

Supporting Information

Understanding High-Temperature Chemical Reactions on Metal Surfaces: A Case Study on Equilibrium

Concentration and Diffusivity of C_xH_y on Cu(111) Surface

Pai Li, Xiongzhi Zeng, Zhenyu Li*

Hefei National Laboratory for Physical Sciences at the Microscale, University of
Science and Technology of China, Hefei, Anhui 230026, China

(Email: zyli@ustc.edu.cn)

1. More computational details

1.1 Vibrational contributions on energy barriers

Three vibrational terms on energy barrier are zero-point-energy (ZPE), vibrational enthalpy, and vibrational entropy contributions:

$$\Delta_{ZPE} = E_{ZPE}^{TS} - E_{ZPE}^{IS} = \sum_{i \in TS} \frac{h\nu_i}{2} - \sum_{i \in IS} \frac{h\nu_i}{2}$$
$$\Delta_{vib_enthalpy} = E_{vib}^{TS}(T) - E_{vib}^{IS}(T) = \sum_{i \in TS} \frac{h\nu_i}{\exp\left(\frac{h\nu_i}{k_B T}\right) - 1} - \sum_{i \in IS} \frac{h\nu_i}{\exp\left(\frac{h\nu_i}{k_B T}\right) - 1}$$
$$\Delta_{vib_entropy} = -TS_{vib}^{TS} + TS_{vib}^{IS}$$
$$TS_{vib} = k_B T \sum_i \left[\frac{h\nu_i}{k_B T \times \left[\exp\left(\frac{h\nu_i}{k_B T}\right) - 1 \right]} - \ln\left(1 - \exp\left(\frac{-h\nu_i}{k_B T}\right)\right) \right]$$

where ν_i is the vibrational frequency of the i -th mode, h is the Planck's constant, k_B is the Boltzmann constant. TS and IS denote the transition and initial state, respectively.

1.2 Widom insertion to calculate the excess chemical potential

Since all hydrocarbon species considered in this study have a favored adsorption site on the top two Cu atom layers (surface hollow sites for H/CH₃, surface bridge sites for C₂, and subsurface interstitial site for C), we choose an insertion space containing surface/subsurface Cu atoms. In the vertical direction, we set the upper limit of the space (the upper limit of the lowest atom for multi-atom species) to be the average height of the top Cu atom layer plus one and half of the Cu-Cu layer spacing, the lower limit (the lower limit of the highest atom for multi-atom species) to be the average height of the second Cu layer minus half of the Cu-Cu layer spacing.

If we define the concentration of a species in this insertion space as the number ratio between species and Cu atoms in the thin volume, since the number of surface/subsurface Cu atoms equals the number of surface hollow sites, the 3D species

concentration in the insertion space is identical to the 2D species coverage for H/CH₃. For other species, such a mapping can be obtained by simply applying a constant factor.

Given a specific species and temperature, we do the following steps:

1. Run GAP-MD for 110,000 steps at the given temperature with a pure-Cu surface model.
2. Abandon the first 10,000 steps and extract one structure from every 10 steps in the remaining 100,000 steps.
3. For each extracted substrate structure, randomly insert the given species into the insertion space, calculate and record the energy change ϕ due to insertion.
4. Repeat step 3 for 100 times.

Finally, we have 1,000,000 samples to calculate the average value of $\exp(-\phi/k_B T)$. The excess chemical potential can thus be acquired according to its definition

$$\mu^{ex} = -k_B T \times \ln \langle \exp(-\phi/k_B T) \rangle$$

For CH/CH₃/C₂, we fixed the C-H/C-C bond length to the optimized value in the adsorption state. The energy difference caused by the bond-length constraints can be estimated via MD simulations of the system with and without bond length constraints at each temperature. It turns out to be small (about 0.14 eV for CH at 1000 K) and it is added as an energy correction to the excess chemical potential.

For monoatomic species (C and H), the insertion is performed by simply giving three evenly distributed random numbers to determine the insertion position. For diatomic species (CH and C₂), the position of one C atom is randomly given. After that, a random direction is given by providing three random numbers according to the normal distribution. The position of the other atom is determined by the bond length and the random direction. For CH₃, a random direction is given to determine the rotational symmetry axis and a random number is given to determine the rotation angle of H atoms around the symmetry axis.

The ideal-gas chemical potential at a reference density is defined as follows. For monomer, it contains only the translational part:

$$\mu_{trans}^{ideal} = k_B T \times \ln(\rho \times \Lambda_{sp})$$

where ρ is the density of substrate atoms which slightly changed due to the volume expansion at a higher temperature, and Λ_{sp} is the thermal de Broglie wavelength:

$$\Lambda_{sp} = (h^2/2\pi m k_B T)^{1/2}$$

For multi-atom species, it additionally contains the vibrational and rotational parts:

$$\mu_{vib}^{ideal} = \sum_i k_B T \times \ln \left(\frac{\exp(\frac{h\nu_i}{2k_B T})}{1 - \exp(\frac{h\nu_i}{k_B T})} \right)$$

$$\mu_{rot}^{ideal} = \begin{cases} k_B T \times \ln \left(\frac{8\pi^2 I k_B T}{\sigma h^2} \right), & \text{if linear} \\ k_B T \times \ln \left[\frac{\sqrt{\pi I_A I_B I_C}}{\sigma} \left(\frac{8\pi^2 k_B T}{h^2} \right)^{3/2} \right], & \text{if nonlinear} \end{cases}$$

where ν_i is the i -th vibrational frequency of a multi-atom species, I is the degenerate moment of inertia for a linear species, $I_{A/B/C}$ are the three principal moments of inertia for a non-linear species and σ is the symmetry number.

1.3 Kirkwood coupling parameter method

For C and C₂, simulations based on the Kirkwood coupling parameter method were also performed. This method is based on a series of artificial systems, in which the interaction between the adsorbate and the Cu environment is scaled with a factor α . We first calculate the excess chemical potential for the $\alpha=0.1$ system using Widom insertion:

$$\mu^{ex}(\alpha = 0.1) = -k_B T \times \ln \left\langle \exp \left(-\frac{\phi \times 0.1}{k_B T} \right) \right\rangle$$

Then, the excess chemical potential for the real system $\alpha=1$ is calculated with thermodynamic integration:

$$\mu^{ex}(\alpha = 1) = \mu^{ex}(\alpha = 0.1) + \int_{0.1}^1 \langle \phi \rangle_{\alpha} d\alpha$$

where $\langle \phi \rangle_{\alpha}$ is the ensemble average of ϕ for the artificial system with scaled factor α .

To get $\langle \phi \rangle_{\alpha}$, we performed a series of MD simulations for different α ranging from 0.1 to 1 (the $\alpha=1$ system is the real system). To get the scaled interaction, we calculate forces for both the complete model (F_{all}) and the model with the adsorbate removed (F_{subs}). Then, the artificial forces are given by

$$F_{artif} = F_{subs} + \alpha \times (F_{all} - F_{subs})$$

For atoms in the adsorbate, F_{subs} is zero. In this way, the interaction between Cu is kept but that between Cu and adsorbate is scaled with the factor α . Meanwhile, the energy difference between two systems with and without the adsorbate (ϕ) is recorded in each MD step to get its ensemble average $\langle \phi \rangle_{\alpha}$.

This artificial-force scheme for MD simulation can be directly used for the C monomer. For C₂, it scales also the C-C interaction which is incorrect. Here, we simply treat C₂ as a rigid body by fixing the C-C bond length in all MD simulations. The excess chemical potential obtained is finally corrected by applying the bond constrain correction at each temperature.

2. GAP training

For each hydrocarbon species, we first run *ab initio* MD at 3000 K for 50 steps. All these frames are used as the first batch of the training set, from which an initial GAP is obtained. Then, GAP-MD simulations are performed to obtain other configurations of the training set. To extract N structures at a temperature between 2000 and 600 K, we run a GAP-MD simulation at this temperature for $1000+100*N$ steps. We uniformly extract one structure in every 100 frames for the last $100*N$ steps. At temperatures below 600 K, the simulation steps are increased to $1000+1000*N$ steps to extract a structure from every 1000 frames. DFT calculations for all these extracted structures are performed with VASP to get their energies and forces. The value N for each temperature is given in Figure S1a.

To train the less-accurate initial GAPs used for generating more structures in the training set, we use only SOAP descriptor¹ with an expansion of the neighbor density up to $l_{\max}=6$, $n_{\max}=10$, and cutoff=5 Å. For the final GAPs used for calculating diffusivity and coverage properties, we use distance_2b descriptor with cutoff=4 Å, angle_3b descriptor with cutoff=3 Å, and Soap descriptor with $l_{\max}=8$, $n_{\max}=12$, and cutoff=5.5 Å. The parameter atom_sigma is set to be 0.3 and 0.5 for systems with and without H involved, respectively.

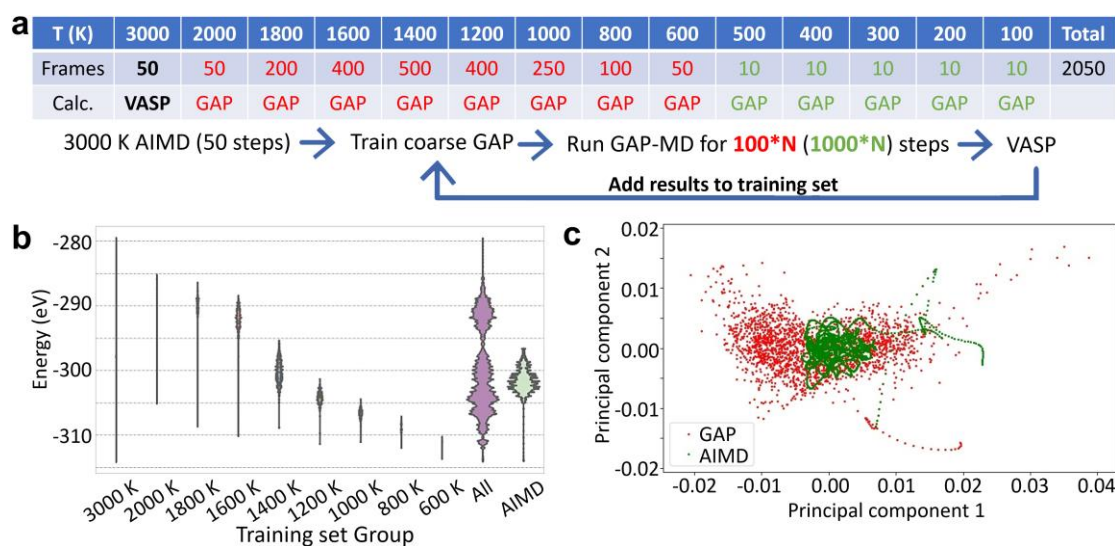


Figure S1. (a) The iterative training schemes. "Frames" in the table gives the number of structures extracted as a training set at each temperature. More structures were collected at temperatures close to the experimental temperature of graphene CVD growth. (b) Energy distribution in each batch of the training set, and in the whole training set (denoted as "All"). An *ab initio* MD trajectory at 1300 K denoted as "AIMD" with the same number of frames to the training set is presented as a comparison. (c) Principal component analysis on structures of the whole training set and that of the 1300 K AIMD trajectory. SOAP descriptor¹ is adopted to convert the structure into a vector for the principal component analysis.

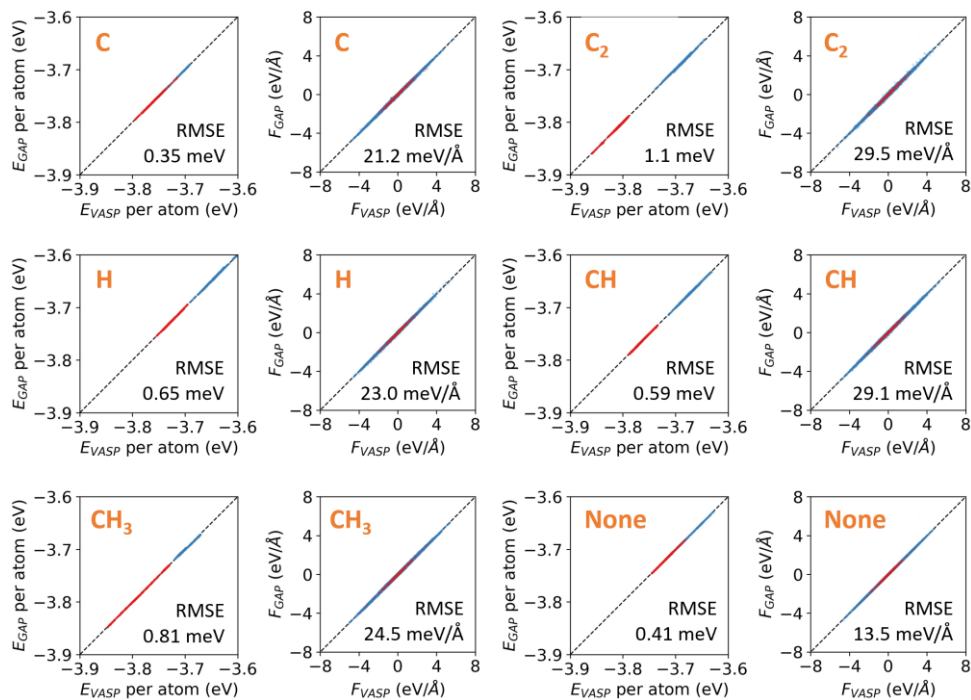


Figure S2. Energy and force error analysis for Cu substrate with adsorbate C/C₂/H/CH/CH₃ or without adsorbate (denoted as "None"). The test set obtained from an AIMD simulation at 1200 K is in red and that at 1400 K is in blue. All these results indicate high accuracy with very low root-mean-square errors (RMSE).

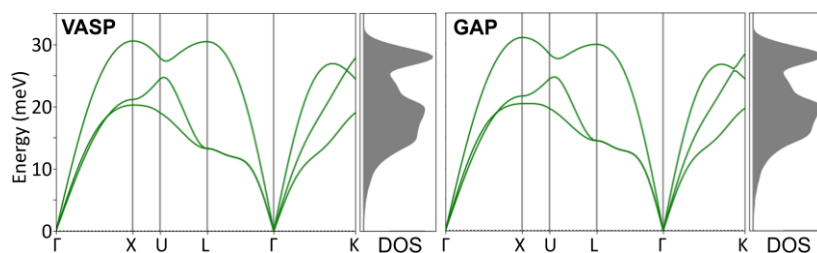


Figure S3. Cu phonon band structures calculated using VASP and GAP, respectively.

3. Crystalline surface model

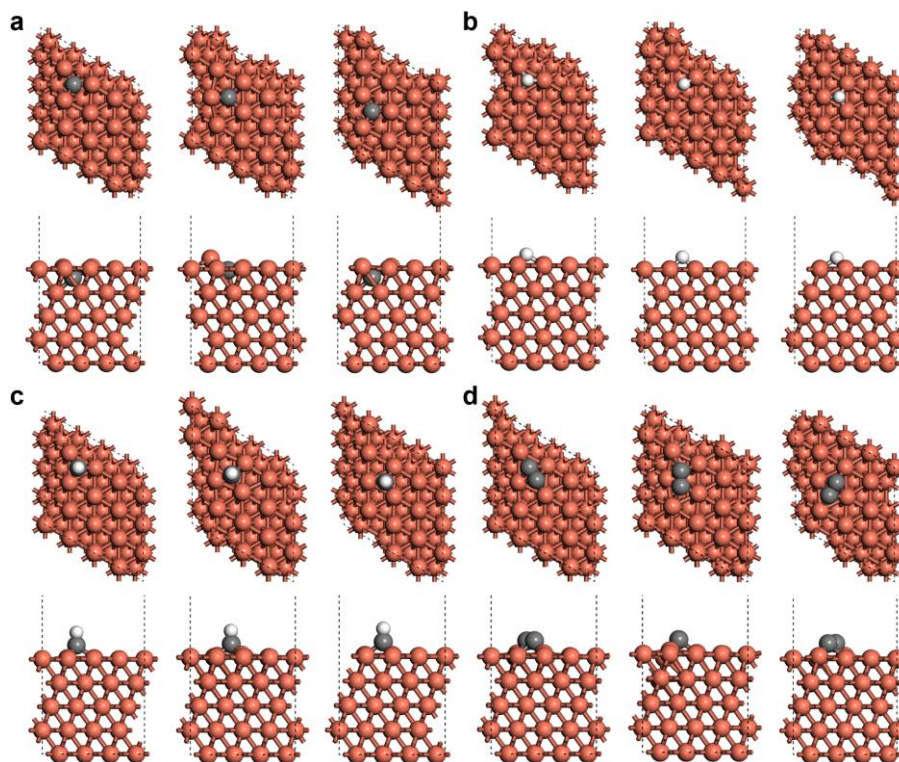


Figure S4. Initial/transition/final states in top and side views for (a) C, (b) H, (c) CH and (d) C₂ diffusion on Cu111 surface. Cu, C, and H atoms are shown in brown, grey, and white, respectively.

Table S1. Vibrational correction terms and the final free energy barriers for C/C₂/H/CH/CH₃ under different temperatures. T is the temperature in K. ΔZPE , ΔC_p and $\Delta(-TS)$ are the difference of ZPE, vibrational enthalpy, and vibrational entropy part between transition and initial states. ΔG_{vib} is the summation of these three terms. "Barrier" is calculated as the summation of the NEB barrier and ΔG_{vib} . The substrate, except for the top layer, is fixed to calculate the vibrational frequencies.

	T	ΔZPE	ΔC_p	$\Delta(-TS)$	ΔG_{vib}	Barrier
C	600	-0.023	-0.033	0.062	0.006	0.504
	800	-0.023	-0.049	0.101	0.029	0.527
	1000	-0.023	-0.065	0.145	0.057	0.555
	1200	-0.023	-0.083	0.192	0.086	0.584
	1400	-0.023	-0.100	0.242	0.119	0.617
	1600	-0.023	-0.116	0.295	0.156	0.654
C ₂	600	-0.033	-0.027	-0.022	-0.082	0.407
	800	-0.033	-0.043	-0.012	-0.088	0.401
	1000	-0.033	-0.059	0.004	-0.088	0.401
	1200	-0.033	-0.075	0.022	-0.086	0.403
	1400	-0.033	-0.092	0.044	-0.081	0.408

	1600	-0.033	-0.108	0.068	-0.073	0.416
	600	-0.024	-0.024	0.031	-0.017	0.113
	800	-0.024	-0.042	0.061	-0.005	0.125
H	1000	-0.024	-0.060	0.096	0.012	0.142
	1200	-0.024	-0.077	0.134	0.033	0.163
	1400	-0.024	-0.095	0.176	0.057	0.187
	1600	-0.024	-0.111	0.220	0.085	0.215
	600	-0.002	-0.049	0.144	0.093	0.206
	800	-0.002	-0.067	0.212	0.143	0.256
CH	1000	-0.002	-0.084	0.284	0.198	0.311
	1200	-0.002	-0.101	0.360	0.257	0.370
	1400	-0.002	-0.118	0.439	0.319	0.432
	1600	-0.002	-0.136	0.520	0.382	0.495
	600	0.004	-0.042	0.08	0.042	0.129
	800	0.004	-0.061	0.129	0.072	0.159
CH ₃	1000	0.004	-0.079	0.181	0.106	0.193
	1200	0.004	-0.096	0.237	0.145	0.232
	1400	0.004	-0.115	0.296	0.185	0.272
	1600	0.004	-0.133	0.358	0.229	0.316

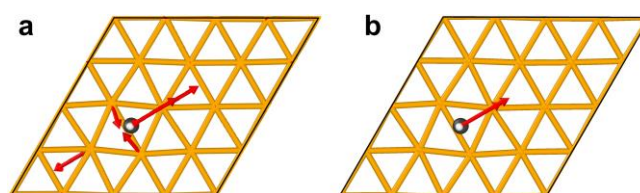


Figure S5. (a) Vibrational mode with an imaginary frequency in the transition state of CH diffusion with all surface atoms unfixed. (b) The corresponding model mode calculated with all Cu atoms fixed, where the vibrational frequency becomes real.

4. MSD and diffusion pattern

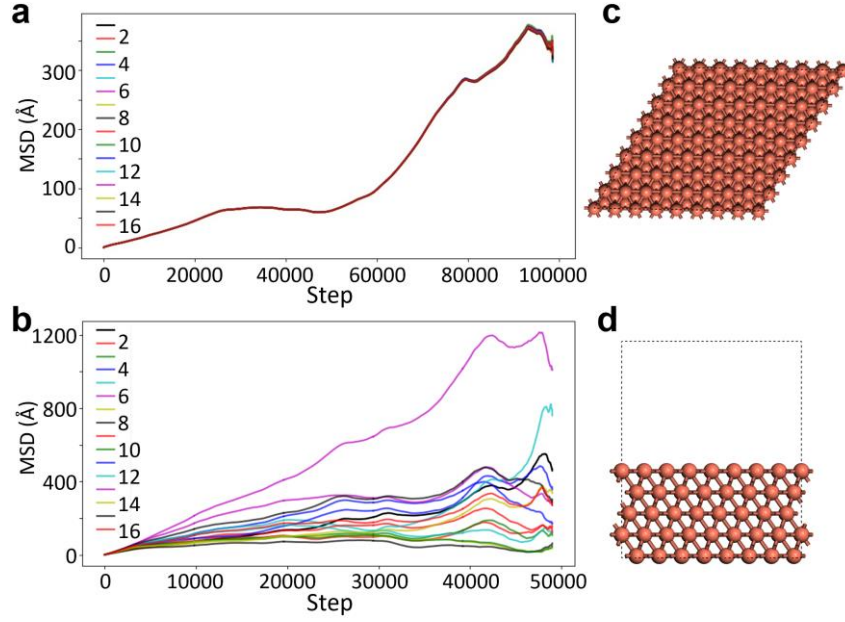


Figure S6. MSD of 16 Cu atoms in the top layer of a 4×4 surface obtained from test MD simulations at (a) 1000 and (b) 1400 K. The MSD of all atoms overlap at 1000 K because of the unphysical slide of the top layer due to the limited size of the surface model. At 1400 K, this phenomenon disappears since the surface is melted. (c) Top and (d) side view of the 8×8 surface model we used for the final MD calculations.

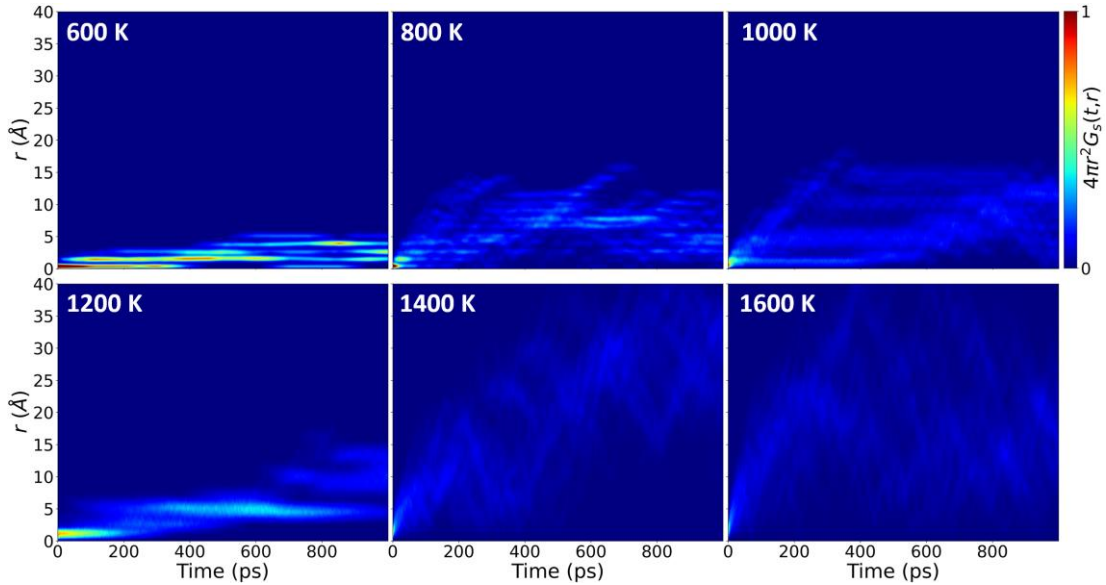


Figure S7. Self-part of the van Hove correlation function for C₂ diffusion at different temperatures. At a temperature lower than the Cu melting point, the correlation function shows a hopping picture. Otherwise, the Cu surface is melted and the diffusion behavior becomes Brownian motion on the liquid surface. The self-part of the Van Hove correlation function is calculated as follows:

$$G_s(r, t) = \frac{1}{4\pi r^2 N_d} \langle \delta(r - |r_i(t_0) - r_i(t + t_0)|) \rangle_{t_0}$$

where $\delta(\cdot)$ is the Dirac delta function, N_d is the number of the diffusing species, $r_i(t)$ is the species position at time t . The angular brackets indicate ensemble average over the initial time t_0 .

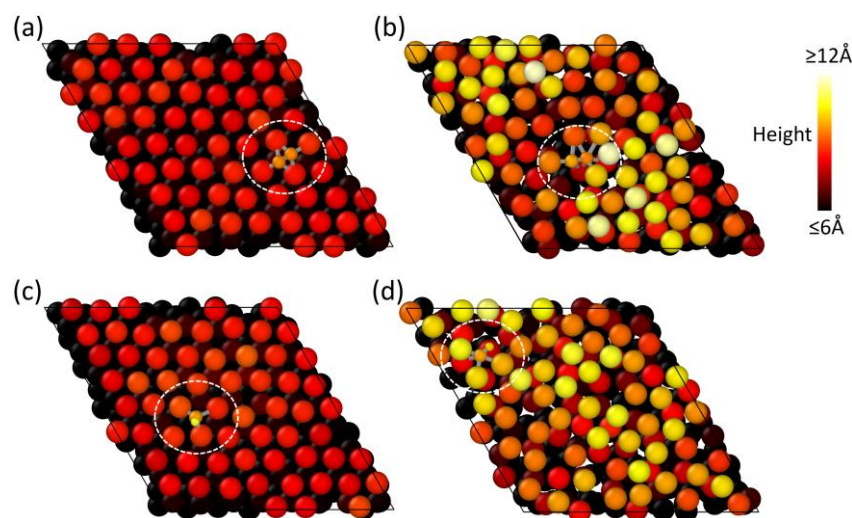


Figure S8. Snapshots of C_2 and CH in MD trajectories. (a) C_2 at 600 K; (b) C_2 at 1600 K; (c) CH at 600 K; (d) CH at 1600 K. Atoms' heights are distinguished with different colors. The largest balls are Cu atoms, the smallest balls are H atoms, and the other balls are C atoms.

6. Thermostat performance and convergence analysis

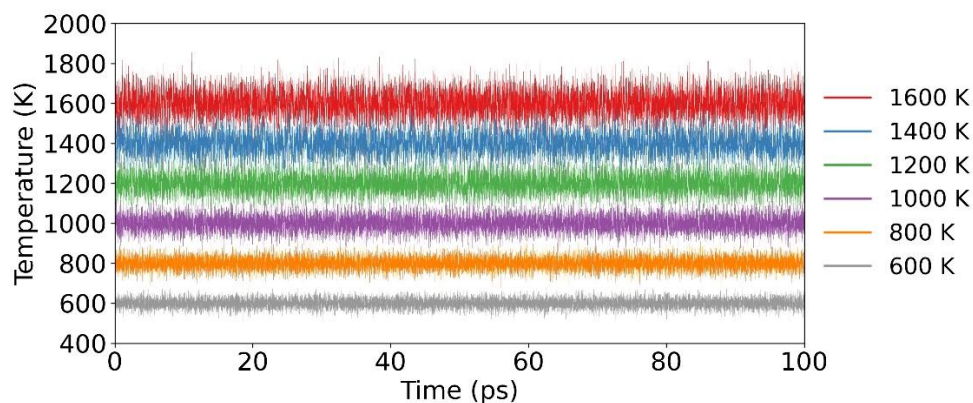


Figure S9. Temperature fluctuations in the last 100 ps for C_2 diffusion MD simulations at different temperatures.

To obtain converged diffusivity, it is suggested that the total mean square displacement (TMSD) in the simulation should be up to a few hundred \AA^2 .² Such a criterion is fulfilled in all our simulations for diffusivity (Figure S10), indicating the trajectories are long enough. For Widom insertion, the typical number of insertion attempts is $10^5 \sim 10^6$.^{3,4} In this study, Widom insertion simulations are performed with 10^6 attempts each. At the same time, we divide each simulation into five parts with equal steps and calculate the diffusivity or excess chemical potential based on each part. The obtained standard deviations are drawn as the error bars in Figure 2 and Figure S11,

which are reasonably small.

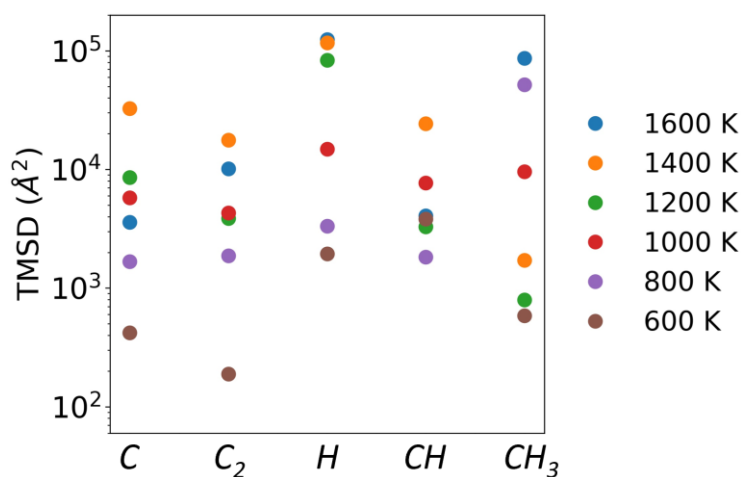


Figure S10. TMSD of adsorbates in MD simulations at different temperatures.

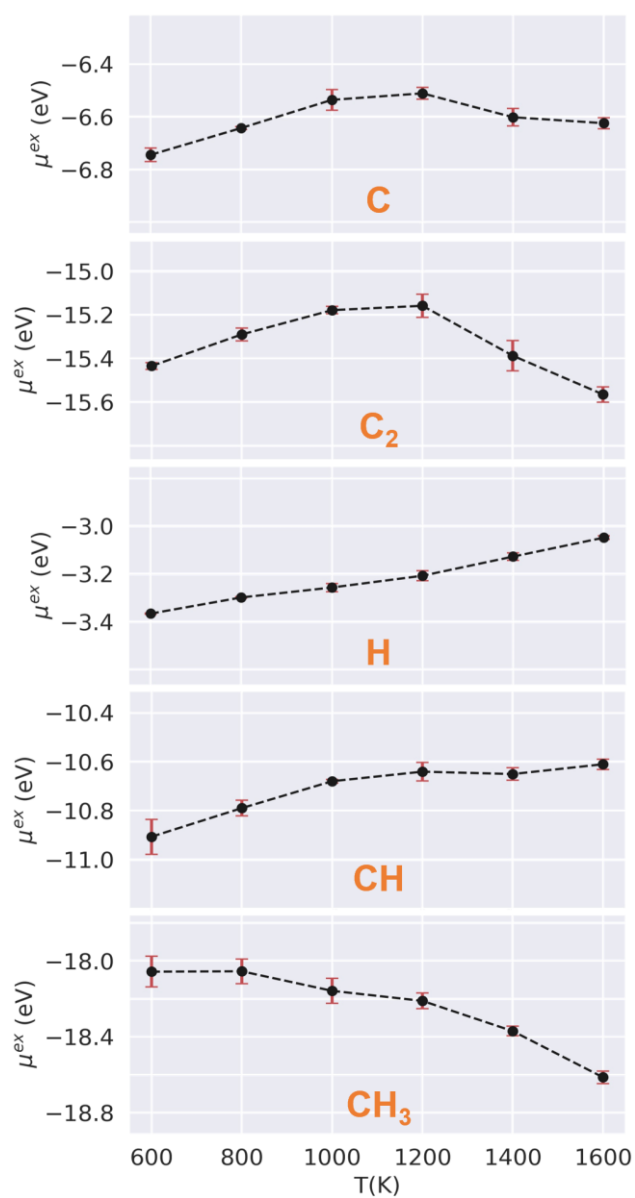


Figure S11. The excess chemical potential of C/C₂/H/CH/CH₃ adsorbates on Cu surface at different temperatures. The error bar gives the standard deviation of five independent sampling groups.

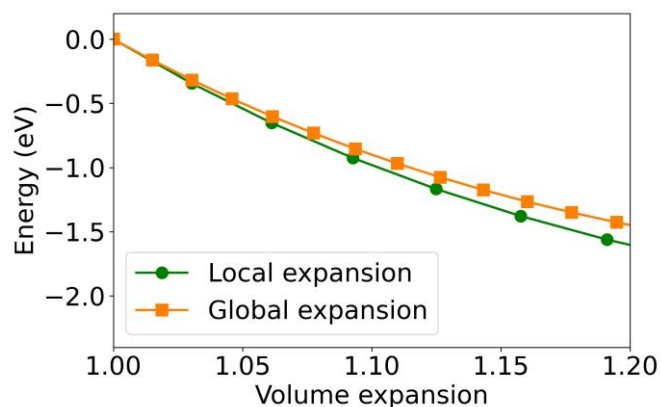


Figure S12. The adsorption energy of C monomer on expanded Cu substrate with local and global expansion models. In the global expansion model, the entire surface slab is expanded in the x/y/z directions. All Cu atoms are fixed and only the C monomer is optimized to get the interaction energy.

References

- (1) Bartók, A. P.; Kondor, R.; Csányi, G. *Phys. Rev. B* **2013**, *87*, 184115.
- (2) He, X.; Zhu, Y.; Epstein, A.; Mo, Y. *npj Comput. Mater.* **2018**, *4*, 18.
- (3) Study, D. *J. Phys. Chem.* **1995**, *99*, 2900.
- (4) Chennamsetty, N.; Bock, H.; Gubbins, K. E. *Mol. Phys.* **2005**, *103*, 3185.