Efficient Global Structure Optimization with a Machine-Learned Surrogate Model

Malthe K. Bisbo and Bjørk Hammer[®]

Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

(Received 18 June 2019; revised manuscript received 20 September 2019; accepted 23 January 2020; published 27 February 2020)

We propose a scheme for global optimization with first-principles energy expressions of atomistic structure. While unfolding its search, the method actively learns a surrogate model of the potential energy landscape on which it performs a number of local relaxations (exploitation) and further structural searches (exploration). Assuming Gaussian processes, deploying two separate kernel widths to better capture rough features of the energy landscape while retaining a good resolution of local minima, an acquisition function is used to decide on which of the resulting structures is the more promising and should be treated at the first-principles level. The method is demonstrated to outperform by 2 orders of magnitude a well established first-principles based evolutionary algorithm in finding surface reconstructions. Finally, global optimization with first-principles energy expressions is utilized to identify initial stages of the edge oxidation and oxygen intercalation of graphene sheets on the Ir(111) surface.

DOI: 10.1103/PhysRevLett.124.086102

In materials science and physical chemistry, the search for optimal structure is a recurring task, e.g., in describing crystalline defects, such as grain boundaries [1] and surface reconstructions [2,3], and in modeling heterogeneous systems such as binary compounds [4,5] and supported nanoparticles [6-8]. Depending on the search strategy and the complexity of a given problem, many thousands of energy and force evaluations may be required for the structural candidates in the course of the search. The results of these calculations constitute a set of structureenergy relation data points which represents a valuable resource that can direct the search. If the energy calculations are done at a first-principles (FP) level using density functional theory or quantum chemical methods, the computational bottleneck lies in performing the individual energy-force evaluations and considerable speed-ups may be achieved by introducing machine-learning (ML) techniques that utilize this resource and provide tools to minimize the total amount of FP calculations.

As a first approach, the ML model can be trained in advance of the search on a database of structures and FP properties. The entire search, including local relaxations, can then be performed on this machine-learned surrogate energy landscape. Here, regression models based on kernels, invariant polynomials, and neural networks have all proven successful in a number of studies [9–12] and are drastically changing the field of fitting force fields [13–17]. Since the models are all interpolative and give reliable results only within their training domain, prefitted models have the drawback that they require the expensive generation of a large, diverse database of training data to be successfully applied to a structure search problem.

A more data efficient approach is to start from a small incomplete training database and then augment it on the fly only with the data deemed most relevant [18-26]. This is the philosophy in the area of active learning [27,28]. It was recently demonstrated in the context of an evolutionary algorithm (EA) structure search framework, where an artificial neural network was trained and used for local relaxation while the EA acted only on FP single-point energy evaluations [8]. Active learning approaches have been extensively applied in molecular dynamics simulations [29-33] with data efficient training databases as a byproduct. It has also been applied in local optimization problems such as local relaxation [34,35] and in minimum energy path determination with the nudged elastic band method [36-38]. Local optimization problems lend themselves particularly well to the construction of surrogate energy landscapes as a Cartesian coordinate representation of atoms may be adopted.

When a surrogate energy landscape is trained via active learning, the issue arises which next computationally expensive FP single-point energy to evaluate. In this work, we present a strategy for global optimization with first-principles energy expressions (GOFEE) that utilizes Bayesian statistics in the context of Gaussian processes (GP) [39]. This framework allows for the estimation of the uncertainty in any prediction on the surrogate energy landscape, which provides the foundation for an acquisition function that guides the search. The virtues of the proposed method are the following: (1) It uses the Bayesian surrogate model to perform cheap relaxations of new candidate structures. (2) In each search iteration multiple new candidates are surrogate relaxed at low computational cost. The model is used to select which of these to evaluate at the expensive FP level. (3) It utilizes two separate kernel widths to improve the surrogate model. A number of examples of considerable speed-ups, obtained using the method, are given for inorganic surface structures. The degree of success achievable by any surrogate based search method is largely dependent on the quality of the surrogate model [22,23]. In this work, the GP regression method is adopted partly because of its tractable simplicity and partly because GPs are expected to behave well, as the number of training examples increases during the search, due to the adaptive predictive power inherent to nonparametric methods.

The task of GPs is to infer a distribution over functions, here $p(E_{sur}|X, \mathbf{E})$, that is consistent with a training set of observed atomic configurations $X = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)^T$ and their corresponding energies $\mathbf{E} = (E_1, E_2, \dots, E_N)^T$. To include the symmetries of the system, \mathbf{x}_i is taken to be the feature vector for the *i*th configuration rather than the Cartesian coordinates. We adopt the global fingerprint feature from Oganov and Valle [40], however, the method is expected to work equally well with other features. For GPs, the distribution is assumed to be normal, which enables the estimation of not only the energy $E_{sur}(\mathbf{x})$ as the mean of the distribution, but also the predictive uncertainty $\sigma_{sur}(\mathbf{x})$. As will be discussed later, the predictive uncertainty is useful in a search context, as it allows the distinction between explored and unexplored regions of the search space.

A Gaussian process is specified by its prior mean function $\mu(\mathbf{x})$ and covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$, which encodes prior assumptions about the target function. Given these, energy and uncertainty predictions for a new structure \mathbf{x}_* are carried out using [39]

$$E_{\text{sur}}(\mathbf{x}_*) = \mathbf{k}_*^T (K + \sigma_n^2 I)^{-1} [\mathbf{E} - \mu(\mathbf{x})] + \mu(\mathbf{x}), \quad (1)$$

$$\sigma_{\rm sur}(\mathbf{x}_*)^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (K + \sigma_n^2 I)^{-1} \mathbf{k}_*, \qquad (2)$$

where K = k(X, X) and $\mathbf{k}_* = k(X, \mathbf{x}_*)$ and the target function is assumed noisy with uncertainty $\sigma_n^2 = 10^{-5} \text{ eV}^2$, which acts as regularization. To include the repulsive atomic core generally present, the prior mean function is taken to be a conservatively chosen repulsive interatomic potential, specifically $\mu(\mathbf{x}) \propto \sum_{ij} (0.7r_{\text{CD},ij}/r_{ij})^{12}$, where r_{ij} and $r_{\text{CD},ij}$ are the distance and covalent distance between atom *i* and *j*. This is especially beneficial in a structure search context, where the fine details of the repulsive part of the potential are not crucial, unlike the near equilibrium part of the potential. The covariance function was chosen to be a sum of two Gaussian covariances

$$k(\mathbf{x}, \mathbf{x}') = \theta_0[(1 - \beta)e^{-(\mathbf{x} - \mathbf{x}')^2/(2\lambda_1^2)} + \beta e^{-(\mathbf{x} - \mathbf{x}')^2/(2\lambda_2^2)}], \quad (3)$$

with characteristic length scale λ_1 and λ_2 , respectively, maximal covariance θ_0 , and weights given by $\beta = 0.01$.

Figure 1 shows, for a simple problem, how the surrogate energy landscape improves as more data are added to the training set. The system considered is naphthalene, constrained, for illustrative purposes, to change only according



FIG. 1. Example of a surrogate guided structure search in a two dimensional search space. The two artificial dimensions are constructed by perturbing naphthalene as depicted in the top left. The FP energy landscape is shown in the top right. In the bottom, the surrogate landscape resulting from N = 3, 4, and 5 training structures is shown. Note that training structures 4 and 5 result from local relaxation in the surrogate landscape.

to the two coordinates specified. The resulting 2D slice of the full energy landscape contains four local minima including naphthalene itself. With only N = 3 training examples near one local minimum, the model is able to predict the locations of the remaining minima to approximately coincide with those of the true energy landscape. In the search we will take advantage of this, and conduct most of the search, specifically all local relaxations, in the surrogate energy landscape, which is orders of magnitude faster than FP calculations. As illustrated in the figure, a structure relaxed with the model can then be evaluated with a single FP calculation and used to update the model.

Relying entirely on the surrogate model to guide the search has the drawback that the data collection process, vital to actively improving the model, is itself model dependent. This interplay has a tendency to cause underexploration of the search space and in turn lead to premature stagnation of the search. The minimum belonging to naphthalene in Fig. 1 is an example that the true depth of a minimum might be underestimated until appropriate data have been collected. To remedy this problem we bias data collection towards unexplored regions of the search space, using the predictive uncertainty $\sigma_{sur}(\mathbf{x})$ as a natural way to quantify this. In practice, this is done by introducing an acquisition function f(x) to be minimized, which relies on both the predicted energy and uncertainty. This function is used to determine which, out of multiple new candidates, generated in each search iteration, to select

for FP evaluation. There exist multiple choices for such an acquisition function [41] including the lower confidence bound

$$f(\mathbf{x}) = E_{\text{sur}}(\mathbf{x}) - \kappa \cdot \sigma_{\text{sur}}(\mathbf{x}), \qquad (4)$$

used in this work due to its simplicity. Here κ is a unitless, tunable parameter determining the emphasis on the predicted uncertainty and thus the degree of exploration in the search.

The surrogate model is central to the GOFEE search method sketched in Fig. 2. The main ideas are the generation of N_c new candidates, derived from the population, in each search iteration and subsequent relaxation of all these using the surrogate model. Out of all surrogaterelaxed structures only the one with the lowest value of the acquisition function is selected for two FP single-point evaluations. Relaxing multiple new candidates in each search iteration takes full advantage of the inexpensiveness of the surrogate model, and can be well parallelized. When updating the surrogate model, the parameters λ_1 , λ_2 , and θ_0 , are automatically determined using multirestart gradient based optimization of the marginal likelihood [39] of the model given the data, a quantity inherent to GP models. The bounds used during hyperparameter optimization are $1 \le \lambda_1, \lambda_2 \le 10^3$, and $1 \le \theta_0 \le 10^6$. As seen from Fig. 3(a) two separate kernel widths are clearly favored. This enhances the models capacity to capture both the long range features of the energy such as large energy funnels, as well as allowing for sufficient resolution of local minima.

As a first example using GOFEE we considered the $SnO_2(110) - (4 \times 1)$ surface, for which the global minimum (GM) structure, see Fig. 3(b), is known [3]. Figure 3(c) shows the cumulative success curves for finding the GM with this method as well as with the well established EA [42] originally used to find the structure [3]. Noting the broken axis, the figure shows a 2 orders of magnitude decrease in the number of FP calculations required to reach, e.g., 80% success. This is largely attributed to the fact that this method relies only on FP for single-point calculations. In Fig. 3(d) the effect of the exploration promoting parameter κ is visualized. The figure shows the feature space coordinates, projected into two dimensions using principal component analysis, of structures visited in three independent GOFEE restarts deploying $\kappa = 1, 2, 4$, respectively. A more detailed figure is shown in Fig. S1 along with the evolution of the prediction error in Fig. S2. We have used $\kappa = 2$ for all applications. GOFEE was further applied to a number of diverse problems including the shape of C_{24} clusters [43], the surface reconstructions of Si(111)- (7×7) [44] and anatase TiO₂(001)-(1 \times 4) [2], and a high pressure bulk phase of CaH_6 [45]. The results for these systems are presented in the Supplemental Material [46], Figs. S3–S9.



FIG. 2. Key elements of GOFEE. (i) Creation and FP evaluation of random initial structures. (ii) Addition of new FP data to training database and training of surrogate model. (iii) Generation of multiple new candidate structures by applying mutation operations to a diversified population of the best structures, currently found in the search. (iv) Local relaxation of the population and all new candidates, N_c in total, using the surrogate model. (v) Selection of the most promising relaxed candidate using the acquisition function. (vi) Single-point FP evaluation of the chosen structure and, to benefit from the rich information content of the FP forces, another on the same structure perturbed slightly along the forces. This is the bottleneck in the search, $t_{\rm FP} > t_{\rm ML}$. Finally, the search is carried out by repeating steps (ii)–(vi). Steps (iii)–(iv) are easily parallelized by handling each new candidate on separate computer cores.

To demonstrate the versatility of the GOFEE method we proceed to address the hitherto prohibitively complex problems of determining the edge structure of graphene patches on Ir(111) as well as that of the oxidized edge. For this application we will use GOFEE with fixed kernel widths, $l_1 = 10$ and $l_2 = 0.5$, as opposed to optimized, as this is more efficient for larger systems as seen for the TiO₂(001)-(1 × 4) example in Fig. S4. For fixed kernel



FIG. 3. The SnO₂(110) – (4 × 1) test system. (a) Crosses: The optimized kernel widths after 200 visited structures of 20 GOFEE restarts. Contour: the log marginal likelihood [39], for fixed λ_1 and λ_2 and optimal θ_0 , averaged over the same 20 datasets. (b) GM of the system. (c) Success curves for finding the GM when using GOFEE with fixed (gray) and optimized (blue) hyperparameters compared to an EA. (d) 2D principal component analysis visualization of structures visited in three different GOFEE restarts with $\kappa = 1, 2, 4$, respectively. The two principal components were found from a dataset of ≈4000 structures from 10 GOFEE restarts, on which they explained 77% of the variance.

widths, θ_0 is given by the analytic expression [39] $\theta_0^2 = \mathbf{y}^T K \mathbf{y} / N$. The resulting structures are used to study the atomistic mechanisms involved in intercalation of oxygen in the system. The intercalation process has been intensively studied experimentally for Ir(111) [47–49] and involves dissociative adsorption and diffusion of oxygen as well as penetration of the graphene edge. Although experiments suggest [47] that the limiting step for the intercalation process is this edge penetration, it is not well understood.

In our contribution to fill this gap, intercalation through the nonoxidized edge was first considered. Figure 4(a) shows the most stable edge structures found when varying the number of carbon atoms present in the cell. Using the energy of carbon within the graphene patch as reference, the energies are compared in Fig. 4(b), showing that the preferred structure is not the perfect edge (A), but instead the structures with one (A-1) and three (A-3) carbon atoms less on the edge, both of which feature pentagonal rings



FIG. 4. (a),(b) The most stable structures from the search and their energies shown for the edges with 0–5 carbon atoms less than the perfect edge. (c),(d) Structures and energies for oxygen added to the perfect edge. Structures (E, G) are made by hand, motivated by the trend from structures (B - D). Finally, lowest energy CI - EB curves are given for (e) inserting a surface adsorbed oxygen atom to the perfect edge (A), (f) inserting the third oxygen atom, and (g) intercalation of an oxygen atom below the edge of structure (E). In the search, the second layer of iridium is kept fixed, while the upper layer is allowed to relax a maximum of half a covalent distance away from the bulk positions.

(see Fig. S10). This can be attributed to these structures avoiding unfavorable C-Ir bonds in the position of largest mismatch between the periodicities of the edge and surface. Despite small gaps in the graphene edge, none of these structures will allow oxygen intercalation, with energy barriers all above 2 eV

The structure of the oxidized graphene edge was also considered. Searches are performed with up to three oxygen atoms in the cell. The resulting structures and energies are depicted in Fig. 4(c) (B - D) and 4(d). They display a preference for oxidizing the edge with the oxidized region partially detaching from the surface when two or more oxygen atoms are present. This results in a significant gap in the edge, likely of accommodating intercalation. Figure 4(c) (E - G) (Fig. S11) further shows the structures resulting from extending this trend up to six oxygen atoms. As energy reference, atomic oxygen adsorbed on the iridium surface is used. Based on the energies, the size of the gaps are thermodynamically self-limiting, as edge oxidation is only thermodynamically favored up to four oxygen atoms. Further oxidation requires breaking of increasingly strong C----Ir bonds.

To study whether these oxidized edge structures are likely to form and contribute to the intercalation process under typical experimental conditions, lowest energy paths were calculated using the climbing image elastic band (CI - EB) nudged elastic band type method [50]. Figures 4(e) and 4(f), (Figs. S12 and S13), respectively, show the energy profiles for inserting the first and third oxygen to the edge revealing the third oxygen to be the more expensive of the two with an energy barrier of 1.23 eV. However, oxygen intercalation experiments typically feature large oxygen coverages on the iridium surface, resulting in weaker bonding of the adsorbed oxygen as this coverage builds up. This effectively lowers the barriers, as the transition state for binding to and opening the graphene edge is expected to remain unchanged. The effect is depicted in Fig. 4(f) showing in red how destabilizing the oxygen, adsorbed in front of the edge, reduces the calculated barrier to 0.76 eV. Figure 4(g) (Fig. S14) shows the energy profile for the intercalation of an oxygen atom through the edge gap of structure E, displaying a barrier of 0.93 eV, which, as discussed above, will also be lower at realistic oxygen coverages.

In conclusion, we have formulated a machine-learning enhanced structure search method, deploying two separate kernel widths, and used it to solve a previously prohibitively hard problem. The underlying PYTHON implementation is available in [51].

We acknowledge support from VILLUM FONDEN (Investigator grant, Project No. 16562).

- A. M. van der Zande, P. Y. Huang, D. A. Chenet, T. C. Berkelbach, Y. You, G.-H. Lee, T. Heinz, D. R. Reichman, D. A. Muller, and J. Hone, Nat. Mater. 12, 554 (2013).
- [2] M. Lazzeri and A. Selloni, Phys. Rev. Lett. 87, 266105 (2001).
- [3] L. R. Merte, M. S. Jørgensen, K. Pussi, J. Gustafson, M. Shipilin, A. Schaefer, C. Zhang, J. Rawle, C. Nicklin, G. Thornton, R. Lindsay, B. Hammer, and E. Lundgren, Phys. Rev. Lett. **119**, 096102 (2017).
- [4] E. Flikkema and S. Bromley, J. Phys. Chem. B **108**, 9638 (2004).
- [5] R. Ferrando, J. Jellinek, and R. L. Johnston, Chem. Rev. 108, 845 (2008).
- [6] H. Zhai and A. Alexandrova, J. Phys. Chem. Lett. 9, 1696 (2018).
- [7] B. Zandkarimi and A. Alexandrova, WIREs Comput. Mol. Sci. 9, e1420 (2019).
- [8] E. L. Kolsbjerg, A. A. Peterson, and B. Hammer, Phys. Rev. B 97, 195424 (2018).
- [9] J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 146401 (2007).
- [10] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
- [11] A. Shapeev, Multiscale Model. Simul. 14, 1153 (2016); https://arxiv.org/abs/1512.06054.

- [12] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. 148, 241722 (2018).
- [13] H. Zhai and A. Alexandrova, J. Chem. Theory Comput. 12, 6213 (2016).
- [14] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. 8, 3192 (2017).
- [15] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, J. Phys. Chem. C 121, 511 (2017).
- [16] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Sci. Adv. 3, e1603015 (2017).
- [17] V.L. Deringer and G. Csányi, Phys. Rev. B 95, 094203 (2017).
- [18] H. Zhai, M.-A. Ha, and A. N. Alexandrova, J. Chem. Theory Comput. 11, 2385 (2015).
- [19] M. Todorović, M. Gutmann, J. Corander, and P. Rinke, npj Comput. Mater. 5, 35 (2019).
- [20] T. Yamashita, N. Sato, H. Kino, T. Miyake, K. Tsuda, and T. Oguchi, Phys. Rev. Mater. 2, 013803 (2018).
- [21] T. L. Jacobsen, M. S. Jørgensen, and B. Hammer, Phys. Rev. Lett. **120**, 026102 (2018).
- [22] V. L. Deringer, C. J. Pickard, and G. Csányi, Phys. Rev. Lett. 120, 156001 (2018).
- [23] V. L. Deringer, D. M. Proserpio, G. Csányi, and C. J. Pickard, Faraday Discuss. 211, 45 (2018).
- [24] Q. Tong, L. Xue, J. Lv, Y. Wang, and Y. Ma, Faraday Discuss. 211, 31 (2018).
- [25] K. Gubaev, E. Podryabinkin, G. Hart, and A. Shapeev, Comput. Mater. Sci. 156, 148 (2019).
- [26] M. Van den Bossche, J. Phys. Chem. A 123, 3038 (2019).
- [27] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, J. Chem. Phys. 148, 241733 (2018).
- [28] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and E. Weinan, Phys. Rev. Mater. 3, 023804 (2019).
- [29] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. 114, 096405 (2015).
- [30] A. A. Peterson, R. Christensen, and A. Khorshidi, Phys. Chem. Chem. Phys. 19, 10978 (2017).
- [31] K. Miwa and H. Ohno, Phys. Rev. Mater. 1, 053801 (2017).
- [32] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. 140, 171 (2017).
- [33] R. Jinnouchi, F. Karsai, and G. Kresse, Phys. Rev. B 100, 014105 (2019).
- [34] E. G. del Río, J. J. Mortensen, and K. W. Jacobsen, Phys. Rev. B 100, 104103 (2019).
- [35] A. Denzel and J. Kästner, J. Chem. Phys. 148, 094114 (2018).
- [36] A. A. Peterson, J. Chem. Phys. 145, 074106 (2016).
- [37] O.-P. Koistinen, F. B. Dagbjartsdóttir, V. Ásgeirsson, A. Vehtari, and H. Jónsson, J. Chem. Phys. 147, 152720 (2017).
- [38] J. A. G. Torres, P. C. Jennings, M. H. Hansen, J. R. Boes, and T. Bligaard, Phys. Rev. Lett. 122, 156001 (2019).
- [39] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (Massachusetts Institute of Technology, Cambridge, 2005).
- [40] M. Valle and A. Oganov, Acta Crystallogr. Sect. A 66, 507 (2010).
- [41] H. Wang, B. van Stein, M. Emmerich, and T. Back, in Proceedings of the 2017 IEEE International Conference

on Systems, Man, and Cybernetics (SMC) (IEEE, 2017), pp. 507–512.

- [42] L. B. Vilhelmsen and B. Hammer, J. Chem. Phys. 141, 044711 (2014).
- [43] W. An, N. Shao, S. Bulusu, and X. C. Zeng, J. Chem. Phys. 128, 084301 (2008).
- [44] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, Phys. Rev. Lett. 50, 120 (1983).
- [45] H. Wang, J. S. Tse, K. Tanaka, T. Iitaka, and Y. Ma, Proc. Natl. Acad. Sci. U.S.A. 109, 6463 (2012).
- [46] See Supplemental Material at http://link.aps.org/supplemental/ 10.1103/PhysRevLett.124.086102 for extended figures for the principal component analysis and the graphene edge application. In addition, further application examples are

supplied, including carbon clusters, silicon surface reconstructions and high pressure bulk calcium hydride.

- [47] E. Grånäs, J. Knudsen, U. A. Schröder, T. Gerber, C. Busse, M. A. Arman, K. Schulte, J. N. Andersen, and T. Michely, ACS Nano 6, 9951 (2012).
- [48] A. J. Martínez-Galera, U. A. Schröder, F. Huttmann, W. Jolie, F. Craes, C. Busse, V. Caciuc, N. Atodiresei, S. Blügel, and T. Michely, Nanoscale 8, 1932 (2016).
- [49] R. Larciprete, S. Ulstrup, P. Lacovig, M. Dalmiglio, M. Bianchi, F. Mazzola, L. Hornekær, F. Orlando, A. Baraldi, P. Hofmann, and S. Lizzit, ACS Nano 6, 9551 (2012).
- [50] E. L. Kolsbjerg, M. N. Groves, and B. Hammer, J. Chem. Phys. 145, 094107 (2016).
- [51] https://gofee.au.dk/.