

ARTICLE OPEN



Accurate large-scale simulations of siliceous zeolites by neural network potentials

Andreas Erlebach¹, Petr Nachtigall¹ and Lukáš Grajciar¹✉

The computational discovery and design of zeolites is a crucial part of the chemical industry. Finding highly accurate while computationally feasible protocol for identification of hypothetical siliceous frameworks that could be targeted experimentally is a great challenge. To tackle this challenge, we trained neural network potentials (NNP) with the SchNet architecture on a structurally diverse database of density functional theory (DFT) data. This database was iteratively extended by active learning to cover not only low-energy equilibrium configurations but also high-energy transition states. We demonstrate that the resulting reactive NNPs retain DFT accuracy for thermodynamic stabilities, vibrational properties, as well as reactive and non-reactive phase transformations. As a showcase, we screened an existing zeolite database and revealed >20k additional hypothetical frameworks in the thermodynamically accessible range of zeolite synthesis. Hence, our NNPs are expected to be essential for future high-throughput studies on the structure and reactivity of siliceous zeolites.

npj Computational Materials (2022)8:174; <https://doi.org/10.1038/s41524-022-00865-w>

INTRODUCTION

Zeolites are of central importance for numerous industrial applications ranging from catalysis through adsorption to ion exchange¹, owing to their highly diverse structures and properties. Theoretically, there are more than two million^{2–5} possible zeolite frameworks but only 240 zeolite frameworks listed in the IZA database^{6,7} have been prepared so far, a discrepancy known as the zeolite conundrum⁵. Therefore, ongoing research focuses on sophisticated synthesis routes, like the ADOR protocol⁸, allowing the preparation of “unfeasible” zeolites that are not accessible by standard solvothermal procedures^{9–11}. Another way to prepare new feasible or unfeasible zeolites is the polymorphous inter-zeolite transformation under elevated temperature or pressure^{12–16}. Finding reliable and computationally feasible protocol for identification of hypothetical zeolites that could be synthesized experimentally is still a great challenge.

In order to guide the ongoing search for new zeolites, computer simulations proved indispensable yet challenging for the (pre-)screening of structures and properties. Such a screening performed by Deem et al. allowed to narrow down the number of possible zeolite frameworks to thermodynamically accessible ones^{3,4}. This Deem database generated by atomistic simulations using analytical force fields contains about 330k hypothetical zeolites. Other computational studies used the IZA and Deem databases to estimate the feasibility of hypothetical zeolites and formulated design rules for their targeted solvothermal synthesis^{17–20}. Central quantity determining the feasibility of zeolites is the correlation between the zeolite density and energy, firstly calculated using atomistic simulations²¹ and then confirmed by experiments²².

Recently, the advent of machine learning in materials science and chemistry enabled the search for more complex correlations of the zeolite structure²³, stability, and properties²⁴. For example, graph similarity analysis of the Deem and IZA databases predicted thousands of possible diffusionless transformations from known to hypothetical zeolite frameworks²⁵. Apart from zeolite synthesis, machine learning studies also used the zeolite databases to find

structure-property correlations, e.g., for mechanical properties²⁶, discovery of new auxetic materials²⁷ and gas adsorption capacities to enable the targeted zeolite synthesis²⁸. However, the critical prerequisite for finding reliable correlations guiding experimental studies is generating accurate structural and energetic data at the atomistic level.

The atomistic simulations indeed provide vital insights into the structure and properties of zeolites²⁹. However, realistic modeling of zeolites with ab initio quality is frequently hampered by the prohibitive costs of first-principles methods. For example, only a few studies used atomistic simulations investigating the collapse of zeolites under high temperatures or pressures^{30,31}. Under high temperatures and low to moderate pressures, zeolites show a two-stage transformation, first to a low-density and subsequently to a high-density amorphous phase^{31–34}. Computational studies of such phase transition used either ab initio simulations employing simple structure models with few atoms and short timescales³⁰ or more realistic structure models and longer timescales but with analytical (reactive) force fields³¹. Large-scale simulations with ab initio quality are therefore of fundamental importance for discovering new zeolites not only by the screening of databases but also through the understanding of reaction pathways of zeolite phase transformations.

Enabling such simulations at a large-scale requires approximate modeling of the potential energy surface (PES) that retains the accuracy of high-level quantum mechanical calculations. In recent years, numerous machine learning potentials (MLP) have been proposed that accurately interpolate the PES providing the necessary speed-up compared to ab initio simulations^{35–39}. Among them are neural network potentials (NNP)³⁵ of different types and architectures, e.g., hierarchical interacting particle NNP (HIP-NN)⁴⁰, tensor field networks⁴¹, and the graph convolutional NNP SchNet^{42,43}. The latter is a message-passing type NNP architecture that uses trainable input representations of atomic environments repeatedly refined by convolutional operations in several iterations to model many-body interactions. Tests on benchmark datasets^{37,42,43} focusing on molecular systems proved

¹Department of Physical and Macromolecular Chemistry, Charles University, Hlavova 8, 128 00, Prague 2, Czech Republic. ✉email: lukas.grajciar@natur.cuni.cz

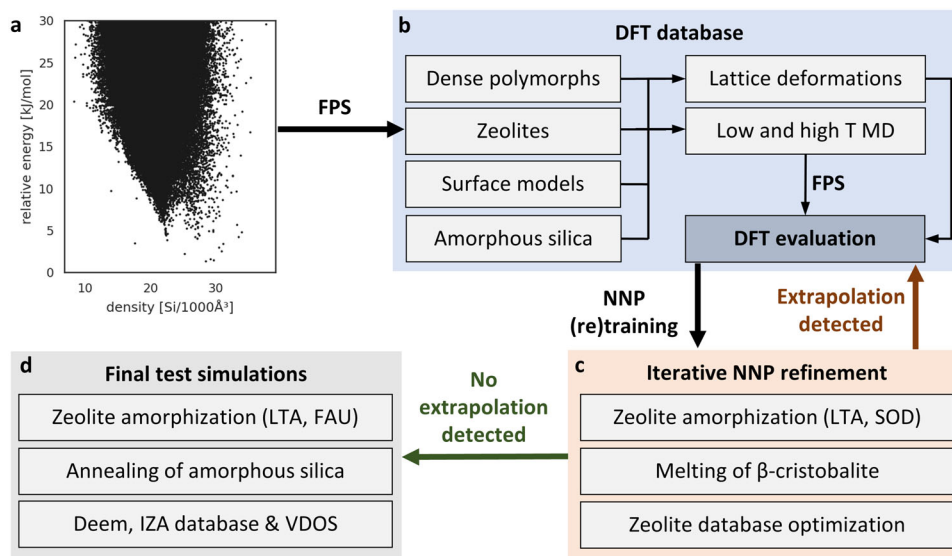


Fig. 1 Computational workflow. Starting from the Deem database (a) containing >330k hypothetical zeolites, a structurally diverse subset of zeolites was first selected by farthest point sampling (FPS). Unit cell deformations and MD simulations were applied to the selected zeolites, dense silica polymorphs, surface models and amorphous silica structures (b), followed by DFT single-point calculations on an FPS sparsified set of configurations. The NNPs trained on the initial database were iteratively refined by active learning using structure optimizations and MD simulations at high temperatures and pressures (c). The final NNP level calculations allowed the reoptimization of the Deem and IZA database, prediction of structure and vibrational properties, and simulations of highly reactive phase transformations such as glass melting and zeolite amorphization, including zeolites (FAU) not present in the training set (d).

the very good accuracy of SchNet NNPs to model energy and forces. However, little is known about transferability and accuracy of SchNet for materials science related questions³⁶, such as diffusion⁴⁴, phase stability⁴⁵ and transitions⁴⁶, phonon properties⁴⁷, and especially its robustness for reactive phase transformations of zeolites. So far, a few studies trained MLPs for PES modeling silica using polymorphs, surface models, amorphous and liquid configurations^{48,49}, including a recently trained MLP accurately modeling the structure and high-pressure phase diagrams of dense silica polymorphs⁵⁰. However, no silica MLP training considered the tremendous structural diversity of zeolites and their reactive phase transformations.

The central aim of this work is the training of reactive SchNet NNPs for accurate and general PES modeling of silica, including the structural diversity of zeolites over a wide density range. Training of an NNP ensemble allows active learning for iterative extension of the reference dataset and refinement of the NNP^{35,51}. The final dataset covers the silica configuration space ranging from low-density zeolites to high-pressure polymorphs, including low-energy equilibrium structures and high-energy transition states. This allows interpolation of the PES for accurate and transferable modeling of siliceous zeolites within the most relevant parts of the configuration space and enables the required large-scale simulations with ab initio accuracy.

The trained NNPs facilitated the reoptimization of the Deem database with high accuracy providing vital input for future machine learning studies to find correlations between structure, stability, and properties of zeolites. The database reoptimization also revealed >20 k additional hypothetical zeolites in the thermodynamically accessible range of zeolite synthesis. In addition, rigorous accuracy tests of the NNPs showed good agreement with DFT and experimental results, including not only equilibrium structures and phonon properties but also silica phase transformations under extreme conditions such as glass melting and the thermal collapse of zeolites. The trained NNPs show an accuracy improvement of about one order of magnitude for modeling energy and forces compared to other PES approximations: two state-of-the-art analytical force fields including the non-

reactive Sanders-Leslie-Catlow (SLC) potential^{52,53} and the reactive silica force field ReaxFF of Fogarty et al., and one tight-binding DFT parameterization GFNO-xTB^{54,55}. Consequently, this work provides a computational tool for accurate, reactive modeling of siliceous zeolites for their targeted design and synthesis.

RESULTS

Database generation and NNP training

Key prerequisite for the training of NNPs is the generation of a diverse dataset covering the variety of atomic structures and densities of zeolites in both low and high-energy parts of the PES to accurately model structure, equilibrium properties, and phase transitions. This was achieved by the computational procedure depicted in Fig. 1. Firstly, a small zeolite subset of ten frameworks was selected from the Deem database by farthest point sampling (FPS) along with SOAP as similarity metric (see Section 4.1) to capture the structural diversity with the least number of configurations (Supplementary Fig. 1). Then, unit cell deformations and MD simulations sampled low and high-energy parts of the PES using the selected zeolites, higher density polymorphs, surface models and amorphous (AM) silica. FPS extracted the most relevant structures from every MD trajectory to reduce the number of required DFT single-point calculations.

After training of an initial NNP ensemble, structure optimizations of the Deem and IZA database along with extrapolation detection using a query-by-committee approach enabled the search for additional frameworks required to sufficiently cover the zeolite configuration space. The sampling of previously unseen transition states included MD simulations for the melting of β-cristobalite, equilibration of liquid silica and the zeolite amorphization (ZA) of Linde Type A (LTA) and Sodalite (SOD). These simulations and NNP retraining on the extended DFT dataset were repeated until no extrapolation was detected. Two reference methods, PBE + D3 and SCAN + D3, were used to train NNPs (Supplementary Table 1) on the structural database (Supplementary Fig. 2). The corresponding NNPs (and their

ensembles) are termed as NNPpbe (eNNPpbe) and NNPscan (eNNPscan), respectively.

Finally, NNP level simulations were performed for: (i) the reoptimization of the Deem and IZA database, (ii) glass melting and ZA using Faujasite (FAU) not included in the training set to evaluate the NNP robustness for reactive phase transformations, and (iii) prediction of equilibrium structures and vibrational properties to compare the NNP results with their reference method and experimental data. Further details on the NNPs training and dataset details are provided in Section 4.

When confronting NNPs results with available experimental results, it must be stressed that NNPs cannot outperform the reference method used for generation of the training data. However, accurate NNPs help us understand the accuracy of the reference level of theory (DFT in this work) with respect to experiment providing that the NNPs retain DFT accuracy. Highly accurate NNPs can be used for the simulations of experimental observables using more realistic models and longer simulation times than possibly allowed with more demanding DFT calculations. Thus, the accuracy of the employed DFT methods and the NNPs is demonstrated first (Section 2.2) while the performance with respect to experimental data is described in following subsections (2.3–2.5).

Accuracy compared to other methods

Before evaluating NNP accuracy, we first benchmark the accuracy of the dispersion corrected PBE and SCAN functionals using available experimental data for the structure and energetics of siliceous zeolites (Supplementary Tables 2 and 3). Two dispersion corrections were considered, namely semiempirical D3 correction by Grimme et al.⁵⁶ and more involved many body dispersion (MBD)⁵⁷ correction (see Section 4.4 for more details). Previous studies showed that dispersion corrected PBE (both D3 and MBD corrected PBE) agrees best with the experimentally determined structures and phase transition enthalpies of siliceous zeolites as compared to several other GGA exchange-correlation functionals, including also the non-local vdW exchange-correlation (XC) functionals⁵⁸. However, no SCAN benchmark data have been reported so far for siliceous zeolites. Therefore, we used a small test to compare the performance between SCAN and PBE. The resulting relative energies are almost the same for both XC functionals and both dispersion corrections and are in very good agreement with observed phase transition enthalpies with mean average deviations (MAD) of about 2–3 kJ mol⁻¹, i.e., within the chemical accuracy. However, SCAN shows slightly better performance than PBE for modeling structural features, for example, with a density MAD of 0.2 Si nm⁻³ (SCAN + D3) versus 0.3 Si nm⁻³ (PBE + D3). In addition, Supplementary Fig. 3 compares the energies of the reaction pathway of the Stone-Wales (SW) defect formation⁵⁹ in a silica bilayer (*in vacuum*) calculated at the PBE + D3, SCAN + D3, and B3LYP level. The B3LYP has been shown⁵⁹ to yield the energy barrier for the first reaction step (for silica bilayer supported on Ru(0001) surface) close to experimental activation energies (about 3–4% deviation) and can be approximated as a reference level of theory also for the SW defect formation in a silica bilayer *in vacuum*. In comparison to the B3LYP reference, the PBE + D3 barriers are slightly underestimated (by up to 0.67 eV—~7% of the highest energy barrier) while SCAN + D3 barriers are slightly overestimated (by up to 0.59 eV—~6% of the highest energy barrier). In light of these results, we conclude that SCAN + D3 provides a consistent, albeit small, performance improvement over dispersion corrected PBE for equilibrium properties and reactive transformations of silica. In addition, earlier benchmarking studies for other systems showed that SCAN + D3 consistently outperforms PBE + D3 not only for equilibrium structures but also for reaction energies and activation barriers⁶⁰. Therefore, SCAN + D3 will be taken as the reference

Table 1. RMSE and MAE of energies [meV/atom] and forces [eV/Å] calculated for all test cases and only for equilibrium configurations (EQ) with respect to SCAN + D3 results.

Level of theory	Energy (EQ)		Forces (EQ)		Energy (all)		Forces (all)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
sNNPscan	2.99	4.20	0.048	0.070	3.90	5.44	0.175	0.303
eNNPscan	2.83	3.95	0.046	0.067	3.83	5.49	0.155	0.265
SLC	88.0	111	2.612	3.431	207	313	3.206	4.166
ReaxFF	56.4	78.4	1.266	2.996	88.8	136	2.789	8.533
GFNO-xTB	57.5	106	0.302	0.787	127	201	0.735	3.391

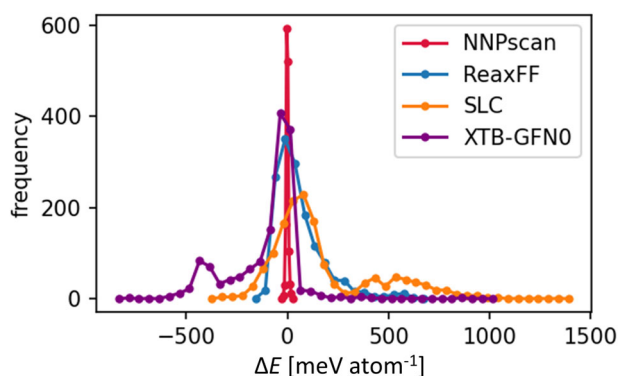


Fig. 2 Error distribution of energies. Energy errors ΔE are given with respect to SCAN + D3 for analytical force fields (SLC, ReaxFF) and tight-binding DFT (GFNO-xTB) as well as NNPscan.

DFT method and NNP trained on the SCAN + D3 data (NNPscan) will be considered as the reference NNP in the following sections, in which we will compare its performance to other PES approximations (analytical force fields, tight-binding DFT, etc.) and experimental data.

The NNP accuracy, together with the accuracy of the commonly used SLC potential, a reactive force field (ReaxFF), and one tight-binding DFT implementation (GFNO-xTB), is evaluated for the set of single-point energy calculations on a test set of structures taken from the NNPscan simulations (Sections 2.3–2.5). This test set contains 1460 configurations including (i) close to equilibrium (EQ) structures randomly chosen from the NNPscan optimized zeolite databases (see Section 2.3), (ii) silica bilayer configurations of the Stone-Wales defect formation (see Section 2.5), and (iii) high-energy structures from the glass melting and ZA simulations (see Section 2.5). These structures were not included in the reference dataset for NNP training. The entire test set can be found in the Zenodo repository (<https://doi.org/10.5281/zenodo.5827897>).

Table 1 summarizes the MAE and RMSE of energies and forces of all methods with respect to SCAN + D3 results (Supplementary Table 4 shows PBE + D3 results). Figure 2 shows the corresponding energy error distributions. In the case of EQ structures, NNPscan energies are in best agreement with SCAN + D3 calculations with an RMSE of <4.2 meV atom⁻¹, which are about the same as training errors of 4.7 meV atom⁻¹ (Supplementary Table 1) showing the good generalization capabilities of the NNPs. Using the ensemble average of six NNPs (eNNPscan) provides only minor improvement to the NNP accuracy. The analytical potentials (SLC, ReaxFF) and GFNO-xTB show higher errors by more than one order of magnitude. Such energy errors (~100 meV atom⁻¹) translate into uncertainties of zeolite phase stability calculations as described in Section 2.3 (see Fig. 3) of about 30 kJ (mol Si)⁻¹. The fairly good agreement of SLC with experimental and DFT results

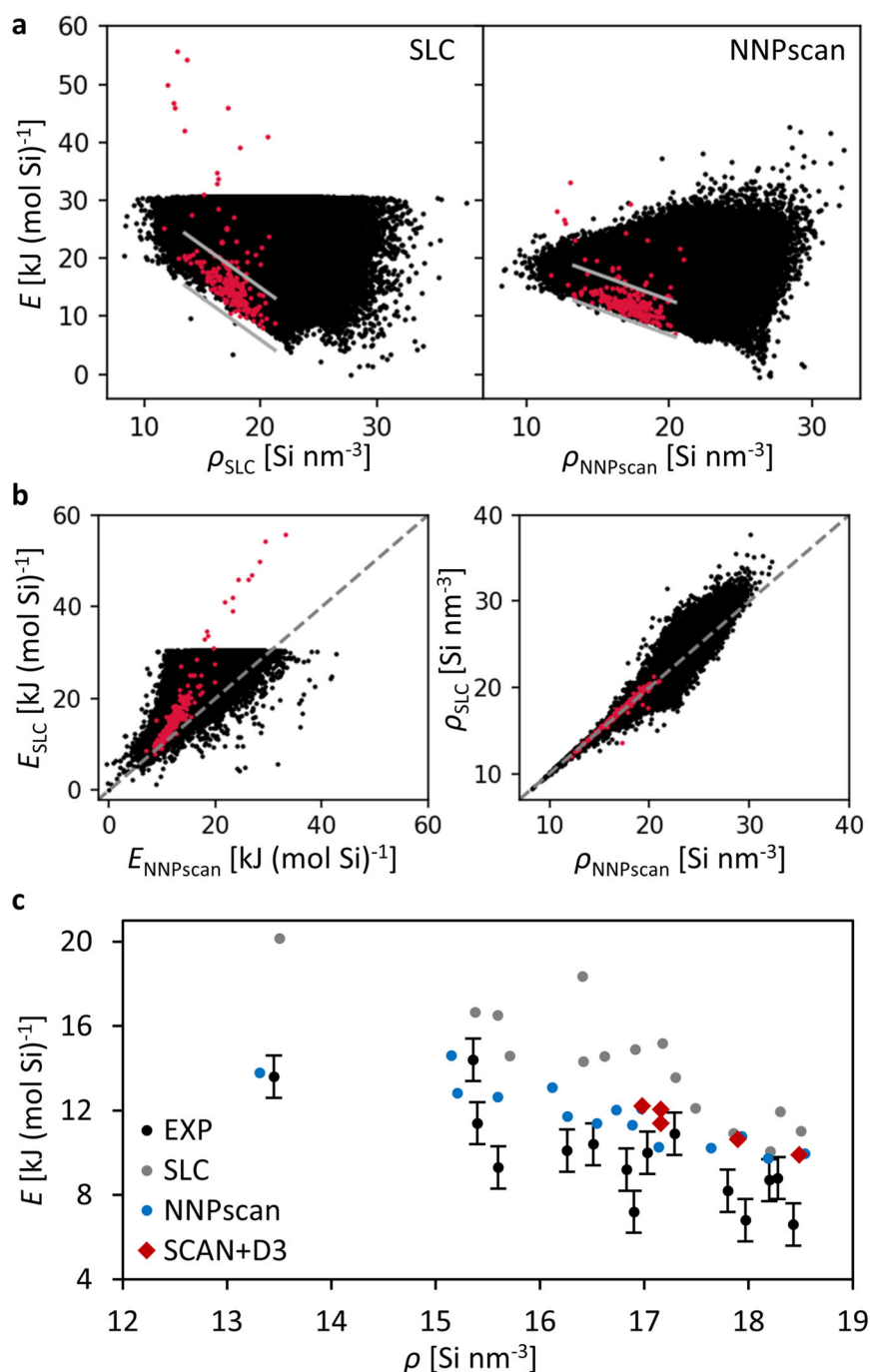


Fig. 3 Calculated energies and densities of siliceous zeolites. Relative energies E are given with respect to α -quartz as a function of framework density ρ (a) of hypothetical (black) and existing (red) zeolite frameworks calculated using SLC and NNPscan. Solid lines in a indicate the energy-density range of purely siliceous zeolites (Supplementary Fig. 4). b Correlations of energies and densities obtained at the NNPscan and SLC level. c Comparison of simulation results with experimental (EXP) values (Supplementary Table 5). Error bars correspond to an estimated experimental accuracy of ± 1 kJ (mol Si)⁻¹¹⁶⁴.

(Supplementary Table 5) applies only to existing siliceous zeolites but not to the rigorous test set including pure silica framework models that cannot be synthesized in their high-silica form (see Section 2.3). Therefore, the trained NNPs provide a sufficiently accurate PES covering much larger configurational space than SLC allowing reliable prediction of zeolite topologies that could be thermodynamically accessible, e.g., via alternative synthesis routes beyond solvothermal methods.

It must be stressed that the trained NNPs approximate energies and forces of the reference level DFT with high accuracy even for

high-energy structures and transition states (Table 1). For example, the NNPscan energies deviate about 10–27 meV atom⁻¹ from their DFT reference for the glass melting trajectories (see Section 2.5). In addition, even the errors of the extrapolated configurations of the ZA simulations (up to 40 meV atom⁻¹) are at least three times lower than the RMSEs of SLC, ReaxFF, and GFN0-xTB. Among the latter, ReaxFF tailored for the elements Si, O, and H provides the lowest energy errors but with an RMSE of 136 meV atom⁻¹¹⁵⁴. Recently, a benchmark study of ReaxFF potentials (parameterized for C, O, H) reported similar energy RMSEs of about 100 meV

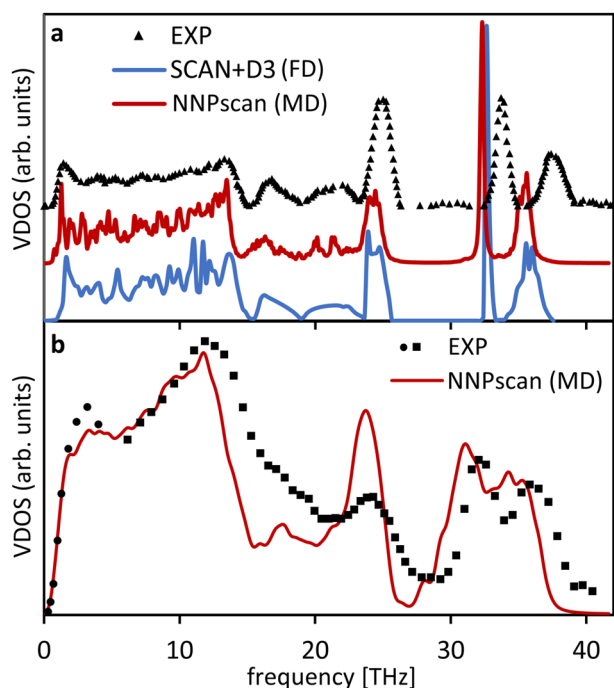


Fig. 4 Vibrational density of states (VDOS). Comparison of SCAN + D3 and NNPscan calculated VDOS for **a** α -cristobalite and **b** silica glass with experimentally observed VDOS (black triangles: ref. ⁶⁶, dots: ref. ⁷¹, squares: refs. ^{72,73}). VDOS were calculated using the velocity-autocorrelation from MD simulations (at 300 K) or the finite-difference (FD) approach.

atom⁻¹ for hydrogen combustion reactions⁶¹. On the other hand, GFNO-xTB allows a more general modeling with a parameterization for 86 elements focusing on equilibrium structures and frequency calculations⁵⁵. Therefore, GFNO-xTB shows a larger energy RMSE for transition state structures (Fig. 2) yet gives higher force accuracy than the silica potentials SLC and ReaxFF. Finally, using an ensemble of six tailor-made silica NNPs provides only a little improvement over single NNP calculations, the latter achieves the best performance in terms of accuracy and computational effort for the reactive modeling of silica.

Zeolite databases

The trained NNPs enable the reoptimization of the Deem and IZA database (available at: <https://doi.org/10.5281/zenodo.5827897>) to provide highly accurate input for investigations of structure-property relationships of existing and hypothetical zeolites. Figure 3a compares the relative energies and framework densities of the NNPscan optimized databases with the results from the SLC analytical potential, a state-of-the-art analytical potential for siliceous zeolites, taken from ref. ³ (www.hypotheticalzeolites.net, accessed: November 29, 2019). For sake of clarity, only the low-density zeolite analogue RWY⁶² (Ga_2GeS_6) is not shown in Fig. 3 (NNPscan: 61 kJ mol⁻¹, 7.86 Si nm⁻³; SLC: 104.2 kJ mol⁻¹, 7.62 Si nm⁻³), which does not exist in a high-silica form due to a large number of three-membered rings in the structure that would induce high ring tension. Please also note that the Deem database only includes hypothetical zeolites with relative energies up to 30 kJ mol⁻¹, which were deemed thermodynamically inaccessible in ref. ⁴ and therefore removed from the database.

Figure 3b shows the (qualitative) correlation between SLC and NNPscan results - the relative energies (left panel) and densities (right panel) of optimized structures from the Deem and IZA databases are compared. The Pearson correlation coefficients are 0.89 and 0.98 for energies and densities, respectively. However,

the SLC results show systematically higher relative energies than NNPscan for zeolites at high energies and densities, probably due to the energetic overestimation of structural features in those zeolites connected with the harmonic three-body bond-bending term of the SLC potential. For example, SLC yielded up to 20 kJ (mol Si)⁻¹ higher energies for three-ring containing zeolites such as OBW, OSO, NAB, and JOZ that can only be synthesized if Be is incorporated in the framework (Supplementary Fig. 4)⁶³. Therefore, the SLC potential probably overestimates the relative energies of hypothetical and existing frameworks that cannot be synthesized as purely siliceous zeolites.

To verify that NNPscan shows improved accuracy compared to SLC, Supplementary Table 5 compares the NNP and SLC results with experimentally available relative enthalpies and densities of 15 siliceous zeolites and five silica polymorphs^{67,64}. Additionally, DFT optimizations were applied to a subset of five zeolites and five polymorphs. Figure 3c qualitatively compares the dependence of relative energies on the framework density of siliceous zeolites calculated at the SLC, NNPscan, and SCAN + D3 level with experimentally determined phase transition enthalpies and densities. Such energy-density correlations were used in previous studies to find frameworks thermodynamically accessible for solvothermal zeolite synthesis^{3,4}. The analytical SLC potential shows relatively good agreement with experimental results in the case of purely siliceous zeolites along with an energy MAE of 4.0 kJ (mol Si)⁻¹. However, SLC systematically overestimates the experimentally observed enthalpies (Fig. 3c and Supplementary Table 5), which may relate to the energetic overestimation of structural features in zeolites due to the bond-bending term as described above. In contrast, the trained NNPscan achieved a substantial accuracy improvement with an energy MAE of 2.2 kJ (mol Si)⁻¹. Structure optimizations at the SCAN + D3 level of a smaller subset give MAE close to NNPscan, namely, 2.7 kJ (mol Si)⁻¹, which is a similar deviation from experiment as reported in previous DFT benchmark studies⁵⁸. The MAEs of atomic densities show a similar trend as that for relative energies, that is, the NNPs provide significantly higher quality than SLC for quantitative structural and energetic predictions of siliceous zeolites with almost same accuracy as dispersion corrected DFT methods (see Section 2.2 and Supplementary Tables S2 and S3).

Therefore, reoptimization of the Deem database using NNPscan provides significantly improved input for the computational design and discovery of new zeolite frameworks by analyzing structure, energy, and density correlations for hypothetical and existing frameworks^{3,4,17-20}. The solid lines in Fig. 3a show the range of relative energies and densities of the 40 zeolite frameworks that have been successfully synthesized in their purely siliceous form (Supplementary Fig. 4)⁶⁵. The SLC calculated (relative) energies and densities range between ~4–24 kJ (mol Si)⁻¹ and 13.5–21.2 Si nm⁻³, respectively. On the other hand, NNPscan optimizations yield a narrower energy range of 6–19 kJ (mol Si)⁻¹ but similar densities of 13.3–20.4 Si nm⁻³ (dashed lines in Supplementary Fig. 4). Hypothetical zeolites within these energy ranges can be considered as thermodynamically accessible by solvothermal synthesis methods. In the case of SLC, this applies to about 33k frameworks of the Deem database. However, due to the systemically overestimated SLC energies, >20k additional hypothetical zeolites (total of about 53k) were obtained from NNPscan calculations that fulfill the stability criterion mentioned above (Supplementary Fig. 5). These results demonstrate the crucial importance of accurate large-scale simulations of equilibrium structures for the discovery of zeolites.

Vibrational properties

In addition to simulations of equilibrium configurations at zero Kelvin, calculations of vibrational properties or free energies at elevated temperatures require accurate modeling of close to

equilibrium structures and forces on atoms. To test the reliability of SCAN + D3 and NNPscan for predicting the vibrational density of states (VDOS), the VDOS of α -cristobalite was calculated at both levels. Figure 4a shows both VDOS along with the experimentally observed one⁶⁶, and Supplementary Table 6 compares the frequencies of each vibrational mode with the experimental results from IR and Raman spectroscopy^{67–70}. Since α -cristobalite was part of the reference database, we performed additional VDOS calculations (NNPscan level) on vitreous silica structures not considered in the NNP training procedure. Three amorphous silica structures were generated using independent simulated annealing MD runs (Section 4). The obtained VDOS for the glass models are virtually identical (Supplementary Fig. 6), demonstrating sufficient sampling of amorphous structures. Figure 4b shows the average of the three calculated VDOS along with experimental results^{71–73}. Please note, the VDOS calculations at the SCAN + D3 level employed the finite-difference approach (FD) using the harmonic approximation while the NNPscan level calculations used MD simulations at 300 K for calculation of the velocity-autocorrelation function (VACF, see Section 4). The latter approach includes anharmonic effects, i.e., the temperature-dependent shift of vibrational frequencies. However, at low temperatures such as 300 K, only minor frequency shifts in the order of 0.1 THz are expected (e.g., as shown before for Al_2O_3 ⁷⁴, MgSiO_3 ⁷⁵) not influencing the comparison of different PES approximations shown in Fig. 4.

In the case of α -cristobalite, the VDOS calculated at the NNPscan and SCAN + D3 level are almost identical with a frequency MAD of about 0.4 THz for NNPscan compared to the SCAN + D3 reference (Supplementary Table 6). Both SCAN + D3 and NNPscan show good agreement with experimentally determined frequencies with MADs of 0.3 and 0.5 THz, respectively. The largest deviations of SCAN + D3 (up to 1.6 THz) and NNPscan (up to 2.0 THz) from experiment were obtained for the high-frequency modes A_2 (at 34.4 THz) and B_2 (at 35.6 THz). We obtained slightly higher frequency errors at the NNPscan level when the FD approach was applied to calculate the harmonic VDOS (Supplementary Fig. 7 and Supplementary Table 6) with an MAD of 0.9 THz from the experiment. These frequency changes are not connected with anharmonic (temperature) effects as described above^{74,75}. Most likely, the FD approach is prone to minor force errors of the few single-point calculations required to compute the VDOS. In contrast, the employed MD approach samples the VACF over a trajectory with several thousand microstates, probably facilitating a certain cancellation of the force errors and resulting in a better agreement with the SCAN + D3 FD results.

Since the MD approach proved more accurate for VDOS calculations at the NNPscan level, this procedure was also applied to the VDOS calculations of vitreous silica. Similar to α -cristobalite, the NNPscan calculated VDOS is in good agreement the experimentally observed one. In the case of the high-frequency doublet, the NNPscan calculations yielded a systematic shift by up to 1.5 THz with respect to the experimental VDOS. This shift is similar to the one observed above for the α -cristobalite VDOS at the SCAN + D3 and NNPscan level. Therefore, these systematically underestimated vibrational frequencies are expected to arise from the limitation of the DFT reference method. In summary, the trained NNPs can accurately model equilibrium structures and properties in line with their DFT reference and are in good agreement with available experimental observations.

Phase transitions

Apart from close to equilibrium properties, considering high-energy parts of the PES including transition states is indispensable for simulations of phase transitions and the thermal stability of zeolites potentially leading to the discovery of new zeolites. To showcase the accuracy of the trained NNPs for the description of

reactive events the Stone-Wales defect formation⁵⁹ in a silica bilayer was chosen as a test case. Figure 5 depicts the reaction path for Stone-Wales defect formation along with DFT and NNP energies (cf. Methods section). The bilayer structure is similar to the hypothetical bilayer structure in the reference dataset which consists of four, five, six and ten-membered rings (Supplementary Fig. 2c). However, no transition states from a six to seven-ring topology were included in the training set. Nonetheless, the NNPscan shows good agreement with its DFT reference. NNPscan deviates <0.207 eV from the corresponding DFT value, which is about 2% of the highest barrier.

Achieving general modeling of reactive and non-reactive zeolite phase transitions beyond the model reaction path described above requires diverse configurations of the high-energy parts of the PES. Two extreme cases of phase transformations were considered to probe the quality of the PES interpolation between the low-energy EQ and high-energy transition states, i.e., via Si-O bond cleavage (cf. Section 4.3): the melting and annealing of amorphous silica and ZA. Figure 6 shows the relative energies with respect to α -quartz for simulations using the NNPs including the melting of β -cristobalite and the amorphization of LTA and FAU zeolites. Note that simulations of β -cristobalite melting and LTA amorphization up to mass density of 2.2 g cm^{-3} (22 Si nm^{-3}) were used for training and active learning procedure, however, these simulations were carried out with different potentials, either ReaxFF or the initial NNPs (see Method section). Figure 6 also depicts results of DFT single-point calculations performed for a subset of structures as accuracy checks.

During the first $0.5 \cdot 10^6$ timesteps of the β -cristobalite melting simulation at 4800 K, only a few defects were created. The steep increase of the energy at about $0.6 \cdot 10^6$ timesteps corresponds to the phase transition into liquid silica. After 10^6 timesteps, the temperature was lowered stepwise down to 2500 K with no considerable changes in the structure during the last 100 000 timesteps. Again, NNPscan and SCAN + D3 results show very good agreement. Similar to the results for the Stone-Wales defect formation energies, NNPscan does not deviate from SCAN + D3 by <2–3% ($\sim 27 \text{ meV atom}^{-1}$).

In contrast to the melting of glass, thermal zeolite amorphization involves not only Si-O bond breaking but also considerable volume changes during the collapse of the framework. To mimic the thermal collapse of LTA and FAU, the structures were equilibrated at 1200 K for 6.5 ns with a stepwise volume reduction every 500 ps such that after 12 equivalent volume steps a mass density of 2.4 g cm^{-3} (24 Si nm^{-3}) was reached (cf. Method section). The target density exceeds the density range of the configurations in the AM and ZA part of the reference dataset by about 10% to demonstrate the NNP accuracy in extrapolated regions of the PES. Figure 6b, c show the energies of the last $2 \cdot 10^6$ timesteps of the trajectories. Figure 6d depicts example structures taken from the MD trajectory. Note that FAU was not included in reference database. In addition, the equilibration time of the last volume step was 1 ns in the case of FAU to ensure full equilibration in the final stage of the framework collapse.

FAU shows no Si-O bond breaking up to the density of amorphous silica (2.2 g cm^{-3}). However, the microporous structure considerably changes, mainly due to the collapse of the large cages. Starting from a mass density of 2.4 g cm^{-3} , the trajectories of FAU show Si-O bond breakage and reformation events during the last 10^6 timesteps. The transition states are five-fold coordinated Si leading to cleavage of Si-O bonds and reorientation of SiO_4 tetrahedra. The same bond cleavage mechanism was obtained for LTA. However, the first bond breaking was obtained at a density of 2.1 g cm^{-3} as indicated by the energy drop in Fig. 6c.

Up to a density of 2.2 g cm^{-3} , the deviation of NNPscan from SCAN + D3 is <11 meV atom^{-1} , which is about 3% of the highest

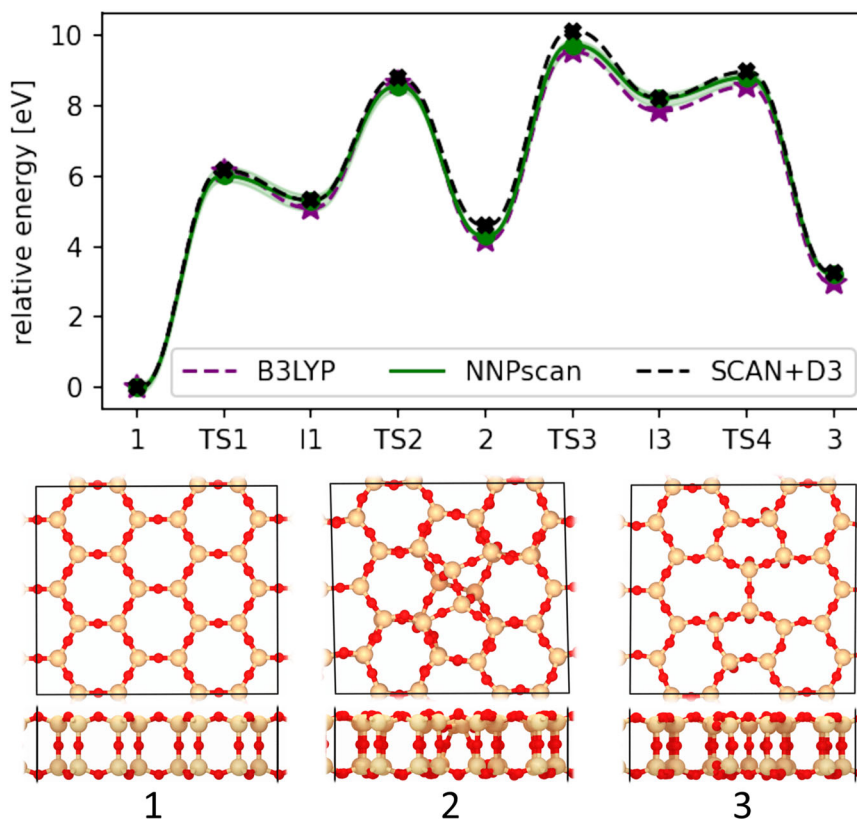


Fig. 5 Energetics of Stone-Wales defect formation. Energies for transition states (TS) and intermediate (I) structures are calculated at the DFT (B3LYP taken from ref. 59) and NNPscan level. Shaded areas indicate the energy range calculated with all six NNPs (solid line: ensemble average).

energy at the phase transition. At densities above 2.2 g cm^{-3} , larger NNP errors (up to 40 meV atom^{-1}) were obtained showing the onset of the extrapolation region in the high-energy part of the PES not covered by the training dataset (Supplementary Fig. 2b). The higher NNP errors are also indicated by the increased spread of the energy predictions of the NNP ensemble enabling future refinement of the NNPs by active learning. Indeed, the NNP energy deviation from the DFT reference considerably increased once the energy spread of the NNP ensemble exceeded $\sim 8 \text{ meV atom}^{-1}$ (Supplementary Fig. 8). However, even the extrapolated NNP energies qualitatively agree with the SCAN + D3 results with a Pearson correlation coefficient of 0.99 and 0.66 for FAU and LTA, respectively, showing a fairly systematic energy shift from the reference values (Supplementary Fig. 9). These results demonstrate that SchNet provides reasonable atomic configurations even in extrapolated regions at densities about 10% above the reference data facilitating a robust sampling of the configuration space for further active learning and NNP refinement.

DISCUSSION

Energy errors of a few meV atom^{-1} and force errors of about $100\text{--}300 \text{ meV \AA}^{-1}$ have been reported previously for state-of-the-art MLP such as moment tensor or Gaussian approximation potentials trained for large-scale simulations of different materials^{38,45–47}. Results reported herein show that SchNet NNPs provide the same quality as other MLPs not only for close-to-equilibrium structures of materials but also for high-energy bond-breaking scenarios. In addition, we have demonstrated that the trained SchNet NNPs retain DFT accuracy and provide at least an order of magnitude higher accuracy compared to analytical force fields and tight-binding DFT.

Previous trained reactive NNPs for silica⁴⁸ used a DFT database containing only two polymorphs (quartz and cristobalite), two surface structures, amorphous, and liquid silica configurations with unit cells comprising <144 atoms. The NNPs of Behler and Parrinello show RMSEs of about 200 meV \AA^{-1} for forces, i.e., somewhat higher compared to the RMSE of the database test set used in this work (147 meV \AA^{-1} , Supplementary Table 1). The DFT database used in this work covers several low- and high-density polymorphs, 2D models, amorphous structures, and the large structural diversity of zeolites using unit cells with up to 400 atoms, including high-energy transition states (Supplementary Fig. 2). Therefore, the NNPs provided in this work aim for a far more general modeling of the silica PES compared to previous studies⁴⁸ using only a small number of dense polymorphs, surface models, and amorphous silica structures.

The glass melting simulations clearly demonstrated the good NNP modeling accuracy for bond-breaking events at 4800 K. During the equilibration at such high temperatures, the MD trajectory showed numerous five-fold coordinated transition states of Si in good agreement with DFT results. These MD simulations were performed using silica glass density (2.2 g cm^{-3}) covered by the reference dataset containing configurations with densities from about 1.6 g cm^{-3} to 2.2 g cm^{-3} ($16\text{--}22 \text{ Si nm}^{-3}$) for high-energy transition states and densities from 1.0 to about 4.5 g cm^{-3} ($10\text{--}45 \text{ Si nm}^{-3}$) for low-energy EQ structures.

For comparison, the density range of the simulated zeolite collapse was 1.3 to 2.4 g cm^{-3} . At densities below 2.2 g cm^{-3} , MD simulations showed bond-breaking events in the case of LTA ($2.1\text{--}2.2 \text{ g cm}^{-3}$) and no bond cleavage in FAU. For both zeolites, the NNP energies and forces showed no extrapolation and agreed well with DFT results at densities $<2.2 \text{ g cm}^{-3}$. Note that FAU was not part of the reference database. Only further

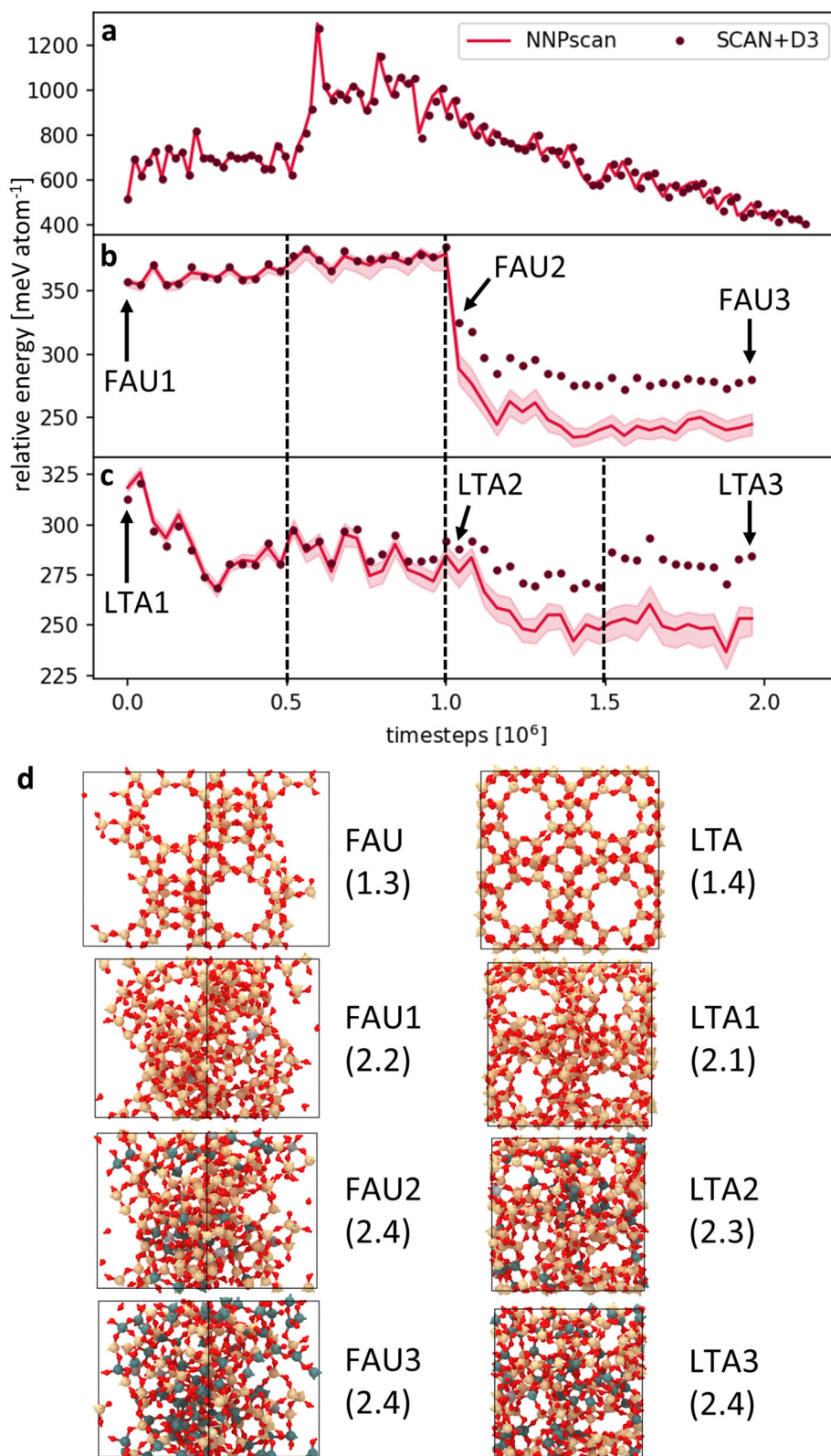


Fig. 6 Reactive silica phase transformations. Comparison of SCAN + D3 and NNPscan energies for **a** melting and annealing of β -cristobalite as well as amorphization of **b** FAU and **c** LTA by compression. Shaded areas in **a–c** show the energy range of all six NNPs (not visible in **a** due to energy scale) and dashed lines in **b** and **c** indicate compression steps. Structures and mass densities (g cm⁻³, in brackets) are depicted in **d**. Si: yellow, O: red, Si (after bond breaking): green.

compression to artificially high densities up to 10% beyond silica glass density resulted in NNP extrapolation. However, the difference between NNP and DFT energies was even in the extrapolation region at least three times lower ($<40 \text{ meV atom}^{-1}$) than the RMSEs of the other PES approximations (e.g., $136 \text{ meV atom}^{-1}$ for ReaxFF) shown in Table 1. In addition, the MD trajectories contain physically reasonable configurations allowing straightforward extension of the DFT dataset and further NNP refinement. Hence, these ZA simulations demonstrate that the SchNet NNPs are transferable and reasonably data-efficient interpolators of the silica PES as exemplified by qualitatively correct description of zeolite amorphization even slightly beyond the interpolation region.

The employed ZA simulation protocol only mimics the thermal zeolite collapse and does not provide realistic modeling of this phase transition. In fact, there are no reports of the thermal collapse for purely siliceous LTA and FAU. Most experimental studies on such phase transformations used Al-containing zeolites showing that zeolites with Si/Al ratios >4 are thermally very stable due to the higher energetic barrier for breaking Si-O bonds than Al-O bonds^{30,34}. Therefore, the ZA simulations required artificial compression to higher densities to obtain a higher degree of amorphization. However, even at lower densities ($<2.2 \text{ g cm}^{-3}$), the ZA simulations showed bond-breaking events in LTA without extrapolation and in agreement with DFT results. These results demonstrate that the NNPs also reliably interpolate reactive parts of the PES that are relevant for transformations between different zeolite structures.

In summary, the trained NNPs allow general and accurate modeling of siliceous zeolites with DFT accuracy. This includes modeling of thermodynamic stabilities, equilibrium properties as well as reactive and non-reactive phase transitions of zeolites by interpolation of all *relevant* parts of the PES. Even in the observed cases of extrapolation, the NNPs showed qualitative agreement with DFT results with energy errors far lower than analytical force fields demonstrating the robustness of SchNet NNPs that allows their straightforward refinement and extension by active learning. Thanks to active learning, the NNPs capture the structural diversity of zeolites that is used for reoptimization of the Deem database with high accuracy. The revised database provides vital input for future machine learning studies on structure-stability-property correlations facilitating the computational—in silico—design and discovery of zeolites. Finally, NNP extension for modeling zeolites containing heteroatoms such as Al or guest molecules such as water is a promising route towards realistic atomistic modeling of zeolites under synthesis and operating conditions^{29,76} with ab initio accuracy.

METHODS

Dataset generation

Generation of the initial DFT datasets used PBE + D3 single-point calculations applied to a diverse set of structures, including silica polymorphs, surface models, hypothetical, and existing zeolites. First, ten hypothetical zeolites were selected from the Deem database by Farthest Point Sampling (FPS)^{77,78} to find the most diverse subsample of atomic environments. The FPS employed the similarity distance metric $d(A, B)$ between two zeolites A, B calculated using the average similarity Kernel $\bar{K}(A, B)$ of the smooth overlap of atomic positions (SOAP)⁷⁹ descriptor (Supplementary methods)⁸⁰:

$$d(A, B) = \sqrt{2 - 2\bar{K}(A, B)}. \quad (1)$$

Apart from the ten selected zeolites, the FPS detected a hypothetical silica bilayer *in vacuo* (72 atoms, 12 Å vacuum layer), which was also added to the dataset (Supplementary Fig. 2c). Additionally, the dataset included a hypothetical α -quartz (001) surface model (120 atoms, 15 Å vacuum layer) terminated with dangling Si-O bonds. The dataset also contained five existing zeolites (CHA, SOD, IRR, MVY, MTF) and six silica polymorphs (α -

quartz, α -cristobalite, tridymite, moganite, coesite, and stishovite) for consideration realistic silica structures.

All selected configurations were optimized at the PBE + D3 level under zero pressure conditions. Next, 210 different unit cell deformations were applied to all optimized structures (Supplementary information). Further sampling of atomic environments close to the optimized configurations used ten ps MD equilibrations (ReaxFF level) at 600 K and 1200 K. The 200 most diverse structures were extracted from every MD trajectory by the FPS described above. The resulting set of structures constitutes the low-energy, close to equilibrium (EQ) part of the silica database (Supplementary Fig. 2).

Sampling of high-energy configurations and transition states used MD simulations (ReaxFF level) for melting and simulating annealing of β -cristobalite ($2 \times 2 \times 2$ supercell). After scaling its mass density from 2.3 to 2.2 g cm^{-3} (silica glass density) and geometry optimization, the structure was equilibrated for 100 ps at 6000 K. Next, the temperature was reduced to 3000 K in three steps along with an equilibration for 100 ps at each temperature step. The equilibration at 3000 K used additional 100 ps to improve the structural sampling. Again, FPS was applied to the MD trajectories to find the 1000 most diverse configurations. To generate low-energy amorphous structures, ten configurations from the 3000 K MD trajectory were optimized (quenched) at constant volume (PBE + D3 level). The lowest energy structure obtained was optimized under zero pressure conditions. Subsequently, the 210 lattice deformations used above (Supplementary information) were applied to the fully optimized unit cell. Structures generated from the simulated annealing of β -cristobalite are denoted as amorphous silica in the reference database (Supplementary Fig. 2).

In contrast to the melting of silica polymorphs, low-density zeolites show a significant volume contraction during melting, that is, thermal collapse. To mimic this process, eight hypothetical zeolites were equilibrated at 1200 K for 100 ps employing ReaxFF. Then, the unit cell volume was scaled stepwise such that after ten equivalent steps, the mass density of silica glass (2.2 g cm^{-3}) was reached. After each contraction step, the zeolites were equilibrated for 100 ps. FPS of the resulting trajectories located 1000 diverse structures for each zeolite that fall into the ZA category of the dataset (Supplementary Fig. 2).

Single-point calculations at the PBE + D3 level were applied to the initial database providing energies and forces for the training of an ensemble of six NNPs allowing their iterative refinement.

NNP refinement

Refinement of the initially trained NNP ensemble requires extrapolation detection for previously unseen configurations. This is achieved by performing simulations using one leading NNP and applying single-point calculations to the trajectory using the remaining five potentials³⁵. If the energy and force predictions deviate by $>10 \text{ meV atom}^{-1}$ or 750 meV \AA^{-1} , respectively, from the NNP ensemble average, additional PBE + D3 single-point calculations were added to the reference database. Simulations and re-training of the NNP ensemble were repeated until no extrapolation was detected during test simulations.

To enhance the structural diversity of the EQ dataset, the Deem ($>331k$ structures, www.hypotheticalzeolites.net, accessed: November 29, 2019) and IZA (235 fully connected frameworks) databases were optimized using constant (zero) pressure conditions. Additionally, β -cristobalite was equilibrated at 4800 K for 1 ns to sample more liquid silica configurations. Extension of the ZA dataset used the same computational protocol for the thermal collapse (up to 2.2 g cm^{-3}) of zeolites described above but for the frameworks LTA and SOD, which were not considered in the initial ZA dataset. The resulting PBE + D3 dataset contains about 33k structures with up to 400 atoms per unit cell. Single-point calculations at the SCAN + D3 level were also applied to the final database to train the NNPs. More details on the DFT database are summarized in the supplementary information (Supplementary Figs. 1 and 2).

Training of NNPs and NNPs used energy and forces of the databases calculated at the PBE + D3 and SCAN + D3 levels, respectively. The resulting test RMSE are $\sim 4.7 \text{ meV atom}^{-1}$ for energies and 147 meV \AA^{-1} for forces (Supplementary Table 1). These errors are about an order of magnitude lower compared to other methods approximating the PES of silica (*cf.* Section 2.2).

Test simulations

The final geometry optimization of the Deem and IZA database was performed at the NNPscan level. To test the NNP quality for reactive events, structures of the Stone-Wales defect formation were taken from ref. 59. The unit cell parameters were optimized at the NNPscan and NNPpbe level keeping the fractional coordinates and the vacuum layer frozen, followed by DFT single-point calculations using the optimized structures. MD simulations (timestep 1 fs, NNPscan level) of the thermal LTA and FAU collapse served as a test case of the final NNPs (FAU was not part of the final reference dataset). These simulations used the same procedure described above, but with 12 compression steps up to a mass density of 2.4 g cm^{-3} and an equilibration time of 500 ps between each step.

Test simulations (NNPscan level) for the annealing of amorphous silica used three different initial structures: β -cristobalite and two vitreous silica structures taken from the ReaxFF simulated annealing described above. Melting of β -cristobalite employed an equilibration for 1 ns at 4800 K. After geometry optimization, the two amorphous structures were equilibrated for 1 ns at 4200 K due to their lower energetic barrier for transition to the liquid state. In all three cases, the temperature was stepwise decreased to 2500 K in 100 K steps and an equilibration time of 25 ps per temperature step. The last structures of the MD trajectories were optimized under zero pressure conditions. The obtained glass configurations were equilibrated for 10 ps at 300 K (NVT ensemble), followed by another 10 ps equilibration using the NVE ensemble. Calculation of the VDOS used the velocity autocorrelation function from the NVE trajectory. In the case of α -cristobalite, the harmonic VDOS was calculated at the NNPscan and SCAN + D3 level using a $3 \times 3 \times 2$ supercell and the finite-difference (FD) approach. In addition, calculation of the anharmonic α -cristobalite VDOS at the NNPscan level employed MD simulations at 300 K with the same computational protocol used for the silica glass structures.

Computational details

DFT simulations at the GGA (PBE)⁸¹ and meta-GGA (SCAN)⁸² level employed the Vienna Ab initio Simulation Package (VASP, version 5.4.4)^{83–86} along with the Projector Augmented-Wave (PAW) method^{87,88}. Calculations at constant volume used a plane-wave energy cutoff of 400 eV while the constant pressure optimizations used cutoff of 800 eV. The \mathbf{k} -point grids had a linear density of at least one \mathbf{k} -point per 0.1 \AA^{-1} along the reciprocal lattice vectors. The consideration of long-range dispersion interactions is essential for accurate modeling of zeolites^{58,89,90}. However, it has been shown that accurate modelling of dispersion in porous materials can be challenging^{91,92} and that the SCAN functional in particular can exhibit non-systematic accuracy for description of dispersion in systems with variable sizes and densities⁹³. Therefore, we considered two types of dispersion corrections with both PBE and SCAN functionals, a simple semiempirical one proposed by Grimme et al. (D3)⁵⁶ (with Becke-Johnson damping)⁹⁴ and more involved density-dependent many-body dispersion (MBD)⁵⁷ correction, and we compared their performance with available experimental data for equilibrium structures and energies of siliceous zeolites (Supplementary Tables S2 and S3). The MBD correction was used with the vdW scaling parameters β of 0.84 for PBE and 1.12 for SCAN as optimized in ref. 93. We found that both dispersion corrections provide virtually the same quality with respect to experimental data on equilibrium structures and energies, similar to the results of a previous study (considering PBE functional only)⁵⁸. Therefore, we opted out for the computationally less demanding Grimme D3 dispersion correction for the dataset generation.

Training of SchNet⁴² NNPs employed the Python package SchNetPack⁴³ and random splits of the reference datasets into training, validation, and test sets at a ratio of 8:1:1 that showed lowest RMSEs for different tested split ratios (Supplementary Fig. 10). Mini-batch gradient descent optimization was applied for training along with a mini-batch size of eight structures and the ADAM optimizer⁹⁵. During NNP training the learning rate lowered stepwise (from 10^{-4} to 10^{-6}) by a factor of 0.5 if the validation loss shows no improvement after 15 epochs. We used the same squared loss function for energy and forces as in ref. 42 along with a trade-off factor of 0.01, that is, with high weight on force errors. The setup of the NNP hyper-parameters used six interaction blocks, 128-dimensional feature vectors, a cutoff radius 6 \AA and a grid of 60 Gaussians for expansion of pairwise distances as input for the filter generating networks. A similar training and hyper-parameter setup provided very good NNP accuracy and training performance in previous works^{42,43}.

Calculations with the trained NNPs employed the atomic simulation environment (ASE)⁹⁶. Simulations at the ReaxFF⁵⁴ level used the large-scale atomic/molecular massively parallel simulator (LAMMPS)^{97,98} and in the case of the Sanders-Leslie-Catlow (SLC) potential^{52,53} the General Utility Lattice Program (GULP)⁹⁹. GFNO-xTB⁵⁵ calculations were performed with the xTB program package (version 6.3.3, available at: <https://github.com/grimme-lab/xtb>).

Unless stated otherwise, all MD simulations used a time step of 0.5 fs and the canonical (NVT) ensemble with the Nosé-Hoover thermostat^{100,101}. Calculation of the harmonic VDOS at the SCAN + D3 and NNPscan level employed the finite-difference (frozen-phonon) approach implemented in Phonopy¹⁰² along with displacements of 0.02 \AA . The calculation of the VDOS and anharmonic vibrational frequencies from MD trajectories used the Python packages pwtools (available at: <https://github.com/elcorto/pwtools>) and DynaPhoPY¹⁰³, respectively. Calculations of the SOAP descriptor were performed with the Python package Dscribe¹⁰⁴.

DATA AVAILABILITY

The Deem and IZA database, the trained NNPs, and the test set used for accuracy evaluation is openly available in a Zenodo repository (<https://doi.org/10.5281/zenodo.5827897>). The Deem database contains >331k hypothetical zeolite frameworks geometrically optimized at the NNPscan level. The NNPscan optimized database of the International Zeolite Association contains 236 existing, fully connected zeolite frameworks. Both databases are SQLite database files of the Atomic Simulation Environment (ASE) containing the ASE Atoms objects with energies and forces calculated at the NNPscan level and are readable with ASE's I/O module. Additionally, the repository contains plain text data as csv files containing zeolite features such as relative energies and densities for both the Deem and IZA database. The remaining data for the reproduction of results is available upon reasonable request.

CODE AVAILABILITY

Source code files for reproduction of results are available upon reasonable request.

Received: 10 January 2022; Accepted: 3 August 2022;

Published online: 19 August 2022

REFERENCES

- Li, Y., Li, L. & Yu, J. Applications of zeolites in sustainable. *Chem. Chem.* **3**, 928–949 (2017).
- Akporiaye, D. E. & Price, G. D. Systematic enumeration of zeolite frameworks. *Zeolites* **9**, 23–32 (1989).
- Deem, M. W., Pophale, R., Cheeseman, P. A. & Earl, D. J. Computational discovery of new zeolite-like materials. *J. Phys. Chem. C* **113**, 21353–21360 (2009).
- Pophale, R., Cheeseman, P. A. & Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **13**, 12407 (2011).
- Blatov, V. A., Ilyushin, G. D. & Proserpio, D. M. The zeolite conundrum: why are there so many hypothetical zeolites and so few observed? A possible answer from the zeolite-type frameworks perceived as packings of tiles. *Chem. Mater.* **25**, 412–424 (2013).
- Baerlocher, C., Meier, W. M. & Olson, D. M. *Atlas of Zeolite Framework Types* (Elsevier, Amsterdam, 2001).
- Baerlocher, C. & McCusker, L. B. *Database of Zeolite Structures*. <http://www.iza-structure.org/databases/> (2020).
- Mazur, M. et al. Synthesis of 'unfeasible' zeolites. *Nat. Chem.* **8**, 58–62 (2016).
- Čejka, J., Morris, R. E., Nachtigall, P. & Roth, W. J. Layered inorganic solids. *Dalton Trans.* **43**, 10274 (2014).
- Eliášová, P. et al. The ADOR mechanism for the synthesis of new zeolites. *Chem. Soc. Rev.* **44**, 7177–7206 (2015).
- Firth, D. S. et al. Assembly–disassembly–organization–reassembly synthesis of zeolites based on *chf*-type layers. *Chem. Mater.* **29**, 5605–5611 (2017).
- Gatta, G. D. & Lee, Y. Zeolites at high pressure: a review. *Mineral. Mag.* **78**, 267–291 (2014).
- Jordá, J. L. et al. Synthesis of a novel zeolite through a pressure-induced reconstructive phase transition process. *Angew. Chem. Int. Ed.* **52**, 10458–10462 (2013).
- Thibaud, J.-M. et al. High-pressure phase transition, pore collapse, and amorphization in the siliceous 1D zeolite, TON. *J. Phys. Chem. C* **121**, 4283–4292 (2017).

15. Alberti, A., Cruciani, G. & Martucci, A. Reconstructive phase transitions induced by temperature in gmelinite-Na zeolite. *Am. Mineral.* **102**, 1727–1735 (2017).
16. Mazur, M. et al. Pressure-induced chemistry for the 2D to 3D transformation of zeolites. *J. Mater. Chem. A* **6**, 5255–5259 (2018).
17. Foster, M. D., Delgado Friedrichs, O., Bell, R. G., Almeida Paz, F. A. & Klinowski, J. Chemical evaluation of hypothetical uninodal zeolites. *J. Am. Chem. Soc.* **126**, 9769–9775 (2004).
18. Li, Y., Yu, J. & Xu, R. Criteria for zeolite frameworks realizable for target synthesis. *Angew. Chem. Int. Ed.* **52**, 1673–1677 (2013).
19. Zimmermann, N. E. R. & Haranczyk, M. History and utility of zeolite framework-type discovery from a data-science perspective. *Cryst. Growth Des.* **16**, 3043–3048 (2016).
20. Perez, J. L. S., Haranczyk, M. & Zimmermann, N. E. R. High-throughput assessment of hypothetical zeolite materials for their synthesizability and industrial deployability. *Z. Kristallogr. - Cryst. Mater.* **234**, 437–450 (2019).
21. Akporiaye, D. E. & Price, G. D. Relative stability of zeolite frameworks from calculated energetics of known and theoretical structures. *Zeolites* **9**, 321–328 (1989).
22. Henson, N. J., Cheetham, A. K. & Gale, J. D. Theoretical calculations on silica frameworks and their correlation with experiment. *Chem. Mater.* **6**, 1647–1650 (1994).
23. Helfrecht, B. A., Semino, R., Pireddu, G., Auerbach, S. M. & Ceriotti, M. A new kind of Atlas of zeolite building blocks. *J. Chem. Phys.* **151**, 154112 (2019).
24. Moliner, M., Román-Leshkov, Y. & Corma, A. Machine learning applied to zeolite synthesis: the missing link for realizing high-throughput discovery. *Acc. Chem. Res.* **52**, 2971–2980 (2019).
25. Schwalbe-Koda, D., Jensen, Z., Olivetti, E. & Gómez-Bombarelli, R. Graph similarity drives zeolite diffusionless transformations and intergrowth. *Nat. Mater.* **18**, 1177–1181 (2019).
26. Evans, J. D. & Coudert, F.-X. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem. Mater.* **29**, 7833–7839 (2017).
27. Gaillac, R., Chibani, S. & Coudert, F.-X. Speeding up discovery of auxetic zeolite frameworks by machine learning. *Chem. Mater.* **32**, 2653–2663 (2020).
28. Lee, S., Kim, B. & Kim, J. Predicting performance limits of methane gas storage in zeolites through an artificial neural network. *J. Mater. Chem. A* **7**, 2709–2716 (2019).
29. Grajciar, L. et al. Towards operando computational modeling in heterogeneous catalysis. *Chem. Soc. Rev.* **47**, 8307–8348 (2018).
30. Peral, I. & Íñiguez, J. Amorphization induced by pressure: results for zeolites and general implications. *Phys. Rev. Lett.* **97**, 225502 (2006).
31. Wondraczek, L. et al. Kinetics of decelerated melting. *Adv. Sci.* **5**, 1700850 (2018).
32. Greaves, G. N. et al. The rheology of collapsing zeolites amorphized by temperature and pressure. *Nat. Mater.* **2**, 622–629 (2003).
33. Greaves, G. N. et al. Zeolite collapse and polyamorphism. *J. Phys. Condens. Matter* **19**, 415102 (2007).
34. Cruciani, G. Zeolites upon heating: factors governing their thermal stability and structural changes. *J. Phys. Chem. Solids* **67**, 1973–1994 (2006).
35. Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.* **56**, 12828–12840 (2017).
36. Mueller, T., Hernandez, A. & Wang, C. Machine learning for interatomic potential models. *J. Chem. Phys.* **152**, 050902 (2020).
37. von Lilienfeld, O. A., Müller, K.-R. & Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **4**, 347–358 (2020).
38. Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
39. Keith, J. A. et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **121**, 9816–9872 (2021).
40. Lubbers, N., Smith, J. S. & Barros, K. Hierarchical modeling of molecular energies using a deep neural network. *J. Chem. Phys.* **148**, 241715 (2018).
41. Thomas, N. et al. Tensor field networks: rotation- and translation-equivariant neural networks for 3D point clouds. Preprint at <http://arxiv.org/abs/1802.08219> (2018).
42. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet—A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
43. Schütt, K. T. et al. SchNetPack: A deep learning toolbox for atomistic systems. *J. Chem. Theory Comput.* **15**, 448–455 (2019).
44. Novoselov, I. I., Yanilkina, A. V., Shapeev, A. V. & Podryabinkin, E. V. Moment tensor potentials as a promising tool to study diffusion processes. *Comput. Mater. Sci.* **164**, 46–56 (2019).
45. Rosenbrock, C. W. et al. Machine-learned interatomic potentials for alloys and alloy phase diagrams. *npj Comput. Mater.* **7**, 1–9 (2021).
46. Sivaraman, G. et al. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *npj Comput. Mater.* **6**, 104 (2020).
47. George, J., Hautier, G., Bartók, A. P., Csányi, G. & Deringer, V. L. Combining phonon accuracy with high transferability in Gaussian approximation potential models. *J. Chem. Phys.* **153**, 044104 (2020).
48. Li, W. & Ando, Y. Comparison of different machine learning models for the prediction of forces in copper and silicon dioxide. *Phys. Chem. Chem. Phys.* **20**, 30006–30020 (2018).
49. Liu, H., Fu, Z., Li, Y., Sabri, N. F. A. & Bauchy, M. Parameterization of empirical forcefields for glassy silica using machine learning. *MRS Commun.* **9**, 593–599 (2019).
50. Erhard, L. C., Rohrer, J., Albe, K. & Deringer, V. L. A machine-learned interatomic potential for silica and its relation to empirical models. *npj Comput. Mater.* **8**, 1–12 (2022).
51. Schran, C., Brezina, K. & Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. *J. Chem. Phys.* **153**, 104105 (2020).
52. Sanders, M. J., Leslie, M. & Catlow, C. R. A. Interatomic potentials for SiO₂. *J. Chem. Soc. J. Chem. Soc.* **19**, 1271–1273 (1984).
53. Schröder, K.-P. et al. Bridging hydroxyl groups in zeolitic catalysts: a computer simulation of their structure, vibrational properties and acidity in protonated faujasites (H-Y zeolites). *Chem. Phys. Lett.* **188**, 320–325 (1992).
54. Fogarty, J. C., Aktulga, H. M., Grama, A. Y., van Duin, A. C. T. & Pandit, S. A. A reactive molecular dynamics simulation of the silica-water interface. *J. Chem. Phys.* **132**, 174704 (2010).
55. Bannwarth, C. et al. Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* **11**, e1493 (2021).
56. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
57. Tkatchenko, A., DiStasio, R. A., Car, R. & Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **108**, 236402 (2012).
58. Fischer, M., Kim, W. J., Badawi, M. & Lebègue, S. Benchmarking the performance of approximate van der Waals methods for the structural and energetic properties of SiO₂ and AlPO₄ frameworks. *J. Chem. Phys.* **150**, 094102 (2019).
59. Klemm, H. W. et al. A silica bilayer supported on Ru(0001): following the crystalline-to-vitreous transformation in real time with spectro-microscopy. *Angew. Chem. Int. Ed.* **59**, 10587–10593 (2020).
60. Brandenburg, J. G., Bates, J. E., Sun, J. & Perdew, J. P. Benchmark tests of a strongly constrained semilocal functional with a long-range dispersion correction. *Phys. Rev. B* **94**, 115144 (2016).
61. Bertels, L. W., Newcomb, L. B., Alaghamandi, M., Green, J. R. & Head-Gordon, M. Benchmarking the performance of the ReaxFF reactive force field on hydrogen combustion systems. *J. Phys. Chem. A* **124**, 5631–5645 (2020).
62. Zheng, N. Microporous and photoluminescent chalcogenide zeolite analogs. *Science* **298**, 2366–2369 (2002).
63. Armstrong, J. A. & Weller, M. T. Beryllosilicate frameworks and zeolites. *J. Am. Chem. Soc.* **132**, 15679–15686 (2010).
64. Piccione, P. M. et al. Thermochemistry of pure-silica zeolites. *J. Phys. Chem. B* **104**, 10001–10011 (2000).
65. Wragg, D. S., Morris, R. E. & Burton, A. W. Pure silica zeolite-type frameworks: a structural analysis. *Chem. Mater.* **20**, 1561–1570 (2008).
66. Wehinger, B. et al. Lattice dynamics of α -cristobalite and the Boson peak in silica glass. *J. Phys.: Condens. Matter* **27**, 305401 (2015).
67. Bates, J. B. Raman spectra of α and β cristobalite. *J. Chem. Phys.* **57**, 4042–4047 (1972).
68. Swainson, I. P., Dove, M. T. & Palmer, D. C. Infrared and Raman spectroscopy studies of the α - β phase transition in cristobalite. *Phys. Chem. Miner.* **30**, 353–365 (2003).
69. Sigaev, V. N. et al. Low-frequency band at 50 cm⁻¹ in the Raman spectrum of cristobalite: identification of similar structural motifs in glasses and crystals of similar composition. *J. Non-Cryst. Solids* **248**, 141–146 (1999).
70. Coh, S. & Vanderbilt, D. Structural stability and lattice dynamics of SiO₂ cristobalite. *Phys. Rev. B* **78**, 054117 (2008).
71. Buchenau, U. et al. Low-frequency modes in vitreous silica. *Phys. Rev. B* **34**, 5665–5673 (1986).
72. Guillot, B. & Guissani, Y. Boson peak and high frequency modes in amorphous silica. *Phys. Rev. Lett.* **78**, 2401–2404 (1997).
73. Carpenter, J. M. & Price, D. L. Correlated motions in glasses studied by coherent inelastic neutron scattering. *Phys. Rev. Lett.* **54**, 441–443 (1985).
74. Tong, Z., Yang, X., Feng, T., Bao, H. & Ruan, X. First-principles predictions of temperature-dependent infrared dielectric function of polar materials by including four-phonon scattering and phonon frequency shift. *Phys. Rev. B* **101**, 125416 (2020).
75. Zhang, D.-B., Sun, T. & Wentzcovitch, R. M. Phonon quasiparticles and anharmonic free energy in complex systems. *Phys. Rev. Lett.* **112**, 058501 (2014).

76. Saha, I., Erlebach, A., Nachtigall, P., Heard, C. J. & Grajciar, L. Reactive neural network potential for aluminosilicate zeolites and water: Quantifying the effect of Si/Al ratio on proton solvation and water diffusion in H-FAU. Preprint at <https://doi.org/10.26434/chemrxiv-2022-d1sj9> (2022).
77. Eldar, Y., Lindenbaum, M., Porat, M. & Zeevi, Y. Y. The farthest point strategy for progressive image sampling. *IEEE Trans. Image Process.* **6**, 1305–1315 (1997).
78. Imbalzano, G. et al. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *J. Chem. Phys.* **148**, 241730 (2018).
79. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
80. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
81. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
82. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
83. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
84. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).
85. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
86. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
87. Blöchl, P. E. Projector Augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
88. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
89. Hay, H., Ferlat, G., Casula, M., Seitsonen, A. P. & Mauri, F. Dispersion effects in SiO₂ polymorphs: an ab initio study. *Phys. Rev. B* **92**, 144111 (2015).
90. Rehak, F. R., Piccini, G., Alessio, M. & Sauer, J. Including dispersion in density functional theory for adsorption on flat oxide surfaces, in metal-organic frameworks and in acidic zeolites. *Phys. Chem. Chem. Phys.* **22**, 7577–7585 (2020).
91. Liu, X., Hermann, J. & Tkatchenko, A. Communication: many-body stabilization of non-covalent interactions: structure, stability, and mechanics of Ag₃Co(CN)₆ framework. *J. Chem. Phys.* **145**, 241101 (2016).
92. Wieme, J., Lejaeghere, K., Kresse, G. & Van Speybroeck, V. Tuning the balance between dispersion and entropy to design temperature-responsive flexible metal-organic frameworks. *Nat. Commun.* **9**, 4899 (2018).
93. Hermann, J. & Tkatchenko, A. Electronic exchange and correlation in van der Waals systems: balancing semilocal and nonlocal energy contributions. *J. Chem. Theory Comput.* **14**, 1361–1369 (2018).
94. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
95. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations (ICLR)*, <http://arxiv.org/abs/1412.6980> (2015).
96. Larsen, A. H. et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
97. Plimpton, S. Fast parallel algorithms for short-range molecular-dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
98. Aktulga, H. M., Fogarty, J. C., Pandit, S. A. & Grama, A. Y. Parallel reactive molecular dynamics: numerical methods and algorithmic techniques. *Parallel Comput.* **38**, 245–259 (2012).
99. Gale, J. D. & Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **29**, 291–341 (2003).
100. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
101. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
102. Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
103. Carreras, A., Togo, A. & Tanaka, I. DynaPhoPy: A code for extracting phonon quasiparticles from molecular dynamics simulations. *Comput. Phys. Commun.* **221**, 221–234 (2017).
104. Himanen, L. et al. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).

ACKNOWLEDGEMENTS

The authors acknowledge Charles University Centre of Advanced Materials (CUCAM) (OP VVV Excellent Research Teams, project number CZ.02.1.01/0.0/0.0/15_003/0000417), the support of Primus Research Program of the Charles University (PRIMUS/20/SCI/004), and of the Czech Science Foundation (20-26767Y). P.N. acknowledges the Czech Science Foundation (grant No. 19-215345). This work was supported by The Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90140).

AUTHOR CONTRIBUTIONS

L.G. and P.N. acquired funding. L.G. and A.E. conceptualized and administered the project. A.E. generated, curated and analyzed computational data. L.G. helped with data analysis. A.E. prepared all the article graphics. L.G. supervised the investigation. A.E. wrote the original article draft. L.G., P.N. and A.E. carried out the manuscript review and editing.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-022-00865-w>.

Correspondence and requests for materials should be addressed to Lukáš Grajciar.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022