Improving accuracy of interatomic potentials: more physics or more data? A case study of silica.

Ivan S. Novikov\textsuperscript{a}, Alexander V. Shapeev\textsuperscript{a,∗}

\textsuperscript{a}Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Nobel St. 3, Moscow 143026, Russia

Abstract

In this paper we investigate the extent to which a charge-equilibration (QEq) model can improve the accuracy of interatomic potentials in general, and a machine-learning potential in particular. To that end, we compare three kinds of interatomic potentials: the Stillinger-Weber (SW) potential, the SW potential combined with the QEq model (SW+QEq) and the Moment Tensor Potential (MTP) fitted to a DFT dataset of $\alpha$-quartz SiO$_2$-based structures. In order to make a meaningful comparison, in addition to the accuracy, we assess the uncertainty of predictions of each potential. It is shown that MTP has the lowest error and uncertainty among the three potentials. Moreover, adding the QEq model to MTP does not make any improvement over the MTP potential alone.

Keywords: charge-equilibration model; Stillinger-Weber potential; machine-learning interatomic potentials; Moment Tensor Potential; uncertainty quantification

1. Introduction

Oxides represent one of the most important class of functional materials deriving their unique properties from the nature of ionic bonding of oxygen. Silicon dioxide (SiO$_2$, silica), although not rated as a functional material, has been extensively studied as it has many industrial applications including semiconductors \textsuperscript{1}, metal casting \textsuperscript{2}, and in the production of glass \textsuperscript{3}.

\textsuperscript{∗}Corresponding author

Email address: a.shapeev@skoltech.ru (Alexander V. Shapeev)
to name a few. In addition to experimental works \[4, 5, 6\], there has been extensive efforts in studying SiO$_2$ computationally. Density functional theory (DFT) is able to correctly describe ionic bonds and Coulombic interaction and hence provides an accurate description of interatomic interaction in the SiO$_2$ system. However, DFT is too computationally expensive to model some of the critical properties of SiO$_2$ such as the solidification of molten silicon, for which the interaction model has to be several orders of magnitude more computationally efficient.

(Semi-)empirical interatomic potentials has thus been the only alternative to DFT for conducting large-scale simulations of SiO$_2$. Examples of such potentials are the Stillinger-Weber (SW) potential \[7, 8\] and the Tersoff potential \[9\]. These potentials have a relatively simple functional form for the short-range interatomic interactions and they do not explicitly capture long-range Coulombic interactions in oxides. For example, the SW potential is a pair-interaction potential with a three-body term penalizing the bond angles that differ from the ones expected to occur in SiO$_2$. Many efforts have been made to put more physics into the model to make it more accurate. For instance, fixed-charge pair potentials combining short-range and long-range interactions have been proposed in \[10, 11\]. However, such the potentials cannot readjust to match the electrostatic environment. In \[12, 13\] the above fixed-charge potentials were extended. In those interatomic interaction models the charges became the parameters which were optimized and the effects of dipole polarization of the oxygen ions were taken into account. The charge equilibration (QEq) model is a more sophisticated model proposed by Rappe and Goddard \[14\]. QEq allows the charges to respond to changes in the electrostatic environment. Some interatomic interaction models, such as modified Tersoff \[15\], ReaxFF \[16\] and COMB \[17\] were constructed on the basis of the QEq model. These potentials have been successfully used in the description of SiO$_2$. However, application of these potentials to large-scale systems may be limited due to the fact that the QEq method requires a significantly larger computational effort than conventional empirical potentials.

Another research direction intended to increase the accuracy of the interatomic potentials is the so-called machine-learning interatomic potentials \[18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41\]. Ideologically, they are different from the empirical interatomic potentials in the way that machine-learning potentials attempt to increase accuracy not by putting more physics into the model, but through a flexible
functional form that allows large amounts of DFT data to be used for the fitting. The first work on this topic was published by Behler and Parinello [18]. They constructed a neural network potential (NNP) and successfully applied it to modelling of silicon; in particular, their potential predicted the radial distribution function of a silicon melt at 3000 K with high accuracy. In [19] on the basis of the idea of Gaussian process regression, the Gaussian approximation potential (GAP) was proposed and successfully applied for prediction of various properties of carbon, silicon and germanium. Thereafter, many works appeared that propose or validate interatomic potentials based on neural networks [25, 26, 27, 21, 28, 29, 30, 31, 32, 33, 34, 42], Gaussian processes [20, 35, 36, 37] and other methods [22, 23, 24, 38]. In the above works only short-range interaction was taken into account. Behler and his collaborators extended their NNP by including electrostatic interactions explicitly in the functional form [43] and testing it for ZnO, not reporting, however, that it improves the accuracy of their potential. A similar model directly predicting atomic charges has recently been proposed in [44]. The field of machine-learning interatomic potentials is ideologically and methodologically close to the field of machine-learning cheminformatics—developing of models predicting the properties of molecules and materials directly, without a molecular simulation [45, 46, 47, 48, 49, 50, 51, 52, 53].

Many studies on machine-learning potentials report [18, 19, 20, 28, 33, 35] that these potentials are more accurate than off-the-shelf empirical potentials. A conceptually interesting study was [20], where the authors considered a growing set of quantities of interest (phonons, elastic constants, defects, surfaces, etc.) and show that it is possible to construct a series of potentials that reproduce this growing set without losing accuracy by increasing the number of parameters in a potential. The most closely related paper to this study is [21]. In this paper the authors show that when the potentials are fitted on a very large configurational space (containing bulk, surfaces, clusters, etc.) for pure gold, a machine-learning potential was significantly more accurate than ReaxFF. However, gold has largely delocalized metallic bonds that are hard to represent explicitly in an empirical potential with high accuracy. On the other hand, SiO$_2$ has ionic/covalent bonding which empirical potentials are expected to represent sufficiently accurately.

The main purpose of this work was to assess to what extent the accuracy can be increased by adding a charge equilibration model (and hence capture the ionic nature of SiO$_2$ bonding better) to an interatomic potential in general, and a machine-learning potential in particular. In order to answer that
question, we compare three different potentials, the Stillinger-Weber (SW) potential, the SW potential with the additional charge equilibration model (SW+QEq) and the Moment Tensor Potential (MTP), belonging to a class of machine-learning potentials. MTP was first proposed in [22] for the case of single-component materials and extended in [23, 24] to multi-component materials.

This paper is organized as follows. In Section 2 we describe the methodology namely, the interatomic potentials, their fitting, and our uncertainty quantification method. In Section 3 we present and discuss the results of numerical experiments. In particular, we compare the training errors in energies, forces and stresses for different models, as well as elastic constants, vacancy formation energies (VFEs), phonon spectra and radial distribution functions (RDFs) obtained by these models. Finally, in Section 4 we give the concluding remarks.

2. Methodology

In order to address the main question of this study, namely to what extent the accuracy of the description of ionic bonds can be improved by adding a charge-equilibration (QEq) model, we introduce the SW potential, the QEq model, and the MTP potential (Section 2.1). Since our main focus is the machine-learning potentials, we consider a local optimization method for finding the parameters (Section 2.2). Finally, to in order to better analyze the results, we introduce an uncertainty quantification method in Section 2.3.

2.1. Interatomic potentials

We start by formulating an SW potential in the form similar to the one from [8]. The potential has originally been developed to describe bulk silicon [54]. It expresses local interaction of atoms and does not explicitly account for the Coulombic long-range interactions. Let $x = \{(x_i, z_i) : i = 1, \ldots, n\}$ be a configuration with $n$ atoms, each atom is encoded by its position $x_i$ and atomic type $z_i$. We assume that the $i$-th atom interacts with its neighbors and we refer to the $j$-th atom as the neighbor of the $i$-th atom if the distance between them is not greater than a cutoff radius $R_{\text{cut}}$. The locality of interaction is expressed by expanding the total interaction energy as a sum of contributions of individual atoms (or, to be precise, individual atomic neighborhoods): $E(x) = \sum_i V_i := \sum_i V(r_i)$, where $r_i = (r_{i1}, \ldots, r_{ij}, \ldots, r_{in})$ is
the neighborhood of the $i$-th atom, $r_{ij} = x_j - x_i$ is the position of the $j$-th atom relative to the $i$-th atom.

We generalize the SW potential by allowing its radial and angular functions to be arbitrary. To that end, we introduce a radial basis

$$
\varphi_{\alpha}(r) = \begin{cases} 
T^\alpha(r)(R_{\text{cut}} - r)^2 & r < R_{\text{cut}} \\
0 & r \geq R_{\text{cut}},
\end{cases}
$$

where $T^\alpha(r)$ is the Chebyshev polynomial of degree $\alpha$ on the interval $[R_{\text{min}}, R_{\text{cut}}]$, the term $(R_{\text{cut}} - r)^2$ is introduced to ensure a smooth cutoff to 0 at $r \geq R_{\text{cut}}$, and $R_{\text{min}}$ is some lower bound on minimal interatomic distances. For illustration, we plot the first six radial basis functions on the interval $[1.5 \, \text{Å}, 5 \, \text{Å}]$ in Figure 1.

Next we define the contribution $V_{i}^{SW}$ of each atom:

$$
V_{i}^{SW} := V_{i}^{SW2} + V_{i}^{SW3},
$$

$$
V_{i}^{SW2} = \frac{1}{2} \sum_j \sum_{\beta} c_{\beta,i} z_j \varphi_{\beta}(|r_{ij}|),
$$

$$
V_{i}^{SW3} = \frac{1}{2} \sum_j \sum_{k<j} \lambda_{z_i z_j z_k} \cos(\tilde{\theta}_{ijk} + \tilde{\theta}_{z_i z_j z_k})^2 \sum_{\beta} \tilde{c}_{\beta,i} z_j z_k \varphi_{\beta}(|r_{ij}|) \sum_{\gamma} \tilde{c}_{\gamma,z_k z_k} \varphi_{\gamma}(|r_{ik}|),
$$

![Figure 1: Radial basis $\varphi_1, \ldots, \varphi_6$.](image)
where $V_{SW}^{i} := V_{SW}^{i}(r_i)$ is the two-body term that describes pairwise interaction between the $i$-th and $j$-th atoms and $V_{SW}^{3} := V_{SW}^{3}(r_i)$ is the three-body term describing the interaction between the $i$-th, $j$-th, and $k$-th atoms. Here $\cos \theta_{ijk} = \frac{r_{ij} \cdot r_{ik}}{|r_{ij}| |r_{ik}|}$ is the cosine of the angle between $r_{ij}$ and $r_{ik}$.

Finally, $c_{\beta, z_i z_j}, \tilde{c}_{\beta, z_i z_j}, \tilde{c}_{\beta, z_i z_j}, \lambda_{z_i z_j z_k}$ and $\tilde{\theta}_{z_i z_j z_k}$ are the parameters that are determined during fitting. We denote them collectively by $\theta_{SW} := (c_{\beta, z_i z_j}, \tilde{c}_{\beta, z_i z_j}, \lambda_{z_i z_j z_k}, \tilde{\theta}_{z_i z_j z_k})$.

Thus, the total interaction energy depends on the atomic configuration $\mathbf{x}$ and on the parameters $\theta_{SW}$: $E_{SW} = E_{SW}(\theta_{SW}; \mathbf{x}) = \sum_{i=1}^{N} V_{SW}^{i}$.

Next we introduce MTP, first proposed in [22] and then generalized to multiple components in [24, 23]. It has the following form:

$$V_{i}^{MTP} := \sum_{\alpha} \xi_{\alpha} B_{\alpha}(r_i),$$

where $\xi_{\alpha}$ are the parameters to be fitted and $B_{\alpha}(r_i)$ are the basis functions. In order to define these functions we introduce the so-called moment tensor descriptors:

$$M_{\mu, \nu}(r_i) = \sum_{j} f_{\mu}(|r_{ij}|, z_i, z_j) r_{ij} \otimes ... \otimes r_{ij},$$

where the symbol “$\otimes$” stands for the outer product of vectors and therefore the angular part $r_{ij} \otimes ... \otimes r_{ij}$ resembles the moments of inertia, $f_{\mu}(|r_{ij}|, z_i, z_j)$ is the radial part of the following form:

$$f_{\mu}(|r_{ij}|, z_i, z_j) = \sum_{\beta} c_{\mu, z_i z_j}^{(\beta)} \varphi_{\beta}(|r_{ij}|),$$

in which $c_{\mu, z_i z_j}^{(\beta)}$ are the parameters to be fitted and $\varphi_{\beta}(|r_{ij}|)$ are the radial basis functions [11]. We construct our basis functions $B_{\alpha}$ as all possible contractions of the moment tensor descriptors [11] to scalar, e.g.:

$$B_0(r_i) = M_{0,0}(r_i),$$
$$B_1(r_i) = (M_{1,2}(r_i) M_{0,1}(r_i)) \cdot M_{1,1}(r_i),$$
$$B_2(r_i) = M_{0,2}(r_i) : M_{2,2}(r_i),$$
$$...$$
where "." is the dot product, "::" is the Frobenius product, $M_{0,0}(r_i)$ is a scalar ($\nu = 0$), $M_{0,1}(r_i)$ and $M_{1,1}(r_i)$ are vectors ($\nu = 1$), while $M_{0,2}(r_i)$, $M_{1,2}(r_i)$ and $M_{2,2}(r_i)$ are matrices ($\nu = 2$). We denote the free parameters of MTP by $\theta_{\text{MTP}} := (\xi_{\alpha}, c^{(\beta)}_{\mu, z_i, z_j})$. The total interaction energy is $E_{\text{MTP}} = E(\theta_{\text{MTP}}; x) = \sum_{i=1}^{n} V_{i}^{\text{MTP}}$.

Finally, we describe the QEq model. This model was proposed in [14] and we use it in the following form:

$$E_{\text{QEq}} := E(\theta_{\text{QEq}}; x, q) = \sum_{i=1}^{n} \left( \chi_{z_i} q_i + \frac{J_{z_i} q_i^2}{2} + \sum_{j<i} \frac{q_i q_j}{|r_{ij}|} \right),$$

where $\chi_{z_i}$ is the electronegativity of the atom of type $z_i$, $J_{z_i}$ describes repulsion between two electrons, $q_i$ and $q_j$ are the partial charges of the $i$-th and the $j$-th atom, respectively. The third sum in (5) describing the Coulombic interaction converges only conditionally, therefore we use the Ewald summation [55] for its calculation. We denote the free parameters of the QEq model by $\theta_{\text{QEq}} := (\chi_{z_i}, J_{z_i})$ and the collection of partial charges in a configuration by $q := (q_1, \ldots, q_n)$. It should be emphasized that (5) is a long-range interaction model with no cutoff radius.

We find the partial charges $q_i$ by solving the following optimization problem (equilibration of charges, see [14] for more details) for each configuration $x$ occurring in a simulation:

$$q^* = \arg \min_{q} E(\theta_{\text{QEq}}; x, q)$$

subject to $\sum_{i=1}^{n} q_i = 0$. (6)

The third interatomic interaction model used in this work is, thus, the combination of the SW potential and the QEq model:

$$E_{\text{SW+QEq}} := E(\theta_{\text{SW+QEq}}; x, q) = E(\theta_{\text{SW}}; x) + E(\theta_{\text{QEq}}; x, q),$$

where $\theta_{\text{SW+QEq}} := (\theta_{\text{SW}}, \theta_{\text{QEq}})$.

2.2. Fitting

We now describe the (machine-learning) optimization problems for finding the free parameters of the models described above. Let us consider a training set which consists of $K$ configurations $x^{(k)} (k = 1, \ldots, K)$. Suppose also that
we have the DFT energies $E^{\text{DFT},(k)}$, forces $f^{\text{DFT},(k)}$, and stresses $\sigma^{\text{DFT},(k)}$, where by $f^{\text{DFT},(k)}$ we mean the collection of forces on each atom of the $k$-th configuration and by $\sigma^{\text{DFT},(k)}$ the collection of (six) stresses of the $k$-th configuration. In order to find the free parameters of the SW potential we solve the following optimization problem:

$$L(\theta^{\text{SW}}) = \sum_{k=1}^{K} \left[ w_e \left( E^{\text{DFT},(k)} - E(\theta^{\text{SW}}; \mathbf{x}^{(k)}) \right)^2 ight. + \left. w_f \sum_i \left| f_i^{\text{DFT},(k)} - f_i(\theta^{\text{SW}}; \mathbf{x}^{(k)}) \right|^2 \right] \rightarrow \min,$$

(8)

where $w_e$, $w_f$ and $w_s$ are some nonnegative weights. We solve a similar problem in order to find the MTP parameters.

Because of partial charges in the combined model (7) we apply a slightly different algorithm for finding the free parameters of SW+QEq. Before each iteration of minimizing the objective function

$$L(\theta^{\text{SW+QEq}}) = \sum_{k=1}^{K} \left[ w_e \left( E^{\text{DFT},(k)} - E(\theta^{\text{SW+QEq}}; \mathbf{x}^{(k)}, q^{*(k)}) \right)^2 ight. + \left. w_f \sum_i \left| f_i^{\text{DFT},(k)} - f_i(\theta^{\text{SW+QEq}}; \mathbf{x}^{(k)}, q^{*(k)}) \right|^2 \right] \rightarrow \min,$$

(9)

we optimize the charges $q^{*(k)}$ for each configuration by solving the problem (9). Thus, we have different equilibrated charges $q^{*(k)}$ during each iteration of solving (9).

Since the models depend nonlinearly on the parameters $\theta$, we use a quasi-Newton optimization method, namely, the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS). To that end, we explicitly implemented the gradients of the loss function with respect to the parameters $\theta$ for each of the methods. In particular, for MTP, we have implemented an efficient back-propagation algorithm which has a favorable scaling when the number of parameters is large.

2.3. Uncertainty quantification

The fitted potential is a random quantity: it depends on the random training set and/or a particular local minimum that the optimization rou-
tine has found. Therefore the predictions of such potentials are also, strictly speaking, random. In order to quantify such randomness (uncertainty), we fit an ensemble of potentials of each type starting from random initial values of parameters and analyze the distribution of predictions by each (type of) potential, not just a single value of the “best” potential. In particular, we analyze the standard deviation (sometimes called predictive variance) of predictions of the ensemble of potentials and compare it to the actual error. As we will see, on our tests the standard deviation gives a good estimation of the actual error in all the quantities of interest considered in this study. Such technique of estimating the uncertainty of predictions is known as query by committee \[56\].

3. Numerical Testing

We fit three potentials, MTP, SW, and SW+QEq, and test how well they predict the elastic constants, phonon spectrum, and RDFs of the α-quartz SiO\textsubscript{2}. α-quartz is the stable crystalline structure of SiO\textsubscript{2} at normal temperature and pressure.

3.1. Training dataset

We composed the training dataset by perturbing and introducing defects to the α-quartz SiO\textsubscript{2}. This structure has a trigonal unit cell with the lattice parameters \(a = b = 5.022\ \text{Å}, c = 5.551\ \text{Å}, \alpha = \beta = 90^\circ, \text{and} \gamma = 120^\circ\). The unit cell contains 3 atoms of silicon and 6 atoms of oxygen.

We first generated 22 (all possible) supercells with 9, 18, 27 and 36 atoms by replicating the unit cell in different axes and applying shear. Next, in addition to pristine crystals, we generated a number of structures with a single O-vacancy defects, which were then relaxed (equilibrated). We used the VASP code \[57, 58, 59, 60\] for DFT calculations with the PBE functional \[61\], the PAW pseudopotentials \[62\], k-point meshes were equivalent to the \(4 \times 4 \times 3\) k-point mesh in the unit cell and a cutoff energy was 400 eV. After the configurations were relaxed, we randomly displaced every atom in each configuration by about 0.1 Å. From each “undisplaced” configuration several configurations with displaced atoms were generated. In addition, in order to predict elastic constants, we added 13 more configurations to the training dataset: 12 of them are relaxed configurations without atomic defects and with lattice vectors extended, compressed, or sheared by 2%, i.e., two configurations with extensions/compressions along each of six directions (\(xx, yy, \text{and} zz\)).
Table 1: Comparison of the average absolute and relative training root-mean-square errors and their standard deviations for the three models on the same dataset. In the last column the average silicon partial charge in the ideal α-quartz crystal and its standard deviation are shown.

| Potential   | energy error meV/atom | force error meV/Å (%) | stress error GPa (%) | Si partial charge in ideal crystal |
|-------------|-----------------------|-----------------------|----------------------|-----------------------------------|
| SW          | 2.61 ± 0.44           | 173.9 ± 14.2          | 0.42 ± 0.09          | -                                 |
| SW+QEq      | 2.47 ± 0.33           | 170.0 ± 13.8          | 0.41 ± 0.08          | 0.74 ± 0.25                       |
| MTP         | 1.63 ± 0.09           | 90.5 ± 4.7            | 0.20 ± 0.02          | -                                 |

zz, yz, xz, xy) and one of them is the relaxed configuration without any defects and extensions/compressions of lattice vectors. Thus, our dataset contains 418 SiO$_2$ atomic configurations with oxygen vacancies, random displacements of atomic positions and shear/compression.

3.2. Comparison of Potentials

We fit three types of potentials: SW (2), SW+QEq (7), MTP (3). For all the potentials we choose $R_{\text{min}} = 1.4$ Å, $R_{\text{cut}} = 5$ Å, and eight radial functions (1). We consider the MTPs with 92 basis functions $B_\alpha$. The weights in the objective functions (8), (9) were $w_e = 1$, $w_f = 10^{-2}$ Å$^2$, and $w_s = 10^{-3}$.

For each type of potentials we fit an ensemble of potentials in order to be able to estimate uncertainty due to randomness of the fitting. The average fitting errors and the standard deviations for each family of potentials are reported in Table 1. We trained 5 MTPs, 20 SW potentials and 20 SW+QEq combined models. We can see that MTP has the lowest error and also lowest uncertainty, i.e., lowest deviation between different MTP instances. Interestingly, the uncertainty in the error is very small—within a few percent of the actual error. For the family of SW potentials and SW+QEq models, in contrast, the fitting errors are about two times larger than those for the family of MTPs and the uncertainty in the error is three to four times larger than the same quantity for the MTPs. Another interesting observation is that the fitting errors of SW and SW+QEq are rather close to each other, in other words, adding QEq to SW does not significantly improve the error. For the
family of SW+QEq potentials we also present the average partial charge on a Si atom in the ideal crystal. We observe that the fitting procedure produces physically meaningful results, although we did not extract and fit to DFT charges; however, the uncertainty in predicting the charge from total energy and forces is rather large.

In order to check the predictive power of the potentials we compared elastic constants, VFEs, phonon spectra and RDFs calculated by these potentials to the results computed with DFT.

![Comparison of energy vs volume curves](image)

Figure 2: Comparison of the energy vs volume curves for SiO$_2$ $\alpha$-quartz unit cell obtained (a) by the MTPs, (b) the SW potentials, and (c) the SW+QEq models to the reference DFT curve (black). The the energy/volume dependence is a convex function for each fitted MTP and the energy vs volume curves are close to the reference DFT curve, whereas this is not the case for SW and SW+QEq.

The elastic constants were calculated for the $\alpha$-quartz SiO$_2$ nine-atom primitive unit cell. Our first finding was that the errors and uncertainties of SW and SW+QEq were unreasonably large (of the order of the magnitude of the constants themselves). In order to understand the large uncertainty of SW and SW+QEq, we studied the dependence of the energy on the cell volume as predicted by the three models. For MTP, the energy vs volume dependence is a convex function for each fitted MTP (see Figure 2) and the energy vs volume curves are close to the reference DFT curve. This is not the case for the SW potentials: the energy vs volume curves were either non-convex functions or their local minima were far from the reference (DFT) curve for many SW and SW+QEq potentials. Due to this reason
half of the trained SW potentials gave negative, or too large, \( C_{11} \) and \( C_{33} \) constants. Moreover, several SW potentials gave negative VFEs. Therefore we excluded those unphysical SW potentials and we were thus left with only four SW potentials with reasonable energy vs volume curves (see Figure 2). Similarly, after fitting 20 SW+QEq models, we excluded, using the same criteria, 15 of them.

| Model | DFT      | MTP      | SW       | SW+QEq   |
|-------|----------|----------|----------|----------|
| \( C_{11} \) | 91.1     | 77.6 ± 23.3 | 70.8 ± 30.3 | 100.8 ± 35.2 |
| \( C_{12} \) | 5.9      | −5.6 ± 2.8 | −5.1 ± 28.9 | 29.9 ± 21.4  |
| \( C_{13} \) | 16.0     | 13.5 ± 10.0 | 7.5 ± 18.8 | 30.1 ± 33.2  |
| \( C_{14} \) | 15.8     | 16.9 ± 4.7 | 17.7 ± 2.8 | 14.9 ± 16.1  |
| \( C_{33} \) | 93.8     | 87.7 ± 9.0 | 99.3 ± 15.0 | 103.0 ± 48.4 |
| \( C_{44} \) | 53.2     | 54.6 ± 10.1 | 51.0 ± 6.5 | 40.4 ± 24.6  |
| \( C_{66} \) | 42.6     | 40.2 ± 6.8 | 37.5 ± 5.7 | 33.6 ± 13.7  |
| bias  | 6.7      | 11.4     | 9.8      | 13.1      |
| uncertainty | 11.4 | 18.6 | 29.7 |

Table 2: Average elastic constants of \( \text{SiO}_2 \) \( \alpha \)-quartz (in GPa) calculated by DFT, MTPs, SW and SW+QEq models, and their standard deviations due to randomness in fitting. Bias is the root-mean-square deviation of the average predicted elastic constants from the reference DFT results. The last line shows the root-mean-square standard deviations (uncertainty) of the predictions. MTP reproduces the elastic constants better than SW and SW+QEq by 50 to 100 percent. The uncertainty in predictions is the lowest for MTP as well.

The elastic constants and their standard deviations for five MTP potentials, four SW potentials and five SW+QEq models are given in Table 2 together with their root-mean-square error compared to the reference DFT values (bias) and root-mean-square standard deviation (uncertainty due to randomness in the fitting). One can see that the uncertainty due to randomness in the fitting is about twice as large as the actual error of the mean prediction (bias). This indicates that the uncertainty thus calculated is a robust, but somewhat overestimated, measure of the actual error. MTP reproduces the elastic constants better than SW and SW+QEq by 50 to 100 percent, and the uncertainty of MTP is also the lowest compared to the other two potentials. An interesting observation is that the combination of SW and QEq yields slightly worse prediction of elastic constants as compared even to the plain SW potential.
Table 3: Average vacancy formation energies (eV) of SiO$_2$ α-quartz as calculated by the groups of MTP, SW and SW+QEq potentials, their standard deviations. In the last line the absolute errors (eV) in vacancy formation energy calculations (biases) are presented. MTP gives the best combination of the error and uncertainty, while SW and SW+QEq exhibit large uncertainty. The average prediction of SW+QEq is close to the exact one, however, with large spread between different instances of SW+QEq.

| Model | DFT | MTP   | SW    | SW+QEq |
|-------|-----|-------|-------|--------|
| VFE   | 2.23 | 2.14 ± 0.04 | 1.88 ± 0.41 | 2.31 ± 0.37 |
| Bias  | 0.09 | 0.35   | 0.08   | 0.08   |

In order to obtain the average VFEs and their standard deviations we relaxed and calculated the energies of two configurations: the first configuration is the $2 \times 2 \times 2$ supercell of 72 atoms and the second configuration is the same supercell with the oxygen atom vacancy. The average VFEs and their standard deviations calculated by the three groups of potentials and the reference DFT VFE are presented in Table 3. MTP gives the best average error, about 0.1 eV, and low standard deviation. SW+QEq also gives the same average error, but this could be just a coincidence—the standard deviation of the error is large, 0.4 eV. The plain SW potential yields a large error, 0.35 eV, however, this is the same as its standard deviation and the standard deviation is about the same as for SW+QEq. We cannot hence conclude, with certainty, that SW+QEq is better than SW alone in reproducing VFEs, but we can conclude that MTP is better than both SW and SW+QEq.

For comparison of phonon spectra and RDFs, we choose the “best” MTP potential, SW potential and SW+QEq model—namely, we choose those that reproduce the elastic constants and VFE better than the other potentials of the same type. The partial charges calculated by the “best” QEq model for most of silicon atoms were approximately equal to 0.9 and the same values for most of oxygen atoms were approximately equal to $-0.45$. It is worth noting that for the structures with oxygen atom vacancies the partial charges distribution is non-trivial: the partial charges for the atoms around the vacancy are smaller than those for the atoms located farther from the vacancy. We illustrate this fact on Figure 3 for one of the structures with vacancies from the training set.

We next test how well the potentials reproduce the phonon spectra. We used the PHONOPY open-source package to plot the spectra. The results shown in Figure 4 were obtained with the $4 \times 4 \times 4$ supercell of the
nine-atom primitive unit cell. The k-path for this system was Γ-M-K-Γ-A-L-H-A|L-M|K-H (see, e.g., [64]), where Γ, M, K, A, L and H are the high-symmetry points in the Brillouin zone. The three potentials, generally, show a good agreement with the reference DFT data except for the very high frequencies. SW+QEq seems to be least accurate and also largely underestimates the low-frequency mode around the A–L–H–A path, which may lead to lower stability in molecular dynamics.

Finally, we compare the radial distribution functions. The RDF calculations were performed in a 2×2×2 supercell with 72 atoms. To obtain the RDFs for the three potentials we ran molecular dynamics (MD) on LAMMPS [65] sampling an NVT ensemble with $T = 300$ K and time step of 1 fs. The reference RDFs were obtained by running MD on VASP. The RDFs are plotted in Figure 5. All the three models correctly predict the Si–Si RDF, with SW+QEq being slightly more accurate than the other two. On the contrary, for the O–O RDFs, the MTP is in a good agreement with DFT, while SW and SW+QEq did not describe the second peak (near $r = 3.5\text{Å}$) correctly.
Figure 4: Phonon spectra of SiO$_2$ α-quartz predicted by MTP, SW and SW+QEq (the figures (a), (b) and (c), respectively) compared to the reference DFT phonon spectrum (black circles). The three potentials, generally, show a good agreement with the reference DFT data except for the very high frequencies. SW+QEq seems to be the least accurate and also largely underestimates the low-frequency mode.

Figure 5: Radial distribution function (a) for Si–Si, and (b) O–O, computed at $T = 300$ K. All the three models correctly predict the Si–Si RDF. MTP is in the best agreement with DFT.
3.3. MTP+QEq potential

We have also tried an MTP+QEq potential, similarly to how we added QEq to SW. However, the results, in view of the above, were largely as expected: MTP+QEq was never more accurate than a plane MTP and had a much larger uncertainty of prediction. The only interesting observation was that the optimization algorithm often yielded a potential with zero charges, which is another indication that machine learning was better at capturing the effective Coulombic interaction than the QEq model in our tests.

4. Conclusion

In this work we investigated the extent to which a charge-equilibration model can improve the accuracy of interatomic potentials in general, and a machine-learning potential, MTP [22, 23, 24], in particular. To that end, we tested three kinds of interatomic potentials: the Stillinger-Weber (SW) potential, the SW potential combined with the charge equilibration model (SW+QEq) and the Moment Tensop Potential (MTP). We only used a local optimization algorithm for the parametrization of the potential, since global optimization algorithms are rarely used with machine-learning potentials (due to an extremely large complexity of such a combined algorithm).

In order to make a meaningful comparison we assessed the uncertainty of predictions of each potential. The uncertainty was due to the fact that typically the parameters of the fitted potentials are only near a local optimum, and an optimization routine typically finds some random local optimum.

Our main conclusion is that MTP has the lowest error and uncertainty among the three potentials and it is sufficient to use a local optimization algorithm for the parametrization of MTP. However, adding the QEq model to MTP did not make any improvement over the single MTP potential. In order to understand the underlying reason, we tested and compared the SW and SW+QEq models. We found that the uncertainties of their predictions were rather high and many instances of such potentials yield unphysical results (such as negative $C_{11}$ or $C_{33}$ elastic constants or negative vacancy formation energies). We emphasize that the empirical potentials are usually fitted directly to the quantities of interest (such as elastic constants, defect formation energies, etc.) with a global optimization algorithm. This indicates that a global optimization algorithms (or maybe some other, more sophisticated algorithm) have to be used to fit MTP+QEq in order to see improvement over MTP, while MTP itself shows a significant improvement over classical
interatomic potentials. It should be added that a charge-equilibration model is typically more computationally expensive than even a machine-learning potential (see the Supplemental Materials for details). It is, thus, not trivial to practically improve machine-learning potentials by adding a charge equilibration model.

5. Acknowledgements

The authors thank our colleague, Konstantin Gubaev, for giving advance access to the code implementing MTP. I.N. also thanks Dmitry Aksenov for valuable advices concerning the usage of DFT (VASP) and application of PHONOPY to the computations of phonon spectra. The work was supported by the Russian Science Foundation (grant number 18-13-00479).

6. Data availability

The data used for the fitting of the interatomic potentials are available to download from http://gitlab.skoltech.ru/Novikov/SiO2_training_set.

References

[1] P. F. Satterthwaite, A. G. Scheuermann, P. K. Hurley, C. E. Chidsey, P. C. McIntyre, Engineering interfacial silicon dioxide for improved metal–insulator–semiconductor silicon photoanode water splitting performance, ACS applied materials & interfaces 8 (20) (2016) 13140–13149.

[2] T. R. Rao, Metal casting: Principles and practice, New Age International, 2007.

[3] Z. H. Stachurski, Fundamentals of amorphous solids: structure and properties, John Wiley & Sons, 2015.

[4] L. Levien, C. T. Prewitt, D. J. Weidner, Structure and elastic properties of quartz at pressure, American Mineralogist 65 (9-10) (1980) 920–930.

[5] D. Strauch, B. Dorner, Lattice dynamics of alpha-quartz. i. experiment, Journal of Physics: Condensed Matter 5 (34) (1993) 6149.
[6] R. Ghita, C. Logofatu, C.-C. Negrila, F. Ungureanu, C. Cotirlan, A.-S. Manea, M.-F. Lazarescu, C. Ghica, Study of SiO$_2$/Si interface by surface techniques, in: Crystalline Silicon-Properties and Uses, InTech, 2011.

[7] T. Watanabe, H. Fujiwara, H. Noguchi, T. Hoshino, I. Ohdomari, Novel interatomic potential energy function for Si, O mixed systems, Japanese journal of applied physics 38 (4A) (1999) L366.

[8] P. Ganster, G. Tréglia, A. Saúl, Atomistic modeling of strain and diffusion at the Si/SiO$_2$ interface, Physical Review B 81 (4) (2010) 045315.

[9] S. R. Billeter, A. Curioni, D. Fischer, W. Andreoni, Ab initio derived augmented tersoff potential for silicon oxynitride compounds and their interfaces with silicon, Physical Review B 73 (15) (2006) 155329.

[10] D. t. Williams, S. Cox, Nonbonded potentials for azahydrocarbons: the importance of the coulombic interaction, Acta Crystallographica Section B 40 (4) (1984) 404–417.

[11] B. Van Beest, G. J. Kramer, R. Van Santen, Force fields for silicas and aluminophosphates based on ab initio calculations, Physical Review Letters 64 (16) (1990) 1955.

[12] P. Tangney, S. Scandolo, An ab initio parametrized interatomic force field for silica, The Journal of chemical physics 117 (19) (2002) 8898–8904.

[13] J. Kermode, S. Cereda, P. Tangney, A. De Vita, A first principles based polarizable O (N) interatomic force field for bulk silica, The Journal of Chemical Physics 133 (9) (2010) 094102.

[14] A. K. Rappe, W. A. Goddard III, Charge equilibration for molecular dynamics simulations, The Journal of Physical Chemistry 95 (8) (1991) 3358–3363.

[15] A. Yasukawa, Using an extended tersoff interatomic potential to analyze the static-fatigue strength of SiO$_2$ under atmospheric influence, JSME international journal. Ser. A, Mechanics and material engineering 39 (3) (1996) 313–320.
[16] A. C. Van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, W. A. God- 
dard, Reaxffsio reactive force field for silicon and silicon oxide systems, The Journal of Physical Chemistry A 107 (19) (2003) 3803–3811.

[17] J. Yu, S. B. Sinnott, S. R. Phillpot, Charge optimized many-body po-
tential for the Si/SiO\textsubscript{2} system, Physical Review B 75 (8) (2007) 085311.

[18] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Physical review letters 98 (14) (2007) 146401.

[19] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Physical review letters 104 (13) (2010) 136403.

[20] W. J. Szlachta, A. P. Bartók, G. Csányi, Accuracy and transferability of gaussian approximation potential models for tungsten, Physical Review B 90 (10) (2014) 104108.

[21] J. R. Boes, M. C. Groenenboom, J. A. Keith, J. R. Kitchin, Neural 
network and reaxff comparison for au properties, International Journal of Quantum Chemistry 116 (13) (2016) 979–987.

[22] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Modeling & Simulation 14 (3) (2016) 1153–1173.

[23] K. Gubaev, E. V. Podryabinkin, A. V. Shapeev, Machine learning of molecular properties: Locality and active learning, The Journal of Chemical Physics 148 (24) (2018) 241727.

[24] K. Gubaev, E. V. Podryabinkin, G. L. Hart, A. V. Shapeev, Accelerating high-throughput searches for new alloys with active learning of interatomic potentials, arXiv preprint arXiv:1806.10567.

[25] N. Artrith, A. M. Kolpak, Grand canonical molecular dynamics simulations of Cu–Au nanoalloys in thermal equilibrium using reactive ANN potentials, Comput. Mater. Sci. 110 (2015) 20–28.

[26] J. Behler, Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations, Phys. Chem. Chem. Phys. 13 (40) (2011) 17930–17955.
[27] J. Behler, Representing potential energy surfaces by high-dimensional neural network potentials, J. Phys. Condens. Matter. 26 (18) (2014) 183001. URL http://stacks.iop.org/0953-8984/26/i=18/a=183001

[28] P. E. Dolgirev, I. A. Kruglov, A. R. Oganov, Machine learning scheme for fast extraction of chemically interpretable interatomic potentials, AIP Advances 6 (8) (2016) 085318.

[29] M. Gastegger, P. Marquetand, High-dimensional neural network potentials for organic reactions and an improved training algorithm, J. Chem. Theory Comput. 11 (5) (2015) 2187–2198.

[30] S. Manzhos, R. Dawes, T. Carrington, Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces, Int. J. Quantum Chem. 115 (16) (2015) 1012–1020.

[31] S. K. Natarajan, T. Morawietz, J. Behler, Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials, Phys. Chem. Chem. Phys. 17 (13) (2015) 8356–8371.

[32] N. Lubbers, J. S. Smith, K. Barros, Hierarchical modeling of molecular energies using a deep neural network, J. Chem. Phys. 148 (24) (2018) 241715. doi:10.1063/1.5011181 URL https://doi.org/10.1063/1.5011181

[33] J. S. Smith, O. Isayev, A. E. Roitberg, ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost, Chem. Sci. 8 (4) (2017) 3192–3203. doi:10.1039/c6sc05720a URL https://doi.org/10.1039/c6sc05720a

[34] B. Kolb, L. C. Lentz, A. M. Kolpak, Discovering charge density functionals and structure-property relationships with PROPhet: A deep learning approach, Sci. Rep. 7 (1). doi:10.1038/s41598-017-01251-z URL https://doi.org/10.1038/s41598-017-01251-z

[35] V. L. Deringer, G. Csányi, Machine learning based interatomic potential for amorphous carbon, Phys. Rev. B 95 (2017) 094203. doi:10.1103/PhysRevB.95.094203 URL https://link.aps.org/doi/10.1103/PhysRevB.95.094203
[36] V. L. Deringer, C. J. Pickard, G. Csányi, Data-driven learning of total and local energies in elemental boron, Phys. Rev. Lett. 120 (15) (2018) 156001.

[37] A. Grisafi, D. M. Wilkins, G. Csányi, M. Ceriotti, Symmetry-adapted machine learning for tensorial properties of atomistic systems, Phys. Rev. Lett. 120 (3) (2018) 036002.

[38] A. Thompson, L. Swiler, C. Trott, S. Foiles, G. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, J. Comput. Phys. 285 (2015) 316 – 330. doi:http://dx.doi.org/10.1016/j.jcp.2014.12.018 URL http://www.sciencedirect.com/science/article/pii/S0021999114008353

[39] V. Botu, R. Ramprasad, Learning scheme to predict atomic forces and accelerate materials simulations, Phys. Rev. B 92 (9) (2015) 094306.

[40] Z. Li, J. R. Kermode, A. De Vita, Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces, Phys. Rev. Lett. 114 (2015) 096405. doi:10.1103/PhysRevLett.114.096405 URL http://link.aps.org/doi/10.1103/PhysRevLett.114.096405

[41] I. Kruglov, O. Sergeev, A. Yanilkin, A. R. Oganov, Energy-free machine learning force field for aluminum, Sci. Rep. 7 (1) (2017) 8512.

[42] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, SchNet: A continuous-filter convolutional neural network for modeling quantum interactions, in: Advances in Neural Information Processing Systems, 2017, pp. 992–1002.

[43] N. Artrith, T. Morawietz, J. Behler, High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide, Physical Review B 83 (15) (2011) 153101.

[44] B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, S. Tretiak, Transferable dynamic molecular charge assignment using deep neural networks, Journal of chemical theory and computation.
[45] M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. Von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, Physical review letters 108 (5) (2012) 058301.

[46] H. Huo, M. Rupp, Unified representation for machine learning of molecules and crystals, preprint.

[47] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke, Finding density functionals with machine learning, Physical review letters 108 (25) (2012) 253002.

[48] S. De, A. P. Bartók, G. Csányi, M. Ceriotti, Comparing molecules and solids across structural and alchemical space, Physical Chemistry Chemical Physics 18 (20) (2016) 13754–13769.

[49] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, O. A. von Lilienfeld, Fast machine learning models of electronic and energetic properties consistently reach approximation errors better than DFT accuracy, arXiv preprint arXiv:1702.05532.

[50] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Muller, A. Tkatchenko, Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space, The journal of physical chemistry letters 6 (12) (2015) 2326–2331.

[51] B. Huang, O. A. Von Lilienfeld, Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity, Journal of Chemical Physics 145 (16).

[52] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, Neural message passing for quantum chemistry, arXiv preprint arXiv:1704.01212.

[53] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, Nature communications 8 (2017) 13890.

[54] F. H. Stillinger, T. A. Weber, Computer simulation of local order in condensed phases of silicon, Physical review B 31 (8) (1985) 5262.
[55] P. P. Ewald, Die berechnung optischer und elektrostatischer gitterpotentiale, Annalen der physik 369 (3) (1921) 253–287.

[56] B. Settles, Active learning, Synthesis Lectures on Artificial Intelligence and Machine Learning 6 (1) (2012) 1–114.

[57] G. Kresse, J. Hafner, Ab initio molecular dynamics for liquid metals, Physical Review B 47 (1) (1993) 558.

[58] G. Kresse, J. Hafner, Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, Physical Review B 49 (20) (1994) 14251.

[59] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Computational materials science 6 (1) (1996) 15–50.

[60] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Physical review B 54 (16) (1996) 11169.

[61] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Physical review letters 77 (18) (1996) 3865.

[62] P. E. Blöchl, Projector augmented-wave method, Physical review B 50 (24) (1994) 17953.

[63] A. Togo, I. Tanaka, First principles phonon calculations in materials science, Scripta Materialia 108 (2015) 1–5.

[64] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al., Commentary: The materials project: A materials genome approach to accelerating materials innovation, Apl Materials 1 (1) (2013) 011002.

[65] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, Journal of computational physics 117 (1) (1995) 1–19.