Global optimization of atomic structure enhanced by machine learning

Malthe K. Bisbo and Bjørk Hammer ⬤

Center for Interstellar Catalysis, Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

Global optimization with first-principles energy expressions (GOFEE) is an efficient method for identifying low-energy structures in computationally expensive energy landscapes such as the ones described by density functional theory (DFT), van der Waals enabled DFT, or even methods beyond DFT. GOFEE is an evolutionary algorithm, that in order to explore configuration space creates several candidates in parallel. These are treated approximately using a machine learned surrogate model of energies and forces, trained on the fly, eliminating the need for expensive relaxations using first-principles methods. Eventually, using Bayesian statistics, GOFEE chooses one candidate and treats that at the full first-principles level. In this paper we elaborate on the importance of the use of a Gaussian kernel with two length scales in the Gaussian process regression surrogate model. We further explore the role of the use in GOFEE of the lower confidence bound for relaxation and selection of candidate structures. In addition, we present details of a sampling scheme for obtaining parent structures in the evolution. Using machine learning clustering of the entire pool of low-energy structures ever calculated, and choosing the most stable member from each cluster, the scheme ensures a highly diverse sample of structures that plays the role of a population. The versatility of the GOFEE method is demonstrated by applying it to identify the low-energy structures of gas-phase fullerene-type 24-atom carbon clusters and of dome-shaped 18-atom carbon clusters supported on Ir(111).

I. INTRODUCTION

The atomic scale understanding of material properties is a fundamental goal of modern computational chemistry and materials science. As material properties to a large extent are governed by the lowest-energy atomic structure, the efficient determination of such optimal structure is an important problem. The problem is however a difficult one, due to the vastness of the configurational space of even small sized systems, a result of the exponential scaling of the number of metastable structures with the number of atoms in a system [1,2]. For simple systems, the use of domain knowledge is often enough to identify the correct structure. The literature however also contains many examples, where this approach fails [3], and for the vast majority of systems, that are more complex and less studied, one has to thoroughly explore the configurational space in order to identify the optimal structure. This is commonly achieved using automated and unbiased search strategies such as random search [1], basin and minima hopping [4,5], particle swarm optimization [6,7], evolutionary algorithms [2,8–10], etc., which have been successfully applied to a wide range of different systems including surface reconstructions [11–14], surface steps [15] grain boundaries [16,17], binary compounds [18,19], isolated [20–22] and supported [3,23–25] nanoparticles, solids [26], etc.

Most often, atomic-scale materials science projects rely on computationally expensive first-principles methods such as density functional theory (DFT), van der Waals enabled DFT, or even more advanced quantum chemical methods.

When using conventional global optimization methods in conjunction with such electronic structure and quantum chemical methods, their performance becomes limited due to the computational cost of the many energy and force evaluations required to sufficiently explore the configurational space. As a potential solution, solving the global optimization problem in machine learned model potentials has proven a capable and less expensive alternative to solving it with the full energy expression, which we for the sake of simplicity within this paper will refer to as DFT or "target potential." Various machine learning methods have been successfully used to create such model potentials, from training databases. These include kernel based methods [27,28] combined with robust representations of atomic structures [29–31], deep neural networks [29,32], and methods based on body-ordered energy decompositions [33,34].

Despite cheap evaluations once the model is trained, the required quantum mechanical calculated databases represent a considerable computational expense. As an example, for molecular dynamics to be successfully carried out with a machine learned potential, the potential must, depending on the temperature adopted, accurately describe all configurations below a certain energy, which requires a both broad and thorough database. Improvements in data efficiency of such databases have recently been achieved by generating the data using active learning [35–42], where starting from a small, incomplete database the model itself, in combination with an acquisition strategy, actively collects all further data, with new data iteratively improving the model throughout the collection process. Active learning can also be applied in a problem

245404-1

specific context, where further savings in training data are possible, because accurate predictions are especially relevant only for subsets of the full configurational space. This includes configurational problems that are local in nature, such as structure relaxation [43,44] and transition state determination [45–47] as well as more ambitious tasks such as molecular dynamics [48–53], chemical reaction networks [54,55], free energy calculation [56], and finally global structure search [3,25,57–66], including our recently proposed global optimization with first-principles energy expressions (GOFEE) structure search method [67]. For global structure search an active learning approach can utilize the fact that accuracy is increasingly important for lower-energy structures, such that higher-energy structures can be screened based on only rough energy predictions. In this paper we further detail the key elements of GOFEE, which relies heavily on exploring the configurational space in an actively learned Gaussian process regression (GPR) surrogate model. This includes elaborating on the effect of using a Gaussian kernel with two separate length scales in the surrogate model as well as the inclusion of a repulsive interatomic potential as a prior for this surrogate model. In addition, we explore the role of the lower confidence bound, of the surrogate model, in relaxing and selecting among new candidate structures in the search. Finally, we discuss a recently proposed improvement to the method [68] in which the entire pool of low-energy structures is sampled using a clustering technique. The resulting highly diverse sample plays the role, normally served by the population, of providing parent structures for the evolutionary process.

The paper is outlined as follows. First a detailed description of the GOFEE search method is given. Throughout, the effects of key elements of the method are illustrated on searches for the reconstructed surface of $TiO_2(001)$-(1×4). Finally the model is applied to isolated and Ir(111) supported carbon clusters, and we report new low-energy structures for the supported system.

II. METHOD

The GOFEE search method combines the evolutionary search strategy with a computationally inexpensive, actively learned surrogate model of the energy landscape, which can be used to carry out significantly more structure queries than would be possible with the target potential. A much smaller number of evaluations is however carried out using the target potential on the structures deemed most promising by the surrogate model. These are in turn used as training data to further improve the surrogate model. A flowchart of the GOFEE search scheme is shown in Fig. 1. The key steps can be summarized as follows.

(i) Generate and evaluate a number of random structures used to serve as initial parent structures and used for initial training of the surrogate model.

(ii) Train the surrogate model based on all DFT evaluated structures accumulated up until this point in the search.

(iii) Handle parent structures by updating a population or extracting a sample from all DFT evaluated structures.

(iv) Generate new candidates by applying stochastic rattle and permutation operations on some parent structures.
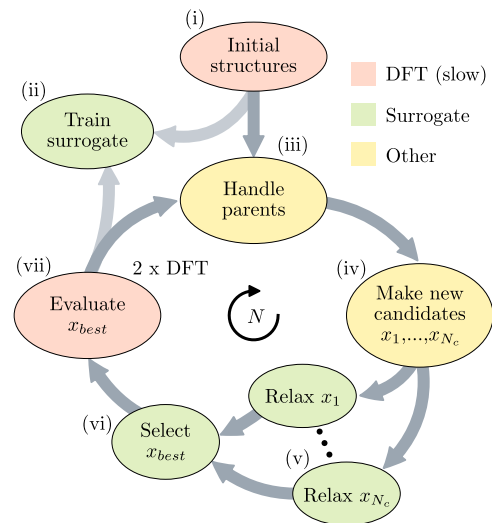

FIG. 1. Dataflow in GOFEE. (i) Random initial structures are generated, evaluated with DFT, and added to a structural database. (ii) The surrogate model is trained based on the database. (iii) Handle parent structures. (iv) A number, N_c, of new candidate structures are generated by applying stochastic changes to some parent structures. (v) All N_c candidates are relaxed in the acquisition function. (vi) The single most promising candidate according to the acquisition function is selected. (vii) A single evaluation, using the target energy expression, is performed for this structure, along with a second evaluation for the structure perturbed slightly along the force. The search is carried out by adding these two newly evaluated structures to the training database and repeating steps (ii)–(vii) N times in total.

(v) Relax all new candidates by local optimization in the acquisition function based on the surrogate model potential.

(vi) Select "most promising" candidate structure from (v) according to the acquisition function.

(vii) Evaluate the selected structure using the DFT potential.

The generation of initial structures in the first step (i) is done by randomly placing atoms within a predefined box and requiring that no bond lengths are shorter than $0.7d_{\text{cov},ij}$, with $d_{\text{cov},ij}$ being the sum of the covalent radii of the involved atoms i and j. In addition each atom is required to have its nearest neighbor within $1.4d_{\text{cov},ij}$, to avoid isolated atoms. After this step, the search is carried out by repeating steps (ii)–(vii).

III. ASSESSING SEARCH PERFORMANCE: SUCCESS CURVES

In the following we will discuss in more detail all the above outlined elements of the search and assess the importance of specific choices that have been made in regards to each of them. To this end, the effect on the search performance is quantified by carrying out multiple independent restarts of the search on problems, for which the global minimum is known. As depicted in Fig. 2, the results are converted into *success curves*, showing, as a function of the number of DFT evaluations spent, how large a fraction of the independent searches have found the global minimum.
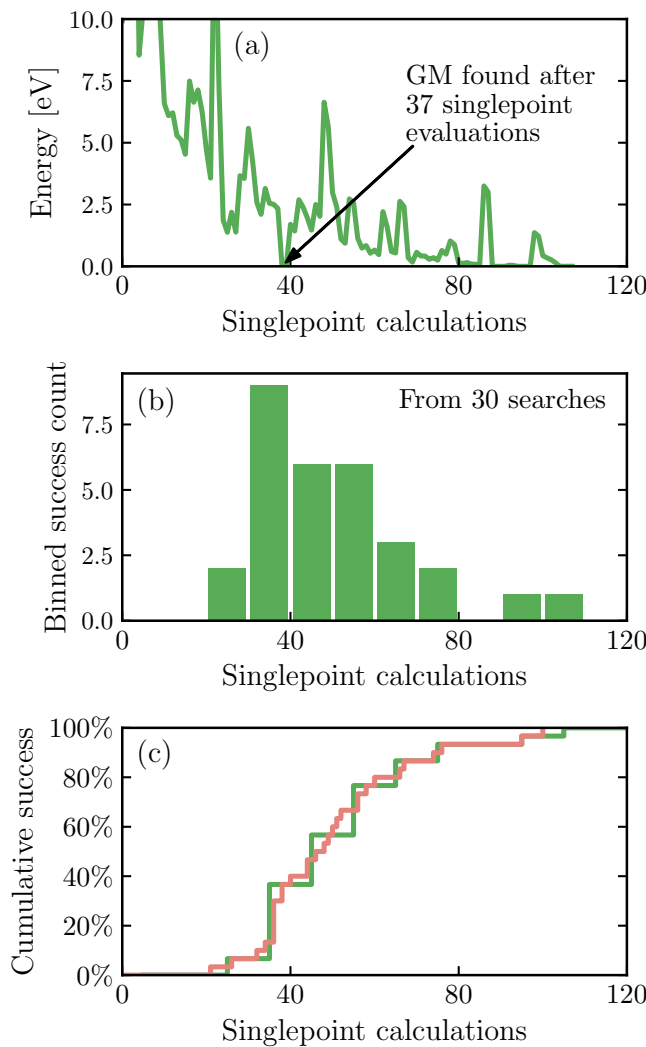
FIG. 2. The construction of success curves based on a known global minimum. (a) DFT energies for the structures in a single search instance, as the search progresses one single-point evaluation at a time. (b) Histogram (with a bin size of 10) over 30 independent search instances, counting at what point in each search the known global minimum is first found. (c) Green curve, normalized cumulative distribution of (b); red curve, cumulative distribution with a bin size of 1 (this is referred to as a success curve).

From the success curves one can directly read, e.g., how many DFT evaluations are required for a single search to have 50% chance of finding the global minimum. In this paper, we use, as our main problem on which to compare search performance, the anatase $TiO_2(001)$-(1×4) surface reconstruction [11], for which the global minimum is already known and is shown in Fig. 3. We will address this problem with two levels of difficulty, requiring the search to find the correct positions for atoms in the uppermost two or three layers (see Fig. 3). We will label these the two-layer and three-layer $TiO_2(001)$-(1×4) problems, respectively. For computational convenience, all TiO_2 studies were carried out using density functional tight-binding (DFTB) theory calculations with the parameters from Ref. [69]. Using DFTB as opposed to DFT for these systems does not introduce any significant differ-
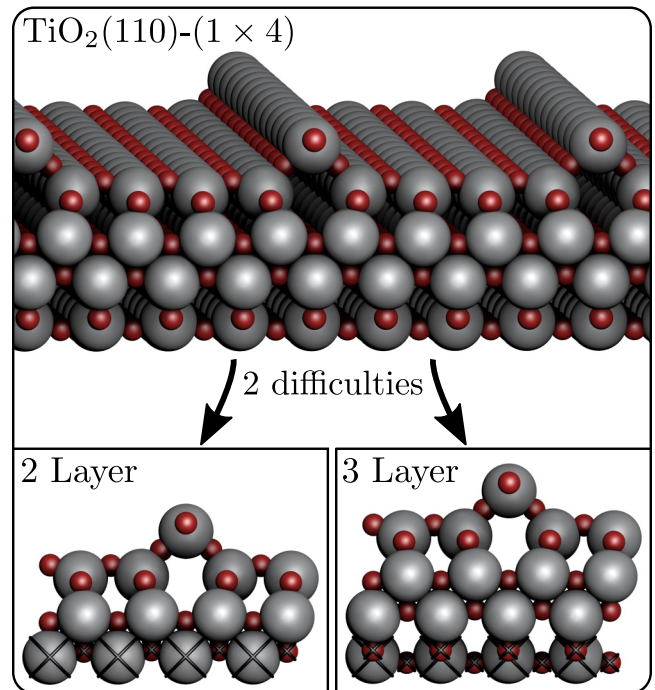

FIG. 3. The anatase $TiO_2(001)$-(1×4) surface reconstruction, featuring rows of protruding titanium atoms. In the structure search context we consider two versions of the problem, having either two or three atomic layers optimized on top of a fixed bulk layer. One unit cell of the global minimum for these problems, labeled the two-layer and three-layer $TiO_2(001)$-(1×4) problems, respectively, are shown in the figure. For the two-layer problem 27 atoms are optimized, whereas 39 atoms are optimized in the three-layer problem.

ence, as the two potentials share the same global minimum and are of comparable difficulty from a structure search perspective. As we move on to discuss each element of the search algorithm in detail, we will make use of success curves to assess the performance of these elements.

A. Surrogate model

The GOFEE search method relies heavily on a computationally inexpensive surrogate model of the energy landscape, to reduce the number of expensive DFT evaluations required to carry out a search. For this purpose we adopt a Bayesian approach, to have convenient access to prediction uncertainties, and specifically choose to use a GPR model, as it is very adequate at learning continuous functions and behaves well even with little training data. This is indeed the condition in the beginning of a search, where only a small number of structures have yet been evaluated. The data comprise a set of observed atomic configurations $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T$ and their energies $\mathbf{E} = (E_1, E_2, \ldots, E_N)^T$. To accommodate learning, it is crucial to take advantage of the basic symmetries of the Hamiltonian. This is achieved by letting $\mathbf{x}_i$ be a suitable representation of the ith configuration. In the GOFEE method we follow the approach of Valle and Oganov [30] and use, for the representation, a Gaussian smeared distribution of interatomic distances and extend it to interatomic angles as well,
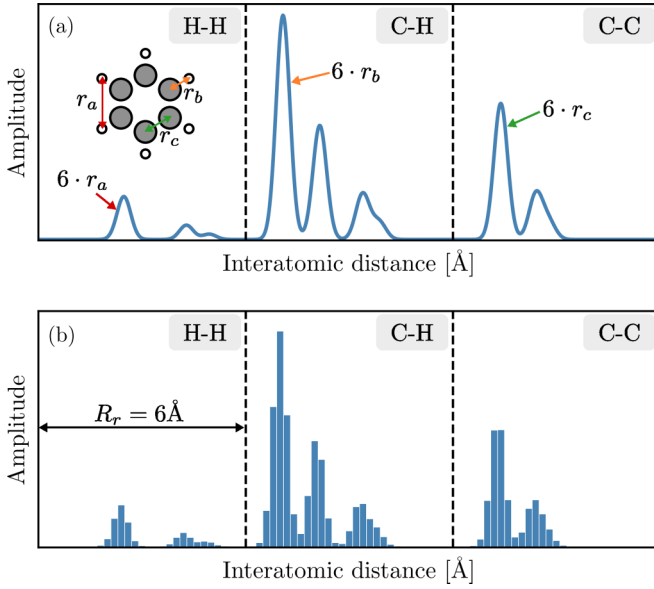
FIG. 4. The radial part of the representation depicted for the benzene molecule. (a) Radial distribution functions for the three unique combinations of atomic species (H-H, H-C, and C-C) are concatenated. Three specific separation distances are highlighted on the benzene model, and their contributions to the distribution functions are highlighted. (b) To turn the distribution functions into a representation, each is discretized into 30 elements. The angular part of the representation can be constructed analogously and is concatenated with the radial part to construct the full representation.

which we will refer to as the radial and angular distribution functions.

As depicted in Fig. 4, for the radial part alone, the representation is constructed by concatenating the distribution functions for each unique combination of atomic species, after which the distribution functions are discretized to achieve a finite dimensional representation. To achieve the full representation, the angular part of the representation is constructed in an equivalent manner, and is concatenated onto the radial part. The relevant radial and angular distribution functions, for each unique combination of atomic species, are defined as follows:

$$F_{A,B}(r) = \begin{cases} D_r \sum_{A_i,B_j} \frac{1}{r_{ij}^2} \exp\left(-\frac{(r-r_{ij})^2}{2l_r^2}\right), & r < R_r, \\ 0, & r \geqslant R_r, \end{cases}$$
(1)

$$F_{A,B,C}(\theta) = D_\theta \sum_{A_i,B_j,C_k} f_c(r_{ij})f_c(r_{ik}) \exp\left(-\frac{(\theta-\theta_{ijk})^2}{2l_\theta^2}\right),$$
(2)

where f_c is the smooth cutoff function,

$$f_c(r) = \begin{cases} 1 + \gamma\left(\frac{r}{R_\theta}\right)^{\gamma+1} - (\gamma+1)\left(\frac{r}{R_\theta}\right)^\gamma, & r < R_\theta, \\ 0, & r \geqslant R_\theta, \end{cases}$$
(3)

with A, B, and C being placeholders for different atomic species. For the cutoff function a sharpness of $\gamma = 2$ is used. The radial and angular cutoff radii adopted are $R_r = 6$ Å and $R_\theta = 4$ Å, respectively, and the widths of the Gaussians used

for smearing are $l_r = 0.2$ Å and $l_\theta = 0.2$ rad. The scaling constants are $D_r = \Omega/(4\pi N_a^2)$ and $D_\theta = \eta\Omega/N_a^3$, respectively, with Ω being the supercell volume and N_a the number of atoms. $\eta = 20$ Å^{-2} is introduced to convert units and scale the angular contributions relative to the radial ones. Finally, each individual distribution function has been discretized into 30 elements as illustrated in Figs. 4(a) and 4(b) for the radial part.

Representing atomic configurations with feature vectors **x** in this form, the GPR model is then tasked with inferring a distribution over functions, $p(E_{sur}|X, \mathbf{E})$, that is consistent with the training data $(X, \mathbf{E})$. As this distribution, by definition of a Gaussian process, is assumed normal, its mean and standard deviation, respectively, can be used to predict both the energy, $E_{sur}(\mathbf{x})$, and the expected uncertainty, $\sigma_{sur}(\mathbf{x})$, on the energy prediction, for a new atomic configuration. The GPR model is defined by its prior mean function $\mu(\mathbf{x})$ and covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$, from which the prior distribution, i.e., the untrained model, is derived. Given these, the posterior distribution, i.e., the trained model, is determined by conditioning the prior distribution on the available training data. The resulting expressions, for energy and uncertainty prediction of a new structure $\mathbf{x}_*$, are [70]

$$E_{sur}(\mathbf{x}_*) = \mathbf{k}_*^T(K + \sigma_n^2 I)^{-1}(\mathbf{E} - \boldsymbol{\mu}) + \mu(\mathbf{x}_*),$$
(4)

$$\sigma_{sur}(\mathbf{x}_*)^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T(K + \sigma_n^2 I)^{-1}\mathbf{k}_*,$$
(5)

where $K = k(X, X)$, $\mathbf{k}_* = k(X, \mathbf{x}_*)$, $\boldsymbol{\mu} = (\mu(\mathbf{x}_1), \mu(\mathbf{x}_2), \ldots, \mu(\mathbf{x}_N))^T$, and the target function is assumed noisy with uncertainty $\sigma_n^2 = 5 \times 10^{-2}$ eV^2, which acts as regularization. Figures 5(a)–5(c) illustrate a GPR model applied to data sampled from a one-dimensional function. Both the prior and posterior distributions are depicted with their mean and standard deviation, along with concrete sample functions from the two distributions.

The GPR model in Figs. 5(a)–5(c) uses the Gaussian covariance:

$$k(\mathbf{x}, \mathbf{x}') = \theta_0 e^{-(\mathbf{x}-\mathbf{x}')^2/(2\lambda^2)},$$
(6)

with characteristic length scale λ and maximal covariance θ_0. In GOFEE we instead adopt a covariance function consisting of a sum of two Gaussian covariance functions with different length scales, which we dub the double Gaussian covariance:

$$k(\mathbf{x}, \mathbf{x}') = \theta_0[(1-\beta)e^{-(\mathbf{x}-\mathbf{x}')^2/(2\lambda_1^2)} + \beta e^{-(\mathbf{x}-\mathbf{x}')^2/(2\lambda_2^2)}],$$
(7)

with characteristic length scales λ_1 and λ_2, respectively, maximal covariance θ_0, and weights given by $\beta = 0.01$. As is common for GP models, all nonfixed hyperparameters are automatically selected by maximizing the marginal log likelihood of the parameters given the observed data. In GOFEE we use multirestart gradient descent to carry out the optimization. The present covariance function is chosen because the optimized length scale of a single Gaussian covariance tends to be significantly larger than the feature space distance between neighboring local minima and comparable to the extent of the training data in the feature space. This limits the resolution of the resulting model, which in addition tends to be overconfident in its predictions. Adding the second Gaussian with a small weight and shorter length scale than the first partly remedies this problem.

Single Gaussian Covariance
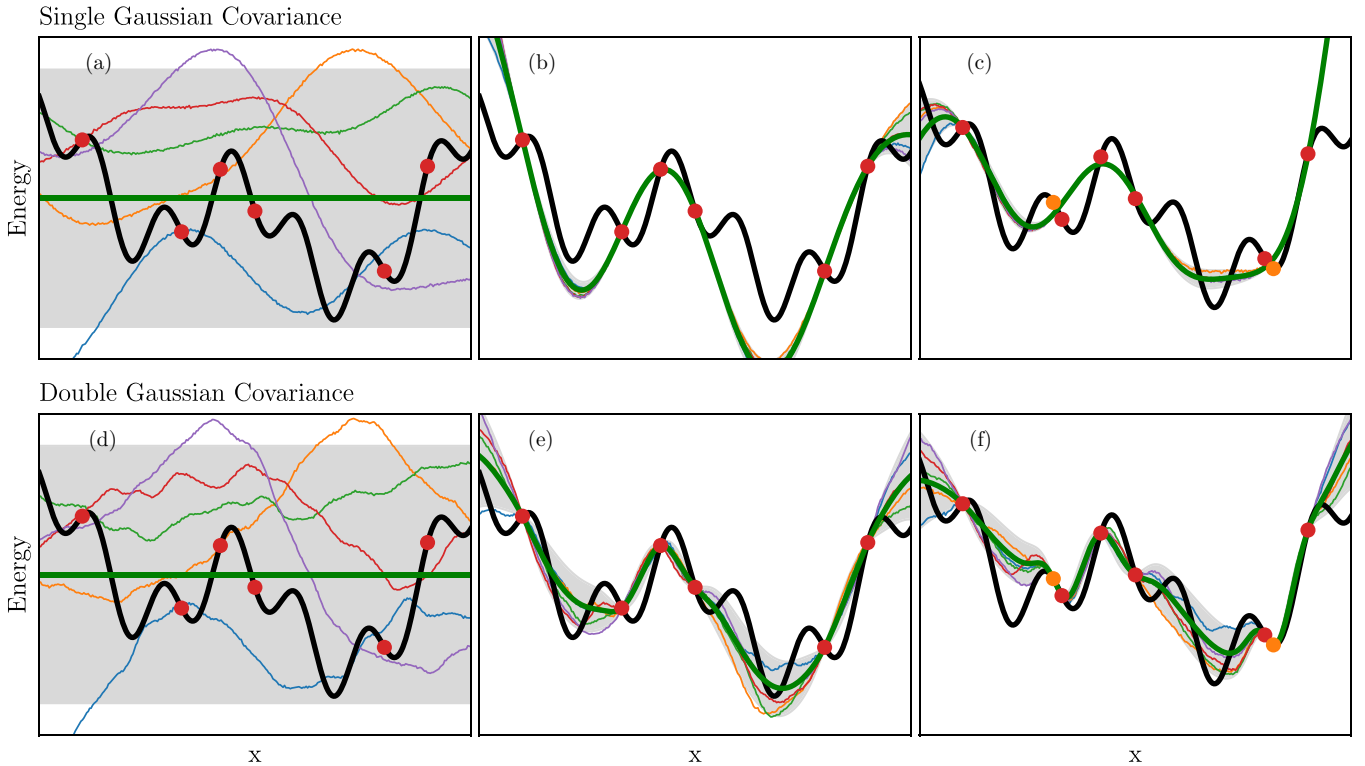

Double Gaussian Covariance

FIG. 5. One-dimensional GPR example, comparing the normal Gaussian kernel (a–c) with the double Gaussian kernel (d–f) used in GOFEE. The target function is represented by the thick black line, the GPR model is represented by the thick green line, and the training data are represented by red dots. The thin lines represent functions sampled from the model distributions. (a, d) Prior to training, both models just predict the mean of the data, however functions sampled for the two kernels are visually dissimilar, with the double Gaussian kernel resulting in short scale wiggles on top of the long scale variations. (b, e) For the trained models, the single Gaussian model tends to overshoot the target, and underestimate short scale uncertainties Even with only a weight of $\beta = 0.01$, the short scale Gaussian in the double Gaussian model remedies these flaws to a large extent. (c, f) Adding further data, as orange dots, requiring even more short scale variation, only reinforces this point.

The effect of this choice of covariance is exemplified in Fig. 5, which in addition to Figs. 5(a)–5(c), showing the model resulting from using the normal Gaussian covariance, Eq. (6), also shows, in Figs. 5(d)–5(f), the result of using the double Gaussian covariance, Eq. (7). The compared covariance functions are identical in all but the shorter length scale Gaussian, which has only a weight $\beta = 0.01$, relative to the long length scale Gaussian. Even this small addition of a shorter length scale is remarkably visible, already in the functions sampled from the priors, Figs. 5(a) and 5(d). Moving on to the trained models, the addition of the shorter length scale results in models that are better at accommodating short scale changes in the target function, and that do not suffer from the underestimation of uncertainties on the short scale, to the extent seen when using the normal Gaussian covariance. We have recently used the double Gaussian covariance in the context of structure optimization with image recognition and reinforcement learning [71], where it proved highly efficient in speeding up the searches.

Because the acquisition function, central to GOFEE, depends on the predicted uncertainties, to supply explorative incentive in the search, the underestimation of uncertainties causes a less lively search, more prone to stagnation. This is reflected in the results in Fig. 6(a), showing that the searches adopting the normal Gaussian kernel are more prone

to stagnate, compared to the double Gaussian kernel. This is indicated by the slope of the success curve almost vanishing after ≈ 700 DFT evaluations.

Besides the covariance function, the second component to the prior distribution of the GPR model is the prior mean function, $\mu(\mathbf{x})$. For this a common and simple choice is to use a constant value equal to the mean, $\bar{E}$, of the training data. It is however useful to keep in mind that the model effectively only needs to learn the difference between the prior mean function and the target function. It can therefore be useful to include general features of the target function, here the total energy, into $\mu(\mathbf{x})$. In GOFEE we add to the data mean a conservatively chosen, repulsive interatomic potential, such that

$$\mu(\mathbf{x}) = \bar{E} + \sum_{ij} \left(\frac{0.7 r_{CD,ij}}{r_{ij}(\mathbf{x})} \right)^{12} \text{eV}, \qquad (8)$$

where $r_{CD,ij}$ is the sum of the covalent radii of the ith and jth atoms and $r_{ij}(\mathbf{x})$ is the distance between them. This is used with the main purpose of not producing data of very high energy due to some very short bonds. Such data can significantly impact the energy scale of the regression problem and negatively affect the resulting GPR model.

Without the repulsive prior, the search naturally, and especially early on, tends to spend more resources sampling
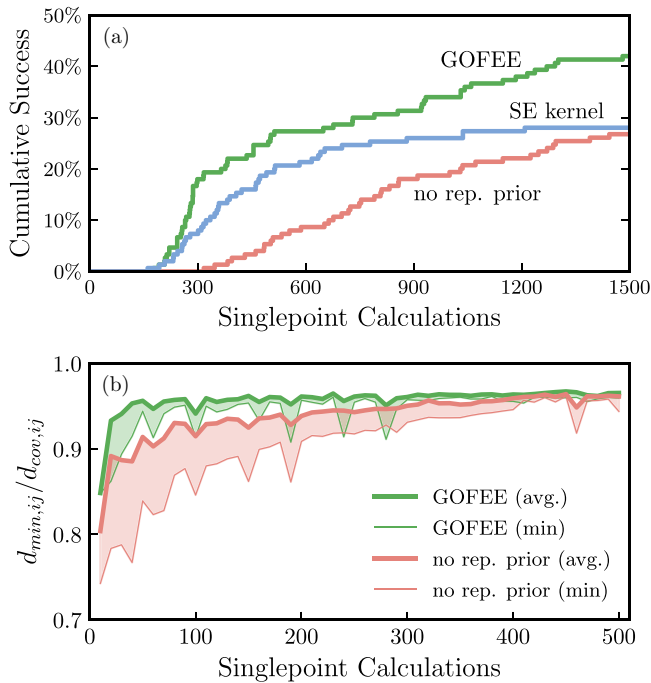
FIG. 6. Benchmark of GOFEE, on the three-layer $TiO_2(001)$-(1×4) problem. Green curve: GOFEE is used as presented in this paper, with dual-point evaluation, the double Gaussian kernel, the same $\kappa_{relax} = \kappa = 2$ used for relaxation and candidate selection, and the repulsive prior. Blue curve: Using the normal Gaussian covariance. Red curve: Omitting the repulsive prior mean function. (b) Minimum bond distance as a fraction of the covalent bond distance, shown for the first 500 structures in GOFEE searches. Thick and thin lines represent, respectively, the mean and minimum value in bins of ten episodes. All lines are averages over ten independent searches.

structures with unreasonably short bonds, which typically have high energies. As a derived effect the surrogate model also has to accommodate a significant amount of high-energy structures, which to some extent compromises the prediction accuracy on the low-energy structures, which are relevant to the search. This derived effect is largest early on, where the high-energy structures make up a larger proportion of the data. As shown in Fig. 6(a), neglecting the repulsive prior results in a less effective search strategy, with the difference being especially apparent early on in the searches. Figure 6(b) illustrates, based on data from GOFEE searches, that the first ≈ 500 structures, sampled in searches without the repulsive prior, contain significantly shorter bonds.

Sticking with the order defined in the flowchart, we will now halt the discussion of the surrogate model for a while, in favor of a description of how the parent structures are established and used as the basis for generating new candidates. We subsequently return to the surrogate model, when it is used for relaxation of these new candidates.

B. Parent structures

A population of breeding parent structures is an important element for any evolution-based search strategy, as it effectively controls what new candidate structures are likely to

be created [2,8–10]. To avoid premature convergence of the search, while still progressing it, the population must both be diverse and prioritize low-energy structures. Usually, the population is maintained in two steps: in every generation, all new offspring structures are added to it, and subsequently the "survival of the fittest" evolutionary pressure is applied to trim the population to a manageable size. In a structure search context, this is often interpreted as the population being reduced to a fixed size with the energetically most favorable, yet sufficiently different, structures as members. In some implementations, the evolutionary search is run in a mode without the concept of generations, and new structures are produced and handled either one at a time or asynchronously. In these cases, the population is updated in much the same manner, having the new structure, S_{new}, replace an existing member, S_{pop}, of the population if S_{new} is more stable than S_{pop} and S_{new} is sufficiently different from all other population members that are also more stable than S_{new}. In our original formulation of GOFEE [67], we enforced the difference criterion by applying a maximum covariance threshold of $k_{max} = 0.995$ eV^2 between any two structures of the population.

Recently, we proposed a new approach, which achieves a diverse set of parent structures via direct sampling of the entire pool of structures treated at the first-principles level [68]. The sample of parent structures is established by first selecting from the pool all structures that have an energy within ΔE_{sample} of the best structure found so far in the search. Next, these structures are clustered into N_{sample} clusters, e.g., by k-means++ clustering, and the best structure within each cluster becomes part of the sample. While the precise clustering technique is not considered important for the method, the stochastic nature of k-means++ clustering might be beneficial as it will cause variations in the composition of the sample, that may counter stagnation during late stages of searches.

The sampling scheme is sketched in Fig. 7 for real GOFEE data using principal component analysis to enable a two-dimensional representation of the data. Figure 7(a) illustrates the initial selection of low-energy structures, here using $\Delta E_{sample} = 5$ eV. Figure 7(b) illustrates the subsequent clustering of the selected structures, based on Euclidean distances in the feature-space, Eqs. (1) and (2), using the k-means++ algorithm. The dots representing the structures are colored according to the clusters the structures end up in and the cluster centers are marked with crosses. Since the clustering is done in the entire feature space, and not in the two first principal component dimensions, the clusters overlap slightly in Fig. 7(b). Once the clusters have been established, the sample is set up by choosing the most stable structure (crosses) from each cluster. Figure 7(c) depicts the actual structures in a sample resulting from applying this scheme. The number of clusters in the clustering step is fixed at the desired number of sample members. Figure 8 explores, for the three-layer $TiO_2(001)$-(1×4) system and using $\kappa = 4$ in the search, how the number of sample members affects the search performance. Comparing $N_{sample} = 5, 10$, and 25, the best performance is achieved using $N_{sample} = 10$. We speculate, however, that for more complex systems, with a larger number of competing energy funnels, having more sample
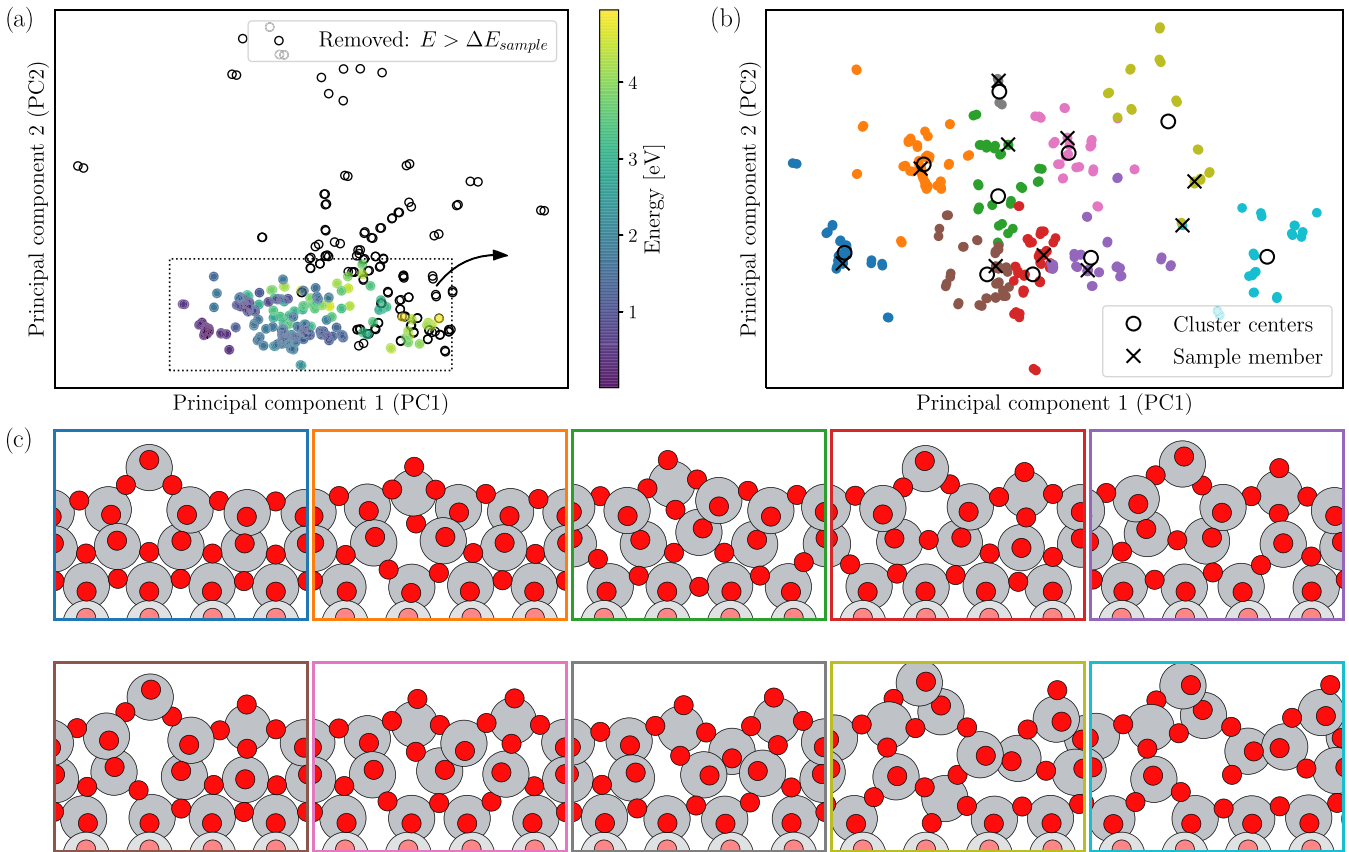
FIG. 7. The sampling scheme for assembling parent structures. (a, b) Points represent previously DFT evaluated structures, visualized with principal component analysis (PCA), with the first two principal components capturing 83% of the variance. (a) First, all evaluated structures, with an energy above ΔE_{sample} of the best structure found so far, are ignored. These are represented by black circles. The remaining structures are colored according to their DFT energy. (b) Second, the remaining structures are clustered in the feature space using the k-means++ algorithm. The resulting clusters are colored, and cluster centers drawn as black circles. Third, the sample is created by selecting the lowest-energy structure in each cluster, highlighted by black crosses. (c) The actual structures resulting from applying the sampling scheme to this dataset, consisting of the 500 first structures of a GOFEE search. As the sampled structures can lie on cluster edges, some could in principle be very similar. The diversity of the sample of parent structures is however not found to suffer greatly from this effect, as exemplified in (c). The adequacy of the PCA visualization is also visible in the fairly reasonable positions of the cluster centers, and only slight smearing of the cluster borders.

members is likely beneficial. In this paper, $N_{\text{sample}} = 10$ sample members have been used for all other presented searches.

Figure 8 also compares the clustering based sampling strategy to the similarity based population strategy for handling the parents. The sample based search generally performs best.

The presently proposed means of using clustering in setting up the set of parent structures for GOFEE differs from our previous use of clustering in conjunction with an algorithm approach. In the work of Jørgensen *et al.* [72] new members for the population were chosen based on energy (fitness) and dissimilarity to lower-energy members of the population and selection of population members to become parents was made with a selection scheme that favored population members that were outliers when clustering the entire pool of calculated structures. In this paper, we construct the set of parent structures—the sample—directly via the clustering, have a uniform selection scheme for picking parents from the sample, and choose the optimal candidate for DFT evaluation via Bayesian statistics, as detailed below.

C. Generating new candidates

After setting up a sample of parent structures, N_c, new candidate structures are generated by applying stochastic changes to structures uniformly drawn from the sample. At present, these include simple operations such as shifting or permuting atoms. More elaborate operations such as crossover [73,74] represent a potential future avenue for improvements. It is worth pointing out that generating multiple new candidates is only beneficial because we have the computationally inexpensive surrogate model, which can be used to relax and compare them all selecting a single one, to be evaluated with DFT. Since sample members have structures that reflect local minima in the surrogate model-based acquisition function at earlier stages of the search, it serves a purpose to reoptimize the sample members with newer versions of the acquisition functions. We hence occasionally add to the N_c generated candidates, the entire sample. Specifically, it is done in every third search iteration and adds to the degree of exploitation done for the best structures found so far. However, since only a single
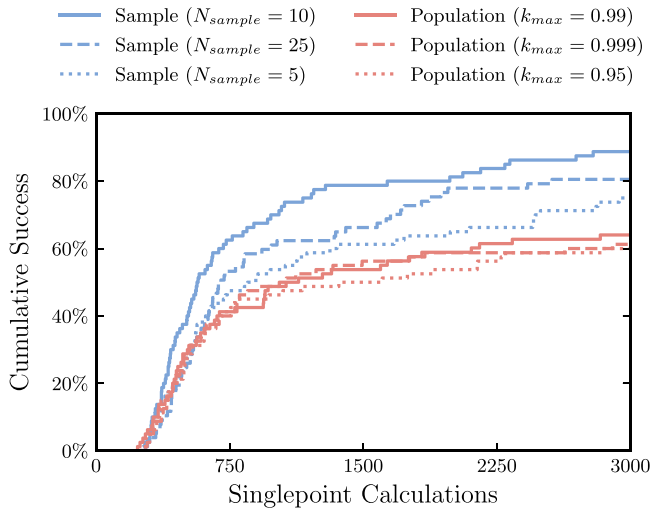
FIG. 8. Comparison of sample strategy and population strategy for handling parents. Success curves are for the three-layer TiO$_2$(001)-(1 × 4) system and for this comparison and $\kappa = 4$ has been adopted to reduce the degree to which the acquisition function is a limiting factor for exploration. Blue curves: Cluster based sampling strategy, with different values of the sample size, $N_{\text{sample}} = 5, 10, 25$. Red curves: Similarity based population strategy, with different values for the similarity threshold, $k_{\text{max}} = 0.95, 0.99, 0.995$.

candidate is selected for DFT evaluation, this measure comes at the expense of less resources being spent on exploration. This particular compromise ensures that at least two-thirds of the resources are spent on some degree of exploration.

We will now turn back to the surrogate model, and discuss how it is used to define an acquisition function for relaxing the new candidates and selecting the most promising one among them. But before we move on, it is important to introduce the acquisition function used in combination with the surrogate model in these steps.

D. Acquisition function

The main goal of the search is to arrive at the global minimum structure. The surrogate model contributes to this task by inexpensively supplying approximate energies of new structures, thus avoiding many unnecessary DFT evaluations. However, the surrogate model is inherently uncertain in its predictions, especially in regions far from its training data. When deciding what new candidate structure to evaluate with DFT, it is useful to also take the uncertainty into account. This is the purpose of the acquisition function [75], $f(\mathbf{x})$. It is used to quantify how desirable, to the overall optimization problem, it is to evaluate a given structure with DFT, taking into account both the predicted energy, $E_{\text{sur}}(\mathbf{x})$, and the predicted uncertainty, $\sigma_{\text{sur}}(\mathbf{x})$. In this paper we have used the lower confidence bound [76,77] of the model predictions,

$$f(\mathbf{x}) = E_{\text{sur}}(\mathbf{x}) - \kappa \cdot \sigma_{\text{sur}}(\mathbf{x}), \qquad (9)$$

due to its simplicity. Here κ controls the emphasis put on the predicted uncertainty. This also effectively controls the degree of exploration in the search strategy. Ideally we would want to identify the structure corresponding to the global minimum of the acquisition function and evaluate this structure with

DFT, repeating this to progress the search. This is however not practically feasible in the present setting, due to the size of the configurational space. Instead we must settle on sampling the configurational space and selecting among these samples, which we refer to as new candidate structures.

E. Relaxing new candidates

In sampling the configurational space, there is general consensus among well-established, all DFT, search strategies such as random search [1], basin hopping [4,5], and evolutionary algorithms [2,8–10], that it is more efficient to explore the configurational space by locally optimizing each sample, compared to evaluating a larger number of samples, without optimizing them. We apply the same strategy to sample the acquisition function and locally optimize each new candidate structure in the acquisition function, which we also refer to as relaxing the structures. The procedure is illustrated schematically in Fig. 9. After such relaxation of a number of structures, the acquisition function, i.e., the κ-dependent lower confidence bound, is further used to select which new candidate structure is most beneficial to evaluate with DFT. This selection step will be further discussed in the next section.

Figure 10(a) shows the success curves, for finding the global minimum structure of the two-layer TiO$_2$(001)-(1 × 4) system. The curves represent the expected performance from running GOFEE with different values of κ. Each curve is produced from 75 independent GOFEE restarts. Comparing the $\kappa = 0$ curve with those for $\kappa = 1, 2$, and 3 strongly suggests using a value of $\kappa > 0$. Figure 10(b) shows snapshots of the corresponding success rates at different points in the search. From Fig. 10(b) it is seen that the pronounced tendency to exploit, associated with small values of κ, results in some, lucky searches finding the GM very quickly. For instance, $\approx$30% of searches with $\kappa = 1$ do find the GM within 200 evaluations. Using $\kappa = 1$, however, results in a success rate of only $\approx$70% within 800 evaluations, compared to the more explorative searches, such as $\kappa = 4$–8, achieving success rates of 90% within 800 evaluations. Even larger values of κ are expected to cause similar or higher success rates, if the searches were allowed to progress for more episodes. Based on this, a choice in the range $\kappa = 2$–4 is suspected to strike a reasonable balance between exploration and exploitation. In this paper we have used $\kappa = 2$ for most searches, but note that, if the computational budget allows, it is likely beneficial to adopt a slightly larger value of, e.g., $\kappa = 4$–6 in searches, carried out for more episodes, as this will improve the thoroughness of the search. As an example, Fig. 10(c) shows success curves for both $\kappa = 2$ and 4 for the three-layer TiO$_2$(001)-(1 × 4) system. For this more complex system, $\kappa = 4$ performs notably better overall, with $\kappa = 2$ only offering the best performance in the the very limited range of $\approx$250–350 single-point calculations.

While in GOFEE the acquisition function with $\kappa > 0$ is used for both relaxation and selection, one could also imagine using different values of κ, i.e., $\kappa_{\text{relax}} \neq \kappa_{\text{select}}$. The posterior mean is the model's best guess of the target potential, and therefore also of the locations of local minima, which ultimately are the structures of interest. Following this line of thought suggests using $\kappa_{\text{relax}} = 0$, while maintaining $\kappa_{\text{select}} > 0$ to maintain explorative incentive when selecting. Looking
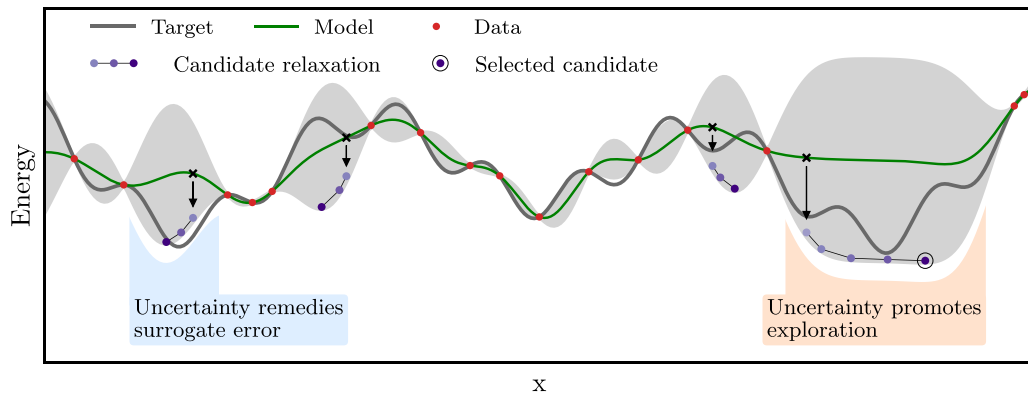
FIG. 9. Artificial one-dimensional PES (gray line) and surrogate model (green curve) estimating the PES based on training data (red dots). The figure depicts the advantage of relying on both surrogate energy and uncertainty for relaxation of new candidates and data acquisition. The new candidates (crosses), in this case $N_c = 4$, are generated in each search episode, and relaxed (purple dots) in the lower confidence bound acquisition function.

at Fig. 11(a), adopting $\kappa_{\text{relax}} = 0$ does however significantly reduce the performance on the three-layer $TiO_2(001)$-(1×4) system. The main reason is likely that leaving out the bias towards uncertain structures during relaxations naturally results in less diversity in the training database containing the evaluated structures. This is supported by Fig. 11(b), which shows learning curves for models resulting from the two strategies. The models, actively learned, using $\kappa_{\text{relax}} = 0$, are significantly less accurate than those with $\kappa_{\text{relax}} = 2$ when tested on the diverse set of 1000 test structures. This suggests that the active learning element of the search method suffers due to reduced diversity in the training database, collected during searches using $\kappa_{\text{relax}} = 0$.

F. Selecting candidate

Having established the acquisition function for selecting among the new candidates, it is interesting to investigate how efficiently it chooses from among the N_c new candidates, and in particular to what extent it is worth it to increase the number of candidates from which the acquisition function chooses. As long as the acquisition function is as good or better than a random selection scheme at selecting the candidate best suited for progressing the search, increasing N_c should statistically benefit the search. However, at some point the computational cost of generating and locally optimizing new candidates in the surrogate model will contribute significantly to the total computation time of the search, which in practice limits the value of N_c. This contribution typically becomes significant at values $N_c > 100$, but is dependent on the specific system and choice of DFT settings. Figure 12(a) shows the success curves for different values of N_c. It illustrates the dramatic importance of generating multiple candidates in each search iteration. When using $N_c = 1$ the problem is barely solvable, with only 10% of searches successful, in 800 DFT evaluations, whereas 90% of searches are successful in only little more than 200 evaluations, when using $N_c = 256$. This dramatic improvement of the search performance, when increasing the number of new candidates generated in each search iteration, highlights the importance of surrogate models being not only accurate, but also efficient to evalu-

ate. Efforts to reduce the computational cost of evaluating machine learned regression models [34,78,79] are therefore highly relevant, along with methods to reduce the number of evaluations spent on local optimization [43,80], here carried out using the model.

Another way to assess the quality of the surrogate based acquisition function is to measure its ability to take advantage of desirable structures, when they appear in the set of new candidate structures. This could be the global minimum structure, which is objectively desirable. One way of quantifying this is thus to measure at each candidate selection phase whether the global minimum structure would have been selected by the acquisition function, were it among the candidates actually generated by the search. As shown in Fig. 12(b), already long before the global minimum is found in any search, and to a remarkable extent, the acquisition function prefers the global minimum over any of the candidates generated in the search. This does not tell us that the model has already identified the global minimum, as there are likely other structures, not among the candidates, that would be selected even over the global minimum. It does however support the acquisition function's ability to successfully identify promising structures even when far from the structures currently explored in the search. In addition it indicates that the search can be further improved by improving the generation of new candidates, which can be achieved either by increasing the number N_c of new candidates, by improving the operations used to generate new candidates, or by improving the way the sample of parent structures is made.

G. Evaluation

As the final step in a search iteration, the selected structure is evaluated using DFT. There are several different ways this evaluation could be carried out. On one end of the spectrum, only a single-point evaluation is carried out for the selected structure, whereas on the other end the structure is locally optimized with DFT to a certain force threshold, or for a certain number of steps. With the computational cost of DFT in mind, we opt for a solution toward the single-point part of the spectrum, but have found it beneficial to use two DFT
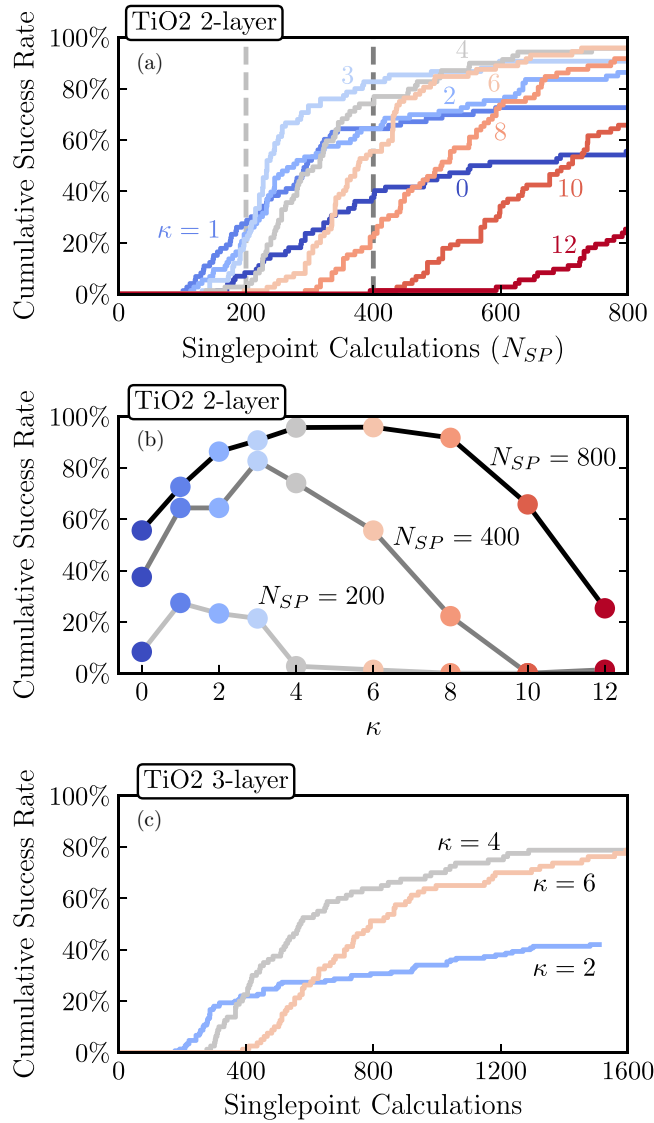
FIG. 10. The effect of κ on the search performance. (a, b) Results for the two-layer $TiO_2(001)$-(1×4) system. (a) Success curves, each based on 75 independent search restarts, for different values of κ, with values increasing when going from blue to red. (b) Snapshots of the success rates, for the curves in (a). The snapshots are taken at $N_{SP} = 200$, 400, and 800 single-point evaluations into the search. (c) Comparison of $\kappa = 2$ and 4 for the larger three-layer $TiO_2(001)$-(1×4) system.

evaluations in total, in what we dub a dual-point evaluation. A dual-point evaluation consists of a single-point evaluation of the structure itself and, in addition, a second evaluation resulting from a single gradient decent step, with the step $\Delta \mathbf{R}$ given by

$$\Delta \mathbf{R} = \begin{cases} r_{\max} \frac{\tilde{F}_{\max}}{\tilde{F}_{\text{flat}}} \hat{\mathbf{u}}, & \text{if} \quad \tilde{F}_{\max} < \tilde{F}_{\text{flat}} \\ r_{\max} \hat{\mathbf{u}}, & \text{if} \quad \tilde{F}_{\max} \geqslant \tilde{F}_{\text{flat}} \end{cases}. \quad (10)$$

Here $\hat{\mathbf{u}}$ is the unit vector in the force direction and $\tilde{F}_{\max}$ represents the largest atomic force of any atom in the structure. This choice $\Delta \mathbf{R}$ ensures that the step length is linear in the size of the largest force of any atom, $\tilde{F}_{\max}$, up to the point that this
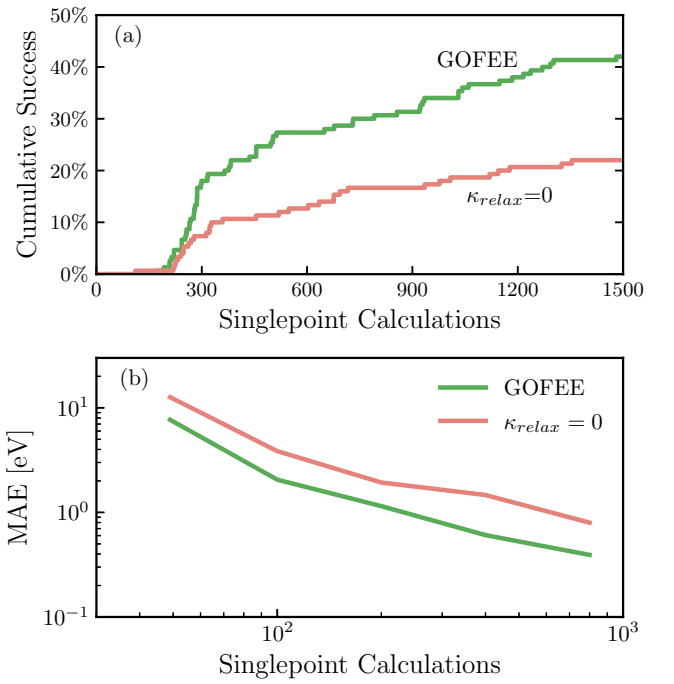

FIG. 11. Benchmark of GOFEE, on the three-layer $TiO_2(001)$-(1×4) problem. (a) Green curve, GOFEE as presented in this paper with $\kappa = \kappa_{\text{relax}} = \kappa_{\text{select}} = 2$; red curve, relaxations and selection are carried out with $\kappa_{\text{relax}} = 0$ and $\kappa_{\text{select}} = 2$, respectively. (b) Learning curves based on the surrogate model trained on the 50–800 first structures visited in GOFEE searches. The models are tested on a diverse set of 1000 structures and the reported MAE values are averaged over ten training sets from independent searches.

force equals $\tilde{F}_{\text{flat}}$, from which point the step size levels off at $r_{\max}$. In GOFEE, the values $r_{\max} = 0.1$ Å and $\tilde{F}_{\text{flat}} = 5$ eV/Å are used.

The dual-point evaluation is used to take advantage of the forces readily available from DFT codes at little additional cost, when having already evaluated the energy. When considering success curves as a function of search iterations, as shown in Fig. 13(a), using the dual-point structures only to improve the GPR model represents a significant improvement compared to using single-point evaluations only. In addition, using the dual-point structures for all elements of the search, i.e., for the GPR model and the sample of parents, results in further improvement. The full effect of the dual-point scheme thus results from the combined benefit to both the GPR model and the population. To correct for the additional single-point evaluation carried out per search iteration in the dual-point scheme, Fig. 13(b) shows the more fair comparison of cumulative success as a function of single-point evaluations. In this comparison, the full dual-point scheme still, but to a lesser degree, performs significantly better than using only single-point evaluations. Figure 13(b) shows how inclusion of the dual-point structures benefits the quality of the sample of parents. Especially for uncertain structures, the dual-point evaluation frequently results in lowering the energy by 0.5 to 1 eV. As novel structures tend to be more uncertain, the dual-point evaluations will thus result in a positive bias towards selecting novel structures of the sample of
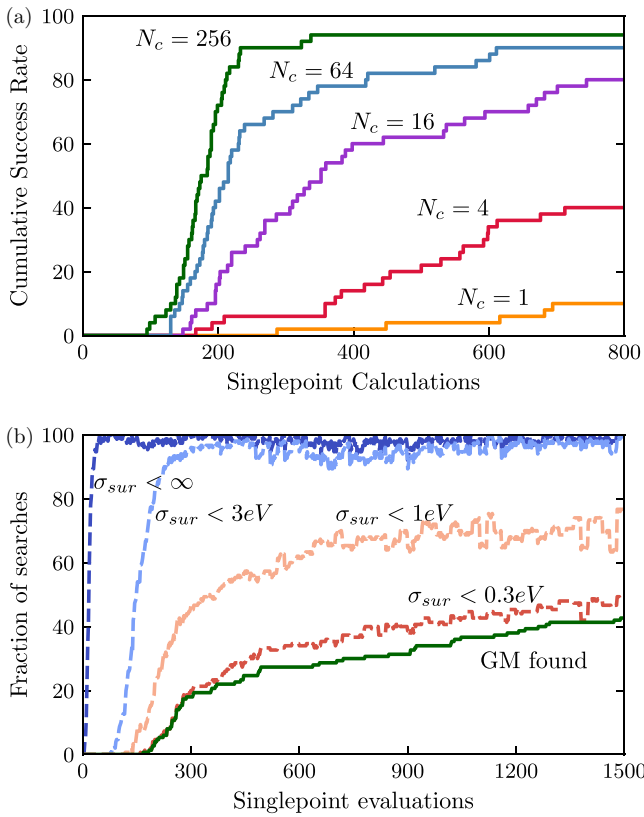
FIG. 12. Importance of the new candidate structures. (a) Success curves for the two-layer $TiO_2(001)$-(1×4) problem, resulting from searches with different numbers, N_c, of new candidates generated in each search iteration. The gain in performance, from, e.g., quadrupling N_c is remarkable, especially in the range of relatively small N_c. (b) The full, green line shows the success curve for the three-layer $TiO_2(001)$-(1×4) problem. The leftmost dashed (dark blue) line shows the fraction of the searches, that, based on the acquisition function, would have selected the global minimum, had it been among the new candidate structures in the searches. The remaining dashed lines only count this artificial selection of the global minimum (GM) successful, if predicted uncertainties on the GM, $\sigma_{sur}(\mathbf{x}_{GM})$, thereon, are less than the various threshold values given in the figure. This supplies information about how certain the model is about its prediction of the global minimum, in cases where it would be selected over the other candidates.

parents. For this effect to be significant pronounced, the step must be fairly large, as achieved by the parameters used in GOFEE. A similar benefit might be achieved by implementing a preference towards novelty [77] in addition to the diversity enforced by the clusters, when constructing the sample of parents.

One can also take advantage of the force information by explicitly including it, when training the GPR model [28,43,44,46,47,81,82]. This is not adopted in GOFEE, although we have explored the approach, the reason being significantly longer training and evaluation times, which makes it unfit for all but small systems, despite improving model predictions. This is also noted in Ref. [81], extending GOFEE with force information.
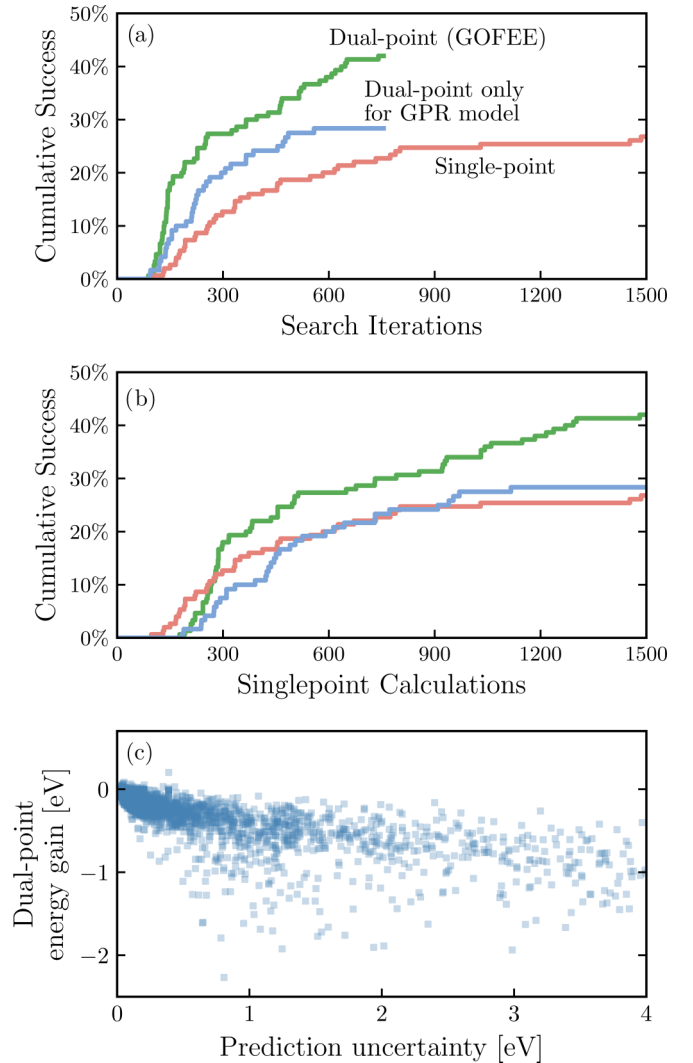

FIG. 13. Benchmark of GOFEE, on the three-layer $TiO_2(001)$-(1×4) problem. (a) Success curves as a function of search iterations. Green curve: GOFEE as presented in this paper, i.e., with dual-point evaluations. Red curve: Single-point evaluations only. Blue curve: With dual-point evaluation but the dual-point structures are used only for the GPR model and not for constructing the sample of parent structures. (b) The same success curves as in (a), but as a function of single-point evaluations used. (c) The difference in target energy between the single- and dual-point structures, plotted as a function of the prediction uncertainty for the single-point structure.

IV. APPLICATIONS

To showcase the versatility of the method we finally present two full-scale DFT structure searches. Specifically, we have applied the method to gas-phase C_{24} clusters [83], and Ir(111) surface bound C_{18} clusters.

A. Carbon clusters

In the case of the C_{24} clusters, the potential energy landscape exhibits multiple local minima, which are very close in energy to the global minimum, but which are geometrically quite different, as seen in Fig. 14, which shows the
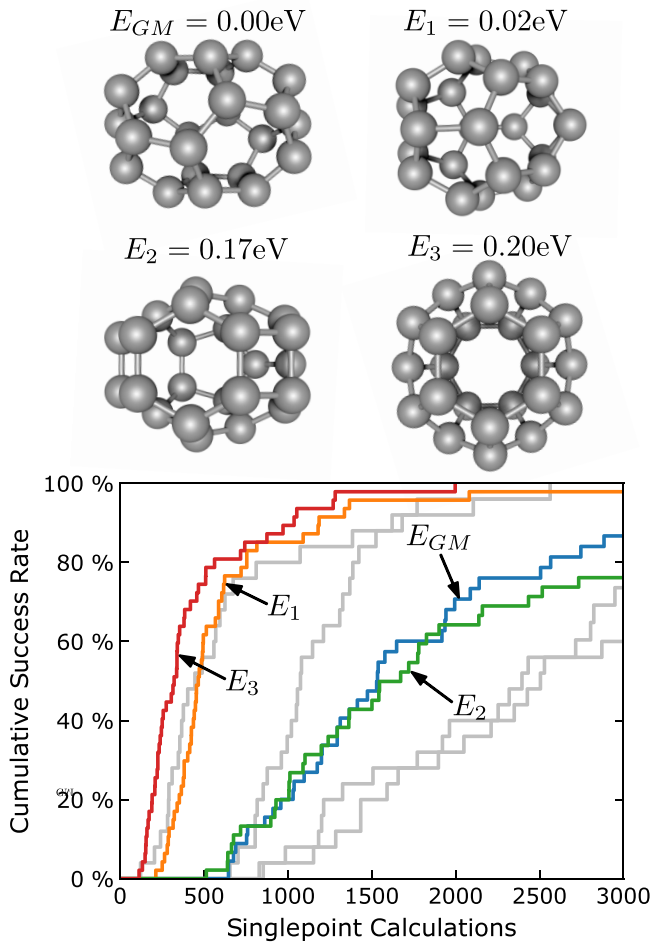
FIG. 14. The geometries of the four lowest-energy structures of C_{24}, along with success curves for finding each structure, based on 50 independent restarts of GOFEE. The colored success curves represent searches run with the new sampling strategy for handling parents and an elaborate relaxation scheme detailed in Ref. [84]. The gray curves represent searches without these additions. Structures 1 and 3 are found in almost 100 % of searches after only 1500 DFT evaluations, whereas the global minimum and structure 2 are found in ≈80 % of searches after 3000 DFT evaluations. These might be more difficult because they both feature two four membered rings, which are typically not energetically favored compared to hexagonal and pentagonal rings. In comparison, structures 1 and 3 contain only one and none, respectively.

geometry of the four lowest-energy structures for the system. The GOFEE search method relies to a very large extent on the surrogate model, trained on the fly, from scratch. One might therefore be concerned that the search can overlook important minima in unexplored regions of the search space, if the surrogate model falsely predicts a higher energy, and therefore never visits the region to update this prediction. However, because the acquisition function also takes into account the uncertainty of the surrogate model, the search is naturally driven towards such unexplored regions, despite unfavorable initial predictions, and the surrogate model will thus eventually be corrected in these regions. As seen from Fig. 14, which also shows the success curves for finding each of the four structures, the search does indeed explore the

low-energy minima thoroughly, as almost ≈80 % of searches find all four structures within 3000 DFT evaluations. This exemplifies, that, given time, the search is able to identify a number of lowest-energy structures. This is for example relevant in cases where it is not computationally tractable to carry out the searches with a sufficiently high level of DFT, but where the global minimum can be identified from among the lowest-energy structures of a less computationally demanding DFT potential used to carry out the search.

B. Carbon clusters on Ir(111)

In recent years, graphene nanostructures, such as nanoribbons and nanoislands, have attracted increasing attention, due to the tunability of their electronic and chemical properties, achieved by varying the shape and size of such structures [85]. Compared to nanoribbons, graphene nanoislands have the largest potential for variation, owing to the larger number of configurational degrees of freedom [86,87]. In addition, they are important for the understanding of graphene growth [88,89] and are promising candidates for quantum dot technology [90]. Graphene nanoislands show much variety in both shape and size, with the favored shapes governed by the substrate, in particular its interaction strength with the islands as well as the degree of lattice mismatch. Several studies have investigated the structures of small to medium sized graphene islands (C_n with $n \leqslant 24$) on various transition metal surfaces [88,91–96]; however, in all cases, the exploration of the configurational space was limited to a small number of manually constructed candidates, for each island size. In order to support and extend this body of research, we have, in the present paper, used GOFEE to carry out unbiased searches for the lowest-energy C_{18} islands on the Ir(111) surface. We choose Ir(111) for its ability to support a particularly high quality graphene layer. For this problem, GOFEE was applied with two layers of iridium atoms all fixed at bulk positions and the positions of all carbon atoms were optimized, starting from randomly initialized positions. To avoid unintended infinite carbon structures extending through the periodic boundary conditions, carbon atoms were constrained to stay at least 1 Å from the edge of the periodic, computational cell, in the plane of the surface. During the search, energy and force evaluations were carried out using the GPAW code in LCAO mode with a dzp basis set. Generalized gradient approximation with the dispersion corrected optPBE-vdW functional was used to describe the exchange–correlation interaction [97]. Subsequently, the best structures were transferred to a slab with four layers, and relaxed with only the bottom two layers fixed.

The resulting lowest-energy C_{18}/Ir(111) structures are shown in Fig. 15. All structures exhibit a domelike shape, with the edge atoms bonding to the surface, and the center atoms detached. The predominant building blocks are hexagonal and pentagonal rings, as is also the case for medium to large carbon clusters in the gas phase. However, compared to isolated clusters, the geometry of supported clusters is also governed by their ability to bond to the support. In the structures found, this is represented by a few cases (4,11,13) of heptagonal rings, but especially by the many structures (0–9,11) featuring single atoms branching off closed rings, to bond in hollow sites of the iridium surface. Such branching atoms are present
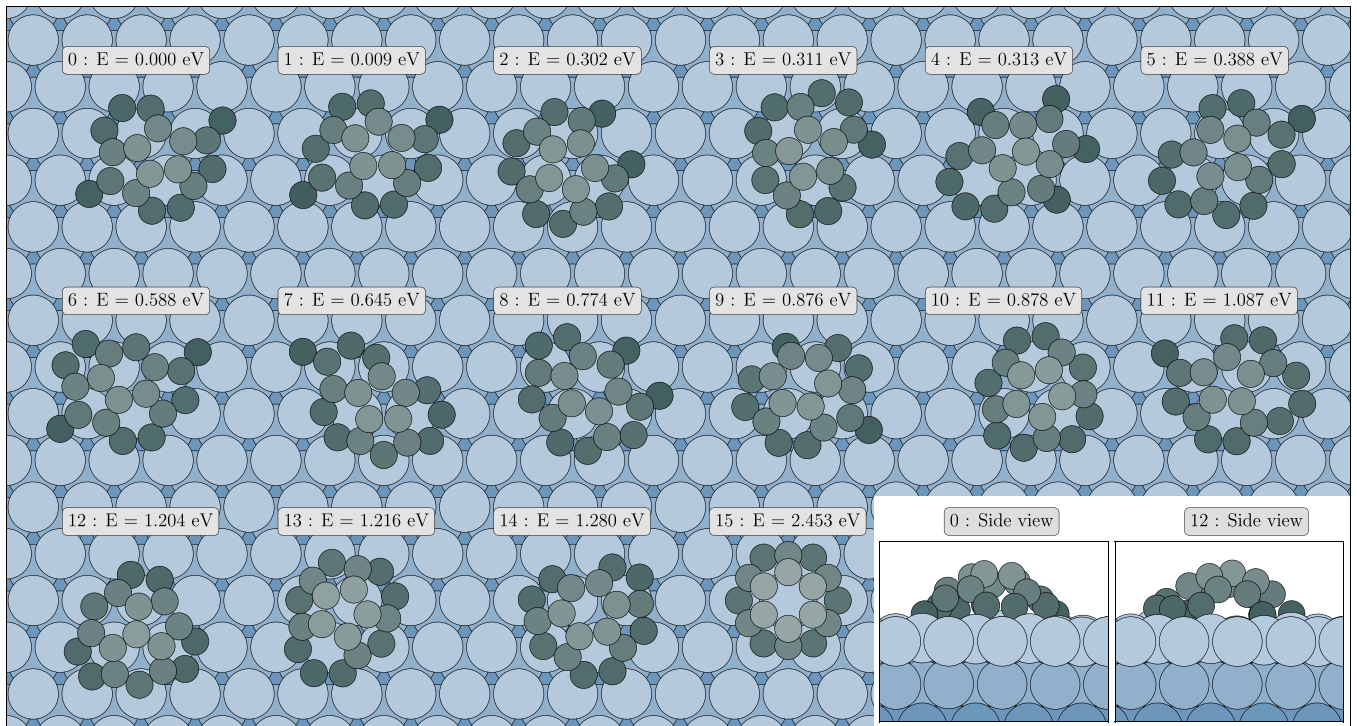
FIG. 15. Top views of the lowest-energy structures of C_{18} supported on Ir(111). Carbon and iridium atoms are shown in gray and blue colors, respectively. The brightness of the colors follows the coordinate perpendicular to the surface. Side views are further shown for structures 0 and 4.

in all structures found within ≈ 0.85 eV of the proposed global minimum. The most predominant edge motives are however still a combination of closed pentagonal and hexagonal rings. In situations where closed rings alone presumably do not offer a favorable bonding to the substrate along the entire edge, the results suggest that the favored alternative is the single branching atoms. The ability of the branching atoms to flexibly bond to the surface is illustrated by a comparison of structures 7 and 9, which feature carbon atoms branching at very different angles to reach the nearest hollow site.

The complexity and diversity of the structures found is striking and highlights the need for an automated and unbiased search strategy as the present one. We speculate that any heuristic or otherwise human inspired search strategy [88] would generally fail in identifying all such structural candidates for similar types of problems.

V. CONCLUSION

In conclusion we have here further documented the GOFEE search method [67]. This includes how the introduc-

tion of an additional, shorter length scale in the kernel can improve short scale resolution and uncertainty estimates of kernel based surrogate models, how the degree of exploration can conveniently be changed, and the dramatic improvements resulting from increasing the number of new candidate structures generated in each search iteration. In addition we have presented a scheme for extracting a diverse sample containing low-energy structures. The search method has been applied to various systems, including surface reconstruction, isolated clusters, and supported clusters.

[1] C. J. Pickard and R. J. Needs, J. Phys.: Condens. Matter **23**, 053201 (2011).

[2] A. Oganov and C. Glass, J. Chem. Phys. **124**, 244704 (2006).

[3] E. L. Kolsbjerg, A. A. Peterson, and B. Hammer, Phys. Rev. B **97**, 195424 (2018).

[4] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1998).

[5] S. Goedecker, J. Chem. Phys. **120**, 9911 (2004).

[6] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Phys. Rev. B **82**, 094116 (2010).

[7] Y. Wang, J. Lv, L. Zhu, and Y. Ma, Comput. Phys. Commun. **183**, 2063 (2012).

[8] R. L. Johnston, Dalton Transactions **22**, 4193 (2003).

[9] L. Vilhelmsen and B. Hammer, J. Chem. Phys. **141**, 044711 (2014).

[10] F. Curtis, X. Li, T. Rose, A. Vázquez-Mayagoitia, S. Bhattacharya, L. Ghiringhelli, and N. Marom, J. Chem. Theory Comput. **14**, 2246 (2018).

[11] M. Lazzeri and A. Selloni, Phys. Rev. Lett. **87**, 266105 (2001).

[12] F. Chuang, C. Ciobanu, C. Predescu, C. Wang, and K. Ho, Surf. Sci. **578**, 183 (2005).

[13] D. E. E. Deacon-Smith, D. O. Scanlon, C. R. A. Catlow, A. A. Sokol, and S. M. Woodley, Adv. Mater. **26**, 7252 (2014).

[14] 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).

[15] U. Martinez, L. B. Vilhelmsen, H. H. Kristoffersen, J. Stausholm-Møller, and B. Hammer, Phys. Rev. B **84**, 205434 (2011).

[16] A. M. van der Zande, P. Y. Huang, D. A. Chenet, T. C. Berkelbach, Y. M. You, G.-H. Lee, T. F. Heinz, D. R. Reichman, D. A. Muller, and J. C. Hone, Nat. Mater. **12**, 554 (2013).

[17] Z.-L. Li, Z.-M. Li, H. Y. Cao, J.-H. Yang, Q. Shu, Y. Y. Zhang, H. J. Xiang, and X. G. Gong, Nanoscale **6**, 4309 (2014).

[18] E. Flikkema and S. Bromley, J. Phys. Chem. B **108**, 9638 (2004).

[19] R. Ferrando, J. Jellinek, and R. L. Johnston, Chem. Rev. **108**, 845 (2008).

[20] I. Demiroglu, K. Yao, H. Hussein, and R. Johnston, J. Phys. Chem. C **121**, 10773 (2017).

[21] M. Aslan, J. Davis, and R. Johnston, Phys. Chem. Chem. Phys. **18**, 6676 (2016).

[22] J. B. A. Davis, S. L. Horswell, and R. L. Johnston, J. Phys. Chem. A **118**, 208 (2013).

[23] H. Zhai and A. Alexandrova, J. Phys. Chem. Lett. **9**, 1696 (2018).

[24] B. Zandkarimi and A. N. Alexandrova, WIREs Comput. Mol. Sci. **9**, e1420 (2019).

[25] M. L. Paleico and J. Behler, J. Chem. Phys. **153**, 054704 (2020).

[26] A. Oganov, C. Pickard, Q. Zhu, and R. Needs, Nat. Rev. Mater. **4**, 331 (2019).

[27] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).

[28] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Sci. Adv. **3**, e1603015 (2017).

[29] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[30] M. Valle and A. Oganov, Acta Crystallogr. Sect. A **66**, 507 (2010).

[31] A. P. Bartok, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).

[32] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).

[33] A. Shapeev, Multiscale Modeling and Simulation **14**, 1153 (2016).

[34] C. van der Oord, G. Dusson, G. Csányi, and C. Ortner, Mach. Learn.: Sci. Technol. **1**, 015004 (2020).

[35] V. L. Deringer, C. J. Pickard, and G. Csányi, Phys. Rev. Lett. **120**, 156001 (2018).

[36] V. L. Deringer, D. M. Proserpio, G. Csányi, and C. J. Pickard, Faraday Discuss. **211**, 45 (2018).

[37] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, J. Chem. Phys. **148**, 241733 (2018).

[38] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, Phys. Rev. Materials **3**, 023804 (2019).

[39] K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, J. Chem. Phys. **148**, 241727 (2017).

[40] C. Schran, K. Brezina, and O. Marsalek, J. Chem. Phys. **153**, 104105 (2020).

[41] J. S. Smith, R. Zubatyuk, B. T. Nebgen, N. Lubbers, K. Barros, A. Roitberg, O. Isayev, and S. Tretiak, Scientific Data **7**, 134 (2020).

[42] H. A. Doan, G. Agarwal, H. Qian, M. J. Counihan, J. Rodríguez-López, J. S. Moore, and R. S. Assary, Chem. Mater. **32**, 6338 (2020).

[43] E. Garijo del Río, J. J. Mortensen, and K. W. Jacobsen, Phys. Rev. B **100**, 104103 (2019).

[44] A. Denzel and J. Kästner, J. Chem. Phys. **148**, 094114 (2018).

[45] A. A. Peterson, J. Chem. Phys. **145**, 074106 (2016).

[46] O.-P. Koistinen, F. B. Dagbjartsdóttir, V. Ásgeirsson, A. Vehtari, and H. Jónsson, J. Chem. Phys. **147**, 152720 (2017).

[47] J. A. Garrido Torres, P. C. Jennings, M. H. Hansen, J. R. Boes, and T. Bligaard, Phys. Rev. Lett. **122**, 156001 (2019).

[48] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015).

[49] A. A. Peterson, R. Christensen, and A. Khorshidi, Phys. Chem. Chem. Phys. **19**, 10978 (2017).

[50] K. Miwa and H. Ohno, Phys. Rev. Materials **1**, 053801 (2017).

[51] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater. Sci. **140**, 171 (2017).

[52] I. Novikov, Y. Suleimanov, and A. Shapeev, Phys. Chem. Chem. Phys. **20**, 29503 (2018).

[53] R. Jinnouchi, F. Karsai, and G. Kresse, Phys. Rev. B **100**, 014105 (2019).

[54] Z. Ulissi, A. Medford, T. Bligaard, and J. Nørskov, Nat. Commun. **8**, 14621 (2017).

[55] S. Stocker, G. Csányi, K. Reuter, and J. Margraf, Nat. Commun. **11**, 5505 (2020).

[56] J. Xu, X.-M. Cao, and P. Hu, J. Chem. Theory Comput. **17**, 4465 (2021).

[57] H. Zhai, M.-A. Ha, and A. N. Alexandrova, J. Chem. Theory Comput. **11**, 2385 (2015).

[58] M. Todorović, M. Gutmann, J. Corander, and P. Rinke, npj Comput. Mater. **5**, 35 (2019).

[59] T. Yamashita, N. Sato, H. Kino, T. Miyake, K. Tsuda, and T. Oguchi, Phys. Rev. Materials **2**, 013803 (2018).

[60] Q. Tong, L. Xue, J. Lv, Y. Wang, and Y. Ma, Faraday Discuss. **211**, 31 (2018).

[61] K. Gubaev, E. Podryabinkin, G. Hart, and A. Shapeev, Comput. Mater. Sci. **156**, 148 (2019).

[62] M. Van den Bossche, J. Phys. Chem. A **123**, 3038 (2019).

[63] M. P. Lourenço, A. dos Santos Anastácio, A. L. Rosa, T. Frauenheim, and M. C. da Silva, Journal of Molecular Modeling **26**, 187 (2020).

[64] J. Roberts, J. R. S. Bursten, and C. Risko, Chem. Mater. **33**, 6589 (2021).

[65] M. Arrigoni and G. K. H. Madsen, npj Comput. Mater. **7**, 71 (2021).

[66] E. Musa, F. Doherty, and B. R. Goldsmith, Current Opinion in Chemical Engineering **35**, 100771 (2022).

[67] M. K. Bisbo and B. Hammer, Phys. Rev. Lett. **124**, 086102 (2020).

[68] L. R. Merte, M. K. Bisbo, I. Sokolović, M. Setvín, B. Hagman, M. Shipilin, M. Schmid, U. Diebold, E. Lundgren, and B. Hammer, Angew. Chemie Int. Ed. (2022), doi: 10.1002/anie.202204244.

[69] G. Dolgonos, B. Aradi, N. H. Moreira, and T. Frauenheim, J. Chem. Theory Comput. **6**, 266 (2010).

[70] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT, Cambridge, MA, 2005).

[71] H. L. Mortensen, S. A. Meldgaard, M. K. Bisbo, M.-Peter V. Christiansen, and B. Hammer, Phys. Rev. B **102**, 075427 (2020).

[72] M. S. Jørgensen, M. N. Groves, and B. Hammer, J. Chem. Theory Comput. **13**, 1486 (2017).

[73] M. Takagi, T. Taketsugu, H. Kino, Y. Tateyama, K. Terakura, and S. Maeda, Phys. Rev. B **95**, 184110 (2017).

[74] M.-P. V. Christiansen, H. L. Mortensen, S. A. Meldgaard, and B. Hammer, J. Chem. Phys. **153**, 044107 (2020).

[75] H. Wang, B. van Stein, M. Emmerich, and T. Back, in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE, New York, 2017), pp. 507–512.

[76] F. Häse, L. M. Roch, C. Kreisbeck, and A. Aspuru-Guzik, ACS Central Science **4**, 1134 (2018).

[77] M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, J. Phys. Chem. A **122**, 1504 (2018).

[78] M. A. Caro, Phys. Rev. B **100**, 024112 (2019).

[79] E. Kocer, J. Mason, and H. Erturk, AIP Adv. **10**, 015021 (2020).

[80] K. Terayama, T. Yamashita, T. Oguchi, and K. Tsuda, npj Comput. Mater. **4**, 32 (2018).

[81] S. Kaappa, E. G. del Río, and K. W. Jacobsen, Phys. Rev. B **103**, 174114 (2021).

[82] S. Kaappa, C. Larsen, and K. W. Jacobsen, Phys. Rev. Lett. **127**, 166001 (2021).

[83] W. An, N. Shao, S. Bulusu, and X. C. Zeng, J. Chem. Phys. **128**, 084301 (2008).

[84] M. K. Bisbo, Improving Atomistic Global Optimization With Machine Learning, Ph.D. thesis, Aarhus University, 2022.

[85] V. Meunier, A. G. Souza Filho, E. B. Barros, and M. S. Dresselhaus, Rev. Mod. Phys. **88**, 025005 (2016).

[86] A. Barnard and I. Snook, Modell. Simul. Mater. Sci. Eng. **19**, 054001 (2011).

[87] H. Chen, Y. Que, L. Tao, Y.-Y. Zhang, X. Lin, W. Xiao, D. Wang, S. Du, S. Pantelides, and H.-J. Gao, Nano Res. **10**, 3722 (2017).

[88] H. Tetlow, I. Ford, and L. Kantorovich, J. Chem. Phys. **146**, 044702 (2017).

[89] Y. Liu, L. Xu, L. Zhang, Z. Dong, S. Wang, and L. Luo, ACS Appl. Mater. Interfaces **12**, 52201 (2020).

[90] J. Shen, Y. Zhu, X. Yang, and C. Li, Chem. Commun. **48**, 3686 (2012).

[91] P. Lacovig, M. Pozzo, D. Alfè, P. Vilmercati, A. Baraldi, and S. Lizzit, Phys. Rev. Lett. **103**, 166101 (2009).

[92] B. Wang, X. Ma, M. Caffio, R. Schaub, and W.-X. Li, Nano Lett. **11**, 424 (2011).

[93] J. Gao, Q. Yuan, H. Hu, J. Zhao, and F. Ding, J. Phys. Chem. C **115**, 17695 (2011).

[94] Q. Yuan, J. Gao, H. Shu, J. Zhao, X. Chen, and F. Ding, J. Am. Chem. Soc. **134**, 2970 (2012).

[95] P. Wu, H. Jiang, W. Zhang, Z. Li, Z. Hou, and J. Yang, J. Am. Chem. Soc. **134**, 6045 (2012).

[96] J. Gao and F. Ding, J. Cluster Sci. **26**, 347 (2015).

[97] J. Klimeš, D. R. Bowler, and A. Michaelides, J. Phys.: Condens. Matter **22**, 022201 (2009).

[98] gofee.au.dk.