

# Bayesian learning of chemisorption for bridging the complexity of electronic descriptors

Siwen Wang<sup>1,2</sup>, Hemanth Somarajan Pillai<sup>1,2</sup> & Hongliang Xin <sup>1</sup>✉

Building upon the *d*-band reactivity theory in surface chemistry and catalysis, we develop a Bayesian learning approach to probing chemisorption processes at atomically tailored metal sites. With representative species, e.g., \*O and \*OH, Bayesian models trained with ab initio adsorption properties of transition metals predict site reactivity at a diverse range of intermetallics and near-surface alloys while naturally providing uncertainty quantification from posterior sampling. More importantly, this conceptual framework sheds light on the orbitalwise nature of chemical bonding at adsorption sites with *d*-states characteristics ranging from bulk-like semi-elliptic bands to free-atom-like discrete energy levels, bridging the complexity of electronic descriptors for the prediction of novel catalytic materials.

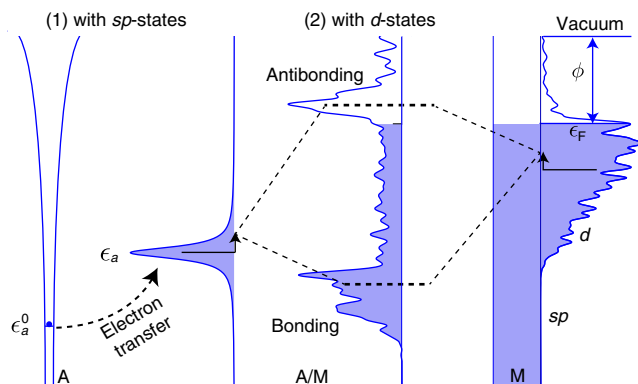
<sup>1</sup>Department of Chemical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA. <sup>2</sup>These authors contributed equally: Siwen Wang, Hemanth Somarajan Pillai. ✉email: [hxin@vt.edu](mailto:hxin@vt.edu)

Adsorption of molecules or their fragments at transition-metal surfaces is a fundamental process for many technological applications, such as chemical sensing, molecular self-assembly, and heterogeneous catalysis. Because of the convoluted interplay between electron transfer and orbital coupling, chemical bonding can be formidably complex. Recent decades have brought major advances in spectroscopic tools<sup>1,2</sup>, which reveal orbitalwise information of chemisorbed systems and concurrently in predicting chemical reactivity at sites of interest via electronic factors, e.g., the number of valence *d*-electrons<sup>3</sup>, density of *d*-states at the Fermi level<sup>4</sup>, *d*-band center<sup>5</sup>, and *d*-band upper edge<sup>6,7</sup>. Compared with a full quantum-mechanics treatment of many-body systems, the simplicity of physics-inspired descriptors comes at a cost of limited generalization, particularly for high-throughput materials screening. Incorporation of multifidelity site features into reactivity models with machine learning (ML) algorithms has shown early promise for the prediction of adsorption energies, with an accuracy comparable to the typical error ( $\sim 0.1\text{--}0.2$  eV) of density functional theory (DFT) calculations<sup>8–16</sup>. However, the approach is largely black-box in nature, prohibiting its physical interpretation. Developing a theory-based, generalizable model of chemisorption that bridges the complexity of electronic descriptors, and predicts the binding affinity of active sites to key reaction intermediates with uncertainty quantification represents one of the biggest challenges in fundamental catalysis.

Here, we present a Bayesian inference approach to probing chemisorption processes at metal sites by learning from ab initio datasets. The model is built upon the basic framework of the *d*-band reactivity theory<sup>5</sup>, while employing a Newns–Anderson-type Hamiltonian<sup>17,18</sup> to capture essential physics of adsorbate–substrate interactions. Such types of simplified Hamiltonians were originally used for describing magnetic properties of impurities in a bulk metallic host<sup>17</sup>, and later extended with success by Newns and Grimley to chemisorption at surfaces<sup>18,19</sup>. A basis set of orbitals consisting of the adsorbate and substrate states was used for solving the hybridization problem within a self-consistent Hartree–Fock scheme<sup>18</sup>. Despite a remarkable success in advancing the basic understanding of adsorption phenomena at surfaces, particularly for *d*-block metals<sup>6</sup>, its application in materials design remains limited due to the lack of accurate model parameters and meaningful error estimates. Bayesian inference produces the posterior probability distribution of model parameters under the influence of observations and prior knowledge<sup>20</sup>. With representative species, e.g., \*O and \*OH, we demonstrate the predictive performance and physical interpretability of Bayesian models for chemical bonding at a diverse range of intermetallics and near-surface alloys, bridging the complexity of electronic descriptors in search of novel catalytic materials.

## Results

**The *d*-band reactivity theory.** Within the basic framework of the *d*-band reactivity theory for transition-metal surfaces, the formation of the adsorbate–metal bond conceptually takes place in two consecutive steps<sup>5</sup>, as illustrated in Fig. 1. First, the adsorbate frontier orbital (or orbitals)  $|a\rangle$  at  $\epsilon_a^0$  couples to the delocalized, free-electron-like *sp*-states of the metal substrate, leading to a Lorentzian-shaped resonance state at  $\epsilon_a$ . Second, the adsorbate resonance state interacts with the localized, narrowly-distributed metal *d*-states, shifting up in energies due to the orthogonalization penalty for satisfying the Pauli principle, and then splitting into bonding and antibonding states. The first step interaction contributes a constant  $\Delta E_0$  albeit often the largest part of chemical bonding. The variation in adsorption energies from one metal to another is determined by the metal *d*-states. This part of



**Fig. 1 Illustration of chemical bonding at transition-metal surfaces within the *d*-band reactivity theory.** An adsorbate A with a valence electron at a discrete energy level  $\epsilon_a^0$  first interacts with the free-electron-like *sp*-states of the substrate M, forming a broadened resonance at  $\epsilon_a$  accompanied with electron transfer. Conceptually, it further overlaps and hybridizes with the narrowly distributed *d*-states, which leads to a splitting into bonding and antibonding states. The work function  $\phi$  and Fermi level  $\epsilon_F$  of M are marked.

the interaction energy  $\Delta E_d$  can be further partitioned into orbital orthogonalization and orbital hybridization contributions<sup>21</sup>. To a first approximation, the orbital hybridization energy can be evaluated by the changes of integrated one-electron energies. The orbital orthogonalization cost is considered simply as proportional to the product of interatomic coupling matrix and overlap matrix,  $VS$ , or equivalently  $\alpha V^2$ , where  $\alpha$  is the orbital overlap coefficient. The absolute value of  $V^2$  can be written as  $\beta V_{ad}^2$ , in which the standard values of  $V_{ad}^2$  relative to Cu are readily available on the Solid State Table<sup>22</sup>. The overall adsorption energy  $\Delta E$  can then be written as the sum of the energy contributions from the *sp*-states  $\Delta E_0$  and the *d*-states  $\Delta E_d$ , with the latter depending on the symmetry and degeneracy of adsorbate frontier orbitals. Another important information from this framework is the evolving density of states projected onto the adsorbate orbital (*s*) upon adsorption,  $\rho_a$ . A full account of the theoretical framework is presented in the “Methods” section.

There are a number of unknown parameters within the basic framework of the *d*-band reactivity theory as discussed above and detailed in “Methods” section, including the energy contribution from the *sp*-band  $\Delta E_0$ , adsorbate resonance energy  $\epsilon_a$  relative to the Fermi level, *sp*-band chemisorption function  $\Delta_0$ , orbital overlap coefficient  $\alpha$ , and orbital coupling coefficient  $\beta$ . By least-squares fitting of the adsorbate density of states and the integrated one-electron energy changes to those from DFT calculations<sup>23,24</sup>, the Schmickler model of electron transfer has been developed to understand H<sub>2</sub> evolution/oxidation and OH<sup>−</sup> adsorption at metal–electrolyte interfaces. However, the deterministic fitting of adsorption properties from a single surface is prone to overfitting or trapping into a locally optimal region, limiting its application in catalysis.

**Bayesian learning.** We instead employ Bayesian learning to infer the vector of model parameters  $\vec{\theta} = (\Delta E_0, \epsilon_a, \Delta_0, \alpha, \beta)'$  from the evidence, i.e., ab initio adsorption properties, along with prior knowledge if available<sup>20</sup>. In Bayes’ view, those parameters are not deterministic point values, but rather probabilistic distributions reflecting the uncertainty of physical variables. The use of parameter distributions as opposed to computationally-derived point values has obvious advantages for uncertainty quantification. In the chemical sciences, Bayesian learning has been used for calibration and validation of thermodynamic models for the uptake

of CO<sub>2</sub> in mesoporous silica-supported amines<sup>25</sup>, designing the Bayesian error estimation functional with van der Waals correlations<sup>26</sup>, and identifying potentially active sites and mechanisms of catalytic reactions<sup>27</sup>, just to name a few. The Bayesian approach allows one to infer the posterior probability distribution  $P(\vec{\theta}|\mathcal{D})$  for latent variables based on the prior  $P(\vec{\theta})$  as well as the likelihood function  $P(\mathcal{D}|\vec{\theta})$  subject to the observation  $\mathcal{D}$ . The mathematical relationship between the prior, observation, and posterior is given by the Bayes' theorem<sup>20</sup>,  $P(\vec{\theta}|\mathcal{D}) = P(\mathcal{D}|\vec{\theta})P(\vec{\theta})/P(\mathcal{D})$ . Our initial belief about likely parameter values is provided by weakly informative priors to minimize potential bias. For example,  $\Delta E_0$  and  $\epsilon_a$  can be estimated from DFT calculations of the adsorbate on a simple metal, e.g., sodium (Na) at the face-centered cubic (fcc) phase. Specifically, we took Normal for floating-point variables unrestricted in sign, Log-Normal for non-negative parameters, and Uniform for others (see the details of Bayesian learning and parameter choices in the "Methods" section). Computing the normalizing constant  $P(\mathcal{D})$ , denominator of the posterior distribution, is impossible in most practical scenarios. To avoid this complication, the Markov chain Monte Carlo (MCMC) method<sup>28</sup>, whose sampling criterion only depends on the relative posterior density of the newly explored point and its preceding point, is used. To compute the transition probability of each MCMC step, we define the sum of the (negative) logarithm of the likelihood functions corresponding to binding energies and projected density of states onto each adsorbate orbital with a hyperparameter  $\lambda$  adjusting the weight of two contributing metrics, see details in the "Methods" section. After a large number of MCMC samplings, burning (discard) of the first half of the trajectory and then thinning (1 out of 5 samplings) were performed before extracting converged values from the joint posterior distributions. The convergence of the MCMC sampling is checked by using parallel chains with different starting parameter sets such that the variance of interchain samplings is close or within 1.2–1.5 times to that of intrachains<sup>28</sup>. The complete code, named *Bayeschem*, is now available at a Github repository <https://github.com/hlxin/bayeschem> for public access.

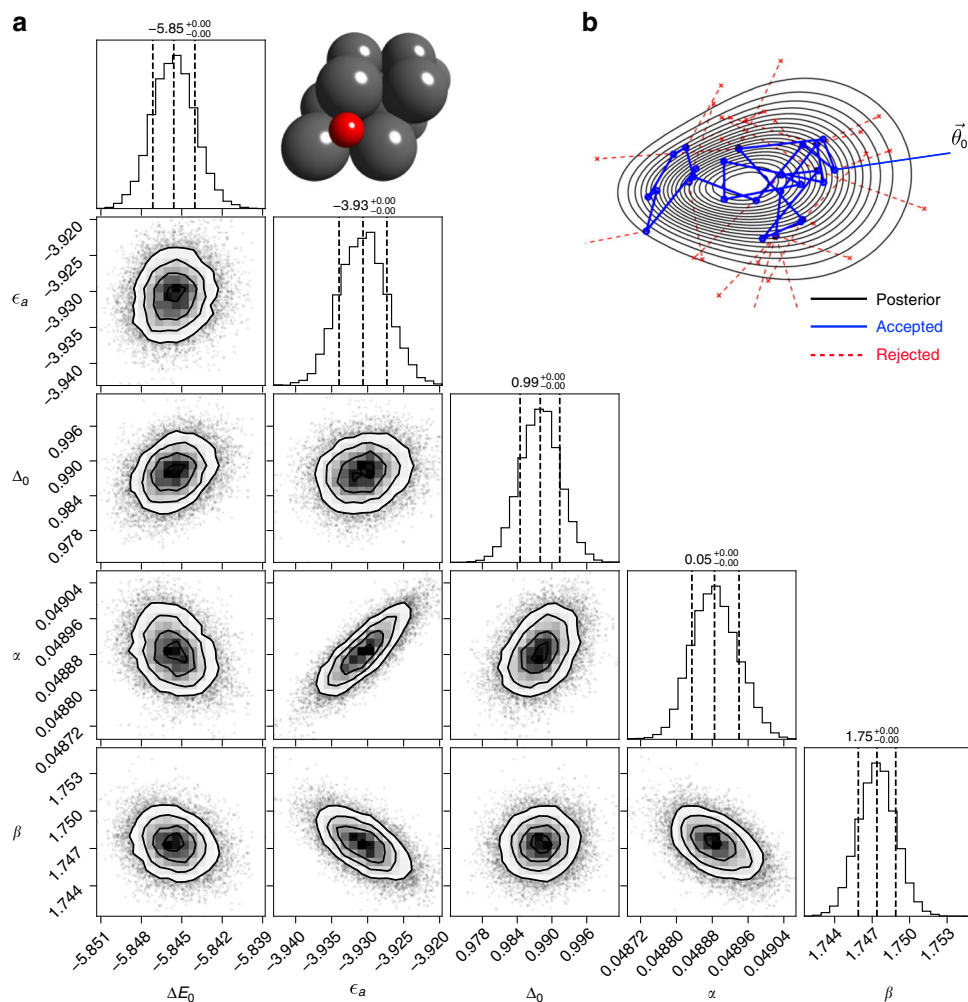
**Model development.** In Fig. 2a, we are showing the co-variance of the joint posterior distribution for each parameter pair and the 1D histogram of model parameters ( $\Delta E_0$ ,  $\epsilon_a$ ,  $\Delta_0$ ,  $\alpha$ , and  $\beta$ ) from MCMC simulations for \*O adsorption at the fcc-hollow site of the {111}-terminated transition-metal surfaces (Cu, Ag, Au, Ni, Pd, Pt, Co, Rh, Ir, and Ru). We assume three degenerate O<sub>2p</sub> orbitals as used before<sup>29</sup> for demonstration of the approach, while later extend it to multiorbital models. To attain converged posterior distributions, 200k MCMC sampling steps with the Metropolis–Hastings algorithm were performed in a multi-dimensional parameter space illustrated in Fig. 2b. In Fig. 2, the approximate contours for 68, 95, and 99% confidence regions are shown at the lower triangle, showing little to no correlation between latent-variable pairs.

With the converged Bayesian sampling, in Fig. 3a, it shows the model-predicted adsorption energies of \*O at the fcc-hollow site of transition-metal surfaces, with a mean absolute error (MAE) ~0.17 eV compared to DFT calculations. The standard deviation of model prediction using the posterior distribution of model parameters ( $\vec{\theta}$ ,  $\vec{\sigma}$ ) is overlaid, providing for the first time uncertainty quantification of adsorption energies within the *d*-band reactivity theory. Figure 3b shows DFT-calculated and model-constructed projected density of states onto the O<sub>2p</sub> orbital using the posterior means of model parameters, taking Pt(111) as an example (see all the surfaces in Supplementary Fig. 1). The

chemisorption function  $\Delta(\epsilon)$  and its Hilbert transform  $\Lambda(\epsilon)$  along with the straight adsorbate line ( $\epsilon - \epsilon_a$ ) are shown for the graphical solution of the Newns–Anderson model<sup>18</sup>. The intersects indicated by solid circles in Fig. 3b represent the O<sub>2p</sub>–Pt<sub>5d</sub> bonding and antibonding states, with the latter above the Fermi level, suggesting a strong covalent interaction of \*O at Pt(111). Given the simplicity of the model, the clearly captured electronic structure of the adsorbate–substrate system and the reactivity trend are satisfying.

To extend the approach for adsorbates with multiple valence orbitals that possibly contribute to bonding, we have explicitly treated O<sub>2p</sub> states with the doubly degenerate *p*<sub>xy</sub> orbitals and the single *p*<sub>z</sub> orbital in Bayesian learning. We infer model parameters ( $\epsilon_a$ ,  $\Delta_0$ , and  $\beta$ ) corresponding to each non-equivalent adsorbate orbital together with an orbital-independent  $\alpha$ <sup>29</sup> and a global parameter  $\Delta E_0$ . The posterior parameter distributions are shown in Supplementary Fig. 2. From the posterior means of model parameters, we can see that the orbital coupling coefficient  $\beta$  of *p*<sub>xy</sub> (1.67 eV<sup>-1</sup>) is smaller than that of *p*<sub>z</sub> (1.77 eV<sup>-1</sup>), consistent with the symmetry analysis, that the *p*<sub>xy</sub> orbitals that are parallel to a surface form  $\pi$  bonds with the *d*-states, while the *p*<sub>z</sub> orbital can interact through a stronger  $\sigma$  bond. A weaker coupling manifests itself in a narrower orbital splitting of  $\pi/\pi^*$  than that of  $\sigma/\sigma^*$ , which has been previously observed using the angle-resolved photoemission spectroscopy on Cu and Ni<sup>30</sup>. In Supplementary Figs. 3 and 4, it shows that the model-constructed projected density of states onto symmetry-resolved orbitals closely resemble the DFT-calculated distributions and the predicted values of \*O adsorption energies have a MAE ~0.17 eV. To demonstrate the robustness and generalizability of the approach, we have also optimized the *Bayeschem* model of \*O at the atop configuration, see Supplementary Figs. 5–7. In this model scheme, an individual set of parameters is obtained for the adsorbate at a given site. Compared to the linear adsorption-energy scaling relations<sup>31</sup> that link adsorption energies of different adsorbates, *Bayeschem* creates the connection between the electronic structure of a surface site and the adsorption energy.

To test the prediction capability of the *Bayeschem* model for unseen systems, we took the \*OH species at the atop adsorption configuration as a case study because of its fundamental importance in understanding the nature of chemical bonding<sup>32</sup>, and practical interests as a key reactivity descriptor in transition metal catalysis<sup>33–35</sup>. Three frontier molecular orbitals, i.e.,  $3\sigma$ ,  $1\pi$ , and  $4\sigma^*$ , are assumed to be involved in chemical bonding<sup>32</sup>. Symmetry-resolved, molecular orbital density of states projected onto OH along with adsorption energies are used as the DFT ground truth *Y* in Eq. (6). With the *Bayeschem* model developed here (see Supplementary Figs. 8–10 for posterior parameter distributions, model-predicted adsorption energies and projected density of states on training samples), we predict \*OH binding energies at a diverse range of intermetallics and near-surface alloys. Specifically, we included A<sub>3</sub>B, A'@A<sub>ML</sub>, A–B@A<sub>ML</sub>, A<sub>3</sub>B@A<sub>ML</sub>, A@A<sub>3</sub>B, and A@AB<sub>3</sub>, where A (A') represents ten fcc/hcp metals used in the model development and B covers *d*-metals across the periodic table (see ref. <sup>36</sup> for structural details and tabulated data). The coupling matrix element *V*<sub>ad</sub> for alloys is assumed to be constant from the Solid State Table<sup>22</sup>. Its dependence on the local chemical environment can be incorporated into the model using the tight-binding approximation<sup>33</sup>. The A sites of above-mentioned surfaces exhibit diverse characteristics of the metal *d*-states ranging from bulk-like semi-elliptic bands to free-atom-like discrete energy levels<sup>37</sup>, as illustrated in Fig. 4a using Pt and Ag<sub>3</sub>Pt as examples. Similar to previous observations of single-atom alloys with coinage metal hosts<sup>37,38</sup>, a reactive guest metal often exhibits peaky signatures within the *d*-band due to the energy misalignment of coupling *d*–



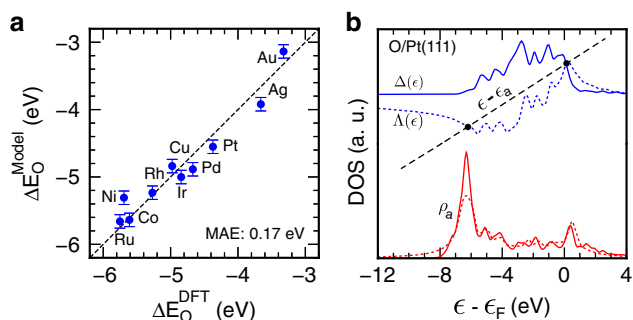
**Fig. 2 Bayesian parameterization.** **a** The co-variance of the joint posterior distribution for each parameter pair and the 1D histogram of model parameters ( $\Delta E_0$ ,  $\epsilon_a$ ,  $\Delta_0$ ,  $\alpha$ , and  $\beta$ ) from MCMC simulations for  $^*O$  adsorption at the fcc-hollow site of the {111}-terminated transition-metal surfaces. A top view of the model structure is shown in inset. **b** Schematic illustration of the MCMC sampling in a multidimensional parameter space.  $\vec{\theta}_0$  represents the initial guess of model parameters.

$d$  orbitals<sup>7</sup>. A direct consequence of such diverse electronic properties of adsorption sites is that no single electronic descriptor can capture the local chemical reactivity accurately. Encouragingly, the *Bayeschem* model, parameterized using ten pristine transition-metal data, predicts  $^*OH$  adsorption energies on 512 alloy surfaces with a MAE 0.16 eV, see Fig. 4b. The standard deviation of predicted  $^*OH$  adsorption energies from the posterior distribution of model parameters is marked for uncertainty quantification. It shows a similar performance to data-driven ML models<sup>8–11</sup> while outperforming the state-of-the-art electronic descriptors, e.g., the  $d$ -band center  $\epsilon_d$  (MAE: 0.20 eV) and upper edge  $\epsilon_u$  (MAE: 0.23 eV). The approach can be easily extended to more complex adsorbates than  $^*O$  and  $^*OH$ , e.g.,  $^*OOH$ , without losing its generalizability in the development workflow.

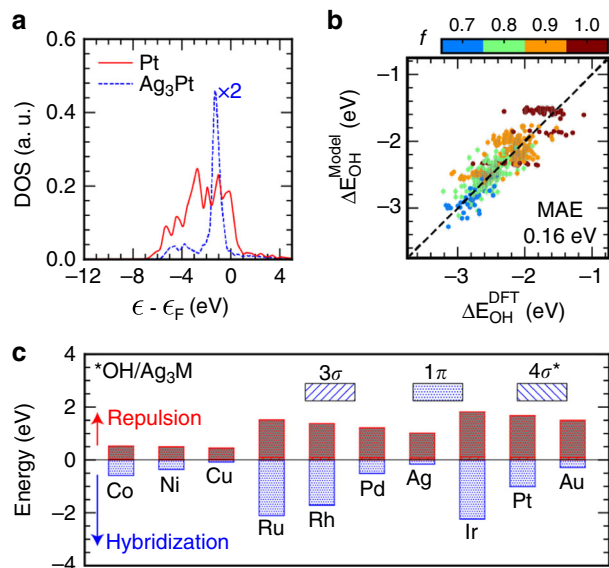
**Orbitalwise interpretation of chemical bonding.** More importantly, the Bayesian framework with built-in physics allows us to quantitatively interrogate the underlying mechanism of chemical bonding, that is difficult to obtain from purely data-driven regression models. Taking  $^*OH$  adsorption at the M (10 fcc/hcp metals) site of {111}-terminated  $Ag_3M$  intermetallics as examples, Fig. 4c shows the partition of  $^*OH$  adsorption energies resulting from the 2<sup>nd</sup> step interaction ( $\Delta E_d$ ) into orbital orthogonalization

and hybridization. As we can see, for 3d, 4d, and 5d series of the guest metal M, the orthogonalization and hybridization contributions decrease in magnitude from left to right across the periodic table, while the hybridization dominates the reactivity trends. The changes in  $\Delta E_d^{hyb}$  can be understood from the simplified  $d$ -band model, with the position and occupancy of adsorbate–substrate antibonding states tracking with the  $d$ -band center or upper edge. The orthogonalization energy is proportional to the filling  $f$  and  $V_{ad}^2$  (see Eq. (4)), which are offsetting each other to a certain extent ( $V_{ad}^2$  decreases while  $f$  increases across 3d, 4d, and 5d series), leading to a less dominant role than the hybridization. The orbitalwise contributions shown in Fig. 4c with different fill patterns suggest that the sole contribution of  $^*OH$  adsorption at  $d$ -metal surfaces is from the  $1\pi$  orbital, while those from  $3\sigma$  and  $4\sigma^*$  are too small to be visible. This is supported by projected molecular orbital density of states in Supplementary Fig. 7, which shows that  $3\sigma$  and  $4\sigma^*$  are forming resonance states after their interactions with the  $sp$ -states of the metal site without noticeable splitting due to  $d$ -states. Thus, they do not contribute to the observed trend of  $^*OH$  adsorption. The Bayesian-optimized orbital coupling coefficients of  $3\sigma$  and  $4\sigma^*$  are rather small (0.12 and 0.001 as shown in Supplementary Fig. 5, respectively), supporting unfavorable orbital overlaps with the  $d$ -states. This rationalizes the observation that





**Fig. 3 Model-predicted adsorption properties.** **a** DFT-calculated  $^*O$  adsorption energies (atomic O as the reference) at transition-metal surfaces vs. model prediction using the posterior distribution of model parameters ( $\vec{\theta}$ ,  $\vec{\sigma}$ ). Error bars represent the standard deviation of model prediction with 1000 random draws from converged trajectories. **b** Projected density of states  $\rho_a$  onto the  $O_{2p}$  orbital from DFT calculations (solid) and model prediction (dashed) using the posterior means of model parameters, taking Pt(111) as an example. The graphical solution to the Newns-Anderson model is also shown, in which  $\Delta(\epsilon)$  and  $\Lambda(\epsilon)$  represent the chemisorption function, and  $\epsilon_a$  is the adsorbate resonance energy level.



**Fig. 4 Model test and interpretation.** **a** The  $d$ -states of a transition-metal site exhibit diverse characteristics ranging from bulk-like semi-elliptic bands to free-atom-like discrete energy levels (Pt and  $Ag_3Pt$  as examples). **b** DFT-calculated vs. model-predicted adsorption energies of  $^*OH$  at the atop site of {111}-terminated intermetallics and near-surface alloys with one standard deviation. **c** Partition of  $^*OH$  adsorption energies at the M site of  $Ag_3M$  into orbital hybridization and orthogonalization of  $3\sigma$ ,  $1\pi$ , and  $4\sigma^*$  orbitals with the metal  $d$ -states. The  $3\sigma$  and  $4\sigma^*$  contributions are too small to be visible.

$^*OH$  prefers the nearly-parallel adsorption geometry on most of the  $d$ -metals to maximize the interaction of the  $1\pi$  orbital with metal  $d$ -states, while  $^*OH$  on Na(111) adsorbs more strongly in a up-straight orientation because of a lack of such directional interactions. This orbitalwise insight of chemical bonding could provide guidance in tailoring orbital-specific characteristics of the metal  $d$ -band for desired catalytic properties through site engineering. Despite an exclusive discussion about the  $d$ -metals, it is possible to extend the *Bayeschem* framework to  $p$ -block metals and alloys see Supplementary Fig. 11, unifying the reactivity theory of metal surfaces.

To conclude, we present the first Bayesian model of chemisorption by learning from ab initio adsorption properties. The model leverages the well-established  $d$ -band reactivity theory and a Newns-Anderson-type Hamiltonian for capturing essential physics of chemisorption processes. We demonstrated that the *Bayeschem* models of descriptor species, e.g.,  $^*O$  and  $^*OH$ , optimized with pristine transition-metal data predicts adsorption energies at a diverse range of atomically-tailored metal sites with a MAE  $\sim 0.1$ – $0.2$  eV while providing uncertainty quantification. Incorporation of physics-based models into data-driven ML algorithms, e.g., deep learning, might hold the promise toward developing highly accurate while interpretable reactivity models. Furthermore, this conceptual framework can be broadly applied to unravel orbital-specific factors governing adsorbate-substrate interactions, paving the path toward design strategies to go beyond adsorption-energy scaling limitations in catalysis.

## Methods

**DFT calculations.** Spin-polarized DFT calculations were performed through Quantum ESPRESSO<sup>39</sup> with ultrasoft pseudopotentials. The exchange-correlation was approximated within the generalized gradient approximation (GGA) with Perdew-Burke-Ernzerhof (PBE)<sup>40</sup>. {111}-terminated metal surfaces were modeled using  $(2 \times 2)$  supercells with four layers and a vacuum of 15 Å between two images. The bottom two layers were fixed while the top two layers and adsorbates were allowed to relax until a force criteria of 0.1 eV/Å. A plane wave energy cutoff of 500 eV was used. A Monkhorst-Pack mesh of  $6 \times 6 \times 1$  was used to sample the Brillouin zone, while for molecules and radicals only the Gamma point was used. Gas phase species of O and OH were used as the reference for adsorption energies of  $^*O$  and  $^*OH$ , respectively. The projected atomic and molecular density of states were obtained by projecting the eigenvectors of the full system at a denser  $k$ -point sampling ( $12 \times 12 \times 1$ ) with an energy spacing 0.01 eV onto the ones of the part, as determined by gas-phase calculations. The convergence of DFT calculations was thoroughly tested to be within 0.05 eV. Further details and tabulated data can be found in the ref. <sup>9</sup>.

**The  $d$ -band reactivity theory.** To revisit the  $d$ -band theory of chemisorption along with new developments, let's consider a metal substrate M in which electrons occupy a set of continuous states with one-electron wavefunctions  $|k\rangle$  and eigen-energies  $\epsilon_k$ , and an isolated adsorbate species A with a valence electron described by an atomic wavefunction  $|a\rangle$  at  $\epsilon_a^0$ , see Fig. 1. When the adsorbate is brought close to the substrate, the two sets of states will overlap and hybridize with each other. The strength of such interactions is determined by the coupling integral  $V_{ak} = \langle a|\hat{H}|k\rangle$ , where  $\hat{H}$  is the system Hamiltonian. Within the Newns-Anderson model of chemisorption<sup>17–19</sup>,  $\hat{H}$  is defined as,

$$\hat{H} = \sum_{\sigma} \left\{ \epsilon_{a\sigma} n_{a\sigma} + \sum_k \epsilon_k n_{k\sigma} + \sum_k (V_{ak} c_{k\sigma}^\dagger c_{a\sigma} + H.c.) \right\}, \quad (1)$$

where  $\sigma$  denotes the electron spin,  $n$  is the orbital occupancy operator, and  $c^\dagger$  and  $c$  represent the creation and annihilation operator, respectively. The first two terms in Eq. (1) are the one-electron energies from the adsorbate and the substrate when they are infinitely separated in space. The last term captures the coupling, or intuitively electron hopping, between the adsorbate orbital  $|a\rangle$  and a continuum of substrate states  $|k\rangle$ . If the one-electron states of the whole system can be described as a linear combination of the unperturbed adsorbate and substrate states, the one-electron Schrödinger equation can be solved using the Green's function approach<sup>18</sup>. In Fig. 1, we illustrate the chemisorption process of a simple adsorbate onto a  $d$ -block metal site characterized by delocalized  $sp$ -states and localized  $d$ -states<sup>21</sup>. The interaction of the adsorbate state at  $\epsilon_a^0$  with the structureless  $sp$ -states, typically accompanied with electron transfer from/to the Fermi sea, results in a broadened resonance (or so-called renormalized adsorbate state) at an effective energy level  $\epsilon_a$ . Conceptually viewing chemical bonding as consecutive steps in Fig. 1, the renormalized adsorbate state then couples with the narrowly distributed  $d$ -states, shifting up in energies due to orbital orthogonalization that increases the kinetic energy of electrons and splitting into bonding and antibonding states. One important information from this framework is the evolving density of states projected onto the adsorbate orbital  $|a\rangle$  upon adsorption

$$\rho_a(\epsilon) = \frac{1}{\pi} \frac{\Delta(\epsilon)}{[\epsilon - (\epsilon_a + \Lambda(\epsilon))]^2 + \Delta(\epsilon)^2}, \quad (2)$$

in which spin is neglected for simplicity. The effective adsorbate energy level,  $\epsilon_a$ , is determined by the image potential of a charged particle in front of conducting surfaces and the Coulomb repulsion between electrons in the same orbital<sup>18</sup>. The chemisorption function  $\Delta(\epsilon)$  includes contributions from the  $sp$ -states and the  $d$ -

states

$$\Delta(\epsilon) = \pi \sum_{\mathbf{k}} V_{\mathbf{ak}}^2 \delta(\epsilon - \epsilon_{\mathbf{k}}) = \Delta_0 + \Delta_{\mathbf{d}}. \quad (3)$$

To simplify the matter, only the 2<sup>nd</sup> step interaction, i.e., the coupling of the renormalized adsorbate state with the substrate  $d$ -states, is explicitly considered in Eq. (2). As a new development in our approach, we include an energy-independent constant  $\Delta_0$  along with  $\Delta_{\mathbf{d}}$  as the chemisorption function  $\Delta(\epsilon)$ . The inclusion of  $\Delta_0$  provides a lifetime broadening of the adsorbate state, serving as a mathematical trick to avoid burdensome sampling of the resonance, i.e., the Lorentzian distribution  $\tilde{\rho}_{\mathbf{a}}$  from the 1<sup>st</sup> step interaction in Fig. 1. Accordingly,  $\epsilon_{\mathbf{a}}$  represents the renormalized adsorbate state. Attributed to the narrowness of a typical metal  $d$ -band,  $\Delta_{\mathbf{d}}$  can be simplified as the projected density of  $d$ -states onto the metal site  $\rho_{\mathbf{d}}(\epsilon)$  modulated by an effective coupling integral squared  $V^2$ , i.e.,  $\Delta_{\mathbf{d}} \approx \pi V^2 \rho_{\mathbf{d}}(\epsilon)$ .  $\Lambda(\epsilon)$  is the Hilbert transform of  $\Delta(\epsilon)$ . In this framework, the interaction energy between the adsorbate and the substrate can be partitioned into two contributions, i.e.,  $\Delta E_0$  and  $\Delta E_{\mathbf{d}}$ .  $\Delta E_0$  is the energy change due to the interaction of the unperturbed adsorbate orbital(s) with the delocalized  $sp$ -states, while  $\Delta E_{\mathbf{d}}$  is the energy contribution from further interactions with the localized  $d$ -states of the substrate. Since all  $d$ -block metals have a similar, free-electron-like  $sp$ -band,  $\Delta E_0$  can be approximated as a surface-independent constant albeit the largest contribution to bonding<sup>21</sup>. To calculate  $\Delta E_{\mathbf{d}}$ , we include both the attractive orbital hybridization  $\Delta E_{\mathbf{d}}^{\text{hyb}}$  and repulsive orbital orthogonalization  $\Delta E_{\mathbf{d}}^{\text{orth}}$ <sup>29,41</sup>:

$$\begin{aligned} \Delta E_{\mathbf{d}}^{\text{hyb}} &= \frac{2}{\pi} \int_{-\infty}^{\epsilon_{\text{F}}} \tan^{-1} \left[ \frac{\Delta(\epsilon)}{\epsilon - \epsilon_{\mathbf{a}} - \Lambda(\epsilon)} \right] d\epsilon - \frac{2}{\pi} \int_{-\infty}^{\epsilon_{\text{F}}} \tan^{-1} \left[ \frac{\Delta(\epsilon)}{\epsilon - \epsilon_{\mathbf{a}}} \right] d\epsilon \\ \Delta E_{\mathbf{d}}^{\text{orth}} &= 2(\langle \tilde{n}_{\mathbf{a}} \rangle + f) \alpha \beta V_{\text{ad}}^2. \end{aligned} \quad (4)$$

The constant 2 considers spin degeneracy of the orbital,  $\langle \tilde{n}_{\mathbf{a}} \rangle$  is the occupancy of the renormalized adsorbate state by integrating the Lorentzian distribution  $\tilde{\rho}_{\mathbf{a}}$  up to the Fermi level  $\epsilon_{\text{F}}$  (taken as 0), and  $f$  is the idealized  $d$ -band filling of the metal atom. The  $\tan^{-1}$  is defined to lie between  $-\pi$  to 0 since  $\Delta_0$  is a nonzero constant across the energy scale  $[-15, 15]$  eV. Thus there is no need to explicitly include localized states even if present below or above the  $d$ -band. In Eq. (4),  $\alpha$  is termed the orbital overlap coefficient, i.e.,  $S \approx \alpha|V|$ , in which the overlap integral  $S$  is linearly proportional to the coupling integral  $V$  for a given orbital. Similarly, the effective coupling integral squared  $V^2$  can be written as  $\beta V_{\text{ad}}^2$ , where  $\beta$  denotes the orbital coupling coefficient and  $V_{\text{ad}}^2$  characterizes the interorbital coupling strength when the bonding atoms are aligned along the  $z$ -axis at a given distance<sup>42</sup>. Its values of  $d$ -block metals relative to that of Cu are readily available on the Solid State Table<sup>22</sup>. It is important to note that  $\beta$  is in the chemisorption function, which determines both the adsorption energy and adsorbate density of states, whereas  $\alpha$  only affects the orbital orthogonalization energy since overlap was not explicitly considered.

**Bayesian learning.** Due to the computationally intensive nature of the MCMC algorithm, there is a need for a more efficient implementation of the Newns–Anderson model than what is obtained by Python and standard libraries like *SciPy* and *NumPy*. We make extensive use of Cython, a C++ extension to the standard Python, to speed up the performance (10–1000 times) of some CPU-intensive functions in the model, e.g., Hilbert transform. To perform MCMC sampling, we use *PyMC*, a flexible and extensible Python package which includes a wide selection of built-in statistical distributions and sampling algorithms<sup>43</sup>, e.g., Metropolis–Hastings. A “burn-in” of the first half of the samplings and then thinning (1 out of 5 samplings) was performed to ensure that subsequent ones are representative of the posterior distribution. Convergence of our MCMC-based sampling was verified using parallel chains<sup>28</sup>. The MCMC sampling results can be directly visualized using *corner*, an open-source Python module. We took *Normal* for floating-point variables unrestricted in sign, *LogNormal* for non-negative parameters, and *Uniform* for others.  $\Delta E_0$  and  $\epsilon_{\mathbf{a}}$  can be estimated from DFT calculations of the adsorbate on a simple metal, e.g., sodium (Na) at the face-centered cubic (fcc) phase. Specifically, for \*O, we used  $\Delta E_0 \sim N(-5.0, 1)$ ,  $\epsilon_{\mathbf{a}} \sim N(-5, 1)$ ,  $\Delta_0 \sim LN(1, 0.25)$ ,  $\beta \sim LN(2, 1)$ , and  $\alpha \sim U(0, 1)$ . For \*OH, we used  $\Delta E_0 \sim N(-3.0, 1)$ ,  $\epsilon_{\mathbf{a}}^{3\sigma} \sim N(-6, 1)$ ,  $\epsilon_{\mathbf{a}}^{1\pi} \sim N(-2, 1)$ , and  $\epsilon_{\mathbf{a}}^{4\sigma^*} \sim N(4, 1)$ . We assume that the predicted adsorption properties from Eqs. (2) and (4) are subject to independent normal errors. Specifically, for the property  $Y$  and the surface  $i$  we have

$$Y_i = \hat{Y}_i(\vec{\theta}) + \sigma \epsilon_i, \quad i = 1, 2, \dots, n, \quad (5)$$

where  $\epsilon_i$  is an independent and standard normal random variable and  $\sigma$  is the standard deviation, allowing for a mismatch between the model prediction  $\hat{Y}_i(\vec{\theta})$  and the DFT ground truth  $Y_i$ . In this approach, we define the likelihood function of the property  $Y$  from  $n$  observations<sup>44</sup>

$$P(Y|\vec{\theta}, \sigma) \propto \sigma^{-n} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \{Y_i - \hat{Y}_i(\vec{\theta})\}^2 \right], \quad (6)$$

where the sum runs over  $n$  training samples for the property  $Y$ , which is either the projected density of states onto an adsorbate orbital or adsorption energies. For adsorption energies,  $Y_i$  and  $\hat{Y}_i$  are scalar values with no ambiguity. For projected density of states, it is a vector of paired values, i.e., the one-electron energy of a state

and its probability density, thus deserving a clarification. The mean squared residuals of model prediction from Eq. (2) for the surface  $i$  is used as  $\{Y_i - \hat{Y}_i(\vec{\theta})\}^2$  in Eq. (6). To compute the transition probability of each MCMC step, we define the sum of the (negative) logarithm of the likelihood functions corresponding to projected density of states onto each adsorbate orbital and binding energies with a hyper parameter  $\lambda$  adjusting the weight of two contributing metrics, i.e.,  $-\ln(P_{\Delta E}) - \lambda \sum \ln(P_{\rho_{\mathbf{a}}})$ . To optimize this parameter, we varied it on a grid of 1.0e–3, 1.0e–2, 1.0e–1, and 1, and found that 1.0e–2 is the optimal value to obtain the best performance in adsorption energy prediction.

## Data availability

The training data of metal surfaces used for model development is available at the Github repository <https://github.com/hlxin/bayeschem> while the test data are from the article <https://doi.org/10.1039/C7TA01812F10.1039/C7TA01812F>.

## Code availability

The complete code of *Bayeschem* is available at a Github repository <https://github.com/hlxin/bayeschem> for public access.

Received: 24 July 2020; Accepted: 12 October 2020;

Published online: 30 November 2020

## References

- Nilsson, A., Pettersson, L. & Nørskov, J. K. *Chemical Bonding at Surfaces and Interfaces*. (Elsevier, Amsterdam, Oxford, 2008).
- Somorjai, G. A. & Li, Y. *Introduction to Surface Chemistry and Catalysis* (Wiley, Hoboken, 2010).
- Calle-Vallejo, F. et al. Number of outer electrons as descriptor for adsorption processes on transition metals and their oxides. *Chem. Sci.* **4**, 1245–1249 (2013).
- Tong, Y. Y., Renouprez, A. J., Martin, G. A. & van der Klink, J. J. In *Studies in Surface Science and Catalysis* (eds Hightower, J. W. et al.) Vol. 101, 901–910 (Elsevier, Amsterdam, 1996).
- Hammer, B. & Nørskov, J. K. Electronic factors determining the reactivity of metal surfaces. *Surf. Sci.* **343**, 211–220 (1995).
- Vojvodic, A., Nørskov, J. K. & Abild-Pedersen, F. Electronic structure effects in transition metal surface chemistry. *Top. Catal.* **57**, 25–32 (2014).
- Xin, H., Vojvodic, A., Voss, J., Nørskov, J. K. & Abild-Pedersen, F. Effects of  $d$ -band shape on the surface reactivity of transition-metal alloys. *Phys. Rev. B Condens. Matter* **89**, 115114 (2014).
- Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-Learning-Augmented chemisorption model for CO<sub>2</sub> electroreduction catalyst screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).
- Li, Z., Wang, S., Chin, W. S., Achenie, L. E. & Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. *J. Mater. Chem. A* **5**, 24131–24138 (2017).
- Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution. *Nat. Catal.* **1**, 696–703 (2018).
- Palizhati, A., Zhong, W., Tran, K., Back, S. & Ulissi, Z. W. Towards predicting intermetallics surface properties with high-throughput DFT and convolutional neural networks. *J. Chem. Inf. Model.* **59**, 4742–4749 (2019).
- Back, S. et al. Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *J. Phys. Chem. Lett.* **10**, 4401–4408 (2019).
- Andersen, M., Levchenko, S. V., Scheffler, M. & Reuter, K. Beyond scaling relations for the description of catalytic materials. *ACS Catal.* **9**, 2752–2759 (2019).
- Gu, G. H. et al. Practical deep-learning representation for fast heterogeneous catalyst screening. *J. Phys. Chem. Lett.* **11**, 3185–3191 (2020).
- Montemore, M. M., Nwaokorie, C. F. & Kayode, G. O. General screening of surface alloys for catalysis. *Catal. Sci. Technol.* **10**, 4467–4476 (2020).
- Esterhuizen, J. A., Goldsmith, B. R. & Lincic, S. Theory-Guided Machine Learning Finds Geometric Structure-Property Relationships for Chemisorption on Subsurface Alloys. *Chem* **6**, 3100–3117 (2020).
- Anderson, P. W. Localized magnetic states in metals. *Phys. Rev.* **124**, 41 (1961).
- Edwards, D. M. & Newns, D. M. Electron interaction in the band theory of chemisorption. *Phys. Lett. A* **24**, 236–237 (1967).
- Grimley, T. B. The indirect interaction between atoms or molecules adsorbed on metals. *Proc. Phys. Soc. Lond.* **90**, 751 (1967).
- Bayes, T. & Price, N. LII. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a

- letter to John Canton, a. m. f. r. S. *Philos. Trans. R. Soc. Lond.* **53**, 370–418 (1763).
21. Hammer, B., Morikawa, Y. & Nørskov, J. K. CO chemisorption at metal surfaces and overlayers. *Phys. Rev. Lett.* **76**, 2141 (1996).
  22. Harrison, W. A. & *Physics. Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond* (Dover Publications, New York, 1989).
  23. Santos, E., Quaino, P. & Schmickler, W. Theory of electrocatalysis: hydrogen evolution and more. *Phys. Chem. Chem. Phys.* **14**, 11224–11233 (2012).
  24. Román, A. M., Dudoff, J., Baz, A. & Holewinski, A. Identifying “optimal” electrocatalysts: impact of operating potential and charge transfer model. *ACS Catal.* **7**, 8641–8652 (2017).
  25. Mebane, D. S. et al. Bayesian calibration of thermodynamic models for the uptake of CO<sub>2</sub> in supported amine sorbents using ab initio priors. *Phys. Chem. Chem. Phys.* **15**, 4355–4366 (2013).
  26. Wellendorff, J. et al. Density functionals for surface science: Exchange–correlation model development with Bayesian error estimation. *Phys. Rev. B Condens. Matter* **85**, 235149 (2012).
  27. Walker, E. A., Mitchell, D., Terejanu, G. A. & Heyden, A. Identifying active sites of the Water–Gas shift reaction over titania supported platinum catalysts under uncertainty. *ACS Catal.* **8**, 3990–3998 (2018).
  28. Gamerman, D. & Lopes, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* 2nd Edn (Chapman and Hall/CRC, London, 2006).
  29. Hammer, B. & Nørskov, J. K. In *Chemisorption and Reactivity on Supported Clusters and Thin Films: Towards an Understanding of Microscopic Processes in Catalysis* (eds Lambert, R. M. & Pacchioni, G.), 285–351 (Springer, Dordrecht, 1997).
  30. Wandelt, K. Photoemission studies of adsorbed oxygen and oxide layers. *Surf. Sci. Rep.* **2**, 1–121 (1982).
  31. Abild-Pedersen, F. et al. Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces. *Phys. Rev. Lett.* **99**, 016105 (2007).
  32. Xin, H. & Linic, S. Communications: exceptions to the *d*-band model of chemisorption on metal surfaces: the dominant role of repulsion between adsorbate states and metal *d*-states. *J. Chem. Phys.* **132**, 221101–221104 (2010).
  33. Xin, H., Holewinski, A. & Linic, S. Predictive structure–reactivity models for rapid screening of Pt-based multimetallic electrocatalysts for the oxygen reduction reaction. *ACS Catal.* **2**, 12–16 (2012).
  34. Tang, M. T., Peng, H., Lamoureux, P. S., Bajdich, M. & Abild-Pedersen, F. From electricity to fuels: descriptors for C1 selectivity in electrochemical CO<sub>2</sub> reduction. *Appl. Catal. B* **279**, 119384 (2020).
  35. Strmcnik, D. et al. Improving the hydrogen oxidation reaction rate by promotion of hydroxyl adsorption. *Nat. Chem.* **5**, 300–306 (2013).
  36. Li, Z., Ma, X. & Xin, H. Feature engineering of machine-learning chemisorption models for catalyst design. *Catal. Today* **280**, 232–238 (2017).
  37. Thirumalai, H. & Kitchin, J. R. Investigating the reactivity of single atom alloys using density functional theory. *Top. Catal.* **61**, 462–474 (2018).
  38. Greiner, M. T. et al. Free-atom-like *d* states in single-atom alloy catalysts. *Nat. Chem.* **10**, 1008–1015 (2018).
  39. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
  40. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
  41. Vojvodic, A., Nørskov, J. K. & Abild-Pedersen, F. Electronic structure effects in transition metal surface chemistry. *Top. Catal.* **57**, 25–32 (2014).
  42. Ma, X. & Xin, H. Orbitalwise coordination number for predicting adsorption properties of metal nanocatalysts. *Phys. Rev. Lett.* **118**, 036101 (2017).
  43. Patil, A., Huard, D. & Fonnesbeck, C. J. PyMC: Bayesian stochastic modelling in python. *J. Stat. Softw.* **35**, 1–81 (2010).
  44. Baggaley, A. W., Sarson, G. R., Shukurov, A., Boys, R. J. & Golightly, A. Bayesian inference for a wave-front model of the neolithization of Europe. *Phys. Rev. E* **86**, 016105 (2012).

## Acknowledgements

S.W., H.S.P., and H.X. acknowledge the financial support from the NSF CAREER program (CBET-1845531). The computational resource used in this work is provided by the advanced research computing at Virginia Polytechnic Institute and State University. H.X. acknowledges the insightful discussion with Prof. John Kitchin from Carnegie Mellon University that inspired the work.

## Author contributions

S.W. and H.S.P. equally contributed to the work. H.X. supervised the research. S.W. and H.X. conceived the idea and designed the general approach. S.W. and H.S.P. conducted DFT calculations and coding. S.W. and H.S.P. performed the detailed analysis. All authors revised the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19524-z>.

Correspondence and requests for materials should be addressed to H.X.

Peer review information *Nature Communications* thanks Christopher Bartel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020