Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning

Evgeny V. Podryabinkin, Evgeny V. Tikhonov, Alexander V. Shapeev, and Artem R. Oganov

Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning

Evgeny V. Podryabinkin,[1, *] Evgeny V. Tikhonov,[1, 2] Alexander V. Shapeev,[1] and Artem R. Oganov[1, 3]

[1]*Skolkovo Institute of Science and Technology,*
Skolkovo Innovation Center, Building 3, Moscow, 143026, Russia
[2]*Moscow State University,*
Leninskie Gory, Moscow 119991, Russia
[3]*Moscow Institute of Physics and Technology,*
9 Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russia
(Dated: February 5, 2019)

In this article we propose a new methodology for crystal structure prediction, which is based on the evolutionary algorithm USPEX and the machine-learning interatomic potentials actively learning on-the-fly. Our methodology allows for an automated construction of an interatomic interaction model from scratch replacing the expensive DFT with a speedup of several orders of magnitude. Predicted low-energy structures are then tested on DFT, ensuring that our machine-learning model does not introduce any prediction error. We tested our methodology on a problem of prediction of carbon allotropes, dense sodium structures and boron allotropes including those which have more than 100 atoms in the primitive cell. All the the main allotropes have been reproduced and a new 54-atom structure of boron have been found at very modest computational efforts.

I. INTRODUCTION

Crystal structure prediction consists in searching for atomic structures with lowest thermodynamic potential[1]. Usually theoretical methods of the prediction involve two components: an algorithm sampling the configuration space and a relaxation algorithm which finds local minima of the relevant thermodynamic potential (e.g., internal energy). For example, the USPEX algorithm[2–4] is based on evolutionary structure optimization. The relaxed structures are sorted according to their energies, and the lowest-energy structures are used for producing the next generation of the evolutionary algorithm. This process of producing and relaxing structures continues generation-by-generation, until the lowest-energy structure remains unchanged for a number of generations.

The success of crystal structure prediction largely depends on the choice of the Hamiltonian. Density functional theory (DFT) offers a sufficient accuracy in reproducing the sophisticated interaction of real atoms in crystals, however, DFT has a high computational cost. Indeed, the complexity of DFT calculations grows cubically with the number of atoms, and in the course of structure relaxation such calculations are repeated many times. Thus, structure relaxation typically takes more than 99.9% of the total CPU time of the prediction when using DFT. Therefore, in practice, prediction of a crystal structure with DFT is usually limited to systems with a few tens of atoms. Furthermore, many types of calculations are unfordable with DFT: for example, crystal structure prediction at finite temperatures, were proper sampling of the phase space is needed for computing entropies.

Empirical interatomic potentials—a very computationally efficient alternative to DFT—can rarely be useful for predicting new materials, because their algebraic form is limited to reproducing physical properties of a few known structures that they were specifically designed for. A promising alternative to the empirical potentials is the so-called *Machine-Learning Interatomic Potentials*. They typically have a flexible functional form that allows for systematic improvement of their accuracy by the cost of computational efficiency. Several approaches to developing machine-learning potentials exist: neural network-based potentials[5–8], Gaussian approximation potentials[9–11] and potentials based on linear regression, e.g.,[12] and[13]. In particular, neural-network potentials have been used to explore transitions between different phases of Si for a range of pressures and temperatures[14]. Machine learning methods have been successfully used for crystal structure prediction problems. For example, an idea to fit a potential while structure search was suggested in[15]; in[16] the authors apply active-learning techniques to predict the surface reconstructions; in[17] Bayesian optimization was used for the problem of prediction of molecular compounds.

II. MACHINE LEARNING INTERATOMIC POTENTIALS

In the present paper we propose a novel approach to the application of machine-learning potentials to crystal structure prediction. It is based on the moment tensor potentials (MTPs)[13] as the machine-learning interatomic interaction model. Briefly, MTPs assume a partitioning of the energy into contributions of each atom: $E = \sum_i V_i$, where i goes through all atoms in the structure. Each $V_i = V(\boldsymbol{u}_i)$ depends on the atomic neighborhood $\boldsymbol{u}_i$ defined as the collection of positions the atoms relative to the i-th atom within a cutoff sphere of radius R_{cut}. MTPs provide an analytical expression for V_i as a linear combination of m basis functions $B_j = B_j(\boldsymbol{u}_i)$ with the fitting

parameters $\theta = (\theta_1, \ldots, \theta_m)$:

$$V_i = \sum_{j=1}^{m} \theta_j B_j(\boldsymbol{u}_i). \qquad (1)$$

The number of basis functions, m, is chosen by empirically balancing the accuracy and computational efficiency of the MTP. The basis functions satisfy all the physical symmetries (in particular, rotation invariance and invariance toward permutation of atoms of the same type) and have explicit expressions for calculation of the forces and stresses. The parameters θ in the simplest case are found from the requirement that the predicted energies are the DFT energies, $E(x^{(k)}) \approxeq E^{\text{dft}}(x^{(k)})$, on a set of configurations that we call the *training set*. This yields a system of linear algebraic equations on the coefficients θ_j:

$$\sum_{j=1}^{m} \theta_j \underbrace{\left[\sum_{i=1}^{N} B_j(u_i^{(k)}) \right]}_{=:b_j(x^{(k)})} = E^{\text{dft}}(x^{(k)}), \qquad (2)$$

which in the matrix notation we write as $\mathsf{A}\theta = \vec{E}^{\text{dft}}$, where

$$\mathsf{A} = \begin{pmatrix} b_1(x^{(1)}) & \ldots & b_m(x^{(1)}) \\ \vdots & \ddots & \vdots \\ b_1(x^{(K)}) & \ldots & b_m(x^{(K)}) \end{pmatrix}. \qquad (3)$$

The system (2) is typically overdetermined, therefore its solution is found from a least-square minimization problem as

$$\theta := (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top \vec{E}^{\text{dft}}. \qquad (4)$$

III. LEARNING ON THE FLY

There are two main ways to train and use a machine-learning potentials. The first (classical) is to train them at the offline stage and use it for calculation of the energy, forces and stresses at the online stage. The biggest challenge associated with this approach is related to transferability of machine-learning potentials: they do not produce reliable predictions outside their training domain. This appears a fatal weakness of them in the context of crystal structure prediction: since we do not know the crystal structures to be predicted, we cannot train a potential on those structures.

The idea of *active learning* comes to rescue. An active learning algorithm[18] can detect when an MTP attempts to extrapolate outside its training domain and include those extrapolative configurations in the training set. To formalize the concept of extrapolation, we consider an arbitrary configuration x^* and note that its predicted energy can be expressed as a linear combination of the energies of configurations from the training set:

$$E(x^*) = \sum_{j=1}^{m} \theta_j b_j(x^*) = (b^*)^\top \cdot \theta =$$

$$= \underbrace{(b^*)^\top (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top}_{=:c} \vec{E}^{\text{dft}} = \sum_{k=1}^{K} c_k E^{\text{dft}}(x^{(k)}). \quad (5)$$

If at least one c_k in (5) is larger than 1 by its absolute value, then we consider the energy calculation as linear *extrapolation* outside of the training domain; otherwise, if $|c_k| \leq 1$ for all k, then we say that *interpolation* within the training set takes place. In the case of interpolation the calculated energy is always bounded by the energy values from the training set (and therefore is expected to be close to the DFT values), whereas extrapolation may yield non-physically low or high values of energy. To simplify the algorithm, we assume that the training set size, K, equals to the number of basis functions, m. Thus, to detect extrapolation while calculating $E(x^*)$ we should additionally calculate c_k, which requires only one additional matrix-vector multiplication $\mathsf{A}(\mathsf{A}^\top \mathsf{A})^{-1} b^* = \mathsf{A}^{-1} b^*$, provided that we store A^{-1} in our computations. We emphasize that no additional DFT calculations are required to detect extrapolation.

Our active learning approach consists of detecting and including the extrapolative configurations to the training set. Thus, for a configuration x^* we compute the extrapolation grade that we define as $\gamma(x^*) = \max_k(|c_k|)$, and compare it to the tunable parameter $\gamma_{\text{tsh}} > 1$, which we call the extrapolation threshold. If the extrapolation grade is sufficiently high, $\gamma(x^*) > \gamma_{\text{tsh}}$, then the expensive DFT calculation is performed and the configuration is "learned", otherwise the energy, forces and stresses are calculated by the MTP (see Fig. 1). We always keep the size of the training set constant, $K = m$. Therefore, after adding x^* to the training set we eliminate $x^{(k^*)}$ from the training set, where k^* is such that c_{k^*} is maximal by its absolute value among all c_k. It can be shown that replacing configuration $x^{(k^*)}$ becomes interpolative after adding x^* to the training set[18].

The described active learning procedure employs a part of the maxvol algorithm[19] for finding the most linearly independent rows in a tall matrix, and it was shown that this leads to the increase of $|\det(\mathsf{A})|$ by the factor of $\gamma(x^*) > \gamma_{\text{tsh}} > 1$. In other words, we select the configurations so that they maximize $|\det(\mathsf{A})|$. Such approach is known as the D-optimality criterion[20] and is commonly used in machine learning and optimal experiment design.

MTP that actively learns on-the-fly can be considered, effectively, as an interatomic interaction model (see Fig. 1). We use this model as a replacement for DFT in USPEX. Note that since the interatomic potential within this model may slightly change when being retrained, the relaxation method requires adaptation—we force the BFGS algorithm used in relaxation[21] to perform one gradient descent iteration once a learning event occurs. It
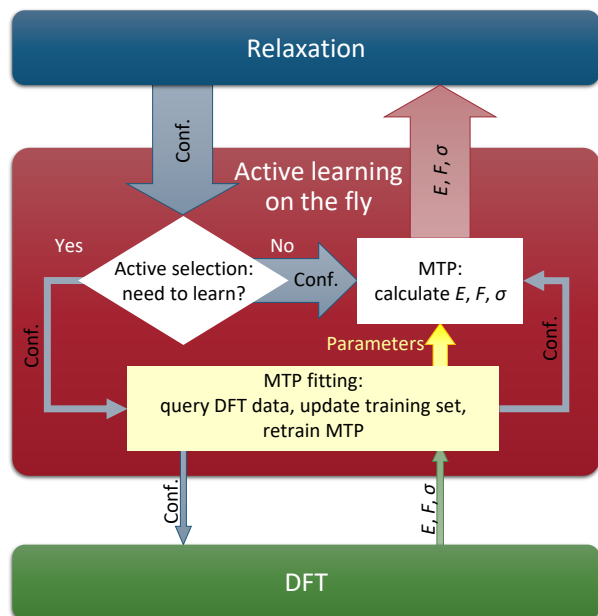
FIG. 1. The scheme of learning on-the-fly. An active selection algorithm estimates the degree of extrapolation for each configuration sampled. If it is high then the configuration is learned. After this, the energy, forces and stresses are calculated by MTP and returned to the relaxation process.

should be emphasized that the described selection algorithm (and fitting procedure) is not limited to configurations with the same number of atoms; it allows us to actively train a single model that works on structures with different numbers of atoms.

When an MTP starts learning on-the-fly from the empty training set, the major part of DFT calculations take place at the initial stage, while practically remain unchanged later (see the blue curve in Figure 2). Such behavior is explained by the following. Our selection method unconditionally selects each configuration until the training set size is less than the number of MTP parameters m. Next, the configuration x^* enters into the training set only if $\gamma(x^*) > \gamma_{\text{tsh}}$, i.e. its extrapolation grade is larger than the threshold. Since initially the configurational space is not well explored, the chance to meet an extrapolative configuration at the initial stage of learning on-the-fly is higher than on a later stage. Thus, on a late stage the training set is selected from a sequence sampled from a large configurational space and hence new extrapolative configurations appear rarely.

One can take advantage of this fact by performing pre-exploration of the configurational space, thus significantly decreasing the number of DFT calculations at the initial stage. Since our selection method does not require DFT data, the pre-exploration is done by sampling random structures subject to a minimal distance constraint. In our tests we generated a pool of 100 000 structures with the different number of atoms. The initial training thus comprises m (same as the number of MTP parame-

ters) configurations selected from the pool. This reduces the number of DFT calculations by about a factor of 5 (compare the blue and green curve in Figure 2).
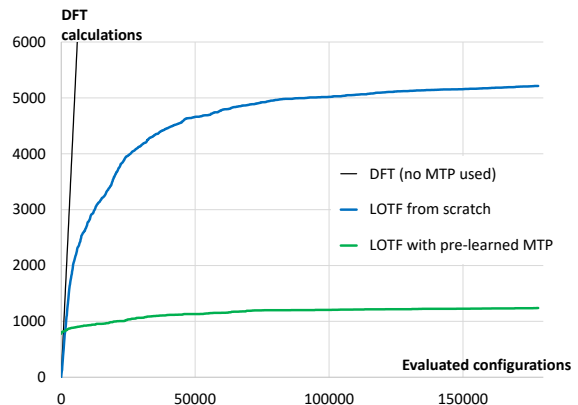

FIG. 2. Comparison of learning curves for pre-trained MTP and for learning from scratch.

IV. CRYSTAL STRUCTURE PREDICTION WITH OUR METHODOLOGY

We have tested our methodology on finding the structures for three chemical elements: (1) carbon allotropes, (2) sodium structures under pressure, (3) boron allotropes. In our tests no *a priori* information about the low-energy structures was used. In all the three cases we employ an MTP with about 800 parameters providing a balance between accuracy and computational efficiency. Since our model reproduces DFT only approximately, the structures predicted with DFT and our model may not coincide. Therefore, the structures with sufficiently low energies (within 100 meV/atom of the lowest energy) found with our method were relaxed with DFT. All DFT calculations were performed with the VASP package[22–24] at the generalized gradient approximation level of theory[25] using the projected augmented wave wave method for describing effects of core electrons[26,27].

A. Carbon structures

In the course of searching of carbon structures with 8 atoms in the unit cell our method has correctly predicted all main allotropes: graphite, diamond and lonsdaleite. More than $1.9 \cdot 10^4$ configurations were evaluated by MTP whereas the number of DFT calculations was about 1300. After learning on-the-fly, the MTP training error was 86 meV/atom. We note that the training error evaluated on the actively selected training set is an overestimation of the actual prediction error since the actively selected

configurations tend to sample more "extreme" parts of the configurational space[18]. The error on the predicted structures was less than 40 meV/atom.

B. Sodium under pressure

Next we searched for sodium structures under pressure in the range 120–300 GPa with up to 20 atoms in the unit cell. For this purpose we executed our method with a single MTP which correctly predicted all the known ground state sodium structures[28–31] (namely, cI16, tI19, hP4, provided in Supplemental Materials) within the specified pressure range. It is remarkable and reassuring that MTP captures non-trivial physics here: hP4-Na is an electride, i.e., can be described as an ionic salt made of Na^+ cores and localized electron pairs, which play the role of anions. We found it remarkable that MTP, based only on nuclear positions, is able to ~~describe~~ recognize and predict this electride phase. About 1500 DFT calculations have been done while learning on-the-fly, whereas the total number of evaluated configurations was more than $2 \cdot 10^6$. In this test the speedup of our method (which is equal to the ratio between the number of DFT and MTP calculations) is 100 times higher than in the previous, which indicates a higher efficiency of our method when searching structures with more atoms.

C. Boron allotropes

In the third test we demonstrate an application of our approach to one of the most challenging, in our opinion, crystal structure prediction problems: finding the allotropes of boron. Indeed, boron has an extremely complex potential energy surface with large amount of local minima (i.e., large number of allotropes) and small energy differences between them. Moreover, some of allotropes have more than 100 atoms in their primitive cell[32–36], therefore, exhaustive search is virtually impossible since the number of possible structures grows exponentially with the number of atoms. ~~Up to now, all theoretical attempts to predict the structure of β-boron experimental knowledge of lattice parameters and of positions of most boron atoms (the fully occupied sites), only varying the occupation of the partially occupied sites. Here we present the first attempt to perform a fully theoretical and unconstrained prediction of this structure, using no empirical information. The success that we have achieved is remarkable and sets the record of complexity of the predicted crystal structure.~~

We started by looking for boron allotropes with 1–60 ~~and 105–108~~ atoms in the primitive cell. We started from structures with one atom in the primitive cell, and increased the number of atoms by one in each subsequent run. At the 12th stage of the search the α-boron 12-atom structure was found with MTP.

Despite the fact that the α-boron structure was correctly identified as the ground state, the training error was 170 meV/atom. This error seems large, therefore, we modified our procedure by adding the second actively learning MTP. The first, reliable, MTP is used to screen out the high-energy structures, whereas the second one was introduced for accurate treatment of configurations with sufficiently low energy per atom. Switching between the reliable and accurate MTPs is done in a two-stage relaxation (see Fig. 3). At the first stage a configuration is relaxed with a reliable MTP and is discarded if its energy is higher than -5.5 eV/atom. Otherwise, the relaxation proceeds with the accurate MTP. Active learning of the accurate MTP is further restricted to low-energy structures by applying to the selection and training procedures the weight

$$W(x) = 1/(E(x)/N - E_{\min} + 0.02)^2,$$

where $E_{\min}$ eV/atom is the minimal DFT per atom energy of the structures from the training set ($E_{\min} = -6.705$ after finding of α-boron) ~~where $E_{\min} = -6.705$ eV/atom is the per atom energy of α-boron~~, and N is the number of atoms in the configuration x. That is, both sides of every equation (2) are multiplied by $W(x^{(k)})$. Being trained only on low-energy configurations the second MTP better reproduces the potential energy surface near deep minima, but is not suitable for high-energy structures.
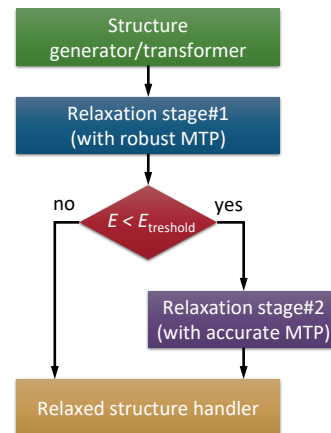

FIG. 3. The two-stage relaxation scheme

The two-stage scheme yields the root-mean-square error of 11 meV/atom for the low-energy structures (we have compared the MTP and DFT energies for the 100 lowest-energy structures we found). The total number of DFT calculations required to train the two MTPs was about 5000 while the total number of evaluated configurations exceeds $4 \cdot 10^8$. The best structures (with the lowest energy) found with our method are shown in Figure 4. The data files for these structures are provided in Supplemental Materials.

α-boron
E^{DFT}= –6.706 eV/atom
Atoms: 12,
Space group: R-$3m$,
Core-hours: 10^3 AL-MTP vs. $3 \cdot 10^3$ DFT
$|E^{DFT} – E^{MTP}|$ = 28.6 meV/atom

β-boron approximant
E^{DFT} = –6.704 eV/atom,
Atoms: 106,
Space group: $P1$,
Core-hours: $7 \cdot 10^3$ AL-MTP vs. $6.6 \cdot 10^7$ DFT
$|E^{DFT} – E^{MTP}|$ = 10.1 meV/atom

γ-boron
E^{DFT} = –6.678 eV/atom,
Atoms: 28,
Space group: $Pnnm$,
Core-hours: $2 \cdot 10^3$ AL-MTP vs. $2.5 \cdot 10^4$ DFT
$|E^{DFT} – E^{MTP}|$ = 58.1 meV/atom

E^{DFT} = –6.667 eV/atom,
Atoms: 54,
Space group: Im-3,
Core-hours: $3 \cdot 10^3$ AL-MTP vs. $3.5 \cdot 10^5$ DFT
$|E^{DFT} – E^{MTP}|$ = 7.3 meV/atom

E^{DFT} = –6.667 eV/atom,
Atoms: 52,
Space group: P-$42m$,
Core-hours: $3 \cdot 10^3$ AL-MTP vs. $3.2 \cdot 10^5$ DFT
$|E^{DFT} – E^{MTP}|$ = 37.3 meV/atom

E^{DFT} = –6.665 eV/atom,
Atoms: 26,
Space group: $Cccm$,
Core-hours: $2 \cdot 10^3$ AL-MTP vs. $2.1 \cdot 10^4$ DFT
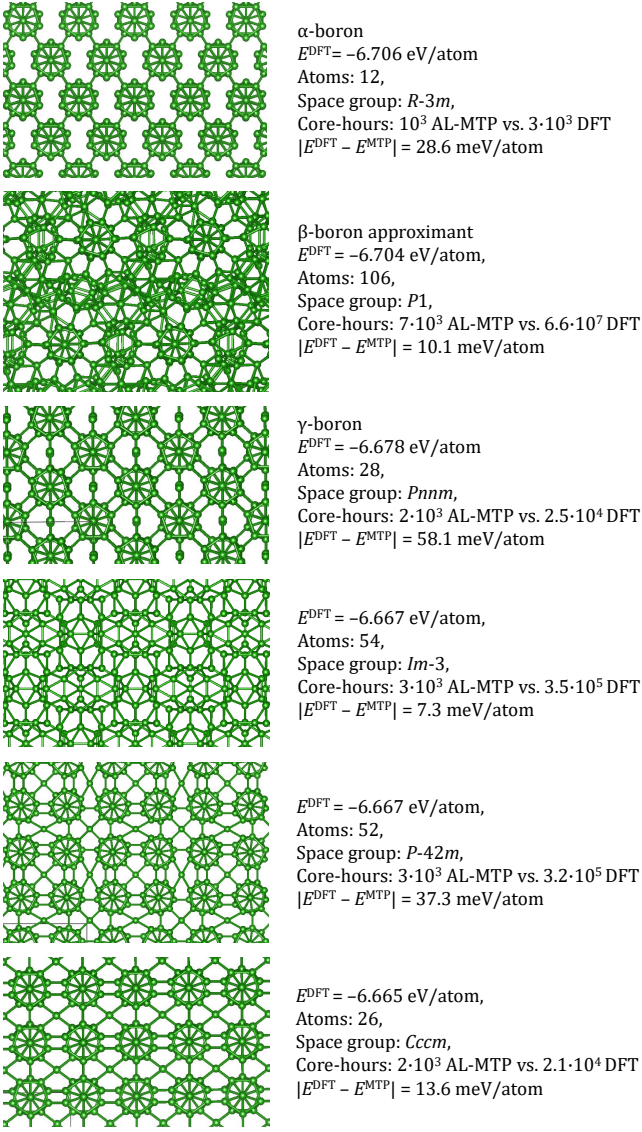$|E^{DFT} – E^{MTP}|$ = 13.6 meV/atom

FIG. 4. The best (lowest-energy) structures found with our method. The estimate for the time required to find these structures on DFT is based on the number of configuration threated by MTP in our search and the time required for VASP to process all these structures on a single core. The actual time spent with DFT can be up to 10 times less than indicated, because at early stages of structure relaxation cheaper computational settings are usually applied.

The USPEX+MTP calculation correctly reproduced the lowest-energy $Cccm$ structure with 26 atoms/cell and also discovered a closely related tetragonal $P\overline{4}2m$ structure with 52 atoms/cell and slightly lower energy. These two structures are topologically very similar to the 52-atom $Pnn2$ structure published in[37] and the 52-atom $Pnnn$ structure recently seen in experiments[38], the fourth established pure boron allotrope. These structures are very similar to the tetragonal $(P4_2/nnm)$ B_{50}, the first structurally characterized form of boron[39], which was later found to be impure and stabilized by nitro-

gen and carbon atoms[40]. DFT predicts that all the three 52-atom structures and the 26-atom structure are dynamically stable at zero temperature and have very close energies, within 8 meV/atom (all four structures are provided in Supplemental Materials). However, MTP "catches" only two of them: the $Cccm$ and P-$42m$ structures. Relaxation of $Pnn2$ and $Pnnn$ structures results in the 26×2-atom $Cccm$ structure. This proves boron to be a hard benchmark problem for crystal structure prediction. With our two-MTP scheme ~~In our calculations~~ we have also predicted correctly the γ-boron structure, recently discovered theoretically and experimentally[33].

~~Another~~ A remarkable result of this work is the discovery of a 54-atom metallic structure of boron, with energy just 39 meV/atom higher than that of α-boron and 11 meV/atom higher than that of γ-boron. This structure has an unexpectedly high symmetry, cubic (space group Im-3), the point group of which is actually the highest group that an icosahedron can have in any crystalline environment (m-3). The B_{12} icosahedron is connected to the neighboring icosahedra only by two-center bonds— this is what allows it to keep the highest symmetry allowed for icosahedron in crystals. As a result, we can identify in this structure the highly symmetric B_{84} units: in their center is one B_{12} icosahedron, with each of its atoms bonded by a two-center bond to a B_6 pentagonal umbrella. These B_6-umbrellas can be viewed as fragments of icosahedra; building this structure with complete icosahedra would lead to an aperiodic and highly strained structure; our B_{54} is its simplest periodic approximant.

We have checked thermal stability of this structure with molecular dynamics simulations. To this end we replicated the unit cell and ran simulations with NPT ensemble of 13500 atoms at $T = 1200K$ with the accurate MTP. After 100 ps we performed relaxation and verified that no structure transformation occurred. The structure has specific volume of 7.61 Å^3/atom, which is close to that of β-boron. It has the following elastic constants: $C_{11} = 418$ GPa, $C_{12} = 102$ GPa, $C_{44} = 160$ GPa, bulk modulus $B = 208$ GPa, shear modulus $G = 160$ GPa and Young's modulus $E = 381$ GPa. It is a brittle metallic structure according to Pugh's criterion[41] (with B/G ratio of 1.3) with the hardness of $H = 25.3$ GPa calculated with the formula proposed in[42]. Low mass of the boron and covalent B-B bonding (leading to strong electron-phonon coupling) make metallic boron a potential superconductor, however, here we found $T_c < 1K$.

D. Prediction of β-boron structure

The most challenging test for our methodology is an attempt to predict the β−boron structure. It has a high configurational entropy and even now its exact structure remain unclear[43]. The minimal number of atoms in the primitive cell of such approximants is in the range of 105–108, whereas microscopically twinned boron phase with

212 atoms in the unit cell has the per atom energy lower than that of $\alpha-$boron[43]. Up to now, all theoretical attempts to predict the structure of $\beta-$boron[32,36,44,45] used experimental knowledge of lattice parameters and of positions of most boron atoms (the fully occupied sites), only varying the occupation of the partially occupied sites. Here we present the first attempt to perform a fully theoretical and unconstrained prediction of this structure, using no empirical information (except for the approximate number of atoms in the supercell). The success that we have achieved is remarkable and sets the record of complexity of the predicted crystal structure.

With our method we have found the best known 108-atom structure of boron, which is a supercell of α-boron (the same allotrope was also found among all the appropriate smaller structures). It confirms that our method is able to treat large structures with more than 100 atoms in the primitive cell. Furthermore, another predicted structure with 108 atoms in the unit cell has structural similarity with β-boron[36] (in particular, it also contains fused icosahedra with point defects and some non-icosahedral atoms, refer to Supplemental Materials) and energy only 8 meV/atom higher than that of α-boron. In addition to the 108-atom α-boron supercell structure, we have also found β-boron approximants with 105, 106, 107, and 108 atoms in the primitive cell whose energies are 14, 2, 8, and 8 meV/atom higher as compared to α-boron. These configurations have structural similarity with the known 106 atomic β-boron approximant[36], in particular, they also contain fused icosahedra with point defects and some non-icosahedral atoms. Moreover, our 106-atom structure has virtually the same energy as provided in[36] (DFT energies differ by less than 1 meV/atom). Both structures are shown in Fig. 5. Interestingly, our calculations found more than hundred β-boron-like structures with close energies, which is a manifestation of its configurational entropy.

ing dataset, thus seamlessly replacing DFT without significant changes to the crystal structure prediction algorithm. We have applied this method to sodium under pressure, and to carbon and boron. For compressed sodium, all high-pressure phases (including host-guest tI19 and electride hP4 phases) were found. For carbon, graphite, diamond, and lonsdaleite were reproduced. For boron we demonstrated that accuracy of 11 meV/atom is achievable. All known pure boron allotropes were found
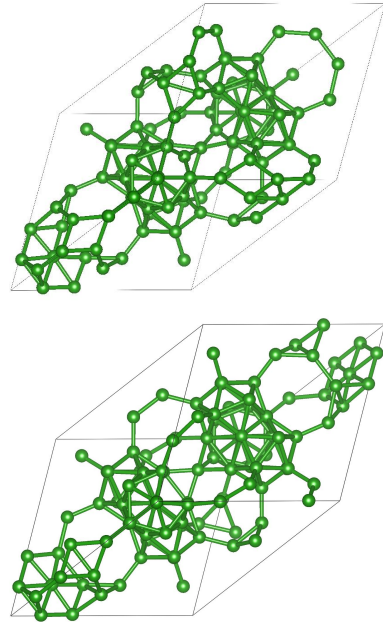

FIG. 5. Comparison of found 106-atom structure (upper figure) and structure from[36] (lower figure).

(including disordered $\beta-$boron with 106 atoms/cell), and a new low-energy metallic allotrope predicted.

V. CONCLUSION

In summary, we have proposed and tested a methodology for crystal structure prediction, based on a machine-learning interatomic interaction model and the evolutionary algorithm USPEX. The new methodology is orders of magnitude more computationally efficient compared to the conventional DFT-based algorithms. Our machine-learning model is automatically trained on-the-fly and does not require manually assembling the train-

[*] E.Podryabinkin@skoltech.ru
[1] A. R. Oganov, *Modern Methods of Crystal Structure Prediction* (2010).
[2] A. R. Oganov and C. W. Glass, The Journal of Chemical Physics **124**, 244704 (2006).
[3] A. R. Oganov, A. O. Lyakhov, and M. Valle, Accounts of Chemical Research **44**, 227 (2011).

[4] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, Computer Physics Communications **184**, 1172 (2013).

[5] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[6] K. J. Jose, N. Artrith, and J. r. Behler, Journal of Chemical Physics **136**, 194111 (2012).

[7] N. Artrith and A. Urban, Computational Materials Science **114**, 135 (2016).

[8] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. **21**, 124 (2017).

[9] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[10] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[11] W. J. Szlachta, A. P. Bartók, and G. Csányi, Phys. Rev. B **90**, 104108 (2014).

[12] A. Thompson, L. Swiler, C. Trott, S. Foiles, and G. Tucker, Journal of Computational Physics **285**, 316 (2015).

[13] A. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).

[14] J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, Phys. Rev. Lett. **100**, 185501 (2008).

[15] S. Wu, M. Ji, C. Wang, M. Nguyen, X. Zhao, K. Umemoto, R. Wentzcovitch, and K. Ho, Journal of physics. Condensed matter : an Institute of Physics journal **26** (2014), 10.1088/0953-8984/26/3/035402.

[16] T. L. Jacobsen, M. S. Jørgensen, and B. Hammer, Phys. Rev. Lett. **120**, 026102 (2018).

[17] M. S. Jorgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, The Journal of Physical Chemistry A **122**, 1504 (2018), pMID: 29314842, https://doi.org/10.1021/acs.jpca.8b00160.

[18] E. V. Podryabinkin and A. V. Shapeev, Computational Materials Science **140**, 171 (2017).

[19] S. Goreinov, I. Oseledets, D. Savostyanov, E. Tyrtyshnikov, and N. Zamarashkin, in *Matrix Methods: Theory, Algorithms, Applications* (Word Scientific, 2010) pp. 247–256.

[20] B. Settles, *Active Learning Literature Survey*, Computer Sciences Technical Report 1648 (University of Wisconsin–Madison, 2009).

[21] R. Fletcher, *Practical Methods of Optimization; (2Nd Ed.)* (Wiley-Interscience, New York, NY, USA, 1987).

[22] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[23] G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996).

[24] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[25] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[26] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[27] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[28] Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, Nature **458**, 182 (2009).

[29] E. Gregoryanz, O. Degtyareva, M. Somayazulu, R. J. Hemley, and H.-k. Mao, Phys. Rev. Lett. **94**, 185502 (2005).

[30] E. Gregoryanz, L. F. Lundegaard, M. I. McMahon, C. Guillaume, R. J. Nelmes, and M. Mezouar, Science **320**, 1054 (2008).

[31] L. F. Lundegaard, E. Gregoryanz, M. I. McMahon, C. Guillaume, I. Loa, and R. J. Nelmes, Phys. Rev. B **79**, 064105 (2009).

[32] T. Ogitsu, E. Schwegler, and G. Galli, Chemical Reviews **113**, 3425 (2013).

[33] A. R. Oganov, J. Chen, C. Gatti, Y. Ma, Y. Ma, C. W. Glass, Z. Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko, Nature **457**, 863 (2009).

[34] Q. An, K. M. Reddy, K. Y. Xie, K. J. Hemker, and W. A. Goddard, Phys. Rev. Lett. **117**, 085501 (2016).

[35] Y. Ma, C. T. Prewitt, G. Zou, H.-k. Mao, and R. J. Hemley, Phys. Rev. B **67**, 174116 (2003).

[36] M. J. van Setten, M. A. Uijttewaal, G. A. de Wijs, and R. A. de Groot, Journal of the American Chemical Society **129**, 2458 (2007).

[37] Q. Zhu, A. R. Oganov, C. W. Glass, and H. T. Stokes, Acta Crystallographica Section B **68**, 215 (2012).

[38] E. A. Ekimov, Y. B. Lebed, N. Uemura, K. Shirai, T. B. Shatalova, and V. P. Sirotinkin, Journal of Materials Research **31**, 2773 (2016).

[39] J. Hoard, S. Geller, and R. Hughes, J. Am. Chem. Soc. **73**, 1892 (1951).

[40] K. Ploog and E. Amberger, J. Less-Comm. Metals **23**, 33 (1971).

[41] S. F. Pugh, Philosophical Magazine **45**, 823 (1954).

[42] X.-Q. Chen, H. Niu, D. Li, and Y. Li, Intermetallics **19**, 1275 (2011).

[43] Q. An, K. M. Reddy, K. Y. Xie, K. J. Hemker, and W. A. Goddard, Phys. Rev. Lett. **117**, 085501 (2016).

[44] M. Widom and M. Mihalkovič, Phys. Rev. B **77**, 064113 (2008).

[45] V. L. Deringer, C. J. Pickard, and G. Csányi, Phys. Rev. Lett. **120**, 156001 (2018).