Computer Physics Communications 236 (2019) 1–7

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Topology-based crystal structure generator

Pavel V. Bushlanov [a,b,*], Vladislav A. Blatov [c,d], Artem R. Oganov [a,b,d,*]

[a] Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 3 Nobel St., Moscow 143026, Russia
[b] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region 141700, Russia
[c] Samara Center for Theoretical Materials Science (SCTMS), Samara State Technical University, Molodogvardeyskaya St. 244, Samara 443100, Russia
[d] International Center for Materials Discovery, Northwestern Polytechnical University, Xi'an, Shaanxi 720072, China

ARTICLE INFO

Article history:
Received 19 February 2018
Received in revised form 30 July 2018
Accepted 25 September 2018
Available online 17 October 2018

Keywords:
Crystal structure prediction
Computational materials discovery
Crystal topology
Random structure generation

ABSTRACT

Crystal structure prediction is a new dynamically developing field. Two main types of crystal structure prediction methods exist: (1) based on global optimization and (2) based on data mining. It seems promising to hybridize data mining and global optimization techniques. The former generally involve no empirical information and are truly predictive, the latter rely on the databases of existing crystal structures, and are relatively fast, but prone to error, because the databases are far from complete. Furthermore, the theorist's dream is to be as little dependent on empirical data as possible and be always capable of predicting new structures, not contained in the databases. Here we present an approach to generate an infinite number of crystal structures from a finite set of idealized periodic nets. The resulting structures are highly ordered, possess nontrivial symmetries, and often low energy. Topologically generated structures can be used for initializing evolutionary crystal structure prediction calculations, or on their own — as an extended data mining approach. The efficiency of the proposed approach in both scenarios is confirmed by a series of tests, which we also present here. As an additional enhancement to evolutionary algorithm we introduce a technique for adjusting fractions of variation operators on the fly. Tests show significant performance improvements due to both of these developments.

0. Introduction

Computational materials discovery is a dream coming true — enabling one to design materials with superior properties, without relying on the traditional trial-and-error approach. A prerequisite for computational materials discovery is the ability to predict crystal structures — and several methods of crystal structure prediction have been developed, e.g. our evolutionary algorithm USPEX [1–3], minima hopping [4–6], particle swarm optimization [7], random sampling [8–10], metadynamics [11–14]. The development reported here is included in the latest version 10.1 of USPEX, and discussed below in the context of evolutionary crystal structure prediction — but can be coupled with other approaches as well.

Evolutionary algorithm is a powerful approach for predicting materials structure and properties. To avoid bias, the initial population of structures is usually created by a random generator, or its improved version — random symmetric structure generator [2,15]. The introduction of symmetry helps to generate a diverse set of structures, many of which have low energies, and this improves efficiency and reliability of calculations. It can be proven that a random structure generator will produce a set of nearly identical, glassy structures when the crystal structure has many degrees of freedom [16]. Introduction of symmetry lowers the number of degrees of freedom and thus increases the diversity and degree of order of structures.

It seems promising – perhaps even more promising than symmetric random generator – to use data mining for creating the initial set of structures, which can then be subjected to natural selection, heredity and mutations, and evolve. A preliminary hybridization between data mining and particle swarm optimization has recently been reported [17]. However, utilizing only experimentally known crystal structure types, as in [17], seems restrictive — especially for complex stoichiometries, where databases are extremely far from complete.

In this respect, it seems especially useful to resort to the topological description of periodic structures, which has been developing during last 20 years [18–20]. This approach allows one to extract periodic architectures from crystal structures, abstracting from their chemical composition and geometric features. The resulting idealized periodic nets are collected in topological databases [21] and can be used for classification of existing crystal structures [22] as well as for generating new periodic motifs [23].

We have developed an approach, which combines the evolutionary algorithm and the topological description of crystal

* Corresponding authors at: Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 3 Nobel St., Moscow 143026, Russia.
E-mail addresses: p.bushlanov@skoltech.ru (P.V. Bushlanov), a.oganov@skoltech.ru (A.R. Oganov).

https://doi.org/10.1016/j.cpc.2018.09.016
0010-4655/

structures, using the idealized periodic nets to produce an initial generation, and also to add these "topological random" structures to every subsequent generation, to maintain diversity of the population. The approach consists of two steps: first, extracting proper nets from a topological database; second, generating crystal structures for any given stoichiometry, from the idealized nets.

Section 1 starts with a sketch of the evolutionary algorithm USPEX and presents one additional new development (evolutionary adjustment of variation operators). Section 2 describes our topological structure generator. Section 3 shows tests of our improved algorithm. Conclusions are presented in Section 4.

1. Sketch of the evolutionary algorithm USPEX

Evolutionary algorithms are a class of population-based global optimization techniques, most commonly involving an iteratively updated set of trial solutions (in our context, these are trial crystal structures) called "generation". Structures of the same generation are ranked by their fitness, and the fittest part of the generation (typically, fittest 60% of the structures) are allowed to produce the next generation of structures by means of variation operators — such as heredity (two parents producing one child structure) or various mutations (where one parent produces one child structure). The probability of choosing each structure from the parent pool is determined by its fitness. Variation operators are constructed such that they transfer a significant amount of structural information from parent structures into the offspring. Percentages of different variation operators are input parameters and efficiency of each calculation depends to some extent on the wise choice of these parameters. Below we present a way to reduce (essentially completely) the dependence on these percentages, so that they are adapted on-the-fly during the calculation — in other words, the calculation itself chooses which variation operators are more promising, and increases their use.

Automatic fraction adjustment: parameter control. Effective variation operator, ideally, should at the same time produce a large fraction of lower-energy (compared to parents and same-generation structures) structures and maintain high diversity among the generated structures. Accounting for both of these requirements (low energy and high diversity) is non-trivial (especially because these two requirements often clash), but fortunately, there is a simple solution.

We want to obtain "dynamics" of percentages of each operator, updated on the fly. When calculating percentage of structures produced by an ith variation operator in the nth generation $X_i(n)$, what really matters is how many structures this variation operator supplies into the parent structure pool. Let $\overline{N_i}(n)$ be the number of structures produced in the nth generation by an ith operator that belong to the parent pool and $N_i(n)$ the total number of structures in that generation produced by that operator. Naively fraction X_i would be just the ratio between $\overline{N_i}$ and the total number of structures in the parent pool, but it is better to take into account ratios $\overline{N_i}(n)/N_i(n)$ as well. After introducing "inertia" term to smoothen "trajectories" our "dynamics" of percentages X_i takes form:

$$\overline{X_i}(n+1) = \frac{\overline{N_i}(n)^2/N_i(n)}{\sum_i \overline{N_i}(n)^2/N_i(n)}, \tag{1}$$

$$X_i(n+1) = \left[X_i(n) + \overline{X_i}(n+1)\right]/2 \tag{2}$$

The parent pool is by construction sufficiently diverse (all structures in it are distinct) and low in energy. Those operators which give better ratios $\overline{X_i}$ are given a higher percentage X_i in the next generation.

Consider example. Let there be 3 variation operators: random, heredity, mutation. Initially fractions were set to: 0.2, 0.5, 0.3 respectively. Then the first generation is created entirely by random and second according these fractions. Suppose each generation contains 10 structures. 6 structures made it to parents pool: 4 obtained from heredity, 1 from mutation and 1 from random. We have $N_h(2) = 5$, $N_m(2) = 3$, $N_r(2) = 2$, $\overline{N_h}(2) = 4$, $\overline{N_m}(2) = 1$, $\overline{N_r}(2) = 1$. Then $\overline{X_h}(3) = \frac{4^2/5}{4^2/5+1^2/3+1^2/2} = \frac{3.2}{4.03} = 0.8$, $\overline{X_m}(3) = \frac{0.33}{4.03} = 0.08$, $\overline{X_r}(3) = \frac{0.5}{4.03} = 0.12$. Finally $X_h(3) = (0.5+0.8)/2 = 0.65$, $X_m(3) = (0.3+0.08)/2 = 0.19$, $X_r(3) = (0.2+0.12)/2 = 0.16$. So in the 3rd generation we will generate 65% of structures with heredity, 19% with mutation and 16% with random.

This mechanism of adjusting percentages of variation operators is also a very effective tool for judging relative effectiveness of different operators. In this paper, we consider random structure generator on the same footing as all variation operators (and will list it together with them, even though it is not, *per se*, a variation operator). Later in this paper we will compare, using this and other tools, symmetric and topological structure generators.

2. Topology-based random structure generation

There are at least three ways to create structures randomly:
(i) Fully random structure generator. Here (just as in all the other types of random structure generators), it is useful to impose some constraints — e.g. minimal atom–atom distances should not be too small (e.g. >0.5 A). Fully random structure generator turns out to fail for large systems: the more degrees of freedom, the more disordered will randomly produced structures be, in the limit leading to a population of physically identical and glass-like structures [24].
(ii) Symmetric random structure generator cures the main problem mentioned above: it reduces the effective number of degrees of freedom, allowing one to obtain ordered, low-energy and diverse structures even when the number of atoms in the unit cell is relatively large. This method was implemented, in different flavors, in USPEX [2,15], PSO method [7], random sampling [9,10]. For very large systems, however, mostly disordered structures will still be produced. Another drawback is that most structures produced this way contain clusters of identical atoms and – more generally – there is no way to make sure that atomic environments are chemically reasonable (no simple way to prohibit unreasonable coordination environments, or encourage favorable ones).
(iii) Topological random structure generator – whereby more realistic structures are generated with the use of databases – is the step forward, which we describe in this paper. One can realize that the search space we are sampling may be regarded as space of all possible crystal structures – rather than space of atomic coordinates. Most of this space is occupied by amorphous structures – which is what the fully random structure generator samples most of the time. Our task here is, instead, to prefer ordered structures over amorphous ones, by means of a method of regularizing the space of crystal structures.

Constructing the method (iii), we could limit ourselves with just the crystal structures reported in databases, such as Inorganic Crystal Structure Database (ICSD) [25,26], Pauling file [27], Crystallography Open Database (COD) [28], Cambridge Crystallographic Data Centre (CCDC) [29], Protein Data Bank (PDB) [30,31], American Mineralogist Crystal Structure Database [32], Pearson's Crystal Data [33]. Using a database of all known crystal structures, one could randomly pick entries from it. However, crystal structure databases do not cover the entire space of relevant crystal structures: new structure types are being continually discovered. In fact, there are an infinity of possible structure types and a discovery of a new structure type is routine in crystallography. On the contrary, the number of known crystal structure topologies is much smaller and it is very rare that a new topology is discovered. Our approach

is to invent a way of generating an infinite number of possible architectures on the basis of a finite set of topological types of crystal structures.

In the topological approach [20] every structure is represented as a periodic graph (net), consisting of nodes and edges. Atom types and actual positions of the atoms are ignored; thus, many different structures have the same topology (but may very well have different space groups). While there are only 230 space groups, which mathematically describe all possible crystal structures, thousands of underlying topologies have been documented for the known crystal structures [21]. More precise definition of underlying topology will be given in the next section. While symmetry is a purely mathematical notion, topology captures information about the highest possible space group and chemical information (structural connectivity) — the latter allows us to pick structures not randomly, but taking preferable coordination numbers into account. Choosing a topology and various occupations of nodes with atomic species, we obtain structures, including essentially *all* known crystal structures and an infinite set of not *yet* observed structures. This helps both to accelerate global optimization and to find low-energy metastable polymorphs. In short, this approach enables sampling of the configuration space in a much more profound and judicious way than straightforward database sampling.

Topology based random structure generation: details. While topology of a structure is a net of nodes and edges representing atoms and bonds regardless of the type of atoms and their positions, *underlying topology* is an even more simplified descriptor of a structure. To construct underlying topology of a structure we need to simplify it by (i) squeezing all complex structural units (e.g. clusters, coordination groups or molecules) into their centers of mass, and (ii) dropping all two- and one-coordinate nodes from the structure. The resulting *underlying net* will have only three- and higher-coordinated nodes. In a special case, when we consider the structure completely, the structural units coincide with atoms and the first step of simplification is not needed. For structure generation, though, some information about geometry is required. Luckily, necessary combination of topological and geometric data can be found in the ToposPro databases of idealized nets, which include both topologies observed in crystal structures and hypothetical ones [21]. These databases contain the most symmetric spatial embedding of each topology and currently include almost 200 000 records. That is from all possible systems with the same topology only the system with the highest symmetry is listed in this database. The entries of this database give us a very good starting point for further structure generation.

The simplest way to generate a structure from a chosen entry in the topological database is to associate each node of the structure with an atom randomly. For example we want to generate a structure for $Mg_4Al_8O_{16}$. We should find in the database an entry with 28 nodes and randomly associate nodes with our 28 atoms. This however will break symmetry of the system. The resulting structure would not be much better than generated with simple random sampling.

So we need a wiser way to associate nodes with atoms. Each entry in a ToposPro database contains information about symmetry group, unit cell parameters and several basis nodes in different Wyckoff positions. Acted upon by symmetry operators, each of the basis nodes produces a set of equivalent nodes.

Thus we have an option to associate atoms of the same element to nodes derived from the same basis node, i.e. in same Wyckoff position. So for our example of $Mg_4Al_8O_{16}$ we should find in the database an entry with three basis nodes, one of which is replicated by symmetry operators to produce 4 nodes, other 8 and the other 16, which corresponds to multiplicities of the Wyckoff positions. Structures obtained this way are quite promising, but this approach is too restrictive: only a few topologies will be usable for a given stoichiometry.

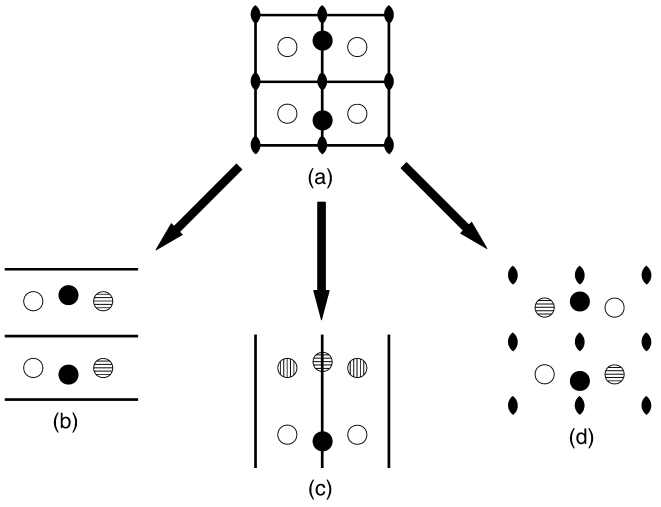

Fig. 1. Possible decompositions of Pmm2 symmetry and the corresponding structures. Here the lines represent symmetry planes, seed-like symbols represent twofold axes. Initial structure (a) has symmetry Pmm2 and contains two basis nodes, one in a general Wyckoff position (i) and one in a special Wyckoff position (h). Depending on which subgroup is chosen we get different structures. Structures (b) and (d) have 3 basis nodes, while structure (c) has 4. Initial structure (a) allows only two compositions A_2B_4 and A_6. Structures (b) and (d) allow A_2B_4, $A_2B_2C_2$ and A_6. Structure (c) allows $A_1B_1C_2D_2$, $A_1B_1C_4$, $A_1B_2C_3$, A_1C_5, $A_2B_2C_2$, A_2B_4, A_3B_3, A_6.

The next trick to be done is to group together basis nodes. Again for our example of $Mg_4Al_8O_{16}$ we can take into consideration entries with four basis nodes, two of which give together 4 nodes and the remaining two give 8 and 16. All possible combinations should be considered. This approach broadens the range of usable topologies, but still does not make full use of topologies.

The most dramatic increase of effectiveness of our method is related to symmetry group decomposition. We have already applied this approach to generate all possible subnets for a given uninodal or binodal (i.e. containing one or two topologically inequivalent nodes) periodic net [23,34]. The generation procedure includes decreasing the space group symmetry of the initial net by passing to a subgroup of the space group and then enumerating all possible ways of breaking the net edges. The complete list of all translation-equivalent (i.e. keeping the translational subgroup **T** of the space group **G**) and class-equivalent (i.e. keeping the non-translational part, which is described by the factor group **G/T**) subgroups of space groups is contained in the International Tables for Crystallography [35]. Here we restrict ourselves with all translation-equivalent and only maximal class-equivalent subgroups obtained by decomposition of **G** by all primitive translations. This approach is substantiated by that in most cases class-equivalent subgroups of high orders result in very complicated structures with a large number of basis atoms, which are rarely realized in nature. Thus for a given topology, besides the unit cell of different symmetries corresponding to translation-equivalent subgroups, we consider the supercells, which are described by maximal class-equivalent subgroups.

Let us look at a simple group where operators are $\mathbf{g_1} = x, y$; $\mathbf{g_2} = -x, y$; $\mathbf{g_3} = x, -y$; $\mathbf{g_4} = -x, -y$. Here $\mathbf{g_1}, \mathbf{g_2}$ form a subgroup and the corresponding factor group is $\mathbf{g_1}, \mathbf{g_3}$. This and other possible decompositions are shown in Fig. 1.

Structures obtained through this algorithm preserve symmetry and are much closer to what nature prefers, compared to what is given by other random structure generators. In our evolutionary calculations we use both topological and symmetric random structure generators and make them compete for share in USPEX algorithm, using the parameter control approach described by eqs. (1) and (2).

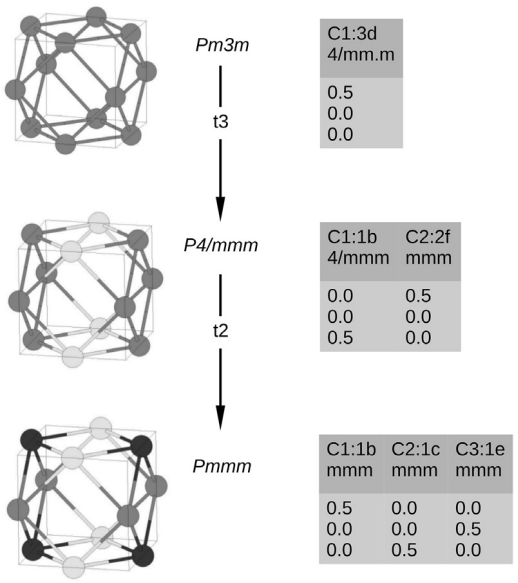

Fig. 2. **reo** topology is represented with Pm-3m-symmetry structure (the O net in ReO_3 has this structure). P4/mmm symmetry is translation equivalent subgroup of Pm-3m. Thus P4/mmm-symmetry structure with **reo** underlying topology can be obtained from Pm-3m-symmetry structure by breaking some point group operations (3-fold axis) by coloring atoms. In the same way Pmmm-symmetry structure with **reo** underlying topology results from P4/mmm-symmetry structure after breaking more point group symmetry operations (4-fold axis becomes 2-fold).

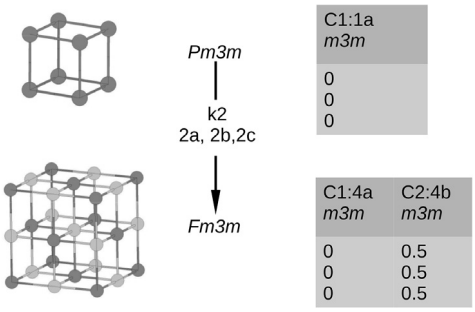

Fig. 3. Primitive cubic topology (**pcu**) is represented with Pm-3m-symmetry structure. Fm-3m symmetry of the NaCl structure is a class-equivalent subgroup of Pm-3m. Thus the NaCl structure (**pcu-b**) can be obtained from the primitive cubic topology: first, double **pcu** unit cell in each direction (this structure still has the Pm-3m symmetry), then break the primitive translations by coloring atoms.

Besides information about node positions, the ToposPro database contains information about bonds, most importantly, the coordination numbers of nodes. So we can now make a preliminary filtering of structures according to our knowledge of chemistry. For example, if we know that silicon prefers 4- and 6-coordinate positions, then we can filter out structures with very different coordination numbers. More than this, preferred coordination numbers for each element can be found early in the calculation, by analyzing low-energy structures. Such analysis and filtering will further increase the efficiency of the whole method. Even with such constraints on the coordination numbers, our method will produce a large set of structures — in contrast to straightforward database techniques (as, e.g., in [17]).

Discussion. The algorithm just described is a simplification of a general approach, where each topology in our database generates an infinite family of structures, each having the same topology, but with different symmetries. Symmetry of a given structure in the family is a subgroup of the space group of the system representing the net topology.

Our approach combines topological database sampling with constructing tree-like structure relationship graphs related to Bärnighausen trees [36]. While original Bärnighausen trees describe relations of existing structures with respect to different distortions and ordered substitutions, we consider only structures derived from the same topology by ordered substitutions. Bearing this in mind we will refer these graphs to as Bärnighausen trees.

An entry of topological database defines positions of atoms, but atom types could be assigned arbitrarily. So by coloring nodes and thus partially breaking symmetry we can obtain a family of structures. Breaking symmetry means that resulting structure will have space group **H** which is a subgroup of space group **G** of the initial structure. Figs. 2 and 3 illustrate this, and use the same notations as in [36].

Topologies and Bärnighausen-like trees together form a regularization of the crystal structures configuration space. This regularization not only allows one to generate structures randomly, but also tells us if two structures are related to each other and even allows to enumerate closest neighbors of a structure in the configuration space.

There is one more extremely important option which comes from topological approach to structure generation and follows from the first step of the procedure for underlying net construction. In place of nodes we can put not just atoms, but also molecules or some complex structural units, i.e. perform the so-called *decoration* procedure [14]. Underlying topology thus becomes a blueprint of the crystal and encodes the method of structure assembly, while structural units become building blocks.

The topological approach should be efficient because it predetermines a rather uniform distribution of the initial structure configurations in the configuration space. Indeed, each optimized (high-symmetry) underlying net embedding should occupy some region (basin) in the configuration space, and these regions should not overlap. This means that starting from different topologies we cover the configuration space with initial structure configurations rather regularly. The richer the topological database, the more detailed is the configuration space coverage and more likely it is find all stable configurations by subsequent evolutionary procedure.

3. Performance tests

Two types of tests were performed:

i. Comparison of standard evolutionary global optimizations, with and without topological random generator and parameter control.
ii. Pure comparison of symmetric and topological random structure generators.

Comparison of symmetric and topological random structure generators was done for three systems Ca_4F_8, $Cu_4In_4S_8$ $Mg_4Al_8O_{16}$ (the former two at *ab initio* level of theory, and the latter one using an empirical forcefield). Comparison of evolutionary global optimizations (much more computationally expensive) was performed for $Mg_4Al_8O_{16}$ system, where the energy was computed using the same empirical forcefield.

For $Mg_4Al_8O_{16}$, structure relaxation and energy evaluation GULP software [37] was used. Simulation details: 28 atoms/cell, 50 structures per generation, external pressure 100 GPa, forcefield is given in Supplementary Materials. Since evolutionary algorithms are stochastic, for proper comparison it is necessary to run calculations with the same input many times and then compare statistics of runs obtained with different approaches; here we performed 100 runs of each type:

(1) "standard" runs, with symmetric random, but without parameter control and topological random,

Table 1
Efficiency of evolutionary USPEX searches for $Mg_4Al_8O_{16}$ with different settings. Figures are given in following format: average number of generations to get to the ground state, average number of structures to get to the ground state (with its standard deviation in parentheses).

	Without parameter control	With parameter control
Symmetric random	27, 1307(1086)	22, 1069(700)
Symmetric and topological		13, 609(569)
Topological random	10, 463(524)	8, 368(247)

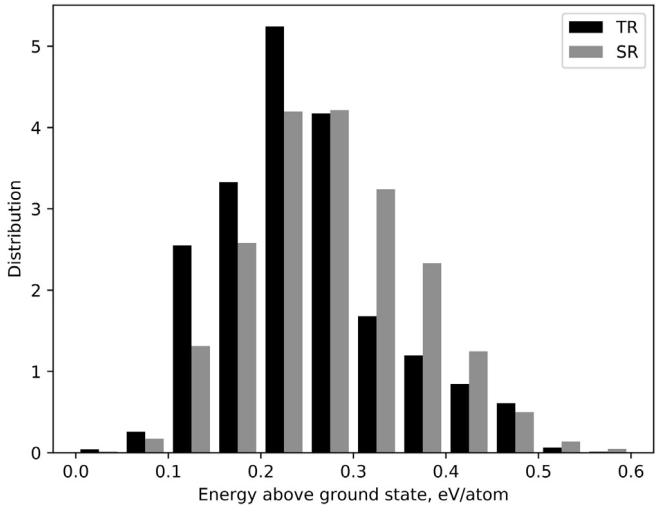

Fig. 4. Statistics for $Mg_4Al_8O_{16}$. Normalized distribution of structures generated by topological (TR) and symmetric (SR) random structures generators by their locally optimized (relaxed) energy. Total number of structures for TR is 8990 and for SR is 21 200.

(2) runs with symmetric random and parameter control, but without topological random,
(3) runs with topological random and parameter control, but without symmetric random,
(4) runs with topological random, but without symmetric random and parameter control,
(5) runs with both symmetric and topological random, and with parameter control.

In total, this gives 500 evolutionary runs, which in practice can be done only using empirical forcefield. Results of these tests are given in Table 1. Clearly, parameter control improves the efficiency of USPEX (here, by ∼20%) and topological random structure generator improves efficiency by almost 3 times. The combined effect of these improvements is a 3.55-fold speedup. We can also see clear superiority of topological random with respect to symmetric random.

Figs. 4–6 represent distribution of structures by their energy for $Mg_4Al_8O_{16}$, Ca_4F_8 and $Cu_2In_2S_8$ systems. In all these (very different by degree of complexity and type of chemical bonding) systems we see that topological random produces much greater percentage of low-energy structures. This advantage increases for more complex systems.

One can see that topological random much more easily produces low-energy structures, even though peaks for both generators are close. We can conclude that it is the ability of a generator to provide a variety of structures close to the ground state that has crucial importance for the efficiency of evolutionary optimization, rather than the energy of the majority of generated structures.

Recently, an idea was proposed [17] to sample a database of structure prototypes (based on actually known crystal structures from the Inorganic Crystal Structure Database) for preparing initial structures, and this was implemented in the CALYPSO code. This approach has some similarities with our method, but is much more limited: our approach, unlike the one of [17] is capable of predicting entirely new structures, not present in any database. Comparing the results, we find differences, too, but unexpectedly, there are large differences also in the results of symmetric random

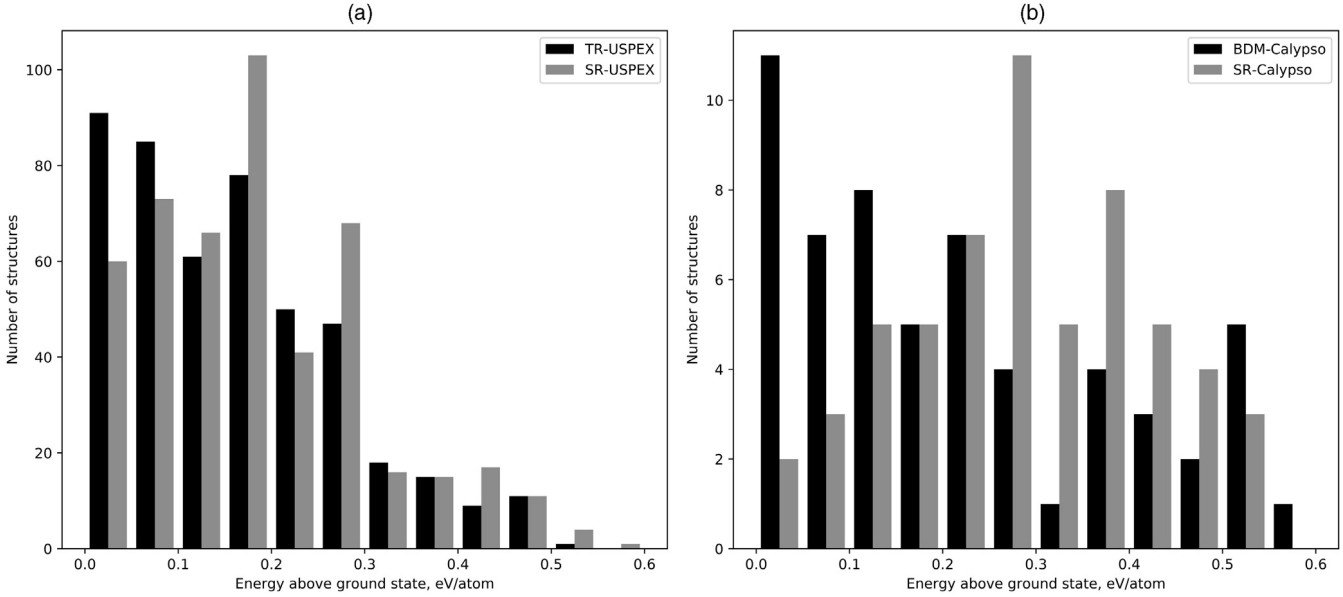

Fig. 5. Statistics for Ca_4F_8. (a) non-normalized energy distributions of structures generated with topological (TR-USPEX) and symmetric (SR-USPEX) random generators. The total number of structures for TR-USPEX is 475 and for SR-USPEX is 500 (b) non-normalized distribution of structures generated with "big data method" (BDM-Calypso) and random structure generator (SR-Calypso) by energy. Panel (b) is based on data from [17]. The total number of structures for BDM-Calypso and SR-Calypso is 58 each.

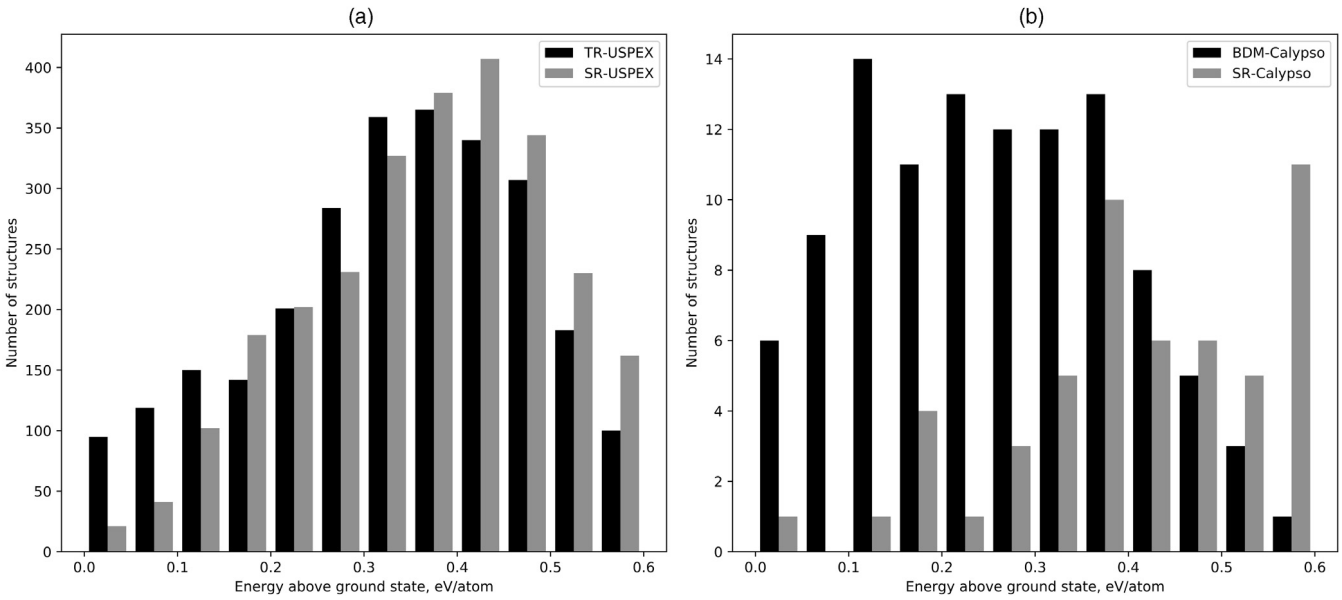

Fig. 6. Statistics for $Cu_4In_4S_8$. (a) non-normalized energy distributions of structures generated with topological (TR-USPEX) and symmetric (SR-USPEX) random generators. The total number of structures for TR-USPEX is 2842 and for SR-USPEX is 3000 (b) non-normalized energy distributions generated with "big data method" (BDM-Calypso) and random structure generator (SR-Calypso). Panel (b) is based on data from [17]. Total number of structures for BDM-Calypso and SR-Calypso is 106 each.

generator implemented in USPEX and in CALYPSO, which are hard to explain, because details of CALYPSO implementation of symmetric random have not been reported (for example, in CALYPSO results, symmetric random generates very few low-energy structures, compared with our symmetric generator) [2,15]

$Cu_4In_4S_8$ is a much more complex system, because directional covalent bonding allows for many local-minimum configurations, and many patterns of Cu–In order/disorder have very close energies. For this system topological random generator works again much better than symmetric random, while still keeping the ability to generate hundreds of diverse structures. The "BDM-Calypso" database algorithm, using only known structure types [17] gives a better energy distribution, but at a high price: only a small number of structures can be used.

Our approach not only increases efficiency of finding the ground-state structure, but also can be used for predicting numerous low-energy metastable structures. For example, for $Cu_2In_4S_8$ we have found not only the stable structure (I-42d space group, chalcopyrite-type, based on zincblende topology), but also metastable P-4m2 (also based on zincblende topology and 1.2 meV/atom higher in energy) and $Pmc2_1$ (based on wurtzite topology and 6.0 meV/atom above the ground state). Both these metastable semiconducting structures may be synthesized and may possess interesting properties.

4. Conclusions

We have presented a new method to create a diverse set of low-energy crystal structures - the topological structure generator - and demonstrated its power in its own right and in conjunction with an evolutionary algorithm. Our topological structure generator is not limited by the existing databases of structure types and is capable of generating an infinite number of new crystal structure types from a finite set of underlying topologies and group–subgroup relations. The topological structure generator is superior to the traditional symmetric random structure generator, and provides a much better starting point for evolutionary global optimization. We have implemented this approach in the USPEX code [1–3] and demonstrated that it is capable of bringing a nearly 3-fold speedup.

Additionally, we have proposed a method for evolving strengths of variation operators on-the-fly ("parameter control"), and proved its effectiveness. Its main idea is to encourage those variation operators that produce a more diverse set of lower-energy structures. This parameter control approach gives an additional ∼20% gain of efficiency in evolutionary searches.

These developments represent a major step forward. Now it is possible to tackle more complex and larger systems, and systematically explore both stable and low-energy metastable phases.

Acknowledgments

A.R.O. and P.V.B. thank Russian Science Foundation (grant 16-13-10459). V.A.B. thanks the Russian Foundation for Basic Research for a partial support within grant 17-43-630619.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cpc.2018.09.016.

References

[1] A.R. Oganov, C.W. Glass, J. Chem. Phys. 124 (2006) 244704.
[2] A.O. Lyakhov, A.R. Oganov, H.T. Stokes, Q. Zhu, Comput. Phys. Comm. 184 (2013) 1172.
[3] A.R. Oganov, A.O. Lyakhov, M. Valle, Acc. Chem. Res. 44 (2011) 227–237.
[4] S. Gödecker, J. Chem. Phys. 120 (2004) 9911–9917.
[5] S. Schönborn, S. Goedecker, S. Roy, A.R. Oganov, J. Chem. Phys. 130 (2009) 144108.
[6] M. Amsler, S. Goedecker, J. Chem. Phys. 133 (2010) 224104.
[7] Y. Wang, J. Lü, L. Zhu, et al., Phys. Rev. B 82 (2010) 094116.
[8] C.M. Freeman, J.M. Newsam, S.M. Levine, C.R.A. Catlow, J. Mater. Chem. 3 (1993) 531–535.
[9] M.U. Schmidt, U. Englert, J. Chem. Soc.–Dalton Trans. 10 (1996) 2077–2082.
[10] C.J. Pickard, R.J. Needs, Phys. Rev. Lett. 97 (2006) 045504.
[11] A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. 99 (2002) 12562–12566.
[12] R. Martoňák, A. Laio, M. Parrinello, Phys. Rev. Lett. 90 (2003) 075503.
[13] R. Martoňák, A. Laio, M. Bernasconi, C. Ceriani, P. Raiteri, F. Zipoli, M. Parrinello, Z. Krist. 220 (2005) 489–498.
[14] R. Martoňák, D. Donadio, A.R. Oganov, M. Parrinello, Nature Mater. 5 (2006) 623–626.
[15] Q. Zhu, A.R. Oganov, C.W. Glass, H. Stokes, Acta Crystallogr. B68 (2012) 215–226.
[16] A.O. Lyakhov, A.R. Oganov, M. Valle, Comput. Phys. Comm. 181 (2010) 1623–1632.

[17] C. Su, et al., J. Phys.: Condens. Matter 29 (2017) 165901.
[18] M. O'Keeffe, M. Eddaoudi, H. Li, T. Reineke, O.M. Yaghi, J. Solid State Chem. 152 (2000) 3–20.
[19] L. Öhrström, K. Larsson, Molecule-Based Materials: The Structural Network Approach, Elsevier, Amsterdam, 2005.
[20] V.A. Blatov, D.M. Proserpio, in: A.R. Oganov (Ed.), Modern Methods of Crystal Structure Prediction, Wiley-VCH, Berlin, ISBN: 978-3-527-40939-6, 2010, pp. 1–28.
[21] V.A. Blatov, A.P. Shevchenko, D.M. Proserpio, Cryst. Growth Des. 14 (7) (2014) 3576–3586.
[22] E.V. Alexandrov, A.V. Virovets, V.A. Blatov, E.V. Peresypkina, Chem. Rev. 115 (2015) 12286–12319.
[23] V.A. Blatov, Acta Crystallogr. 2007 (A63) (2007) 329–343.
[24] A.R. Oganov, M. Valle, J. Chem. Phys. 130 (2009) 104504.
[25] G. Bergerhoff, R. Hundt, R. Sievers, I.D. Brown, J. Chem. Inf. Model. 23 (1983) 66–69.
[26] A. Belsky, M. Hellenbrandt, V.L. Karen, P. Luksch, Acta Crystallogr., Sect. B: Struct. Sci. 58 (2002) 364–369.
[27] P. Villars, M. Berndt, K. Brandenburg, K. Cenzual, J. Daams, F. Hulliger, T. Massalski, H. Okamoto, K. Osaki, A. Prince, H. Putz, S. Iwata, J. Alloys Compd. 367 (2004) 293–297.
[28] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N.R. Serebryanaya, P. Moeck, R.T. Downs, A. Le Bail, Nucleic Acids Res. 40 (2012) D420–D427.
[29] F.H. Allen, S. Bellard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B.G. Hummelink-Peters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers, D.G. Watson, Acta Crystallogr., Sect. B Struct. Crystallogr. Cryst. Chem. 35 (1979) 2331–2339.
[30] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, Eur. J. Biochem. 80 (1977) 319–324.
[31] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[32] R.T. Downs, M. Hall-Wallace, Am. Mineral. 88 (2003) 247–250.
[33] P. Villars, K. Cenzual, Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds (on DVD), Release 2016/17, ASM International®, Materials Park, Ohio, USA.
[34] V.A. Blatov, D.M. Proserpio, Acta Crystallogr. A65 (2009) 202–212.
[35] International Tables for Crystallography, H. Wondratschek, U. Müller (Eds.), Symmetry Relations Between Space Groups, Vol. A1, Kluwer, Dordrecht, 2004.
[36] U. Muller, Symmetry Relations Between Crystal Structures, Oxford University Press, ISBN: 978-0-19-966995-0, 2013.
[37] J.D. Gale, JCS Faraday Trans. 93 (1997) 629.