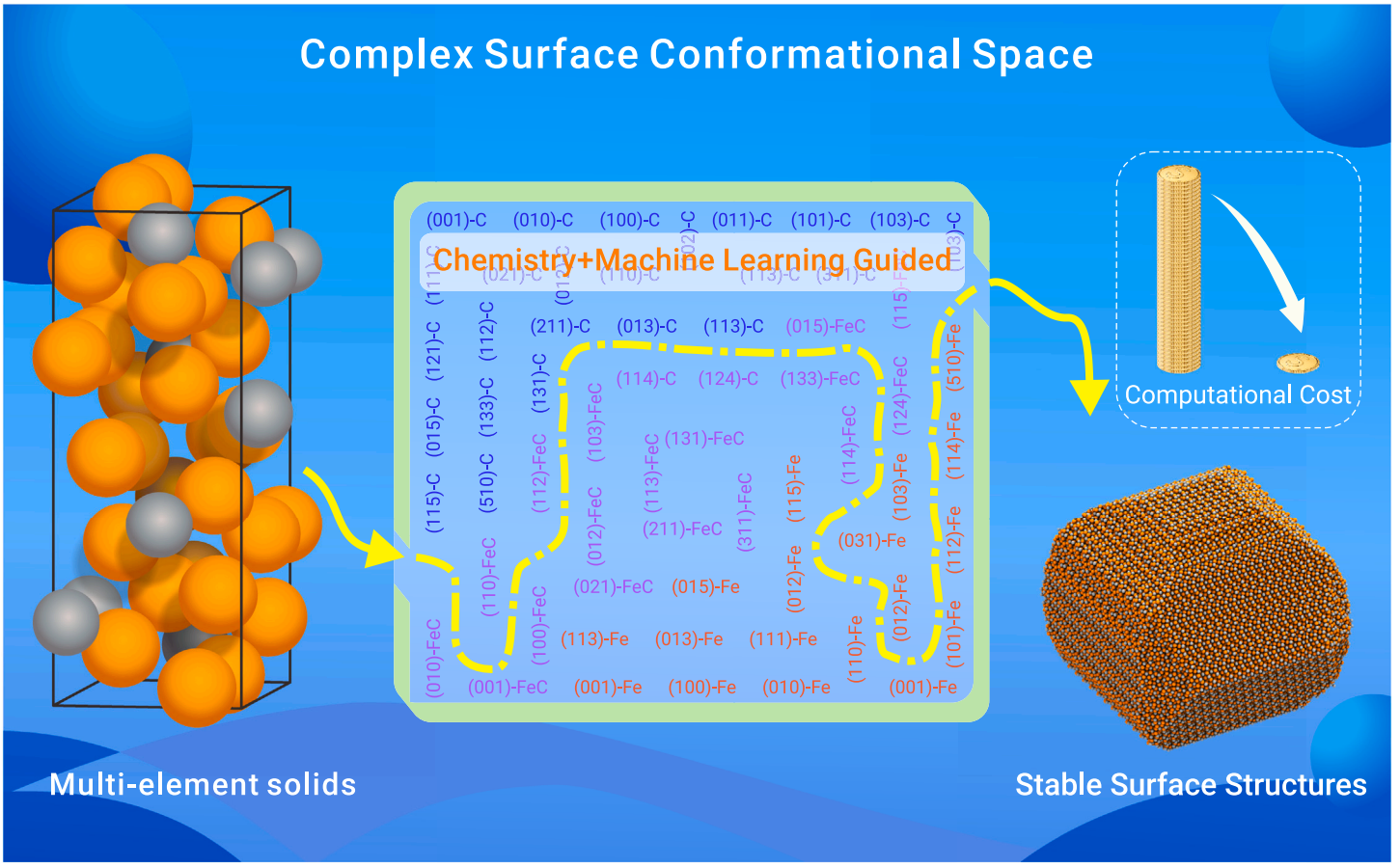
ARTICLE

Machine learning predicts atomistic structures of multielement solid surfaces for heterogeneous catalysts in variable environments

Huan Ma,[1,2,3] Yueyue Jiao,[1,2,3] Wenping Guo,[3,*] Xingchen Liu,[1,2,*] Yongwang Li,[1,2,3,4] and Xiaodong Wen[1,2,3,4,*]
*Correspondence: ybyygu@gmail.com (W.G.); liuxingchen@sxicc.ac.cn (X.L.); wxd@sxicc.ac.cn (X.W.)
Received: August 4, 2023; Accepted: January 2, 2024; Published Online: January 8, 2024; https://doi.org/10.1016/j.xinn.2024.100571
© 2024 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- The catalytic activity of multielement solid surfaces relies on their atomic-level structure under working conditions.
- An active learning scheme for identifying stable surface structures of multielement solids under working conditions is proposed.
- The morphological evolution of Fe_7C_3 nanoparticles in various chemical environments is accompanied by the drastic redistribution of exposed surface sites.
- *In silico* prediction of the surface sites of multielement catalysts in heterogeneous catalysis becomes computationally achievable.

ARTICLE

Machine learning predicts atomistic structures of multielement solid surfaces for heterogeneous catalysts in variable environments

Huan Ma,[1,2,3] Yueyue Jiao,[1,2,3] Wenping Guo,[3,*] Xingchen Liu,[1,2,*] Yongwang Li,[1,2,3,4] and Xiaodong Wen[1,2,3,4,*]

[1]State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan 030001, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]National Energy Center for Coal to Liquids, Synfuels China Co., Ltd., Beijing 101400, China

[4]Beijing Advanced Innovation Center for Materials Genome Engineering, Industry–University Cooperation Base between Beijing Information S&T University and Synfuels China Co., Ltd., Beijing 100101, China

*Correspondence: ybyygu@gmail.com (W.G.); liuxingchen@sxicc.ac.cn (X.L.); wxd@sxicc.ac.cn (X.W.)

Received: August 4, 2023; Accepted: January 2, 2024; Published Online: January 8, 2024; https://doi.org/10.1016/j.xinn.2024.100571

© 2024 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Citation: Ma H., Jiao Y., Guo W., et al., (2024). Machine learning predicts atomistic structures of multielement solid surfaces for heterogeneous catalysts in variable environments. The Innovation **5(2)**, 100571.

Solid surfaces usually reach thermodynamic equilibrium through particle exchange with their environment under reactive conditions. A prerequisite for understanding their functionalities is detailed knowledge of the surface composition and atomistic geometry under working conditions. Owing to the large number of possible Miller indices and terminations involved in multielement solids, extensive sampling of the compositional and conformational space needed for reliable surface energy estimation is beyond the scope of *ab initio* calculations. Here, we demonstrate, using the case of iron carbides in environments with varied carbon chemical potentials, that the stable surface composition and geometry of multielement solids under reactive conditions, which involve large compositional and conformational spaces, can be predicted at *ab initio* accuracy using an approach that combines the bond valence model, Gaussian process regression, and *ab initio* thermodynamics. Determining the atomistic structure of surfaces under working conditions paves the way toward identifying the true active sites of multielement catalysts in heterogeneous catalysis.

INTRODUCTION

The surfaces of solids play central roles in many physical and chemical applications, such as photoelectric devices, anticorrosion coatings, fuel cells, and heterogeneous catalysis.[1–6] In heterogeneous catalysis, in particular, the properties and performances of surfaces depend on their composition and atomic-scale structures in working environments, which usually involve finite temperatures and pressures. Under such conditions, particle exchange may occur between the surface and the gaseous or liquid environment by chemical interactions such as oxidation[7] or reduction,[8] leading to the alteration of surface structures, which can significantly affect how the surfaces function, such as catalytic activity and selectivity.[9,10] The rational design of solids with effective surfaces is hindered mainly by the ambiguity of the structure of surfaces due to the environment.

Determining the atomic-scale structure of multielement solid surfaces under variable reactive conditions remains a substantial challenge for both experiments and theory. This can be attributed to the complex nature of structural evolution caused by air and temperature sensitivity.[11–14] Experimentally, scanning electron microscopy (SEM) is an effective technique for obtaining three-dimensional (3D) images of morphological characteristics or elemental compositions combined with energy-dispersive X-ray data, but this method cannot reach atomic resolution.[15] Transmission electron microscopy (TEM) has a higher resolution than SEM, but it is limited to 2D imaging and can cause sample damage due to the high dose of radiation.[15,16] Although environmental SEM and environmental TEM allow the introduction of a gaseous environment, the resolution degrades due to the collision of incident electrons with gas molecules.[17,18]

Theoretically, under conditions of a defined environment at finite temperature and pressure, the thermodynamic quantity that determines the equilibrium geometry and composition of a solid surface is the Gibbs free energy. The foundation for calculating the Gibbs free energy of a surface under given conditions by using modern electronic structure theories was laid by Reuter and Scheffler in *ab initio* thermodynamics theory.[19,20] Despite its success in predicting the thermodynamic properties of preselected surfaces of elementary substances and compounds,[21–24] the determination of all surface terminations of an inorganic compound under reactive conditions, which are collectively responsible for the interfacial behaviors of the surface such as catalysis, requires adequate sampling from the entire surface compositional and conformational space. This approach is intractable for solids of high compositional and structural complexity because of the prohibitively large computational expenses associated with evaluating the diverse surface structures that are possible in terms of Miller indices, surface stoichiometry, and surface terminations. Machine learning interatomic potentials could dramatically reduce the cost of surface structure sampling, but it suffers from low transferability and generalizability, so models must be generated for specific tasks.[25] Thus, in many theoretical studies of multielement systems, surface properties have been investigated using only several empirically selected low-index facets and/or bulk-terminated surface structures,[26–29] leading to significant "material gaps."[30]

In this work, we propose a general and transferable approach for the fast prediction of the thermodynamically stable surface structures of complex multielement solids in different gaseous environments. This is made possible by combining the bond valence model,[31] Gaussian process regression (GPR), active learning, and *ab initio* thermodynamics. We demonstrate the validity of our approach compared with full-accuracy density functional theory (DFT) in predicting the stable surface structures of iron carbides at given carbon chemical potentials ($\Delta\mu_C$) imposed by the gas phase surroundings.[11,32] Because atomic-scale surface structures of iron carbides determined from experiments in reactive gaseous environments are lacking, direct comparisons of the theoretical predictions with experimental data are not possible. However, using o-Fe_7C_3, an active phase in Fe-based Fischer–Tropsch synthesis (FTS),[33] as a demonstration case, we obtain the evolution of stable surface structures and surface sites of nanoparticles with changes in $\Delta\mu_C$, and the predicted shape under various conditions resembles that of the experiments. We anticipate that our approach will help bridge the gap between our understanding of the atomistic structure of multielement solid surfaces *ex situ* and under reactive conditions in heterogeneous catalysis and other complicated gaseous environments.

MATERIALS AND METHODS

DFT methods

All of the geometry optimizations and energy calculations presented were performed at the DFT level using the Vienna *ab initio* simulation package,[34,35] with the Perdew-Burke-Ernzerhof generalized gradient approximation[36] and the projector-augmented wave[37,38] method. Surface calculations were performed on symmetric slabs generated by using the Python Materials Genomics (pymatgen) package.[39] More details of the DFT methods can be found in the supplemental information.

Geometric descriptor of the degree of surface openness

The degree of surface undercoordination (d_{suc}) is a geometric descriptor that quantifies the openness of arbitrary surface structures.[40] d_{suc} for a given slab can be expressed as follows:

$$d_{suc} = \frac{\sum\left[\left(\overline{V}_i^{bulk} - \overline{V}_i^{slab}\right)\Big/\overline{V}_i^{bulk}\right]^2}{2A} \quad \text{(Equation 1)}$$

where A is the surface area of one side of the slab, $\overline{V}_i^{slab}$ is the generalized atomic valence (GAV) of atom i in the slab, and $\overline{V}_i^{bulk}$ is the GAV that atom i should possess in the bulk. The GAV of an atom i can be calculated as follows:

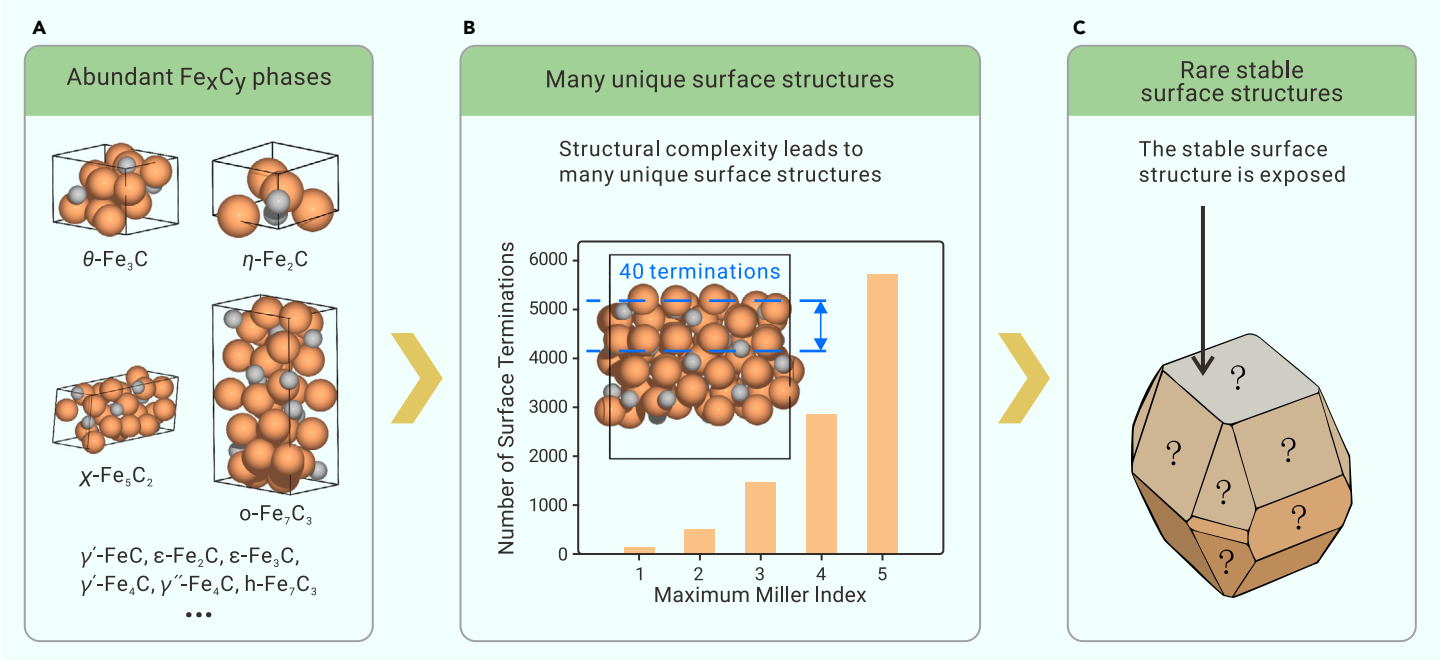


Figure 1. Computational challenge of determining the surface properties of complex multielement solids (A) The abundant bulk phases of Fe carbides. (B) For each phase, the compositional and structural complexity leads to a large surface compositional and conformational space. (C) Only a small number of surface compositions and conformations are expected to be thermodynamically stable, and they appear on a crystal particle under specific conditions.

$$\overline{V}_i = \sum \frac{V_j}{V_{j,\max}} \qquad \text{(Equation 2)}$$

where V_j is the AV of the individual first-nearest-neighbor atom j and $V_{j,\max}$ is the maximum AV of the element corresponding to atom j in the bulk. The AV can quantitatively describe the atomic coordination under the weight of interatomic distance.[31] The AV of an atom j is the sum of the individual bond valences

$$V_j = \sum \exp\left(\frac{R_{jk} - d_{jk}}{b}\right) \qquad \text{(Equation 3)}$$

where b is an empirical constant usually equal to 0.37 Å, d_{jk} is the geometric distance between atoms j and k, R_{jk} is a parameter that is empirically determined by different bond types, and in this work, R_{jk} is determined by the sum of the covalent radii of atoms.

Surface structure dataset

Considering computational cost, all of the surface terminations of surface orientations up to a maximum Miller index (MMI) of 2 and surface orientations with characteristic peaks in the simulated X-ray diffraction (XRD) were included in the surface structure dataset. The dataset sizes were 265, 420, 345, and 312 for θ-Fe_3C, χ-Fe_5C_2, γ-FeP_4, and β-$FeSi_2$, respectively. All of the surface structures in the dataset were used to test the active learning scheme

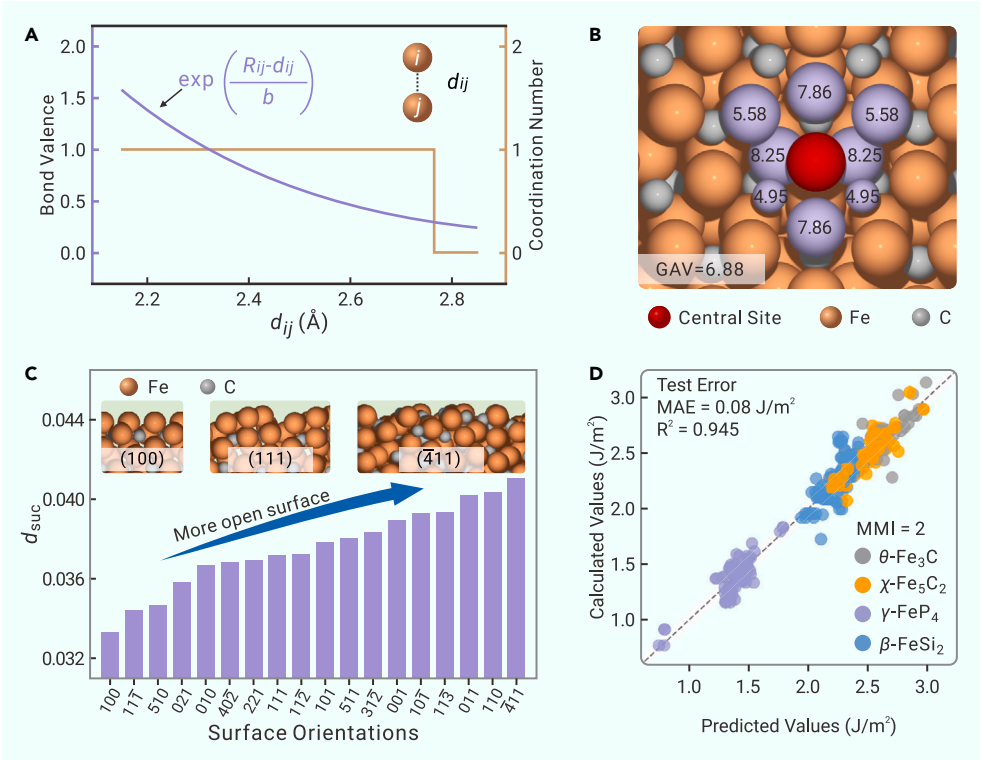


Figure 2. Capturing surface structure features and predicting surface stability using d_{suc} (A) In contrast to the conventional coordination number, the bond valence can quantitatively represent the atomic coordination as a function of interatomic distance. (B) The GAV of an atom (red) introduces a weight to each bonded atom (purple) corresponding to its atomic valence (labeled on the atom). GAV = $(2 \times 8.25 + 2 \times 7.86 + 2 \times 5.58)/8.41 + (2 \times 4.95)/5.75 = 6.88$. (C) d_{suc} of different χ-Fe_5C_2 surface structures. Each bar represents the surface termination with the smallest d_{suc} of the corresponding surface orientation. (D) Comparison of the surface free energies predicted from GPR with those calculated from DFT for test sets of θ-Fe_3C, χ-Fe_5C_2, γ-FeP_4, and β-$FeSi_2$. The surfaces (including only stoichiometric terminations) are limited to those with MMIs of 2 or those with higher Miller indices, which appear as characteristic peaks in the simulated XRD. Models are trained separately for different systems.

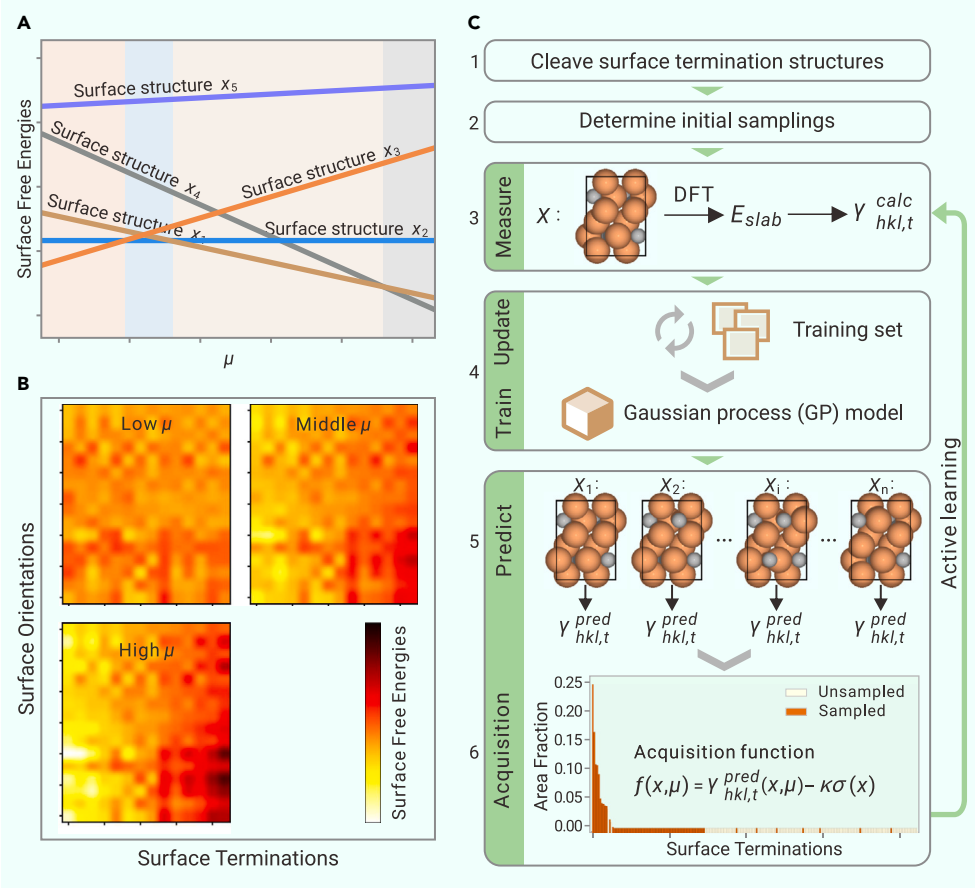


Figure 3. Identifying thermodynamically preferential surface structures at given chemical potentials with an active learning framework (A) Schematic drawing of the surface free energies depending on chemical potential. (B) Schematic drawing of the "surface structure free energy surface". (C) Active learning scheme for identifying thermodynamically preferential surface structures.

proposed in this work. The dataset composed of terminations with stoichiometric ratio of the bulk was randomly split into training and test sets (6:4) three times to test the ability to use descriptors combined with different regression models to predict surface free energies without considering external environments. The training set sizes were 34, 36, 28, and 45 for θ-Fe_3C, χ-Fe_5C_2, γ-FeP_4, and β-$FeSi_2$, respectively.

RESULTS AND DISCUSSION

Surface compositional and conformational space

For multielement solids, their compositional and structural complexity significantly increases the range of the surface conformational space, resulting in many unique surface configurations. For instance, among the abundant Fe carbide phases, o-Fe_7C_3 (the unit cell structure is given in Figure S1) contains 175 symmetrically distinct surface orientations and 5,729 unique surface terminations (which will take approximately 250,000 h to simulate with DFT on a 24-core computing cluster), even when the MMI reaches 5 (Table S1). A gaseous environment can lead to a change in the composition of surface structures through chemical interactions, which could greatly increase the compositional and conformational space. Fortunately, only a small number of surface structures are expected to be thermodynamically stable under specific conditions (Figure 1).

Predicting the surface stability of compounds without considering external environments

The stability of a surface is described by its surface free energy, which is essentially the excess energy of the surface atoms due to broken chemical bonds. Previously, we showed that the simple geometric descriptor of surface undercoordination (d_{suc}) based on the bond valence model allows accurate prediction of the surface stability of arbitrary metal surfaces by quantifying the disruption of intermolecular bonds.[40] Here, we first assess the suitability of using d_{suc} to capture the structural features of the native surfaces of multielement solids without considering the external environment. As shown in the example of χ-Fe_5C_2, d_{suc} can be used to effectively quantify the degree of surface openness; the larger the value of d_{suc}, the more open the surface structure (Figure 2C). This is attributed to its basis in the fundamental bonding theory and the fact that the bond valence model can capture more comprehensive geometric information. The bond valence provides the ability to quantitatively describe atomic interactions under the weight of interatomic distance (Figure 2A),[31] and the GAV can capture the atomic coordination environment with the contribution of the next-nearest-neighbors (Figure 2B).[40]

To further verify the ability of the descriptor to predict the surface stability of multielement solids, the surface free energies of the stoichiometric terminations of four binary iron compounds, θ-Fe_3C, χ-Fe_5C_2, γ-FeP_4, and β-$FeSi_2$ (the unit cell structures are given in Figure S1), and several bimetallic alloys (FeTM and Fe_3TM, TM = Co, Ni, Pd, and Pt) were predicted by DFT. We took the d_{suc} of the given bulk phase and the surface free energies from DFT as the inputs and target values, respectively, to train various machine learning models, with which predictions could be made. The results show that for bimetallic alloys, d_{suc} can predict the surface free energies well using a simple linear model (Figure S2). However, for θ-Fe_3C, χ-Fe_5C_2, γ-FeP_4, and β-$FeSi_2$, even the GPR and neural network (NN) models cannot achieve very accurate prediction (Figures 2D and S3–S5; Table S2). We speculate that this is because these alloys, despite being multielement solids, maintain the same atomic packing (body-centered or face-centered cubic) as pure metals. However, for complex multielement solids such as θ-Fe_3C and χ-Fe_5C_2, the C atoms are located in the trigonal prismatic sites of the distorted hexagonally closed-packed structure of Fe atoms.[41] This structural feature leads to the coexistence of multiple inequivalent Fe sites accompanied by a variety of Fe–C bond lengths and bond angles, so that there is no longer a simple relationship between the surface stability and surface undercoordination. For complex multielement solids with a small training set, GPR is expected to be a more reliable model than the linear and NN models. However, the prediction results still greatly deviate from the DFT data. Thus, for systems with limited training sets, models based on pure geometric analysis can be used to estimate surface free energies but are not able to predict accurate values; for this purpose, explicit DFT calculations are indispensable.

Active learning and identification of stable surface structures at given chemical potentials

Note that the above analysis is only for stoichiometric terminations (symmetric and stoichiometric slabs). Under realistic conditions, the surface may undergo preferential nonstoichiometric termination through particle exchange with its environment to reach thermodynamic equilibrium. Breaking stoichiometry requires including the chemical potential (μ) as a reservoir for nonstoichiometric systems. For Fe_mX_n binary compounds in a gas-phase environment with a chemical potential of μ_X, the surface stability depends on μ_X, which connects the working environment (Figure 3A). Details of the derivations following *ab initio* thermodynamics theory can be found in the supplemental information. Here, we show how the most stable surface structures of a multielement particle can be determined by traversing the surface compositional and conformational space; we take advantage of the simple geometric descriptor d_{suc} and a machine learning approach.

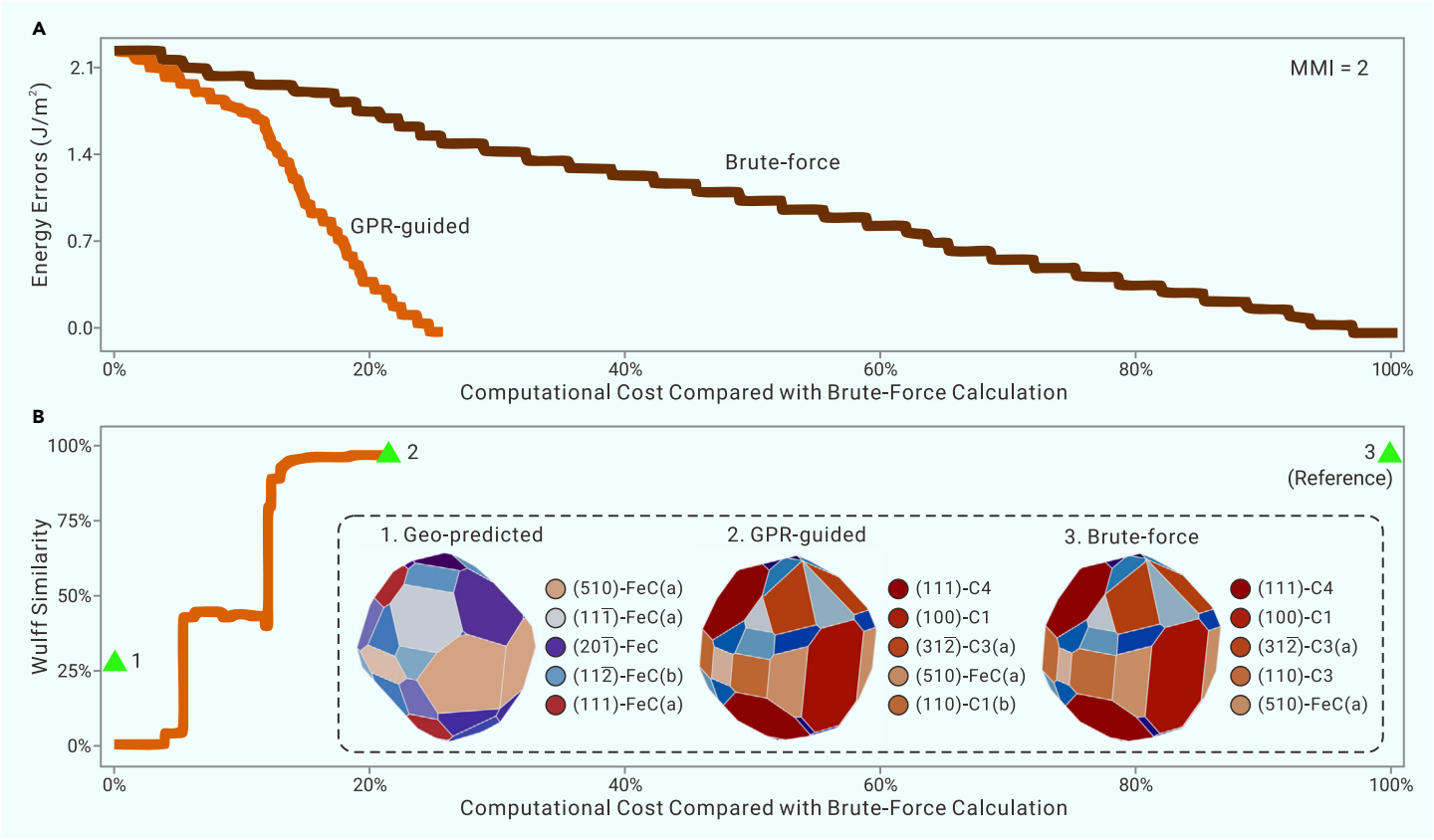


Figure 4. Comparison of GPR-guided sampling with brute-force DFT at a specific $\Delta\mu_C$ for χ-Fe_5C_2 (A) Error curves of the surface free energy for identifying the stable structure of each surface orientation at a $\Delta\mu_C$ of −7.55 eV. The energy errors represent the MAEs of the surface free energies of all of the stable surface structures determined in each iteration with respect to the corresponding brute-force DFT results. (B) Wulff similarity curves for comparing the surface structures that appear on crystal shapes at a $\Delta\mu_C$ of −7.55 eV, with the brute-force DFT as the reference. The Wulff similarity represents the cosine similarity of 2 vectors composed of area fractions on the crystal shape determined in each iteration, derived from pure geometric analysis, GPR-guided sampling, and brute-force DFT, corresponding to the 3 triangle markers. Pure geometric analysis is only used for stoichiometric terminations. The legend on the right side of each crystal shape shows the top 5 exposed facets.

We find that the compositional and conformational space in this problem can be regarded as a "surface structure free energy surface," which is composed of the surface orientation, surface termination, chemical potential, and surface free energy (Figure 3B). Searching this space is somewhat similar to the conventional global search task on a potential energy surface. The difference is that the system size of the former is variable, whereas that of the latter is constant. Inspired by the GPR-based global optimization framework proposed by Bisbo and Hammer,[42] we propose a new scheme for identifying thermodynamically preferential surface structures by accelerating the exploration of the surface compositional and conformational space via active learning (Figure 3C); for these structures, exact DFT calculations of the surface free energies can be carried out.

In this scheme, the surface free energies of a small number of surface structures with low-index facets are first predicted via DFT. A machine learning model is then trained based on the existing DFT data. Considering its high performance in estimating statistical uncertainties and its high adaptability to dynamic sampling and active learning, GPR is chosen as the regression model; it has been widely applied to solve surface science problems over the years.[43–45] The surface free energy dependence of μ is affected by both the surface undercoordination and the composition. Thus, the d_{suc} of the entire surface and the d_{suc} and ΔN of different types of atoms ($d_{suc,Fe}$, $d_{suc,X}$, ΔN_{Fe}, ΔN_X) are treated as model inputs, where ΔN is the number of excess or deficit atoms per unit surface area and reflects the degree to which the surface deviates from the bulk stoichiometry. A larger absolute value of ΔN is associated with greater deviation. The details of the GPR model for predicting surface free energy can be found in the supplemental information.

It is not wise to blindly perform a large number of calculations to generate a set of data to train the model. Thus, we use the active learning framework to gradually refine the model by performing selective sampling of candidate structures and dynamically updating the training set until the convergence criterion is reached. This surrogate model is used to predict the surface free energies of unsampled structures, and then an acquisition function based on surface free energies is used to determine which of the candidate structures is the most promising and should be sampled by DFT. Based on Bayesian probability theory, uncertainty estimations can be obtained and used to increase the advantages of unsampled but stable structures on the energy landscape. The Wulff construction can determine the crystal shape under equilibrium conditions,[46] and an atomistic Wulff structure is useful for identifying active sites on a nanoparticle.[47] By combining the Wulff construction, the likelihood of a surface structure appearing on a crystal particle can also be included. Although the model prediction errors are generally between 0.10 and 0.15 J/m^2 (mean absolute error [MAE]), this poor accuracy does not significantly affect the final energies. This is because the final energies are determined at the DFT level. More details of the scheme for identifying thermodynamically stable surface structures are included in the supplemental information.

Performance of the active learning scheme

In this section, we illustrate that the accuracy of the scheme is comparable to that of brute-force DFT, but the computational cost can be greatly reduced by rational sampling through our active learning scheme using Bayesian inference as the structure selection metric. For testing purposes, we restrict our sampling to all of the surfaces up to an MMI of 2 and orientations that appear as characteristic peaks in the simulated XRD pattern. The results of traversing the surface compositional and conformational space with DFT calculations are regarded as the reference, which predicts similar crystal shapes for θ-Fe_3C and χ-Fe_5C_2 to those detected from SEM characterizations (Figures S6 and S7)[48,49] and a facet termination of β-$FeSi_2$(100) that is identical with that obtained by experimental observation (Figure S8).[50] We first chose only a single $\Delta\mu_C$ (−7.55 eV) to verify our approach to χ-Fe_5C_2, one of the active phases of Fe carbide in FTS, under conditions in which its crystal shape has a complex facet composition. We used this scheme to identify the stable surface structure with the lowest surface

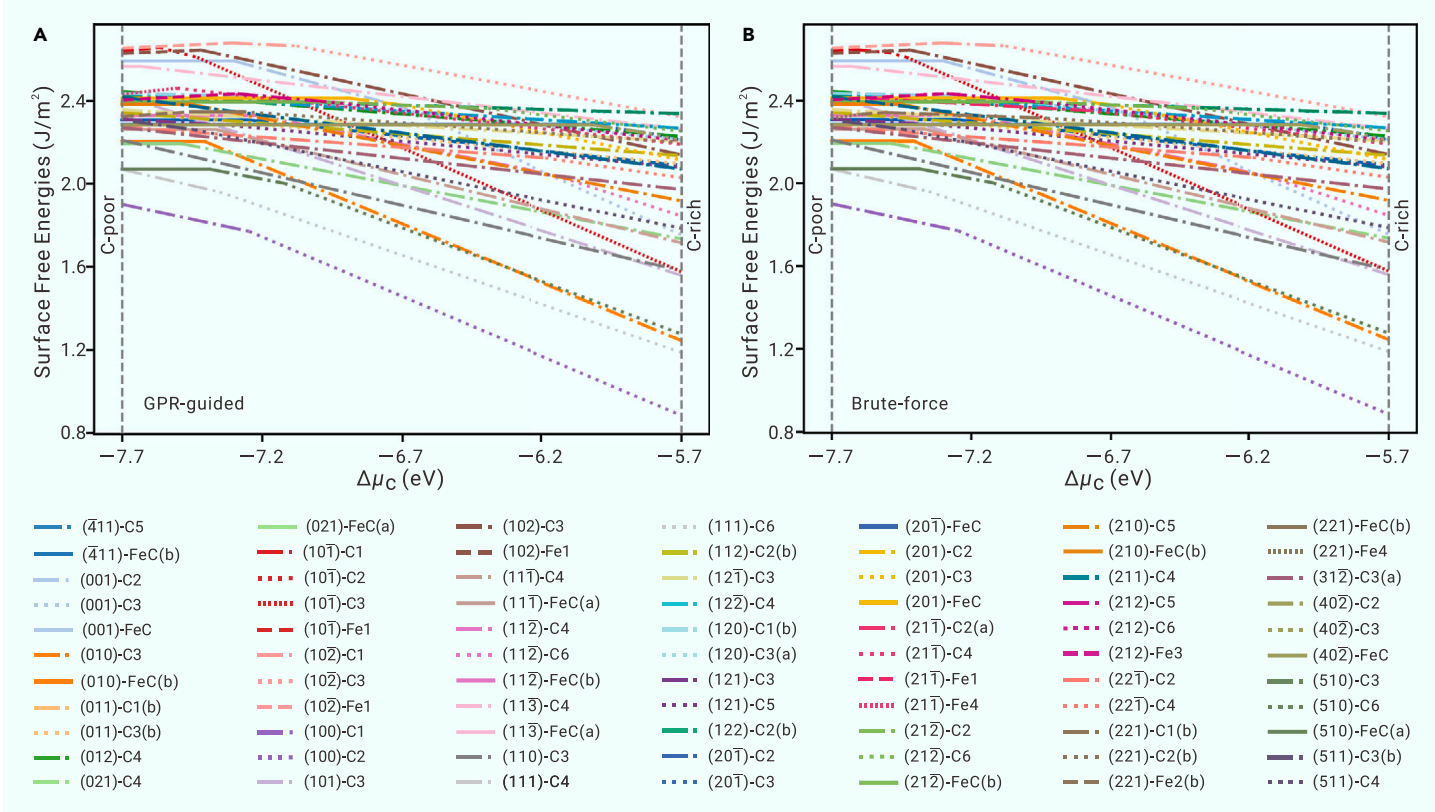


Figure 5. Comparison of GPR-guided sampling with brute-force DFT at various $\Delta\mu_C$ values for χ-Fe_5C_2 (A and B) Surface free energies of the stable terminations as a function of $\Delta\mu_C$ from (A) GPR-guided sampling and (B) brute-force DFT. For clarity, surface terminations with chemical potential ranges less than 0.05 eV are not shown. The surfaces (including all possible terminations) are limited to those with MMIs of 2 or those with higher Miller indices, which appear as characteristic peaks in the simulated XRD.

free energy for each surface orientation. The error in the surface free energy of the identified stable surface structures compared with that of brute-force DFT can reach 0.01 J/m^2 (MAE), which suggests that the surface terminations with the lowest surface free energy for almost all of the surface orientations have been identified. However, only 25.2% of the structure dataset of χ-Fe_5C_2 needs to be calculated in our scheme (Figure 4A). The GPR model trained with a limited dataset is also accurate enough (MAE of 0.12 J/m^2 for model prediction) to be embedded in the active learning framework to guide DFT sampling for reconstructed or defect surfaces without noticeable errors (Figure S9).

Furthermore, as a more practical extension, we identify the stable surface structures that appear on the crystal shape. As shown in Figure 4B, the ratios of surface exposure obtained by our active learning scheme are in good agreement with those from brute-force DFT, with a high Wulff similarity of 96.2% (for the definition of the cosine similarity of two Wulff shapes, see the supplemental information); however, the computational cost is only 20.9% that of brute-force DFT. Most of the stable surface structures on the crystal shape are successfully reproduced; the only exception among the top five exposed facets is that the metastable termination C1(b) is incorrectly regarded as the most stable termination of the (110) facet instead of the C3 termination predicted by brute-force DFT. However, the difference in the surface free energy between the terminations C1(b) and C3 is only approximately 0.05 J/m^2, so the stabilities of these two terminations are very close. From a statistical point of view, surface structures that are close in energy may have similar populations according to the Boltzmann distribution. We also examined the sampling of multiple candidates in each iteration, proving that our scheme allows parallel computing on computing clusters (Figure S10). As a comparison, the crystal shape from pure geometric analysis has a low Wulff similarity of only 28.6% to the shape obtained by the brute-force DFT. The ratios of the surface exposures of crystal shapes from different approaches are given in Table S3.

When considering a range of $\Delta\mu_C$, we can use this scheme to obtain the surface phase diagram. The calculated surface free energies of the most stable terminations at different $\Delta\mu_C$ can be used to construct the surface phase diagram of χ-Fe_5C_2, as shown in Figure 5. The diagram based on GPR-guided sampling is very close to that based on brute-force DFT but is obtained at only 48% of the computational cost. The slight deviations for the (10 $\bar{1}$), (21 $\bar{1}$), and (221) facets can be attributed to the existence of surface terminations with very close surface free energies. In addition, we investigated θ-Fe_3C, γ-FeP_4, and β-$FeSi_2$, and high consistency between the surface phase diagrams of our scheme and those of brute-force DFT was reached (Figures S11–S13).

Identifying the stable surface structures and atomic sites of an o-Fe_7C_3 nanoparticle

Limited by the space resolution of existing experimental techniques, it is rarely possible to observe the atomic-scale surface structures and crystal shapes of single-phase Fe carbide in gaseous environments via high-resolution electron microscopy. As far as we are aware, among the Fe carbides, the only phase with a morphology at various carbon chemical potentials that has been determined experimentally is Fe_7C_3. Liu et al.[51] reported that o-Fe_7C_3 is most likely the Fe_7C_3 phase that contributes to the experimental XRD pattern under FTS conditions.[33] Multiple diffraction peaks attributed to high-index facets in the simulated XRD data hint at the complexity of the preferentially exposed facets of the o-Fe_7C_3 crystal (Figure S14). However, the existing theoretical work on the stable surface structures of o-Fe_7C_3 is unsystematic and limited to low-index facets.[26,51] Using the approach proposed in this work, we can extensively explore a large compositional and conformational space with 5,729 candidate structures with MMIs up to 5 while reducing the computational cost by 2 orders of magnitude (Figure 6B). We studied the thermodynamic stability of the surface structures of this material and the evolution of crystal shapes under different conditions, which are vital for tuning the surface composition and structure and synthesizing materials with tailored shapes. The calculated stable terminations and crystal shapes at different $\Delta\mu_C$ values are shown in Figure 6A. The optimized surface structures of all of the stable terminations are available at https://github.com/spdkit/ML-Assisted_Surface_Structure_Prediction.

In the studied range of $\Delta\mu_C$ (−7.75 to −5.75 eV), the stable surface structures of o-Fe_7C_3 are mainly carbon-rich terminations. At low $\Delta\mu_C$ (−7.75 eV), the facet composition of the crystal shape is more complex and has no obvious predominant facets. This can be attributed to the narrow distribution of the surface free energy (0.20 J/m^2 across the top 5 exposed surface structures). Despite the

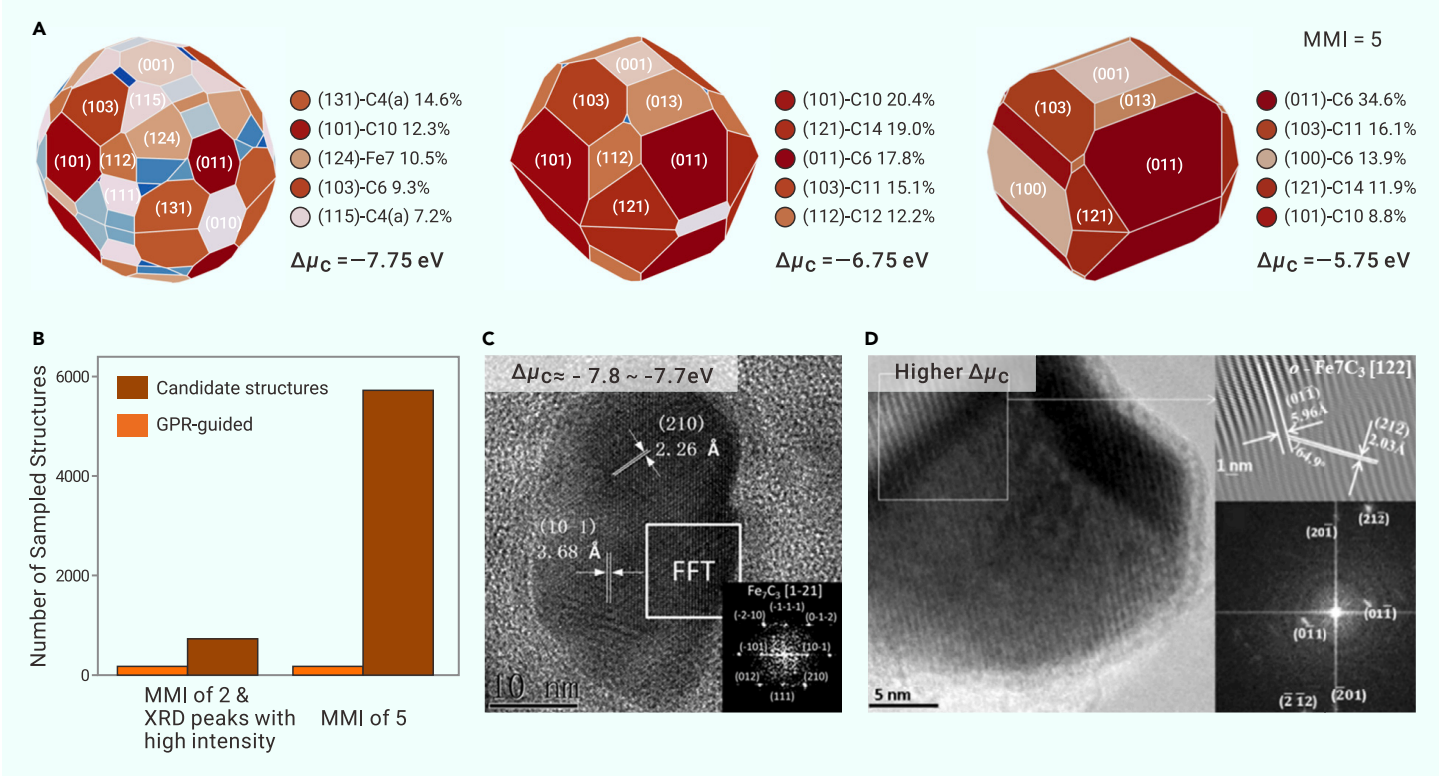


Figure 6. Evolution of stable terminations and crystal shapes of o-Fe_7C_3 under different conditions (A) Calculated stable terminations and crystal shapes at different $\Delta\mu_C$. The surfaces (including all possible terminations) are limited to an MMI of 5. The legend on the right side of each crystal shape shows the top 5 exposed facets. (B) Number of sampled structures for identifying the stable surface structures that appear on the crystal shape at a single $\Delta\mu_C$ in o-Fe_7C_3 surface compositional and conformational spaces of different sizes using GPR-guided sampling. (C) and (D) Experimental HRTEM images reported in the literature. (C) Reproduced with permission from reference 33. Copyright 2018, American Chemical Society. (D) Reproduced with permission from reference 52. Copyright 2018, Elsevier.

difficulties in discriminating all of the exposed surfaces in the *in operando* conditions, the ellipsoidal shape and (101) facet have been clearly detected at similar $\Delta\mu_C$ values by high-resolution TEM (HRTEM) characterization (Figure 6C) in previously published experimental work.[33] Four of the five facets with the largest contributions are high-index facets, namely, (131), (124), (103), and (115), and they cover 41.6% of the total surface area, emphasizing the significance of investigating high-index facets. When $\Delta\mu_C$ increases to −6.75 eV, the surface free energy of the facets is broadened, resulting in a decrease in the number of exposed facets, with the (101) facet having the largest proportion. At higher $\Delta\mu_C$, our theoretical prediction shows a polyhedral shape with sharp edges and the (011) facet being predominant. Similar shape features and the (011) facet have also been detected by HRTEM characterization in earlier reported experimental work for a higher chemical potential (Figure 6D).[52]

An extremely fascinating transformation of the nanoparticle structure with changing $\Delta\mu_C$ is the change in surface atom coordination. Metal sites with varied coordinations are known to exhibit quite different coordination chemistries in regard to reacting gas molecules. Therefore, the catalytic performance and other interfacial properties of a solid nanoparticle are mainly determined by the distribution of the metal coordination sites.[53,54] In FTS, for example, previous studies[55,56] have reported that active sites with low coordination numbers, such as vacancy sites, are likely responsible for the large decrease in the CO dissociation barrier. However, surface Fe sites with high Fe−C coordination numbers can favor CH_4 formation.[55,57] In this case, with the morphology and surface structures of all the exposed facets determinable at *ab initio* accuracy, our approach makes it possible to track the changes in the atomic-site distribution of an entire nanoparticle with variable external environments.

As a demonstration, we constructed a realistic Fe_7C_3 nanoparticle with a diameter of 13 nm at high and low $\Delta\mu_C$ (Figures 7A and 7C), which was experimentally synthesized[33,52,58,59] and used in FTS with the carbon chemical potential tuned by the H_2/CO ratio in syngas. This approach is important when matching theory with experiments in which functional materials or catalysts usually exist in the form of nanoparticles, and the evolution of their precise atomistic structures with the external environment may be unclear. We used the GAV to reflect the coordination of the surface atoms and the structural features of the surface sites (Figure 7B) and performed t-distributed stochastic neighbor embedding (t-SNE)[60] visualization for 18,753 Fe sites on the surface of the nanoparticles at different $\Delta\mu_C$ (Figures 7D and 7E). The sites are color-coded by their GAVs. The site fingerprint consists of a series of bond valence values between the central site and 12 nearest-neighbor Fe atoms and 6 nearest-neighbor carbon atoms. The proximity in this reduced space indicates a similarity in the structural features of the surface sites. More details of the t-SNE visualization can be found in the supplemental information.

As shown in Figure 7D, the types of Fe sites on the surface of the nanoparticles are diverse in terms of the GAV, and their distribution changes significantly with $\Delta\mu_C$. Edge and corner sites generally have low GAVs and are grouped into their own clusters. However, we find that some sites belonging to crystal facets (e.g., high-index (131) and (124) facets) also have low coordination and are grouped into clusters containing edge and corner sites, which indicates that these sites may play similar roles as edge and corner sites in FTS processes. Overall, from low $\Delta\mu_C$ to high $\Delta\mu_C$, more C atoms from gaseous environments are deposited on the surfaces of the nanoparticles. Taking the (011) facet as an example, with increasing $\Delta\mu_C$, the types of sites on the facet can be grouped into different clusters with GAVs centered at 5.6, 6.6, and 7.2 at a $\Delta\mu_C$ of −7.75 eV and centered at 6.1, 7.1, and 8.9 when $\Delta\mu_C$ increases to −5.75 eV (Figures 7D and 7E). The separation of these clusters in the latent space indicates a drastic change in the distribution of bond valence (and thus coordination) between the Fe sites and their neighbors due to the change in surface terminations from stoichiometric to carbon rich. Moreover, with increasing $\Delta\mu_C$, the exposed area of high-index facets and the number of edges and corners decrease due to the reduction in the number of exposed facets (Figure 6A). These findings explain why the coordination of Fe sites on the surface increases as $\Delta\mu_C$ increases, as shown in Figure 7B, which suggests that o-Fe_7C_3 nanoparticles at high $\Delta\mu_C$ may have higher intrinsic activity in catalyzing CH_4 formation in FTS.

CONCLUSION

We propose a new active learning scheme for identifying the surface structures of multielement solids under realistic conditions based on the bond

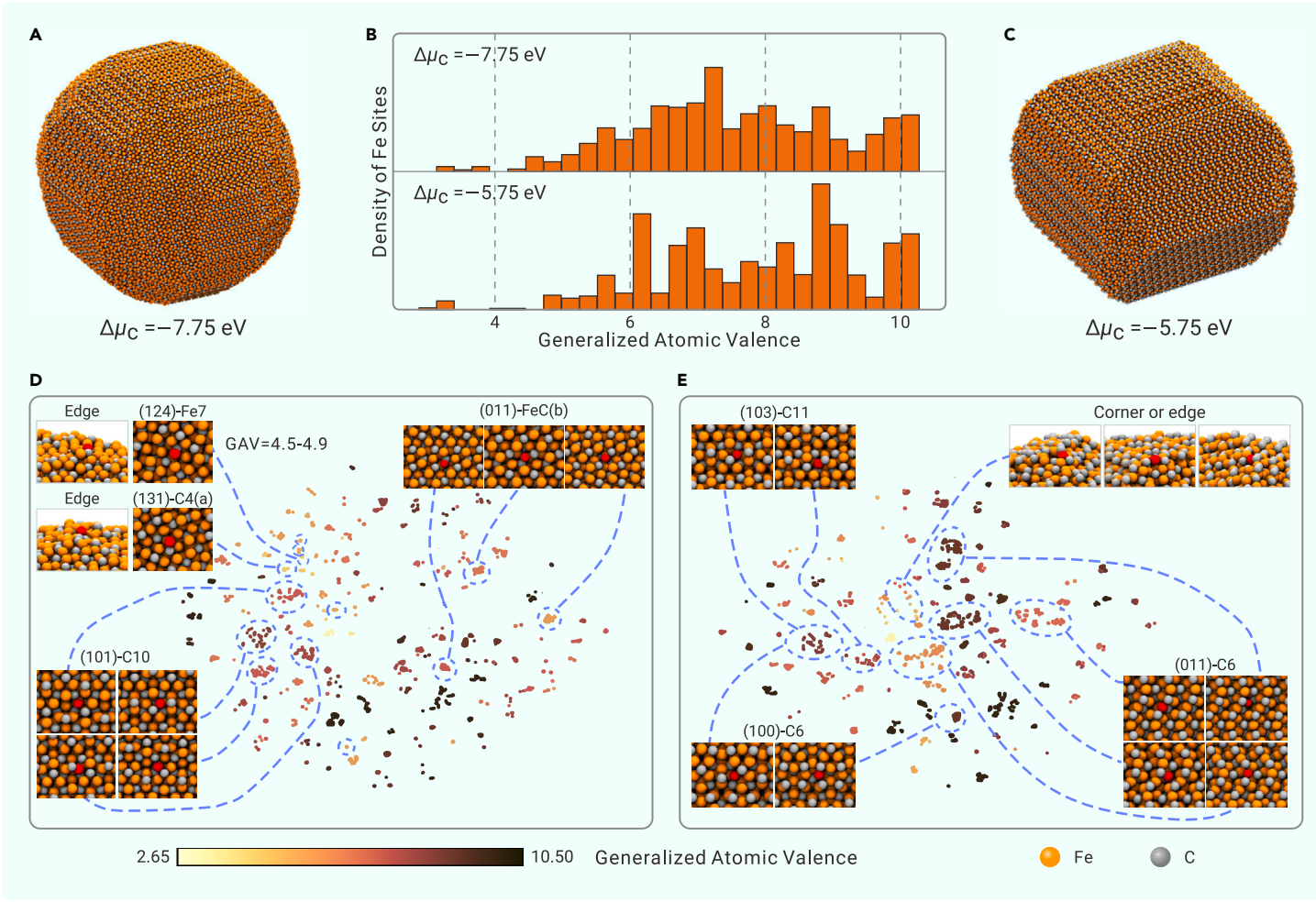


Figure 7. Analysis of the atomic-site distribution on the surface of o-Fe_7C_3 nanoparticles (A and C) o-Fe_7C_3 nanoparticles with diameters of 13 nm at high and low $\Delta\mu_C$. (B) Histogram of the GAV of the surface Fe sites of nanoparticles at different $\Delta\mu_C$. (D and E) Latent space visualization with t-SNE for surface Fe sites corresponding to the nanoparticles in (A) and (C). The sites are color-coded by their GAV.

valence model, GPR, and *ab initio* thermodynamics. This scheme is able to efficiently and accurately identify the most thermodynamically stable structures that appear in the crystal shapes of multielement solids in different gaseous environments, which has been a major obstacle in understanding their structure–performance relationships under working conditions. Using this approach, we demonstrate via o-Fe_7C_3 that the evolution of the crystal shapes of multicomponent solids in various chemical environments is accompanied by a drastic redistribution of exposed surface sites, which may profoundly influence their catalytic performance. This approach can be extended to systems with more constituent elements. The deciphered atomistic structure of surfaces in actual environments paves the way toward identifying the true active sites of multielement solid catalysts in heterogeneous catalysis and understanding the actual surface structures of materials in contact with gaseous environments.

REFERENCES

1. Zhou, Y., Fuentes-Hernandez, C., Shim, J., et al. (2012). A universal method to produce low-work function electrodes for organic electronics. Science ***336***: 327–332. https://doi.org/10.1126/science.1218829.
2. Hou, L., Liu, X., Ge, X., et al. (2023). Designing of anisotropic gradient surfaces for directional liquid transport: Fundamentals, construction, and applications. Innovation ***4***: 100508. https://doi.org/10.1016/j.xinn.2023.100508.
3. Xu, X., Shi, S., Tang, Y., et al. (2021). Growth of NiAl-layered double hydroxide on graphene toward excellent anticorrosive microwave absorption application. Adv. Sci. ***8***: 2002658. https://doi.org/10.1002/advs.202002658.
4. Stamenkovic, V., Mun, B.S., Mayrhofer, K.J.J., et al. (2006). Changing the activity of electrocatalysts for oxygen reduction by tuning the surface electronic structure. Angew. Chem. Int. Ed. ***45***: 2897–2901. https://doi.org/10.1002/anie.200504386.
5. Nørskov, J.K., Bligaard, T., Rossmeisl, J., et al. (2009). Towards the computational design of solid catalysts. Nat. Chem. ***1***: 37–46. https://doi.org/10.1038/nchem.121.
6. Wu, T., Sun, M., and Huang, B. (2021). Atomic-strain mapping of high-index facets in late-transition-metal nanoparticles for electrocatalysis. Angew. Chem. Int. Ed. ***60***: 22996–23001. https://doi.org/10.1002/anie.202110636.
7. Huš, M., Kopač, D., Bajec, D., et al. (2021). Effect of surface oxidation on oxidative propane dehydrogenation over chromia: An ab initio multiscale kinetic study. ACS Catal. ***11***: 11233–11247. https://doi.org/10.1021/acscatal.1c01814.
8. Cao, J., Rinaldi, A., Plodinec, M., et al. (2020). In situ observation of oscillatory redox dynamics of copper. Nat. Commun. ***11***: 3554. https://doi.org/10.1038/s41467-020-17346-7.
9. Ruan, P., Chen, B., Zhou, Q., et al. (2023). Upgrading heterogeneous Ni catalysts with thiol modification. Innovation ***4***: 100362. https://doi.org/10.1016/j.xinn.2022.100362.
10. Li, H., Jiang, Y., Li, X., et al. (2023). C2+ selectivity for CO2 electroreduction on oxidized Cu-based catalysts. J. Am. Chem. Soc. ***145***: 14335–14344. https://doi.org/10.1021/jacs.3c03022.
11. de Smit, E., and Weckhuysen, B.M. (2008). The renaissance of iron-based Fischer–Tropsch synthesis: On the multifaceted catalyst deactivation behaviour. Chem. Soc. Rev. ***37***: 2758–2781. https://doi.org/10.1039/b805427d.
12. Alaba, P.A., Abbas, A., Huang, J., et al. (2018). Molybdenum carbide nanoparticle: Understanding the surface properties and reaction mechanism for energy production towards a sustainable future. Renew. Sustain. Energy Rev. ***91***: 287–300. https://doi.org/10.1016/j.rser.2018.03.106.
13. Fischer, N., and Claeys, M. (2020). In situ characterization of Fischer–Tropsch catalysts: A review. J. Phys. D Appl. Phys. ***53***: 293001. https://doi.org/10.1088/1361-6463/ab761c.
14. Li, H., Jiao, Y., Davey, K., et al. (2023). Data-driven machine learning for understanding surface structures of heterogeneous catalysts. Angew. Chem. Int. Ed. ***62***: e202216383. https://doi.org/10.1002/anie.202216383.
15. Padhi, S., and Behera, A. (2022). Biosynthesis of silver nanoparticles: Synthesis, mechanism, and characterization. In Agri-Waste and Microbes for Production of Sustainable Nanomaterials, K.A. Abd-Elsalam, R. Periakaruppan, and S. Rajeshkumar, eds. (Elsevier), pp. 397–440. https://doi.org/10.1016/b978-0-12-823575-1.00008-1.
16. Ke, X., Bittencourt, C., and Van Tendeloo, G. (2015). Possibilities and limitations of advanced transmission electron microscopy for carbon-based nanomaterials. Beilstein J. Nanotechnol. ***6***: 1541–1557. https://doi.org/10.3762/bjnano.6.158.

17. Wang, Z.L., and Lee, J.L. (2008). Electron microscopy techniques for imaging and analysis of nanoparticles. In Developments in Surface Contamination and Cleaning, R. Kohli and K.L. Mittal, eds. (William Andrew Publishing), pp. 395–443. https://doi.org/10.1016/B978-0-323-29960-2.00009-5.
18. Hansen, T.W., and Wagner, J.B. (2012). Environmental transmission electron microscopy in an aberration-corrected environment. Microsc. Microanal. ***18***: 684–690. https://doi.org/10.1017/S1431927612000293.
19. Reuter, K., and Scheffler, M. (2001). Composition, structure, and stability of RuO2(110) as a function of oxygen pressure. Phys. Rev. B ***65***: 035406. https://doi.org/10.1103/PhysRevB.65.035406.
20. Reuter, K., and Scheffler, M. (2003). Composition and structure of the RuO2(110) surface in an O2 and CO environment: Implications for the catalytic formation of CO2. Phys. Rev. B ***68***: 045407. https://doi.org/10.1103/PhysRevB.68.045407.
21. Opalka, D., Scheurer, C., and Reuter, K. (2019). Ab initio thermodynamics insight into the structural evolution of working IrO2 catalysts in proton-exchange membrane electrolyzers. ACS Catal. ***9***: 4944–4950. https://doi.org/10.1021/acscatal.9b00796.
22. Timmermann, J., Kraushofer, F., Resch, N., et al. (2020). IrO2 surface complexions identified through machine learning and surface investigations. Phys. Rev. Lett. ***125***: 206101. https://doi.org/10.1103/PhysRevLett.125.206101.
23. Lee, Y., Scheurer, C., and Reuter, K. (2022). Epitaxial core-shell oxide nanoparticles: First-principles evidence for increased activity and stability of rutile catalysts for acidic oxygen evolution. ChemSusChem ***15***: e202200015. https://doi.org/10.1002/cssc.202200015.
24. Li, H., and Reuter, K. (2022). Ab initio thermodynamic stability of carbide catalysts under electrochemical conditions. ACS Catal. ***12***: 10506–10513. https://doi.org/10.1021/acscatal.2c01732.
25. Kocer, E., Ko, T.W., and Behler, J. (2022). Neural network potentials: A concise overview of methods. Annu. Rev. Phys. Chem. ***73***: 163–186. https://doi.org/10.1146/annurev-physchem-082720-034254.
26. Rivera de la Cruz, J.G., Sabbe, M.K., and Reyniers, M.-F. (2017). First principle study on the adsorption of hydrocarbon chains involved in Fischer–Tropsch synthesis over iron carbides. J. Phys. Chem. C ***121***: 25052–25063. https://doi.org/10.1021/acs.jpcc.7b05864.
27. Ulissi, Z.W., Tang, M.T., Xiao, J., et al. (2017). Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO2 reduction. ACS Catal. ***7***: 6600–6608. https://doi.org/10.1021/acscatal.7b01648.
28. Sun, S.P., Zhu, J.L., Gu, S., et al. (2019). First principles investigation of the surface stability and equilibrium morphology of MoO3. Appl. Surf. Sci. ***467–468***: 753–759. https://doi.org/10.1016/j.apsusc.2018.10.162.
29. Zuo, E., Dou, X., Chen, Y., et al. (2021). Electronic work function, surface energy and electronic properties of binary Mg-Y and Mg-Al alloys: A DFT study. Surf. Sci. ***712***: 121880. https://doi.org/10.1016/j.susc.2021.121880.
30. Esposito, D. (2018). Mind the gap. Nat. Catal. ***1***: 807–808. https://doi.org/10.1038/s41929-018-0188-0.
31. Brown, I.D. (2009). Recent developments in the methods and applications of the bond valence model. Chem. Rev. ***109***: 6858–6919. https://doi.org/10.1021/cr900053k.
32. Opeyemi Otun, K., Yao, Y., Liu, X., et al. (2021). Synthesis, structure, and performance of carbide phases in Fischer–Tropsch synthesis: A critical review. Fuel ***296***: 120689. https://doi.org/10.1016/j.fuel.2021.120689.
33. Chang, Q., Zhang, C., Liu, C., et al. (2018). Relationship between iron carbide phases (ε-Fe2C, Fe7C3, and χ-Fe5C2) and catalytic performances of Fe/SiO2 Fischer–Tropsch catalysts. ACS Catal. ***8***: 3304–3316. https://doi.org/10.1021/acscatal.7b04085.
34. Kresse, G., and Furthmüller, J. (1996). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys. Rev. B ***54***: 11169–11186. https://doi.org/10.1103/PhysRevB.54.11169.
35. Kresse, G., and Furthmüller, J. (1996). Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput. Mater. Sci. ***6***: 15–50. https://doi.org/10.1016/0927-0256(96)00008-0.
36. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. Phys. Rev. Lett. ***77***: 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.
37. Blöchl, P. (1994). Projector augmented-wave method. Phys. Rev. B ***50***: 17953–17979. https://doi.org/10.1103/PhysRevB.50.17953.
38. Kresse, G., and Joubert, D. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. Phys. Rev. B ***59***: 1758–1775. https://doi.org/10.1103/PhysRevB.59.1758.
39. Ong, S.P., Richards, W.D., Jain, A., et al. (2013). Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. Comput. Mater. Sci. ***68***: 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028.
40. Ma, H., Jiao, Y., Guo, W., et al. (2020). Predicting crystal morphology using a geometric descriptor: A comparative study of elemental crystals with high-throughput DFT calculations. J. Phys. Chem. C ***124***: 15920–15927. https://doi.org/10.1021/acs.jpcc.0c03537.
41. Liu, X.-W., Zhao, S., Meng, Y., et al. (2016). Mössbauer spectroscopy of iron carbides: From prediction to experimental confirmation. Sci. Rep. ***6***: 26184. https://doi.org/10.1038/srep26184.
42. Bisbo, M.K., and Hammer, B. (2020). Efficient global structure optimization with a machine-learned surrogate model. Phys. Rev. Lett. ***124***: 086102. https://doi.org/10.1103/PhysRevLett.124.086102.
43. Ulissi, Z.W., Singh, A.R., Tsai, C., et al. (2016). Automated discovery and construction of surface phase diagrams using machine learning. J. Phys. Chem. Lett. ***7***: 3931–3935. https://doi.org/10.1021/acs.jpclett.6b01254.
44. Ulissi, Z.W., Medford, A.J., Bligaard, T., et al. (2017). To address surface reaction network complexity using scaling relations machine learning and DFT calculations. Nat. Commun. ***8***: 14621. https://doi.org/10.1038/ncomms14621.
45. Mamun, O., Winther, K.T., Boes, J.R., et al. (2020). A Bayesian framework for adsorption energy prediction on bimetallic alloy catalysts. npj Comput. Mater. ***6***: 177. https://doi.org/10.1038/s41524-020-00447-8.
46. Ringe, E., Van Duyne, R.P., and Marks, L.D. (2011). Wulff construction for alloy nanoparticles. Nano Lett. ***11***: 3399–3403. https://doi.org/10.1021/nl2018146.
47. Honkala, K., Hellman, A., Remediakis, I.N., et al. (2005). Ammonia synthesis from first-principles calculations. Science ***307***: 555–558. https://doi.org/10.1126/science.1106435.
48. Davydov, V., Rakhmanina, A., Allouchi, H., et al. (2012). Carbon-encapsulated iron carbide nanoparticles in the thermal conversions of ferrocene at high pressures. Fullerenes, Nanotubes, Carbon Nanostruct ***20***: 451–454. https://doi.org/10.1080/1536383X.2012.655649.
49. Malina, O., Jakubec, P., Kašlík, J., et al. (2017). A simple high-yield synthesis of high-purity Hägg carbide (χ-Fe5C2) nanoparticles with extraordinary electrochemical properties. Nanoscale ***9***: 10440–10446. https://doi.org/10.1039/C7NR02383A.
50. Romanyuk, O., Hattori, K., Someta, M., et al. (2014). Surface structure and electronic states of epitaxial β-FeSi2(100)/Si(001) thin films: Combined quantitative LEED, ab initio DFT, and STM study. Phys. Rev. B ***90***: 155305. https://doi.org/10.1103/PhysRevB.90.155305.
51. Liu, Q.-Y., Shang, C., and Liu, Z.-P. (2021). In situ active site for CO activation in Fe-catalyzed Fischer–Tropsch synthesis from machine learning. J. Am. Chem. Soc. ***143***: 11109–11120. https://doi.org/10.1021/jacs.1c04624.
52. Cuong, L.T., Dung, N.D., Tuan, T.Q., et al. (2018). In situ observation of phase transformation in iron carbide nanocrystals. Micron ***104***: 61–65. https://doi.org/10.1016/j.micron.2017.10.009.
53. Liu, P., Qin, R., Fu, G., et al. (2017). Surface coordination chemistry of metal nanomaterials. J. Am. Chem. Soc. ***139***: 2122–2131. https://doi.org/10.1021/jacs.6b10978.
54. Gu, G.H., Lim, J., Wan, C., et al. (2021). Autobifunctional mechanism of jagged Pt nanowires for hydrogen evolution kinetics via end-to-end simulation. J. Am. Chem. Soc. ***143***: 5355–5363. https://doi.org/10.1021/jacs.0c11261.
55. Huo, C.-F., Li, Y.-W., Wang, J., et al. (2009). Insight into CH4 formation in iron-catalyzed Fischer–Tropsch synthesis. J. Am. Chem. Soc. ***131***: 14713–14721. https://doi.org/10.1021/ja9021864.
56. Chun, H.-J., and Kim, Y.T. (2021). Theoretical study of CO adsorption and activation on orthorhombic Fe7C3(001) surfaces for Fischer–Tropsch synthesis using density functional theory calculations. Energies ***14***: 563. https://doi.org/10.3390/en14030563.
57. Liu, Q.-Y., Shang, C., and Liu, Z.-P. (2022). In situ active site for Fe-catalyzed Fischer–Tropsch synthesis: Recent progress and future challenges. J. Phys. Chem. Lett. ***13***: 3342–3352. https://doi.org/10.1021/acs.jpclett.2c00549.
58. Tajima, S., and Hirano, S.-i. (1990). Synthesis and magnetic properties of Fe7C3 particles with high saturation magnetization. Jpn. J. Appl. Phys. ***29***: 662. https://doi.org/10.1143/JJAP.29.662.
59. Starchikov, S.S., Zayakhanov, V.A., Vasiliev, A.L., et al. (2021). Core@shell nanocomposites Fe7C3/FexOy/C obtained by high pressure-high temperature treatment of ferrocene Fe(C5H5)2. Carbon ***178***: 708–717. https://doi.org/10.1016/j.carbon.2021.03.052.
60. van der Maaten, L.J.P. (2014). Accelerating t-SNE using tree-based algorithms. J. Mach. Learn. Res. ***15***: 3221–3245.

ACKNOWLEDGMENTS

This work was financially supported by the National Science Fund for Distinguished Young Scholars of China (grant no. 22225206), the National Key R&D Program of China (no. 2022YFA1604103), the National Natural Science Foundation of China (nos. 21972157 and 21972160), the CAS Project for Young Scientists in Basic Research (YSBR-005), the Key Research Program of Frontier Sciences CAS (ZDBS-LY-7007), the Major Research Plan of the National Natural Science Foundation of China (92045303), the Informatization Plan of the Chinese Academy of Sciences (grant no. CAS-WX2021SF0110), and the Youth Innovation Promotion Association CAS (2020179). Funding support was also received from the Beijing Advanced Innovation Center for Materials Genome Engineering, Synfuels China Co., Ltd., and the Institute of Coal Chemistry, Chinese Academy of Sciences.

AUTHOR CONTRIBUTIONS

H.M. and Y.J. contributed to the DFT calculations and machine learning work. H.M., W.G., X.L., and X.W. contributed to the data analysis and cowrote the manuscript. W.G., X.L., Y.L., and X.W. supervised the project. All of the authors conceived the research and reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.1016/j.xinn.2024.100571.

LEAD CONTACT WEBSITE

https://sklcc.sxicc.ac.cn/sysgk/syszr/