An introductory example of machine learning enhanced global optimization

Søren A. Meldgaard,[1] Esben L. Kolsbjerg,[1] and Bjørk Hammer[1, a)]
Department of Physics and Astronomy and Interdisciplinary Nanoscience Center (iNANO), Aarhus University, 8000, Aarhus, Denmark.

(Dated: 14 October 2019)

We show how to speed up global optimization of molecular structures using machine learning methods. To represent the molecular structures we introduce the *auto-bag* feature vector that combines: i) a local feature vector for each atom, ii) an unsupervised clustering of such feature vectors for many atoms across several structures, and iii) a count for a given structure of how many times each cluster is represented. During subsequent global optimization searches, accumulated structure-energy relations of relaxed structural candidates are used to assign local energies to each atom using supervised learning. Specifically, the local energies follow from assigning energies to each cluster of local feature vectors and demanding the sum of local energies to amount to the structural energies in the least squares sense. The usefulness of the method is demonstrated in basin hopping searches for 19-atom structures described by single- or double-well Lennard-Jones type potentials and for 24-atom carbon structures described by density functional theory (DFT). In all cases, utilizing the local energy information derived on-the-fly enhances the rate at which the global minimum energy structure is found.

I. INTRODUCTION

The field of atomic-scale structure search is crucial in a wide span of disciplines ranging from catalysis over material science to molecular biology. For an efficient search for the structural global minimum in an energy landscape of many dimensions it requires optimization techniques of global character. Such methods include random search[1], basin hopping[2], and evolutionary algorithms[3–11]. Common for these methods is that they facilitate an increased exploration of configuration space compared to e.g. molecular dynamics driven searches, that excel on the exploitational search in already identified funnels of the energy landscape.

First principles methods, that take no input from experiment and involve no empirical parameters, are the methods of choice for atomic-scale structure search. Notably, density functional theory (DFT) is being used for structure optimization, as it has proven highly accurate in reproducing observed structures, rationalizing experimental observations where the underlying structures were elusive[12,13], and in some cases even predicting new structures[14]. The high accuracy of DFT builds on solving the quantum mechanical problem of single-particle Hamiltonians for electrons in effective potentials and on representing the electrons through a self-consistently found all-electron spatial density. These elements of DFT must be attended for each energy and force evaluation during structural searches and make the method computational highly demanding.

Recently, machine learning (ML) procedures have been introduced to map the atomic interactions from the elaborate DFT framework to predefined functional expressions such as force fields[15], or to general fitting functions involving e.g. the local atomic environments[16–19] or the atomic bonds[20]. These ML-based methods demonstrate tremendous speed ups at only moderate loss of accuracy. Protocols have been established in which the accuracy is monitored while the ML-based potentials are being used in order to capture failure and re-adjust the ML models[21–24].

With the introduction of ML methods, a local energy concept often emerges, which is not present in DFT unless extra measures are taken[25]. A local energy concept is useful in the context of structure optimization as it opens for directing the structural search more efficiently by focusing on improving the high energy regions of structures while preserving structural arrangements in low energy regions[26].

In the present work, we add to the current developments in the use of ML methods in chemical physics by introducing a simple means of representing atomic structures with the *auto-bag* feature vector and by formulating a simple regression framework, that enables the extraction of local atomic energies. The method is demonstrated to work with very little input data at the DFT level, which potentially makes it interesting when probing new systems, where DFT data is scarce.

The paper has been written with newcomers to ML-enhanced DFT-based structural searches in mind. It thus starts by describing the method in the context of a basin hopping-enabled structural search for the optimum 2D structure using only classical Lennard-Jones (LJ) potentials for the atomic interactions. The basin hopping method and the LJ potentials are sufficiently simple, that students and researchers with no prior training in the field may readily code these elements and reproduce the presented results. Also, the use of the LJ potentials has the added benefit that the ML-enabled local energies can be directly compared to the true LJ-based local energies, which would not be possible in a DFT-framework. The paper proceeds with a demonstration of the applicability of method when used in conjunction with DFT in a

a)Electronic mail: hammer@phys.au.dk.

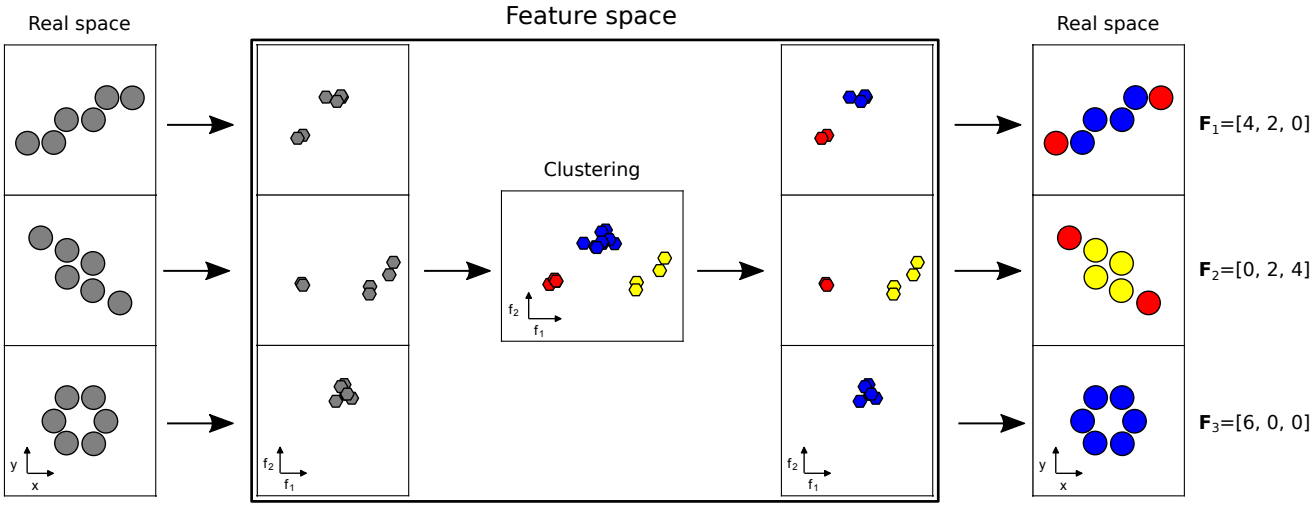

FIG. 1. Schematic illustration of the method for extracting the *auto-bag* feature vector. From left to right: Structures with atoms (grey disks) in various local environments. Atomic feature vectors (grey hexagons) are extracted. The feature vectors for all structures are clustered (colored hexagons). Each structure may subsequently be illustrated either in feature space (structure-wise plots of colored hexagons) or in real space (colored disks). By counting the abundance of members of each cluster in a given structure, I, the *auto-bag* feature vector, $\mathbf{F}_I$, is obtained.

search for the optimum 2D-cluster shape of a 24 carbon atom cluster.

II. METHOD

A. Representation

When machine learning methods are introduced, the first concern is a proper representation of the data on which the learning is made. Representation is highly domain specific. In text processing, the *bag-of-words* vector, that counts the occurrence of known words and neglects grammar and word order, is often used to map an entire text to a simple vector of integers. Likewise, performing customer segmentation in a retail or web shop, a customer (as identified by a credit card or a login) may be represented by his or her historical spending in various predefined product categories in what we may dub a *bag-of-spendings* vector.

In the chemical physics domain, structures are naturally described by a list of atomic identities and corresponding cartesian coordinates, and – for space filling matter – by the super cell vectors. However, such a representation is not adequate for machine learning[27], since it is not invariant to translations, rotations, or permutation of identical atoms – operations that do otherwise leave the physical properties of the compounds unchanged.

Several representations have been proposed that deal with this deficit. One of the simplest such representations is one in which all interatomic separations are evaluated, sorted in ascending order, and kept in "bags" of AA, AB, BB, ... bond-type, where A, B, ... are the atomic identities. The resulting *bag-of-bonds* vector thus represents the entire structure[28]. Several other global descriptors have been proposed, including the fingerprint feature vector[29] and the coulomb matrix representation[30,31]. Presently, much work in the field relies on local feature vectors such as the Behler-Parrinello symmetry functions[16], SOAP[32] and others[33–35]. These feature vectors describe the local environment of each atom in a given structure and represent entire structures by the collection of such atomic feature vectors.

To keep the complexity of the method introduced in this work at a minimum and to construct a method that requires only little data to provide useful predictions, we propose the *auto-bag* feature vector representation, which will be detailed in the following. The method has two elements, (i) the automated identification of bags of atomic environments, and (ii) the representation of a structure as the count of the occurrence of those environments in that given structure.

Fig. 1 presents the build-up of the *auto-bag* feature vector. Its starts by evaluating atomic feature vectors for each atom in a collection of structures. Next, the feature vectors are grouped in "bags", and finally for each structure the abundance of group members may be counted and collected as a vector of integers. The groups act as the classification in line with the "known words", the "predefined product categories", and the "AA, AB, and BB bond-types" in the bag-of-words, bag-of-spendings, and bag-of-bonds methods discussed above. However, by using a machine learning technique, *clustering*, to identify the groups, they need not be predefined, and hence the naming: the *auto-bag* feature vector.

In Fig. 1, the *local* feature vector describing each atomic environment is considered two-dimensional for illustration purposes, but it may in fact be chosen as sim-

ple as a single number, e.g.:

$$f_i(\mathbf{r}_1, \ldots, \mathbf{r}_N) = f_i^\alpha \tag{1}$$

where N is the number of atoms and f_i^α is a function describing the local density around atom i within some cutoff distance (α being a label). A simple example of such a local feature is the radial symmetry function as given by Eq. (A1) in the appendix. However, a radial symmetry function will not uniquely define the local environment. It will for instance not be able to differentiate between configurations with two atoms located a distance r from atom i but with different bond angles. To encapsulate angular information the local feature vector can be expanded to

$$\mathbf{f}_i(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \begin{bmatrix} f_i^\alpha \\ f_i^\beta \end{bmatrix} \tag{2}$$

where f_i^β is an angular symmetry function given by Eq. (A2) in the appendix (β being a label). For a even more detailed description of the local environment several radial and angular symmetry functions can be used to yield a higher dimensional local feature vector as described in the appendix.

The grouping of local feature vectors illustrated in the middle of Fig. 1 must be done using an unbiased method in order for the method to work autonomously. This is possible using unsupervised machine learning techniques known as clustering methods. A clustering method acts to identify relations and propose a categorization of data without prior definition of the categories – hence the adjective "unsupervised". Many clustering schemes have been proposed and some have even been demonstrated to enable speed up of structural search[36,37]. In the present context, we have found the simple clustering method, the K-means algorithm[38], sufficient to fulfill the needs of establishing the auto-bag feature vector. K-means takes the number of desired clusters as input and is non-deterministic in that its random initialization may cause new results in repeated uses on the same data. This property is not a problem in connection with structure optimization where a certain level of stochastic behavior may even be desirable[37]. To stabilize the method using K-means, we did, however, use the K-means++ initialization[39] to achieve a reasonable local optimum and prevent empty clusters.

Once the clustering is done, a structure will be described by a global feature vector:

$$\mathbf{F}(\mathbf{f}_1, \ldots, \mathbf{f}_N) = [n_1, \ldots, n_c, \ldots, n_C], \tag{3}$$

where n_c is the number of atoms in cluster c with C clusters in total. The global feature vector respects invariance to permutation of identical atoms, and inherits translational and rotational symmetry from the local feature vectors.

B. Local energies

We now propose to parameterize the system energy as

$$E(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_{i=1}^{N} \epsilon(\mathbf{f}_i) \approx \sum_{i=1}^{N} \varepsilon_{c(i)} = \sum_{c=1}^{C} n_c \varepsilon_c, \tag{4}$$

where $\epsilon(\mathbf{f}_i)$ is the local energy of atom i as defined by its local feature vector, ε_c is a common local energy assigned to all members of cluster c, and $c(i)$ is the cluster index for atom i. We note that writing the total energy of a structure as a sum of local energies is an approximation[16,26]. By forcing the local energies to be identical for all members of a given cluster, the approximation becomes even more severe. However, choosing cluster-wise local energies means that the method has fewer free parameters and that extraction of meaningful energies can be done with the simple method that follows below.

Introducing an index, I, that enumerates the structures we have the relation:

$$E_I = \sum_{c=1}^{C} n_{Ic} \varepsilon_c \tag{5}$$

defining the unknown local energies ε_c as a function of the global feature-energy relations, $(\mathbf{F}_I, E_I) = ([n_{I1}, \ldots, n_{Ic}, \ldots, n_{IC}], E_I)$. By observing multiple structures a matrix problem emerges:

$$\begin{bmatrix} n_{11} & \ldots & n_{1C} \\ \vdots & \ddots & \vdots \\ n_{S1} & \ldots & n_{SC} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_C \end{bmatrix} = \begin{bmatrix} E_1 \\ \vdots \\ E_S \end{bmatrix}, \tag{6}$$

where n_{Ic} is number of atoms in cluster c, for structure I, with a total of S structures observed. Eq. (6) can be restated as

$$\mathbf{X}\varepsilon = \mathbf{E}, \tag{7}$$

which ordinary least squares estimate the solution to by minimizing

$$\mathcal{E} = ||\mathbf{X}\varepsilon - \mathbf{E}||^2, \tag{8}$$

i.e the sum of squared residuals. Eq. (8) is minimized by

$$\varepsilon = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{E}. \tag{9}$$

However, depending on the rank of $\mathbf{X}$, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ can potentially be singular. To overcome this problem ridge regression is used by altering Eq. (8) to

$$\mathcal{E} = ||\mathbf{X}\varepsilon - \mathbf{E}||^2 + \lambda||\varepsilon||^2, \tag{10}$$

where λ is a positive parameter chosen by the user. Eq. (10) is minimized by

$$\varepsilon = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{E}, \tag{11}$$

where $\mathbf{I}$ is the identity matrix. Eq. (11) always exists and has the added benefit of preventing overfitting by regularization on the free parameters ε.

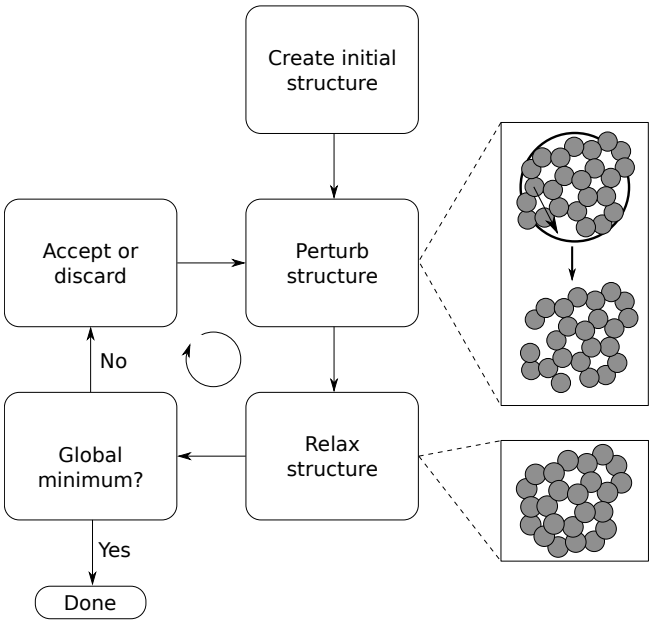

FIG. 2. Illustration of the basin hopping framework.

C. Optimization

To demonstrate the applicability of the auto-bag feature and the local energies we will derive and use these during global optimization runs. Fig. 2 illustrates the layout of a basin hopping search for the global minimum energy structure of a set of atoms.

First, a random structure is initiated and relaxed into the nearest local minimum energy structure according to the atomic forces. Next, the structure undergoes some perturbative modification. Many strategies may exist for this step. In the present context we have chosen a simple procedure that we shall refer to as the *fireworks perturbation*. With this procedure, a random number, N_{at}, of atoms are repositioned uniformly within a disk centered at the center of mass of the structure (Fig. 2, black ring). The radius of the disk is a parameter analogous to the rattle distance in a rattle mutation. N_{at} follows the normalized geometric series:

$$P(x) = \left(\frac{1}{2}\right)^x \Bigg/ \sum_{x=1}^{N} \left(\frac{1}{2}\right)^x , \tag{12}$$

where N is the total number of atoms. With Eq. (12), $P(1) \approx \frac{1}{2}$, $P(2) \approx \frac{1}{4}$, ..., meaning that with about 50% likelihood only one atom is repositioned, with about 25% likelihood two atoms are repositioned, and so on. When the optimization is run without use of the local energies, the atoms are chosen randomly. However, when the local energies are exploited, the N_{at} atoms are drawn one by one with a likelihood that also follows the normalized geometric series. I.e. every time an atom is chosen, there will be about 50% chance that it is the most unstable stable atom not yet chosen, about 25% chance that it is the second most unstable and so on.

Once the structure has undergone the perturbation, it is relaxed according to the forces and a local minimum energy structure is identified. This new structure replaces the previous structure according to the Metropolis-Hastings criterion:

$$A = \min\{1, \exp[\beta(E_{k-1} - E_k)]\}, \tag{13}$$

where A is the probability of acceptance, $\beta = 1/k_B T$, k_B is the Boltzmann constant and T is a temperature parameter. Here E_k is the potential energy of the newly found structure and E_{k-1} is the potential energy of the previous structure. If the structure is accepted it serves as the starting point of the next perturbation, otherwise it is discarded.

III. LENNARD-JONES SYSTEM

As a first demonstration of the method, we consider a 2D structure of 19 atoms described by the classical Lennard-Jones (LJ) interaction potential. Using this simple potential has several benefits, it is easily programmable meaning that the reader may code it and reproduce our results. Local energies can be uniquely assigned to LJ atoms meaning that the approximate machine learned local energies following our method above may be benchmarked, and, importantly, the use of the LJ potential is fast which allows for fast testing and the production of converged statistics on the efficiency.

The LJ pair-potential is given by:

$$V(r) = \varepsilon_0 \left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6} \right]. \tag{14}$$

ε_0 sets the energy scale and further turns out to be the depth of the well in the pair-potential, while r_0 sets the length-scale and coincides with the equilibrium distance of the LJ dimer. Upon training the model by solving Eq. (11) for various data sizes and with varying number of clusters, the error on predicting local energies of structures not included in the training set is seen in Fig. 3. The cutoff radius is chosen as $r_c = 2\,r_0$ and a single radial Behler-Parrinello feature vector, Eq. (A1), with $r_s = 0$ and $\eta = 0.05$ is employed. Only relaxed structures are used for both training and testing, resulting in energies ranging from $-22\,\varepsilon_0$ to $-45\,\varepsilon_0$. All relaxations and perturbations where confined to a plane causing the resulting structures to become strictly 2D.

Figure 3 shows that including more data in the training generally leads to lower mean absolute errors (MAEs). However, with only few clusters, $C = 5$, the possible improvement of the MAE stagnates when utilizing around 100 different structures. This behavior is as expected since restricting the number of local environments will naturally provide a lower bound on the error if less clusters than unique local environments are used. Applying more clusters improves the lower bound on the error at

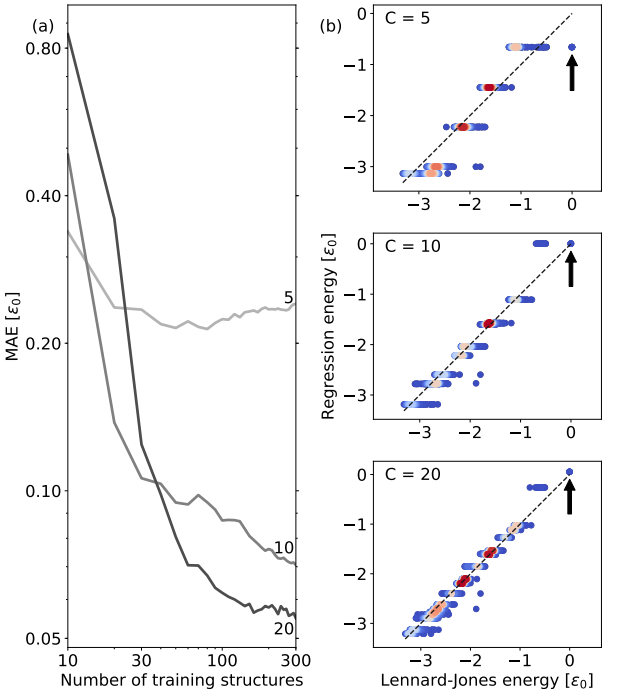

FIG. 3. (a): Local energy learning curve for 19 atoms using varying number of clusters. The curve shows the mean absolute error (MAE) on the predicted energies of a set of test structures not available during training. (b): Prediction of local energies for unseen structures upon training on 300 configurations with 5, 10 or 20 clusters. The color indicates the density of data points. The three black arrows indicate atoms with no neighbors.

the expense of an increased error for small training samples due to additional free parameters. Increasing the number of clusters prevents dissimilar atoms from being forced into the same cluster and thus leads to a more accurate energy prediction. This is seen in the transition from 5 to 10 to 20 clusters where atoms not participating in any chemical bonds (see arrows in Fig. 3) initially belong to a non-zero energy cluster, then a zero energy cluster with non-zero energy atoms and finally a zero energy cluster with only zero energy atoms.

A further inspection of the local energies is seen in Fig. 4 where several structures have been colored according to predicted local energy from the fully trained model with 10 clusters. As the LJ energy correlates with the coordination of atoms and all atoms are placed approximately equidistantly to their nearest neighbors it is easily verified that the order of the local energies is correct for the global minimum shown show in Fig. 4 (h). In Fig. 4 (f) an indication of the applicability of local energies is seen as the only misplaced atom is predicted to be the most unstable. An inspection of the atom just above the most unstable atom reveals additional insight into the model as this atom, despite having four neighbors, is more unstable than the other atoms with four neighbors. The same tendency is observed in Fig. 4 (b) and Fig. 4 (d) where in all cases the most stable four-neighbor atom has three or more second-nearest neighbors, whereas the more unstable atom only has two. For atoms with two and five neighbors the same effect is observed. Thus the ML model is able to correctly order very similar local environments.

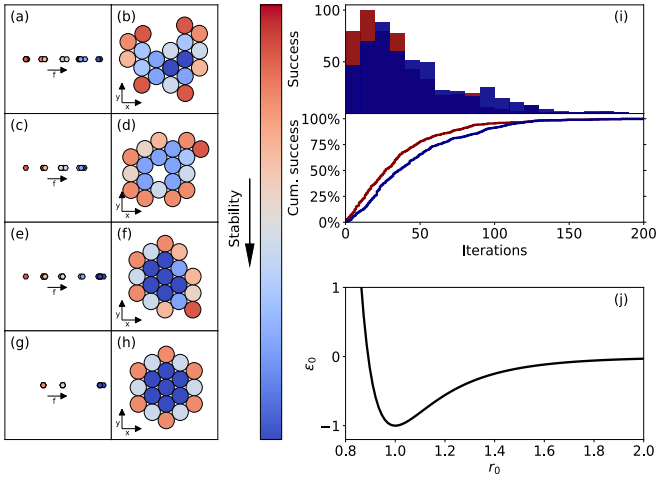

FIG. 4. (a)-(h): Lennard-Jones structures and feature vectors colored to express the local energies based on ML model trained on 300 random structures. (i): Success curve for finding the global minimum without machine learning (blue) and with (red). (j): Lennard-Jones potential.

Having shown that local energies are possible to learn it remains to be seen if they can be utilized in a optimization setting. Hence a ML model with 10 clusters is trained on-the-fly during a basin hopping (BH) search and used to predict the local energies. 10 clusters are chosen based on optimizing the performance of the search as early as possible, while also obtaining an acceptable final error. In each BH step, a number of atoms according to Eq. (12) are perturbed. The atoms are chosen randomly in the benchmark run – or dependent on their local energies according to the ML model as detailed in the discussion of Eq. (12). The radius of the disk in the perturbation is $3\,r_0$. Due to the small number of local minima for the LJ system the search is executed at $T = 0$K. From repeating the search 500 times, the cumulative success is seen in Fig. 4 (i). We stress that each run starts with an untrained ML model. Only a minor increase in the success rate is observed, presumably due to the low complexity of the system. To test this hypothesis the search is repeated for a double well LJ system:[40]

$$V(r) = \varepsilon_0 \left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6} - \exp\left(-\frac{(r-1.7r_0)^2}{2\sigma^2}\right)\right], \tag{15}$$

with $\sigma^2 = 0.02\,r_0^2$. The global minimum is identical to that of the ordinary LJ potential and the same perturbation as before is used. As the pair-potential has two

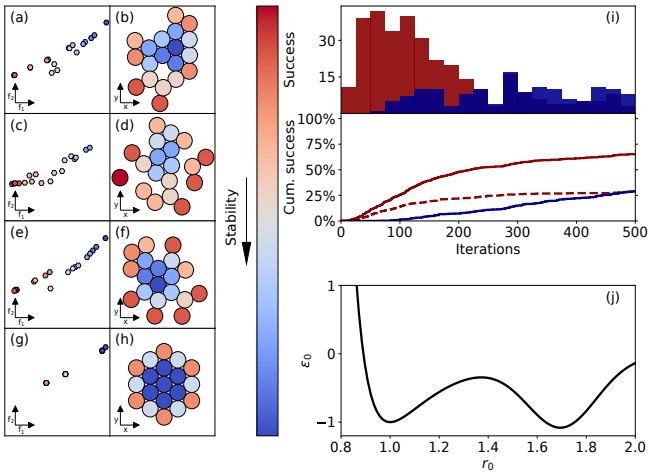

FIG. 5. (a)-(h): Double-well Lennard-Jones structures and feature vectors colored to express the local energies based on ML model trained on 300 random structures. (i): Success curve for finding the global minimum without machine learning (blue) and with 1D feature (Eq. (1), red dashed) and 2D feature (Eq. (2), red). (j): Double-well Lennard-Jones potential.

minima, more structures may evolve as meta-stable local minima in configuration space, and as a consequence, finding the global minimum becomes a harder ordeal. As a measure to increase the success rate of finding the global minimum energy structure, the BH runs were performed at a finite temperature of $k_BT = 0.1\,\varepsilon_0$. Yet, the search remains a challenge as testified to by Fig. 5 where the benchmark run now takes 500 BH iterations to find the global minimum with $\approx 20\%$ likelihood, while it took about 200 BH iterations to find it with almost 100% certainty with the standard LJ pair-potential.

As the 19 atom structure represents a harder problem with the double well LJ pair-potential, it becomes easier to demonstrate the beneficial effects of the machine learning approach introduced in this work. The dashed red line of Fig. 5 shows how greater success is achieved when local energies are derived based on a one-dimensional feature vector and exploited in the BH search. However, even more striking is the success rates achieved when a two-dimensional feature vector is employed, as shown by the full red line in Fig. 5. Now, the $\approx 20\%$ success level in finding the global minimum is achieved after a mere 100 BH iterations, representing a five-fold rate increase over the benchmark run. The two-dimensional feature vector contains an angular component of the Behler-Parrinello type with parameters: $\eta = 0.005$, $\lambda = 1$ and $\xi = 1$. Extending the feature vector with an angular component clearly outperforms the one-dimensional one. We attribute this to a more rich variety of local environments being present in the relaxed structures of the double well potential compared to the standard Lennard-Jones potential. The increased performance seen with Lennard-Jones type potentials motivates a search using a quantum mechanical energy expression as with DFT, where the both the energy landscape and the local environments can be much more complex.

IV. DENSITY FUNCTIONAL THEORY SYSTEM

Owing to the non-local Hamiltonian of quantum systems it is uncertain whether useful local energy information can be extracted. To investigate this we are using DFT to describe a 3D system of 24 carbon atoms utilizing a basis of linear combination of atomic orbitals for computational efficiency as available in the GPAW[41,42] code with ASE[43] managing the atomic structures and optimization. To describe the exchange and correlation effects the PBE functional[44] was chosen. The computational cell was constructed with no atoms closer than 6 Å from the non-periodic cell boundaries. The optimization task is conducted using the auto-bag feature based on 10 clusters. The local feature vector is expanded to 13 dimensions using standard Behler-Parrinello symmetry functions with cutoff radius $r_c = 2$ Å, $r_s = 0$, and default remaining parameters taken from ref.[45] (see Table I in the appendix). To further prevent stagnation a parallel tempering scheme[46] is employed where four basin hopping searches at different temperatures are performed simultaneously. Every five iterations temperature swaps between simulations with adjacent temperatures (i, j) are attempted and accepted with probability

$$A = \min\{1, \exp[(\beta_k^i - \beta_k^j)(E_k^i - E_k^j)]\}, \qquad (16)$$

where $\beta_k^i = 1/k_BT_k^i$ and E_k^i is the potential energy as before. The subscript refers to the structure at iteration k and the superscript is an index on the parallel runs. Stagnated structures will then eventually acquire a higher temperature allowing them to escape local minima. Temperatures are chosen as $k_BT = [0.200\,\text{eV}, 0.293\,\text{eV}, 0.425\,\text{eV}, 0.620\,\text{eV}]$, keeping a constant ratio between adjacent temperatures as suggested in the literature[47]. The temperatures are chosen to span both low temperatures allowing for exploitation as well as high temperatures for efficient exploration. The highest temperature is chosen to give a 20% chance of accepting a structure 1 eV higher than the current energy, and the lowest temperature such that almost only lower energy structures are accepted. In order to evaluate the efficiency of the local energies a full parallel tempering minimum search is conducted. In this work we presume that the D_{6h} structure is the global minimum. To direct our search towards planar structures the same perturbation as for the two-dimensional LJ system is used but with a disk of radius 4 Å. While the perturbation action was 2D, structural relaxation was done without any constraints leading to the structure becoming quasi-2D and occasionally 3D. As a benchmark the same parallel tempering algorithm is run with atoms to be perturbed chosen stochastically.

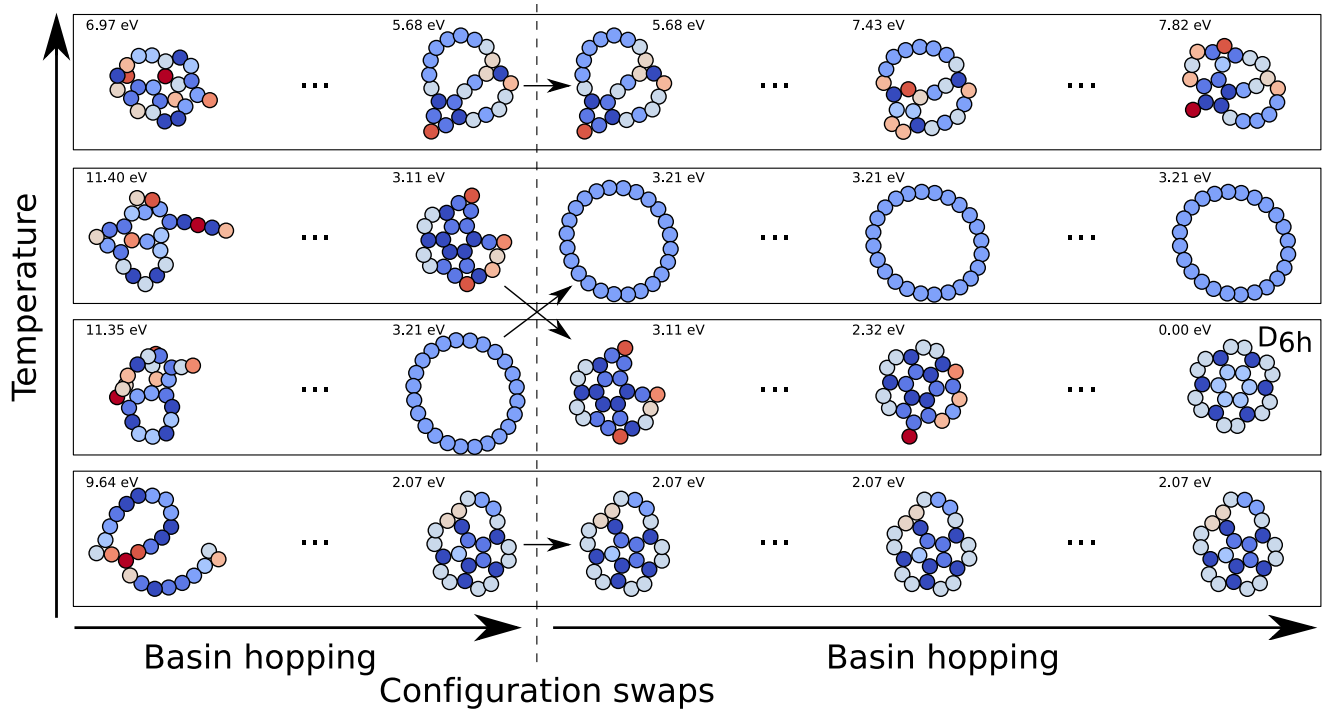

FIG. 6. Initial, intermediate and final structures found by a parallel tempering search with local energies depicted as the model predicted them on-the-fly. Configurational swaps were made possible for every five basin hopping iterations, but for clarity only one such swap is shown. Likewise most basin hopping steps are omitted. The total length of the run was 90 iterations. DFT based total energies are given relative to the presumed global minimum.

In Fig. 6 a parallel tempering run in search for C_{24} structures with DFT potential energy is illustrated. Structures are shown at selected iterations, and the swapping action is shown at one given iteration. The structures and energies reported show how the highest temperature run remains agile and keeps exploring new structures through the run, while the lower temperature runs exploit found structures and perform refinements (or remain stuck). Eventually, the run with the second lowest temperature identifies the presumed global minimum energy structure (D_{6h} with $E \equiv 0.00$ eV) and the calculation was stopped. Atoms are colored according to the ML model energy prediction at the given time of the search. Since the ML model is refined on-the-fly, as more training data is accumulated, it is not possible to compare colors for different iterations, especially not for the initial and earliest iterations. However, in general it is observed that atoms pointing out of the structures are drawn in red or reddish colors meaning that they are the more unstable atoms. This is for instance seen in the structure prior to the global minimum. Here the unsaturated carbon atom is shown to be extremely unstable. In the same structure it is also observed that one 5-membered and 7-membered ring contain unstable atoms, whereas 6-membered rings are shown to be stable. The other 5-membered ring of that structure is composed of more stable atoms according to the modeled local energies. This is an effect of an atom in this ring binding to the low coordinated high energy atom (colored dark red) and shows how the autobag feature captures well the details in the local environments of atoms.

The cumulative success for 40 runs is seen in Fig. 7 displaying a convincing boost in success rate for the ML enhanced run. Note the improvement in cumulative success already in the beginning of the run, demonstrating the limited amount of data necessary to achieve reasonable local energies. In order to reach 50% success the ML enhanced approach required 81 attempts which is 69 iterations less than the ordinary algorithm that required 150 iterations. Using the computationally inexpensive machine learning model described in this work thus produced a speed up factor of almost 2 when applied to a full atomic-scale structural search within a DFT setting.

V. SUMMARY

In this work we have introduced the *auto-bag* feature vector that combines a local feature vector for each atom in a structure, unsupervised machine learning (K-means) to establish clusters of such local feature vectors, and supervised machine learning (ridge regression) to extract atomic energies. The method was first demonstrated to be capable of extracting the local energies for a pairwise classical potential of the Lennard-Jones form, where the local energies are well-defined. Next, these machine

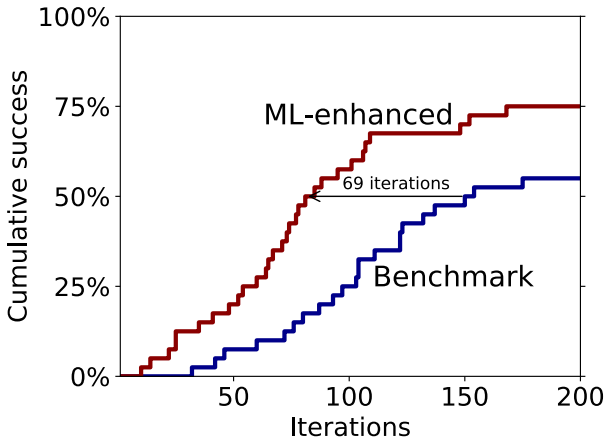

FIG. 7. Cumulative success rate over 40 independent parallel tempering search runs for C_{24} described with DFT. The blue curve represents the benchmark situation, where perturbed atoms are picked randomly. The red curve shows the results when perturbed atoms are picked according to their machine learned local energies, Eq. (11), using the auto-bag feature. Stagnation in fullerene type 3D structures more stable than the D_{6h} 2D structure occurred rarely (2.5% and 5% for benchmark and ML runs, respectively) and was counted as failure.

learned local energies were used to speed up the search for the global minimum energy structure of 19 atoms described by the standard Lennard-Jones potential or a more challenging double well Lennard-Jones potential. Finally, the methodology was applied to a quantum mechanical potential, density functional theory, description of structures of 24 carbon atoms. Here the local energies might be ill-defined, yet our results show that the stochastic search for the global minimum energy structure using the method of parallel tempering may be sped up considerably when perturbing in the basin hopping steps preferentially atoms that are predicted to be more unstable. The elements of the method are rather simple and are expected to show similar behavior for structural searches within chemical physics in general, including multi-component systems, molecules, nanoparticles, and solids.

ACKNOWLEDGMENTS

Grants from VILLUM FONDEN (Investigator grant, project number 16562) and the Danish Research Council (0602-02566B) have supported this research.

Appendix A: Local feature vector

To describe an atomic environment we use the symmetry functions proposed by Behler and Parrinello[16] which ensure rotational and translational invariance. The feature vector of atom i is composed of pairwise- and triple-atom interactions given as

$$f_i^{\alpha}(\mathbf{r}_1,\ldots,\mathbf{r}_N)=\sum_{j\neq i} e^{-\eta(r_{ij}-r_s)^2/r_c^2} f_c(r_{ij}), \qquad \text{(A1)}$$

and

$$f_i^{\beta}(\mathbf{r}_1,\ldots,\mathbf{r}_N)=2^{1-\xi}\sum_{j,k\neq i(j\neq k)}(1-\lambda\cos\phi_{ijk})^{\xi} \times e^{-\eta(r_{ij}^2+r_{ik}^2+r_{jk}^2)/r_c^2} f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk}), \qquad \text{(A2)}$$

respectively. Here j and k denotes the index of other atoms, r_{ij} is the distance between atom i and atom j, ϕ_{ijk} is the valence angle between atom i, j and k, centered at atom i and η, r_s, ξ and λ are parameters. By using multiple sets of parameters one can achieve a detailed description of the local environment, resulting in a high-dimensional local feature vector. See ref.[48] for more information on choosing the parameters. The interactions are only accounted for within a sphere of radius r_c by the use of the cutoff function

$$f_c(r)=\begin{cases}0.5(1+\cos\pi r/r_c), & \text{if } r\leq r_c\\ 0 & \text{if } r\geq r_c\end{cases} \qquad \text{(A3)}$$

which is a smoothly decaying function approaching zero at $r=r_c$. For the 13-dimensional feature vector the parameters can be seen in Table I.

TABLE I. Behler-Parrinello parameters

Radial symmetry functions							
η	0.05	2	4	8	20	40	80
r_s	0						

Angular symmetry functions			
ξ	1	2	4
λ	1	-1	
η	0.005		

[1] C. J. Pickard and R. J. Needs, Journal of Physics: Condensed Matter **23**, 053201 (2011).
[2] D. J. Wales and J. P. K. Doye, The Journal of Physical Chemistry A **101**, 5111 (1997).
[3] B. Hartke, The Journal of Physical Chemistry **97**, 9973 (1993).
[4] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[5] R. L. Johnston, Dalton Trans. (2003).
[6] A. N. Alexandrova and A. I. Boldyrev, Journal of Chemical Theory and Computation **1**, 566 (2005).
[7] A. R. Oganov and C. W. Glass, The Journal of Chemical Physics **124**, 244704 (2006).
[8] J. Marques and F. Pereira, Chemical Physics Letters **485**, 211 (2010).
[9] S. Bhattacharya, S. V. Levchenko, L. M. Ghiringhelli, and M. Scheffler, Phys. Rev. Lett. **111**, 135501 (2013).
[10] L. B. Vilhelmsen and B. Hammer, Phys. Rev. Lett. **108**, 126101 (2012).
[11] L. B. Vilhelmsen and B. Hammer, The Journal of Chemical Physics **141**, 044711 (2014).

[12]M. Lazzeri and A. Selloni, Phys. Rev. Lett. **87**, 266105 (2001).
[13]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).
[14]J. Lv, Y. Wang, L. Zhu, and Y. Ma, Nanoscale **6**, 11692 (2014).
[15]V. Botu, R. Batra, J. Chapman, and R. Ramprasad, The Journal of Physical Chemistry C **121**, 511 (2017).
[16]J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[17]T. L. Pham, H. Kino, K. Terakura, T. Miyake, and H. C. Dam, The Journal of Chemical Physics **145**, 154103 (2016).
[18]V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017).
[19]L. Zhang, J. Han, H. Wang, R. Car, and W. E, Phys. Rev. Lett. **120**, 143001 (2018).
[20]K. Yao, J. E. Herr, S. N. Brown, and J. Parkhill, The Journal of Physical Chemistry Letters **8**, 2689 (2017).
[21]S. Q. Wu, M. Ji, C. Z. Wang, M. C. Nguyen, X. Zhao, K. Umemoto, R. M. Wentzcovitch, and K. M. Ho, Journal of Physics: Condensed Matter **26**, 035402 (2014).
[22]A. A. Peterson, The Journal of Chemical Physics **145**, 074106 (2016).
[23]M. Van den Bossche, H. Grönbeck, and B. Hammer, Journal of Chemical Theory and Computation **14**, 2797 (2018).
[24]E. L. Kolsbjerg, A. A. Peterson, and B. Hammer, Phys. Rev. B **97**, 195424 (2018).
[25]M. Yu, D. R. Trinkle, and R. M. Martin, Phys. Rev. B **83**, 115113 (2011).
[26]X. Chen, M. S. Jørgensen, J. Li, and B. Hammer, Journal of Chemical Theory and Computation **83** (2018).
[27]B. Huang and O. A. von Lilienfeld, The Journal of Chemical Physics **145**, 161102 (2016).
[28]K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, Journal of Chemical Theory and Computation **9**, 3404 (2013).
[29]A. R. Oganov and M. Valle, The Journal of Chemical Physics **130**, 104504 (2009).
[30]M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[31]J. E. Moussa, Phys. Rev. Lett. **109**, 059801 (2012).
[32]A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013).
[33]G. Ferré, T. Haut, and K. Barros, The Journal of Chemical Physics **146**, 114107 (2017).
[34]O. T. Unke and M. Meuwly, The Journal of Chemical Physics **148**, 241708 (2018).
[35]T. D. Huan, R. Batra, J. Chapman, S. Krishnan, L. Chen, and R. Ramprasad, npj Computational Materials **3**, 37 (2017).
[36]M. S. Jørgensen, M. N. Groves, and B. Hammer, Journal of Chemical Theory and Computation **13**, 1486 (2017).
[37]K. H. Sørensen, M. S. Jørgensen, A. Bruix, and B. Hammer, The Journal of Chemical Physics **148**, 241734 (2018).
[38]S. Lloyd, IEEE Transactions on Information Theory **28**, 129 (1982).
[39]D. Arthur and S. Vassilvitskii, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (2007) p. 1027.
[40]M. Rechtsman, F. Stillinger, and S. Torquato, Phys. Rev. E **73**, 011406 (2006).
[41]J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).
[42]J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, Journal of Physics: Condensed Matter **22**, 253202 (2010).
[43]A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, Journal of Physics: Condensed Matter **29**, 273002 (2017).
[44]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[45]A. Khorshidi and A. A. Peterson, Computer Physics Communications **207**, 310 (2016).
[46]R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[47]D. A. Kofke, The Journal of Chemical Physics **117**, 6911 (2002).
[48]J. Behler, The Journal of Chemical Physics **134**, 074106 (2011).