Commun. Comput. Phys.
doi: 10.4208/cicp.101210.241210a

Vol. **10**, No. 5, pp. 1089-1112
November 2011

Simulation of Three-Dimensional Strained Heteroepitaxial Growth Using Kinetic Monte Carlo

Tim P. Schulze[1,*] and Peter Smereka[2]

[1] *Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA.*
[2] *Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA.*

Received 10 December 2010; Accepted (in revised version) 24 December 2010

Communicated by Weinan E

Available online 2 August 2011

Abstract. Efficient algorithms for the simulation of strained heteroepitaxial growth with intermixing in 2+1 dimensions are presented. The first of these algorithms is an extension of the energy localization method [T. P. Schulze and P. Smereka, An energy localization principle and its application to fast kinetic Monte Carlo simulation of heteroepitaxial growth, J. Mech. Phys. Sol., 3 (2009), 521–538] from 1+1 to 2+1 dimensions. Two approximations of this basic algorithm are then introduced, one of which treats adatoms in a more efficient manner, while the other makes use of an approximation of the change in elastic energy in terms of local elastic energy density. In both cases, it is demonstrated that a reasonable level of fidelity is achieved. Results are presented showing how the film morphology is affected by misfit and deposition rate. In addition, simulations of stacked quantum dots are also presented.

AMS subject classifications: 82B80, 82B21, 82D25, 65C05

Key words: Heteroepitaxy, strained thin film, kinetic Monte Carlo.

1 Introduction

The computational cost of simulating heteroepitaxial growth with misfit strain using kinetic Monte Carlo (KMC) is orders of magnitude greater than that for strain-free growth due to the need to update the long-range elastic deformation of the film as the simulation proceeds. Until recently, this has prevented the widespread use of KMC for such simulations, especially in 2+1 dimensions. In this paper, we extend, from 1+1 to 2+1 dimensions, methods introduced in earlier work [1,22], refine somewhat a key result upon which those methods are based, and introduce new approximations to further enhance computational performance.

[*] Corresponding author. *Email addresses:* schulze@math.utk.edu (T. P. Schulze), psmereka@umich.edu (P. Smereka)

The rationale for KMC simulations aimed at understanding the growth and relaxation of crystals is based on molecular dynamics (MD) simulations and transition state theory. The essence of this model is that the system spends most of its time randomly oscillating within the N-particle configuration space about a local minimizer $x_m \in \mathbb{R}^{3N}$ of the system potential energy, $U(x)$, with rare transitions between these basins of attraction. The harmonic approximation to transition state theory estimates the rate $r_{a\to b}$ at which the transition occurs as

$$r_{a\to b}=K\exp\Big(-\frac{\Delta U}{k_BT}\Big), \tag{1.1}$$

where ΔU is the minimum energy barrier that must be overcome in moving from the initial, locally minimizing configuration, x_a, to a neighboring one, x_b, in configuration space, K is a weakly temperature-dependent oscillation "frequency" and k_BT is an energy scale defined by the temperature of the film.

These observations suggest an alternative model where the Newtonian dynamics is replaced by a continuous time Markov-chain, with the system making relatively rare, random transitions between states that represent local minimizers, x_a, in the system's configuration space at rates $r_{a\to b}$ calculated from (1.1). More specifically, the energy barrier

$$\Delta U=U(x_s)-U(x_a), \tag{1.2}$$

requires locating both the initial local minimum, x_a, and the saddle point, x_s (where $\nabla U=0$ and all but one of the principal curvatures are positive), separating the basins of attraction. Note that these local minima and saddle points are, in principle, determined by the motion of all of the particles simultaneously within the configuration space. When this sort of scheme is carried out in detail, it is referred to as off-lattice KMC or on-the-fly KMC [2,6,7]. While this is much faster than the corresponding MD simulation, or even accelerated MD simulations based on similar considerations [21,24], it is still prohibitively expensive in that one could not hope to simulate the growth of a crystal on physically relevant space and time scales.

For single-crystal, homoepitaxial systems, an often-used and greatly simplified model immediately suggests itself. In the simplified approach, the states are approximated using occupation arrays structured in the form of a perfect lattice-most often simple cubic, but face centered cubic and other lattices are also used; the allowed transitions are restricted to a limited catalog of characteristic events (e.g., single particle moves to neighboring sites); and the transition rates are parameterized based on the local lattice configuration. Indeed, it is this type of model that people generally refer to when they speak of KMC.

A well known example is nearest-neighbor, bond-counting KMC. In this model, atoms are restricted to positions on a simple cubic lattice, and the surface of the film, $h_{ij} \in \mathbb{Z}$, is often assumed single valued (the solid-on-solid assumption). Only surface atoms can move, and they move by hopping to a randomly chosen neighboring site in one of the four orthogonal directions. The hopping rate for the surface atom at site (i,j) is taken to be

$$r_{ij} = K\exp\Big(-\frac{\gamma n_{ij}}{k_B T}\Big), \tag{1.3}$$

where n_{ij} is the number of bonds this atom has with its nearest neighbors and γ is the bond energy. While this model is idealized, it captures the essential physical effects of homoepitaxial growth, such as surface diffusion and nucleation. Furthermore, the model satisfies detailed balance, which implies that, in the absence of deposition, the model will approach an equilibrium solution (in a statistical sense). Notice that the rates for this model are independent of the particle's destination; there exist many variations on this model that account for such non-nearest neighbor effects. Finally, it is important to realize that bond-counting KMC is orders of magnitude faster than off-lattice KMC, a gap in performance we seek to bridge by introducing intermediate approximations.

Bond counting models are inappropriate when the basins of attraction are modified by long-range elastic deformation. In particular, the misfit strain in heteroepitaxy falls into this latter category. Our approach will be similar to that proposed by Orr et al. [16], in which they modify (1.3) to

$$r_{ij} = K\exp\Big[\frac{(-\gamma n_{ij} + \Delta W)}{k_B T}\Big], \tag{1.4}$$

where W is the total elastic energy of the system in mechanical equilibrium and

$$\Delta W = W(\text{with surface atom } (i,j)) - W(\text{without surface atom } (i,j)). \tag{1.5}$$

In this way the contribution of the long-range, elastic interactions is included in the hopping rates. It is not hard to verify that this model also satisfies detailed balance. This formulation has been the basis of KMC models used for simulating heteroepitaxial growth in a number of studies [1,9,12,19,22].

In this paper, we extend to three dimensions a modification of (1.4) proposed by Baskaran et al. [1] to account for intermixing of the film and substrate material. We shall consider two species of atoms denoted type 1 and type 2. For most of our simulations we will consider the situation in which atoms of type 2 are deposited on a substrate of type 1. We will let $\gamma_{\alpha\beta}$ denote the bond strength between atoms of type α and type β. The hopping rate of a surface atom at site (i,j) is given by

$$r_{ij} = K\exp\Big[\frac{(-E_D - B + \Delta W)}{k_B T}\Big], \tag{1.6}$$

where

$$B = B_{11} + B_{22} + B_{12} - (a+4b)\gamma_{12}, \quad \text{with} \quad B_{\alpha\beta} = \big(aN^{(1)}_{\alpha\beta} + bN^{(2)}_{\alpha\beta}\big)\gamma_{\alpha\beta}.$$

We will let $N^{(1)}_{\alpha\beta}$ denote the total number of bonds of type α and β connecting the atom at site (i,j) and its nearest neighbors. $N^{(2)}_{\alpha\beta}$ is analogously defined but for next-to-nearest neighbors instead. We observe for an isolated atom of type 2 on a substrate of type 1 that $B=0$. This implies that E_D is the energy barrier for the diffusion of a type 2 adatom on a type 1 substrate in the absence of elastic strain. The parameters a and b allow one to change the shape of the islands. For all our simulations we take $a=0.5$ and $b=1.8$.

The elastic interactions are accounted for using a ball and spring model with longitudinal and diagonal springs having spring constants k_L and k_D respectively. The elastic effects arise because the natural bond length of materials 1 and 2 are different. We will denote these lengths as a_1 and a_2. The misfit is then $\mu=(a_2-a_1)/a_1$. The details of this model can be found in Russo and Smereka [19] and Baskaran et al. [1]

In the next section, we review the energy localization method for fast simulation of KMC with strain developed in [1,22]. We present results of this method extended to 2+1 dimensions along with some convergence checks. We then introduce two approximations of this method which have the advantage of being faster and reasonably accurate. Finally, we present several examples of heteroepitaxial growth. A sequence of simulations is displayed which shows the effect of increasing the lattice misfit, followed by another one showing the effect of the deposition rate over the range from 0.01 monolayers/second to 10.0 monolyaers/second. We also present a series of annealing studies in which a single three dimensional island is annealed. It is shown that as the volume of the island increases one observes a transition not unlike the pyramid-to-dome transition observed for germanium on silicon.

2 Energy localization method

In this section we review our algorithm for simulating the ball and spring model described above. First, we recall the steps required for the implementation of an arbitrary KMC model:

1. Compute all of the rates $\{r_{ij}\}$.
2. Compute the partial sums
$$p_{IJ}=\sum_{i=1}^{I}\sum_{j=1}^{J}r_{ij}.$$
3. Generate a random number $r\in[0,p_{MM})$, for an M by M surface.
4. Execute the corresponding event.
5. Go to Step 1.

When elastic effects are included, the first step is by far the most time consuming, as each calculation of ΔW involves solving a large linear system. Considerable effort has been expended developing efficient methods for performing this calculation. In particular, multigrid methods with artificial boundary conditions have been developed to quickly solve for the elastic field [1,19,20].

In more recent work [22] we introduce three ideas aimed at speeding up this basic algorithm, which we extend to three dimensional growth in this paper. The first idea is to implement a rejection scheme, which reduces the number of rate calculations from $\mathcal{O}(M^2)$ to $\mathcal{O}(1)$. In [22] this was based upon an empirical observation of the energy barrier data. Below, we demonstrate that this observation extends to three dimensional films, but, more importantly, offer additional insight into why one should expect such an observation to hold. Next, we briefly review an energy localization principle, where we previously showed that very accurate estimates of ΔW can be obtained using local rather global updates of the elastic field. The final ingredient in the previous work was an expanding box method-a simple iterative technique based on successive over-relaxation (SOR) in a series of nested domains. Collectively, we refer to the full implementation of these three techniques as the energy localization method. In Section 3, we introduce further improvements of this method.

2.1 Reduced rejection scheme

As in the earlier, two-dimensional study [22], it is observed through extensive numerical calculations that

$$C_L w_{ij} < \Delta W < C_U w_{ij}, \tag{2.1}$$

where $C_L = C_L(k_L, k_D)$, $C_U = C_U(k_L, k_D)$ and w_{ij} is the total elastic energy contained in the springs attached to the moving atom *before* it is removed. We find, for example, that for the values of k_L and k_D used in this paper $C_U \approx 1.5$. These results are illustrated in Fig. 1, where we plot ΔW versus w_{ij} for every atom on a surface featuring a number of islands. Notice if one just removed the surface atom but did not allow the springs to relax, one would find that $\Delta W = w_{ij}$. This implies that the difference $\Delta W - w_{ij}$ corresponds to the work as the crystal relaxes after the atom is removed.

For our present purposes, C_U is used to determine an upper bound on the rates:

$$\hat{r}_{ij} = K \exp\left[\frac{(-E_D - B + C_U w_{i,j})}{k_B T}\right]. \tag{2.2}$$

The rate tables consist of these upper bounds on the rates. The results presented in this paper actually use a slightly smaller value of C_U, shown in Fig. 1. The rougher approximation gives a lower rejection rate and a somewhat faster computation. The number of events that are under-sampled as a result is extremely small; a calculation with the larger value of C_U was performed to verify that the results were unaffected by this choice. Further, the surface shown in Fig. 1 is a more extreme case, giving rise to more outlying data points, than that which is typically encountered. When an atom is selected, the true rate r_{ij} is then calculated and the move is accepted with a probability equal to $r_{ij}/\hat{r}_{ij}$. In future work, it may also be possible to exploit the lower bound as well to automatically accept a certain fraction of the candidate moves without performing any elastic updates.

One can begin to understand the nature of these bounds by considering a much simpler calculation using a well know result due to Eshelby [5]. Consider the situation of

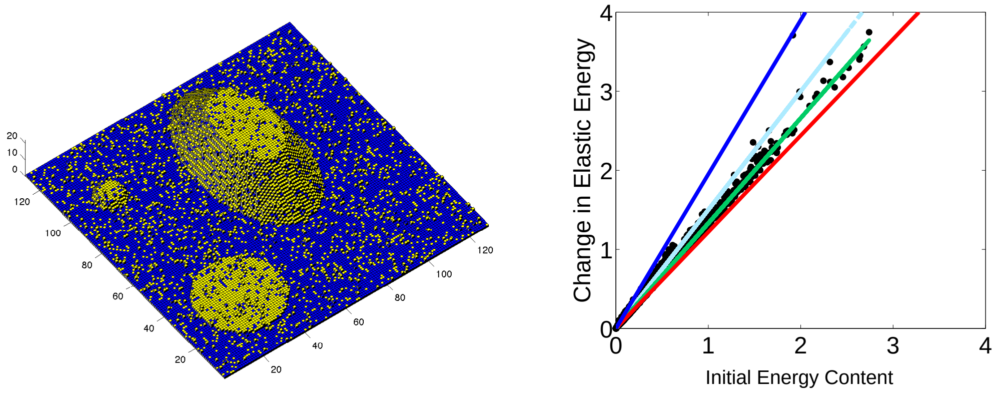


Figure 1: The figure on the right is a scatter plot of ΔW versus w_{ij} for every atom on the surface shown on the left. The blue and red curves are approximate upper and lower bounds, while the green curve is an approximate best fit, and the cyan curve is a less conservative upper bound used in the numerical procedures (see text).

an infinite, three-dimensional, elastically isotropic material with a varying stress field denoted t_{ij}. The elastic energy density can be decomposed into two pieces

$$w = U + V,$$

where

$$U = \frac{1}{18\kappa}(t_{\ell\ell})^2 \quad \text{and} \quad V = \frac{1}{4\mu}\Big(t_{ij} - \frac{1}{3}\delta_{ij}t_{\ell\ell}\Big)^2,$$

κ is the bulk modulus, μ is a Lamé coefficient, and σ is the Poisson ratio.

If one considers a slowly varying stress field and removes a relatively small spheroid of volume τ then Eshelby's calculations can be reworked to give the change in elastic energy as

$$\Delta W = \tau(AU + BV),$$

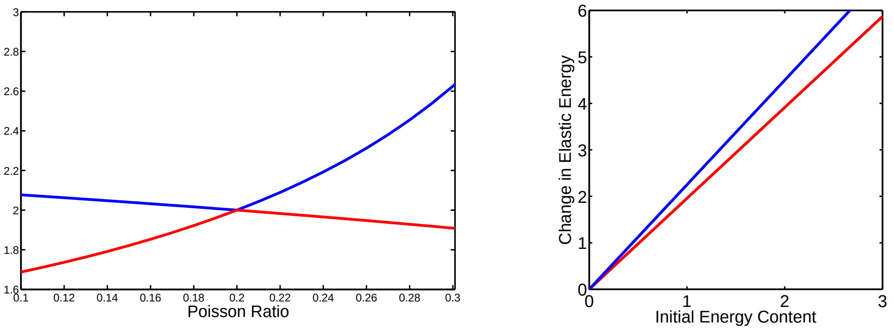


Figure 2: In the figure on the left, the bounds S_U (blue) and S_L (red) are plotted as a function of the Poisson ratio. As a result, the relationship between the change in elastic energy ΔW and the local energy density w_{ij} is bounded between the linear curves shown on the right.

where

$$A=\frac{3(1-\sigma)}{2-4\sigma} \quad \text{and} \quad B=\frac{15(1-\sigma)}{7-5\sigma}.$$

It is convenient to define

$$S_U=\max_{t_{ij}}\frac{AU+BV}{U+V}=\max(A,B) \quad \text{and} \quad S_L=\min_{t_{ij}}\frac{AU+BV}{U+V}=\min(A,B).$$

Since U and V are both positive it follows that

$$S_L\leq\frac{\Delta W}{\tau w}\leq S_U.$$

For reasonable σ, S_L and S_U are quite close (see Fig. 2).

2.2 The principle of energy localization

The fact that the change in elastic energy can be accurately calculated using local calculations is somewhat surprising, but can be understood from the work in [22] which we summarize here.

The argument is based on linear continuum elasticity. The exact energy barrier ΔE is defined to be

$$\begin{aligned}\Delta W=&W\big(\text{with surface atom at site}(i,j)\big)-W\big(\text{without surface atom at site }(i,j)\big)\\ =&\lim_{\rho\to\infty}\left[E(\mathbf{u};\Omega_\rho)-E(\mathbf{u}^a;\Omega_\rho^a)\right],\end{aligned}$$

where the energy barrier depends on the displacement fields for the initial configuration $\mathbf{u}$ integrated over a finite region, Ω_ρ, with characteristic size ρ and a second displacement field $\mathbf{u}^a$ for a slightly modified surface (representing the atom-off case) integrated over a correspondingly modified domain, Ω_ρ^a.

For the local approximation, we redefine the atom-off solution on a domain Ω_ρ^a with lower boundary constrained so that it agrees with the atom-on solution on some arc Γ_ρ:

$$\mathbf{u}_\rho^a=\mathbf{u}, \quad \mathbf{x}\in\Gamma_\rho.$$

Our approximation for the elastic energy barrier is then

$$\Delta W_L=W(\mathbf{u};\Omega_\rho)-W(\mathbf{u}_\rho^a;\Omega_\rho^a),$$

and we are able to prove that this approximate energy barrier satisfies the estimate

$$\Delta W=\Delta W_L\big(1+\mathcal{O}(\rho^{-2})\big), \quad \text{as } \rho\to\infty.$$

Naively, one might expect an approximation based on a simple truncation of the energy integrals,

$$\Delta W_T = W(\mathbf{u};\Omega_\rho) - W(\mathbf{u}^a;\Omega_\rho^a),$$

would be better. However, under the same assumptions used to prove the above estimate, we are able to show

$$\Delta W = \Delta W_T\big(1+\mathcal{O}(H\rho^{-1})\big), \quad \text{as } \rho\to\infty,$$

which scales much worse as the size of the local region ρ grows and gets worse still as the film thickness H increases.

2.3 Expanding box method

Following our earlier work [22], we use SOR to solve for the displacement field within localized regions. Since we are starting from a system which is relaxed, the first iteration of the displacement field within a localized region will have negligible effect for lattice points that are more than one lattice spacing from the change. Similarly, the second iteration will have significant impact up to two lattice spacings from the site of the change. In this way, the effect of any localized change continues to propagate into the lattice at the rate of one site per iteration. For this reason, we apply SOR on a region that expands no faster than this. In view of the energy localization observations, the boundary values for the displacement field are taken to be pre-correction values. Extensive experimentation with this technique indicates two applications of SOR for each box size is optimal for typical calculations.

To assess the accuracy of a local solution, the residual, defined as

$$R = AU - F, \tag{2.3}$$

is computed. We define the global residual error as

$$R_G = \frac{\|R\|_2}{\|F\|_2},$$

where $\|\cdot\|_2$ is the discrete L^2 norm. The local residual error in a region Ω_ρ is defined as

$$R_L = \frac{1}{\mu a_s k_L}\max_{\Omega_\rho}|R|, \tag{2.4}$$

where both R_G and R_L are dimensionless. A local calculation is considered successful if R_L is sufficiently small. In our previous work we prove, within the context of continuum elasticity, that the residual error decreases as the box size ρ is increased, scaling like $1/\rho^2$. This implies that the atoms just outside the last shell have a small net force and are therefore somewhat out of equilibrium.

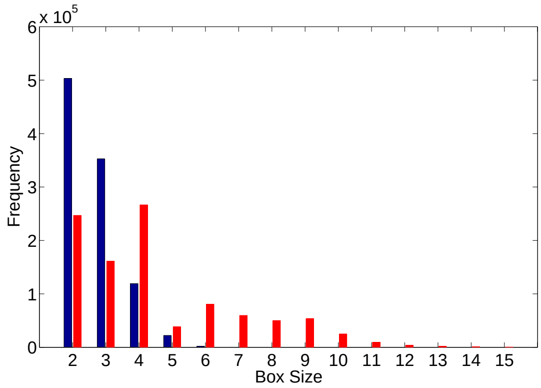


Figure 3: A histogram illustrating the frequency of various box sizes for which the expanding box method converges at two tolerance levels, $R_L=0.01$ (blue) and $R_L=0.1$ (red).

In cases where the size of the expanding box exceeds a threshold, $\rho_{\max}$, we abandon the local calculation in favor of a global one. If the threshold is set too small, the solution reverts to a global calculation too frequently; if it is set too large, then a local update can take longer than an application of the Fourier-multigrid method. In Fig. 3, a histogram illustrating the frequency of various box sizes for which the expanding box method converges at two tolerance levels, $R_L=0.01$ (blue) and $R_L=0.1$ (red) is shown. Typically, global solutions were needed for especially difficult configurations or as the result of accumulating errors in the displacement field after many local updates.

2.4 The algorithm

The complete algorithm is summarized below.

1. Select an event by choosing a uniformly distributed random number $r\in[0,\hat{R})$, with $\hat{R}=r_{dep}+\sum\hat{r}_{ij}$. This interval represents an overestimate of the sum of rates for atoms hopping plus the rate of deposition. The event to which r corresponds is located using a binary tree search [3].

2. If the event is a deposition, locally update the height and connection arrays and attempt a local elastic solve; revert to a full elastic solve if the expanding box exceeds size S. Update the rate estimates using (2.2) in the same region in which the elastic field was updated. Return to Step 1.

3. If the event selected is a hop, then take into account elastic effects:

   (a) Make a copy of the displacement field $\mathbf{u}_\rho$ (atom on). Follow the same procedures in Step 2 to compute the displacement field with the atom removed, $\mathbf{u}^a_\rho$ (atom off).

   (b) Once the elastic field has been updated (locally or globally as necessary), calculate the energy barrier and actual rate r_{ij} $(\leq\hat{r}_{ij})$.

   (c) Use rejection to decide whether or not to make the move. Note that the atom-off calculation must be performed whether or not this move is made.

   (d) If the move is rejected, no change is made to the displacement field. Return to Step 1.

(e) If the move is accepted, a hop is made. Update the displacement field in the vacated position using $\mathbf{u}^a_\rho$. Perform a second local/global calculation in the atom's new position thereby updating $\mathbf{u}$.

4. One event has been completed. Return to Step 1.

2.5 An example

The results shown in this paper use $T=800K$, energy barrier for diffusion $E_D=1.1eV$, $\gamma_{11}=0.18eV$, $\gamma_{12}=0.16eV$ and $\gamma_{22}=0.16eV$ $a_s=5.5\mathring{A}$. $k_L=62eV/a_s^2$,and $k_D=30eV/a_s^2$. These values come from Lee et al. [10], and were chosen to model the growth of InAs on GaAs. All of the simulations where done on a 128×128 film with periodic boundary conditions. Further, unless otherwise specified, the results shown below will correspond to a misfit of 0.07.

Fig. 4 shows a sequence of snapshots from a single simulation using the energy localization method in three dimensions. The coverage ranges from 0.05 to 2.0 monolayers.

Our basic tool for comparing the results of different simulations is a radially averaged autocorrelation function. First we define $\tilde{h}=h-\bar{h}$, where $\bar{h}$ is the mean surface height and compute the discrete form of

$$I(u,v)=\iint \tilde{h}(x-u,y-v)\tilde{h}(x,y)dxdy,$$

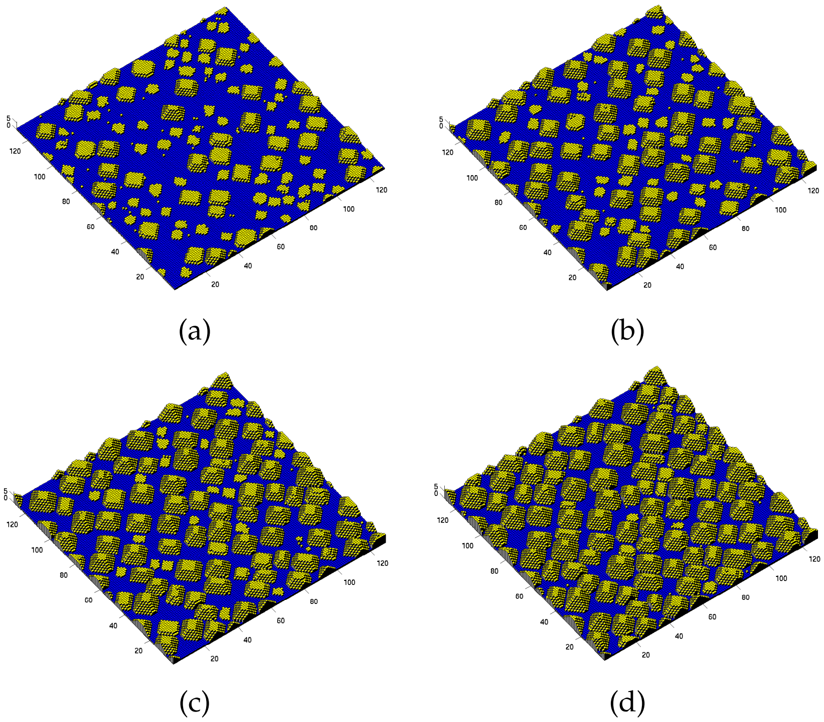
followed by

$$\bar{g}(R)=\frac{1}{2\pi R}\int I\big(u(r,\theta),v(r,\theta)\big)\delta(r-R)drd\theta,$$

where one uses a suitably mollified delta function. This gives a fairly robust measure of film characteristics at different length scales. For a given set of operating and material parameters, we frequently do some additional ensemble averaging over a set of four to eight simulations. The result is then characteristic of larger ensembles while remaining sensitive to changes in materials parameters. In Fig. 5, we plot $\bar{g}(R)$ for the four surfaces shown in Fig. 4, while Fig. 6 demonstrates the convergence of $\bar{g}(R)$ as the tolerance for the relative residual is reduced. The roughness of the film is $\sqrt{\bar{g}(0)}$, the island size is approximately the distance to the first zero of $\bar{g}(r)$, and the island spacing is approximately the distance between the first two local maxima, the region where $\bar{g}(r)<0$ being indicative of a bare substrate.

3 Faster methods

In this paper, we introduce two additional ideas for speeding up the simulations described above. While these can be combined, we study them separately in order to establish their validity. First, we consider a coarse grained random walk and follow this with a method that seeks to estimate the energy difference ΔW using the local energy density w_{ij}.



(a)

(b)

(c)

(d)

Figure 4: Snapshots of film growth for a flux of 1.0 ML/sec and a misfit $\mu = 0.07$ for different coverages in monolayers: 0.5 (a), 1.0 (b), 1.5 (c), and 2.0 (d).

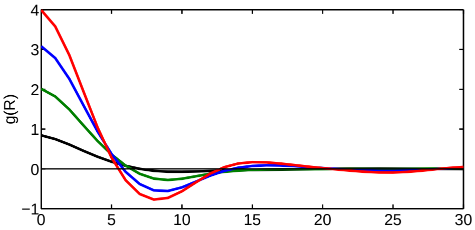


Figure 5: Autocorrelation curves for the film surfaces shown in Fig. 4 (0.5 monolayers (ML) black, 1.0 ML green, 1.5 ML blue, 2.0 ML red).

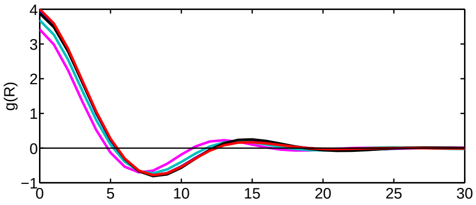


Figure 6: Convergence Check. Ensemble averaged auto correlation function for different tolerances: $\varepsilon = 1.0$ (magenta), $\varepsilon = 0.5$ (cyan), $\varepsilon = 0.1$ (black), $\varepsilon = 0.01$ (blue), and $\varepsilon = 0.005$ (red).

3.1 Coarsened random walks

In the basic implementation described above, each event requires a similar amount of computation, dominated by a local elastic update, which sometimes reverts to a global elastic update. However, the vast majority of these events are associated with adatom motion. Clearly, there is much to be gained if this particular type of processes can be handled quickly using a specialized approach. It turns out that the elastic contribution to the energy barrier is relatively small for adatoms, and that it varies relatively little, in an absolute sense, for atoms on the substrate (see Fig. 7). This suggests that, as an approximation, one could ignore this variation.

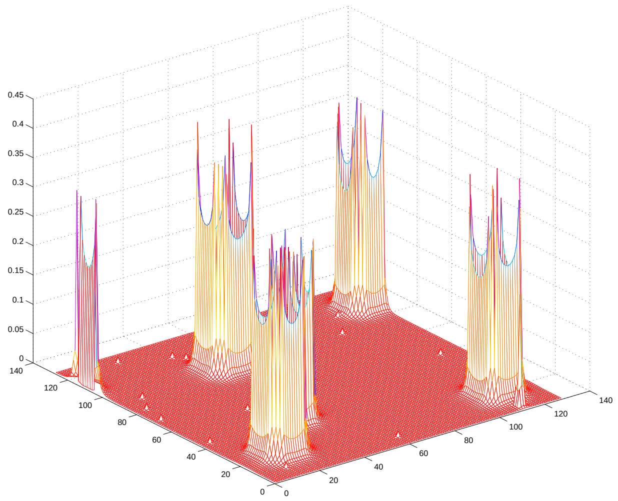


Figure 7: Plot of the elastic contribution to the energy barrier on a surface featuring several islands.

We aim to exploit these observations using some form of a coarse-grained hop for the adatoms. Since moving an adatom without updating the elastic field by what amounts to a standard KMC technique is essentially a zero-cost event compared to the cost of motion requiring elastic updates, in the present case we can simply implement the ″coarse-grained″ motion by simulating a short random walk using, as an approximation, the rate calculated at the adatom's initial position, and terminating the walk if the atom encounters another atom or a vacancy. Even in the absence of elastic contributions, one makes a potentially serious error in implementing such a procedure, in that one neglects potential interactions with other adatoms and/or vacancies *during* the coarse grained motion. Previous implementations of this idea [4] suggest that this interaction error can be neglected if the scale of the coarse grained motion is sufficiently small. It is clear, for example, that if adatoms are always allowed to hop until they attach to existing islands, the nucleation of new islands will never occur. If, however, the coarse grained motion is limited to a number of hops that is small compared to the typical island spacing, adatom interactions omitted by the coarse grained motion are found to be negligible. We will verify that this holds in the present case as well. What is essential is that each pair of adatoms has sufficient opportunity to form a dimer.

One way of implementing this is to compute an effective rate for an n-hop process, and include this in the list of events in place of the usual one-hop process. There are two potential problems with this. First, because there is the possibility that the coarse grained motion ends early due to interaction with a surface inhomogeneity, one must find a way of using the remaining hops in a way that does not corrupt the underlying stochastic process. In [4] this is accomplished by distributing the balance of the moves over other adatoms. Notice that this amounts to a reduced hop limit for the last adatom selected, even when it has not yet encountered any surface inhomogeneity. In the present implementation this is somewhat undesirable due to the expense of updating the elastic field at the end of coarse-grained moves. Thus, we adopt an alternative way of maintaining the correct balance between the various independent processes that involves assigning separate clocks to adatoms. This idea has been used successfully in other applications of coarse-grained random walks [18].

The second issue is that the waiting-time distribution for an n-hop move is not exponential, nor is it additive. For example, the two-hop waiting time distribution, assuming the events are independent, is

$$f(t_1,t_2)=f(t_1)f(t_2)=R^2e^{-R(t_1+t_2)},$$

where $f(t)=Re^{-Rt}$ is the usual exponential distribution with mean waiting time $1/R$. Treating this as one process, we can compute the exact distribution for the total waiting time

$$P\big(\{(t_1,t_2)|t_1+t_2=t\}\big)\equiv f_2(t)=\int_0^t f(t_1,t-t_1)dt_1=tR^2e^{-Rt}.$$

While one can verify that $\langle t\rangle=2/R$, the variance is not the same (i.e., if one samples the usual exponential distribution with the reduced rate, the mean is correct but the fluctuations are different). Further, there is no simple way to combine a two-hop process with other n-hop processes in the way one can with the one-hop processes, (i.e., while the mean waiting-times remain additive the distribution functions get increasingly complicated).

In the present application, the second issue is easily avoided by sampling the one-particle waiting time distribution repeatedly and adding up the result. While this would be too costly in terms of computational time for many KMC methods, the cost remains negligible compared to the cost of the elastic computations. Similarly, one can also afford to simulate the correct distribution for the atom's final location by simply generating the random walk one step at a time. This also allows one to check for surface inhomogeneity along the way.

3.1.1 Using multiple clocks

Both of the issues outlined above benefit from the idea of simulating some of the individual Poisson processes independently, while the remaining processes are handled collectively in the usual way. Let T_0^K represent the time reached by the bulk of the processes, which we refer to as the ″system″ and denote with a zero subscript, by a sequence of K random time increments, which we denote using a superscript:

$$T_0^K = \sum_{k=1}^{K} t^k.$$

If at some point in the overall process an adatom forms, we assign it its own clock, initialized using the current system time. The adatom's clock will move in bursts of several random steps, so that it becomes desynchronized, under the assumption that it does not interact with the system during its short random walk. After a while, several adatoms will be assigned to separate clocks T_n^K, none of which are equal to the system clock. At the beginning of each simulation time-step k, a single clock is advanced using a random increment

$$t^k = R_n \log r,$$

where r is uniformly distributed between zero and one, and R_n is either the hopping rate of an adatom or the sum of all of the ″system″ rates. If the event is an adatom, this can immediately be repeated until either the hop-limit is reached or a non-uniform environment is encountered. One might think it is most accurate to always choose the clock which is lagging, but careful testing has revealed that it is more accurate to choose the clock that minimizes its expected reading at the conclusion of the next step.

3.1.2 An example

One can get a sense for why the last statement is true by considering a simplified scenario in 1+1 dimensions, with an immobile substrate and an immobile island/wall at both ends of the domain. We consider just two processes: a slow deposition and a fast hop, with irreversible attachment. Suppose we do regular KMC. The first event is always a deposition. In most samples, this atom subsequently walks to and sticks to the wall *before* the second atom drops. Now consider the KMC algorithm with a separate clock for each process. The first event is the same. Then either the adatom makes one hop or a second atom is deposited. Since the adatom clock was just split off from the deposition clock, they have the same initial reading, so a straight comparison of the clocks gives us no preference for stepping one over the other at the next step. In that case, suppose we choose randomly. If we choose the adatom hop, its clock advances a small amount, so that the deposition clock is *slightly* lagging, but the configuration is essentially the same with the adatom displaced just one site. If we now step the process with the slowest clock (or if we chose the deposition event at the last step, when the clocks were equal), a second adatom is deposited, advancing its clock *far* ahead of the first adatom because deposition is a slow process. Either way, after the second atom drops there are now two atoms on the surface, with the first one's clock way behind both the second atom's clock and the deposition clock. So the first atom will make a bunch of hops before its clock catches up, giving it a rather large chance of nucleating an island rather than attaching to a wall.

This enhanced nucleation effect is exaggerated in 1+1-dimensional growth, but even in 2+1-dimensional growth, we find that choosing the lagging clock systematically distorts the results slightly in favor of enhanced nucleation. Thus, this example suggests that when comparing clocks to decide which process to update next, add to each clock the current mean waiting time. While not exact, the improvement is significant. This will be illustrated further in the results presented below.

3.1.3 Clock merging principle

In principle, each one of the surface sites could be assigned their own clocks, but only the ones that are allowed to take multiple time steps need their own clocks. Once an adatom encounters another atom or vacancy, we no longer wish to advance it multiple times, so it can once again be merged with the system clock. Some care must be taken in the merging, however, which exploits the well known property of Poisson processes that the waiting time distribution at the current time is independent of how long you have waited to the current time. This leads to the following clock merging principle:

When a clock passes another clock on a given random increment, they can be merged at the value of the lagging clock.

This is equivalent to the claim that the sequence of waiting times generated by successively sampling the one-hop distribution $f(t)$ is not corrupted if we force the system to make a stop at a particular time T^* and then resume the process. Suppose the process is at time T^k. If we sample $f(t)$ and get a sample t_1 where $T^k+t_1<T^*$, the distribution is clearly unaffected. If, on the other hand, $T^k+t_1>T^*$, replace t_1 with $t^*=T^*-T^k$, and sample $f(t)$ a second time. The probability of this happening is

$$P(t_1>t^*)=\int_{t^*}^{\infty} f(t)dt=e^{-Rt^*}.$$

Since the second sample is independent of the first, the distribution for $t>t^*$ becomes

$$P(t_1>t^*)f(t_2)=e^{-Rt^*}Re^{-Rt_2},$$

corresponding to a total $t=t^*+t_2$ distributed by $f(t)$.

Thus, two things can happen to an atom which is no longer an adatom and therefore a candidate for merging when it is its turn to move: 1) it may take a time step that fails to catch it up to the system clock, in which case it remains an independent process, or 2) it may take a step that surpasses the system clock, in which case it is merged with the system (unless it happens to have become an adatom again).

4 The local energy method

As indicated earlier, numerical experiments (Fig. 1) demonstrate that the elastic energy barrier is strongly correlated with the elastic energy contained in the springs immediately attached to the atom being removed. This observation was first made in our earlier work [22], where we used it to estimate upper bounds for hopping rates, which were then used as part of a rejection scheme. When the upper and lower bounds are close, it is reasonable to simply replace the rejection scheme with the approximation

$$\Delta W \approx C(k_L,k_D)w_{ij}, \tag{4.1}$$

where C is now chosen to achieve a best fit rather than a bound. It is perhaps useful to note that ΔW is precisely equal to w_{ij} plus the work required to reinsert the atom and restore the original configuration. In view of this, it is not entirely surprising that these quantities would be comparable and that some sort of scaling law would hold. In future work we plan to use the work of Eshelby [5] to obtain bounds on the latter contribution in the case of semi-infinite elastic solid with a flat free boundary. A crucial aspect of this calculation is a knowledge of the Green's function for the half space problem. For the isotropic case this is known (see, for example, [8]). For the anisotropic case this was recently studied in [17].

Employing (4.1) instead of calculating ΔW by the energy localization method achieves a roughly fifty percent reduction in computational cost, as one no longer needs to do atom-off calculations. In this scheme we still employ the expanding box method to update the elastic field after the atom is moved to its new location.

5 Comparison of the methods

In Fig. 8 we plot the ensemble averages of four autocorrelation curves using the test parameters identified above for a) the unapproximated energy localization method, b) the local energy approximation, and c) the coarse grained random walks with up to twenty-five hops per iteration. The fact that the autocorrelation curves agree closely indicates that both approximations are capturing the essential features of the film surface correctly. In particular, the roughness, island size and island spacing appear to agree well. The local energy approximation was a bit more than twice as fast, while the big hop method was about 25% faster. This latter number improves significantly for simulations in the submonolayer regime, where a much larger fraction of the events correspond to adatoms moving on the substrate. This performance boost is further enhanced in the low deposition regime, which features a lower island density and larger regions of bare substrate. If, for example, the deposition rate is reduced by a factor of 100, the performance of the coarse graining method is roughly a factor of ten better than the original energy localization method during the first one tenth of a monolayer of growth.

The quality of the coarse-grained results improves as the step size is lowered. This is illustrated in Fig. 9, where we plot the ensemble averaged, unapproximated autocorrelation curve along the approximations with stepsizes of one, five and twenty-five.

It would be possible to combine the two approximations for an additional increase in computational speed. For this paper, however, we will use the local energy approximation to compute our remaining results, as we feel the error it introduces is more uniform

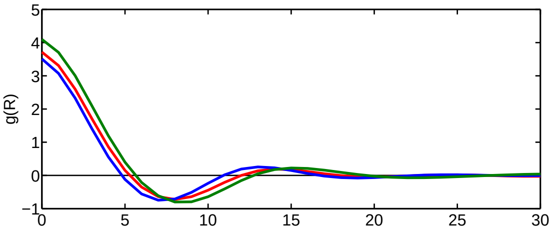


Figure 8: Autocorrelation curves using the test parameters identified in Section 2.5 for (a) the unapproximated energy localization method (red), (b) the local energy approximation (blue), and (c) the coarse grained random walks with up to 25 hops per iteration (green).

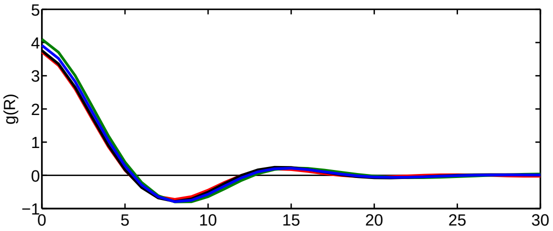


Figure 9: Comparison of the coarse-grained random walk method using varying step sizes-25 (green), 5 (blue), 1 (black)-with the energy localization method (red).

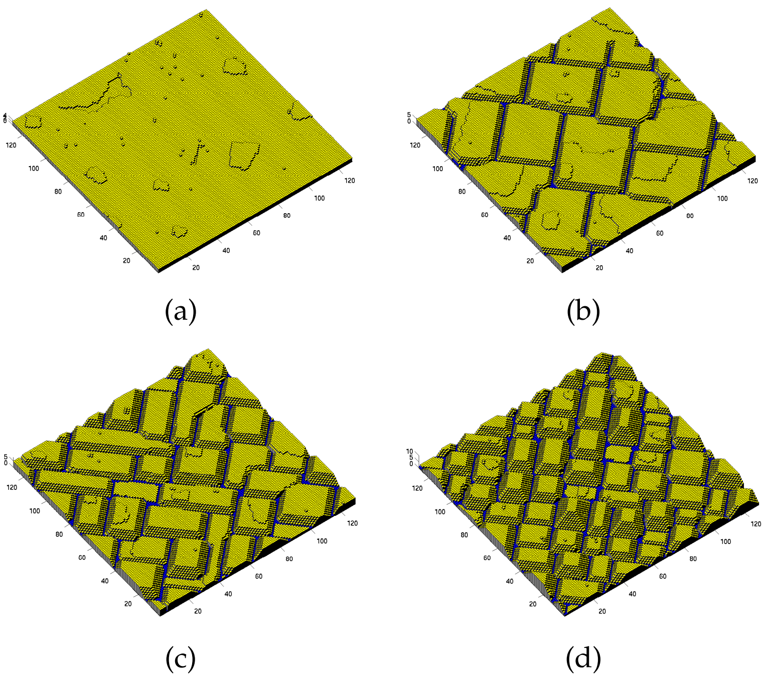
from case to case, and it seems more likely that this will become the method of choice for future calculations based on the present model.

6 Simulation of heteroepitaxial growth

In this section we shall apply the local energy method to study various aspects of heteroepitaxial growth. In the first series of simulations the misfit is varied.

6.1 Misfit and flux variation

In Fig. 10, we show four surfaces after three monolayers of growth for misfits varying from 0.03 to 0.05, while the remaining parameters are held fixed. The corresponding autocorrelation functions, averaged over four samples, are shown in Fig. 11. These figures show that the average island size decreases with increasing misfit. It is interesting to note that while relatively small and large misfits yield square islands there seems to be an intermediate range that favors rectangular islands. In addition, these results seem to indicate that a critical misfit is needed to form islands after three monolayers of deposition. This, however, is a nonequilibrium effect. Increasing the temperature will drive the system to equilibrium faster. Similarly, lowering the deposition rate will allow the system more time to relax. Fig. 12 shows simulations that support these comments. In particular, we consider the case in which the misfit is 0.03, but the deposition and temperature are



(a)

(b)

(c)

(d)

Figure 10: Snapshots of a film at three monolayers of growth with a flux of 1.0 ML per second with a misfits $\mu = 0.03$ (a), 0.035 (b), 0.04 (c), and 0.045 (d).

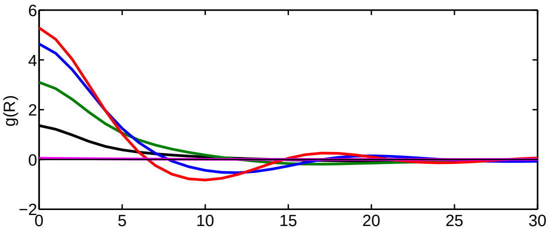
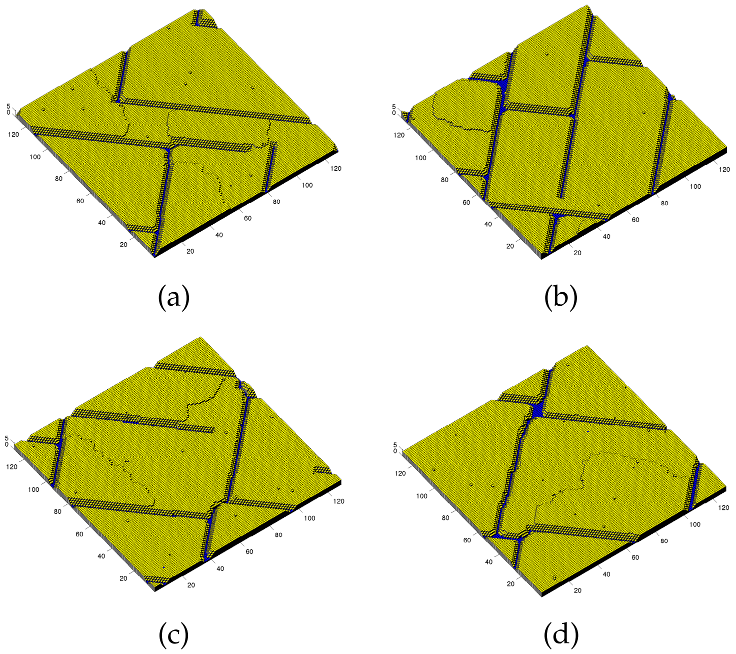


Figure 11: Autocorrelation curves for the surfaces shown in Fig. 10 and the lower left panel of Fig. 13: misfit 0.03 (magenta), 0.035 (black), 0.04 (green), 0.045 (blue), and 0.05 (red).

varied. In the upper two images the flux has been reduced to 0.1 and 0.01 monolayers per second allowing the film more time to relax, which results in the formation of islands. In the lower two images the flux is one monolayer per second, but we have increased the temperature to 900 and 950K, resulting in the formation of islands.

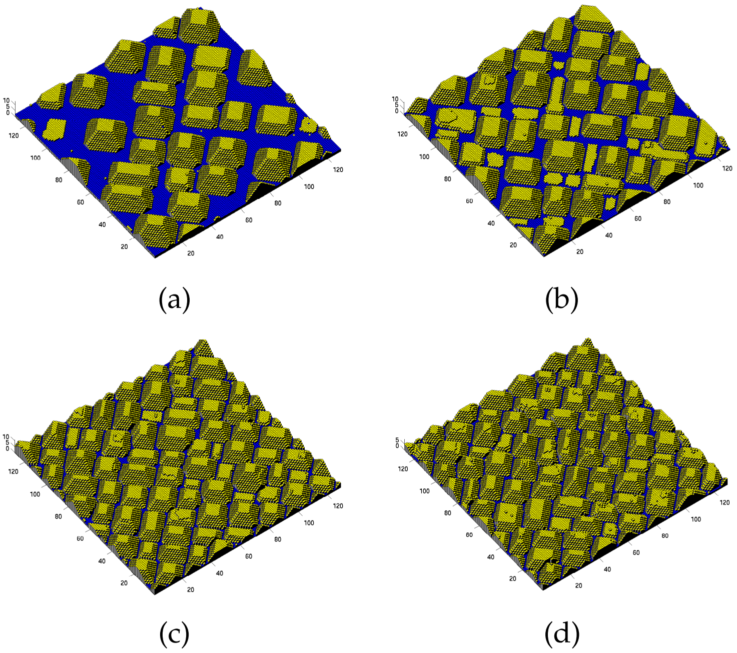
Fig. 13 displays a sequence of simulations where the flux is varied over four orders of magnitude; the values used were 10^{-2}, 10^{-1}, 1, and 10 monolayers per second. The results indicate the morphology of the film changes quite slowly when varying the flux. Not surprisingly, smaller values of the flux yield larger and more well separated islands. This suggests there is a large entropic bottleneck for the formation of relatively large, three-dimensional islands.



(a) (b)

(c) (d)

Figure 12: Snapshots of a film at three monolayers of growth with misfit μ= 0.03. (a): flux = 0.1 ML/second, temperature 800K; (b): flux=0.01 ML/second, temperature 800K; (c): flux= 1.0 ML/second, temperature 900K; and (d): flux=1.0 ML/second, temperature 950K.



(a) (b)

(c) (d)

Figure 13: Snapshots of a film at three monolayers of growth with a misfit of 0.05 and fluxes equal to 10^{-2} ML/second (a), 10^{-1} ML/second (b), 1 ML/second (c), and 10 ML/second (d).

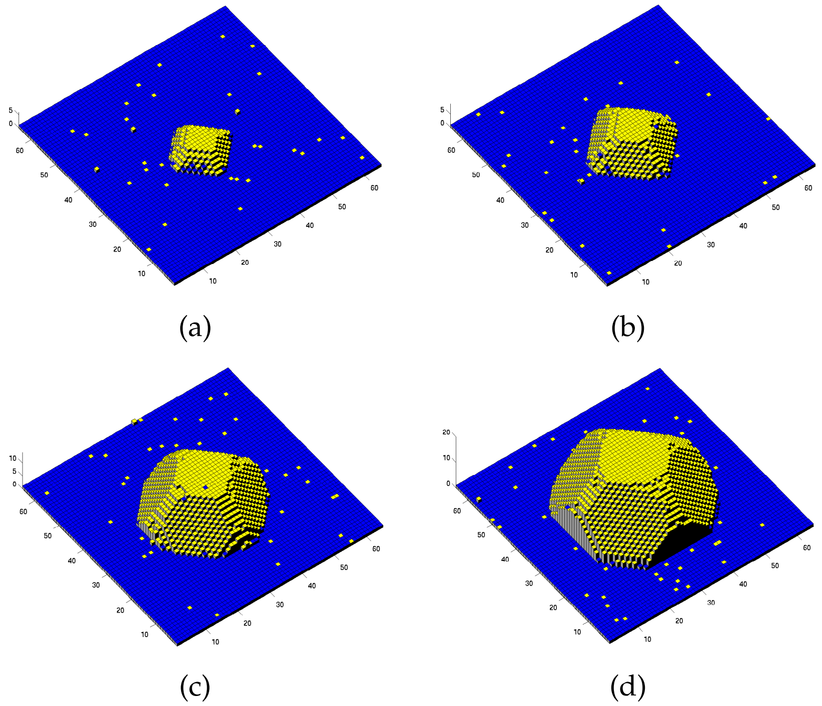


Figure 14: Annealed shapes in which the total number of atoms is varied with misfit equals 0.04 with no deposition.

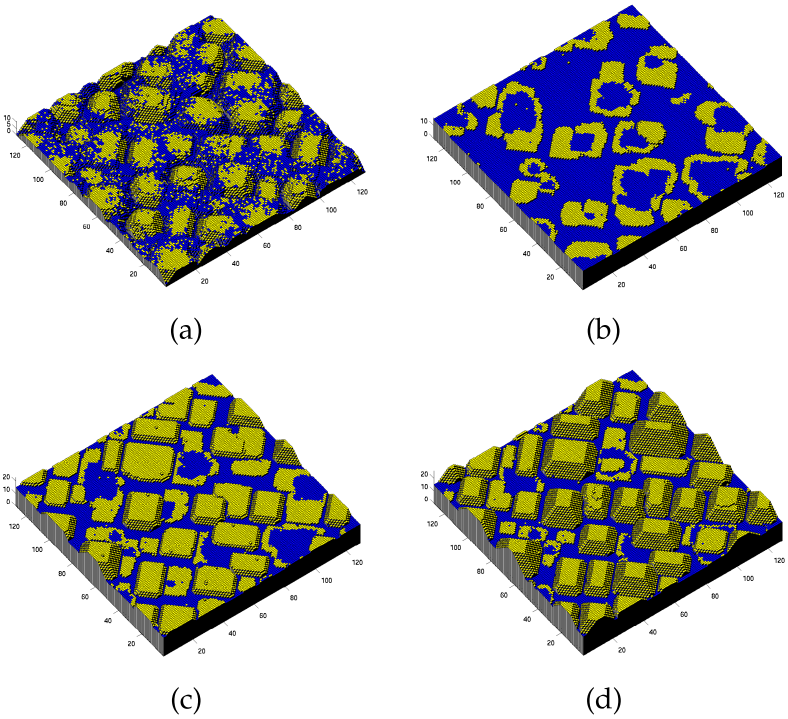
6.2 Equilibrium three dimensional islands

The next sequence of pictures was generated by annealing, for two billion KMC timesteps, a number of isolated, initially cube-shaped islands with different initial sizes. There is an interesting transition in the typical morphology as the size of the islands increases, not unlike the pyramid-to-dome transition observed in the deposition of germanium on silicon [13]. The overall trend is for additional facets to form as the islands become larger.

6.3 Stacked quantum dots

Fig. 15 is a simulation of capping, where a number of layers of substrate material are deposited after islands have formed. This process can be repeated to build structures reminiscent of stacked quantum dots, see for example [11,14]. In more detail, we choose the misfit to be 0.05 and a deposition flux of 0.1 ML/sec. We deposit three monolayers of film material, followed by ten monolayers of capping/substrate material. This is repeated once and then a final three layers of film material is deposited, so that in total there are three layers of islands, separated by two layers of capping. The state of the film after the initial three monolayers of growth is shown in lower-left panel of Fig. 13.

The upper-left panel of Fig. 15 shows this same surface after two layers of capping. At this stage, the islands have been eroded somewhat due to intermixing; also note that



(a) (b)

(c) (d)

Figure 15: Capping simulations with a misfit of 0.05 and flux of 0.1 monolayers per second. The initial condition for this simulation is shown in Fig. 11(c). (a) is after two monolayers of capping. (b) is after 10 monolayers of capping and 0.5 monolayers of deposition. (c) and (d) show the film after 1.5 and 3 monolayers of deposition respectively.

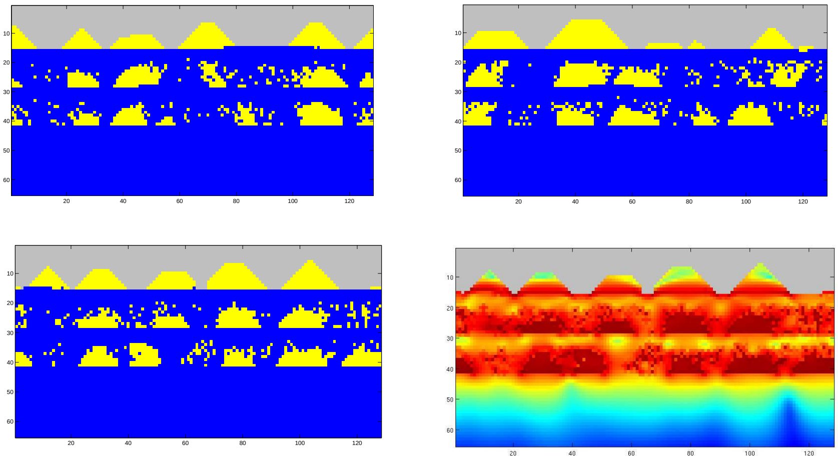


Figure 16: Slices of the capping simulation at thee monolayers and ten monolayers of capping; three monolayers of deposition then ten monolayers of capping and then misfit 0.05, flux 0.1. The last plot is a plot of elastic energy density. The color is scaled to the logarithm of the energy density.

the capping material prefers to grow in the low strain regions between well separated islands. The upper-right portion of Fig. 15 shows the situation after ten layers of capping plus another half layer of the film material. The morphology at this stage is quite interesting. Notice that the final monolayer of capping material is incomplete, with several islands on the surface. These islands, however, are not necessarily aligned with the underlying quantum-dot structures. When the film material is subsequently deposited, it initially continues the growth of these islands in the manner of layer-by-layer growth, and only later do the islands align with the underlying, capped, quantum dots. Importantly, it does not appear that the new layer of quantum dots nucleates over the existing layer, as has been previously suggested [15,23], but rather there is a somewhat haphazard nucleation process, followed by a subtle migration of the dots to the aligned positions. The lower portion of Fig. 15 shows the latter stages of the evolution in more detail.

Fig. 16 shows three cross-sections of the same growth sequence after its completion along with the elastic energy density for the third cross section. These figures demonstrate the alignment of the dots via the migration mechanism mentioned above. The final image shows a cross-section of the elastic energy density, revealing that the exposed layer of dots is very relaxed, whereas the buried dots are under considerable stress.

7 Summary and conclusions

In this work, we have extended our previous energy localization and expanding box methods from 1+1 to 2+1 dimensional simulations. We have also introduced two new approximating methods aimed at further improvements in computational performance: a coarse grained random walk for adatoms and a local energy approximation. We went on to compare the performance and accuracy of the three methods. Collectively, these methods are demonstrating the viability of using KMC to simulate the growth of films in situations where elastic effects dominate.

The coarse-graining method must be used with care, as it can quite easily alter nucleation statistics. This method is most useful in the submonolayer regime, where it can significantly reduce computation times. Most of our results in this paper are aimed at the growth of quantum dot structures that emerge after the deposition of several layers of growth. For this, we relied on the local energy approximation.

We presented a brief analysis based on linear elasticity that offers insight into several aspects of these methods. In particular, it now seems clear why one should expect bounds of the type (2.1) to hold, whereas these bounds were based entirely on computational observations in our earlier work. From these calculations, one can also see why there appears to be a particular value of the poisson ratio near which the upper and lower bounds collapse. In this paper, the material parameters we worked with appear to be near this special case, allowing the local energy method to work well. For more general material parameters, a generalized approximation based on scalar invariants of the local stress tensor could be used. This will be explored in future work, where we will also aim to make these arguments quantitative by adapting the work of Eshelby to a half-space geometry-we will seek to derive the upper and lower bounds used in the computation and/or the fitting paramers used to approximate the energy barrier.

Acknowledgments

This research was supported, in part, by NSF grants DMS-0810113 (PS) and DMS-0707443 (TPS) and DMS-0854870 (FRG-TPS and PS).

References

[1] A. Baskaran, J. Devita, and P. Smereka, Kinetic Monte Carlo simulation of strained heteropitaxy griwth with intermixing, Cont. Mech. Thermo., 22 (2010), 1–26.

[2] M. Biehl, M. Ahr, W. Kinzel, and F. Much, Kinetic Monte Carlo simulations of heteroepitaxial growth, Thin Solid Films, 428 (2003), 52–55.

[3] J. L. Blue, I. Biechl, and F. Sullivan, Faster Monte Carlo simulations, Phys. Rev. E, 51 (1995), 876.

[4] J. P. Devita, L. M. Sander, and P. Smereka, Multiscale kinetic Monte Carlo for simulating epitaxial growth, Phys. Rev. B, 72 (2005), 205421.

[5] J. D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion, and related problems, Proc. Roy. Soc. Lond. A, 241 (1957), 376–396.

[6] W. Guo, T. P. Schulze, and W. E, Simulation of impurity diffusion in a strained nanowire using off-lattice KMC, Commun. Comput. Phys., 2 (2007), 164–176.

[7] G. Henkelman, B. P. Uberuaga, and H. Jonsson, A climbing image nudged elastic band method for finding saddle points and minimum energy paths, J. Chem. Phys., 113 (2000), 9901–9904.

[8] M. Kachanov, B. Shafiro, and I. Tsukrov, Handbook of Elasticity Solutions, Kluwer Academic Publishers, Dordrecht, 2003.

[9] C. H. Lam, C. K. Lee, and L. M. Sander, Competing roughening mechanisms in strained heteroepitaxy: a fast kinetic Monte Carlo study, Phys. Rev. Lett., 89 (2002), 16102.

[10] J. Y. Lee, M. J. Noordhoek, P. Smereka, H. McKay, and J. M. Millunchick, Filling of hole arrays with InAs quantum dots, Nanotechnology, 20 (2009), 285305.

[11] B. Lita, R. S. Goldman, J. D. Phillips, and P. K. Bhattacharya, Nanometer-scale studies of vertical organization and evolution of stacked self-assembled InAs/GaAs quantum dots, Appl. Phys. Lett., 74 (1999), 2824–2826.

[12] M. T. Lung, C. H. Lam, and L. M. Sander, Island, pit, and groove formation in strained heteroepitaxy, Phys. Rev. Lett., 95 (2005), 086102.

[13] G. Medeiros-Ribeiro, M. Bratkovski, T. I. Kamins, D. A. A. Ohlberg, and R. S. Williams, Shape transition of germanium nanocrystals on a silicon (001) surface from pyramids to domes, Science, 279 (1998), 353–355.

[14] J. M. Millunchick, R. D. Twesten, D. M. Follstaedt, S. R. Lee, E. D. Jones, Y. Zhang, S. P. Ahrenkiel, and A. Mascarenhas, Lateral composition modulation in AlAs/InAs short period superlattices grown on InP (001), Appl. Phys. Lett., 70 (1997), 1402–1404.

[15] X. B. Niu, Y. J. Lee, R. E Caflisch, and C. Ratsch, Optimal capping layer thickness for stacked quantum dots, Phys. Rev. Lett., (2008), 086103.

[16] B. G. Orr, D. A. Kessler, C. W. Snyder, and L. M. Sander, A model for strain-induced roughening and coherent island growth, Europhys. Lett., 19 (1992), 33–38.

[17] E. Pan and F. G. Yuan, Three dimension Green's functions in anisotropic bimaterials, Int. J. Solids Struct., 37 (2000), 5329–5351.

[18] M. Plapp and A. Karma, Multiscale finite-difference-diffusion-Monte-Carlo method for simulating dendritic solidification, J. Comput. Phys., 165 (2000), 592–619.

[19] G. Russo and P. Smereka, Kinetic Monte Carlo simulation of strained epitaxial growth in three dimensions, J. Comput. Phys., 214 (2006), 809–828.

[20] G. Russo and P. Smereka, A multigrid-Fourier method for the computation of elastic fields with application to heteroepitaxy, Multiscale Model. Simu., 5 (2006), 130–148.

[21] M. R. Srensen and A. F. Voter, Temperature-accelerated dynamics for simulation of infrequent events, J. Chem. Phys., 112 (2000), 9599–9606.

[22] T. P. Schulze and P. Smereka, An energy localization principle and its application to fast kinetic Monte Carlo simulation of heteroepitaxial growth, J. Mech. Phys. Sol., 3 (2009), 521–538.

[23] J. Tersoff, C. Teichert, and M. G. Lagally, Self-organization in growth of quantum dot superlattices, Phys. Rev. Lett., 76 (1996), 1675–1678.

[24] A. F. Voter, F. Montalenti, and T. C. Germann, Extending the time scale in atomistic simulation of materials, Annu. Rev. Mater. Res., 32 (2002), 321–346.