

# Active machine learning model for the dynamic simulation and growth mechanisms of carbon on metal surface

Received: 16 February 2023

Accepted: 18 December 2023

Published online: 06 January 2024

 Check for updatesDi Zhang<sup>1,2</sup>✉, Peiyun Yi<sup>1,3</sup>, Xinmin Lai<sup>1,3</sup>, Linfa Peng<sup>1,3</sup>✉ & Hao Li<sup>2</sup>✉

Substrate-catalyzed growth offers a highly promising approach for the controlled synthesis of carbon nanostructures. However, the growth mechanisms on dynamic catalytic surfaces and the development of more general design strategies remain ongoing challenges. Here we show how an active machine-learning model effectively reveals the microscopic processes involved in substrate-catalyzed growth. Utilizing a synergistic approach of molecular dynamics and time-stamped force-biased Monte Carlo methods, augmented by the Gaussian Approximation Potential, we perform fully dynamic simulations of graphene growth on Cu(111). Our findings accurately replicate essential subprocesses—from the preferred diffusion of carbon monomer/dimer, chain or ring formations to edge-passivated Cu-aided graphene growth and bond breaks by ion impacts. Extending our simulations to carbon deposition on metal surfaces like Cu(111), Cr(110), Ti(001), and oxygen-contaminated Cu(111), our results align closely with experimental observations, providing a practical and efficient approach for designing metallic or alloy substrates to achieve desired carbon nanostructures and explore further reaction possibilities.

Controllable synthesis of carbon nanomaterials, such as single-crystalline, large-area graphene, specifically chiral carbon nanotubes, or carbon films with certain  $sp^2$  or  $sp^3$  content, is a crucial challenge for realizing their potential applications in future electronics or energy devices<sup>1–3</sup>. The substrate-catalyzed deposition has been considered to be one of the most promising ways to achieve the controllable growth of two or three-dimensional covalently bonded network of carbon atoms<sup>4,5</sup>, because the surface composition and facet identity of the underlying substrate greatly affect their crystallinity<sup>6,7</sup>, orientation<sup>8</sup>, edge geometry<sup>9</sup>, and compressive strain<sup>10</sup>. Apart from the well-explored metallic catalysts like Cu and Ni, there have been reports on the efficient production of wafer-scale single-crystalline graphene on less common substrates such as high-index Cu facets<sup>7</sup>, Cu-Ni alloys<sup>11</sup>,  $Cu_2O$  surface<sup>12</sup>, hydrogen-terminated Ge<sup>13</sup>, and hexagonal

boron nitride<sup>14</sup>. While the growth mechanisms on common surfaces have been extensively studied, there is limited knowledge regarding the dynamic and atomic-level factors that govern the quality of graphene on high-index or composite surfaces, including nucleation and growth kinetics. This research gap significantly hinders the development of theory-guided design approaches for novel catalyzed metal substrates in the growth of carbon nanostructures.

Searching for metallic or alloy catalysts experimentally poses a considerable challenge, primarily due to the extensive range of potential substrates and the sensitivity of the carbon nanomaterial growth process to various experimental parameters. Thus, there is plenty of room for theoretical simulations, where many atomic details are readily available. As a standard approach to study a variety of chemical processes, *ab initio* density functional theory (DFT)

<sup>1</sup>State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, 200240 Shanghai, People's Republic of China. <sup>2</sup>Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan. <sup>3</sup>Shanghai Key Laboratory of Digital Manufacture for Thin-walled Structures, Shanghai Jiao Tong University, 200240 Shanghai, People's Republic of China. ✉e-mail: [zhangdi2015@sjtu.edu.cn](mailto:zhangdi2015@sjtu.edu.cn); [penglinfa@sjtu.edu.cn](mailto:penglinfa@sjtu.edu.cn); [li.hao.b8@tohoku.ac.jp](mailto:li.hao.b8@tohoku.ac.jp)

calculations can compute the step-by-step energy levels of a sequence of states from carbon feedstock, carbon monomer<sup>15</sup>, dimers<sup>16</sup>, or small clusters<sup>17</sup> to final graphene or carbon films on metal<sup>18,19</sup> or alloy surfaces<sup>11</sup>, atomic steps<sup>20–22</sup>, dislocations<sup>4</sup>, grain boundaries<sup>9,23</sup>, and subsurface<sup>24</sup>. Theoretical simulations, such as kinetic Monte Carlo (KMC) and ab initio molecular dynamics (AIMD), have significantly enhanced the understanding of graphene growth on certain metal surfaces. However, it is still challenging to go beyond these atomic details to obtain the fully dynamic and overall picture of carbon growth on an arbitrary metal or alloy substrate. For instance, a DFT-based KMC model<sup>25</sup> was employed to study the diffusion-limited growth of carbon dimers on Cu(111) and Cu(100). Nevertheless, KMC simulations have limitations in capturing the complete dynamics of carbon growth on a dynamically evolving substrate due to the difficulty in constructing event tables for the off-lattice simulations<sup>26</sup>. On the other hand, AIMD suffers from limited time and length scales despite advancements in computational power. Conventional methods such as MD/MC simulations with empirical interatomic potentials face difficulties in accurately calculating interactions between graphitic layers<sup>27</sup>, experimentally observed sp<sup>2</sup>/sp<sup>3</sup> fractions<sup>3</sup>, and the interactions between carbon species and metal surfaces<sup>28</sup> due to the limitations of their fixed functional form. Therefore, there is an ongoing and urgent demand for a robust design model capable of accurately depicting the growth mechanisms of carbon on metallic surfaces.

Machine-learning potentials (MLPs) based on artificial neural networks<sup>29–31</sup> or kernel-based methods<sup>32–34</sup> have been suggested to be a powerful method to address the limited accuracy and transferability of classic force fields and maintain a DFT-level accuracy. This is well demonstrated in the studies on understanding the growth mechanisms of high sp<sup>3</sup> content in tetrahedral amorphous carbon<sup>3,35</sup>, the pressure-induced phase transition in silicon<sup>36</sup>, and complex aqueous systems<sup>37</sup>. Despite these significant achievements in data-driven MD simulations, the construction of an accurate MLP remains a difficult task because it can take years to sample in broad regions of phase space, especially for deposition simulations, in which high-energy incidents can lead to locally strongly disordered structures and the potential should be highly flexible<sup>35</sup>. One solution to this issue is the on-the-fly learning techniques<sup>37,38</sup>, which allow one to sample a small region of configuration space under specific thermodynamic and boundary conditions. In the production of the on-the-fly training set, the selection of the most diverse structures is crucial for achieving high efficiency and accuracy<sup>39</sup>. So far, several selective principles have been proposed to construct a training set with predictive uncertainties, including the Bayesian inference<sup>38,40</sup>, spilling factor<sup>41</sup>, Maxvol method<sup>42</sup>, and CUR decomposition<sup>43</sup>. Notably, the CUR decomposition method stands out because it does not rely on a prior assumption of a Gaussian distribution for potential energy, as is often the case in Bayesian inference. Additionally, with its foundation in a configuration-averaged metric, the CUR decomposition method has consistently shown impressive accuracy and robustness, especially when applied to carbon-based and metal materials<sup>43</sup>. However, relying only on the configuration-averaged metric for selecting new structures during deposition simulation could omit structures that exhibit significant variations only in the local areas surrounding the deposited atom. To enhance the efficiency and effectiveness in the on-the-fly training of deposition processes, a well-defined selection protocol is required. On the other hand, the dynamics of carbon growth on a metal substrate can be governed by important rare events<sup>44</sup>, such as surface diffusion and graphene nucleation. Therefore, how to enhance the training efficiency of MLPs by coupling enhanced sampling methods to classical dynamics needs further study.

This work presents an approach to generate MLPs with minimum human effort through a data-driven automatic learning framework suited for carbon growth on metal or alloy surfaces. To achieve this

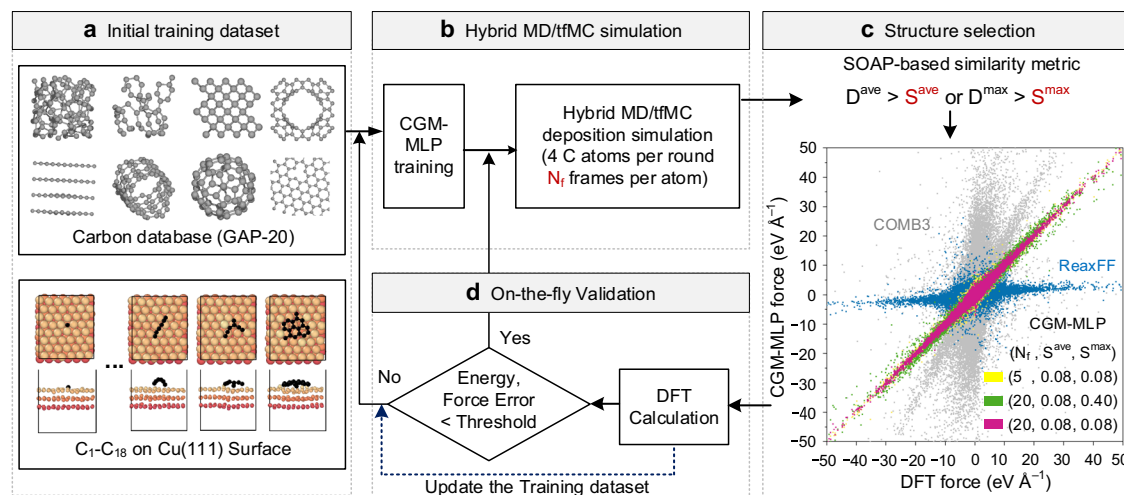
task, we make use of (1) the Gaussian Approximation Potential (GAP) machine learning model, which has been widely validated to capture the deposition process correctly when benchmarked against experiments, including the phonon dispersion relations, thermal expansion, and Raman spectra at different temperatures, and carbon depositions<sup>3,33–35</sup>; (2) an enhanced sampling method named time-stamped force-biased Monte Carlo (tfMC) method<sup>27,45,46</sup> to accelerate the relaxation process after carbon depositions, thus including the important rare events in the training database; (3) an effective strategy for the selection of representative training data based on the descriptor of smooth overlap of atomic positions (SOAP)<sup>47</sup>; (4) a well-developed carbon training set<sup>34</sup>; (5) an automated screening, fitting, and validation procedure. The resulting potential is then applied to study the deposition growth of carbon atoms on a Cu(111) surface. This approach can correctly capture the critical processes of carbon growth on Cu(111), such as the formation and migration of subsurface carbon monomer and surface dimer<sup>5</sup>, the appearance of 1-dimensional carbon nanoarches<sup>18</sup>, graphene nucleation involved with edge-passivated Cu atoms and carbon chains, and the precipitated growth process<sup>4</sup>. Our simulations of the initial nucleation on different metallic surfaces, specifically carbon deposition on Cu(111), Cr(110), Ti(001), and O-contaminated Cu(111), exhibit consistency with experimental observations and DFT calculations. This study provides a practical and efficient methodology for designing metallic or alloy substrates to achieve desired carbon nanostructures and explore further opportunities.

## Results

Figure 1 illustrates the schematic of an active-learning algorithm by which an MLP is generated on the fly during the MD/tfMC simulation of graphene growth on a Cu(111) surface. The carbon-growth-on-metal machine-learning potential is henceforth dubbed CGM-MLP. The construction of CGM-MLP begins with a well-developed database for carbon potential functions generated by Rowe et al.<sup>34</sup>, namely GAP-20, and several equilibrium configurations of carbon clusters on Cu(111) surface (i.e., C<sub>1</sub>–C<sub>18</sub> in Fig. 1a). By including all of these structures, the initial training dataset provides a good starting point for generating an MLP capable of accurately describing Cu-C interactions. The potential energy and forces of the structures in the initial training set are calculated using DFT to obtain the initial iteration of the MLP using the GAP methodology<sup>32</sup>. Subsequently, the generated MLP is employed in a hybrid MD/tfMC simulation to simulate carbon deposition on the Cu(111) surface.

During the MD/tfMC simulations, four carbon atoms are deposited in each round of training. From these simulations, a subset of structures is generated with a sampling rate of N<sub>f</sub> frame per carbon atom. A screening process is then employed, using SOAP-based similarity metrics (see Methods for the definitions) to refine the dataset further. The similarity measure is the heart of the active-learning algorithm, determining whether a new structure should be added to the training set and ensuring a balanced representation of different regions of the phase space. Notably, this work introduces an additional measure,  $D^{\max}$ , which, though resembling  $D^{\text{ave}}$  in form, plays a pivotal role in refining the on-the-fly training set. These measures specifically capture the maximum and average SOAP-based distances between atoms in newly observed and previously chosen structures, respectively.

By considering both maximum and average SOAP distances, we can effectively capture variations in local environments around the deposited atoms, which may be overlooked if only the average distance<sup>43</sup> is considered (see Supplementary Fig. 4). Hence if a new structure's  $D^{\text{ave}}$  or  $D^{\max}$  exceeds the specified screening parameter ( $S^{\max}$  or  $S^{\text{ave}}$ ), it will be included the new structure in the training set. To evaluate the effectiveness of the screening parameters, we selected a subset of structures from the MD/tfMC simulations using different



**Fig. 1 | Schematic illustrations of carbon-growth-on-metal machine-learning potential (CGM-MLP) generated by active learning on-the-fly during hybrid molecular dynamics and time-stamped force-biased Monte Carlo (MD/tfMC) simulations. a** The initial training dataset includes representative carbon structures from Gaussian Approximation Potential (GAP-20)<sup>34</sup> and C<sub>1</sub>-C<sub>18</sub> carbon clusters on Cu(111) surfaces. **b** The CGM-MLP trained from this dataset is then used in a deposition simulation employing a hybrid MD/tfMC method<sup>27</sup>. **c** A smooth overlap of atomic positions (SOAP-based) algorithm is used to select the most

representative structures from the MD/tfMC simulations. The inset figure presents the force correlation plots by using different quality control parameters, namely  $N_f$  (the number of structures sampled for each deposited carbon atom),  $S^{\max}$ , and  $S^{\text{ave}}$  (i.e., the thresholds for the maximum and average SOAP distances,  $D^{\text{ave}}$  and  $D^{\max}$ ). The definitions of the similarity matrix  $D^{\text{ave}}$  and  $D^{\max}$  are available in the “Methods” section. Source data and code are provided. **d** DFT benchmarks energy and force, and if the error is below a threshold, MD/tfMC continues. Otherwise, the training dataset is updated with newly selected structures.

values of  $N_f$ ,  $S^{\max}$ , and  $S^{\text{ave}}$ . Meanwhile, approximately 500 structures were randomly selected from the held-out structures to serve as the testing sets. By increasing  $N_f$  and lowering  $S^{\max}$ , as shown in Fig. 1c, the CGM-MLP exhibits improved force accuracies and a significant improvement compared to classical empirical potentials, such as COMB3<sup>48</sup> and ReaxFF<sup>49</sup>. In Supplementary Fig. 3, we also provide an energy correlation plot and additional tests of the parameter  $S^{\max}$  (Supplementary Fig. 4) for various structure types (Supplementary Table 1). The results demonstrate that a predictive energy mean absolute error ( $E_{\text{MAE}}$ , with the detailed formula provided in Supplementary Method 3) of below  $0.05 \text{ eV atom}^{-1}$  and a force MAE of less than  $0.5 \text{ eV \AA}^{-1}$  are attainable for all relevant structures in a Cu-C system when  $S^{\max}$  is set to less than 0.08 (Supplementary Method 4). Specifically, the energy and force MAE of the CGM-MLP trained with  $N_f = 20$  and  $S^{\text{ave}}, S^{\max} = 0.08$  converge to approximately  $0.013 \text{ eV atom}^{-1}$  and  $0.43 \text{ eV \AA}^{-1}$ , respectively. Given the intrinsic challenges associated with complex hybridizations in carbon and the long-range interactions beyond the MLPs’ cutoff, the energy and force errors we observed are believed to approach the peak accuracy achieved by MLPs for systems containing amorphous or defective carbon<sup>34,50</sup>. A more detailed test of long-range interactions for metal-carbon systems can be found in Supplementary Discussion. 1 and Supplementary Fig. 5. As a practical and efficient means to explore a wide range of reaction pathways during carbon growth, it is believed that the achieved energy and force MAE are appropriate for the research goals. Given the distinct effects of long-range interactions across different metal-carbon systems, strategies such as adopting varying force or energy thresholds for distinct deposition stages, or setting system-dependent total-energy training convergence thresholds, can be considered to further enhance accuracy. Yet, addressing the challenges posed by long-range interactions remains crucial for refining the precision and adaptability of simulations, especially those centered on metal-catalyzed low-density carbon structures.

During the error estimation step (Fig. 1d), we use an energy MAE of  $<0.05 \text{ eV atom}^{-1}$  and a force MAE of  $<0.5 \text{ eV \AA}^{-1}$  to determine the inclusion of newly generated structures in the training set or if carbon atom deposition can proceed without further training. Additional details in the training process can be found in Supplementary Fig. 6.

The decision to end the iterations depends on the desired number of carbon atoms to be deposited on the metal surface. In this study, the number of carbon atoms to be deposited is predetermined as 100 carbon atoms. Additionally, if further growth processes or the simulation of carbon-based films are of interest, the predetermined number of carbon atoms can be increased accordingly.

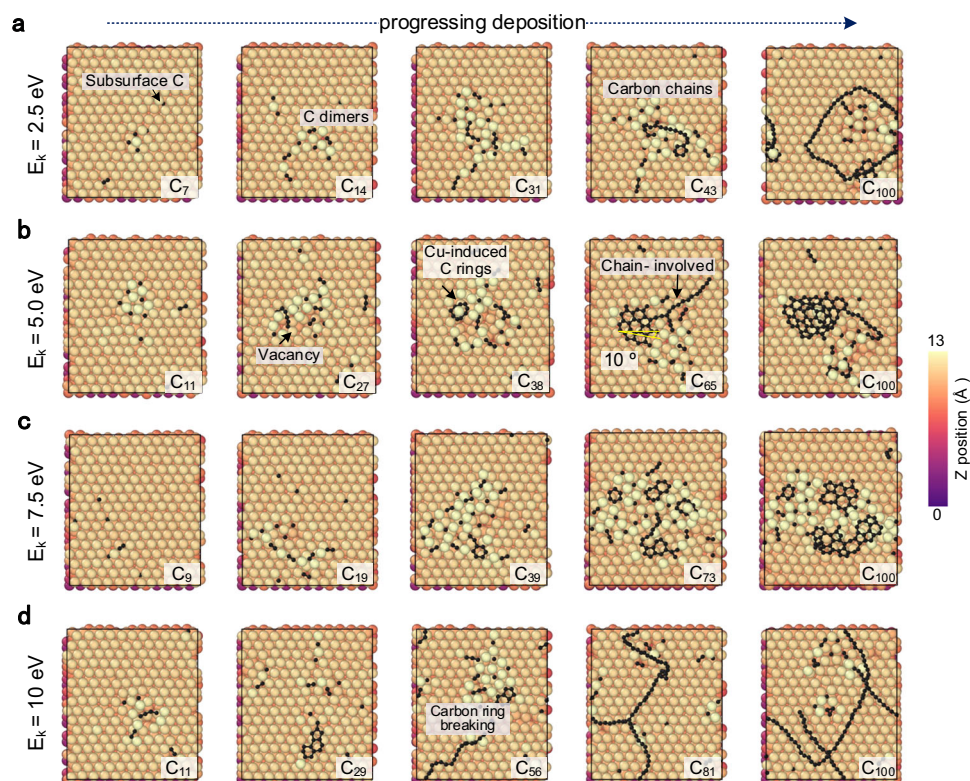
By leveraging the high accuracy of the CGM-MLP and incorporating rare atomistic events in the MD/tfMC method<sup>31</sup>, we successfully replicate essential subprocesses associated with graphene nucleation and carbon growth on metallic surfaces, as exemplified in Fig. 2.

### Carbon monomer and dimer

In the first stage, impinging gas phase carbon atoms readily adsorb on the Cu(111) surface and subsequently diffuse to the sublayer of the Cu(111) surface spontaneously (Fig. 2a, C<sub>7</sub>). This process is consistent with previous static DFT calculations, which suggested that C monomers prefer to remain at an octahedral site in the subsurface layer of Cu(111). The adsorption energy of the subsurface Cu atoms increases by about  $0.5 \text{ eV}$ <sup>15</sup>. In all deposition simulations with different impinging energies, no carbon atom enters the bulk of the Cu substrate in the initial stage. This is consistent with previous findings that carbon exhibits very low solubility in copper<sup>1</sup>. During the ongoing deposition process, subsurface carbon atoms have the potential to intermittently form bonds. Once carbon dimers are formed in the subsurface, they can subsequently migrate back to the surface regions (Fig. 2a, C<sub>14</sub> or Fig. 2b, C<sub>11</sub>). Notably, our simulations indicate that carbon monomers do not always migrate to the subsurface layer. Instead, if the surface copper atoms can absorb the kinetic energy of the incident carbon atoms and transform them into adsorbed copper atoms, vacancies may be created on the Cu(111) surface. In cases where the deposited carbon atoms are near the adsorbed copper atoms, they tend to remain on the surface rather than migrate to the subsurface.

### Carbon-chain formation

Owing to the continued addition of carbon, the extrusive carbon dimers start to connect to the carbon atoms stabilized by adsorbed Cu or off-surface Cu atoms. The short carbon chains appear on Cu(111) surface (Fig. 2a, C<sub>43</sub>). When the incident energy of carbon is as low as



**Fig. 2 | CGM-MLP driven simulations of graphene growth on Cu(111) with different carbon incident kinetic energies ( $E_k$ ).** **a** 2.5 eV, **b** 5.0 eV, **c** 7.5 eV, and **d** 10 eV. Carbon atoms are colored black, and copper atoms are color-coded according to their height coordination. For the labels  $C_x$ ,  $x$  denotes the number of deposited carbon atoms. Different graphitization degrees can be observed when the carbon incident energy varies during the simulations. The fully dynamical

simulations have correctly reproduced many subprocesses, such as the spontaneous diffusion of carbon monomers and dimers<sup>45</sup>, the stabilization effect of adsorbed Cu atoms for carbon rings, the energetically favorable property of carbon chains<sup>18</sup>, and the carbon ring breaking process<sup>18</sup>, achieving excellent agreements with previous DFT studies<sup>19</sup>. To ensure reproducibility, a repeated run is available in Supplementary Fig. 12.

2.5 eV, it is more probable for carbon chains to grow longer and detach from the Cu(111) surface, rather than forming carbon rings on Cu(111) (Fig. 2a,  $C_{100}$ ). An explanation for the preferred form of carbon chains has been furnished by a previous DFT study<sup>18</sup>, which compared the stability of carbon nanoarches and compact carbon nanoislands of equal sizes. Their calculations revealed that the formation of carbon chains was energetically favorable if there were fewer nucleation sites, such as adsorbed Cu atoms and Cu steps. The fully dynamic simulations have reproduced the formation of carbon chains (Fig. 2d) and shown that different carbon chains could share one binding site, as shown in Fig. 2d,  $C_{81}$ .

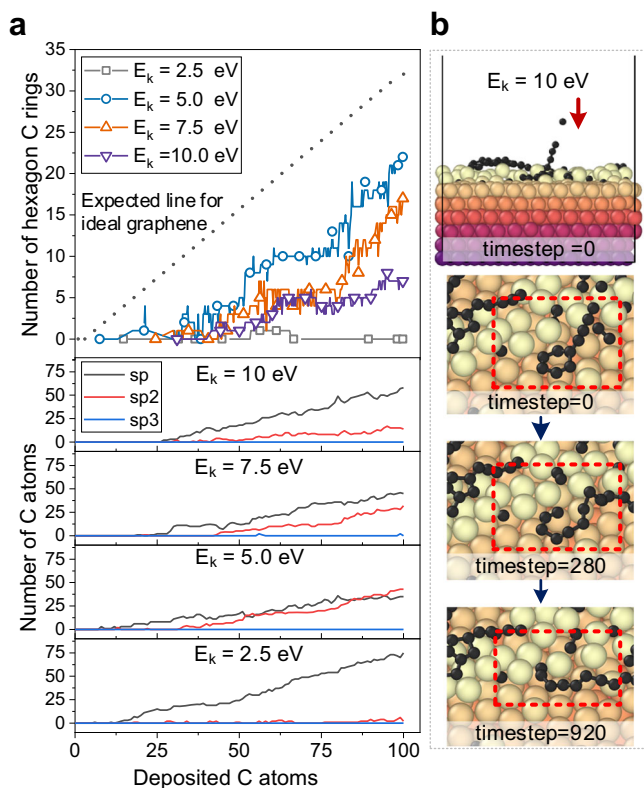
### Carbon ring and graphene island formation

For a moderate irradiation energy, i.e., 5 eV or 7.5 eV, hexagonal carbon rings can be observed at the initial stage of the deposition simulations, as shown in Fig. 2b,  $C_{38}$  and Fig. 2c,  $C_{39}$ . Due to the more vigorous bombardment of carbon atoms, the number of off-surface Cu atoms may increase. The presence of adsorbed copper atoms or Cu-C bridges surrounding the initial hexagonal carbon rings has been previously observed in STM images<sup>52</sup>. This observation aligns well with experimental findings<sup>53</sup> that the initial carbon rings tend to nucleate at surface steps or impurities on Cu(111). Furthermore, it is consistent with experiments showing that well-ordered graphene structures only form above 790 °C, attributed to the motion of Cu steps induced by sublimation during growth<sup>22</sup>. Our research has also demonstrated, through dynamic simulations (the animation is available in Supplementary Movie 1), the significance of surface Cu steps in graphene growth. We also show that the critical incident energy of creating a graphene nucleation site on a flat Cu(111) surface may be within 5–7.5 eV.

Our simulations also reveal that the thermodynamics and kinetics analyses of surface growth of graphene on Cu(111) should take into account the involvement of adsorbed Cu atoms, such as C-Cu bridges<sup>15</sup>, Cu atoms passivating graphene edges, and edged carbon chains (Fig. 2b,  $C_{65}$ ). In addition, the preferred growth orientation of graphene growth was also reproduced by our simulation model. As shown in Fig. 2b,  $C_{65}$ , we measured the growth orientation of the final graphene island by using the same benchmark angle proposed by Ding and co-workers<sup>19</sup>, and the angle was measured as about 10°, at which the potential energy of graphene rotating on the Cu(111) surface reached a minimum<sup>19</sup>. As highlighted by researchers<sup>25,54</sup>, traditional KMC simulations on static metal surfaces may overlook crucial reaction processes, and AIMD simulations often have limited timescales to capture the complete reaction pathway involving Cu atoms. Previous COMB3 and ReaxFF potentials were unable to correctly simulate the growth of graphene on Cu(111) (see Supplementary Discussion. 3, Supplementary Fig. 9, and Supplementary Fig. 10). This is precisely why the CGM-MLP-based MD/tfMC method emerges as the most promising and efficient approach to elucidate the growth mechanism of carbon on metallic surfaces.

### Carbon ring breaking

To gain atomistic insight into the observed suppression of graphene island formation at an incident energy of 10 eV in Fig. 2d, we performed further analyses including the evolution of hexagonal carbon rings and hybridization analysis on the carbon atoms deposited on Cu(111) surfaces, as shown in Fig. 3a. At an incident energy of 2.5 eV, the deposited carbon atoms primarily form carbon chains, resulting in a higher proportion of  $sp^2$ -hybridized carbon compared to  $sp^2$ - or  $sp^3$ -carbon. With increasing incident energy, the content of  $sp^2$ -



**Fig. 3 | Carbon structure analysis and observation of carbon ring breakage by high-energy bombardments.** **a** Evolution of hexagonal carbon rings and hybridization analysis with increasing deposited carbon atoms. Coordination numbers with a 1.8 Å cutoff radius for the nearest neighbors were used to determine the hybridization orbitals. Source data are provided as a Source Data file.  $E_k$  is the incident kinetic energy of carbon atoms. The maximum number of hexagonal carbon rings and optimal sp<sup>2</sup>-content are achieved at a moderate incident energy of 5 eV. Lower or higher incident energies hinder graphene formation due to insufficient adsorbed Cu atoms or carbon ring breakage, respectively. **b** Our simulations capture the transformation from carbon rings to chains, as highlighted by the dashed red boxes.

carbon initially increases and then decreases, indicating that high-energy bombardments can induce more nucleation sites for stabilizing carbon rings. Previous DFT studies<sup>18</sup> also indicated that without the adsorbed copper atoms to stabilize the hexagonal carbon rings, they would change to linear chains easily. To provide direct evidence of the process, Fig. 3b presents the transformation from carbon ring fragmentation into carbon chains when subjected to high-energy bombardments. This transformation begins with a carbon ring on a Cu(111) surface being subjected to a 10 eV carbon atom bombardment (timestep 0). Subsequent images illustrate the gradual dissociation of carbon rings and the formation of carbon chains resulting from the impact of high-energy carbon atoms. This observation aligns with the findings in Fig. 3a, suggesting that high-energy bombardment impedes the formation of carbon rings and, consequently, the nucleation of graphene. Additionally, we have included an animated demonstration in Supplementary Movie 2, showcasing the powerful capability of the MLP-based MD/tfMC method in capturing dynamic atomistic processes.

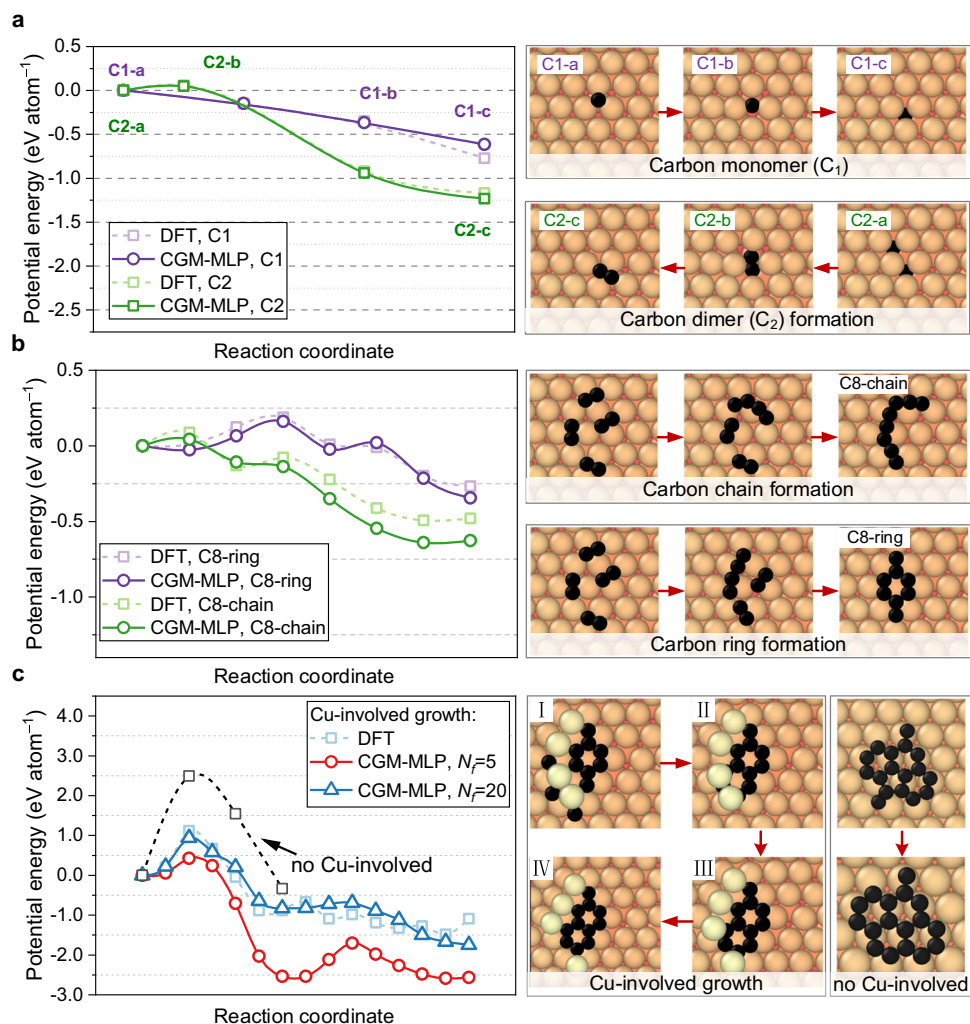
The CGM-MPL enables us to perform a fully dynamic simulation of graphene growth on the metal surface and then allows us to comprehensively understand possible growth mechanisms on a dynamical catalytic surface instead of a stagnant catalytic surface. Based on the growth animations obtained from our simulations, some new insights on how the copper atoms get involved in the growth of graphene on

Cu(111) surfaces can be provided, offering possible reaction paths for the following energetic analysis.

In Fig. 4, we proposed a growth model involving the passivation of Cu atoms to the graphene edge and employed the DFT-based, or CGM-MPL-based climbing-image nudged elastic band (CI-NEB) method to demonstrate the accuracy of the machine-learning potentials in thermodynamically or kinetically. As shown in Fig. 4a, the C monomer prefers to remain at an octahedral site in the sub-surface layer. Compared to C adatom on the surface, the DFT and CGM-MPL predict an increase in the binding energy of about 0.5 eV<sup>15</sup>. Then, combining two monomers, i.e., forming a carbon dimer (C2-a – C2-c), overcomes a small energy barrier of only 0.06 eV atom<sup>-1</sup> (0.12 eV in total) and releases 1.15 eV atom<sup>-1</sup>. These energies obtained from the CGM-MPL are in excellent agreement with those from DFT calculations<sup>15</sup>. Next, in Fig. 4b, we studied the two possible reaction paths from four carbon dimers to a C8 chain or a C8 ring. There is less than 0.1 eV atom<sup>-1</sup> energy barrier (0.7 eV in total) during the carbon-chain reaction process, which releases an energy of 0.5 eV atom<sup>-1</sup> obtained from the DFT calculation. In the ring-reaction path, there is a larger energy barrier compared with the carbon-chain path, i.e., approximately 0.2 eV atom<sup>-1</sup> (1.5 eV in total), and this ring-forming reaction releases 0.25 eV atom<sup>-1</sup> obtained from DFT. From a kinetic standpoint, carbon-chain formation is more favorable than carbon-ring formation, as the energy barrier for carbon-ring formation is approximately twice as high. From a thermodynamic perspective, the carbon-chain reaction process releases an energy of 0.5 eV atom<sup>-1</sup> while the ring-forming reaction releases 0.25 eV atom<sup>-1</sup>. The higher energy release in the carbon-chain reaction suggests greater stability and a lower energy state of the chain configuration. Therefore, the spontaneous formation of carbon rings on a flat Cu(111) surface is rarely observed due to the unfavorable thermodynamics and kinetics of graphitic ring formation.

To stabilize the hexagonal carbon ring, transforming surface copper atoms to adsorbed copper atoms is suggested to be necessary during the growth of graphene on the Cu(111) surface. At high deposition temperatures or under energetic carbon bombardments, the surface Cu atoms can quickly move to an adatom site, offering the initial nucleation site for hexagonal carbon rings. Furthermore, it is observed that the binding energy of initial graphene clusters increases with the addition of passivated Cu atoms (Supplementary Discussion 2 and Supplementary Fig. 7). Therefore, the growth of graphene on a flat Cu(111) is supposed to start from the simple structure shown in Fig. 4c, I. Firstly, the deposited carbon atoms migrate to the other side of adsorbed Cu atoms, which offer a nucleation site for graphene and form the metal bridge<sup>52</sup>. Then, the following snapshots from Fig. 4c, II to IV, show a repeatable cycle of incorporating three carbon atoms onto pristine graphene edges. It can be seen in Fig. 4c that the whole process only has a small energy barrier of less than 0.95 eV atom<sup>-1</sup> (2.82 eV in total), which is much less than the energy barrier without the incorporation of the Cu atoms (2.47 eV atom<sup>-1</sup>)<sup>20</sup>. In Supplementary Fig. 8, to ensure a transparent and comprehensive error assessment, we also provided a comparison of energy barriers as calculated by DFT and CGM-MPL.

The differences between the potential energies along the reaction coordinates calculated by CGM-MPL (solid lines) and DFT (dashed lines) are generally below 0.2 eV atom<sup>-1</sup>. This level of accuracy represents a significant improvement compared to classical empirical potentials. Although there are inherent limitations of CGM-MPL, such as energy errors due to cutoff radius and the complexity of carbon interactions, the potential reaction pathways can be further identified and refined by the subsequent DFT-NEB calculations. In addition, Fig. 4c demonstrates the effect of enhancing the sampling rate  $N_f$  on the inclusion of more transition states in the training set, resulting in energy barriers that approach DFT calculations.



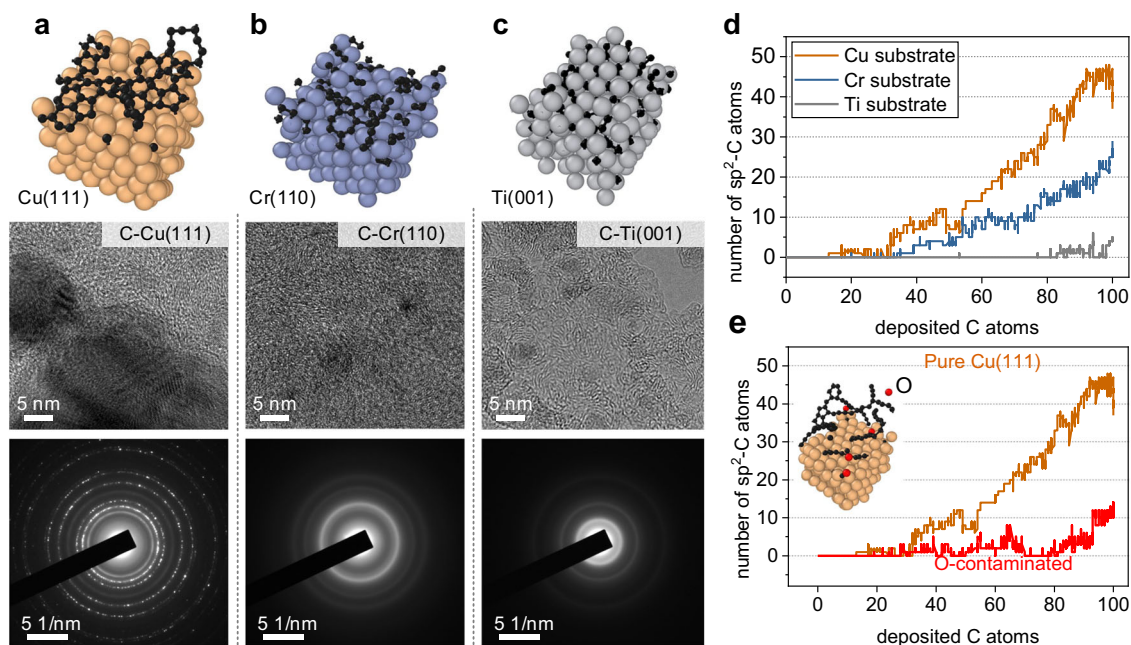
**Fig. 4 | Minimum energy paths of carbon diffusion and graphene nucleation obtained using carbon-growth-on-metal machine-learning potential (CGM-MLP) and DFT-based climbing-image nudged elastic band (CI-NEB) calculations. a** Monomer and dimer diffusion. C<sub>x</sub>-a/b/c indicates the correspondence between the energy points on the left side and the structures on the right side. **b** Conversion from dimers to carbon chains or rings, and **c** Graphene growth with

passivated edges by Cu atoms. Carbon atoms (black), Cu atoms (orange), and adsorbed Cu atoms (wheat). Incorporating edge-passivated Cu atoms reduces the energy barrier for graphene growth (0.95 eV atom<sup>-1</sup>) compared to pristine edges (2.47 eV atom<sup>-1</sup>, the dashed line)<sup>20</sup>. N<sub>f</sub> is the number of structures sampled for each deposited carbon atom. Source data are provided as a Source Data file.

To demonstrate the transferability of our CGM-MLP training framework, we extended our study by training additional MLPs, namely Cr-C and Ti-C MLPs, as well as considering the C-Cu system with surface oxygen contamination. Using these CGM-MLPs, we conducted simulations of carbon deposition on metallic surfaces, as depicted in Fig. 5. Our simulation results reveal that compared to carbon growth on Cu(111), fewer carbon rings are observed on the Cr(110) surfaces, and almost no carbon rings are observed on Ti(001). To validate our simulations, we employed a magnetron sputtering system to deposit approximately 30 nm of carbon film on Cu, Cr, and Ti surfaces. High-resolution transmission electron microscopy (HRTEM) images and selected area electron diffraction (SAED) patterns in Fig. 5 show that the carbon film deposited on Cu(111) exhibits the highest crystalline degree, followed by the Cr(110) surface, while the catalytic effect on the Ti(001) surface is the weakest, matching well with the simulations. It should be noted that the diffraction rings observed in the HRTEM and SAED images may arise from both the metal substrate and the carbon films. However, the metallic layers used in this study are quite thin, with an approximate thickness of 10 nm. Moreover, based on the surface morphologies, it is believed

that the catalytic effect on the formation of nanocrystalline carbon follows the order of Cu, Cr, and Ti. Generally, under the same deposition temperature and energy, the crystalline degree of carbon films mainly depends on the initial nucleation rate and the metal-carbon interface. Our simulation results demonstrate that the initial nucleation rate (Fig. 5d) of these three metals follows the order of Cu, Cr, and Ti, which is consistent with experimental observations and previous DFT calculations<sup>6</sup>.

Using the Cu-C-O MLP-based MD/tfMC method, we performed deposition simulations of carbon on O-contaminated Cu(111). Detailed growth processes and discussions are available in Supplementary Fig. 11. After depositing 100 carbon atoms, the absence of oxygen contamination led to the formation of small graphene islands on Cu(111) (Fig. 5a). However, in the presence of oxygen, the final structure consisted of cross-linked carbon chains (Fig. 5e inset). Therefore, the observed C-Cu-O bridges and terminated oxygen atoms (Supplementary Fig. 11) significantly reduced the initial nucleation rate of graphene on Cu(111), which is in agreement with previous experimental observations<sup>35</sup>. More discussions can be found in Supplementary Discussion 4.



**Fig. 5 | Representative metallic surfaces for the growth of carbon nanostructures.** **a** pure Cu(111), **b** Cr(110), and **c** Ti(001) surface. Below each surface, High-resolution transmission electron microscopy (HRTEM) images and selected area electron diffraction (SAED) images of carbon nanostructures prepared by

magnetron sputtering deposition are provided. **d** The number of  $sp^2$ -C as a function of deposited carbon atoms on different metal substrates and **e** O-contaminated Cu(111). Source data are provided as a Source Data file.

## Discussion

In summary, this study represents a pioneering advancement in integrating MLPs and MD/tfMC, offering a transferable and efficient strategy for designing metallic or alloy substrates to achieve desired carbon nanostructures. The CGM-MLPs effectively combine the accuracy of first-principles methods with the efficiency of classical force fields. Moreover, the tfMC method overcomes the timescale limitations of traditional AIMD or classic MD approaches. Among various MC methods, tfMC stands out as a trajectory-quasi approach for extending the timescale in simulating off-lattice structures. In addition, the automatic training framework of CGM-MLPs incorporates a specialized query strategy for constructing an on-the-fly training set in deposition simulations, emphasizing the significance of considering the local environment surrounding the deposited atoms.

These developments have enabled the straightforward theoretical study of carbon growth mechanisms on complex metal surfaces. We demonstrate the efficacy of our model through its application to the well-established system of graphene growth on a Cu(111) surface, showcasing high accuracy compared to previous static DFT predictions and experimental observations. Our simulations reproduce the microscopic physical processes correctly and efficiently providing a wide range of possible reaction pathways. Several critical processes and physical insights into how Cu adatoms stabilize graphene nucleation and how graphene grows with its edge passivated by Cu atoms or carbon chains are revealed. DFT-based CI-NEB calculations validate the model's high accuracy in thermodynamics and kinetics, presenting low-energy-barrier processes for graphene growth on Cu(111). Additionally, to demonstrate transferability, we expand our investigation by training additional carbon-metal MLPs, specifically Cr-C and Ti-C MLPs, while also considering the C-Cu system with surface oxygen contamination. HRTEM and SEAD observations confirm the simulated catalytic effect on nucleation and the formation of nanocrystalline carbon on Cu, Cr, and Ti. The machine-learning-driven deposition models presented in this study may open up opportunities for investigating multi-element metallic or alloy substrates in the

growth of diverse carbon nanostructures, such as graphene, CNTs, graphite- or diamond-like carbon films.

## Methods

### Hybrid molecular dynamics and time-stamped force-biased Monte Carlo methods

The hybrid MD/tfMC simulations were implemented in the large-scale atomic/molecular massively parallel simulator (LAMMPS)<sup>56</sup> with a plugin of the Quantum Mechanics and Interatomic Potentials (QUIP) software package, which can be found at <https://github.com/libAtoms/QUIP>. Carbon atoms were deposited along the  $-Z$  direction on the metal substrate. Periodic boundary conditions were applied to the  $X$ - and  $Y$ - directions. To avoid energy recycling through the periodic boundaries, a 2 Å thick velocity rescaling wall with a lateral displacement greater than 6 Å from the initial position of each incident atom was applied to the substrate<sup>51</sup> (see Supplementary Fig. 1 for more details). The deposition simulations consisted of a cycle of MD and tfMC methods. Carbon atoms with varying kinetic energies were deposited during the MD simulations, followed by a 2 ps NVT MD simulation with a timestep of 0.5 fs. The 2 ps duration was sufficient to reach equilibrium, as evidenced by the temperature and total energy of the system remaining constant over an observable time.

During tfMC simulations, in each step, all atoms in the selected group are displaced using the stochastic tfMC algorithm<sup>46</sup>. However, the tfMC method differs from MD algorithms by employing a force-bias probabilistic description of atomic motion (see Supplementary Method 1). Two critical parameters in tfMC are the temperature  $T$  and the maximum allowed displacement  $\Delta$ . The parameter  $\Delta$  is determined through a series of simulation tests at different deposition temperatures, following the criterion proposed by Timonova et al.<sup>46</sup>. This criterion ensures that a perfect crystal remains perfect after tfMC simulation and a short MD equilibration. The parameter values  $\Delta = 0.18$  Å and  $T = 573$  K were carefully determined and utilized in our previous work<sup>27</sup> and were also employed in this study.

## Active machine-learning-driven modeling of carbon growth on the metal surface

The MLP training starts from a training set including numerous carbon allotropes from an accurate and transferable machine-learning potential for carbon<sup>34</sup> and simple carbon clusters (C<sub>1</sub>–C<sub>18</sub>) on Cu(111) surface. The GAP fitting in this work also uses a 2-body (2b), 3-body (3b) SOAP descriptor to describe the nature of bonding interactions for carbon and copper atoms. A GAP-fitted 2b component with a cutoff of 3.7 Å is used as the 2-body part of our model. A 3b descriptor and the SOAP descriptor represent the many-body contributions to the potential energy. The 3b term is a symmetrized transformation of the Cartesian coordinates of triplets of atoms with a cutoff of 2.5 Å. In the construction of the SOAP descriptor, an expansion of the neighbor density up to  $l_{\max} = 4$ ,  $n_{\max} = 12$ , a cutoff of 3.7 Å,  $\sigma_{\text{force}} = 0.01 \text{ eV \AA}^{-1}$ ,  $\sigma_{\text{energy}} = 0.001 \text{ eV}$ , and  $\zeta = 4$ , which are systematically tested for the optimization of the GAP model previously<sup>34</sup> and full details of hyper-parameters can also be found in Supplementary Method 3 and Supplementary Fig. 2.

Throughout the training procedure, we introduce two SOAP-based screening parameters,  $D^{\max}$  and  $D^{\text{ave}}$ , specifically designed to optimize the training set. Their respective equations are presented below:

$$D^{\text{ave}} = \text{ave}_{i < N_1} \left( \min_{j < N_0} (|\vec{A}(a_i) - \vec{A}(a_j)|) \right) \quad (1)$$

$$D^{\max} = \max_{i < N_1} \left( \min_{j < N_0} (|\vec{A}(a_i) - \vec{A}(a_j)|) \right) \quad (2)$$

where  $\mathbf{A}(a_i)$  and  $\mathbf{A}(a_j)$  represent the SOAP vectors of the atom  $i$  and  $j$  position ( $\mathbf{a}_i$  and  $\mathbf{a}_j$ ), respectively. The  $\|\cdot\|$  denotes the Euclidean distance between the two SOAP vectors. The source code for computing  $D^{\max}$  and  $D^{\text{ave}}$  is available at [https://github.com/sjtudizhang/CGM-MLP/Calculate\\_similarity.py](https://github.com/sjtudizhang/CGM-MLP/Calculate_similarity.py)<sup>57</sup>. First, the environment fingerprint of each particle is encoded using the SOAP descriptor. The Euler distance between SOAP vectors is calculated as a similarity measure between atoms from different structures. The minimum Euler distance between the particles in one structure and all particles in the other structure is determined to define the similarity measure between the two structures. Average ( $D^{\text{ave}}$ ) and maximum ( $D^{\max}$ ) values of this minimum Euler distance are calculated for different atoms. Configurations are selected based on the condition that both  $D^{\text{ave}}$  and  $D^{\max}$  with previously selected configurations exceed the screen parameters ( $S^{\text{ave}}$  and  $S^{\max}$ ). Detailed tests of the screen parameters can be found in Supplementary Fig. 3 and Supplementary Fig. 4.

### DFT and CI-NEB calculations

All the energies and forces of the structures in the training database are calculated by the QUICKSTEP scheme in CP2K<sup>58,59</sup>. Electronic wave functions were described at the Gamma point using a mixed-basis scheme with Goedecker–Teter–Hutter (GTH) pseudopotentials<sup>60</sup> and a cutoff and the relative cutoff energy of 300 Ry and 60 Ry. Shorter range Double- $\zeta$  quality basis sets optimized for GTH Perdew–Burke–Ernzerhof (PBE)<sup>61</sup> pseudopotential were used. Dispersion corrections are included by the well-established Grimme’s D3 method<sup>62</sup>. For CI-NEB calculations, we used a total of 4 or 8 replica geometries along the path with a spring constant of 0.001 to restrain the replicas. Other details can be found in Supplementary Method 2.

### Carbon film preparation and characterizations

Carbon films were prepared using a Miba-Teer UDP850 closed-field unbalanced magnetron sputtering system. Metallic seed layers and

carbon films were directly deposited onto TEM Cu grids for TEM analysis. The vacuum chamber was evacuated to less than  $4.0 \times 10^{-3}$  Pa. Cu, Cr, or Ti seed layers were deposited on Cu grids using a power density of approximately  $2.2 \text{ W cm}^{-2}$  for 2 min. Carbon films were then deposited on the metallic seed layers using a power density of  $3.6 \text{ W cm}^{-2}$  on two graphite targets for 5 min. The carbon films had a thickness of approximately 30 nm. The deposition temperature was maintained at 300 °C, and substrate bias voltages were disabled to avoid ion bombardments. HRTEM images were obtained using a JEOL 2100 F device at an acceleration voltage of 200 kV.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All of the final training datasets are available at <https://github.com/sjtudizhang/CGM-MLP><sup>57</sup>. Source data and molecular dynamics trajectories generated in this study have been deposited in the figshare database<sup>63</sup> under accession code: <https://doi.org/10.6084/m9.figshare.24591774.v2>.

### Code availability

The training code for CGM-MLP is available from the website <https://libatoms.github.io/GAP/>. The GAP fitting Shell script, Python code for calculating the structural similarity, and C++ code for carbon ring analyses are available at <https://github.com/sjtudizhang/CGM-MLP><sup>57</sup>.

### References

- Li, X. et al. Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
- Gomez-Gualdrón, D. A., McKenzie, G. D., Alvarado, J. F. J. & Balbuena, P. B. Dynamic evolution of supported metal nanocatalyst/carbon structure during single-walled carbon nanotube growth. *ACS Nano* **6**, 720–735 (2012).
- Caro, M. A., Deringer, V. L., Koskinen, J., Laurila, T. & Csányi, G. Growth mechanism and origin of high sp<sup>3</sup> content in tetrahedral amorphous carbon. *Phys. Rev. Lett.* **120**, 166101 (2018).
- Zhang, X., Li, H. & Ding, F. Self-assembly of carbon atoms on transition metal surfaces—chemical vapor deposition growth mechanism of graphene. *Adv. Mater.* **26**, 5488–5495 (2014).
- Riikonen, S., Krashennikov, A. V., Halonen, L. & Nieminen, R. M. The role of stable and mobile carbon adspecies in copper-promoted graphene growth. *J. Phys. Chem. C* **116**, 5802–5809 (2012).
- Li, X., Li, L., Zhang, D. & Wang, A. Ab initio study of interfacial structure transformation of amorphous carbon catalyzed by Ti, Cr, W transit. layers. *ACS Appl. Mater. Interfaces* **9**, 41115–41119 (2017).
- Wu, M. et al. Seeded growth of large single-crystal copper foils with high-index facets. *Nature* **581**, 406–410 (2020).
- Hayashi, K., Sato, S., Ikeda, M., Kaneta, C. & Yokoyama, N. Selective graphene formation on copper twin crystals. *J. Am. Chem. Soc.* **134**, 12492–12498 (2012).
- Murdock, A. T. et al. Controlling the orientation, edge geometry, and thickness of chemical vapor deposition graphene. *ACS Nano* **7**, 1351–1359 (2013).
- Li, B.-W. et al. Orientation-dependent strain relaxation and chemical functionalization of graphene on a Cu(111) foil. *Adv. Mater.* **30**, 1706504 (2018).
- Wu, T. et al. Fast growth of inch-sized single-crystalline graphene from a controlled single nucleus on Cu–Ni alloys. *Nat. Mater.* **15**, 43–47 (2016).



12. Zhou, H. et al. Chemical vapour deposition growth of large single crystals of monolayer and bilayer graphene. *Nat. Commun.* **4**, 2096 (2013).
13. Lee, J.-H. et al. Wafer-scale growth of single-crystal monolayer graphene on reusable hydrogen-terminated germanium. *Science* **344**, 286–289 (2014).
14. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
15. Li, P. & Li, Z. Theoretical insights into the thermodynamics and kinetics of graphene growth on copper surfaces. *J. Phys. Chem. C* **124**, 16233–16247 (2020).
16. Chen, H., Zhu, W. & Zhang, Z. Contrasting behavior of carbon nucleation in the initial stages of graphene epitaxial growth on stepped metal surfaces. *Phys. Rev. Lett.* **104**, 186101 (2010).
17. Yuan, Q. et al. Magic carbon clusters in the chemical vapor deposition growth of graphene. *J. Am. Chem. Soc.* **134**, 2970–2975 (2012).
18. Wesep, R. G. V., Chen, H., Zhu, W. & Zhang, Z. Communication: stable carbon nanoarches in the initial stages of epitaxial growth of graphene on Cu(111). *J. Chem. Phys.* **134**, 171105 (2011).
19. Zhang, X., Xu, Z., Hui, L., Xin, J. & Ding, F. How the orientation of graphene is determined during chemical vapor deposition growth. *J. Phys. Chem. Lett.* **3**, 2822–2827 (2012).
20. Shu, H., Chen, X., Tao, X. & Ding, F. Edge structural stability and kinetics of graphene chemical vapor deposition growth. *ACS Nano* **6**, 3243–3250 (2012).
21. Artyukhov, V. I., Liu, Y. & Yakobson, B. I. Equilibrium at the edge and atomistic mechanisms of graphene growth. *Proc. Natl Acad. Sci. USA* **109**, 15136–15140 (2012).
22. Wofford, J. M., Nie, S., McCarty, K. F., Bartelt, N. C. & Dubon, O. D. Graphene islands on Cu Foils: the interplay between shape, orientation, and defects. *Nano Lett.* **10**, 4890–4896 (2010).
23. Rasool, H. I. et al. Continuity of graphene on polycrystalline copper. *Nano Lett.* **11**, 251–256 (2011).
24. Xu, J. & Saeys, M. First principles study of the stability and the formation kinetics of subsurface and bulk carbon on a Ni catalyst. *J. Phys. Chem. C* **112**, 9679–9685 (2008).
25. Qiu, Z., Li, P., Li, Z. & Yang, J. Atomistic simulations of graphene growth: from kinetics to mechanism. *Acc. Chem. Res.* **51**, 728–735 (2018).
26. Dong, J., Zhang, L., Wu, B., Ding, F. & Liu, Y. Theoretical study of chemical vapor deposition synthesis of graphene and beyond: challenges and perspectives. *J. Phys. Chem. Lett.* **12**, 7942–7963 (2021).
27. Zhang, D., Peng, L., Li, X., Yi, P. & Lai, X. Controlling the nucleation and growth orientation of nanocrystalline carbon films during plasma-assisted deposition: a reactive molecular dynamics/Monte Carlo study. *J. Am. Chem. Soc.* **142**, 2617–2627 (2020).
28. Klaver, T. P. C., Zhu, S.-E., Sluiter, M. H. F. & Janssen, G. C. A. M. Molecular dynamics simulation of graphene on Cu (100) and (111) surfaces. *Carbon* **82**, 538–547 (2015).
29. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
30. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).
31. Khaliullin, R. Z., Eshet, H., Kühne, T. D., Behler, J. & Parrinello, M. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nat. Mater.* **10**, 693–697 (2011).
32. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
33. Deringer, V. L. & Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B* **95**, 094203 (2017).
34. Rowe, P., Deringer, V. L., Gasparotto, P., Csányi, G. & Michaelides, A. An accurate and transferable machine learning potential for carbon. *J. Chem. Phys.* **153**, 034702 (2020).
35. Caro, M. A., Csányi, G., Laurila, T. & Deringer, V. L. Machine learning driven simulated deposition of carbon films: From low-density to diamondlike amorphous carbon. *Phys. Rev. B* **102**, 174201 (2020).
36. Deringer, V. L. et al. Origins of structural and electronic transitions in disordered silicon. *Nature* **589**, 59–64 (2021).
37. Schran, C. et al. Machine learning potentials for complex aqueous systems made simple. *Proc. Natl Acad. Sci. USA* **118**, e2110077118 (2021).
38. Jinnouchi, R., Miwa, K., Karsai, F., Kresse, G. & Asahi, R. On-the-fly active learning of interatomic potentials for large-scale atomistic simulations. *J. Phys. Chem. Lett.* **11**, 6946–6955 (2020).
39. Botu, V., Batra, R., Chapman, J. & Ramprasad, R. Machine learning force fields: construction, validation, and outlook. *J. Phys. Chem. C* **121**, 511–522 (2017).
40. Jinnouchi, R., Lahnsteiner, J., Karsai, F., Kresse, G. & Bokdam, M. Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with bayesian inference. *Phys. Rev. Lett.* **122**, 225701 (2019).
41. Miwa, K. & Ohno, H. Molecular dynamics study on  $\beta$ -phase vanadium monohydride with machine learning potential. *Phys. Rev. B* **94**, 184109 (2016).
42. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
43. Bernstein, N., Csányi, G. & Deringer, V. L. De novo exploration and self-guided learning of potential-energy surfaces. *npj Comput. Mater.* **5**, 99 (2019).
44. Vandermause, J. et al. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput. Mater.* **6**, 20 (2020).
45. Neyts, E. C., Shibuta, Y., van Duin, A. C. T. & Bogaerts, A. Catalyzed growth of carbon nanotube with definable chirality by hybrid molecular dynamics–force biased Monte Carlo simulations. *ACS Nano* **4**, 6665–6672 (2010).
46. Timonova, M., Groenewegen, J. & Thijssse, B. J. Modeling diffusion and phase transitions by a uniform-acceptance force-bias Monte Carlo method. *Phys. Rev. B* **81**, 144107 (2010).
47. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
48. Liang, T. et al. Classical atomistic simulations of surfaces and heterogeneous interfaces with the charge-optimized many body (COMB) potentials. *Mater. Sci. Eng.: R: Rep.* **74**, 255–279 (2013).
49. Monti, S., Li, C. & Carravetta, V. Reactive dynamics simulation of monolayer and multilayer adsorption of glycine on Cu(110). *J. Phys. Chem. C* **117**, 5221–5228 (2013).
50. Shaidu, Y. et al. A systematic approach to generating accurate neural network potentials: the case of carbon. *npj Comput. Mater.* **7**, 52 (2021).
51. Marks, N. A., Cover, M. F. & Kocer, C. Simulating temperature effects in the growth of tetrahedral amorphous carbon: the importance of infrequent events. *Appl. Phys. Lett.* **89**, 131924 (2006).
52. Wu, P., Zhang, W., Li, Z., Yang, J. & Hou, J. G. Communication: coalescence of carbon atoms on Cu (111) surface: Emergence of a stable bridging-metal structure motif. *J. Chem. Phys.* **133**, 071101 (2010).
53. Nie, S., Wofford, J. M., Bartelt, N. C., Dubon, O. D. & McCarty, K. F. Origin of the mosaicity in graphene grown on Cu(111). *Phys. Rev. B* **84**, 155425 (2011).
54. Bhowmik, S., Govind & Rajan, A. Chemical vapor deposition of 2D materials: a review of modeling, simulation, and machine learning studies. *iScience* **25**, 103832 (2022).

55. Hao, Y. et al. The role of surface oxygen in the growth of large single-crystal graphene on copper. *Science* **342**, 720–723 (2013).
56. Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
57. Zhang, D., Yi, P., Lai, X., Peng, L. & Li, H. Active machine learning-driven model for the dynamic simulation and growth mechanisms of carbon on metal surface. *Zenodo* 105281/zenodo10159885, (2023).
58. Kühne, T. D., Krack, M., Mohamed, F. R. & Parrinello, M. Efficient and accurate car-parrinello-like approach to born-oppenheimer molecular dynamics. *Phys. Rev. Lett.* **98**, 066401 (2007).
59. VandeVondele, J. et al. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167**, 103–128 (2005).
60. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703–1710 (1996).
61. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
62. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
63. Zhang, D. Source data and dynamic trajectory files. *Figshare*, <https://doi.org/10.6084/m9.figshare.24591774.v2> (2023).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 22309109 and No. 52225504) (D.Z. and L.F.P.). It was also supported by Research Projects of the National Key Laboratory (No. JCKYS2022603C009) (P.Y.Y., L.F.P., X.M.L.) and the State Key Laboratory of Mechanical System and Vibration (No. MSVZD202009) (P.Y.Y., L.F.P., X.M.L.). D.Z. acknowledges the support of the China National Postdoctoral Program for Innovative Talents from the China Postdoctoral Science Foundation (BX2021178) (D.Z.). The computations in this paper were run on the Siyuan-1 cluster supported by the Center for High-Performance Computing at Shanghai Jiao Tong University. H.L. acknowledges the support received from JSPS KAKENHI (Grant No. JP23K13703) (H.L.) and the Iwatani Naoji Foundation. Additionally, the Center for Computational Materials Science at the Institute for Materials Research, Tohoku University, is acknowledged for providing access to MASAMUNE-IMR (Project No. 202212-SCKXX-0204) (H.L.). The Institute for Solid State Physics (ISSP) at the University of Tokyo is acknowledged for granting access to their supercomputers.

## Author contributions

D.Z. conceived the idea and performed all the simulations and analyses. P.Y.Y., X.M.L., L.F.P., and H.L. supervised the research. All authors contributed to the manuscript before submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-44525-z>.

**Correspondence** and requests for materials should be addressed to Di Zhang, Linfa Peng or Hao Li.

**Peer review information** *Nature Communications* thanks Bora Karasulu, Jonathan Vandermause, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

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

© The Author(s) 2024