https://doi.org/10.15407/ujpe71.6.554

B. SEMENIUK,[1] O. FEIA [1, 2, 3, 4]

[1] Kyiv Academic University
(36, Vernadsky Blvd., Kyiv 03142, Ukraine; e-mail: o.feia@kau.edu.ua)
[2] G.V. Kurdyumov Institute for Metal Physics, Nat. Acad. Sci. of Ukraine
(36, Academician Vernadsky Blvd., Kyiv 03142, Ukraine)
[3] Leibniz Institute for Solid State and Materials Research
(IFW Dresden, 20, Helmholtz Str., 01069 Dresden, Germany)
[4] National University "Kyiv Aviation Istitute"
(1, Lubomyr Guzar Ave., Kyiv 03058, Ukraine)

IMPLEMENTATION OF A METHODOLOGY FOR CRYSTAL STRUCTURE PREDICTION USING GENETIC ALGORITHMS INTEGRATED INTO THE PYTHON ASE LIBRARY

This work is dedicated to the development and implementation of a methodology for crystal structure prediction using genetic algorithms integrated into the Python ASE library. Crystal structure prediction plays a critical role in materials science, chemistry, and nanotechnology, enabling the discovery of novel compounds with tailored properties. By combining the flexibility of ASE with the speed of classical relaxers and the accuracy of DFT-based methods, our approach significantly reduces computational costs while maintaining predictive reliability. The methodology was validated on polymorphs of silica (SiO_2), where our system successfully recovered both global and local minima of the energy landscape. We also explore the integration of neural network relaxers such as MACE and AIMNet2 to further accelerate the search process. This study lays the groundwork for efficient, scalable, and accurate predictive modeling of crystalline materials.

Keywords: crystal structure prediction, genetic algorithm, energy landscape, crystalline silica polymorphs, SiO_2, structure relaxation, ASE, GULP.

1. Introduction

The continuous advancement of human civilization demands innovative technologies, which require the creation of advanced materials for their successful implementation. Yet progress in this area is often slowed by conventional material discovery methods that are financially demanding, time-consuming, and test researchers' perseverance. Technological advances have enabled the automation of this process through the integration of new components, such as computer simulations and structure prediction [1, 2], into the conventional materials research workflow.

A major component of this cycle is the technology for structure prediction and synthesis. The central challenges extend beyond the speed and accuracy of prediction algorithms to the practical relevance of the predicted structures. Ideally, such predictions should facilitate the direct experimental re-

Citation: Semeniuk B., Feia O. Implementation of a methodology for crystal structure prediction using genetic algorithms integrated into the Python ASE library. *Ukr. J. Phys.* **71**, No. 6, 554 (2026). https://doi.org/10.15407/ujpe71.6.554.

alization of the proposed compounds in a laboratory setting [2]. The field of computational prediction of materials is fast growing, and quickly incorporates new methods, such as machine learning tools, neural networks, *etc.* [3, 4]. Most algorithms for this purpose are based on finding the global and local minima of a multidimensional energy landscape (Fig. 1) [5].

The energy landscape can be visualized as a continuous surface composed of infinitely many points, where each point corresponds to a particular energy value and, in turn, to a unique atomic configuration. In this representation, the vertical axis reflects the system's energy, while the horizontal dimensions span the structural parameters that can vary (Fig. 1). For example, these parameters can encompass the lattice vectors and the atomic positions within the unit cell, whose collective variation defines the accessible regions of the landscape.

In the context of structure prediction, the central aim is to identify the energy minima, both global and local, on the potential energy landscape. The global minimum corresponds to the most thermodynamically stable structure, which is also the most likely candidate for successful experimental synthesis. Local minima, on the other hand, represent metastable configurations that, while less stable, may still be experimentally accessible under specific thermodynamic or kinetic conditions.

To identify these minima, a variety of search algorithms are employed, such as basin hopping [6], minima hopping [7], metadynamics [8], simulated annealing [9], and particle swarm optimization [10] ((Fig. 2, *a*–*e*). In our work, we use a genetic algorithm (Fig. 2, *f*) [11], which enables efficient exploration and analysis of large portions of the energy landscape with reasonable precision.

The present work aims to develop an efficient methodology for exploring energy landscapes by combining genetic algorithms implemented within the Python ASE (Atomic Simulation Environment) [12] framework with the computational packages GULP [13] and Quantum ESPRESSO [14]. This integrated approach is designed to provide a computationally affordable yet accurate route for identifying the global minima of complex energy surfaces.

2. Computational Methodology

The successful application of genetic algorithms relies on the use of computational "calculators". These tools must be able not only to evaluate the total energy of candidate structures but also to perform structural relaxations, i.e., to optimize atomic configurations so as to locate the nearest local energy minimum on the landscape.

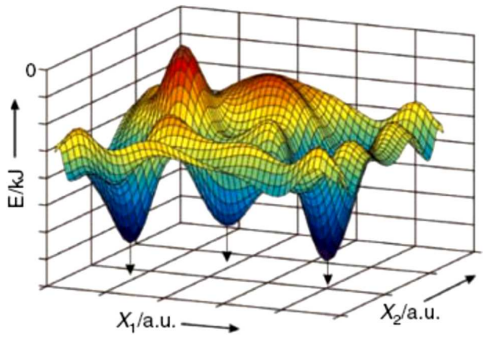

Fig. 1. Multidimensional energy landscape [5]

In our implementation, the genetic algorithm is provided by the Python-based ASE [12] library, while structural relaxations are carried out using two complementary calculators: GULP [13] and Quantum ESPRESSO [14]. GULP employs classical Coulomb-based potentials [13], which results in lower accuracy but allows for much faster evaluations. By contrast, Quantum ESPRESSO relies on density functional theory (DFT) [15, 16], delivering significantly higher accuracy at the expense of substantially greater computational cost. By leveraging the strengths of both tools, our approach uses GULP for rapid initial screening of candidate structures, followed by refinement with Quantum ESPRESSO to ensure reliable and accurate determination of the energy minima.

The genetic algorithm proceeds as an iterative process [12] (Fig. 3), where each iteration is termed a generation (or population). The search begins with an initial population of randomly generated structures. Once their energies are evaluated, the resulting distribution corresponds to a set of randomly scattered points across the energy landscape.

In the subsequent step, the candidate structures are evaluated and ranked according to their energies. Each structure is first relaxed, and then its final energy is determined. Structures are ordered on the basis of these values, with the lowest-energy configuration representing the most stable arrangement. The next generation is constructed from a mixture of selected fractions: a subset of the best-performing structures carried over from the previous generation,

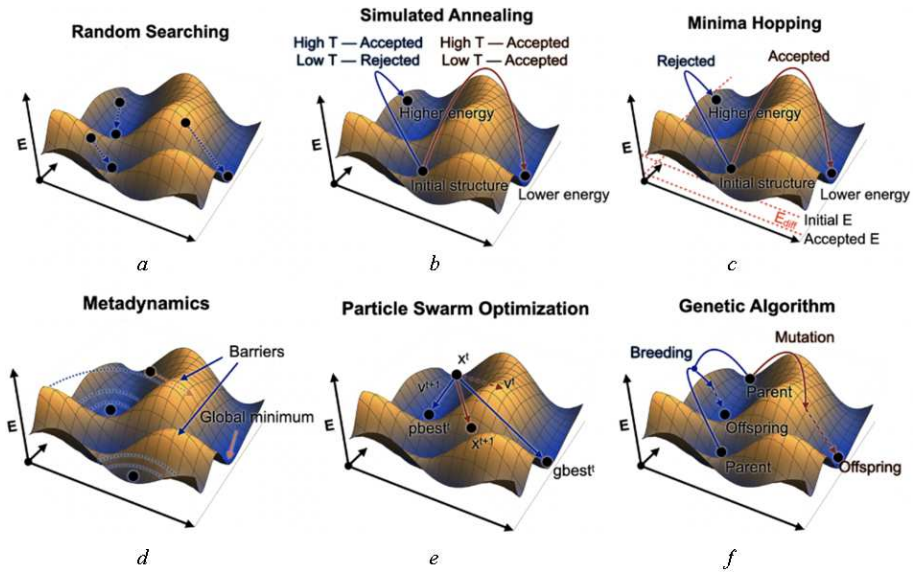

Fig. 2. Different methods of energy landscape global minima optimization [11]

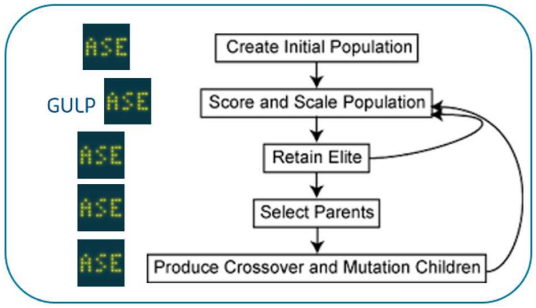

Fig. 3. Iterative genetic algorithm [11, 12]

a portion of mutated variants produced by applying mutation operators to existing structures, and a fraction of newly generated random structures, analogous to those in the initial population. This relaxation–evaluation–selection cycle is iteratively repeated, gradually guiding the search toward lower-energy regions of the landscape.

Mutation operators play a crucial role in this evolutionary scheme by enabling exploration of the local minima around parent structures. Through controlled perturbations, they prevent premature convergence, allowing the algorithm to escape local traps and to sample the surrounding regions of the energy landscape more effectively. In our implementation, three types of mutation operators were employed. The *pairing* (or *cut-and-slice*) operator combines two parent structures by cutting them along a plane and exchanging their parts to generate offspring, thereby enabling the transfer of stable atomic arrangements from one structure to the next generation. The *strain* mutation operator introduces random modifications to the unit cell parameters, facilitating exploration of neighboring regions of the energy landscape. Finally, the *softmutation* operator perturbs atomic positions along low-frequency vibrational modes, estimated using a simple harmonic bond model, thus promoting efficient sampling of structurally relevant configurations [17–20].

During the execution of the algorithm, GULP is mainly employed to carry out rapid evaluations across multiple generations (populations). The search is terminated once a predefined convergence criterion is met, here defined as the repeated appearance of identical structures in the energy–fitness diagram, a point that will be elaborated upon later. This procedure yields a pool of candidate structures with symmetries and geometries closely resembling those of the target (reference) phase. These selected structures are subsequently refined through final relaxation using Quantum ESPRESSO. Importantly, this stage requires only a few dozen calculations, in contrast to the hundreds performed during the initial GULP-based exploration, thus achieving substantial savings in overall computational cost.

For the DFT calculations, we used as input the crystal structures pre-relaxed with GULP. Struc-

tural optimizations were carried out within Quantum ESPRESSO v6.8 [14, 21, 22] using the BFGS quasi-Newton algorithm, minimizing both the total energy and atomic forces. The exchange–correlation functional was treated within the Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation (GGA) [23], implemented using PBEsol pseudopotentials with projected augmented plane wave method [24]. A plane wave cutoff energy of 80 Ry was employed, and both relaxations and self-consistent field calculations were performed using an $8 \times 8 \times 8$ Monkhorst–Pack k-point mesh for the sampling of the Brillouin zone.

3. Results and Discussion

To validate and confirm the effectiveness of our methodology, we selected a material that clearly demonstrates the presence of global and numerous local minima in the energy landscape. Silicon dioxide (SiO_2) [25] was chosen for this purpose due to its rich polymorphism (Fig. 4). In particular, the stable phase under ambient conditions – α-quartz – serves as the global minimum, while metastable phases such as β-quartz, α- and β-cristobalite, tridymite, keatite, coesite, and stishovite represent local minima.

The results of successive populations are conveniently visualized using energy–fitness diagrams (Fig. 5), where the energy of each structure is plotted against its generation number and position within that generation. In these diagrams, each structure is represented by a colored dot: the initial randomly generated population appears in green; the best-performing structures are highlighted in purple, blue, yellow, and red (corresponding to those produced by different mutation operators); while black dots indicate structures discarded due to excessively large unit cell volumes.

Comparison of all four plots reveals a consistent pattern: distinct horizontal lines formed by structures across different generations, each situated at a characteristic energy level. These lines correspond to the local and global minima of the energy landscape. In this case, the lowest line represents α-quartz, while the higher-lying lines correspond to metastable phases such as β-quartz, α- and β-cristobalite, tridymite, keatite, and stishovite. All of these structures were successfully recovered, with the exception of coesite, which was not identified due to its relatively large unit

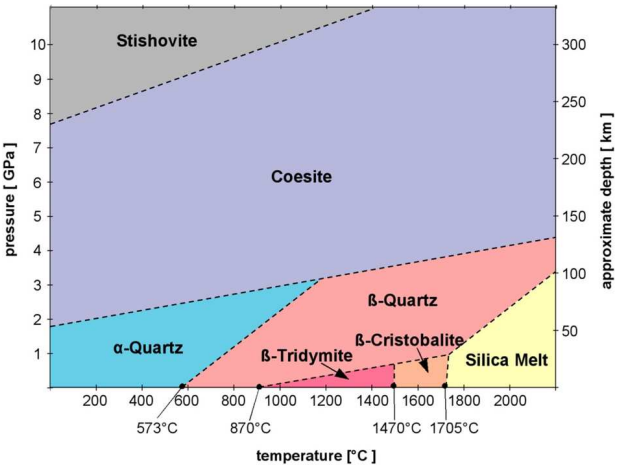

Fig. 4. Phase diagram of silica [25]

cell size, as such extensive calculations were not included in the present study. This outcome highlights the capability of the proposed methodology to capture not only the thermodynamically stable phase but also a broad spectrum of metastable polymorphs, thereby underscoring its effectiveness for structure prediction in complex energy landscapes.

For a clear comparison, the predicted structures were compiled into Table 1 alongside the corresponding reference structures from the Materials Project database [26]. To ensure consistency, each structure was carefully re-relaxed in Quantum ESPRESSO using identical computational parameters. The structural similarity was then quantified by calculating the cosine distance between the fingerprint vectors [28], where a value of zero denotes perfect agreement. On the basis of this analysis, we confirm that the structures generated by our methodology are indistinguishable from the known reference phases.

As a second test case, we investigated the Mg–Al–O system, focusing on magnesium aluminate spinel $Mg(AlO_2)_2$. This compound was selected due to its structural complexity and the additional challenge posed by its ternary composition, which makes structure prediction particularly demanding. Although the global minimum could not be identified in this case, our approach revealed several previously unknown low-energy metastable structures, making this system a noteworthy example to discuss (Table 2).

According to the Materials Project database, only two experimentally confirmed structures are reported for this composition (mp-3536 and mp-5857; see Ta-

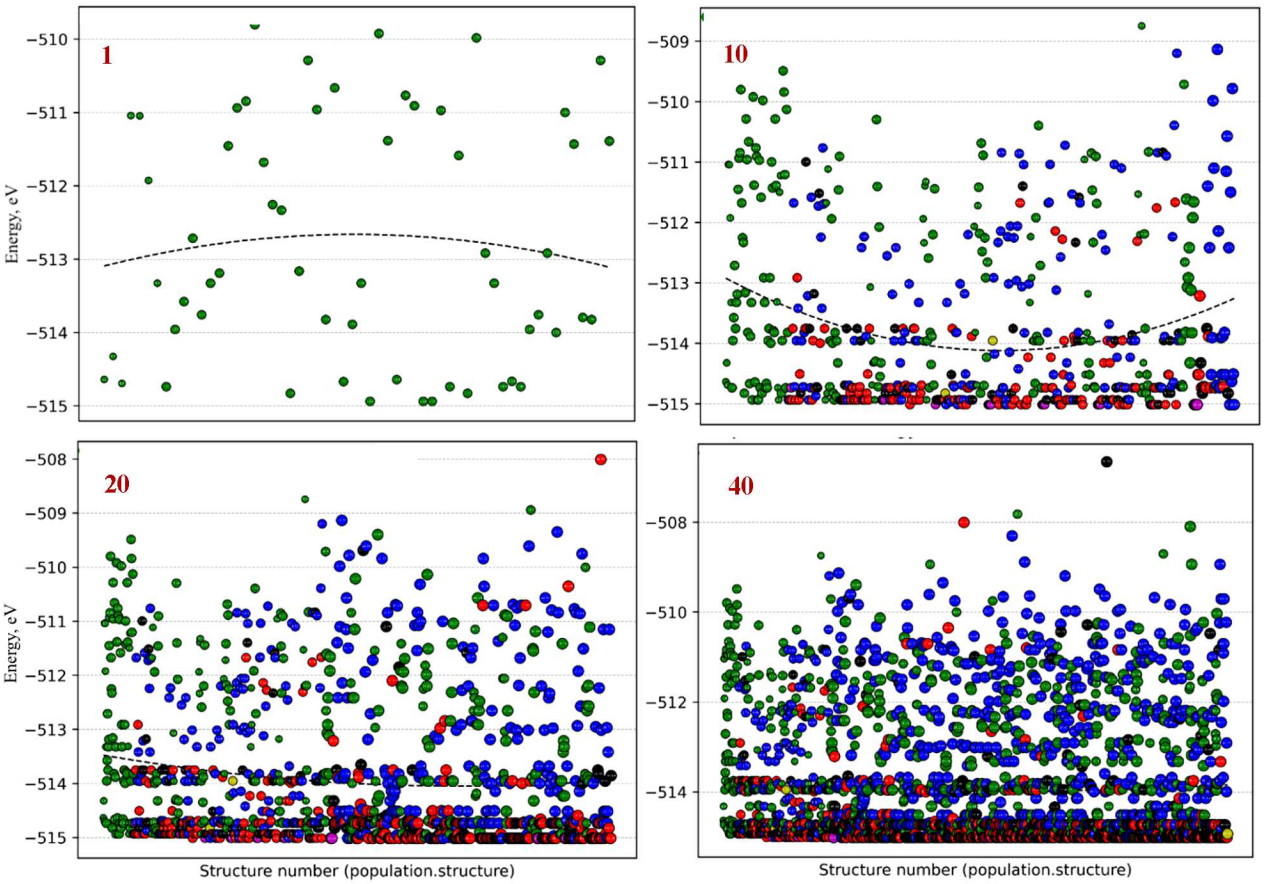

Fig. 5. Results of silica population generation. Green circles represent randomly created structures; pink – best structures from previous population; blue, yellow and red – produced by pairing, softmutation and strain mutation variational operators; black – structures, removed from search because of inadequate parameters. Energy units are electron-volts

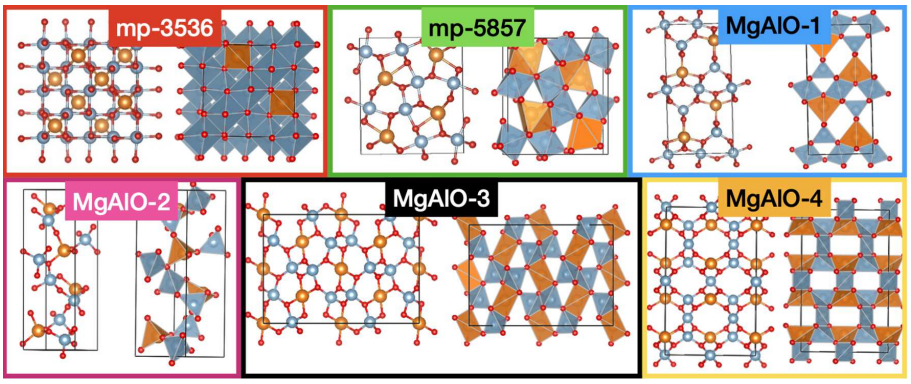

Fig. 6. The crystal structures corresponding to Table 2. Orange spheres represent Mg atoms, blue spheres correspond to Al atoms, and red spheres denote O atoms. In addition, orange polyhedra are drawn around Mg sites and blue polyhedra around Al sites. All crystal structure visualizations were generated using VESTA [27]

ble 2 and Fig. 6). The lowest-energy phase (mp-3536) crystallizes in the $Fd\bar{3}m$ (227) space group and is composed of $[MgO_4]$ tetrahedra and $[AlO_6]$ octahedra. Its conventional unit cell contains 56 atoms, corresponding to 14 atoms in the primitive cell. The second experimentally confirmed structure (mp-5857) adopts orthorhombic Pnma symmetry. Remarkably, one of the structures obtained in our search, denoted

Table 1. **Results of evolutionary search (in pairs of structures – top row) and relaxation of structures from Materials Project (bottom)**

Name	$a\,b\,c$ (Å)	$\alpha\ \beta\ \gamma$ (°)	Space group	Energy/atom (Ry)	Cosine distance
α-quartz	4.961 4.961 5.452	90 90 120	152	–42.767192	0.0
mp-7000	4.96 4.96 5.452	90 90 120	152	–42.767615	
β-quartz	5.081 5.081 5.562	90 90 120	180	–42.76719	0.0039
mp-6922	5.081 5.081 5.562	90 90 120	180	–42.767187	
β-tridymite	5.254 5.254 8.575	90 90 120	194	–42.766982	0.0
mp-559091	5.254 5.254 8.575	90 90 120	194	–42.766976	
β-cristobalite	7.430 7.430 7.430	90 90 90	227	–42.766908	0.0355
mp-8352	7.430 7.430 7.430	90 90 90	227	–42.766929	
Stishovite	4.192 4.192 2.6805	90 90 90	136	–42.762639	0.0069
mp-6947	4.193 4.193 2.680	90 90 90	136	–42.762690	
Fluorite	4.543 4.543 4.543	90 90 90	225	–42.684103	0.02
mp-10064	4.545 4.545 4.545	90 90 90	225	–42.684113	
Hydrophilite	4.082 5.039 4.497	90 90 120	60	–42.581702	0.0025
mp-10498	4.082 5.041 4.497	90 90 120	60	–42.758837	

Table 2. **Mg–Al–O system – structures from Materials Project (first two lines) and results of evolutionary search, labeled as MgAlO–N**

Name	$a\,b\,c$ (Å)	$\alpha\ \beta\ \gamma$ (°)	Space group	Energy/atom (Ry)
mp-3536	8.109 8.109 8.109	90 90 90	227	–54.741383
mp-5857	8.672 2.794 9.969	90 90 90	62	–54.735592
MgAlO-1	3.010 14.010 7.006	90 90 90	63	–54.735508
MgAlO-2	4.903 15.213 5.291	90 111.16 90	9	–54.729543
MgAlO-3	14.769 10.307 5.041	90 90 90	43	–54.729471
MgAlO-4	15.922 5.064 9.756	90 90 90	22	–54.726328

MgAlO-1 (Table 2, Fig. 6), is nearly degenerate in energy with mp-3536 - its energy is only 8×10^{-5} Ry higher – and crystallizes in the Cmcm (63) space group. In the mp-5857 structure, Mg atoms are coordinated by seven O atoms, while Al atoms exhibit a coordination number of six. By contrast, in MgAlO-1 the Mg atoms are five-coordinated, but [AlO_6] octahedral units are still preserved. In the higher-energy structures MgAlO-2, MgAlO-3, and MgAlO-4 (Table 2, Fig. 6), such octahedral building blocks are absent; instead, these phases are composed exclusively of [MgO_4] and [AlO_4] tetrahedra. This suggests that, for $Mg(AlO_2)_2$, the coexistence of Mg- and Al-centered tetrahedra is energetically unfavorable, whereas the presence of [AlO_6] octahedra contributes significantly to stabilizing the structure, as evidenced by both the experimentally confirmed phases and our MgAlO-1 prediction.

There are several possible reasons why the global minimum of the Mg–Al–O system was not reached, despite the identification of numerous low-energy states. The most likely explanation is the size of the unit cell: with 56 atoms, the search space becomes exceedingly large and difficult to sample exhaustively. A second factor is the limitation of GULP, which relies on classical Coulomb-based interactions and may therefore fail to capture the full range of physical effects required for accurately describing this system. This suggests that a more advanced yet computationally efficient calculator would be necessary to improve the search for the true global minimum.

4. Conclusions and Future Improvements

We have implemented a methodology for predicting crystal structures by integrating Python with the GULP and Quantum ESPRESSO relaxers. This approach was successfully validated on binary chemical systems, but its performance remains limited for more complex ternary systems such as Mg–Al–O. The use of two relaxers substantially reduced the overall computational cost and proved effective for identifying both global and local minima. At the same time, it introduced new challenges—most notably, the need to generate input files for GULP, which reduces the overall usability of the workflow.

While GULP provides rapid but less accurate evaluations, Quantum ESPRESSO, as a DFT-based method, offers much higher accuracy at the expense of considerable computational resources [29]. Recent advances in neural-network potentials offer a promising alternative: by being trained on DFT datasets, they can approach DFT-level accuracy while drastically reducing computational demands, albeit with accuracy constrained by the quality of the training set. To address these limitations, we integrated machine-learning pseudopotentials into our workflow. Specifically, we adopted MACE [30] and AIMNet2 [31] as replacements for GULP within the genetic algorithm framework, and we are currently optimizing their performance for structure prediction.

We are grateful to all our colleagues for fruitful discussions at KAU seminar meetings— Dr. Prof. Oleksandr Kordyuk, Dr. Prof. Evhen Len, Dr. Volodymyr Bezguba, and to our collaborator from Carnegie Mellon University, Dr. Prof. Olexandr Isayev. OF thanks Ulrike Nitzsche for technical assistance with the supercomputer cluster of IFW Dresden. This work was supported by the German Federal Ministry of Research, Technology and Space (BMFTR) through the GU-QuMat project (01DK24008). This work is also supported by a grant from the National Academy of Sciences of Ukraine for young scientists' research laboratories "Looking for topological superconductivity in cuprates and iron-based superconductors".

1. D.G. Pettifor. Computer-aided materials design: bridging the gaps between physics, chemistry and engineering. *Phys. Ed.* **32**, 164 (1997).
2. S.G. Louie, Y.H. Chan, F.H. da Jornada, Z. Li, D.Y. Qiu. Discovering and understanding materials through computation. *Nat. Mater.* **20**, 728 (2021).
3. T. Xie, J.C. Grossman.Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
4. S. Feng, H. Zhou, H. Dong. Using deep neural network with small dataset to predict material defects. *Mater. Des.* **162**, 300 (2019).
5. M. Jansen. The energy landscape concept and its implications for synthesis planning. *Pure Appl. Chem.* **86**, 883 (2014).
6. D.J. Wales, J.P. Doye. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111 (1997).
7. B. Schaefer, S. Mohr, M. Amsler, S. Goedecker. Minima hopping guided path search: an efficient method for finding complex chemical reaction pathways. *J. Chem. Phys.* **140**, 214102 (2014).
8. A. Barducci, M. Bonomi, M. Parrinello. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 826 (2011).
9. S. Kirkpatrick, C.D. Gelatt Jr, M.P. Vecchi. Optimization by simulated annealing. *Science* **220**, 671 (1983).
10. Y. Wang, J. Lv, L. Zhu, Y. Ma. CALYPSO: A method for crystal structure prediction. *Comput. Phys. Commun.* **183**, 2063 (2012).
11. Z. Falls, P. Avery, X. Wang, K.P. Hilleke, E. Zurek. The XtalOpt evolutionary algorithm for crystal structure prediction. *J. Phys. Chem. C* **125**, 1601 (2021).
12. A.H. Larsen, J.J. Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dulak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, E.D. Hermes, P.C. Jennings, P.B. Jensen, J. Kermode, J.R. Kitchin *et al.* The atomic simulation environment–a Python library for working with atoms. *J. Phys. Condens. Matter.* **29**, 273002 (2017).
13. J.D. Gale, A.L. Rohl. The general utility lattice program (GULP). *Mol. Simul.* **29**, 291 (2013).
14. P. Giannozzi *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantumsimulations of materials. *J. Phys. Condens. Matter.* **21**, 395502 (2009).
15. P. Hohenberg, W. Kohn. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864 (1964).
16. W. Kohn, L.J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).
17. D.M. Deaven, K.M. Ho. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **75**, 288 (1995).
18. S. Lysgaard, D.D. Landis, T. Bligaard, T. Vegge. Genetic algorithm procreation operators for alloy nanoparticle catalysts. *Top. Catal.* **57**, 33 (2014).
19. P.B. Jensen, S. Lysgaard, U.J. Quaade, T. Vegge. Designing mixed metal halide ammines for ammonia storage using density functional theory and genetic algorithms. *Phys. Chem. Chem. Phys.* **16**, 19732 (2014).
20. M. Van den Bossche, H. Gronbeck, B. Hammer. Tight-binding approximation-enhanced global optimization. *J. Chem. Theor. Comput.* **14**, 2797 (2018).

21. P. Giannozzi *et al.* Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter.* **29**, 465901 (2017).
22. P. Giannozzi *et al.* Quantum ESPRESSO toward the exascale. *J. Chem. Phys.* **152**, 154105 (2020).
23. J.P. Perdew, K. Burke, M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
24. P.E. Blochl. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
25. K. Royce, C. Baars, H. Viles. Defining damage and susceptibility, with implications for mineral specimens and objects: introducing the mineral susceptibility database. *Stud. Conserv.* **68**, 298 (2023).
26. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013),
27. K. Momma, F. Izumi. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Cryst.* **44**, 1272 (2011).
28. N.E. Zimmermann, A. Jain. Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity. *RSC Adv.* **10**, 6063 (2020).
29. J.S. Smith, O. Isayev, A.E. Roitberg. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192 (2017).
30. I. Batatia, D.P. Kovacs, G. Simm, C. Ortner, G. Csanyi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process* **35**, 11423 (2022).
31. R. Zubatyuk, J.S. Smith, J. Leszczynski, O. Isayev. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **5**, eaav6490 (2019).

Received 15.10.25

Б. Семенюк, О. Фея

ПРОГНОЗУВАННЯ КРИСТАЛІЧНОЇ СТРУКТУРИ З ВИКОРИСТАННЯМ ГЕНЕТИЧНИХ АЛГОРИТМІВ БІБЛІОТЕКИ ASE НА МОВІ PYTHON

Робота присвячена розробці та впровадженню методики прогнозування кристалічної структури з використанням генетичних алгоритмів, інтегрованих у бібліотеку ASE на мові Python. Поєднуючи гнучкість ASE зі швидкістю класичних релаксаторів і точністю методів DFT, наш підхід значно знижує обчислювальні витрати, водночас зберігаючи прогностичну надійність. Методика була перевірена на поліморфах кремнезему (SiO_2), де наша система дала змогу успішно обчислити як глобальні, так і локальні мінімуми енергетичного ландшафту. Також досліджено інтеграцію релаксаторів на основі нейронних мереж, таких як MACE та AIMNet2, для подальшого пришвидшення процесу пошуку. Це дослідження закладає основу для ефективного, масштабованого й точного прогностичного моделювання кристалічних матеріалів.

Ключові слова: прогнозування кристалічної структури, генетичний алгоритм, енергетичний ландшафт, поліморфи кристалічного кремнезему, SiO_2, релаксація структури, ASE, GULP.