Accelerating the search for global minima on potential energy surfaces using machine learning

S. F. Carr' , R. Garnett' , and C. S. Lo'

Citation: The Journal of Chemical Physics **145**, 154106 (2016); doi: 10.1063/1.4964671
View online: http://dx.doi.org/10.1063/1.4964671
View Table of Contents: http://aip.scitation.org/toc/jcp/145/15
Published by the American Institute of Physics

Accelerating the search for global minima on potential energy surfaces using machine learning

S. F. Carr,[1,2,a)] R. Garnett,[2,b)] and C. S. Lo[1,c)]
[1]*Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA*
[2]*Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA*

(Received 14 July 2016; accepted 26 September 2016; published online 18 October 2016)

Controlling molecule-surface interactions is key for chemical applications ranging from catalysis to gas sensing. We present a framework for accelerating the search for the global minimum on potential surfaces, corresponding to stable adsorbate-surface structures. We present a technique using Bayesian inference that enables us to predict converged density functional theory potential energies with fewer self-consistent field iterations. We then discuss how this technique fits in with the Bayesian Active Site Calculator, which applies Bayesian optimization to the problem. We demonstrate the performance of our framework using a hematite (Fe_2O_3) surface and present the adsorption sites found by our global optimization method for various simple hydrocarbons on the rutile TiO_2 (110) surface. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4964671]

I. INTRODUCTION

Surface science, encompassing real-world applications ranging from catalysis to gas sensing, is predicated on gas-phase molecules adsorbing to the thermodynamically most-favorable "active" sites on solid surfaces. From a mathematical point of view, this boils down to an optimization problem—efficiently performing global energy minimization in a multidimensional space corresponding to the potential energy hypersurface, while simultaneously ensuring that the identities of the molecule and surface remain unchanged.

Over the last few decades, researchers have been exploring many different first-principles approaches to solving this problem, including approaches based on simulated annealing,[1,2] genetic and evolutionary algorithms,[3–8] stochastic tunneling,[9] basin hopping,[10–12] tabu search,[13–15] heuristical sampling,[16] and others. This journal and its peer publications have published articles reviewing and comparing some of the existing global optimization techniques.[17–19] One of the most cited methods in recent publications is *minima hopping*, a method proposed here by Goedecker.[11] Minima hopping, a method in the basin hopping family, is based on a two-step procedure: (1) a molecular dynamics simulation to stochastically perturb the atoms and (2) a density functional theory (DFT) step to relax the atoms to a local minimum. This procedure is run several times to ensure that all possible binding sites have been explored. *Constrained minima hopping* (CMH) (2014)[12] is an improvement to minima hopping that enforces molecular identity by placing constraints on the distance between atoms. Chemical applications of minima hopping range from the synthesis of fullerenes,[20] superconductors,[21] and photovoltaics[22] to the reactive pathways of synthesis gas.[23] Unfortunately, minima hopping is computationally expensive—often requiring several hundred DFT calculations to converge to the solution.

We recently published initial results of the Bayesian Active Site Calculator (BASC) at the 33rd International Conference on Machine Learning.[24] BASC formulates the problem at hand as a low-dimensional objective function that is capable of parameterizing all possible configurations of a rigid molecule on the surface. BASC then uses *Bayesian optimization* to search for the global minimum of that objective function. The atoms of the surface and molecule may deform from their free-space configurations during adsorption, but we found that the true minimum is closely approximated by the minimum in the rigid-molecule space, at least for the physisorption observed with carbon monoxide (CO) on the stoichiometric, Fe-terminated surface of Fe_2O_3 we studied in that work.

In this paper, we expand on BASC by presenting a framework for predicting the outcome of each DFT evaluation in fewer self-consistent field (SCF) iterations. Further, we demonstrate the accuracy and process speedup of our method on two model surfaces: the Fe_2O_3 surface from our previous work and the single-oxygen 110 (rutile) surface of TiO_2. On Fe_2O_3, we benchmarked the performance of BASC against constrained minima hopping for the adsorption of carbon monoxide (CO)—a representative molecule in surface science.[25] On TiO_2, we used BASC to determine the active sites and energies of adsorption for four molecules—CO, CO_2, CH_2O, and CH_4—for comparison to results previously published in the literature. The results of the Fe_2O_3 performance benchmarks are in Section V, and the results of the TiO_2 calculations are in Section VI.

a)Electronic mail: shane.carr@wustl.edu
b)Electronic mail: garnett@wustl.edu
c)Electronic mail: clo@wustl.edu

II. ACCELERATION OF DFT CALCULATIONS

Generally, situations where adsorbate-surface interactions are of interest involve complex electronic structure. These are treated by first-principles methods, such as density functional theory (DFT),[26,27] which involve the numerical solution of the Kohn–Sham equation.

A DFT calculation is initiated by making an educated guess for the *basis set* and then running an iterative procedure to converge the basis set parameters to their final values. The latter procedure is commonly known as the *self-consistent field* (SCF) loop. Performing an SCF calculation is computationally expensive, with the classical eigenvalue problem scaling as $O(N^3)$ with the number of atoms and electrons, with ongoing efforts to develop more efficient $O(N^2)$ and linear scaling DFT methods.[28]

Even though a large number of potential energy calculations are performed during BASC's global optimization, the changes to the system are small. Since BASC's objective function assumes a fixed surface, it is only atoms of the adsorbate molecule that change from calculation to calculation. We take advantage of this observation to accelerate the energy minimization procedure in DFT and utilize a popular electronic structure code, GPAW,[29–31] to perform these calculations.

A. Analysis of SCF convergence traces

From our study of CO adsorbed to Fe_2O_3, we observed that the difference between the calculated energy at the first SCF iteration, O_1, and the calculated energy at final convergence, E_∞, is normally distributed, as shown in Figure 1. This plot was generated from 154 631 SCF traces, each of which converged to 10^{-3} eV; this corresponds to a grid of parameters ranging over 1.5 Å $\leq z \leq$ 2.5 Å and all values of the periodic variables. Of those traces, 73% of their $\Delta_1 \equiv E_\infty - O_1$ values fall within 0.1 eV of the mean. This suggests that E_∞ may be inferred probabilistically from O_1.

We show in Figure 2 how these SCF traces exhibit high similarity for molecule-surface adsorption to Fe_2O_3 (001), as well as TiO_2 (110).

At each iteration i, define the difference between the observed energy at i, O_i, and the final converged energy, E_∞, to be

$$\Delta_i \equiv E_\infty - O_i. \tag{1}$$

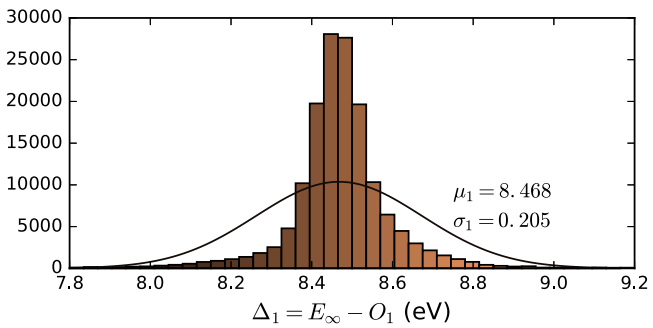

FIG. 1. Histogram of Δ_1 values for CO on Fe_2O_3, based on 154 631 traces.

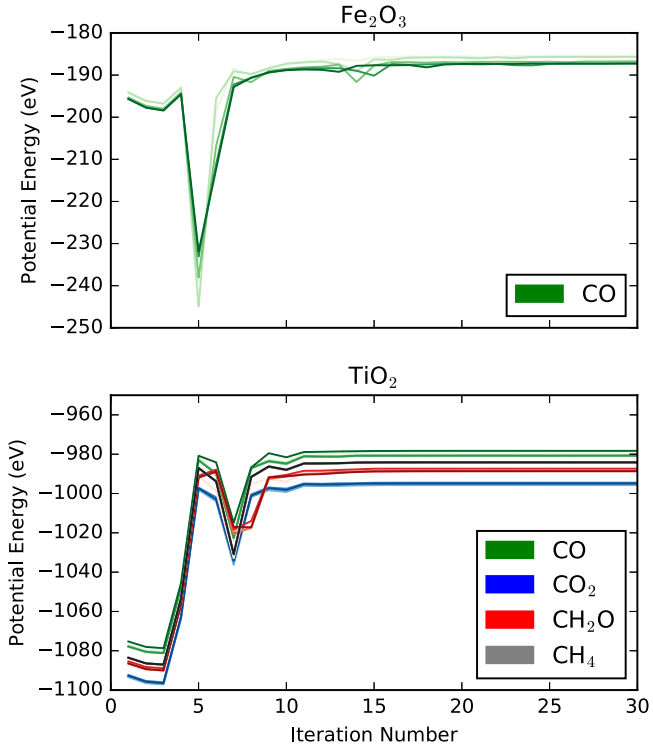

FIG. 2. Sample SCF traces for molecular adsorption; each molecule-surface combination plotted consists of five distinct traces—showing the reproducibility of such calculations regardless of initial configuration.

These traces are obtained using the *pseudo partial waves* basis set.[32] As expected, the traces for each system follow a similar pathway from the initial guess to the final value. For the case of TiO_2, Δ_1 can be as high as 90 eV, but the corresponding standard deviation remains well below 1 eV, as shown in Table I, even though CO, CO_2, CH_2O, and CH_4 exhibit a range of polarities and charge distributions. This is further demonstrated in Figure 3, which shows the standard deviation σ_i for Δ_i ($i \leq 30$), for each of the molecule-surface systems explored here.

B. Application of Bayesian inference

If we precompute several sample traces for a particular surface-adsorbate combination, we can infer likely values for the converged energy, E_∞, using Bayes' rule. The probability of observing E_∞, given O_i and pre-computed "training data"

TABLE I. Mean and standard deviation of Gaussian distributions fitted over Δ_1 for the systems we studied. The TiO_2 systems had eight samples each; the Fe_2O_3 corresponds to the same 154 631 traces as Figure 1.

	μ_i (eV)	σ_1 (eV)
CO on Fe_2O_3	8.468	0.205
CO on TiO_2	97.114	0.149
CO_2 on TiO_2	97.700	0.076
CH_2O on TiO_2	97.874	0.065
CH_4 on TiO_2	99.377	0.070

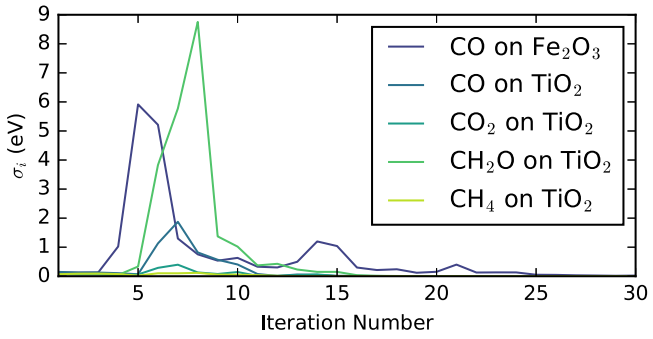

FIG. 3. Standard deviation (σ_i) of the Gaussian distributions fitted over Δ_i (Equation (1)) for molecular adsorption on Fe_2O_3 and TiO_2, with $1 \le i \le 30$.

(e.g., SCF traces), $\mathcal{D}$, is

$$p(E_\infty \mid O_i, \mathcal{D}) \propto p(O_i \mid E_\infty, \mathcal{D}) p(E_\infty \mid \mathcal{D}). \tag{2}$$

We model the likelihood term, $Pr(O_i \mid E_\infty, \mathcal{D})$, with a normal (Gaussian) distribution of Δ_i, as fitted to training traces,

$$p(O_i \mid E_\infty, \mathcal{D}) = E_\infty - \exp\left[-\frac{1}{2\sigma_i^2}(O_i - \mu_i)^2\right] \tag{3}$$

$$= E_\infty - \mathcal{N}(O_i; \mu_i, \sigma_i^2) \tag{4}$$

$$= \mathcal{N}(O_i; E_\infty - \mu_i, \sigma_i^2) \tag{5}$$

$$= \mathcal{N}(E_\infty; O_i + \mu_i, \sigma_i^2), \tag{6}$$

where μ_i and σ_i^2 are fit to the observed traces in $\mathcal{D}$.

The prior probability term $Pr(E_\infty \mid \mathcal{D})$ we select also uses a Gaussian fitted over the E_∞ values from the same training data that we used to fit the Δ_i distributions,

$$p(E_\infty \mid \mathcal{D}) = \mathcal{N}(\mu_0, \sigma_0^2). \tag{7}$$

Plugging Equations (6) and (7) into Equation (2), we can express the posterior probability $Pr(E_\infty \mid O_i, \mathcal{D})$ in closed form, with the help of conjugate priors,[33]

$$p(E_\infty \mid O_i, \mathcal{D}) = \mathcal{N}(\mu_p, \sigma_p^2), \tag{8}$$

$$\mu_p = \left(\frac{\mu_0}{\sigma_0^2} + \frac{O_i + \mu_i}{\sigma_i^2}\right) \Big/ \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_i^2}\right), \tag{9}$$

$$\sigma_p^2 = 1 \Big/ \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_i^2}\right). \tag{10}$$

We can choose to gain several observations O_i of an unknown trace. As we gain each observation, we can iteratively apply Equation (8) and update and refine our belief about E_∞.

Looking back at Figure 3, we see that iterations 1–5 give us the most information (i.e., small σ_i) about E_∞ (only iterations 1–4 for Fe_2O_3), with subsequent iterations giving us less information until we reach convergence around iteration 15–20. Indeed, as shown in Figure 4, it turns out that the prediction of E_∞ based on just the first iteration of SCF gives us a very good understanding of the potential energy surface.

The methods outlined here work well for accelerating the prediction of potential energies, but they do not yield a substantial improvement in speed for other properties of the system, such as atomic forces. This means that procedures like structural relaxation do not exhibit appreciable benefit from pre-computation of sample traces, compared to the original BASC method for global optimization.

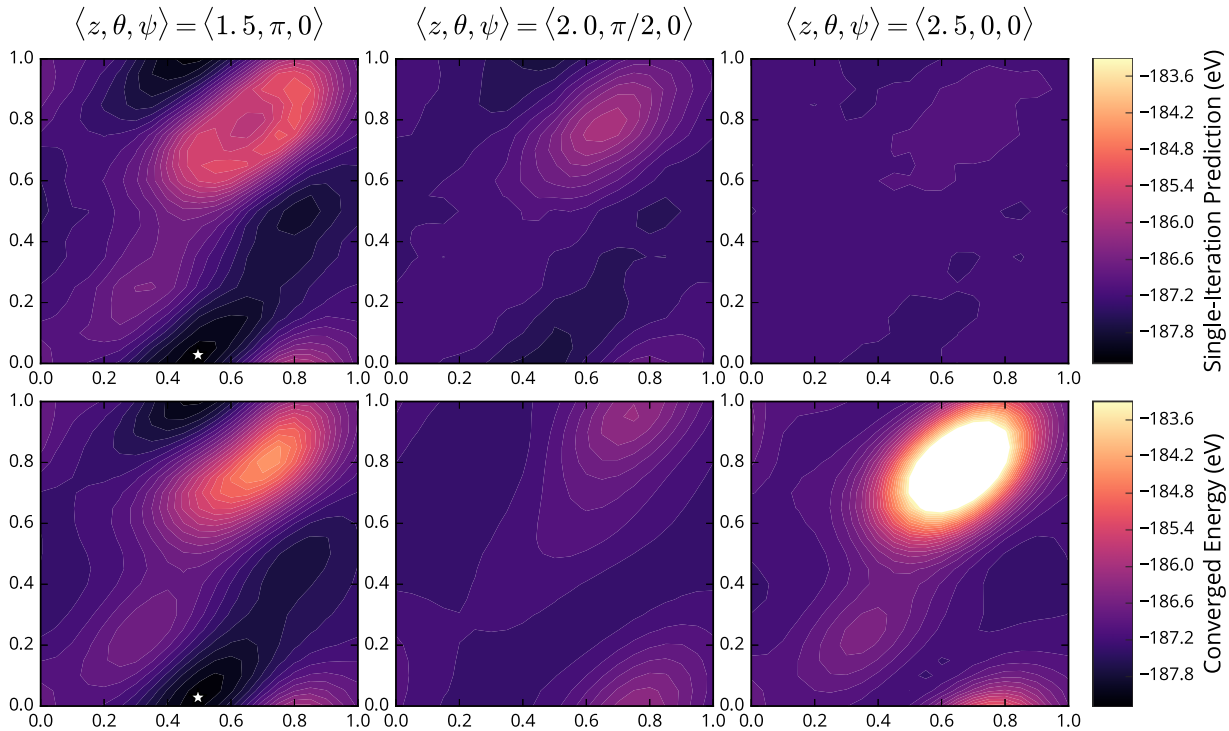

FIG. 4. Slices along the x, y plane of the objective function for three configurations of CO on Fe_2O_3. The top row is the prediction of Equation (8) on a single SCF iteration, while the bottom row is the converged value. The data used to train Equation (8) do not include any of the traces from this figure. The global minimum, projected into x, y, occurs at the center of the dark area in the leftmost column (shown as a white star).

III. INTEGRATION INTO BASC OPTIMIZATION FRAMEWORK

In our previous work, we focused on the global optimization step of BASC. We used a rigid molecule assumption to construct a low-dimensional objective function on which we used Bayesian optimization[34–36] to converge to the global minimum. We explained that a final structural optimization step was necessary to relax the rigid molecule assumption.

With the insight from Section II, we can formulate BASC as a three-step procedure:

1. Obtain training traces.
2. Run Bayesian optimization, using the training traces, a single SCF iteration, and Equation (8) to accelerate each calculation.
3. Perform a final structure relaxation on the rigid-molecule global minimum identified in step 2.

For the results shown below, We trained our accelerated DFT calculations on eight 50-iteration sample traces for each system.

IV. EXPERIMENTAL DETAILS

We used GPAW's implementation of density functional theory (DFT) for all calculations. In the Bayesian optimization loop, we used the accelerated GPAW discussed in Section II; for all other calculations, we used GPAW run to convergence. For all atomic manipulations, we used the Atomic Simulation Environment (ASE).[37]

When performing structural relaxations, we converged the nuclear coordinates using L-BFGS with the criterion $\max_a |\vec{F}_a| = f_{\max} = 0.05$ eV. All calculations were spin polarized.

We obtained crystal structures for Fe_2O_3 and TiO_2 from Materials Project,[38,39] corresponding to material IDs "mp-24972" and "mp-2657" (both the primitive cell).

Since we used Fe_2O_3 as only a performance benchmark, we erred on the side of a cheaper functional (local-density approximation (LDA) exchange-correlation functional, Ref. 40) and a smaller surface area for faster calculations. We cut a 1-by-1 semi-periodic surface slab in the (001) direction, with $a = 5.138$ Å, $b = 5.484$ Å, thickness of two unit cell layers, and 15 Å of vacuum; this corresponds to a total of 20 atoms in the unit cell. When relaxing the empty surface, we held the bottom half of the slab fixed, while the top half was allowed to relax.

Since our intent with TiO_2 was to obtain more realistic results that could be compared to values in the published literature, we used the vdW-DF functional,[41] which corrects for *van der Waals* interactions between the molecule and the surface, and the Hubbard U[42] for the Ti atoms. We set $U_{\text{eff}} = 2.5$ eV, a figure suggested in the work of Stausholm-Moeller *et al.*[43] For the k-points, we used Monkhorst–Pack sampling with a minimum density of 2.5 points per Å, rounded up to the nearest even number.

We also used a larger surface area for the TiO_2 (110) surface. We first cut a 1-by-1 cell with $a = 6.579$ Å, $b = 2.970$ Å, five unit cell layers thick, 15 Å of vacuum, and one surface-facing oxygen atom on each side of the slab, corresponding to 30 atoms. After the initial relaxation in which all atoms were allowed to move, we removed the bottom three unit cell layers of atoms, copied the cell once in the y direction, and then performed a $\sqrt{2} \times \sqrt{2}$ transformation, obtaining a final surface slab with $a = b = 8.863\,39$ Å and 48 atoms. Here, only the topmost atomic layer was allowed to relax, with the remaining layers held fixed, as we observed that relaxation of the subsurface atomic layer did not appear to change the results.

The aforementioned settings for TiO_2 correspond to the following dictionary of options for GPAW:

```
{
    "spinpol": True,
    "xc": "PBE",
    "kpts": {"density": 2.5, "even": True},
    "setups": {"Ti": "paw:d,2.5"}
}
```

V. RESULTS FOR Fe_2O_3 (PERFORMANCE BENCHMARK)

BASC and constrained minima hopping (CMH) find the same active site of CO on Fe_2O_3, with C-down over an iron atom, as shown in Figure 5. The agreement with CMH is evidence of the accuracy of BASC. The relaxed C–O bond length for BASC and CMH is 1.14 Å. On average, CMH identifies an angle of 15.8° between the CO axis and the vertical z axis, compared with 19.8° for BASC.

Figure 6 shows the relative performance of BASC and CMH. The x axis shows the number of SCF calculations, which we assume to be roughly proportional to computation time. Depending on the choice of XC functional, the size of the system, the number of central processing units (CPUs), and numerous other variables, one SCF calculation typically takes from several seconds to several minutes of wall clock time. The three steps of BASC are all incorporated into the plot: the gathering of training data (400 calculations), the Bayesian optimization procedure (200 calculations), and

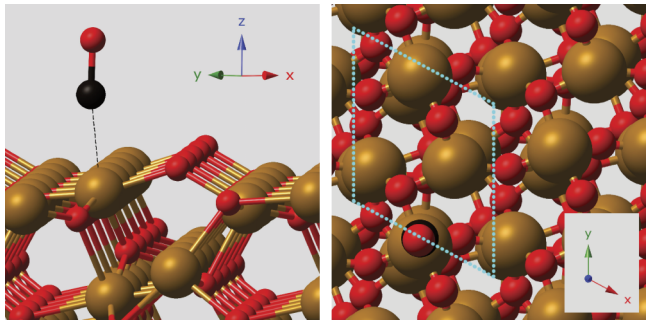

FIG. 5. Adsorption site of CO on the Fe_2O_3 surface. Both CMH and BASC find this result, with the C closest to an Fe atom on the surface. The top view is overlaid with the unit cell boundary; the CO molecule is visible in the lower left corner. Color scheme: C (black), O (red), and Fe (gold). (a) Side view. (b) Top view.

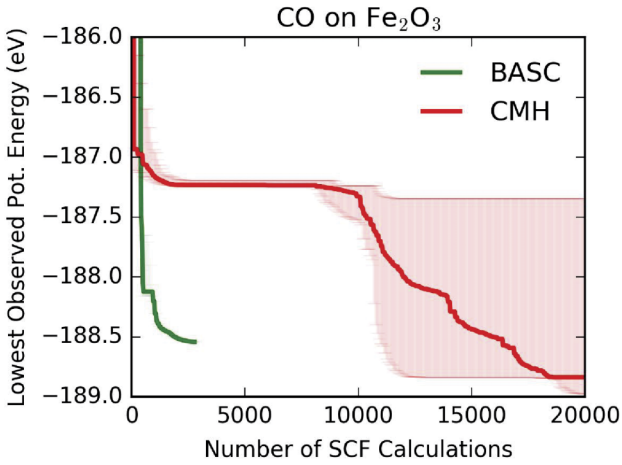

FIG. 6. Comparing the number of calculations for BASC and CMH to find the minimum of the potential energy surface. The lines indicate the minimum potential energy observed up to that point in the calculation in the average case (measured in eV). For CMH, the error bars indicate the range of values at that calculation number over four runs; the solid line shows the median of those runs.

finally the relaxation step (2176 calculations). Note that the bulk of CMH's calculations is in its local relaxation steps.

Figure 6 reveals a difference of 0.295 Å between the BASC solution and the CMH solution. Since CMH performs ten or more structural relaxations with a variety of CO configurations, the surface atoms are able to relax to a greater extent than with the BASC's single relaxation. As a result, BASC does not achieve as high of an energy of adsorption, but its surface atoms retain a configuration closer to that of the empty surface.

VI. RESULTS FOR TiO_2

BASC finds thermodynamically stable adsorption sites for all four molecules. The energies of adsorption are shown in Table II. The orientations of the molecules are shown in Figure 7, with the angle between the molecule's primary axis and the z axis also listed in Table II. Below, we compare BASC's results with previous experimental and theoretical studies.

TABLE II. Energies of adsorption for CO, CO_2, CH_2O, and CH_4 on TiO_2, corresponding to the binding sites found by BASC after having undergone the final relaxation step (see Figure 7). E_{ads} is the energy of adsorption, with $E_{ads} = E_{sys} - E_{surf} - E_{vac}$, where E_{sys} is the energy of the surface with the molecule, $E_{surf} = -954.578$ eV is the energy of the surface without the molecule, and E_{vac} is the energy of the molecule by itself in the center of a cube with side lengths of 20 Å. θ is the angle between the z axis and the primary axis of the molecule; for CO, CO_2, and CH_2O, this is the C–O bond, and for CH_4, this is the C–H bond to the uppermost H atop.

	E_{vac} (eV)	E_{sys} (eV)	E_{ads} (eV)	θ
CO	−26.172	−981.148	−0.398	3.62°
CO_2	−40.830	−995.964	−0.555	24.5°
CH_2O	−33.915	−989.183	−0.690	16.2°
CH_4	−29.573	−984.433	−0.281	11.8°

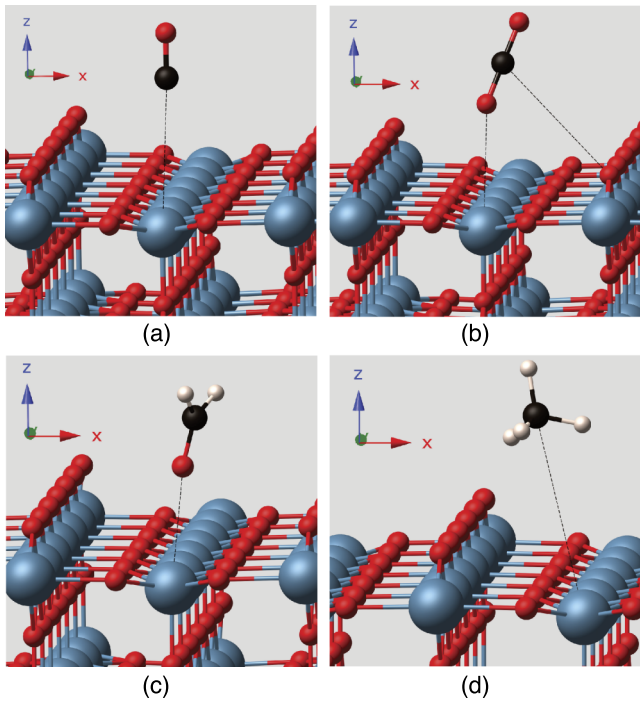

FIG. 7. The binding sites found by BASC on TiO_2 (110, rutile). The corresponding energies are shown in Table II. The dashed black lines are provided for perspective. Color scheme: C (black), O (red), H (White), and Ti (blue). (a) CO (carbon monoxide). (b) CO_2 (carbon dioxide). (c) CH_2O (formaldehyde). (d) CH_4 (methane).

A. Carbon monoxide

Sorescu and Yates[44] used first-principles calculations to evaluate two types of adsorption of CO on rutile TiO_2: Ti–O–C ("O down") and Ti–C–O ("C down"). They found that the "C down" orientation resulted in a stronger energy of adsorption, in agreement with BASC and with previous experimental results.

B. Carbon dioxide

Sorescu *et al.*[45] studied CO_2 adsorption on TiO_2. BASC's result corresponds to their Configuration I, which they also found experimentally to be a configuration along the pathway of CO–to–CO_2 oxidation. BASC's configuration has the CO_2 molecule tilted 26.9° relative to the surface plane.

C. Formaldehyde

Liu *et al.*[46] performed first-principles calculations on four hand-selected adsorption sites for CH_2O on rutile TiO_2, two chemisorption (A_1 and A_2) and two physisorption (A_3 and A_4). BASC's result corresponds to the A_3 adsorption site, which is the stronger of the two physisorption sites. However, site A_1, in which the bond angles of the CH_2O change, corresponds to a stronger adsorption energy than A_3. This is an excellent illustration of the rigid molecule assumption. As discussed in our previous work, adding a parameter for such structural changes should make BASC capable of finding site A_1.

D. Methane

There have been comparatively fewer studies of larger organic molecules on TiO_2. Sushko, Gal, and Shluger[47] found using Hartree–Fock calculations that methane demonstrates weak adsorption when positioned above the Ti atom, consistent with BASC. They found, however, that the configuration with two H atoms facing the surface was slightly stronger than the one with three H atoms facing the surface in a pyramid. This discrepancy may be due to differences in methodology (Hartree–Fock vs. DFT), as well as the acknowledged margin of error in BASC's accelerated SCF calculation.

VII. CONCLUSION

In this paper, we introduced a method to accelerate the DFT calculations required for BASC, by taking advantage of the similarity in structure between calculations.

In our performance benchmarks, we found that BASC outperformed constrained minima hopping (CMH) by a factor of 10-to-1, while still finding the same binding site for CO on Fe_2O_3. In our study of TiO_2, we found that BASC's prediction had agreement with results from the previous literature for all of the molecules we studied. We believe that BASC will be a useful tool for researchers in the field of surface science.

The code for BASC is available online at https://gitlab.com/caml/basc.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Award No. CBET–1349713 and IIA–1355406. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant No. ACI–1053575.[48]

[1]S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983), http://science.sciencemag.org/content/220/4598/671.full.pdf.
[2]E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[3]D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[4]T. S. Bush, C. R. A. Catlow, and P. D. Battle, J. Mater. Chem. **5**, 1269 (1995).
[5]G. Bhm, Biophys. Chem. **59**, 1 (1996).
[6]S. Woodley, P. Battle, J. Gale, and C. A. Catlow, Phys. Chem. Chem. Phys. **1**, 2535 (1999).
[7]R. L. Johnston, Dalton Trans. **2003**, 4193.
[8]C. Wehmeyer, G. Falk von Rudorff, S. Wolf, G. Kabbe, D. Schrf, T. D. Khne, and D. Sebastiani, J. Chem. Phys. **137**, 194110 (2012).
[9]W. Wenzel and K. Hamacher, Phys. Rev. Lett. **82**, 3003 (1999).
[10]D. J. Wales and J. P. Doye, J. Phys. Chem. A **101**, 5111 (1997).
[11]S. Goedecker, J. Chem. Phys. **120**, 9911 (2004).
[12]A. A. Peterson, Top. Catal. **57**, 40 (2014).
[13]S. Stepanenko and B. Engels, J. Comput. Chem. **28**, 601 (2007).
[14]S. Stepanenko and B. Engels, J. Comput. Chem. **29**, 768 (2008).
[15]S. Stepanenko and B. Engels, J. Phys. Chem. A **113**, 11699 (2009).
[16]C. J. Pickard and R. J. Needs, J. Phys.: Condens. Matter **23**, 053201 (2011).
[17]A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).
[18]S. E. Schnborn, S. Goedecker, S. Roy, and A. R. Oganov, J. Chem. Phys. **130**, 144108 (2009).
[19]G. Rossi and R. Ferrando, J. Phys.: Condens. Matter **21**, 084208 (2009).
[20]S. De, A. Willand, M. Amsler, P. Pochet, L. Genovese, and S. Goedecker, Phys. Rev. Lett. **106**, 225502 (2011).
[21]J. A. Flores-Livas, M. Amsler, T. J. Lenosky, L. Lehtovaara, S. Botti, M. A. L. Marques, and S. Goedecker, Phys. Rev. Lett. **108**, 117004 (2012).
[22]S. Botti, J. A. Flores-Livas, M. Amsler, S. Goedecker, and M. A. L. Marques, Phys. Rev. B **86**, 121204 (2012).
[23]N. Yang, A. J. Medford, X. Liu, F. Studt, T. Bligaard, S. F. Bent, and J. K. Nrskov, J. Am. Chem. Soc. **138**, 3705 (2016).
[24]S. F. Carr, R. Garnett, and C. Lo, in *Proceedings of the 33rd International Conference on Machine Learning* (2016).
[25]J. T. Yates, Surf. Sci. **299**, 731 (1994).
[26]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[27]F. Jensen, *Introduction to Computational Chemistry* (Wiley, 2007).
[28]D. Osei-Kuffuor and J. L. Fattebert, Phys. Rev. Lett. **112**, 4 (2014).
[29]J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).
[30]J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Duak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, and H. Hansen, J. Phys.: Condens. Matter **22**, 253202 (2010).
[31]A. H. Larsen, M. Vanin, J. J. Mortensen, K. S. Thygesen, and K. W. Jacobsen, Phys. Rev. B **80**, 195112 (2009).
[32]We performed preliminary calculations with an atomic orbital-like basis set (i.e., double-zeta potential), but the choice of basis set did not dramatically change the statistical predictions of the converged energies; therefore, we proceeded with the grid-based wavefunctions, implemented in GPAW, for computational efficiency.
[33]C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2007).
[34]J. Mockus, V. Tiesis, and A. Zilinskas, Towards Global Optim. **2**, 2 (1978).
[35]D. R. Jones, M. Schonlau, and W. J. Welch, J. Global Optim. **13**, 455 (1998).
[36]J. Snoek, H. Larochelle, and R. P. Adams, *Advances in Neural Information Processing Systems 25* (Curran Associates, Inc., 2012), pp. 2951–2959.
[37]S. R. Bahn and K. W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002), atomic Simulation Environment (ASE).
[38]A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, and G. Ceder, APL Mater. **1**, 011002 (2013).
[39]S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson, Comput. Mater. Sci. **97**, 209 (2015).
[40]J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[41]M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).
[42]V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[43]J. Stausholm-Moeller, H. H. Kristoffersen, B. Hinnemann, G. K. Madsen, and B. Hammer, J. Chem. Phys. **133**, 144708 (2010).
[44]D. C. Sorescu and J. T. Yates, J. Phys. Chem. B **102**, 4556 (1998).
[45]D. C. Sorescu, J. Lee, W. A. Al-Saidi, and K. D. Jordan, J. Chem. Phys. **134**, 104707 (2011).
[46]H. Liu, X. Wang, C. Pan, and K. M. Liew, J. Phys. Chem. C **116**, 8044 (2012).
[47]M. L. Sushko, A. Y. Gal, and A. L. Shluger, J. Phys. Chem. B **110**, 4853 (2006).
[48]J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkens-Diehr, Comput. Sci. Eng. **16**, 62 (2014).