Kronberg, Rasmus; Lappalainen, Heikki; Laasonen, Kari

Hydrogen Adsorption on Defective Nitrogen-Doped Carbon Nanotubes Explained via Machine Learning Augmented DFT Calculations and Game-Theoretic Feature Attributions

Published in:
Journal of Physical Chemistry C

DOI:
10.1021/acs.jpcc.1c03858

Published: 29/07/2021

Document Version
Publisher's PDF, also known as Version of record

THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC

Article

Hydrogen Adsorption on Defective Nitrogen-Doped Carbon Nanotubes Explained via Machine Learning Augmented DFT Calculations and Game-Theoretic Feature Attributions

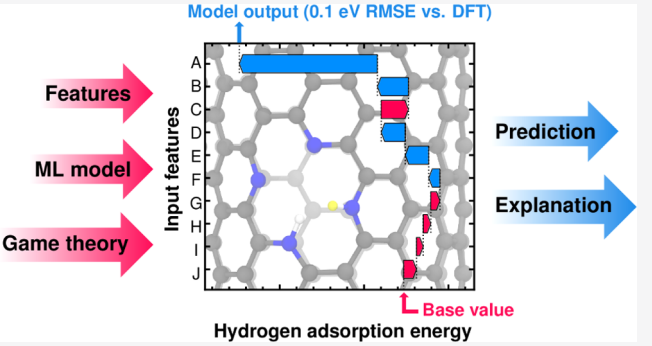
Rasmus Kronberg, Heikki Lappalainen, and Kari Laasonen*

Cite This: https://doi.org/10.1021/acs.jpcc.1c03858

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Complex machine learning (ML) models applied within computational chemistry and materials science tend to be seen as black boxes, yielding property predictions given some input features. While the purpose of ML methods is often to circumvent computationally expensive first-principles calculations, the fact that the inner workings of the models are not understood conceals chemical insight and knowledge regarding the underlying data and physical correlations within it. Knowing what a model is learning from the data and how outputs are formed is also useful in facilitating the justification and wider adoption of ML solutions. Here, we present an important contribution in this direction by exploring and explaining the hydrogen adsorption properties of defective nitrogen-doped carbon nanotubes (NCNTs) through density functional theory simulations and machine learning-based data analysis. As the main highlight, we demonstrate the application of a recent game-theoretic approach to deconvolute and interrogate the trained ML models, revealing how various structural, chemical, and electronic features contribute toward the hydrogen affinities of roughly 6500 different NCNT adsorption sites. The employed method of Shapley additive explanations (SHAP) attributes locally accurate importances to the investigated features, unraveling high spin polarization, narrow highest occupied molecular orbital−lowest unoccupied molecular orbital (HOMO−LUMO) gap, small dopant−adsorption site separation, and diverse angle and coordination effects as particularly impactful for increasing hydrogen adsorption strengths. The SHAP method is shown capable of promoting a deep understanding of complex feature−activity relationships, facilitating research efforts such as rational catalyst design for energy conversion applications.



1. INTRODUCTION

Rational materials development is pivotal in facilitating a sustainable deployment of modern energy conversion technologies empowering envisioned concepts such as the hydrogen economy.[1] To substitute current best available technologies, novel materials must satisfy criteria of physicochemical activity, operational stability, cost-effectiveness, and environmental friendliness, a multiobjective optimization task that has led to the introduction of materials of ever-increasing complexity. This important effort demands that atomistic structure−activity relations are comprehensively elucidated so that a targeted fine-tuning of the distinct structural and compositional properties is enabled. However, decoupling different microscopic features and their individual contributions to the macroscopic response of a complex material is complicated and time-consuming.[2]

While computational chemistry and density functional theory (DFT) have mitigated this problem by providing complementary tools to assess nanoscale structural and electronic properties of, e.g., catalyst materials,[3] the uncertain correspondence between models and complex experimental systems presents a considerable challenge.[2,4] As experimental materials are seldom uniform or monodisperse in nature, computations should ideally consider an ensemble of probable model systems to determine which configurations and moieties might contribute most to the macroscopic observables.[5] Clearly, as the number of degrees of freedom grows rapidly, high-throughput tools for screening and analyzing individual configurations and features are highly desired, a hurdle optimally tackled by machine learning (ML)-based approaches.[6]

Machine learning models are applied within computational chemistry and materials science to an increasing extent for the purpose of accelerated catalyst development,[7−9] force-field training,[10−12] and density functional fitting,[13−15] to name a few

Received: April 30, 2021
Revised: June 18, 2021

examples. However, a problem of most ML models is their apparent black-box nature.[16] Complex ML algorithms such as artificial neural networks and ensemble models are often observed to be capable of learning and predicting target properties with high precision, but interpreting the models and explaining why they work as they do is challenging and frequently omitted. Interpretability aspects of ML models have consequently gained significant interest to increase the level of understanding of how ML models compose outputs based on input features, giving rise to the subfield of explainable artificial intelligence (XAI).[17,18] Resolving feature weights is essential to understand which features have the greatest impact on the target variable and therefore deserve the most attention. A model that is interpretable and explainable is likely to be also more convincing and easily adoptable as it provides insights concerning the physics and chemistry behind the investigated data as well as how the model could be potentially improved. For a more thorough overview of state-of-the-art developments within ML-driven materials science, the reader is referred to the recent comprehensive reviews presented by Marques et al.[19] and others.[20,21] The former is particularly recommended for its extensive discussion on the importance of model interpretability.

In this work, we present a significant contribution toward this direction by applying a recent game-theoretic approach based on Shapley values to accurate feature attribution within machine learning augmented computational catalyst design. As our target, we consider hydrogen adsorption on defective nitrogen-doped carbon nanotubes (NCNTs), a noble metal-free system with electrocatalytic potential for reactions such as hydrogen and oxygen evolution (HER and OER),[22] as well as oxygen reduction (ORR).[23] HER constitutes the cathodic half-reaction of electrolytic water-splitting, composed of an initial hydrogen adsorption process, the Volmer reaction ($H^+ + e^- \rightarrow H^*$), followed by desorption via two alternative paths, either the Tafel ($2H^* \rightarrow H_2$) or Heyrovský ($H^* + H^+ + e^- \rightarrow H_2$) step. Considering that adsorbed hydrogen serves as a key intermediate in the HER, hydrogen adsorption energies are frequently used as descriptors in high-throughput catalyst screening, alluding to whether the reaction could be facile from a thermodynamic perspective.[24] In view of the Sabatier principle, heterogeneously catalyzed reactions should exhibit optimal adsorption characteristics, as too strong adsorption will decrease the driving force of the desorption step, while a too low adsorbate affinity will hinder the required stabilization and activation of the adsorbing species on the catalyst surface. For the HER, this balance has been empirically shown to call for a hydrogen adsorption free energy of $\Delta G \sim 0$ eV, although recent theoretical evidence has suggested somewhat more positive values.[25] Nevertheless, it is clear that adsorption energies of excessive magnitudes will have a detrimental effect on HER rates.

Although heteroatom doping has been shown to have a promoting effect on the insufficient catalytic activity of pure carbon nanomaterials,[26] experimentally measured activation overpotentials gauging the surplus energy needed to drive HER on NCNTs remain inferior to state-of-the-art platinum-based catalysts.[22,27,28] This performance gap suggests suboptimal hydrogen adsorption properties, and there is evidently still plenty of understanding to be gained regarding the activity-determining features of NCNTs and how a rational fine-tuning of these could enable further performance improvements. To this end, we perform DFT calculations at both generalized gradient approximation (GGA) and hybrid Hartree−Fock/DFT (HF/DFT) levels of theory, probing the hydrogen affinity of roughly 6500 different adsorption sites on 16 NCNT systems. Potentially important structural, chemical, and electronic features affecting the target variable are parsed from the DFT output and used to train random forest machine learning models for the regression task of predicting the DFT adsorption energies. The fitted models and acquired data are analyzed using Shapley additive explanations (SHAP),[17,29] revealing intricate details on how distinct features impact the model output at both local (individual sample) and global (average impact across data set) levels. Furthermore, we examine higher-order interactions between features that explain how multivariate effects can contribute to feature−activity relationships.

2. THEORETICAL METHODS

2.1. Random Forest Ensemble Learning. A random forest (RF) is a machine learning method that operates by generating an ensemble of decision trees based on training data composed of several features for each sample and outputting the mean prediction of the individual estimators for a target quantity.[30,31] To reduce correlation between individual trees, the random subspace method is applied during tree generation such that only a stochastically chosen subset of all input features is considered when splitting each node. A second level of randomness is injected by using a bootstrap sample of the training data for each tree, i.e., training samples are selected with replacement so that each data point may be chosen a variable amount of times. On average, $1/e \sim 37$ % of the training samples will not be used for the construction of a given tree, forming an "out-of-bag" (OOB) set of samples that can be used for internal validation. By training a random forest of stochastically distinct decision trees, both numerical (regression) and categorical (classification) target variables of unseen test data can be predicted robustly with a reduced risk of overfitting compared to conventional decision trees or, e.g., deep neural networks.[32]

2.2. Feature Attribution Methods. *2.2.1. Impurity-Based Metrics.* Estimating feature importances is a delicate matter. By default, most random forest regressor implementations calculate feature importances based on the mean decrease in impurity (MDI) as measured by variance reduction, i.e., the drop in the mean-squared error (MSE) when splitting a node. More generally, an impurity measure $i(n)$ is a figure of merit of any node n such that the smaller the impurity, the purer the node is. With the full training set initially represented by a single node, successive binary splits s_n that maximize the decrease in impurity are made, partitioning the data into increasingly homogeneous groups until the terminal nodes (leaves) cannot be made any purer. Denoting the impurity decrease as $\Delta i(s_n)$, the importance I_j of a given feature j for predicting the target variable is then defined as the sum of weighted impurity decreases for all nodes where j is used, averaged over all trees $\{\hat{y}_t\}_{t=1}^{T}$ in the random forest[30,33]

$$I_j = \frac{1}{T}\sum_{t=1}^{T}\sum_{n\in \hat{y}_t|j} w(n)\Delta i(s_n) \tag{1}$$

where the weight $w(n)$ is the fraction of samples reaching node n. Thus, features used at the top of a decision tree contribute to the final prediction of a larger portion of the input samples. Consequently, eq 1 corresponds to a normalized feature

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

importance estimate accounting for the expected fraction of samples feature j contributes to in combination with the decrease in impurity gained upon splitting them. For classification tasks, one of the most common impurity measures is the Gini score,[34] and when employed, eq 1 is correspondingly known as the Gini importance. A further overview of the different flavors of impurity measures is, however, beyond the scope of the present brief discussion.

Unfortunately, impurity-based feature importances suffer from a major drawback as they tend to inflate the importances of high cardinality numerical features, i.e., variables that may attain a broad spectrum of unique values in contrast to, e.g., binary or finite categorical features.[33,35] This bias toward continuous high cardinality features arises from their inherent ability to offer more potential cut-points that may, by chance, produce a large impurity decrease when the node is split. Consequently, even randomly generated numbers can be ranked misleadingly as highly important as long as the model is complex enough to be able to overfit using these, resulting in severe inconsistencies in feature attributions. This is also because impurity-based feature importances are evaluated at the model training time and are thus biased toward the training data and do not appropriately reflect the usefulness of the features to make generalizable predictions. Impurity-based attributions fail also in treating strongly correlated features, yielding underestimated importance values for actually impactful multicollinear variables that offer access to equivalent information via different channels.

2.2.2. Shapley Additive Explanations. An effective mitigation to the above issues has been recently presented by Lundberg and Lee[29] and is based on Shapley values known from cooperative game theory.[36] Shapley values measure the contribution of each member of a coalition in a collaborative game, providing a theoretically well-founded division of credit among the members based on the average of all contributions made by an individual. In the context of machine learning, each feature can be considered to represent a single player of a coalition of features that cooperate toward forming a prediction. Consequently, the contribution of each player can be identified as a measure of the feature importances, the Shapley values. To calculate the Shapley value of a feature j, one considers a subset $\mathcal{S}$ of all features $\mathcal{F}$ and evaluates a general model f both with the feature j present in the subset ($\mathcal{S} \cup \{j\}$) and with the feature absent ($\mathcal{S}$). The marginal contribution of the feature in a specific coalition is then $f(\mathcal{S}\cup\{j\}) - f(\mathcal{S})$, and averaging over all possible permutations in which a coalition $\mathcal{S}$ can be formed yields the Shapley value ϕ_j as

$$\phi_j = \sum_{\mathcal{S}\subseteq\mathcal{F}\setminus\{j\}} \frac{|\mathcal{S}|!(|\mathcal{F}| - |\mathcal{S}| - 1)!}{|\mathcal{F}|!}[f(\mathcal{S}\cup\{j\}) - f(\mathcal{S})] \tag{2}$$

where $|\cdot|$ denotes the number of elements in a set. To estimate feature importances based on Shapley values in practice, we apply the concept of Shapley additive explanations (SHAP), as implemented in the efficient `SHAP` Python package optimized for tree ensemble methods.[17] The SHAP values are a special case of eq 2 in which the general set function $f(\mathcal{S})$ is defined as a conditional expectation function of the machine learning model $\hat{f}(\mathbf{x})$ given a subset $\mathcal{S}$ of the features, $\mathbb{E}[\hat{f}(\mathbf{x})|\mathbf{x}_\mathcal{S}]$. In other words, each feature is sequentially introduced into the conditional expectation function and the induced change in the expected output is defined as the SHAP value of that feature after averaging over all possible feature permutations. While this task is formally NP-hard with a notorious exponential $O(TLF2^F)$ scaling, where T, L, and F are the number of trees, terminal nodes (leaves), and input features, respectively, the computational complexity is mitigated in the decision tree-specific implementation of `SHAP` by recursively tracking the fraction of all possible feature subsets that end up in each leaf node. This is essentially equivalent to evaluating the conditional expectation function for all 2^F feature subset permutations simultaneously, yielding a more tractable $O(TLD^2)$ polynomial complexity where D is the maximum depth of any tree.

The SHAP values satisfy three important properties necessary for feature attribution methods: local accuracy, missingness, and consistency.[29] Local accuracy, or additivity, entails that the model output $\hat{f}(\mathbf{x})$ for each prediction equals exactly the sum of the individual feature contributions plus a base value corresponding to the expected output (weighted average of all predictions) in the absence of input features

$$\hat{f}(\mathbf{x}) = \mathbb{E}[\hat{f}(\mathbf{x})] + \sum_{j\in\mathcal{F}} \phi_j(\hat{f}, \mathbf{x}) \tag{3}$$

The property of missingness, on the other hand, requires that features absent in the original input, i.e., $f(S \cup \{j\}) = f(S)$, have no attributed impact ($\phi_j = 0$). Perhaps most importantly, consistency ensures that a change in a model that *increases* some feature's contribution does not *decrease* the importance of this feature. The requirement of consistency is not satisfied by other feature importance evaluation methods based on standard concepts such as the MDI.[17] Indeed, it has been shown that Shapley values are the only feature attribution method that satisfies all of the above three properties simultaneously, enabling reliable and meaningful comparisons of importances across the feature space.

It is notable that the SHAP values are local feature attributions explaining the importance of input features for individual predictions. Focusing on a single observation i, the SHAP values reflect the contribution of each feature toward the model output for that sample. In analogy with previous *global* model interpretations measuring the average impact of each feature across the whole data set, the local SHAP explanations can also be combined for a global understanding of the model structure and the importance of each feature for the full data set. This is most simply achieved by averaging the magnitudes of all SHAP values for each feature j

$$\langle|\phi_j|\rangle = \frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}} |\phi_j^{(i)}| \tag{4}$$

where $|\mathcal{D}|$ is the number of samples in data set $\mathcal{D}$.

Although the SHAP values are a rather recent feature attribution method, their desirable properties have facilitated a successful application of the concept in the context of machine learning augmented computational chemistry for explaining model predictions of band gaps and exciton binding energies in two-dimensional (2D) materials,[37] optic dielectric constants of crystals,[38] nonequilibrium proton-coupled electron transfer dynamics in proton wires,[39] exchange−correlation functional fitting,[14] and heterogeneous ice nucleation on diverse solid substrates.[40]

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

2.2.3. SHAP Interaction Values. While the SHAP values attribute to each input feature the impact that feature has on the model output, they do not directly reveal information on the possible interactions between features. Indeed, the local effect of a feature on the model prediction may alter depending on the values of the other features, i.e., the context of the prediction. Consequently, it may be highly informative to separate interaction effects of features from the main effect a feature has on the model output. To this end, Shapley interaction indices Φ_{ij} extend Shapley values by allocating credit among all pairs of features. Analogous to eq 2, SHAP interaction values are defined as[17]

$$\Phi_{ij} = \sum_{\mathcal{S}\subseteq\mathcal{F}\setminus\{i,j\}} \frac{|\mathcal{S}|!(|\mathcal{F}| - |\mathcal{S}| - 2)!}{2(|\mathcal{F}| - 1)!} \nabla_{ij}(\mathcal{S}) \tag{5}$$

where $i \neq j$ and

$$\nabla_{ij}(\mathcal{S}) = f(\mathcal{S}\cup\{i, j\}) - f(\mathcal{S}\cup\{i\}) - [f(\mathcal{S}\cup\{j\}) - f(\mathcal{S})] \tag{6}$$

with $f(\mathcal{S}) = \mathbb{E}[\hat{f}(\mathbf{x})|\mathbf{x}_{\mathcal{S}}]$ as before. In other words, the marginal effect of feature j in the absence of feature i within the square brackets in eq 6 is subtracted from the effect of j when i is present, averaging over all possible permutations of feature coalitions $\mathcal{S}$ to which i and j are introduced. Consequently, Φ is an $F \times F$ matrix, the diagonal elements of which correspond to the main effect of each feature, respectively. These are obtained by subtracting all off-diagonal interaction effects from the corresponding SHAP values ϕ_i

$$\Phi_{ii} = \phi_i - \sum_{j\neq i} \Phi_{ij} \tag{7}$$

As the interaction effects are split equally between all pairs of features, $\Phi_{ij} = \Phi_{ji}$, the SHAP interaction matrix is symmetric and the total interaction of a certain feature pair is given by $\Phi_{ij} + \Phi_{ji} = 2\Phi_{ij}$.

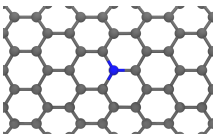
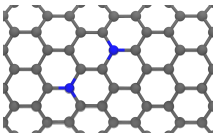
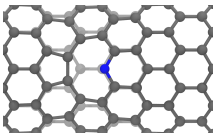
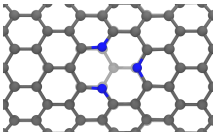
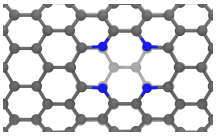
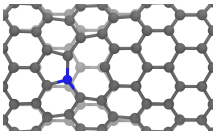
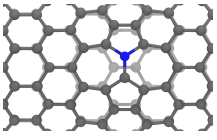
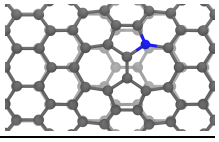
3. COMPUTATIONAL DETAILS

3.1. Model Systems. Hydrogen adsorption was studied on 16 different nitrogen-doped and defective CNTs, each consisting of up to 224 atoms in a periodically repeated simulation cell with dimensions of approximately 3.1 × 3.1 × 1.7 nm^3. Two different CNT types were considered, namely, the achiral zigzag (14, 0) and armchair (8, 8) CNTs, both having experimentally realistic diameters of ca. 1.1 nm. Roughly 2.0 nm of vacuum was applied around the nanotubes to decouple periodically repeated images of the systems in the *xy*-plane. Of the various dopant and defect configurations, two substitutional graphitic, three pyridinic, and three pyrrolic nitrogen moieties were explored with single vacancy, double vacancy, and Stone−Wales rotation-type defects, resulting in dopant and vacancy concentrations ranging between 0.4−1.8 and 0−0.9 atom %, respectively. The dopant and defect types were selected based on previous experimental and computational studies[41−45] on probable doping configurations in NCNTs. The investigated substrates are summarized in Table 1.

Given the defined reference configurations, the hydrogen affinity was probed considering all possible sites in the positive half-space containing the nitrogen/defect configuration, amounting to roughly 100 sites per NCNT. Having determined the most stable adsorption site for each nanotube, the minimum energy adsorbed systems were selected as substrates for a second round of adsorption energy calculations, thus now including two hydrogen adsorbates. This procedure was repeated for the cases of three and finally four adsorbed hydrogen intermediates to assess the impact of hydrogen coverage. A total of 6517 distinct adsorption configurations were optimized, and the (differential) adsorption energies ΔE were evaluated conventionally following eq 8

$$\Delta E = E_{\mathrm{NCNT}+n\mathrm{H}} - E_{\mathrm{NCNT}+(n-1)\mathrm{H}} - \frac{1}{2}E_{\mathrm{H_2}} \tag{8}$$

Table 1. Specification and Illustration of the Studied Model NCNT Systems[a]

N moiety	Defect	(n, m)	Image
Graphitic, N_1	None	(14, 0) (8, 8)	
Graphitic, N_2	None	(14, 0) (8, 8)	
Pyridinic, N_1	V_1	(14, 0) (8, 8)	
Pyridinic, N_3	V_1	(14, 0) (8, 8)	
Pyridinic, N_4	V_2	(14, 0) (8, 8)	
Pyrrolic, N_1	V_1	(14, 0) (8, 8)	
Pyrrolic, N_{1a}	SW	(14, 0) (8, 8)	
Pyrrolic, N_{1b}	SW	(14, 0) (8, 8)	

[a]Blue and gray spheres correspond to carbon and nitrogen atoms, respectively. The abbreviations V_1, V_2, and SW denote a single vacancy, double vacancy, and Stone−Wales rotation, respectively. For the Stone−Wales defect, two distinct nitrogen positions resembling indole (N_{1a}) and indolizine (N_{1b}) structures were considered.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

Table 2. Mathematical Formulation and Description of All 25 Input Features Used in the RF Models[a]

feature	definition	unit	description
x_k	N_k/N_{NCNT}	atom %	atomic concentration of k = N, V, or H[b]
min d_k	$\min_k \sqrt{r^2\theta_k^2 + z_k^2}$	Å	shortest curvilinear distance along the NCNT surface between S and k = N or H[c]
min l	$\min_k \lvert\mathbf{R}_S - \mathbf{R}_k\rvert$	Å	shortest S to $k \in$ NN bond length[d]
max l	$\max_k \lvert\mathbf{R}_S - \mathbf{R}_k\rvert$	Å	longest S to $k \in$ NN bond length
RMSD	$\sqrt{\langle(\Delta\mathbf{R}_k)^2\rangle}$	Å	adsorption-induced root-mean-squared displacement of atomic positions
RmaxSD	$\sqrt{\max_k(\Delta\mathbf{R}_k)^2}$	Å	adsorption-induced root-maximum-squared displacement of atomic position
χ	$\arctan\left(\frac{\sqrt{3}m}{2n+m}\right)$	rad	chiral angle of (n,m) NCNT
min φ_k	$\min_{j\neq i} \arccos(\hat{\mathbf{u}}_{ik}\cdot\hat{\mathbf{u}}_{jk})$	rad	smallest angle where k = S or nearest N defines the vertex and $i,j \in$ NN of k
max φ_k	$\max_{j\neq i} \arccos(\hat{\mathbf{u}}_{ik}\cdot\hat{\mathbf{u}}_{jk})$	rad	largest angle where k = S or nearest N defines the vertex and $i,j \in$ NN of k
α_k	$\arccos(\hat{\mathbf{z}}\cdot\hat{\mathbf{v}}_k)$	rad	angular displacement of S with respect to the nearest k = N or H
$\overline{\mathrm{CN}}_k$	$\sum_i \frac{\mathrm{CN}_i}{\mathrm{CN}_{i,\max}}$		generalized coordination number of k = S or nearest N with $i \in$ NN
$\Delta\overline{\mathrm{CN}}_k$			adsorption-induced change in $\overline{\mathrm{CN}}_k$
Z			atomic number of the adsorption site
M	$2S + 1$		spin multiplicity of the system
q	$n_{\mathrm{val}} - (n_\uparrow + n_\downarrow)$	e	residual charge on the adsorption site (iterative Hirshfeld partitioning)
μ	$n_\uparrow - n_\downarrow$	e	spin polarization on the adsorption site (iterative Hirshfeld partitioning)
E_g	$E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}}$	eV	energy gap (for open-shell systems the SOMO−LUMO gap)[e]

[a]Note that all features except those inferring to adsorption-induced changes are calculated on the reference configurations, i.e., the NCNT structures before adsorption, NCNT + $(n-1)$H. A glossary of auxiliary variables used in defining each feature is presented in Table S1. [b]N, V, H = nitrogen, vacancy, hydrogen. [c]S = adsorption site. [d]NN = nearest-neighbor sites. [e]SOMO = singly occupied molecular orbital.

where E_i is the ground-state energy of system i obtained from the DFT calculation. We note that the free energy of adsorption ΔG can be easily estimated from eq 8 by adding a constant finite-temperature correction amounting to roughly 0.3 eV for hydrogen adsorption on NCNTs.[46] Here, we will nevertheless focus on the ground-state energy differences to avoid introducing any unnecessary approximations.

We note that a few of the optimized adsorption configurations may be similar based on symmetry arguments despite the fact that a large amount of adsorption sites far away from the defect moieties were eliminated by considering only sites in the positive half-space. However, dopant−defect configurations more complex than, e.g., simple graphitic sites effectively break the symmetry of the CNTs, resulting in an appreciable reduction in the number of equivalent sites. The inherent symmetries are further reduced upon the adsorption of the initial hydrogen, which inevitably perturbs the structures of the NCNTs subject to three further hydrogen adsorption events, each in turn also reducing the symmetry. Thus, while some configurations considered in our data set may indeed be similar, we note that these correspond to a minority.

3.2. Electronic Structure Calculations. Spin-polarized DFT calculations were carried out using the hybrid Gaussian and plane-wave (GPW) method,[47] as implemented in the CP2K/Quickstep software package.[48] The plane-wave expansion of the electron density was truncated at an optimized cutoff value of 600 Ry, while the orbital functions of the valence electrons were expanded in short-range, molecularly optimized, and polarized double-ζ quality Gaussian basis sets (MOLOPT-SR-DZVP).[49] The remaining ionic cores were represented by dual-space norm-conserving Goedecker−Teter−Hutter (GTH) pseudopotentials.[50−52]

The exchange−correlation energy was described using the Perdew−Burke−Ernzerhof (PBE)[53] GGA functional including dispersion corrections based on the DFT-D3 scheme of Grimme et al.[54] with rational Becke−Johnson damping.[55] A direct minimization of the electronic Kohn−Sham energy functional was conducted by the orbital transformation (OT) method[56] employing an energy convergence threshold of 2.7×10^{-5} eV. Geometries were optimized by relaxing the atomic positions using the BFGS algorithm until the force on any atom was less than 2.3×10^{-2} eV Å^{-1}. Net atomic charges and electronic spin moments were calculated self-consistently from the optimized electron density using the iterative Hirshfeld partitioning procedure.[57] This population analysis scheme enables a rational selection of the promolecule and has been shown to compare favorably with alternative population analyses based on the electrostatic potential or topological partitioning.[58−60]

To validate the performed GGA calculations, a subset of 256 adsorption configurations were selected from the full GGA data set of structures relaxed using PBE and reconverged at the hybrid HF/DFT level of theory. Specifically, for each of the 64 NCNT-coverage combinations, four adsorption configurations corresponding to the maximum, minimum, median, and midrange adsorption energy values were chosen. Notably, the midrange energy corresponds to the configuration exhibiting an adsorption energy closest to the average of the maximum and minimum energy values. Hybrid single-point calculations were subsequently performed on these PBE-optimized configurations using the truncated, long-range corrected PBE0[61] functional (PBE0-TC-LRC) introduced by Guidon et al.,[62] where the standard $1/r$ Hartree−Fock exchange (HFX) term is truncated using a cutoff radius of 8 Å. The auxiliary density matrix method (ADMM)[63] was further

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

applied to reduce the cost of computing the exact HFX energy. In this approach, a supplementary polarized pFIT3 basis set was used to project the original density matrix into an auxiliary one for which the HFX energy can be computed more efficiently, thus facilitating accelerated hybrid calculations despite the large system sizes.

3.3. Machine Learning. The employed machine learning workflow was implemented using the tools provided in the `scikit-learn`[64] Python package. For the purpose of understanding how different features of defective NCNTs contribute to their inherent hydrogen adsorption properties, random forests were trained on the calculated PBE(0) DFT data employing 25 different structural, chemical, and electronic features parsed from the CP2K output (Table 2). The target variable of the regression task was the adsorption energy ΔE. We note that all features except those inferring to adsorption-induced changes were calculated on the reference configurations, i.e., the NCNT structures prior to adsorption, NCNT + $(n - 1)$H, to emphasize intrinsic catalyst characteristics more interesting from an experimental point of view. The limited PBE0 data set was applied to train complementary RF models and tentatively assess the impact of exact Hartree–Fock exchange on the feature attributions. Adsorption energy outliers were removed from both data sets based on the local outlier factor (LOF) method. Instances where geometry optimization resulted in spontaneous formation of H_2 gas were also excluded, yielding the final GGA and hybrid data set sizes of 6477 and 254 samples, respectively.

10 × 5-fold nested cross-validation (CV) was applied to obtain an unbiased estimate of the random forest generalization performance. The benefit of the nested CV strategy is that the full data set can be exhaustively tested employing different subsets of the data in turn for model training and validation. The utilization of different train–test splits provides also a way to estimate the model stability. This workflow is depicted in Figure 1 and entails the following steps. After DFT calculations and data preprocessing, the acquired data set $\mathcal{A}$ is split in accordance with the outer CV loop strategy. As we employ 10-fold outer CV, the data set is split into 10 subsets, each of which is, in turn, held out as a test set $\mathcal{B}_i$, while the remaining data forms the outer loop training (trainval) set $\mathcal{A}\backslash\mathcal{B}_i$. For each outer fold, the trainval set is next split in accordance with the inner CV loop strategy. For 5-fold inner CV, the set $\mathcal{A}\backslash\mathcal{B}_i$ is split into five parts, each of which is, in turn, used as a validation set $\mathcal{C}_j$, while the rest forms the inner loop training set $(\mathcal{A}\backslash\mathcal{B}_i)\backslash\mathcal{C}_j$. Using the inner loop training set, the model is trained multiple times using different hyperparameter settings and the performance of a particular setting is evaluated against the validation set. For the hyperparameter tuning, we employ randomized parameter sampling to optimize the number of features considered for the random subspace method, the maximum depth of the decision trees, and the minimum number of samples required at a node to be split. The root-mean-squared error, RMSE = $\langle(\hat{y}_i - y_i)^2\rangle^{1/2}$, is used as the loss function, and each random forest is composed of 500 decision trees for an adequate compromise between computational cost and accuracy.

Having trained a random forest for each inner CV fold and hyperparameter setting, the RMSEs are averaged over all folds and the set of hyperparameters that minimizes the loss is chosen. Leaving the inner loop, a random forest is trained using the selected hyperparameters on the whole trainval set

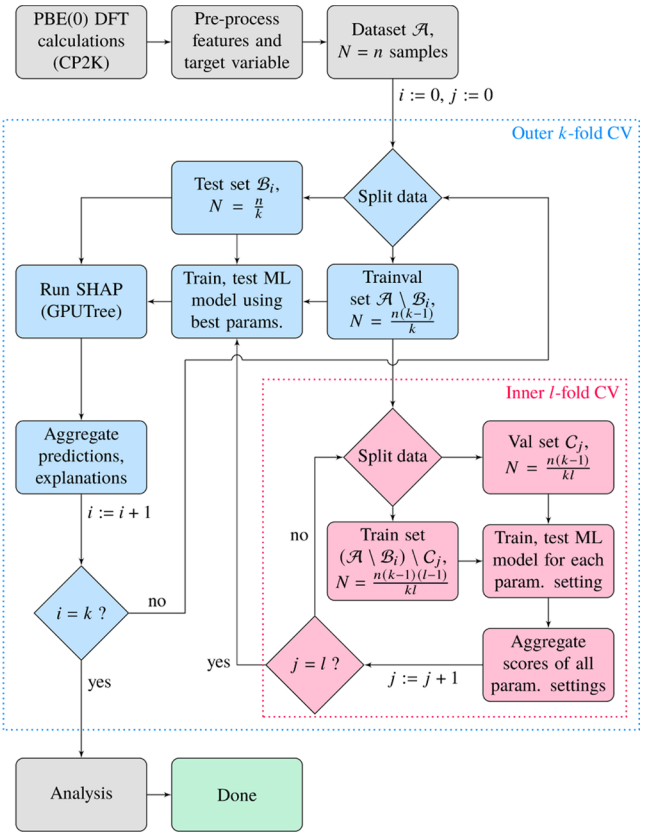


Figure 1. Employed machine learning and SHAP analysis workflow with n = 6477 and 254 samples for the PBE and PBE0 data sets, respectively, and k = 10 and l = 5 for 10-fold outer and 5-fold inner cross-validation.

$\mathcal{A}\backslash\mathcal{B}_i$ and evaluated on the test set $\mathcal{B}_i$ held out before entering the inner loop. The trained model and applied test data are subsequently explained using the GPU accelerated version of the TreeSHAP algorithm, and the obtained feature attributions including SHAP interaction values are aggregated together with the model predictions on the held-out samples. Repeating this nested cycle for each outer CV split allows us to evaluate the model performance against all samples in the collected data set, as well as to explain the model output for each of these instances. Importantly, the nested CV approach ensures that no data leakage occurs and trained models are always validated and tested on unseen data. Due to the heterogeneous nature of the data set, we note that all CV folds were stratified by the adsorption energies (rounded to nearest integer) for balanced and representative train–test splits containing roughly equal proportions of configurations with similar hydrogen affinities. A further validation of the specified RF size and nested CV strategy is presented in the Supporting Information, showcasing a robust generalization performance and invariance of the SHAP results against increasing the number of inner CV folds and decision trees in the RF.

4. RESULTS AND DISCUSSION

4.1. Predictions and Performance. All features detailed in Table 2 and adsorption energies were parsed from the PBE-level CP2K output and associated atomic coordinate files. As illustrated by the distribution in Figure 2a, the adsorption energies follow a skew normal distribution with a mean and standard deviation of 0.74 and 0.38 eV, respectively, suggesting

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX–XXX

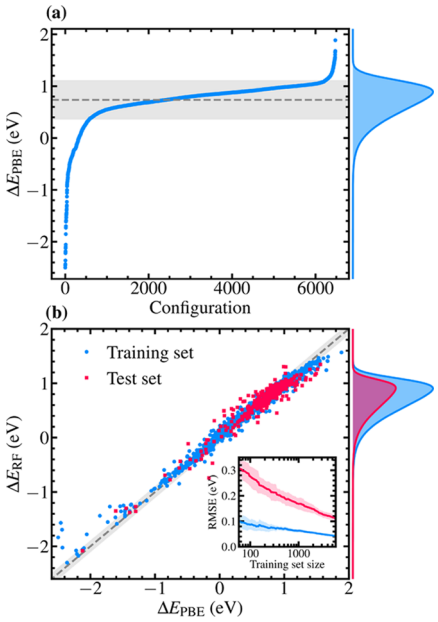


Figure 2. (a) Sorted adsorption energies within the GGA data set. The mean and standard deviation of 0.74 ± 0.38 eV are marked by the dashed line and shaded area, respectively. (b) Parity plot of the adsorption energies predicted by the RF model versus the true PBE values of the training and test set samples of an example outer CV split. Ideal correlation is marked by the dashed line, and the shaded area corresponds to ± the test set RMSE averaged over all CV splits. The skew normal distributions on the right side of the panels illustrate the distribution of samples within the full data and stratified example training and test sets. The test set distribution (red) has been arbitrarily scaled for visual purposes. The inset in (b) shows cross-validated learning curves for the training and test sets using a single hyperparameter setting. The shaded areas denotes the standard deviation of the respective RMSE losses.

that most adsorption sites on the studied NCNTs have low hydrogen affinities. The maximum adsorption energy included in the data set is 1.89 eV, while the minimum value is −2.50 eV. Figure 2a displays clearly the nonuniform nature of the data demanding the previously discussed need for stratified cross-validation. The predictions made using the optimized RF models were aggregated over the course of the nested CV procedure. As an illustration of the model performance, a parity plot displaying the RF predictions as a function of the true PBE adsorption energies is shown in Figure 2b for one example outer CV fold. In addition to the unseen test data, the model is intentionally tested also on the training data to contrast the model accuracy on the two data subsets. We stress that the performance on the test data is nonetheless what determines the true generalization performance of the models.

The inset of Figure 2b shows learning curves for the training and test sets where the RMSE loss is depicted as a function of the training set size. This data highlights the continuous decrease in the training and test set errors as the number of training samples is incremented. Despite the marked difference between the training and test set RMSE losses, it is clear from the curves that the test set RMSE keeps decreasing as the number of training samples is increased. We expect that a further increase of the data set beyond the current size would improve the generalization performance and reduce the discrepancy between the training and test sets, although the number of GGA samples required would be rather large due to the evidenced inverse power law behavior. Importantly, however, overfitting cannot be observed based on the learning curves as the model keeps improving (learning) with increasing data size.

The correlation between the PBE adsorption energies and the values predicted by the RFs on the test set samples is high, exhibiting average MAE and RMSE values of 0.06 ± 0.01 and 0.11 ± 0.01 eV, respectively, and a coefficient of determination (r^2) of 0.92 ± 0.01. Expectedly, the predictions on the training set are very precise, reaching a chemical accuracy (<1 kcal mol^{-1} ∼ 43 meV) based on both the MAE and RMSE losses with an r^2-score of 0.99. Reassuringly, although the majority of samples reside roughly within the 0.5 ± 1.0 eV range, the RF appears also capable of predicting adsorption energies of less frequent configurations displaying markedly stronger binding with a reasonable accuracy. This is of paramount importance for obtaining reliable feature attributions, as it shows that at least some of the input features are indeed important for the hydrogen affinity and that the model is capable of learning this dependence sufficiently well across the whole data set.

The performance metrics for the training and test sets averaged over all outer CV folds are summarized in Figure 3.

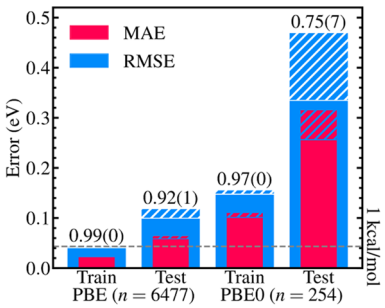


Figure 3. Unbiased generalization performance of the RF models based on 10 × 5-fold nested CV on the GGA and hybrid HF/DFT data sets. The solid bars denote a lower bound of the respective averaged errors, while the hatched parts indicate the variability as twice the standard deviation across the outer CV folds. The average coefficients of determination with standard deviations are annotated above the bars. The limit of chemical accuracy is marked for reference by the dashed line.

We include here also the results on the hybrid data set, the results of which are discussed in full detail in the Supporting Information. Briefly, the model generalization performance on the PBE0 data is expectedly worse (MAE 0.29 ± 0.03 eV, RMSE 0.40 ± 0.07 eV) compared to using the GGA data set due to the 96 % smaller data size. The hybrid data and trained models are, however, useful for a tentative assessment of the impact of the exact Hartree−Fock exchange on the attributed feature importances.

4.2. SHAP Analysis of Random Forest Outputs. *4.2.1. Global and Local Explanations.* In the following, SHAP values are computed to unravel the most important features of defective NCNTs that contribute to their intrinsic hydrogen adsorption properties. By virtue of the employed nested cross-validation workflow, the SHAP analysis is performed on all outer CV test set splits using the RF models tuned in the respective inner CV loops. As the outer loop test sets are completely separate from the outer CV training data, which is also used in the inner CV hyperparameter tuning, we ensure that the feature attributions are calculated for samples

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

of data unseen by a particular model. Thus, as discussed above, we are able to utilize the full data set for the analysis while completely avoiding bias associated with the training instances to which the models could theoretically overfit. This approach would not be possible using conventional impurity-based feature attributions that are computed on the training samples during tree construction.

As a summary of the feature attributions, the mean magnitude of the SHAP values of each feature is computed. Sorting the results, the 10 most impactful features are displayed in Figure 4a, highlighting the global feature

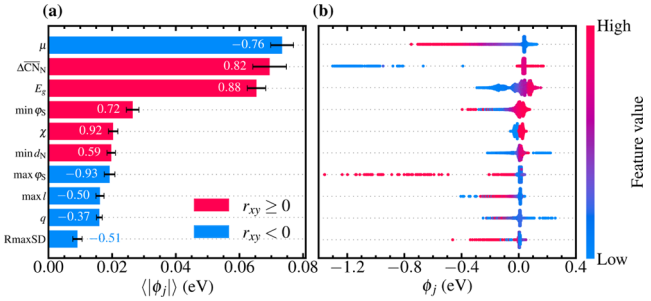


Figure 4. (a) Global SHAP importances ranking the 10 most impactful features in forming adsorption energy predictions. Bars correspond to cross-validated averages, and the error bars indicate ± 1 standard deviation across the outer CV folds. Correlation coefficients of SHAP and feature values are annotated on each accordingly color-coded bar. (b) SHAP values (local explanations) of the 10 globally most important features for all test set observations. The vertical dispersion of data points depicts high densities of SHAP values, i.e., abundance of configurations with similar ϕ_j. All data points are color-coded according to the feature values they represent.

importances. Clearly, three features emerge as particularly important when considering the whole data set: the spin polarization (local magnetic moment) on the adsorption site, the adsorption-induced change in the generalized coordination number of the nearest nitrogen dopant, and the highest occupied molecular orbital−lowest unoccupied molecular orbital (HOMO−LUMO) gap of the system. Inspecting the correlation between the SHAP values and the feature values, an overall increase of the adsorption energy is indicated as a function of increasing $\Delta\overline{\mathrm{CN}}_N$ and E_g, while a negative correlation is seen for μ. Other globally important features include bond angles and lengths between the adsorption site and its nearest neighbors, containing information on ring strain. Furthermore, the distance of the adsorption site to the nearest N-dopant, adsorption site charge, and NCNT chirality have considerable impacts toward the RF model output.

Although the feature attributions based on the mean magnitudes of the SHAP values provide a simple and concise picture of the global feature importances, local information contained within the SHAP values is lost in the process of evaluating the modulus and averaging following eq 4. It is therefore informative to leverage the property of SHAP values as *local* feature attributions, revealing the impact of features for individual data samples and predictions. We re-emphasize that local accuracy (additivity) is a desirable property of feature attribution methods. This property is not reflected by default impurity-based importances, limiting the accessible level of accuracy to the global scale presented in Figure 4a. Consequently, we illustrate in Figure 4b all SHAP values of the 10 globally most impactful features using violin-like scatter plots, where the vertical dispersion of the values encodes the abundance of similar attributed importances for a given feature. This visualization reveals that the high global impact of $\Delta\overline{\mathrm{CN}}_N$ is largely caused by a cluster of configurations exhibiting low feature values, responsible for negative marginal contributions as low as −1.3 eV in some model outputs. These samples are specifically characterized by negative $\Delta\overline{\mathrm{CN}}_N$ values around −1, suggesting adsorption-induced bond breaking at nitrogen dopant sites as the source of substantially negative adsorption energies. This is a manifestation of possible instabilities in the CNT structure due to substitutional nitrogen doping, resulting in highly reactive structures. We note that this tendency is particularly characteristic of N_1V_1-pyrrolic nitrogen moieties, accounting for nearly all 200 configurations in which bond breaking at the nitrogen site is observed. Such instabilities are likely to be deleterious for efficient HER catalysis. Bond breakage modulates also the potential energy surface close to the nitrogen dopant, opening up new adsorption sites to which preadsorbed hydrogen atoms may preferentially bind. Subsequently, hydrogen affinities calculated for these sites also contain contributions from bond rearrangement and hydrogen site hopping, thus complicating the analysis of adsorption energetics at pyrrolic nitrogen moieties.

In contrast, the SHAP values of the spin polarization on the adsorption site are rather continuously distributed, showcasing

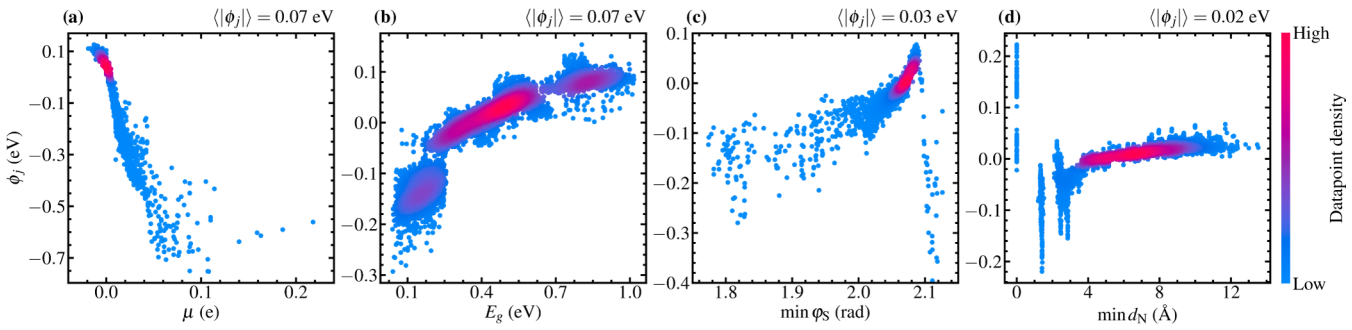


Figure 5. SHAP values of the (a) spin polarization on the adsorption site, (b) system energy gap, (c) minimum angle at the adsorption site, and (d) distance between the adsorption site and the nearest nitrogen dopant. The density of SHAP values is indicated by the color scale, and the global importances of the respective features are annotated above the panels for reference. Each data point in (b) has been shifted (jittered) along the E_g-axis by a small random amount (max. ±0.05 eV) to better highlight the overall trend between the SHAP and feature values. We stress that no more than 64 different values of the band gap were explicitly considered, and the jittering of the data is only a visual change that importantly does not affect the qualitative information conveyed by the plot. For the clustered raw data, please see Figure S4.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

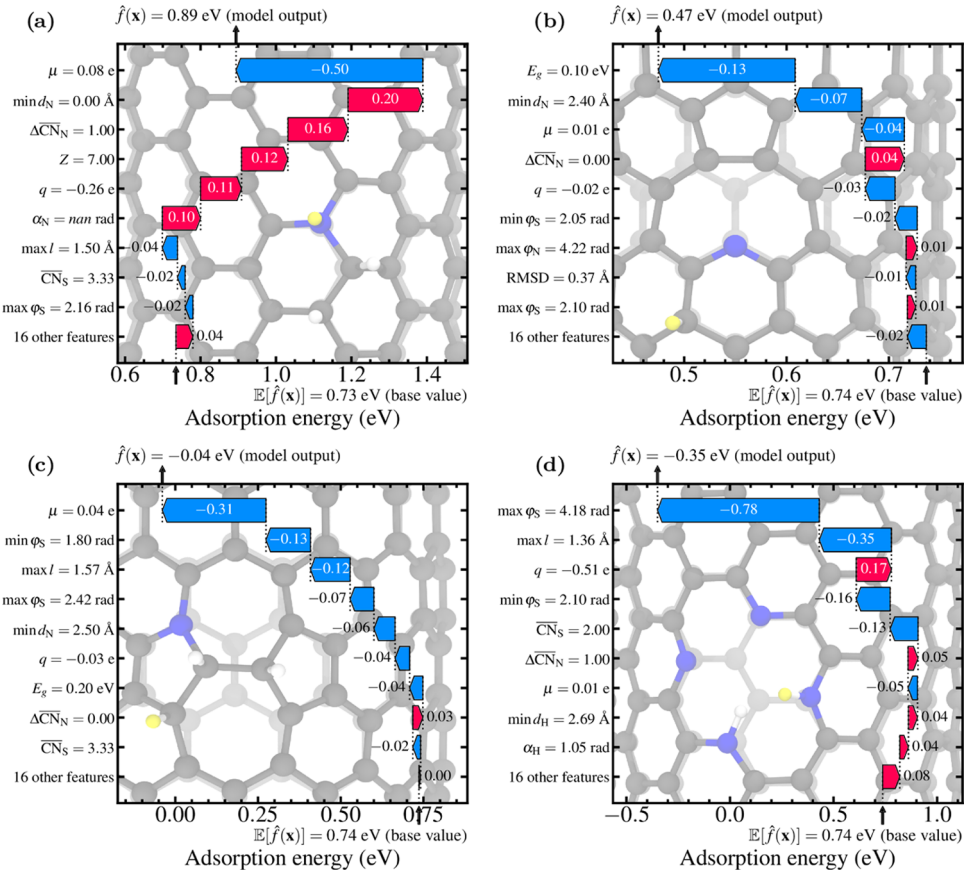


Figure 6. Case examples of model output formation as described by the SHAP values. Local explanations for adsorption at (a) (8,8) graphitic, (b) (14,0) N_1V_1-pyridinic, (c) (14,0) N_{1b}SW-pyrrolic, and (d) (8,8) N_4V_2-pyridinic dopant configurations are illustrated. The adsorbing hydrogen is marked in yellow, while white, gray, and blue spheres correspond to preadsorbed hydrogen, carbon, and nitrogen atoms, respectively. The base value $\mathbb{E}[\hat{f}(\mathbf{x})]$ of the model is marked below the panels, and the final output is obtained by summing the marginal contributions, yielding the model prediction $\hat{f}(\mathbf{x})$) annotated above each panel.

a clear negative correlation with the feature values of μ. The same applies for the energy gap E_g, although with the opposite correlation. Furthermore, the chiral angle χ of the NCNT has a quite moderate impact on the adsorption energy predictions, but there is a clear distinction between the effects of zigzag and armchair nanotubes, with the former having an overall negative impact on the output and the latter a positive. Intriguingly, the largest angle at the adsorption site exhibits a similar discontinuous behavior as $\Delta\overline{CN}_N$, with a separate group of configurations with high feature values having a decisively increasing effect on the predicted adsorption strengths. The high feature values turn out to be associated with reactive 2-fold coordinated nitrogen sites characterized by maximum angles around $4\pi/3$ rad (240°), as well as to some extent seven-membered rings and stretched six-ring structures.

Although summarizing individual SHAP values in the spirit of Figure 4b facilitates a deeper analysis of the local feature effects, the precise relation between the attributed importances and the feature values cannot be elucidated. This is particularly important for more complex dependencies between SHAP and feature values, such as those seen for the minimum angle at the adsorption site min φ_S as well as the distance between the adsorption site and the nearest nitrogen dopant min d_N. In the former case, high feature values may correspond to both near-zero and considerably negative SHAP values, while in the latter, small values of min d_N may be attributed with both positive and negative importances. As a direct analysis of these effects, the dependence plots of the SHAP values as a function of the respective feature values are presented in Figure 5. Here, we highlight four important features, μ and E_g with a monotonic dependence as a function of the feature values, as well as min φ_S and min d_N, the attributed importances of which show a more complex trend. The dependence plots of the remaining features are shown and discussed in Figure S4.

As inferred above, high values of the spin polarization at the adsorption site result in decreased adsorption energies, while large energy gaps have a low or slightly positive impact on the model output. A similar increasing dependence is observed for the minimum angle at the adsorption site in Figure 5c for angles smaller than $2\pi/3$ rad (120°). This indicates that adsorption sites at ideal six-membered ring structures have a small increasing effect on the adsorption energy predictions, while decreasing the angle toward $3\pi/5$ rad (108°) results in stronger adsorption. This value corresponds to the angle of a five-membered ring, which, opposed to the six-membered case, experiences pronounced reactivity-increasing internal strain. Interestingly, for min φ_S values larger than 120°, an abrupt decrease in the SHAP values is evidenced, suggesting that stretched six-ring structures display high hydrogen affinities. Such moieties are mostly associated with 2-fold coordinated adsorption sites observed at the edges of vacancy structures connected to pyridinic and pyrrolic dopant configurations.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

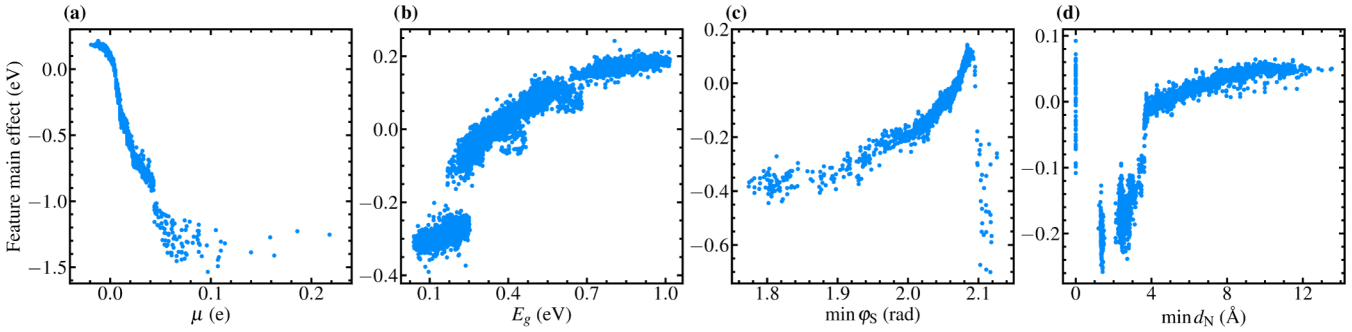


Figure 7. Main feature effects on the model output as captured by the diagonal elements of the SHAP interaction matrices of the (a) spin polarization on the adsorption site, (b) system energy gap, (c) minimum angle at the adsorption site, and (d) distance between the adsorption site and the nearest nitrogen dopant. The data points in (b) have been jittered slightly along the E_g-axis for visual purposes (see Figure 5).

Figure 5d unravels finally the intriguing effect of the relative position of the adsorption site with respect to the nearest nitrogen dopant. Indeed, the erratic trend illustrated in Figure 4b of low feature values having both negative and positive impacts on the model outputs is explained by the totally different hydrogen affinities of the nitrogen dopant sites and their (next) nearest neighbors. Here, the former have a small or decisively positive impact on the adsorption energy predictions, while the latter sites tend to be considerably activated by the presence of nitrogen. Moving farther away from the dopant is observed to result in near-zero attributed importances.

4.2.2. Decomposition of Individual Adsorption Energy Predictions. It is informative to inspect a few case examples of individual adsorption configurations to highlight the explanatory power of the locally accurate SHAP feature attributions. Figure 6 displays four distinct adsorption configurations and the largest marginal contributions toward forming each model output. As a first example, Figure 6a highlights two competing effects observed on the graphitic substitutionally doped NCNT, namely, adsorption at a 3-fold coordinated nitrogen dopant site and strong spin polarization. The adsorption to the nitrogen site is directly captured by four features, namely, zero adsorption site to dopant separation, an atomic number of 7 of the adsorption site, a unit increase in the generalized coordination number of the nearest nitrogen, and undefined angular displacement of the adsorption site with respect to the nearest dopant. Additionally, the nitrogen site is indirectly recognized by the residual charge that tends to be substantially negative for the more electronegative nitrogen. Recalling the additivity of SHAP values, these features sum up to a positive contribution of roughly 0.7 eV in the model output for the sample configuration. Importantly, this is a clear indication of the chemistry within our data—direct adsorption at a nitrogen site is generally unfavorable, corroborating the discussion associated with Figure 5d. However, the nitrogen site exhibits a pronounced spin polarization likely induced by the neighboring preadsorbed hydrogens that favorably perturb the local electronic structure of the adjacent atoms. Such clustering of sequentially adsorbing hydrogens has been previously demonstrated on pristine CNTs.[65] Consequently, the final output of the model collapses back toward the base value, totaling an adsorption energy of 0.89 eV.

Figure 6b exemplifies, on the other hand, the influence of the energy gap of the catalyst. In this example, no local features of the sample NCNT appear to influence the adsorption energy prediction substantially, except for the close, next-nearest-neighboring proximity of the adsorption site with respect to the nitrogen dopant. Thus, the low value of the band gap most significantly affects the hydrogen affinity, resulting in a model output of 0.47 eV. Conversely, the cumulative effect of multiple features illustrated in Figure 6c contributes toward a considerably more negative −0.04 eV adsorption energy prediction. Also here, hydrogen adsorbs to a nitrogen next-nearest-neighbor site, which in addition is located adjacent to a Stone−Wales defect. Consequently, the maximum and minimum angles correspond to those of a five- and seven-membered ring structure, respectively, and the maximum bond length at the adsorption site is markedly stretched, resulting in a net strain-induced shift of the model output by ca. −0.3 eV. In addition, the site exhibits a high local magnetic moment and a relatively small energy gap, which further decrease the adsorption energy prediction.

Some of the most important effects contributing toward high hydrogen affinities are attributed to the coordination of the adsorption site. Figure 6d breaks down this effect for the case of hydrogen adsorption to a pyridinic nitrogen dopant. In accordance with the preceding discussion, the direct adsorption to a nitrogen site is often weak, and also in this case, positive SHAP values of 0.05 and 0.17 eV are attributed to $\Delta\overline{CN}_N$ and the negative residual charge on the adsorption site associated with the electronegative nitrogen. However, the compressed bond lengths, as well as the large maximum and minimum angles at the adsorption site induced by the 2-fold coordination, contribute toward an increased adsorption strength of −0.35 eV, demonstrating the overriding effect of site coordination.

In the analyzed example cases, the predicted adsorption energies deviate from the true PBE values by 0.06 eV on average, underscoring the accuracy of the trained RF models. We note that the two samples discussed last exhibit adsorption energies close to the regime where the thermodynamic prerequisite for efficient HER catalysis is approximately met. Most importantly, the above analysis unravels why these example configurations show optimal hydrogen affinities, providing a path to actually understand the output of a complex ML model as well as the chemistry underlying the studied data. For a compact visualization, we note that multiple prediction decompositions as those displayed in Figure 6 can be combined to form a heatmap of SHAP values presented in Figure S5. Hierarchical clustering of the samples based on the attributed importances provides a way of obtaining a quick overview of the data and features resulting in similar model outputs. For example, the most negative energies are predicted

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

for adsorption configurations possessing high spin polarizations and structural instabilities as encoded by strained bond lengths and angles as well as adsorption-induced bond breaking and large atomic displacements. Conversely, the most positive model outputs are seen for substrates with large energy gaps and for direct adsorption to nitrogen dopants.

4.2.3. Interaction Effects. The vertical scatter of SHAP values in Figure 5 indicates that different importances can be attributed to a specific feature value depending on the adsorption configuration. This is a manifestation of interaction effects between features easily missed by studying importances at a global level. Quantifying these interactions is informative, unraveling hidden correlations that facilitate an explanation of why a given feature may be attributed with different importances despite the value of the feature remaining unchanged. In the present case, a feature value previously associated with a high/low hydrogen affinity might have an opposite impact on the model output depending on the context formed by the other features. Such insights are exceptionally valuable from the viewpoint of catalyst design, suggesting that observed structure−activity trends may not always reflect simple one-to-one relations. To analyze feature interactions within our data, we consider first the feature main effects. To this end, we focus on the diagonal elements of the SHAP interaction matrices for all outer CV test set samples and the same features as illustrated in Figure 5 to contrast the results in the absence of interactions. This data is visualized in Figure 7.

Expectedly, removing the interaction effects yields a less disperse representation of the feature dependencies. For values up to 0.1 e, the attributed impact of the local spin polarization in Figure 7a follows a monotonically decreasing trend, leveling out as the feature values increase further. The SHAP main effect of the energy gap in Figure 7b increases, on the other hand, as a function of the gap width, as does the minimum angle at the adsorption site in Figure 7c for values less than or equal to the ideal angle of a six-membered ring. The exclusion of interaction effects clarifies the picture of how different feature values impact adsorption energy predictions. Although in the three aforementioned cases the same information could be rather well inferred also from Figure 5, Figure 7d exemplifies the importance of explicitly considering interaction effects. Indeed, examining min d_N-values between 1 and 3 Å (nearest- and next-nearest-neighbor sites) reveals a clear difference compared to Figure 5d. Specifically, nonlinear interaction effects appear to have a significantly decreasing impact on the hydrogen affinities of the next-nearest sites of the nitrogen dopant.

As an overall assessment of the interactions perturbing the main feature effects, a global analysis of the SHAP interaction values is performed. In analogy with eq 4 and Figure 4a, we consider all pairwise feature combinations and sum their absolute interaction values over all test set samples k in each outer CV fold. The obtained result is a symmetric $F \times F$ matrix with off-diagonal elements $\sum_k|\Phi_{ij}^{(k)}| = \sum_k|\Phi_{ji}^{(k)}|$. Setting the diagonal main effects to zero yields a global interaction matrix presented in Figure 8, where the 10 most strongly interacting feature pairs are highlighted in sorted order. Four pairwise interactions emerge as particularly dominant, namely, between E_g and μ, E_g and χ, min d_N and E_g, as well as min d_N and q. We will focus on these feature pairs in the following analysis and plot the corresponding SHAP interaction values as a function of the feature values of each pair in Figure 9. We reiterate that

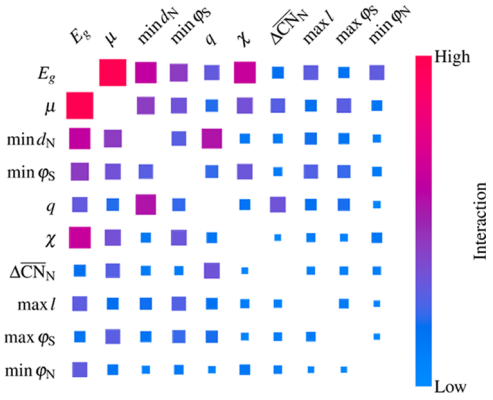


Figure 8. Global measure of SHAP interaction effects for the 10 most strongly interacting features. Each element of the matrix corresponds to the sum of absolute interaction values of a feature pair over all test set samples in each outer cross-validation split. The color scale indicates the magnitude of the global interaction, which is also proportional to the size of each square.

summing all pairwise interaction values of a given feature in all possible pairs and including the main effect yields the SHAP value of that feature in accordance with eq 7.

Figure 9a illustrates the intriguing impact of min d_N on the model output depending on the value of E_g. Notably, for adsorption to nitrogen sites (min d_N = 0), small values of the energy gap have a further increasing effect on the adsorption energies, while large E_g have a small negative contribution. This trend is reversed for nearest, next-nearest, and third-nearest sites, where small energy gaps may have a considerably decreasing effect on the predicted adsorption energies. However, for more distant adsorption sites, the behavior resets, i.e., small energy gaps have a slight positive contribution to the adsorption energy. Overall, the variations in the feature interaction values are larger for NCNTs with small E_g, suggesting a more dispersed dopant-induced modulation of the electronic structure promoted by the increased electron mobility. This increases especially the activity of sites closer to the dopant.

The feature interaction values of min d_N and q show a clear separation between adsorption sites with positive and negative residual charges, with the former having a slight decreasing contribution to the SHAP values of min d_N and the latter a more positive impact. Clearly, excess negative charge on the adsorption sites tends to decrease the hydrogen affinity, while electron deficiency (p-type doping) increases the adsorption strength. Due to the electron-withdrawing property of nitrogen, dopant sites are exclusively negatively charged, while adjacent carbons are positive. Next-nearest-neighboring sites, on the other hand, show a less homogeneous charge distribution, bearing both positive and negative residual charges depending on the specific NCNT configuration and its perturbing effect on the π-conjugated electronic structure. Importantly, we emphasize that this interaction between min d_N and q at the next-nearest and to some extent third-nearest carbon sites is what drives the attributed importances of min d_N toward zero in Figure 5d in contrast to the main feature effect. Therefore, the close proximity of an adsorption site to the nitrogen dopant may not always guarantee a high hydrogen affinity but requires in addition a favorable local charge distribution. This can be achieved by engineering catalysts with

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

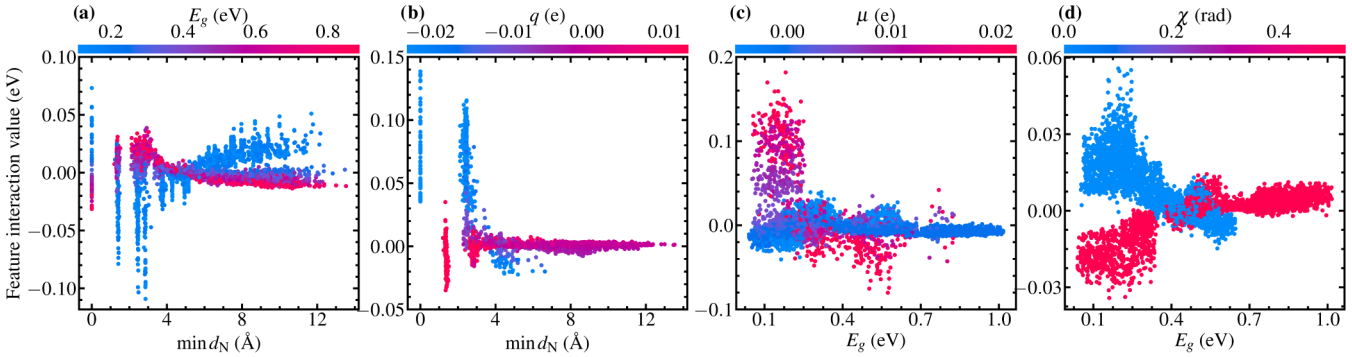


Figure 9. Pairwise SHAP interaction effects between (a) the dopant−adsorption site separation and the NCNT energy gap, (b) the dopant−adsorption site separation and the residual charge on the adsorption site, (c) the energy gap and the spin polarization on the adsorption site, and (d) the energy gap and the NCNT chirality. The data points are colored based on the values of the feature with respect to which the interaction effect is measured. The data points in (c) and (d) have been jittered slightly along the E_g-axis for visual purposes (see Figure 5).

high concentrations of p-type doping configurations such as observed for pyridinic moieties.[22]

Considering the energy gap and the local spin polarization, the observed interaction effect is negligible for small or slightly negative values of μ. However, as the spin polarization increases, the interaction effect becomes substantial, especially at small values of the energy gap. Here, high local spin moments lower the predicted adsorption strengths, thus working against the main effect of E_g illustrated in Figure 7b. For intermediate values of E_g, the interaction effect is intriguingly reversed, suggesting on average slightly more negative predicted adsorption energies for sites with pronounced spin polarization. The increased electron mobility promoted by a small energy gap allows the excess spin introduced by N-doping to favorably delocalize in the system. Consequently, the spin effect becomes more distributed and is effectively damped, while for larger energy gaps, delocalization is spatially constrained, yielding sites lesser in number but with larger spin polarization and thus increased hydrogen affinities.

A somewhat similar behavior is also evidenced in the case of the pairwise interaction between E_g and the NCNT chirality. Given that only the armchair ($\chi = \pi/6$ rad) and zigzag ($\chi = 0$ rad) variants of the NCNTs were considered, the feature and interaction values show a binary separation with opposed trends as a function E_g. For more conductive NCNTs, the zigzag structure induces a positive shift in the SHAP values of E_g, although the effect decreases toward zero as the energy gap widens. Conversely, for low values of the gap, the interaction effect of armchair NCNTs is seen to have a negative contribution toward the effect of E_g on the model output, with the interaction again diminishing as E_g increases. These results together with the feature importances of χ illustrated in Figure 4b point toward a small but distinct and complex chirality effect on adsorption energies, as discussed by Gíslason and Skúlason.[66] A more detailed analysis is, however, complicated by the fact that only two different NCNT structures were considered herein. The required more comprehensive sampling of different chiralities including NCNTs with varying diameters is consequently left as a topic for future research efforts.

4.3. Discussion. The high importance of the adsorption site spin polarization coincides with previous notions of catalytically activating electron transfer and spin modulation effects in heteroatom-doped carbon nanomaterials.[23,26,67−70] While such effects have been mainly investigated and identified for the ORR, our work reveals that similar electronic effects are also likely to influence the HER as inferred using the adsorption energy as an activity descriptor. Most importantly, in contrast to the qualitative reasoning presented in previous research, the quantitative feature attributions shown here unravel local feature−activity trends at an unprecedented level of detail. For example, regarding the spin polarization effect, we observe that an almost linear decrease in attributed SHAP values occurs as the magnetic moment on the adsorption site increases, but only up to roughly 0.1 e where the effect levels out (Figure 7a). Thus, engineering the spin properties of the catalyst may be desired only up to a certain degree after which negligible improvements in hydrogen affinities are obtained. Of course, the risk of overbinding remains relevant as well.

Moreover, the attributed impact of the HOMO−LUMO energy gap can be elaborated considering fundamental concepts of chemical reactivity. The energy gap and the chemical hardness η of a molecular structure can be straightforwardly associated by treating the HOMO−LUMO gap as the difference between the ionization potential and electron affinity in accordance with Koopmans' theorem and applying a finite-difference approximation,[71] yielding $E_g \sim 2\eta$. The principle of maximum hardness states that molecular structures tend toward an equilibrium state that maximizes η. Consequently, soft systems with low η undergo electronic structure-perturbing chemical transformations more readily, implying that molecular catalysts with small E_g reflect low kinetic stabilities. Conversely, the rearrangement of electrons upon a chemical reaction involving large-energy-gap materials is hindered due to the energetic unfavorability of electron injection to a high-lying LUMO and/or electron extraction from a low-lying HOMO. This principle has been demonstrated for the reactivity of polycyclic aromatic hydrocarbons and ORR catalysis on nitrogen-doped graphene.[67,72] Herein, we show that the attributed SHAP values specify this dependence for hydrogen adsorption on NCNTs to follow approximately an inverse power law, $\Delta E \sim -E_g^{-\gamma}$. As the band gap is inherently a global property of a specific system, we emphasize that this finding suggests more conductive (soft) NCNTs to exhibit *on average* higher hydrogen affinities. Therefore, all local sites are not necessarily activated toward hydrogen adsorption despite small values of E_g and special moieties, such as 3-fold coordinated nitrogen sites, may remain inactive due to their exceptional local properties.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

The importance of the energy gap is furthermore emphasized by its connections to the minimum dopant−adsorption site separation and the spin polarization. While high electron conductivity is desirable to minimize ohmic losses in the electrocatalysis of HER, small values of E_g are shown here to promote also the delocalization of spin density. Subsequently, an increased number of adsorption sites may attain pronounced magnetic moments, thus reflecting increased hydrogen affinities. However, for a given multiplicity, a more delocalized spin density implies decreased magnitudes of the spin polarization. Thus, decreasing the energy gap may contribute to a transition from few very reactive adsorption sites adjacent to a nitrogen dopant to multiple more dispersed ones showcasing intermediate hydrogen affinities. Overall, the main effect of the dopant−adsorption site separation on the adsorption energies follows the form of an interatomic interaction potential. This emphasizes that the localization of the optimal electronic structure-perturbing effect of the dopant is limited to 4 Å, i.e., nearest, next-nearest, and third-nearest carbon sites. Consequently, even though our analysis does not indicate a direct effect of dopant and vacancy concentrations on hydrogen affinities, HER rates are nevertheless promoted by increasing the number of defect sites that enter the kinetic equations through the pre-exponential factor.

In addition to electronic effects, the way in which hydrogen affinities are influenced by structural features including ring angles, bond lengths as well as site coordinations can be understood based on the presented results. While five-membered ring structures have a clear improving effect on adsorption strengths, the stability of dopant configurations remains an important consideration. Particularly, N_1V_1-pyrrolic configurations are found to easily undergo adsorption-induced bond breakage at the nitrogen moiety. Bond rearrangement-resistant five-membered ring structures associated with N-doped Stone−Wales defects and five-rings adjacent to studied N_1V_1-pyridinic dopant configurations are thus more likely to constitute active, yet sufficiently stable hydrogen-adsorbing sites. We note that the importance of five-membered ring structures for HER catalysis has been previously demonstrated theoretically in nondoped open-ended nanotubes where a truncation of the CNT promotes the formation of segregated and fused five-rings at the CNT edge exhibiting more than 0.3 eV lower HER barriers than six-ring sites.[73]

Although some of the features considered herein are perhaps not the most useful for the purpose of high-throughput materials screening as they require DFT calculations, we emphasize that this issue is of secondary importance considering the motivation of the present work. Indeed, we reiterate that the primary objective of this research is to gain understanding of fundamental physicochemical relationships underlying the investigated data using a novel feature attribution methodology. Thus, we highlight the "reverse" side of ML-driven materials science where the aim is to identify and concretize complex connections between input features and property predictions, not only to generate more or less unintelligible outputs without minding the inner workings of the deployed model. Nonetheless, future holistic studies in the spirit of explainable artificial intelligence will also benefit from the consideration of more easily accessible features to capture all aspects of state-of-the-art ML augmented computational materials science. This entails the development of fast and highly accurate, yet interpretable and transparent ML models relying on computationally inexpensive descriptors based on, for example, atomic structure and composition.

5. CONCLUSIONS

Robust and locally accurate feature attributions provided by Shapley additive explanations[17,29] present a tremendous contribution toward understanding black-box machine learning models as well as fundamental structure−property relationships within studied data. Such insights are of high value, for example, within rational catalyst development, where the number of activity-impacting structural and chemical degrees of freedom grows combinatorially as investigated materials become increasingly complex. Tackling large data sets using ML models is attractive, and analyzing the output using the SHAP methodology produces reliable feature importances that help in narrowing down which attributes of the investigated catalyst candidates are most important with respect to activity descriptors. Consequently, the catalyst search space can be rationally expanded and directed toward the most relevant systems and active sites, subject to more accurate calculations or experimental characterization.

In this work, we demonstrated the application of SHAP for the purpose of explaining the hydrogen adsorption properties of defective nitrogen-doped carbon nanotubes. Roughly 6500 adsorption configurations on 16 different NCNTs were analyzed using random forest ensemble learning, and the employed input features were attributed importances based on the SHAP (interaction) values. This quantitative analysis unraveled high spin polarization, narrow HOMO−LUMO gap, small dopant-site separation, and diverse angle and coordination effects as particularly impactful with respect to increasing hydrogen affinities of NCNT adsorption sites. The approach was also able to capture catalyst stability issues as reflected by bond breaking at pyrrolic nitrogen moieties, emphasizing the tradeoff between defect- and dopant-induced activation and excessive destabilization. Notably, our results corroborate previous analyses of activity modulation effects in heteroatom-doped carbon nanomaterials that have been mainly reported for ORR catalysis. The present work adds to this by showing that similar features also impact the nature of hydrogen adsorption and HER on NCNTs, demonstrating locally accurate feature−activity trends at an unprecedented level of detail.

We expect the SHAP feature attribution method to find widespread adoption within computational chemistry and materials science owing to its universal applicability. Making applied machine learning models more transparent and interpretable is highly desirable for answering questions such as why does a particular model work and how does it form predictions. Importantly, this information contributes to the development of physical and chemical knowledge of the system under investigation, ultimately paving way for a transition within ML augmented computational chemistry from proof-of-principle learning capability demonstrations toward using ML to actually explain and understand fundamental phenomena underlying complex systems and data sets.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.1c03858.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

Glossary of auxiliary variables used to define employed features; further validation of the applied ML workflow; comparison of electronic features produced by PBE and PBE0 calculations; RF performance and global feature importances for the hybrid data set; supplemental SHAP dependence plots and heatmap for all employed input features and GGA samples (PDF)

AUTHOR INFORMATION

Corresponding Author

Kari Laasonen − *Research Group of Computational Chemistry, Department of Chemistry and Materials Science, Aalto University, FI-00076 Aalto, Finland;* orcid.org/0000-0002-4419-7824; Email: kari.laasonen@aalto.fi

Authors

Rasmus Kronberg − *Research Group of Computational Chemistry, Department of Chemistry and Materials Science, Aalto University, FI-00076 Aalto, Finland;* orcid.org/0000-0002-6257-5956

Heikki Lappalainen − *Research Group of Computational Chemistry, Department of Chemistry and Materials Science, Aalto University, FI-00076 Aalto, Finland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.1c03858

Notes

The authors declare no competing financial interest.

The Python code and preprocessed GGA and hybrid data sets supporting this paper are published online at https://github.com/rkronberg/ncnt-random-forest/.

ACKNOWLEDGMENTS

R.K. acknowledges funding in the form of a doctoral scholarship by the School of Chemical Engineering of Aalto University. The authors thank Emil Kreutzman for helpful discussions on the machine learning aspects of the work. The authors wish to acknowledge CSC, IT Center for Science, Finland, for providing the required computational resources.

REFERENCES

(1) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, *355*, No. eaad4998.

(2) Magnussen, O. M.; Groß, A. Toward an atomic-scale understanding of electrochemical interface structure and dynamics. *J. Am. Chem. Soc.* **2019**, *141*, 4777−4790.

(3) Zheng, Y.; Jiao, Y.; Jaroniec, M.; Qiao, S. Z. Advancing the electrochemistry of the hydrogen-evolution reaction through combining experiment and theory. *Angew. Chem., Int. Ed.* **2015**, *54*, 52−65.

(4) Jørgensen, M.; Grönbeck, H. Scaling relations and kinetic Monte Carlo simulations to bridge the materials gap in heterogeneous catalysis. *ACS Catal.* **2017**, *7*, 5054−5061.

(5) Guo, D.; Shibuya, R.; Akiba, C.; Saji, S.; Kondo, T.; Nakamura, J. Active sites of nitrogen-doped carbon materials for oxygen reduction reaction clarified using model catalysts. *Science* **2016**, *351*, 361−365.

(6) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547−555.

(7) 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* **2017**, *5*, 24131−24138.

(8) Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Himanen, L.; Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Comput. Mater.* **2018**, *4*, No. 37.

(9) Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **2018**, *1*, 696−703.

(10) Jinnouchi, R.; Karsai, F.; Kresse, G. On-the-fly machine learning force field generation: Application to melting points. *Phys. Rev. B* **2019**, *100*, No. 014105.

(11) Deringer, V. L.; Caro, M. A.; Csányi, G. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. *Nat. Commun.* **2020**, *11*, No. 5461.

(12) Ko, T. W.; Finkler, J. A.; Goedecker, S.; Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **2021**, *12*, No. 398.

(13) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Phys. Rev. B* **2012**, *85*, No. 235149.

(14) Mitrofanov, A.; Korolev, V.; Andreadi, N.; Petrov, V.; Kalmykov, S. Simple Automatized Tool for Exchange-Correlation Functional Fitting. *J. Phys. Chem. A* **2020**, *124*, 2700−2707.

(15) Margraf, J. T.; Reuter, K. Pure non-local machine-learned density functional theory for electron correlation. *Nat. Commun.* **2021**, *12*, No. 344.

(16) Li, Z.; Wang, S.; Xin, H. Toward artificial intelligence in catalysis. *Nat. Catal.* **2018**, *1*, 641−642.

(17) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56−67.

(18) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573−584.

(19) Schmidt, J.; Marques, M. R.; Botti, S.; Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **2019**, *5*, No. 83.

(20) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **2017**, *3*, No. 54.

(21) Schleder, G. R.; Padilha, A. C.; Acosta, C. M.; Costa, M.; Fazzio, A. From DFT to machine learning: recent approaches to materials science−a review. *J. Phys. Mater.* **2019**, *2*, No. 032001.

(22) Davodi, F.; Tavakkoli, M.; Lahtinen, J.; Kallio, T. Straightforward synthesis of nitrogen-doped carbon nanotubes as highly active bifunctional electrocatalysts for full water splitting. *J. Catal.* **2017**, *353*, 19−27.

(23) Gong, K.; Du, F.; Xia, Z.; Durstock, M.; Dai, L. Nitrogen-doped carbon nanotube arrays with high electrocatalytic activity for oxygen reduction. *Science* **2009**, *323*, 760−764.

(24) Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **2006**, *5*, 909−913.

(25) Lindgren, P.; Kastlunger, G.; Peterson, A. A. A challenge to the $G \sim 0$ interpretation of hydrogen evolution. *ACS Catal.* **2020**, *10*, 121−128.

(26) Liu, X.; Dai, L. Carbon-based metal-free catalysts. *Nat. Rev. Mater.* **2016**, *1*, No. 16064.

(27) Tuomi, S.; Pakkanen, O. J.; Borghei, M.; Kronberg, R.; Sainio, J.; Kauppinen, E. I.; Nasibulin, A. G.; Laasonen, K.; Kallio, T.; et al. Experimental and computational investigation of hydrogen evolution reaction mechanism on nitrogen functionalized carbon nanotubes. *ChemCatChem* **2018**, *10*, 3872−3882.

(28) Li, S.; Yu, Z.; Yang, Y.; Liu, Y.; Zou, H.; Yang, H.; Jin, J.; Ma, J. Nitrogen-doped truncated carbon nanotubes inserted into nitrogen-doped graphene nanosheets with a sandwich structure: a highly

efficient metal-free catalyst for the HER. *J. Mater. Chem. A* **2017**, *5*, 6405−6410.
(29) Lundberg, S. M.; Lee, S.-I. In *A Unified Approach To Interpreting Model Predictions*, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017; pp 4765−4774.
(30) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(31) Mitchell, J. B. Machine learning methods in chemoinformatics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 468−481.
(32) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947−1958.
(33) Louppe, G.; Wehenkel, L.; Sutera, A.; Geurts, P. In *Understanding Variable Importances in Forests of Randomized Trees*, NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013; pp 431−439.
(34) Raileanu, L. E.; Stoffel, K. Theoretical comparison between the gini index and information gain criteria. *Ann. Math. Artif. Intell.* **2004**, *41*, 77−93.
(35) Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, No. 25.
(36) Shapley, L. S. A Value for n-Person Games. In *Contributions to the Theory of Games*; 1953; Vol. *2*, pp 307−317.
(37) Liang, J.; Zhu, X. Phillips-Inspired Machine Learning for Band Gap and Exciton Binding Energy Prediction. *J. Phys. Chem. Lett.* **2019**, *10*, 5640−5646.
(38) Morita, K.; Davies, D. W.; Butler, K. T.; Walsh, A. Modeling the dielectric constants of crystals using machine learning. *J. Chem. Phys.* **2020**, *153*, No. 024503.
(39) Goings, J. J.; Hammes-Schiffer, S. Nonequilibrium Dynamics of Proton-Coupled Electron Transfer in Proton Wires: Concerted but Asynchronous Mechanisms. *ACS Cent. Sci.* **2020**, *6*, 1594−1601.
(40) Fitzner, M.; Pedevilla, P.; Michaelides, A. Predicting heterogeneous ice nucleation with a data-driven approach. *Nat. Commun.* **2020**, *11*, No. 4777.
(41) Ayala, P.; Arenal, R.; Rümmeli, M.; Rubio, A.; Pichler, T. The doping of carbon nanotubes with nitrogen and their potential applications. *Carbon* **2010**, *48*, 575−586.
(42) Susi, T.; Kotakoski, J.; Arenal, R.; Kurasch, S.; Jiang, H.; Skakalova, V.; Stephan, O.; Krasheninnikov, A. V.; Kauppinen, E. I.; Kaiser, U.; Meyer, J. C. Atomistic description of electron beam damage in nitrogen-doped graphene and single-walled carbon nanotubes. *ACS Nano* **2012**, *6*, 8837−8846.
(43) Arenal, R.; March, K.; Ewels, C. P.; Rocquefelte, X.; Kociak, M.; Loiseau, A.; Stéphan, O. Atomic configuration of nitrogen-doped single-walled carbon nanotubes. *Nano Lett.* **2014**, *14*, 5509−5516.
(44) Susi, T.; Pichler, T.; Ayala, P. X-ray photoelectron spectroscopy of graphitic carbon nanomaterials doped with heteroatoms. *Beilstein J. Nanotechnol.* **2015**, *6*, 177−192.
(45) Murdachaew, G.; Laasonen, K. Oxygen evolution reaction on nitrogen-doped defective carbon nanotubes and graphene. *J. Phys. Chem. C* **2018**, *122*, 25882−25892.
(46) Holmberg, N.; Laasonen, K. Ab initio electrochemistry: Exploring the hydrogen evolution reaction on carbon nanotubes. *J. Phys. Chem. C* **2015**, *119*, 16166−16178.
(47) Lippert, G.; Hutter, J.; Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997**, *92*, 477−488.
(48) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An electronic structure and molecular dynamics software package-Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, No. 194103.
(49) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, No. 114105.
(50) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, No. 1703.
(51) Hartwigsen, C.; Gœdecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, No. 3641.
(52) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145−152.
(53) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, No. 3865.
(54) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, No. 154104.
(55) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(56) Weber, V.; VandeVondele, J.; Hutter, J.; Niklasson, A. M. Direct energy functional minimization under orthogonality constraints. *J. Chem. Phys.* **2008**, *128*, No. 084113.
(57) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126*, No. 144111.
(58) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007**, *444*, 205−208.
(59) Van Damme, S.; Bultinck, P.; Fias, S. Electrostatic potentials from self-consistent Hirshfeld atomic charges. *J. Chem. Theory Comput.* **2009**, *5*, 334−340.
(60) Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-theoretic approaches to atoms-in-molecules: Hirshfeld family of partitioning schemes. *J. Phys. Chem. A* **2018**, *122*, 4219−4245.
(61) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(62) Guidon, M.; Hutter, J.; VandeVondele, J. Robust periodic Hartree-Fock exchange for large-scale simulations using Gaussian basis sets. *J. Chem. Theory. Comput.* **2009**, *5*, 3010−3021.
(63) Guidon, M.; Hutter, J.; VandeVondele, J. Auxiliary density matrix methods for Hartree-Fock exchange calculations. *J. Chem. Theory. Comput.* **2010**, *6*, 2348−2364.
(64) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(65) Andreoni, W.; Curioni, A.; Kroes, J. M. H.; Pietrucci, F.; Gröning, O. Exohedral hydrogen chemisorption on a carbon nanotube: the clustering effect. *J. Phys. Chem. C* **2012**, *116*, 269−275.
(66) Gíslason, P. M.; Skúlason, E. Catalytic trends of nitrogen doped carbon nanotubes for oxygen reduction reaction. *Nanoscale* **2019**, *11*, 18683−18690.
(67) Zhang, L.; Xia, Z. Mechanisms of oxygen reduction reaction on nitrogen-doped graphene for fuel cells. *J. Phys. Chem. C* **2011**, *115*, 11170−11176.
(68) Wang, S.; Zhang, L.; Xia, Z.; Roy, A.; Chang, D. W.; Baek, J.-B.; Dai, L. BCN graphene as efficient metal-free electrocatalyst for the oxygen reduction reaction. *Angew. Chem., Int. Ed.* **2012**, *51*, 4209−4212.
(69) Jeon, I.-Y.; Zhang, S.; Zhang, L.; Choi, H.-J.; Seo, J.-M.; Xia, Z.; Dai, L.; Baek, J.-B. Edge-selectively sulfurized graphene nanoplatelets as efficient metal-free electrocatalysts for oxygen reduction reaction: the electron spin effect. *Adv. Mater.* **2013**, *25*, 6138−6145.
(70) Dai, L. Carbon-based catalysts for metal-free electrocatalysis. *Curr. Opin. Electrochem.* **2017**, *4*, 18−25.
(71) Pearson, R. G. The principle of maximum hardness. *Acc. Chem. Res.* **1993**, *26*, 250−255.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX

(72) Aihara, J.-I. Reduced HOMO- LUMO gap as an index of kinetic stability for polycyclic aromatic hydrocarbons. *J. Phys. Chem. A* **1999**, *103*, 7487−7495.
(73) Holmberg, N.; Laasonen, K. Theoretical insight into the Hydrogen evolution activity of open-ended carbon nanotubes. *J. Phys. Chem. Lett.* **2015**, *6*, 3956−3960.

https://doi.org/10.1021/acs.jpcc.1c03858
J. Phys. Chem. C XXXX, XXX, XXX−XXX