Charting the Atomic C Interaction with Transition Metal Surfaces

Oriol Piqué,[†,¶] Iskra Z. Koleva,[‡,¶] Albert Bruix,[†] Francesc Viñes,[†,*] Hristiyan A. Aleksandrov,[‡,*] Georgi N. Vayssilov,[‡] and Francesc Illas[†]

[†] *Departament de Ciència de Materials i Química Física & Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, c/ Martí i Franquès 1, 08028 Barcelona, Spain*

[‡] *Faculty of Chemistry and Pharmacy, University of Sofia, 1126 Sofia, Bulgaria*

*Emails: francesc.vines@ub.edu and haa@chem.uni-sofia.bg

Abstract

Carbon interaction with transition metal (TM) surfaces is a relevant topic in heterogeneous catalysis, either for its poisoning capability, for the recently attributed promoter role when incorporated in subsurface, or for the formation of early transition metal carbides, which are increasingly used in catalysis. Herein we present a high-throughput systematic study, adjoining thermodynamic *plus* kinetic evidence obtained by extensive density functional calculations on surface models (324 diffusion barriers located on 81 TM surfaces in total), which provides a navigation map of these interactions in a holistic fashion. Correlation between previously proposed electronic descriptors and ad/absorption energies has been tested, with *d*-band centre being found the most suitable one, although machine learning protocols underscore also the importance of the surface energy and the site coordination number. Descriptors have also been tested for diffusion barriers, with ad/absorption energies and the difference in energy between minima being the most appropriate ones. Furthermore, multivariable, polynomial, and random forest regressions show that both thermodynamic and kinetic data are better described when using a combination of different descriptors. Therefore, looking for a single perfect descriptor may not be the best quest, while combining different ones may be a better path to follow.

Keywords: Carbon Atoms, Transition Metal Surfaces, Adsorption, Absorption, Diffusion, Descriptors, Machine Learning, Phase Diagrams

[¶] Both authors equally contributed

1. Introduction

Nowadays, transition metals (TMs) are ubiquitous in many areas of science and technology, from Solid State Physics through Applied Chemistry up to Materials Science, with relevance in diverse industrial fields, such as Nanotechnology and Heterogeneous Catalysis. In fact, late TMs, including noble coinage (Ni, Cu, Ag, Au) and Pt-group (Pt, Pd, Rh, Ir, Re, Os) metals, are in widespread use as heterogeneous catalysts[1] for a large diversity of reactions of industrial interest, *e.g.* ammonia synthesis, exhaust gas treatments, or the Fischer-Tropsch reaction, to name just a few.[2,3] However, early TMs are far too active for such purposes according to *Le Sabatier* principle,[4] adsorbing chemical moieties too strongly. Their heterogeneous catalysis interest though lies into the TM carbides (TMCs) formation, since the carburization of these metals diminishes their chemical activity, to the point of making TMCs viable replacements of the above-commented late TMs in catalysis,[5] featuring improved catalytic performances, selectivities, or poison resistances compared to Pt-group TMs.[6,7]

The relative simplicity of TM systems and the industrial importance of their applications has prompted state-of-the-art research aimed at unveiling their catalytic activity improvement, desirably coupled with materials cost reduction, as the one achieved when using Earth-abundant TMCs. To this end nanostructuration, alloys usage, even the design of nanoalloys have been contemplated as plausible strategies.[8,9] The rational design of novel metal and alloy catalysts, backed up by precise *ab initio* quantum chemistry calculations on proper catalyst models, has meant a great leap forward in the quest for new, improved activity TM catalysts.[9,10]

Still, however, one main drawback of such catalysts is that, in the course of the catalysed reaction, these get gradually deactivated over time and use, with the concomitant loss in efficiency and economic profit. The origin of this loss is the presence of poisoning agents, where carbon excels among others.[11,12] Thus, the interaction of C with TMs is indeed a fundamental field of study in Heterogeneous Catalysis. Historically, from the catalytic deactivation processes point of view, it is quite usual that carbon entities, going from C atoms to small C_n aggregates, are formed on the TM catalyst surface due to secondary reactions of the on-going surface catalysed process, generally involving organic molecules. These carbon entities can spread through the surface and eventually agglomerate forming diverse types of carbon deposits, from graphene through graphite, up to amorphous C aggregates. These surface carbon structures can cover the catalyst, *de*

facto restricting the access of reactants to the TM catalytic surface active sites, and effectively poisoning it.

Despite this, recent experiments and computational simulations have changed the paradigm view of the low C content from a poisoner to a promoter role. For instance, subsurface C into Pd catalysts appears to favour the selective alkyne hydrogenation to olefins,[13] and its presence, explained by simulations based on density functional theory (DFT), shows how such subsurface C can be easily present at the surfaces of late TM surfaces and nanoparticles (NPs).[14-17] Moreover, subsurface C plays an important role in the synthesis of graphene and carbon nanotubes (CNT), where segregation of C atoms diluted in pre-molten or molten TMs appears when cooling down the system, which induces the growth of graphene layers, even CNTs.[18,19]

Furthermore, subsurface C has been found to bias the appearance of other substitutional or interstitial carbon residues in Pd,[20] which may display even higher reactivity towards surface O and H adatoms than surface C,[21] and to act as a gate opener for H absorption.[22] Such subsurface moieties mediated chemistry is non-exclusive to neither C nor Pd. Indeed, subsurface C has been proposed to be a key player in the electrochemical conversion of CO_2 on Cu surfaces,[23] and interstitial C in Au NPs has been found to affect the metal electronic structure, being the ultimate responsible of the three-times increment of the measure turnover frequency in the chemoselective hydrogenation of 3-nitrostyrene.[24] Apart from subsurface C, it is worth mentioning that, subsurface O can also affect the on-going surface activity, see *e.g.* the recent key role of subsurface O on Cu (111) surface in CO_2 capture and activation applications, with critical implications in environmental chemistry.[25] Subsurface chemistry has often been ignored, but is now growingly attracting attention within the scientific community, seen as a change of paradigm in what surface chemical activity is concerned.[26]

Motivated by these results, the primal aim of this work is to deliver, for the first time, a broad, detailed, and holistic atomistic view of the interaction of C with TM surfaces. This is achieved by studying, by *ab initio* DFT calculations on proper surface slab models, the stability of C atoms on the Miller surfaces with index order of 1 for all those TMs featuring a face-centred cubic (*fcc*), body-centred cubic (*bcc*), or hexagonal close-packed (*hcp*) bulk crystallographic structure —see Figures S1 and S2 of Section S1 of the Supporting Information (SI). For such surfaces, most stable surface and subsurface sites are identified and their bond strengths seized, so as to gain thermodynamic pictures and stability phase diagrams, as done earlier for *fcc* TM (111) surfaces.[15] Furthermore, all sorts of C diffusion

energy barriers, E_b, are explored, including surface, subsurface, sinking, and emerging diffusion energy barriers for each metal surface, denoted E_{sur}, E_{sub}, E_{sink}, and E_{emer}, respectively —see Figure S3 in Section S1 of the SI for a depiction of the different barrier types. This systematic study will provide a navigation chart of the C tendency to poison surface active sites, to aggregate into C_n moieties, and will also provide insights on the possible formation of TMCs. Finally, the acquired amount of data allows for further analysis based on artificial intelligence (AI) and machine-learning (ML) regression algorithms, addressed at defining subgroup types of similar behaviour concerning C bond strength and diffusivity, the main physicochemical descriptors defining these, as well as regressions of adsorptive and diffusive properties as a function of physicochemical descriptors.

2. Computational Details

Present DFT calculations have been performed using the Vienna *ab initio* simulation package (VASP) code,[27] imposing periodic symmetry. Relaxed geometries and total energies were acquired using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional,[28] known to accurately describe TMs bulks and surfaces and also the interaction of C with them.[16,29] Moreover, previous works show that relative stabilities and diffusion energy barriers, E_b, are rather unbiased by the used exchange-correlation functional, with small variations between 1-5 kJ mol^{-1} depending on the used functional.[16] The valence electrons density was expanded in a plane wave basis set with a 415 eV cutoff for the kinetic energy, while the projector augmented wave method was used to describe the interactions between core and valence electrons.[30] Calculations were carried out non spin-polarized except for magnetic Ni, Co, and Fe TMs. Geometry optimizations were performed until all forces acting on relaxed atoms became lower than 0.03 eV $Å^{-1}$, and the electronic convergence threshold was set to 10^{-6} eV.

The TMs most stable surfaces with Miller index with maximum order of 1 have been modelled using slabs, generally including most stable surfaces, with no step defects. These are the (001), (011), and (111) Miller indices surfaces of *fcc* and *bcc* TMs; and the (0001), $(10\bar{1}0)$, and $(11\bar{2}0)$ of *hcp* TMs, following Miller-Bravais indices notation for *hcp* metals, being a total number of 81 modelled surfaces.[29] The simulation of extended surfaces has been performed using supercell slab models constructed from previously PBE optimized bulks,[31,32] see Figure S2 in Section S1 of the SI for a depiction of the explored adsorption/absorption sites. The supercell size depends on the specific surface termination

being modelled; (3×3) supercells composed of 54 atoms were used for *fcc* (111), *hcp* $(10\bar{1}0)$, *bcc* (001), and *bcc* (111) surfaces, while (2×2) supercells composed of 48 atoms were used for *fcc* (001), *fcc* (011), *hcp* (0001), *hcp* $(11\bar{2}0)$, and *bcc* (011) surfaces. All surface slab models contain six atomic layers; with nine atoms per layer for (3×3) slabs or eight atoms per layer for (2×2) slabs. Consequently, the adsoprtion/absorption of one C atom implies a similar coverage of $^1/_9$ or $^1/_8$ monolayers (ML), respectively, defined as the number of C atoms with respect the number of surface metal atoms on one side of the slab.

After optimization of the pristine surfaces one C atom was adsorbed/absorbed with the three bottom layers of the slab fixed, while the other three upper layers were allowed to relax during the geometry optimization together with the interacting adsorbed/absorbed C atom, a procedure known as 3+3 approximation. The reciprocal space was sampled with 3×3×1 **Γ**-centred Monkhorst Pack[33] **k**-point grid and calculations were performed using a Gaussian smearing of 0.2 eV energy width to speed up convergence, yet final energies were extrapolated to 0 K (no smearing).

The adsorption/absorption energies, Eads/abs have been calculated as

$$E_{ads/abs} = -E_{C/metal} + E_C + E_{metal} \quad (1),$$

where $E_{C/metal}$ is the total energy of the metal slab with the C atom either adsorbed or absorbed, E_C is the energy of the isolated carbon atom in vacuum, and E_{metal} is the energy of the optimized clean TM substrate. Within this definition, stable adsorption/absorption correspond to positive $E_{ads/abs}$ values. Zero-point energies imply negligible variations between the stability of the different sampled sites, below 0.1 kJ·mol^{-1} according to test calculations, and so have been disregarded in the final analysis.

The surface, subsurface, sinking, and emerging E_b, these are, the E_{sur}, E_{sub}, E_{sink}, and E_{emer}, were determined by using the climbing image nudged elastic band (CI-NEB) procedure using four images in between initial and final states.[34] The approximate transition state (TS) geometries were posteriorly refined using a Quasi-Newton optimization algorithm until forces acting on atoms were under 0.03 eV Å^{-1}. All TS were characterized by vibrational frequency analysis performed *via* construction and diagonalization of the corresponding block of the Hessian matrix, with elements estimated by analytical gradients from finite displacements of 0.03 Å length, certifying their saddle point nature with only one imaginary frequency. The E_b values were calculated subtracting the TS energy from the initial diffusive energy state.

As far as the AI algorithms used in the analysis are concerned, the group analysis was carried out using the k-means (KM) approach, as implemented in the *sklearn* python library.[35] The number of clusters for each case was determined using the elbow method,[36] consisting in the evaluation of the cluster inertia (or distortion) curve shape —defined as the sum of squared distances of samples to their closest cluster center plotted against the number of clusters, and selecting the elbow of the curve, *i.e.* the minimum number of clusters that already yields a sought accuracy. This number of clusters is then used in the cluster groups descriptions, see more details in Section S2 of the SI. KM clustering was used to define subgroup types of TMs with similar activity with respect C atoms, or similar diffusions. KM was also used to correlate these subgroups with the main descriptors defining these.

Concerning the tested ML regression algorithms, those were also implemented using the *sklearn* python library. Aside, $E_{ads/abs}$ and E_b magnitudes were correlated with a list of TM features or physicochemical descriptors, commented in the forthcoming sections, using three different ML regression models: multivariable linear regression (MLR), decision tree regression (DTR), and random forest regression (RFR). Models were refined by removing unnecessary features using the leave-one-out procedure, and hyperparameters were tuned by conducting a grid search involving different parameter combinations and selecting the best preforming ones.

3. Results and Discussion

3.1. Thermodynamic Analysis

3.1.1. Energy landscape

Let us first start with an analysis of the thorough study of C interaction with the explored 81 transition metal surface models, so as to provide a general view of the C interaction with TMs. A full list of adsorption and absorption energies of the gained minima is present in Tables S1-S3 of Section S3 in the ESI. For a better readability, the top panel of Figure 1 shows a summarized overview of the results displaying the most stable position adsorption, E_{ads}, or absorption, E_{abs}, energy values for each surface termination of each metal —thus presenting three different values per metal— *versus* the C height, h, defined as the vertical distance between the C atom and the TM surface plane in each particular position, compiled in Tables S4-S6 of Section S3 of the SI. So as to have clear defined references, an in-plane situation is shown at zero h, along with the graphite cohesive energy, E_{coh}, obtained from the literature and obtained through equivalent optimizations as the ones here

presented in terms of employed DFT exchange-correlation functional, plane-wave cutoff, **k**-points density, etc.[37] As can be seen in Figure 1, the display defines four different quadrants, which imply four different behaviors of C atoms when interacting with TMs surfaces, depending on whether an adsorption or absorption situation is preferred, and whether the interaction of C with the TM is stronger or weaker compared to that in graphite.

A first analysis can be carried out with the data color-coded according to the TM crystallographic structure; this is, *fcc*, *hcp*, or *bcc*. At first glimpse, one notices that nearly all *bcc* values are found to be above the surface limit, implying a preferential adsorption, except for three cases —V, Nb, and Ta (011) surfaces, all being group V of the Periodic Table— where carbon prefers to lie subsurface. The $E_{ads/abs}$ above E_{coh} implies that C has more affinity for *bcc* TMs than for other C atoms, a feature that happens in many cases, and in the above-commented subsurface cases, demonstrating a clear thermodynamic tendency for C to penetrate these metal surfaces. These results are indeed in perfect agreement with these metals propensity to form TMCs.[5] However, since the majority of points imply surface positions, the carburization of *bcc* TMs, implying the interstitial placement of C, appears to be a non-trivial process. Notice that *bcc* crystallographic structures are perfect templates to place C in their octahedral interstitials, thus featuring the final rocksalt crystallographic structure of TMCs. Therefore, crystal reconstruction appears not to be a restricting aspect in many of these cases exhibiting rocksalt TMCs, such as VC, NbC, TaC, CrC, δ-MoC, and FeC,[5] although, for other more stable polymorphs, when necessary, the C incorporation is likely to involve a crystallographic phase transition. Previous studies on Zn oxidation and other TMs hydrogenation reflect a preferential subsurface stabilization of O and H atoms, including superficial crystallographic reconstructions,[38] featured at higher atomic coverages, and so, such type of mechanisms could well explain the eventual subsurface incorporation of C to form TMCs, even when implying a crystal transformation.

Going beyond *bcc* TMs, *fcc* TM values are found in their majority in the lower part of the panel, *i.e.* adsorption or absorption energies weaker than in graphite, meaning that C has more clustering affinity than for these metals. This is in fair agreement with the known fact of *fcc* TMs being poisoned due to the formation of graphene/graphite layers on them, blocking their surface active sites. Nevertheless, this preference to form graphene layers may be interesting in the context of nanotechnology, *e.g.* in graphene synthesis from chemical vapor deposition or segregation.[18] There are a couple of cases —Ni and Rh

(001)— where the C interaction is above the graphite limit, which presents them as a the least favorable for graphene synthesis, and more favorable concerning carbide formation, although their preference is on the surface, or in-plane. Note, however, that such strongly attached C can block still active sites, or perturb the very chemical nature of nearby metal surface atoms. Finally, there are six *fcc* values that are clearly in the subsurface region: The two upper-most data points correspond to Ni and Pd (111), in agreement with previous calculations and experiments that certify the existence of carbide phases of these TMs.[16] Moreover, the four lowest values correspond to Cu (111), Ag (111), Ag (011), and Au (011); indicating, as recently showed in the literature, that C has a tendency to penetrate subsurface on such noble metals.[15]

Last but not least, and regarding *hcp* TMs, their values are scattered, as is their position in the Periodic Table. Still, grouping is observed in the subsurface, highly attached C region, where C appears to present a tendency to penetrate such TMs. In fact, these values correspond to Sc, Y, Ti, Zr, Hf, Zn, and Cd. The first five belong to groups III and IV of the Periodic Table, known to display a very high chemical activity and a propensity to form rocksalt TMCs. Thus, clearly the high absorption energy could well contribute in compensating the energy demands for an eventual phase transition towards a rocksalt structure. On the other hand, Zn and Cd are d^{10} elements, with an *a priori* low chemical activity according to *d*-band center arguments, see below. However, the strong interaction calculated here is in line with the existence of such carbides, as reported in previous studies.[39]

The data in Figure 1 has been used to carry out a KM clustering to better understand how this data can be grouped within the $E_{ads/abs}$ and h space. As shown in the bottom panel of Figure 1, the existence of three different clusters is best detected once the elbow method is applied, see Figure S4 of Section S4 of the SI. Data points in the first cluster, C1, have in common a rather weak interaction with the TM surfaces compared to the graphite E_{coh}, with the group centre or centroid —marked by a fictionalcrossed point in Figure 1— located at an E_{abs} of 424 kJ·mol^{-1}, and with a h of -52 pm below the surface level. For these systems, C atoms would thermodynamically tend to merge into graphitic C aggregates on the surface, eventually poisoning the catalyst surface by site-blocking. However, at low C concentrations, there could be C adatoms or interstitial C atoms, affecting the electronic structure of the surrounding metal atoms, particularly, when their mobility and eventual aggregation would be hindered, see below.

The second cluster, C2, is, on the contrary, characterized by a very strong interaction between C and the TM surfaces, and a general clear preference for the subsurface region, reflected in the centroid being located at an E_{abs} of 843 kJ·mol^{-1}, and h of -117 pm. This implies that these TMs and surfaces are suitable for their carburization, with the sole exception of Ni and Pd (111) surfaces, which feature an E_{abs} smaller than E_{coh}, and so, similarly to C1 members, would imply an eventual formation of surface aggregated carbonaceous structures, particularly when kinetically allowed, in perfect agreement with the reported graphene growth by segregation reported on both surfaces.[40,41]

Finally, the third cluster, C3, groups most of the data points, where C atoms display an interaction with TM surfaces of similar strength to the cohesive energy of graphite, plus and general clear preference for staying at the surface. This is reflected with the group center located at E_{ads} of 753 kJ·mol^{-1}, very close to the E_{coh} of graphite of 757 kJ·mol^{-1}, and a height as well of 53 pm. It is in such situations where the subtle imbalance of interaction strengths and kinetic movement inhibition may finally determine whether such C isolated moieties exist as such at low C concentrations, or they eventually merge forming graphitic layers on the catalyst surface, and such imbalance could be potentially affected by the DFT uncertainty of ca. 20 kJ·mol^{-1}. Even if this is a general trend, there are situations where C atoms have a significant affinity for TM surfaces, but now generally preferring to stay on the surface, thus occupying active sites that would be no longer be available for any other reactants, acting as poisons by site-blocking, exemplified by the subset of points with an E_{ads} larger than 850 kJ·mol^{-1}. The exposed KM clustering allows defining thus certain general behaviors, but is not exempt of singularities; apart from the above, a couple of outliers can be caught from a visual inspection, for instance, Cu (001) and Cu (011) are points that could belong to Cluster 1, although the inertia calculation assigns them to Cluster 3.

Aside from the previous analysis, we inspected suitable reaction conditions at which such C moieties can be present on the inspected TM surfaces. To this end, thermodynamic phase diagrams have been built for the different studied TM surfaces, considering pressure and temperature working conditions that would turn pristine TM surfaces into early C-containing surfaces, either having C on surface or subsurface. To this end, acetylene (C_2H_2) is considered as a carbon source, and the TM surface and molecular chemical potentials are equaled, see completed details on the phase diagram construction procedure found in the literature.[15,42,43] Figure 2 shows the exemplary phase diagrams corresponding to all *fcc* TMs (111) surfaces, since such TM surfaces are most relevant to Heterogeneous

Catalysis, but all the other phase diagrams for *fcc*, *hcp*, and *bcc* surfaces can be found in Figures S5-S7 of Section S5 of the SI. Focusing on the cases revealed on Figure 2, the shown lines define, for each metal, temperatures, T, and C_2H_2 partial pressures, $p_{C_2H_2}$, where C atoms adsorbed or absorbed would be thermodynamically equally stable to a pristine TM surface case, see other details in the literature.[15] Any T and $p_{C_2H_2}$ conditions above the shown curves imply a preference of having C adsorbed, C^{sur}, or absorbed, C_{sub}, while conditions below the curve imply a preferential TM pristine surface situation. In this particular case, C^{sur} moieties are expected for Pd, Ir, Rh, and Pt (111) surfaces, and C_{sub} for Ni (111) at, *e.g.*, a standard working $p_{C_2H_2}$ of 10^5 Pa, and a regular catalytic working temperature of 600 K. Only Cu, Ag, and Au display their nobility in this aspect, Ag being the most C-resisting one, known and explained by the Ag deeper *d*-band center,[44] joined to a weaker C–Ag coupling, which prevents antibonding states being above the Fermi level, eventually destabilizing the C interaction towards Ag.[45] In any case, Cu and Ag (111) surface would prefer to incorporate such C moieties, while surface C would be observed on Au (111) surfaces.

The Group XI —Cu, Ag, and Au— nobility is also shown in (001) and (011) TM surfaces, see Figure S5 of Section S5 of the SI, revealing an enhancement of the surface chemical activity for (011) and (001) surfaces, being the latter the most chemically active, to the point of Cu (001) surfaces being C-poisoned at standard conditions of pressure and working temperatures above *ca.* 750 K. The rest of *fcc* TMs behave similarly among them, featuring systematically C^{sur} situations on (001) surfaces and most of (011) surfaces, with the exception of Ag, Au, and Pd (011) surfaces, where C_{sub} would be preferred.

When addressing *hcp* TMs, see Figure S6 of Section S5 of the SI, three different behaviors are found: Late TMs with d^{10} electronic configuration, these are, Group XII Zn and Cd, behave like noble Group XI elements, thus not being C-poisoned at standard working conditions. Other late TMs, like Group VII (Re, Tc), Group VIII (Ru, Os), and Group IX Co display a chemical affinity more similar to *fcc* Pt-group TMs, while early TMs such as Group III (Sc, Y) and Group IV (Ti, Zr, and Hf) show a high affinity towards C. As can be seen in Figure S6, regardless of the exposed surface, C_{sub} situations are preferred for the very early Groups III and IV TMs (Sc, Y, Ti, Zr, and Hf), and very late XII TMs (Zn, Cd), whereas C^{sur} situations are preferred on middle and late TMs of Groups VII (Re, Tc), VIII (Ru, Os), and IX (Co). The sole exceptions are Ti and Hf $(10\bar{1}0)$ surfaces, where C^{sur} moieties are more stable. Finally, *bcc* TMs, belonging to early Groups

V (V, Nb, Ta), VI (Cr, Mo, W), and VIII (Fe) reveal a high affinity towards C, see Figure S7 of Section S5 of the SI, going for a C-moiety presence in any working conditions, and only avoiding them at ultrahigh vacuum conditions and high temperatures. Such C moieties are systematically C^{sur} for (001) and (111) surfaces, and as well for (011) surfaces, with the exception of Group V TMs (V, Nb, Ta), where C_{sub} are more stable, going for the C incorporation within the TM.

3.1.2. Descriptors of C behaviour at TMs surfaces

The above trends seem to point out that the chemical activity is somewhat influenced by the position along the *d* series of the Periodic Table. Thus, an important aspect resides in the search for descriptors of the ad/absorption energies accounting for the observed trends. Here, different descriptors proposed in the literature, either energetic or electronic, are evaluated so as to linearly correlate ad/absorption energies with them. In particular, surface energy, γ,[46] work function, ϕ,[47] *d*-band center, ε_d,[48] corrected *d*-band center, ε_d^W,[49] and the highest Hilbert transform *d*-band peak, ε_u,[50] are considered —details on their correct calculation can be found in the literature.[29,44] Briefly, the description of their independent evaluations —see Figures S8-S12 in Section S6 of the SI for all the analyses details— reveals that the linear adjustment adequacy decreases in the order $\varepsilon_d > \varepsilon_d^W > \varepsilon_u > \gamma > \phi$. In the case of *d*-band center based descriptors, it is clear that any attempt of improvement on ε_d is detrimental; still, the expected trend is captured, in the sense that the higher in energy is ε_d, ε_d^W, or ε_u, the stronger the C attachment is. Likewise, the larger the surface energy, γ, the stronger the attachment energy of C; although the correlation on this energy-based descriptor is poorer when compared to those based on the electronic structure. Finally, the work function, ϕ, is a very bad descriptor; not only because of the very small regression coefficient, R, of 0.17, but also because one would expect that the larger the work function, the stronger the bonding, as a result of a TM→C charge transfer, observed in late TMs, and expected for earlier and more reducing TMs. However, the observed trend in Figure S12 of Section S6 of the SI is just the opposite.

Given the above analysis, ε_d could be regarded as the most successful descriptor, at least, when describing the exhibited thermodynamic data. The most stable $E_{ads/abs}$ *vs.* ε_d results are shown in Figure 3 grouping results as per the different featured crystallographic structures. Notice that an evaluation with one regression line for each type of surface termination, shown in Figure S13 of Section S6 of the SI, reveals that the trends for the three surfaces of each crystallographic have similar slopes, and such slopes are very

different for the different examined crystal structures. These results reinforce the consideration of crystal packing as a determining aspect concerning the C interaction with TM surfaces. Focusing on Figure 3, however, one readily notices that linear trends can only be valid for *hcp* and *fcc* structures, which present regression coefficients, *R*, of 0.94 and 0.90 respectively. On the other hand, the *bcc* TMs trend line is not representative, presenting a poor regression coefficient of solely 0.09; showing that there is no correlation in these cases. This puts the accent in that the ε_d descriptor, typically tested on coinage and Pt-group TMs,[48] all featuring to *fcc* and *hcp* close packed situations, does not account for a possible packing effect that can translate in changes on the TM surface chemical activity.

Still, the trends based on energetic and electronic structure descriptors, *e.g.* on ε_d and γ, pave the way to inspect whether there exists combined effects of them, tackled here by evaluating them through multivariable regressions. As shown in Figure 4, combining ε_d with γ leads to a better correlation than using them alone. Furthermore, one could make combinations of a larger degree, *e.g.* having $\gamma \cdot \varepsilon_d$ terms or even squared values for each descriptor, which would be second degree combinations, or even third degree combinations; including, *e.g.*, $\gamma^2 \cdot \varepsilon_d$ or ε_d^3 terms. By considering these, one observes a gradual improvement of the multivariable adjustment, as observed by an increase of the *R* value up to 0.90, and the approach of the adjustment to the ideal one. These observations strongly suggest that, when looking for adsorption and/or absorption energy descriptors, one should not restrain to finding the perfect one, as it may well not exist, as the interaction is simultaneously influenced by different surface properties. Thus, one should look for combinations of descriptors, each of them bringing a different piece of information of the metal surface one works with.

The above analysis indeed laid the foundations for a proper and deeper multi-variable analysis, feasible by applying different AI machine learning (ML) regression methodologies. To this end, the E_{ads} or E_{abs} values of most stable positions are expressed as a function of the above-introduced descriptors (ε_d, ε_d^W, ε_u, γ, and ϕ), but regarding as well the crystallographic structure (CS) of the TM case, given the different behaviors of *fcc*, *hcp*, and *bcc* TMs, and including as well the surface coordination number (CN), retrieved from the literature,[51] so as to differentiate the different studied surfaces for the same metal, plus the bulk shortest interatomic distance, δ,[31,32] as a geometry measure distinct for every TM, even when having the same CS and CN. Aside, the number of TM *d* electrons, n_d, are accounted, as they represent a natural trend across the *d* series. Finally, to represent the

different adsorption or absorption sites, the number of TM atoms neighboring the C atom are defined, CN_{site}, allowing for site specificity. This accounts for a total of ten descriptors related to the TM surface. Note that, since the same adsorbing or absorbing moiety is regarded always, this is, C atoms, no descriptors from the ad/absorbed species are necessary.[52]

Within this set of descriptors and features, we evaluated three different ML algorithms, including MLR, DTR, and RFR, see details of them in Section S7 of the SI, and explicative images of the DTR procedure in Figures S14 and S15 in Section S7 of the SI. A first analysis, using all the aforementioned features and default parameters of the algorithms —*e.g.* 100 trees in RTR, and a maximal number of allowed features to be used in each tree to be equal to the total number of considered features— provided by the *sklearn* python library was performed using a shuffle split cross validation (CV) employing 20 splits for the set of the 81 surfaces with C in its most stable position. For the CV, for each size of the training set, a number of random data points are taken which represent 80% of the data set. The remaining 20% are random points also taken, but to be used for the test set.

The analysis results, in terms of mean absolute error (MAE) ± standard deviation, yielded test set values of 66.7 ± 12.5, 49.1 ± 8.4, and 43.5 ± 8.4 $kJ \cdot mol^{-1}$ for MLR, DTR, and RFR, respectively, for the largest training set size. Thus, focusing on RFR, the regression algorithm that delivered the smaller error, one can successively remove those features that were less relevant in terms of minimizing the test set MAE. To this end, we used the leave-one-out procedure, which consists in removing one feature at a time and assessing the impact on the test set MAE to decide whether it is worth including or not. In this case, three parameters emerge as most relevant from the analysis; not surprisingly, ε_d and γ, as outlined above, but also the number of metal atoms neighboring the adsorbed or absorbed C, *i.e.* the site coordination, CN_{site}, which brings site-specificity to the analysis.

Once the candidate descriptors are shortlisted according to the leave-one-out approach, the RFR algorithm hyperparameters were tuned by performing CV evaluations on different combinations of parameters and selecting those that minimized the MAE, exemplified in Figure 5 by the learning curve of RFR, displaying the training and test errors when increasing the number of samples in the training set. Results showed that, by using ε_d, γ, and CN_{site} only as input features, a RFR with 50 decision trees considering two features for each split already provided a very good training set MAE of 13.7 ± 0.9 $kJ \cdot mol^{-1}$, although

the test set MAE accuracy is of 39.1 ± 8.9 kJ·mol^{-1}. Notice, still that curves are not converged, and lower MAEs could expected by widening and diversifying the number of cases, usign, *e.g.* vicinal surfaces or sites at nanoparticle models, or even when differentiating adsorption from absorption situations.

Notice that the obtained final accuracy for the test set is still far from using such a ML model to carry out forecasts in a quantitative fashion, where, desirably, one would require MAEs below the DFT accuracy, estimated to be *ca.* 20 kJ·mol^{-1}. However, still, it can be useful for a rapid evaluation and qualitative assessment. Further than that, the most interesting factor is that, from the RFR algorithm, one can seize the importance of each key feature. In this case, the importance factors are 0.6, 0.3, and 0.1 for ε_d, γ, and CN_{site}, respectively, quite in line to the above discussion referring to ε_d as the main descriptor, but weighting the importance as well of γ, as already above detected in the descriptor analysis, where combinations of them were found to improve the regression quality, see Figure 4. Still, their definition is somewhat modulated by the site coordination, a feature not identified in the previous analysis.

After having isolated the main physicochemical descriptors through the RFR method, one could well carry out a KM analysis as done in Figure 1, but now identifying similarities in the descriptor space instead of on the target $E_{ads/abs}$ and h properties. This three-dimensional grouping is shown in Figure 6 (top panel). The components of these three clusters have been projected into the $E_{ads/abs}/h$ space in Figure 6 (bottom panel). Surprisingly, one observes that clusters in $E_{ads/abs}/h$ space shown in Figure 1 mostly coincide with gained clusters in descriptors space, with only few exeptions. Pd (111) surface of Cluster 1, appearing in Cluster 2; the Pd (011) surface of Cluster 2, appearing in Cluster 3, and the Au (111) surface of Cluster 3, appearing in Cluster 1. In any case, the resulting grouping underscores that groups of systems with similar $E_{ads/abs}$ and h also exhibit similar descriptor values. From the values in Table 1, one notes that, even when accounting for the standard deviation, the feature average values mostly do not overlap with each other for the different clusters, indicating that their representation is mostly unique, which reinforces that such descriptor values are key in defining the features groups shown in Figure 1.

3.2. Influence of Kinetics

3.2.1. The energy barriers landscape

The above analysis dealt only with the energetic stability and, therefore, to reach a holistic overview would require analyzing the C moieties diffusion kinetics on all the studied TMs, as, *e.g.*, the subsurface presence may be kinetically hindered when the sinking energy barrier, E_{sink}, would be too large, and the same applies to the possible C emergence for E_{emer}, hindering an eventual surface coke formation from subsurface C atoms. Aside, coke formation could be as well hindered by surface diffusion, governed by the E_{sur} kinetic energy barrier, E_b. Finally, we evaluate here the possible diffusion at the subsurface level, defined by E_{sub}, and questioning whether lateral diffusion preferentially happens on the surface. Notice that such barrier information has been found valuable in a catalytic context, *e.g.*, serve to obtain the mean lifetimes of such species, and so, in which time frame they can affect the surface on-going catalyzed process, as recently demonstrated on *fcc* (111) surfaces by kinetic Monte Carlo simulations.[17]

The four different types of diffusion barriers, illustrated in Figure S3 of Section S1 of the SI, have been obtained by CI-NEB algorithm and properly characterized by vibrational analysis. For each TM type of surface, different diffusive paths have been investigated, including non-trivial ones for certain complicated diffusions. The explored paths are listed in Tables S7-S9 of Section S8 of the SI. For each case, the lowest E_b values have been collected, accounting for a total of 324 diffusion energy barriers, summarized along their diffusion path in Table S10 of Section S8 of the SI. Notice that mean values, including standard deviation, reveal, as expected, E_{sur} diffusion energy barriers of 86.7 ± 55.8 kJ·mol^{-1}, being slightly lower than E_{sub}, of 94.5 ± 63.1 kJ·mol^{-1}. Still, the difference is not excessively large, and already at this stage one could envisage that surface and subsurface diffusions are similarly possible. This striking finding can be easily explained, as, on one hand, surface diffusion TSs get stabilized thanks to a more freedom of flexibility of surface metal atoms; however, subsurface diffusion TSs get stabilized thanks to a higher metal coordination, see Figures S16-S18 of Section S8 of the SI. Other than that, sinking energy barriers, E_{sink}, have a sensibly larger value of 117.1 ± 79.0 kJ·mol^{-1}, while emerging energy barriers, E_{emer}, are noticeably smaller, of the order of 57.9 ± 56.3 kJ·mol^{-1}, succinctly implying that is more difficult for C adatoms to dissolve in the metal matrix than to segregate towards the surface.

Still, the large standard deviation of the points expresses a great variety of situations. For instance, the largest E_{sur} of 263.2 kJ·mol^{-1} corresponds to Ta (001) surface, where such C atoms would be rather immobile, at variance with Cu (111), where, with an E_{sur} of 8.4 kJ·mol^{-1}, C atoms would be rather mobile; a point that favors the observed graphene

synthesis by deposition on such surfaces.[18,53] Even if subsurface diffusion is less favored, quite inhibited, *e.g.* on W (111), with an E_{sub} of 303.7 kJ·mol^{-1}, it is rather easy on Zn (10$\bar{1}$0) surface, with an E_{sub} of 2.4 kJ·mol^{-1} only. Similarly, one can think that C sinking into the subsurface region is quite difficult, and it is indeed on W (001), with an E_{sink} of 326.6 kJ·mol^{-1}, while C sinking through the Zn (11$\bar{2}$0) surface is essentially barrierless, with an E_{sink} of 0.8 kJ·mol^{-1}. Finally, C emergence to the surface is rather easy on Co (10$\bar{1}$0) surface, with an E_{emer} of 0.8 kJ·mol^{-1}, while it can be quite difficult on Ta (011), where C segregation would be rather impeded with an E_{emer} of 229.0 kJ·mol^{-1}.

This large set of data, which can be quite different, can also be gathered and analysed in a four-quadrant plot, in a similar fashion as done for the thermodynamics in Figure 1. To do so, Figure 7 shows the $\log_{10}(E_{sub}/E_{sur})$ *vs.* $\log_{10}(E_{sink}/E_{emer})$, having thus 81 points corresponding to the same amount of explored TM surfaces. Thus, zero value for $\log_{10}(E_{sub}/E_{sur})$ implies that E_{sub} = E_{sur}, and so, surface and subsurface diffusion are kinetically equally feasible. Likewise, zero value for $\log_{10}(E_{sink}/E_{emer})$ implies that surface C sinking diffusion is kinetically as likely as the subsurface C emerging. These two zero values delimit the four quadrants in Figure 7. Values larger than zero for $\log_{10}(E_{sub}/E_{sur})$ implies that E_{sub} values are higher than E_{sur}, and hence, for these cases, C atoms would diffuse more easily along the surface than through the slab subsurface region, points located on the right side of the zero limit. The opposite behavior is expected for $\log_{10}(E_{sub}/E_{sur})$ smaller than zero, located on the left side. When it concerns $\log_{10}(E_{sink}/E_{emer})$ values, data above the zero limit imply that emerging is preferred over sinking, *i.e.* a preferential location on the surface. On the contrary, values below zero imply a preference towards C penetrating the subsurface region.

The top panel of Figure 7 features data colored depending on the crystallographic structure of the TM, revealing that *fcc* and *bcc* TMs data points behave similarly, being dispersed on the top half of the plot, implying that for such metals E_{sink} is often larger than E_{emer}, and going for a kinetically allowed surface presence of C adatoms. Still, few cases are the early bottom part of the plot. These belong to *fcc* (111) Ni, Pd, Pt, Cu, Ag, and Au (111) surface models, where the above situation reverses, implying a more difficult segregation towards the surface of dissolved C atoms in the bulk metal matrix. Notice that the (111) surface termination is the most stable and abundant one on such TM metals,[29,51] and so such subsurface C effect should not be discarded on any on-going surface catalyzed process. Aside, neither *fcc* nor *bcc* TMs show a clear preference between E_{sur} and E_{sub},

with every metal being somehow unique in this regard. Finally, *hcp* TMs show sizable dispersion, but are mainly found in the bottom right part of the quadrants plot. This quadrant gathers situations with an E_{sink} barrier lower than E_{emer}, and E_{sub} being larger than E_{sur}. Hence, one observes that most of *hcp* TMs feature kinetically favorable penetration of isolated C atoms towards the subsurface region, but with restricted diffusion along the subsurface, as the diffusion across the surface is preferred.

Furthermore, the bottom panel shows the KM analysis again using three clusters as derived from Figure 1. There, it is clear that Cluster 1 fits values of the right-upper quadrant, implying E_{emer} barriers smaller than E_{sink}, and E_{sur} smaller than E_{sub}. Altogether, these features imply a certain preference for C atoms to be located on the surface, and diffusing over it, as shown by the cluster centre, where E_{sub} is two times larger than E_{sur}, and E_{sink} 39.8 times larger than E_{emer}. Beyond that, Cluster 3 also nicely fits one quadrant, but now implying an easier C emergence compared to the sinking, but, more importantly, featuring E_{sub} values sensibly smaller than E_{sur} ones, as shown by the cluster centre, located at a point where E_{sur} values are five times larger than E_{sub}, and E_{sink} 12.8 times larger than E_{emer}. This quadrant points out certain TM surfaces where C diffuses more easily going through the subsurface region, probably due to the aforementioned stabilization of the diffusion TS states by a higher coordination. In this sense, C moieties would move around in the subsurface region, and only emerge from time to time, as if they were the targets of a whack-a-mole game. Finally, Cluster 2 contains the most numerous group, displaying, in general, well balanced E_{sub}/E_{sur} and E_{sink}/E_{emer} ratios near unity, as shown by the center, being the ratios 1.4 and 0.9, respectively. Only a certain preference is found for a surface diffusion over the subsurface one, but, aside from that, the general behavior would be that all diffusion processes should be regarded feasible, implying a somewhat free mobility on the surface, through the subsurface region, and with exchange of C atoms in between surface and subsurface sites.

3.2.2. Descriptors controlling the energy barriers

Similar to the analysis carried out for the adsorption and absorption energies, the attention is now placed into finding descriptors of the diffusion energy barriers. The ones used for the thermodynamic minima, these are ε_d, ε_d^W, ε_u, γ, and ϕ, are listed in Figures S19-S23 of Section S9 of the SI. There, even when using different regressions for each barrier type, the obtained results were unsatisfactory, reaching, at most, a regression coefficient R of 0.69 for E_{sink} *vs.* γ. Still, the limited applicability of such descriptors can be though as

natural as such descriptors were developed and applied to seize the interaction strength, not the heights of energy barriers. Still, trends are regularly observed, *i.e.* the larger the ε_d, ε_d^W, or ε_u the larger the E_b, and so it applies for γ, and ϕ, although for these inverse trends are observed for E_{emer} and E_{sub}, respectively.

In any case, the previous discussion underscores the point of similar trends achieved for adsorption or absorption energies, and diffusion energy barriers. Within this context, one has to remark the work of Nilekar and colleagues[54], who showed in their seminal work that the diffusion energy barrier depends on the adsorption strength of the adsorbed moiety, being indeed basically a twelve per cent of the latter, with an R of 0.85. Figure 8 shows this trend for the four different diffusion processes here studied, revealing that, indeed, the stronger the adsorption or absorption energy, the larger the diffusion energy barrier, and in all the cases with similar slopes of *ca.* 0.16. However, the regression coefficients, R, are still quite modest, being at most 0.45 for E_{sur} energy barriers.

Another extended linear correlation used for energy barriers are Brønsted–Evans–Polanyi (BEP) relations, linearly expressing a reaction step energy barrier with the step energy variation, ΔE.[55-57] In this work we do not deal with reaction steps, as we are focusing on diffusion processes, but the same principle applies. To this end, the BEP relations were analyzed, yet only for C sinking and emerging processes, as surface and subsurface diffusions feature a ΔE of zero. Figure 9 shows their BEP analysis with a fairly good correlation for E_{sink}, with a regression coefficient R of 0.83; however, the regression for E_{emer} shows more dispersion, with a lower R of 0.63. Notice that for sinking process, the linear regression is close to the limit of a very late TS, where $E_b = \Delta E$, signaled in Figure 9 with a dashed black line. Another regression constrain is an earliest TS situation, where E_b equals zero regardless of the value of ΔE. Still, there is also a number of cases located in between, so no evident trend can be claimed. At variance with the description of $E_{ads/abs}$ with respect to ε_d, the BEP correlation coefficients when splitting the data into different crystallographic groups does not substantially improve the outcome, indicating this time that the crystal structure is not a determining factor in the BEP correlations for neither E_{sink} nor E_{emer}; see Figures S24 and S25 in Section S9 of the SI.

Mimicking the above analysis of E_{ads} or E_{abs} dependence on ε_d and γ, and given the best correlations observed for E_{sink} as described by γ and ΔE, we carried out a multivariable regression involving combinations of ΔE and γ descriptors up to a second order degree — since third degree yielded no improvement— shown in Figure 10. There, as happened in

the thermodynamic evaluation in Figure 4, one observes that *i)* the combination of both factors is beneficial, with an improved R of 0.86, while *ii)* the incorporation of descriptors higher orders translates into a mild improvement. As happened with the thermodynamic analysis, this underscores that the description of kinetic processes should be tackled considering a combination of different descriptors, instead of looking for a single, determining one, as appears that different aspects influence the kinetic energy barriers.

Thus, following the same procedure as for the thermodynamic analysis, we applied here different machine learning (ML) regression algorithms to have tools to forecast diffusion energy barriers and locate key descriptors governing such processes. Note that now the data set is four times larger —324 E_b values— than the 81 cases of E_{ads} or E_{abs}. The features and descriptors used are the same of the thermodynamic analysis, but now adding most stable position $E_{ads/abs}$ and ΔE. Moreover, the barrier type, BT —E_{sur}, E_{sub}, E_{sink}, or E_{emer}—, and the CN of the initial and final sites, named CN_{IS} and CN_{FS}, respectively, are also included. The target goal here is to predict E_b, and to this end MLR, DTR, and RFR have been used. A first analysis was carried out following the procedure and hyperparameters used for the thermodynamic data, yielding MAEs of 34.3 ± 2.4, 42.5 ± 4.7, and 33.1 ± 2.8 kJ·mol^{-1} for MLR, DTR, and RFR, respectively. Again, RFR is posed as the best performing ML algorithm.

By refining the number of descriptors by means of the leave-one-out procedure, the less relevant features were removed, and left only those that had a significant enough impact in terms of error minimization, which are ΔE, $E_{ads/abs}$, ϕ, CN_{FS}, and CN_{IS}, with weights of 0.43, 0.27, 0.15, 0.09, and 0.06, respectively. Notice how ΔE is the most important feature, as expected from the BEP analysis, and particularly true for E_{sink} and E_{emer} barriers, although is not as predominant as ε_d was for $E_{ads/abs}$. The next in the list is indeed the most stable position $E_{ads/abs}$, in line with the above discussion where it was found to affect the barrier heights. Notice that the two most important descriptors are thermodynamic parameters, which can be used to estimate kinetic energy barriers. Other factors affecting the energy barriers are, unexpectedly, ϕ, even if by scratch does not show good linear correlations with the kinetic data, and CN_{IS} and CN_{FS}, which shows that the coordination number of the sites involved in the diffusion play minor, yet still relevant role in the height of the barriers.

As a last step, we selected the best performing hyperparameters of the RFR algorithm by performing CV evaluations on different combinations of parameters and choosing those

that lead to minimal MAE, see the learning curve in Figure 11. Results showed that, by using only ΔE, $E_{ads/abs}$, ϕ, CN_{FS}, and CN_{IS} as the input features, a RFR with 30 decision trees which considers three features for each split, see Figure 11, provided the best accuracy, with a MAE of 32.6 ± 2.3 kJ·mol^{-1} for the test set, while the training set offers a MAE of solely 13.1 ± 0.4 kJ·mol^{-1}. The dimensionality reduction in Figure 7 avoids using a KM clustering in the feature space, also because different features are simultaneously affecting the different E_b values. Other than that, the analysis done in Figure 11 can actually be carried out differentiating the four different diffusive processes. This has been done and discussed in Figures S26-S29 and Table S11 of Section S10 of the SI. Briefly, RFR persisted as the best performing algorithm, with MAEs very similar to that obtained when considering all E_b altogether, with best performance found for E_{sur} barriers, with a test set MAE of 25.2 ± 6.6 kJ·mol^{-1}, and the less accurate case being E_{sub}, with a test set MAE of 38.2 ± 7.1 kJ·mol^{-1}. Compared to the thermodynamic data, the RFR offers more accuracy for diffusion energy barriers, which get closer to the typical DFT accuracy of 10-20 kJ·mol^{-1}, although one is still far from achieving an accuracy that would support quantitative estimations. However, the main descriptors affecting the C diffusions have been unfolded, and estimates can be argued upon them, allowing for a rapid qualitative assessment and sieving process.

4. Conclusions

In this work, by performing high-throughput periodic DFT calculations on proper slab models of 81 TM surfaces using the PBE *xc* functional, we provide compelling thermodynamic *plus* kinetic information on the interplay of C atom on and in TM surfaces, providing detailed information of adsorption and absorption minima sites and energies, and diffusive energy barriers along surface and subsurface regions, as well as in between both regions, summarized in Table S12 of Section S11 of the SI. The provided thermodynamic and kinetic results are in line with experimental observations regarding carbide formation in *bcc* and early *hcp* TMs, also revealing the possibility of a surface reconstruction, being the key step of such process in some of the cases. Moreover, results are also in line with the well-known tendency of C to form graphene layers on top of *fcc* TMs surfaces, effectively poisoning them for heterogeneous catalysis purposes. Thermodynamic phase diagrams have been built for all the explored surfaces, delimiting temperature and ethylene partial pressure working conditions at which the presence of C atoms would be favorable.

Further than that, electronic and energetic descriptors have been analyzed on ad- and absorptive minima and energy barriers, where the *d*-band center has been found to be the most successful one when correlating ad/absorption energies, in a particular reliable fashion for *hcp* and *fcc* TMs, whereas presenting significant deviations *bcc* TMs. For diffusion energy barriers, the adsorption or absorption strength of the departing minimum, and the difference in energy between minima, ΔE, in the cases of sinking and emerging diffusions. Finally, k-means clustering has been used to delimit three types of TM thermodynamic and kinetic behavior towards C atoms, while machine learning random forest regression revealed that, both for the gained thermodynamic and kinetic data, a combination of descriptors yields to a better description of minima and energy barriers, although the accuracy is so far only valid for a rapid qualitative assessment. In any case, the analysis underscores the need of looking for a few number of descriptors biasing the interaction strength and diffusion possibilities, rather than trying to get a single perfect descriptor, which may not be the wisest option as different physicochemical aspects do indeed affect such processes.

Associated Content

Supporting Information

The following Supporting Information is available free of charge on the ACS Publications website at DOI:

Models and definitions, K-means clustering details, ad/absorption energies and heights, K-means inertia evaluation, built phase diagrams, evaluation based on isolated descriptors, details on the employed machine learning regression algorithms, diffusion energy barriers and their evaluation based on single descriptors, their evaluation based on machine learning regressions, and a table delivering a summary of C adsorption/absorption preferences.

Acknowledgments

The authors are thankful to the *Ministerio de Economía y Competitividad* (MEC) for the RTI2018-095460-B-I00 and *María de Maeztu* MDM-2017-0767 grants, to the *Generalitat de Catalunya* for the 2017SGR13 grant, and to the COST Action CA18234. Authors are thankful to *Red Española de Supercomputación* (RES) for the supercomputing time

(QCM-2018-1-0005, QCM-2018-2-0008, QCM-2019-1-0017, QS-2019-3-0017, QS-2020-1-0013). The study was also supported by the project EXTREME, funded by the Bulgarian Ministry of Education and Science, D01-76/30.03.2021 through the programme "European Scientific Networks". I. Z. K. is grateful to L'Oréal-UNESCO For Women in Science National Fellowship Program, 2020. H. A. A. is grateful to European Regional Development Fund and the Operational Program "Science and Education for Smart Growth" under contract UNITe No. BG05M2OP001-1.001-0004-C01. O. P. thanks the Spanish MICIUN for a PhD grant (PRE2018-083811). A. B. thanks *Generalitat de Catalunya* for his Beatriu de Pinós grant (2018BP00190). Authors are also thankful for insightful scientific discussions with MSc. Raúl Santiago.

Notes

The authors declare no competing financial interest.

Table 1. Average values for ε_d, γ, and CN, for each of the clusters shown in Figure 6, plus their standard deviation.

	ε_d / eV	γ / J·m^{-2}	CN_{site}
C1	-0.1 ± 1.7	2.6 ± 0.7	3.0 ± 0.9
C2	-5.4 ± 2.5	0.7 ± 0.3	5.5 ± 0.8
C3	2.5 ± 1.7	1.6 ± 0.5	5.8 ± 0.4

Figure 1. Most stable $E_{ads/abs}$ situations on every TM surface with respect to h. In the top panel, values in blue correspond to *fcc* TMs, green to *bcc* TMs, and purple to *hcp* TMs. In the bottom panel, three different colors are used to mark off the three different clusters determined through a KM analysis, with centers marked crossed.

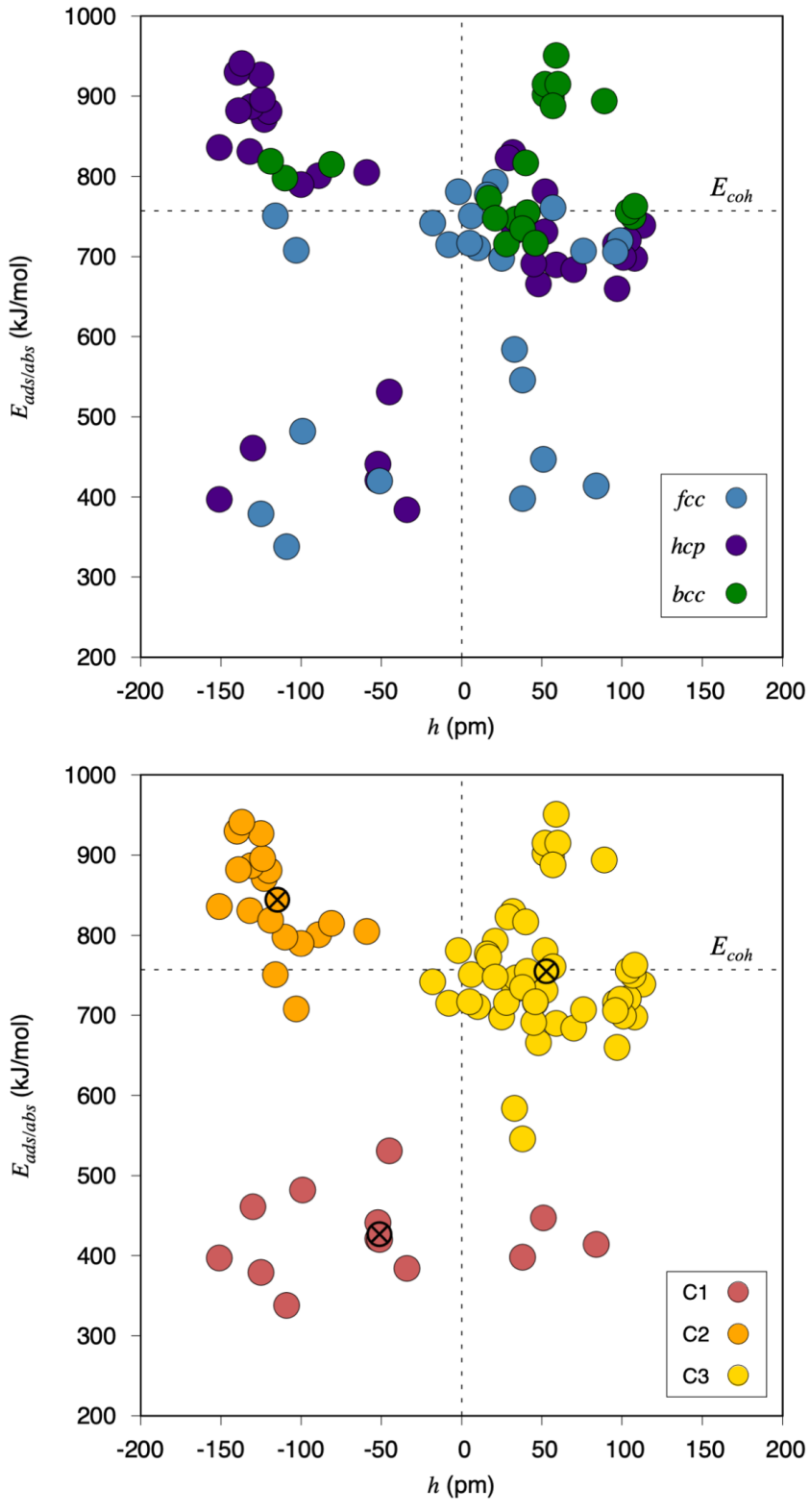


Figure 2. Phase diagrams for *fcc* TMs (111) surfaces as a function of the acetylene partial pressure, $p_{C_2H_2}$, and temperature, *T*. Diagrams are obtained for a constant partial pressure of H_2, $p_{H_2} = 10^{-7}$ Pa. Regions above or below each curve indicate conditions at which the C-containing or pristine surfaces, respectively, are thermodynamically preferred.

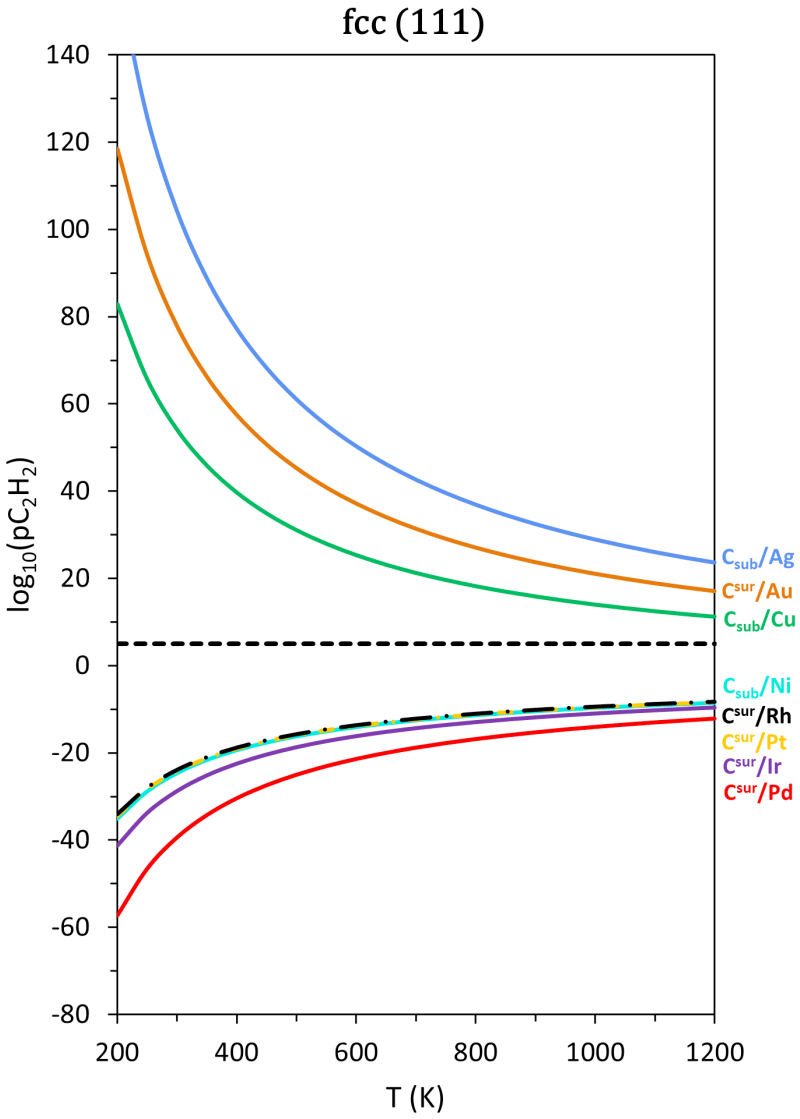


Figure 3. Most stable $E_{ads/abs}$ situations with respect to ε_d, and the corresponding linear correlations. Values in blue correspond to *fcc* TMs, green to *bcc* TMs, and purple to *hcp* TMs.

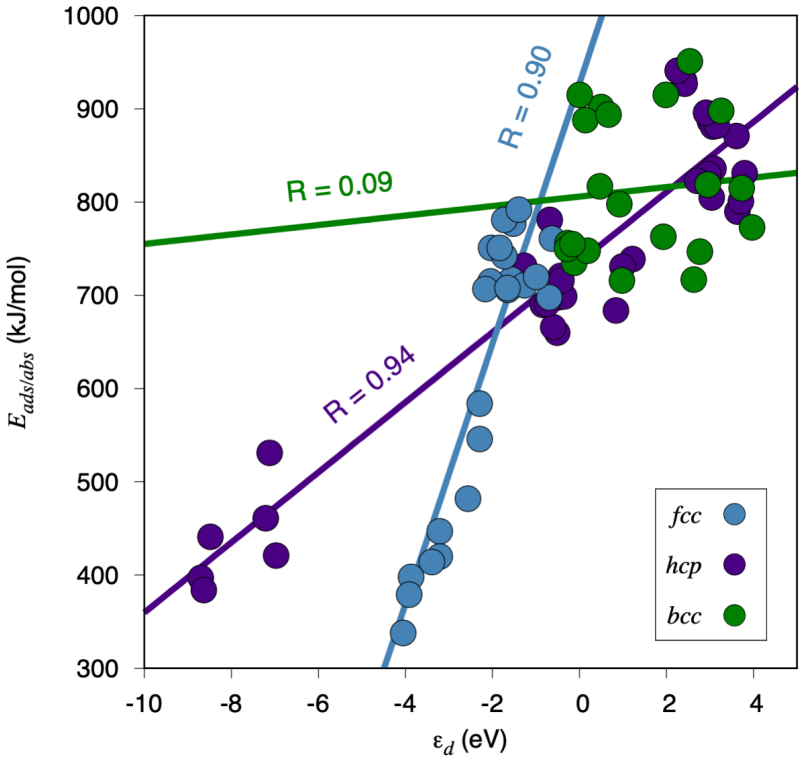


Figure 4. Most stable $E_{ads/abs}$ situations *vs.* the predicted $E_{ads/abs}$ from different linear, multivariable, and polynomial regressions involving γ, ε_d, or a combination of them. The black dashed line represents a perfect correlation.

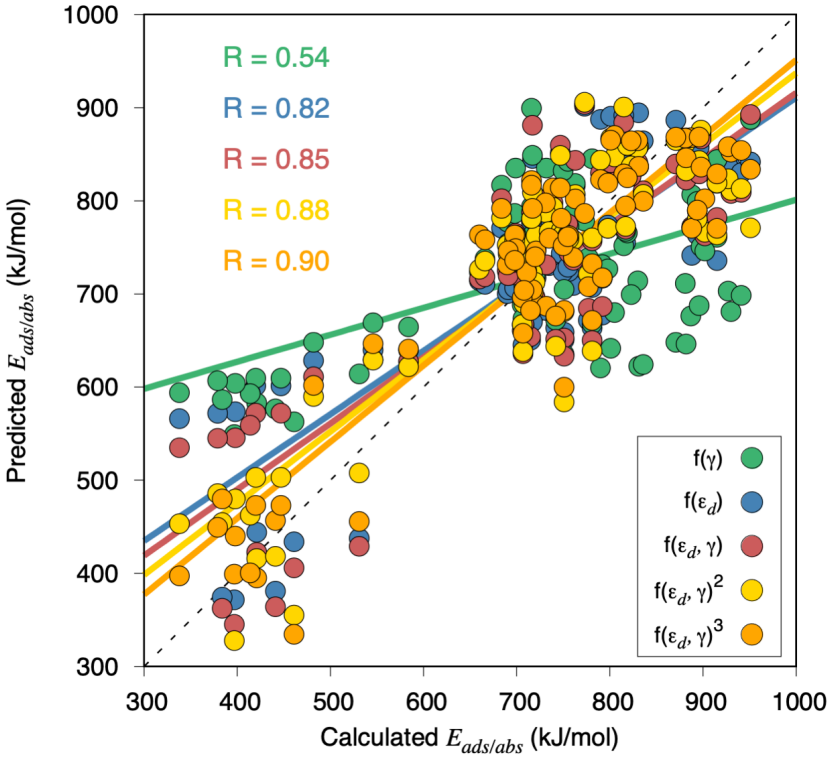


Figure 5. MAE evolution for training (blue) and the test (green) sets as a function of the number of samples contained in the training set for the prediction of $E_{ads/abs}$ using a RFR algorithm. Colored areas around the lines account for the error dispersion resulting from the cross validation using 20 runs.

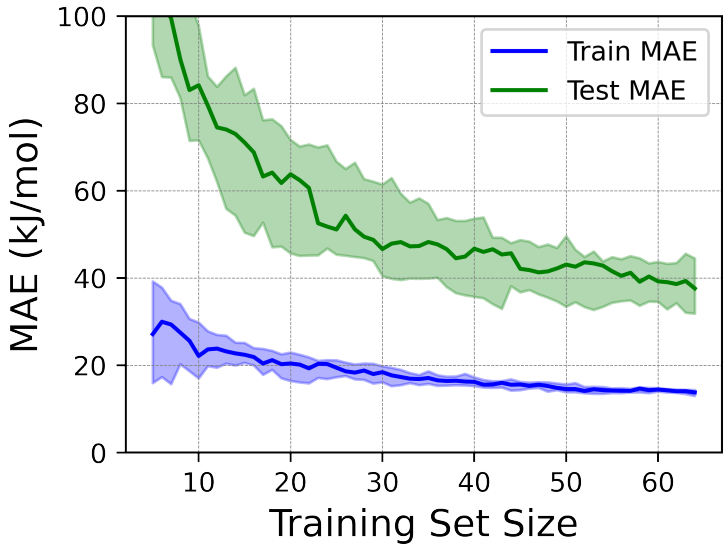


Figure 6. Top panel: the three-dimensional KM clustering of the most important features extracted from the RFR, namely ε_d, γ, and CN_{site}. Bottom panel: a quadrant plot as in Figure 1, but showing the clusters from the top panel projected in the $E_{ads/abs}/h$ space.

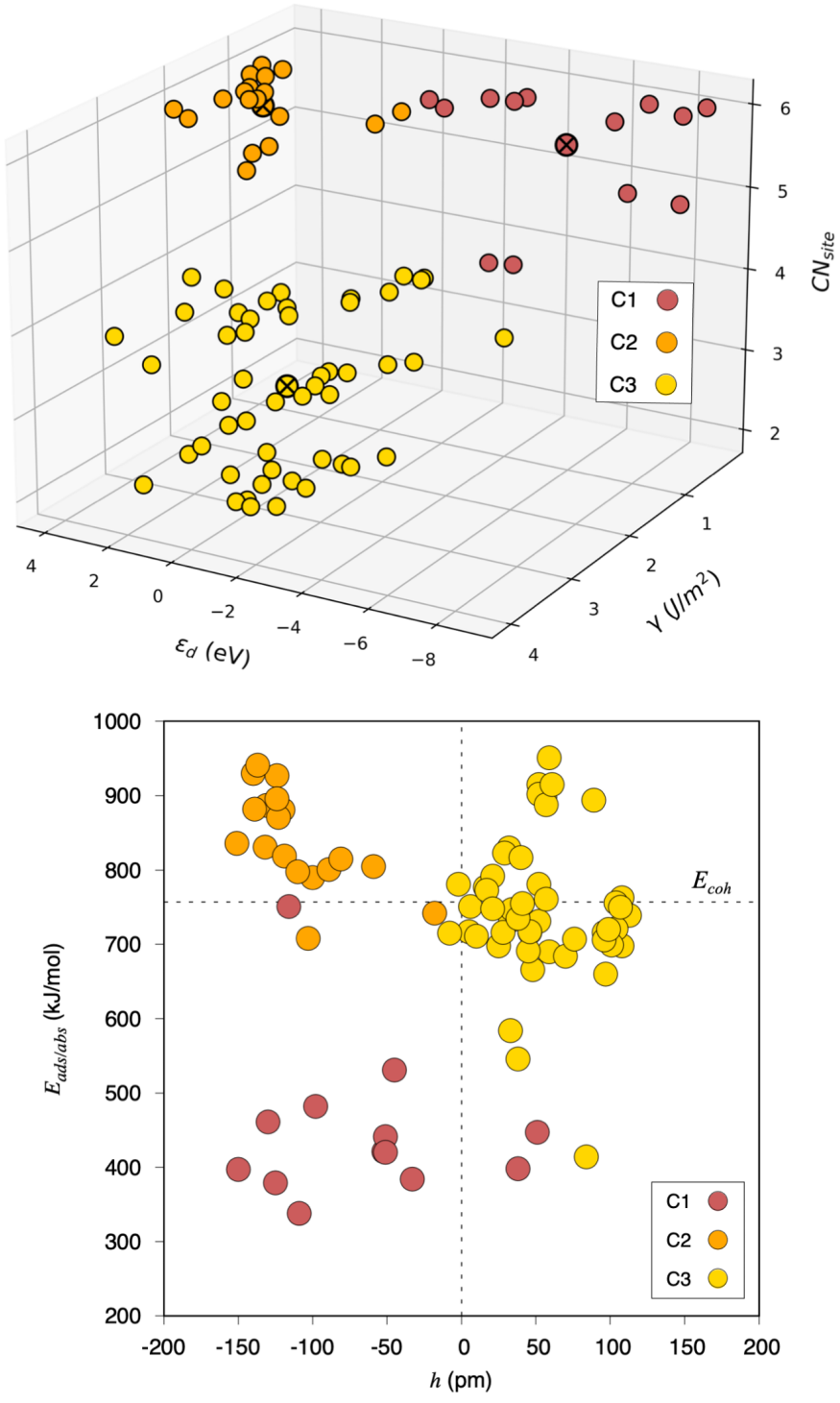


Figure 7. The base 10 logarithm of E_{sub}/E_{sur} with respect to the base 10 logarithm of E_{sink}/E_{emer}. In the top panel, values in blue correspond to *fcc* TMs, green to *bcc* TMs, and purple to *hcp* TMs. In the bottom panel, three different colors are used to mark off the three different clusters determined through a KM analysis.

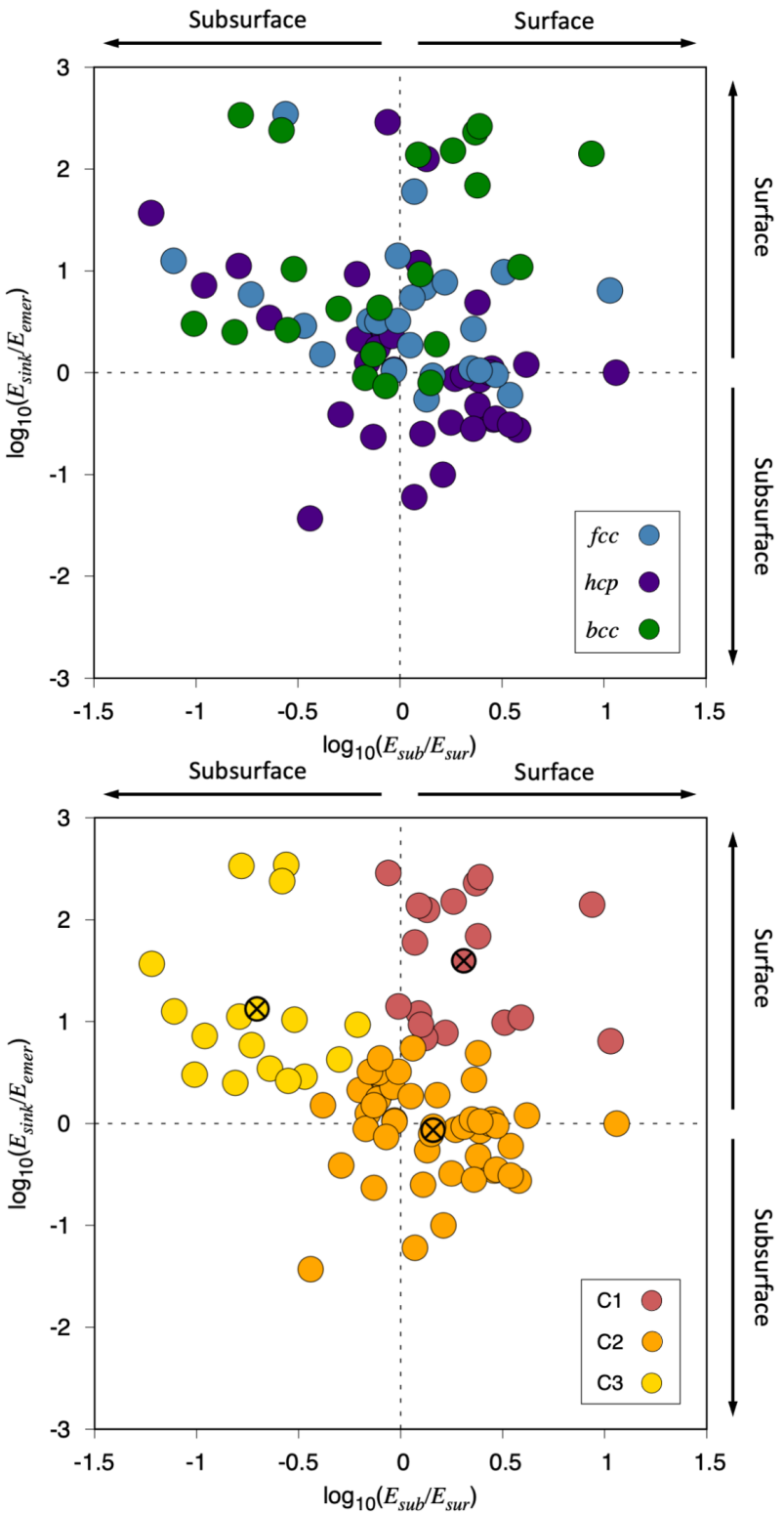


Figure 8. Linear evolutions of the different E_b diffusion energy barriers —E_{sur}, E_{sub}, E_{sink}, and E_{emer}— with respect to most stable initial position E_{ads} or E_{abs}.

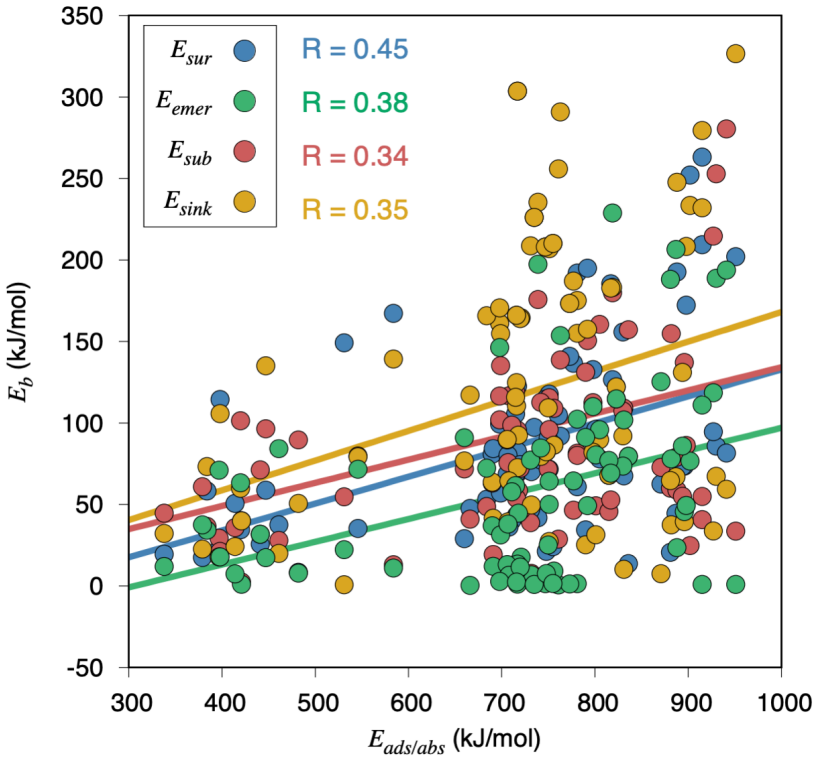

E_{sur}
E_{emer}
E_{sub}
E_{sink}
R = 0.45
R = 0.38
R = 0.34
R = 0.35
E_b (kJ/mol)
$E_{ads/abs}$ (kJ/mol)
350
300
250
200
150
100
50
0
-50
300
400
500
600
700
800
900
1000

Figure 9. BEP linear evolutions of E_b with respect to ΔE, and the corresponding linear correlations for each explored barrier type. Dashed lines define latest TS limit, where $E_b = \Delta E$, or earliest TS limit, where $E_b = 0$ regardless of ΔE.

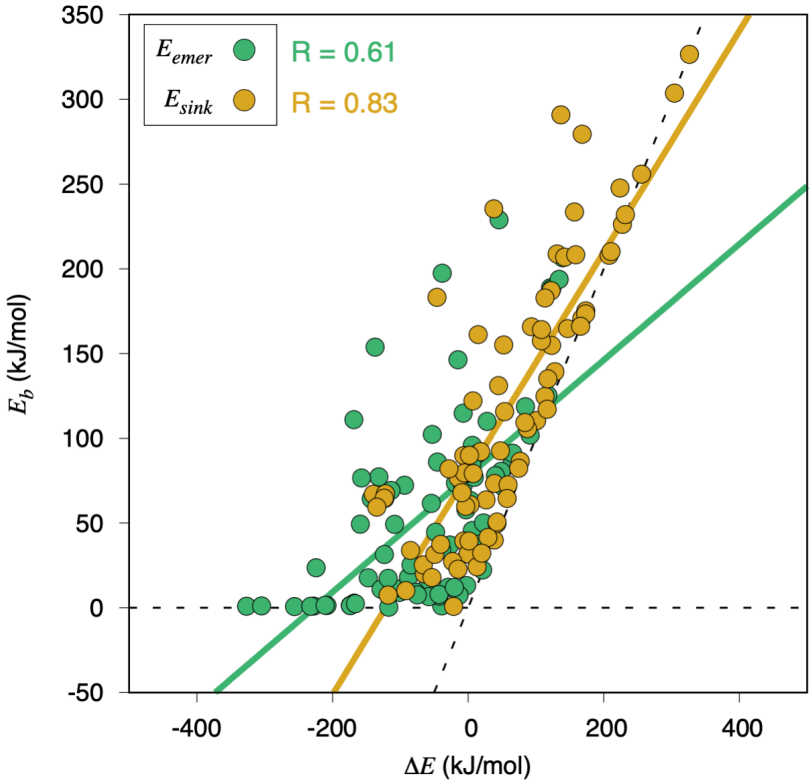
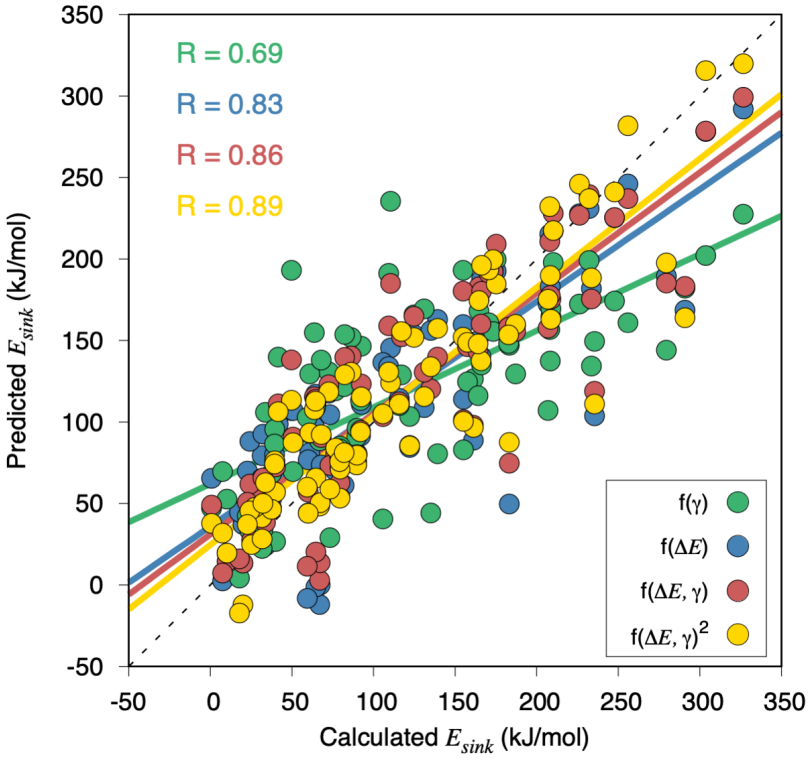
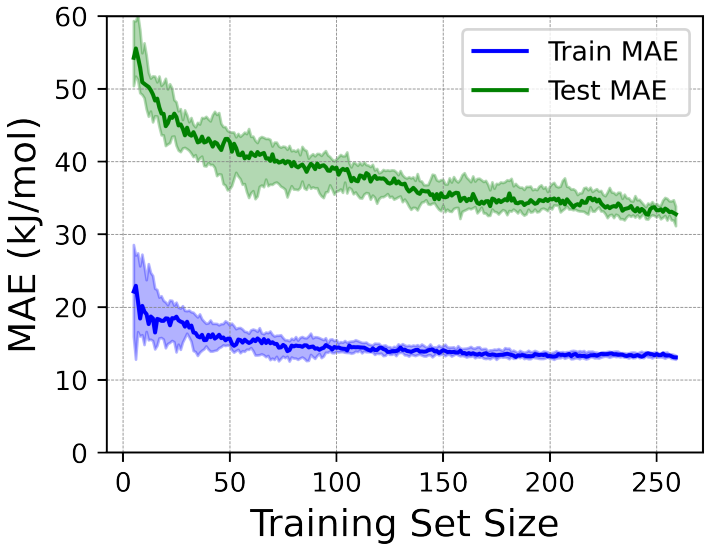


Figure 10. Calculated E_{sink} *vs.* the predicted E_{sink} from different linear, multivariable, and polynomial regressions involving γ, ΔE, or different degrees of combinations of them. The black dotted line represents the ideal correlation.



Figrue 11. MAE evolution for training (blue) and the test (green) sets as a function of the number of samples contained in the training set for the prediction of E_b using a RFR algorithm. Colored areas around the lines account for the error dispersion resulting from the cross validation using 20 runs.



References

(1) Greeley, J.; Nørskov, J. K.; Mavrikakis, M. Electronic Structure and Catalysis on Metal Surfaces. *Annu. Rev. Phys. Chem.* **2002**, *53*, 319–348.

(2) Medford, A. J.; Vojvodic, A.; Hummelshøj, J. S.; Voss, J.; Abild-Pedersen, F.; Studt, F.; Bligaard, T.; Nilsson A.; Nørskov, J. K. From the Sabatier Principle to a Predictive Theory of Transition-Metal Heterogeneous Catalysis. *J. Catal.* **2015**, *328*, 36–42.

(3) Haruta, M.; Tsubota, S.; Kobayashi, T.; Kageyama, H.; Genet, M. J.; Delmon, B. Low-Temperature Oxidation of CO over Gold Supported on TiO_2, α-Fe_2O_3, and Co_3O_4. *J. Catal.* **1993**, *144*, 175–192.

(4) Sabatier, P. *La catalyse en chimie organique*. Béranger, Paris, France, **1920**, 2nd edn.

(5) Quesne, M. G.; Roldán, A.; de Leeuw N. H.; Catlow, C. R. A. Bulk and Surface Properties of Metal Carbides: Implications for Catalysis. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6905–6916.

(6) Hwu H. H.; Chen, J. G. Surface Chemistry of Transition Metal Carbides. *Chem. Rev.* **2005**, *105*, 185-212.

(7) Levy, R. B.; Boudart, M. Platinum-Like Behavior of Tungsten Carbide in Surface Catalysis. *Science* **1973**, *181*, 547-549.

(8) Zaera, F. Shape-Controlled Nanostructures in Heterogeneous Catalysis. *ChemSusChem* **2013**, *6*, 1797–1820.

(9) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. Towards the Computational Design of Solid Catalysts. *Nat. Chem.* **2009**, *1*, 37–46.

(10) Kozlov, S. M.; Kovács, G.; Ferrando, R.; Neyman, K. M. How to Determine Accurate Chemical Ordering in Several Nanometer Large Bimetallic Crystallites from Electronic Structure Calculations. *Chem. Sci.* **2015**, *6*, 3868–3880.

(11) Davis, S. M.; Somorjai, G. A. *The Chemical Physics of Solid Heterogeneous Catalysis: Fundamental Studies of Heterogeneous Catalysis.* Elsevier, Amsterdam, Netherlands, **1982**, 1st edn..

(12) Vajda, S.; Pellin, M. J.; Greeley, J. P.; Marshall, C. L.; Curtiss, L. A.; Ballentine, G. A.; Elam, J. W.; Catillon-Mucherie, S.; Redfern, P. C.; Mehmood, F.; Zapol, P. Subnanometre Platinum Clusters as Highly Active and Selective Catalysts for the Oxidative Dehydrogenation of Propane. *Nat. Mater.* **2009**, *8*, 213–216.

(13) Teschner, D.; Borsodi, J.; Wootsch, A.; Révay, Z.; Hävecker, M.; Knop-Gericke, A.; Jackson, S. D.; Schlögl, R. The Roles of Subsurface Carbon and Hydrogen in Palladium-Catalyzed Alkyne Hydrogenation. *Science* **2008**, *320*, 86–89.

(14) Viñes, F.; Loschen, C.; Illas, F.; Neyman, K. M. Edge Sites as a Gate for Subsurface Carbon in Palladium Nanoparticles. *J. Catal.* **2009**, *266*, 59–63.

(15) Piqué, O.; Koleva, I. Z.; Viñes, F.; Aleksandrov, H. A.; Vayssilov, G. N.; Illas, F. Subsurface Carbon: A General Feature of Noble Metals. *Angew. Chem. Int. Ed.* **2019**, *58*, 1744–1748.

(16) Janthon, P.; Viñes, F.; Sirijaraensre, J.; Limtrakul, J.; Illas, F. Carbon Dissolution and Segregation in Platinum. *Catal. Sci. Technol.* **2017**, *7*, 807–816.

(17) Martínez, B.; Piqué, O.; Prats, H.; Viñes, F.; Illas, F. Towards Understanding the Role of Carbon Atoms on Transition Metal Surfaces: Implications for Catalysis. *Appl. Surf. Sci.* **2020**, *513*, 145765.

(18) Reina, A.; Thiele, S.; Jia, X.; Bhaviripudi, S.; Dresselhaus, M. S.; Schaefer, J. A.; Kong, J. Growth of Large-Area Single- and Bi-Layer Graphene by Controlled Carbon Precipitation on Polycrystalline Ni Surfaces. *Nano Res.* **2009**, *2*, 509–516.

(19) Aguiar-Hualde, J.-M.; Magnin, Y.; Amara, H.; Bichara, C. Probing the Role of Carbon Solubility in Transition Metal Catalyzing Single-Walled Carbon Nanotubes Growth. *Carbon* **2017**, *120*, 226–232.

(20) Rinaldi, A.; Tessonnier, J.-P.; Schuster, M. E.; Blume, R.; Girgsdies, F.; Zhang, Q.; Jacob, T.; Abd Hamid, S. B.; Su, D. S.; Schlögl, R. Enantioselective Organocatalytic Diels–Alder Trapping of Photochemically Generated Hydroxy-o-Quinodimethanes. *Angew. Chem. Int. Ed.* **2011**, *50*, 3313–3317.

(21) Maciejewski, M.; Baiker, A. Incorporation and Reactivity of Carbon in Palladium. *Pure Appl. Chem.* **1995**, *67*, 1879–1884.

(22) Neyman, K. M.; Schauermann, S. Hydrogen Diffusion into Palladium Nanoparticles: Pivotal Promotion by Carbon. *Angew. Chem. Int. Ed.* **2010**, *49*, 4743–4746.

(23) Liu, S. P.; Zhao, M.; Gao, W.; Jiang, Q. Mechanistic Insights into the Unique Role of Copper in CO_2 Electroreduction Reactions. *ChemSusChem* **2017**, *10*, 387–393.

(24) Sun, Y.; Cao, Y.; Wang, L.; Mu, X.; Zhao, Q.; Si, R.; Zhu, X.; Chen, S.; Zhang, B.; Chen, D.; Wan, Y. Gold Catalysts Containing Interstitial Carbon Atoms Boost Hydrogenation Activity. *Nat. Commun.* **2020**, *11*, 4600.

(25) Favaro, M.; Xiao, H.; Cheng, T.; Goddard, W. A.; Yano, J.; Crumlin, E. J. Subsurface Oxide Plays a Critical Role in CO_2 Activation by Cu(111) Surfaces to form Chemisorbed CO_2, the First Step in Reduction of CO_2. *Proc. Natl. Acad. Sci.* **2017**, *114*, 6706–6711.

(26) Chen, T.; Foo, C.; Tsang, S. C. E. Interstitial and Substitutional Light Elements in Transition Metals for Heterogeneous Catalysis. *Chem. Sci.* **2021**, *12*, 517–532.

(27) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.

(28) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(29) Vega, L.; Ruvireta, J.; Viñes, F.; Illas, F. Jacob's Ladder as Sketched by Escher: Assessing the Performance of Broadly Used Density Functionals on Transition Metal Surface Properties. *J. Chem. Theory Comput.* **2018**, *14*, 395–403.

(30) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(31) Janthon, P.; Luo, S.; Kozlov, S. M.; Viñes, F.; Limtrakul, J.; Truhlar, D. G.; Illas, F. Bulk Properties of Transition Metals: A Challenge for the Design of Universal Density Functionals. *J. Chem. Theory Comput.* **2014**, *10*, 3832–3839.

(32) Vega, L.; Viñes, F. Generalized Gradient Approximation Adjusted to Transition Metals Properties: Key Roles of Exchange and Local Spin Density. *J. Comput. Chem.* **2020**, *41*, 2598–2603.

(33) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976**, *13*, 5188–5192.

(34) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(35) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn Res.* **2011**, *12*, 2825–2830.

(36) Chaker, Z.; Chervy, P.; Boulard, Y.; Bressanelli, S.; Retailleau, P.; Paternostre, M.; Charpentier, T. Systematic Method for the Exploration, Representation, and

Classification of the Diphenylalanine Solvatomorphic Space. *J. Phys. Chem. B* **2021**, *125*, 9454–9466.

(37) Jiang, X.; Århammar, C.; Liu, P.; Zhao, J.; Ahuja, R. The R3-Carbon Allotrope: A Pathway Towards Glassy Carbon Under High Pressure. *Sci. Rep.* **2013**, *3*, 1877.

(38) Matsushima, H.; Taranovskyy, A.; Haak, C.; Gründer, Y.; Magnussen, O. M. Reconstruction of Cu(100) Electrode Surfaces during Hydrogen Evolution. *J. Am. Chem. Soc.* **2009**, *131*, 10362–10363.

(39) Tsouloucha, A.; Kerkines, I. S. K.; Mavridis, A. Theoretical Investigation of Titanium Carbide, TiC: $X^3\Sigma^+$, $a^1\Sigma^+$, $A^3\Delta$, and $b^1\Delta$ States. *J. Phys. Chem. A* **2003**, *107*, 6062–6072.

(40) Mok, H. S.; Ebnonnasir, A.; Murata, Y.; Nie, S.; McCarty, K. F.; Ciobanu, C. V.; Kodambaka, S. Kinetics of Monolayer Graphene Growth by Segregation on Pd(111). *Appl. Phys. Lett.* **2014**, *104*, 101606.

(41) Zhao, W.; Kozlov, S. M.; Höfert, O.; Gotterbarm, K.; Lorenz, M. P. A.; Viñes, F.; Papp, C.; Görling, A.; Steinrück, H.-P. Graphene on Ni(111): Coexistence of Different Surface Structures. *J. Phys. Chem. Lett.* **2011**, *2*, 759–764.

(42) Teschner, D.; Révay, Z.; Borsodi, J.; Hävecker, M.; Knop-Gericke, A.; Schlögl, R.; Milroy, D.; Jackson, S. D.; Torres, D.; Sautet, P. Understanding Palladium Hydrogenation Catalysts: When the Nature of the Reactive Molecule Controls the Nature of the Catalyst Active Phase. *Angew. Chem. Int. Ed.* **2008**, *47*, 9274–9278.

(43) Rogal, J.; Reuter, K. Educational Notes RTO-EN-AVT-142, **2007**, *2*, 1–18.

(44) Vega, L.; Martínez, B.; Viñes, F.; Illas, F. Robustness of Surface Activity Electronic Structure-Based Descriptors of Transition Metals. *Phys. Chem. Chem. Phys.* **2018**, *20*, 20548–20554.

(45) Chen, B. W. J.; Kirvassilis, D.; Bai, Y.; Mavrikakis, M. Atomic and Molecular Adsorption on Ag(111). *J. Phys. Chem. C* **2019**, *123*, 7551–7566.

(46) Zhuang, H.; Tkalych, A. J.; Carter, E. A. Surface Energy as a Descriptor of Catalytic Activity. *J. Phys. Chem. C* **2016**, *120*, 23698–23706.

(47) Michaelson, H. B. The Work Function of the Elements and Its Periodicity. *J. Appl. Phys.* **1977**, *48*, 4729–4733.

(48) Hammer, B.; Nørskov, J. K. Electronic Factors Determining the Reactivity of Metal Surfaces. *Surf. Sci.* **1995**, *343*, 211–220.

(49) Vojvodic, A.; Nørskov, J. K.; Abild-Pedersen, F. Electronic Structure Effects in Transition Metal Surface Chemistry. *Top. Catal.* **2014**, *57*, 25–32.

(50) Xin, H.; Vojvodic, A.; Voss, J.; Nørskov, J. K.; Abild-Pedersen, F. Effects of *d*-band Shape on the Surface Reactivity of Transition-Metal Alloys. *Phys. Rev. B*, **2014**, *89*, 115114.

(51) Ruvireta, J.; Vega, L.; Viñes, F. Cohesion and Coordination Effects on Transition Metal Surface Energies. *Surf. Sci.* **2017**, *664*, 45–49.

(52) Praveen, C. S.; Comas-Vives, A. Design of an Accurate Machine Learning Algorithm to Predict the Binding Energies of Several Adsorbates on Multiple Sites of Metal Surfaces. *ChemCatChem*, **2020**, *12*, 4611–4617.

(53) Gao, L.; Guest, R.; Guisinguer, N. P. Epitaxial Graphene on Cu(111). *Nano Lett.* **2010**, *10*, 3512–3516.

(54) Nilekar, A. U.; Greeley, J.; Mavrikakis, M. A. A Simple Rule of Thumb for Diffusion on Transition-Metal Surfaces. *Angew. Chem. Int. Ed.* **2006**, *45*, 7046–7049.

(55) Fajín, J. L. C.; Viñes, F.; Cordeiro, M. N. D. S.; Illas, F.; Gomes, J. R. B. Effect of the Exchange-Correlation Potential on the Transferability of Brønsted–Evans–Polanyi Relationships in Heterogeneous Catalysis. *J. Chem. Theory Comput.* **2016**, *12*, 2121–2126.

(56) Brønsted, J. N. Acid and Basic Catalysis. *Chem. Rev.* **1928**, *5*, 231−338.

(57) Evans, M. G.; Polanyi, M. Inertia and Driving Force of Chemical Reactions. *Trans. Faraday Soc.* **1938**, *34*, 11−24.

Table of Contents