THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

Cite This: J. Phys. Chem. A 2018, 122, 1504−1509

pubs.acs.org/JPCA

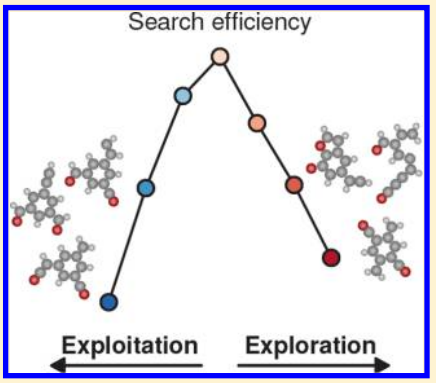
Exploration versus Exploitation in Global Atomistic Structure Optimization

Mathias S. Jørgensen,[†] Uffe F. Larsen,[†] Karsten W. Jacobsen,[‡] and Bjørk Hammer*[,†]

[†]Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus, Denmark

[‡]Center for Atomic-Scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

ABSTRACT: The ability to navigate vast energy landscapes of molecules, clusters, and solids is a necessity for discovering novel compounds in computational chemistry and materials science. For high-dimensional systems, it is only computationally feasible to search a small portion of the landscape, and hence, the search strategy is of critical importance. Introducing Bayesian optimization concepts in an evolutionary algorithm framework, we quantify the concepts of *exploration* and *exploitation* in global minimum searches. The method allows us to control the balance between probing unknown regions of the landscape (exploration) and investigating further regions of the landscape known to have low-energy structures (exploitation). The search for global minima structures proves significantly faster with the optimal balance for three test systems (molecular compounds) and to a lesser extent also for a crystalline surface reconstruction. In addition, global search behaviors are analyzed to provide reasonable grounds for an optimal balance for different problems.



INTRODUCTION

Recent advances in computational resources have provided researchers with the ability to model complex, high-dimensional problems. This includes modeling of the potential energy landscape of materials such as inorganic surfaces,[1] organic systems,[2] and nanostructures.[3] Given that the energy landscape of a system with N atoms is a $3N - 6$ dimensional surface, sampling the landscape of many-atomic systems to find stable configurations is most efficiently done using global optimization techniques. In recent years, a variety of techniques have been developed including minima-hopping,[4] meta-dynamics,[5] Monte Carlo-based methods,[6] particle swarm optimization,[7] basin-hopping,[8] and various implementations of evolutionary algorithms.[9−16] During a search, whenever a new configuration is found, it must be evaluated with quantum chemistry calculations (e.g., density functional theory or density functional tight binding). This evaluation is rather expensive and proves to be the bottleneck of any search technique. Thus, the choice of which configurations to evaluate must be a careful one. In many global structure optimization techniques, the choice is purely stochastic based on only one (as in basin hopping) or a few (as in evolutionary algorithms) other structures from the search.

Efficient global structure optimization balances *exploration* and *exploitation*. The former promotes the ability to jump far in the energy landscape by focusing on disparate regions of the search space, while the latter promotes the ability to search further regions of the landscape known to have low-energy structures. This balance is very intricate because the two abilities seem to oppose each other. In evolutionary algorithms, one tries to satisfy this balance by evolving a population of the most fit structures. One key challenge is how to maintain diversity of the population to avoid stagnation in local minima while still exploiting energetically stable structures. Previous work has demonstrated how to use all known structures from an ongoing search to rationally draw promising parent structures from the population.[17] This is similar to Bayesian optimization,[18] which estimates the most promising structure based on all other structures found in the search. Bayesian optimization has been successfully applied to systems with few degrees of freedom.[19−21] However, because Bayesian optimization needs to sample the entire energy landscape before estimating the most promising structure, this is no longer a viable approach with more than a few degrees of freedom. In this work, we combine the sampling ability of an evolutionary algorithm with elements from Bayesian optimization to build a global structure optimization method that can predict promising regions in configuration space even for high-dimensional problems while quantifying the degree of exploration and exploitation in the search.

OPTIMIZATION FRAMEWORK

Normally, an atomistic structure is assigned a single value for its potential energy as a result of a calculation at some level of theory. Within the Bayesian framework, we ascribe to each structure a probability distribution of energies (presented schematically in Figure 1). As will become clear in a moment,

Received: January 5, 2018
Published: January 9, 2018

DOI: 10.1021/acs.jpca.8b00160
J. Phys. Chem. A 2018, 122, 1504−1509

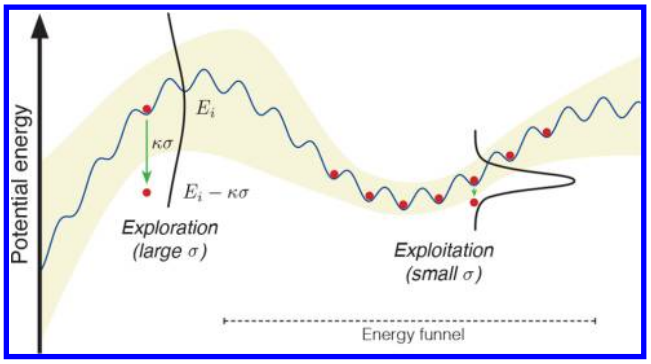


Figure 1. Bayesian fitness function illustrated in a schematic energy landscape (blue line). Red dots represent known local minimum structures, and the shaded yellow area represents the size of the energy variance σ. A structure with energy E_i has an optimistic energy $E_i - \kappa\sigma$ in the Bayesian picture.

structures in regions of configuration space with many other similar structures have narrow distributions, while structures in isolated regions have wide distributions. In other words, this predictive distribution is a measure of the uncertainty of the energy of a structure in the energy landscape. Using a Gaussian process[22] to model the potential energy surface (i.e., any observation noise and the *prior* distribution over model parameters are assumed to be Gaussian), the functional forms of the predictive distributions are Gaussian. The similarity of two structures is defined, as in other kernel methods, in an inner product space of transformed feature vectors $\mathbf{x}$ by a kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\phi}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_j) \tag{1}$$

For any kernel function that can be written in the form of eq 1, the mean and variance of the predictive distribution of structure $N + 1$ are calculated based on N other structures as[23]

$$m(\mathbf{x}_{N+1}) = \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{t} \tag{2}$$

$$\sigma^2(\mathbf{x}_{N+1}) = k(\mathbf{x}_{N+1}, \mathbf{x}_{N+1}) - \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k} \tag{3}$$

where $\mathbf{t}$ is a vector with the target values (energies) of the other N structures, $\mathbf{k}$ is a kernel vector with elements $k(\mathbf{x}_n, \mathbf{x}_{N+1})$ for $n = 1, ..., N$, and $\mathbf{K}$ is the covariance matrix with elements $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The predicted mean m may, in principle, be used to estimate the energy of a structure before local optimization. However, in many global structure optimization problems, this energy does not correlate with the energy after local optimization because the geometry of structures resulting from random structure perturbations are very different from the locally optimized geometries. Hence, in this work, we will only be using the energy variance (eq 3) to enhance the global search. We use the Laplacian kernel, $k(\mathbf{x}_i, \mathbf{x}_j) = \theta_0 \exp[-\theta_1 d(\mathbf{x}_i, \mathbf{x}_j)]$ with $d(\mathbf{x}_i, \mathbf{x}_j) = |\mathbf{x}_i - \mathbf{x}_j|$, which fulfills eq 1 and has proven efficient in kernel regression of atomistic structures.[24,25] The hyperparameter θ_1 determines the width of the kernel (the length scale of correlation between structures), while the meaning of θ_0 is evident in eq 3 because $\sigma^2(\mathbf{x}_{N+1}) \to \theta_0$ when a structure is infinitely far away from all other structures, i.e., $k(\mathbf{x}_n, \mathbf{x}_{N+1}) \to 0$ for $n = 1, ..., N$. Thus, θ_0 can be interpreted as the maximum energy variance of a structure. Both hyperparameters can be easily found by maximization of the log likelihood function for a multivariate Gaussian distribution.[23] Our feature vectors $\mathbf{x}$ contain sorted, exponentially scaled interatomic distances, $z = \exp[-r_{ij}/(r_{\mathrm{cov},i} + r_{\mathrm{cov},j})]$, where $r_{\mathrm{cov},i/j}$ are the covalent radii of atoms i and j. The vectors are sorted according to the type of distance (e.g., A−A, B−B, and A−B distances for atoms of type A and B) and the size of the z entries, which is similar to the Bag of Bonds descriptor.[26]

It is now evident that if a structure is surrounded by other structures in configuration space the predicted energy variance σ^2 is low. If, on the other hand, it is isolated, the variance is high. With this variance, we can directly apply a Bayesian acquisition function to describe the fitness F_i of a structure i in an evolutionary algorithm (EA) framework. Here we use a weighted version of the upper confidence bound (UCB) function:

$$F_i = -(E_i - \kappa\sigma_i) \tag{4}$$

where E_i is the potential energy of structure i and κ is a factor that determines the relative weight of the calculated energy and the predicted uncertainty in the fitness function. Thus, κ is a quantitative measure of the degree of exploration in the global minimum search, and by varying κ, we can now control the balance between exploitation of low-energy structures and exploration of uncertain structures. In practice, F_i is calculated for structure i from the other N structures that have been found in the current state of a search (i.e., structure i is structure number $N + 1$ in eq 3).

Methods and Test Details. We demonstrate the effect of balancing exploitation and exploration by performing a series of independent EA runs searching for the most stable molecular compounds for systems of size 17−22 atoms of type C, O, N, and H. In addition, an anatase $TiO_2(001)$-(1×4) surface reconstruction (the "added molecule model"[27]) is searched for with five TiO_2 units atop a fixed, one-layer slab consisting of four TiO_2 units that are periodic in the x- and y-directions. The EA is described in detail elsewhere[28,29] and is implemented in the atomic simulation environment (ASE).[30,31] For each value of κ, statistics on the search is obtained from 300 independent EA runs. Each run is initialized with a population of $N_{\mathrm{pop}} = 5$ structures with random atomic coordinates (obeying a minimum and maximum distance to other atoms equal to 0.7 and 1.4 times the sum of covalent radii, respectively, for the molecular systems and 0.6 and 1.2 for the TiO_2 surface). The molecular systems are confined to two dimensions. During the search, the population always contains the N_{pop} structures with the highest F_i values. New structures are produced by drawing structures randomly from the populations and pairing them with the cut-and-splice crossover operator. Every time a new structure is produced, the fitness F_i is updated for all structures by recalculating their variances one at a time via eq 3 in a manner reminiscent of the "leave-one-out cross-validation" technique. In addition, the hyperparameters θ_0 and θ_1 are recalculated. All local structure optimizations and energy evaluations are performed within the density functional tight binding scheme using the DFTB+ software[32] and parameters from refs 33 and 34.

RESULTS AND DISCUSSION

Detailed search results for a system with stoichiometry $C_{10}H_6O_2$ are presented below. The cumulative success of all runs for different values of κ is shown in Figure 2. For the standard search with $\kappa = 0$, 37% of the 300 runs have found the global minimum (inset of Figure 2) after 1000 new structures (EA attempts). Increasing the degree of exploration in the search enhances the performance of the EA toward a success of 65% for $\kappa = 4$ where the balance between exploitation and

DOI: 10.1021/acs.jpca.8b00160
J. Phys. Chem. A 2018, 122, 1504−1509

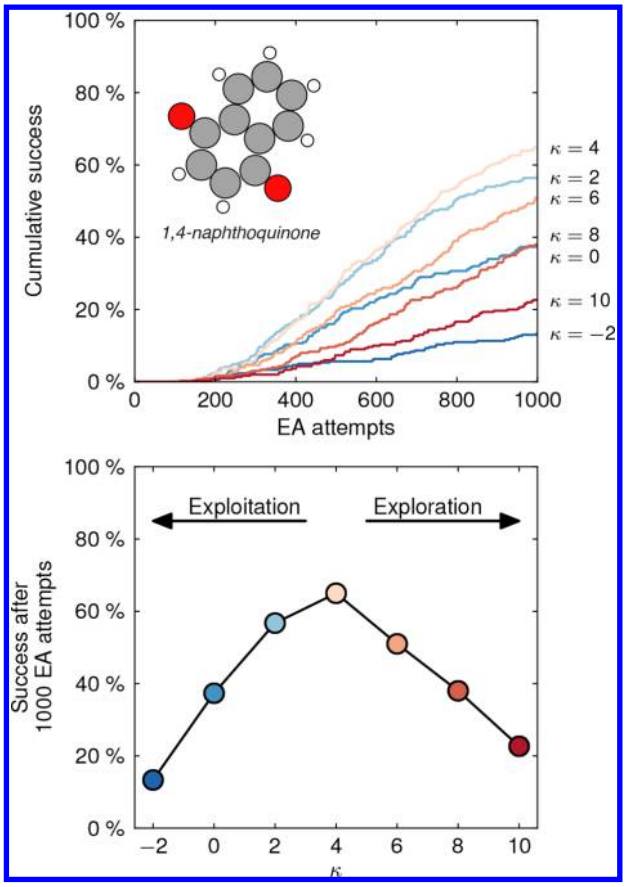


Figure 2. Top: Performance of the EA search for various values of κ. Bottom: Percentage of successful runs (i.e., runs that have found the global minimum) after 1000 EA attempts as a function of the κ values. $\kappa = 4$ is the optimal balance between exploration and exploitation. The global minimum structure is 1,4-naphthoquinone.

exploration is optimal. In the case of an unknown global minimum structure, a higher success rate converts into fewer required restarts of the search before finding or confirming the global minimum, thus making the overall search faster.

The term $\kappa\sigma$ in eq 4 can prohibit a search from spending an excessive amount of time in the same energy funnel by penalizing the increasing number of similar structures. Energy funnel is a frequently used term to describe a higher-level feature of the energy landscape containing several local minima[35] (like the one depicted in Figure 1). In Figure 3, we analyze the evolution of the populations by plotting the average population energy versus the population diversity (calculated as the average pairwise distance $d(\mathbf{x}_i, \mathbf{x}_j)$ between all structures of the population and normalized to $[0;1]$) as the search progresses for different κ. Initially, all populations have similar energies and diversities from the random initialization. After a few EA attempts, the populations start to diverge into different directions. After 10 EA attempts (marked by black dots in the figure), the average population energies for the lowest κ values are similar, while the high κ populations are already favoring uncertain, high-energy structures. The population diversity increases with κ as expected. After 500 attempts, the diverging effects are much more pronounced with populations at extreme κ, including low values, having considerably higher average energies. We interpret this as follows: Global searches with high κ populations spend little time deep down in energy funnels because the density of similar structures is higher here.

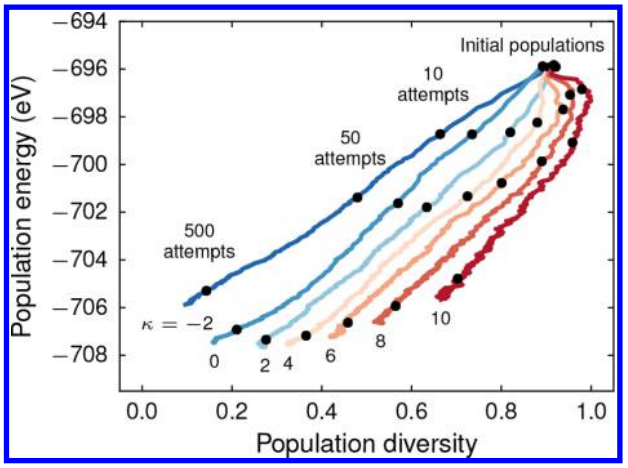


Figure 3. Average energy of the population structures versus the population diversity evolved over time (EA attempts) for various κ values. Black dots mark the states after 0 (initial populations), 10, 50, and 500 attempts.

Consequently, the populations contain either single structures from disparate funnels or structures from high-energy regions of the energy surface. Searches with low κ populations tend to spend too much time exploiting structures in the first funnels that are found, which prevents them from moving on to funnels of lower energy. This is supported by the fact that low κ populations stagnate much faster than high κ populations (Figure 4), where stagnation is here defined as the fraction of

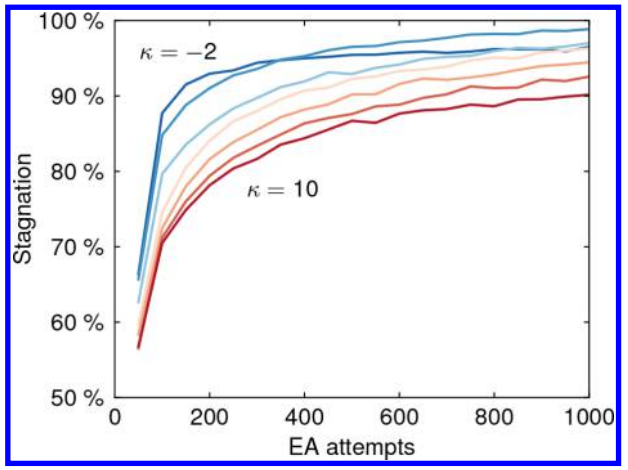


Figure 4. Stagnation of EA populations for various κ values spanning −2 (blue) to 10 (red) uniformly. The stagnation has been averaged over the past 50 attempts.

the population that remains unchanged for N_{stag} EA attempts (we use $N_{stag} = N_{pop}$). After 1000 attempts, the most successful populations ($\kappa = 4$) have the lowest average energies while still being more diverse than low κ populations.

Three other systems have been examined in this work (Figure 5). The success rate for all of them can be improved significantly for $\kappa > 0$ with maxima at around $\kappa = 4$. Another way of increasing the diversity of a population may be to increase the population size N_{pop}. However, if a search is stagnated in a local minimum, increasing N_{pop} will only include more structures from that same minimum. For the three molecular systems studied in this work (compositions $C_{10}H_6O_2$, C_9H_7N, and $C_{12}H_9N$), increasing N_{pop} has no or

DOI: 10.1021/acs.jpca.8b00160
J. Phys. Chem. A 2018, 122, 1504−1509

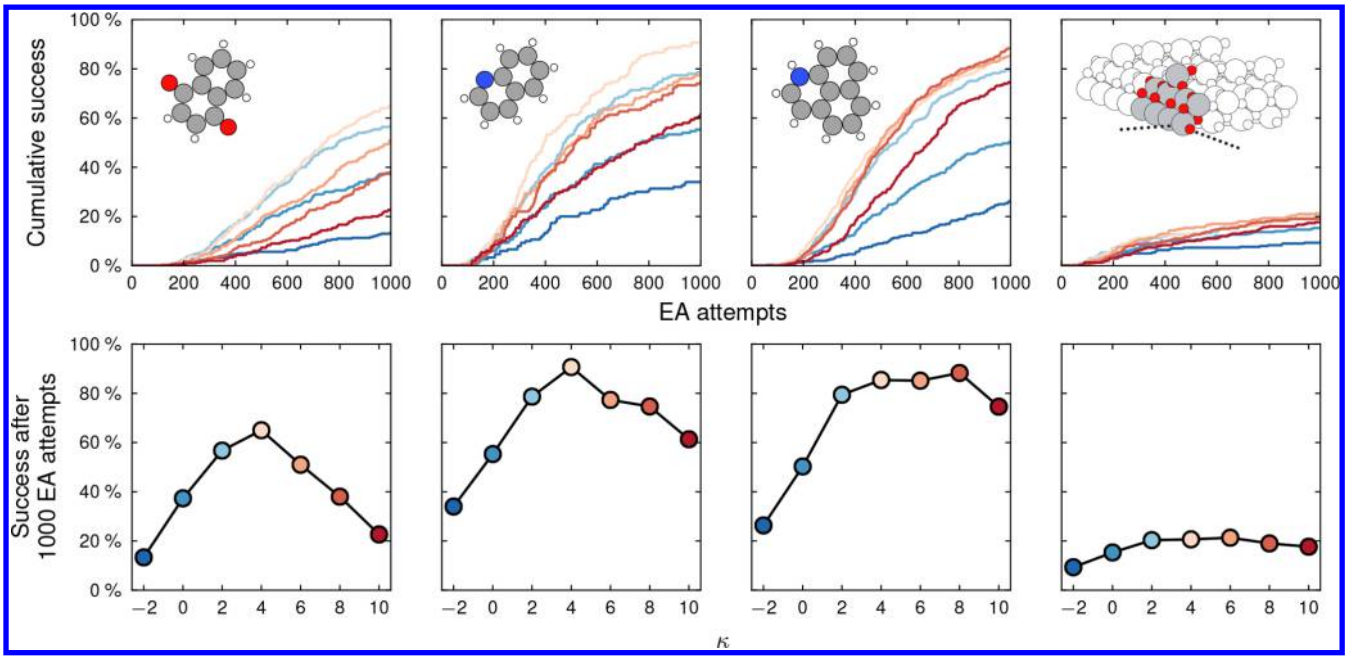


Figure 5. Top: EA performances for four different systems for various κ values. Bottom: Percentage of successful runs after 1000 EA attempts as a function of the κ values.

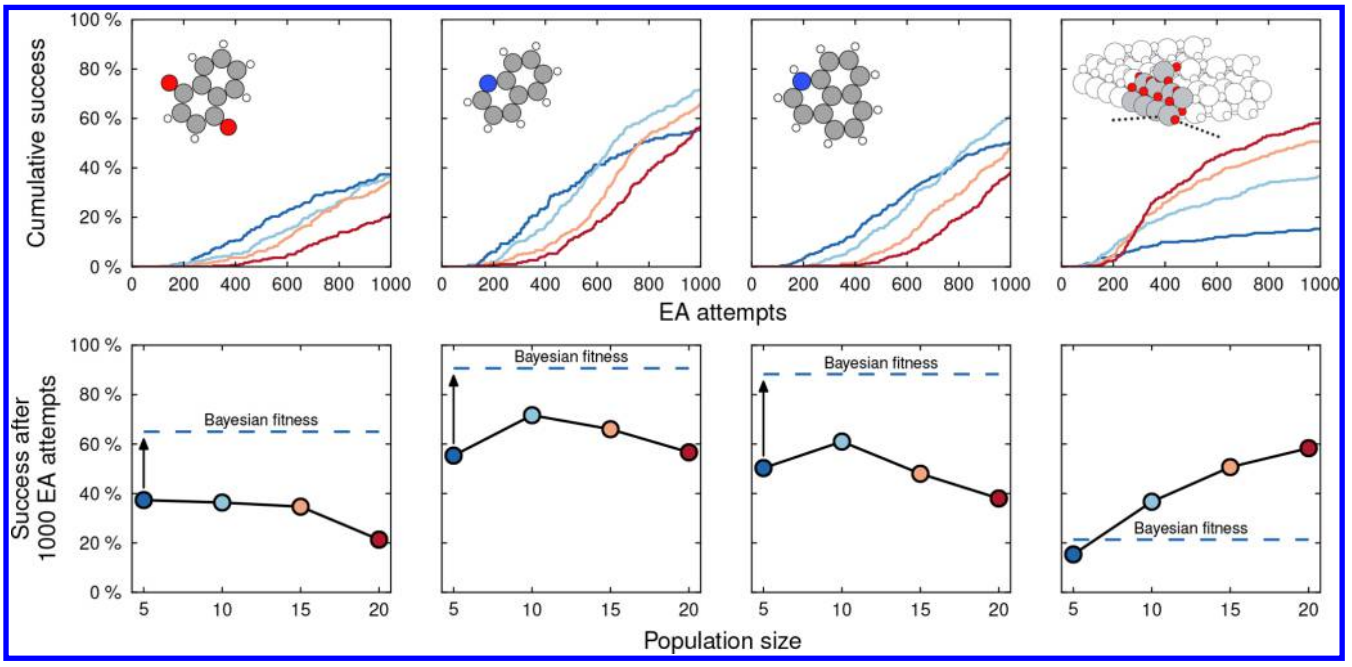


Figure 6. Top: EA performances for four different systems for various population sizes N_{pop}. Bottom: Percentage of successful runs after 1000 EA attempts as a function of the population sizes N_{pop}. The dashed line marks the maximum success of N_{pop} = 5 with a Bayesian fitness function.

little effect compared to increasing κ (Figure 6). The situation is different for a fourth problem, the search for the anatase $TiO_2(001)$-(1×4) surface reconstruction. While a success rate maximum is still found at around $\kappa = 4$, increasing N_{pop} is now the better strategy with a constant increase in success from N_{pop} = 5 to 20 (Figure 6). Moreover, when trying to apply the Bayesian fitness function to the search with $N_{pop} = 20$, we find that the success rate is unchanged for κ in the range of 0−4. This suggests that searching a configuration space of TiO_2 surface structures is a different type of global optimization problem than searching the configuration space of organic compounds. To examine this further, we define an average distance of the population structures to the global minimum

$$\overline{d_{\text{norm}}^{\text{GM}}} = \frac{1}{N_{\text{pop}}} \sum_i d_{\text{norm}}(\mathbf{x}_i, \mathbf{x}_{\text{GM}}) \tag{5}$$

where i runs over population members and d_{norm} is a normalized distance to allow for comparison between different systems[29]

$$d_{\text{norm}}(\mathbf{x}_i, \mathbf{x}_j) = \frac{d(\mathbf{x}_i, \mathbf{x}_j)}{\frac{1}{2}\sum_k [x_i(k) + x_j(k)]} \tag{6}$$

DOI: 10.1021/acs.jpca.8b00160
J. Phys. Chem. A 2018, 122, 1504−1509

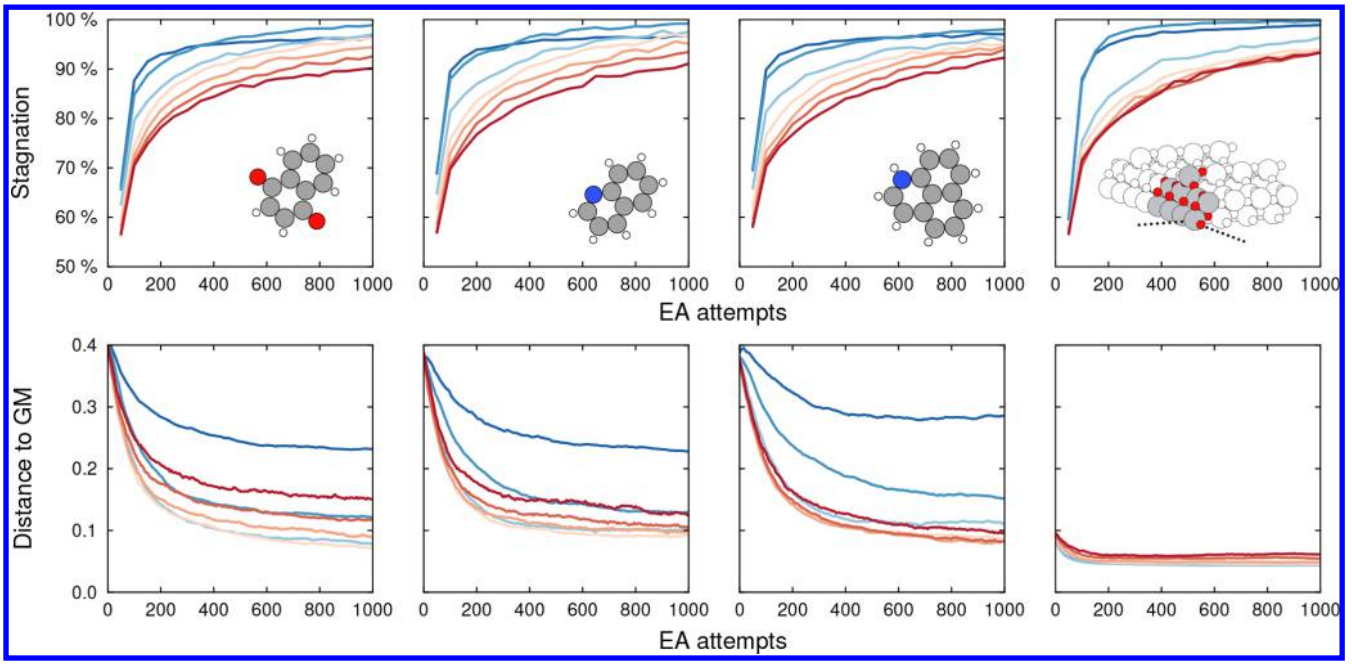


Figure 7. Top: Stagnation of EA populations for various κ values spanning −2 (blue) to 10 (red) uniformly. The stagnation has been averaged over the past 50 attempts. Bottom: The average distance $\overline{d_{\mathrm{norm}}^{\mathrm{GM}}}$ from structures in the population to the global minimum structure.

where k indexes each entry in the feature vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. Plotting $\overline{d_{\mathrm{norm}}^{\mathrm{GM}}}$ as a function of EA attempts (bottom of Figure 7) shows that the TiO_2 searches much more quickly find a funnel with structures similar to the global minimum structure. Inspection of the structures in the early parts of the searches reveals that the high degree of similarity originates from the surface layer of TiO_2 units having been directed toward global minimum positions by the template layer, and many attempts are spent afterward trying to find the protruding TiO_2 ridge. Thus, intelligent exploitation of these structures is required (e.g., via appropriate mutation operators or by exploiting more parent structures from that funnel) rather than a more exploratory search that would tend to disrupt the TiO_2 units settled in surface layer positions.

CONCLUSIONS

In summary, we have quantified the global optimization concepts of *exploration* and *exploitation* within a Bayesian framework. By combining an evolutionary algorithm with ideas from Bayesian optimization, our global minimum search can ensure a favorable balance between exploration of disparate regions in configuration space and exploitation of low-energy structures. Balancing exploration and exploitation makes the global structure minima search for one class of problems (the search for molecular compounds) significantly faster compared to the non-Bayesian search, while the speed-up for another class of problems (the search for a crystalline surface reconstruction) is less significant. This points to a need for assessing the benefit of controlling the exploration−exploitation balance for other classes of global structure optimization problems. In addition, with the introduction of quantitative measures for population diversity and population stagnation, we have been able to track and interpret search behaviors for different exploration−exploitation balances.

AUTHOR INFORMATION

Corresponding Author

*E-mail: hammer@phys.au.dk.

ORCID

Mathias S. Jørgensen: 0000-0001-7248-3185

Bjørk Hammer: 0000-0002-7849-6347

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge support from the Danish Council for Independent Research Natural Sciences (Grant No. 0602-02566B) and from VILLUM FONDEN (Investigator Grant Project No. 16562 and Project No. 9445).

REFERENCES

(1) Merte, L. R.; Jørgensen, M. S.; Pussi, K.; Gustafson, J.; Shipilin, M.; Schaefer, A.; Zhang, C.; Rawle, J.; Nicklin, C.; Thornton, G.; et al. Structure of the $SnO_2(110)$-(4×1) Surface. *Phys. Rev. Lett.* **2017**, *119*, 096102.

(2) Obersteiner, V.; Scherbela, M.; Hörmann, L.; Wegner, D.; Hofmann, O. T. Structure Prediction for Surface-Induced Phases of Organic Monolayers: Overcoming the Combinatorial Bottleneck. *Nano Lett.* **2017**, *17*, 4453−4460.

(3) Viñes, F.; Lamiel-Garcia, O.; Illas, F.; Bromley, S. T. Size Dependent Structural and Polymorphic Transitions in ZnO: From Nanocluster to Bulk. *Nanoscale* **2017**, *9*, 10067−10074.

(4) Goedecker, S. Minima hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *J. Chem. Phys.* **2004**, *120*, 9911.

(5) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(6) Ozolins, V.; Majzoub, E. H.; Wolverton, C. First-Principles Prediction of a Ground State Crystal Structure of Magnesium Borohydride. *Phys. Rev. Lett.* **2008**, *100*, 135501.

(7) Call, S. T.; Zubarev, D. Y.; Boldyrev, A. I. Global Minimum Struture Searches via Particle Swarm Optimization. *J. Comput. Chem.* **2007**, *28*, 1177−1186.

DOI: 10.1021/acs.jpca.8b00160
J. Phys. Chem. A 2018, 122, 1504−1509

(8) Wales, D. J. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(9) Hartke, B. Global Geometry Optimization of Clusters Using a Genetic Algorithm. *J. Phys. Chem.* **1993**, *97*, 9973−9976.
(10) Deaven, D. M.; Ho, K. M. Molecular Geometry Optimization with a Genetic Algorithm. *Phys. Rev. Lett.* **1995**, *75*, 288−291.
(11) Johnston, R. L. Evolving Better Nanoparticles: Genetic Algorithms for Optimising Cluster Geometries. *Dalton Trans.* **2003**, 4193−4207.
(12) Alexandrova, A. N.; Boldyrev, A. I. Search for the $Li_n^{0/+1/-1}$ (n = 5 − 7) Lowest-Energy Stuctures Using the Ab Initio Gradient Embedded Genetic Algorithm (GEGA). Elucidation of the Chemical Bonding in the Lithium Clusters. *J. Chem. Theory Comput.* **2005**, *1*, 566−580.
(13) Oganov, A. R.; Glass, C. W. Crystal Structure Prediction Using Ab Initio Evolutionary Techniques: Principles and Applications. *J. Chem. Phys.* **2006**, *124*, 244704.
(14) Marques, J. M. C.; Pereira, F. B. An Evolutionary Algorithm for Global Minimum Search of Binary Atomic Clusters. *Chem. Phys. Lett.* **2010**, *485*, 211−216.
(15) Bhattacharya, S.; Levchenko, S. V.; Ghiringhelli, L. M.; Scheffler, M. Stability and Metastability of Clusters in a Reactive Atmosphere: Theoretical Evidence for Unexpected Stoichiometries of Mg_MO_x. *Phys. Rev. Lett.* **2013**, *111*, 135501.
(16) Revard, B. C.; Tipton, W. W.; Yesypenko, A.; Hennig, R. G. Grand-Canonnical Evolutionary Algorithm for the Prediction of Two-Dimensional Materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 054117.
(17) Jørgensen, M. S.; Groves, M. N.; Hammer, B. Combining Evolutionary Algorithms with Clustering toward Rational Global Structure Optimization at the Atomic Scale. *J. Chem. Theory Comput.* **2017**, *13*, 1486−1493.
(18) Snoek, J.; Larochelle, H.; Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. *NIPS* **2012**, *25*, 2960−2968.
(19) Toyoura, K.; Hirano, D.; Seko, A.; Shiga, M.; Kuwabara, A.; Karasuyama, M.; Shitara, K.; Takeuchi, I. Machine-Learning-Based Selective Sampling Procedure for Identifying the Low-Energy Region in a Potential Energy Surface: A Case Study on Proton Conduction in Oxides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 054112.
(20) Carr, S. F.; Garnett, R.; Lo, C. S. Accelerating the Search for Global Minima on Potential Energy Surfaces Using Machine Learning. *J. Chem. Phys.* **2016**, *145*, 154106.
(21) Todorović, M.; Gutmann, M. U.; Corander, J.; Rinke, P. Efficient Bayesian Inference of Atomistic Structure in Complex Functional Materials. *arXiv:1708.09274*, 2017.
(22) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; MIT Press, 2006.
(23) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer, 2006.
(24) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
(25) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404−3419.
(26) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326−2331.
(27) Lazzeri, M.; Selloni, A. Stress-Driven Reconstruction of an Oxide Surface: The Anatase TiO_2(001)-(1 × 4) Surface. *Phys. Rev. Lett.* **2001**, *87*, 266105.
(28) Vilhelmsen, L. B.; Hammer, B. Systematic Study of Au_{12} Gold Clusters on MgO(100) F Centers using Density-Functional Theory. *Phys. Rev. Lett.* **2012**, *108*, 126101.
(29) Vilhelmsen, L. B.; Hammer, B. A Genetic Algorithm for First Principles Global Structure Optimization of Supported Nano Structures. *J. Chem. Phys.* **2014**, *141*, 044711.
(30) Bahn, S. R.; Jacobsen, K. W. An Object-Oriented Scripting Interface to a Legacy Electronic Structure Code. *Comput. Sci. Eng.* **2002**, *4*, 56−66.
(31) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dulak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. The Atomic Simulation Environment-A Python Library for Working with Atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.
(32) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678−5784.
(33) Dolgonos, G.; Aradi, B.; Moreira, N. H.; Frauenheim, T. An Improved Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) Set of Parameters for Simulation of Bulk and Molecular Systems Involving Titanium. *J. Chem. Theory Comput.* **2010**, *6*, 266−278.
(34) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.
(35) Pickard, C. J.; Needs, R. J. Ab Initio Random Structure Searching. *J. Phys.: Condens. Matter* **2011**, *23*, 053201.

DOI: 10.1021/acs.jpca.8b00160
J. Phys. Chem. A 2018, 122, 1504−1509