# Low-Scaling Algorithm for Nudged Elastic Band Calculations Using a Surrogate Machine Learning Model

José A. Garrido Torres, Paul C. Jennings, Martin H. Hansen, Jacob R. Boes, and Thomas Bligaard

Phys. Rev. Lett. **122**, 156001 — Published 15 April 2019

DOI: 10.1103/PhysRevLett.122.156001

# Low-Scaling Algorithm for Nudged Elastic Band Calculations Using a Surrogate Machine Learning Model

José A. Garrido Torres,[1,2] Paul C. Jennings,[1,2] Martin H. Hansen,[1,2] Jacob R. Boes,[1,2] and Thomas Bligaard[2,*]

[1]*Stanford University, Department of Chemical Engineering, Stanford, California 94305, USA*

[2]*SUNCAT Center for Interface Science and Catalysis, Stanford Linear Accelerator Center, Menlo Park, California 94025, USA.*

(Dated: March 22, 2019)

## Abstract

We present the incorporation of a surrogate Gaussian Process Regression (GPR) atomistic model to greatly accelerate the rate of convergence of classical Nudged Elastic Band (NEB) calculations. In our surrogate model approach, the cost of converging the elastic band no longer scales with the number of moving images on the path. This provides a far more efficient and robust transition state search. In contrast to a conventional NEB calculation, the algorithm presented here eliminates any need for manipulating the number of images to obtain a converged result. This is achieved by inventing a new convergence criteria that exploits the probabilistic nature of the GPR to use uncertainty estimates of all images in combination with the force in the saddle point in the target model potential. Our method is an order of magnitude faster in terms of function evaluations than the conventional NEB method with no accuracy loss for the converged energy barrier values.

The Nudged Elastic Band (NEB) algorithm is the most popular method for calculating transition states in chemical systems [1–3]. This algorithm is used to find minimum energy pathways (MEP) for the transition between reactants and products, identifying the energy associated with the barrier separating these two states. Many variants of the NEB algorithm have been proposed in the last two decades [3–10]. All of these algorithms rely on an elastic band consisting of interpolated images of the atomic structure, known as moving images. The images are hooked by a spring constant and their positions are optimized by following the gradient of the potential energy surface (PES) while obeying the forces imposed by these springs. A climbing image (CI), without spring forces

and an added force traveling up the gradient along the tangent of the path, can also be included in order ensure the highest energy point is included in the band [3]. The optimization of the path is performed through an iterative process in which all the images are moved and evaluated in each iteration. The coupled iterative nature of the process is very costly, requiring several hundred function calls for the forces even for systems containing few images and degrees of freedom, e.g. describing a single particle diffusion with 10 images.

Further, force evaluations can be computationally very expensive for the first-principle electronic structure calculations. For this purpose, there has been significant work done to build machine learning (ML) surrogate models for atomistic systems [11–16]. These methods function by producing a surrogate model of the PES, which closely approximates the target potential in the region of interest, significantly reducing the number of necessary function calls to achieve convergence. Among all of these models, the critical steps are: (1) moving the atomic positions along the surrogate PES using traditional algebraic or derivative-based solvers, (2) evaluating analytically the forces at the new positions and (3) updating the model with the evaluated point(s) in order to improve the predicting capabilities of the surrogate model. This iterative process is performed until convergence is reached. The premise underlying this protocol is that the optimization cost of the PES surrogate is essentially negligible compared to the cost of an electronic structure calculation.

The aforementioned strategy has served to accelerate NEB calculations using neural networks (NN) as proposed by Peterson *et al.* [12] and using GPR by Jónsson *et al.* [13]. Both approaches have demonstrated the ability to reduce the high computational cost of the classical NEB methods. In particular, Jónsson and coworkers introduced a GPR-assisted algorithm that evaluates the geometry of the image presenting the highest uncertainty of the optimized predicted path each time the NEB is converged in the surrogate model PES. This is known as the one-image-evaluation (OIE) method and surpasses in performance the all-images-evaluated (AIE) method, which relies on calculating all the images of the predicted NEB at each iteration [13]. However, even in this case, all moving images must be evaluated at least once to ensure that the convergence criteria have been satisfied. To the best of our knowledge this also holds true for the other NEB algorithms proposed to date.

One of the main advantages of using GPR is that, as a probabilistic model, the uncertainty estimate for the predictions can be

quantified. In this letter, we demonstrate that the efficiency of the current NEB algorithms can be substantially improved by choosing an acquisition function that optimally utilizes the prediction obtained by the GPR model, i.e. the Gaussian posterior distribution. Here, we propose an algorithm that uses the GPR estimates to detect the geometries of interest to efficiently probe the PES towards a converged MEP. Our algorithm follows the same principles as the aforementioned OIE method proposed by Jónsson and coworkers [13] to avoid calculating all the images of the predicted path in each iteration. The method presented here exploits the regression estimates to define a convergence criterion which is independent of the number of NEB images, therefore solving one the major problems of the previous classical and machine learning NEB methods. This algorithm is implemented in CatLearn [17], which is an open-source Python package for machine learning applications specific to atomic systems. This is, by design, built to interface with the Atomistic Simulation Environment (ASE) [18] and therefore can be easily interfaced with the majority of the electronic-structure calculators, such as CASTEP [19], GPAW [20], Quantum Espresso [21], SIESTA [22], and VASP [23, 24].

Our GPR model considers the positions of the atoms as the descriptors $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ and is trained with their corresponding energies ($\mathbf{e}$) and first derivative observations ($\boldsymbol{\delta}_i$), combining both observations into a vector $\mathbf{y} = [\mathbf{e}\ \boldsymbol{\delta}_1\ \ldots\ \boldsymbol{\delta}_N]$.

The predicted function is *a priori* defined as the Gaussian process:

$$f(x) \sim \mathcal{GP}(P(\mathbf{x}), k(\mathbf{x},\ \mathbf{x}')), \qquad (1)$$

where $k(\mathbf{x},\ \mathbf{x}')$ is the kernel (covariance function) and $P(\mathbf{x})$ is the prior function.

When incorporating first derivative observations to the GP, the covariance matrix takes the form

$$\mathbf{K}(\mathbf{x}) = \begin{pmatrix} \mathbf{K}(\mathbf{x},\mathbf{x}) & \mathbf{K}_{gd}(\mathbf{x},\mathbf{x}) \\ \mathbf{K}_{gd}(\mathbf{x},\mathbf{x})^\top & \mathbf{K}_{dd}(\mathbf{x},\mathbf{x}) \end{pmatrix},$$

with elements of the block matrix being the covariance between the coordinates ($\mathbf{K}(\mathbf{x},\ \mathbf{x})$), and partial derivatives of the covariance with respect to the first coordinate ($\mathbf{K}_{gd}(\mathbf{x},\mathbf{x})$), second coordinate ($\mathbf{K}_{gd}(\mathbf{x},\mathbf{x})^\top$), and the first and second set of coordinates ($\mathbf{K}_{dd}(\mathbf{x},\mathbf{x})$). A more detailed explanation of incorporating derivative observations into GPR can be found in Ref. [25].

Our dataset is defined as $\mathcal{D} = \left\{\left\{\mathbf{x}_n,\ \mathbf{e}_n,\ \boldsymbol{\delta}_n,\ \boldsymbol{\theta}\ \right\}\right\}_{n=1}^{N}$, where $\boldsymbol{\theta}$ contains the set of hyperparameters of the model. The predicted mean and variance of the GP are given by

$$\mathbb{E}[f(\mathbf{x})|\mathcal{D}] = \mathbf{k}(\mathbf{x})[\mathbf{K}(\mathbf{x}) + \sigma_n^2\mathbf{I}]^{-1}\mathbf{y} \qquad (2)$$

and

$$\mathbb{V}[f(\mathbf{x})|\mathcal{D}] = k(\mathbf{x},\,\mathbf{x}) - \mathbf{k}(\mathbf{x})[\mathbf{K}(\mathbf{x}) + \sigma_n^2\mathbf{I}]^{-1}\mathbf{k}(\mathbf{x}), \tag{3}$$

respectively, where $\mathbf{I}$ is the identity matrix and $\sigma_n^2$ is a regularization parameter. The predicted mean (Eq. 2) provides the prediction of the energy for a given position whilst the predicted variance (Eq. 3) offers an estimate of the uncertainty of the same process. The predicted forces are computed using finite differences with a step size of $10^{-4}$ Å. The model parameters selected to describe the GP are included in the Supplemental Material [26].

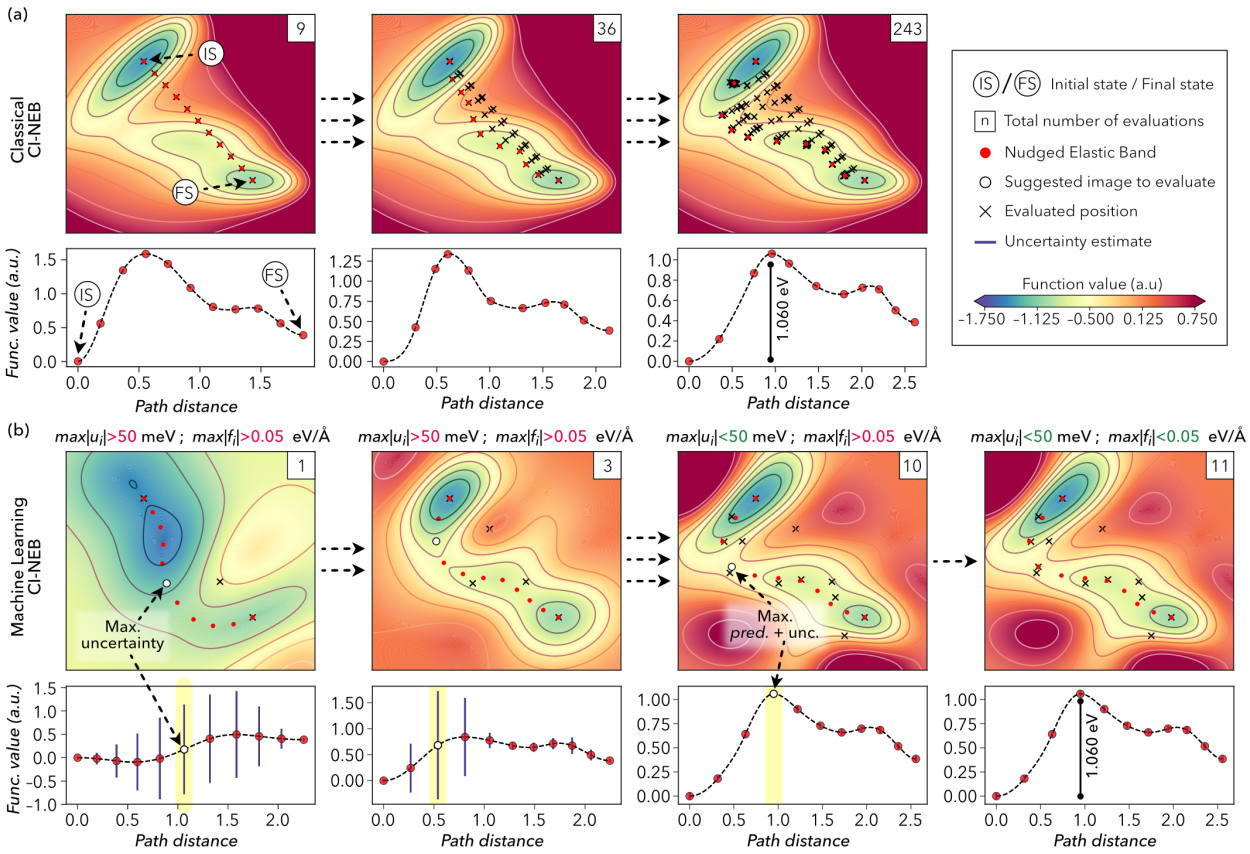


FIG. 1. Comparison between the (a) classical CI-NEB and (b) machine learning NEB (ML-NEB) methods. The performance of both algorithms is illustrated in the two-dimensional Müller-Brown PES. Predicted MEP is included at the bottom of each PES to show the evolution of the energy profile for the elastic band with respect to the number of function calls.

A comparison between the classical NEB and our machine learning accelerated (ML-NEB) methods on the two-dimensional Müller-Brown potential is shown in Figure 1. In this example, we used 9 moving images to describe the transition from the ini-

tial state (IS) to the final state (FS). In the classical NEB method, final convergence is achieved when the maximum forces of the structure of the $i$th NEB image ($\max|F_i^{NEB}|$) perpendicular to the path are below the convergence criteria. This convergence criterion ($\max|F_i^{NEB}|$<0.05 eV/Å) is satisfied after 243 force calls with an energy barrier of 1.060 eV (Figure 1a). The same energy barrier value is obtained by our ML-NEB method after only 11 function calls (see Fig. 1b). In Fig. 1b we illustrate the evolution of the predicted PES and energy profile along the reaction coordinate (red circles) from the IS to the FS obtained after 1, 3, 10 and 11 iterations of our surrogate machine learning model. Our algorithm starts by evaluating an image along the initial interpolated path that is located at one third distance from the end-point with highest energy. This prevents numerical problems during the optimization of the NEB due to a poor initial representation of the predicted PES when the model is trained with only the two end-points of the transition. The model is retrained with the energy and forces of the previously evaluated configurations each time a function evaluation is performed. After training the model, the initial path is optimized on the predicted PES using a velocity-Verlet molecular dynamics algorithm (MDMin, as implemented in ASE). Once the elastic band is converged, the energy and uncertainty estimate (blue bars in Fig. 1b) for each image along the path are stored. On the basis of these predicted values, an acquisition function suggests the next structure to evaluate (see white circles in Fig. 1b). In this example, the acquisition function targets the image along the predicted path with maximum uncertainty until the uncertainty of all the images ($max|u_i|$) is decreased below 0.05 eV. Once this uncertainty convergence criterion is reached, the acquisition function targets the highest energy image (including the uncertainty estimate), until the maximum force of all the relaxed atoms for the last evaluated image goes below the convergence criteria ($max|f_i|$<0.05 eV/Å). This ensures that the saddle-point is obtained with the same accuracy as the classical CI-NEB method.

We demonstrate the performance of our algorithm on three different atomic systems (see Fig. 2a-c) using the Effective Medium Theory (EMT) [27]. We apply our algorithm using three different acquisition functions: The first (Acq. 1) alternates between evaluating the image with the maximum uncertainty and the image with the maximum expected energy value for the transition in each iteration of the surrogate model. This quasi-random sampling mechanism is performed until both convergence criteria are satisfied ($max|u_i|$<0.05 eV and $max|f_i|$<0.05 eV/Å).

The second acquisition function (Acq. 2) is as described above for the example in Figure 1b. The last (Acq. 3) is made of a combination of the two previous acquisition functions, behaving the same as Acq. 2 until the uncertainty convergence criterion is satisfied, and then transitions to Acq. 1 until finding a saddle-point.

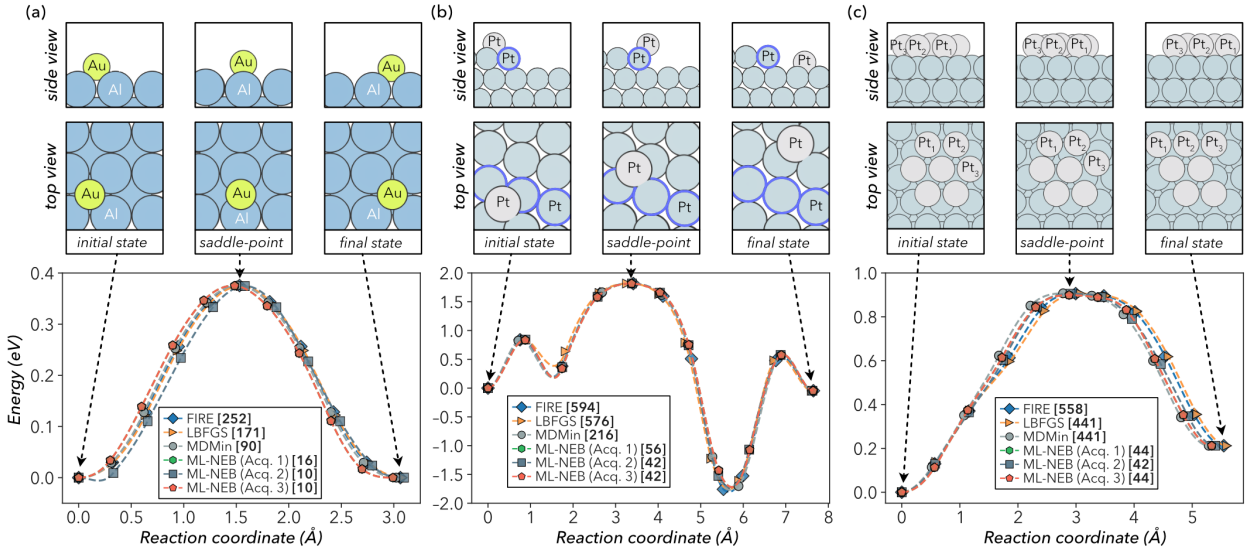


FIG. 2. MEP for the (a) diffusion of a Au atom on an Al(111) surface, (b) diffusion of a Pt adatom on a stepped Pt surface across the two terraces and (c) rearrangement of a Pt heptamer island adsorbed on a Pt(111) surface obtained for the algorithms: FIRE, LBFGS, MDMin, and ML-NEB (using the three acquisition functions presented in the main text). The number of function calls required for each algorithm to converge are shown in bold between brackets. The top and side views of the optimized initial and final states for each transition along with their corresponding saddle-points are included at the top of each composition. The undercoordinated atoms of the Pt step-edge in (b) are highlighted in blue.

Included in Fig. 2 are the optimized paths for three different transitions using FIRE [28], LBFGS [29], and MDMin [18] as implemented in ASE, along with the ML-NEB implementation using the three acquisition functions described above. The different algorithms provide virtually identical estimates of the maximum transition state energy. The same energy barrier values are also obtained when using the classical and ML-NEB algorithms, within numerical precision. The ML-NEB method performs consistently better in terms of function evaluations than the classical algorithms. In particular, when using Acq. 2, the ML-NEB algorithm requires approximately 5-25 times fewer function calls

to achieve convergence than the classical algorithms (see values in brackets in Fig. 2). The improved performance of the acquisition function which makes the most use of the uncertainty estimates also illustrates the potential for GP to accelerate the NEB over other machine learning algorithms over other machine learning algorithms that do not offer an uncertainty estimate.

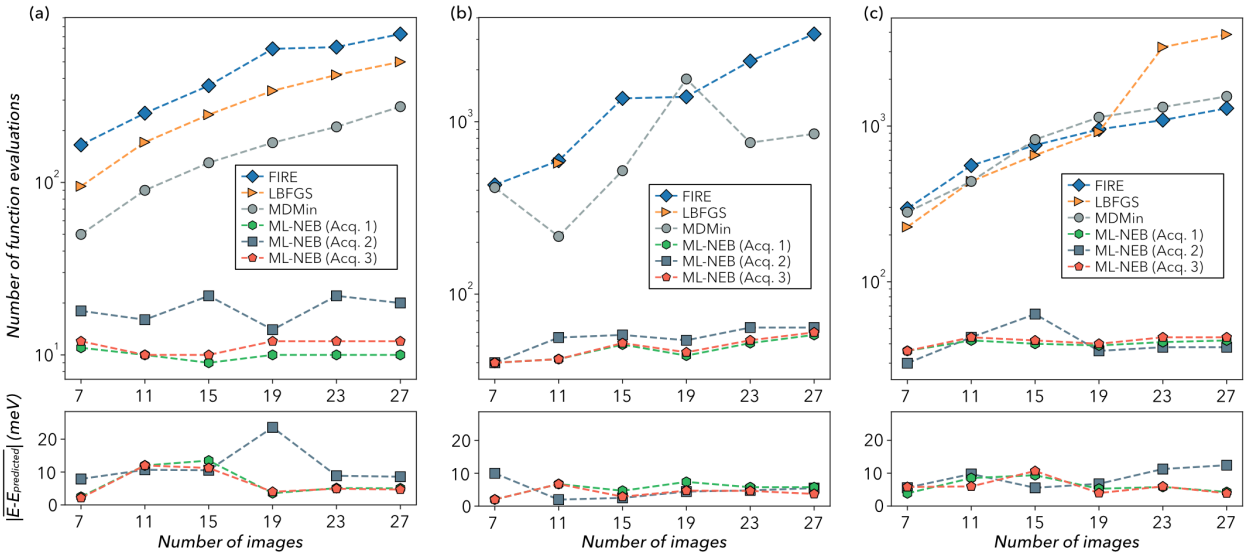


FIG. 3. Comparison of the number of function evaluations required to achieve convergence with increasing number of images for the different classical and machine learning accelerated methods. The benchmark is performed with the classical method (using the FIRE, LBFGS and MDMin algorithms) and the ML-NEB method (using the three acquisition functions described in the main text). The lower panels show the average error of the predicted energy along the path obtained by the three acquisition functions with respect to the target value of the function at the same geometric positions as the ones predicted by the ML-NEB.

The performance of the ML-NEB method is also tested on the previous systems by varying the number of NEB images (see Fig. 3). The number of function calls required to optimize the paths increases exponentially when using the classical implementation of the CI-NEB method. In contrast, the number of function evaluations required by the ML-NEB algorithm is independent of the number of moving images chosen to optimize the path. This allows for the number of images to be optimally chosen whilst performing the NEB optimization at no added cost and can be done by applying similar principles to those proposed by Hammer *et al.* [10] for the classical NEB.

In order to quantify the error magnitude of the GPR estimates, we calculated the energy of the predicted images on the target potential (EMT) using the same geometries as the images along the optimized path. We define the average error of each path as the absolute value of the difference between the energy calculated analytically and GPR predicted energy for the $i$th image along the predicted path. For the three acquisition functions, the maximum error of the predictions lies below the uncertainty convergence criterion imposed (0.05 eV). The two acquisition functions that exploit the maximum uncertainty estimate before targeting the saddle-point, Acq. 2 and 3, performed better than Acq. 1 which alternates targets between the maximum energy and the maximum uncertainty estimates in terms of function evaluations and the accuracy of the predicted path.

For stability, the calculations performed using FIRE, MDMin and ML-NEB converged for all three systems. However, we note that the LBFGS algorithm seems to struggle to find an optimal minimum for the transition represented in Fig. 2b, except when using 11 images. We have also encountered convergence issues with MDMin when performing Density Functional Theory (DFT) calculations for validation [26]. Our algorithm has also been tested on more complex reactions involving bond breaking/forming using DFT [30, 31] as implemented in VASP, also included in the Supplemental Material [26]. Through this variety of examples, our ML-NEB method shows great improvement with respect to the classical optimization in terms of robustness, accuracy, and computational cost.

A good description of a NEB path ultimately relies on including a sufficient number of images. Trying to describe the MEP with a small number of images can lead to convergence problems when optimizing the band on complex energy landscapes [32]. Here, we have presented a machine learning surrogate model that uses the GPR estimates to obtain a converged NEB path which is independent of the number of moving images composing the path. This offers a dramatic improvement in terms of the robustness and efficiency with respect to the classical NEB methods.

In this work, we propose three different acquisition functions in an effort to optimize the decision making protocol in order to obtain an accurate predicted path using the smallest possible number of function calls. We show that the learning rate is driven by the form of the acquisition function and a good selection is dependent on a balance between exploration (reducing the uncertainty of the predicted path) and exploitation (trying to converge the saddle-point). The result of this work is an algorithm which not only surpasses

existing methods in saving function calls, but also improves the robustness in converging an accurate path with respect to the other algorithms, by decoupling the cost in number of function evaluations from the number of moving images on the NEB.

*Acknowledgment:* This work was supported by the U.S. Department of Energy, Chemical Sciences, Geosciences, and Biosciences (CSGB) Division of the Office of Basic Energy Sciences, via Grant DE-AC02-76SF00515 to the SUNCAT Center for Interface Science and Catalysis.

* bligaard@stanford.edu

[1] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and quantum dynamics in condensed phase simulations* (World Scientific, 1998) pp. 385–404.

[2] G. Henkelman and H. Jónsson, The Journal of chemical physics **113**, 9978 (2000).

[3] G. Henkelman, B. P. Uberuaga, and H. Jónsson, The Journal of chemical physics **113**, 9901 (2000).

[4] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras, The Journal of chemical physics **117**, 4651 (2002).

[5] J.-W. Chu, B. L. Trout, and B. R. Brooks, The Journal of chemical physics **119**, 12708 (2003).

[6] S. A. Trygubenko and D. J. Wales, The Journal of chemical physics **120**, 2082 (2004).

[7] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty, The Journal of chemical physics **120**, 7877 (2004).

[8] N. González-García, J. Pu, À. González-Lafont, J. M. Lluch, and D. G. Truhlar, Journal of chemical theory and computation **2**, 895 (2006).

[9] D. Sheppard, P. Xiao, W. Chemelewski, D. D. Johnson, and G. Henkelman, The Journal of chemical physics **136**, 074103 (2012).

[10] E. L. Kolsbjerg, M. N. Groves, and B. Hammer, The Journal of chemical physics **145**, 094107 (2016).

[11] A. Khorshidi and A. A. Peterson, Computer Physics Communications **207**, 310 (2016).

[12] A. A. Peterson, The Journal of chemical physics **145**, 074106 (2016).

[13] O.-P. Koistinen, F. B. Dagbjartsdóttir, V. Ásgeirsson, A. Vehtari, and H. Jónsson, The Journal of Chemical Physics **147**, 152720 (2017).

[14] M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, The Journal of Physical Chemistry A **122**, 1504 (2018).

[15] X. Chen, M. S. Jørgensen, J. Li, and B. Hammer, Journal of chemical theory and computation (2018).

[16] E. Garijo del Río, J. Jørgen Mortensen, and K. W. Jacobsen, ArXiv e-prints (2018), arXiv:1808.08588 [physics.comp-ph].

[17] "A machine learning environment for atomic-scale modelling in surface science and catalysis," `https://github.com/SUNCAT-Center/CatLearn`, [Accessed: May-2018].

[18] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, Journal of Physics: Condensed Matter **29**, 273002 (2017).

[19] M. Segall, P. J. Lindan, M. a. Probert, C. J. Pickard, P. J. Hasnip, S. Clark, and M. Payne, Journal of Physics: Condensed Matter **14**, 2717 (2002).

[20] J. e. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. Hansen, *et al.*, Journal of Physics: Condensed Matter **22**, 253202 (2010).

[21] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, *et al.*, Journal of physics: Condensed matter **21**, 395502 (2009).

[22] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, Journal of Physics: Condensed Matter **14**, 2745 (2002).

[23] G. Kresse and J. Furthmüller, Physical review B **54**, 11169 (1996).

[24] G. Kresse and D. Joubert, Physical Review B **59**, 1758 (1999).

[25] J. Wu, M. Poloczek, A. G. Wilson, and P. Frazier, in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017) pp. 5267–5278.

[26] See Supplemental Material `https://url-supplemental-material` for DFT benchmark and Gaussian Process details, which contains Refs. [33–36].

[27] K. Jacobsen, J. Norskov, and M. J. Puska, Physical Review B **35**, 7423 (1987).

[28] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Physical review letters **97**, 170201 (2006).

[29] D. C. Liu and J. Nocedal, Mathematical programming **45**, 503 (1989).

[30] P. Hohenberg and W. Kohn, Physical review **136**, B864 (1964).

[31] W. Kohn and L. J. Sham, Physical review **140**, A1133 (1965).

[32] D. Sheppard and G. Henkelman, Journal of computational chemistry **32**, 1769 (2011).

[33] P. E. Blöchl, Physical review B **50**, 17953 (1994).

[34] J. P. Perdew, K. Burke, and M. Ernzerhof,

Physical review letters **77**, 3865 (1996).

[35] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, The Journal of chemical physics **132**, 154104 (2010).

[36] S. Grimme, S. Ehrlich, and L. Goerigk, Journal of computational chemistry **32**, 1456 (2011).