Self-learning kinetic Monte Carlo method: Application to Cu(111)

Oleg Trushin
Institute of Microelectronics and Informatics, Academy of Sciences of Russia, Yaroslavl 150007, Russia

Altaf Karim, Abdelkader Kara, and Talat S. Rahman*
Department of Physics, Cardwell Hall, Kansas State University, Manhattan, Kansas 66506, USA
(Received 13 April 2005; published 1 September 2005)

We present a method of performing kinetic Monte Carlo simulations that does not require an *a priori* list of diffusion processes and their associated energetics and reaction rates. Rather, at any time during the simulation, energetics for all possible (single- or multiatom) processes, within a specific interaction range, are either computed accurately using a saddle-point search procedure, or retrieved from a database in which previously encountered processes are stored. This self-learning procedure enhances the speed of the simulations along with a substantial gain in reliability because of the inclusion of many-particle processes. Accompanying results from the application of the method to the case of two-dimensional Cu adatom-cluster diffusion and coalescence on Cu(111) with detailed statistics of involved atomistic processes and contributing diffusion coefficients attest to the suitability of the method for the purpose.

DOI: 10.1103/PhysRevB.72.115401 PACS number(s): 68.43.Fg, 68.43.Hn, 68.43.Jk, 68.47.De

I. INTRODUCTION

The past decade has witnessed a surge in research activities which aim at bridging the gap in length and time scales at which a range of interesting phenomena take place. Some examples of such activities pertain to studies of epitaxial growth and nanostructuring of materials. The aim in such work is to utilize information obtained at the microscopic level to predict behavior at macroscopic scales. There are thus several key tasks to be undertaken, each of which is a challenge in itself. The first of these is an accurate determination of the energetics and dynamics of the system at the microscopic level. For selected systems this may be achieved through *ab initio* electronic structure calculations[1] which are becoming increasingly feasible for complex systems, even though they remain computationally intensive. A reasonable alternative, albeit not as reliable or accurate, is the application of one of several genres of many-body interatomic potentials.[2] With these interatomic potentials it has been possible to carry out computational and theoretical studies of a range of surface phenomena using techniques like molecular statics and molecular dynamics. Molecular dynamics simulations in particular are capable of revealing the essential details of microscopic phenomena as they unfold as a function of temperature, pressure, and other global variables but the application is limited in time and length scales. Since most thermally activated atomistic processes occur in the range of picoseconds, they are best captured with time steps in femtoseconds which limits total simulation time to a few microseconds. These times are many orders of magnitude smaller than processes happening in the laboratory. For example, epitaxial growth and surface morphological changes take place in minutes and hours and are controlled by atomic processes which are infrequent compared to atomic vibrational times of picoseconds. The challenge in molecular dynamics simulations is to find reliable ways that capture infrequent processes and extend to longer time scales with reasonable computational resources.

An alternative to molecular dynamics (MD) simulations for examining surface phenomena is offered by the kinetic Monte Carlo (KMC) technique in which the rates of various eligible atomic processes are provided as input.[3–5] If this input is accurate and complete, KMC simulations are in a good position to mimic experiments. Since the task of accumulating a complete set of atomic processes is nontrivial, standard KMC simulations are typically performed with a set of the most obvious simple atom or concerted processes as input, and all others either ignored or included in approximate ways (e.g., bond-counting models) or added in an *ad hoc* manner to fit experimental data. With a reduced set of barriers, activation energies become effective values rather than actual values, which may be compared with those obtained from experimental data but may not reveal the intervening microscopic processes. This is obviously problematic. Furthermore, it has been shown that unusual multiple-atom processes may play an important role in providing mass transport on surfaces such as Cu(100),[6,7] and Ir(111).[8,9] Any realistic simulation should have a provision for uncovering such processes and including their energetics in the evolution of the system.

To overcome these limitations of the two most common approaches for simulating temperature-dependent morphological evolutions of surfaces and interfaces, several accelerated schemes have been presented in recent times.[10–12] In a set of studies, Voter *et al.*[13,14] have concentrated on enhancing the time scales achievable in MD simulations through three different strategies: parallel replica, temperature-accelerated dynamics, and hyperdynamics. Fichthorn and coworkers, in related work, apply the bond-boost method[15] to extend the time scales in their simulations. The basic principle in these methods is to make the system evolve faster, sampling a larger phase space, through either smartly connected parallel processors, or by application of a boost so that the system can overcome energy barriers with relative ease, or by raising the temperature of the system. At the very least, infrequent processes may be revealed through such ac-

celerated schemes. The main issue is the assurance of one-to-one correspondence between the temporal evolution of the accelerated and nonaccelerated systems and whether the approach actually leads to a large computational speedup for a particular system of interest. The reader is referred to the original papers for further details and suitability of the techniques for specific cases.

Another promising scheme has focused on the completeness issue of KMC methods by allowing the system to evolve according to single- and multiple-atom processes of its choice. The key to the method is the generation of saddle points in the potential-energy surface and benefits from the advances that Jonsson and co-workers[16,17] have made in procedures for extracting diffusion paths and energy barriers using efficient search procedures. Once a large (sufficient) number of saddle points have been identified, the expectation is that the system will evolve naturally according to its inherent mechanisms. The method we propose here is in principle related to the latter approach, with a very important difference. We employ a pattern recognition scheme which allows efficient storage and subsequent retrieval of information from a database of diffusion processes, their paths, and their activation energy barriers. The procedure presented here is thus efficient and reliable. The removal of redundancies and repetitions in the calculations of energetics of system dynamics speeds up the simulations by several orders of magnitude, making it feasible for a range of applications. Since the generation of the database and its future usage through recognition patterns is akin to the simulation procedure learning from itself, we call the technique proposed here self-learning KMC (SLKMC). While the proposed technique can be applied to any surface systems, our interest is in the examination of atomistic phenomena as related to growth on fcc(111) surfaces. This is a challenging surface since the lack of surface corrugation makes the energy landscape relatively flat with a number of diffusion processes which are equally competitive. Some such atomistic processes may include those with multiatoms which are typically ignored in standard KMC techniques. In this paper we focus our attention on some characteristics of the proposed technique and its application to homoepitaxy on fcc(111) surfaces through consideration of the diffusion and coalescence of two-dimensional Cu adatom islands on Cu(111). The structure of the paper is as follows. In the next section we present some essentials of the self-learning KMC framework. This is followed in Sec. III with results of the applications of the method to examine morphological evolution of two-dimensional Cu islands on Cu(111). Section IV contains our conclusions.

II. ESSENTIALS OF SELF-LEARNING KINETIC MONTE CARLO METHOD

Although the principle of the proposed technique is generally applicable, we need a specific surface geometry to illustrate its details. For reasons mentioned above our interest is homoepitaxy on fcc(111) surfaces. We provide in this section some details of the model system, together with an outline of the standard kinetic Monte Carlo method for completeness. This is followed by a summary of the pattern recognition and labeling scheme that we invoke to obtain a self-learning KMC methodology.

A. Model system

To mimic the fcc(111) surface we consider a two-layer substrate, with periodic boundary conditions in the *XY* plane (which is parallel to the surface), which uniquely identifies the fcc and hcp hollow sites on the surface. The system of interest (such as an adatom island, vacancy island, or any other nanostructure whose morphological evolution or diffusion is to be determined) is placed on top of the substrate. In this initial study only occupancy of fcc sites (i.e., hollow sites with no atom in the layer below) on the substrate is allowed. While there is experimental justification for assuming fcc-site occupancy for Cu adatoms on Cu(111),[18] we are aware that on Ir(111) atoms may also occupy hcp sites (hollow sites with an atom in the layer below).[19] In fact, even for homoepitaxial growth on Cu(111) under certain other experimental conditions hcp-site occupancy has been reported.[20] Furthermore, adatoms, dimers, and other smaller clusters may use the hcp site as an intermediate[21] one during their motion. The method we are proposing can easily be generalized to include hcp occupancy. We are also assuming that the diffusion is via hopping . This restriction can be removed in future work. For the moment our interest is in the in-plane [two-dimensional (2D)] motion of adatoms, vacancies, and their clusters on Cu(111), for which diffusion is expected to proceed via hopping.

B. Some ingredients of kinetic Monte Carlo method

The goal of the kinetic Monte Carlo method is to mimic real experiments through sophisticated simulations. For these simulations to be realistic, one has to implement increasingly complex scenarios requiring intensive use of state-of-the art software and hardware. At the heart of a KMC simulation of the time evolution of a given system lie the mechanisms that are responsible for determining the microscopic processes to be performed at any given time. To illustrate the point, consider a system containing N particles at a given time with N_e possible types of processes. Let us also associate with each process type (i), n_i, the number of particles in the system that are candidates for this process type, the activation energy barrier ΔE_i, and a prefactor ν_i. The microscopic rate associated with process i, within transition state theory,[22] is then

$$r_i = \nu_i \exp(-\Delta E_i/kT), \tag{1}$$

where k is the Boltzmann constant and T the surface temperature. The total rate R of the system is further given by

$$R = \sum_{i=1}^{N_e} R_i, \tag{2}$$

where $R_i = n_i r_i$ is the macroscopic rate associated with process type i.

In KMC simulations, the acceptance of a chosen process is always set to 1. However, the choice of a given process is

dictated by the rates. First, a process type is chosen according to its probability $p_i=R_i/R$, and then a particle is randomly chosen from the set n_i to perform this process.

The essential elements of the KMC method are thus the processes i and their activation energy barriers ΔE_i whose determination requires a knowledge of the interatomic interaction which may be obtained from first principles or from model potentials. The validity of the method also hinges on that of the transition state theory whose applicability and limitations have been discussed in detail in a recent review.[23] In this paper, all activation energies are determined using interaction potentials based on the embedded-atom method (EAM) as developed by Foiles *et al.*[24] This is a semiempirical, many-body interaction potential. Although the EAM potentials neglect the large gradient in the charge densities near the surface and use atomic charge density for solids, for the six fcc metals Ag, Au, Cu, Ni, Pd, and Pt and their alloys, it has done a successful job of reproducing many of the characteristics of the bulk and the surface systems.[24]

To get back to the issue of the determination of diffusion processes, their paths, and their activation energy barriers, we should note that several interesting and appealing approaches have been proposed in the past few years. These methods include the nudged elastic band (NEB) method,[16] the step and slide method,[25] eigenvector following,[26] and temperature-accelerated MD.[27] Each of these methods has its own computational demand and measure of accuracy whose balance dictates the choice of the approach. For the studies presented in this paper, we find the simple "drag" method to be adequate, as we shall see. This is, of course, a rudimentary method in which the moving entity is dragged in very small steps toward the probable (aimed) final state. The dragged atom is constrained in the direction toward the aimed position while the other two degrees of freedom (perpendicular to this direction) and all degrees of freedom of the rest of the atoms in the system are allowed to relax. The other atoms are thus free to participate in the move, thereby activating many-particle processes (in which neighbor adatoms start to follow the central leading atom). In connection with the SLKMC method, the central atom is always dragged toward one of its vacant fcc sites. A more general way to map out the potential-energy surface is to use the grid method which has been successful in finding nontrivial diffusion paths and saddle points.[28]

C. Self-learning kinetic Monte Carlo method

As we have already mentioned, the limitation of the standard KMC method is its reliance on an *ad hoc* choice of processes and hence lack of completeness. For these reasons and also because of experimental observations of complex and unforeseen processes, the predictive power of the KMC method is in question. A rethinking of the way we perform the KMC simulations has become a necessity. Simulations with an *a priori* chosen catalog of processes need to be replaced by a continuous identification of possible processes as the environment changes. For these innovations in the KMC procedure, the local environment is the key issue and its complexities need to be exploited. With this in mind we are

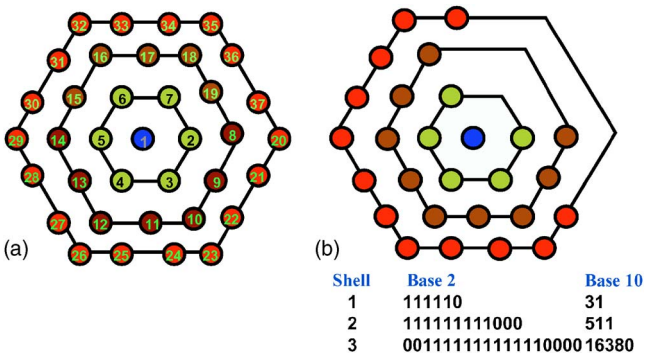

FIG. 1. (Color online) (a) The three-shell indexing around the central atom labeled 1; (b) signature of a particular 2D cluster configuration in base 2 and base 10.

proposing a methodology in which the base ingredient is the collection of local environments of undercoordinated atoms found automatically during the simulations and labeled and stored for subsequent usage in the simulation. As a concrete example of our approach we have chosen the fcc(111) surface which has a sixfold symmetry. For simplicity, we assume that any process in this system will involve a central (undercoordinated) atom and atoms in the next three shells as illustrated in Fig. 1. The motif in Fig. 1 is to serve as a "cookie cutter" and is placed on all active atoms in the system to define their local environment. We further assume, without loss of generality that any process may be described in terms of the central atom moving to a neighboring vacancy accompanied by the motion of any other atom or atoms in the three surrounding shells. The labeling of the surrounding atoms is done in binary and a base ten number is then associated with the first-shell configuration. The same procedure is followed for atoms in the second and third shells. Hence, for an atom in the system to be active (i.e., the central atom for a given process), it should have a vacancy in its first shell (or an occupancy number less than 63 for the cookie cutter), as illustrated in Fig. 1(b).

Once the atoms are classified as active and nonactive and encrypted within the three-shell scheme, we proceed by determining all possible processes associated with every active atom. Next the determination of the activation energy and prefactor is performed for all processes. Examples of how processes are labeled and stored in the database are given in Fig. 2. In this figure, full circles represent occupied sites and open circles vacancy sites. Figure 2(a) illustrates the "diffusion along a step" process where the central atom labeled 1 moves to the vacant site 2 along the step formed by atoms numbered 30, 15, 6, 7, 19, and 37 in the cookie cutter. The initial configuration for this process is recorded in base 10 as (48,3968,261 120) in the database and shown with the base 2 label in the figure. The move in Fig. 2(a) is recorded as atom 1 going to position 2 (1,2) and the activation energy barrier for the process in Fig. 2(a) is found to be 0.31 eV. Similarly, for the multiatom process illustrated in Fig. 2(b), the initial configuration in base 10 is recorded along with the sequence of motion of atoms involved in the process, which in this case is 1 going to 4, 6 to 1, and 15 to 5, which is recorded as (1,4;6,1;15,5). This multiatom process was found during the

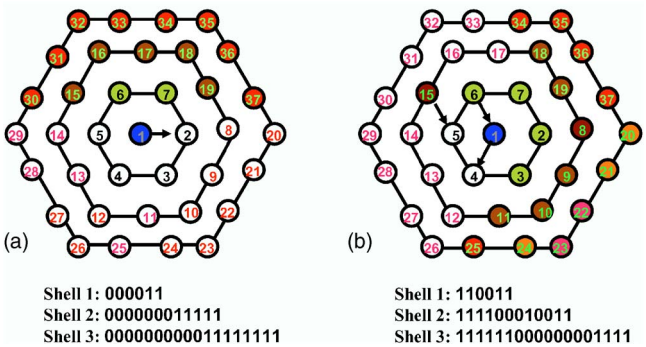

FIG. 2. (Color online) Sample (a) single-atom and (b) multiple-atom processes involved in the diffusion of 2D clusters presented with their specific labels for our database.

coalescence of two islands and will be discussed later. Its activation energy barrier of 0.595 eV is also recorded with the label.

The bottleneck for the simulation is the determination of the activation energy and the prefactor for all possible processes. Even when we make the widely used assumption that all the processes have the same prefactor, the calculation of the activation energy is very expensive if one needs accurate values. Note that since the activation energy is in the exponential, any small variation in the activation energy results in a substantial change in the relative probabilities and hence the outcome of the whole simulation. In standard KMC simulations these energy barriers are provided as input. If, however, as we and others[11] are proposing, these barriers are calculated on the fly, the process will be sped up if provisions are made to avoid recalculations. In the method proposed here this is achieved through the storage of activation energy barriers tagged to specific atomic processes in the database. This is the basis of our KMC method in which self-learning is achieved by the system through the ability to (1) calculate activation energies on the fly; (2) store them in a database; and (3) recognize and retrieve them using the labeling described above. Step 1 is not new. It was already proposed by Jonsson *et al.*[11] and Voter and Montalenti.[13] Steps 2 and 3 are, we believe, unique to our approach and help remove redundancies in the calculations. At any given time, after all the processes have been sorted out, a search for the activation energies in the database is launched. If a new process is encountered, the actual calculation is performed and this process with its activation energy is added to the database. Once the processes are classified and macroscopic rates are calculated, we proceed to perform one Monte Carlo step in which a randomly selected process is executed. The entire simulation process is summarized in the flow chart (Fig. 3). At later times in the simulation, when the system encounters environments for which some of the possible processes have been met earlier, a retrieval process of the activation energy from the database substitutes for the actual calculation. This gives a tremendous gain in the execution times as evident in our application to the diffusion of 2D Cu clusters on Cu(111). With modest computational resources, it was possible to carry out the simulation for a number of MC steps large enough to provide good statistics. The exact number of steps may vary from problem to problem.

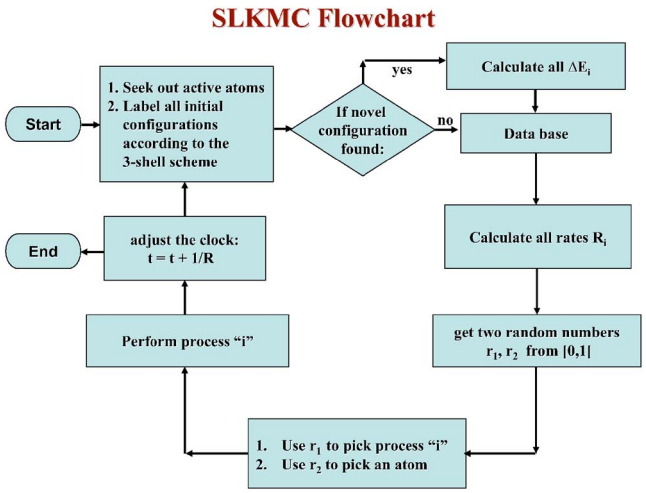

FIG. 3. (Color online) Flow chart for SLKMC simulation.

In the next section, we discuss some key features of the database collected during an extended simulation along with the results obtained from applying the SLKMC method to postdeposition analysis of homoepitaxy on Cu(111).

III. APPLICATION OF SLKMC METHOD TO MORPHOLOGICAL EVOLUTION OF 2D ISLANDS ON Cu(111)

Since the devil is generally in the details, we present below results of the application of the SLKMC method to study Cu cluster diffusion and coalescence on Cu(111). After giving some specifics of the model system, we present an analysis of the database which includes an evaluation of the accuracy of the calculated energy barriers and other factors affecting the simulation (CPU) time. We also comment on the presence and importance of multiatom processes. This is followed by the results and discussion of the diffusion and coalescence of 2D clusters on Cu(111).

A. Model systems

In the first example, i.e., the study of the diffusion of 2D adatom islands of Cu(111) we have chosen four specific sizes (19, 26, 38, and 100 Cu atoms) for which we already have results for comparison with a KMC simulation using a fixed database of logical processes involving single-atom periphery diffusion.[29] For the second application to the process of cluster coalescence, our model system consists of two adatom islands, one consisting of 78 atoms and the other 498 atoms placed on top of the two-layer substrate.

B. Examination of the collected database

To check the reliability of the data in the created database, we have compared in Table I the energy barriers that we obtained for some typical diffusion processes presented in Fig. 4, using both the drag and the NEB methods. We also include in the table values available in the literature. The comparison in the table attests to the reliability of the drag method as compared to the more time-consuming NEB pro-

TABLE I. Diffusion energy barriers for selected mechanisms as shown in Fig. 4.

Process	Drag method (eV)	NEB method (eV)	Ref. 31 (eV)
1*a*	0.68	0.66	
2*a*	0.53	0.52	
3*a*		0.65	
4*a*	0.25	0.25	
1*b*	0.60	0.59	0.59
2*b*	0.58	0.56	0.54
3*b*	0.68	0.67	0.67
4*b*	0.32	0.30	0.29

cedure. For example, with the drag method we were able to achieve speedup of at least an order of magnitude in the CPU time for the calculation of the energy barriers, as compared to one in which we applied the spherical repulsion method[30] to obtain the final states for a given initial state followed by application of the NEB method for the calculation of the activation energies.

As an illustration of the richness of the database that we collect, we plot in Fig. 5 the energy distribution of about 5000 diffusion processes which have been accumulated during a simulation containing several hundreds of millions of Monte Carlo steps. Note from Fig. 5 that the distribution is very wide, covering activation energies as small as a few tens of a meV to about 1 eV. Unlike the highly energetically corrugated surfaces like Cu(100),[31] energy barriers cannot be classified into groups. Note that in the calculations of the energy barriers differences are introduced when the effect of next nearest neighbors of the local environment is included in the calculation, as we have done. Note also that the accumulation of the database does not proceed uniformly with time, as reflected in the inset of Fig. 6. The SLKMC simulation starts, in this case, by accumulating about 400 different processes during the very first MC step, after which the database is "quasisaturated" for a certain period of CPU

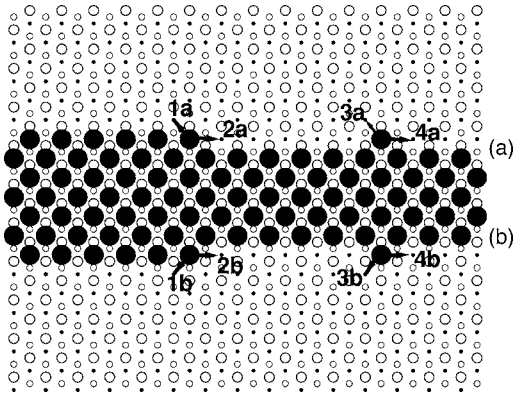

FIG. 4. Selected single-atom processes on the two types of steps *A* (100 microfacetted) and *B* (111 microfacetted), on fcc(111) surface. Process 1 is kink-detachment rounding, 2 is kink detachment along step, 3 is adatom detachment from step, and 4 is adatom diffusion along step. The labels *a* and *b* refer to steps *A* and *B*, respectively.

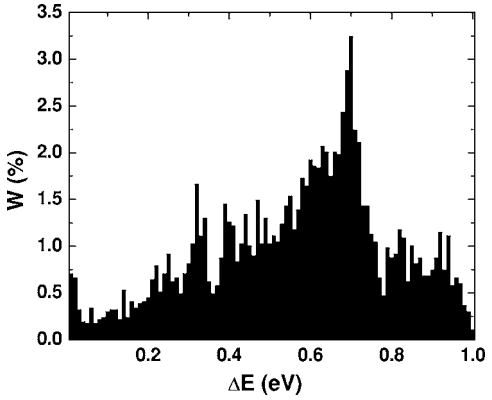

FIG. 5. Distribution (percentage) of activation energies of stored processes in the database during a SLKMC simulation.

time. This is followed by another phase of accumulation of about 600 processes, and so on. It is clear from the slope in Fig. 6 that when the simulation runs with a quasisaturated database, the number of KMC steps per CPU time increases dramatically. During a heavy buildup of the database, the yield is about 80 KMC steps per second and can go up to several thousands of KMC steps per second as the database saturates. The onset of new events in the database after a certain duration of simulation does raise the issue of measures that would assure that the database is complete. So far we have found the database to saturate after runs of about $100–500\times10^6$ MC steps. Actually, for the systems under study we have rarely found new processes to set in after 10×10^6 time steps.

One of the most important features of the method, as we have seen, is its ability to treat many-particle processes, the so-called concerted atomic motion. The recent version of the code allows inclusion of simultaneous displacements for atoms up to the third shell. From our simulations of several types of local environments (straight steps with kinks, compact islands, fractal-like islands) we found that in some cases

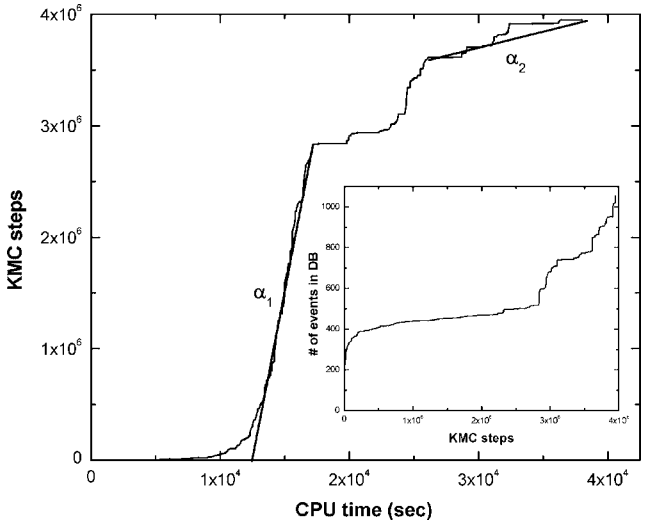

FIG. 6. Variations in the number of KMC steps per CPU time (i.e., performance) and the buildup of the databse as a function of the number of KMC steps (inset).

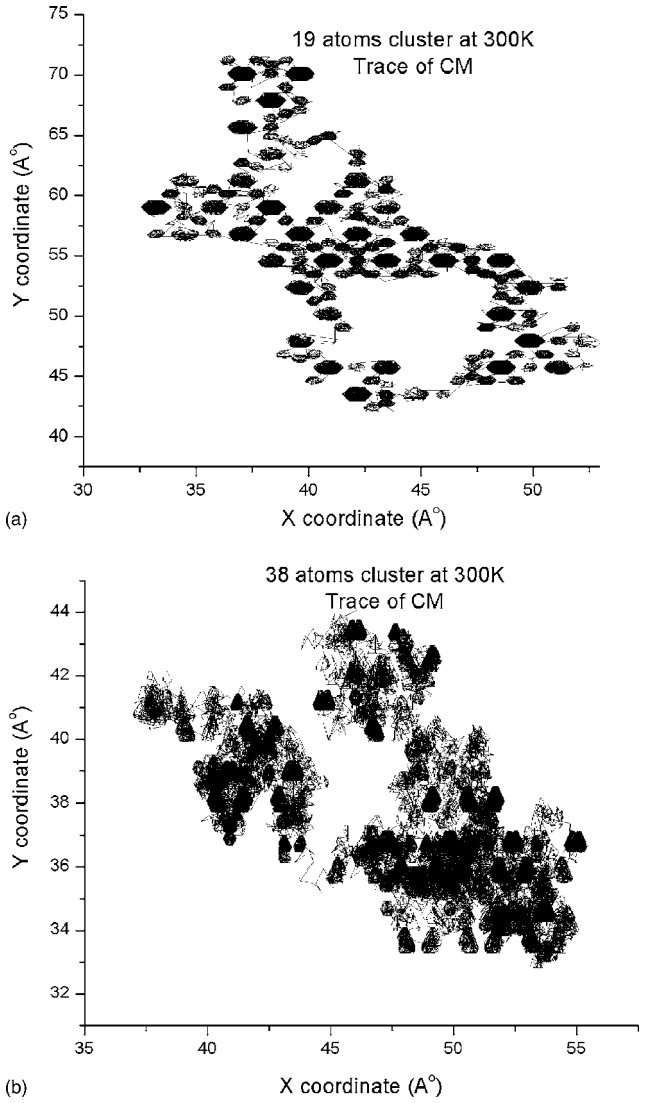

FIG. 7. Trace of center of mass of 19- and 38-atom Cu clusters on Cu(111) at 300 K as obtained from SLKMC simulations (10^7 steps).

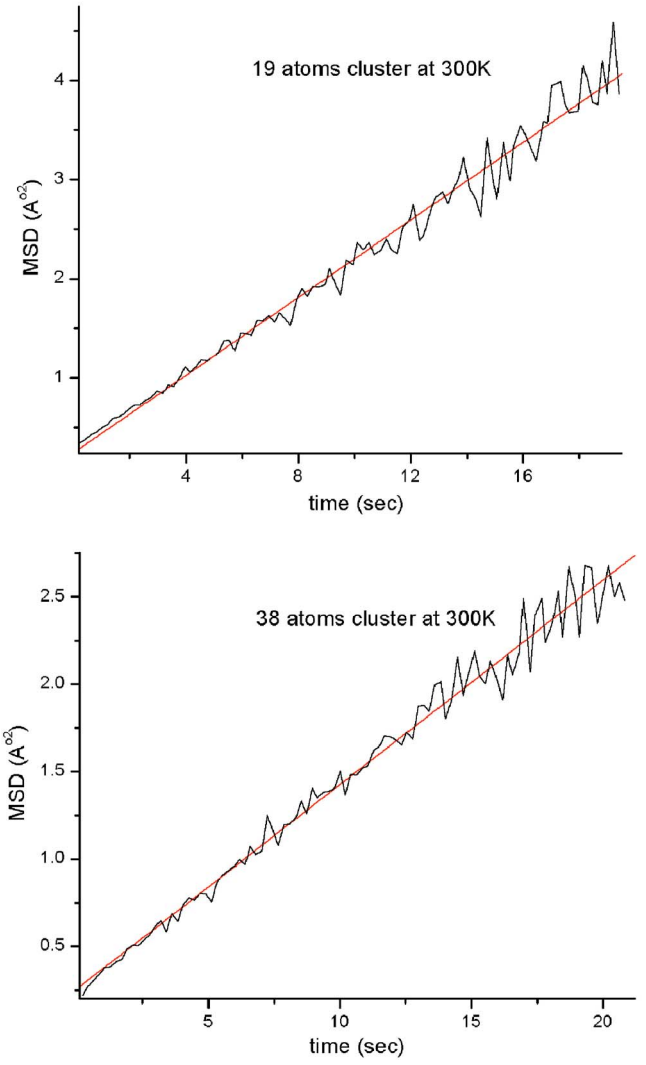

FIG. 8. (Color online) Mean square displacement (MSD) for 19- and 38 atom Cu clusters on Cu(111), as function of time at 300 K.

many-particle processes play an important role in providing atomic transport.[29] They are especially important in the case of low-coordination systems, like fractal islands. In such cases atoms are weakly bound and prefer to perform concerted motion rather than single atomic jumps. Furthermore, their importance increases with decreasing size of the cluster. In fact molecular dynamics simulations of a ten-atom Cu island on Cu(111) at 700 and 900 K show that the island moves by concerted displacement rather than through single-atom motion.[32] We next move onto examination of the results for two specific applications of the SLKMC method.

C. Morphological evolution

1. Diffusion of 2D islands

As a first application of the SLKMC method, we present results for the diffusion of 2D Cu islands on Cu(111) of four sizes: 19, 26, 38, and 100 atoms. These simulations were performed using $(10–100)\times 10^6$ MC steps at 300 and 500 K. During the simulation, the position of the center of mass was recorded at each MC step along with the performed process. After 10×10^6 MC steps, the islands have moved far enough that their diffusion coefficient may be extracted from the mean square displacement of the center of mass. In Fig. 7 we show the trace of the position of the center of mass on the (x,y) plane for both 19- and 38-atom clusters at 300 K. Note the dark spots for both cases indicating a stick-slip type of motion of the center of mass. The corresponding mean

TABLE II. Diffusion coefficient for 2D Cu islands on Cu(111) (Å^2/sec).

Cluster size	300 K	500 K
19	0.196	1.67×10^5
26	0.170	8.05×10^4
19	0.117	4.27×10^4
19	0.016	1.02×10^4

TABLE III. Frequencies of diffusion processes for the 19-atom cluster at two temperatures.

	Energy barrier (eV)		300 K		500 K	
Process	NEB	Drag	KMC	SLKMC	KMC	SLKMC
Step edge *A* [4(a)]	0.25	0.25	0.62	0.6797	0.42	0.511
Step edge *B* [4(b)]	0.30	0.31	0.17	0.0954	0.24	0.1403
Kink detach along step *A* [2(a)]	0.52	0.52	0.0	0.0	0.0020	0.0016
Kink detach along step *B* [2(b)]	0.56	0.54	0.0	0.0	0.0	0.0008
Kink detach along step (small) *A*	0.61	0.62	0.026	0.0106	0.012	0.0
Kink detach along step (small) *B*	0.68	0.69	0.0016	0.0007	0.0023	0.0018
Kink incorp. *A*	0.22	0.22	0.0	0.0001	0.0020	0.0025
Kink incorp. *B*	0.27	0.29	0.0	0.0	0.0	0.0009
Kink incorp. (small) *A*	0.01	0.01	0.025	0.0	0.011	0.002
Kink incorp. (small) *B*	0.08	0.11	0.0	0.0	0.0012	0.0
AA corner detachment		0.44		0.0007		0.0063
Kink detach out of step *B*	0.59	0.60	0.0	0.0091	0.0	0.0098
Kink fall into step *A*	0.07	0.10	0.0	0.0007	0.0	0.0016
Kink fall into step *B*	0.01	0.02	0.0	0.0109	0.0	0.0101
BB corner detachment		0.34		0.0322		0.0451
All multiple-atom processes				0.00015		0.0042
KESE *A*	0.37		0.0		0.0011	
Corner rounding at *AA* stage 1	0.31	0.33	0.0.	0.0001	0.0	0.0017
Corner rounding at *AA* stage 3	0.01	0.01	0.0	0.0	0.0	0.0017
Corner rounding at *BB* stage 1	0.37	0.39	0.0	0.0	0.0	0.0002
Corner rounding at *BB* stage 3	0.05	0.07	0.0	0.0	0.0	0.0002
Corner rounding at *AB* stage 1	0.32	0.33	0.066	0.0579	0.11	0.0894
Corner rounding at *AB* stage 2	0.08	0.11	0.0053	0.0023	0.024	0.0158
Corner rounding at *BA* stage 1	0.40	0.42	0.0047	0.0013	0.023	0.0095
Corner rounding at *BA* stage 2	0.015	0.02	0.067	0.0884	0.12	0.1348
AB corner detachment toward *B* step		0.62		0.0003		0.0017
AB corner detachment toward *A* step		0.69		0.0		0.0002

square displacements, for these two islands, as a function of time show a linear behavior (within statistical errors) and are shown in Fig. 8. The extracted slope from the mean square displacement plot gives the diffusion coefficient. In Table II, we report the diffusion coefficient for the four cluster sizes at 300 and 500 K. Note that the diffusion coefficient increases exponentially with temperature. The decrease of the diffusion coefficient with the cluster size follows a power law ($D=N^{-1.57}$ at 300 K and $N^{-1.64}$ at 500 K), which is in good agreement with previous results.[33] The virtue of our calculation is that the atomic processes leading to cluster diffusion were picked by the system itself during the simulation. The frequencies of the contributing processes vary with cluster size and, more importantly, with surface temperature (see Table III). Detailed descriptions of the processes in Table III are found in Ref. 35.

2. Island coalescence

As a second example of application of the SLKMC method, we present here results of simulation of the coalescence process in which two adatom islands join together to form a larger island with an equilibrium shape on Cu(111). This simulation was performed at 300 K using a small island containing 78 atoms and with an arbitrary shape, put close to a larger island containing 498 atoms with a circular shape. Successive snapshots of the system during the SLKMC simulation are shown in Fig. 9, for a total number of 40×10^6 KMC steps. From this figure, one notes that a neck between the two islands forms during the first 100 000 KMC steps, corresponding to a physical time of 0.25 s. After this time, the neck grows until the two islands form an elongated single island after about 10 s. Finally, the shape of the island evolves to a quasitriangle with mostly (111) steps (*B* type), which is a result of the asymmetry in the activation energy barriers associated with *A*- and *B*-type steps (see Table I). In order to get an insight into the mechanisms involved in the neck formation, we have analyzed the frequency distribution of key processes during the first and second 100 000 KMC steps. Three types of processes appear prominent in the coalescence of these two clusters: kink detachment on an *A*-type step (2*a* in Fig. 4), the reverse of 2*a* (labeled Rev. 2*a* in Table IV) also called kink incorporation, and diffusion along an *A*-type step (4*a* in Fig. 4). Listed in Table IV are the frequencies for these processes. We note from Table IV that during the formation of the neck, kink detachment and kink

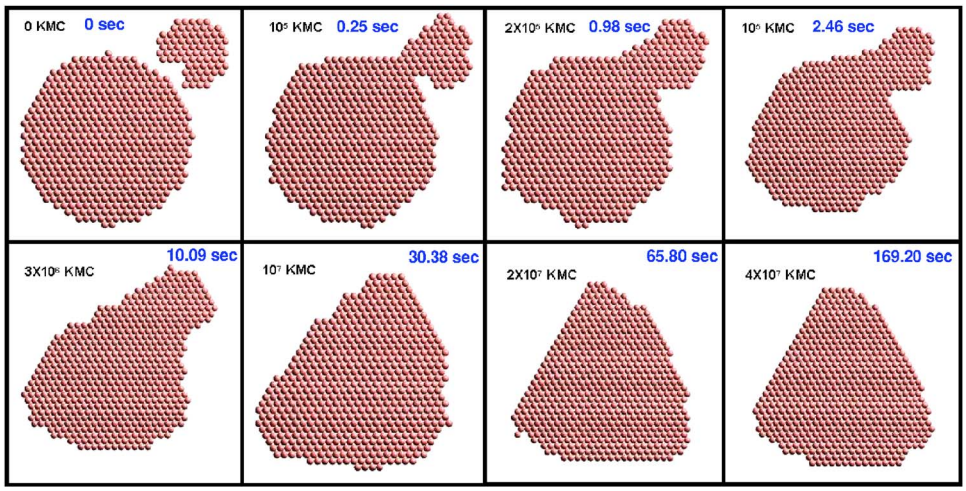

FIG. 9. (Color online) Coalescence of a small Cu cluster (78 atoms) with a larger one (498 atoms) on Cu(111) at 300 K, using SLKMC simulation (10^7 steps).

incorporation count for about 15% of the performed processes, another 70% involve diffusion along *A*-type steps, and other single- and multiple-atom processes including kink rounding and two-atom diffusion along steps constitutes the remaining 15%. For the second 100 000 KMC steps, the simulation is mostly dominated by diffusion along the *A*-type step (about 96%), with about 4% from various mechanisms. The important fact to note here is that kink-detachment and kink-incorporation contributions drop to almost zero after the neck has been formed. Detailed analysis of similar simulations involving islands of various sizes and shapes are actually in the processes of being performed and will be published elsewhere. A similar process for our simulations of cluster island coalescence are in qualitative agreement with the observations made by Giesen[34] using scanning tunneling microscopy.

IV. CONCLUSION

We have addressed the issue of completeness of KMC simulations by proposing a method in which the system finds, calculates, and collects the energetics of all possible diffusion processes that the moving entities are capable of performing. What separates our technique from others recently proposed is the provision for storing and retrieving the environment-dependent activation energy barriers from a database. Examination of the database shows that the simulation proceeds much faster when the set of processes is quasisaturated and that after sampling such regions the system has the ability to trigger the participation of new diffusion processes requiring enhanced CPU time for the calculation of new activation energy barriers. The system eventually settles down; the number of MC steps needed to do this depends on the system and the number of entries already in the database (about 10^7–10^8 steps). With the use of the pattern recognition scheme we are able to identify and calculate the frequency of occurrence of individual single- and multiple-atom diffusion processes that actually participate in the evolution of a particular entity. The microscopic details of the processes involved in surface morphological evolution can thus be documented for a system that has the freedom to evolve on their own accord. We show this through application to the diffusion and coalescence of 2D adatom islands on Cu(111) for which the simulation began with an empty database. Once a substantial accumulation has occurred, the simulation time speeds up by orders of magnitude and allows the calculation of system dynamics for time scales relevant to those phenomena happening in the laboratory. Interestingly, the two simple examples that we have presented here show that only a few dozen diffusion processes are in the end vital for a diffusion event. The question of course is, which ones? Our approach answers this question. As we have already alluded to, the task of calculating diffusion prefactors is still ahead of us. This is particularly important since we find many competing processes to differ only slightly in energy and differences in their vibrational entropy contributions to the prefactors can make a difference in the ultimate evolution of the film morphology. Another important result from our simulations with the open database is that dynamical evolution of the system with prejudged diffusion processes may yield erroneous results. Also, the pattern recognition schemes are a prudent way to develop a database of diffusion processes and their energetics. It does involve a lot of work in the beginning but once the database is compiled, it can be used for any type of simulation of the system. Of course, for realistic simulations of thin films we need to incorporate exchange and other processes which involve motion in 3D. We have already alluded to the importance of the inclusion of occupancy of the hcp site. Efforts are currently under way to include hcp sites in the pattern recognition scheme. In fact, preliminary results have already been obtained for the diffusion of small clusters (2–10 atoms) in

TABLE IV. Frequency of selected processes during the coalescence of two islands.

Process	Barrier (eV)	Frequency $[(0–1)\times 10^5$ steps]	Frequency $[(1–2)\times 10^5$ steps]
2*a*	0.530	7.41	0.03
Rev. 2*a*	0.220	8.43	0.04
4*a*	0.25	69.66	95.88
Others		14.50	4.05

which the SLKMC code allowed both fcc and hcp occupancy.[35,36] In the same vein, this work focuses on homoepitaxy. But this is not a limitation of the method, as with well-defined changes the SLKMC method can be adopted to extend to heteroepitaxy. Another development worth mentioning is the ability to perform simulations off lattice. For such a scheme we are introducing lattice discretization only for the pattern recognition part. In other words we apply rigid lattice geometry only locally around the central active atom. Inclusion of these and related changes in the SLKMC code is opening the way for the application of the technique to a multitude of phenomena.

AKNOWLEDGMENTS

We thank James Evans, Chandana Ghosh, and Ahlam Alrawi for helpful discussions. This work was supported by NSF Grants No. CRDF RU-P1-2600-YA-04, No. ERC 0085604, and No. ITR 0428826.

*Corresponding author. Email address: rahman@phys.ksu.edu

[1] B. D. Yu and M. Scheffler, Phys. Rev. Lett. **77**, 1095 (1996).
[2] F. Ercolessi, M. Parrinello, and E. Tosatti, Philos. Mag. A **58**, 213 (1988); M. S. Daw, S. M. Foiles, and M. I. Baskes, Mater. Sci. Rep. **9**, 251 (1993).
[3] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
[4] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[5] A. F. Voter, Phys. Rev. B **34**, 6819 (1986).
[6] O. S. Trushin, P. Salo, and T. Ala-Nissila, Phys. Rev. B **62**, 1611 (2000).
[7] P. Salo, J. Hirvonen, I. T. Koponen, O. S. Trushin, J. Heinonen, and T. Ala-Nissila, Phys. Rev. B **64**, 161405(R) (2001).
[8] S. C. Wang, U. Kurpick, and G. Ehrlich, Phys. Rev. Lett. **81**, 4923 (1998).
[9] S. C. Wang and G. Ehrlich, Phys. Rev. Lett. **79**, 4234 (1997).
[10] M. R. Sorensen and A. F. Voter, J. Chem. Phys. **112**, 9599 (2000).
[11] G. Henkelman and H. Jonsson, J. Chem. Phys. **115**, 9657 (2001).
[12] G. Henkelman and H. Jonsson, Phys. Rev. Lett. **90**, 116101 (2003).
[13] A. F. Voter, F. Montalenti, and T. C. Germann, Annu. Rev. Mater. Res. **32**, 321 (2002).
[14] F. Montalenti and A. F. Voter, J. Chem. Phys. **116**, 4819 (2002).
[15] R. A. Miron and K. A. Fichthorn, J. Chem. Phys. **119**, 6210 (2003); Phys. Rev. Lett. **93**, 128301 (2004).
[16] H. Jónsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne *et al.* (World Scientific, Singapore, 1998).
[17] G. Henkelman and H. Jonsson, J. Chem. Phys. **115**, 7010 (1999).
[18] M. Giesen and H. Ibach, Surf. Sci. **529**, 135 (2003).
[19] S. C. Wang and G. Ehrlich, Surf. Sci. **239**, 301 (1990).
[20] J. Camarero, J. D. L. Figuera, J. J. D. Miguel, R. Miranda, J. Alvarez, and S. Ferrer, Surf. Sci. **459**, 191 (2000).
[21] J. Repp, G. Meyer, K. H. Rieder, and P. Hyldgaard, Phys. Rev. Lett. **91**, 206102 (2003).
[22] S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes* (McGraw-Hill, New York, 1941); D. A. King, J. Vac. Sci. Technol. **17**, 241 (1980).
[23] T. Ala-Nissila, S. C. Ying, and R. Ferrando, Adv. Phys. **51**, 949 (2002).
[24] S. M. Foiles, M. I. Baskes, and M. S. Daw, Phys. Rev. B **33**, 7983 (1986).
[25] R. A. Miron and K. A. Fichthorn, Mol. Simul. **30**, 273 (2004).
[26] L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999), and references therein.
[27] A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).
[28] T. S. Rahman, A. Kara. A. Karim, and A. Al-Rawi, in *Collective Diffusion on Surfaces: Correlation Effects and Adatom Interactions*, edited by M. C. Tringides and Z. Chvoj (Kluwer Academic Publishers, Dordrecht, 2001), p. 327.
[29] T. S. Rahman, A. Kara, A. Karim, and O. Trushin, in *Modeling of Morphological Evolution at Surfaces and Interfaces*, edited by J. Evans, C. Orme, M. Asta, and Z. Zhang, MRS Symposia Proceedings No. 859E (Materials Research Society, Warrendale, Pennsylvania 2004).
[30] O. S. Trushin, P. Salo, T. Ala-Nissila, and S. C. Ying, Phys. Rev. B **69**, 033405 (2004).
[31] M. Karimi, T. Tomkowski, G. Vidali, and O. Biham, Phys. Rev. B **52**, 5364 (1995).
[32] A. Al-Rawi and T. S. Rahman (unpublished).
[33] A. Bogicevic, C. Liu, J. Jacobsen, B. Lundqvist, and H. Metiu, Phys. Rev. B **57**, R9459 (1998).
[34] M. Giesen, Surf. Sci. **441**, 391 (1999).
[35] C. Ghosh, A. Kara, and T. S. Rahman (unpublished); C. Ghosh, Ph.D. thesis, Kansas State University, 2003.
[36] P. Vikulov, O. Trushin, A. Kara, and T. S. Rahman (unpublished).