



Full length article

# Enabling automated high-throughput Density Functional Theory studies of amorphous material surface reactions

Martin Siron<sup>a,b</sup>, Nita Chandrasekhar<sup>b</sup>, Kristin A. Persson<sup>a,c,\*</sup>

<sup>a</sup> Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA

<sup>b</sup> Intel Corporation, Santa Clara, CA 95054, USA

<sup>c</sup> Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA



## ARTICLE INFO

### Keywords:

Density Functional Theory  
High-throughput  
Etching  
Site reduction  
Amorphous  
Surfaces

## ABSTRACT

Amorphous materials are increasingly considered for electronic and energy applications. However, modeling site-specific processes on amorphous materials is prohibitively expensive due to the large configuration space, and the large simulation cell needed to properly capture material properties. In this work, we develop a high-throughput workflow, powered by semiempirical methods in combination with Density Functional Theory, to simulate amorphous surfaces and utilize statistical sampling methods to assess surface reaction kinetics. We employ the Smooth Overlap of Atomic Positions (SOAP) to featurize surface species and apply a Bayesian Gaussian Mixture clustering model to perform a site reduction analysis. Finally, we automate the generation of input images for Improved-Tangent Nudged Elastic Band (IT-NEB) to simulate an etching reaction on the reduced sites. The resulting etching barriers are found to follow experimental etching trends for both amorphous silicon and amorphous carbon. Notably, our method reproduces etching barriers using a significantly reduced amount of sites (6 sites) out of all potential sites on the amorphous material surfaces. The successful site reduction method described in this study opens amorphous materials to high-throughput explorations of interfacial chemistry and surface properties.

## 1. Introduction

An increasing number of high-value energy applications, such as photovoltaics, batteries, transistors, phase change memory devices, etc., employ functional, amorphous materials either as active or supportive materials [1–5]. Notably, in the performance and/or processing of amorphous materials, interfacial chemistry, and reactivity provide an elusive yet key piece of information to understanding and evaluating material properties. However, while today's accelerated materials design paradigm increasingly leverages modeling and in-silico first-principles screening of material properties, it has been less utilized to evaluate materials' suitability for manufacturing and processing [6]. Furthermore, non-crystalline materials have been less explored through modeling, due to the dramatically increased structural complexity and simulation time needed to properly capture amorphous material properties. Established methods to model the amorphous materials' atomic structure include methods that compute interactions between atoms, such as Ab-initio Molecular Dynamics (AIMD) melt-quench simulations [1,7]. These methods have also been modified to use parametrized classical force fields, such as Tersoff potentials, which are material and application specific [8]. Other methods include reversing experimental

data, such as using Reverse Monte Carlo techniques on diffraction or structural data [9]. Hybrid methods combine both experimental and ab-initio methods [10,11]. An active area of research involves using relatively small, periodic structure samples to train machine learning (ML) models on force data for use in more complex molecular dynamics simulations [12,13]. Similar methods based on Gaussian process regression have also been developed by Bartok et al. [14]. These methods generally suffer the same fate: they are either material specific, costly, or both. Moreover, modeling of interfacial or surface reactions compounds the challenge due to the large chemical and structural diversity of surface sites. In principle, an a priori intractable number of local surface environments with potentially different reaction kinetics need consideration. Previous work in this field includes a notable contribution by Caro et al. which addresses this issue [15]. The work introduced a featurizer to transform and reduce a structure's bulk sites to a numerical vector, and then compared these vectors to those of pre-selected chemical motifs, employing previous chemical intuition specifically for amorphous carbon systems. Additionally, this work was able to map the adsorption reactivity of surface sites by reducing sites via clustering, motivating the need for more chemically

\* Corresponding author at: Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA.  
E-mail address: [kapersson@lbl.gov](mailto:kapersson@lbl.gov) (K.A. Persson).

<https://doi.org/10.1016/j.commsci.2023.112192>

Received 31 January 2023; Received in revised form 5 April 2023; Accepted 7 April 2023

Available online 23 April 2023

0927-0256/© 2023 Elsevier B.V. All rights reserved.

generalized methods. Other notable contributions can be found in the modeling of amorphous catalysts. Common commercial catalysts such as mesoporous silica have been difficult to model within the paradigm of density functional theory due to the synthesis-dependent structures, the lack of long-range order, and the lack of experimental fidelity given that a small number of sites can dominate the observed activity [16]. Previously, modeling these materials involved the creation of statistically significant cluster models and the sampling of various sites on these clusters. Indeed, reaction activation energies have been shown to correlate with different local chemical configurations, further motivating a method for choosing chemically different and important sites [17].

In this work, we present a material agnostic software framework, capable of addressing general amorphous surface reactivity, and demonstrate its usefulness on an example amorphous surface process: plasma etching. Plasma etching, of high importance to semiconductor manufacturing, is commonly used in patterning where a proper etch rate is critical to transistor processing and performance [18]. Dry etching in particular is ubiquitous in the semiconductor manufacturing process due to its flexibility in generating different etch profiles and tunable material selectivity [19]. Plasma environments consist of ions, electrons, and radicals of gas mixtures which promote a wide variety of physical and chemical reactions [20], which are difficult to characterize due to the dynamic evolution of the surface. For example, the plasma gas may exhibit multiple species including catalysts [21], the etchant may in some materials diffuse below the surface, attacking sublayers [22] and in other materials, the surface may undergo reconstruction and/or passivate or all the above simultaneously. Finally, the details of the materials in terms of surface phases, dopants, and grain boundaries can significantly affect the process as well [23]. As a result, the details of reaction mechanisms associated with plasma processes have only been elucidated for a few material/etchant combinations [24–27].

To advance first-principles modeling and screening into the realm of amorphous material processing and manufacturing, we developed a low-cost methodology to model 200–300 atom amorphous surfaces consisting of any combination of the first 80 elements of the periodic table. In our workflow, amorphous material surfaces are initially generated using a liquid-quench molecular dynamics simulation [28] within the approximations of Parametrized Model 6 (PM6) [29], then, followed by a relaxation using Perdew–Burke–Ernzerhof (PBE) [30] Density Functional Theory (DFT). To reduce the number of representative sites of study on the amorphous material surface, we performed site analysis featurization using the Smooth Overlap of Atomic Positions (SOAP) [31] combined with a Bayesian Gaussian Mixture clustering model.

To test the robustness and utility of the structure surface generation and site reduction technique on plasma etching, we adopted a simplified etching mechanism whereby an etch gas cleaves a bonded surface site to form a volatile etch product. An automated method for generating NEB inputs for systematic modeling of etching reaction kinetics is also provided. In addition, the model was generalized to other combinations of etch products and amorphous materials and was able to capture experimental trends with a systemic error across the combinations tested. In summary, the methodology presented in this study effectively reduces computational cost without losing overall material accuracy for calculated properties thereby allowing high-throughput exploration of amorphous material surfaces. Furthermore, we note that the developed workflows can be generally adopted and utilized for modeling many site-specific reactions such as adsorption, desorption, bond cleavage, etching, and deposition.

## 2. Methods

### 2.1. Data and software availability

The workflow presented here is developed as part of the atomate open-source software package [32]. The workflow makes use of the

FireWorks [33], pymatgen [34], sklearn [35] and matminer [36] software packages which are freely available. Density Functional Theory (DFT) calculations were performed with the CP2K simulation package [37]. The initial input structure for the MD simulation was generated using PACKMOL [38] with density inputs generated using MP-Morph [28]. A graphical description of the workflow is presented in Fig. 1.

### 2.2. Amorphous surface generation

All density functional theory electronic structure calculations were performed using the CP2K atomistic simulation package [37]. The complete workflow includes the generation of the surface amorphous structure from the bulk structure including relaxation and volume minimization. The computational parameters highlighted below are the default parameters employed by the workflow.

To generate the bulk amorphous materials we implemented a liquid quench process, similar to previously published methods [28]. Initial structures were generated with PACKMOL [38] and heated at 3000 K, with a 2 fs timestep for 700 steps, for a total of 1.4 ps. After 700 steps, a new molecular dynamics simulation using the resulting structure from the previous step was run at 2500 K. This was repeated until a temperature of 500 K was reached. Molecular dynamics simulations were performed in the NVT ensemble and using the Nosé thermostat to control for the temperature [39,40]. The total simulation time was 8.4 ps for each material. The Parametrized Method 6 (PM6) was used to solve for forces at every timestep [41] with an electronic convergence criterion of  $1 \times 10^{-5}$  Hartree. The Orbital Transformation method was utilized to speed up electronic convergence using the built-in Quickstep module [42]. For PM6, a cutoff of 20 Å was used for both the Coulomb and exchange parameters. We employed gamma point only Brillouin zone integration, and periodic boundary condition. An auxiliary basis set cutoff energy of 500 Hartree, and a reference grid cutoff of 80 Hartree was utilized.

Following the liquid-quench simulation, all subsequent Density Functional Theory calculations employed the Perdew–Burke–Ernzerhof [30] general gradient approximation functional. Calculations utilized the Orbital Transformations method built in the Quickstep module of CP2K [42]. In addition, all subsequent calculations employed Gaussian plane wave with the short-ranged double- $\zeta$  valence plus polarization molecularly optimized basis sets (DZVP-MOLOPT-SR) and norm-conserving Goedecker–Teter–Hutter (GTH) pseudopotentials assigned to all atom types [43,44]. Atomic coordinates were optimized using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

A structure relaxation calculation was performed, keeping the volume of the original cell constant. An auxiliary basis set planewave energy cutoff of 600 Hartree, with a reference grid cutoff of 80 Hartree was used. An overall convergence criterion was set such that a maximum displacement of  $3 \times 10^{-3}$  Å and maximum force of  $4.5 \times 10^{-3}$  H/Å was obtained. A self-consistent electronic convergence criterion of  $1 \times 10^{-7}$  Hartree was utilized. To arrive at the final bulk structure, the density which minimizes the overall energy of the system is identified by homogeneously straining the lattice from 80% its current dimension to 120%, sampling 25 total different densities. Static tests are performed at each new density. For these, an electronic convergence criterion of  $1 \times 10^{-7}$  Hartree was utilized. The structure corresponding to the lowest energy is then chosen. Next, six slab models are generated by cutting the bulk at the end of each facade of the cell and adding 20 Å of vacuum. The bottom 2 Å are kept frozen. The surface is allowed to relax, generating a set of amorphous material slabs. For these, the same convergence criterion from the post-PM6 DFT calculation is employed. Further calculation details including CP2K input files can be found in the Supplementary Information. The approach considered in this work aims to produce structures with high-throughput, however, we note that not all materials can be properly modeled by this method. We note

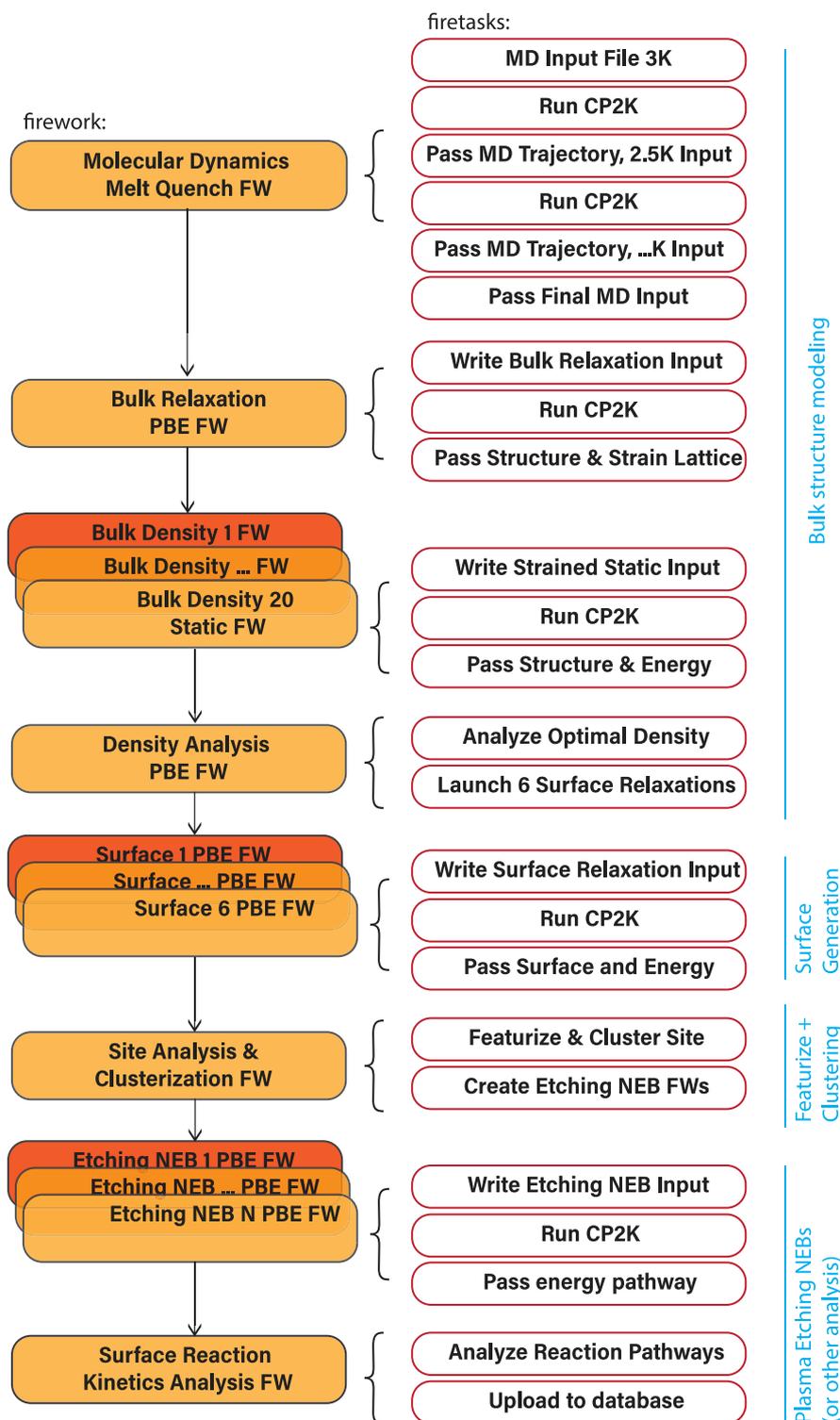


Fig. 1. General workflow structure. The workflow consists of individual computing jobs (known as Fireworks, in filled boxes), each of which is a sequence of computing tasks (known as Firetasks, in white boxes). The workflow can be divided in four sections: the modeling of bulk amorphous materials, the generation of amorphous material structure, the featurization and clustering of surface sites and finally the etching reaction modeling of these sites.

that the high quench rate of 300 K/ps may not be appropriate for more complex multi-element amorphous materials. Additionally, the representation of forces from semi-empirical PM6, which was developed to capture the formation energy of molecules, should be interpreted with caution. Further, surfaces may need additional molecular dynamics calculation to overcome local potential energy barriers, and the melt-quench protocol might need to be tuned depending on the material

system being modeled. While the workflow does not by default provide an annealing step nor allow the surface to relax using molecular dynamics simulations rather than DFT during the surface generation process, we emphasize that a user could model any material using any computational approaches desired for any material which may require more stringent modeling, and pass the resulting structure onto the site featurization and clustering portion of the workflow.

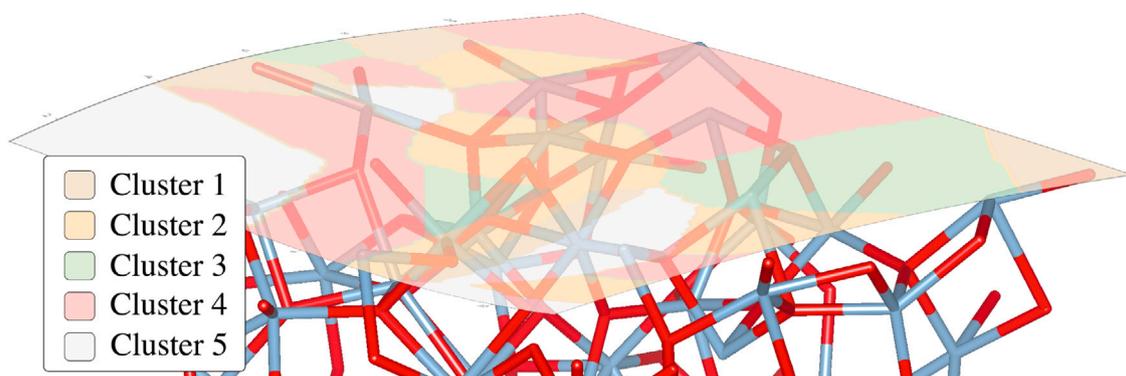


Fig. 2. Pictorial representation of clustering mechanism for surface sites of an amorphous  $\text{SiO}_2$  surface generated using the methods described in this work. The sheet of colors above represent similar chemical environments detected using the site identification and clustering mechanism described in this work.

To compare amorphous silicon etch rates against crystalline, we generated a crystalline silicon (110) surface, starting from the bulk crystalline Si structure in the Materials Project database [45]. The surface was generated using Pymatgen's slab generator feature [46], selecting a (110) miller index. The slab was relaxed using a kpoint mesh of  $2 \times 2 \times 1$ , and the same convergence criterion was utilized as the amorphous surface DFT calculations. No species were frozen. The slab was approximately 14 Å in height, 11 Å in width, and with 20 Å of vacuum, similar to the length specifications used for the a-Si surface.

### 2.3. Site identification & clustering

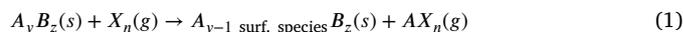
Amorphous surfaces exhibit a much broader range of surface sites as compared to crystalline surfaces, which can become computationally prohibitive to explore. In this work, we utilized two methods to reduce the number of potential sites, while still retaining a representative range of local site information. The workflow employs two surface site detection methods: the Shrake–Rupley [47] rolling sphere algorithm, which computes the solvent accessible surface area (SASA) of each atom on the top half of the slab, and by height, finding surface sites which are within 1.2 Å of the top-most site. For the SASA computing algorithm, the calculations were performed using the OpenSource FreeSASA package [48]. Sites with SASA values above  $15 \text{ \AA}^2$  were utilized. While the workflow has two methods of identifying surface sites, we employed the more simple, height-based site finding approach due to the unusual bulkiness of an etching reaction and because our surfaces were relatively smooth. The Smooth Overlap of Atomic Positions (SOAP) [36,49] was applied across all surface sites of all six surfaces. SOAP effectively encodes the atomic environment of a particular atom into a vector [31] and has previously been used to assess the 'similarity' of amorphous structure sites [15,50]. For a generic system, a cutoff radius of 3 Å was chosen, along with 3 radial contributions considered, a maximum angular contribution of 3, and Gaussians with a standard deviation of 0.6 Å. The featurization vectors along with the site coordinates and the slab to which the vectors belong were saved and passed down through the remaining steps of the workflow for future analysis. We note that the featurization can be customized to any algorithm available in the Matminer package.

To reduce the number of sites for further study, sites were clustered by similarity. By default, the workflow employs the Bayesian Gaussian Mixture clustering technique. The number of bins chosen for this method was set to three times the number of species present on the material, however, the binning can be easily tailored. After sites were clustered, sites closest to the center of each cluster by using the Mahalanobis distance [51] in SOAP space were chosen for further sampling. Each site, and its associated cluster across all potential surfaces, were passed down to the workflow for future statistical analysis. A pictorial representation of a surface whose sites have been featurized and then clustered using the methods described in this work is presented in

Fig. 2. In the supplementary information, we outline chemical motifs extracted from the featurization and clustering of an a-C and a- $\text{SiO}_2$  slab. For a-C, we found that generated clusters could be discerned by the sites' coordination environment, and for a- $\text{SiO}_2$  we found that clusters could be discerned by the concentration of Oxygen in the local environment. This demonstrates the ability of this method to capture chemical motifs.

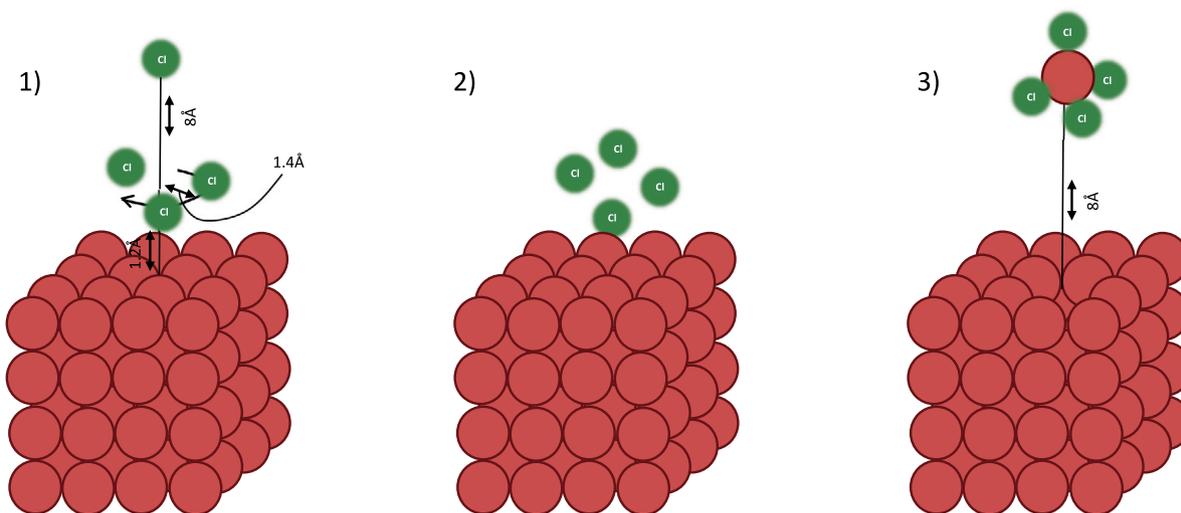
### 2.4. Modeling etching reactions

To test and validate the workflow and methodology, we screened etchants and compared etching barriers across material systems. Generally such calculations have excluded amorphous materials due to the plethora of surface chemical sites and coordination environment. To model the etching reaction, we assume that the rate determining step of dry etching is the dissociation of a surface species, A, from the material, using an etching agent  $X_n$  and its formation into a volatile etching product  $AX_n$ :



This simple etching mechanism where the formed product desorbs from the surface is expected to mirror the rate-limiting activation energy for the overall process, thereby providing a good metric for experimental etch rates in plasma conditions. We note that similar heuristic, model etch barriers have been employed previously, resulting in reasonable agreements with experimental etch rate trends [52–56]. However, we also emphasize that highly dynamic and/or unstable surfaces, that are uncommonly reactive with the plasma species, are unlikely to adhere to such model reactions, or may exhibit other reaction rate limiting steps, and for such cases, we recommend that efforts should be undertaken to model more realistic environments.

In the following, we demonstrate the methodology using a homogeneous composition volatile etch product (i.e.  $\text{SiF}_4$ ) but emphasize that the workflow can accept heterogeneous composition volatile etch products, i.e.  $\text{SiF}_3\text{Cl}$ , where F and Cl are etchant species present in the plasma. Nudged Elastic Band (NEB) calculations were initiated by automatically generating three input images for the calculation. For a given site (A), and a given etch product (i.e.  $A-X_4$ ), where A is an atom on the surface to be etched, and the plasma gas contains species X and forms etch product  $AX_4$ , one of the X atoms will be labeled as the "etchant atom". This atom will be appended directly above the 'A' site, 8 Å above the surface, far enough to prevent any potential interaction. The other three remaining atoms (from  $X_4$ ) for this etchant will be attached to A. To attach the remaining atoms systematically, a circle, 1.2 Å above the "A" site, and 1.4 Å in radius, is drawn. This circle marks possible sites to symmetrically append the remaining atoms which are needed to form a volatile etching product. At each point on the circle, homogeneously sampled by 100 points, the distance



**Fig. 3.** Schematic of input replicas for a single site of a slab etched by chlorine plasma, forming a  $\text{Cl}_4$  byproduct. In the first replica, 3 Cl atoms are positioned along a 1.4 Å radius circle, placed 1.2 Å above the site. A Cl is placed 8 Å above the site. In the second replica, the ‘attacking’ Cl atom is brought 2 Å away from the etched site. Finally, in the final input replica, the etched site and Cl atoms are moved 8 Å into the vacuum. The molecule is rearranged to match the input geometry while minimizing rotation.

to the closest surface atom is calculated. To minimize steric hindrance when appending the remaining atoms, the set of points representing the maximum average distance of any remaining atoms to be appended is chosen. For the second image, the etchant atom is brought 2 Å away from the ‘A’ site. Finally, in the final image, the ‘A’ site is lifted along with all the ‘X’ atoms. The lifted set of molecules is rearranged to match a relaxed etching product. The rearrangement is done using the Kabsch algorithm [57] built into Pymatgen to minimize artificial rotations. Fig. 3 demonstrates the pictorial representation of a set of input NEB images for a slab site etched forming an  $\text{Cl}_4$  etch product.

NEB calculations were performed using the Improved Tangent-Nudged Elastic Band method [58]. For these, the Brillouin zone was sampled by the Gamma point only and the PBE functional was used [30]. Orbital transformations as part of CP2K were allowed with a self-consistent electronic convergence criterion of  $1 \times 10^{-6}$  Hartree, allowing for an overall maximum displacement of  $1 \times 10^{-4}$  Å and maximum force of  $4.5 \times 10^{-4}$  H/Å for convergence. Calculations were run with 15 replicas, by allowing CP2K to linearly interpolate between input images. We note that no rotations or transformations were applied to the volatile etch product, instead the endpoints of the NEBs were allowed to relax. Most NEB calculations converged energetically after approximately 300 steps, however, a limit of 500 steps was imposed to control computational costs. Resulting curves for these dry etching simulations were analyzed using the Minimum Distance Nearest Neighbor class built in Pymatgen to detect the replica steps at which the etching occurred [59]. A coordination number analysis was used to extract etching barriers by detecting the commencement and end of the etching reaction in the NEB, discarding any of the post- or pre-etching replicas typically associated with minor surface rearrangements (more information about the etching detection and coordination number analysis can be found in the Supplementary Information).

### 3. Results & discussion

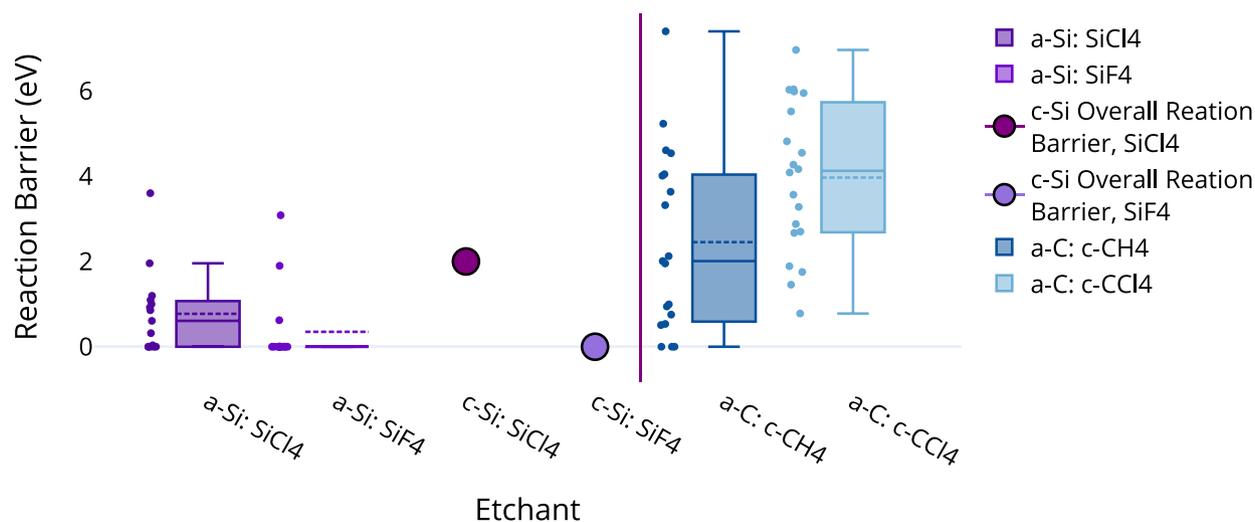
We benchmarked our methodology and workflow on trends in surface reactivity and etching, for two systems: amorphous carbon and silicon, motivated by their broad application areas. Amorphous carbon (a-C) is a commonly studied material in plasma etching processes, due to its increased selectivity over photoresist, its high transmittance, easy deposition, and removability [60]. Amorphous silicon (a-Si) is popularly used in the semiconductor industry, including for photovoltaics, and as a transistor material. The selectivity of crystalline silicon (c-Si) has been explored [61] using different etchants such as  $\text{CHF}_3$ ,  $\text{SF}_6$ ,

$\text{CF}_4$  commonly resulting in  $\text{SiF}_4$  formation. Amorphous Si has also been etched under chlorine-based plasma, and its etching rate is compared to both crystalline and poly-crystalline silicon [23]. In the case of amorphous carbon, we expect that etching with  $\text{H}_2$  plasma to form  $\text{CH}_4$ , would be faster than chlorine-based etching, as literature suggests hydrogen etching is often used [56]. In addition, previous experiments with chlorine plasma have found a-C to be rather etch-resistant [62]. Conversely, it has been found that chlorine-based etchants etch a-Si faster than c-Si, while a fluorine etchant reacts with silicon faster than a chlorine-based one [20,23]. In the following, we present and discuss results for the etching of a-Si and c-Si forming etch products  $\text{SiCl}_4$  and  $\text{SiF}_4$  as well as the etching of a-C forming etch products  $\text{CH}_4$  and  $\text{CCl}_4$ .

#### 3.1. Amorphous silicon and carbon structures

Initial amorphous structures were built using 200 atoms for amorphous silicon (a-Si), and amorphous carbon (a-C). While larger simulation cells could have been employed, we chose these default parameters to give reasonable results while not proving too expensive to compute. For a-C, initial input densities were generated using MPMorph [28], by averaging the densities for all crystalline materials in a given amorphous structure’s composition in the Materials Project database, leading to  $3.3 \text{ g cm}^{-3}$ . In the case of a-Si, an input density of  $2.2 \text{ g cm}^{-3}$  was chosen. Comparing the a-C structure against published experimental and molecular dynamics results, we found similar radial distribution function (RDF) peak locations of 1.5, 2.5, and 3.7 Å [63]. Our 2nd and 3rd RDF peaks were found at lower intensities compared to experimental results, showing a decrease in mid-range order, nonetheless the peak positions matched well. Using the Brunner Nearest Neighbor algorithm built into Pymatgen we detected that around 85%–90% of our carbon species were four-fold coordinated ( $\text{sp}^3 \text{ C}$ ) [64]. We also note a bond angle distribution of 75–150°, centered at 110°, similar to previously reported results [65]. For a-Si, we calculated an RDF with peaks at 2.3 and 3.8 Å and bond angle distribution ranging from 50 to 150° centered at 110°. Again, these peak distance and bond angle distribution results are similar to previous molecular dynamics simulations and experimental observations on a-Si [65]. Similarly, using the Brunner algorithm, we calculated that only under 4% of the species in our simulated a-Si structure were under-coordinated. The static density optimization eventually resulted in an a-Si structure with a density of  $2.2 \text{ g cm}^{-3}$ , similar to other experimental observations. To further benchmark the technique for amorphous bulk generation, amorphous  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$  were modeled. More details on the input set for structure generation and on the result of the benchmark of the other materials tested can be found in the Supplementary Information.

## Comparison of a-C and a-Si Etching by Different Etchants



**Fig. 4.** Comparison of etching reaction barrier of a-C, a-Si, and c-Si by different etchants. In this plot, the amorphous materials are presented by both a swarm and box plot. The swarm section represents all of the tested amorphous surface sites for which an etching reaction was modeled. This information is similar to what is portrayed in the box plot. The box plot's top and most bottom whiskers are the 2% and 98% interval. The box itself is the first and third quartile. The bold line is the median value and the dashed line is the mean. For all the analysis we compare the mean and the variance of these sets. For crystalline Silicon (c-Si), the average of 3 unique sites are taken and displayed as a scatter plot.

### 3.2. Etching reactions and comparison with experimental trends

Prior to any site reduction, etching reactions were modeled on all distinct 15 surface sites of a-Si forming two volatile etch products,  $\text{SiF}_4$ , and  $\text{SiCl}_4$ . All but one site converged (a  $\text{SiCl}_4$  site). Three unique surface sites were modeled in crystalline silicon. Etching barriers for both Si and C materials are shown in Fig. 4, including statistical error bars across representing the 2% and 98% percentile of the etch barriers, boxes representing the first and third quartile, the mean as a dashed line, and the median as a bold line. As expected, sites showed a distribution of etching barriers depending on their local environments. Arguably, mean barriers may not provide the most direct metric for etching rates, as, for example, sites with low barriers may etch first and the structure may rearrange or be left with more resistant etch sites. However, a structure/etch product combination with lower-barrier sites should generally etch faster than one with many high-barrier sites. Thus, such a metric is useful for comparison across material and etchant gas composition. While etching sites between the crystalline and amorphous silicon varied, the overall trend across all 15 sites of a-Si and three sites of c-Si was such that, fluorine plasma forming  $\text{SiF}_4$  exhibited a lower average barrier on both materials than chlorine plasma forming  $\text{SiCl}_4$ . Additionally, while c-Si showcased a higher etching barrier by chlorine than a-Si, the trend was reversed for fluorine. It was found that the reaction barrier for c-Si was lower by fluorine etching than chlorine. In the case of c-Si, the etching by fluorine species forming  $\text{SiF}_4$  product was found to be barrier-less, indeed the same barrier has been previously reported for this combination [20]. The etch barriers calculated for a-Si by  $\text{SiCl}_4$  ranged from 0 to 3.6 eV with a mean of 0.77 eV, while in  $\text{SiF}_4$  the barrier ranged from 0 to 3.1 eV with a mean of 0.35 eV. For the case of a-Si etched by  $\text{SiF}_4$ , only 3 sites had barriers greater than 0.5 eV, the other 11 sites were either barrier-less or had insignificant etching barriers. Meanwhile, 8 sites for a-Si etched by  $\text{SiCl}_4$  exhibited significant, above 0.5 eV barriers. In the case of crystalline Si, the etching barrier by  $\text{SiCl}_4$  was 1.89 eV while it was barrier-less by  $\text{SiF}_4$ .

The reactivity of two etchants: hydrogen and chlorine plasma were explored on amorphous carbon, creating  $\text{CH}_4$  and  $\text{CCl}_4$  etching products, respectively. Three  $\text{CH}_4$  forming sites, and two  $\text{CCl}_4$  forming sites

did not converge. Input replica images for all surface sites of a-C were set up similarly to the method described for a-Si. Both amorphous carbon products,  $\text{CH}_4$  and  $\text{CCl}_4$ , exhibited larger standard deviations than their counterparts in a-Si. Amorphous carbon etching barriers by  $\text{CH}_4$  were described by a standard deviation of 2.12 eV as compared to 1.78 eV for  $\text{CCl}_4$ . The product  $\text{CH}_4$  exhibited a mean etching barrier rate of 2.45 eV while  $\text{CCl}_4$  was noticeably higher at 3.96 eV. It is common practice in etching of a-C to first hydrogenate the material, and notably, the model surface employed here was not hydrogenated. Hydrogenation could lower the barrier for both etchants, however, our reported mean etching barrier of 2 eV matches with reported literature etching barriers of 1.6–2.5 eV [56]. We also observe that dense amorphous carbon is often used as a popular photoresist in plasma etching and thus we would expect a-C to exhibit high etching energy barriers, as validated by our simulation results in comparisons to a-Si [66]. The resulting etching barriers for a-C with the two etchants can be found in Fig. 4.

While the etching barriers here are for individual sites, a weighted average of the energy barriers provides a useful metric to quickly compare etching barriers of the same etchant, on a different material, forming similar neutral volatile etch products (for example  $\text{SiCl}_4$  vs.  $\text{CCl}_4$ ) or comparing two etchants etching the same material ( $\text{SiCl}_4$  vs.  $\text{SiF}_4$ ). While the etch rates derived from these barriers are not additive, the trends captured by the weighted average of these barriers are in agreement with experimental observations rendering these methods useful in determining selectivity trends based on chemical etching mechanisms and helpful in determining appropriate etching gases based on the materials to be etched. Additionally, we emphasize that while the modeled NEB involved multiple sequential reactions: (i) the adsorption of multiple initial plasma atoms, (ii) the adsorption of the attacking atom, and (iii) the removal of a surface site, typically the adsorption reactions were found to be exothermic and barrier-less while the desorption of the site atom exhibited a non-negligible transition energy. Hence, in the cases examined here, we found the desorption to be the rate limiting step of the etching process. Improvements to the approximations employed here include an advanced sampling scheme as well as statistical ensemble methods to ensure the entire

### Generalization of SOAP & GM model across materials & etchants, 6 point model

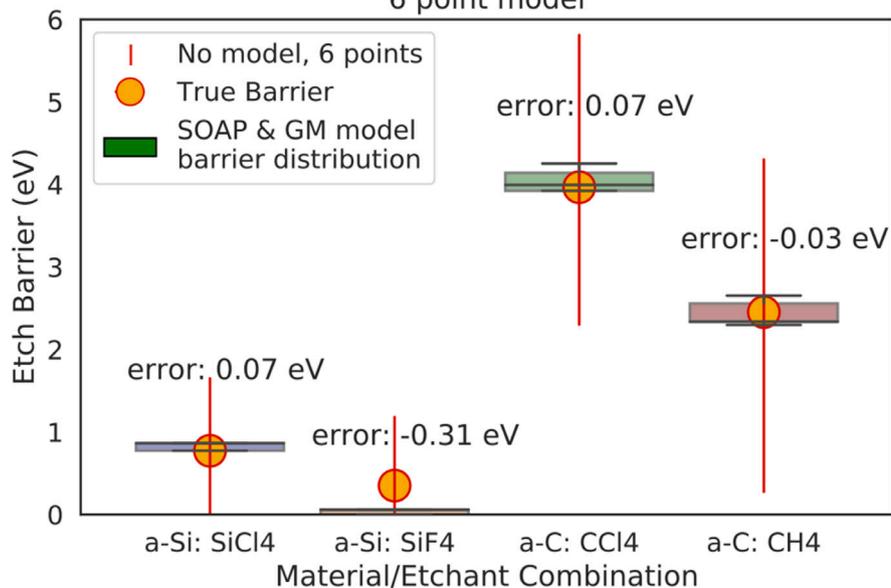


Fig. 5. For the graph above, a SOAP model with cutoff radius of 3 Å for a-Si and 2 Å for a-C, maximum radial contribution 3, maximum angular contribution of 2 and a Gaussian width of 0.4 was chosen. The model was applied to a-C and a-Si surfaces separately to obtain site-feature vectors. Once features were obtained, Bayesian Gaussian Mixture models were generated 500 times using each of these features to group the sites into labels. For each Gaussian Mixture model, the two closest sites to the cluster center using Mahalanobis distance were chosen, and the proper weighted average based on number of cluster samples was taken. The SOAP & 3-cluster Gaussian Mixture model, with each cluster sampled twice, led to narrow energy distributions of potential barrier energy prediction for etching by these etchants of these materials. The errors are 70, -310, 70, and -30 meV respectively. For each of these material/volatile etch product combination the model performs better than no model (the red vertical line demonstrates the range of modeled barriers if 6 sites on the surface had been chosen at random). The yellow circle represents the true barrier, or the average of all sites on the surface.

range of etching sites and barriers are accounted for. A useful application of these methods is to quickly screen etching gases/materials for selectivity and improved etch rates finding the best combination computationally at the start of process development, shrinking development time, and optimizing costs. Further analysis, such as performing importance learning sampling [67,68], can always be performed by users separately.

#### 3.3. Benchmarking of model

The simulation results from the etching of the fifteen sites on the a-Si generated surface, forming SiCl<sub>4</sub> product were analyzed by featurization and clustering of the surface sites, as described in the Site Identification & Clustering section. We specifically chose the a-Si/SiCl<sub>4</sub> surface/volatile etch product combination due to the completeness of the data (all but one site tested on this surface finished to convergence) and because not as many sites were barrier-less (in comparison to a-Si/SiF<sub>4</sub>). The result of the SOAP featurization and Bayesian Gaussian Mixture clustering model (SOAP & GM) was compared to using no model. The potential distribution of predicted energy barriers if six sites were chosen at random and averaged to calculate an overall etching barrier (no model) was benchmarked against picking the two closest points to the cluster centers in three different clusters from the SOAP & GM model. Because a Bayesian Gaussian Mixture model is not always unique in its cluster creation, the model was simulated 500 times and each result was recorded. Comparing the results between the two cases, we found a clear trend: applying the SOAP & GM model provided a much narrower distribution of predicted etching energy barriers. Furthermore, using the model to reduce all fifteen sites on the surface to calculating only six sites, gave a narrow error of 70-meV from the overall etching barrier across all the sites, with 75% of the prediction falling within a narrow energy bandwidth of 20-meV. In comparison, a site-specific distribution of etching barriers (e.g. no model) provided an energy bandwidth of 3800 meV, with 75% of the prediction within an error bound of 1600 meV. Thus, utilizing the model to describe

and cluster the sites and their environments decreased the error over twentyfold.

We evaluate the versatility of the workflow by comparing the number of clusters, as well as the clustering algorithm, between a Gaussian Mixture model, a Bayesian Gaussian Mixture model, and a KMeans model. Five hundred simulations were performed for each clustering algorithm, and the two closest points to each cluster center were chosen and the results compared to having no model. For metrics of comparison, we chose the model's predicted bandwidth of the etching barriers and the mean of the predicted barriers relative to the mean of all etching barriers on the surface ("true barrier"). The number of clusters was varied while using a Bayesian Gaussian Mixture clustering algorithm. It was found that while increasing from three to four clusters improved the mean error, the resulting bandwidth of predicted etching barriers increased as well. In addition, further increasing the cluster number did not improve the mean error or the bandwidth. In terms of the clustering algorithm, the Bayesian Gaussian Mixture model performed the best: it had the least deviation from the true barrier and the smallest bandwidth of predicted etching barriers. In addition, a Bayesian Gaussian Mixture model with different types of covariances: full, diagonal, and spherical, was tested. The full Bayesian Gaussian Mixture also performed the best. Detailed comparison of the resulting barrier bandwidth and the error from the true barrier of the covariance types tested for the Bayesian Gaussian Mixture model, as well as from variations of the clustering model, including the Bayesian Gaussian Mixture, Gaussian Mixture, and KMeans model can be found in the SI.

Different parameters of SOAP including the cutoff radius, the highest radial contribution, the highest angular contribution considered, and the width of the radial basis function were also tested. A small cut-off radius on the order of typical bond length was found to work the best across all simulations. Similarly, a small number of radial and angular contributions performed better. Finally, a Gaussian width of 0.5–0.7 performed best. Detailed comparison of the resulting energy bandwidth of the etching barriers and errors upon variation of

the SOAP parameters can be found in the SI. While the size of the dataset limited the tuning of the SOAP hyperparameters, nonetheless the model's performance is indicative of the smoothness of the SOAP algorithm. In Fig. 5 the optimal SOAP parameters found along with the Bayesian Gaussian Mixture 3-cluster model were generalized across all material/volatile etch product combinations calculated. For each combination, the resulting model performed better than no model. The errors were narrow at 70, -310, 70, and -30 meV for a-Si/SiCl<sub>4</sub>, a-Si/SiF<sub>4</sub>, a-C/CCl<sub>4</sub>, a-C/CH<sub>4</sub> respectively. In addition, in each case, the resulting modeled etch-barrier distribution was over tenfold narrower than the etch-barrier distribution with no model.

We emphasize that the workflow easily allows switching the site featurizer to any other relevant Matminer Base Featurizer class. For demonstration purposes, SOAP was compared to other radial basis-based featurizers, including the Generalized Gradient Distribution Function (GRDF), Angular Fourier Series (AFS) [69], and AGNI Fingerprint [70]. The cutoff parameter was also tuned for both the GRDF and AFS featurizers. While other featurizers could perform better than SOAP on a-Si, etched by SiCl<sub>4</sub>, with some tuning of the hyperparameters, these featurizers were not as generalizable to other material/volatile etch product combinations. Within the set of tested featurizers, SOAP proved to be the most generalizable across material composition and volatile etch product tested. More detailed information on how these featurizers, and their tuned hyper-parameters affected the resulting energy barrier distribution and error can be found in the SI.

#### 4. Conclusions

We developed a methodology to reduce potential sites to study on amorphous material surfaces. The approach relies on featurizing surface sites using the Smooth Overlap of Atomic Positions and clustering them using a Bayesian Gaussian Mixture model. We applied this method to dry plasma etching, an important process for the semiconductor industry. Using this method, we demonstrated that we were able to reduce the fifteen surface sites for a-Si and 22 surface sites for a-C down to 6 calculations to achieve an etching barrier prediction with reasonable error. The methodology developed to analyze surface sites on amorphous surfaces can be used for any simple, site-specific computational study, specifically accelerating in-silico prediction of amorphous surface properties and reactivities, previously a costly computational feat to achieve.

#### CRedit authorship contribution statement

**Martin Siron:** Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Nita Chandrasekhar:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision. **Kristin A. Persson:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Martin Siron, Nita Chandrasekhar reports financial support and equipment or supplies were provided by Intel Corporation. Martin Siron, Kristin Persson reports administrative support, article publishing charges, and writing assistance were provided by University of California Berkeley. Martin Siron, Nita Chandrasekhar reports a relationship with Intel Corporation that includes: employment.

#### Data availability

Data will be made available on request.

#### Acknowledgments

M.S. acknowledges support from the National Science Foundation Graduate Research Fellowship, USA under grant no. DGE 1752814. M.S. also acknowledges great support from Karson Knutson during the time of his internship at Intel Corporation.

#### Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.commatsci.2023.112192>. CP2K details for the PM6 MD + PBE DFT calculations for generating amorphous surfaces, as well as NEBs for a-C, a-Si, and c-Si etching forming various volatile etch products.

#### References

- [1] A.L. Lacaite, D.J. Wouters, Phase-change memories, *Phys. Status Solidi A* 205 (10) (2008) 2281–2297.
- [2] L.-F. Cui, R. Ruffo, C.K. Chan, H. Peng, Y. Cui, Crystalline-amorphous coreshell silicon nanowires for high capacity and high current battery electrodes, *Nano Lett.* 9 (1) (2009) 491–495, PMID: 19105648.
- [3] M. Shur, M. Hack, Physics of amorphous silicon based alloy field-effect transistors, *J. Appl. Phys.* 55 (10) (1984) 3831–3842.
- [4] D.E. Carlson, C.R. Wronski, Amorphous silicon solar cell, *Appl. Phys. Lett.* 28 (11) (1976) 671–673.
- [5] S. Hu, M.H. Richter, M.F. Lichterman, J. Beardslee, T. Mayer, B.S. Brunshwig, N.S. Lewis, Electrical, photoelectrochemical, and photoelectron spectroscopic investigation of the interfacial transport and energetics of amorphous TiO<sub>2</sub>/Si heterojunctions, *J. Phys. Chem. C* 120 (6) (2016) 3117–3129.
- [6] S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, *Nature Mater.* 12 (3) (2013) 191–201.
- [7] R. Car, M. Parrinello, Structural, dynamical, and electronic properties of amorphous silicon: An ab initio molecular-dynamics study, *Phys. Rev. Lett.* 60 (1988) 204–207.
- [8] J. Tersoff, New empirical approach for the structure and energy of covalent systems, *Phys. Rev. B* 37 (1988) 6991–7000.
- [9] R.L. McGreevy, Reverse Monte Carlo modelling, *J. Phys.: Condens. Matter* 13 (46) (2001) R877–R913.
- [10] A. Pandey, P. Biswas, D.A. Drabold, Inversion of diffraction data for amorphous materials, *Sci. Rep.* 6 (1) (2016) 33731.
- [11] G. Opletal, T. Petersen, B. O'Malley, I. Snook, D.G. McCulloch, N.A. Marks, I. Yarovsky, Hybrid approach for generating realistic amorphous carbon structure using metropolis and reverse Monte Carlo, *Mol. Simul.* 28 (10–11) (2002) 927–938.
- [12] V.L. Deringer, N. Bernstein, A.P. Bartók, M.J. Cliffe, R.N. Kerber, L.E. Marbella, C.P. Grey, S.R. Elliott, G. Csányi, Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics, *J. Phys. Chem. Lett.* 9 (11) (2018) 2879–2885, PMID: 29754489.
- [13] N. Artrith, Machine learning for the modeling of interfaces in energy storage and conversion materials, *J. Phys.: Energy* 1 (3) (2019) 032002.
- [14] A.P. Bartók, M.C. Payne, R. Kondor, G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, *Phys. Rev. Lett.* 104 (2010) 136403.
- [15] M.A. Caro, A. Aarva, V.L. Deringer, G. Csányi, T. Laurila, Reactivity of amorphous carbon surfaces: Rationalizing the role of structural motifs in functionalization using machine learning, *Chem. Mater.* 30 (21) (2018) 7446–7455.
- [16] B.R. Goldsmith, B. Peters, J.K. Johnson, B.C. Gates, S.L. Scott, Beyond ordered materials: Understanding catalytic sites on amorphous solids, *ACS Catal.* 7 (11) (2017) 7543–7557.
- [17] B. Peters, S.L. Scott, Single atom catalysts on amorphous supports: A quenched disorder perspective, *J. Chem. Phys.* 142 (10) (2015) 104708.
- [18] V.M. Donnelly, A. Kornblit, Plasma etching: Yesterday, today, and tomorrow, *J. Vac. Sci. Technol. A* 31 (5) (2013) 050825.
- [19] G.S. Oehrlein, J.F. Rembetski, Plasma-based dry etching techniques in the silicon integrated circuit technology, *IBM J. Res. Dev.* 36 (2) (1992) 140–157.
- [20] F. Gou, A.W. Kleyn, M.A. Gleeson, The application of molecular dynamics to the study of plasma-surface interactions: CF<sub>x</sub> with silicon, *Int. Rev. Phys. Chem.* 27 (2) (2008) 229–271.
- [21] L.M. Ephraïm, Selective etching of silicon dioxide using reactive ion etching with CF<sub>4</sub>-H<sub>2</sub>, *J. Electrochem. Soc.* 126 (8) (1979) 1419–1421.
- [22] H. Park, D. Woo, J.M. Lee, S.J. Park, S. Lee, H.J. Kim, E. Yoon, G.-D. Lee, The influence of hydrogen concentration in amorphous carbon films on mechanical properties and fluorine penetration: a density functional theory and ab initio molecular dynamics study, *RSC Adv.* 10 (2020) 6822–6830.

- [23] E.A. Ogryzlo, D.E. Ibbotson, D.L. Flamm, J.A. Mucha, Doping and crystallographic effects in Cl-atom etching of silicon, *J. Appl. Phys.* 67 (6) (1990) 3115–3120.
- [24] J. Li, S.J. Kim, S. Han, Y. Kim, H. Chae, Etching characteristics of hydrogenated amorphous carbon with different sp<sup>2</sup>/sp<sup>3</sup> hybridization ratios in CF<sub>4</sub>/O<sub>2</sub> plasmas, *Plasmas Plasma Process. Polym.* e2100075.
- [25] H.-H. Doh, J.-H. Kim, S.-H. Lee, K.-W. Whang, Mechanism of selective SiO<sub>2</sub>/Si etching with fluorocarbon gases (CF<sub>4</sub>, C<sub>4</sub>F<sub>8</sub>) and hydrogen mixture in electron cyclotron resonance plasma etching system, *J. Vac. Sci. Technol. A* 14 (5) (1996) 2827–2834.
- [26] D. Edelson, D.L. Flamm, Computer simulation of a CF<sub>4</sub> plasma etching silicon, *J. Appl. Phys.* 56 (5) (1984) 1522–1531.
- [27] C. Cardinaud, G. Turban, Mechanistic studies of the initial stages of etching of Si and SiO<sub>2</sub> in a CHF<sub>3</sub> plasma, *Appl. Surf. Sci.* 45 (2) (1990) 109–120.
- [28] M. Aykol, K.A. Persson, Oxidation protection with amorphous surface oxides: Thermodynamic insights from Ab initio simulations on aluminum, *ACS Appl. Mater. Interfaces* 10 (3) (2018) 3039–3045, PMID: 29297220.
- [29] J.J.P. Stewart, Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *J. Mol. Model.* 13 (12) (2007) 1173–1213.
- [30] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* 77 (18) (1996) 3865–3868.
- [31] A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments, *Phys. Rev. B* 87 (2013) 184115.
- [32] K. Mathew, J.H. Montoya, A. Faghaninia, S. Dwarakanath, M. Aykol, H. Tang, I.-h. Chu, T. Smidt, B. Bocklund, M. Horton, J. Dagdelen, B. Wood, Z.-K. Liu, J. Neaton, S.P. Ong, K. Persson, A. Jain, Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows, *Comput. Mater. Sci.* 139 (2017) 140–152.
- [33] A. Jain, S.P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, K.A. Persson, FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications, *Concurr. Comput.: Pract. Exper.* 27 (17) (2015) 5037–5059.
- [34] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, *Comput. Mater. Sci.* 68 (2013) 314–319.
- [35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [36] L. Ward, A. Dunn, A. Faghaninia, N.E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G.J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, *Comput. Mater. Sci.* 152 (2018) 60–69.
- [37] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele, Cp2k: atomistic simulations of condensed matter systems, *WIREs Comput. Mol. Sci.* 4 (1) (2014) 15–25.
- [38] L. Martínez, R. Andrade, E.G. Birgin, J.M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations, *J. Comput. Chem.* 30 (13) (2009) 2157–2164.
- [39] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, *J. Chem. Phys.* 81 (1) (1984) 511–519.
- [40] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.* 52 (2) (1984) 255–268.
- [41] J.J.P. Stewart, Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *J. Mol. Model.* 13 (12) (2007) 1173–1213.
- [42] J. VandeVondele, J. Hutter, An efficient orbital transformation method for electronic structure calculations, *J. Chem. Phys.* 118 (10) (2003) 4365–4369.
- [43] J. VandeVondele, J. Hutter, Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases, *J. Chem. Phys.* 127 (11) (2007) 114105.
- [44] S. Goedecker, M. Teter, J. Hutter, Separable dual-space Gaussian pseudopotentials, *Phys. Rev. B* 54 (1996) 1703–1710.
- [45] K. Persson, Materials data on Si (SG:227) by materials project, 2014, An optional note.
- [46] W. Sun, G. Ceder, Efficient creation and convergence of surface slabs, *Surf. Sci.* 617 (2013) 53–59.
- [47] A. Shrake, J. Rupley, Environment and exposure to solvent of protein atoms. Lysozyme and insulin, *J. Mol. Biol.* 79 (2) (1973) 351–371.
- [48] S. Mitternacht, FreeSASA: An open source C library for solvent accessible surface area calculations, *F1000Res* 5 (2016) 189.
- [49] L. Himanen, M.O.J. Jäger, E.V. Morooka, F. Federici Canova, Y.S. Ranawat, D.Z. Gao, P. Rinke, A.S. Foster, DScribe: Library of descriptors for machine learning in materials science, *Comput. Phys. Comm.* 247 (2020) 106949.
- [50] J. Mavračić, F.C. Mocanu, V.L. Deringer, G. Csányi, S.R. Elliott, Similarity between amorphous and crystalline phases: The case of TiO<sub>2</sub>, *J. Phys. Chem. Lett.* 9 (11) (2018) 2985–2990, PMID: 29763315.
- [51] P.C. Mahalanobis, 1936.
- [52] S. Kondati Natarajan, S.D. Elliott, Modeling the chemical mechanism of the thermal atomic layer etch of aluminum oxide: A density functional theory study of reactions during HF exposure, *Chem. Mater.* 30 (17) (2018) 5912–5922.
- [53] T. Chowdhury, R. Hidayat, T.R. Mayangsari, J. Gu, H.-L. Kim, J. Jung, W.-J. Lee, Density functional theory study on the fluorination reactions of silicon and silicon dioxide surfaces using different fluorine-containing molecules, *J. Vac. Sci. Technol. A* 37 (2) (2019) 021001.
- [54] S. Park, H. Jung, K.-A. Min, J. Kim, B. Han, Unraveling the selective etching mechanism of silicon nitride over silicon dioxide by phosphoric acid: First-principles study, *Appl. Surf. Sci.* 551 (2021) 149376.
- [55] A. Jenichen, C. Engler, Etching of GaAs(100) surfaces by Cl<sub>2</sub>: quantum chemical calculations on the mechanisms, *J. Phys. Chem. B* 104 (34) (2000) 8210–8216.
- [56] T.A.R. Hansen, J.W. Weber, P.G.J. Colsters, D.M.H.G. Mestrom, M.C.M. van de Sanden, R. Engeln, Synergistic etch rates during low-energetic plasma etching of hydrogenated amorphous carbon, *J. Appl. Phys.* 112 (1) (2012) 013302.
- [57] W. Kabsch, A solution for the best rotation to relate two sets of vectors, *Acta Crystallogr. Sect. A* 32 (5) (1976) 922–923.
- [58] G. Henkelman, H. Jónsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, *J. Chem. Phys.* 113 (22) (2000) 9978–9985.
- [59] N.E.R. Zimmermann, M.K. Horton, A. Jain, M. Haranczyk, Assessing local structure motifs using order parameters for motif recognition, interstitial identification, and diffusion path characterization, *Front. Mater.* 4 (2017) 34.
- [60] J.K. Kim, S.I. Cho, N.G. Kim, M.S. Jhon, K.S. Min, C.K. Kim, G.Y. Yeom, Study on the etching characteristics of amorphous carbon layer in oxygen plasma with carbonyl sulfide, *J. Vac. Sci. Technol. A* 31 (2) (2013) 021301.
- [61] R.G. Poulsen, Plasma etching in integrated circuit manufacture—A review, *J. Vac. Sci. Technol.* 14 (1977) 266.
- [62] Y. Zhang, M. Imamura, K. Ishikawa, T. Tsutsumi, H. Kondo, M. Sekine, M. Hori, Reaction mechanisms between chlorine plasma and a spin-on-type polymer mask for high-temperature plasma etching, *Japan. J. Appl. Phys.* 57 (10) (2018) 106502.
- [63] F. Li, J.S. Lannin, Radial distribution function of amorphous carbon, *Phys. Rev. Lett.* 65 (1990) 1905–1908.
- [64] G.O. Brunner, A definition of coordination and its relevance in the structure types AlB<sub>2</sub> and NiAs, *Acta Crystallogr. Sect. A* 33 (1) (1977) 226–227.
- [65] S.J. Clark, J. Crain, G.J. Ackland, Comparison of bonding in amorphous silicon and carbon, *Phys. Rev. B* 55 (1997) 14059–14062.
- [66] A. Pranda, S.A. Gutierrez Razo, Z. Tomova, J.T. Fourkas, G.S. Oehrlein, Role of the dense amorphous carbon layer in photoresist etching, *J. Vac. Sci. Technol. A* 36 (2) (2018) 021304.
- [67] C.A. Vandervelden, S.A. Khan, B. Peters, Importance learning estimator for the site-averaged turnover frequency of a disordered solid catalyst, *J. Chem. Phys.* 153 (24) (2020) 244120.
- [68] S.A. Khan, C.A. Vandervelden, S.L. Scott, B. Peters, Grafting metal complexes onto amorphous supports: from elementary steps to catalyst site populations via kernel regression, *React. Chem. Eng.* 5 (2020) 66–76.
- [69] A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, I. Tanaka, Representation of compounds for machine-learning prediction of physical properties, *Phys. Rev. B* 95 (2017) 144110.
- [70] V. Botu, R. Ramprasad, Adaptive machine learning framework to accelerate ab initio molecular dynamics, *Int. J. Quantum Chem.* 115 (16) (2015) 1074–1083.