Universal Structure Predictor: Evolutionary Xtallography

Code by **P.V. Bushlanov, A.I. Samtsevich, S.V. Lepeshkin, A.R. Oganov**

based and improved upon earlier version by
A.R. Oganov, Z. Allahyari, C.W. Glass, A.O. Lyakhov, Q. Zhu, G.-R. Qian, H.T. Stokes, V. Stevanovic, S. Lepeshkin, E. Mazhnik, and others

with contributions from
M. Galasso.

MANUAL

Version 2021.0, October 27, 2021.

https://uspex-team.org

Contents

1 Features, aims and history of USPEX

1.1 Overview

From the beginning in 2004, non-empirical crystal structure prediction was the main aim of the USPEX project. In addition to this, USPEX also allows one to predict a large set of robust metastable structures and perform several types of simulations using various degrees of prior knowledge. Starting from 2010, our code explosively expanded to other types of problems, and from 2012 includes many complementary methods.

The problem of crystal structure prediction is very old and does, in fact, constitute the central problem of theoretical crystal chemistry. In 1988 John Maddox[1] wrote that:

> "One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition... Solids such as crystalline water (ice) are still thought to lie beyond mortals' ken".

It is immediately clear that the problem at hand is that of global optimization, *i.e.*, finding the global minimum of the free energy of the crystal (per mole) with respect to variations of the structure. To get some feeling of the number of possible structures, let us consider a simplified case of a fixed cubic cell with volume V, within which one has to position N identical atoms. For further simplification let us assume that atoms can only take discrete positions on the nodes of a grid with resolution δ. This discretization makes the number of combinations of atomic coordinates C finite:

$$C = \frac{1}{(V/\delta^3)} \frac{(V/\delta^3)!}{[(V/\delta^3)-N]!N!} \tag{1}$$

If δ is chosen to be a significant fraction of the characteristic bond length (e.g., $\delta = 1$ Å), the number of combinations given by eq. 1 would be a reasonable estimate of the number of local minima of the free energy. If there are more than one type of atoms, the number of different structures significantly increases. Assuming a typical atomic volume ∼10 Å^3, and taking into account Stirling's formula ($n! \approx \sqrt{2\pi n}(n/e)^n$), the number of possible structures for an element A (compound AB) is 10^{11} (10^{14}) for a system with 10 atoms in the unit cell, 10^{25} (10^{30}) for a system with 20 atoms in the cell, and 10^{39} (10^{47}) for a system with 30 atoms in the unit cell. These numbers are enormous and practically impossible to deal with even for small systems with a total number of atoms $N \sim 10$. Even worse, complexity increases exponentially with N. It is clear then, that point-by-point exploration of the free energy surface going through all possible structures is not feasible, except for the simplest systems with ∼1-5 atoms in the unit cell.

USPEX[2;3] employs an evolutionary algorithm devised by A.R. Oganov and C.W. Glass, with major subsequent contributions by A.O. Lyakhov, Q. Zhu, G.R. Qian, P. Bushlanov,

Z. Allahyari, S. Lepeshkin and A. Samtsevich. Its efficiency draws from the carefully designed variation operators, while its reliability is largely due to the use of state-of-the-art *ab initio* simulations inside the evolutionary algorithm. The strength of evolutionary simulations is that they do not require any system-specific knowledge (except the chemical composition) and are self-improving, i.e. in subsequent generations increasingly good structures are found and used to generate new structures. This allows a "zooming in" on promising regions of the energy (or property) landscape (Fig. 1). Furthermore, by carefully designing variation operators, it is very easy to incorporate additional features into an evolutionary algorithm.

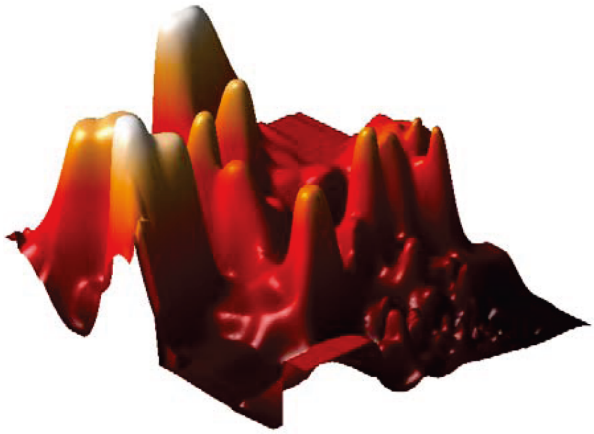

Figure 1: **2D projection of the reduced landscape of Au_8Pd_4, showing clustering of low-energy structures in one region.** The landscape was produced using the method of Oganov & Valle (2009).

A major motivation for the development of USPEX was the discovery of the post-perovskite phase of $MgSiO_3$ (Fig. 2), which was made in 2004[4;5] and has significantly changed models of the Earth's internal structure. In mid-2005 we had the first working version of USPEX. By September 2010, when USPEX was publicly released, the user community numbered nearly 200, over 800 users in May 2012, over 2100 in December 2014, and over 4500 in December 2018.

The popularity of USPEX is due to its extremely high efficiency and reliability. This was shown in the First Blind Test for Inorganic Crystal Structure Prediction[6], where USPEX outperformed the other methods it was tested against (simulated annealing and random sampling). Random sampling (a technique pioneered for structure prediction by Freeman and Schmidt in 1993 and 1996, respectively, and since 2006 revived by Pickard[7] under the name AIRSS) is the simplest, but also the least successful and computationally the most expensive strategy. Even for small systems, such as GaAs with 8 atoms/cell, these advantages are large (random sampling requires on average 500 structure relaxations to find the ground state in this case, while USPEX finds it after only $\sim$ 30 relaxations! (Fig. 3)). Due to the exponential scaling of the complexity of structure search (eq. 1), the advantages of USPEX increase exponentially with system size. For instance, 2 out

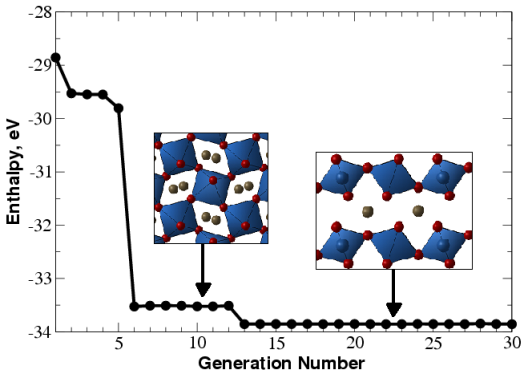

Figure 2: **Prediction of the crystal structure of $MgSiO_3$ at 120 GPa (20 atoms/cell).** Enthalpy of the best structure as a function of generation is shown. Between the 6^{th} and 12^{th} generations the best structure is perovskite, but at the 13^{th} generation the global minimum (post-perovskite) is found. This simulation was performed in 2005 using one of the first versions of USPEX combined with *ab initio* calculations. It used no experimental information and illustrates that USPEX can find both the stable and low-energy metastable structures in a single simulation. Each generation contains 30 structures. This figure illustrates the slowest of ∼10 calculations performed by the very first version of USPEX — and even that was pretty fast!

of 3 structures of SiH_4 predicted by random sampling to be stable[7], turned out to be unstable[8]; and similarly random sampling predictions were shown[9] to be incorrect for nitrogen[10] and for SnH_4 (compare predictions[11] of USPEX and of random sampling[12]).

For larger systems, random sampling tends to produce almost exclusively disordered structures with nearly identical energies, which decreases the success rate to practically zero, as shown in the example of $MgSiO_3$ post-perovskite with 40 atoms/supercell — random sampling fails to find the correct structure even after 120,000 relaxations, whereas USPEX finds it after several hundred relaxations (Fig. 4).

Random sampling runs can easily be performed with USPEX — but we see this useful mostly for testing. Likewise, the Particle Swarm Optimization (PSO) algorithm for cluster and crystal structure prediction (developed by A.I. Boldyrev and re-implemented by Wang, Lu, Zhu and Ma) has been revamped and implemented on the basis of USPEX with minor programming work as a corrected PSO (corPSO) algorithm, which outperforms previous versions of PSO. Still, any version of PSO is rather weak and we see the PSO approach suitable mostly for testing purposes, if anyone wants to try. A very powerful new method, complementary to our evolutionary algorithm, is evolutionary metadynamics[13], a hybrid of Martoňák's metadynamics and the Oganov-Glass evolutionary approach. This method is powerful for global optimization and for harvesting low-energy metastable structures, and even for finding possible phase transition pathways. For detailed investigations of phase transition mechanisms, additional methods are implemented: variable-cell NEB method[14] and transition path method[15] in the version[16].

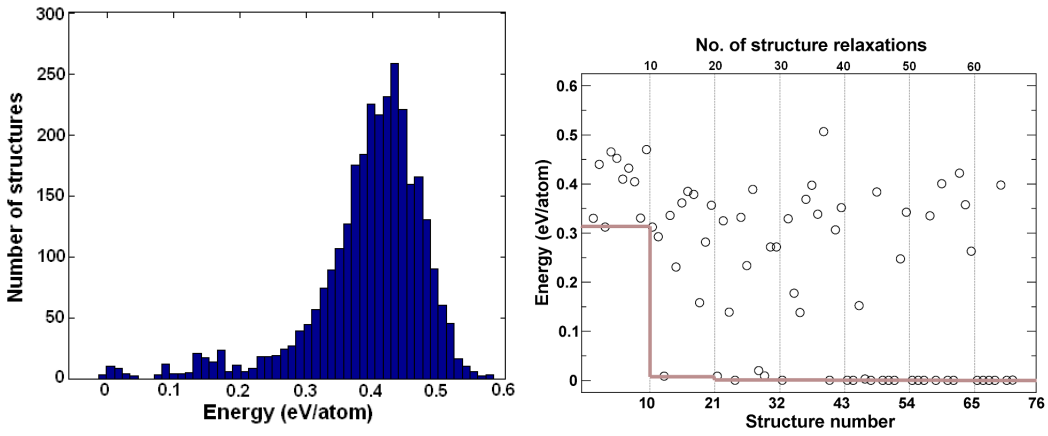

Figure 3: **Structure prediction for GaAs.** a) Energy distribution for relaxed random structures, b) progress of an evolutionary simulation (thin vertical lines show generations of structures, and the grey line shows the lowest energy as a function of generation). All energies are relative to the ground-state structure. The evolutionary simulation used 10 structures per generation. In addition, the lowest-energy structure of the previous generation survived into the next generation.

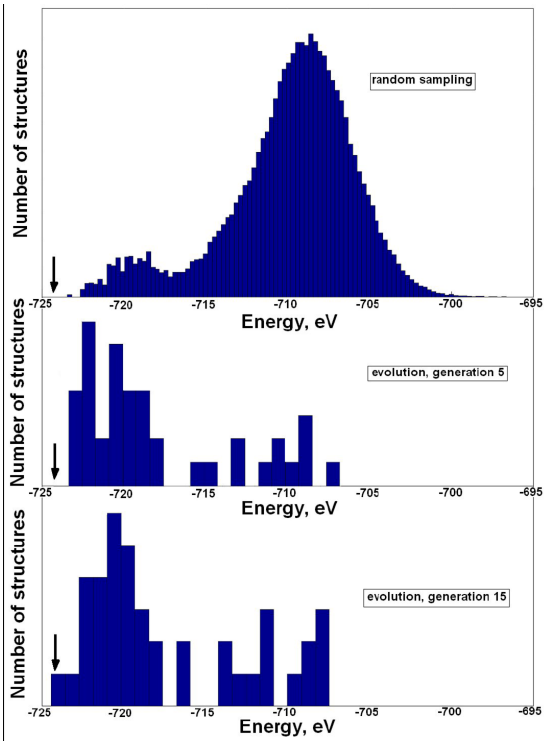

Figure 4: **Sampling of the energy surface: comparison of random sampling and USPEX for a 40-atom cell of $MgSiO_3$ with cell parameters of post-perovskite.** Energies of locally optimized structures are shown. For random sampling, 1.2×10^5 structures were generated (none of which corresponded to the ground state). For USPEX search, each generation included 40 structures and the ground-state structure was found within 15 generations. The energy of the ground-state structure is indicated by the arrow. This picture shows that "learning" incorporated in evolutionary search drives the simulation towards lower-energy structures.

1.2 Features of USPEX

- Prediction of the stable and metastable structures knowing only the chemical composition. Simultaneous searches for stable compositions and structures are also possible.
- Incorporation of partial structural information is possible:
  - constraining search to fixed experimental cell parameters, or fixed cell shape, or fixed cell volume;
  - starting structure search from known or hypothetical structures ;
  - assembling crystal structures from predefined molecules.
- Efficient constraint techniques, which eliminate unphysical and redundant regions of the search space. Cell reduction technique (Oganov & Glass, 2008).
- Niching using fingerprint functions (Oganov & Valle, 2009; Lyakhov, Oganov, Valle, 2010).
- Initialization using fully random approach, or using space groups and cell splitting techniques (Lyakhov, Oganov, Valle, 2010). Use of powerful topological structure generator (Bushlanov, Blatov, Oganov, 2018).
- On-the-flight analysis of results — determination of space groups (and output in CIF-format), calculation of the hardness, order parameters, *etc.*
- Prediction of the structure of nanoparticles and surface reconstructions.
- Restart functionality, enabling calculations to be continued from any point along the evolutionary trajectory.
- Powerful visualization and analysis techniques implemented in the STM4 code (by M. Valle), fully interfaced with USPEX.
- USPEX is interfaced with VASP, GULP, LAMMPS, ABINIT, Quantum Espressocodes. See full list of supported codes in Subsection 2.5.
- Submission of jobs from local workstation to remote clusters and supercomputers is possible.
- Options to optimize physical properties other than energy — *e.g.*, hardness (Mazhnik & Oganov, 2019), density (Zhu et al., 2011), band gap and dielectric constant (Zeng et al., 2014), and many other properties.
- Starting from version 9.4.1, USPEX has a Python-based runner of the code (`USPEX` Python module), providing a number of useful command line options.

1.3 Key USPEX papers

1. Oganov A.R., Glass C.W. (2006). Crystal structure prediction using evolutionary algorithms: principles and applications. *J. Chem. Phys.*, **124**, 244704.

2. Oganov A.R., Stokes H., Valle M. (2011). How evolutionary crystal structure prediction works — and why. *Acc. Chem. Res.*, **44**, 227–237.

3. Lyakhov A.O., Oganov A.R., Stokes H., Zhu Q. (2013). New developments in evolutionary structure prediction algorithm USPEX. *Comp. Phys. Comm.*, **184**, 1172–1182.

4. Zhu Q., Oganov A.R., Glass C.W., Stokes H. (2012). Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Cryst. B*, **68**, 215–226.

5. Zhu Q., Li L., Oganov A.R., Allen P.B. (2013). Evolutionary method for predicting surface reconstructions with variable stoichiometry. *Phys. Rev. B*, **87**, 195317.

6. Zhu, Q., Sharma V., Oganov A.R., Ramprasad R. (2014). Predicting polymeric crystal structures by evolutionary algorithms. *J. Chem. Phys.*, **141**, 154102.

7. Lyakhov A.O., Oganov A.R., Valle M. (2010). Crystal structure prediction using evolutionary approach. In: Modern methods of crystal structure prediction (ed: A.R. Oganov), Berlin: Wiley-VCH.

8. Oganov A.R., Ma Y., Lyakhov A.O., Valle M., Gatti C. (2010). Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. *Rev. Mineral. Geochem.*, **71**, 271–298.

9. Duan, D., Liu, Y., Tian, F., Li, D., Huang, X., Zhao, Z., Yu, H., Liu, B., Tian, W., Cui, T. (2014). Pressure-induced metallization of dense $(H_2S)_2H_2$ with high-T_c superconductivity. *Sci. Rep.*, **4**, 6968.

10. Bushlanov P.V., Blatov V.A., Oganov A.R. (2019). Topology-based crystal structure generator. Comp. Phys. Comm., DOI: 10.1016/j.cpc.2018.09.016.

11. Lepeshkin S.V., Baturin V.S., Uspenskii Yu.A., Oganov A.R. (2019). Method for simultaneous prediction of atomic structure of nanoclusters in a wide area of compositions. J. Phys. Chem. Lett. 10, 102-106.

12. Allahyari Z., Oganov A.R. (2018). Multi-objective optimization as a tool for materials design. In:.Handbook of Materials Modeling (ed. W. Andreoni, S. Yip). Volume 2 Applications: Current and Emerging Materials. Springer Verlag.

1.4 Version history

v.1 — Evolutionary algorithm without local optimization. Real-space representation, interface with VASP. Experimental version. October 2004.

v.2 — CMA-ES implementation (CMA-ES is a powerful global optimization method developed by N. Hansen). Experimental version. January 2005.

v.3 — Evolutionary algorithm with local optimization.

v.3.1 — Working versions, sequential. Major basic developments.

3.1.4-3.1.5 — First production version. Based largely on heredity with slice-shifting and with minimum-parent contribution (hard-coded to be 0.25). May 2005.

3.1.8 — Adaptive k-point grids. 15/10/2005.

3.1.11 — Restart from arbitrary generation. Experimental version. 04/11/2005.

3.1.12 — Production version based on v.3.1.11, variable slice-shift mutation. 11/11/2005.

3.1.13 — Adaptive scaling volume. 29/11/2005.

3.1.14 — Basic seed technique. 29/11/2005 (debugged 6/12/2005).

v.3.2 — Massively parallel version.

v.4 — Unified parallel/sequential version.

4.1.1 — Lattice mutation. 20/12/2005 (debugged 10/01/2006).

4.2.1 — Interfaced with SIESTA. Initial population size allowed to differ from the running population size. 24/01/2006 (debugged 20/04/2006).

4.2.3 — Relaxation of best structures made optional. Version with fully debugged parallelism. 25/04/2006.

4.4.1 — Interfaced with GULP. 08/05/2006.

v.5 — Completely rewritten and debugged version, clear modular structure of the code.

5.1.1 — Atom-specific permutation, code interoperability, on-the-fly reading of parameters from `INPUT_EA.txt`. 20/12/2006.

5.2.1 — SIESTA-interface for Z-matrix, rotational mutation operator. 01/03/2007.

v.6 — Production version.

6.1.3 — To efficiently fulfill hard constraints for large systems, an optimizer was implemented within USPEX. 07/06/2007.

6.2 — Development version.

6.3.1–6.3.2 — Introduced angular constraints for cell diagonals. Completely rewritten remote submission. Improved input format. Further extended standard tests. 07/12/2007.

6.3.3 — X-com grid interface (with participation of S. Tikhonov and S. Sobolev). 05/03/2008.

6.4.1 — Fingerprint functions for niching. 07/04/2008.

6.4.4 — Space group recognition. Fast fingerprints (from tables). 05/05/2008.

6.5.1 — Split-cell method for large systems. Easy remote submission. Variable number of best structures (clustering). 16/07/2008.

6.6.1 — A very robust version — improved fingerprint and split-cell implementations. 13/08/2008.

6.6.3 — Heredity with multiple parents implemented. 01/10/2008.

6.6.4 — Added a threshold for parents participating in heredity (niching). 03/10/2008.

6.6.6 — First implementation of multicomponent fingerprints. 04/12/2008.

6.6.7, 6.7.1 and 6.7.2 — Implemented quasi-entropy to measure the diversity of the population. 10/12/2008.

v.7 — Production version, written to include variable composition.

7.1.1–7.1.7 — Series of improved versions. Version 7.1.7 has been distributed to ∼200 users. Variable composition partly coded, most known bugs fixed, improved tricks based on energy landscapes. Improved cell splitting, implemented pseudo-subcells. Implemented multicomponent fingerprints (much more sensitive to the structure than one-component fingerprints). 28/04/2009 (version finalized 28/05/2009).

7.2.5 — First fully functional version of the variable-composition method. Introduced transmutation operator and compositional entropy. 06/09/2009.

7.2.7 — Thoroughly debugged, improved restart capabilities, improved seeding, introduced perturbations within structure relaxation. 25/09/2009, further improved in versions 7.2.8/9.

7.3.0 — Full fingerprint support in the variable-composition code, including niching. "Fair" algorithm for producing the first generation of compositions. 22/10/2009.

7.4.1 — Introduced coordinate mutation based on local order[17]. Heredity and transmutation are also biased by local order. Introduced computation of the hardness and new types of optimization by hardness and density. 04/01/2010.

7.4.2 — Implementation of multiple-parents heredity biased by local order. 15/01/2010.

7.4.3 — Implementation of new types of optimization (to maximize structural order and diversity of the population). Implemented antiseeds, eliminated parameters `volTimeConst`, `volBestHowMany`. 24/01/2010.

v.8 — Production version, written to include new types of optimization.

8.1.1–8.2.8 — Development versions. Local order and coordinate mutation operator. Softmutation operator. Calculation and optimization of the hardness. Optimization of the dielectric susceptibility. Prediction of the structure of nanoparticles and surfaces. Implementation of point groups. Greatly improved overall performance. Option to perform PSO simulations (not recommended for applications, due to PSO's inferior efficiency — so use only for testing purposes). Parameter `goodBonds` transformed into a matrix and used for building nanoparticles. 22/09/2010.

8.3.1 — Optimization of dielectric constants, cleaned-up input. 08/10/2010.

8.3.2 — For clusters, introduced a check on connectivity (extremely useful), `dynamicalBestHM=2` option improved, as well as mechanism for producing purely softmutated generations. Improved fingerprints for clusters. Interface to Quantum Espresso and CP2K codes. 11/10/2010.

8.4 — Improved antiseed functionalities and several improvements for nanoparticles. Development branches for surface reconstructions, pseudo-metadynamics, molecular crystals.

8.5.0 — Initialization of the first random generation using the space group code of H. Stokes added. New formulation of metadynamics implemented and finalized, for now in a separate code. Several debugs for varcomp, antiseeds, nanoparticles, computation of hardness. 18/03/2011.

8.5.1 — Space group initialization implemented for cases of fixed unit cell, variable composition, and subcells. 20/04/2011.

8.6.0 — Added space group determination program from H. Stokes. Merger with the updated code for molecular crystals (including space group initialization). Fixed a bug for SIESTA (thanks to D. Skachkov). 06/05/2011.

8.6.1–8.7.2 — Improved symmetric initialization for the case of a fixed cell. Implemented optimization of dielectric constants (using GULP and VASP), band gap (using VASP), and DOS at the Fermi level (VASP). Graphical output enabled. Improved softmutation (by using better criteria for mode and directional degeneracies) and heredity (by using energy-order correlation coefficient and cosine formula for the number of trial slabs) operators. Most variables now have default values, which enables the use of very short input files. Shortened and improved the format of log-files. 13/11/2011.

8.7.5 — Graphical output now includes many extra figures. Added utility to extract all structures close to convex hull for easier post-processing. 21/03/2012.

v.9 — Production version, made more user-friendly and written to include new types of functionality and to set the new standard in the field.

9.0.0 — Evolutionary metadynamics and VCNEB codes added to USPEX package, added tensor version of metadynamics, added additional figures and post-processing tools, cleaned the code output. A few parameters removed from the input. Improved softmutation. April 2012.

9.1.0 — Release version. Cleaned up, documented. The user community is >800 people. Released 28/05/2012.

9.2.0 — Working GEM. Constant development of the GEM code. Space group determination tolerance is now an input parameter. Improved default for number of permutations. July-August 2012.

9.2.1–9.2.3 — Improved GEM, more diverse populations and supercell sizes, improved mode selection. September-October 2012.

9.2.4–9.2.6 — (9.2.4 is a release version). Intelligent defaults for most input parameters. Improved symmetric initialization for clusters. Order-enhanced heredity for nanoparticles. New parameter to tune the tolerance for the space group determination. New property (quasientropy) can be optimized. Fully integrated VCNEB code. November-December 2012.

9.2.7. — Release version. Enabled optimization of order for alloys, without structure relaxation (for easy creation of quasirandom structures, based on the more general definition than the so-called "special quasi-random structures"). Symmetry generation was improved (particularly important for fixed-cell calculations). For fixed-cell calculations, one can now specify the cell parameters, not only in the form of a 3×3 matrix, but also as a row of six values (three lengths in Angstroms and three angles in degrees). For the maximum number of permutation swaps

(parameter `howManySwaps`), we have introduced an intelligent default. Added new tests, and cleaned and reran the old ones. Added interface to CASTEP (thanks to Z. Raza, X. Dong and AL). User community 1160 people. 30/12/2012.

9.3.0–9.3.3 — Fixed a bug in generation of random symmetric structures (this bug appeared in 9.2.7). Significantly simplified input and output. Created file `OUTPUT.txt` with the most important information. Enabled split-cell trick for molecular crystals. Improved variable-composition calculations by allowing one to specify initial compositions. Added interface to CASTEP and LAMMPS. Added new test cases. 20/03/2013.

9.3.4 — Release version, cleaned up. 25/03/2013.

9.3.5 — Added code for prediction of 2D-crystals. 19/04/2013.

9.3.6 — Incorporated plane groups for 2D-crystals. 29/04/2013.

9.3.8 — Incorporated plane groups for 1D-polymer crystals, improved variables of stoichiometry for surfaces. 19/06/2013.

9.3.9 — Released version. Significantly improved version, improved user-friendliness, new functionalities (2D-crystals, GEM) made more robust, improvements in the variable-composition algorithm (and enabled support for single-block calculations, *i.e.* fixed-composition searches with variable number of atoms in the cell), fully functional surface calculations, new optimization types (can optimize band gaps, dielectric constants, and newly invented figure of merit of dielectric materials). Interfaces with LAMMPS and ATK are documented in new test cases. Continuously updated with minor debugs (last debug 10/02/2014). 19/07/2013.

9.4.1 — A major upgrade, greatly improved user-friendliness (automatic estimate of volumes and of percentages of variation operators for each case), new functionalities (optimization of elastic properties and Chen's model of hardness, prediction of polymeric structures, anti-compositions, automatic analysis of statistics, improved seed technique), first release of GEM (generalized evolutionary metadynamics), provided a set of real-life examples of USPEX calculations, test cases, documentation. More than 2100 users. Released 30/12/2014.

9.4.2 — Release version, compatible with Octave 3.4. Convex hull code rewritten. Interface with MOPAC implemented. Default values for `goodBonds`, `valences`, `IonDistances` enabled. More robust for ternary, quaternary, and more complex variable-composition searches. Robust TPS implementation (only for developers, will be available for users soon). More than 2200 users. Released 21/03/2015.

9.4.3 — Release version. It includes fixing a number of bugs (which should slightly speed up performance), interface with MOPAC, improved documentation. Released 10/08/2015.

9.4.4 — Release version. It includes fix for space group determination and other problems reported by users, improved documentation and examples, full Octave 3.4 compatibility and partial Octave 3.6/3.8/4.0 support. This version should be nearly bug-free and is a milestone towards a very major upgrade, which will be made available in version 10. Released 05/10/2015.

10.1 - Experimental version, released as a virtual machine on 01/08/2018.

10.2 - Release version. Distributed as a compiled code, so users no longer need to have Matlab. Has no known bugs and contains a very large number of new features and improvements. Random topological structure generator and automatic parameter control greatly speed up cal-

culations. Prediction of (collinear) magnetic materials is enabled. Many new types of fitness implemented: magnetization, birefringence, thermoelectric figure of merit ZT, fracture toughness. Fitness can be minimized or maximized, and we can input mathematical expressions as fitness. Pareto optimization of several fitnesses is enabled. USPEX has been interfaced with Gaussian, MOPAC, DFTB, ORCA, FHI-aims, ABINIT. Symmetry determination is now done using SPGLIB. Symmetrization can also be optionally performed during calculation of physical properties, making it cheaper and more robust. 80 layer symmetry groups are utilized for generating initial population of 2D-structures. Variable-composition prediction of 2D-crystals is enabled. For surface structure prediction, we have enabled all possible surface supercells and output the surface phase diagram. Variable-cell NEB method has been greatly improved (made a few times faster). Utility "pmpaths" of V. Stevanovic has been added to predict the likeliest phase transition mechanisms (which can then be directly input into the VCNEB code). Released 19/01/2019.

10.3 - Similar to version 10.2, with some critical fixed-bugs in job submission, and some minor fixed-bugs in the code. Released 15/10/2019.

10.4 - Several improvements, bug fixes, new features. Local and remote submission files are improved. VCNEB and PSO codes are improved. Interfaces to QE, DMACRYS, VASP, and DFTB+ are improved. Interfaces to Abinit and CRYSTAL codes are added. Half-metallicity fitness is added. Added Mazhnik-Oganov model for calculation of hardness and fracture toughness. All python scripts are translated to python3 - no more python2 needed for USPEX. Released 05/11/2020.

10.5 - minor bug fixes. Added optimization of more than one quantity expressed by formula. Improved interface with DFTB+ and GULP5.2. Added "USPEX -u" feature, which allows one to update the package to the latest version without downloading the entire USPEX distribution. Enabled machine learning calculation of the elastic moduli (their optimization is available as optType=1201-1207), using graph convolutional neural network, and added Example35 to show how to use this feature. Released 08/07/2021.

2021.0 - This version is a beginning of reimplementation of previously developed techniques (with improvements, bug fixes, and more transparent code structure to facilitate further developments). In this version we have reimplemented structure/compound prediction for bulk crystals. Released 27/10/2021.

NOTE: All regimes of 10.5 version will be reimplemented in subsequent releases.

2 Getting started

2.1 How to obtain USPEX

USPEX can be downloaded at:

https://uspex-team.org

In the download page you will need to register and soon thereafter will receive a password for downloading USPEX. There are separate packages for USPEX itself, manual files and files for its installation.

2.2 Necessary citations

Whenever using USPEX, in all publications and reports you must cite the original papers, for example, in the following way:

> "Crystal structure prediction was performed using the USPEX code[2;13;18], based on an evolutionary algorithm developed by Oganov, Glass, Lyakhov and Zhu and featuring local optimization, real-space representation and flexible physically motivated variation operators".

Consult the `OUTPUT.txt` file for some of the most important references.

2.3 Bug reports

Like any large code, USPEX may have bugs. If you see strange behavior in your simulations, please report it to us in USPEX Google group at:

https://groups.google.com/forum/#!forum/uspex

Please describe your problem in details and attach `INPUT.txt`, `OUTPUT.txt`, `log` and other related files when you report a problem. You can also send the questions or problem descriptions to (bugreport@uspex-team.org)).

2.4 On which machines USPEX can be run

USPEX can be used on linux platform — all you need is one CPU and installed python and — using its special remote submission mechanism, USPEX will be able to connect to any remote machine and use it for calculations.

2.5 Codes that can work with USPEX

Trial structures generated by USPEX are relaxed and then evaluated by an external code interfaced with USPEX. Based on the obtained ranking of relaxed structures, USPEX generates new structures — which are again relaxed and ranked. Our philosophy is to use existing well-established *ab initio* (or classical forcefield) codes for structure relaxation and energy calculations. Currently, USPEX is interfaced with:

- VASP — `https://www.vasp.at/`
- GULP — `http://nanochemistry.curtin.edu.au/gulp/`
- LAMMPS — `http://lammps.sandia.gov/`
- Quantum Espresso — `http://www.quantum-espresso.org/`
- ABINIT — `https://www.abinit.org`

We chose these codes based on 1) their efficiency for structure relaxation; 2) robustness; and 3) popularity. Of course, there are other codes that can satisfy these criteria, and in the future we can interface USPEX to them.

2.6 How to install USPEX

USPEX v.2021.0 is compiled using Cython compiler. After you download the archive with USPEX, you need to unpack and install it using *pip* installer.

```
pip install USPEX-2021.0.1-<ver>-<ver>-linux_x86_64.whl
```

Chose one of *whl* files in archive depending on version of python you are using.

The file **random_cell** should be copied manually into any directory listed in user's *PATH* environment setting. For example into `~/.local/bin/` folder.

2.7 How to run USPEX

To run USPEX, you need to have **Python** and to have the executable of the external code on the compute nodes that you use for relaxing structures and computing their energies (see Subsection 2.5 for a list of supported codes). To set up your calculation, find an example (included in archive) similar to what you want to do, and start by editing `input,uspex`. The variables of this crucial file are described in Section 4 below. Then, gather the files needed for the external code performing structure relaxation in the `Specific/` folder — the executable (*e.g.*, `vasp`), and such files as (if you use VASP)

INCAR_1, INCAR_2, ..., INCAR_*N*, and POTCAR_*A*, POTCAR_*B*, ..., where *A*, *B*, ... are the symbols of the chemical elements of your compound.

To run USPEX just type:

```
uspex -r
```

File log will contain information on the progress of the simulation and, if any, errors (send these to us, if you would like to report a bug). File OUTPUT.txt will contain details of the calculation and an analysis of each generation.

3 Overview of input and output files

Input/output files depend on the external code used for structure relaxation.

An important technical element of our philosophy is the multi-stage strategy for structure relaxation. Final structures and energies must be high-quality, in order to correctly drive evolution. Most of the newly generated structures are far from local minimum and their high-quality relaxation is extremely expensive. This cost can be offset if the first stages of relaxation are done with cruder computational conditions — only at the last stages is there a need for high-quality calculations. The first stages of structure relaxation can be performed with cheaper approaches or lower computational conditions (basis set, k-points sampling, pseudopotentials) or level of approximation (forcefields *vs.* LDA *vs.* GGA) and even different structure relaxation code (see Subsection 2.5 for a list of supported codes) during structure relaxation of each candidate structure. We strongly suggest you initially optimize the cell shape and atomic positions at constant unit cell volume, and only then perform full optimization of all structural variables. While optimizing at constant volume, you do not need to worry about Pulay stresses in plane-wave calculations — it is OK to use a small basis set; however, for variable-cell relaxation you will need a high-quality basis set. For structure relaxation, you can often get away with a small set of k-points — but don't forget to sufficiently increase this at the last stage(s) of structure relaxation, to get accurate energies. Use your (and our) wisdom, be a strategist, and remember that poor relaxation can ruin your results.

3.1 Input files

Suppose that the directory where the calculations are performed is ∼`/StructurePrediction`. This directory will contain:

- file `input.uspex`, thoroughly described in Section 4.
- Subdirectory ∼`/StructurePrediction/Specific/` with VASP, SIESTA or GULP (*etc.*) executables, and enumerated input files for structure relaxation — `INCAR_1`, `INCAR_2`, ..., and pseudopotentials. You can actually alter this filenames (see 4.7)
- Subdirectory ∼`/StructurePrediction/Seeds` — contains files with seed structures. Seed structures should be in VASP5 POSCAR format grouped into folders. Which folder will be used for which generation is specified in 4.5.9. For now this feature allows only atomic sructures to be supplied as seeds. But we plan to extend it on molecular structures.
- Files with molecule definitions (see 4.9).

3.1.1 Specific/ folder

Executables and enumerated input files for structure relaxation (using external codes, like VASP, SIESTA, GULP, ...) should be put in subdirectory ∼/StructurePrediction/Specific/

- For VASP, put files INCAR_1, INCAR_2, . . . , *etc.*, defining how relaxation and energy calculations will be performed at each stage of relaxation (we recommend at least 3 stages of relaxation), and the corresponding POTCAR_* files with pseudopotentials. *E.g.*, INCAR_1 and INCAR_2 perform very crude structure relaxation of both atomic positions and cell parameters, keeping the volume fixed, INCAR_3 performs full structure relaxation under constant pressure with medium precision, INCAR_4 performs very accurate calculations. Each higher-level structure relaxation starts from the results of a lower-level optimization and improves them. POTCAR files of all relevant elements should also be in Specific/ folder, for instance POTCAR_C, POTCAR_O, *etc.*
- For GULP, files goptions_1, goptions_2, . . . , and ginput_1, ginput_2, . . . must be present. The former specify what kind of optimization is performed, the latter specify the details (interatomic potentials, pressure, temperature, number of relaxation iterations, *etc.*).
- For Quantum Espresso, files qEspresso_options_1, qEspresso_options_2, . . . , must be present. All files should be the normal QE input files with all parameters except atom coordinates, cell parameters and k-points (these will be written by USPEX at the end of the file). We recommend performing a multi-step relaxation. For instance, qEspresso_options_1 does a crude structure relaxation of atomic positions with fixed cell parameters, qEspresso_options_2 does full structure relaxation under constant external pressure with medium precision; and qEspresso_options_3 does very accurate calculations.

INCAR_* files in Specific/ folder for VASP To use USPEX correctly, you should carefully edit the files in Specific/ folder to control the structure relaxation in USPEX. We take example of VASP as an external code:

- Your final structures have to be well relaxed, and energies — precise. The point is that your energy ranking has to be correct (to check this, look at E_series.pdf file in the output).
- Your POTCAR files: To yield correct results, the cores of your pseudopotentials (or PAW potentials) should not overlap by more than 10–15%.
- To have accurate relaxation at low cost, use the multistage relaxation with at least three stages of relaxation for each structure, *i.e.* at least three INCAR files (INCAR_1, INCAR_2, INCAR_3, . . .). We usually set 4–5 stages of relaxation.

- Your initial structures will be usually very far from local minima, in such cases it helps to relax atoms and cell shape at constant volume first (ISIF=4 in INCAR_1,2), then do full relaxation (ISIF=3 in INCAR_3,4), and finish with a very accurate single-point calculation (ISIF=2 and NSW=0 in INCAR_5).

  Exceptions: when you do fixed-cell predictions, and also in evolutionary metadynamics (except full relaxation) you must have ISIF=2.

- When your volume does not change, you can use default plane wave cutoff. When you optimize cell voluem (ISIF=3), you must increase it by 30–40%, otherwise you get a large Pulay stress. Also your convergence criteria can be loose in the beginning, but have to be tight in the end: *e.g.*, EDIFF=1e-2 and EDIFFG=1e-1 in INCAR_1, gradually tightening to EDIFF=1e-4 and EDIFFG=1e-3 in INCAR_4. The maximum number of iterations (NSW) should be sufficiently large to enable good relaxation, but not too large to avoid wasting computer time on poor configurations. The larger your system, the larger NSW should be.

- Choosing an efficient relaxation algorithm can save a lot of time. In VASP, we recommend to start relaxation with conjugate gradients (IBRION=2 and POTIM=0.02) and when the structure is closer to local minimum, switch to IBRION=1 and POTIM=0.3.

- Even if you study an insulating system, many configurations that you will sample are going to be metallic, so to have well-converged results, you must use "metallic" treatment — which works both for metals and insulators. We recommend the Methfessel-Paxton smearing scheme (ISMEAR=1). For a clearly metallic system, use ISMEAR=1 and SIGMA=0.1–0.2. For a clearly insulating system, we recommend ISMEAR=1 and SIGMA starting at 0.1 (INCAR_1) and decreasing to 0.05.

Here we provide an example of INCAR files for carbon with 16 atoms in the unit cell, with default ENCUT=400 eV in POTCAR:

```
INCAR_1:
    PREC=LOW
    EDIFF=1e-2
    EDIFFG=1e-1
    NSW=65
    ISIF=4
    IBRION=2
    POTIM=0.02
    ISMEAR=1
    SIGMA=0.10
```

```
INCAR_2:
    PREC=NORMAL
    EDIFF=1e-3
    EDIFFG=1e-2
    NSW=55
    ISIF=4
    IBRION=1
    POTIM=0.30
    ISMEAR=1
    SIGMA=0.08
```

```
INCAR_3:
    PREC=NORMAL
    EDIFF=1e-3
    EDIFFG=1e-2
    ENCUT=520.0
    NSW=65
    ISIF=3
    IBRION=2
    POTIM=0.02
    ISMEAR=1
    SIGMA=0.07
```

```
INCAR_4:
    PREC=NORMAL
    EDIFF=1e-4
    EDIFFG=1e-3
    ENCUT=600.0
    NSW=55
    ISIF=3
    IBRION=1
    POTIM=0.30
    ISMEAR=1
    SIGMA=0.06
```

```
INCAR_5:
    PREC=NORMAL
    EDIFF=1e-4
    EDIFFG=1e-3
    ENCUT=600.0
    NSW=0
    ISIF=2
    IBRION=2
    POTIM=0.02
    ISMEAR=1
    SIGMA=0.05
```

3.2 Output files

These are stored in the folder ∼/StructurePrediction/results1. If this is a new calculation, results2, results3, ... (if the calculation has been restarted or run a few times), there will be a separate results* folder for each calculation. **Caution:** When looking at space groups in the file Individuals, keep in mind that USPEX often underdetermines space group symmetries, because of finite precision of structure relaxation and relatively tight space group determination tolerances. You should visualize the predicted structures. To get full space group, either increase symmetry tolerances (but this can be dangerous), or re-relax your structure with increased precision.

The subdirectory ∼/StructurePrediction/results1 contains the following files:

- OUTPUT.txt — summarizes input variables, structures produced by USPEX, and their characteristics.
- parameters.uspex — this is a copy of the input.uspex file used in this calculation with some defaults explicitly writen, for your reference.
- Individuals — gives details of all produced structures (energies, unit cell volumes, space groups, variation operators that were used to produce the structures, k-points mesh used to compute the structures' final energy, degrees of order, *etc.*). File BESTIndividuals gives this information for the best structures from each generation.
- convex_hull — only for variable-composition calculations, this file gives all thermodynamically stable compositions, and their enthalpies (per atom).
- gatheredPOSCARS — relaxed structures (in the VASP5 POSCAR format).
- BESTgatheredPOSCARS — the same data for the best structure in each generation.

- `gatheredPOSCARS_unrelaxed` — gives all structures produced by USPEX before relaxation.
- `enthalpies_complete.csv` — gives the enthalpies for all structures in each stage of relaxation.
- `origin` — shows which structures originated from which parents and through which variation operators.
- `gatheredPOSCARS_order` — gives the same information as `gatheredPOSCARS`, and in addition for each atom it gives the values of local order parameters (Ref.[17]).
- `goodStructures_POSCARS` and `extended_convex_hull_POSCARS` (for fixed- and variable-composition calculations correspondingly) report all of the different structures in order of decreasing stability, starting from the most stable structure and ending with the least stable.
- `compositionStatistics` is a file containing statistics of the compositions in terms of which variation operators produced these compositions.
- graphical files (`*.svg`) — for rapid visual assessment of the results:
  - `Energy_vs_N.svg` (`Fitness_vs_N.svg`) — energy (fitness) as a function of structure number;
  - `Energy_vs_Volume.svg` — energy as a function of volume;
  - `Variation-Operators.svg` — energy of the child *vs.* parent(s); different operators are marked with different colors (this graph allows one to assess the performance of different variation operators) also show evolution of each operator's strength.
  - `E_series.svg` — correlation between energies from relaxation steps i and $i+1$; helps to detect problems and improve structure relaxation.
  - For variable compositions there is an additional graph `extendedConvexHull.svg`, which shows the enthalpy of formation as function of composition.

4 Input options. The input.uspex file

The input.uspex file has json-like syntax and hierarchical structure representing modular nature of USPEX program. Typical input.uspex files are given in the Appendix 6.1. Below we describe the most important parameters of the input. Most of the parameters have reliable default values (this allows you to have extremely short input files!). Those options that have no default should always be specified. Please consult online utilities at https://uspex-team.org/online_utilities/ — these help to prepare input and analyze some of results. Section 5 of this Manual briefly discusses these utilities.

4.1 The input.uspex file syntax

The input.uspex file consists of main section and a number of definition sections. Main section is obligatory and preceds definition sections. Definition sections are optional. There could be any number of definition sections. Each definition section starts with define name line.

```
{...}

#define name1
{...}

#define name2
{...}

#define name3
{...}

...
```

Main input section is described in 4.2-4.6.

Definition sections correspond to parameters of ab initio calculations (4.7), submission parameters (4.8) and molecule definitions (4.9).

Each section consists of pairs of keywords and corresponding values.

```
{
key1 : value1
key2 : value2
key3 : value3
```

...
}

Such pair correspond either to some parameter name and its value, or to parameters block.

Parameters can have numeric, string, sequence or map values.

numGenerations : 50

stages: [vasp1 vasp2 vasp3]

ionDistances: {'C C': 2.0 'C H': 1.2 'H H': 0.7}

Input parameters are grouped into blocks. Each block reflects some aspect of calculation. So it should be quite intuitive to prepare such input file.

Blocks of parameters looks like.

compositionSpace: {
symbols: ...
blocks: ...
}

There are two types of sequences: lists [] and tuples (). Tuples are expected to have fixed length when lists can be of arbitrary length.

4.2 General calculation parameters

All USPEX modes have the following outline. There is a loop. At each loop iteration program deals with a number of systems also called individuals. The set of these individuals is called population or generation. Loop iteration itself is also referred to as generation.

The loop is controlled by two input parameters: *numGenerations* and *stopCrit*.

▷ *variable* numGenerations

Meaning: Maximum number of generations allowed for the simulation. The simulation can terminate earlier, if the same best structure remains unchanged for stopCrit generations.

Default:

Format:

numGenerations : 50

▷ *variable* stopCrit

Meaning: The simulation is stopped if the best structure did not change for stopCrit generations, or when numGenerations have expired — whichever happens first.

Default:

Format:

```
stopCrit :  20
```

USPEX job with population consists of two parts: creation and relaxation.

Creation is controlled by block *optimizer*. Depending on the type of optimization this block looks different. For global optimum search it is *GlobalOptimizer*. In later releases $VCNEB$ optimizer type will be available for local optimization of phase transition pathways.

▷ *variable* optimizer

Meaning: Specifies the type of calculation and its parameters

Available types:

- GlobalOptimizer — global search algorithm, see 4.3.

Default:

Format:

```
optimizer :  {type:  GlobalOptimizer ...}
```

For population relaxation USPEX employs a powerful two-level parallelization scheme, making its parallel scalability exemplary. The first level of parallelization is performed within structure relaxation codes, the second level of parallelization distributes the calculation over the individuals in the same population (since structures within the same generation are independent of each other).

▷ *variable* stages

Meaning: List of relaxation stages to be applied to each system in the population.

The names of definition sections should be used in this sequence. See 4.7 for information on writing such section.

Default:

Format:

```
stages :  [gulp gulp gulp]
```

▷ *variable* numParallelCalcs

Meaning: Specifies how many structure relaxations you want to run in parallel.

Default:

Format:

```
numParallelCalcs :  10
```

4.3 Global optimization search

Global optimization is one of possible calculation types in USPEX (see *optimizer* in 4.2). As the title suggests, it is search for global optimum of some property over certain configuration space. Such search is done by sampling this space iteration by iteration. Sample for this space is called population. Target configuration space is described with parameters block *target*, property to be optimized with *optType* parameter and sampling method with *selection* block.

```
{
type: GlobalOptimizer
target: {...}
selection: {...}
optType: ...
}
```

▷ *variable* target

Meaning: Specifies the target space of search

Available targets:

- Crystal — crystal structure prediction, see 4.5

Default:

Format:

```
target :  {type:  Crystal ...}
```

▷ *variable* selection

Meaning: Specifies the evolutionary algorithm to be used

Available algorithms:

- USPEXClassic — full evolutionary algorithm, see 4.6

Default:

Format:

```
selection :  {type:  USPEXClassic ...}
```

▷ *variable* optType

Meaning: This parameter specifies the property (or properties) that you want to optimize.

See 4.4

Default: no default, must be specified

Format:

```
optType :  enthalpy
```

There is helper parameter *stopFitness*.

▷ *variable* stopFitness

Meaning: Specifies the fitness value so that the calculation will stop after reaching fitness $\leq$ stopFitness.

Default:

Format:

```
stopFitness:  -90.912
```

4.4 Optimization type

Optimization type specification is designed as powerful method of writing functions inside input file. Such functions looks like

(func arg1 arg2 ...)

Here, $arg1$, $arg2$ could be functions themselves. For example:

(pareto (aging enthalpy) (negate radialDistributionUtility.structureOrder))

This means that parameter to be optimized will be calculated for each structure by following procedure. Enthalpy will be taken from structutre directly, parameter *structureOrder* will be calculated using *radialDistributionUtility* module. Then *aging* function will be applied to enthalpy and *negate* function to *structureOrder*. Then finally optimization parameter will be determined via *pareto* function with two arguments: aged enthalpy and structure order taken with negative sign.

For now the following functions are available:

Function	Description
negate	takes opposite value
aging	accumulate penalties to structures sampled more than once (not tunable at current release)
pareto	calculate Pareto fronts.

Available base parameters:

Name	Description
enthalpy	enthalpy of system
enthalpyCCH	enthalpy per block above convex hull (for variable composition)
cellUtility.volume	unit cell volume
radialDistributionUtility.structureOrder	degree of order
radialDistributionUtility.quasientropy	structural quasientropy

Providing only base parameter as *optType* is also a valid option.

4.5 Crystal structure prediction

When *target* type in 4.3 is set to *Crystal*, USPEX performs global optimization of crystal properties. Historically, the first property which USPEX optimized was enthalpy and search was for stable structures. This mode is usually referred to as crystal structure prediction.

Typical *target* block in this mode looks like.

```
{
type: Crystal
compositionSpace: {symbols : [Mg Al O] blocks: [[4 8 16]]}
conditions: {externalPressure : 100}
}
```

Besides obligatory *compositionSpace* block there are a number of optional parameter blocks and standalone parameters.

▷ *variable* compositionSpace

Meaning: Describes the identity of each type of atom or molecule and specifies the number of species of each type.

Default: no default

Format:

```
compositionSpace :  {...}
```

See 4.5.1.

▷ *variable* cellUtility

Meaning: Properties of the unit cell. Such as cell parameters, initial volume etc.

Default: by default variable cell search is done, volume is estimated from conditions

Format:

```
cellUtility :  {...}
```

See 4.5.2.

▷ *variable* radialDistributionUtility

Meaning: Radial distribution of atomic distances is used for fingerprinting and order calculation. This block contains its parameters.

Default:

Format:

```
radialDistributionUtility :  {...}
```

See 4.5.3.

▷ *variable* conditions

Meaning: Properties of environment in general (currently only external pressure).

Default: zero pressure

Format:

```
conditions :  {...}
```

See 4.5.4.

▷ *variable* ionDistances

Meaning: Sets the minimum interatomic distance matrix between different atom types. Structures with distances lower than `IonDistances` will be strictly discarded.

Default: the `IonDistances` between atom A and B are estimated as $0.22 \times (V_A^{1/3} + V_B^{1/3})$ but not larger than 1.2 Å, and $0.45 \times (V_A^{1/3} + V_B^{1/3})$ in molecular calculations, where V_A and V_B are the default volumes of atom A and B estimated in USPEX.

Format:

```
ionDistances :  {'Mg Mg':  1.0 'Mg Si':  1.0 'Mg O': 0.8 'Si Si':  1.0
'Si O': 0.8 'O O': 1.0}
```

NOTE: If the compound in the example above is $MgSiO_3$, the matrix reads as follows: the minimum Mg–Mg distance allowed in a newly generated structure is 1.0 Å, the minimum Mg–

Si, Si–Si and O–O distances are also 1.0 Å, and the minimum Mg–O and Si–O distances are 0.8 Å. You can use this keymatrix to incorporate further system-specific information: *e.g.*, if you know that C atoms prefer to be very far apart and are never closer than 3 Å in your system, you can specify this information. Beware, however, that the larger these minimum distances, the more difficult it is to generate structures fulfilling these constraints (especially for large systems), so strive for a compromise and remember that `IonDistances` must be **much** smaller than the actual distances in the crystal: realistic distances will be achieved by structure relaxation. What `IonDistances` trick does is to avoid structures which cannot be relaxed correctly. Commonly used computational methods (pseudopotentials, PAW, LAPW, and many parametric forcefields) fail when the interatomic distances are too small.

▷ *variable* `randSym`

Meaning: Properties of symmetric random structure generator.

Default:

Format:

```
randSym :  {...}
```

This generator produces structures with a space symmetry imposed. See 4.5.5.

▷ *variable* `randTop`

Meaning: Properties of topological random structure generator.

Default:

Format:

```
randTop :  {...}
```

This generator produces structures with known topology used as blueprint.

▷ *variable* `heredity`

Meaning: Properties of heredity variation operator.

Default:

Format:

```
heredity :  {...}
```

This operator joins layers cut from parent structures into child one.

▷ *variable* `permutation`

Meaning: Properties of permutation variation operator.

Default:

Format:

```
permutation : {...}
```

This operator swaps identities of compositional elements (atoms or molecules), preserving total composition. See 4.5.6.

▷ *variable* transmutation

Meaning: Properties of transmutation variation operator.

Default:

Format:

```
transmutation : {...}
```

This operator alters identities of compositional elements (atoms or molecules), breaking total composition. See 4.5.7.

▷ *variable* softmodemutation

Meaning: Properties of softmodemutation variation operator.

Default:

Format:

```
softmodemutation : {...}
```

This operators move atoms along the eigenvector of soft vibrational mode. See 4.5.8.

▷ *variable* seeds

Meaning: Seeds are the user-provided structures to accelerate the search.

Default:

Format:

```
seeds : {...}
```

See 4.5.9.

4.5.1 Compositional space

Examples.

Fixed-compositioin search for $Mg_4Al_8O_{16}$.

```
{
symbols : [Mg Al O]
blocks: [[4 8 16]]
```

}

Fixed-composition search with two molecules of type *mol_1* and two of type *mol_2* in unit cell.

{
symbols : [mol_1 mol_2]
blocks: [[2 2]]
}

Variable-composition search for the $MgO - Al_2O_3$ system, where the minimum number of atoms is 8 and maximum 20.

{
symbols : [Mg Al O]
blocks: [[1 0 1] [0 2 3]]
minAt: 8
maxAt: 20
}

Variable-composition search for the $MgO - Al_2O_3$ system, where the minimum number of atoms is 8 and maximum 20. With additional constraint that all the sampled compositions should contain between 1 and 3 compositional blocks Al_2O_3.

{
symbols : [Mg Al O]
blocks: [[1 0 1] [0 2 3]]
range: [(0 10) (1 3)]
minAt: 8
maxAt: 20
}

▷ *variable* symbols

Meaning: Describes the identity of each atomic or molecular type.

The value is a list of chemical elements or molecule names.

Default: no default

Format:

```
symbols :  [mol_1 mol_2 O]
```

Molecule names should be defined in molecule definition, see section 4.9

▷ *variable* blocks

Meaning: Specify compositional building blocks.

Default: no default

Format:

```
blocks :  [[1 0 1] [0 2 3]]
```

▷ *variable* range

Meaning: Specify ranges for each block.

Default: If maxAt is set (variable-composition) then each block can be taken from 0 to maxAt devided by size of block times. If maxAt is not set (fixed-composition) each block is taken one time.

Format:

```
range :  [(1 10) (1 10)]
```

▷ *variable* minAt

Meaning: Minimum number of atoms in the primitive cell.

Default:

Format:

```
minAt :  8
```

▷ *variable* maxAt

Meaning: Maximal number of atoms in the primitive cell.

Default:

Format:

```
maxAt :  20
```

4.5.2 Unit cell properties

▷ *variable* cellVolume

Meaning: Initial volume of the unit cell.

Default: for cell volumes you don't have to specify values — USPEX has a powerful algorithm to make reasonable estimates at any pressure.

Format:

```
cellVolume :  125.0
```

(1) This volume is only used as an initial guess to speed up structure relaxation and does not affect the results, because each structure is fully optimized and adopts the volume corresponding to the (free) energy minimum.

(2) You can also use online program https://uspex-team.org/online_utilities/volume_estimation. The users can also input the volumes manually.

(3) If you study molecular crystals under pressure, you might sometimes need to increase the initial volumes somewhat, in order to be able to generate initial random structures.

▷ *variable* cellVectors

Meaning: Unit cell vectors.

Default: by default cell vectors are variable

Format:

```
cellVectors:  [[2.474 0.0 0.0] [0.0 8.121 0.0] [0.0 0.0 6.138]]
```

This parameter is used in fixed-cell clalculations.

▷ *variable* cellParameters

Meaning: Unit cell paramaters.

Default: by default cell patameters are variable

Format:

```
cellParameters:  {a:  2.474 b:  8.121 c:  6.138 alpha:  90.0 beta:  90.0
gamma:  90.0}
```

This parameter is used in fixed-cell clalculations.

4.5.3 Radial distribution based fingerprint

Please read reference[17] for details on fingerprint functions.

▷ *variable* Rmax

Meaning: Distance cutoff (in Å).

Default: 10.0

Format:

```
Rmax :  10
```

▷ *variable* delta

Meaning: Discretization (in Å) of the fingerprint.

Default: 0.08

Format:

```
delta :  0.08
```

▷ *variable* sigma

Meaning: Gaussian broadening of interatomic distances.

Default: 0.03

Format:

```
sigma :  0.03
```

▷ *variable* tolerance

Meaning: Specifies the minimal cosine distances between structures that qualify them as non-identical — for participating in the production of child structures and for survival of the fittest, respectively. This depends on the precision of structure relaxation and the physics of the system (for instance: for alloy ordering problems, fingerprints belonging to different structures will be very similar, and these tolerance parameters should be made small).

Default: 0.008

Format:

```
tolerance :  0.0012
```

4.5.4 Conditions

▷ *variable* externalPressure

Meaning: Specifies external pressure at which you want to find structures, in GPa.

Default: 0.0

Format:

```
externalPressure:  0.00001
```

NOTE: Please: do not specify it in relaxation files in the Specific/ folder.

4.5.5 Symmetric random generator

▷ *variable* `nsym`

Meaning: Possible symmetry groups for random symmetric structure generator crystals (spacegroups), layer plane groups for 2D-crystals, plane groups for surfaces, or point groups for clusters. A certain number of structures will be produced using randomly selected groups from this list, using randomly generated lattice parameters and atomic coordinates. During this process special Wyckoff sites can be produced from general positions (Fig. 7).

Default:

- For 3D crystals: `2-230`

Format:

```
nsym:  '16-74'
```

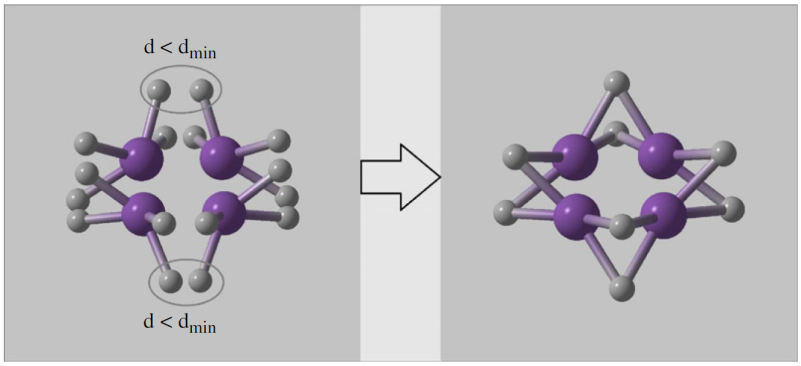

Figure 5: **Example of random symmetric structure generation and merging atoms onto special Wyckoff positions (for detail, see Ref.[13]).**

▷ *variable* splitInto

Meaning: Defines the number of identical subcells or pseudosubcells in the unit cell. If you do not want to use splitting, just use the value 1, or delete the block. Use splitting only for systems with $>$ 25–30 atoms/cell.

Default: 1

Format:

```
splitInto:  [2 4]
```

Subcells introduce extra translational (pseudo)symmetry. In addition to this, each subcell can be built using a special space groups algorithm developed by A.R. Oganov and H.T. Stokes and implemented by H.T. Stokes (see Reference[13]).

4.5.6 Permutation

▷ *variable* howManySwaps

Meaning: For permutation, the number of pairwise swaps will be randomly drawn from a uniform distribution between 1 and howManySwaps.

Default: 0.5×(maximum number of possible swaps). If atoms a and b, and atoms c and d are swappable, then the total number of possible swaps is $\min(Na, Nb) + \min(Nc, Nd)$, and the default for howManySwaps is $0.5 \times [\min(Na, Nb) + \min(Nc, Nd)]$. In most cases, it is a good idea to rely on this default.

Format:

```
howManySwaps:  5
```

▷ *variable* specificSwaps

Meaning: Specifies which atom types you allow to swap in permutation.

Default: No specific swaps and all atoms are permutable.

Format:

```
specificSwaps:  [1 2]
```

NOTE: In this case, atoms of type 1 can be swapped with atoms of type 2. If you want to try all possible swaps, just leave a blank line inside this keyblock, or delete the block.

4.5.7 Transmutation

In this operator, a randomly selected atom is transmuted into another chemical species present in the system -the new chemical identity is chosen randomly.

▷ *variable* howManyTrans

Meaning: Maximum percentage of atoms in the structure that are being transmuted (0.1 = 10%). The fraction of atoms that will be transmuted is drawn randomly from a homogeneous distribution bounded from 1 to the fractional parameter `howManyTrans`.

Default: `0.2`

Format:

```
howManyTrans:  0.2
```

4.5.8 Softmutation

▷ *variable* degree

Meaning: The maximum displacement in softmutation in Å. The displacement vectors for softmutation are scaled so that the largest displacement magnitude equals `mutationDegree`.

Default: 3×(average atomic radius)

Format:

```
degree:  1.0
```

4.5.9 Seeds

The use of seeds technique requires aditional input files to be provided (see 3.1).

▷ *variable* generations

Meaning: List of generations into which seeds will be injected in search.

Default: No default

Format:

```
generations:  [0 5 10]
```

▷ *variable* seedsFolders

Meaning: List of paths to folders which contain seeds files to be injected in search. In same order as *generations* above.

Default:

Format:

```
seedsFolders:  ['./Seeds/1' './Seeds/2' './Seeds/3']
```

4.6 Evolutionary algorithm USPEX

Here we describe parameters of evolutionary algorithm, one of the possible algorithms driving global optimization search (see 4.3). Other algorithms available in USPEX code for years (evolutionary metadynamics and particle swarm optimization) are under re-implementation in this version.

▷ *variable* popSize

Meaning: The number of structures in each generation; size of initial generation can be set separately, if needed.

Default: $2 \times N$ rounded to the closest 10, where N is the number of atoms/cell (or `maxAt` for variable composition). The upper limit is 60. Usually, you can trust these default settings.

Format:

```
popSize:  20
```

▷ *variable* initialPopSize

Meaning: The number of structures in the initial generation.

Default: equal to `populationSize`.

Format:

```
initialPopSize:  20
```

NOTE: In most situations, we suggest that these two parameters be equal. Sometimes (especially in variable-composition calculations) it may be useful to specify `initialPopSize` to be larger than `populationSize`. It is also possible to have a smaller `initialPopSize`, if one wants to produce the first generation from seed structures.

▷ *variable* bestFrac

Meaning: Fraction of the current generation that shall be used to produce the next generation.

Default: `0.7`

Format:

```
bestFrac:  0.7
```

NOTE: This is an important parameter, values between 0.5–0.8 are reasonable.

▷ *variable* howManyDiverse

Meaning: Defines how many good and diverse structures will survive into the next gen-

eration.

Default: 0.15×popSize

Format:

```
howManyDiverse:  3
```

We perform clustering of the population into specified number of groups here. And take the best representative from each group

▷ *variable* optType

Meaning: This keyblock specifies the property that you wish to use for selecting potential parents for new generation.

See 4.4

Default: same as in global optimizer section.

Format:

```
optType :  (aging enthalpy)
```

It can differ from property to be optimized. Usually aging procedure is additionally applied.

▷ *variable* fractions

Meaning: This parameter defines allowed percentages of each variation operator. The first value is minimum percentage of structures in each generation produced by this operator. The second — maximum percentage. The third value is weight (not the percentage) of the operator in 1st generation.

Default:

Format:

```
fractions :  {
heredity:  (0.3 0.7 0.5)
softmodemutation:  (0.1 0.3 0.2)
randTop:  (0.1 0.5 0.3)
}
```

The fractions of operators evolve during the calculation, so that the more successful operators gain weight at the expense of the less successful operators, but within the limits specified here.

4.7 Details of *ab initio* calculations sections

Typical ab initio definition section looks like

```
#define vasp1
{
type : vasp
commandExecutable : 'vasp'
kresol : 0.12
}
```

It contains type of interface for external program package, common parameters (like *commandExecutable*, see 4.7.1) and package specific parameters (like *kresol*).

The definiton name (*vasp*1 in the example above) should be used in *stages* parameter of general parameters section 4.2.

Available types of interfaces are:

`vasp` — VASP interface, see 4.7.2.

`gulp` — GULP interface, see 4.7.3.

`lammps` — LAMMPS interface, see 4.7.4.

`qe` — Quantum Espresso interface, see 4.7.5.

`abinit` — Abinit interface, see 4.7.6.

4.7.1 Common interface parameters.

▷ *variable* `commandExecutable`

Meaning: Specifies executable for a given code.

Default: no default, has to be specified by the user.

Format:

```
commandExecutable :  'mpirun vasp'
```

▷ *variable* `taskManager`

Meaning: Specifies name of task manager to be used for submission.

Default: By default no task manager is used.

Format:

```
taskManager :  TM
```

This name (*TM* in the example above) should be defined as another section, see 4.8. Usually you define one task manager and use it in several stages, but it can be useful to use different task managers for different stages.

▷ *variable* tag

Meaning: Optional identifier for stage.

Default: equal to index number of stage.

Format:

```
tag :  final
```

NOTE: This is useful if you want to highlight or distinguish a stage. This tag will appear as suffix for a job in job submission system and in calculation folder created for this job. If this parameter is set then files in Specific/ folder should use it in their suffixes instead of index number of stage. For eaxample $INCAR\_final$ istead of just $INCAR\_4$.

4.7.2 VASP interface parameters.

▷ *variable* sleepTime

Meaning: Specifies sleep delay between checks of completion for this stage in seconds.

Default: 30

Format:

```
sleepTime :  2
```

▷ *variable* kresol

Meaning: Specifies the reciprocal-space resolution for k-points generation (units: $2\pi\text{Å}^{-1}$).

Default:

Format:

```
kresol :  0.12
```

NOTE: Using different values for each step of structure relaxation, starting with cruder (*i.e.*, larger) values and ending with high resolution dramatically speeds up the calculations, especially for metals, where very many k-points are needed. This keyblock is important for ab initio calculations (through some codes, e.g. VASP and SIESTA, now have similar tricks)).

4.7.3 GULP interface parameters.

▷ *variable* sleepTime

Meaning: Specifies sleep delay between checks of completion for this stage in seconds.

Default: 10

Format:

```
sleepTime :  2
```

4.7.4 LAMMPS interface parameters.

▷ *variable* sleepTime

Meaning: Specifies sleep delay between checks of completion for this stage in seconds.

Default: 10

Format:

```
sleepTime :  2
```

4.7.5 Quantum Espresso interface parameters.

▷ *variable* sleepTime

Meaning: Specifies sleep delay between checks of completion for this stage in seconds.

Default: 30

Format:

```
sleepTime :  2
```

▷ *variable* kresol

Meaning: Specifies the reciprocal-space resolution for k-points generation (units: $2\pi\text{Å}^{-1}$).

Default:

Format:

```
kresol :  0.12
```

NOTE: Using different values for each step of structure relaxation, starting with cruder (*i.e.*, larger) values and ending with high resolution dramatically speeds up the calculations, especially for metals, where very many k-points are needed. This keyblock is important for ab initio calculations (through some codes, e.g. VASP and SIESTA, now have similar tricks)).

4.7.6 ABINIT interface parameters.

▷ *variable* sleepTime

Meaning: Specifies sleep delay between checks of completion for this stage in seconds.

Default: 30

Format:

```
sleepTime :  2
```

▷ *variable* kresol

Meaning: Specifies the reciprocal-space resolution for k-points generation (units: $2\pi\text{Å}^{-1}$).

Default:

Format:

```
kresol :  0.12
```

NOTE: Using different values for each step of structure relaxation, starting with cruder (*i.e.*, larger) values and ending with high resolution dramatically speeds up the calculations, especially for metals, where very many k-points are needed. This keyblock is important for abinitio calculations (through some codes, e.g. VASP and SIESTA, now have similar tricks)).

4.8 Task manager definition

Task manager definition looks like

```
#define TM
{
type : SBATCH
header : '#!/bin/sh\n#SBATCH -p debug\n#SBATCH -N 1\n#SBATCH -n 1 \n\n'
}
```

Type could be:

- SBATCH — for slurm submission system,
- QSUB — for torque submission system,
- BSUB — for lsf submission system.

Header should contain part of submission script for supercomputer which you are going to use. Usually such header should contain information on submission queue, number of nodes and cores which the job is going to consume. To determine exact content of such header consult with your supercomputer user's guide or administrator.

4.9 Molecules definitions

Each molecule definition looks like.

```
#define mol_name
{
filename : 'MOL_FILE'
}
```

The defined molecule name (like *mol_name*) should then be used in *compositionSpace* block 4.5.1.

5 Online utilities

We have created a number of useful online utilities, which can be used for preparing USPEX input and for post-processing. The utilities are available at:

https://uspex-team.org/online_utilities/

Below you can find information about each one of them.

5.1 Structure characterization

Here we have 4 utilities:

- Fingerprints — the utility calculates and plots fingerprint function, which is a crystal structure descriptor, a 1D-function related to the pair correlation function and diffraction patterns. It does not depend on absolute atomic coordinates, but only on interatomic distances. Small deviations in atomic positions will influence the fingerprint only slightly, *i.e.* it is numerically robust.
- Multifingerprint — the utility calculates average quasi-entropy, A-order(average atomic order parameter) and S-order(whole-structure order parameter) for a set of structures. Also it filters unique structures by cosine distances difference ≥ 0.003, identifies the symmetry of these structures and lists them in the `uniq_gatheredPOSCARS` file.
- POSCAR2CIF — determines space group and prepares a `CIF` file from a `POSCAR` file.
- CIF2POSCAR — prepares a `POSCAR` file from a `CIF` file.
- XSF2POSCAR — prepares a `POSCAR` file from a `XSF` (XCRYSDEN) file.

5.2 Properties calculations

Here we have 2 utilities:

- Hardness — the utility is to calculate hardness based on the Mazhnik-Oganov model.
- EELS — the utility calculates the Electron Energy Loss Spectrum (EELS). Written by Priya Johari.

5.3 Molecular crystals

Here we have 2 utilities:

- MOL precheck — the utility allows you to check `MOL_1` files before running USPEX for molecular crystals (`calculationType`=310/311/110).
- Zmatrix — the utility converts `XYZ` file to USPEX `MOL_1` file.

5.4 Surfaces

Substrate — a program which prepares a substrate from a `POSCAR`/`CIF` file, given Miller indices, thickness of the layer, and shift. The resulting `POSCAR` file can be used for `calculationType`=200/201 as a substrate for surface calculations.

5.5 Miscellaneous

Here we have the following:

- Input generator — USPEX `INPUT.txt` generator. The utility can help beginners to create a correct input for USPEX calculations.
- Volume estimation — the utility estimates volumes of non-molecular and molecular crystals for USPEX (for `INPUT.txt` file).
- USPEX manual — online version of this manual.
- USPEX examples — archives with USPEX examples.

6 Appendices

6.1 Sample INPUT.txt files

6.1.1 Fixed-composition USPEX calculation

```
{
    optimizer: {
        type: GlobalOptimizer
        target: {
            type: Crystal
            conditions: {externalPressure: 100}
            compositionSpace: {symbols: [Mg Al O]
                               blocks: [[4 8 16]]}
        }
        optType: enthalpy
        selection: {
            type: USPEXClassic
            popSize: 50
            optType: (aging enthalpy)
            fractions: {
                heredity: (0.1 1.0 0.5)
                softmodemutation: (0.1 1.0 0.2)
                randSym: (0.05 1.0 0.1)
                randTop: (0.05 1.0 0.1)
                permutation: (0.05 1.0 0.1)
            }
        }
    }
    stages: [gulp gulp gulp gulp gulp]
    numParallelCalcs: 10
    numGenerations: 50
    stopCrit: 28
}

#define gulp
{type : gulp, commandExecutable : 'gulp', goptions : './Specific/
   goptions'}
```

6.1.2 Variable-composition USPEX calculation

```
{
    optimizer: {
        type: GlobalOptimizer
        target: {
            type: Crystal
            compositionSpace: {symbols: [Mo B]
                               blocks: [[1  0] [0  1]]
                               minAt: 8
                               maxAt: 18
            }
        }
        optType: enthalpy
        selection: {
            type: USPEXClassic
            popSize: 80
            initPopSize: 200
            optType: (aging enthalpy)
            fractions: {
                heredity: (0.1 1.0 0.5)
                softmodemutation: (0.1 1.0 0.2)
                randSym: (0.05 1.0 0.1)
                randTop: (0.05 1.0 0.1)
                permutation: (0.05 1.0 0.1)
            }
        }
    }
    stages: [gulp gulp gulp]
    numParallelCalcs: 10
    numGenerations: 60
    stopCrit: 20
}

#define gulp
{type : gulp, commandExecutable : 'gulp', goptions : './Specific/
   goptions'}
```

6.2 List of space groups

1	$P1$	2	$P\text{-}1$	3	$P2$	4	$P2_1$
5	$C2$ $(A2)^*$	6	Pm	7	Pc $(Pa)^*$	8	Cm $(Am)^*$
9	Cc $(Aa)^*$	10	$P2/m$	11	$P2_1/m$	12	$C2/m$ $(A2/m)^*$
13	$P2/c$ $(P2/a)^*$	14	$P2_1/c$ $(P2_1/a)^*$	15	$C2/c$ $(A2/a)^*$	16	$P222$
17	$P222_1$	18	$P2_12_12$	19	$P2_12_12_1$	20	$C222_1$
21	$C222$	22	$F222$	23	$I222$	24	$I2_12_12_1$
25	$Pmm2$	26	$Pmc2_1$	27	$Pcc2$	28	$Pma2$
29	$Pca2_1$	30	$Pnc2$	31	$Pmn2_1$	32	$Pba2$
33	$Pna2_1$	34	$Pnn2$	35	$Cmm2$	36	$Cmc2_1$
37	$Ccc2$	38	$Amm2$ $(C2mm)^*$	39	$Aem2$ $(C2mb)^*$	40	$Ama2$ $(C2cm)^*$
41	$Aea2$ $(C2cb)^*$	42	$Fmm2$	43	$Fdd2$	44	$Imm2$
45	$Iba2$	46	$Ima2$	47	$Pmmm$	48	$Pnnn$
49	$Pccm$	50	$Pban$	51	$Pmma$	52	$Pnna$
53	$Pmna$	54	$Pcca$	55	$Pbam$	56	$Pccn$
57	$Pbcm$	58	$Pnnm$	59	$Pmmn$	60	$Pbcn$
61	$Pbca$	62	$Pnma$	63	$Cmcm$	64	$Cmce$ $(Cmca)^*$
65	$Cmmm$	66	$Cccm$	67	$Cmme$ $(Cmma)^*$	68	$Ccce$ $(Ccca)^*$
69	$Fmmm$	70	$Fddd$	71	$Immm$	72	$Ibam$
73	$Ibca$	74	$Imma$	75	$P4$	76	$P4_1$
77	$P4_2$	78	$P4_3$	79	$I4$	80	$I4_1$
81	$P\text{-}4$	82	$I\text{-}4$	83	$P4/m$	84	$P4_2/m$
85	$P4/n$	86	$P4_2/n$	87	$I4/m$	88	$I4_1/a$
89	$P422$	90	$P42_12$	91	$P4_122$	92	$P4_12_12$
93	$P4_222$	94	$P4_22_12$	95	$P4_322$	96	$P4_32_12$
97	$I422$	98	$I4_122$	99	$P4mm$	100	$P4bm$
101	$P4_2cm$	102	$P4_2nm$	103	$P4cc$	104	$P4nc$
105	$P4_2mc$	106	$P4_2bc$	107	$I4mm$	108	$I4cm$
109	$I4_1md$	110	$I4_1cd$	111	$P\text{-}42m$	112	$P\text{-}42c$
113	$P\text{-}42_1m$	114	$P\text{-}42_1c$	115	$P\text{-}4m2$	116	$P\text{-}4c2$
117	$P\text{-}4b2$	118	$P\text{-}4n2$	119	$I\text{-}4m2$	120	$I\text{-}4c2$
121	$I\text{-}42m$	122	$I\text{-}42d$	123	$P4/mmm$	124	$P4/mcc$
125	$P4/nbm$	126	$P4/nnc$	127	$P4/mbm$	128	$P4/mnc$
129	$P4/nmm$	130	$P4/ncc$	131	$P4_2/mmc$	132	$P4_2/mcm$
133	$P4_2/nbc$	134	$P4_2/nnm$	135	$P4_2/mbc$	136	$P4_2/mnm$
137	$P4_2/nmc$	138	$P4_2/ncm$	139	$I4/mmm$	140	$I4/mcm$
141	$I4_1/amd$	142	$I4_1/acd$	143	$P3$	144	$P3_1$
145	$P3_2$	146	$R3$	147	$P\text{-}3$	148	$R\text{-}3$
149	$P312$	150	$P321$	151	$P3_112$	152	$P3_121$
153	$P3_212$	154	$P3_221$	155	$R32$	156	$P3m1$
157	$P31m$	158	$P3c1$	159	$P31c$	160	$R3m$
161	$R3c$	162	$P\text{-}31m$	163	$P\text{-}31c$	164	$P\text{-}3m1$
165	$P\text{-}3c1$	166	$R\text{-}3m$	167	$R\text{-}3c$	168	$P6$
169	$P6_1$	170	$P6_5$	171	$P6_2$	172	$P6_4$
173	$P6_3$	174	$P\text{-}6$	175	$P6/m$	176	$P6_3/m$
177	$P622$	178	$P6_122$	179	$P6_522$	180	$P6_222$
181	$P6_422$	182	$P6_322$	183	$P6mm$	184	$P6cc$
185	$P6_3cm$	186	$P6_3mc$	187	$P\text{-}6m2$	188	$P\text{-}6c2$
189	$P\text{-}62m$	190	$P\text{-}62c$	191	$P6/mmm$	192	$P6/mcc$
193	$P6_3/mcm$	194	$P6_3/mmc$	195	$P23$	196	$F23$
197	$I23$	198	$P2_13$	199	$I2_13$	200	$Pm\text{-}3$
201	$Pn\text{-}3$	202	$Fm\text{-}3$	203	$Fd\text{-}3$	204	$Im\text{-}3$
205	$Pa\text{-}3$	206	$Ia\text{-}3$	207	$P432$	208	$P4_232$
209	$F432$	210	$F4_132$	211	$I432$	212	$P4_332$
213	$P4_132$	214	$I4_132$	215	$P\text{-}43m$	216	$F\text{-}43m$
217	$I\text{-}43m$	218	$P\text{-}43n$	219	$F\text{-}43c$	220	$I\text{-}43d$
221	$Pm\text{-}3m$	222	$Pn\text{-}3n$	223	$Pm\text{-}3n$	224	$Pn\text{-}3m$
225	$Fm\text{-}3m$	226	$Fm\text{-}3c$	227	$Fd\text{-}3m$	228	$Fd\text{-}3c$
229	$Im\text{-}3m$	230	$Ia\text{-}3d$				

*In parentheses there non-standard space groups used in the code.

6.3 List of layer groups

Triclinic							
1	$p1$	2	$p\text{-}1$				
Monoclinic / inclined							
3	$p112$	4	$p11m$	5	$p11a$	6	$p112/m$
7	$p112/a$						
Monoclinic / orthogonal							
8	$p211$	9	$p2_111$	10	$c211$	11	$pm11$
12	$pb11$	13	$cm11$	14	$p2/m11$	15	$p2_1/m11$
16	$p2/b11$	17	$p2_1/b11$	18	$c2/m11$		
Orthorhombic							
19	$p222$	20	$p2_122$	21	$p2_12_12$	22	$c222$
23	$pmm2$	24	$pma2$	25	$pba2$	26	$cmm2$
27	$pm2m$	28	$pm2_1b$	29	$pb2_1m$	30	$pb2b$
31	$pm2a$	32	$pm2_1a$	33	$pb2_1a$	34	$pb2n$
35	$cm2m$	36	$cm2e$	37	$pmmm$	38	$pmaa$
39	$pban$	40	$pmam$	41	$pmma$	42	$pman$
43	$pbaa$	44	$pbam$	45	$pbma$	46	$pmmn$
47	$cmmm$	48	$cmme$				
Tetragonal							
49	$p4$	50	$p\text{-}4$	51	$p4/m$	52	$p4/n$
53	$p422$	54	$p42_12$	55	$p4mm$	56	$p4bn$
57	$p\text{-}4m2$	58	$p\text{-}42_1m$	59	$p\text{-}4m2$	60	$p\text{-}4b2$
61	$p4/mmm$	62	$p4/nbm$	63	$p4/mbn$	64	$p4/nmm$
Trigonal							
65	$p3$	66	$p\text{-}3$	67	$p312$	68	$p321$
69	$p3m1$	70	$p31m$	71	$p\text{-}31m$	72	$p\text{-}3m1$
Hexagonal							
73	$p6$	74	$p\text{-}6$	75	$p6/m$	76	$p622$
77	$p6mm$	78	$p\text{-}m2$	79	$p\text{-}62m$	80	$p6/mmm$

6.4 List of plane groups

Number	Group
1	p1
2	p2
3	pm
4	pg
5	cm
6	pmm
7	pmg
8	pgg
9	cmm
10	p4
11	p4m
12	p4g
13	p3
14	p3m1
15	p31m
16	p6
17	p6m

6.5 List of point groups

List of all crystallographic and the most important non-crystallographic point groups in Schönflies and Hermann-Mauguin (international) notations.

Crystallographic point groups:

Hermann-Maugin	Schönflies	In USPEX
1	C_1	C1 or E
2	C_2	C2
222	D_2	D2
4	C_4	C4
3	C_3	C3
6	C_6	C6
23	T	T
$\overline{1}$	S_2	S2
M	C_{1h}	Ch1
mm2	C_{2v}	Cv2
$\overline{2}$	S_4	S4
$\overline{3}$	S_6	S6
$\overline{6}$	C_{3h}	Ch3
m$\overline{3}$	T_h	Th
2/m	C_{2h}	Ch2
mmm	D_{2h}	Dh2
4/m	C_{4h}	Ch4
32	D_3	D3
6/m	C_{6h}	Ch6
432	O	O
422	D_4	D4
3m	C_{3v}	Cv3
622	D_6	D6
$\overline{4}$3m	T_d	Td
4mm	C_{4v}	Cv4
$\overline{3}$m	D_{3d}	Dd3
6mm	C_{6v}	Cv6
m$\overline{3}$m	O_h	Oh
$\overline{4}$2m	D_{2d}	Dd2
$\overline{6}$2m	D_{3h}	Dh3
4/mmm	D_{4h}	Dh4
6/mmm	D_{6h}	Dh6
m$\overline{3}$m	O_h	Oh

Important non-crystallographic point groups

Hermann-Maugin	Schönflies	In USPEX
5	C_5	C5
5/m	S_5	S5
$\overline{5}$	S_{10}	S10
5m	Cv_{5v}	Cv5
$\overline{10}$	Ch_{5h}	Ch5
52	D_5	D5
$\overline{5}$m	D_{5d}	Dd5
$\overline{10}$2m	D_{5h}	Dh5
532	I	I
5$\overline{3}$m	I_h	Ih

6.6 Table of univalent covalent radii used in USPEX

Table of covalent radii (in Å) used in USPEX (for hardness calculations, *etc.*):

Z	Element	radius	Z	Element	radius	Z	Element	radius
1	H	0.31	30	Zn	1.22	63	Eu	1.98
2	He	0.28	31	Ga	1.22	64	Gd	1.96
3	Li	1.28	32	Ge	1.20	65	Tb	1.94
4	Be	0.96	33	As	1.19	66	Dy	1.92
5	B	0.84	34	Se	1.20	67	Ho	1.92
6	Csp^3	0.76	35	Br	1.20	68	Er	1.89
	Csp^2	0.73	36	Kr	1.16	69	Tm	1.90
	Csp	0.69	37	Rb	2.20	70	Yb	1.87
7	N	0.71	38	Sr	1.95	71	Lu	1.87
8	O	0.66	39	Y	1.90	72	Hf	1.75
9	F	0.57	40	Zr	1.75	73	Ta	1.70
10	Ne	0.58	41	Nb	1.64	74	W	1.62
11	Na	1.66	42	Mo	1.54	75	Re	1.51
12	Mg	1.41	43	Tc	1.47	76	Os	1.44
13	Al	1.21	44	Ru	1.46	77	Ir	1.41
14	Si	1.11	45	Rh	1.42	78	Pt	1.36
15	P	1.07	46	Pd	1.39	79	Au	1.36
16	S	1.05	47	Ag	1.45	80	Hg	1.32
17	Cl	1.02	48	Cd	1.44	81	Tl	1.45
18	Ar	1.06	49	In	1.42	82	Pb	1.46
19	K	2.03	50	Sn	1.39	83	Bi	1.48
20	Ca	1.76	51	Sb	1.39	84	Po	1.40
21	Sc	1.70	52	Te	1.38	85	At	1.50
22	Ti	1.60	53	I	1.39	86	Rn	1.50
23	V	1.53	54	Xe	1.40	87	Fr	2.60
24	Cr	1.39	55	Cs	2.44	88	Ra	2.21
25	Mn l.s.	1.39	56	Ba	2.15	89	Ac	2.15
	h.s	1.61	57	La	2.07	90	Th	2.06
26	Fe l.s	1.32	58	Ce	2.04	91	Pa	2.00
	h.s.	1.52	59	Pr	2.03	92	U	1.96
27	Co l.s.	1.26	60	Nd	2.01	93	Np	1.90
	h.s.	1.50	61	Pm	1.99	94	Pu	1.87
28	Ni	1.24	62	Sm	1.98	95	Am	1.80
29	Cu	1.32				96	Cm	1.69

Source: Cordero *et al.*, Dalton Trans. 2832-2838, 2008[19].

6.7 Table of default chemical `valences` used in USPEX

Table of chemical `valences` used in USPEX (for hardness calculations, *etc.*):

Z	Element	valence	Z	Element	valence	Z	Element	valence
1	H	1	35	Br	1	69	Tm	3
2	He	0.5	36	Kr	0.5	70	Yb	3
3	Li	1	37	Rb	1	71	Lu	3
4	Be	2	38	Sr	2	72	Hf	4
5	B	3	39	Y	3	73	Ta	5
6	C	4	40	Zr	4	74	W	4
7	N	3	41	Nb	5	75	Re	4
8	O	2	42	Mo	4	76	Os	4
9	F	1	43	Tc	4	77	Ir	4
10	Ne	0.5	44	Ru	4	78	Pt	4
11	Na	1	45	Rh	4	79	Au	1
12	Mg	2	46	Pd	4	80	Hg	2
13	Al	3	47	Ag	1	81	Tl	3
14	Si	4	48	Cd	2	82	Pb	4
15	P	3	49	In	3	83	Bi	3
16	S	2	50	Sn	4	84	Po	2
17	Cl	1	51	Sb	3	85	At	1
18	Ar	0.5	52	Te	2	86	Rn	0.5
19	K	1	53	I	1	87	Fr	1
20	Ca	2	54	Xe	0.5	88	Ra	2
21	Sc	3	55	Cs	1	89	Ac	3
22	Ti	4	56	Ba	2	90	Th	4
23	V	4	57	La	3	91	Pa	4
24	Cr	3	58	Ce	4	92	U	4
25	Mn	4	59	Pr	3	93	Np	4
26	Fe	3	60	Nd	3	94	Pu	4
27	Co	3	61	Pm	3	95	Am	4
28	Ni	2	62	Sm	3	96	Cm	4
29	Cu	2	63	Eu	3	97	Bk	4
30	Zn	2	64	Gd	3	98	Cf	4
31	Ga	3	65	Tb	3	99	Es	4
32	Ge	4	66	Dy	3	100	FM	4
33	As	3	67	Ho	3	101	Md	4
34	Se	2	68	Er	3	102	No	4

6.8 Table of default goodBonds used in USPEX

Table of default goodBonds used in USPEX (for hardness calculations, *etc.*):

Z	Element	goodBonds	Z	Element	goodBonds	Z	Element	goodBonds
1	H	0.20	35	Br	0.10	69	Tm	0.20
2	He	0.05	36	Kr	0.05	70	Yb	0.20
3	Li	0.10	37	Rb	0.05	71	Lu	0.20
4	Be	0.20	38	Sr	0.10	72	Hf	0.30
5	B	0.30	39	Y	0.20	73	Ta	0.40
6	C	0.50	40	Zr	0.30	74	W	0.30
7	N	0.50	41	Nb	0.35	75	Re	0.30
8	O	0.30	42	Mo	0.30	76	Os	0.30
9	F	0.10	43	Tc	0.30	77	Ir	0.30
10	Ne	0.05	44	Ru	0.30	78	Pt	0.30
11	Na	0.05	45	Rh	0.30	79	Au	0.05
12	Mg	0.10	46	Pd	0.30	80	Hg	0.10
13	Al	0.20	47	Ag	0.05	81	Tl	0.20
14	Si	0.30	48	Cd	0.10	82	Pb	0.30
15	P	0.30	49	In	0.20	83	Bi	0.20
16	S	0.20	50	Sn	0.30	84	Po	0.20
17	Cl	0.10	51	Sb	0.20	85	At	0.10
18	Ar	0.05	52	Te	0.20	86	Rn	0.05
19	K	0.05	53	I	0.10	87	Fr	0.05
20	Ca	0.10	54	Xe	0.05	88	Ra	0.10
21	Sc	0.20	55	Cs	0.05	89	Ac	0.20
22	Ti	0.30	56	Ba	0.10	90	Th	0.30
23	V	0.30	57	La	0.20	91	Pa	0.30
24	Cr	0.25	58	Ce	0.30	92	U	0.30
25	Mn	0.30	59	Pr	0.20	93	Np	0.30
26	Fe	0.25	60	Nd	0.20	94	Pu	0.30
27	Co	0.25	61	Pm	0.20	95	Am	0.30
28	Ni	0.15	62	Sm	0.20	96	Cm	0.30
29	Cu	0.10	63	Eu	0.20	97	Bk	0.30
30	Zn	0.10	64	Gd	0.20	98	Cf	0.30
31	Ga	0.25	65	Tb	0.20	99	Es	0.30
32	Ge	0.50	66	Dy	0.20	100	FM	0.30
33	As	0.35	67	Ho	0.20	101	Md	0.30
34	Se	0.20	68	Er	0.20	102	No	0.30

Bibliography

[1] J. Maddox. Crystals from first principles. *Nature*, 335:201, 1988.

[2] A.R. Oganov and C.W. Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of Chemical Physics*, 124:244704, 2006.

[3] C.W. Glass, A.R. Oganov, and N. Hansen. USPEX — evolutionary crystal structure prediction. *Comp. Phys. Comm.*, 175:713–720, 2006.

[4] A.R. Oganov and S. Ono. Theoretical and experimental evidence for a post-perovskite phase of MgSiO3 in Earth's D" layer. *Nature*, 430(6998):445–448, July 2004.

[5] M. Murakami, K. Hirose, K. Kawamura, N. Sata, and Y. Ohishi. Post-perovskite phase transition in MgSiO3. *Science*, 304(5672):855–858, 2004.

[6] A.R. Oganov, J.C. Schon, M. Jansen, S.M. Woodley, W.W. Tipton, and R.G. Hennig. *Appendix: First Blind Test of Inorganic Crystal Structure Prediction Methods*, pages 223–231. Wiley-VCH Verlag GmbH & Co. KGaA, 2010.

[7] C.J. Pickard and R.J. Needs. High-pressure phases of silane. *Phys. Rev. Lett.*, 97:045504, Jul 2006.

[8] M. Martinez-Canales, A.R. Oganov, Y. Ma, Y. Yan, A.O. Lyakhov, and A. Bergara. Novel structures and superconductivity of silane under pressure. *Phys. Rev. Lett.*, 102:087005, Feb 2009.

[9] Y. Ma, A.R. Oganov, Y. Xie, Z. Li, and J. Kotakoski. Novel high pressure structures of polymeric nitrogen. *Phys. Rev. Lett.*, 102:065501, 2009.

[10] C.J. Pickard and R.J. Needs. High-pressure phases of nitrogen. *Phys. Rev. Lett.*, 102:125702, Mar 2009.

[11] G. Gao, A.R. Oganov, P. Li, Z. Li, H. Wang, T. Cui, Y. Ma, A. Bergara, A.O. Lyakhov, T. Iitaka, and G. Zou. High-pressure crystal structures and superconductivity of stannane (SnH4). *Proceedings of the National Academy of Sciences*, 107(4):1317–1320, 2010.

[12] C.J. Pickard and R.J. Needs. Structures at high pressure from random searching. *physica status solidi (b)*, 246(3):536–540, 2009.

[13] A.O. Lyakhov, A.R. Oganov, H.T. Stokes, and Q. Zhu. New developments in evolutionary structure prediction algorithm USPEX. *Comp. Phys. Comm.*, 184:1172–1182, 2013.

[14] G.R. Qian, X. Dong, X.-F. Zhou, Y. Tian, A.R. Oganov, and H.-T. Wang. Variable cell nudged elastic band method for studying solid-solid structural phase transitions. *Computer Physics Communications*, 184(9):2111–2118, 2013.

[15] C. Dellago, P.G. Bolhuis, F.S. Csajka, and D. Chandler. Transition path sampling and the calculation of rate constants. *The Journal of Chemical Physics*, 108(5):1964–1977, 1998.

[16] S.E. Boulfelfel, A.R. Oganov, and S. Leoni. Understanding the nature of "superhard graphite"'. *Scientific Reports*, 2(471):1–9, 2012.

[17] A.R. Oganov and M. Valle. How to quantify energy landscapes of solids. *The Journal of Chemical Physics*, 130:104504, 2009.

[18] A.R. Oganov, A.O. Lyakhov, and M. Valle. How evolutionary crystal structure prediction works — and why. *Accounts of Chemical Research*, 44(3):227–237, 2011.

[19] B. Cordero, V. Gomez, A.E. Platero-Prats, M. Reves, J. Echeverria, E. Cremades, F. Barragan, and S. Alvarez. Covalent radii revisited. *Dalton Trans.*, 21:2832–2838, 2008.