

## ARTICLE OPEN



## Crystal structure prediction at finite temperatures

Ivan A. Kruglov<sup>1,2,3</sup>, Alexey V. Yanilkin<sup>1,2</sup>, Yana Propad<sup>1</sup>, Arslan B. Mazitov<sup>1,2</sup>, Pavel Rachitskii<sup>1</sup> and Artem R. Oganov<sup>4</sup>

Crystal structure prediction is a central problem of crystallography and materials science, which until mid-2000s was considered intractable. Several methods, based on either energy landscape exploration or, more commonly, global optimization, largely solved this problem and enabled fully non-empirical computational materials discovery. A major shortcoming is that, to avoid expensive calculations of the entropy, crystal structure prediction was done at zero Kelvin, reducing to the search for the global minimum of the enthalpy rather than the free energy. As a consequence, high-temperature phases (especially those which are not quenchable to zero temperature) could be missed. Here we develop an accurate and affordable solution, enabling crystal structure prediction at finite temperatures. Structure relaxation and fully anharmonic free energy calculations are done by molecular dynamics with a forcefield (which can be anything from a parametric forcefield for simpler cases to a trained on-the-fly machine learning interatomic potential), the errors of which are corrected using thermodynamic perturbation theory to yield accurate results with full ab initio accuracy. We illustrate this method by applications to metals (probing the  $P$ - $T$  phase diagram of Al and Fe), a refractory covalent solid (WB), an Earth-forming silicate  $\text{MgSiO}_3$  (at pressures and temperatures of the Earth's lower mantle), and ceramic oxide  $\text{HfO}_2$ .

*npj Computational Materials* (2023)9:197; <https://doi.org/10.1038/s41524-023-01120-6>

## INTRODUCTION

The search for a stable crystal structure is a very challenging task, amounting to finding the structure with the lowest free energy among an astronomical number of possible structures, for each of which the free energy has to be computed. Today, this can only be done at zero Kelvin<sup>1–5</sup>: relaxing crystal structures and computing their free energies at finite temperatures would require extensive statistical sampling, increasing the number of configurations by several orders of magnitude. The challenge of crystal structure prediction at finite temperatures must be addressed because high-temperature phases often possess interesting properties and are of fundamental and practical interest. Such elements as titanium, iron, calcium, as well as many compounds (for example, cubic paraelectric  $\text{BaTiO}_3$  or ultrahard borides, carbides, and nitrides) have high-temperature phases which have no stability fields at zero temperature and some may not even be metastably quenched at low temperatures. Another important example is planetary science: properties of planet-forming materials at extreme conditions of planetary interiors are responsible for the observed seismic profiles and dynamical processes occurring within planets. For example, the exact structure of iron in the Earth's core<sup>6–10</sup> and exact location of the perovskite–post-perovskite  $\text{MgSiO}_3$  boundary at conditions of the lower mantle<sup>11–15</sup> are still under debate.

Here we develop a new method, T-USPEX, extending crystal structure prediction beyond the zero Kelvin regime. It is based on a previously developed evolutionary crystal structure prediction method USPEX<sup>1–3</sup>, which predicts the lowest-energy crystal structure for a given chemical composition. Due to the need for statistical sampling in the calculation of the free energy, the computing costs of T-USPEX may increase by several, roughly 3–5, orders of (in comparison with USPEX). However, this can be offset when taking advantage of two factors.

First, the energy landscape is dramatically simplified as temperature increases, greatly simplifying the combinatorial

problem. As a consequence, the global minimum is found in fewer generations. This was confirmed, e.g., by results on molecular crystals in ref. <sup>16</sup>, and our tests for  $\text{MgSiO}_3$  system at 110 GPa and different temperatures (Supplementary Fig. 1 and chapter “Free energy surface calculations” in Supplementary Materials). This decrease must be very fast (at zero temperature, the number of minima  $C$  is very large, let's say,  $\gg 10^6$  for medium-size systems, whereas near the melting temperature there is only one or very few local minima). This has led Oganov<sup>17</sup> to conjecture exponential decrease of the number of local minima  $C$ :

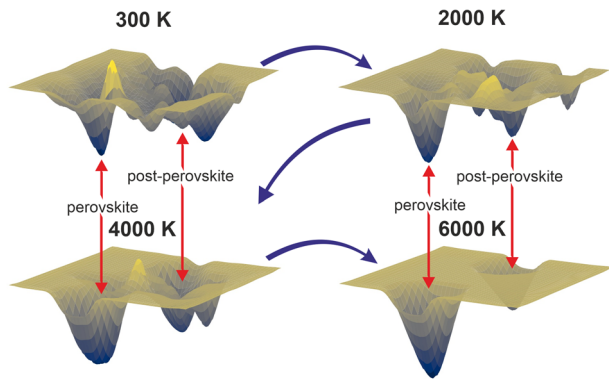
$$C = \exp\left(\alpha \frac{T^* - T}{T^*}\right), \quad (1)$$

where  $\alpha$  is a constant, and  $T^*$  is a characteristic temperature (higher than the melting temperature), at which only one free energy minimum exists. We have verified this conjecture for  $\text{MgSiO}_3$  at 110 GPa, where fitted values of the parameters were:  $\alpha = 2.97$ ,  $T^* = 7935\text{K}$  (see Supplementary Materials), while the melting temperature of  $\text{MgSiO}_3$  at such conditions is about  $6000\text{K}$ <sup>18</sup>. Such rapid decrease of the number of minima (Fig. 1) makes the problem much simpler: the number of structure relaxations and free energy calculations needed to find the global minimum decreases strongly.

Second, each structure relaxation and free energy calculation can be made much cheaper if one replaces ab initio calculations with a machine learning (ML) forcefield<sup>19–28</sup>, which speeds up the calculations by at least two orders of magnitude<sup>29</sup> with only a small reduction of accuracy (usual mean average errors are  $\sim 10\text{meV/atom}$ <sup>29</sup>). Moreover, with ML one can consider large supercells, thus considering long-wavelength phonons, which are so important for thermodynamics. ML forcefields have been applied to many different problems<sup>19–23,30–35</sup>, for a comparison of different ML approaches, see ref. <sup>36</sup>. ML forcefields were used to build the phase diagram of uranium<sup>34</sup>, accelerate crystal structure prediction<sup>29</sup>, calculate thermal conductivity<sup>37</sup>, lattice dynamics<sup>38</sup> and so on. Here we mostly use the MTP potential<sup>39</sup>, which showed

<sup>1</sup>Moscow Institute of Physics and Technology, 9 Institutsky lane, Dolgoprudny 141700, Russia. <sup>2</sup>Dukhov Research Institute of Automatics (VNIIA), Moscow 127055, Russia.

<sup>3</sup>Emerging Technologies Research Center, XPANCEO, Dubai Investment Park First, Dubai, United Arab Emirates. <sup>4</sup>Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 30-1 Bolshoy blvd, Moscow 121205, Russia. ✉email: [ivan.kruglov@phystech.edu](mailto:ivan.kruglov@phystech.edu); [a.oganov@skoltech.ru](mailto:a.oganov@skoltech.ru)



**Fig. 1 Free energy surfaces of MgSiO<sub>3</sub> calculated with T-USPEX.** Free energy surfaces were calculated at 110 GPa and 300 K, 2000 K, 4000 K, and 6000 K. Positions of perovskite and post-perovskite MgSiO<sub>3</sub> phases are shown with red arrows.

outstanding performance<sup>36</sup>. Structure relaxation at given  $P$ – $T$  conditions is equivalent to taking statistical averages for lattice translation vectors and for atomic fractional coordinates, while Gibbs free energy is calculated for each structure as:

$$G = PV + F = PV + F_0 + \int_0^1 (U(\lambda) - U_0) d\lambda, \quad (2)$$

where  $P$  is the external pressure,  $V$  the equilibrium volume, and  $F$  is the Helmholtz free energy which is computed by thermodynamic integration using adiabatic switching from a reference system with known free energy  $F_0$  and potential  $U_0$  (here we use the Einstein crystal as reference system) to the system with potential  $U(\lambda = 1)$  whose free energy we wish to compute<sup>40,41</sup>.

It is very important that we correct the errors of the ML potential to obtain free energies equivalent to full ab initio values  $F_{AI}$  using thermodynamic perturbation theory, to second order<sup>42</sup>:

$$F_{AI} \simeq F + \frac{1}{N_{at}} \left[ \langle U_{AI} - U \rangle - \frac{1}{2k_B T} \langle [U_{AI} - U]^2 \rangle \right], \quad (3)$$

where  $U_{AI}$  is the total energy of the ab initio system,  $k_B$  and  $T$  are Boltzmann constant and temperature, respectively, and angle brackets denote ensemble averages. Free energy calculations (2) are done on large supercells with  $\sim 10,000$  atoms, in the  $NVT$  ensemble (typical duration of a run is 40 ps). Free energy corrections (3) are done on smaller  $\sim 60$ -atom supercells (this required a classical  $NVT$  run of 10 ps duration, where DFT energies and ensemble averages in (3) were computed for 100 equally spaced snapshots). Convergence of values of free energies corrections is documented in Supplementary Materials section 8).

To summarize, crystal structure prediction at finite temperatures is difficult because for each candidate structure extensive sampling is needed to compute the entropy. However, two factors were identified that give hope—the reduction of the complexity of the free energy surface on increasing temperature, and the possibility of computing highly precise (to within 1 meV/atom) free energies in an economic way via ML forcefield using thermodynamic integration and thermodynamic perturbation theory corrections. This allowed us to develop an efficient and reliable approach for crystal structure prediction at finite temperatures, and we illustrate its performance on five substances—aluminum (calculating its phase diagram and searching for possibly missed unknown phases), iron (to figure out the stable crystal structure at conditions of the Earth’s inner core), MgSiO<sub>3</sub> (calculating its important phase diagram and searching for possibly overlooked phases at conditions of the lowermost mantle of the Earth), WB (testing our approach for a system with complex chemical bonding and predicting stability of a temperature-

induced phase), and HfO<sub>2</sub> (studying its high-temperature phases and looking for potential ferroelectric phases).

## RESULTS

### T-USPEX scheme

The scheme of the T-USPEX algorithm is shown in Fig. 2. The initial population of structures was created using symmetric<sup>5</sup> and topological<sup>43</sup> random structure generators. Each produced structure was then relaxed, first using density functional theory (DFT) calculations at zero Kelvin, and then using a forcefield-based molecular dynamics in the  $NPT$  ensemble for a relatively small  $\sim 60$ -atom supercell (types of interatomic potentials, used in this work, will be discussed below in the section “Interatomic potentials”). After a preliminary forcefield-based  $NPT$ -MD run, we obtained equilibrium lattice vectors. Then, for equilibrium lattice vectors averaged atomic coordinates were calculated from  $NVE$ -MD run.

Even the best ML forcefields give rise to non-negligible errors in pressure estimations, and this will affect the densities of the final structures and the  $PV$ -term in the free energy. For this reason, we calculate the pressure correction: we perform a molecular dynamics simulation in the  $NVE$  ensemble using the forcefield for 20 ps, and intermediate structures are saved every 1 ps—for each of which we calculate the pressure using DFT, and pressure correction is calculated as an ensemble average.

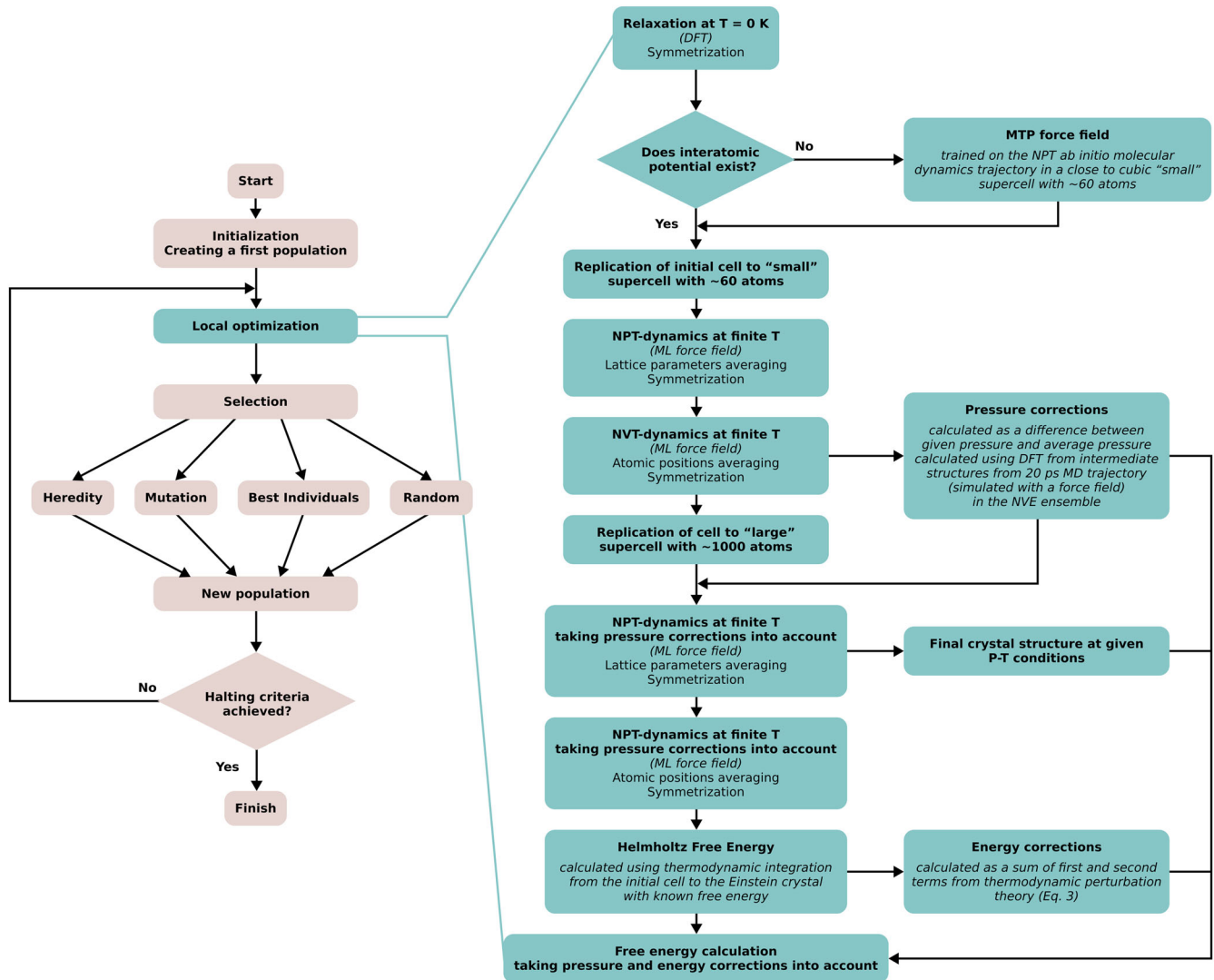
Then we created a large,  $\sim 1000$ -atom supercell and relaxed the structure using  $NPT$  (for 40 ps) followed by  $NVE$  (for 20 ps) molecular dynamics, taking pressure correction into account. This gave us final averaged supercell vectors and fractional coordinates of each atom. From them, using ASE<sup>44</sup> and Spglib<sup>45</sup> libraries, we obtained space group, primitive cell vectors and all atomic coordinates of the final structure. Structures with  $P1$  symmetry and unusually large unit cells were counted as defective and were discarded.

For the remaining structures, we proceeded to compute Gibbs free energy. First, using  $\sim 10,000$ -atom supercells, we computed the Helmholtz free energy using thermodynamic integration as the most accurate method<sup>46</sup>. To obtain accurate Gibbs free energies, these Helmholtz free energies were combined with the  $PV$ -term and 1st- and 2nd-order corrections of thermodynamic perturbation theory. To determine these corrections, we ran  $NVE$ -MD using the same forcefield, and took statistically independent (separated by  $\sim 1$  ps) snapshots of the MD trajectory, at which we computed DFT energy. While the 1st-order correction is quite large, the 2nd-order correction in all cases here is tiny, indicating extremely high reliability of the final free energies and robustness of our procedure.

Next, we test our method on five very different systems—aluminum, iron, WB, MgSiO<sub>3</sub>, and HfO<sub>2</sub>—in order to check the possible presence of unknown stable phases and refine phase diagrams.

### Phase diagram of aluminum

For aluminum, three phases are experimentally known to be stable at different  $P$ – $T$  conditions:  $fcc$  (at normal conditions),  $hcp$  and  $bcc$ . First, using DFT we determined that at zero Kelvin  $fcc$ -Al transforms to  $hcp$ -Al at 162 GPa (Supplementary Fig. 1), and  $hcp$ -Al transforms to  $bcc$ -Al at 373 GPa (Supplementary Fig. 2). These values agree well with previous theoretical calculations<sup>47–49</sup>, but in experiment  $fcc$ -Al transforms to  $hcp$ -Al at  $217 \pm 10$  GPa<sup>50</sup> at 300 K. Then, we ran T-USPEX calculations for Al at 0 GPa and 300 K, 100 GPa and 2000 K, 400 GPa and 2000 K. Each generation consisted of 20 randomly generated structures, and used the MTP potential trained on results of ab initio molecular dynamics (see “Methods”). We saw that during subsequent  $NPT$  molecular dynamics calculations (used for



**Fig. 2** The scheme of the T-USPEX algorithm. Left panel shows general workflow of the developed algorithm, and right panel shows process of crystal structure relaxation and free energy calculation.

relaxing structures at finite temperatures) almost all randomly generated structures transformed either to the *fcc*, *hcp* and/or *bcc* structures, or to defective structures with large unit cells (the latter structures were discarded). The transformation of a randomly generated AI structure to *bcc*-AI in a T-USPEX calculation is shown in Fig. 3. It is worth noting that at the first step the number of atoms in the unit cell decreased. This happens after symmetrization (these steps are shown in Fig. 2), when lattice vectors and the number of atoms in the unit cell are transformed to a standard setting corresponding to the space group with higher symmetry (with Spglib<sup>45</sup>).

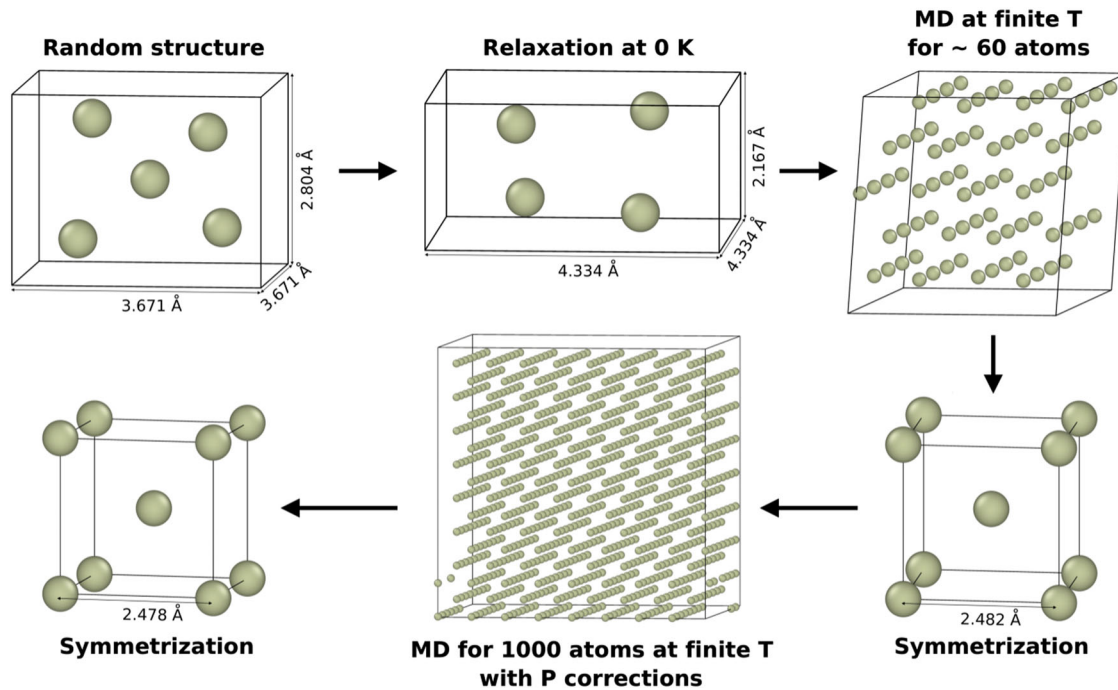
The model potential has two kinds of errors (relative to *ab initio* calculations)—a small error in density (which we compensate by a shift of pressure to bring pressure in exact agreement with *ab initio* result) and (at a given density) a small error in the free energy (which we correct for by thermodynamic perturbation theory). The pressure correction turned out to be several GPa at most (Supplementary Fig. 3); at the simulated pressure of 400 GPa it varied from 5 to 12 GPa. In the calculation at 0 GPa and 300 K 15 structures (out of 20) transformed to *fcc*-AI with free energies equal to  $(-3.8474 \pm 0.0002)$  eV/atom and densities  $(2.660 \pm 0.0044)$  g/cm<sup>3</sup>, 1 structure to *I4/mmm* phase with free energy of  $-3.7490$  eV/atom (clearly, this structure is not competitive), the other 4 structures

being defective supercells (Supplementary Table 1). We find that in the P–T region of interest (temperature up to 5000 K and pressures up to 400 GPa) only *fcc*, *hcp*, and *bcc* structures have stability fields.

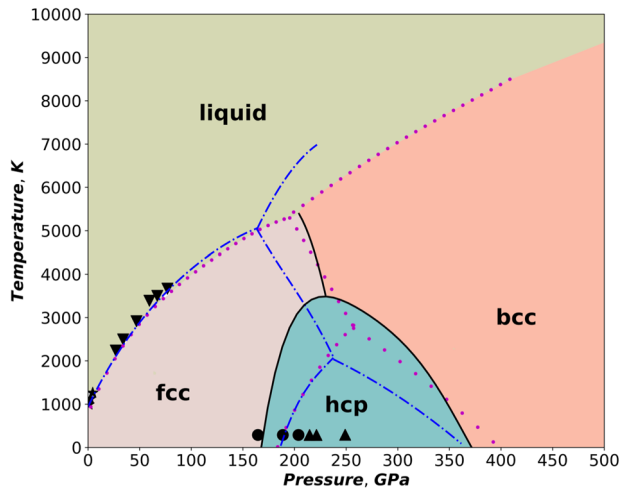
In order to calculate the P–T phase diagram, we used the *fcc*, *hcp*, and *bcc* structures as seeds and performed T-USPEX calculations close to the previously calculated phase transition boundaries<sup>47,49</sup>. These calculations can be viewed as computer experiments akin to real experiments sampling the phase diagram. The resulting P–T phase diagram for AI is presented in Fig. 4 (see also Supplementary Fig. 4). Clearly, the *hcp* phase occupies a much larger region of the phase diagram than in previous studies<sup>47–49</sup>. Our fully anharmonic result should be more reliable than previous calculations<sup>47–49</sup> based on the quasiharmonic approximation. The melting curve was taken from refs.<sup>47,49,51,52</sup>.

### Iron in the Earth's inner core

Until now, an open question is which phase of iron (*bcc* or *hcp* or perhaps some new structure) is stable at conditions of the Earth's solid inner core (pressures in the range 330–350 GPa and temperatures of approximately 5400–5700 K). Available experimental results are not fully consistent<sup>6–10</sup>. Computational studies



**Fig. 3** Relaxation of randomly generated Al structure in the T-USPEX algorithm. After a series of consequent MD relaxations and symmetrizations, random Al structure transforms to bcc-Al.



**Fig. 4** P-T phase diagram of MgSiO<sub>3</sub>. Vertical dashed line shows the pressure of the Earth's core–mantle boundary.

up to now either checked a few simple structures at relevant temperatures and pressures, or performed crystal structure prediction, but at zero Kelvin. Here, we perform crystal structure prediction at actual P–T conditions of the Earth's inner core, namely at the pressure of 330 GPa and temperatures of 5400 K, 5700 K and 6000 K. We trained a single ML forcefield for iron at each P–T point. An iterative approach was used to prepare a suitable training set. First, we generated random crystal structures of iron with up to 16 atoms in the unit cell at 330 GPa. Then, we collected new configurations for our training set from ab initio MD trajectories at 330 GPa and 5700 K using the generated random structures as initial configurations. The sampled configurations were pruned to remove identical structures, and their energies,

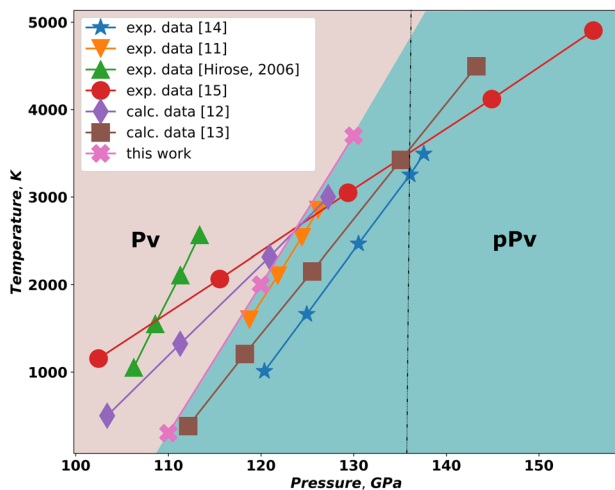
interatomic forces, and stresses were calculated within DFT. Finally, using active learning, the existing training set was updated with new data, and the interatomic potential was re-trained. This process was repeated iteratively until no new configurations were actively sampled from molecular dynamics runs. The resulting training set contained 494 configurations, and the ML forcefield had the mean absolute error (MAE) on energies prediction of 29 meV/atom, while the MAEs for forces and stresses were 0.43 eV/Å (1%) and 3.2 GPa (1.7%), respectively. The non-negligible MAE in energy prediction is the price for the generality of the forcefield in a wide range of configurations, but this error was canceled by corrections due to thermodynamic perturbation theory. On the other hand, high accuracy of interatomic forces and stresses prediction was essential for relaxations and free energy calculations.

Using T-USPEX and pre-trained MTP potential, we searched for stable Fe phase at 330 GPa and 5700 K. In the first generation consisting of 20 random structures, 3 structures relaxed into the *hcp* phase with the Gibbs free energy  $6.0652 \pm 0.0008$  eV/atom, 4—into *fcc* phase with the free energy of  $6.0859 \pm 0.0005$  eV/atom, 2—into newly found  $R\bar{3}m$  phase (which is a close-packed structure, its crystal structure parameters are shown in Supplementary Table 10) with free energy of  $6.0768 \pm 0.0019$  eV/atom (see Supplementary Table 3). We also found that at 330 GPa and 5400 K all initially random structures relaxed either to *hcp* or to *fcc* phase, and *hcp* phase had the lowest free energy (T-USPEX results are shown in Supplementary Table 4). Thus, our calculations showed that *hcp* structure is stable for iron at P–T conditions corresponding to the Earth's inner core, in agreement with results from ref. <sup>8</sup>. In order to check the stability of *bcc* phase, we used it as a seed in the T-USPEX calculation. At 330 GPa and all temperatures we studied (5400 K, 5700 K, 6000 K) it relaxed to the *hcp* phase. Thus, at our level of theory (PBE-DFT) we rule out the stability of the *bcc* phase of pure iron in the inner core of the Earth—however, impurities may still stabilize it<sup>9</sup>.

### Phase diagram of $\text{MgSiO}_3$ at the Earth's core–mantle boundary

Using T-USPEX, we explored the main Earth-forming compound— $\text{MgSiO}_3$  (comprising ~40 vol.% of our entire planet, or ~80 vol.% of its lower mantle)—searching for phases stable at conditions of the lowermost mantle. Throughout most of the lower mantle this compound exists as bridgmanite, a perovskite-type modification, while in the lowermost D' layer a post-perovskite polymorph is stable<sup>11,12</sup>. Here we followed a scheme somewhat different from what we used above, to illustrate the flexibility of our approach. For crystal structure relaxation and free energy calculations, we used the classical interatomic potential developed in ref. 53, which showed good agreement with experimental crystal structures and many properties. Here, using this interatomic potential, we ran T-USPEX for 20 generations (each containing 20 structures) at 135 GPa and 2700 K. Instead of applying expensive pressure and free energy corrections within crystal structure prediction runs, we ran highly efficient T-USPEX searches based solely on this interatomic potential, then selected structures with the lowest free energies and for these performed very accurate crystal structure and free energies calculations as described above.

In the T-USPEX search at 135 GPa and 4000 K, already in the first generation, we saw post-perovskite (pPv)  $Cmcm$ - $\text{MgSiO}_3$  as the most stable structure, and in the second generation—bridgmanite, the phase with  $Pnma$ -perovskite (Pv) structure, and no other competitive structures with up to 80 atoms in the primitive cell.



**Fig. 5** P–T phase diagram of  $\text{MgSiO}_3$ . Vertical dashed line shows the pressure of the Earth's core–mantle boundary.

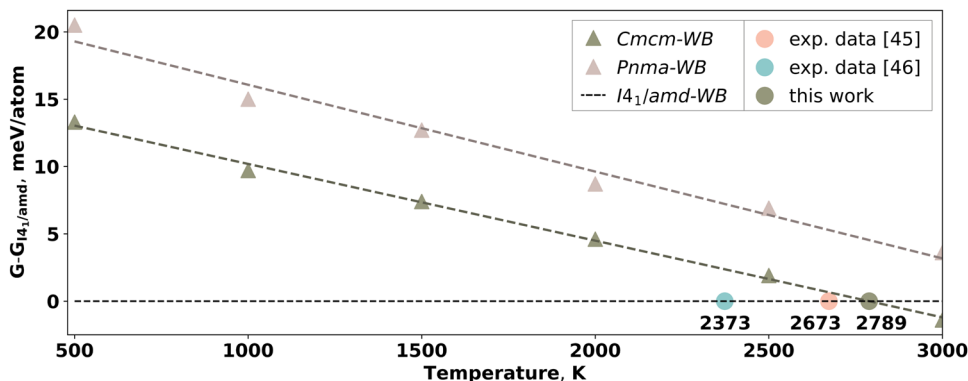
Then we obtained reliable structures and their free energies using pressure and free energy corrections as described above; this was done at 120 GPa and 1500 K, 2000 K and 2500 K; at 130 GPa and 500 K, 1500 K, 3000 K, 4000 K. At 120 GPa pPv phase was stable up to 2000 K, at 130 GPa—up to 3700 K. Using these values, the phase transition boundary between Pv and pPv  $\text{MgSiO}_3$  phases was calculated (Fig. 5). The result is right in the middle of the previous theoretical<sup>12,13</sup> and experimental<sup>11,14,15,43</sup> works (see Fig. 5) and has a positive Clapeyron slope of +5.88 MPa/K. This value is important for calculating the heat flow from the core into the mantle and for determining the thermal structure of the lowermost mantle<sup>19</sup>. Here, we give its most accurate theoretical calculation, with full account of the anharmonic terms in the free energy.

### Temperature-induced phase transition in WB

Next, we applied our method to WB - a hard refractory ultrahard compound, undergoing a temperature-induced phase transition and characterized by a nontrivial chemical bonding and lack of an accurate and transferable forcefield. The temperature-induced phase transition from the low-temperature  $\alpha$ -phase into the high-temperature  $\beta$ -phase takes place at ~2100–2700 K<sup>44–46</sup>. We ran T-USPEX at 0 GPa and 2000 K with MTP forcefield, and obtained both  $I4_1/amd$ -WB ( $\alpha$ -phase) and  $Cmcm$ -WB ( $\beta$ -phase) and a hitherto unknown low-energy  $Pnma$  structure of WB to have the lowest free energies (within a few meV/atom of each other, see Supplementary Table 2). In order to find stability fields of the discovered tungsten monoborides, their equilibrium crystal structures and free energies were calculated with higher accuracy in the temperature range of 300–3000 K at zero pressure. Our calculations show that the  $\alpha$ - $\beta$  transition takes place at 2789 K (Fig. 6), in agreement with experimental results<sup>45</sup>, while  $Pnma$ -WB remains a low-energy metastable phase at all temperatures.

### High-temperature phase transitions in $\text{HfO}_2$

We also applied T-USPEX to predict high-temperature phase transitions in a ceramic material -  $\text{HfO}_2$ . This material is particularly interesting due to the recently found ferroelectricity in its thin films<sup>54</sup>. Experimental studies<sup>55</sup> showed that at ambient pressure and room temperature, it adopts a monoclinic  $P2_1/c$  structure, which then transforms at ~2000 K to tetragonal  $P4_2/nmc$  phase and after at ~2800 K to cubic  $Fm\bar{3}m$  phase. At higher pressures,  $\text{HfO}_2$  transforms to  $Pbca$  and  $Pnma$  polymorphs. Yet, none of the experimentally found phases can explain ferroelectricity in  $\text{HfO}_2$ . Possible explanations include metastable  $Pca2_1$  or  $Pmn2_1$  structures<sup>56–60</sup>. Previously, machine learning interatomic potentials for  $\text{HfO}_2$  were constructed based on PBE<sup>61</sup> and PBEsol<sup>62</sup> exchange-



**Fig. 6** Thermodynamics of WB polymorphs. Here we show the calculated Gibbs free energies of  $Cmcm$  and  $Pnma$  structures of WB (relative to  $I4_1/amd$ ).

correlation functionals, yet for the closely related compound  $\text{ZrO}_2$  it was shown that  $\Delta$ -ML approach (based on the RPA approximation) gives the most accurate results for phase transition boundaries<sup>63</sup>. Here, using the T-USPEX method (within PBE and PBEsol approximations), we searched for stable  $\text{HfO}_2$  phases at 300 K, 1000 K and 2000 K. The results are shown in Supplementary Tables 5–9 and Supplementary Figs. 6 and 7.

Using T-USPEX (within the PBE approximation) at 300 K for  $\text{HfO}_2$  we found the experimentally known  $P2_1/c$  structure as the most stable one. We also found likely candidates for the ferroelectric phase:  $Pca2_1$  (~25 meV/atom above ground state) and  $Pmn2_1$  (~30 meV/atom above ground state). At higher temperatures (1000 K and 2000 K) we found anatase  $I4_1/amd$  phase to be the most stable one, and experimentally known  $P4_2/nmc$  phase to be metastable (detailed results are given in Supplementary Tables 5–7). Supplementary Fig. 6 shows that at  $T > 300$  K anatase  $I4_1/amd$  phase is the stable phase within PBE functional, which contradicts experiment. Given small free energy differences between phases, one can expect that small errors of the PBE exchange-correlation functional can lead to incorrect phase stability. It is known<sup>64</sup> that PBEsol is more accurate for  $\text{HfO}_2$ , so we decided to use it.

Doing T-USPEX calculations for  $\text{HfO}_2$  within PBEsol exchange-correlation functional (Supplementary Tables 8 and 9), we find  $P2_1/c$ - $\text{HfO}_2$  to be the most stable phase at temperatures up to about 2000 K (Supplementary Fig. 7), while at higher temperatures  $Fm\bar{3}m$ - $\text{HfO}_2$  is thermodynamically stable. The  $P4_2/nmc$ - $\text{HfO}_2$  phase is found to be metastable, while in experiment it has stability field between  $P2_1/c$  and  $Fm\bar{3}m$  phases. Anatase-type phase is no longer stable within PBEsol functional. The tiny free energy differences between these phases are calculated by T-USPEX with remarkable numerical robustness. Clearly (just as for zero-temperature crystal structure prediction) the global minimum depends on the choice of the exchange-correlation functional when energy differences are small.

## DISCUSSION

Until now, finite-temperature crystal structure prediction was thought to be prohibitively expensive. Here we showed a computationally efficient approach solving this problem, and illustrated its power by applying it to five very different substances: a simple metal (Al at high pressures and temperatures), a transition metal (Fe in Earth's core), a geologically important silicate ( $\text{MgSiO}_3$  at conditions of the lowermost mantle), a hard refractory compound WB with a temperature-induced phase transition and a ceramic oxide ( $\text{HfO}_2$ ). We find that the *hcp* phase of aluminum has a wider stability field than previously thought, and the temperature-induced  $\alpha$ - $\beta$  transition in WB occurs at 2789 K. It is also found that iron has *hcp* structure at conditions of the Earth's inner core, and the much debated (and important for constraining Earth's thermal structure) Clapeyron slope of the post-perovskite phase transition in  $\text{MgSiO}_3$  is 5.88 MPa/K. For  $\text{HfO}_2$  we find a number of competitive phases with small free energy differences, making the choice of exchange-correlation functional critical for correct predictions. We also show how this method can be used, in the spirit of a computational experiment akin to a real experiment, to probe phase diagrams, resolve long-standing disputes, and open new horizons in the search for unique materials with improved properties. Our method takes advantage of increasing simplification of the topology of the free energy landscape on increasing temperature, and of reliable machine learning forcefields and thermodynamic perturbation theory which allow highly accurate and affordable calculations of free energies. This development enables crystal structure prediction at any temperature for ordered materials. Further work is needed to extend T-USPEX approach to configurationally or magnetically disordered systems in a numerically robust and computationally affordable way.

## METHODS

### DFT calculation details

Crystal structure relaxations, energy calculations at 0 K and ab initio molecular dynamics were performed using density functional theory (DFT) as implemented in the VASP code<sup>65–67</sup> within Perdew–Burke–Ernzerhof (PBE)<sup>68</sup> functional. Projector-augmented wave (PAW)<sup>69</sup> method was used to describe core electrons and their interaction with valence electrons. The plane wave kinetic energy cutoff was set at 500 eV and  $\Gamma$ -centered k-points meshes with a resolution of  $2\pi \times 0.05 \text{ \AA}^{-1}$  were used for sampling the Brillouin zone. For metals, the electronic free energy was calculated and included in the total free energy. Ab initio molecular dynamics was performed in the NPT ensemble during 5 ps with 1 fs timestep. The convergence criterion for electronic step was set to  $10^{-6}$  eV. Molecular dynamics with classical or ML forcefields was performed using the LAMMPS code<sup>70</sup>.

### Interatomic potentials

The choice of interatomic potential plays an important role in T-USPEX calculations. If, for a given system, a good classical forcefield is known, then it can be applied in molecular dynamics simulations (as we did for  $\text{MgSiO}_3$ ). However, for most systems there is no classical forcefield which can work in a vast configurational space (created by random structures from T-USPEX). In this case, we use a machine learning forcefield trained on ab initio data. We used MTP machine learning forcefield<sup>23</sup>, due to its outstanding performance<sup>33</sup> and ability to perform active learning<sup>27</sup>, i.e., re-train the forcefield whenever a structure different from its original training set is encountered. We considered MTP potential to be reliable if the error on energies was about few meV/atom and the error on forces not more than 10%. For Fe, we trained one ML forcefield and then applied it to all structures. For the other cases (Al, WB,  $\text{HfO}_2$ ), we separately trained a ML forcefield for each structure at a given P–T-point, which allowed us to check and confirm numerical stability of our procedure.

### DATA AVAILABILITY

Data will be available on request.

### CODE AVAILABILITY

The code will be merged into the next release of USPEX (<https://uspex-team.org/>).

Received: 1 February 2023; Accepted: 29 August 2023;  
Published online: 21 October 2023

## REFERENCES

1. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX—evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713–720 (2006).
2. Oganov, A. R., Lyakhov, A. O. & Valle, M. How evolutionary crystal structure prediction works—and why. *Acc. Chem. Res.* **44**, 227–237 (2011).
3. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
4. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201 (2011).
5. Wang, Y., Lv, J., Zhu, L. & Ma, Y. CALYPSO: a method for crystal structure prediction. *Comput. Phys. Commun.* **183**, 2063–2070 (2012).
6. Saxena, S. K. et al. Synchrotron x-ray study of iron at high pressure and temperature. *Science* **269**, 1703–1704 (1995).
7. Belonoshko, A. B., Ahuja, R. & Johansson, B. Stability of the body-centred-cubic phase of iron in the Earth's inner core. *Nature* **424**, 1032–1034 (2003).
8. Tateno, S., Hirose, K., Ohishi, Y. & Tatsumi, Y. The structure of iron in earth's inner core. *Science* **330**, 359–361 (2010).
9. Vočadlo, L. et al. Possible thermal and chemical stabilization of body-centered-cubic iron in the Earth's core. *Nature* **424**, 536–539 (2003).

10. Hrubak, R., Meng, Y. & Shen, G. Experimental evidence of a body centered cubic iron at the Earth's core condition. Preprint at <https://arxiv.org/ftp/arxiv/papers/1804/1804.05109.pdf> (2018).
11. Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y. Post-perovskite phase transition in MgSiO<sub>3</sub>. *Science* **304**, 855–858 (2004).
12. Oganov, A. R. & Ono, S. Theoretical and experimental evidence for a post-perovskite phase of MgSiO<sub>3</sub> in Earth's D'' layer. *Nature* **430**, 445–448 (2004).
13. Tsuchiya, T., Tsuchiya, J., Umemoto, K. & Wentzcovitch, R. M. Phase transition in MgSiO<sub>3</sub> perovskite in the earth's lower mantle. *Earth Planet Sci. Lett.* **224**, 241–248 (2004).
14. Ono, S. & Oganov, A. R. In situ observations of phase transition between perovskite and CaIrO<sub>3</sub>-type phase in MgSiO<sub>3</sub> and pyrolitic mantle composition. *Earth Planet Sci. Lett.* **236**, 914–932 (2005).
15. Tateno, S., Hirose, K., Sata, N. & Ohishi, Y. Determination of post-perovskite phase transition boundary up to 4400 K and implications for thermal structure in D'' layer. *Earth Planet Sci. Lett.* **277**, 130–136 (2009).
16. Francia, N. F., Price, L. S., Nyman, J., Price, S. L. & Salvalaglio, M. Systematic finite-temperature reduction of crystal energy landscapes. *Cryst. Growth Des.* **20**, 6847–6862 (2020).
17. Oganov, A. R. Crystal structure prediction: reflections on present status and challenges. *Faraday Discuss.* **211**, 643–660 (2018).
18. Belonoshko, A. B. et al. High-pressure melting of MgSiO<sub>3</sub>. *Phys. Rev. Lett.* **94**, 195701 (2005).
19. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
20. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
21. Shapeev, A. V. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
22. Li, Z., Kermode, J. R. & de Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
23. Kruglov, I., Sergeev, O., Yanilkin, A. & Oganov, A. R. Energy-free machine learning force field for aluminum. *Sci. Rep.* **7**, 8512 (2017).
24. Byggmästar, J., Hamedani, A., Nordlund, K. & Djurabekova, F. Machine-learning interatomic potential for radiation damage and defects in tungsten. *Phys. Rev. B* **100**, 144105 (2019).
25. Dragoni, D., Daff, T. D., Csányi, G. & Marzari, N. Achieving DFT accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron. *Phys. Rev. Mater.* **2**, 013808 (2018).
26. Liu, Ybin et al. Machine learning interatomic potential developed for molecular simulations on thermal properties of β-Ga<sub>2</sub>O<sub>3</sub>. *J. Chem. Phys.* **153**, 144501 (2020).
27. Liu, H., Qian, X., Bao, H., Zhao, C. Y. & Gu, X. High-temperature phonon transport properties of SnSe from machine-learning interatomic potential. *J. Phys. Condens. Matter* **33**, 405401 (2021).
28. Pun, G. P. P., Yamakov, V., Hickman, J., Glaessgen, E. H. & Mishin, Y. Development of a general-purpose machine-learning interatomic potential for aluminum by the physically informed neural network method. *Phys. Rev. Mater.* **4**, 113807 (2020).
29. Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V. & Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* **99**, 064114 (2019).
30. Artrith, N., Morawietz, T. & Behler, J. High-dimensional neural-network potentials for multicomponent systems: applications to zinc oxide. *Phys. Rev. B* **83**, 153101 (2011).
31. Behler, J. Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys. Condens. Matter* **26**, 183001 (2014).
32. Szlachta, W. J., Bartók, A. P. & Csányi, G. Accuracy and transferability of Gaussian approximation potential models for tungsten. *Phys. Rev. B* **90**, 104108 (2014).
33. Dolgirev, P. E., Kruglov, I. A. & Oganov, A. R. Machine learning scheme for fast extraction of chemically interpretable interatomic potentials. *AIP Adv.* **6**, 085318 (2016).
34. Kruglov, I. A., Yanilkin, A., Oganov, A. R. & Korotaev, P. Phase diagram of uranium from ab initio calculations and machine learning. *Phys. Rev. B* **100**, 174104 (2019).
35. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
36. Zuo, Y. et al. A performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* <https://doi.org/10.1021/acs.jpca.9b08723> (2019).
37. Korotaev, P., Novoselov, I., Yanilkin, A. & Shapeev, A. Accessing thermal conductivity of complex compounds by machine learning interatomic potentials. *Phys. Rev. B* **100**, 144308 (2019).
38. Ladygin, V. V., Korotaev, P. Y., Yanilkin, A. V. & Shapeev, A. V. Lattice dynamics simulation using machine learning interatomic potentials. *Comput. Mater. Sci.* **172**, 109333 (2020).
39. Novikov, I. S., Gubaev, K., Podryabinkin, E. V. & Shapeev, A. V. The MLIP package: moment tensor potentials with MPI and active learning. *Machine Learn. Sci. Technol.* **2**, 025002 (2020).
40. Straatsma, T. P. & Berendsen, H. J. C. Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* **89**, 5876–5886 (1988).
41. Freitas, R., Asta, M. & de Koning, M. Nonequilibrium free-energy calculation of solids using LAMMPS. *Comput. Mater. Sci.* **112**, 333–341 (2016).
42. Landau, L. D. & Lifshitz, E. M. Course of theoretical physics. Elsevier, 2013.
43. Hirose, K., Simmyo, R., Sata, N. & Ohishi, Y. Determination of post-perovskite phase transition boundary in MgSiO<sub>3</sub> using Au and MgO pressure standards. *Geophys. Res. Lett.* **33**, L01310 (2006).
44. Kvashnin, A. G. & Samtsevich, A. I. Phase transitions in tungsten monoborides. *JETP Lett.* **111**, 343–349 (2020).
45. Portnoi, K. I., Romashov, V. M., Levinskii, Y. V. & Romanovich, I. V. Phase diagram of the system tungsten-boron. *Sov. Powder Metall. Met. Ceram.* **6**, 398–402 (1967).
46. Rudy, E. Experimental phase equilibria of selected binary, ternary, and higher order systems. Part 5. the Phase Diagram W-B-C. <https://apps.dtic.mil/sti/citations/ADA472013> (1970).
47. Sjostrom, T., Crockett, S. & Rudin, S. Multiphase aluminum equations of state via density functional theory. *Phys. Rev. B* **94**, 144101 (2016).
48. Tambe, M. J., Bonini, N. & Marzari, N. Bulk aluminum at high pressure: a first-principles study. *Phys. Rev. B Condens. Matter Mater. Phys.* **77**, 172102 (2008).
49. Kudasov, Y. B. et al. Lattice dynamics and phase diagram of aluminum at high temperatures. *J. Exp. Theor. Phys.* **117**, 664–671 (2013).
50. Akahama, Y., Nishimura, M., Kinoshita, K., Kawamura, H. & Ohishi, Y. Evidence of a fcc-hcp transition in aluminum at multimegabar pressure. *Phys. Rev. Lett.* **96**, 045505 (2006).
51. Grigoriev, I. & Meilikhov, E. *Handbook of Physical Quantities (November 25, 1996 ed.)* (Energoatomizdat, Moscow, 1991; CRC Press, Boca Raton, Florida, United States, 1996).
52. Boehler, R. & Ross, M. Melting curve of aluminum in a diamond cell to 0.8 Mbar: implications for iron. *Earth Planet Sci. Lett.* **153**, 223–227 (1997).
53. Oganov, A. R., Brodholt, J. P. & Price, G. D. Comparative study of quasiharmonic lattice dynamics, molecular dynamics and Debye model applied to MgSiO<sub>3</sub> perovskite. *Phys. Earth Planet. Inter.* **122**, 277–288 (2000).
54. Börscke, T. S., Müller, J., Bräuhäus, D., Schröder, U. & Böttger, U. Ferroelectricity in hafnium oxide thin films. *Appl. Phys. Lett.* **99**, 102903 (2011).
55. Ohtaka, O. et al. Phase relations and volume changes of hafnia under high pressure and high temperature. *J. Am. Ceram. Soc.* **84**, 1369–1373 (2001).
56. Hyuk Park, M. et al. Ferroelectricity and antiferroelectricity of doped thin HfO<sub>2</sub>-based films. *Adv. Mater.* **27**, 1811–1831 (2015).
57. Huan, T. D., Sharma, V., Rossetti, G. A. & Ramprasad, R. Pathways towards ferroelectricity in hafnia. *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 064111 (2014).
58. Sang, X., Grimley, E. D., Schenk, T., Schroeder, U. & Lebeau, J. M. On the structural origins of ferroelectricity in HfO<sub>2</sub> thin films. *Appl. Phys. Lett.* **106**, 162905 (2015).
59. Materlik, R., Kunne, C. & Kersch, A. The origin of ferroelectricity in Hf1–xZrxO<sub>2</sub>: a computational investigation and a surface energy model. *J. Appl. Phys.* **117**, 134109 (2015).
60. Batra, R., Huan, T. D., Jones, J. L., Rossetti, G. & Ramprasad, R. Factors favoring ferroelectricity in hafnia: a first-principles computational study. *J. Phys. Chem. C* **121**, 4139–4145 (2017).
61. Sivaraman, G. et al. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *NPJ Comput. Mater.* **6**, 1–8 (2020).
62. Verdi, C., Karsai, F., Liu, P., Jinnouchi, R. & Kresse, G. Thermal transport and phase transitions of zirconia by on-the-fly machine-learned interatomic potentials. *NPJ Comput. Mater.* **7**, 1–9 (2021).
63. Liu, P., Verdi, C., Karsai, F. & Kresse, G. Phase transitions of zirconia: machine-learned force fields beyond density functional theory. *Phys. Rev. B* **105**, L060102 (2022).
64. Fan, S. et al. Vibrational fingerprints of ferroelectric HfO<sub>2</sub>. *NPJ Quantum Mater.* **7**, 1–8 (2022).
65. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
66. 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).
67. 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).
68. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
69. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
70. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

## ACKNOWLEDGEMENTS

I.A.K. gratefully acknowledges the financial support from the Ministry of Science and Higher Education (Agreement No. 075-15-2021-606) and from the Foundation for Assistance to Small Innovative Enterprises in Science and Technology (the UMNiK program). A.B.M. thanks the Russian Science Foundation (grant No. 19-73-00237) for financial support. The work of A.R.O. is supported by the Russian Science Foundation (grant 19-72-30043).

## AUTHOR CONTRIBUTIONS

I.A.K., A.V.Ya., A.R.O., and Ya.V.P. developed the T-USPEX code. I.A.K. did calculations of Al and MgSiO<sub>3</sub> phase diagrams, A.B.M. trained interatomic potential for Fe, Ya.V.P. calculated transition temperature between WB phases, P.Yu.R. estimated the number of local minima in MgSiO<sub>3</sub>. I.A.K., Ya.V.P., A.B.M., and A.R.O. wrote the manuscript. All the authors provided critical feedback and helped shape the research.

## 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/s41524-023-01120-6>.

**Correspondence** and requests for materials should be addressed to Ivan A. Kruglov or Artem R. Oganov.

**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) 2023