

## ***Ab Initio* Approach for Thermodynamic Surface Phases with Full Consideration of Anharmonic Effects: The Example of Hydrogen at Si(100)**

Yuanyuan Zhou<sup>1,\*</sup>, Chunye Zhu<sup>1,2</sup>, Matthias Scheffler<sup>1</sup>, and Luca M. Ghiringhelli<sup>1</sup>

<sup>1</sup>The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Berlin-Dahlem 14195, Germany

<sup>2</sup>School of Advanced Manufacturing, Guangdong University of Technology, Jieyang 515200, China



(Received 31 January 2022; accepted 6 May 2022; published 17 June 2022)

A reliable description of surfaces structures in a reactive environment is crucial to understand materials' functions. We present a first-principles theory of replica-exchange grand-canonical-ensemble molecular dynamics and apply it to evaluate phase equilibria of surfaces in a reactive gas-phase environment. We identify the different surface phases and locate phase boundaries including triple and critical points. The approach is demonstrated by addressing open questions for the Si(100) surface in contact with a hydrogen atmosphere. In the range from 300 to 1000 K, we find 25 distinct thermodynamically stable surface phases, for which we also provide microscopic descriptions. Most of the identified phases, including few order-disorder phase transitions, have not yet been observed experimentally. Furthermore, we show that the dynamic Si-Si bonds forming and breaking is the driving force behind the phase transition between  $3 \times 1$  and  $2 \times 1$  adsorption patterns.

DOI: [10.1103/PhysRevLett.128.246101](https://doi.org/10.1103/PhysRevLett.128.246101)

Knowledge of the morphology and structural evolution of material surfaces in a given reactive atmosphere is a prerequisite for understanding the mechanism of, e.g., heterogeneous-catalysis reactions and electrocatalysis due to the structure-property-performance relationship [1,2]. In general, reliable tracking of phase equilibria is of technological importance for the rational design of surface properties [3].

Studying phase equilibria with first-principles theory is a formidable challenge, especially for highly anharmonic systems, including therein systems with multiple minima separated by shallow barriers. Pioneering efforts have been performed to estimate the *ab initio* melting line of bulk systems, by direct-coexistence [4] as well as thermodynamic-integration techniques [5,6]. Similar to *ab initio* atomistic thermodynamics (aiAT) [7–10], the state-of-the-art approach has been successful in studying the thermodynamic of surfaces and clusters in reactive environments. Crucially, these techniques require the prior knowledge of the set of relevant phases populating the phase diagram and typically do not yield to the discovery of yet unexpected phases and phase transitions. Furthermore, only differences in free energies among the considered phases are used to assess the stability regions of the phases in the phase

diagram and do not model singularities of a response function (e.g., heat capacity) to locate phase boundaries and distinguish phase transitions from smooth transitions.

Recently, we have introduced a replica-exchange (RE) grand-canonical (GC) algorithm [11] to determine the thermodynamic stability of surfaces in reactive atmospheres, including all anharmonic contributions without approximations. The REGC approach performs an unbiased sampling of the configurational and compositional grand-canonical ensemble and takes as input only the potential-energy function together with the desired chemical-potential ( $\mu$ ) and  $T$  ranges. No prior knowledge about the phase diagram of interest is needed.

In this Letter, we introduce a fully *ab initio* approach to determine temperature-pressure ( $T$ ,  $p$ ) phase diagrams of surfaces in reactive atmospheres, with the accurate location of phase boundaries, as well as triple and critical points. For this purpose, our REGC method is extended by including the evaluation of the constant-volume heat capacity as a function of  $T$  and  $p$ . Furthermore, we show how the crucial limitation of the GC approach, formally defined only for a constant-volume ensemble, is circumvented by sampling different (here, two) simulation cells, each in a (constant-volume) REGC ensemble, and then connecting all the phases in both cells via a common reference phase, which belongs to both cells.

In order to explain the insight that can be revealed by our approach, we studied the silicon (100) surface in a hydrogen atmosphere. The chemistry of hydrogen on silicon surfaces has important applications, such as the passivation of surfaces, etching, and CVD growth. Furthermore, the dissociative adsorption of molecular hydrogen on the

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

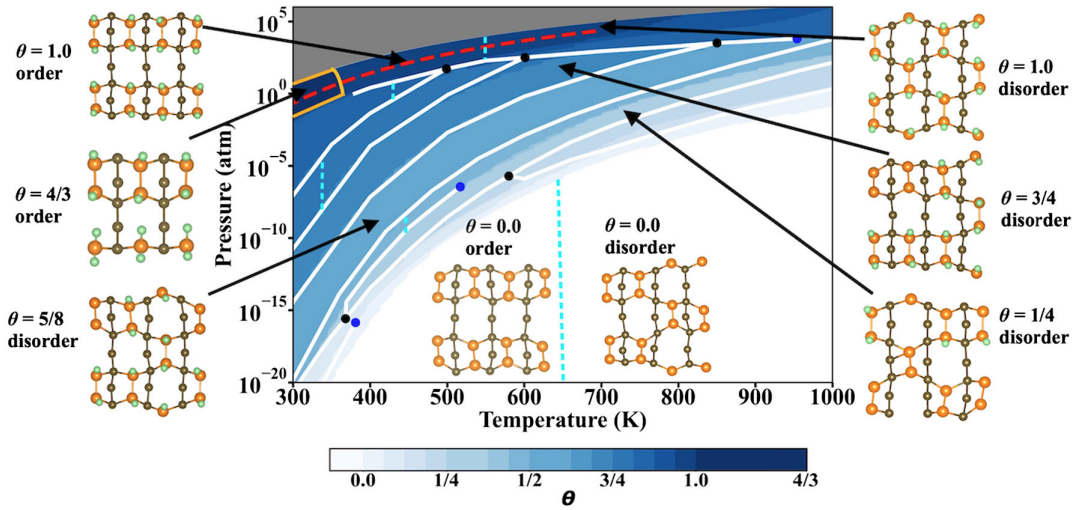


FIG. 1. Phase diagram of the Si(100) surface in a  $D_2$  gas phase. Each color represents a thermodynamic phase with a certain deuterium coverage. The white area is the stability region of the pristine surface. The dark blue area surrounded by the orange frame is the stability region of the phase with maximum coverage of  $4/3$ , sampled in the  $3 \times 3$ , and its the neighbor phases are sampled in the  $4 \times 4$  supercell. The white lines indicate the phase boundaries, identified by the analysis of the heat capacity. The critical points are marked as blue dots while the triple points are marked as black dots. The atomic-structure images show the top view of eight representative phases with diverse deuterium coverage  $\theta$ . The golden spheres are the top silicon atom, the green spheres are deuterium atoms, and the dark spheres are Si atoms in deeper layers. The five cyan dashed lines mark the boundaries of order-disorder phase transitions. The red dashed line indicates  $\mu_D = -0.1$  eV, which marks the  $(T, p_{D_2})$  path analyzed in Fig. 2. The grey area at high pressures corresponds to the region at larger chemical potentials than those sampled in this Letter.

Si(100) surface has become a paradigm in the study of adsorption systems [12–14]. Three distinct phases have been experimentally observed [15]: (i) a  $2 \times 1$  monohydride phase at 600 K (at coverage  $\theta = 1$ , see, e.g., structure at the top left of Fig. 1), where the dimers of the reconstructed pristine Si(100) surface are preserved, (ii) a  $1 \times 1$  dihydride phase below 300 K where the dimer bonds are broken [16], and (iii) at around 400 K, a  $3 \times 1$  phase [17], interpreted as alternating rows on monohydrides and dihydrides ( $\theta = 4/3$ , see structure left center in Fig. 1). Despite the extensive observations under ultrahigh vacuum, there is still a lack of systematic measurements to address directly the thermodynamically stable hydrogen-terminated Si(100) structures, when the surface is in equilibrium with an atmosphere of molecular hydrogen at a given temperature and pressure of the reacting atmosphere.

Our approach entails two steps: data acquisition and data (post) processing. The acquisition of the data is performed by running the REGC [11] simulation, here coupled with first-principles Born-Oppenheimer molecular dynamics (MD). Data processing is performed via the multistate Bennet acceptance ratio (MBAR) approach [18], which is a low-variance estimator of ensemble-averaged thermodynamic observables. The technique is based on the Boltzmann reweighting and exploits the fact that the configurational density of states is temperature independent, while the probability to observe a given configuration depends on temperature via the Boltzmann factor,

$\exp(-\beta U)$ , where  $\beta = 1/k_B T$  is the inverse temperature and  $U$  is the GC potential function of the system, i.e.,  $U = E - \sum_i \mu_i N_i$ .  $E$  is the potential energy of a given configuration,  $\mu_i$  the chemical potential of each species exchanged with the reservoir, and  $N_i$  the number of particles of that species. Importantly, all the observables that are processed via MBAR can be identified *a posteriori*, i.e., after the data acquisition is completed. MBAR post-processes the data sampled at all temperatures and chemical potential from the REGC MD run and estimates ensemble values of the desired observables at any given  $T$  and  $\mu$ , not necessarily among the sampled ones.

The evaluation of the heat capacity as a function of temperature and pressure of the reacting gas enables us to identify phase-transition lines [narrow stripes in the  $(T, p)$  space where the function  $C_V(T, p)$  shows ridges] as well as triple and critical points. This concept represents an important advancement compared with aiAT, which considers only the differences in free energy between the different phases and locates boundaries where such difference is zero. As we will show below, there are cases where the surface restructures (e.g., changes its coverage) as a function of  $(T, p)$ , but there is no associated peak in  $C_V$ , and the transition is therefore smooth.  $C_V$  is calculated via its statistical-mechanics definition:

$$C_{V,(T,p)} = \frac{\langle E^2 \rangle_{(T,p)} - \langle E \rangle_{(T,p)}^2}{k_B T^2}, \quad (1)$$

where  $E$  is the DFT total energy (i.e., all potential-energy contributions and kinetic energy) of the system. The ensemble averages of  $E$  and  $E^2$  are evaluated at each thermodynamical state point  $(T, p)$  of interest, where  $p$  is the pressure of the gas in the grand-canonical reservoir, compatible with the sampled  $T$  and  $\mu$  (see below for more details). This definition is equivalent to the thermodynamic definition  $C_V = (\partial E / \partial T)_V$ , as  $E$  is  $-\partial \ln Z / \partial \beta$ , with  $Z$  the configurational partition function.

In MBAR, the ensemble average of any observable  $A$  that is a function of the configuration  $\mathbf{R}_n$  of the system at a given state point  $(\mu, \beta)$  is statistically evaluated as

$$\langle A \rangle_{\mu, \beta} = \sum_{n=1}^{\Omega} \frac{A(\mathbf{R}_n) c_{\mu, \beta}^{-1} q(\mathbf{R}_n; \mu, \beta)}{\sum_{l, m} \Omega_{l, m} c_{\mu_m, \beta_l}^{-1} q(\mathbf{R}_{l, m}; \mu_m, \beta_l)} \quad (2)$$

where  $\Omega$  is the total number of samples in all replicas and  $\Omega_{l, m}$  is the number of samples in each sampled state point  $(m, l)$ , where  $m$  and  $l$  are the index number of the selected chemical potentials and temperatures, respectively.  $q(\mathbf{R}_n; \mu, \beta) = \exp[-U(\mathbf{R}_n; \mu, \beta)]$  is the grand-canonical density function, and  $c(\mu, \beta)$  is the *partition function* estimated by MBAR [11, 19]. The expectation values of  $\langle E \rangle_{\mu, \beta}$  and  $\langle E^2 \rangle_{\mu, \beta}$  are evaluated via Eq. (2) where  $A(\mathbf{R}_n)$  is  $E(\mathbf{R}_n)$  and  $E^2(\mathbf{R}_n)$ , respectively.

The first step for the creation of the phase diagram is to evaluate via MBAR the relative free energy as a function of  $T$  and  $\mu$  of all phases, identified within the REGC MD sampling, as detailed in the Supplemental Material [20] and Ref. [11]. Here, we limit ourselves to the case of only one reactive species exchanged with the reservoir. The different phases are identified by means of structural parameters, e.g., the coverage and the reconstruction or adsorption patterns. The  $(T, \mu)$  states are mapped into the more intuitive  $(T, p)$  states by choosing the reactive species to be an ideal gas in the reservoir, i.e.,  $\beta\mu = \ln \Lambda^3 + \ln(\beta p)$ , where  $\Lambda$  is the thermal wavelength of the species with mass  $m$ :  $\Lambda = h / \sqrt{2\pi m k T}$ .

Phase boundaries (white lines in Fig. 1) are estimated by connecting the  $(T, p)$  state points where the curve  $C_V(p)$  at constant  $T$ , or, symmetrically,  $C_V(T)$  at constant  $p$ , shows a peak [25]. A critical point is a thermodynamic condition where the free-energy barrier between phases is zero. Here, it is indicated by the  $C_V$  peak becoming shallower with increasing  $T$  and  $p$  then disappearing at  $(T_c, p_c)$  (critical temperature and pressure). The triple points are the intersections of three phase boundaries. Practical details on how  $C_V$  peaks, triple, and critical points are given in the Supplemental Material [20].

In our Letter, the calculated heat capacity maintains a finite value because (i) the simulated system has a finite size and (ii) the  $C_V$  is calculated by integrating over a (small)  $\delta p \cdot \delta T$  area where singular point divergences are smeared.

In all our simulations, we use deuterium instead of hydrogen. This is a common approach to lower the

vibrational frequencies and thus allow for a longer time step in the MD runs. For classical particles, this does not change the phase diagram as the configurational partition function depends only on the shape of the potential energy. Furthermore, we note that similar results are observed in experiments after exposure of silicon surfaces [including the (100) surface] to both H and D [17, 26]. This also suggests that the difference in quantum-nuclear effects given by the different isotopes' masses is negligible in the range of temperatures considered in this Letter. We define *surface region* as the region of thickness 3.0 Å along the surface normal, starting from the average  $z$  coordinate of the topmost Si atoms of the surface. The simulation time per replica was 60 ps (3 000 REGC steps) for Si(100)-(4 × 4) resulting in a combined simulation time of around 6 ns. All Born-Oppenheimer *ab initio* molecular dynamics (AIMD) trajectories are performed in a canonical ( $NVT$ ) ensemble running for 0.02 ps and using a 1-fs time step. The stochastic velocity rescaling thermostat [27] was used to sample the  $NVT$  ensemble with  $\tau$  parameter of 20 fs. All DFT calculations are performed with the all-electron, full-potential electronic-structure package FHI-aims [28]. The Perdew-Burke-Ernzerhof (PBE) [29] exchange-correlation functional is used with a tail correction for van der Waals interactions (Tkatchenko-Scheffler scheme) [30]. Benchmark studies of model systems and the numerical settings are given in the Supplemental Material [20].

For the evaluation of  $C_V$  (Eq. (1)),  $\langle E \rangle_{T, \mu}$  and  $\langle E^2 \rangle_{T, \mu}$  are calculated at  $(T, \mu)$  with  $T$  ranging from 300 K to 1 000 K (with a spacing of 1 K) and  $\mu$  ranging from  $-0.08$  eV to  $-0.06$  eV (with a 0.01 eV spacing). The convergence of the sampling was monitored by checking, in each simulated  $(T_i, \mu_j)$  state, the average of the distribution of the number of deuterium atoms ( $\langle N_D \rangle$ ) and its variance remained constant along the last 80% of generated data along the MD trajectories. All energies in both cells subtract the energy of unreconstructed (1 × 1) pristine Si(100) for jointly postprocessing (see the Supplemental Material [20] for more details).

Figure 1 summarizes our main finding, i.e., the phase diagram of Si(100) in the temperature range between 300 and 1 000 K and in D<sub>2</sub> the pressure range between 10<sup>-20</sup> and 10<sup>5</sup> atm. Eight selected phases are indicated in the insets, but we identified 25 phases, differing in coverage and/or bond connectivity. The phases are characterized by three structural descriptors: the number of chemisorbed deuterium atoms  $N_D^c$ , the coordination histogram  $H_{\text{coord}}$  of the top-layer Si atoms and the dimerization type. A deuterium atom is regarded as chemisorbed onto the silicon surface when the distance to the closest Si atom is smaller than 2.0 Å. The value of the cutoff is determined as the first minimum in the D-Si radial distribution function of surface structures sampled by REGC (see the Supplemental Material [20]).  $H_{\text{coord}}$  is obtained by constructing a coordination histogram (distribution of Si-atoms coordination

number) for each surface configuration, i.e., the number of Si atoms bonded to each Si atom and the number of D atoms bonded to each Si atom. The dimerization type is either *order* when all the Si-Si dimers are formed on the same side (see, e.g., top left structure in Fig. 1), or *disorder* (e.g., top right structure). At each  $(T, p_{D_2})$ , the phases are distinguished by the different values of any of the three descriptors, and the phase at lowest free energy is reported on the  $T - p_{D_2}$  phase diagram.

An order-disorder transition is identified at constant coverage for  $\theta = \{0, 1/2, 13/16, 7/8, 1\}$ . For these coverages, the low-temperature, ordered phases consist of Si-Si dimer bonds arranged in stripes, while in the disordered phases, the dimer bonds break and form dynamically between top-layer Si atoms. The total number of dimer bonds remains on average the same across both transitions but the topology is different. Also at other coverages a order-disorder phase transition can exist, but the mentioned five coverages have  $(p, T)$  stability regions for such transitions. As mentioned, in the disordered phases, the dimer bonds are dynamic. For instance, the average lifetime of the bonds at 800 K is  $1.48 \pm 0.05$  ps and  $1.63 \pm 0.06$  ps for  $\theta = 0$  and  $\theta = 1$ , respectively. The longer bond lifetime for  $\theta = 1$  tells us that adsorbed D stabilizes the dimer bond, i.e., it makes the dynamics of dimer bonds slower (see the Supplemental Material [20] for details on this statistical analysis). As anticipated, another important feature of the phase diagram is that not all changes in coverage are accompanied by a phase boundary, i.e., some surface-structure transformations are smooth, for instance, between  $\theta = 7/8$  and  $\theta = 15/16$  above  $\sim 650$  K or between  $\theta = 3/8$  and  $\theta = 1/2$  above the critical point at  $\sim 500$  K. In the phase diagram, one can notice the change in color marking the change in coverage, but without a phase boundary in between. Also all the order-disorder phase transitions we found are not accompanied by a  $C_V(T)$  peak and are therefore smooth transitions.

We note that, in the present approach, the heat capacity cannot be evaluated between phases sampled in two different simulations. Therefore we are unable to determine whether there is a phase boundary between these phases. In the studied system, this is the case of the  $\theta = 4/3$  phase, simulated in the  $3 \times 3$  supercell, as compared with the other phases, all simulated in the  $4 \times 4$  supercell. For this reason, we have marked the boundary with a distinctive orange frame.

Finally, we address the relative stability of the different surface reconstructions and coverages. We fix our attention at a specific value of  $\mu_D = -0.1$  eV (red dashed line in Fig. 1), chosen to be inside the range of deuterium chemical potentials where both  $\theta = 1$  and  $\theta = 4/3$  are stable across a wide range of temperatures. As shown in Fig. 2, according to the REGC results, the higher saturation coverage  $\theta = 4/3$  is thermodynamically more stable below 359 K than that of  $\theta = 1$ , and the latter phase becomes

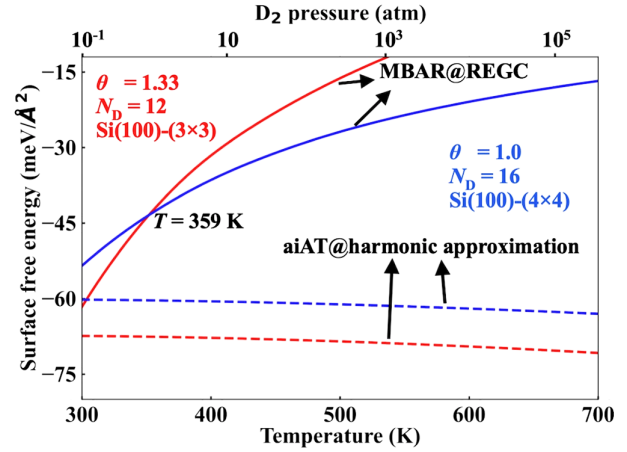


FIG. 2. Surface free energies of the Si(100) with different D coverage  $\theta = 4/3$  (red lines) and  $\theta = 1.0$  (blue lines) as a function of temperature at  $\mu_D = -0.1$  eV.  $\mu_D$  is chosen inside the range where both coverages are stable across a range of relevant  $T$ . This requirement is met with  $\mu_D$  in the range between  $-0.1$  eV and  $-0.01$  eV. The red or blue dashed lines are the surface free energy calculated by the *ab initio* thermodynamic method at harmonic approximation. The red or blue solid lines are the surface free energy calculated by the REGC method. The reference is the unreconstructed bare surface.

thermodynamically more stable when  $T > 359$  K. The predicted results of REGC are consistent with the experimental observations that a  $3 \times 1$  low-energy electron diffraction pattern is produced at  $380 \pm 20$  K, and a mild annealing to 600 K yields a very sharp  $2 \times 1$  pattern [26,31,32].

As a comparison, we show the prediction via aiAT for the same two phases, where the vibrational free energy is modeled via the harmonic approximation. The aiAT results show that two phases coexist in the temperature range from 300 K to 600 K, with only less than  $10$  meV/Å<sup>2</sup> free energy difference between two phases. The discrepancy between the two methods is ascribed to the anharmonic contributions. These manifest themselves in terms of a complex dynamics of the Si dimers, both in the pristine surface and in the high-coverage surfaces (see discussion of the order parameter dimerization type). Such surface restructuring is completely missed within the harmonic approximation, resulting in aiAT not only failing to account for the phase transition between  $\theta = 4/3$  and  $\theta = 1$  phases, but also in erroneously predicting that both coverages become more and more stable with respect to the pristine surface at increasing temperature. In fact, for both  $\theta = 4/3$  and  $\theta = 1$  phases, the slopes of MBAR@REGC surface free energies as a function of  $T$  and relative to the pristine surface are positive. The dynamical restructuring of Si-Si dimers has more (configurational) freedom in the  $4 \times 4$  supercell than that of the  $3 \times 3$  supercell; thus the  $\theta = 1$  phase becomes more stable at high temperature. This is also consistent with the shorter average Si-Si dimer bond

lifetime in the pristine compared with the terminated structures. No uniform  $1 \times 1$  dihydride phase is found to be stable at the studied conditions, which is consistent with experimental observation [16,33].

In conclusion, we have presented a first-principles theory of REGC molecular dynamics to evaluate and characterize the atomistic structure, composition, and geometry of surfaces in reactive environments, at technologically relevant  $(T, p)$  conditions, including vibrational free energies and all anharmonic effects. The capacity and strength of the approach is demonstrated by studying the phase diagram of the Si(100) surface in a deuterium atmosphere. The established  $(2 \times 1)$  and the controversial  $(3 \times 1)$  adsorption structures are found to be thermodynamically stable around  $T = 600$  and  $400$  K, respectively. Furthermore, a new, dynamic type of Si(100) surface reconstruction is identified at higher temperature, where Si-Si dimer bonds are dynamically restructuring. Specifically, three new adsorption patterns are found to be stable at higher temperature and pressure; these are a  $(2 \times 2)$  and two different  $(2 \times 4)$  monohydride structures. The approach not only rigorously accounts for the anharmonic vibrational contributions to the free energy but also quantitatively addresses phase boundaries including triple and critical points of a surface at catalytic  $(T, p)$  conditions. Several of the identified phases have not been found experimentally so far. Examples include low coverage and disorder phases. A possible reason for the lacking experimental results is that the REGC method accesses stable and dynamic surface restructuring in thermodynamic equilibrium with the given  $T$  and  $p_{D_2}$  while the experimental studies have been performed in a UHV chamber. Furthermore, it may be experimentally difficult to achieve full thermodynamic equilibrium within the applied measuring times. Our computational REGC approach provides an accurate and robust roadway for fully first-principles predictions of thermodynamic properties relevant for a plethora of important applications, such as heterogeneous catalysis, dopant profiles, surface segregation, and crystal growth.

In compliance with the FAIR-data principles, all the REGCMD trajectories are available in the NOMAD Repository and Archive, at the Ref. [34].

We thank Sergey Levchenko and Haiyuan Wang for useful discussions about the setup of the surface modeling. This project has received funding from the European Union's Horizon 2020 research and innovation program (No. 951786, the NOMAD Center of Excellence, and No. 740233, TEC1p) and the EPSRC Centre-to-Centre Project (Grant No. EP/S030468/1). C.Z. acknowledges funding from the National Natural Science Foundation of China under Grant No. 11704355. We acknowledge computational resources from the North German Supercomputing Alliance (HLRN).

\*zhou@fhi-berlin.mpg.de

- [1] E. Fabbri, M. Nachttegaal, T. Binninger, X. Cheng, B.-J. Kim, J. Durst, F. Bozza, T. Graule, R. Schäublin, L. Wiles *et al.*, *Nat. Mater.* **16**, 925 (2017).
- [2] A. Bergmann, T.E. Jones, E. Martinez Moreno, D. Teschner, P. Chernev, M. Gliech, T. Reier, H. Dau, and P. Strasser, *Nat. Catal.* **1**, 711 (2018).
- [3] K. Reuter, *Catal. Lett.* **146**, 541 (2016).
- [4] A. A. Correa, S. A. Bonev, and G. Galli, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1204 (2006).
- [5] X. Wang, S. Scandolo, and R. Car, *Phys. Rev. Lett.* **95**, 185701 (2005).
- [6] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1110 (2019).
- [7] C. M. Weinert and M. Scheffler, *Mater. Sci. Forum* **10**, 25 (1986).
- [8] M. Scheffler, *Physics of Solid Surfaces-1987*, edited by J. Koukal (Elsevier, Amsterdam, 1988).
- [9] K. Reuter and M. Scheffler, *Phys. Rev. B* **65**, 035406 (2001).
- [10] S. Bhattacharya, S. V. Levchenko, L. M. Ghiringhelli, and M. Scheffler, *Phys. Rev. Lett.* **111**, 135501 (2013).
- [11] Y. Zhou, M. Scheffler, and L. M. Ghiringhelli, *Phys. Rev. B* **100**, 174106 (2019).
- [12] E. Pehlke and M. Scheffler, *Phys. Rev. Lett.* **74**, 952 (1995).
- [13] P. Kratzer, E. Pehlke, M. Scheffler, M. B. Raschke, and U. Höfer, *Phys. Rev. Lett.* **81**, 5596 (1998).
- [14] C. Filippi, S. B. Healy, P. Kratzer, E. Pehlke, and M. Scheffler, *Phys. Rev. Lett.* **89**, 166102 (2002).
- [15] H. Neergaard Waltenburg and J. Yates, *Chem. Rev.* **95**, 1589 (1995).
- [16] J. J. Boland, *Surf. Sci.* **261**, 17 (1992).
- [17] Y. J. Chabal and K. Raghavachari, *Phys. Rev. Lett.* **53**, 282 (1984).
- [18] M. R. Shirts and V. S. Pande, *J. Chem. Phys.* **122**, 144107 (2005).
- [19] M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).
- [20] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.128.246101> for the benchmarks of surface model; the convergence of the REGC simulations; an example of bond coordination histogram and radial distribution function of Si-D and Si-Si; phase diagrams; heat capacity peaks at constant temperatures or at constant pressures;  $T$ - $p$  maps of coordination histogram, dimer type and heat capacity; tables contain the different coordination histograms of thermodynamical phases; vibrational free energy in harmonic approximations; Ref [21–24].
- [21] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, *Phys. Rev. Lett.* **108**, 236402 (2012).
- [22] A. Ramstad, G. Brocks, and P. J. Kelly, *Phys. Rev. B* **51**, 14504 (1995).
- [23] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, *Comput. Phys. Commun.* **180**, 1961 (2009).
- [24] L. Verlet, *Phys. Rev.* **159**, 98 (1967).

- [25] Y. Zhou, Surface phase diagrams including anharmonic effects via a replica-exchange grand-canonical method, Doctoral thesis, Technische Universität Berlin, Berlin, 2020.
- [26] Y. J. Chabal and K. Raghavachari, *Phys. Rev. Lett.* **54**, 1055 (1985).
- [27] G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- [28] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Comput. Phys. Commun.* **180**, 2175 (2009).
- [29] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **78**, 1396(E) (1997).
- [30] A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.* **102**, 073005 (2009).
- [31] T. Sakurai and H. D. Hagstrum, *Phys. Rev. B* **14**, 1593 (1976).
- [32] S. Maruno, H. Iwasaki, K. Horioka, S.-T. Li, and S. Nakamura, *Phys. Rev. B* **27**, 4110 (1983).
- [33] C. C. Cheng and J. T. Yates, Jr., *Phys. Rev. B* **43**, 4041 (1991).
- [34] [10.17172/NOMAD/2022.05.30-1](https://arxiv.org/abs/2022.05.30).