Cluster Generation via Deep Energy-Based Model

A. Yu. Artsukevich
artsukevich@gmail.com

S. V. Lepeshkin
S.Lepeshkin@skoltech.ru

1 Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, 3 Nobel St., Moscow, 143026, Russian Federation
2 Vernadsky Institute of Geochemistry and Analytical Chemistry, Russian Academy of Sciences, Kosygina, 19, Moscow, 119991, Russia
3 Lebedev Physical Institute, Russian Academy of Sciences, 53 Leninskii prosp., 119991, Moscow, Russian Federation

Abstract

We present a new approach for the generation of stable structures of nanoclusters using deep learning methods. Our method consists in constructing an artificial potential energy surface, with local minima corresponding to the most stable structures and which is much smoother than "real" potential in the intermediate regions of the configuration space. To build the surface, graph convolutional networks are used. The method can extrapolates the potential surface to cases of structures with larger number of atoms than was used in training. Thus, having a sufficient number of low-energy structures in the training set, the method allows to generate new candidates for the ground-state structures, including ones with larger number of atoms. We applied the approach to silica clusters \((\text{SiO}_2)_n\) and for the first time found the stable structures with \(n = 28\ldots51\). The method is universal and does not depend on the atomic composition and number of atoms.

1 Introduction

The particles that are sized in order of several nanometers can be found among the most interesting research objects in different scientific domains like condensed matter physics, materials science and other related fields. Being settled
in between molecules and bulk solids, these particles (nanoparticles, nanoclusters or quantum dots) exhibit many unique properties, contrasting to ones of bulk solids, related to the same material. These properties are closely related to atomic structure of the particles, especially in case of small particles and nanoclusters. Despite the importance of knowing the atomic structure, it’s experimental determination remains rather difficult. For this reason the main body of structural information about nanoclusters is gathered with first-principles calculations.

Nowadays the widely acceptable approach to predicting the cluster structure is based on global optimization. A number of corresponding methods have been developed - basin [1] and minima hopping [2], simulated annealing [3, 4], metadynamics [5], quasirandom structure search [6] and evolutionary algorithm [7]. Among all, evolutionary algorithms are widely used in structure prediction since they are population-based, can find various global and local minima with various initial guesses, and often show more robust searching without being trapped in local minima.

Existing evolutionary strategies generally involve two key steps: the initialization of structure population for given chemical composition and the update of the population after the evaluation of target property. The evolutionary algorithm is iterative and for the start it requires creation of an initial population of structures. It is realized by special quasirandom initialization [8], which does not rely on any knowledge about already generated structures.

The use of previously generated structures is significant for the task of cluster structure prediction, where the variable, responsible for the number of atoms in a cluster, is added to the stoichiometric variable. It appears feasible that clusters with close number of atoms belong to the same structural motif [9–11]. Therefore a method allowing to predict structures with use of already known structures is of interest.

Deep learning methods are capable of predicting accurate properties and potential energy of chemical compounds [12–14]. In particular, generative machine learning models are a powerful data-driven approach to sampling from a learned distribution of molecular configuration. Furthermore, some generative models allow to sample 3d molecular configurations, that is one of the most informative way of representation. Models that enable sampling of 3d configurations can be divided into autoregressive [15, 16, 17, 18], generative adversarial network [19], variational autoencoder [20], reinforcement learning [21, 22, 23], equivariant normalizing flows [24], equivariant diffusion models [25].

Molecules are stable compounds whose atoms are covalently bonded to each other. In contrast, clusters are a loosely organized group of atoms that may have a much more complex (and diverse) structure. In result, it is complicated to apply the majority of molecules generation approaches to cluster generation. It seems the only approaches are suitable that generate the whole structure at the same time. The energy-based models are one of straightforward way to ap-
proach generation. They allow to access diverse samples, admit a naive structure representation, and the most important - permit physical interpretation of energy function that may also be called an artificial interatomic potential surface. This specific artificial potential energy can be interpreted as 'smoothed' version of the actual interatomic potential surface. The both of these surfaces have sustainable low-energy cluster structures as local minima with an equal energy. On the other hand, artificial potential surface has no local minima, corresponding with high-energy cluster structures. This option significantly simplifies cluster generation.

In this work we establish an energy-based model and generate silica clusters. An artificial potential of the output satisfies two requirements: the value of the potential coincides with the physical value of the energy for stable structures, whereas the stable structure is a local minimum of this potential surface. The application of the indicated approach to silica clusters \((SiO_2)_n\) has resulted in the discovery of the stable structures with \(n = 28 \ldots 51\).

2 Methods

The method is to built a model that describes a potential surface where each cluster structure, used in training the model, is a local minimum on this surface. To train the model, clusters with different number of atoms are taken. For each chemical compound of the cluster, structures are selected with energies that do not differ significantly from the energy of the global minimum. Unlike the potential surface described using DFT, quantum empirical methods, or their approximations [26–29], the surface constructed through the proposed method has a smaller number of local minima. This property of the potential surface is also hold for clusters with more atoms than were used in training.

The dominance of low-energy local minima on the potential surface for clusters with a large number of atoms can be explained by the idea that the model learns the specific patterns for low-energy structures and extrapolates them. Based on that, it additionally follows that the method cannot generate structures that do not have common patterns with structures with fewer atoms. Structure generation consists in finding local minima of the potential surface, for which the Langevin [30] algorithm is used.

The description of the method is divided into the description of the potential model training, section 2.1, and the description of generation process using the trained potential, section 2.2.

2.1 Artificial potential surface

The training data consists solely of stable cluster structures and their energies. Using this data, it is possible to determine the values of the potential surface \(U\) for the equilibrium structure of cluster. To describe the potential surface for non-equilibrium structures of cluster atoms, the fact of stability of the initial
structure is used: the forces, acting on randomly displaced atoms of the cluster, point in the direction of the equilibrium position of the atoms.

Descriptions of the energy surface and forces acting on atoms are possible with a variety of models \[12\]-\[14\]. The chosen architecture of SchNet \[12\] is relatively simple but powerful, in addition further changes have been made to the architecture. The main differences from the original SchNet: layer norm \[31\] for learning stability, variational dropout \[32\] to prevent overfitting, the activation function LeakySiLU \[13\] - a differentiable analogue of LeakyReLU, radius base functions are implemented by trigonometric functions. For a more detailed description of the architecture used, see Appendix A.

The cluster structure is described by determining the atomic charges \(Z = (z_1, \ldots, z_n)\) and their coordinates \(R = (r_1, \ldots, r_n)\), where \(n\) is the number of cluster atoms. The set of these quantities is denoted by \(X = (Z, R)\). When an atom is taken out of its equilibrium position, it changes only the matrix of atomic coordinates. The charge matrix is always constant. Denote by \(X_0\) - an arbitrary stable structure from the training sample, \(U_0\) - the energy corresponding to it. In the learning process, for each structure, its deformation \(X_\sigma \sim \mathcal{N}(X|X_0, \sigma^2I)\) is randomly constructed, where the value of deformation \(\sigma\) is distributed according to the law \(p(\sigma) \propto 1/\sigma\), \(\sigma \in [\sigma_{\text{min}}, \sigma_{\text{max}}]\). The range of taken \(\sigma\) values is based on the notion that \(\sigma_{\text{max}} \sim L\), where \(L = 1\text{Å}\) is the characteristic interatomic distance, and \(\sigma_{\text{min}} \ll L\).

Finding the weights of the artificial potential model is carried out by minimizing the loss function

\[
\mathcal{L} = w_{\text{property}}\mathcal{L}_{\text{property}} + w_{\text{cosine}}\mathcal{L}_{\text{cosine}} + w_2\mathcal{L}_2
\]

which is the weighted sum of the term describing the energy

\[
\mathcal{L}_{\text{property}} = \mathbb{E}_{X_0}||U(X_0) - U_0||^2
\]

term describing the direction of the gradients of the potential

\[
\mathcal{L}_{\text{cosine}} = \mathbb{E}_{X_0}\mathbb{E}_{\sigma \sim p(\sigma)}\mathbb{E}_{X_\sigma | X_0}\left[1 - \frac{\nabla U(X_\sigma) \cdot (X_\sigma - X_0)}{||\nabla U(X_\sigma)|| ||X_\sigma - X_0||}\right]
\]

and a term describing the stability of the model in the equilibrium region

\[
\mathcal{L}_2 = ||\nabla U(X_\sigma) - (X_\sigma - X_0)||^2.
\]

The weights \(w_{\text{property}}\) and \(w_{\text{cosine}}\) are chosen approximately equal, and the value \(w_2 \ll \min(w_{\text{property}}, w_{\text{cosine}})\). The loss function \(\mathcal{L}_2\) is used solely to eliminate the instability of potential energy gradients in the equilibrium region \[33\].

In general, the graph convolution networks are not able to generalize from small to large graphs \[34\]. In the case of the cluster generation, the more size of the structure, the less often the model finds a good local minima. This feature can be indicated with a concept of a threshold number of atoms in a cluster,
above which the model does not find a good local minima. The threshold value depends on the size of the clusters used to build the model and can be increased by use of the bulk-cut nanocrystal \cite{35}.

### 2.2 Generation

Structure generation implies finding a local minimum of the potential surface $U(x)$. Finding a local minimum is carried out by the Langevin algorithm \cite{30}

\[
R_{k+1} = R_k - \alpha_k \nabla_R \tilde{U}(X_k) + \beta_k \xi,
\]

\[
Z_{k+1} = Z_k
\]

\[
\xi \sim \mathcal{N}(0, I), \quad k = 0, \ldots, N_g - 1
\]

(5)

here $N_g$ is a number of generation steps, $\tilde{U}(X)$ is a potential function that is the sum of the artificial potential $U(x)$ and the regularization potential $\delta U$ described below. The initial state $X_0 = (Z_0, R_0)$ consists of a given charge vector $Z_0$ and the positions of the atoms sampled from normal distribution $R_0 \sim \mathcal{N}(0, \Sigma_n^2 I)$, where $\Sigma_n^2$ - the variance of the position of atoms depending on the number of atoms $n$ is determined by the training set. Exponential parameter decay rate is used

\[
\alpha_k = \alpha_{\max} \left( \frac{\alpha_{\min}}{\alpha_{\max}} \right)^{k/(N_g - 1)}, \quad \beta_k = \beta_{\max} \left( \frac{\beta_{\min}}{\beta_{\max}} \right)^{k/(N_g - 1)}
\]

(6)

Good to note that the exponential decay rate does not guarantee finding the global minimum, but according to our observations it leads to the generation of rather good structures in a small number of steps. We associate the noise parameters with the characteristic parameters of the interatomic distance, namely, $\beta_{\max} \sim 1\text{Å}$ and $\beta_{\min} \ll 1\text{Å}$. The ratios $\alpha_{\min}$ and $\alpha_{\max}$ depend on the potential and require a selection for each potential from the condition for obtaining qualitative structures.

One of the conditions for the Langevin algorithm to find the global minimum of the potential $U(X)$ is its asymptotic behavior \cite{30}

\[
U(X) \to \infty \text{ and } |\nabla_R U(X)| \to \infty \quad \text{if } |R| \to \infty
\]

(7)

This condition is violated by any deep learning model. Any model is defined on the structures of clusters whose atoms rest in some finite volume. This limitation arises due to the fact that the training uses clusters that do not describe all possible volumes. Besides that, it arises when we use the cutoff thresholds by distance in model architectures. Violation of the condition (7) is observed as the generation of an unstructured cloud of atoms or separate groups of structures.

This problem is solved by introducing a regularizing potential $\delta U(x)$ satisfying the property (7). Furthermore, the potential $\delta U(X)$ should slightly distort $U(X)$ on the cluster structures and eventually can be neglected when training
the artificial potential. The algorithm below is designed for constructing the regularizing potential $\delta U(x)$. The symbol $\mathcal{N}$ denotes the set of different atoms that make up the structure. The structure in our algorithm is represented as a fully connected weighted graph $\mathcal{G}$. The nodes of the graph are the atoms of the structure. The edge connecting the atom $i$ with the atom $j$ has the weight $r_{ij}$ corresponding to the distance between these atoms. For each pair of atoms $I, J \in \mathcal{N}$, found in the graph $\mathcal{G}$, we leave only the edges connecting these atoms. On the resulting subgraph, we construct a minimal spanning tree and denote the edges of this tree by $E_{IJ}$. If the weight distribution of $E_{IJ}$ edges is described by a unimodular distribution, the potential energy is defined

$$
\delta U_{IJ}(X) = \sum_{i,j \in E_{IJ}} \left( 2 \log(r_{ij} + \epsilon) - \frac{(r_{ij} - \zeta_{IJ})^2}{2 \eta_{IJ}^2} \right),
$$

\( \zeta_{IJ} = \frac{R_{IJ}^2 - 2 \eta_{IJ}^2}{R_{IJ}}, \ \epsilon \ll R_{IJ} \) (8)

where $R_{IJ}$ is the average weight of the edge $E_{IJ}$, $\eta_{IJ}$ is the characteristic standard deviation of the weights of the edges $E_{IJ}$. The regularizing potential $\delta U(x)$ for the structure is defined as the total of the potentials for each pair of atoms $I$ and $J$.

$$
\delta U(X) = \frac{C}{2} \sum_{I,J \in \mathcal{N}} \delta U_{IJ}(X)
$$

(9)

where $C$ is a constant. The list of pairs of atoms used in the potential (9) and the value of $R_{IJ}$ are determined on the cluster structures used for learning. The values of $C$ and $\eta_{IJ}$ are identified from the analysis of the structures obtained during generation.

It should be noted that the potential (9) does not provide the conditions for finding the global minimum [30], which manifests itself in the appearance of linear structures during generation. The log $r$ term in the function (8) can make the Langevin algorithm unstable. The advantage of the potential (9) is that it practically disappears for stable structures, since it does not require selection of parameters depending on the number of atoms.

3 Experiments

The method is examined on the example of the generation of silicon oxide cluster structures. Training examples with composition $Si_nO_{2n\pm1}$, $n = 10, \ldots, 27$ are obtained by cluster optimization in USPEX code [8, 36, 37]. In addition, we use the structures that have been published in the literature [9, 10, 29]. Two models of artificial potential are trained. The first model is trained on cluster structures with the chemical composition $Si_nO_{2n\pm1}$, $n = 10, \ldots, 21$. The search for structure clusters is carried out for the compositions $Si_nO_{2n}$, $n = 22, \ldots, 30$. The generated structures are compared with those found by USPEX and published in the literature. The second model is trained on cluster structures with
the chemical composition $Si_{n}O_{2n\pm 1}$, $n = 10, \ldots, 30$. The search for cluster structures is carried out for chemical compositions $Si_{n}O_{2n}$, $n = 28, \ldots, 50$. As a result, new structures of silicon oxide clusters are proposed. Clusters of this size have not been previously considered in the literature.

3.1 DFT approximation with MNDO

During our previous calculations \cite{37, 38} we found that $Si - O$ clusters are well described within the semi-empirical MNDO approach that is implemented, for example, in the MOPAC package \cite{39}. This is confirmed by figure 1 where the energy of the first 1000 isomers, found during the global optimization of $Si_{15}O_{30}$ cluster within MNDO approach, is demonstrated along the energies of the corresponding structures, recalculated within the DFT approach (using B3LYP/6-311+G(d,p) implemented in Gaussian code \cite{40, 41}). Therefore, applying this approach can significantly reduce the relaxation time of one structure and allows to generate a huge number of “realistic” (low-energy) structures of silicon-oxide clusters.

3.2 Training samples

The training examples are composed of cluster structures with chemical compositions $Si_{n}O_{2n\pm 1}$, $n = 10, \ldots, 27$ generated by the variable-composition evolutionary algorithm implemented in the USPEX code \cite{8, 36, 37}, combined with the semi-empirical MNDO approach as implemented in the MOPAC package. For each composition, cluster structures were taken with energies not exceeding the minimum energy of more than 6 eV. The total number of structures in this sample reach 22500. Figure 2 shows typical structures obtained. Bulk-cut
nanocrystal with the chemical composition $Si_{95}O_{205}$ was used to regularize the model. It was obtained from a fragment of a silicon oxide crystal by cutting out a ball.

![Common motifs of structures used to train the artificial potential model](image)

Figure 2: Common motifs of structures used to train the artificial potential model

### 3.3 Parameters of the artificial potential model

The parameters of the model are chosen to be $N = 6$ interacting blocks, $H = 4$ heads in each block, $D = 256$ dimension of the hidden vectors of representation of atoms, $FF = 256$ dimension of the inner Atom-wise feedforward layer. Distance encoding is carried out using $K = 32 \sin$ functions and the same number of $\cos$ functions, the scale parameter is chosen equal to $R_{rbf} = 15.0 \AA$, the frequency range from $f_{min} = 0.125$ to $f_{max} = 8.0$. We take $R_{cut} = 7.5 \AA$ as the distance cutoff threshold with a slope of $S_{cut} = 2.0$. To train the model, RMSProp optimization algorithm is used.

The training of the artificial potential model is divided into several steps. Firstly, the model is trained to describe only the energy of clusters using the loss function [2]. Secondly, the task of restoring the cluster structure, the loss function [3], is added. Finally, jumps in the modulus of the gradient around the local minimum, the loss function [4], are removed.
3.4 Generation of cluster structures with composition from $Si_{22}O_{44}$ to $Si_{30}O_{60}$

The training examples are composed only of the structures described in section 3.2 with chemical composition $Si_nO_{2n\pm1}$, where $n = 10, \ldots, 21$. The energies of the selected clusters do not exceed the minimum energy of more than 4 eV. The total number of structures used for training clusters is $15 \cdot 10^3$.

Approximately $74 \cdot 10^3$ generations of cluster structures are carried out for the chemical compositions $Si_nO_{2n}$, $n = 22, \ldots, 30$. All structures found using the artificial potential are optimized using MNDO/MOPAC. For each composition, the 100 lowest energy structures are selected and these structures are optimized by USPEX code for 5 to 10 generations to eliminate defects.

Examples of the retrieved cluster structures are shown in figure 3. It appears that these structures are an extrapolation of the examples used to train the potential (figure 2). All found structures are less energetically favorable than the structures given in [10], except for $Si_{22}O_{44}$. Table 1 confronts the energies of the generated structures with those found during the USPEX code optimization (section 3.2). The comparison indicates that the artificial potential has found structures that are more advantageous in terms of energy.

In the paper [10] shows that up to $Si_{22}O_{44}$ the global minima is formed by columnar-like structures. Starting from $Si_{23}O_{46}$ to $Si_{27}O_{54}$, disk-like structures form the global minimum. The training sample was composed of structures up to $Si_{21}O_{42}$ (figure 3), hence, from columnar-like structures. The method extrapolated this form up to $Si_{27}O_{54}$. Thus, one global minimum $Si_{22}O_{44}$, already proposed in [10], was indicated. However, disk-like structures were not found by the method. The exception is $Si_{29}O_{56}$ (figure 3) that could be extrapolated from $Si_{17}O_{34}$ (figure 2). The fact that there are no disk-like structures demonstrates the limitations of the method. Therefore if characteristic patterns are absent in the training set, they will not be found for structures with a large number of atoms.

The advantage of the method is its versatility. It does not require the development of special potentials for each chemical composition of the cluster [26–29]. The fact that the method extrapolates patterns that have appeared to be optimal on clusters with fewer atoms allows to propose more advantageous structures in less computational time. This is confirmed by the fact that the method found more optimal structures than USPEX code.

3.5 Generation of cluster structures with composition from $Si_{28}O_{56}$ to $Si_{50}O_{100}$

The training examples are composed of the cluster structures described in section 3.2 with chemical formulas $Si_nO_{2n\pm1}$, $n = 10, \ldots, 27$. 992 structures with chemical composition $Si_nO_{2n\pm1}$, $n = 22, \ldots, 30$ obtained in the section 3.4 are added to these training examples. For each atomic composition, cluster struc-
Figure 3: Examples of clusters found by the model trained on $Si_nO_{2n\pm 1}$, $n = 10, \ldots, 21$

| Formula   | USPEX code (eV) | Model (eV)   |
|-----------|-----------------|-------------|
| $Si_{22}O_{44}$ | -16198.153     | -16202.7047 |
| $Si_{23}O_{46}$ | -16935.2157    | -16938.18108|
| $Si_{24}O_{48}$ | -17674.42      | -17675.00966|
| $Si_{25}O_{50}$ | -18411.1303    | -18412.54393|
| $Si_{26}O_{52}$ | -19149.0923    | -19149.72417|
| $Si_{27}O_{54}$ | -19886.1902    | -19886.9401  |

Table 1: Comparison of structures energies found by the USPEX code and the model. The energies are calculated using the semi-empirical MNDO method implemented in the MOPAC package.

Structures are taken with energies not exceeding the minimum energy of more than 3 eV. In addition, clusters with chemical formulas $Si_nO_{2n}$, where $n = 10, \ldots, 27$ described in [9, 10, 29] papers, which were obtained using the heuristic potential [28, 29] and the basing hopping [1] global minimum search algorithm. These clusters are global minima for each composition. The cluster energies are calculated using MNDO. The total number of structures in this sample is 19.
Approximately $120 \cdot 10^3$ generations of cluster structures are carried out for the chemical compositions $Si_nO_{2n}$, $n = 28, \ldots, 50$. All structures found using the artificial potential are optimized using MNDO. For a number of structures, the geometry is changed manually. For each chemical composition, several structures with the lowest energy are selected. These structures are optimized by Gaussian B3LYP/6-31G(d,p) \cite{40, 41}. Figure 4 shows structures with the lowest energy.

From the examples shown in figure 4, we can conclude that the artificial potential model allows us to remember and extrapolate the patterns characteristic of low-energy structures. In particular, the obtained $Si_{28}O_{56}$, $Si_{30}O_{62}$ can be interpreted as an extrapolation of the structures found in \cite{10}. The $Si_{30}O_{72}$ and $Si_{45}O_{90}$ structures, for example, represent an extrapolation of the $Si_{27}O_{54}$ structure found by USPEX code optimization.
Figure 4: Examples of clusters found by the model trained on $Si_{n}O_{2n\pm 1}$, $n = 10, \ldots, 30$. 

$Si_{28}O_{56}$  $Si_{29}O_{58}$  $Si_{30}O_{60}$  $Si_{31}O_{62}$

$Si_{32}O_{64}$  $Si_{33}O_{66}$  $Si_{34}O_{68}$  $Si_{35}O_{70}$

$Si_{36}O_{72}$  $Si_{37}O_{74}$  $Si_{38}O_{76}$  $Si_{39}O_{78}$

$Si_{40}O_{80}$  $Si_{41}O_{82}$  $Si_{42}O_{84}$  $Si_{43}O_{86}$

$Si_{44}O_{88}$  $Si_{45}O_{90}$  $Si_{46}O_{92}$  $Si_{47}O_{94}$

$Si_{48}O_{96}$  $Si_{50}O_{100}$  $Si_{51}O_{102}$
4 Conclusion

We apply energy-base model in combination with graph convolutional networks to construct artificial interatomic potential that reproduces ground-state and low-energy structures correctly. It is significantly smoother than "real" potential in the intermediate regions of the configuration space. Carrying out global optimization in such efficient energy landscape gives much faster convergence compared to the correct (DFT) landscape. Implementation of this approach requires it the initial generation of a sufficient number of good cluster structures, for which other methods of structure generation should be used (e.g. evolutionary algorithms). The model allows to generate larger clusters than ones in the training set.

We applied this approach to silica clusters \((SiO_2)_n\). The model was trained on structures calculated by us and from literature data containing up to \(n=27\) silicon atoms. As a result, we propose novel ground-state candidate structures of \((SiO_2)_n\) clusters with \(n\) up to 51.

References

[1] David J Wales and Jonathan PK Doye. “Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms”. In: The Journal of Physical Chemistry A 101.28 (1997), pp. 5111–5116. arXiv: cond-mat/9803344.

[2] Stefan Goedecker. “Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems”. In: The Journal of chemical physics 120.21 (2004), pp. 9911–9917.

[3] LT Wille. “Searching potential energy surfaces by simulated annealing”. In: Nature 324.6092 (1986), pp. 46–48.

[4] J Christian Schön and Martin Jansen. “First Step Towards Planning of Syntheses in Solid-State Chemistry: Determination of Promising Structure Candidates by Global Optimization”. In: Angewandte Chemie International Edition in English 35.12 (1996), pp. 1286–1304.

[5] Roman Martoňák et al. “Simulation of structural phase transitions by metadynamics”. In: Zeitschrift für Kristallographie-Crystalline Materials 220.5-6 (2005), pp. 489–498. arXiv: cond-mat/0411559.

[6] Chris J Pickard and RJ Needs. “Ab initio random structure searching”. In: Journal of Physics: Condensed Matter 23.5 (2011), p. 053201. arXiv: 1101.3987 [cond-mat.mtrl-sci].

[7] Colin W Glass, Artem R Oganov, and Nikolaus Hansen. “USPEX - Evolutionary crystal structure prediction”. In: Computer Physics Communications 175.11-12 (2006), pp. 713–720.

[8] Andriy O Lyakhov et al. “New developments in evolutionary structure prediction algorithm USPEX”. In: Computer Physics Communications 184.4 (2013), pp. 1172–1182.
[9] Edwin Flikkema and Stefan T Bromley. “Dedicated Global Optimization Search for Ground State Silica Nanoclusters: \((\text{SiO}_2)_N\) \((N = 6 - 12)\)”. In: The Journal of Physical Chemistry B 108.28 (2004), pp. 9638–9645.

[10] Stefan T Bromley and Edwin Flikkema. “Columnar-to-Disk Structural Transition in Nanoscale \((\text{SiO}_2)_N\) Clusters”. In: Physical review letters 95.18 (2005), p. 185505.

[11] Stanislav K Ignatov et al. “How Many Isomers Do Metallic Clusters Have? Case of Magnesium Clusters of up to 55 Atoms”. In: The Journal of Physical Chemistry A 125.30 (2021), pp. 6543–6555.

[12] Kristof Schütt et al. “SchNet: A continuous-filter convolutional neural network for modeling quantum interactions”. In: Advances in neural information processing systems 30 (2017). arXiv: 1706.08566 [stat.ML].

[13] Johannes Klicpera, Janek Groß, and Stephan Günnemann. “Directional Message Passing for Molecular Graphs”. In: International Conference on Learning Representations (ICLR). 2020. arXiv: 2003.03123 [cs.LG]

[14] Johannes Gasteiger, Florian Becker, and Stephan Günnemann. “GemNet: Universal Directional Graph Neural Networks for Molecules”. In: Advances in Neural Information Processing Systems 34 (2021), pp. 6790–6802. arXiv: 2106.08903 [physics.comp-ph]

[15] Niklas W. A. Gebauer, Michael Gastegger, and Kristof T. Schütt. Generating equilibrium molecules with deep neural networks. 2018. arXiv: 1810.11347 [stat.ML]

[16] Niklas Gebauer, Michael Gastegger, and Kristof Schütt. “Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules”. In: Advances in neural information processing systems 32 (2019). arXiv: 1906.00957 [stat.ML]

[17] Yibo Li, Jianfeng Pei, and Luhua Lai. “Structure-based de novo drug design using 3D deep generative models”. In: Chemical science 12.41 (2021), pp. 13664–13675.

[18] Niklas WA Gebauer et al. “Inverse design of 3d molecular structures with conditional generative neural networks”. In: Nature communications 13.1 (2022), pp. 1–11. arXiv: 2109.04824 [cs.LG]

[19] Moritz Hoffmann and Frank Noé. Generating valid Euclidean distance matrices. 2019. arXiv: 1910.03131 [cs.LG]

[20] Vitali Nesterov, Mario Wieser, and Volker Roth. 3DMolNet: A Generative Network for Molecular Structures. 2020. arXiv: 2010.06477 [q-bio.BM]

[21] Gregor Simm, Robert Pinsler, and José Miguel Hernández-Lobato. “Reinforcement Learning for Molecular Design Guided by Quantum Mechanics”. In: International Conference on Machine Learning. PMLR. 2020, pp. 8959–8969. arXiv: 2002.07717 [stat.ML]

[22] Gregor N. C. Simm et al. Symmetry-Aware Actor-Critic for 3D Molecular Design. 2020. arXiv: 2011.12747 [stat.ML]

[23] Søren Ager Meldgaard et al. “Generating stable molecules using imitation and reinforcement learning”. In: Machine Learning: Science and Technology 3.1 (2021), p. 015008. arXiv: 2107.05007 [physics.chem-ph].
[24] Victor Garcia Satorras et al. *E(n) Equivariant Normalizing Flows*. 2021. arXiv: [2105.09016 [cs.LG]]

[25] Emiel Hoogeboom et al. *Equivariant Diffusion for Molecule Generation in 3D*. 2022. arXiv: [2203.17003 [cs.LG]]

[26] S Tsuneyuki et al. “First-Principles Interatomic Potential of Silica Applied to Molecular Dynamics”. In: *Physical Review Letters* 61.7 (1988), p. 869.

[27] BWH Van Beest, Gert Jan Kramer, and RA Van Santen. “Force fields for silicas and aluminophosphates based on ab initio calculations”. In: *Physical Review Letters* 64.16 (1990), p. 1955.

[28] Edwin Flikkema and Stefan T Bromley. “A new interatomic potential for nanoscale silica”. In: *Chemical physics letters* 378.5-6 (2003), pp. 622–629.

[29] Edwin Flikkema and Stefan T Bromley. “Defective to fully coordinated crossover in complex directionally bonded nanoclusters”. In: *Physical Review B* 80.3 (2009), p. 035402.

[30] Saul B Gelfand and Sanjoy K Mitter. “Recursive Stochastic Algorithms for Global Optimization in $\mathbb{R}^d$”. In: *SIAM Journal on Control and Optimization* 29.5 (1991), pp. 999–1018.

[31] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. *Layer Normalization*. 2016. arXiv: [1607.06450 [stat.ML]]

[32] Kirill Neklyudov et al. “Structured Bayesian Pruning via Log-Normal Multiplicative Noise”. In: *Advances in Neural Information Processing Systems* 30 (2017). arXiv: [1705.07283 [stat.ML]]

[33] Yilun Du and Igor Mordatch. *Implicit Generation and Generalization in Energy-Based Models*. 2020. arXiv: [1903.08689 [cs.LG]]

[34] Gilad Yehudai et al. *From Local Structures to Size Generalization in Graph Neural Networks*. 2021. arXiv: [2010.08853 [cs.LG]]

[35] Petter Persson, J Christof M Gebhardt, and Sten Lunell. “The Smallest Possible Nanocrystals of Semiionic Oxides”. In: *The Journal of Physical Chemistry B* 107.15 (2003), pp. 3336–3339.

[36] Artem R Oganov and Colin W Glass. “Crystal structure prediction using ab initio evolutionary techniques: Principles and applications”. In: *The Journal of chemical physics* 124.24 (2006), p. 244704. arXiv: [0911.3186 [cond-mat.mtrl-sci]]

[37] SV Lepeshkin et al. “Method for simultaneous prediction of atomic structure and stability of nanoclusters in a wide area of compositions”. In: *The journal of physical chemistry letters* 10.1 (2018), pp. 102–106. arXiv: [1812.06568 [cond-mat.mtrl-sci]]

[38] Sergey Lepeshkin et al. “Super-oxidation of silicon nanoclusters: magnetism and reactive oxygen species at the surface”. In: *Nanoscale* 8.44 (2016), pp. 18616–18620.

[39] Michael JS Dewar et al. “Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model”. In: *Journal of the American Chemical Society* 107.13 (1985), pp. 3902–3909.
Appendix

A Model architecture

The architecture of the artificial potential model is similar to the architecture of SchNet [12]. The principal design of the model is shown in figure 5a; figure 5b shows the interaction block diagram, figure 5c shows the convolution block diagram.

![Figure 5: Model architecture](image)

Each structure is represented as a set of $n$ atoms with their charges $Z = (z_1, \ldots, z_n)$ and their positions $R = (r_1, \ldots, r_n)$. Each atom in the model is represented as a vector $h^l_i \in \mathbb{R}^D$, where $l = 1, \ldots, N$ is the layer number, $i = 1 \ldots n$ is the atom number, $N$ is the number of interaction layers, $D$ is the dimension of the presentation space. The structure is described as a matrix composed of representation vectors of each atom $H^l = (h^l_1, \ldots, h^l_n)$. As the initial approximation of the vector representation of atoms, the trainable vectors corresponding to the charges of the atoms $E : z_i \rightarrow h^0_i$. Since the charges are integers, the mapping $E$ is implemented through a table.
Based on the positions of the atoms, we calculate the matrix of mutual interatomic distance (IAD) $IAD_{ij} = \| r_i - r_j \|$. This matrix is invariant with respect to rotation and permutation of the atoms. The one-dimensional value of the distance between atoms must be embedded in a high-dimensional space. It can be realised in many ways [12, 13], from which we choose a simple trigonometric expansion with a uniform frequency step. Radius base function (RBF) is described by the formula

$$RBF(r) = [\sin(f_{0}r/R), \ldots, \sin(f_{K-1}r/R), \cos(f_{0}r/R), \ldots, \cos(f_{K-1}r/R)]$$

(10)

where $f_k = f_{\text{min}} \left( \frac{f_{\text{max}}}{f_{\text{min}}} \right)^{k/(K-1)}$, $f_{\text{min}}$ - minimum frequency, $f_{\text{max}}$ - maximum frequency, $K$ - number of periods, $R_{rbf}$ - distance scaling parameter.

The multi-head convolution block implements $H$ graph convolution blocks in parallel. Differentiable distance cutoff mask is calculated from the IAD matrix

$$DDCM_{ij} = 1 - \sigma(S_{\text{cut}} \cdot (IAD_{ij} - R_{\text{cut}}))$$

(11)

where $S_{\text{cut}}$ is a transition slope, $R_{\text{cut}}$ is a threshold distance. The convolution block is described by the formula

$$e_i = \sum_j (f_{rij} \cdot DDCM_{ij}) \odot v_j$$

(12)

Here $v_i$ is the projection of the vector representation of the atom, $f_{rij}$ is the vector encoding the distance between the atoms $i$ and $j$ (see figure 5c), $\odot$ - element-wise multiplication.

As an activation function, the model uses LeakySiLU, a differentiable analogue of LeakyReLU, defined by the formula

$$\text{LeakySiLU}(x) = x \cdot \left[ (1 - ns)\sigma(sx) + ns \right]$$

(13)

here $\sigma(x)$ is the sigma function, $ns$ is the slope of the straight line at $x \to -\infty$, $s$ is the slope of transition from the negative to the positive axis.

The energy of the structure is represented as the sum of the average energy of the atom $E_0(z)$ and the correction $\delta E(Z)$ arising due to the interaction of atoms

$$E = \sum_{z \in Z} E_0(z) + \delta E(Z)$$

(14)

The values of $E_0(z)$ are calculated using linear regression assuming that $\delta E(Z)$ is absent. The task of the model is to calculate the correction $\delta E(Z)$. 
B Other found cluster structures

Examples of the clusters found using the artificial potential model trained on $Si_nO_{2n\pm 1}$, $n = 10, \ldots, 30$
