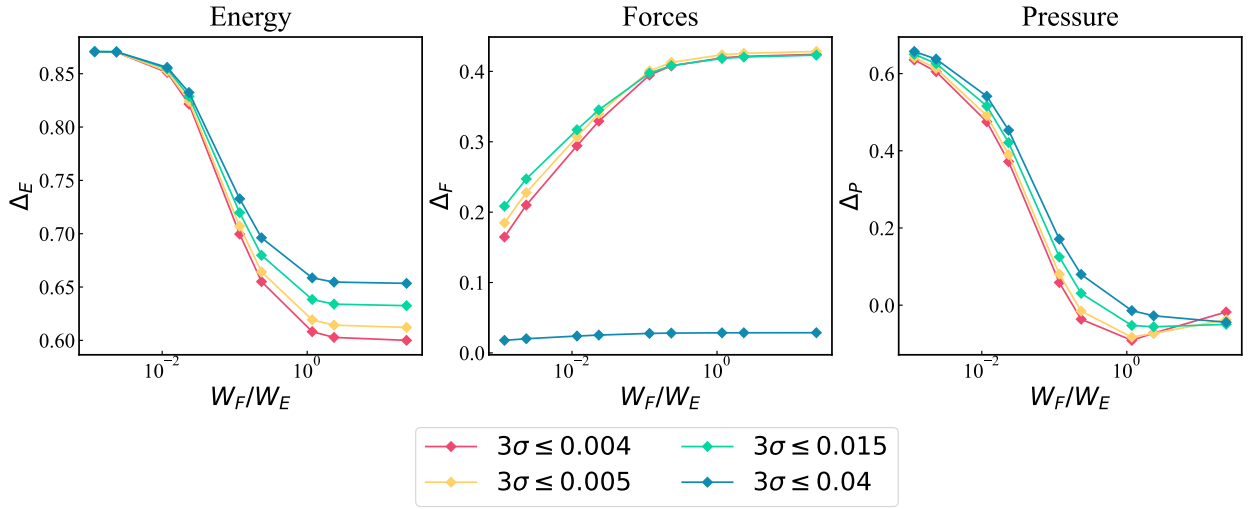

Figure 7: Δ_X for energy, forces, and pressure, as a function of the force/energy weight ratio in the loss function used for training (expressed in atomic units), for models trained on different versions of the corrected dataset. Each dataset corresponds to a different value of the $3\sigma_F$ threshold used to accept/discard the correction. The thresholds in the legend are reported in Ha/Bohr. For pressure, a threshold of 1.5×10^{-5} a.u. is used. The weight on the pressure was set to zero for all models.

A.3 MACE models comparison

Here, we further support the conclusions made in Sec. 3.2 by carrying out the same analysis using MACE.[40] This is a completely different MLIP architecture, because it is a framework based on equivariant message-passing neural networks with high-body order messages. The scope of this complementary analysis is to demonstrate that the behavior observed with the KRR models is general and only due to the training set consistency, and not linked to any particular implementation of the MLIP. The models consider 256 inner invariant features and a cutoff of 3 Å (effectively doubled when considering the message passing step). For each dataset, both training and test sets are the same as the one considered before for the KRR models in Sec. 3.2. The Adam optimizer[65] is used to find the models parameters. In particular, at each step the gradient of the loss function is evaluated on a subset of the training set ("batch") of $N_b = 4$ configurations. The results for the different datasets are reported in Fig. 8, where we show the behavior of Δ_X, defined in Eq. (12), for energy, forces, and pressure. They are consistent with those previously discussed in Sec. 3.2,

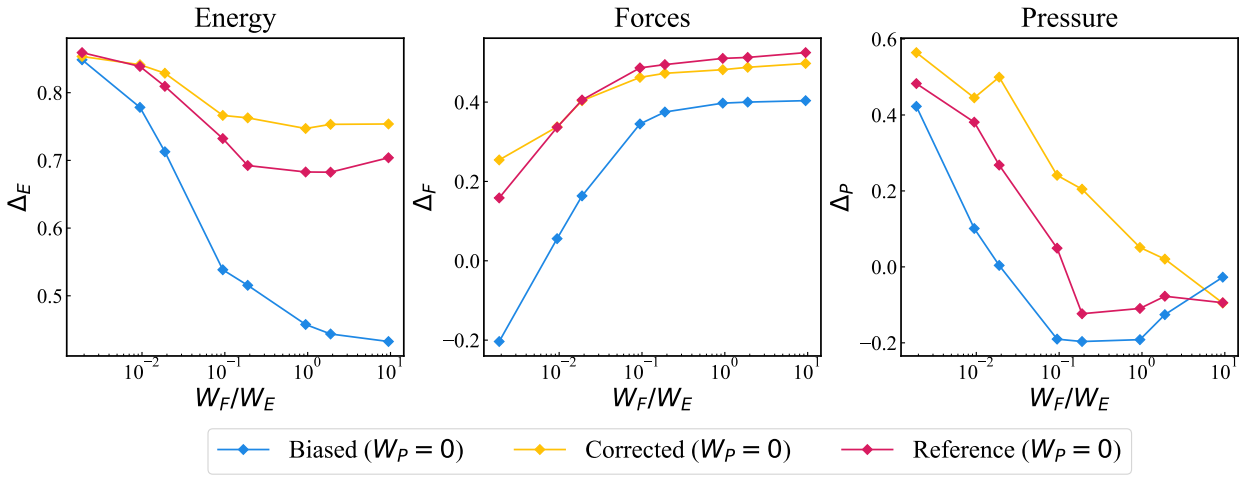

Figure 8: Relative RMSE variation Δ_X (Eq. 12) for energy, forces, and pressure, as a function of the force/energy weight ratio in the loss function (expressed in atomic units) for MACE MLIPs trained on the biased, corrected and reference datasets, respectively. The weight on the pressure is set to zero for all models.

showing the effect of the SCE and the importance of correcting it regardless of the particular choice for the MLIP framework.

A.4 Molecular dynamics simulation details

In this Appendix, we give further details on the MD simulations discussed in Sec. 3.3.

At each step, the energy, forces, and pressure were calculated at the DFT level using the QUANTUM ESPRESSO package in its GPU accelerated version[66–68] with the PBE functional, and then summed with those predicted by the different MLIPs. For the DFT simulations, a 60 Ry plane-wave cutoff with a projector augmented wave pseudopotential[69] was used together with a $4 \times 4 \times 4$ Monkhorst-Pack **k**-point grid. For the MD simulations we used a time step of 0.25 fs and a Langevin thermostat[46,70] with damping $\gamma = 0.13$ fs^{-1}.

After equilibration, we ran each simulation for about 4 ps to obtain the equilibrium quantities. The behavior of pressure as a function of the simulation time for different densities and temperatures and for each MLIP is shown in Fig. 9.

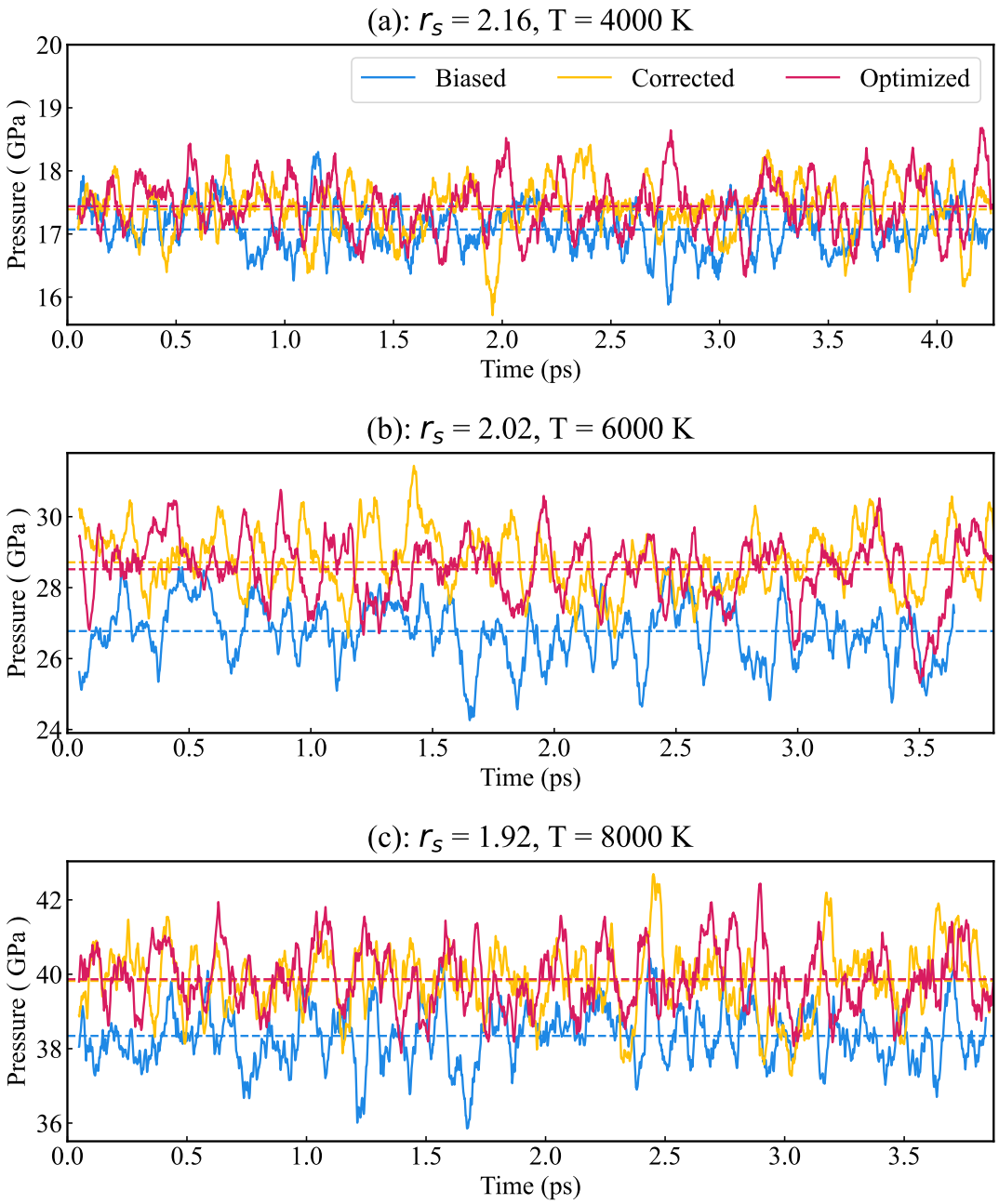

Figure 9: Pressure as a function of simulation time for a system of 128 hydrogen atoms at different thermodynamic conditions: (a) $r_s = 2.16$ and temperature $T = 4000K$, (b) $r_s = 2.02$ and temperature $T = 6000K$, and (c) $r_s = 1.92$ and temperature $T = 8000K$. Each solid line corresponds to an MLIP trained on the different datasets while the dashed lines are the average values.

References

(1) Blank, T. B.; Brown, S. D.; Calhoun, A. W.; Doren, D. J. Neural network models of potential energy surfaces. *J. Chem. Phys.* **1995**, *103*, 4129–4137.

(2) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(3) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114*, 3371–3383.

(4) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930–17955.

(5) Manzhos, S.; Dawes, R.; Carrington, T. Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces. *Int. J. Quantum Chem.* **2015**, *115*, 1012–1020.

(6) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C.* **2017**, *121*, 511–522.

(7) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31*, 1902765.

(8) Wood, M. A.; Cusentino, M. A.; Wirth, B. D.; Thompson, A. P. Data-driven material models for atomistic simulation. *Phys. Rev. B* **2019**, *99*, 184305.

(9) Dral, P. O. Quantum Chemistry in the Age of Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336–2347.

(10) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361–390.

(11) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037–10072.

(12) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186.

(13) Jia, W.; Wang, H.; Chen, M.; Lu, D.; Lin, L.; Car, R.; Weinan, E.; Zhang, L. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning. SC20: International conference for high performance computing, networking, storage and analysis. 2020; pp 1–14.

(14) Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods*; Cambridge University Press, 2004.

(15) Schleder, G. R.; Padilha, A. C. M.; Acosta, C. M.; Costa, M.; Fazzio, A. From DFT to machine learning: recent approaches to materials science–a review. *J. Phys.: Mater.* **2019**, *2*, 032001.

(16) Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. The ab-initio density matrix renormalization group in practice. *The Journal of Chemical Physics* **2015**, *142*, 034102.

(17) Jensen, H. J. A.; Jørgensen, P.; Ågren, H. Efficient optimization of large scale MCSCF wave functions with a restricted step algorithm. *Journal of Chemical Physics* **1987**, *87*, 451–466.

(18) Werner, H.-J.; Knowles, P. J. A second order multiconfiguration SCF procedure with optimum convergence. *The Journal of chemical physics* **1985**, *82*, 5053–5063.

(19) Rath, Y.; Booth, G. H. Interpolating numerically exact many-body wave functions for accelerated molecular dynamics. *Nature Communications* **2025**, *16*, 2005.

(20) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **2001**, *73*, 33–83.

(21) Becca, F.; Sorella, S. *Quantum Monte Carlo approaches for correlated systems*; Cambridge University Press, 2017.

(22) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

(23) Tirelli, A.; Tenti, G.; Nakano, K.; Sorella, S. High-pressure hydrogen by machine learning and quantum Monte Carlo. *Phys. Rev. B* **2022**, *106*, L041105.

(24) Tenti, G.; Nakano, K.; Tirelli, A.; Sorella, S.; Casula, M. Principal deuterium Hugoniot via quantum Monte Carlo and Δ-learning. *Phys. Rev. B* **2024**, *110*, L041107.

(25) Tiihonen, J.; Clay, I., Raymond C.; Krogel, J. T. Toward quantum Monte Carlo forces on heavier ions: Scaling properties. *J. Chem. Phys.* **2021**, *154*, 204111.

(26) Nakano, K.; Casula, M.; Tenti, G. Efficient calculation of unbiased atomic forces in ab initio variational Monte Carlo. *Phys. Rev. B* **2024**, *109*, 205151.

(27) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015.

(28) Christensen, A. S.; Von Lilienfeld, O. A. On the role of gradients for machine learning of molecular energies and forces. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 045018.

(29) Helled, R.; Mazzola, G.; Redmer, R. Understanding dense hydrogen at planetary conditions. *Nat. Rev. Phys.* **2020**, *2*, 562–574.

(30) Hu, S. X.; Militzer, B.; Goncharov, V. N.; Skupsky, S. First-principles equation-of-state table of deuterium for inertial confinement fusion applications. *Phys. Rev. B* **2011**, *84*, 224109.

(31) Craxton, R. S.; Anderson, K. S.; Boehly, T. R.; Goncharov, V. N.; Harding, D. R.; Knauer, J. P.; McCrory, R. L.; McKenty, P. W.; Meyerhofer, D. D.; Myatt, J. F.; Schmitt, A. J.; Sethian, J. D.; Short, R. W.; Skupsky, S.; Theobald, W.; Kruer, W. L.; Tanaka, K.; Betti, R.; Collins, T. J. B.; Delettrez, J. A.; Hu, S. X.; Marozas, J. A.; Maximov, A. V.; Michel, D. T.; Radha, P. B.; Regan, S. P.; Sangster, T. C.; Seka, W.; Solodov, A. A.; Soures, J. M.; Stoeckl, C.; Zuegel, J. D. Direct-drive inertial confinement fusion: A review. *Phys. Plasmas* **2015**, *22*, 110501.

(32) Hu, S. X.; Goncharov, V. N.; Boehly, T. R.; McCrory, R. L.; Skupsky, S.; Collins, L. A.; Kress, J. D.; Militzer, B. Impact of first-principles properties of deuterium–tritium on inertial confinement fusion target designs. *Phys. Plasmas* **2015**, *22*, 056304.

(33) Bonitz, M.; Vorberger, J.; Bethkenhagen, M.; Böhme, M. P.; Ceperley, D. M.; Filinov, A.; Gawne, T.; Graziani, F.; Gregori, G.; Hamann, P.; Hansen, S. B.; Holzmann, M.; Hu, S. X.; Kählert, H.; Karasiev, V. V.; Kleinschmidt, U.; Kordts, L.; Makait, C.; Militzer, B.; Moldabekov, Z. A.; Pierleoni, C.; Preising, M.; Ramakrishna, K.; Redmer, R.; Schwalbe, S.; Svensson, P.; Dornheim, T. Toward first principles-based simulations of dense hydrogen. *Phys. Plasmas* **2024**, *31*, 110501.

(34) Monacelli, L.; Casula, M.; Nakano, K.; Sorella, S.; Mauri, F. Quantum phase diagram of high-pressure hydrogen. *Nature Physics* **2023**, *19*, 845–850.

(35) Duvall, G. E.; Graham, R. A. Phase transitions under shock-wave loading. *Rev. Mod. Phys.* **1977**, *49*, 523–579.

(36) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(37) Bartók, A. P.; Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051–1057.

(38) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(39) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. commun.* **2022**, *13*, 2453.

(40) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. Adv. Neural Inf. Process. 2022; pp 11423–11436.

(41) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.

(42) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(43) Assaraf, R.; Caffarel, M. Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces. *The Journal of Chemical Physics* **2003**, *119*, 10536–10552.

(44) Umrigar, C. J. Two aspects of quantum monte carlo: Determination of accurate wavefunctions and determination of potential energy surfaces of molecules. *Int. J. Quantum Chem.* **1989**, *36*, 217–230.

(45) Assaraf, R.; Caffarel, M. Computing forces with quantum Monte Carlo. *J. Chem. Phys.* **2000**, *113*, 4028–4034.

(46) Attaccalite, C.; Sorella, S. Stable Liquid Hydrogen at High Pressure by a Novel Ab Initio Molecular-Dynamics Calculation. *Phys. Rev. Lett.* **2008**, *100*, 114501.

(47) Sorella, S.; Capriotti, L. Algorithmic differentiation and the calculation of forces by quantum Monte Carlo. *J. Chem. Phys.* **2010**, *133*, 234111.

(48) Filippi, C.; Assaraf, R.; Moroni, S. Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. *J. Chem. Phys.* **2016**, *144*, 194105.

(49) Nakano, K.; Raghav, A.; Sorella, S. Space-warp coordinate transformation for efficient ionic force calculations in quantum Monte Carlo. *J. Chem. Phys.* **2022**, *156*, 034101.

(50) Nakano, K.; Attaccalite, C.; Barborini, M.; Capriotti, L.; Casula, M.; Coccia, E.; Dagrada, M.; Genovese, C.; Luo, Y.; Mazzola, G.; Zen, A.; Sorella, S. TurboRVB: A many-body toolkit for ab initio electronic simulations by quantum Monte Carlo. *J. Chem. Phys.* **2020**, *152*, 204121.

(51) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-mechanical condensed matter simulations with CRYSTAL. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1360.

(52) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048–5079.

(53) Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **1957**, *10*, 151–177.

(54) Azadi, S.; Cavazzoni, C.; Sorella, S. Systematically convergent method for accurate total energy calculations with localized atomic orbitals. *Phys. Rev. B* **2010**, *82*, 125112.

(55) Sorella, S. Green Function Monte Carlo with Stochastic Reconfiguration. *Phys. Rev. Lett.* **1998**, *80*, 4558–4561.

(56) Toulouse, J. Introduction to the calculation of molecular properties by response theory. `https://hal.science/hal-03934866/document`, 2018.

(57) Sorella, S.; Devaux, N.; Dagrada, M.; Mazzola, G.; Casula, M. Geminal embedding scheme for optimal atomic basis set construction in correlated calculations. *J. Chem. Phys.* **2015**, *143*, 244112.

(58) Holzmann, M.; Ceperley, D. M.; Pierleoni, C.; Esler, K. Backflow correlations for the electron gas and metallic hydrogen. *Phys. Rev. E* **2003**, *68*, 046707.

(59) Slootman, E.; Poltavsky, I.; Shinde, R.; Cocomello, J.; Moroni, S.; Tkatchenko, A.; Filippi, C. Accurate Quantum Monte Carlo Forces for Machine-Learned Force Fields: Ethanol as a Benchmark. *Journal of Chemical Theory and Computation* **2024**, *20*, 6020–6027.

(60) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(61) Tenti, G.; Jäckl, B.; Nakano, K.; Rupp, M.; Casula, M. Hydrogen liquid-liquid transition from first principles and machine learning. 2025; `https://arxiv.org/abs/2502.02447`.

(62) Clay, R. C.; Desjarlais, M. P.; Shulenburger, L. Deuterium Hugoniot: Pitfalls of thermodynamic sampling beyond density functional theory. *Phys. Rev. B* **2019**, *100*, 75103.

(63) Xie, S. R.; Rupp, M.; Hennig, R. G. Ultra-fast interpretable machine-learning potentials. *npj Comput. Mater.* **2023**, *9*, 162.

(64) Nakano, K.; Morresi, T.; Casula, M.; Maezono, R.; Sorella, S. Atomic forces by quantum Monte Carlo: Application to phonon dispersion calculations. *Phys. Rev. B* **2021**, *103*, L121110.

(65) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations (ICLR), San Diego, California, USA, May 7–9. 2015.

(66) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Corso, A. D.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentz-

covitch, R. M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens.Matter* **2009**, *21*, 395502.

(67) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Corso, A. D.; de Gironcoli, S.; Delugas, P.; DiStasio, R. A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; de-la Roza, A. O.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys.: Condens.Matter* **2017**, *29*, 465901.

(68) Giannozzi, P.; Baseggio, O.; Bonfà, P.; Brunato, D.; Car, R.; Carnimeo, I.; Cavazzoni, C.; de Gironcoli, S.; Delugas, P.; Ferrari Ruffino, F.; Ferretti, A.; Marzari, N.; Timrov, I.; Urru, A.; Baroni, S. Quantum ESPRESSO toward the exascale. *J. Chem. Phys.* **2020**, *152*, 154105.

(69) H.PBE-KJPAW_PSL.1.0.0.UPF pseudopotential from `http://pseudopotentials.quantum-espresso.org/legacy_tables/ps-library/h`.

(70) Ricci, A.; Ciccotti, G. Algorithms for Brownian dynamics. *Mol. Phys.* **2003**, *101*, 1927–1931.