

Bridging accuracy and efficiency: assessing universal ML potentials for niobium-oxygen clusters

Ilya S. Popov^{1,a}, Albina A. Valeeva^{1,b}, Andrey N. Enyashin^{1,c}

¹Institute of Solid State Chemistry, Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia

^apopov@ihim.uran.ru, ^banibla_v@mail.ru, ^cenyashin@ihim.uran.ru

Corresponding author: I. S. Popov, popov@ihim.uran.ru

ABSTRACT Machine Learning Interatomic Potentials (MLIPs) promise to combine the accuracy of DFT with the speed of classical force fields. However, their reliability for complex, multi-component systems requires rigorous validation. Here, we perform a targeted evaluation of three leading universal MLIPs using niobium oxide clusters (Nb_nO_m , $n \leq 6$, $m \leq 6$) as a challenging test case. The Nb–O system is very well suited for this task due to its complex electronic interactions, manifested in existence of the bulk phase with 25% vacancy-ordered lattice and, at nanoscale, by a diverse range of non-stoichiometric clusters. We employ a dataset of global minima structures identified via DFT-based evolutionary search as a strict reference. A comparative analysis is then performed by executing evolutionary searches with the MLIPs. By directly comparing predicted structures, energies, and relative stability, we provide a comprehensive assessment of the accuracy and limitations of current universal potentials for modeling complex nanoscale oxides.

KEYWORDS DFT, Machine learning potential, MLIP, Evolutionary algorithm, Niobium oxide, NbO, Nanoparticle, Cluster, USPEX.

ACKNOWLEDGEMENTS This work was supported by the Russian Science Foundation (project 19-73-20012-II, <https://rscf.ru/en/project/19-73-20012/>) and was performed at the Institute of Solid State Chemistry UB RAS.

FOR CITATION Popov I.S., Valeeva A.A., Enyashin A.N. Bridging accuracy and efficiency: assessing universal ML potentials for niobium-oxygen clusters. *Nanosystems: Phys. Chem. Math.*, 2025, **16** (5), 619–627.

1. Introduction

The rapid development of universal machine learning interatomic potentials (MLIPs) [1–3] has opened new avenues for high-throughput materials screening and the modeling of complex processes. Trained on extensive datasets [4–8] comprising millions of structures, these potentials claim to predict energies and forces for a wide range of chemical systems with accuracy close to that of the density functional theory (DFT) calculations, but at a fraction of the computational cost. Despite the emergence of general metrics and rankings (such as Matbench Discovery [9]), the reliability of these potentials for specific, particularly complex multi-component systems, remains questionable and requires rigorous validation against well-characterized reference data.

In this work, we perform a targeted evaluation of three top-ranked universal MLIPs according to the Matbench Discovery ranking (Summer 2025), using niobium oxide clusters (Nb_nO_m) as a reference test system. The choice of this system is motivated by a number of unique factors. Niobium monoxide (NbO) possesses a non-trivial crystal lattice (space group Pm-3m), which is derived from the B1 (NaCl) structure type with 25% ordered vacancies in both sublattices [10–13]. The stability of this vacancy-ordered crystal structure is due to complex electronic interactions involving the formation of Nb–Nb bonds via delocalized d-electrons [13, 14]. At nanoscale, the phase equilibria of the Nb–O system become even more complex: in processes such as magnetron sputtering, numerous clusters of diverse stoichiometric composition (NbO , NbO_2 , Nb_2O_5 , Nb_3O_7 , and others) are observed [15–17], whose precise molecular structures have long remained an unresolved problem. Thus, the Nb–O system serves as an excellent testing ground for MLIPs, as it combines complex chemical bonding, a wide range of stoichiometries and sizes, as well as the presence of surfaces.

For this validation, we employ a previously curated reference dataset [18], generated via an exhaustive search for the global energy minima of Nb_nO_m clusters ($n \leq 6$, $m \leq 6$) using the USPEX evolutionary algorithm in combination with DFT calculations. The first part of the paper briefly summarizes the results of the creation and characterization of this reference dataset. In the second part, we conduct a large-scale comparative study by performing an evolutionary structural search for the same set of clusters using the three selected universal MLIPs. Through a direct comparison of the structures, energies, and predicted relative stability obtained by the DFT+USPEX and MLIPs+USPEX methods, we provide a comprehensive assessment of the accuracy, reliability, and limits of applicability of modern universal potentials for predicting the properties of complex nanoscale oxides.

2. Models and methods

The evolutionary search for the thermodynamically most stable Nb_nO_m clusters of a given composition was performed using the USPEX package [19, 20]. USPEX has been successfully applied previously to determine the structures of Si_n [21], Si_nO_m [20], $(\text{TiO}_n)_n$ [22], Ce_nO_m [23], Fe_nO_m [23], B_nP_n [24], Pd_nBi_m [25], Al_6CM_n ($M = \text{Li}, \text{Na}, \text{K}$; $n = 2, 4, 6$) [26], as well as Mg_n clusters using other evolutionary algorithms [27], among others. The criterion for “natural selection” in USPEX was the lowest value of the total energy of the crystal/molecular structure. Although evolutionary algorithms are effective in locating global and local minima on the potential energy surface, they require significant computational resources. The set of investigated clusters included Nb_nO_m compositions with $1 \leq n \leq 6$ and $0 \leq m \leq 6$. The population size per generation was 20 clusters (plus the most stable configurations from previous generations). The algorithm was terminated after 9 consecutive generations failed to yield a more stable structure. The initial population was generated by randomly selecting point groups and creating structures based on them. Subsequent generations were created based on the following ratio of variation operators: 50% heredity, 20% generation of random structures based on space groups, and 30% mutation. The contribution of variation operators was subsequently automatically adjusted by the program during the calculations to improve the performance of the evolutionary algorithm.

To create the reference set of structures, geometry optimization and total energy calculations for the Nb_nO_m clusters generated by USPEX were performed using the Density Functional Theory (DFT) method, as implemented in the VASP package [28]. Spin-polarized calculations for the clusters were conducted using the Gamma-point approximation, with a vacuum region of at least 12 Å separating the clusters. The exchange-correlation potential was described within the Generalized Gradient Approximation (GGA) using the Perdew-Burke-Ernzerhof (PBE) parameterization. The plane-wave basis set cutoff energy was set to 450 eV. The convergence criteria for the electronic self-consistent cycle and the ionic relaxation cycle were set to 10^{-5} eV and 10^{-4} eV, respectively.

An analogous procedure for searching for the most stable Nb_nO_m clusters was conducted using a combination of USPEX and MLIPs. The initial plan was to use three potentials leading the Matbench Discovery ranking as of summer 2025: eSEN-30M-OAM [29], ORB v3 [30], and SevenNet [31]. However, because the developers of the eSEN-30M-OAM potential denied access to this model, adhering to discriminatory restrictions against citizens of Russia, Belarus, and China, we used another potential from the same developers – UMA [32]. Two UMA potential variants uma-m-1.1 and uma-s-1.1 were used, differing in model size, performance, and forecast accuracy. From here on, the larger and smaller models will be referred to as UMA(m) and UMA(s), respectively. Both potentials were applied with the parameter `task_name = “omat”`. In the case of the ORB v3 potential, the `orb_v3_conservative_inf_omat` model was used as the most accurate one within this family. For the SevenNet potential, the `7net-mf-ompa` model was employed with the parameter `modal = “omat24”`.

The Atomic Simulation Environment (ASE) library [33] was used as an interface for the selected MLIPs to perform geometry optimization of the Nb_nO_m clusters generated by the USPEX program. Since USPEX does not natively support the ASE library, two primary solutions were possible. The first involved writing a custom interface between USPEX and ASE. In this work, an alternative approach was implemented. Within the USPEX calculation parameters, the LAMMPS program [34] was specified for geometry optimization and energy evaluation. The command to launch LAMMPS was set to execute a custom Python script. This script performed the following steps: it converted the LAMMPS input files generated by USPEX into ASE-compatible formats, executed the ASE calculation, and subsequently converted the ASE output files back into the format of LAMMPS output files.

Atomic structure visualization of the clusters was performed using the OVITO program [35]. All graphs were generated using the matplotlib library [36].

3. Results and discussion

3.1. NbO nanoparticles generated by the USPEX and DFT

The search for stable Nb_nO_m clusters that could be based on coordination polyhedra different from those found in crystalline Nb_3O_3 and Nb_4O_4 polymorphs was conducted using the USPEX program. The evolutionary algorithm was run once for each chemical composition with given n and m indices, without repeated runs, due to the high computational costs and the long execution time of each calculation.

To illustrate the operation of the evolutionary algorithm, let us consider the study of Nb_6O_6 clusters as an example. Fig. 1 shows the total energies of all optimized Nb_6O_6 clusters generated by the USPEX algorithm, excluding clusters with very high total energy values. Horizontal lines separate configurations belonging to 13 sequentially generated populations, arranged from bottom to top. The configuration with the lowest total energy in a given generation is highlighted with a yellow circle. Configurations resulting from the action of heredity, random structure generation, and mutation variation operators, as well as the best configurations from previous generations, are denoted by red, green, blue, and gray dots, respectively. In the first generation, configuration number 13, characterized by an octahedral Nb atom framework, was the most stable. In the second and third generations, more stable configurations were identified, including Nb frameworks in the shape of a triangular prism or two tetrahedra sharing a common edge. Other even more stable clusters were not found in the fourth generation. The most stable configuration, number 93, was discovered in the fifth generation, after

which no new, more stable configurations were found. A total of 298 Nb_6O_6 configurations were investigated before the evolution stopping criteria were met. A configuration based on random generation proved to be the most stable only in the first generation, where other variation operators are not applied. Further reduction of the cluster's total energy was achieved by applying heredity and mutation operators to configurations obtained earlier. Nevertheless, random generation was continued to explore the configuration space and introduce greater diversity into the Nb_nO_m populations.

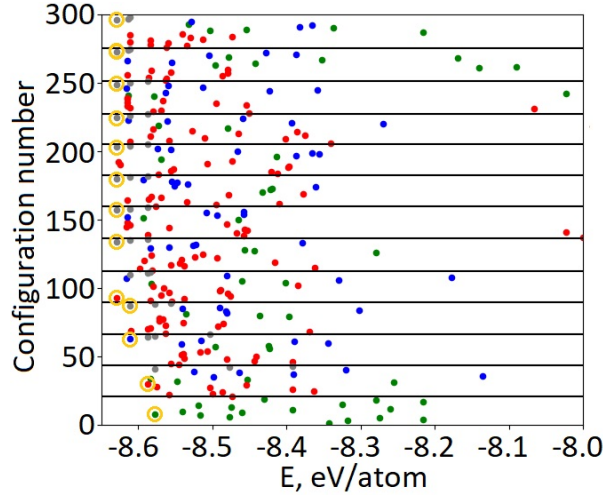


FIG. 1. Total energies of Nb_6O_6 clusters obtained during the USPEX evolutionary algorithm based on DFT calculations

The configurations of the most stable Nb_nO_m clusters for each of the investigated compositions are shown in Fig. 2. All clusters are characterized by dense atomic packing and minimal surface area. Clusters with an equal number of Nb atoms can be described by one or two types of Nb frameworks, to which O atoms are sequentially attached. For example, clusters with two Nb atoms form a dimer surrounded by oxygen atoms. Clusters with three Nb atoms tend to form a triangle of Nb atoms. Four Nb atoms in a cluster lead to the formation of a tetrahedral framework. For clusters with five Nb atoms, the frameworks were either a triangular bipyramid or a distorted square pyramid, depending on the number and position of the O atoms. Nb_6O_m clusters are characterized by an Nb atom framework forming either an octahedron or two tetrahedra sharing a common edge. The most favorable positions for O atoms in most clusters are located above the centers of Nb-Nb edges. A significant excess of oxygen leads to the formation of $-\text{O}-\text{O}-$ bridges within the clusters, which possess a high excess energy. Nevertheless, such “peroxide-containing” clusters were still found to be more stable than the corresponding oxide clusters with the same number of atoms and individual O_2 molecules. In general, clusters in the studied size range tend to form symmetric or regular geometric configurations. It is likely that further growth in cluster size will lead to a reduction in cluster symmetry and the formation of more amorphous structures, as has been observed in other binary systems [22, 24].

A comparative study of the relative stability of clusters of different compositions was carried out using two descriptors [20]: the disproportionation energy (E_{disp}) of two identical clusters with the transfer of one Nb or O atom from one cluster to another, and the dissociation energy (E_{diss}) of a cluster into two fragments. Higher values of E_{disp} and E_{diss} indicate greater stability of the cluster and a lower tendency for decomposition or disproportionation. The results of the E_{disp} and E_{diss} calculations are shown in Fig. 3. Although the constructed E_{disp} and E_{diss} energy landscapes differ, the positions of the most of their main maxima coincide. According to our results, the most stable clusters have the compositions NbO , NbO_2 , Nb_2O_5 , Nb_2O_6 , and Nb_4O_6 . The existence of NbO , NbO_2 , Nb_2O_5 , and Nb_2O_6 clusters has been previously registered experimentally, along with heavier clusters such as Nb_3O_7 , Nb_4O_9 , and Nb_5O_{12} [15, 16], which are beyond the scope of our study.

$$E_{disp}(\text{Nb}) = E(\text{Nb}_{n+1}\text{O}_m) + E(\text{Nb}_{n-1}\text{O}_m) - 2E(\text{Nb}_n\text{O}_m) \quad (1)$$

$$E_{disp}(\text{O}) = E(\text{Nb}_n\text{O}_{m+1}) + E(\text{Nb}_n\text{O}_{m-1}) - 2E(\text{Nb}_n\text{O}_m) \quad (2)$$

$$E_{disp} = \min(E_{disp}(\text{Nb}), E_{disp}(\text{O})) \quad (3)$$

$$E_{diss} = \min(E(\text{Nb}_{n-x}\text{O}_{m-y}) + E(\text{Nb}_x\text{O}_y) - E(\text{Nb}_n\text{O}_m)) \quad (4)$$

An estimate of the probability of detecting different configurations of particles of the same composition was made using the Boltzmann distribution as a thumb approximation. The corresponding distributions for the compositions Nb_2O_5 , Nb_2O_6 , and Nb_4O_6 are shown in Fig. 4. The graphs for Nb_2O_5 and Nb_2O_6 compositions at temperatures 1500–3000 K

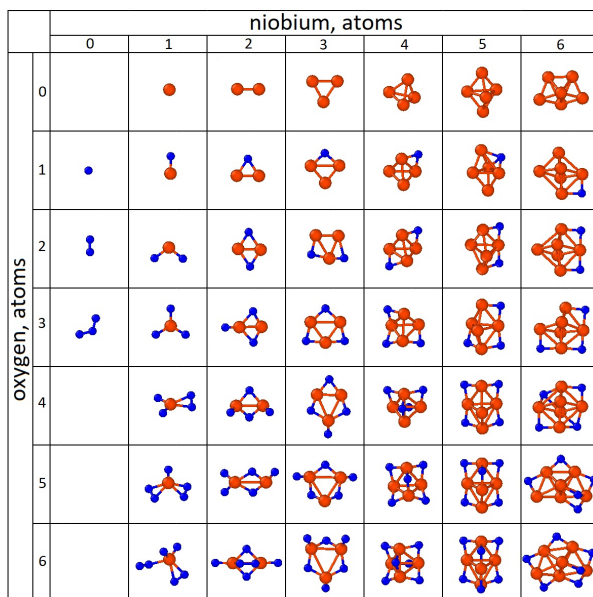


FIG. 2. Structure of the most stable Nb_nO_m clusters as a function of n and m indices, identified using the USPEX evolutionary algorithm coupled with DFT calculations. Nb and O atoms are represented by large orange and small blue spheres, respectively

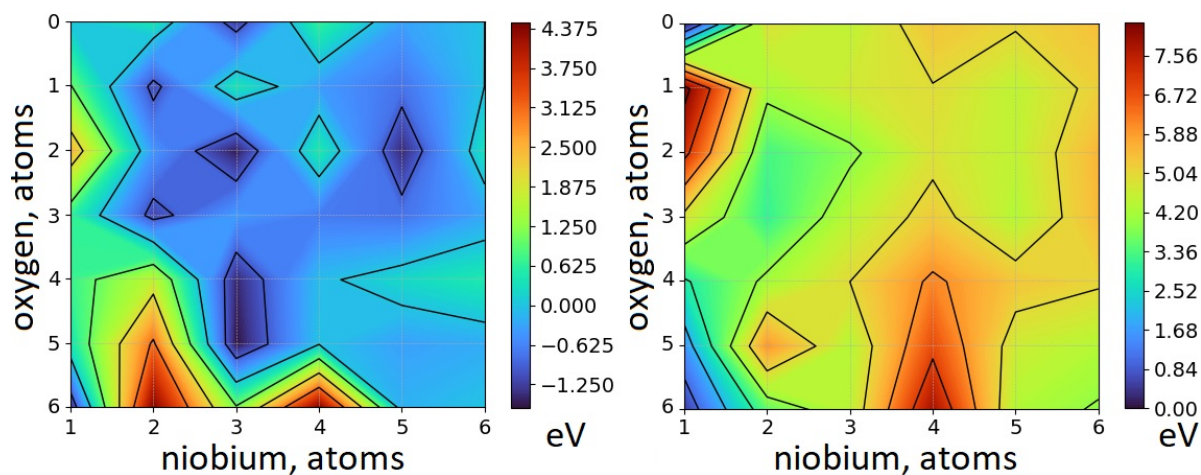


FIG. 3. Comparative analysis of the thermodynamic stability of Nb_nO_m clusters using the disproportionation energy, E_{disp} (left), and the dissociation energy, E_{diss} (right). Higher positive values indicate greater cluster stability. DFT calculations

demonstrate the coexistence of several cluster configurations, with the most stable configuration being predominant. It is quite probable that under the conditions of a high-temperature synthesis, several isomers of these clusters will coexist in the medium. The opposite picture is observed for the Nb_4O_6 composition. Due to the significant energy gap between the most stable cluster and the others, the most stable configuration, which possesses high symmetry, will dominate over the entire temperature range, while other isomers are expected only in trace amounts.

3.2. NbO nanoparticles generated by the USPEX and MLIPs

MLIPs were used in conjunction with evolutionary algorithms to search for the most stable structures of Nb_nO_m clusters, mirroring the approach previously taken with the DFT + evolutionary algorithms combination. The settings for the evolutionary algorithms remained unchanged. For the UMA(m) and SevenNet potentials, a single search without repeats was performed for each composition. For the UMA(s) and ORB potentials, the evolutionary algorithm was run three times to verify the reaching of the global minimum and to assess reproducibility. The structures of the clusters found during the global optimization process with ORB v3 and UMA(m) are shown in Figures S1–S2.

The results of comparing the most stable structures predicted by the MLIPs with those found by DFT are presented in Fig. 5. In the case of the UMA(m) potential, 25 out of 39 most stable Nb_nO_m structures with a total number of atoms between 3 and 12 match the DFT results. However, this potential is the most resource-consuming and computationally

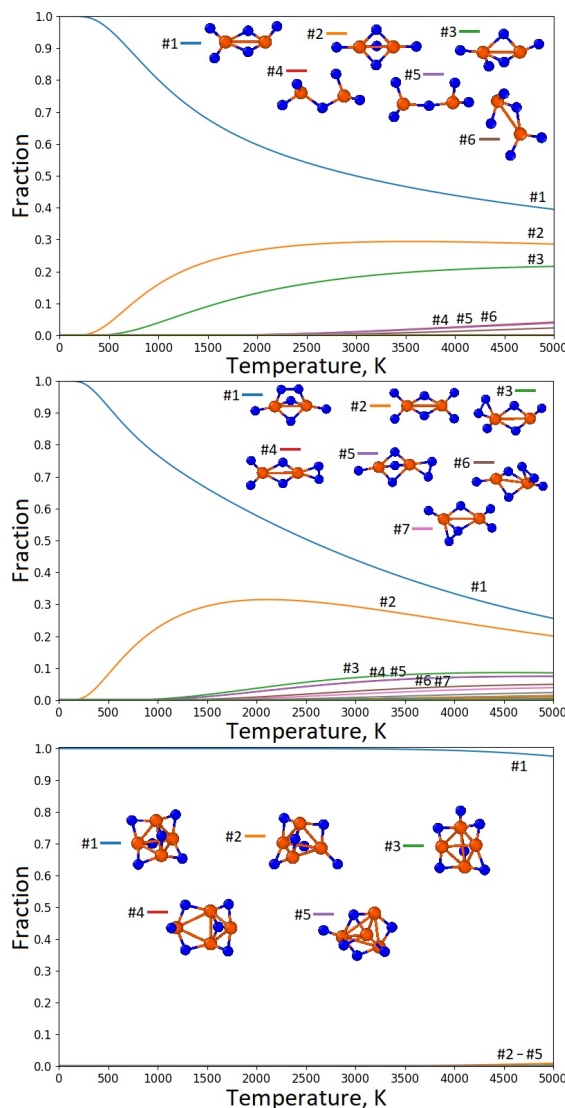


FIG. 4. Estimated probability of detecting different isomeric clusters using the Boltzmann distribution for: Nb_2O_5 (top), Nb_2O_6 (center), and Nb_4O_6 (bottom)

“heavy” compared to the others. The UMA(s) potential demonstrates the lower resource demands and the higher calculation speeds but shows a notable deterioration in results: 16, 14, and 14 out of 39 structures found in each independent run match the DFT results. When selecting the single most stable configuration from each set of three replicates, 14 out of 39 structures remain correctly identified. The structures correctly predicted by UMA(s) are, for the most part, also correctly predicted by UMA(m). The only exception is the Nb_5O_4 composition, where UMA(m) fails to find the most stable configuration according to DFT calculations.

The ORB potential correctly identifies the most stable configurations for 25, 22, and 19 structures in the first, second, and third independent runs, respectively. Combining the results from all independent runs and retaining only one most stable structure per composition yields 22 out of 39 correct structures. However, if the most stable clusters from all three replicates are considered together, they contain 28 out of 39 correctly identified configurations. The SevenNet potential correctly identifies the most stable configurations for 19 out of 39 Nb_nO_m compositions. This value is lower than that of the ORB potential, and almost all compositions correctly predicted by SevenNet are also correctly predicted by ORB.

Consequently, the best results can be achieved by using the UMA(m) and ORB potentials jointly. Their combined predictions allow for the correct identification of 29 out of 39 compositions. This value can be further increased by considering not only the most stable configurations but also several energy-rich configurations. Notably, incorporating the most stable configuration from three independent USPEX+ORB runs already elevates this count to 32 out of 39.

The most stable structures identified via DFT+USPEX were re-optimized using MLIPs. During structural relaxation with MLIP potentials, structural rearrangements were observed in the following cases: ORB v3 – the Nb_6 cluster adopts an octahedral shape; SevenNet – the same process occurs for Nb_6 , along with structural changes in NbO_6 ; UMA(m) – modification of the NbO_6 cluster; UMA(s) – opening of the Nb_2O_2 cycle and deformation of the Nb_5 bipyramid into

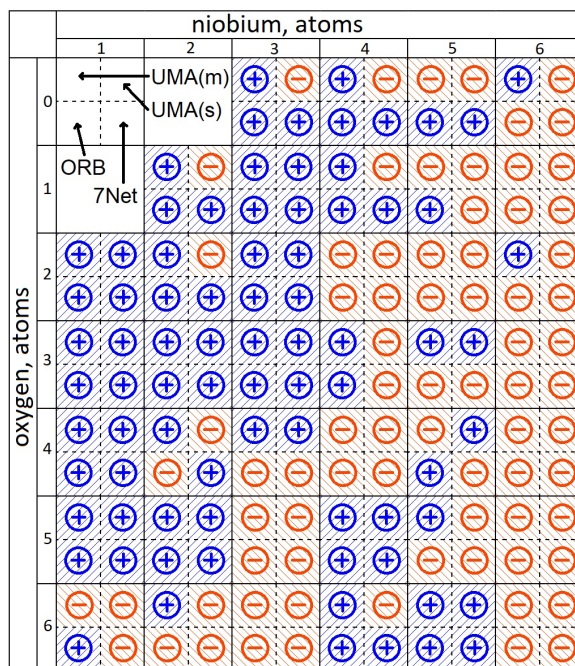


FIG. 5. Comparison of the most stable cluster configurations found using MLIPs and USPEX with the corresponding results from DFT and USPEX. A “+” sign indicates structural agreement, while a “-” sign indicates a discrepancy

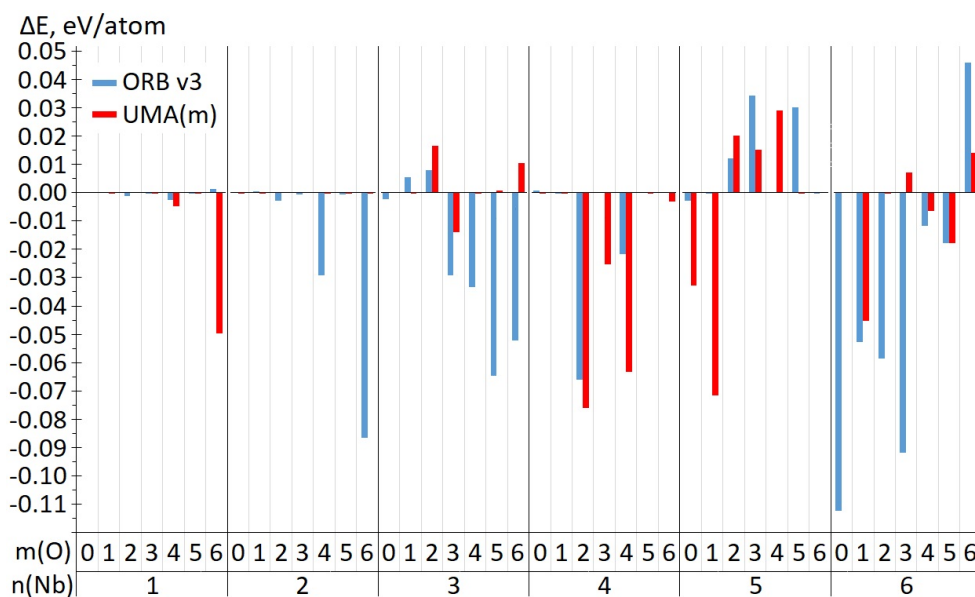


FIG. 6. Energy difference between the most stable clusters found by USPEX+MLIPs and USPEX+DFT. The global minimum structures found via the USPEX+DFT approach were re-optimized and their energies were recalculated using the MLIPs (ORB v3 and UMA(m) models)

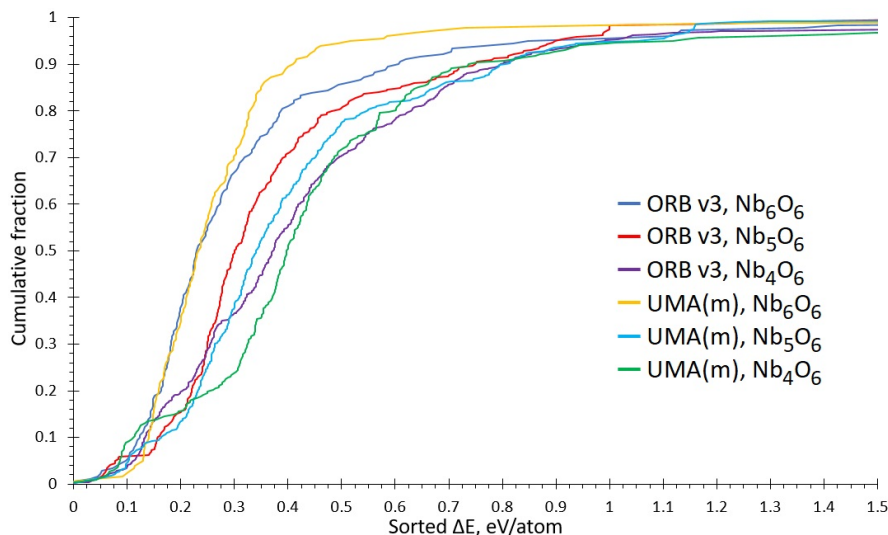


FIG. 7. Cumulative fraction of configurations in the evolutionary search that differ in total energy by no more than ΔE from the most stable configuration found

a planar configuration. In these cases, the total energy was taken from single-point calculations without relaxation. In all other instances, only minor changes in bond lengths and angles were observed. The re-optimization with MLIPs was performed to evaluate where the global minimum of the potential energy surface found via DFT is located on the MLIP potential energy surface and how significantly it differs from the MLIP global minima. The total energy differences are presented in Fig. 6 for ORB v3 and UMA(m), and in Figures S3–S4 for SevenNet and UMA(s). When the ΔE value in Fig. 6 is positive, a more stable structure is identified on the MLIP potential energy surface than that found in the USPEX+MLIP evolutionary search. A negative value indicates that such a configuration could have been discovered during the USPEX+MLIP evolutionary search. For ORB v3 and UMA(m) potentials, the vast majority of ΔE values do not exceed 0.1 eV/atom. In the case of SevenNet and UMA(s) potentials, the deviations can be more substantial, indicating their lower accuracy.

Subsequently, an approximate estimation was performed for the fraction of configurations that lag in total energy by a certain value ΔE from the most stable configuration. For this purpose, USPEX+ORB and USPEX+UMA(m) calculations for Nb_6O_6 , Nb_5O_6 , and Nb_4O_6 compositions were arbitrarily selected. During the analysis, the “keep best” configurations repeated every generation were removed from the sampling, but other possible duplicates were not eliminated. The results are presented in Fig. 7. The obtained data indicate that the number of configurations with ΔE less than 0.1 eV ranges between 5–10% of all configurations considered during the search. To accelerate the global minimum search for a given composition, one can perform such a search using USPEX+MLIPs, followed by DFT geometric re-relaxation for the top 5–10% most stable configurations. However, even within this computational scheme, a significant time saving would be achieved compared to pure DFT calculations for global minimum searches via evolutionary algorithms. Such a combined approach – rapid screening of a large number of configurations using MLIPs followed by the selection of a small number of the most stable candidates for re-evaluation with a more accurate DFT method – holds significant promise for discovering new material phases with desired properties by substantially reducing research time.

4. Conclusion

Using the USPEX evolutionary algorithm in combination with DFT calculations, we have identified the thermodynamically most stable structures of Nb_nO_m clusters ($1 \leq n \leq 6$, $0 \leq m \leq 6$). It was found that clusters with the same number of Nb atoms can be described by one or two types of Nb frameworks, to which O atoms are sequentially attached. The most stable compositions across the studied range were identified as NbO, NbO_2 , Nb_2O_5 , Nb_2O_6 , and Nb_4O_6 .

Furthermore, a comprehensive assessment of four universal machine learning interatomic potentials (MLIPs) – UMA(m), UMA(s), ORB and SevenNet – was performed. The results indicate varying levels of accuracy and computational cost. The UMA(m) potential demonstrated the highest accuracy, correctly predicting 25 out of 39 stable structures, albeit with the highest computational cost. The ORB potential showed strong performance and accuracy close to the UMA(m) potential. A synergistic approach, using the combined predictions of the UMA(m) and ORB potentials, allowed for the correct identification of 32 out of 39 compositions.

This study confirms that while even the best universal MLIPs may not yet be perfectly reliable for standalone global minimum search in complex systems like Nb–O clusters, they are highly valuable in a hybrid computational scheme. The most promising strategy involves the rapid preliminary screening of a vast configurational space using MLIPs, followed

by targeted re-evaluation of a limited number of low-energy candidates with high-precision DFT. This approach should significantly accelerate the discovery of new molecular species and crystal phases.

References

- [1] Wan K., He J., Shi X. Construction of High Accuracy Machine Learning Interatomic Potential for Surface/Interface of Nanomaterials – A Review. *Adv. Mater.*, 2024, **36**(22), P. 2305758.
- [2] Wang G., Wang C., Zhang X., Li Z., Zhou J., Sun Z. Machine learning interatomic potential: Bridge the gap between small-scale models and realistic device-scale simulations. *iScience*, 2024, **27**(5), P. 109673.
- [3] Mortazavi B., Zhuang X., Rabczuk T., Shapeev A.V. Atomistic modeling of the mechanical properties: the rise of machine learning interatomic potentials. *Mater. Horiz.*, 2023, **10**, P. 1956–1968.
- [4] Barroso-Luque L., Shuaibi M., Fu X., Wood B.M., Dzamba M., Gao M., Rizvi A., Zitnick C.L., Ulissi Z.W. Open Materials 2024 (OMat24) Inorganic Materials Dataset and Models. *arXiv*, 2024, [**cond-mat.mtrl-sci**], arXiv:2410.12771.
- [5] Ramlaoui A., Siron M., Djafar I., Musielewicz J., Rossello A., Schmidt V., Duval A. LeMat-Traj: A Scalable and Unified Dataset of Materials Trajectories for Atomistic Modeling. *arXiv*, 2025, [**cs.LG**], arXiv:2508.20875.
- [6] Jain A., Ong S.P., Hautier G., Chen W., Richards W.D., Dacek S., Cholia S., Gunter D., Skinner D., Ceder G., Persson K.A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 2013, **1**, P. 011002.
- [7] Schmidt J., Cerqueira T.F.T., Romero A.H., Loew A., Jäger F., Wang H.-C., Botti S., Marques M.A.L. Improving machine-learning models in materials science through large datasets. *Materials Today Physics*, 2024, **48**, P. 101560.
- [8] Saal J.E., Kirklin S., Aykol M., Meredig B., Wolverton C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *Jom.*, 2013, **65**(11), P. 1501–1509.
- [9] Riebesell J., Goodall R.E.A., Benner P., Chiang Y., Deng B., Ceder G., Asta M., Lee A.A., Jain A., Persson K.A. A framework to evaluate machine learning crystal stability predictions. *Nat. Mach. Intell.*, 2025, **7**, P. 836–847.
- [10] Okatz A.M., Keesom P.H. Specific heat and magnetization of the superconducting monoxides: NbO and TiO. *Phys. Rev. B*, 1975, **12**, P. 4917.
- [11] Bowman A.L., Wallace T.C., Yarnell J.L., Wenzel R.G. The crystal structure of niobium monoxide. *Acta Crystallogr.*, 1966, **21**, P. 843.
- [12] Kurmaev E.Z., Moewes A., Bureev O.G., Nekrasov I.A., Cherkashenko V.M., Korotin M.A., Ederer D.L. Electronic structure of niobium oxides. *J. Alloys Compd.*, 2002, **347**, P. 213–218.
- [13] Hu Z., Qian G., Li S., Yang L., Chen X., Weng M., Tan W., Pan F. Discovery of space aromaticity in transition–metal monoxide crystal Nb₃O₃ enabled by octahedral Nb₆ structural units. *Sci. Bull.*, 2019, **65**, P. 367–372.
- [14] Miura A., Takei T., Kumada N., Wada S., Magome E., Moriyoshi C., Kuroiwa Y. Bonding Preference of Carbon, Nitrogen, and Oxygen in Niobium-Based Rock-Salt Structures. *Inorg. Chem.*, 2013, **52**, P. 9699.
- [15] Music D., Schmidt P., Mráz S. Adsorption of film-forming species on NbO and NbO₂ surfaces. *J. Vac. Sci. Technol. A*, 2017, **35**, P. 061512.
- [16] Fielicke A., Meijer G., Helden G. Infrared Spectroscopy of Niobium Oxide Cluster Cations in a Molecular Beam: Identifying the Cluster Structures. *J. Am. Chem. Soc.*, 2003, **125**, P. 3659–3667.
- [17] Deng H.T., Kerns K.P., Castleman A.W. Formation, Structures, and Reactivities of Niobium Oxide Cluster Ions. *J. Phys. Chem.*, 1996, **100**, P. 13386–13392.
- [18] Popov I.S., Valeeva A.A., Enyashin A.N. Identifying stable Nb–O clusters using evolutionary algorithm and DFT: A foundation for machine learning potentials. *Chem. Phys.*, 2025, **590**, P. 112533.
- [19] Lyakhov A.O., Oganov A.R., Stokes H., Zhu Q. New developments in evolutionary structure prediction algorithm USPEX. *Comp. Phys. Comm.*, 2013, **184**, P. 1172–1182.
- [20] Lepeshkin S.V., Baturin V.S., Uspenskii Y.A., Oganov A.R. Method for simultaneous prediction of atomic structure of nanoclusters in a wide area of compositions. *J. Phys. Chem. Lett.*, 2019, **10**, P. 102–106.
- [21] Heydariyan S., Nouri M.R., Alaei M., Allahyari Z., Niehaus T.A. New candidates for the global minimum of medium-sized silicon clusters: A hybrid DFTB/DFT genetic algorithm applied to Si_n, n = 8–80. *J. Chem. Phys.*, 2018, **149**, P. 074313.
- [22] Olvera-Neria O., García-Cruz R., Gonzalez-Torres J., García-Cruz L.M., Castillo-Sánchez J.L., Poulain E. Strongly Bound Frenkel Excitons on TiO₂ Nanoparticles: An Evolutionary and DFT Approach. *Int. J. Photoenergy*, 2024, **2024**, P. 4014216.
- [23] Yu X., Oganov A.R., Zhu Q., Qi F., Qian G. The stability and unexpected chemistry of oxide clusters. *Phys. Chem. Chem. Phys.*, 2018, **20**, P. 30437.
- [24] Mahdaviyar Z. Prediction of unexpected B_nP_n structures: promising materials for non-linear optical devices and photocatalytic activities. *Nanoscale Adv.*, 2021, **3**, P. 2846.
- [25] Sandu M.P., Kovtunov M.A., Baturin V.S., Oganov A.R., Kurzina I.A. Influence of the Pd:Bi ratio on Pd–Bi/Al₂O₃ catalysts: structure, surface and activity in glucose oxidation. *Phys. Chem. Chem. Phys.*, 2021, **23**, P. 14889.
- [26] Zhou T., Ma L., Chen H. Electronic structure and stability of Al₆CM_n (M = Li, Na, K; n = 2, 4, 6) clusters. *Comput. Theor. Chem.*, 2020, **1178**, P. 112780.
- [27] Steshin I.S., Pantelev S.V., Petukhov I.V., Ignatov S.K. Parametrization of Gaussian approximation potential for the global optimization of magnesium clusters Mg_N (N ≤ 100). *Phys. Chem. Chem. Phys.*, 2025, **27**, P. 18960–18977.
- [28] Kresse G., Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 1996, **54**, P. 11169.
- [29] Fu X., Wood B.M., Barroso-Luque L., Levine D.S., Gao M., Dzamba M., Zitnick C.L. Learning Smooth and Expressive Interatomic Potentials for Physical Property Prediction. *arXiv*, 2025, [**physics.comp-ph**], arXiv:2502.12147.
- [30] Rhodes B., Vandenhoute S., Šimkus V., Gin J., Godwin J., Duignan T., Neumann M. Orb-v3: atomistic simulation at scale. *arXiv*, 2025, [**cond-mat.mtrl-sci**], arXiv:2504.06231.
- [31] Kim J., Kim J., Kim J., Lee J., Park Y., Kang Y., Han S. Data-efficient multi-fidelity training for high-fidelity machine learning interatomic potentials. *arXiv*, 2025, [**cond-mat.mtrl-sci**], arXiv:2504.06231.
- [32] Wood B.M., Dzamba M., Fu X., Gao M., Shuaibi M., Barroso-Luque L., Abdelmaqsoud K., Gharakhanyan V., Kitchin J.R., Levine D.S., Michel K., Sriram A., Cohen T., Das A., Rizvi A., Sahoo S.J., Ulissi Z.W., Zitnick C.L. UMA: A Family of Universal Models for Atoms. *arXiv*, 2025, [**cs.LG**], arXiv:2506.23971.
- [33] Larsen A.H., Mortensen J.J., Blomqvist J., Castelli I.E., Christensen R., Dułak M., Friis J., Groves M.N., Hammer B., Hargus C., Hermes E.D., Jennings P.C., Jensen P.B., Kermode J., Kitchin J.R., Kolsbjerg E.L., Kubal J., Kaasbjerg K., Lysgaard S., Maronsson J.B., Maxson T., Olsen T., Pastewka L., Peterson A., Rostgaard C., Schiøtz J., Schütt O., Strange M., Thygesen K.S., Vegge T., Vilhelmsen L., Walter M., Zeng Z., Jacobsen K.W. The atomic simulation environment – a Python library for working with atoms. *J. Phys.: Condens. Matter*, 2017, **29**, P. 273002.

- [34] Thompson A.P., Aktulga H.M., Berger R., Bolintineanu D.S., Brown W.M., Crozier P.S., Veld P.J., Kohlmeyer A., Moore S.G., Nguyen T.D., Shan R., Stevens M.J., Tranchida J., Trott C., Plimpton S.J. LAMMPS – a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.*, 2022, **271**, P. 10817.
- [35] Stukowski A. Visualization and analysis of atomistic simulation data with OVITO – the Open Visualization Tool. *Modelling Simul. Mater. Sci. Eng.*, 2010, **18**, P. 015012.
- [36] Hunter J.D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 2007, **9**(3), P. 90–95.

Submitted 9 October 2025; accepted 15 October 2025

Information about the authors:

Ilya S. Popov – Institute of Solid State Chemistry, Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia; ORCID 0000-0001-8312-7071; popov@ihim.uran.ru

Albina A. Valeeva – Institute of Solid State Chemistry, Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia; ORCID 0000-0003-1656-732X; anibla_v@mail.ru

Andrey N. Enyashin – Institute of Solid State Chemistry, Ural Branch of Russian Academy of Sciences, Yekaterinburg, Russia; ORCID 0000-0001-6195-7971; enyashin@ihim.uran.ru

Conflict of interest: the authors declare no conflict of interest.