



Machine-learning-enabled optimization of atomic structures using atoms with fractional existenceCasper Larsen ¹, Sami Kaappa,^{1,2} Andreas Lyngbe Vishart,^{3,4} Thomas Bligaard,^{3,4} and Karsten Wedel Jacobsen ¹¹*CAMD, Department of Physics, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark*²*Computational Physics Laboratory, Tampere University, P.O. Box 692, FI-33014 Tampere, Finland*³*CatTheory, Department of Physics, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark*⁴*ASM, Department of Energy Conversion and Storage, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark*

(Received 24 November 2022; accepted 2 June 2023; published 15 June 2023)

We introduce a method for global optimization of the structure of atomic systems that uses additional atoms with fractional existence. The method allows for movement of atoms over long distances bypassing energy barriers encountered in the conventional position space. The method is based on Gaussian processes, where the extrapolation to fractional existence is performed with a vectorial fingerprint. The method is applied to clusters and two-dimensional systems, where the fractional existence variables are optimized while keeping the atomic positions fixed on a lattice. Simultaneous optimization of atomic coordinates and existence variables is demonstrated on copper clusters of varying size. The existence variables are shown to speed up the global optimization of large and particularly difficult-to-optimize clusters.

DOI: [10.1103/PhysRevB.107.214101](https://doi.org/10.1103/PhysRevB.107.214101)**I. INTRODUCTION**

The atomic-scale structure is of critical relevance to the physical and chemical properties of materials and nanoparticles. In the low-temperature limit, the most stable atomic configuration is found by minimizing the total energy, but the optimization problem is difficult because of many metastable states, and in many cases, the total energy evaluations are computationally time-consuming.

To address these problems several algorithms of automated structure prediction have been proposed [1], including random searches [2], genetic searches [3–6], basin hopping [7], and particle swarm optimizations [8]. Central to most of these machine learning models is that they rely on carrying out large numbers of time-consuming calculations with density functional theory (DFT) or other quantum chemistry methods. Gaussian processes have been shown to be effective in constructing fast surrogate potential energy surfaces (PESs) [9,10], which can be explored by random searching and updated by a Bayesian search machine learning model, as demonstrated with the so-called global optimization with first-principles energy expressions (GOFEE) algorithm in Ref. [11]. This methodology is generalized to include training on forces in the BEACON (Bayesian Exploration of Atomic Configurations for Optimization) code [12]. GOFEE can decrease the number of energy evaluations necessary to find the global minimum by up to several orders of magnitude [13].

The efficiency of random searching can be improved by inclusion of hyperdimensions [14], which make it possible to circumvent barriers in configuration space. However, the energy function has to be defined for the extra hyperdimensions, and it is not clear how to do this for potential energy surfaces based on quantum mechanical calculations.

An alternative way to increase the dimensionality of configuration space and circumvent barriers is to interpolate between chemical elements (“ICE”) as implemented in the

ICE-BEACON code [15]. Here, additional dimensions are introduced so that an atom can be a fractional mixture of two chemical elements. In this paper, we apply the idea of expanded dimensionality in a different way by introducing extra variables which allow the atoms to have partial existence. The idea is that additional atoms of fractional existence can act as candidate sites for real atoms, allowing existence to be transferred from less to more favorable sites over arbitrarily long distances bypassing energy barriers in the conventional position space. Since some of the atoms end up with very little or no existence, we shall refer to the additional atoms as ghost atoms, and we will refer to the approach as ghost-BEACON. An illustration of the algorithm is shown in Fig. 1.

II. RESULTS**A. Fingerprint-enabled removal of energy barriers**

In the model, a system with N atoms is treated as a surrogate system with $N^* > N$ atoms, where every atom (with index i) is given a fractional existence $q_i \in [0, 1]$ with the constraint that the fractions sum to the number of real atoms $\sum_i^{N^*} q_i = N$. The system is thus characterized by $3N^*$ spatial coordinates and N^* existence variables. The existence variables are incorporated into a structural fingerprint with radial and angular parts that resemble the corresponding distribution functions. The radial part reads

$$\rho^R(r) = \sum_{\substack{i,j \\ i \neq j}} q_i q_j \frac{1}{r_{ij}^2} f_c(r_{ij}) e^{-|r-r_{ij}|^2/2\delta_R^2}, \quad (1)$$

where r is the distance variable, r_{ij} are the interatomic distances, f_c is a cutoff function, and δ_R is a length parameter. The angular part has a similar form. (See the details of the machine learning model and the fingerprint in Appendix B.)

The radial fingerprint is, in general, quadratic in the existence variables. However, let us consider a situation where

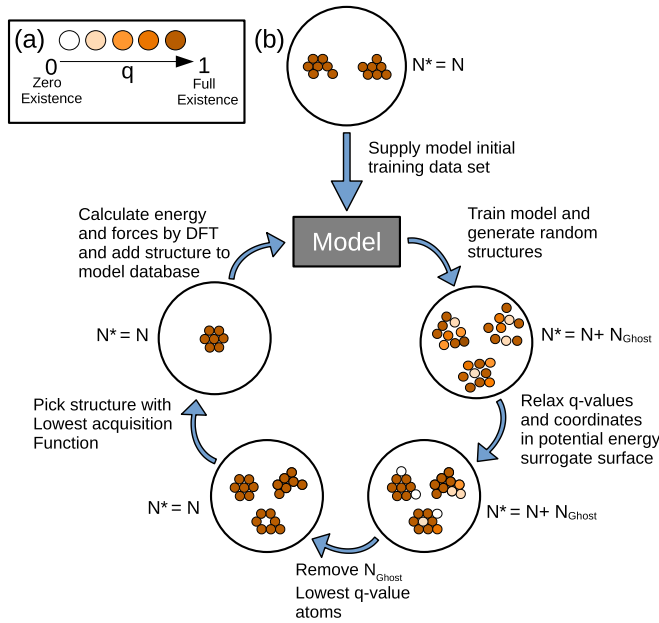


FIG. 1. (a) Illustration of different levels of fractional atomic existence q of a copper atom ranging from zero existence at $q = 0$ to full existence at $q = 1$. The low to zero existence limit is referred to as a ghost atom. (b) Illustration of the ghost-BEACON algorithm with seven real atoms and two ghost atoms. N^* is the total number of atoms, N is the number of real atoms, and N_{ghost} is the number of ghost atoms. The fractional existence values q are spread across all atoms during the relaxation. Except for the inclusion of existence variables, the cycle is identical to that of GOFEE [11].

all atoms either fully exist ($q = 1$) or are completely removed ($q = 0$) except for two atoms, say, numbers 1 and 2, whose distance is larger than the cutoff distance. In that case, the fingerprint becomes linear in q_1 and q_2 . If we furthermore assume that the surroundings of the two atoms are identical, the transfer of existence from atom 2 to atom 1 ($q_2 = 1 - q_1$) leaves the fingerprint *completely unchanged* during the transfer. This means that any machine-learning model based on the fingerprint shows no energy barrier for the process. This analysis also holds if the angular fingerprint is included. (Shown explicitly in the Supplemental Material, Fig. S1 [16]).

To illustrate the removal of energy barriers further, we show in Fig. 2(a) a system with seven copper atoms accompanied by a ghost atom with the energies calculated with an effective-medium-theory (EMT) interatomic potential [17,18]. We investigate the energy profile of moving an atom from a less favorable site (site 8) to a more favorable one (site 1) by following the trajectory shown in blue, which is the minimal-energy path found with a nudged-elastic-band (NEB) calculation [19,20]. We compare this motion to the alternative path of existence transfer. A Gaussian-process surrogate model is trained on eight points along the NEB trajectory. The black curve in Fig. 2(b) shows the EMT energies along the NEB path, while the blue curve is the surrogate energy along the same path. The curves exhibit two energy barriers in the energy landscape corresponding to atom 8 bypassing atoms 5 and 2. The yellow curve in Fig. 2(b) shows the energy during the transfer of existence from atom 8 to 1 with the reaction

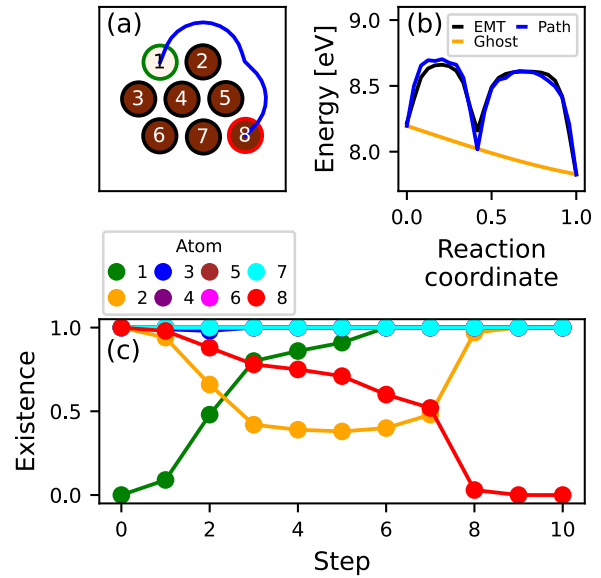


FIG. 2. (a) The 2D test system with eight atoms, labeled from 1 to 8. In this configuration, atom 1 is a ghost atom, and atoms 2–8 are real. The blue curve shows the real-space minimum-energy path, where atom 8 is moved to the empty site, site 1. (b) Different energy profiles while moving the atom from site 8 to site 1 in (a). The black curve shows the EMT energies along the minimum-energy path, and the blue curve shows surrogate energies along the same path. The yellow curve shows the energy profile in the case where no atoms are moved, but the existence is transferred from atom 8 to atom 1. (c) The variation of the existence variables during the transfer of existence from atom 8 to 1.

coordinate $q_1 = 1 - q_8$ and all other existence variables fixed. The energy is almost linear with no potential barrier.

Figure 2(c) visualizes the energy minimization process where, initially, $q_1 = 0$ and $q_i = 1$ for $i = 2, 3, \dots, 8$. During the relaxation, the existence of atom 8 decreases, while the existence of atom 1 increases. The process also involves atoms 2 and 3, which temporarily lose some of their existence. At the end of the relaxation, the existence has been completely transferred from atom 8 to atom 1.

We further illustrate the property of the surrogate PES when varying the existence variables in Fig. 3. Atom 8 is now moved along the indicated linear path in Fig. 3(c) with different amounts of existence q_8 , where the remaining existence is taken up by atom 1, $q_1 = 1 - q_8$. Atom 8 is more weakly interacting with the rest of the cluster when its existence is reduced, but the bonding distance remains essentially the same. An atom with varying existence will tend to position itself at similar geometries, making the transfer of existence more relevant. Atoms without existence do not interact and therefore cannot move, making it unlikely that they take part in optimization. For efficient structure optimizations, it is therefore necessary to introduce a lower bound for the existence variables and, consequently, increase the total existence.

The extension of the machine learning model to the fractional existence space is an extrapolation that cannot be controlled by the addition of data points. The quality of the model therefore depends on the way the existence fractions

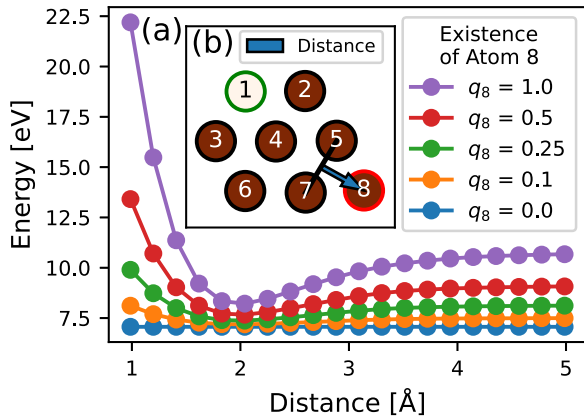


FIG. 3. (a) Energy curve of copper atom 8 as a function of the distance between copper atom 8 and the remaining cluster along the direction of the blue arrow depicted in (b) for different existence fractions of atom 8. Training is done with EMT for 10 different distances of atom 8. All existence not carried in atom 8 is placed in atom 1 ($q_1 = 1 - q_8$). The energy curves are seen to exhibit a minimum at approximately the same distance.

are included in the fingerprint and the choice of hyperparameters for the machine learning model.

B. Optimization of atomic existence fractions on a grid

We now turn to structural optimizations in which the energies and forces are based on DFT as specified in Appendix A. The optimization algorithm is as follows (see Fig. 1): Given a database of structures with DFT calculated energies and forces, a surrogate PES is constructed using a Gaussian process in which the structures are described by the fingerprint. All systems in the database have N atoms, but the surrogate model can be used to make predictions for systems with N^* atoms with fractional existence. The surrogate PES is explored with random searching, that is, with 40 local relaxations based on random initial configurations. The relaxations can be performed in either the atomic coordinates or the fractional existence variables or both. If the existence variables take on fractional values after relaxation, the N largest fractions are set to 1, and the remaining ones are set to 0. The relaxed structures are evaluated with an acquisition function—we use the lower-confidence bound (LCB)—and the structure with the lowest value is added to the DFT database. This procedure is iteratively repeated while keeping track of the low-energy structures obtained. The full simulation procedure is repeated to obtain the statistics of the performance. The details of the algorithm, including the computational parameters, can be found in Appendix B.

We first consider some examples in which the atomic positions are fixed and only the existence variables are optimized. Figures 4(a)–4(d) show four different systems: Figure 4(a) shows a single layer of carbon atoms on a periodic triangular lattice with an equilibrium interatomic distance of 1.42 \AA , corresponding to that of graphene. The system contains a total of 72 atoms with 48 real atoms, which is the number of atoms corresponding to a layer of graphene. Figure 4(b) shows a dense layer of carbon atoms on a periodic rectangular grid

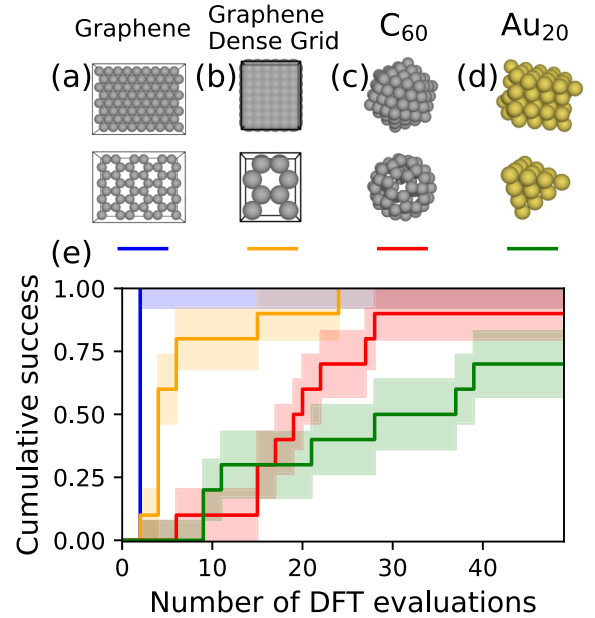


FIG. 4. (a)–(d) Atomic grids (top) and global minimum-energy structures (bottom) of (a) carbon (48 atoms) on a periodic triangular lattice (72 atoms), (b) carbon (8 atoms) on a dense rectangular lattice (48 atoms total), (c) C_{60} on a 147-atom icosahedral grid, and (d) Au_{20} on a 64-atom fcc grid. (e) Success curves for finding the global minimum-energy structure for each setup shown in (a)–(d). Only the existence variables are optimized while keeping the atomic positions fixed on the grid. The uncertainties are Bayesian estimates.

with interatomic distance $a = 0.710 \text{ \AA}$ in one direction and $0.5\sqrt{3}a$ in the other direction. The total number of atoms is 48, with 8 real atoms again corresponding to the density of graphene. Figure 4(c) shows an icosahedron of carbon atoms with 147 atoms in total and 60 real atoms with an interatomic distance of 1.44 \AA between atoms belonging to the same icosahedral layer, roughly agreeing with the bond lengths for a buckyball. Figure 4(d) shows a cluster of fcc gold containing a total of 64 atoms and 20 real atoms.

Each optimization has an initial training set of two random sets of existence variables: one in which the atoms are chosen at random and one in which the atoms are chosen at random but so that the final structure is connected. The obtained minimum-energy structures for the four systems are shown in the lower panels of Figs. 4(a)–4(d). The minimum-energy structure for Figs. 4(a) and 4(b) is a graphene layer, for Fig. 4(c) it is a C_{60} buckyball, and for Fig. 4(d) it is the tetrahedral Au_{20} cluster [21]. The statistics of the optimizations are shown in the success curves in Fig. 4(e). In all four cases 10 independent simulations have been performed, and the success curves show the fraction of simulations which have found the lowest-energy structure as a function of the number of DFT calculations being performed. Success is registered once a structure satisfies the correct nearest-neighbor distribution for all atoms in the cluster compared to the global minimum. Success curve uncertainty is calculated by a Bayesian approach as documented in Appendix C.

The algorithm succeeded in finding the global optimum within 50 DFT calculations in 10/10 runs for both grid types

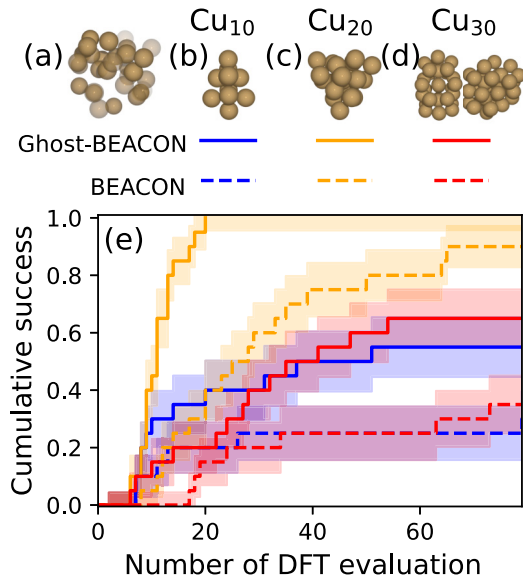


FIG. 5. (a) Random initial configuration of Cu₂₀ with 10 ghost atoms. Existence fractions are represented by the transparencies of the atoms. (b) and (d) Global minimum structures of Cu₁₀ and Cu₂₀ and (c) and (e) the two lowest-energy minima of Cu₃₀, which are so close in energy that they are almost inseparable. (e) Success curves of 20 independent runs without ghost atoms (BEACON) and with 50% ghost atoms added. Each iteration of the BEACON cycle was based on 40 surrogate relaxations. Each run had an initial training set of two random structures.

of graphene and in 9/10 and 7/10 attempts for C₆₀ and Au₂₀, respectively.

C. Simultaneous optimization of atomic coordinates and existence fractions

The method also allows for simultaneous optimization of atomic coordinates and existence fractions, as we shall now illustrate with copper clusters of varying size. We compare the performance of BEACON, which optimizes only the atomic coordinates, and the present approach, ghost-BEACON, which optimizes both atomic coordinates and existence variables. We consider copper clusters with sizes of 10, 20, and 30 atoms, and in each case we add 50% ghost atoms and perform 20 independent simulations. The resulting minimum-energy structures are shown in Fig. 5 together with the success curves, where success is declared when a structure is within 0.1 eV of the lowest energy encountered across all runs of a given cluster size. Further analysis shows that the declared successful structures for Cu₁₀ are all identical, while in the case of Cu₂₀ two distinct structures are identified. In the case of Cu₃₀ several structures have low energies, most of them slight alterations of the structures shown in Fig. 5(d).

We first note that the number of DFT calculations necessary to determine low-energy structures does not vary monotonically with cluster size. The Cu₁₀ cluster requires considerably more computational effort than Cu₂₀. This might seem surprising because the number of variables to consider in the optimization of course increases with cluster size. However, it should be recalled that we are doing random searching

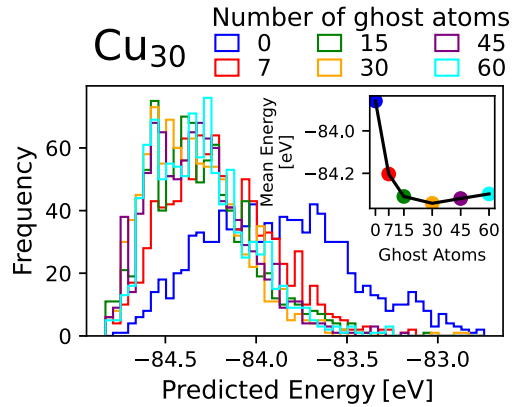


FIG. 6. Distribution of the energies obtained with 1000 relaxations on a surrogate PES for Cu₃₀. The inset shows the variation of the average energies as a function of the number of ghost atoms.

on the surrogate PES (with or without the existence variables) starting from random initial configurations, and the basin of attraction for the different local minima might vary substantially. This is the case for Cu₁₀ and Cu₂₀, for which the third- and second-lowest-energy structures, respectively, are found more frequently than the ground state (shown with success curves in Figs. S2 and S3). The use of an LCB acquisition function promotes the necessary exploration of the configuration space. The exploration is especially pronounced in the early phases of the optimizations, as shown in Fig. S4.

Overall, the presence of ghost atoms is seen to improve the searches considerably, in particular in the cases where BEACON does not easily identify the ground state.

The structures in Figs. 5(b)–5(d) are different from the ones found using empirical potentials or tight-binding molecular dynamics [22–24]. They are also different and lower in energy than the structures found using DFT in Ref. [25], as verified by relaxing all candidate structures with DFT.

D. On the number of ghost atoms

The main function of the ghost atoms is to open new relaxation pathways, as discussed above. To analyze this more, we construct a surrogate PES for Cu₃₀ from a training set consisting of 151 configurations, including some of the identified low-energy structures. We perform 1000 relaxations on the potential energy surface from random initial configurations for different choices of ghost atoms. The distributions of the obtained relaxed surrogate energies are shown in Fig. 6. We see that when ghost atoms are introduced, the distribution is shifted to lower energies as an indication that the relaxations are not trapped as much in higher-lying local minima as is the case for BEACON. The inset in Fig. 6 shows the average energies of the distributions. Clearly, the main effect comes from introducing just a few ghost atoms into the system, and the effect quickly levels off with the number of ghost atoms. The fact that very few ghost atoms improve the efficiency is also seen for Cu₁₀ and Cu₂₀ and is also observed in the success curves (Figs. S3 and S5).

E. Concluding remarks

Increased efficiency of global optimization of N -atom structures using $N^* > N$ atoms with fractional existence has been demonstrated. Several modifications and extensions of the approach presented here are possible. In the present implementation, the sum of the existence variables is constrained to be the number of real atoms in the system. However, one could generalize this to treat open systems with a variable number of atoms controlled by a chemical potential.

The ghost-BEACON code with demo examples as well as the optimized structures and grids generated during the current study are available in the GPATOM package and can be accessed via GitLab [26].

ACKNOWLEDGMENTS

We acknowledge support from the VILLUM Center for Science of Sustainable Fuels and Chemicals, which is funded by the VILLUM Fonden research grant (9455).

APPENDIX A: DENSITY FUNCTIONAL THEORY

The DFT calculations are performed using GPAW [27,28] and the Atomic Simulation Environment [29,30]. We apply the Perdew-Burke-Ernzerhof [31] exchange-correlation functional but note that, in principle, any other functional or alternative quantum chemical method can be used. The plane wave cutoff is 700 eV, and the Fermi temperature is 0.1 eV. Only the Γ point is used for k -point sampling except for graphene on a dense grid (Fig. 4) where (3, 2, 1) k points are used. The convergence criterion for relaxations is that all atomic forces are smaller than 0.01 eV/Å.

APPENDIX B: MACHINE LEARNING MODEL

1. Fingerprint

The fingerprint is based on the one used in BEACON [12] with the inclusion of existence fractions $q_i \in [0, 1]$ for each atom i . The fingerprint is denoted by $\rho(\mathbf{x}, Q)$, where \mathbf{x} is the full set of Cartesian coordinates and Q is the full set of existence fractions. $\rho(\mathbf{x}, Q)$ is a concatenation of a radial part $\rho^R(r; \mathbf{x}, Q)$ and an angular part $\rho^\alpha(\theta; \mathbf{x}, Q)$, which for a single-element system are given by

$$\rho^R(r; \mathbf{x}, Q) = \sum_{\substack{i,j \\ i \neq j}} q_i q_j \frac{1}{r_{ij}^2} f_c(r_{ij}; R_c^R) e^{-|r-r_{ij}|^2/2\delta_R^2}, \quad (\text{B1})$$

$$\rho^\alpha(\theta; \mathbf{x}, Q) = \sum_{\substack{i,j,k \\ i \neq j \neq k}} [q_i q_j q_k f_c(r_{ij}; R_c^\alpha) f_c(r_{jk}; R_c^\alpha) e^{-|\theta-\theta_{ijk}|^2/2\delta_\alpha^2}], \quad (\text{B2})$$

$$f_c(r_{ij}; R_c) = \begin{cases} 1 - (1 + \gamma) \left(\frac{r_{ij}}{R_c}\right)^\gamma + \gamma \left(\frac{r_{ij}}{R_c}\right)^{1+\gamma} & \text{if } r_{ij} \leq R_c, \\ 0 & \text{if } r_{ij} > R_c, \end{cases} \quad (\text{B3})$$

where the indices i , j , and k run over all atoms. Here, r_{ij} is the distance between atoms i and j ; θ_{ijk} is the angle between atoms i , j , and k , while f_c is a smooth cutoff function going to zero at the radial and angular cutoff radii $R_c^R = 5r_{\text{cov}}$ and $R_c^\alpha = 3r_{\text{cov}}$, respectively with r_{cov} being the covalent radius

of the element. γ is a parameter set to 2. The constants $\delta_R = 0.4$ Å and $\delta_\alpha = 0.4$ rad are identical for all systems. The radial and angular fingerprint consist of 200 and 100 grid points, respectively.

2. Gaussian process in the Ghost-BEACON framework

The energies and forces, $\mu = (E, -F)$, are calculated with the standard expression for a Gaussian process [32,33]:

$$\mu(\mathbf{x}, Q) = \mu_p(\mathbf{x}, Q) + K(\rho[\mathbf{x}, Q], P)C(P, P)^{-1}[y - \mu_p(X)], \quad (\text{B4})$$

where $\mu_p(\mathbf{x}, Q)$ and $\rho(\mathbf{x}, Q)$ are the prior mean and the fingerprint, respectively, K and C are the covariance matrix without and with regularization, P is a matrix containing the training data fingerprints, y is the training data targets, and $\mu_p(X)$ is the prior function applied to all structures in the training data. The uncertainty of the predicted energy is given by

$$\Sigma(\mathbf{x}, Q) = \{\tilde{K}(\rho[\mathbf{x}, Q], \rho[\mathbf{x}, Q]) - K(\rho[\mathbf{x}, Q], P)C(P, P)^{-1}K(P, \rho[\mathbf{x}, Q])\}^{1/2}, \quad (\text{B5})$$

where $\tilde{K}(\rho[\mathbf{x}, Q], \rho[\mathbf{x}, Q])$ represents the covariance matrix for the fingerprint. We observe from Eqs. (B4) and (B5) that K , \tilde{K} , and $\mu_p(\mathbf{x}, Q)$ are the only terms including the existence fractions. The details of the construction of K , C , and y are reported in Ref. [12].

We use a squared exponential kernel function

$$k(\rho_1, \rho_2) = \sigma^2 \exp\left(\frac{-D(\rho_1 - \rho_2)^2}{2l^2}\right), \quad (\text{B6})$$

where $D(\rho_1 - \rho_2)$ is the Euclidean distance between two fingerprints and σ^2 and l denote the prefactor and length scale hyperparameters, respectively. The routine for the optimization of the hyperparameters is described in the Supplemental Material [16].

3. Prior function

For the simultaneous optimization of positions and existence fractions of Figs. 5 and 6, a repulsive prior modified to include the existence fractions is used [11,12]:

$$\mu_p(\mathbf{x}, Q) = \mu_c + \sum_{\substack{i,j \\ i \neq j \\ r_{ij} < 2\tilde{r}_{\text{cov}}}} q_i q_j \left(\frac{2\sigma_p \tilde{r}_{\text{cov}}}{r_{ij}}\right)^{12}, \quad (\text{B7})$$

where $\sigma_p = 0.4$ is a repulsive constant, \tilde{r}_{cov} is an atomic radius set to be $0.8r_{\text{cov}}$ of the element, and μ_c is a constant prior updated throughout the run (see the Supplemental Material [16]). The potential drops to zero if atom centers are farther than $2\tilde{r}_{\text{cov}}$ apart. This prior represents our physical knowledge that atoms repel when squeezed too closely together while hindering low existence atoms from interfering with the clustering of high existence atoms. For all other simulations the repulsive part is omitted, and the prior is simply set to $\mu_p = \mu_c$.

4. Acquisition function

To select which of the relaxed structures to include in the DFT database, we use a lower confidence bound acquisition function:

$$f(\mathbf{x}) = \mu(\mathbf{x}) - \kappa \Sigma(\mathbf{x}), \quad (\text{B8})$$

where $\kappa = 2$ is a constant and $\mu(\mathbf{x})$ and $\Sigma(\mathbf{x})$ are the predicted energy and uncertainty of Eqs. (B4) and (B5) [11,12]. The dependence on Q is omitted as the acquisition function is always evaluated on structures without ghost atoms.

5. Structure exclusion

To hinder inclusion of identical or highly similar structures in the database, structures closer than d_{fp} from already known structures in fingerprint space are discarded.

We set $d_{\text{fp}} = 5$ for optimization of both atomic coordinates and existence values and $d_{\text{fp}} = 0.05$ for optimization on a grid. If all candidates are discarded, a random structure is added to the database instead.

6. Random structure generation

All random configurations not placed on a grid are set up using a cubic box with a volume 5 times the sum of the volumes of atomic spheres with radius r_{cov} surrounded by an additional 7.5 Å of vacuum to complete the unit cell. The atoms initially placed randomly in the box are repelled until all atom centers are at least $1.6r_{\text{cov}}$ away from each other. This procedure ensures a similar initial atomic packing fraction independent of the number of atoms in the BEACON/ghost-BEACON runs. The initial training sets were made in the same way, omitting the ghost atoms. For atoms on a grid the positions are always the same, and only the initial fractions differ.

7. Random fraction generation

The random sampling of the initial existence values is done using the Dirichlet-rescale algorithm [34,35]. This allows for a uniform distribution of the existence values satisfying

the constraints $q_i \in [q_{\text{min}}, 1]$ and $\sum_i q_i = N + (N^* - N)q_{\text{min}}$, where $0 \leq q_{\text{min}} < 1$ is a lower existence bound.

8. Surrogate surface relaxations

As atoms with $q_i = 0$ become immobile, the following procedure is adopted for simultaneous optimization of coordinates and existence fractions:

(1) Initialize a system of random atomic positions and existence fractions between $q_{\text{min}}^{\text{init}} (> 0)$ and 1 with a total existence of $N + (N^* - N)q_{\text{min}}^{\text{init}}$.

(2) Relax the system on the surrogate PES for n_{relax} steps.

(3) Decrease the lower limit in n_D steps of $q_{\text{min}}^{\text{init}}/n_D$, and at each level, perform a relaxation with n_d steps.

(4) Relax the system for n_p steps with all existence variables fixed to 0 or 1 to effectively remove the ghost atoms.

The simultaneous relaxations of existence and positions is done with $n_{\text{relax}} = 200$, $q_{\text{min}}^{\text{init}} = 0.05$, $n_D = 5$, $n_d = 20$, and $n_p = 100$. All nonghost BEACON runs are performed with $n_{\text{relax}} = 400$ with $N^* = N$ and all fractions fixed to 1. The relaxations on a grid are performed with $n_{\text{relax}} = 200$ and $q \in [0, 1]$.

All relaxations are performed using sequential least squares programming [36] as implemented in the SCIPY package [37] with the convergence criterion set to 0.001.

APPENDIX C: SUCCESS CURVE UNCERTAINTY ESTIMATES

We model a success curve as $n + m$ independent attempts at finding the global optimum, with n and m denoting the numbers of successful and unsuccessful attempts, respectively. Using Bayes theorem with a uniform prior, the posterior probability of success p_s becomes a beta distribution $B(p_s | \alpha = n + 1, \beta = m + 1)$. We use the mode of this distribution, $\text{mode}(p_s) = n/(n + m)$, as the value of the success curve. For the uncertainty, we use the square root of the variance,

$$\sqrt{\text{var}(p_s)} = \sqrt{\frac{(n + 1)(m + 1)}{(n + m + 2)^2(n + m + 3)}}. \quad (\text{C1})$$

-
- [1] J. Zhang and V. A. Glezakou, *Int. J. Quantum Chem.* **121**, 044114 (2020).
- [2] C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter* **23**, 053201 (2011).
- [3] L. B. Vilhelmsen and B. Hammer, *J. Chem. Phys.* **141**, 044711 (2014).
- [4] S. V. Lepeshkin, V. S. Baturin, Y. A. Uspenskii, and A. R. Oganov, *J. Phys. Chem. Lett.* **10**, 102 (2019).
- [5] S. Lysgaard, D. D. Landis, T. Bligaard, and T. Vegge, *Top. Catal.* **57**, 33 (2014).
- [6] M. Jäger, R. Schäfer, and R. L. Johnston, *Nanoscale* **11**, 9042 (2019).
- [7] D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A* **101**, 5111 (1997).
- [8] Z. Chen, W. Jia, X. Jiang, S.-S. Li, and L.-W. Wang, *Comput. Phys. Commun.* **219**, 35 (2017).
- [9] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).
- [10] M. Todorović, M. U. Gutmann, J. Corander, and P. Rinke, *npj Comput. Mater.* **5**, 35 (2019).
- [11] M. K. Bisbo and B. Hammer, *Phys. Rev. Lett.* **124**, 086102 (2020).
- [12] S. Kaappa, E. G. del Río, and K. W. Jacobsen, *Phys. Rev. B* **103**, 174114 (2021).
- [13] M.-P. V. Christiansen, N. Rønne, and B. Hammer, *J. Chem. Phys.* **157**, 054701 (2022).
- [14] C. J. Pickard, *Phys. Rev. B* **99**, 054102 (2019).
- [15] S. Kaappa, C. Larsen, and K. W. Jacobsen, *Phys. Rev. Lett.* **127**, 166001 (2021).
- [16] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.107.214101> for additional information on the machine learning model and the ghost-BEACON

- algorithm, as well as supporting data on the performance of ghost-BEACON.
- [17] K. W. Jacobsen, J. K. Nørskov, and M. J. Puska, *Phys. Rev. B* **35**, 7423 (1987).
- [18] K. Jacobsen, P. Stoltze, and J. Nørskov, *Surf. Sci.* **366**, 394 (1996).
- [19] G. Mills and H. Jónsson, *Phys. Rev. Lett.* **72**, 1124 (1994).
- [20] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations: Proceedings of the International School of Physics “Computer Simulation of Rare Events and Dynamics of Classical and Quantum Condensed-Phased Systems,” Lerici, Villa Marigola, 7 July–18 July 1997*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998), pp. 385–404.
- [21] J. Li, X. Li, H.-J. Zhai, and L.-S. Wang, *Science* **299**, 864 (2003).
- [22] M. Büyükata and J. C. Belchior, *J. Braz. Chem. Soc.* **19**, 884 (2008).
- [23] J. P. Doye and D. J. Wales, *New J. Chem.* **22**, 733 (1998).
- [24] M. Kabir, A. Mookerjee, and A. K. Bhattacharya, *Eur. Phys. J. D* **31**, 477 (2004).
- [25] U. J. Rangel-Pena, R. L. Camacho-Mendoza, S. González-Montiel, L. Feria, and J. Cruz-Borbolla, *J. Cluster Sci.* **32**, 1155 (2021).
- [26] <https://gitlab.com/gpatom/ase-gpatom/>.
- [27] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, *Phys. Rev. B* **71**, 035109 (2005).
- [28] J. Enkovaara *et al.*, *J. Phys.: Condens. Matter* **22**, 253202 (2010).
- [29] A. H. Larsen *et al.*, *J. Phys.: Condens. Matter* **29**, 273002 (2017).
- [30] Atomic Simulation Environment, <https://wiki.fysik.dtu.dk/ase/>.
- [31] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [32] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).
- [33] J. Wu, M. Poloczek, A. G. Wilson, and P. Frazier, *Advances in Neural Information Processing Systems* (Curran Associates Inc., New York, 2017), pp. 5267–5278.
- [34] D. Griffin, I. Bate, and R. I. Davis, *IEEE Real-Time Systems Symposium, RTSS 2020, Houston, Texas, USA, December 1–4, 2020* (IEEE, Piscataway, NJ, 2020), pp. 76–88.
- [35] D. Griffin, I. Bate, and R. I. Davis, dgdguk/drs: v2.0.0, Zenodo (2020), <https://doi.org/10.5281/zenodo.4118058>.
- [36] D. Kraft, *ACM Trans. Math. Software* **20**, 262 (1994).
- [37] P. Virtanen *et al.*, *Nat. Methods* **17**, 261 (2020).