May 2008

Coarse-grained lattice kinetic Monte Carlo simulation of systems of strongly interacting particles

Jianguo Dai
University of Pennsylvania

Warren D. Seider
University of Pennsylvania, seider@seas.upenn.edu

Talid Sinno
University of Pennsylvania, talid@seas.upenn.edu

Coarse-grained lattice kinetic Monte Carlo simulation of systems of strongly interacting particles

Abstract
A general approach is presented for spatially coarse-graining lattice kinetic Monte Carlo (LKMC) simulations of systems containing strongly interacting particles. While previous work has relied on approximations that are valid in the limit of weak interactions, here we show that it is possible to compute coarse-grained transition rates for strongly interacting systems without a large computational burden. A two-dimensional square lattice is employed on which a collection of (supersaturated) strongly interacting particles is allowed to reversibly evolve into clusters. A detailed analysis is presented of the various approximations applied in LKMC coarse graining, and a number of numerical closure rules are contrasted and compared. In each case, the overall cluster size distribution and individual cluster structures are used to assess the accuracy of the coarse-graining approach. The resulting closure approach is shown to provide an excellent coarse-grained representation of the systems considered in this study.

Keywords
lattice theory, Monte Carlo methods

Coarse-grained lattice kinetic Monte Carlo simulation of systems of strongly interacting particles

Jianguo Dai, Warren D. Seider, and Talid Sinno[a)]
Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

(Received 12 December 2007; accepted 1 April 2008; published online 15 May 2008)

A general approach is presented for spatially coarse-graining lattice kinetic Monte Carlo (LKMC) simulations of systems containing strongly interacting particles. While previous work has relied on approximations that are valid in the limit of weak interactions, here we show that it is possible to compute coarse-grained transition rates for strongly interacting systems without a large computational burden. A two-dimensional square lattice is employed on which a collection of (supersaturated) strongly interacting particles is allowed to reversibly evolve into clusters. A detailed analysis is presented of the various approximations applied in LKMC coarse graining, and a number of numerical closure rules are contrasted and compared. In each case, the overall cluster size distribution and individual cluster structures are used to assess the accuracy of the coarse-graining approach. The resulting closure approach is shown to provide an excellent coarse-grained representation of the systems considered in this study. © *2008 American Institute of Physics*.
[DOI: 10.1063/1.2913241]

I. INTRODUCTION

The lattice kinetic Monte Carlo (LKMC) method is an efficient approach for simulating dynamical evolution in microscopic systems such as microstructural evolution in crystals, point defect clustering in silicon,[1–3] phase segregation in metallic alloys,[4,5] and surface morphological evolution during vapor deposition.[6–9] The primary advantage of LKMC is that it effectively circumvents the vibrational motion of atoms that is responsible for the temporal bottleneck in molecular dynamics. Nonetheless, fully atomically resolved LKMC, in which one lattice site is assigned to each LKMC degree of freedom, is still highly constrained in scope and is limited to the micron scale for many problems of interest.

Recently, there has been much interest in the development of coarse-grained LKMC (CGLKMC) simulations in which multiple atomic lattice sites are spatially grouped together into coarse-grained cells. The coarse-grained system temporally evolves by a sequence of coarse-grained events that are related to the microscopic events in the atomic system.[10–17] A key step in formulating such approaches is the coarse-graining methodology and, most importantly, the *closure rule*, both of which dictate how the single-atom events are averaged within a coarse-grained cell.

In this work, we present a detailed analysis of a LKMC coarse-graining approach based on extensions to the framework previously published by Vlachos and co-workers.[10–14] The study is carried out by using a two-dimensional square lattice on which a system of strongly interacting particles is allowed to evolve (reversibly) into clusters. A novel coarse-graining framework and several new closure approaches are developed to account for the heterogeneous environments within each coarse cell that arise because of strong interparticle interactions. We consider averaging approaches to accurately account for both intracell interactions (i.e., between particles within the same coarse cell) and intercell interactions (between particles in different coarse cells). In so doing, we demonstrate the limitations of existing coarse-graining closure rules, such as the local mean-field (LMF) approximation,[18] which assumes that particles in a given coarse cell are homogeneously distributed.[13] In each case, the overall evolution of the particle cluster size distribution and the individual cluster structures that are generated are used to assess the accuracy of the proposed coarse-graining approaches by comparing the predictions with the results of the "atomically" resolved LKMC (i.e., fine-grid LKMC or FGLKMC) simulations.

The remainder of this paper is organized as follows. In Sec. II, we describe the FGLKMC model that is used as the basis for coarse graining. In Sec. III, the basic elements of the coarse-graining framework are discussed. Approaches to treat strong interparticle interactions by using new numerical closure approximations are discussed in Secs. IV and V. Test applications of the developed coarse-graining framework are discussed in Sec. VI. Finally, a brief discussion of the computational gains associated with coarse graining and conclusions are presented in Secs. VII and VIII, respectively.

II. FINE-GRID LKMC MODEL FOR 2D SYSTEMS OF INTERACTING PARTICLES (FGLKMC)

The LKMC simulations employed in this work are based on a broadly applicable bond-counting approach for the calculation of the energy barrier associated with single-atom hops.[16,19] We have demonstrated in previous work that it is possible to treat continuous systems within LKMC by effectively including configurational entropy associated with off-

[a)]Author to whom correspondence should be addressed. Electronic mail: talid@seas.upenn.edu.

lattice relaxations into a bond-counting LKMC model.[1] In the present work, we limit our discussion to an idealized two-dimensional square lattice system with mobile interacting particles, although the proposed approach should be readily generalizable to three-dimensional lattices with different symmetries.

In the bond-counting model employed here, the energy barrier associated with a particle hop originating from a site, x, is calculated based on Arrhenius dynamics.[12] In Arrhenius dynamics, the energy barrier, which is equivalent to the total binding energy at that site, is a function of the initial but not the final configuration, i.e.,

$$E_x = \sum_{y \neq x} J|x-y|\sigma_x\sigma_y, \tag{1}$$

where the sum is taken over all sites in the system, $J|x-y|$ is the interaction potential between sites x and y, and σ_x is the occupancy state (0 or 1) at site x. It is often convenient to calculate the binding energy associated with a particular *particle* (i.e., only for occupied sites where $\sigma=1$) rather than a site (which can either be occupied or not). The binding energy of a given particle is defined by

$$E_i^b = \sum_{\substack{j=1 \\ j \neq i}}^{N_p} J(i,j), \tag{2}$$

where $J(i,j)$ is the potential energy between particles i and j, and the sum is taken over all particles in the system within interaction range. Note that for particle i residing on site x, $E_i^b = E_x$. The hopping rate from site x is therefore given by

$$\hat{r}_x = \sigma_x \nu_0 \exp\left(-\frac{E_x}{k_B T}\right), \tag{3}$$

where ν_0 is the attempt frequency, k_B is the Boltzmann constant, and T is the system temperature. We also define a conditional hopping rate which is the transition rate from site x given that site x is filled, such that $r_x \equiv (\hat{r}_x | \sigma_x = 1)$, and therefore

$$r_x = \nu_0 \exp\left(-\frac{E_x}{k_B T}\right). \tag{4}$$

III. GENERAL CONCEPTS FOR COARSE-GRAINING LKMC SIMULATIONS

The principal concept within a spatially CGLKMC framework is that several adjacent fine-grid lattice sites are grouped to form a single coarse cell. The collection of coarse cells forms a new coarse lattice, which in this work is assumed to be uniformly spaced, i.e., we apply a single level of coarse graining throughout the domain. The relationship between the fine-grid and a coarse-grid lattice is shown in Fig. 1 for the two-dimensional square lattice. Each coarse cell contains q fine-grid (lattice) sites and is characterized by an occupancy number, $0 \leqslant \eta \leqslant q$.

The central assumptions implicit in constructing a CGLKMC approximation have been discussed in detail by Vlachos and co-workers.[10–14] In brief, it is assumed that the microscopic processes within a coarse cell are fast compared to the overall system dynamics, i.e., that local quasiequilibrium is maintained within a coarse cell at all times. This quasiequilibrium is a function only of the local mesoscopic system properties such as local mass, temperature, and potential energy[20] and does not explicitly depend on time. The system degrees of freedom corresponding to these fast microscopic processes therefore are "averaged out" of the overall dynamical equations that govern the CGLKMC system. The mesoscopic properties that describe the state of each coarse cell are, however, explicit functions of time, evolving (relatively slowly) according to the remaining degrees of freedom in the system. In essence, a timescale separation is achieved, similar in nature to a pseudo-steady-state approximation made in continuous systems.[21]

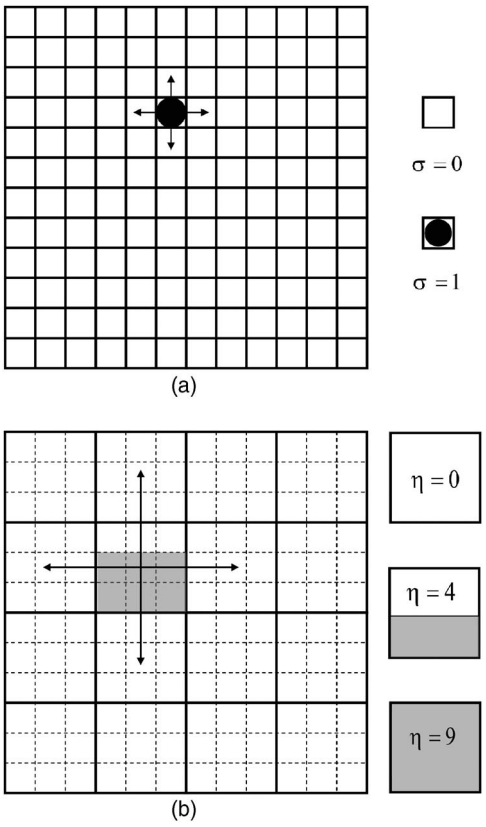

FIG. 1. Schematic of (a) fine-grid lattice and (b) corresponding coarse-grained lattice ($q=9$).

A. General expressions for the coarse hopping rate

The essence of the coarse-graining procedure is to compute the overall jump rate for a particle from an origination cell k to an adjacent destination cell l_d on the coarse lattice, under the assumption that the local microscopic system within each coarse cell is in equilibrium with respect to the current local mesoscopic conditions (namely, the number of particles and the temperature in the coarse cell). The total rate for this event is given by

$$\bar{r}(k \to l_d) = \sum_{x \in D_k} \frac{1}{q}\left(\sum_{y \in D_{ld}} (1-\sigma_y)\sigma_x r_x\right) = \sum_{x \in D_k} \langle (1-\sigma_y)\sigma_x r_x \rangle_y = q\langle (1-\sigma_y)\sigma_x r_x \rangle_{xy}, \tag{5}$$

where $\langle\ \rangle_y$ is the average over the y sites in cell l_d, and $\langle\ \rangle_{xy} \equiv \langle\ \rangle$ represents the average over all x and y sites. By assuming that all spatial correlations can be accounted for in the calculation of the average transition rate (ATR), $\langle r_x \rangle$, $\langle \sigma_x \rangle$, and $\langle \sigma_y \rangle$ in Eq. (5) can be replaced by η_k/q and η_{ld}/q, respectively, where η_k and η_{ld} are the occupancy numbers of cell k and cell l_d. The average coarse hopping rate from cell k to cell l_d is then given by

$$\bar{r}(k \to l_d) = \frac{\eta_k}{q}\left(1 - \frac{\eta_{ld}}{q}\right) q \langle r_x \rangle. \tag{6}$$

By applying Eqs. (4)–(6), the rate equation for the coarse-grained hop from cell k to cell l_d is therefore given by

$$\bar{r}(k \to l_d) = \frac{\eta_k}{q}(q - \eta_{ld}) \nu_c \left\langle \exp\left(-\frac{E_x}{k_B T}\right) \right\rangle, \tag{7}$$

where it is understood that the term in the $\langle\ \rangle$ is evaluated only if site x is occupied. In fact, this conditional average is simply equivalent to the average over the particles on the coarse cell k—we therefore use particle-averaging notation in the remainder of this paper. Also note that in Eq. (7), $\nu_c = \nu_0/q$, which reflects the fact that the hop length on the coarse lattice is equal to the distance between the centers of the coarse cells. For example, for q sites per coarse cell, the length of one hop effectively increases from a on the fine-grid lattice to $a\sqrt{q}$ on the coarse-grid lattice.

As shown in Eq. (7), once the average of the rates corresponding to all possible microconfigurations is computed, the hopping rate in the coarse-grained system is only a function of the coarse variables, η_k and η_{ld}, as required in a CGLKMC simulation. How this average is computed defines the particular type of coarse-graining approach. In the following section, we discuss the validity of various assumptions used in the calculation of this average and present a new numerical approach that leads to excellent results even in the limit of strong interactions.

B. Common approximations in spatial coarse graining of lattice systems

Two approximations have been generally employed in literature to simplify the calculation of the ensemble average in Eq. (7). In the first approximation, it is assumed that the ATR can be approximated by an average over the *binding energies* of each particle in each microconfiguration, i.e.,

$$\left\langle \exp\left(-\frac{E_i^b}{k_B T}\right) \right\rangle \approx \exp\left(-\frac{\langle E_i^b \rangle}{k_B T}\right). \tag{8}$$

The approximation is useful because the average binding energies (ABEs) per particle, $\langle E_i^b \rangle$, are the direct input to CGLKMC simulations. We refer to this approximation as the ABE, which has been applied in all previous LKMC coarse-graining models in literature, with the exception of the multigrid approach proposed by Katsoulakis *et al.*[12] In this approximation, the coarse-grained rate is thus given by

$$\bar{r}(k \to l_d) = \frac{\eta_k}{q}(q - \eta_{ld}) \nu_c \exp\left(-\frac{\langle E_i^b \rangle}{k_B T}\right). \tag{9}$$

It is intuitive to assume that particles with high binding energies should contribute less to the ATR from a coarse cell. On the other hand, in the ABE approximation [Eq. (8)], all particles are assumed to equally contribute to the ABE, leading to an error that should increase with the interaction strength. In the exact general expression, Eq. (7), the average is directly taken over the transition rates, which naturally biases the overall average toward particles (or sites) with low binding energies. We refer to the direct averaging of transition rates as the ATR framework.

Nonetheless, the ABE approximation is mathematically consistent with the ATR framework. With $\tilde{E}_i^b = E_i^b/k_B T$, the weak interaction limit is represented by $\tilde{E}_i^b \ll 1$. Expanding the transition rate into a Taylor series at this limit gives

$$\exp(-\tilde{E}_i^b) \approx 1 - \tilde{E}_i^b + O((\tilde{E}_i^b)^2). \tag{10}$$

Accordingly, a first-order approximation of the ensemble average of the transition rate in Eq. (7) (neglecting the pre-exponential factor) is

$$\langle \exp(-\tilde{E}_i^b) \rangle \approx \langle 1 - \tilde{E}_i^b \rangle = 1 - \langle \tilde{E}_i^b \rangle \approx \exp\langle -(\tilde{E}_i^b) \rangle, \tag{11}$$

where $\langle \tilde{E}_i^b \rangle$ is the approximate ensemble average of the dimensionless binding energy for one particle. The preceding shows the origin of the ABE approximation in Eq. (9). Note that as the dimensionless binding energy approaches zero, the two averaging frameworks (ABE and ATR) become identical.

1. Interaction decoupling approximations

The second approximation used to simplify the calculation of the averages in Eq. (7) or Eq. (9) is the assumption that particle interactions within the same cell and those on different cells can be separated. This is tantamount to assuming that the total binding energy associated with one particle in a coarse-grid lattice can be decomposed into two parts: The *intracell* interaction, which represents binding to other particles within the same coarse cell, and the *intercell* interaction, which represents binding to particles in neighboring coarse cells. Most generally, the binding energy per particle in a given coarse cell k is parametrically given by

$$\begin{aligned} E_i^b = E_i^b(k, &\{l_{1\alpha}, \alpha = 1, \ldots, n_1\}, \\ &\{l_{2\alpha}, \alpha = 1, \ldots, n_2\}, \ldots, \{l_{N\alpha}, \alpha = 1, \ldots, n_N\}, \end{aligned} \tag{12}$$

where l_i represents coarse cells at neighbor shell i, α is the index representing a particular coarse cell in neighbor shell i, n_i is the number of coarse cells at each (coarse lattice) neighbor shell i, and N is the total number of neighbor shells that must be considered, which is set by the range of the fine-grid potential. The separation assumption suggests the following general expression for the per-particle binding energy:

$$E_i^b = E_i^{\text{intra}} + E_i^{\text{inter}}. \tag{13}$$

Moreover, the intercell contribution can be further separated into contributions over each of the neighbor shells on the coarse lattice so that

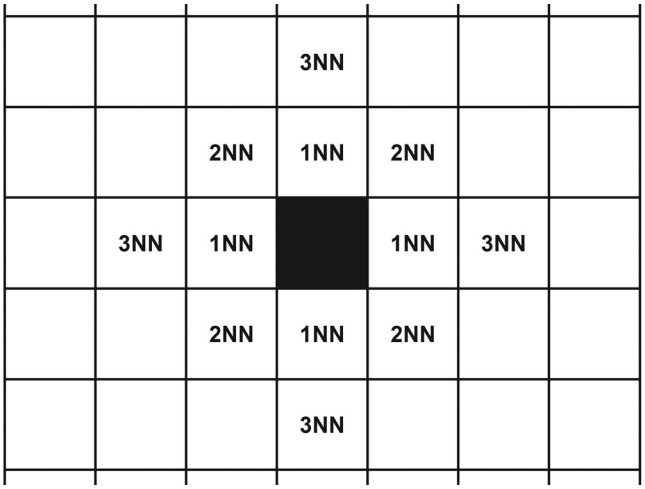

FIG. 2. Schematic showing the neighboring coarse cells on a cubic two-dimensional coarse-grid lattice. The dark cell in the center is the reference cell.

$$E_i^{\text{inter}} = \sum_{j=1}^{N} E_i^{\text{inter},l_j}, \tag{14}$$

where the index l_j refers to specific coarse cells at the jth neighbor shell of coarse cell k. A schematic of the neighbor shells around a reference coarse cell in a cubic, two-dimensional coarse lattice is shown in Fig. 2.

Note that the two components in Eq. (13) are also, in general, functions of particle distributions in multiple cells but the implied separation allows for systematic decoupling approximations to be made. For example, while the intracell particle distribution is still, in principle, a function of particle distributions in all neighboring coarse cells, particles in the same cell are (on average) much closer to each other than to those in neighboring cells. As a result, an important decoupling assumption can be made: The intracell binding energies between particles in reference cell k are only weakly affected by the surrounding cells and the cell can be considered to be isolated as shown in Fig. 3(a), i.e.,

$$E_i^{\text{intra}} = E_i^{\text{intra}}(k,\{l_{1\alpha},\alpha=1,\ldots,n_1\},$$
$$\{l_{2\alpha},\alpha=1,\ldots,n_2\},\ldots,\{l_{N\alpha},\alpha=1,\ldots,n_N\} \sim E_i^{\text{intra}}(k). \tag{15}$$

Similar approximations can be made for the various intercell contributions. Consider the first-nearest-neighbor (1NN) intercell interactions $E_i^{\text{inter},\,l_1}$. As before, these interactions are coupled to both the intracell interaction and to intercell interactions at other neighbor shells on the coarse lattice, i.e.,

$$E_i^{\text{inter},l_1} = E_i^{\text{inter},l_1}(k,\{l_{1\alpha},\alpha=1,\ldots,n_1\},$$
$$\{l_{2\alpha},\alpha=1,\ldots,n_2\},\ldots,\{l_{N_\alpha},\alpha=1,\ldots,n_N\}. \tag{16}$$

By assuming that the coupling to intercell distributions between cell k and other neighbor shells is weak, the interaction between cell k and coarse cells at the second (and other) neighbor shells is not expected to strongly affect the interactions between cell k and the first neighbor shell. The validity of this approximation will be discussed later. However, the coupling to the intracell particle distribution is generally strong and must be included. As will be demonstrated later, omission of the coupling between intercell and intracell interactions leads to substantial errors in strongly interacting systems. With these considerations, a simplified parametric expression for the 1NN intercell interaction is given by

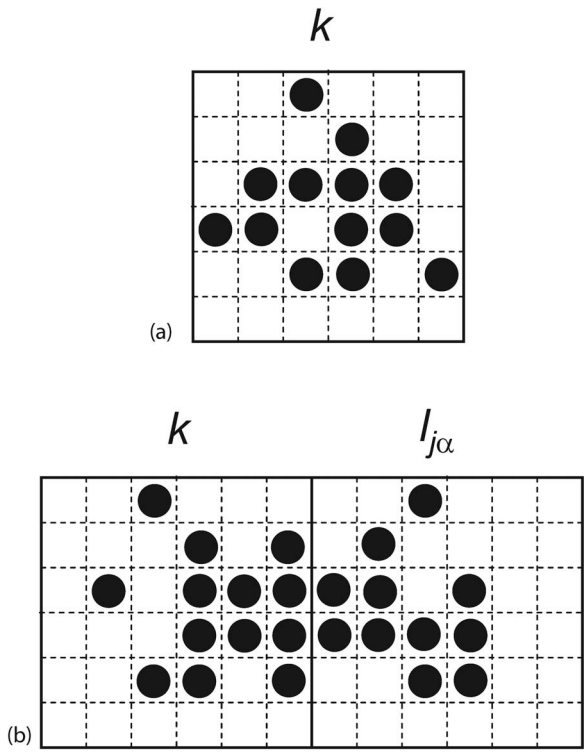

FIG. 3. Coupling between particles in the (a) intracell case and (b) intercell case. Coarse-graining level in both cases is $q=36$.

$$E_i^{\text{inter},l_1} = E_i^{\text{inter},l_1}(k,\{l_{1\alpha},\alpha=1,\ldots,n_1\}). \tag{17}$$

Similar decoupling approximations may be made for the intercell interactions at each neighbor shell, i.e.,

$$E_i^{\text{inter},l_j} = E_i^{\text{inter},l_j}(k,\{l_{j\alpha},\alpha=1,\ldots,n_j\}). \tag{18}$$

Finally, the calculation of the coarse binding energies in Eqs. (17) and (18) can be further simplified by assuming that within a given (coarse lattice) neighbor shell j, the total intercell interaction can be computed by assuming that the single cell contributions are independent and additive, i.e.,

$$E_i^{\text{inter},l_j} = \sum_{\alpha} E_i^{\text{inter},l_j}(k,l_{j\alpha}). \tag{19}$$

The coupling considered in Eq. (19) is schematically shown in Fig. 3(b), where only pairs of cells are considered.

C. A brief summary of previous LKMC coarse-graining approaches

In general, closure rules for computing the average in Eq. (7) can be categorized into two types: (1) Methods in which the average is analytically computed and (2) methods in which local fine-grid (or multigrid[13]) simulations are used to compute the average by numerical sampling.

The LMF method, discussed and analyzed in detail by Chatterjee and Vlachos,[13] is perhaps the simplest approach for analytically coarse graining a LKMC simulation, and will be used as the base case for discussing the present work. The basic assumption in the LMF model is that each microconfiguration within a coarse cell has equal probability and that no spatial correlations exist between the site locations and the binding energies. In other words, each particle has an equal probability of residing on any fine-grid site within a coarse cell and is unaffected by interactions with other particles (except for same-site exclusion).

For very short-range interactions relative to the coarse cell dimensions, the LMF is a poor approximation. Under these conditions, and ignoring cell boundary effects, the distribution of pair interactions can be analytically computed.[22] Chatterjee and Vlachos have recently applied the quasichemical approximation[13,22] (QCA) to the calculation of ABEs within coarse cells for a situation in which only first nearest-neighbor interactions are considered within each coarse cell and intercell interactions are neglected. The latter assumption implies that there will be no driving force for particles to further aggregate once each individual coarse cell is filled, severely limiting the application of the QCA framework in its present form to clustering problems.

Clearly, while computationally efficient, analytical approaches are limited to fairly simple systems. Numerical methods, on the other hand, apply fine-grid simulators to stochastically sample the local configurational space to compute the coarse-grained potential. The fine-grid sampling can be implemented online (i.e., on the fly during the simulation) or offline (once before the LKMC simulation). For example, in Ref. 13, a FGLKMC simulator was embedded in an online fashion within a CGLKMC simulation to compute the particle interaction distribution function in each coarse cell. This multigrid LKMC simulation significantly improves the accuracy of CGLKMC for systems with very short interaction ranges because it effectively applies ATR averaging over the intracell processes. On the other hand, the extension of this approach to long interaction ranges, which include substantial intercell interactions, is a challenge (which we focus on in the present work). Other numerical coarse-graining approaches have also been developed. For example, in Refs. 15–17 wavelet transforms were employed to ABEs in Monte Carlo simulations.

IV. A NEW STOCHASTIC CLOSURE APPROACH FOR LKMC COARSE GRAINING

A. Microscopic Wang–Landau Monte Carlo simulations

In this work, we stochastically (numerically) compute the ABE or ATR for a single particle in each coarse cell by using the recently developed Wang–Landau Monte Carlo[23] (WLMC) method. This is tantamount to computing a coarse-grained potential from the fine-grid one. The WLMC approach is applied because large energy differences between the various configurations (e.g., multiple isolated clusters in one coarse cell) often lead to severe sampling bottlenecks in a standard Metropolis Monte Carlo. Microscopic kinetic Monte Carlo simulations can also be applied to compute the coarse potential, but once again, large variations in the binding energy distribution can lead to inefficient sampling.

For decoupled intracell interactions [Eq. (15)], η particles are initialized in a single coarse cell with a coarse-graining level q, as shown in Fig. 3(a). Throughout the WLMC simulation, particles are confined to the same coarse cell. A system energy E_s is defined for a given microscopic configuration such that

$$E_s = E_0 - 0.5 \times \sum_{i=1}^{N_P} E_i^b \equiv E_0 - E_{\text{tot}}^b, \tag{20}$$

where E_0 is an arbitrary reference energy, and E_{tot}^b therefore is the total binding energy in the coarse cell at a given microscopic configuration. The configurational density-of-states (DOS) function for the system energy, $g(E_s)$, and the visit histogram, $h(E_s)$, are initialized to unity and zero, respectively. In the present work, both $g(E_s)$ and $h(E_s)$ are discretized by using 0.01 eV energy bins.

WLMC moves are performed by moving a randomly selected particle to a vacant location picked at random from all sites within the coarse cell. The WLMC acceptance/rejection criterion for accepting a move from system energy level E_s^1 to E_s^2 is given by

$$p(E_s^1 \rightarrow E_s^2) = \min\left[\frac{g(E_s^1)}{g(E_s^2)}, 1\right]. \tag{21}$$

Each time an energy level E_s is visited (rejections are treated as revisits of the current configuration), the DOS value, $g(E_s)$, is multiplied by a factor $f > 1$ so that $g(E_s) = g(E_s)f$. The multiplicative factor is initialized with a value of exp(1) in the present simulations. Concurrently, the visit histogram is updated by $h(E_s) = h(E_s) + 1$.

The simulation proceeds until a minimum flatness criterion is achieved in the function $h(E_s)$, which is taken here to be 85% (defined here by the maximum deviation from the mean). Once the flatness criterion is achieved, the value of f is reduced according to the rule $f_{i+1} = \sqrt{f_i}$ (any monotonically decreasing function will do in practice), where i represents the number of simulation "stages," and $h(E_s)$ is reset to zero for the next stage. The criterion used to end the simulations in the present work is $f = 1.000\,001$. Wang and Landau[23] have shown that the above procedure leads to a converged value for the DOS, from which the probability distribution function (PDF) at any temperature can be computed. The PDF for the lattice system is given by

$$p(E_s) = g(E_s)\exp\left(-\frac{E_s}{k_B T}\right). \tag{22}$$

Any ensemble-averaged property of the system can be computed from the system energy PDF by

$$\langle A \rangle = \int p(E_s) A(E_s) dE_s, \tag{23}$$

where $A(E_s)$ is the value of the property for a configuration with system energy level E_s, and $\langle A \rangle$ is the ensemble average of the property A. Implicit in Eq. (23) is the assumption that the property A for each particle is uniquely valued at a given

system energy level E_s. In general, this is not true, and an average value of A must be computed for each system energy level. In the current application, we are interested in the ABE (or ATR) associated with a single particle. This average is generated by averaging over all particles for each WLMC configuration, and again over many configurations. For example, in the ABE framework, the averaging over particles is taken first at each configuration (with system energy E_s),

$$\langle E_1^b \rangle|_{E_s} = \frac{1}{\eta_k} \sum_{i=1}^{\eta_k} E_i^b. \tag{24}$$

The average over configurations is generated on the fly during the WLMC simulation, while the DOS and PDFs are being converged. The final ensemble ABE for one particle within one coarse cell is therefore given by

$$E_i^b(\mathrm{ABE}) = \langle \langle E_1^b \rangle|_{E_s} \rangle = \int p(E_s) \langle E_1^b \rangle|_{E_s} dE_s. \tag{25}$$

Similarly, in the ATR framework, the particle-averaged property of interest for each configuration is the ATR, which is given by

$$\left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \Big|_{E_s} = \frac{1}{\eta_k} \sum_{i=1}^{\eta_k} \exp\left(-\frac{E_i^b}{k_B T}\right). \tag{26}$$

Accordingly, the ensemble ATR is given by

$$\left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle = \left\langle \left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \Big|_{E_s} \right\rangle = \int p(E_s) \times \left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \Big|_{E_s} dE_s. \tag{27}$$

Since the inputs to a CGLKMC simulation are always binding energies, not transition rates, an effective binding energy for one particle in the ATR framework is computed from the averaged transition rate by

$$E_1^b(\mathrm{ATR}) = -k_B T \ln\left(\left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \right). \tag{28}$$

Equations (26)–(28) show that the ATR framework is nothing more than an exponentially weighted ABE scheme. Once again, this makes physical sense—particles that are strongly bound should contribute less to the total effective transition rate from a given coarse cell.

The preceding approach is equally applicable to the computation of intercell interactions. For calculating the intercell binding energy, two interacting coarse cells are constructed and the movements of particles in each coarse cell are also confined to their respective coarse cells. The relative position of the two coarse cells depends on the interaction shell being considered; in Fig. 3(b), for example, the 1NN interaction on the (square) coarse lattice is shown. Once again, the number of interaction shells that need to be considered within the coarse lattice entirely depends on the fine-grid interaction potential function.

It should be noted that the sampling approach discussed above fully accounts for the effect of particle interactions—in other words, particle trajectories are generated in the WLMC simulations on the basis of a system energy that is a function of the total binding energy. By allowing the particle trajectories to be influenced by particle-particle interactions, the system is expected to exhibit clustering, which is a manifestation of nonideality in the context of solution theory.[24] This thermodynamic analogy will be explored in detail elsewhere—here, we simply use this concept to refer to the interacting particle model as the "nonideal solution" (NIS) model.

By contrast, the LMF closure is an "ideal solution" model in which the particle *trajectories* evolve as if the particles are free of interactions (other than same-site exclusion). In the LMF model therefore, the mean interaction energy in a cell is computed on the basis of interactionless trajectories. WLMC sampling can be applied for computing ABEs within the LMF model, although this is only useful for testing purposes given that the LMF binding energies can be computed more efficiently by using analytical methods.[10] In the LMF model, the system energy, as defined by Eq. (20) is always equal to the reference energy, E_0, because particles are assumed not to interact as they move within their respective coarse cells. Under these conditions, every WLMC move is accepted [see Eq. (21)] and there is no meaningful PDF; the ABE per particle is directly computed by averaging over particles and configurations. The averaging in this special case can be performed within both the ABE and ATR frameworks.

V. CALCULATION OF COARSE INTERACTION POTENTIALS FOR A 2D LATTICE

In this section, the WLMC sampling method discussed in Sec. IV is applied to compute the coarse-grained interactions that are used as inputs to CGLKMC simulations. The intracell and intercell interactions are separately treated. In the latter case, different approaches are tested for including interaction coupling.

A. Intracell interactions

As discussed in Sec. III, we assume in the following that the intracell interaction is independent of the particle distributions in neighboring coarse cells because particles within a single coarse cell are, on average, much closer to each other than to other particles in surrounding cells. For a given coarse-graining level q, the coarse cell is initialized with an occupancy number η, and WLMC simulation is performed until the equilibrium criterion is met. In the following example, 20 particles are put in a single coarse cell ($q=36$). A constant (fine-grid) interaction potential is applied between particles up to the 6NN shell ($L=6$) with $\beta J \equiv J/(k_B T)=1.5$, where J is the (constant in this particular case) interaction strength between two particles. The converged DOS function and the PDF obtained from the WLMC simulation are shown in Fig. 4.

The corresponding ABE (ABE framework) and effective binding energy (from the ATR framework) for one particle as a function of system energy are shown in Fig. 5. The ABE, $E_1^b(\mathrm{ABE})$, is inversely proportional to the system energy, which is expected from consideration of Eq. (20) and that

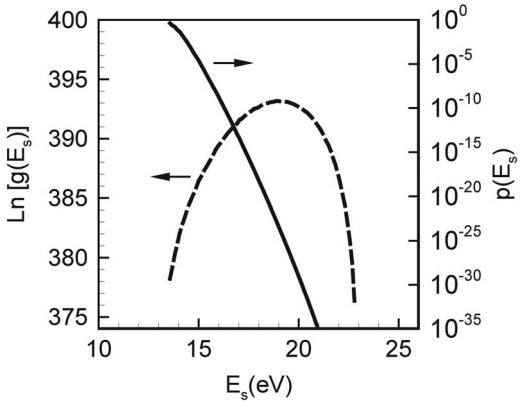

FIG. 4. DOS function $g(E_s)$—dashed line, and PDF $p(E_s)$—solid line, for the system energy in a single coarse cell obtained from WLMC simulation. System information: $q=36$, $\eta=20$, $\beta J=1.5$, and $L=6$NN.

$(\eta/2)E_1^b(\mathrm{ABE})=E_{\mathrm{tot}}^b$. On the other hand, the ATR binding energy, $E_1^b(\mathrm{ATR})$, is a strongly nonlinear function of the system energy and reaches a constant value as the system energy increases. Furthermore, $E_1^b(\mathrm{ATR})$ is much lower than $E_1^b(\mathrm{ABE})$ for all occupancy levels because the ATR is mainly dominated by the particles with low binding energy.

The intracell ABE experienced by one particle is then computed for each occupancy number, $1 \leqslant \eta \leqslant q$, by using a sequence of WLMC simulations. The results obtained by using both the ABE and ATR frameworks are shown in Fig. 6. Also, the results for the two different closure methods are shown: LMF model and NIS model.

Whether ATR or ABE averaging is employed, the ABE predicted by the NIS model is always higher than the corresponding LMF value. Physically, this is because interacting particles tend to cluster under the simulation conditions, which leads to a higher binding energy per particle for all occupancy ratios. The NIS model curves converge to the LMF model at $\eta=0$ and $\eta=q$, the latter being a result of geometric confinement. For the particular conditions shown in Fig. 6, the nonideality of the system is strong and significant clustering is present across most of the occupancy range.

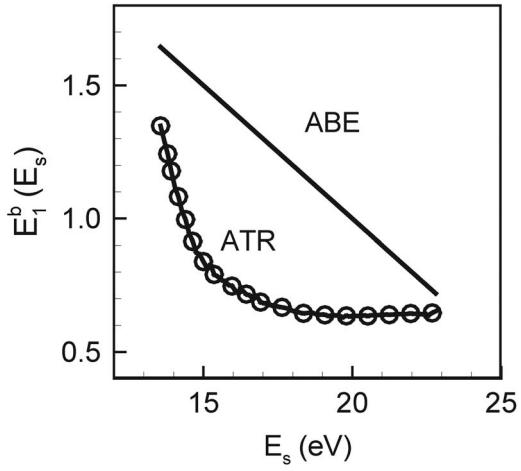

FIG. 5. Intracell ABE (per particle) as a function of system energy obtained with the ABE and ATR averaging frameworks in the NIS closure model. System information: $q=36$, $\eta=20$, $\beta J=1.5$, and $L=6$NN.

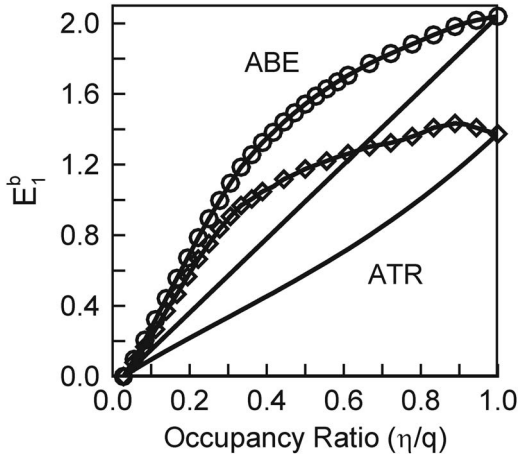

FIG. 6. Intracell ABE as a function of occupancy ratio. Solid lines—LMF model; symbols and lines—NIS model. Upper curves—$E_1^b(\mathrm{ABE})$—ABE obtained by using the ABE framework. (b) Lower curves—$E_1^b(\mathrm{ATR})$—ABE by using the ATR framework. System information: $q=36$, $\beta J=1.5$, and $L=6$NN.

The effective binding energies obtained in the ATR framework are lower than those obtained by using the ABE framework at each occupancy level. Once again, this arises because strongly bound particles in the ATR scheme contribute less to the overall average. Note also that while the ABE per particle in the ABE-LMF case is a linear function of occupancy level, it deviates from linear behavior in the ATR-LMF case because of the exponential-weighted averaging. In each framework, the nonideal behavior of the system is obvious and significant compared to LMF (ideal solution).

B. Intercell interactions

The calculation of coarse intercell binding energies by using the WLMC framework discussed in the previous sections is mechanistically similar to the intracell case. In this section, we present a sequence of different closure rules to test approaches for capturing the coupling between the intracell and intercell interactions. The generation of different closure rules is illustrated in Fig. 7. Consider two neighboring coarse cells, k and l, where cell k is the origination cell for a particle hop. In the LMF closure rule, particles in both cells are assumed to be uniformly distributed (as in the intracell case) and the coarse interaction potential can be ana-

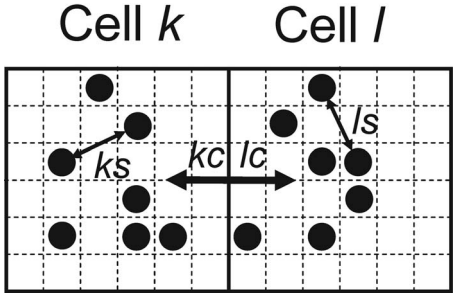

FIG. 7. Schematic of various types of intercell interactions. Cell k represents the origination cell for a particle hop, while cell l is a neighboring cell.

TABLE I. Combinatorial representation of closure rules for intercell interactions. Some rules are not valid because one-way interactions lead to ambiguously defined system energies (see text.)

	Cell l			
	lb	*lc*	*ls*	*lr*
kb	Each particleinteracts with everyother particle.	Particles in cell k interact with everyparticle. Particles incell l only interact with particles in cell k.	Not defined.	Particles in cell k interact with every particle. Particles incell l have random movements.
kc	Particles in cell k only interact with particles in cell l. Particles in cell l may interact with every particle.	Particles in cell k only interact with particles in cell l. Particles in cell l only interact with particles in cell k.	Not defined.	Particles in cell k only interact with particles in cell l. Particles in cell l have random movements.
ks	Not defined.	Not defined.	Particles in cell k only interact with particles in cell k. Particles in cell l only interact with particles in cell l.	Particles in cell k only interact with particles in cell k. Particles in cell l have random movements.
kr	Particles in cell k have random movements. Particles in cell l interact with every other particle.	Particles in cell k have random movements. Particles in cell l only interact with particles in cell k.	Particles in cell k have random movements. Particles in cell l only interact with particles in cell l.	Particles in cell k have random movements. Particles in cell l have random movements. (LMF)

lytically computed. This is equivalent to assuming that the particle trajectories are random, save for volumetric exclusion.

At the other extreme is the fully coupled closure rule where particles in both cells interact with each other as well as with particles in the other cell—we denote this closure rule by the notation *kb-lb*, where *b* stands for "both" self- and cross interactions, and k and l denote the cell indices. Similarly, the LMF closure rule is described by the notation *kr-lr*, where *r* denotes "random," or interactionless trajectories within a cell. Other possible types of interaction models include "self" (s), in which particles within a single cell only interact with each other, or "cross" (c), where particles only interact with those in different cells.

There are several possible combinations that can be applied to generate a closure rule for intercell interactions; these are summarized in Table I. For each valid closure method defined in Table I, the system energy at every fine-grid configuration is required to compute the DOS function and PDF. The system energy in each case is defined as

$$kb\text{-}lb: \quad E_s = E_0 - 0.5\sum_{i\in D_k}^{\eta_k}\sum_{\substack{j\in D_k\\ j\neq i}}^{\eta_k} J(i,j) - 0.5\sum_{i\in D_l}^{\eta_l}\sum_{\substack{j\in D_l\\ j\neq i}}^{\eta_l} J(i,j) - \sum_{i\in D_k}^{\eta_k}\sum_{j\in D_l}^{\eta_l} J(i,j), \tag{29}$$

$$kb\text{-}lc: \quad E_s = E_0 - 0.5\sum_{i\in D_k}^{\eta_k}\sum_{\substack{j\in D_k\\ j\neq i}}^{\eta_k} J(i,j) - \sum_{i\in D_k}^{\eta_k}\sum_{j\in D_l}^{\eta_l} J(i,j), \tag{30}$$

$$kb\text{-}lr: \quad E_s = E_0 - 0.5\sum_{i\in D_k}^{\eta_k}\sum_{\substack{j\in D_k\\ j\neq l}}^{\eta_k} J(i,j) - \sum_{i\in D_k}^{\eta_k}\sum_{j\in D_l}^{q}\frac{\eta_l}{q} J(i,j), \tag{31}$$

$$kc\text{-}lb: \quad E_s = E_0 - 0.5\sum_{i\in D_l}^{\eta_l}\sum_{\substack{j\in D_l\\ j\neq i}}^{\eta_l} J(i,j) - \sum_{i\in D_k}^{\eta_k}\sum_{j\in D_l}^{\eta_l} J(i,j), \tag{32}$$

$$kc\text{-}lc: \quad E_s = E_0 - \sum_{i\in D_k}^{\eta_k}\sum_{j\in D_l}^{\eta_l} J(i,j), \tag{33}$$

$$kc\text{-}lr: \quad E_s = E_0 - \sum_{i\in D_k}^{\eta_k}\sum_{j\in D_l}^{q}\frac{\eta_l}{q} J(i,j), \tag{34}$$

$$ks\text{-}ls: \quad E_s = E_0 - 0.5\sum_{i\in D_k}^{\eta_k}\sum_{\substack{j\in D_k\\ j\neq i}}^{\eta_k} J(i,j) - 0.5\sum_{i\in D_l}^{\eta_l}\sum_{\substack{j\in D_l\\ j\neq i}}^{\eta_l} J(i,j), \tag{35}$$

$$ks\text{-}lr: \quad E_s = E_0 - 0.5\sum_{i\in D_k}^{\eta_k}\sum_{\substack{j\in D_k\\ j\neq i}}^{\eta_k} J(i,j), \tag{36}$$

$$kr\text{-}lb: \quad E_s = E_0 - 0.5\sum_{i\in D_l}^{\eta_l}\sum_{\substack{j\in D_l\\ j\neq i}}^{\eta_l} J(i,j) - \sum_{i\in D_k}^{q}\sum_{j\in D_l}^{\eta_l}\frac{\eta_k}{q} J(i,j), \tag{37}$$

$$kr\text{-}lc: \quad E_s = E_0 - \sum_{i\in D_k}^{q} \sum_{j\in D_l}^{\eta_l} \frac{\eta_k}{q} J(i,j), \tag{38}$$

$$kr\text{-}ls: \quad E_s = E_0 - 0.5 \sum_{i\in D_l}^{\eta_l} \sum_{\substack{j\in D_l \\ j\neq i}}^{\eta_l} J(i,j), \tag{39}$$

$$kr\text{-}lr: \quad E_s = E_0. \tag{40}$$

In the above expressions, $J(i,j)$ represents the pair interaction potential between particles i and j on the fine-grid lattice. Note that there is a subtlety regarding closure rules with "one-way" bonding, which are denoted by "not defined" in Table I. Consider, for example, the closure rule *ks-lb* in which particles in cell k only interact with each other, but those in cell l interact with both each other and particles in cell k. It is not possible to define the total system energy in this case because the cross interactions between cells k and l are effectively one-way and binding energies for these interactions are ill posed.

Closure rules involving random (i.e., interactionless) particle trajectories are a special case. For example, the last rule in the list above, *kr-lr* [Eq. (40)], which is simply the LMF closure, assumes that particle trajectories are unaffected by the interparticle interactions and therefore that the mean interaction is computed on the basis of random (uniformly distributed) particle positions. To compute the total system energy for closure rules involving random movements of particles in one cell only (i.e., *kr* or *lr*), we assume that each site in the interactionless coarse cell is uniformly occupied by a fractional particle (η_k/q or η_l/q) and that the binding energy between that site with particles on other cells is scaled by that fraction; see, for example, the last term in Eq. (31). In this case, the particle positions in cell l are not evolved during the simulation but their interactions with particles in cell k are included in the calculation of the total system energy. Once again, it should be emphasized that closure rules based on interactionless particles (such as the LMF approximation) only assume that the particle trajectories are interactionless—the interaction energies themselves are still included in the system energy.

For each closure method in Table I, WLMC simulations were used to compute the intercell binding energy for one particle in cell k interacting with particles in cell l. In a WLMC simulation, particles are confined to move within a given coarse cell; and for each configuration, the ABE for one particle in cell k to particles in cell l is given by

$$\langle E_1^b \rangle|_{E_s} = \frac{1}{\eta_k} \sum_{i\in D_k}^{\eta_k} \sum_{j\in D_l}^{\eta_l} J(i,j). \tag{41}$$

Similarly, the ATR, for a given configuration, for one particle in cell k due to binding with particles in cell l is

$$\left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \Big|_{E_s} = \frac{1}{\eta_k} \sum_{i\in D_k}^{\eta_k} \exp\left(-\frac{E_i^b}{k_B T}\right), \tag{42}$$

where E_i^b is the binding energy associated with particle i in cell k due to interactions with particles in cell l and is given by $E_i^b = \Sigma_{l\in D_l}^{\eta_l} J(i,j)$.

For closure rules involving random movements of particles (i.e., interactionless particles), the sampling in the interactionless cell is taken over all sites in the coarse cell where each site is assumed to contain η/q particles (i.e., a uniform distribution). For example, if the particles in cell k are interactionless and therefore uniformly distributed, the ABE per particle and ATR, respectively, due to interactions with particles in cell l are given by

$$\langle E_1^b \rangle|_{E_s} = \frac{1}{q} \sum_{i\in D_k}^{q} \sum_{j\in D_l}^{\eta_l} \frac{\eta_k}{q} J(i,j) \tag{43}$$

and

$$\left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \Big|_{E_s} = \frac{1}{q} \sum_{i\in D_k}^{q} \exp\left(-\frac{E_i^b|_{E_s}}{k_B T}\right), \tag{44}$$

where

$$E_i^b|_{E_s} = \sum_{j\in D_i}^{\eta_l} \frac{\eta_k}{q} J(i,j). \tag{45}$$

Similar expressions hold for the case where the particles in cell l are assumed to be interactionless:

$$\langle E_1^b \rangle|_{E_s} = \frac{1}{\eta_k} \sum_{i\in D_k}^{\eta_k} \sum_{j\in D_l}^{q} \frac{\eta_l}{q} J(i,j) \tag{46}$$

and

$$\left\langle \exp\left(-\frac{E_1^b}{k_B T}\right) \right\rangle \Big|_{E_s} = \frac{1}{\eta_k} \sum_{i\in D_k}^{\eta_k} \exp\left(-\frac{E_i^b|_{E_s}}{k_B T}\right), \tag{47}$$

where

$$E_i^b|_{E_s} = \sum_{j\in D_l}^{q} \frac{\eta_l}{q} J(i,j). \tag{48}$$

Obviously, in every case, the intercell binding energy is a function of the occupancy level in both cells. For each pair of occupancy levels, WLMC is used to compute the DOS and PDFs by sampling over all fine-grid configurations. For example, Fig. 8 shows the DOS and PDF as a function of system energy for the *kb-lb* closure rule in which 20 particles are placed in each of two neighboring cells ($q=36$).

The WLMC simulations can be used to calculate the intercell binding energy surface as a function of occupancy numbers, which are required as input to CGLKMC simulations. Typical binding energy surfaces are shown in Fig. 9 for the LMF model by using both the ABE and ATR frameworks to compute the ABE surface. Note that the ABE averaging framework leads to substantially higher binding energies for all occupancy pairs (upper surface). As noted earlier, this is because high binding energy configurations in the ATR framework contribute less to the overall average, as would be

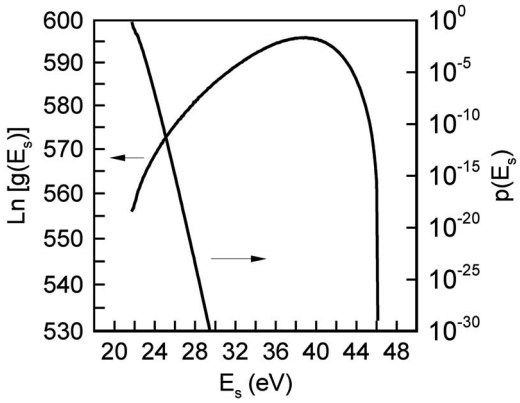

FIG. 8. DOS and PDFs for a system of two 1NN coarse cells in the *kb-lb* closure model. System information: $q=36$, $\beta J=1.5$, $L=6$NN, and $\eta_k=\eta_l=20$.

expected. In both frameworks, the maximum binding energy occurs at $\eta_l=q$ regardless of η_k. As in the intracell case, the curvature of the lower surface in Fig. 9 arises because the final binding energy is effectively computed within the exponentially weighted ATR framework.

Shown in Fig. 10 is the intercell binding surface for the *kb-lb* closure model in which full coupling between cell k and cell l is included. Both the binding energy surfaces predicted by the ABE and ATR frameworks for the *kb-lb* closure rule are plotted. Once again the ATR averaging leads to lower binding energies than those predicted by ABE, although the difference is much smaller in this case. Note also that for high η_l and low η_k, the intercell binding energy predicted by the *kb-lb* closure rule is much higher than that predicted by LMF, reflecting the increased proximity of the particles in the two cells because of particle coupling across the shared cell boundary (see Fig. 7). As we will demonstrate later, this coupling is important for accurately coarse-graining clustering dynamics at later times when the average cluster is comprised of many monomers.

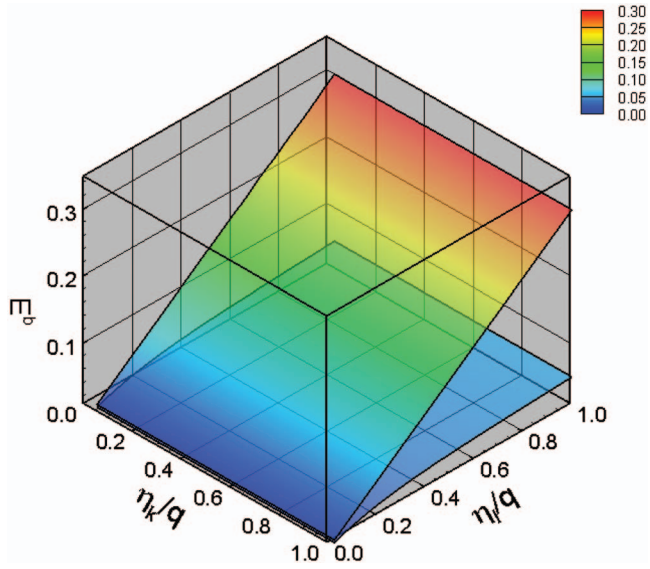

FIG. 9. (Color) Intercell binding energy (1NN coarse cells) for the LMF model (*kr-lr* closure rule) in the ABE framework (upper surface) and ATR framework (lower surface). System information: $q=36$, $\beta J=1.5$, and $L=6$NN.

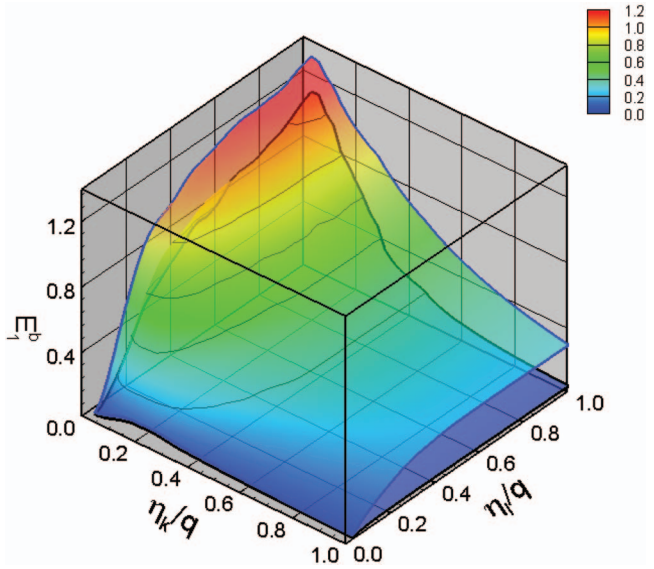

FIG. 10. (Color) Intercell binding energy (1NN coarse cells) for the *kb-lb* closure rule in the ABE framework (upper surface) and ATR framework (lower surface). System information: $q=36$, $\beta J=1.5$, and $L=6$NN.

Finally, the intercell binding energy surfaces for the same system obtained with the *kb-lr* closure rule in the ABE and ATR frameworks are shown in Fig. 11(a). The distribution of binding energies in both the ATR and ABE schemes is qualitatively similar to those in Fig. 10, but is somewhat lower across much of the surface. The *kb-lr* model also accounts for particle coupling between cells k and l, but does not presuppose any particular configuration for the particles in cell l, assuming instead that those particles are uniformly distributed. The motivation for such a closure rule comes from assuming that secondary coupling interactions arising from other cells surrounding cell l act equally strongly on the particles as those in cell k, leading to a roughly homogeneous distribution therein. A schematic of this effect is shown in Fig. 11(b). Note that the presence of such secondary coupling effects in general implies that the *kb-lb* intercell interactions are likely to be overestimated. The predictions of CGLKMC by using both the *kb-lb* and *kb-lr* closure models are compared in Sec. VI.

The intercell binding energies for other interaction shells in the coarse lattice can be computed in the same way as for the 1NN case discussed above. The intercell binding energies quickly decrease with each interaction shell, and for the 6NN interaction potential used in the above examples, only the 2NN coarse cell interactions need to be considered. In fact, the reduced interaction distance on the coarse lattice provides an important contribution to the overall computational savings achieved by coarse graining.

VI. CGLKMC SIMULATIONS OF DIFFUSION AND PARTICLE AGGREGATION

In this section, the ABE and ATR frameworks, along with different closure methods, are tested by using two systems. The first is a weakly interacting system in which we study bulk diffusion, while the second is a strongly interact-

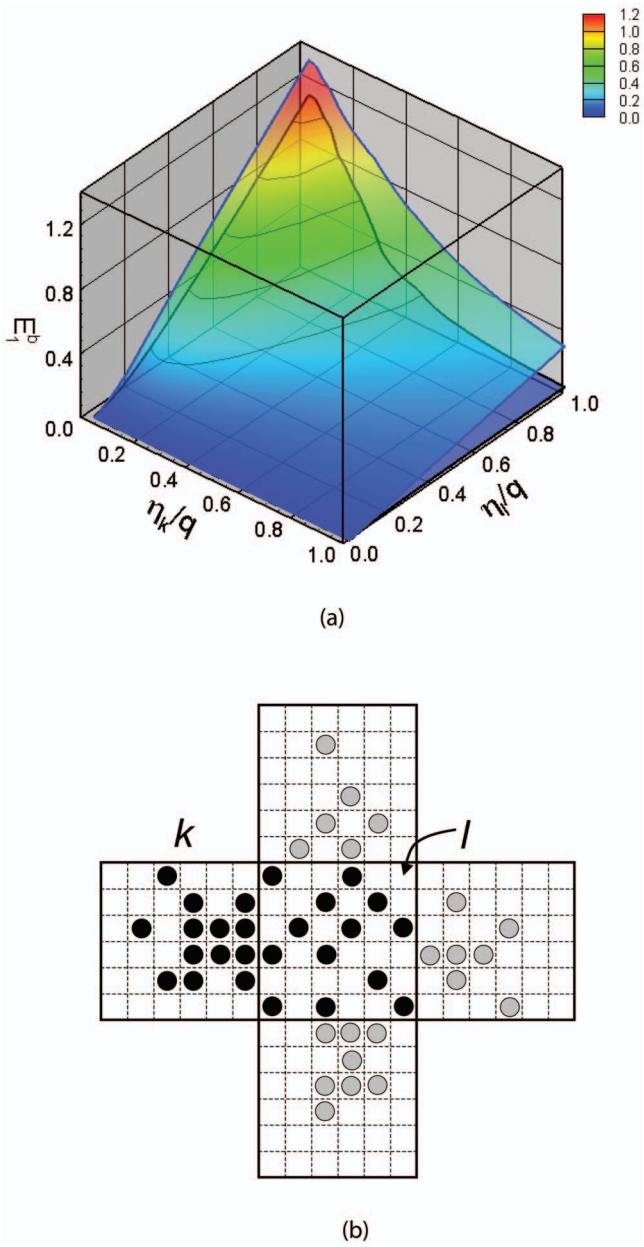

FIG. 11. (Color) (a) Intercell binding energy (1NN coarse cells) for the *kb-lr* closure rule in the ABE framework (upper surface) and ATR framework (lower surface). System information: $q=36$, $\beta J=1.5$, and $L=6$NN. (b) Schematic of a homogeneous neighbor cell (center cell with dark particles) due to implicit secondary interactions with other cells (peripheral cells with light particles).

ing one in which the focus is on cluster formation. In both cases, CGLKMC results are compared to FGLKMC predictions.

A. Weakly interacting system (diffusion)

In order to assess the accuracy of CGLKMC simulation with the closure rules and averaging frameworks discussed in the previous sections, we begin by considering a two-dimensional system subject to a one-dimensional concentration gradient, motivated by the example in Ref. 6. The concentration gradient is imposed along the x-direction while the y-direction is assumed to be infinitely long through the use of periodic boundary conditions. The boundary conditions in

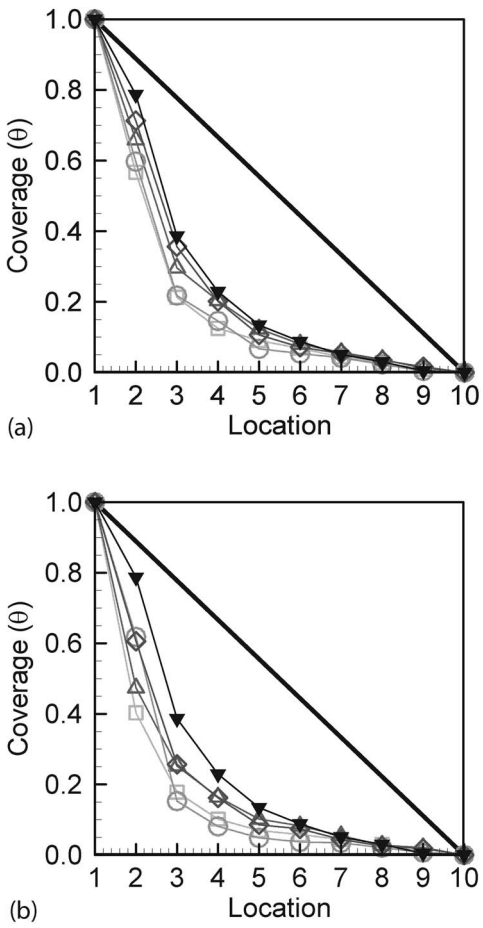

FIG. 12. Steady state local coverage profile for short, weak interactions at a coarse-graining level of (a) $q=9$ and (b) $q=36$. Black gradients—FGLKMC; squares—ABE-LMF; deltas—ATR-LMF; circles—ABE-*kb-lr*; diamonds—ATR-*kb-lr*.

the x-direction are given by $\theta(1)=1$ and $\theta(10)=0$; i.e., the domain lies between $1 \leqslant x \leqslant 10$. At steady state, the continuum equation for the concentration θ is given by

$$\frac{\partial}{\partial x}\left(D\frac{\partial \theta}{\partial x}\right)=0, \tag{49}$$

where D is the effective particle diffusivity, which is a function of the local concentration and the particle interaction potential.

The performance of CGLKMC is tested by using weak, short-ranged interactions at two coarse-graining levels (i.e., $q=9$ and $q=36$). A 1NN interaction potential with $\beta J=1$ is used in the FGLKMC simulations; the corresponding coarse potential is computed by using WLMC simulations, as discussed in the previous sections. The steady-state local coverage profile at a coarse-graining level of $q=9$ is shown in Fig. 12(a) along with the FGLKMC results. The steady-state condition is determined by monitoring the local coverage until a stationary profile is established. In each case, the profile is periodically monitored and the steady state identified as the point at which the profile is constant to within 3% over a given number of events (about 10^6). Good agreement between each CGLKMC prediction and the FGLKMC results is observed, but the profiles obtained by using the ATR

framework (with both LMF and *kb-lr* closures) are marginally closer to the FGLKMC results than those from the ABE framework. The performance of CGLKMC at $q=36$ for the same simulation conditions is shown in Fig. 12(b). Note that the plotted concentrations in both Figs. 12(a) and 12(b) are all mapped onto the $q=36$ coarse lattice so that a direct comparison can be made. The results for the $q=36$ case are similar to those obtained at $q=9$ although the higher level of coarse graining exhibits a larger error for all closure rules and averaging frameworks. No significant difference from the *kb-lr* results is observed when the *kb-lb* closure rule is applied to either case because the interaction potential is weak.

For short, weak interactions, therefore, the two averaging frameworks (ABE and ATR) are similar although the ATR framework appears to provide a systematically better approximation. Also, the new closure rules appear to give a slightly better result than the LMF model, but the differences are small. More conclusive results are demonstrated in the following section for the case of strongly interacting particles.

B. Strongly interacting system (aggregation)

For strongly interacting particles, the one-dimensional diffusion model described in Sec. VI A is no longer appropriate because the particles quickly form large clusters and the diffusivities of the clusters become extremely low, effectively halting the concentration profile evolution. Hence, we use a different system geometry in which the focus is on the aggregation dynamics of the particles, and the evolution of the cluster size distributions and morphologies is studied and compared to FGLKMC simulations.

We begin by investigating the effect of the closure rule and the averaging framework on the particle aggregation dynamics by using a coarse-grained system with $q=36$. In the following simulations, 900 particles are initially uniformly distributed on a two-dimensional square lattice containing 129 600 sites. A 6NN fine-grid interaction potential with constant interaction strength, $\beta J=1.5$, is applied for all cases unless otherwise noted. A sequence of WLMC simulations is used to compute the coarse interaction potential for each closure rule at the $q=36$ coarse-graining level, which is then used as input into the CGLKMC aggregation simulations. During the aggregation simulations, the cluster size distribution and morphology are periodically monitored and compared to the results of the FGLKMC runs. A cluster is defined as a group of particles that are connected at the 2NN neighbor distance on the coarse-grid lattice. This definition is self-consistent with the range of the coarse potential applied in this work. Connectivity between particles on neighboring coarse cells is assumed to hold at all occupancy levels because no configurational information within a coarse cell is available. Note that all lattice configurations from the FGLKMC simulation are mapped onto the coarse lattice system to enable a quantitatively consistent comparison between the two simulations.

It is immediately obvious that none of the closure methods coupled with the ABE averaging framework give a good

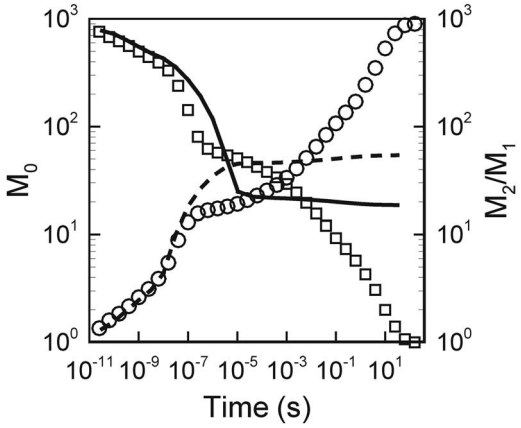

FIG. 13. Cluster size distribution obtained from FGLKMC and CGLKMC ($q=36$). The ABE-LMF framework is poor for strong interactions. Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=36$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

representation of the FGLKMC clustering dynamics. For example, two moments of the cluster size distribution are shown in Fig. 13 for the ABE-LMF combination; in particular, we use the average cluster size, M_2/M_1, and the total cluster number, M_0, to represent the transient evolution of the cluster size distribution. All moments of the size distribution are defined as $M_n=\Sigma_s s^n X_s$, where X_s is the number of clusters of size s, and n is the moment order. As shown in Fig. 13, the aggregation dynamics predicted by CGLKMC are initially in agreement with the FGLKMC results but then rapidly level off. The apparent kinetic arrest is likely due to the relatively low intercell binding energy predicted by the ABE-LMF model, which is the principal driving force for cluster aggregation after coarse cells begin to fill. Snapshots of the corresponding particle distributions after about 100 s for both simulations are shown in Fig. 14. While the FGLKMC simulation has evolved into a single large cluster, the CGLKMC system exhibits a distribution of isolated clusters, each of size $\sim q$. Note that the occupancy level in each coarse cell (up to 36 particles) is denoted by both grayscale shading and sphere size—high occupancy is denoted by large dark spheres, while low occupancy is denoted by small light spheres.

Cluster dynamics predicted by the ABE-*kb-lr* model are shown in Fig. 15. Some improvement is observed compared to the ABE-LMF case, but the agreement between the CGLKMC and FGLKMC results is still poor at later times. Further tests (not shown) demonstrate that none of the closure rules in Table I coupled with the ABE averaging approximation lead to a satisfactory representation of the fine-grid results simply because the ABE framework is not valid for a strongly interacting system.

The effect of using the ATR averaging framework was investigated next by using the same closure rules discussed above. As shown in Fig. 16, the ATR-LMF model is still poor because the LMF does not capture any of the coupling present in the intracell and intercell particle distributions. By contrast, *excellent* results are generated with the ATR-*kb-lr* (and the closely related ATR-*kc-lr*) models as shown in

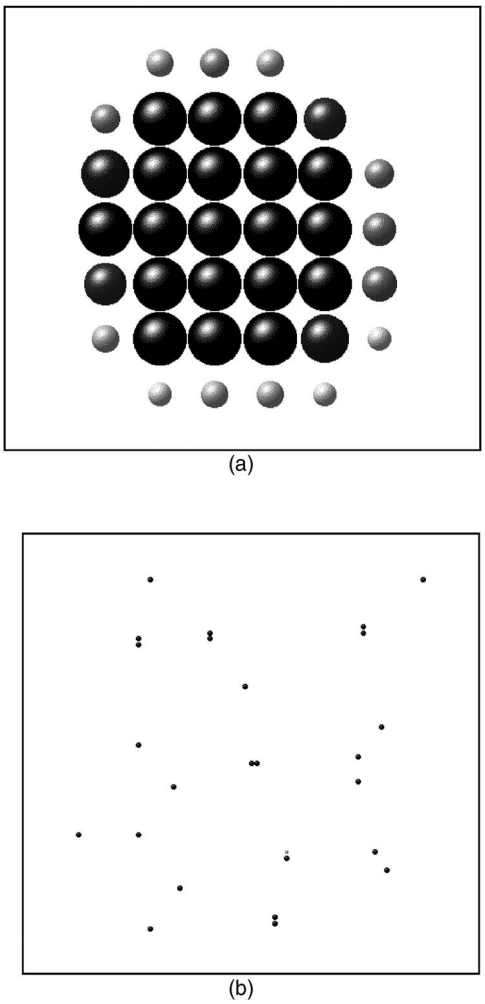

FIG. 14. Cluster morphology predicted by (a) FGLKMC and (b) ABE-LMF CGLKMC. Both particle distributions are presented on the $q=36$ coarse lattice. A single cluster with a dense core is formed in the FGLKMC simulation at $t\sim 100$ s. Isolated clusters with sizes that are integer multiples of q observed in CGLKMC. Note that zoom levels in (a) and (b) are different due to the different microconfigurations.

Figs. 17 and 18. The predictions of these two models are essentially identical and we will focus on the ATR-*kb-lr* results. As previously discussed, the *kb-lr* closure rule attempts to capture the coupling between cells k and l, while assuming that the cell l population is uniformly (or randomly) distributed because of (equal) interactions with other surrounding cells. As shown in Fig. 18, this model not only provides accurate dynamic evolution of the distribution moments but also captures the details of the cluster morphology throughout the growth process.

The *kc-lr* closure rule (i.e., no particle-particle interactions on the origination cell k) leads to very similar results to the *kb-lr* case because the particle self-interactions within the origination cell are effectively introduced by coupling to cell l. In other words, particles in the origination cell are going to be drawn to the cell boundary closest to cell l regardless of whether self-interactions between them are present.

Based on these results, it is clear that both the ATR averaging framework and a realistic closure model are essential for generating accurate coarse-grained models for strongly interacting particles. In the latter regard, it is instructive to consider one additional case, namely, the ATR-*kb-lb* model. Recall that the *kb-lb* model represents "full" coupling between the origination cell k and each surrounding cell l—i.e., each particle in cells k and l is free to interact with all other particles in those two cells, under the implicit assumption that all other surrounding cells are empty. The predicted cluster size distribution evolution for *kb-lb* model is shown in Fig. 19. Agreement with the FGLKMC data is excellent to $t\sim 0.01$ s, but poor thereafter, where the CGLKMC dynamics suddenly slow down, only to accelerate about two time decades later. As discussed above, this behavior is likely due to overprediction of the coupling between cells k and l. None of the other closure rules in Table I lead to satisfactory results.

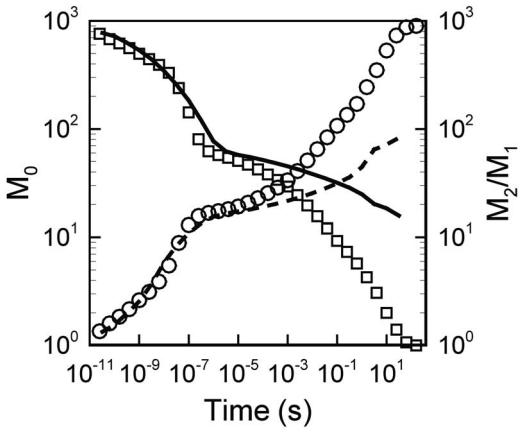

FIG. 15. Cluster size distribution obtained from FGLKMC and ABE-*kb-lr* framework CGLKMC ($q=36$). Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=36$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

Although the effect of coupling to other cells is not explicitly taken into account in any of the closure rules presented in our study, the successful *kb-lr* model attempts to capture secondary coupling by forcing the (interacting) particles on neighboring cells to assume random positions. Note again that secondary coupling is not expected to have a sig-

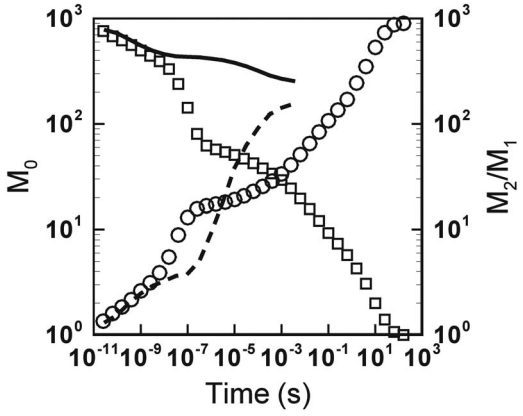

FIG. 16. Cluster size distribution obtained from FGLKMC and ATR-LMF framework CGLKMC ($q=36$). Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=36$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

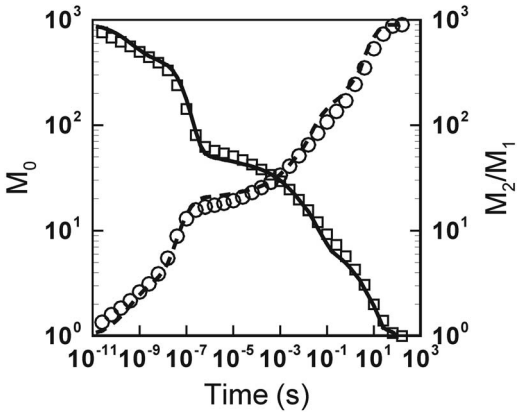

FIG. 17. Cluster size distribution obtained from FGLKMC and ATR-*kb-lr* framework CGLKMC ($q=36$). Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=36$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

nificant impact on the intracell interactions because these are generally much stronger than intercell interactions due to particle proximity considerations. In summary, the ATR-*kb-lr* (and the closely related ATR-*kc-lr*) model accurately captures all aspects of the aggregation processes in the two-dimensional system considered here. Both the averaging framework and the closure rule are important in obtaining a good representation when the interparticle interaction strength is high.

C. Further testing of the ATR-*kb-lr* model

Two additional tests of the ATR-*kb-lr* model are presented in this section to further demonstrate the robustness of this coarse-graining framework. In particular, different coarse-graining levels and fine-grid interaction potentials are tested as described in the following sections.

1. Effect of coarse-graining level

The ATR-*kb-lr* framework was applied to CGLKMC simulations with $q=64$. All other parameters were maintained at the values specified in Sec. VI B. Note that additional WLMC simulations are needed to generate the coarse interaction potentials for each of the new coarse-graining levels. The CGLKMC simulation prediction of the cluster size distribution evolution for the $q=64$ case is shown in Fig. 20. Again, the clustering dynamics are captured across the entire simulation timescale for both coarse-graining levels, as are the final cluster structures (not shown). Note that the overall quality of the CGLKMC results does not appear to appreciably deteriorate with increasing coarse-graining level.

2. Effect of interaction potential shape

In this test, a decaying, rather than a constant, interaction potential is applied. The particle interaction distance still extends to the 6NN shell but the interaction exponentially decays to about 40% of its 1NN strength as it approaches the 6NN shell. Again, all other parameters are kept unchanged and the corresponding coarse interaction is computed by using the WLMC method. The performance of the CGLKMC simulation with this interaction potential is shown in Fig. 21. Once again, excellent agreement between FGLKMC and CGLKMC ($q=9$) is observed, although a little deviation is apparent at $t\sim10^{-6}$ s, where the cluster dynamics transition from growth to coalescence. The overall aggregation dynamics predicted by this interparticle potential are about 100 times faster than those predicted by the constant potential applied in the previous cases.

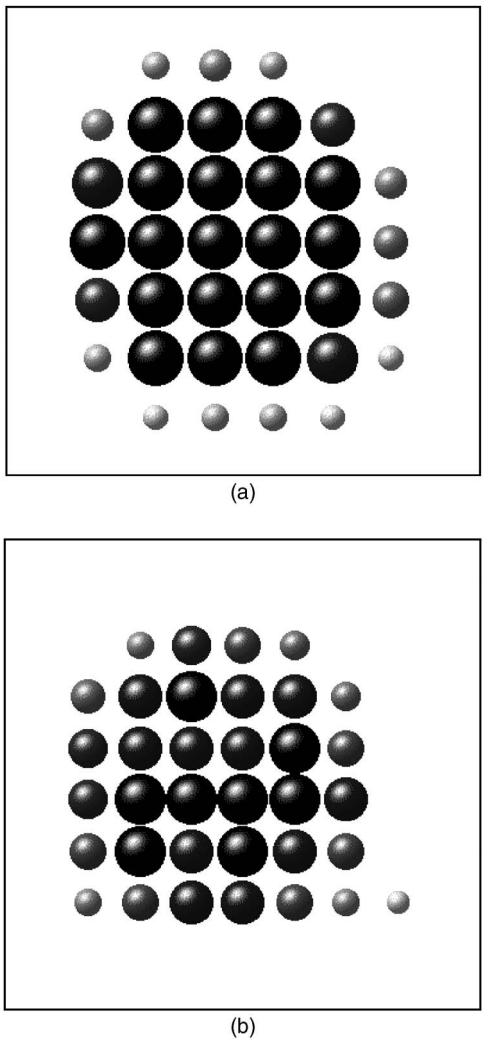

FIG. 18. Cluster morphology predicted by (a) FGLKMC and (b) ATR-*kb-lr* CGLKMC at $t\sim100$ s. Both particle distributions are presented on the $q=36$ coarse lattice. A single cluster with a dense core is formed in both simulations.

VII. COMPUTATIONAL EFFICIENCY OF CGLKMC SIMULATIONS

The single most important driving force for developing CGLKMC models is to extend the length and timescale accessible with LKMC by reducing the computational cost associated with particle hopping events. The savings in computational time in a CGLKMC simulation arise from three main coarse-graining aspects. The first and probably dominant one is the reduction in the configurational space.[17] The second type of savings is because the particle jump length in a coarse-grained system can be much larger than in the fine-

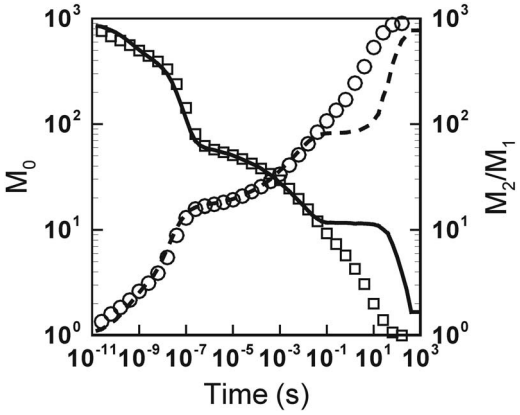

FIG. 19. Cluster size distribution obtained from FGLKMC and ATR-*kb-lb* framework CGLKMC ($q=36$). Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=36$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

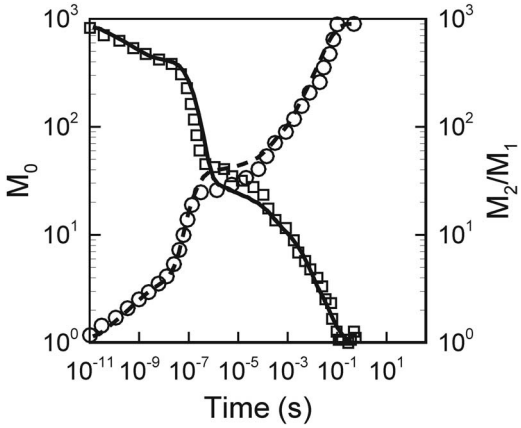

FIG. 21. Cluster size distribution obtained from FGLKMC and ATR-*kb-lr* framework CGLKMC ($q=9$) with a decaying interaction potential. Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=9$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

grid one; specifically the coarse jump length in each direction is proportional to the number of fine-grid sites in one coarse cell along that direction. Finally, the CGLKMC algorithm is also simplified compared to the fine-grid one in that the range of the potential is reduced by coarse graining. For example, in the previous sections, the 6NN fine-grid interaction potential was effectively reduced to a 2NN interaction potential on the coarse lattice, greatly reducing the cost associated with updating the rates after each particle hop.

To quantitatively assess the savings in CPU time due to coarse graining, we used a large system in which 4624 particles were distributed across 665 856 lattice sites in two dimensions. A comparison between the performance of the FGLKMC and two different levels of CGLKMC ($q=9$ and $q=36$) is shown in Fig. 22, which shows the evolution of CPU time versus simulation time throughout the simulation. It is interesting to note that the computational advantage is not uniform throughout the simulation. Specifically, the maximum advantage is realized at intermediate simulation times where the $q=36$ CGLKMC simulation is about 100 times faster than the FGLKMC and 10 times faster than the $q=9$ CGLKMC simulation. At this stage, the reduction in configurational space is most advantageous. The advantage shrinks at later times because the difference in configurational space between the simulations is not as large, i.e., there are only a few large clusters in the system and the main advantage of CGLKMC is simply the larger jump distance per event, as well as the reduced CPU cost associated with the rate database update.

There is a significant overhead cost associated with the off-line computation of the coarse interaction potentials, most of which is created in the computation of the coarse intercell potentials. One WLMC simulation is required for a given occupancy combination in the origination and destination cells. Of course, not every single possibility needs to be computed because interpolation can be applied to fill in the interaction surfaces, particularly if they are smooth functions of the occupancy level in a coarse cell. At $q=64$, which is the largest degree of coarse graining considered in this work, we computed the ABE at every third value of occupancy and interpolated for the others. This results in approximately

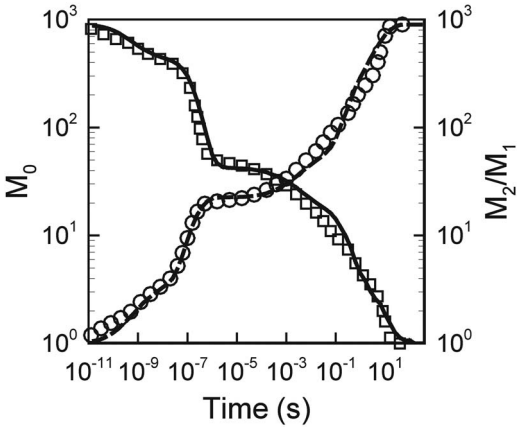

FIG. 20. Cluster size distribution obtained from FGLKMC and ATR-*kb-lr* framework CGLKMC ($q=64$). Symbols—FGLKMC: Circles—average cluster size (M_2/M_1); squares—total number of clusters (M_0). Lines—CGLKMC ($q=64$): Dashed line—average cluster size (M_2/M_1); solid line—total number of clusters (M_0).

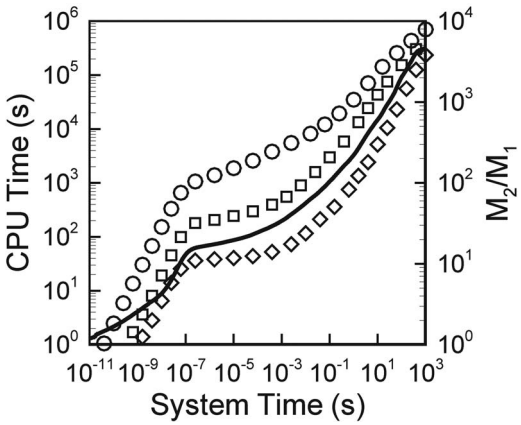

FIG. 22. Timing profile of CGLKMC simulations. Circles—FGLKMC; squares—CGLKMC ($q=9$); diamonds—CGLKMC ($q=36$); solid line—average cluster size (M_2/M_1).

about 400 WLMC being required to completely define the 1NN coarse potential. Another 400 simulations are required for the coarse potential at the 2NN interaction range. A total of about 4–6 h of CPU time (using the same processors that were employed in the CGLKMC analysis in Fig. 22) is required to compute the coarse interaction potential at $q=64$. This is significant when compared to the CPU savings implied in Fig. 22, particularly for smaller systems. However, it should be noted that the WLMC simulations are ideally suited for execution in a parallel setting and therefore do not consume much actual time relative to the CGLKMC (which is much more difficult to parallelize).

VIII. CONCLUSIONS

A novel CGLKMC simulation framework was developed and shown to provide excellent representation of particle aggregation in two-dimensional systems of strongly interacting particles. The accurate modeling of aggregation in strongly interacting systems has posed challenges for previously published LKMC coarse-graining approaches. Two particular elements were shown to be critical for successful coarse graining in strongly interacting systems. The first is the use of exponentially weighted ABEs rather than simple averages. Physically, the former is an average over the transition rates, which effectively biases the binding energies in a given coarse cell toward less strongly bound configurations that are more likely to generate particle hops. In the limit of weak interactions, both averaging schemes were shown to converge to the same estimate for the ABE within a coarse cell.

The second important component of our coarse-graining framework is the consideration of intra- and intercell coupling by an appropriate choice of closure rule, which has been largely ignored in the literature to date. Several closure rules were presented and tested that extend beyond the LMF closure, which has been widely applied in the literature. In the LMF closure approach, all particles are assumed to be interactionless and all configurations are equally likely. In the present work, both intracell and intercell particle interactions are considered. Intracell interactions are shown to lead to "nonideal" behavior in which the particles form clusters, and therefore become heterogeneously distributed within the coarse cell. Intercell interactions lead to correlations between particles in adjacent cells and additional heterogeneity. Moreover, coupling to cells beyond the intercell system can lead to further bias that must be considered. It was demonstrated that ignoring these "secondary" coupling effects leads to overprediction of the intercell coupling. The optimal closure rule assumes that the neighbor cells are homogeneous (LMF), but allows for full coupling within the origin cell. This hybrid closure rule accurately predicts the cluster evolution dynamics for a variety of interaction potentials and for all coarse-graining levels considered in our study.

The coarse-graining approach and closure rules presented here extend previous work to the general case of arbitrary potential interaction range and strength. The multigrid approach in Ref. 12 was the first to effectively include rate averaging (rather than binding energy averaging) and the effect of particle interactions on this average, but did not consider intercell interactions and is therefore limited to very short interactions. By contrast, the commonly applied LMF closure rule addresses intercell interactions but does not consider the effect of particle interactions on the particle distributions (i.e., nonideality).

Some limitations of the present approach should be noted. First, the computational cost for the off-line WLMC simulations used to compute the coarse-grained interaction potential is significant, especially for high levels of coarse graining and small LKMC systems. Second, the present approach is not highly transferable. In particular, it is not clear how to apply the present coarse-graining framework to nonisothermal processes. As the system temperature changes, the particle distribution function in each coarse cell, and therefore the coarse interaction potential, changes. It is not clear whether the computed coarse interactions at one temperature can be scaled in some way to apply at another temperature. This issue will be addressed in future work. Extension of the current methodology to three dimensions and to other lattice structures should be relatively straightforward. These extensions will also be addressed in future work. Finally, we note that extension of the ideas presented here to off-lattice, or continuous space systems is not immediately obvious. We have, however, demonstrated in previous work[1–3] that certain off-lattice systems can be effectively mapped onto a lattice representation by using model-to-model regression. Once this is achieved, it should be readily possible to apply the ideas presented in this work to provide further coarse graining.

ACKNOWLEDGMENTS

Financial support from the CBET Division of the National Science Foundation (CBET-0730971) is gratefully acknowledged.

[1] J. Dai, J. M. Kanter, S. S. Kapur, W. D. Seider, and T. Sinno, Phys. Rev. B **72**, 134102 (2005).
[2] J. Dai, W. D. Seider, and T. Sinno, Mol. Simul. **32**, 305 (2006).
[3] J. Dai, W. D. Seider, and T. Sinno, Mol. Simul. **33**, 733 (2007).
[4] J. M. Roussel and P. Bellon, Phys. Rev. B **73**, 085403 (2006).
[5] N. R. Zangenberg, J. L. Hansen, J. Fage-Pederson, and A. N. Larsen, Phys. Rev. Lett. **89**, 125901 (2001).
[6] J. B. Adams, Z. Y. Wang, and Y. H. Li, Thin Solid Films **365**, 201 (2000).
[7] R. Pinacho, P. Castrillo, M. Jaraiz, I. Martin-Bragado, J. Barbolla, H.-J. Gossmann, G. H. Gilmer, and J. L. Benton, J. Appl. Phys. **92**, 1582 (2002).
[8] O. Biham, I. Furman, M. Karimi, G. Vidali, R. Kennett, and H. Zeng, Surf. Sci. **400**, 29 (1998).
[9] L. G. Wang and P. Clancy, Surf. Sci. **473**, 25 (2001).
[10] M. A. Katsoulakis and D. G. Vlachos, J. Chem. Phys. **119**, 9412 (2003).
[11] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, J. Chem. Phys. **121**, 11420 (2004).
[12] M. A. Katsoulakis, A. J. Majda, and D. G. Vlachos, J. Comput. Phys. **186**, 250 (2003).

[13] A. Chatterjee and D. G. Vlachos, J. Chem. Phys. **124**, 064110 (2006).
[14] A. Chatterjee and D. G. Vlachos, J. Comput.-Aided Mater. Des. **14**, 253 (2007).
[15] A. E. Ismail, G. C. Rutledge, and G. Stephanopoulos, J. Chem. Phys. **118**, 4414 (2002).
[16] A. E. Ismail, G. Stephanopoulos, and G. C. Rutledge, J. Chem. Phys. **118**, 4424 (2002).
[17] A. E. Ismail, G. C. Rutledge, and G. Stephanopoulos, Comput. Chem. Eng. **29**, 689 (2005).
[18] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).
[19] A. La Magna, S. Coffa, and L. Colombo, Nucl. Instrum. Methods Phys. Res. B **148**, 262 (1999).
[20] P. Perrot, *A to Z of Thermodynamics* (Oxford University Press, New York, 1998).
[21] W. M. Deen, *Analysis of Transport Phenomena* (Oxford University Press, New York, 1998).
[22] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover, New York, 1986).
[23] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[24] J. W. Tester and M. Modell, *Thermodynamics and Its Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1996).