Modeling and Control of Epitaxial Thin Film Growth

Thesis by
Martha A. Gallivan

Technical Report
for the
Division of Engineering and Applied Science

California Institute of Technology
Pasadena, California

2003

(Thesis defended September 18, 2002)

Acknowledgements

I consider myself fortunate to have interacted with many faculty members during my time at Caltech. In particular, I would like to thank Kaushik Bhattacharya and Michael Ortiz, for their interest and involvement throughout; Harry Atwater, for generously welcoming me into his lab; Dave Goodwin, for his deliberate critiques; and Richard Murray, for his unwavering support and encouragement, and his infectious enthusiasm. I would also like to thank Professors Melany Hunt at Caltech and Panagiotis Christofides at UCLA for graciously agreeing to serve on my thesis committee.

I cannot imagine going through this experience without the friends I have made at Caltech. In particular, I thank Kristi Morgansen and Doug MacMartin for their support during the conclusion of my Ph.D. and job search. Loyal friends Clancy Rowley and Jeff Eldredge were there from the very beginning, with countless lunches along the way.

I would also like to thank my parents, for sharing with me their love of learning; and my husband Justin, for always being there.

Abstract

Thin film deposition is a manufacturing process in which tolerances may approach the size of individual atoms. The final film is highly sensitive to the processing conditions, which can be intentionally manipulated to control film properties. A lattice model of surface evolution during thin film growth captures many important features, including the nucleation and growth of clusters of atoms and the propagation of atomic-height steps. The dimension of this probabilistic master equation is too large to directly simulate for any physically realistic domain, and instead stochastic realizations of the lattice model are obtained with kinetic Monte Carlo simulations.

In this thesis simpler representations of the master equation are developed for use in analysis and control. The static map between macroscopic process conditions and microscopic transition rates is first analyzed. In the limit of fast periodic process parameters, the surface responds only to the mean transition rates, and, since the map between process parameters and transition rates is nonlinear, new effective combinations of transition rates may be generated. These effective rates are the convex hull of the set of instantaneous rates.

The map between transition rates and expected film properties is also studied. The dimension of a master equation can be reduced by eliminating or grouping configurations, yielding a reduced-order master equation that approximates the original one. A linear method for identifying the coefficients in a master equation is then developed, using only simulation data. These concepts are extended to generate low-order master equations that approximate the dynamic behavior seen in large Monte Carlo simulations. The models are then used to compute optimal time-varying process parameters.

The thesis concludes with an experimental and modeling study of germanium film growth, using molecular beam epitaxy and reflection high-energy electron diffraction. Growth under continuous and pulsed flux is compared in experiment, and physical parameters for the lattice model are extracted. The pulsing accessible in the experiment does not trigger a change in growth mode, which is consistent with the Monte Carlo simulations. The simulations are then used to suggest other growth strategies to produce rougher or smoother surfaces.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Thin film deposition is a manufacturing process in which precursor material is delivered to a surface on which it subsequently attaches, forming a solid layer of material. The resulting film may be as thin as a few atomic layers, or as thick as several micrometers. The deposition of a thin film is a critical step in the manufacture of integrated circuits, which has led to substantial advances in deposition technology over the last 50 years [53]. Other applications requiring thin films include solar cells, mechanical coatings, and, more recently, microelectromechanical systems and microfluidic devices.

The term *thin film deposition* encompasses a wide range of processes [53]. The medium surrounding the deposition surface may be a gas, liquid, or an ultra high vacuum. The precursor material may be produced by evaporation, through gas-phase reactions, or by sputtering a target with ions. A bias voltage may be applied to the substrate to generate a plasma in the gas phase, or to induce deposition in an electrolyte solution. The choice of a particular process depends on the material to be deposited, on the initial substrate, and on the manufacturing tolerances.

As device size in integrated circuits continues to shrink, increasingly stringent manufacturing tolerances on thin films are required. Films must be more uniform across large wafers, and contain fewer defects at the micrometer and nanometer scales. The incorporation of larger numbers of layers in a single device also means that each layer must contain fewer defects to maintain the same yield on the final device [53].

1.1 Sensing and control of thin film deposition

A common source of variation in film properties is a drift in deposition conditions over time. Sensors are often used to measure and compensate for this uncertainty, by either measuring reactor conditions like pressure or gas concentration, or by directly measuring film properties. A wide variety of film and surface sensors are used in practice, including pyrometry, x-ray and electron diffraction, and optical spectroscopy [3, 25]. These sensors provide information ranging from temperature to chemical composition to crystal structure.

Surface temperature is one critical deposition parameter, particularly at low

growth temperatures where the rate-limiting step to deposition is surface reaction. In this regime, a uniform temperature must be maintained precisely across the entire wafer to ensure uniform thickness in the final film. Thermal control is also needed to track aggressive temperature trajectories that enable higher throughput in single-wafer processing, as summarized in a survey article on rapid thermal processing [50]. A typical control strategy includes pyrometer sensing to determine wafer temperature, with subsequent adjustment of power to radiant heaters. The author has also studied thermal control, particularly in the area of reactor design, to obtain a system with good thermal response, uniformity, and temperature observation [17].

Control has also been applied to the delivery of precursor material to the wafer [16, 61]. Thickness and chemical composition have been controlled using gas-phase absorption measurements and ellipsometry. Both feedback and feedforward control strategies have been developed and implemented, based on bulk low-order models of precursor transport. In both of the examples cited, the sensors are closely tied to the control objectives, which consist of step changes in surface composition.

When the transport of precursors in the deposition chamber does not admit a simple low-order description, inlet conditions may be computed by applying optimization methods to the partial differential equations for a chemically reacting flow. This approach has been developed and demonstrated in simulation for a chemical vapor deposition process [47]. Optimal inlet concentrations are computed to achieve a desired concentration profile at the surface in the presence of a time-varying substrate temperature.

1.2 Models of thin film evolution

In the examples previously cited, the models describing the evolution of the control objective are based on continuum descriptions of heat transfer, fluid flow, and chemical kinetics. However, other potential control objectives include film properties like roughness, species segregation, and faceting. Control analysis and design require a mathematical representation describing how the precursors arrange themselves on the surface, but low-order models for this type of behavior do not generally exist.

Theoretically, one could build a complete mathematical model of thin film evolution using quantum mechanical principles, but the large number of atoms involved would make computation prohibitive. In practice, a model is developed to describe a particular quantity of interest in the most compact manner that adequately describes the relevant physical phenomena. Depending on the application, the model may describe the motion of a collection of individual atoms, or a continuum approximation may be used to describe behavior at macroscopic lengthscales. In the later case, the model may be self-contained, or it may require input from models at smaller scales. For example, quantum mechanical computations can be used to provide atomic potentials for molecular dynamics simulations, which subsequently provide rates of atomic transitions between sites on a crystal lattice.

1.3 Surface morphology

The focus of this thesis is on the evolution of the *surface* of a film. As material is deposited on the surface (or removed during the reverse etching process), the mean thickness of the film changes. However, the surface height profile, or surface morphology, also evolves as atoms are added or removed. In many applications, an atomically flat surface is the desired morphology, to provide a smooth interface in a layered device. Other control objectives include regularly spaced clusters of atoms for a quantum dot array [36], or a pyramidal, faceted surface to enhance optical absorption [63].

When atoms are deposited on a crystalline surface, the resulting film may be influenced by the initial surface. If the underlying crystal structure of the film is compatible with that of the initial surface, and if the temperature is sufficiently high to enable reorganization of the deposited atoms, the film may be crystalline and aligned with the underlying substrate. This process is referred to as epitaxial film growth.

The evolution of a crystalline film is dominated by the presence of atomic-height steps on the surface, and the attachment of atoms to these steps. As atoms are deposited on a surface, they may diffuse along the surface or desorb back to the surrounding gas. Surface diffusion and desorption enable a reorganization of the surface during growth to lower energy configurations. Typically, the rates of desorption and surface diffusion depend on the number of bonds each atom has, such that an atom with more bonds is less likely to move. Consequently, a mobile atom or cluster may eventually migrate to a step edge, which then increases the number of bonds, and lowers its mobility. During deposition, mobile species are constantly being created as atoms attach to the surface, leading to a net positive flux of atoms into step edges, and therefore resulting in the propagation of surface steps.

Step edges may exist in the initial growth surface, or they may be created during growth. Parallel arrays of steps are usually present on the initial wafer, since it will never be cut perfectly along the nominal crystallographic direction. Steps also originate from defects either in the initial surface, or in the growing film. The growth process itself may also produce steps, when mobile species collide with each other and form clusters. The edges of these clusters provide additional step edges, whose propagation leads to further growth of the cluster.

The origin of steps, and the relative importance of step propagation and step formation, strongly influence the evolution of surface morphology. Figure 1.1 is a scanning tunneling microscopy image of a silver surface. Two steps span across the image, originating from substrate miscut, while clusters of atoms have formed between the steps. A very different surface morphology is shown in the atomic force microscopy image of Figure 1.2. In this case, screw dislocations provide the initial source of steps. As atoms migrate and attach to the steps, the steps wind up into a spirals, which are anchored at the dislocation core. No cluster nucleation is observed on this surface. Growth instead proceeds as atoms absorb on the terraces and migrate to the step edges.

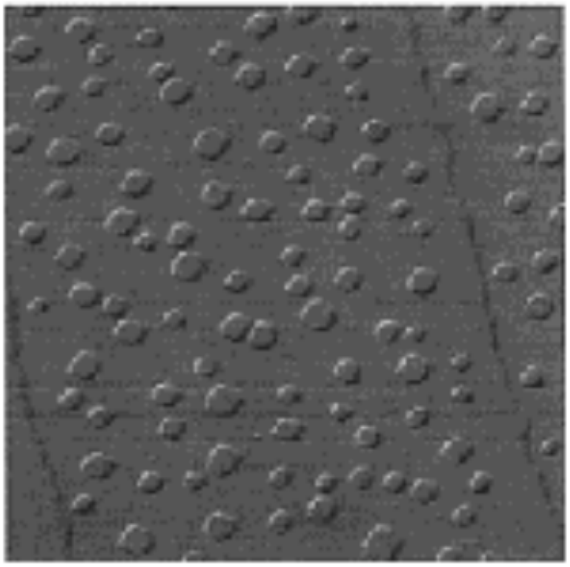


Figure 1.1: Scanning tunneling microscopy image of a Ag surface. The scanned area is 176 nm×176 nm. Both islands and steps are visible. Reproduced from Zhang et al., Surface Science, v. 406, 1998, pages 718-193, with permission from Elsevier Science.

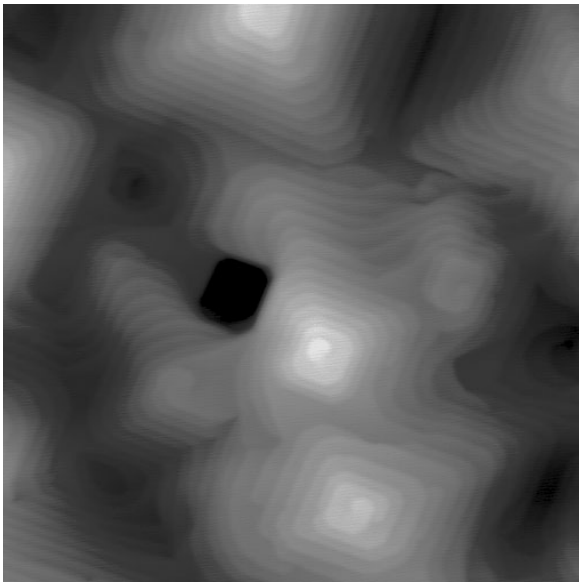


Figure 1.2: An atomic force microscopy image of a $YBa_2Cu_3O_{7-x}$ thin film exhibiting spiral growth. The scanned area is approximately 2μm×2μm. Image courtesy of Jurgen Musolf, Superconductor Technologies.

We would like to have a mathematical model that describes the range of morphologies characteristic of epitaxial growth. When the surface features of interest are much larger than the distance between atoms, a continuum model may suffice. A partial differential equation for the surface height describes morphology evolution in terms of a current of atoms along the surface, according to

$$\frac{\partial h}{\partial t} = F - \nabla \cdot j, \tag{1.1}$$

where $h(\mathbf{x}, t)$ is the height at in-plane location $\mathbf{x}$ and time t, F is the flux of atoms onto the surface, j is the current of atoms along the surface, and $\nabla \cdot j$ is the divergence of the surface current. This approach was developed by Mullins [40] almost fifty years ago and has been adapted more recently to surfaces evolving during deposition [30]. In general, j is a function of h and its spatial derivatives, while F is typically noisy in space and time, and is otherwise spatially uniform. The right-hand side of equation (1.1) often includes the linear terms $\nabla^2 h$ and $\nabla^4 h$, which represent evaporation and surface diffusion, respectively. Additional nonlinear terms may also be incorporated.

When the length scale of surface features of interest approaches the distance between atoms, continuum assumptions are no longer valid. Physical models that describe the interactions between discrete atoms include lattice models, in which atoms are constrained to sites on a rigid lattice; molecular dynamics, in which atomic interactions are described by potential energy wells; and *ab-initio* quantum mechanics, through Schroedinger's equation. The computational demands associated with molecular dynamics and *ab-initio* methods make impractical the simulation of many atoms over the minutes and hours associated with film growth.

The lattice model provides a good compromise between high fidelity and computational tractability, and describes many dynamic features seen in epitaxial surface evolution, like the propagation of atomic-height steps and the nucleation of clusters. It is often manifested through stochastic realizations via a Monte Carlo algorithm, in which atoms may occupy or vacate lattice sites based on the value of a random number. This approach has gained popularity since the 1970's, when it became feasible to compute these realizations for surfaces of physically realistic size [21, 28].

The computational demands of the Monte Carlo simulations are often high, and alternative mathematical representations of the lattice model have been developed. Ordinary differential equations may be formulated to describe cluster formation through a series of rate equations for clusters of various sizes [59], according to

$$\frac{dn_j}{dt} = U_{j-1} - U_j, \tag{1.2}$$

where n_j is the number of clusters with j atoms, and U_j is the rate at which single atoms join clusters of size j to form clusters of size $j+1$. An exact description of U_j requires a solution of the full lattice model, but approximations have been developed in terms of the n_j and fitting parameters called capture numbers. While

this model is compact, it is designed only for the submonolayer regime, before the clusters grow large enough to coalesce.

Level-set methods have been proposed recently to describe the propagation and merger of atomic-height steps [22], including the growth of clusters and their coalescence. The level-set function ϕ is described by a partial differential equation whose level contours represent atomic steps, according to

$$\frac{d\phi}{dt} + \mathbf{v} \cdot \nabla\phi = 0, \tag{1.3}$$

where $\mathbf{v}$ captures the velocity of the level contours. This model provides a compact mathematical framework and captures a wide range of morphology dynamics, but, like the Monte Carlo simulations, its computational demands are high.

1.4 Model reduction

It would be desirable to have a mathematical formulation of the lattice model that captures the range of behaviors seen in Monte Carlo simulations with reduced computational demands. One might also be willing to accept some error in the model, as long as it is not too large, if the computational requirements were greatly diminished. Additionally, if one does not wish to capture the location of every atom, but instead is only interested in a few spatially-averaged metrics, like roughness and step density, then certain aspects of the lattice model might be neglected with no associated error. Such ideas have been developed within control theory. In particular, one may specify the inputs and outputs to the system, and, using a mathematical representation of the system, compute a model of reduced dimension whose map between inputs and outputs is close to that of the original system. Previous work by the author [18, 19] and others [46] has been motivated by this goal, but has not been successful in producing a predictive model useful for optimization or controller design.

Algorithms and error bounds for model reduction of linear time-invariant systems have been developed and are widely applied [39]. These ideas have also been extended to nonlinear systems, although in this case the error bounds do not extend [33]. In recent years much research effort has been directed toward model reduction of complex fluid flows [26]. In this class of systems, one spatially discretizes the partial differential equations that represent the physics to obtain a high-order nonlinear ordinary differential equation. Simulations are then performed, and a small number of dominant spatial modes are extracted from this data using a technique called proper orthogonal decomposition. These modes are projected back onto the original partial differential equation, yielding a low-dimensional nonlinear ordinary differential equation. While this method has been successfully applied to describe low-order behavior in fluid systems, error bounds on the reduction do not exist. In fact, preservation of the stability of the original system cannot be guaranteed, such that the low-order model may be unstable, even when the original system is stable.

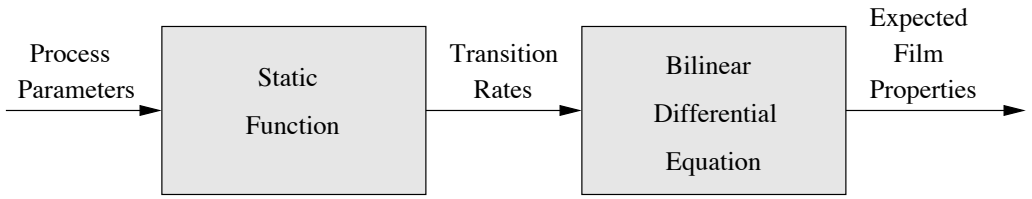


Figure 1.3: Block diagram of the lattice model. The system is divided into a static nonlinear function that maps macroscopic process parameters to microscopic transition rates, and dynamic, bilinear block for the evolution of the probability distribution.

We consider these ideas in the context of the lattice model, and contrast the mathematical structures of the lattice model and the equations for fluid motion. There is a probabilistic differential equation for the lattice model, which has an uncountably-infinite number of states. This is in contrast to fluid systems, which are infinite-dimensional since they are partial differential equations, but which may be made finite-dimensional through discretization using a finite domain and a minimum viscous lengthscale. The number of lattice configurations may also be made finite by assuming a lattice of finite extent, although the number is still generally too large for direct simulation of any physically-realistic system. Consequently, stochastic realizations with a Monte Carlo algorithm are used for simulation instead of numerically integrating the differential equation, so that unlike the fluids example, the state is not directly obtained in simulation. Therefore, it is not straightforward to obtain dominant modes of the state from simulation, and even if we could, we would not be able to write down the full equations on which to project the modes. In the work of this thesis, the Monte Carlo simulations provide input-output data, from which low-order models are constructed, but proper orthogonal decomposition is not used to obtain the states. In contrast, an earlier study is included as Appendix B, in which proper orthogonal decomposition is used to generate dominant spatial modes.

1.5 Thesis overview and contributions

This thesis is about using control and control-oriented modeling to alter thin film evolution. The lattice model, described in Chapter 2, is taken as accurate model of the physics, and is manifested through master equations, kinetic Monte Carlo simulations, and reduced-order models. Viewing the processing parameters as inputs to the system, and expected film properties as outputs, we formulate the lattice model as a control system. A corresponding block diagram is shown in Figure 1.3. The system may be decomposed into two pieces, the first of which is a static nonlinear function that maps the process parameters to microscopic transition rates, and the second of which is a dynamical system describing the evolution of the probability distribution. Because the transition rates multiply the probability vector in the differential equation, the block is referred to as a bilinear

system.

We study the first block in Chapter 3 and use an averaging analysis to show that fast periodic processing parameters may be used to produce new effective transition rates, increasing the space of possible inputs to the second block. This analysis exploits the fact that the transition rates change instantaneously with the processing parameters, while the dynamic block does not. The set of effective transition rates is quantified as the convex hull of the initial set, and may be computed with linear programming techniques.

In Chapter 4, we focus on the second block of Figure 1.3. We develop an approach to extract the dominant dynamics of the kinetic Monte Carlo (KMC) simulations into a finite-dimensional master equation. The method requires only linear computation, and is demonstrated in several solid-on-solid models. This compact representation enables the application of control and optimization techniques, which are impractical for use with the computationally intensive, noisy KMC simulations. We demonstrate the utility and predictive power of our reduced-order models by applying gradient-based optimization and computing optimal temperature profiles.

The final chapter contains a combined experimental and modeling study of a specific material system: germanium grown by molecular beam epitaxy (MBE), using reflection high-energy electron diffraction (RHEED) as an *in situ* diagnostic. We explore the effects of periodically varying the flux. In the range of inputs accessible in our MBE chamber, we do not observe roughening or smoothening due to time-varying process parameters. However, we do extract activation energies for use in KMC simulations. The simulations predict that the instantaneous fluxes of MBE are not high enough to generate a significant effect relative to the mean conditions, but that higher fluxes, as in pulsed laser deposition, could be roughening, while temperature modulation could yield smoother surfaces.

The analysis and modeling approaches developed in this thesis provide a general framework for obtaining compact representations of surface morphology during thin film growth. These models are then demonstrated to capture key features seen in kinetic Monte Carlo simulations, and to be useful for gradient-based optimization. However, the primary goal and contribution of this work is not to propose a particular deposition strategy to minimize surface roughness, but instead to develop general tools that exploit the underlying mathematical structure of the master equation and that are applicable to a wide range of physical systems. The approach developed in this thesis yields reduced-order predictive models that can be used not only in optimization, but are also compatible with stability and bifurcation analysis, and feedback controller design.

Chapter 2

Control-Oriented Formulation of the Lattice Model

This thesis is concerned with the evolution of surface morphology during epitaxial growth. It is known that the evolution is strongly dependent on the process parameters, which we view as inputs that can be intentionally manipulated to influence the surface properties. Dynamic features of particular interest are the propagation of steps and the nucleation and growth of islands, which stem from nearest-neighbor interactions in a distributed system of atoms. We require a model that accurately captures the atomic-scale effects of the process conditions, and that describes film properties of interest on the timescales of film growth. A lattice model for crystal growth, introduced in Chapter 1 and described in detail in this chapter, provides a good balance between these requirements. We thus take the lattice model as our representation of the physics.

2.1 Lattice model

The two key components of the lattice model are the rigid lattice to which atoms are constrained, and the mechanisms and rates of atomic transitions between points on the lattice. The lattice represents the underlying crystal structure of the material, and is assumed to be fixed. The transition mechanisms are defined in terms of lattice configurations, in which unique configurations are distinguished by differences in the occupancy of the lattice sites. Each unique transition mechanism is associated with a set of configuration pairs, in which the first element of each pair may transition into the second element. Three common transition mechanisms—adsorption, desorption, and surface diffusion—are illustrated in Figure 2.1. The dependence on the process conditions enters through the rates of transition associated with each transition mechanism.

The lattice model captures many key features of thin film growth and has been used extensively over the past thirty years as the basis for kinetic Monte Carlo (KMC) simulations [6, 21]. Monte Carlo simulations provide useful predictions of thin film growth, but the rule-based simulations are not conducive to mathematical analysis. For example, the simulations are not invertible; the evolution of a

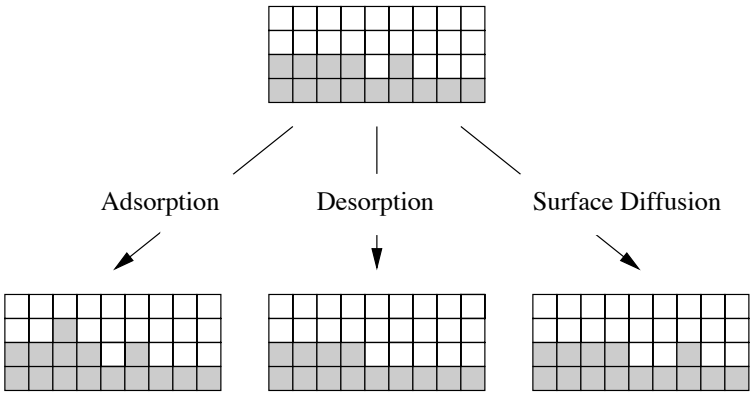


Figure 2.1: Illustration of three common transition mechanisms for a two-dimensional lattice associated with a one-dimensional substrate.

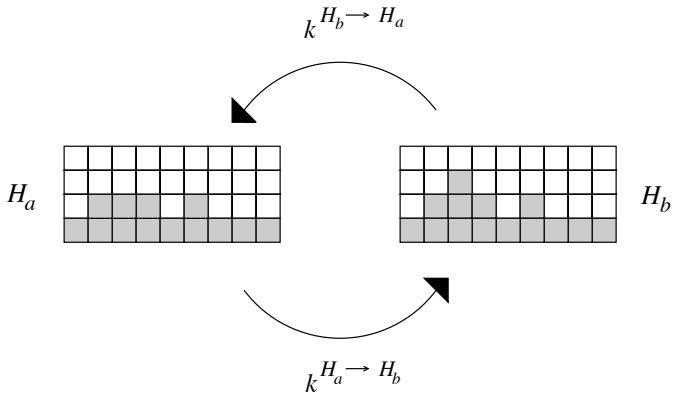


Figure 2.2: Two possible configurations of a two-dimensional lattice associated with a one-dimensional substrate, with corresponding transitions and transition rates.

film grown under particular process conditions can be computed, but the process parameters required to achieve a film with particular film properties cannot.

Each KMC simulation is a stochastic realization of the lattice model, which may also be described by a master equation [15]. The master equation defines the evolution in time of the probabilities of each lattice configuration. Let the symbol H denote a particular configuration, and $k^{H_a \rightarrow H_b}$ the transition rate from H_a to H_b. Figure 2.2 illustrates two of the many possible configurations for a one-dimensional substrate. In the figure, a transition mechanism from H_a to H_b is the adsorption of an atom at the third site from the left. A transition from H_b to H_a may occur through desorption at this site.

The number of configurations grows exponentially with the size of the lattice. For a substrate of 100×100 sites and a maximum height of only 10, the number of configurations is $10^{100 \times 100}$. Although this number can be reduced by physical assumptions and symmetry arguments, it is clearly impossible to directly simulate

the master equation for any realistic domain size. We use the structure of the master equation in our analysis, but continue to perform simulations with the KMC method.

The master equation may be expressed in terms of sums over all the configurations

$$\frac{d}{dt}P_H(t) = \sum_{H'} k^{H'\to H} P_{H'}(t) - \sum_{H'} k^{H\to H'} P_H(t), \tag{2.1}$$

where t is time and $P_H(t)$ is the time-dependent probability of configuration H. The first term on the right-hand side reflects transitions from other configurations H' into configuration H, and the second term represents transitions out of configuration H. Any mean property, such as island density or root-mean-square roughness, may be expressed as a linear combination of the probabilities:

$$\langle Y\rangle(t) = \sum_{H} P_H(t) Y(H), \tag{2.2}$$

where Y is the quantity of interest, $Y(H)$ is the value of Y associated with configuration H, and $\langle Y\rangle(t)$ denotes the time-dependent expected value of Y.

Although there are a large number of configurations and possible transitions between configurations, $k^{H_a\to H_b}$ may only take the m distinct values associated with the m unique transition mechanisms, or a sum of multiple values, or zero if no allowable transition between two configurations exists. The master equation may be recast as a sum over the m transition mechanisms. We use the symbol $k_i^{H\to H'}$ to denote the transition rate associated with a particular transition mechanism, and express the master equation as

$$\frac{d}{dt}P_H(t) = \sum_{i=1}^{m} \left(\sum_{H'} k_i^{H'\to H} P_{H'}(t) - \sum_{H'} k_i^{H\to H'} P_H(t) \right). \tag{2.3}$$

Taking k_i to be the rate of transition through mechanism i, note that $k_i^{H\to H'}$ may take only two values: zero, if no transition from H to H' is allowed via mechanism i, or k_i, if a transition through mechanism i is allowed.

We next rewrite equation (2.3) as a vector equation for the probability vector $\mathbf{x}$:

$$\frac{d}{dt}\mathbf{x} = \sum_{i=1}^{m} \left(k_i N_i^{in}\mathbf{x} - k_i N_i^{out}\mathbf{x} \right), \tag{2.4}$$

where N_i^{in} and N_i^{out} are matrices that represent the allowable transitions either into or out of a configuration through mechanism i. N_i^{in} and N_i^{out} contain mostly zero elements, since a transition mechanism will likely not exist between two randomly selected configurations. The nonzero elements take positive integer values, determined by the number of distinct configurations that a particular configuration can transition into or out of. As the last step we combine N_i^{in} and N_i^{out} into

a single matrix $N_i = N_i^{in} - N_i^{out}$ to arrive at our final form of the master equation

$$\frac{d}{dt}\mathbf{x} = \sum_{i=1}^{m} k_i(\mathbf{u}) N_i \mathbf{x}. \tag{2.5}$$

Equation (2.5) is equivalent to equation (2.1)—only the notation is different. We define $\mathbf{u}$ to be the vector of process parameters and explicitly state the dependence of k_i on our input vector $\mathbf{u}$.

2.2 Kinetic Monte Carlo simulations

A major goal of this thesis is to replace the stochastic kinetic Monte Carlo simulations with a compact differential equation. However, KMC simulations are used throughout the work for the generation of reduced models, and for comparison with the predictions of reduced models. The details of these KMC simulations are described in this section.

2.2.1 Lattice and transitions

In this thesis we focus on the evolution of surface morphology during thin film growth. With this goal in mind, we consider a lattice model that, with a small number of parameters, is able to capture the interplay between the nucleation of clusters, growth of islands, and propagation of atomic-high steps. We use a cubic lattice in which each site has six neighboring sites—one on each side for a total of four, one above, and one below. Lattice sites are not defined by their spatial location, but instead by their connections to other sites. This formulation allows for deformation of the lattice under stress, and in particular enables the incorporation of crystal defects like dislocations. A dislocation is added to the lattice by simply reassigning neighbors along a branch cut.

In the simulations we make the solid-on-solid (SOS) approximation, in which no vacancies in the crystal are permitted; equivalently, every atom must have a neighbor below it. Periodic boundaries are used to simulate an infinite surface, since an actual film is much larger than any simulation domain used in a KMC simulation.

We consider a single-species system, in which every atom in the lattice has the same properties. Allowable transitions include the adsorption of an *adatom*—short for adsorbed atom—onto a surface site, the desorption of an atom from the surface, and the diffusion of an atom along the surface. The transition rate for adsorption is independent of the structure of the particular surface site, while desorption and surface diffusion rates are dependent on the local coordination of the surface atom. In general, adatoms are the most mobile species because they have no side bonds.

The transition rates are strongly dependent on the process parameters. In the simulations, we use the simplest physically realistic models for these rates. For example, we often assume that the adsorption rate k_{ads} is equal to the flux of

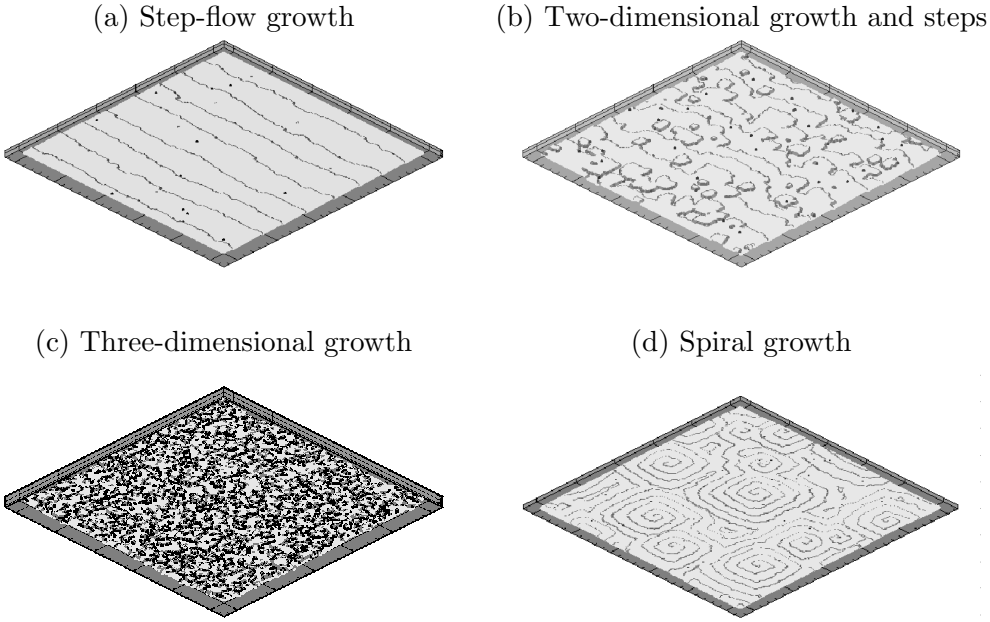


Figure 2.3: Kinetic Monte Carlo simulations of various growth modes.

incident precursors F

$$k_{\text{ads}} = F, \tag{2.6}$$

although it is also possible that only some fraction is actually adsorbed, and furthermore that the fractional value is temperature-dependent. Desorption is based on a bond-counting model

$$k_{\text{des},i} = \nu \exp\left(-\frac{E_{\text{des},0} + i\Delta E}{k_b T}\right), \tag{2.7}$$

where i is the number of occupied neighboring sites, ν is a vibrational frequency, $E_{\text{des},0}$ is the depth of the energy well when the atom has no side neighbors, ΔE is the extra energy for each side neighbor, k_b is Boltzmann's constant, and T is the temperature. Surface diffusion is represented similarly, although the parameter values for the energies and vibrational frequency may be different. Throughout the thesis, the particular transition rates and parameters are noted and described in the context of each KMC simulation

This model captures many dominant features of surface morphology, including those described in the introduction and pictured in Figures 1.1 and 1.2. Four kinetic Monte Carlo simulations we performed are shown in Figure 2.3. In this figure, each surface consists of a domain of at least 200×200 atoms. The dark dots denote atoms with at least one empty side bond, indicating that they are either adatoms or are on the edge of a step. The light dots are atoms with four side bonds, and are part of an atomically flat terrace. The first simulation consists of a surface with steps due to miscut. The adatoms are sufficiently mobile that they

quickly migrate to step edges after deposition. This keeps the adatom density low, preventing nucleation of clusters. In the second simulation, the surface diffusion rate is somewhat less. Some adatoms form clusters, while others attach to steps. In the third image, the mobility is even lower. Many clusters form, and form on top of other clusters, while the steps from miscut play a negligible role. In the final simulation, screw dislocations have been introduced into the initial lattice instead of straight steps. The surface mobility is high enough that adatoms do not form clusters, but instead attach to the steps originating from the core of the dislocation.

In contrast to our simple model, many other physical phenomena may also be described by the lattice model, including a more complex lattice, multi-species chemistry, stress, surface reconstruction, and step edge energy barriers. The KMC simulations of GaAs by Joyce et al. [28] include many of these examples. Extensions of the methods developed in this thesis to these additional effects could be a promising avenue for future work.

2.2.2 Incrementing time

Monte Carlo simulations are often used for the computation of equilibrium properties, in which the evolution in time is not a quantity of interest. The term *kinetic* is added when the system's evolution in time is also captured. In any Monte Carlo simulation, one first specifies an initial configuration, and then determines all of the direct transitions to other configurations. One of these transitions, with its corresponding configuration, is selected based on a random number. Depending on the particular algorithm, the transition rates may be used in selecting the transition, or the transition may be rejected based on these rates. If the transition is selected, then the system is moved to the new configuration. This process of selection and possible execution of transitions is then repeated many times.

We use the kinetic Monte Carlo algorithm described by Fichthorn and Weinberg [15] to capture the correct evolution in time. At each Monte Carlo step, all possible transitions out of the current configuration are enumerated, along with their corresponding transition rates. One of these transitions is then selected using a random number, where the probability of a transition being selected is proportional to its transition rate. This selection criterion is represented mathematically as

$$\frac{\sum_{j=0}^{r-1} k_j}{\sum_{j=0}^{R} k_j} \leq \xi_1 < \frac{\sum_{j=0}^{r} k_j}{\sum_{j=0}^{R} k_j}, \tag{2.8}$$

where event r is selected if the above statement is true. Note that r may take any integer value from 1 to R, and that R is dependent on the current configuration. Each possible transition has an associated transition rate k_j. The event selected depends on the $\{k_j\}$, and on a uniformly distributed random number ξ_1 between zero and one.

In this algorithm, transition r is never rejected. Before actually executing the

transition, a time step is computed according to

$$\Delta t = -\frac{\ln \xi_2}{\sum_{j=0}^{R} k_j}, \tag{2.9}$$

where ξ_2 is a second uniformly distributed random number between zero and one, and Δt is the time interval between the previous transition and the execution of transition r. The transition to the new configuration is then executed, and the entire process is repeated.

The transition rates are dependent on the process parameters, which are functions of time. This is reflected in the KMC simulations only approximately, since we do not know *a priori* what the values of Δt will be for a given simulation, and since the input is effectively held constant over the interval of time between individual events. However, in simulations of large surfaces, Δt is much smaller than the timescale over which the inputs are varied, so the resulting input trajectory for each KMC simulation is a good approximation to the desired continuously varying input.

2.3 Stochastic differential equations

The master equation is one mathematical representation of the lattice model, and is a deterministic differential equation describing the evolution of the probability distribution. Alternatively, a stochastic equation may be formulated for the height at each surface site, whose simulation is equivalent to the KMC method. This equation has been analyzed in the context of the solid-on-sold lattice model [4, 60] in the limit of large system size, in which the surface height becomes a continuous variable.

Conditions under which one may pass from discrete to continuous variables in a master equation have been developed recently by Gillespie [20]. In this work it is noted that when the transition rates vary smoothly in the discrete variable, and furthermore when the discrete variable becomes large, it may be possible to pass to a continuous representation. A notable example used in this work is the connection between the master equation for chemical reactions, and the large-size limit of deterministic chemical kinetics. Because the reaction rate in a chemical reaction is typically a monotonic, polynomial function of the number of species (the discrete variable), in the limit of a large number of species, the integer number of species may be replaced with a continuous variable. However, the solid-on-solid transition rules do not satisfy Gillespie's criterion for passage to continuous time. The discrete variable is the height at each surface sites, while transition rates depend on the local surface coordination. For example, a change in height of one atom may drastically change the transition rate for surface diffusion, while a change in height of 0.5 has no physical meaning.

The work of Vvedensky et al. [4, 60] instead utilizes an expansion based on the system size to obtain passage to the continuous limit. However, they conclude that computation of the resulting equation for a mechanism like surface diffusion re-

quires a treatment of noise correlations that ultimately provides no computational advantage over actually performing Monte Carlo simulations.

We have not identified a use for the stochastic differential equations in understanding and controlling the evolution of surface morphology, and thus focus in the remainder of the thesis on the master equation representation, using Monte Carlo realizations as our simulation algorithm.

Chapter 3

Fast Periodic Inputs

We now analyze the effects of periodic process conditions for the particular case when the process condition period is short relative to the timescales of film growth. In this situation the film is not able to respond fast enough to keep up with changes in the process parameters. Instead, the film evolves as if the transition rates were replaced by constant *effective* transition rates. We employ the method of averaging [62] to compute these effective transition rates. Similar application of the averaging theorem has been applied to mechanical systems with periodic inputs [5].

3.1 Derivation of effective rates

The method of averaging may be applied to a differential equation of the form

$$\frac{d}{dt}\mathbf{z} = \epsilon f(\mathbf{z}, t), \tag{3.1}$$

where $\mathbf{z} \in \mathbb{R}^n$, ϵ is a constant, and f is a function with continuous first and second derivatives. When f is periodic in t with period τ, such that $f(\mathbf{z}, t+\tau) = f(\mathbf{z}, t)$, its average is defined as

$$\bar{f}(\hat{\mathbf{z}}) \equiv \frac{1}{\tau}\int_0^\tau f(\hat{\mathbf{z}}, t)dt, \tag{3.2}$$

and the averaged equation is defined as

$$\frac{d}{dt}\hat{\mathbf{z}} = \epsilon \bar{f}(\hat{\mathbf{z}}), \tag{3.3}$$

with $\hat{\mathbf{z}} \in \mathbb{R}^n$. The averaging theorem relates the difference between $\mathbf{z}$ and $\hat{\mathbf{z}}$ to the size of ϵ. Specifically,

$$|\mathbf{z}(t) - \hat{\mathbf{z}}(t)| = O(\epsilon) \tag{3.4}$$

on a timescale of $O(\frac{C}{\epsilon})$ if $|\mathbf{z}(t_0) - \hat{\mathbf{z}}(t_0)| < O(\epsilon)$ for some initial time t_0. The constant C is independent of ϵ. Refer to Wiggins [62] for further discussion and a proof of the averaging theorem. Notice that when ϵ is small, the averaged equation (3.3) is a good approximation for the original equation (3.1).

The averaging theorem may be applied to the master equation (2.5) when the

process parameters are periodic, i.e., $\mathbf{u} = \mathbf{u}(\omega t)$ with frequency $\omega = \frac{2\pi}{\tau}$. Restating equation (2.5) to emphasize the dependence on time, we obtain

$$\frac{d}{dt}\mathbf{x} = \sum_{i=1}^{m} k_i\left(\mathbf{u}(\omega t)\right) N_i \mathbf{x}. \tag{3.5}$$

Before applying the method of averaging, equation (3.5) must be in the form of equation (3.1) with a small parameter ϵ. With this goal we rescale time by ω. Defining a new time $s \equiv \omega t$ and $\epsilon \equiv \frac{1}{\omega}$, equation (3.5) becomes

$$\frac{d}{ds}\mathbf{x} = \epsilon \sum_{i=1}^{m} k_i\left(\mathbf{u}(s)\right) N_i \mathbf{x}. \tag{3.6}$$

We also define a function $g(\mathbf{x}, s)$

$$g(\mathbf{x}, s) \equiv \sum_{i=1}^{m} k_i\left(\mathbf{u}(s)\right) N_i \mathbf{x} \tag{3.7}$$

and rewrite equation (3.6) as

$$\frac{d}{ds}\mathbf{x} = \epsilon g(\mathbf{x}, s). \tag{3.8}$$

Equation (3.8) is now in the form of equation (3.1), to which the method of averaging may be applied. Note that when the frequency ω is sufficiently high, ϵ is small, and the averaged version of the master equation will be a good approximation to the full equation. We next compute the average of function g, as in equation (3.2), observing that the period in scaled time s is 2π:

$$\begin{aligned}
\bar{g}(\hat{\mathbf{x}}) &= \frac{1}{2\pi}\int_0^{2\pi} g(\hat{\mathbf{x}}, s) ds && (3.9)\\
&= \frac{1}{2\pi}\int_0^{2\pi} \sum_{i=1}^{m} k_i(\mathbf{u}(s)) N_i \hat{\mathbf{x}} ds && (3.10)\\
&= \sum_{i=1}^{m} \left(\frac{1}{2\pi}\int_0^{2\pi} k_i(\mathbf{u}(s)) ds\right) N_i \hat{\mathbf{x}}, && (3.11)
\end{aligned}$$

where $\hat{\mathbf{x}} \in \mathbb{R}^n$ is the averaged state in analogy to $\hat{\mathbf{z}}$ of equation (3.2). Following equation (3.3), the averaged version of equation (3.6) is

$$\frac{d}{ds}\hat{\mathbf{x}} = \epsilon \bar{g}(\hat{\mathbf{x}}) = \epsilon \sum_{i=1}^{m} \left(\frac{1}{2\pi}\int_0^{2\pi} k_i(\mathbf{u}(s)) ds\right) N_i \hat{\mathbf{x}}. \tag{3.12}$$

We may express this more compactly by defining the effective transition rate $k_{\text{eff},i}$

as

$$k_{\text{eff},i} \equiv \frac{1}{\tau}\int_0^\tau k_i(\mathbf{u}(\omega t))dt = \frac{1}{2\pi}\int_0^{2\pi} k_i(\mathbf{u}(s))ds. \tag{3.13}$$

An effective transition rate is simply the average value of the transition rate over a period, and is not a function of time. The averaged version of the master equation in scaled time s is

$$\frac{d}{ds}\hat{\mathbf{x}} = \epsilon\sum_{i=1}^{m} k_{\text{eff},i}N_i\hat{\mathbf{x}}. \tag{3.14}$$

In physical time t the averaged version of equation (3.5) is then

$$\frac{d}{dt}\hat{\mathbf{x}} = \sum_{i=1}^{m} k_{\text{eff},i}N_i\hat{\mathbf{x}}. \tag{3.15}$$

The timescale on which the approximation is valid is $O(\frac{C}{\epsilon})$ in scaled time s but is $O(C)$ in physical time t, independent of the frequency.

The numerical value of the constant C is dependent on the transition rates k_i and the matrices N_i. In the limit of an infinite number of configurations, C may approach zero. However, in the simulations we consider film growth on a finite domain, where the number of configurations is truncated and finite. (This finite domain approximation is valid as long as the surface correlation length is short relative to the domain size.) In this situation C may be small, but is constant and is only a lower bound on the timescale. We explore the timescales of applicability further in the simulations of Section 3.2.

Notice that the averaged master equation (3.15) has the same form as equation (3.5); the transition rates k_i are simply replaced by $k_{\text{eff},i}$. Application of fast periodic process conditions is equivalent to a film growth process with constant transition rates $k_{\text{eff},i}$. If these effective rates are not attainable with constant process conditions, then altered film evolution may be possible. In practice, the process parameters are bounded, so we let $\mathbf{u}_{\min}$ and $\mathbf{u}_{\max}$ be the minimum and maximum values of $\mathbf{u}$, and ask the question: do there exist effective transition rates attainable with periodic process conditions in the range $[\mathbf{u}_{\min}, \mathbf{u}_{\max}]$ that are not attainable with constant parameters in $[\mathbf{u}_{\min}, \mathbf{u}_{\max}]$?

No single transition rate $k_{\text{eff},i}$ can be outside the set of the k_i's attainable with constant parameters, since the effective rate is simply the time average over the instantaneous rate. However, new *combinations* of effective transition rates might be obtained. The ratio of the transition rates of various mechanisms strongly affects the evolution and final properties of a film—for example, the ratio of flux to surface diffusion is a key parameter in the evolution of island density and surface roughness [13, 64].

Model 1	Model 2
$k_1^*(F^*) = F^*$	$k_1^*(F^*) = F^*$
$k_2^*(T^*) = \nu_2^* \exp\left(-\frac{E_2^*}{T^*}\right)$	$k_2^*(T^*) = \nu_2^* \exp\left(-\frac{E_2^*}{T^*}\right)$
$k_3^*(T^*) = \nu_3^* \exp\left(-\frac{E_3^*}{T^*}\right)$	$k_3^*(F^*, T^*) = F^* \nu_3^* \exp\left(-\frac{E_3^*}{T^*}\right)$

Table 3.1: Transition rates for Models 1 and 2. The dimensionless transition rates k_i^* are functions of the dimensionless process parameters F^* and T^*.

3.2 Demonstration of effective rates in Monte Carlo simulations

The effects of periodic modulation of the process parameters are demonstrated through two simple models of film growth. Both models are based on a cubic lattice, have periodic boundary conditions, and disallow vacancies in the crystal. Additionally, the models contain only one type of atomic species. Three types of transitions may occur in each model: (1) adsorption of an atom from the gas onto the surface, (2) diffusion of an atom with no side neighbors along the surface, and (3) loss of an atom with no side neighbors to the gas. Each atom may have at most four side neighbors. An atom with one or more side neighbors may undergo no transition and thus is permanently incorporated into the film.

Two process parameters are considered in each model: the flux F of precursors to the surface and the surface temperature T, so that $\mathbf{u} = \{F, T\}$. We assume that we have complete control over the flux and temperature within preset upper and lower bounds, such that $\mathbf{u}$ may be constant or a periodic function of time. In this study we use dimensionless quantities. The timescale is taken from the maximum flux $F_{\max}$ and the energy scale is $k_b T_{\max}$, where k_b is Boltzmann's constant and $T_{\max}$ is the upper bound on temperature. The lengthscale is set by the lattice spacing a. A dimensionless quantity will be denoted by an asterisk in the remainder of this section—for example, transition rate $k_i^* = k_i F_{\max}$, activation energy $E_j^* = \frac{E_j}{k_b T_{\max}}$, and height $h^* = \frac{h}{a}$.

The transition rates for Models 1 and 2 are given in Table 3.1. The first transition mechanism is adsorption; its rate is equal to the flux F^* in both models (unity sticking coefficient). The second mechanism is surface diffusion. It is a thermally activated process with proportionality constant ν_2^* and activation energy E_2^*, and is again the same in both models. The only difference between Models 1 and 2 is the transition rate for the third mechanism, which results in the removal of an adatom from the surface. In the first model, the third mechanism is a thermally activated process, which is only dependent on the temperature, with a proportionality constant ν_3^* and activation energy E_3^*. However, in Model 2 the third mechanism is an etching process, which is thermally activated as in Model 1

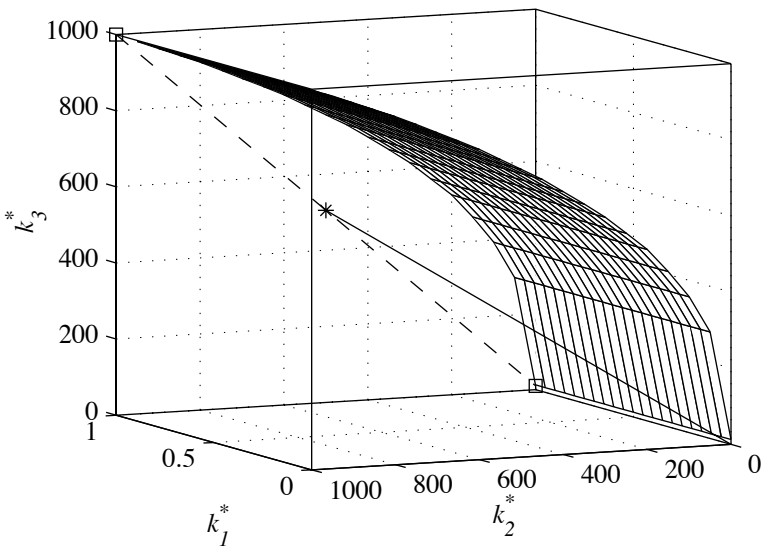


Figure 3.1: The surface of transition rates achievable with constant process parameters for Model 1. The solid line guides the eye to the asterisk, which marks the set of effective transition rates for the input in Figure 3.2. The instantaneous constant transition rates used to generate the set of effective rates are marked with squares.

but is also proportional to the flux. In Model 1, atoms desorb at high temperature due to thermal effects, but in Model 2, high temperature and high flux must coincide to remove atoms from the surface.

First consider the evolution of a film described by Model 1, which we investigate through kinetic Monte Carlo simulations. The physical parameters are determined by selecting the values of the transition rates at the upper and lower bounds of the flux and temperature: $k_1^*(F_{\min}^*, T^*) = 0$, $k_2^*(F^*, T_{\min}^*) = 0.01$, $k_2^*(F^*, T_{\max}^*) = 1000$, $k_3^*(F^*, T_{\min}^*) = 10$, and $k_3^*(F^*, T_{\max}^*) = 1000$. As a final constraint we select $T_{\min}^* = \frac{2}{3}$ to obtain physically realistic activation energies.

Before analyzing periodic flux and temperature, we visualize the set of transition rates attainable with constant flux and temperature. Because there are only three transition mechanisms, we may plot the transition rates against each other and obtain the two-dimensional surface of transition rates. This surface is shown in Figure 3.1. The surface is two-dimensional because there are two process parameters and is bounded because the process parameters have upper and lower bounds.

Any combination of transition rates which is not on the surface of Figure 3.1 cannot be obtained with constant flux and temperature. However, periodic inputs may produce a combination of effective transition rates which is not on the surface. Consider the input pictured in Figure 3.2. The flux is set to the maximum value, but the temperature alternates between its upper and lower bounds. The set of effective rates associated with this periodic input may be computed with equation

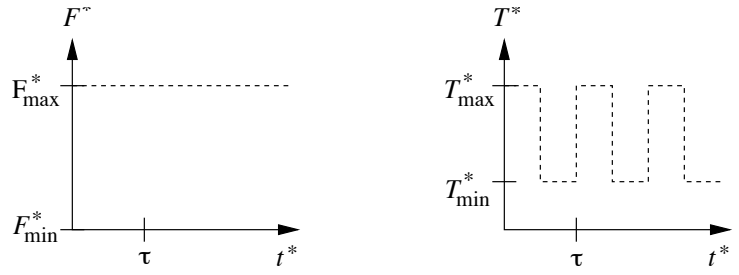


Figure 3.2: A set of periodic process parameters for Model 1.

(3.13), and is denoted by the asterisk in Figure 3.1. These rates are not achievable with constant flux and temperature, and result in a decrease in desorption relative to diffusion and adsorption. An effective transition rate is the time average of the instantaneous rate over a period, and thus for the input in Figure 3.2, the effective rate is the average of the transition rates at $\{F^*_{\max}, T^*_{\max}\}$ and $\{F^*_{\max}, T^*_{\min}\}$. These instantaneous rates are marked by the squares in Figure 3.1. Any point along the dashed line connecting the squares can be achieved by altering the duty cycle of the modulation of Figure 3.2.

Periodic process parameters enable new effective transition rates, which may ultimately result in altered film properties. We use kinetic Monte Carlo simulations to contrast the mean thickness h^* and root-mean-square roughness W^* obtained under constant and periodic inputs. Each simulation is run from $t^* = 0$ to $t^* = 100$, which results in a thickness of up to 100 layers. Simulations are performed for representative values of constant flux and temperature, as well as for the set of effective transition rates considered above. Figure 3.3 displays roughness versus thickness for all of the simulations. Notice that the final combination of thickness and roughness for the effective case, marked by A, could not be achieved by constant inputs in the fixed time interval we considered. The periodic input suppresses desorption relative to diffusion, enabling a thicker, smoother film. To achieve the same thickness in the same amount of time with constant inputs, the roughness must double to nearly 1.5, as marked by B in Figure 3.3; this film is grown under constant parameters with $F^* = 1.0$, $T^* = 0.73$. It is also evident from Figure 3.3 that several constant parameter curves remain near to the effective rate curve, particularly for growth at $F^* = 1.0$ and $T^* = 0.94$, which is marked by C in Figure 3.3. However, in our finite time interval this curve does not extend to point A. Thus, with constant flux and temperature we might be able to produce a film similar to that produced with the periodic inputs, but the growth time would be longer.

We will now analyze Model 2, whose transition rates are shown in Table 3.1. The physical parameters for Model 2 are selected by first setting $k^*_1(F^*_{\min}, T^*) = 0$, $k^*_2(F^*, T^*_{\max}) = 1000$, and $k^*_3(F^*_{\max}, T^*_{\max}) = 10,000$. We also want $k^*_2(F^*, T^*_{\min})$ and $k^*_3(F^*_{\max}, T^*_{\min})$ to be negligible, and consequently set $\nu^*_2 = 10^{13}$, $\nu^*_3 = 10^{12}$, and $T^*_{\min} = \frac{1}{2}$.

Figure 3.4 contains the surface of transition rates achievable with constant

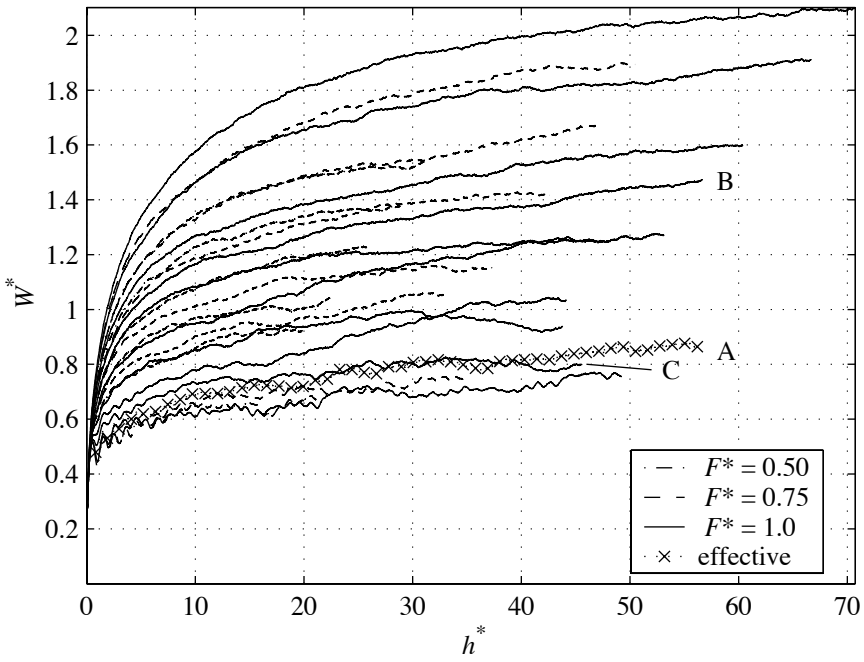


Figure 3.3: KMC simulations of Model 1: Roughness W* versus thickness h* for various constant process parameters and for the set of effective transition rates. The final values of various simulations are marked: (A) effective rate, (B) $F^* = 1.0$, $T^* = 0.73$, (C) $F^* = 1.0$, $T^* = 0.94$.

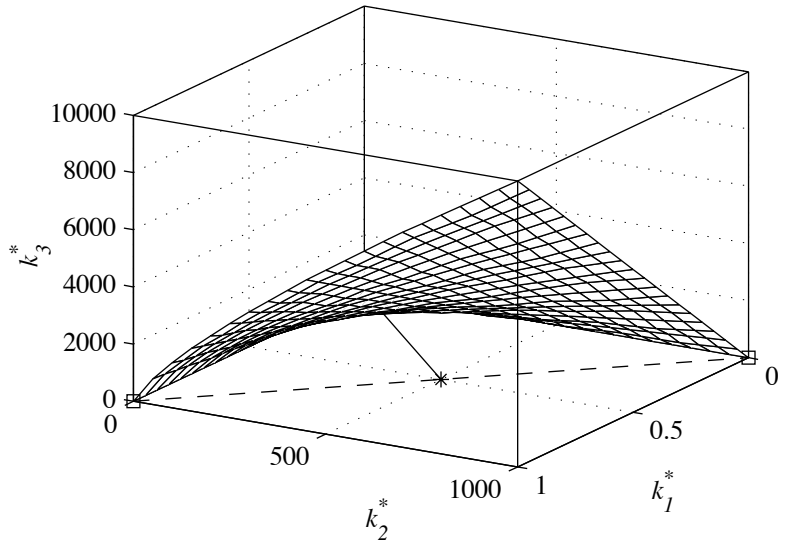


Figure 3.4: The surface of transition rates achievable with constant process parameters for Model 2. The solid line guides the eye to the asterisk, which denotes the set of effective transition rates for the periodic input in Figure 3.5. The instantaneous constant transition rates used to generate the effective rates are marked with squares.

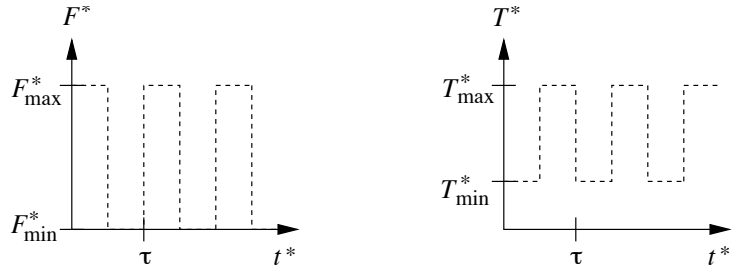


Figure 3.5: A set of periodic process parameters for Model 2.

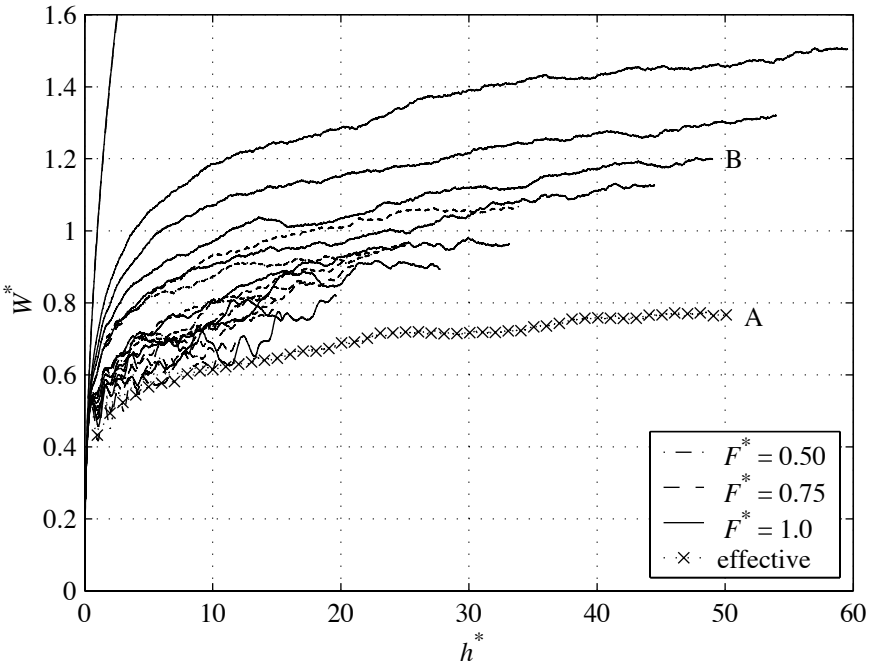


Figure 3.6: KMC simulations of Model 2: Roughness W* versus thickness h* for various constant process parameters and for the effective transition rates. The final values of various simulations are marked: (A) effective rate, (B) $F^* = 1.0$, $T^* = 0.79$.

inputs. Because the desorption rate is a function of both flux and temperature, the shape of the constant input surface is qualitatively quite different from the surface associated with Model 1 (Figure 3.1). We again select a periodic input composed of two constant parameter settings, this time alternating between high flux at low temperature, and low flux at high temperature, as pictured in Figure 3.5. The rates associated with these two constant settings are marked with squares in Figure 3.4; the asterisk denotes the set of effective rates associated with the input in Figure 3.5. Variations in the duty cycle of this input produce other effective rates, which lie along the dashed line in Figure 3.4. The set of effective rates is dramatically different from any combination of rates achievable with constant inputs, which suggests that new film properties may also be obtained.

Kinetic Monte Carlo simulations of Model 2 are shown in Figure 3.6 for a range of constant parameters and for the effective transition rates of Figure 3.5. We again focus on the combinations of roughness and thickness that can be obtained up to time $t^* = 100$. The effective rate curve is qualitatively different from the constant input curves—it lies almost completely outside the region containing the constant input curves. Figure 3.7 shows the final surfaces of a film grown under the set of effective transition rates (point A in Figure 3.6) and a much rougher film of the same final thickness grown under the constant process parameters of $F^* = 1.0$ and $T^* = 0.79$ (point B in Figure 3.6). A significant fraction of surface A is covered by

smooth terraces, while surface B is dominated by step edges. The physical mechanism for the smoothing associated with modulation is simple—under constant process conditions the temperature must be raised to induce smoothing, which results in the loss of atoms due to desorption. The use of periodic parameters enables the suppression of desorption, while still allowing smoothing through diffusion.

We derived effective transition rates associated with fast periodic process conditions, but have not yet addressed what we mean by *fast*. Let us now consider modulated growth at various frequencies and compare to growth at the corresponding set of effective transition rates. We do not yet restrict the modulation to be fast. In the context of the KMC algorithm, these two cases differ in the correlation in time between individual events. As an example we consider modulated growth in which only adsorption occurs in the first half of a period, and only surface diffusion occurs during the second half. When the period is long and many atoms are adsorbed in the first half of the period, the typical time between adsorption events will be shorter than for growth at the constant effective rates, increasing the maximum instantaneous adatom density and island nucleation rate, and potentially altering the overall evolution of the film. However, as the modulation period approaches zero, at most one event will occur during each period, and the distribution of adsorption events in time will approach the distribution of the constant effective transition rates.

In the limit of an infinite number of surface sites, the time between events approaches zero, so it is not practical to modulate faster than the time between individual events. However, we may still be able to replicate the evolution of the constant effective rates using modulated growth of finite period. One such exception occurs when sites are spatially separated beyond the correlation length of the film. We again focus on the example in which only adsorption occurs in the first half of a period, and only surface diffusion occurs during the second half. If two adsorption events occur sequentially at *adjacent* sites, a dimer will be created, potentially leading to the growth of an island. But if instead the two adsorption events occur sequentially, but separated by a large distance, no island will be nucleated. A second exception occurs when a sequence of events may be lumped into a single aggregate event, such as the adsorption of an adatom followed by a series of diffusion events. When such aggregate events dominate the evolution, we can replace the individual events by the aggregate event in the set of allowable transitions, ultimately increasing the interevent time.

The fast limit is explored in simulation for Models 1 and 2 of Section 3.2, using the periodic parameters of Figures 3.2 and 3.5. The modulation period is decreased until the roughness vs. thickness curves approach the curve associated with the constant effective transition rates. These simulations are plotted in Figures 3.8 and 3.9. In both cases the curve for a period of $\tau^* = 0.01$ is near the effective rate curve. Contrast this with the typical interevent time for effective rate growth, which must necessarily be less than the typical time between adsorption events only. Letting N be the number of surface sites, the time between adsorption events is $\frac{F^*}{N}$ or 0.000025 for our 200×200 domain at the maximum flux.

As a final point we stress that the period required for effective rate behavior is

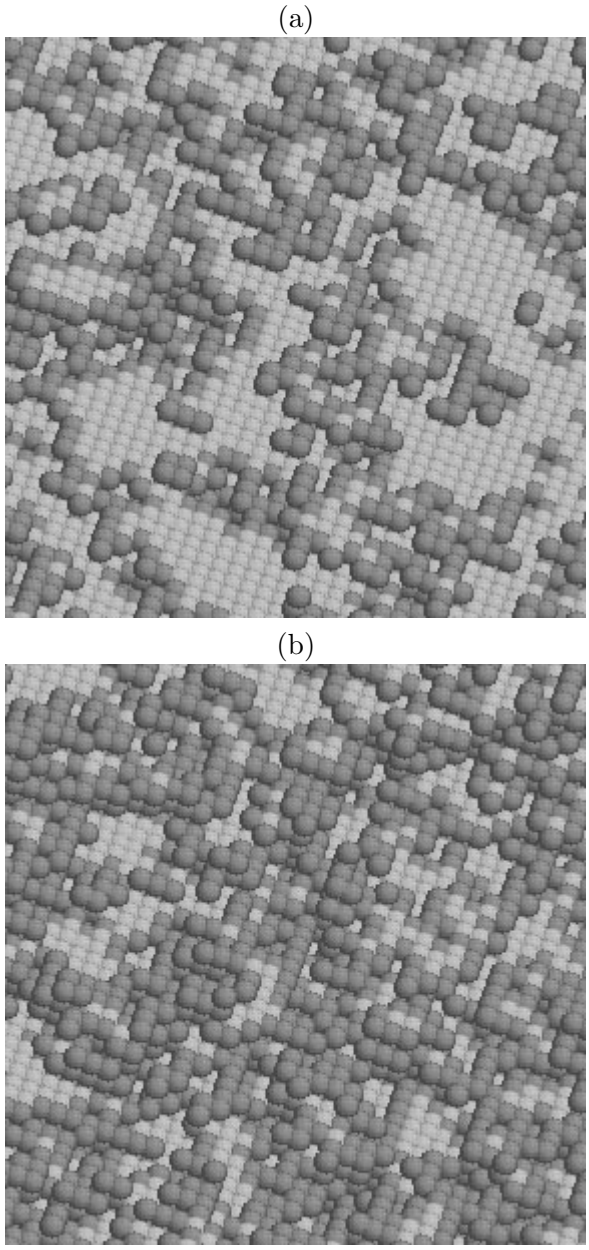


Figure 3.7: KMC simulations of Model 2. Only a portion of the 200×200 domain is shown. (a) Final surface of a film grown at the effective transition rates, marked by A in Figure 3.6. (b) Final surface of a film grown at the constant parameters $F^* = 1.0$, $T^* = 0.79$, marked by B in Figure 3.6. Each atom with four side neighbors is light-colored—atoms with at least one empty side bond are dark.

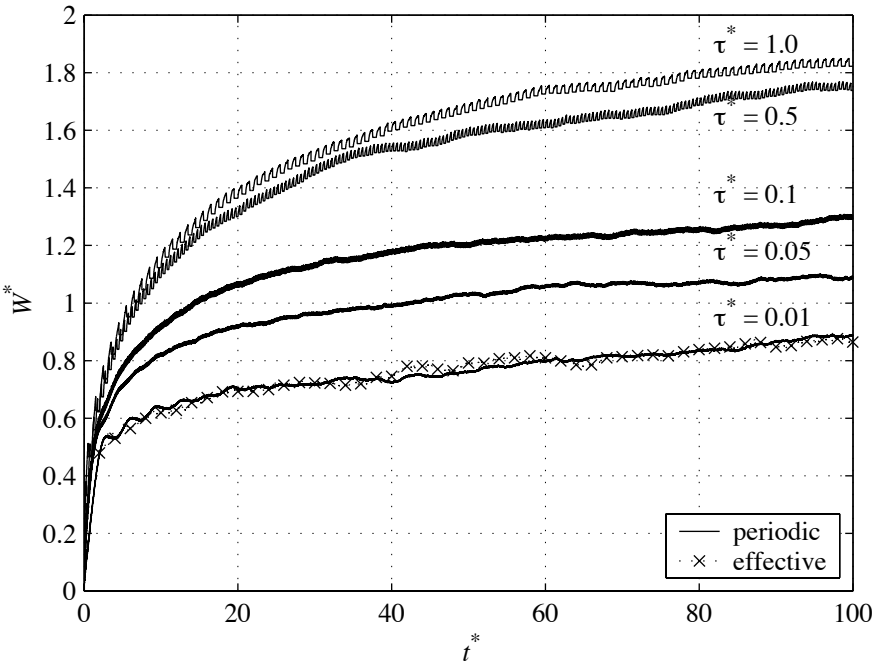


Figure 3.8: KMC simulations of Model 1 on a 200×200 domain. A variety of modulation periods are simulated, as well as growth at the set of constant effective transition rates.

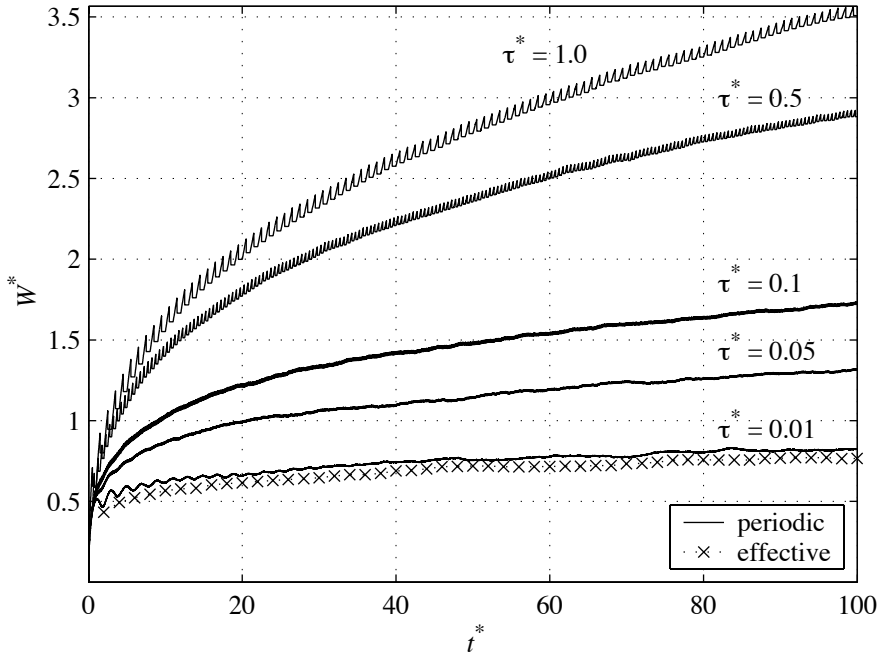


Figure 3.9: KMC simulations of Model 2 on a 200×200 domain. A variety of modulation periods are simulated, as well as growth at the set of constant effective transition rates.

highly dependent on the transition mechanisms and rates, and thus the feasibility of process parameter modulation must be assessed on an individual basis. However, throughout the last decade various film growth processes have been developed to deposit films under periodic process conditions. To modulate the flux of species to the surface, two primary methods have been employed: (1) the use of valves to switch flow between the chamber and a vent line [27], and (2) sequential exposure to flux as the substrate rotates through different environments [44]. Methods for temperature modulation include resistive heating [48], laser irradiation [57], and supplemental cooling with water and liquid nitrogen [14].

3.3 Convex hull interpretation of effective rates

Having shown by example that periodic process parameters can produce film properties unattainable with constant conditions, we will now precisely define the set of all effective transition rates in terms of the set of all possible instantaneous transition rates. Recall that in the examples effective rates were constructed that were on lines connecting instantaneous transition rates. This was accomplished by modulating between two settings of the process parameters. By alternating between three or more settings, additional effective transition rates might be constructed; and, in a more general setting, any continuous-time function of the process parameters could be approximated with a large number of constant segments.

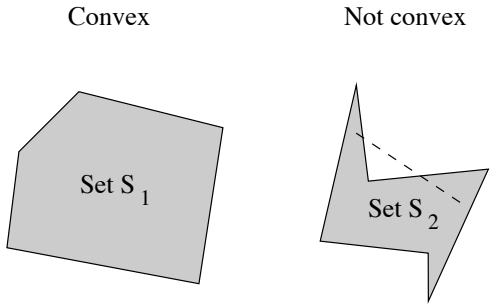


Figure 3.10: Examples of convex and nonconvex sets. The set on the right is not convex, because a line segment connecting two points in the set is not in the set.

We can express the effective transition rate as a sum over the constant segments

$$k_{\text{eff},i} = \frac{1}{\tau}\sum_{j=1}^{r} k_i(\mathbf{u}_j)(\alpha_j \tau) = \sum_{j=1}^{r} \alpha_j k_i(\mathbf{u}_j), \tag{3.16}$$

where τ is the period, α_j is the fraction of the period spent at process condition $\mathbf{u}_j$, and r is the number of different process settings per period. By definition, $\sum_{j=1}^{r} \alpha_j = 1$.

The relationship between k_i and $k_{\text{eff},i}$ may be expressed more precisely in terms of convex sets [49]. A set, or collection of points, is a convex set if for every line segment connecting two points in the set, the entire line segment is also in the set. This definition is consistent with everyday use of the word "convex." Consider the sets shown in Figure 3.10. Set S_1 is convex, while Set S_2 is not, because the dashed line connecting points in S_2 is not contained in S_2.

Clearly, not all sets are convex. However, beginning with a nonconvex set S one can create a convex set, $\text{con}(S)$, called the convex hull of S. This set can be constructed iteratively by connecting every two points in S with a line segment, and then including all of the line segments in a second set along with all the points in the original set. Every two points in the second set are again connected with line segments and are included, along with the second set, in a third set. This procedure is repeated until no new points are added. When this is achieved, the set $\text{con}(S)$ is obtained, which is by definition a convex set.

The concept of making a set convex is illustrated by Figure 3.11. The sets on the left are not convex, and their convex hulls are shown in the right column. The convex hull of a set may be the same dimension as the original set, or it may be of greater dimension.

The procedure described above involves taking weighted averages of two elements of S, and adding those points to the set. Another way to express the convex hull of a set S is in terms of three or more points in S. When the number of elements of S taken at a time is not limited, only a single step (no iterations) is

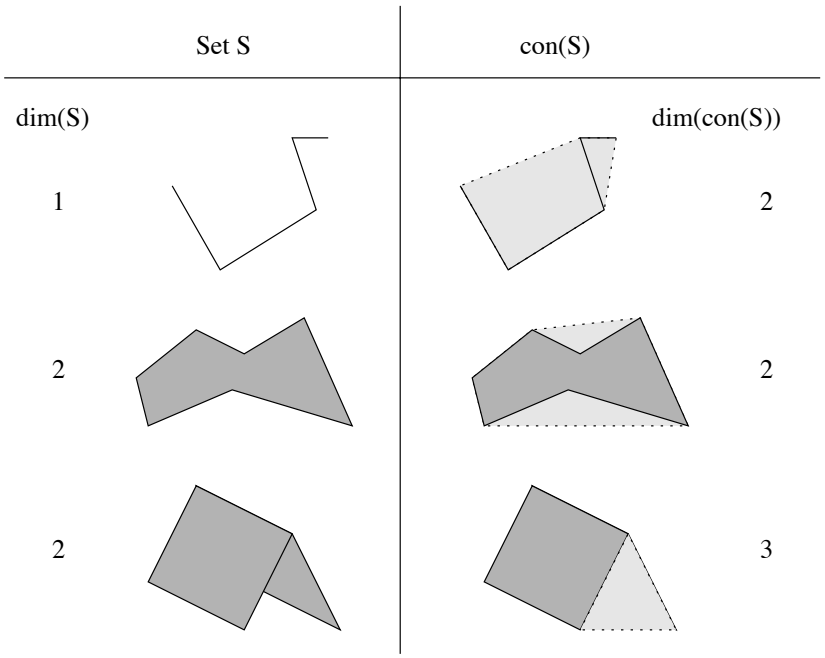


Figure 3.11: Examples of nonconvex sets and their convex hulls. The dimensions of these sets are also shown. The dark grey refers to the original set, which is always part of the convex hull, while the light grey is the part of the convex hull that is not in the original set.

required to construct the convex hull, so that x is in $\text{con}(S)$ if and only if

$$x = \sum_{j=1}^{r} \alpha_j s^j, \tag{3.17}$$

where s^j are points in set S, α_j are positive constants, and $\sum_{j=1}^{r} \alpha_j = 1$. Now compare equation (3.17) to equation (3.16). If we define $\{\mathbf{k}\}$ to be the set of all possible instantaneous transition rates, where each $\mathbf{k}$ is the m-dimensional vector of the k_i's, and define $\{\mathbf{k}_{\text{eff}}\}$ equivalently to be the set of all effective transition rates, then

$$\{\mathbf{k}_{\text{eff}}\} = \text{con}\left(\{\mathbf{k}\}\right). \tag{3.18}$$

That is, the set that includes all possible effective transition rates that are achievable through periodic process parameters is the convex hull of the set of instantaneous transition rates. This result precisely defines the set of all possible effective transition rates, given the dependence of the transition rates on the inputs, and the upper and lower bounds on the inputs.

The convex set description of the effective transition rates is useful because a number of theorems have been proven. We use one such theorem in the next section to construct simple periodic process parameters for any achievable effective transition rate.

3.4 Linear programming solution for periodic inputs

Computing the effective rates associated with a particular input requires only the evaluation of equation (3.13), but the inverse problem is not as straightforward: given a set of desired effective transition rates, compute the periodic process parameters to produce it, if they exist. Carathéodory's theorem, a theorem in convex analysis, provides an answer [49]. Specifically, the theorem states that the maximum number of points required to generate any possible weighted average is $n+1$, where n is the dimension of the space containing the original set. In our case the dimension of the space of rates is m, the number of transition mechanisms, so any achievable set of effective transition rates can be obtained by periodic process parameters which are composed of $m+1$ constant segments. In the examples in Section 3.2, three transition mechanisms are considered, so to achieve any possible set of effective rates, no more than four segments are needed. We considered process parameters with only two segments and obtained effective rates unattainable under constant conditions.

The proof of Carathéodory's theorem [49] suggests a two-step method to construct a desired effective rate as the weighted average of $m+1$ constant-parameter rates. First, a weighted average of a large number of constant-parameter rates is computed which is equal to the desired effective transition rates. The number of rates included in the average is then reduced to $m+1$. If the desired effective rates cannot be expressed as a weighted average of constant-parameter rates, no solution to the first step will be found. The computation requires discretization of

the surface of constant-input rates, after which the fraction of the period spent at each point is computed as a linear program [31]. Matlab code for this problem is included as Appendix A.

Chapter 4

Model Reduction and System Identification

The infinite dimension of the solid-on-solid master equation complicates both simulation and analysis of its dynamics. Stochastic simulations using Monte Carlo techniques can provide good predictions of physical properties of interest, but an equivalent solution to the analysis problem has not been identified. In this chapter, we examine several approaches to reduce the dimension of the system by reducing the number of configurations. The resulting reduced-order model is still a master equation, whose expected properties approximate the expected properties associated with the original system. A system identification technique is then developed, enabling the coefficients of either the original or the reduced master equation to be extracted from simulation data using a single linear computation. The chapter concludes with the introduction of an alternate form of the lattice model, in which the number of adsorption events executed replaces time as the dependent variable. This model advances in discrete steps and may provide a better low-order representation for typical behavior seen in Monte Carlo simulations. An analogous discrete identification process is developed, again using only linear computations.

The example systems explored in this chapter consist of one-dimensional surfaces with a small number of sites. They are chosen to capture important features associated with surface evolution, while remaining computationally tractable for the analysis developed in this chapter. The study of larger, more realistic domains is reserved for Chapter 5.

4.1 Configuration reduction in the lattice model

If the initial state of the system is unconstrained, such that any configuration may initially be occupied, then it is not possible to eliminate any configurations from the master equation. However, if the set of initial configurations is restricted to some subset, then the number of configurations in a master equation may be larger than the minimum number needed to capture the evolution of the system. For example, any configuration that is initially occupied with zero probability will continue to be occupied with zero probability if there exists no path of transitions from an initially occupied configuration. In other words, the set of configurations may be partitioned into subsets that are connected via transition mechanisms, and

if no configuration in any subset is initially occupied, than no configuration in the subset will ever be occupied. All configurations in the subset may therefore be eliminated from the master equation.

Another situation where the number of configurations is not minimal is when there exist redundant configurations or sets of configurations. In this case multiple configurations may always evolve in the exactly the same way, and may therefore be grouped into a single configuration. The initial state will again influence this grouping. For example, two configurations that are identically connected to the remaining configurations will not evolve identically if their initial probabilities differ.

In the preceding two examples, it is possible to reduce the number of configurations in a master equation while incurring no error in the evolution of expected properties. However, it may be that these criteria are only approximately satisfied, and that a reduced-order master equation may be obtained that approximately predicts the evolution of expected properties of the original system. Such ideas have been developed within the economics community for probabilistic discrete-time Markov chains [34, 55], which may be viewed as a discrete-time analog of the continuous-time master equation. In particular, it was recognized by Lange [34] that if states always evolve in a fixed ratio, they may be grouped into a single state via a linear coordinate transformation. Simon and Ando [55] later considered systems that can be decomposed into weakly coupled subsystems. Transitions within each subsystem are fast, while transitions between subsystems are slow. When this is true, the short-time behavior may be estimated by ignoring any coupling between subsystems, while at long times the subsystems may be assumed to be internally at equilibrium (and therefore with states at fixed ratios), so the dynamics are approximated by transitions between subsystems. Thus, in the short-time limit, the system dimension is reduced to the number of states in a subsystem, while at long times the system dimension is reduced to the number of subsystems. Ando and Simon [55] also note that states may evolve in fixed ratios either because they are strongly coupled, or because they are similarly coupled to the rest of the system. In either case, one might consider the coordinate transformation of Lange [34].

We view the reduction of master equation configurations as primarily an issue of timescales, and first define a time, t_f, as the maximum time over which we are interested in the evolution. There may be transitions that rarely occur over this timescale, but which over much longer times would be important, as described in the preceding paragraph. We may thus choose to eliminate these transitions from the master equation, and, by eliminating the transitions, remove the possibility of ever transitioning into certain configurations. These configurations may then be removed from the master equation.

There may also be a minimum timescale over which we wish to resolve the evolution of the master equation. For example, in the solid-on-solid master equation adatoms may hop from site to site at a timescale much faster than the time in which islands nucleate and steps propagate. One may thus choose a time step $\Delta t < t_f$ that is small enough to capture the evolution of morphological features,

but not of individual atomic transitions. In the following development, we assume for convenience that $n_t \equiv t_f/\Delta t$ is an integer. To guarantee that extreme oscillations are not occurring between the time steps, one should examine the eigenvalues of the master equation to ensure sufficient damping of modes with period smaller than Δt. However, once Δt has been established, the configuration set may be reduced by identifying configurations that achieve their equilibrium ratios with each other in times less than Δt. These configurations may be then grouped into a single configuration via a coordinate transformation, in which one new coordinate is the sum of the probabilities over the group, while others represent differences between the coordinates, which are exactly zero in the fixed-ratio limit.

We quantify the approximation introduced by the configuration reduction by considering the map between a probability distribution at time $j\Delta t$ and at the next time step $(j+1)\Delta t$, where $j = 0, 1, ...n_t$. The input vector $\mathbf{u}$ is assume to be fixed between the time steps, with constant value $\mathbf{u}_j$ in the j^{th} interval, although one could also use fast periodic inputs to obtain constant effective transition rates over the interval. A discrete-time version of the master equation is thus obtained:

$$\mathbf{x}_d[j+1] = A_{\mathbf{u}_j}\mathbf{x}_d[j], \tag{4.1}$$

defining

$$A_{\mathbf{u}_j} \equiv \exp\left(\Delta t \sum_{i=1}^{m} k_i(\mathbf{u}_j)N_i\right), \tag{4.2}$$

and with matrix $A_{\mathbf{u}_j} \in \mathbb{R}^{n\times n}$, and discrete-time probability vector $\mathbf{x}_d[j] \in \mathbb{R}^n$ representing the probability distribution at $t = j\Delta t$.

Next consider a map associated with some approximation to the original system, for example due to the removal of a transition. This matrix is referred to as $\tilde{A}_{\mathbf{u}_j} \in \mathbb{R}^{n\times n}$, and is associated with another probability vector $\tilde{\mathbf{x}}_d[j] \in \mathbb{R}^n$, to obtain an approximating system

$$\tilde{\mathbf{x}}_d[j+1] = \tilde{A}_{\mathbf{u}_j}\tilde{\mathbf{x}}_d[j]. \tag{4.3}$$

We wish to quantify the error associated with the approximation, and compare the probability distributions using the one-norm, $\|\cdot\|_1$. This norm is the sum of the absolute value of the elements for a vector, and the maximum column sum of the absolute value for a matrix. It is particularly useful for probabilistic systems, since the one-norm of a probability vector is one, and the one-norm of discrete-time

stochastic matrices like $A_{\mathbf{u}_j}$ is also one. Comparing the two systems, we obtain

$$\begin{aligned}
\|\mathbf{x}_d[j+1] - \tilde{\mathbf{x}}_d[j+1]\|_1 &= \|A_{\mathbf{u}_j}\mathbf{x}_d[j] - \tilde{A}_{\mathbf{u}_j}\tilde{\mathbf{x}}_d[j]\|_1 && (4.4)\\
&= \|A_{\mathbf{u}_j}\mathbf{x}_d[j] - \tilde{A}_{\mathbf{u}_j}\mathbf{x}_d[j] + && (4.5)\\
&\quad \tilde{A}_{\mathbf{u}_j}\mathbf{x}_d[j] - \tilde{A}_{\mathbf{u}_j}\tilde{\mathbf{x}}_d[j]\|_1 && (4.6)\\
&\leq \|A_{\mathbf{u}_j}\mathbf{x}_d[j] - \tilde{A}_{\mathbf{u}_j}\mathbf{x}_d[j]\|_1 + && (4.7)\\
&\quad \|\tilde{A}_{\mathbf{u}_j}\mathbf{x}_d[j] - \tilde{A}_{\mathbf{u}_j}\tilde{\mathbf{x}}_d[j]\|_1 && (4.8)\\
&\leq \|A_{\mathbf{u}_j} - \tilde{A}_{\mathbf{u}_j}\|_1\|\mathbf{x}_d[j]\|_1 + && (4.9)\\
&\quad \|\tilde{A}_{\mathbf{u}_j}\|_1\|\mathbf{x}_d[j] - \tilde{\mathbf{x}}_d[j]\|_1. && (4.10)
\end{aligned}$$

Using the fact that $\|\mathbf{x}_d[j]\|_1 = 1$ and $\|\tilde{A}_{\mathbf{u}_j}\|_1 = 1$, defining $\epsilon_{\mathbf{u}_j} \equiv \|A_{\mathbf{u}_j} - \tilde{A}_{\mathbf{u}_j}\|_1$, and assuming that $\mathbf{x}_d[0] = \tilde{\mathbf{x}}_d[0]$, we obtain a recursive equation for the bound

$$\|\mathbf{x}_d[j+1] - \tilde{\mathbf{x}}_d[j+1]\|_1 \leq \epsilon_{\mathbf{u}_j} + \|\mathbf{x}_d[j] - \tilde{\mathbf{x}}_d[j]\|_1 \tag{4.11}$$

at time $(j+1)\Delta t$. Thus, with each additional time step, the maximum additional error incurred is $\epsilon_{\mathbf{u}_j}$. Because we are only considering a maximum time of t_f, we may impose a maximum acceptable value on $\epsilon_{\mathbf{u}_j}$ to ensure that the error after $t_f/\Delta t$ time steps is sufficiently small.

We now explore physical situations in the solid-on-solid lattice model in which an approximate equation, associated with $\tilde{A}_{\mathbf{u}_j}$, exists such that $\epsilon_{\mathbf{u}_j}$ is small for the range of $\mathbf{u}_j$ that is of interest, and such that the approximate version admits a configuration reduction.

4.1.1 Weak coupling and configuration elimination

Let us first consider the weak-coupling limit [55], in which the states may be arranged such that

$$A_{\mathbf{u}_j} = A^*_{\mathbf{u}_j} + \Delta A_{\mathbf{u}_j}, \tag{4.12}$$

where $A^*_{\mathbf{u}_j} \in \mathbb{R}^{n\times n}$ and $\Delta A_{\mathbf{u}_j} \in \mathbb{R}^{n\times n}$. We define $\epsilon_{\mathbf{u}_j} \equiv \|\Delta A_{\mathbf{u}_j}\|_1$. Furthermore,

$$A^*_{\mathbf{u}_j} = \begin{bmatrix} (A^*_{\mathbf{u}_j})_1 & & & \\ & (A^*_{\mathbf{u}_j})_2 & & \\ & & \ddots & \\ & & & (A^*_{\mathbf{u}_j})_N \end{bmatrix} \tag{4.13}$$

is a block-diagonal matrix, with N decoupled subsystems described by the $(A^*_{\mathbf{u}_j})_r, r = 1, ...N$. The matrix $A^*_{\mathbf{u}_j}$ is a stochastic matrix, such that all elements are positive and the columns sum to 1, as is (consequently) each block $(A^*_{\mathbf{u}_j})_r$. The columns of $\Delta A_{\mathbf{u}_j}$ thus sum to 0. We consider the coupling between the subsystems to be small if $\epsilon_{\mathbf{u}_j}$ is small, and may then use the block diagonal matrix as our approximate system, such that $\tilde{A}_{\mathbf{u}_j} \equiv A^*_{\mathbf{u}_j}$. Although the system described by $\tilde{A}_{\mathbf{u}_j}$ is not of

reduced dimension, its dimension may be reduced without error by eliminating all subsystems that are not represented in the initial condition.

We now consider the idea of decoupling in a solid-on-solid lattice model with four sites on a one-dimensional surface. The boundary conditions are periodic. Ideally, one would list all possible configurations, and then apply the reduction ideas previously described, but even for a small surface, the number of configurations is infinite—one associates with each surface site an integer height, which may take a value from 0 to ∞. However, for finite times and with some basic physical assumptions, one may eliminate from consideration certain configurations that are highly improbable. For example, one might select a finite maximum and minimum height, thus making the number of configurations finite.

Instead of considering all configurations and then eliminating some, we select an initial configuration, and then add configurations to which the system is likely to transition. We refer to this process as enumeration of configurations. The set of configurations identified during the enumeration will depend both on the initial configuration and the transition mechanisms. This process is performed iteratively. One first defines a set of possible initial configurations, and then applies the microscopic transition rules to determine all configurations linked to the initial set by one transition. These configurations are added to the configuration set. This process is repeated until no new configurations are added.

When adsorption is one of the transition mechanisms, configurations will continue to be added at each iteration, and some additional criterion is required to close the enumeration process. One could consider a finite time, and therefore a finite number of adsorption events. Alternatively, one may reject individual transitions that, while not impossible, are highly unlikely. We wish to distinguish transitions based on their importance during the enumeration process, and only add configurations to the configuration set if they contribute significantly to the dynamics via $A_{\mathbf{u}}$. Configurations are not rejected explicitly during the enumeration process, but instead transition paths are rejected, which may subsequently reduce the number of configurations added at each iteration.

We use as our criterion the level of coupling between configurations, and consider, in this example, configurations with the structure illustrated by Figure 4.1 to be weakly coupled. If there is a separation of transition rates such that the rates of transitions entering configuration B are much less than any single transition leaving it, and such that the rate leaving is much greater than $1/\Delta t$, then at the end of each time step, the system will be found in configuration B with very small probability. We may then approximate the connectivity graph as shown in Figure 4.1, by effectively assuming that the transition rate leaving B is infinite. Configuration B is thus completely decoupled from the other configurations in the approximate system.

As a brief aside, we contrast this assumption with a singular perturbation assumption, in which configuration B would be assumed to be at its equilibrium value [32, 45]. While this might also produce a good estimate of the system, the resulting model would be nonlinear in the inputs and state matrices, and would therefore not be a master equation. We will exploit the structure of the master

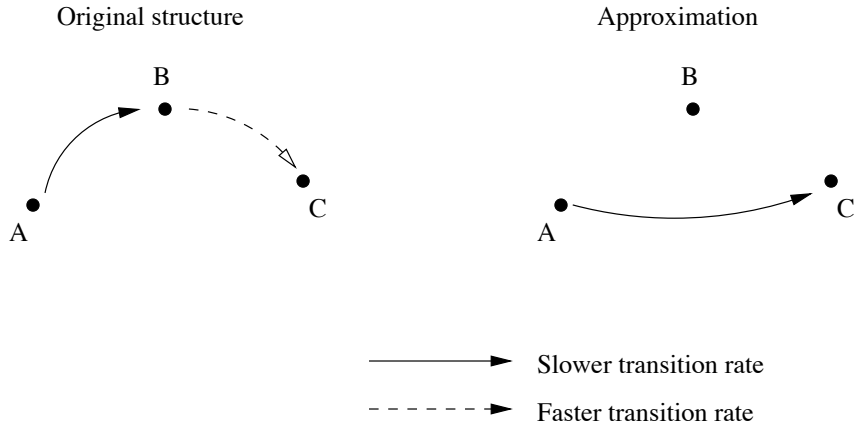


Figure 4.1: One example of a weakly coupled configuration. Transitions from A to B are much slower than transitions from B to C. When the minimum timescale of interest is much greater than that of the fast transition, the system may be approximated by the structure on the right, in which B is completely decoupled from A and C.

equation in the work that follows and therefore do not further explore a singular perturbation approach.

At each iteration in the enumeration process, we formulate a master equation and corresponding $A_{\mathbf{u}}$ for a preselected Δt and for the range of $\mathbf{u}$ of interest. Consider now a time step $\Delta t = 0.01\ s$ and particular inputs such that $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{dif},0} = 10^2, 10^3$, and $10^4\ s^{-1}$. The enumeration process is illustrated by Figure 4.2, which we initiate with a flat configuration, labeled as configuration A. In the first iteration, only one new configuration is identified, which is labeled as configuration B. We do not distinguish here between translations and reflections of configurations, since they all evolve identically in this example. However, the multiplicities are accounted for in constructing the master equations. Configuration B is accessed through an adsorption event, while no new configurations are found through adatom diffusion. We then apply the weak-coupling criterion to a master equation based on configurations A and B, using the microscopic transition rules for adsorption and diffusion. For $k_{\text{dif},0} = 10^3\ s^{-1}$,

$$A_{\mathbf{u}(k_{\text{dif},0}=10^3\ s^{-1})} = \begin{bmatrix} 0.961 & 0 \\ 0.039 & 1 \end{bmatrix}. \tag{4.14}$$

The associated master equation includes only an adsorption transition from A to B, and is thus independent of the diffusion rate. We now search $A_{\mathbf{u}}$ for decoupled, or nearly decoupled, subsystems. We could consider the off-diagonal element of 0.039 to be small, resulting in a decoupled system with no transitions. However, we wish to consider growth over multiple layers, which, with $\Delta t = 0.01$ s, implies hundreds of time steps. Thus, an error of 0.039 added at each time step would not be small. We therefore conclude that configuration B should be accepted, and

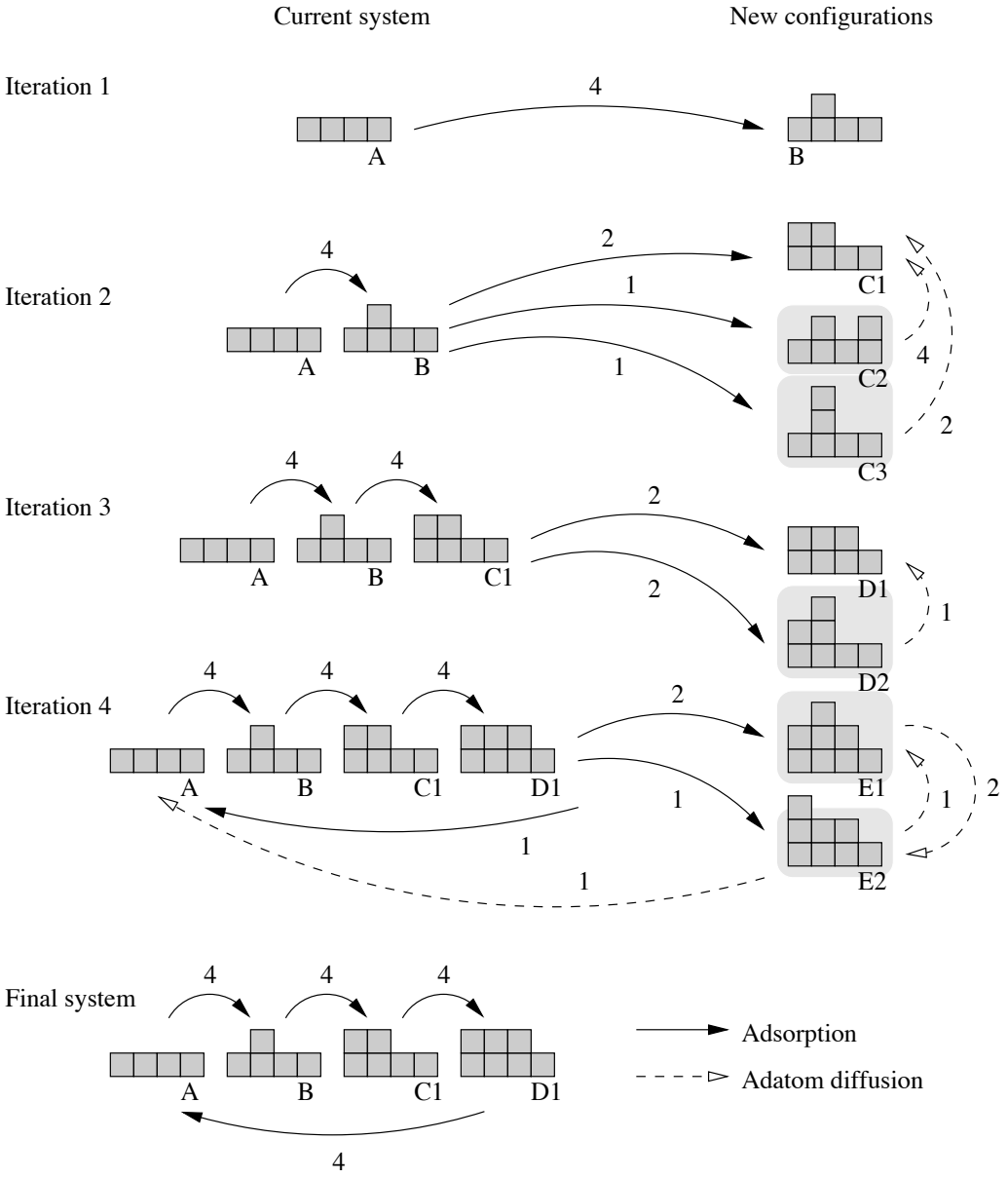


Figure 4.2: Enumeration of configurations for the four-site adsorption-diffusion example. At each iteration, the current configuration set is expanded by executing each possible microscopic transition on each configuration in the initial set. These configurations may then be rejected if they fail the test pictured by Figure 4.1. Configurations that fail the test are surrounded by the gray boxes, while configurations that pass are included in the next iteration. The process ceases when no new configurations are added in an iteration. The integers are the coefficients in the state matrices of the master equation.

Iteration	1	2	3	4
$\epsilon_{k_{\text{dif},0}=10^2\ s^{-1}}$	0	1.3×10^{-2}	2.5×10^{-2}	4.7×10^{-2}
$\epsilon_{k_{\text{dif},0}=10^3\ s^{-1}}$	0	1.4×10^{-3}	3.9×10^{-3}	1.1×10^{-2}
$\epsilon_{k_{\text{dif},0}=10^4\ s^{-1}}$	0	1.4×10^{-4}	3.8×10^{-4}	1.1×10^{-3}

Table 4.1: Error associated with the approximating map at each iteration in the enumeration process.

move on to a second iteration.

In the second iteration, three new configurations are identified through adsorption transitions from configuration B. No new configurations are added via diffusion, although diffusion transitions are possible between the new C configurations, as seen in Figure 4.2. We again formulate the master equation, this time for a configuration set of A, B, C1, C2, and C3, yielding

$$A_{\mathbf{u}(k_{\text{dif},0}=10^3\ s^{-1})} = \begin{bmatrix} 0.961 & 0 & 0 & 0 & 0 \\ 0.038 & 0.961 & 0 & 0 & 0 \\ 0.001 & 0.038 & 1 & 1 & 1 \\ 0.000 & 0.000 & 0 & 0.000 & 0 \\ 0.000 & 0.000 & 0 & 0 & 0.000 \end{bmatrix} \tag{4.15}$$

for $k_{\text{dif},0} = 10^3\ s^{-1}$. We observe that the fourth and fifth configurations (C2 and C3), are nearly zero, since every element in the fourth and fifth rows is small. However, C2 and C3 are not completely decoupled from the system, since their values will contribute to the probability of C1 at the next time step. The physical situation is not that there are only slow transitions into C2 and C3, but that configurations C2 and C3 quickly transition into C1, since diffusion is much faster than adsorption. Configurations C2 and C3 are shaded in gray in Figure 4.2, indicating that they would be eliminated according to the criterion of Figure 4.1. We thus construct an approximating master equation in which transitions into C2 and C3 are redirected into C1. This is equivalent to assuming infinite transition rates for these events. The resulting approximating map for $k_{\text{dif},0} = 10^3\ s^{-1}$ is

$$\tilde{A}_{\mathbf{u}(k_{\text{dif},0}=10^3\ s^{-1})} = \begin{bmatrix} 0.961 & 0 & 0 & 0 & 0 \\ 0.038 & 0.961 & 0 & 0 & 0 \\ 0.001 & 0.039 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0.000 & 0 \\ 0 & 0 & 0 & 0 & 0.000 \end{bmatrix}. \tag{4.16}$$

The difference between the exact and approximating maps, as measured by the the previously defined $\epsilon_{\mathbf{u}}$, is given in Table 4.1, along with the ϵ values for the other diffusion rates, and for subsequent iterations in the enumeration process. The values of ϵ are the maximum one-norm error that can be added during one time step for a particular set of inputs. As the diffusion rate increases, the assumption that transitions from C2 and C3 to C1 are infinitely fast becomes a better

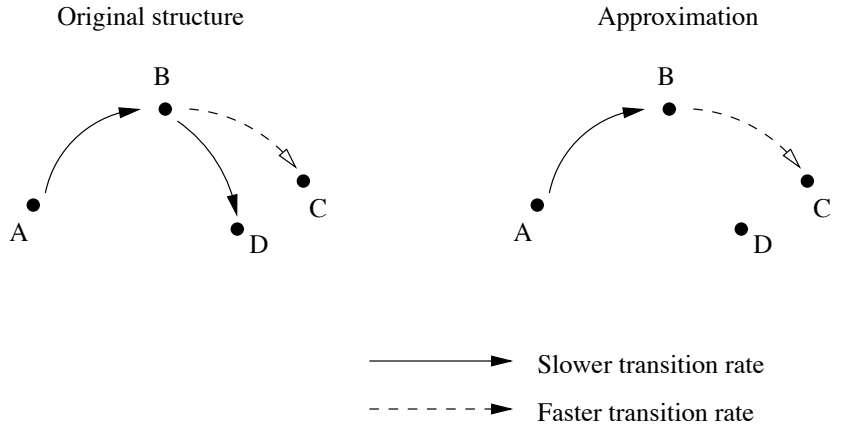


Figure 4.3: Another example of a weakly coupled configuration. If configuration B is rarely occupied, due to fast transitions out of it, then transitions from B to D through a slower transition mechanism are unlikely.

assumption. If one chooses to make this type of assumption at each iteration, then the enumeration process closes with the four-configuration system in Figure 4.2.

Note that the values ϵ in Table 4.1 do not represent an overall error in the evolution of the final system, as compared to the original infinite-configuration system. We simply use the values of ϵ as guidelines for selecting important configurations, by assessing at each iteration the difference between a system based on the current configurations, in relation to a system in which the new configuration is included.

The final four-configuration system of Figure 4.2 represents the limiting behavior of high diffusion, in which the surface evolves in compact, two-dimensional islands, returning to a flat configuration at the end of each monolayer of growth.

We also mention two other simple models, and their limiting cases. If we had considered desorption instead of surface diffusion, in the limit of high desorption, the limiting case would have been only the single flat surface, since each adatom would desorb before another adsorbed. Alternatively, if the surface had contained a step due to miscut, and the surface evolved under adsorption and either surface diffusion or desorption, the limiting case would be the propagation of the step, with configurations corresponding to various locations of the step.

We would ultimately like to capture behaviors more complex than these limiting cases, and now consider a different criterion in the enumeration process, as illustrated in Figure 4.3. In contrast to Figure 4.1, configuration B is now kept in the system, even though transitions leaving it are much faster than those coming in. However, configuration B has a small probability (as measured by the elements of its corresponding row of $A_{\mathbf{u}}$), so transitions out of B through the slower mechanism are quite rare. The enumeration process, using this criterion, is shown in Figure 4.4. The first two iterations are not shown, but neither result in the rejection of any transition or configuration. However, configurations C2 and C3 are deemed to be unlikely, since they may only be entered by adsorption, but quickly transition

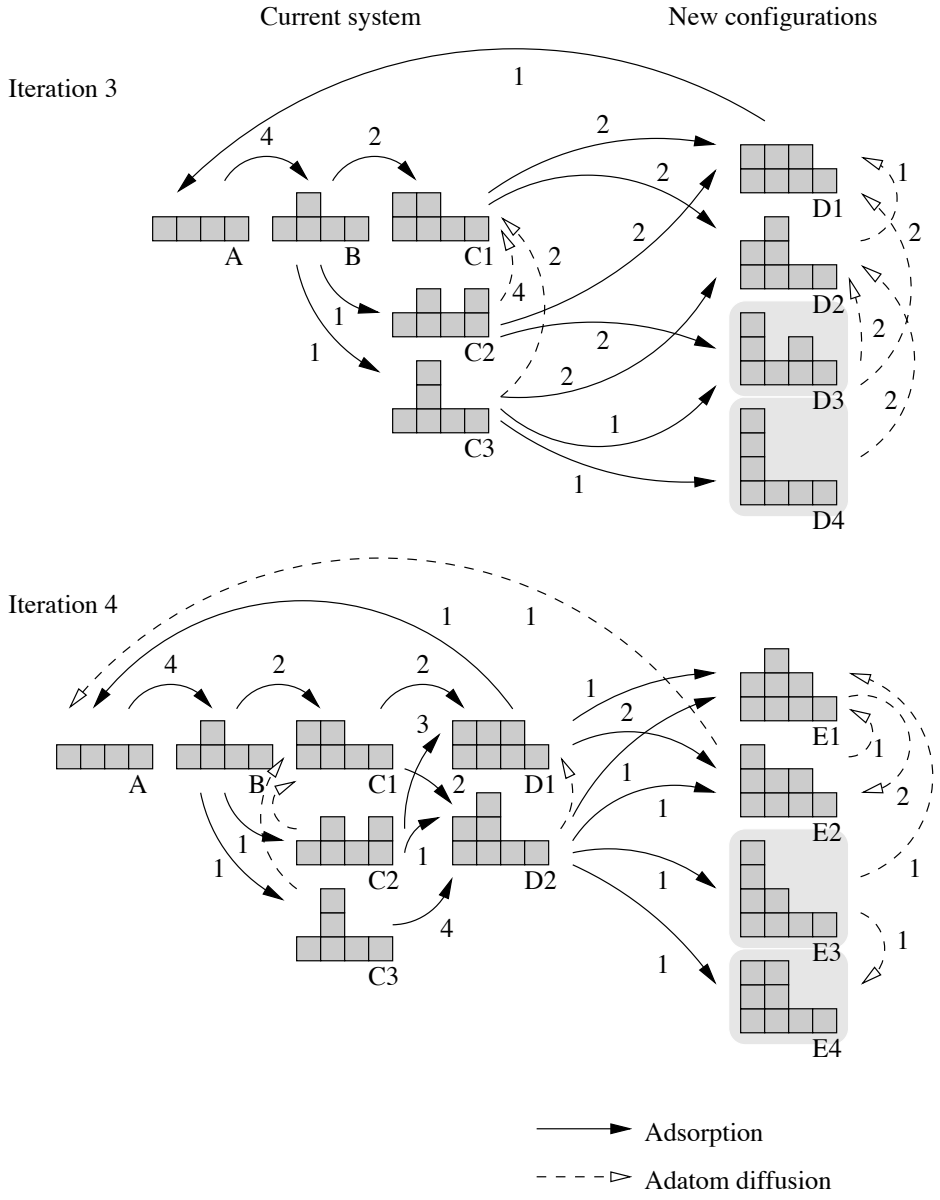


Figure 4.4: Enumeration of configurations for the four-site adsorption-diffusion example, with the rejection criterion pictured in Figure 4.3. During the first two iterations, no configuration is rejected, but the shaded boxes denote configurations rejected in iterations 3 and 4. The integers are the coefficients in the state matrices of the master equation.

Iteration	1	2	3	4
$\epsilon_{k_{\text{dif},0}=10^2\ s^{-1}}$	0	0	3.8×10^{-3}	2.1×10^{-2}
$\epsilon_{k_{\text{dif},0}=10^3\ s^{-1}}$	0	0	9.6×10^{-7}	3.0×10^{-3}
$\epsilon_{k_{\text{dif},0}=10^4\ s^{-1}}$	0	0	9.6×10^{-9}	3.0×10^{-4}

Table 4.2: Error associated with the approximating map at each iteration in the enumeration process of Figure 4.4.

out through surface diffusion. Thus, new configurations in the third iteration that are only accessed through adsorption from C2 and C3 are examined for decoupling, as reflected by ϵ in Table 4.2. Note that the values of ϵ in Table 4.2 are less than the values for Table 4.1 for the same inputs and transition rates. The criterion of Figure 4.3 is less severe than that of Figure 4.1, and ultimately leads to a closure of the enumeration process with the nine-configuration system shown in Figure 4.5. While this system still only captures the smooth two-dimensional growth mode, it is able to capture the dependence of adatom density on the diffusion rate.

4.1.2 Grouping of configurations

After a configuration set has been determined through enumeration, one might wish to further reduce the number of configurations. We consider the system of Figure 4.5, and, instead of looking for configurations that we may decouple from the system, we instead look for configurations that evolve in a fixed ratio, either because they are tightly coupled to each other and come to equilibrium with each other at a timescale less than Δt, or because they are similarly coupled to the rest of the system. Examination of Figure 4.5 suggests that E1 and E2 might be grouped, since they are closely coupled via fast surface diffusion. C2 and C3 are not closely coupled to each other, but might be expected to evolve similarly, since they are both accessed when half a monolayer has been deposited, and since they are both accessed via adsorption, and transition out through surface diffusion. We consider both (E1, E2) and (C2, C3) as candidates for grouping.

We estimate the ratio in which the configuration pairs evolve by examination of the corresponding rows of A. For the highest values of surface diffusion, we observe a ratio of C2/C1 = 2, and E3/E2 = 1.5, while at lower diffusion the ratios are less constant. The appropriateness of these groupings is quantified through ϵ. We formulate an approximating master equation in which C1 and C2, and E2 and E3, are force to evolve in the fixed ratio associated with high surface diffusion. This is the coordinate transformation of Lange [34], which incidentally is equivalent to projecting the original equation onto only the new coordinates that represent probabilities (not differences of probabilities). The values of $\epsilon_{\mathbf{u}}$ for three values of the diffusion rate are given in Table 4.3. Note that in this case, because we started with a finite configuration set, $\epsilon_{\mathbf{u}}$ now represents the error associated with the reduction. This new system now has seven configurations, not nine, since two pairs have been grouped. At the lowest diffusion rate considered, the error associated with the grouping could be greater than one after hundreds of time

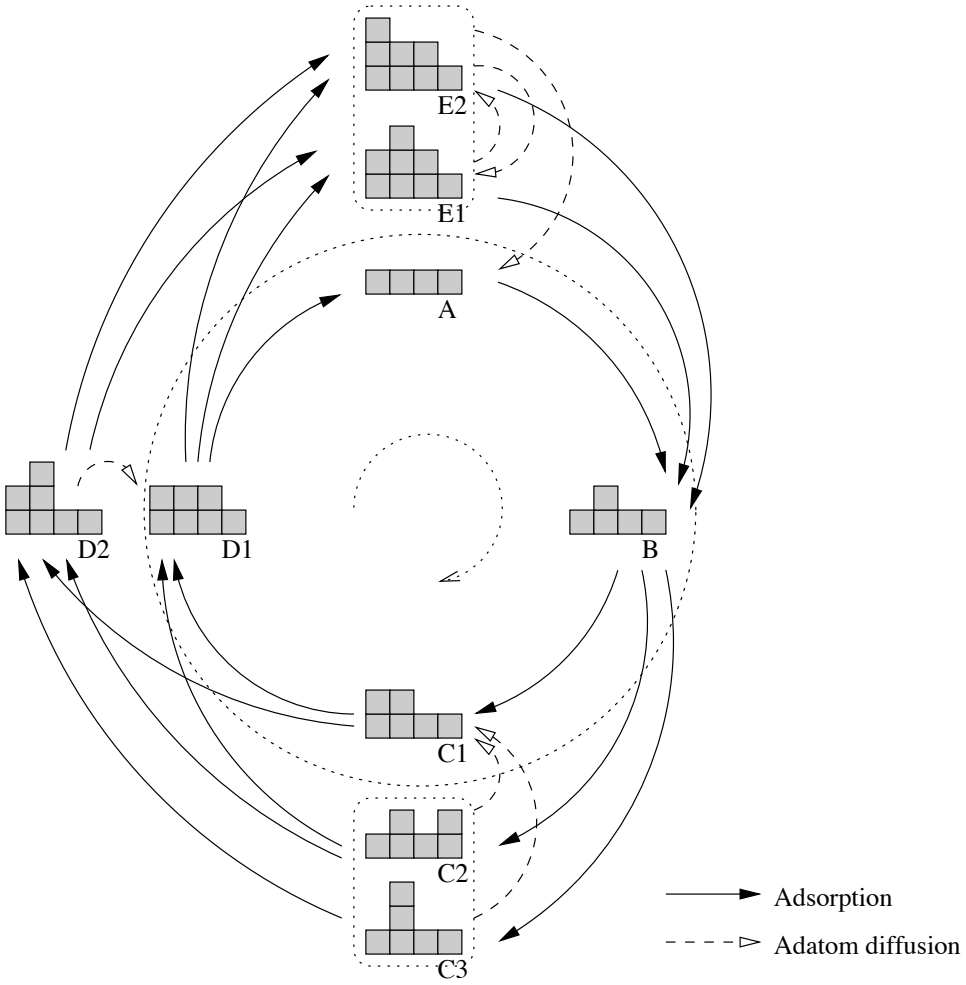


Figure 4.5: Configurations and connectivities for the four-site master equation under adsorption and adatom surface diffusion, as determined by the enumeration process pictured by Figure 4.4. The configurations grouped within the boxes may evolve in a fixed ratio, and are subsequently grouped into a single configuration.

$\epsilon_{k_{\mathrm{dif},0}=10^2\ s^{-1}}$	5.0×10^{-2}
$\epsilon_{k_{\mathrm{dif},0}=10^3\ s^{-1}}$	8.0×10^{-5}
$\epsilon_{k_{\mathrm{dif},0}=10^4\ s^{-1}}$	5.9×10^{-9}

Table 4.3: Error associated with grouping configurations.

steps, although the error will be small at the higher rates. Also realize that the inputs can be changed at each time step, so the diffusion rate could potentially be lower briefly to $10^2\ s^{-1}$, as long as it is not maintained there.

4.1.3 Example: six-site lattice model

We now examine a slightly more complex system: a six-site lattice with adatom adsorption and surface diffusion, in the limit of $k_{\text{dif},0} \gg k_{\text{ads}}$. Kinetic Monte Carlo simulations of this system are shown in Figure 4.6, for $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{ads}} = 1 \times 10^4\ s^{-1}$. Three measures of surface properties are plotted: fractional monolayer coverage, root-mean-square roughness, and adatom density. The mean of 1000 KMC realizations is shown, along with two individual realizations. Note that although the roughness and adatom density reach a minimum when the first layer is complete, the mean shows little oscillation due to variations in the time at which each layer is completed.

The number of possible configurations in the KMC simulations is infinite, since there is no restriction on the maximum height. We begin the enumeration procedure for this system, with the goal of determining a finite approximating configuration set. The enumeration process begins with a flat surface and follows according to the criterion of Figure 4.3, in similar fashion to the four-site example of Figure 4.4. The enumeration closes with the nineteen configurations of Figure 4.7.

A master equation is constructed by applying the microscopic transition rules for adsorption and surface diffusion to the configurations of Figure 4.7, and to all translations and reflections of these configurations. The state dimension is then reduced back to nineteen by observing that all configurations differing by only a translation or reflection are equally probable, and by applying an appropriate coordinate transformation. The resulting nineteen-state master equation is then numerically integrated for transition rates of $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{ads}} = 1 \times 10^4\ s^{-1}$. The expected surface properties of coverage, roughness, and adatom density are plotted in Figure 4.8, and are compared to the mean properties of the KMC simulations shown previously in Figure 4.6.

The error associated with the reduction to a finite, nineteen-configuration set is small. We now further reduce the number of configurations by grouping configurations that evolve together. In particular, we group the configurations in the dashed boxes of Figure 4.7. These groupings denote configurations that have the same coverage and differ only by the location of the adatom. We suggest based on physical arguments that when the probability of one such configuration increases, so will the others, both because they are coupled to each other and are similarly coupled to the rest of the configurations. The evolution of configurations H1, H2, and H3 is plotted in Figure 4.9, again with $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{ads}} = 1 \times 10^4\ s^{-1}$ and an initially flat surface. We group the configurations according to the coordinate transformation proposed earlier, in which one new coordinate for each group is the sum of the probabilities for each, while the other coordinates represent differences between the configurations. We take the mean ratio of the configurations from simulations like Figure 4.9, perform the coordinate transformation, and then

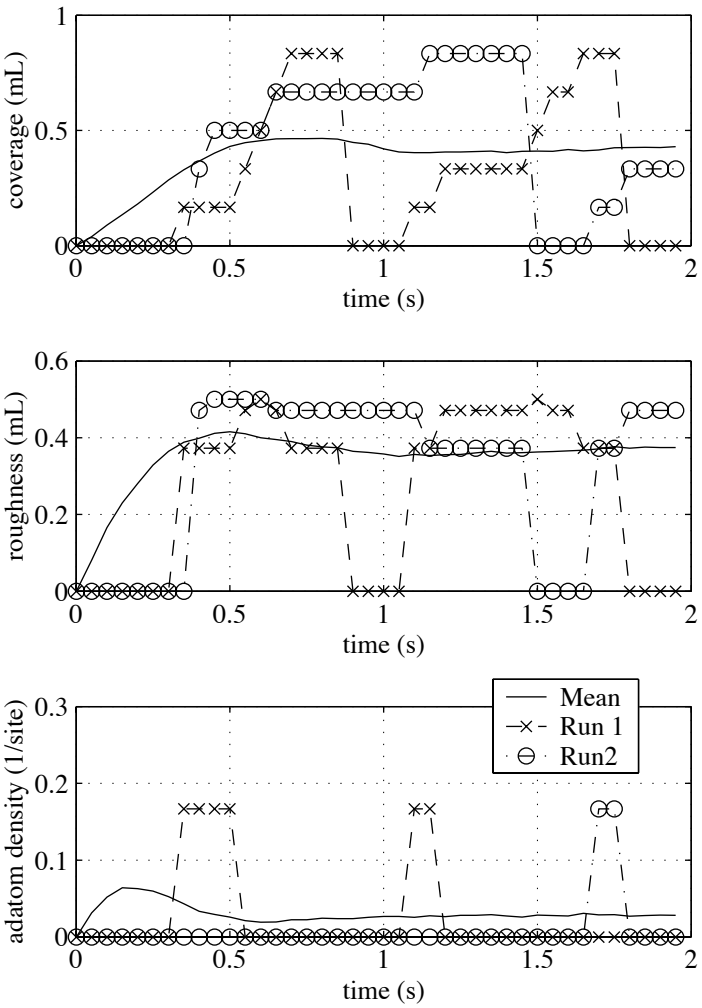


Figure 4.6: Kinetic Monte Carlo simulations of a six-site one-dimensional lattice with $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{ads}} = 1 \times 10^4\ s^{-1}$. Surface measures of monolayer coverage, root-mean-square roughness, and adatom density are shown for two individual realizations and for the mean over 1000 realizations.

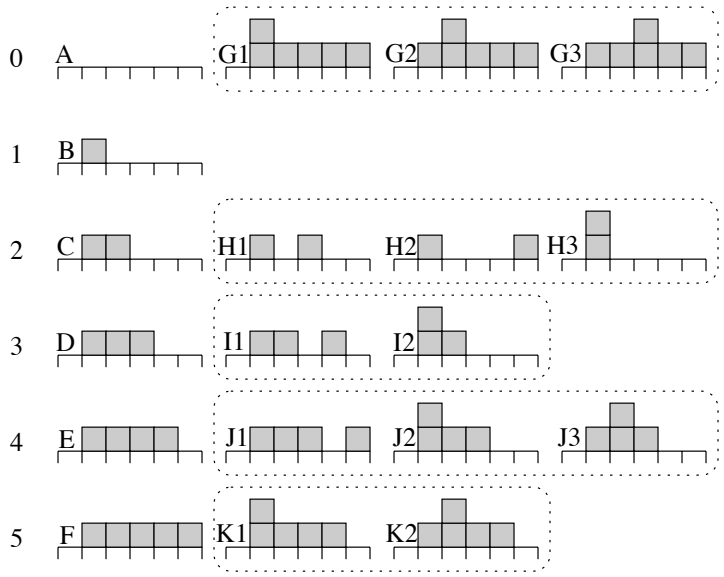


Figure 4.7: Nineteen configurations for a six-site solid-on-solid master equation, in the limit of high surface diffusion. Configuration grouped within the dashed boxes evolve similarly, and are grouped into a single configuration.

truncate the coordinates corresponding to differences in probabilities.

The remaining eleven configuration are pictured in Figure 4.10. This figure illustrates not only the configurations, but also their connectivities through adsorption and adatom diffusion. The source of the oscillatory behavior in the KMC simulations stems from the cyclic paths generated by adsorption. This eleven-state model provides a good approximation to the original nineteen-state system, as shown in Figure 4.11.

4.2 System identification for master equations

In the previous examples, the state matrices and the output matrix were constructed based in microscopic transition rules, using an approximating configuration set. We now develop a method to determine the coefficients in a master equation using simulation data. This idea is then extended to identify coefficients in a reduced master equation. Once the reduced configuration set has been determined, the system identification procedure provides an alternative to directly computing reduced state matrices, e.g., via a coordinate transformation.

The computation of the potentially large number of coefficients in the master equation is aided by the linearity of the system. Because we are using data to compute the coefficients, we convert the master equation to discrete time with a

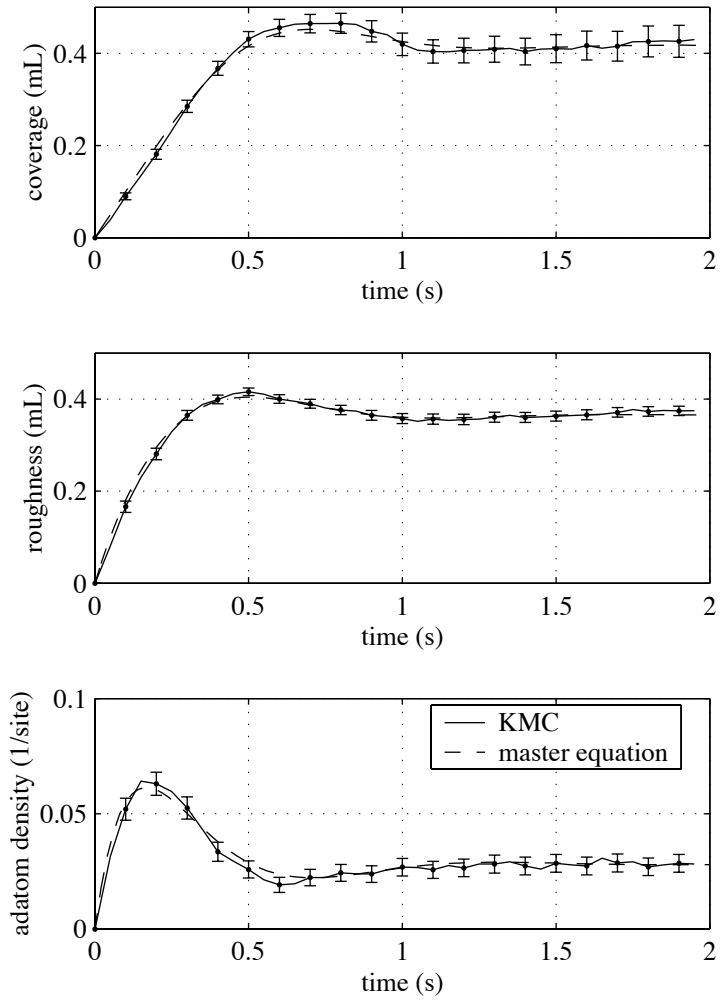


Figure 4.8: Comparison between the mean over 1000 KMC simulations, and the expected value predicted by the nineteen-state master equation. In both cases, $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{ads}} = 1 \times 10^4\ s^{-1}$. Error bars denote a 95% confidence interval on the mean of the KMC simulations.

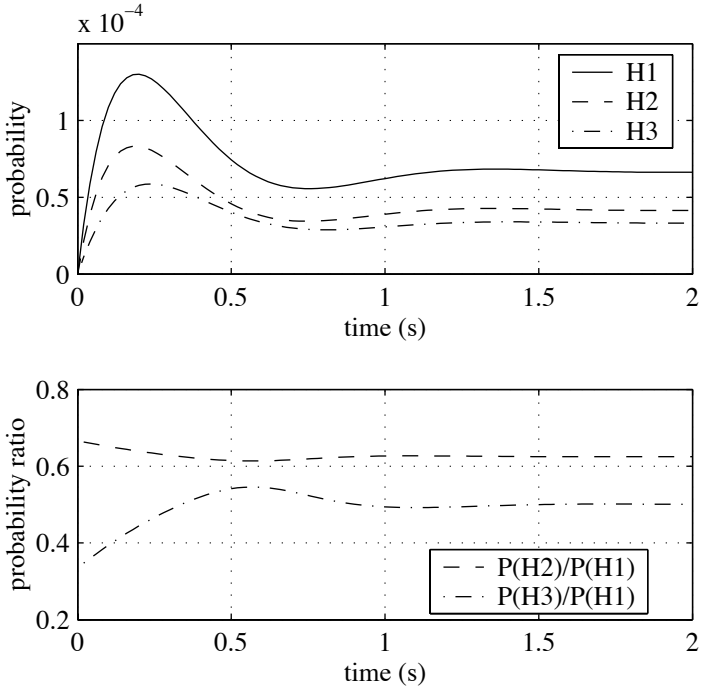


Figure 4.9: The evolution of three configurations that evolve similarly. The configurations are pictured in Figure 4.7.

small time step Δt, and approximate equation (2.5) with

$$\mathbf{x}_d[j+1] = (I + \Delta t \sum_{i=1}^{m} k_i(\mathbf{u}) N_i)\mathbf{x}_d[j] \tag{4.17}$$

$$\mathbf{y}_d[k] = C\mathbf{x}_d[j], \tag{4.18}$$

where $\mathbf{x}_d[k] \in \mathbb{R}^n$ is the discrete-time probability vector at $t = j\Delta t$, $j = 0, 1, \ldots$, and I is an n-dimensional identity matrix. We use a notation similar to that of the previous section, but note here that Δt should be sufficiently small to capture the fastest behavior, so that the first-order Taylor expansion of the matrix exponential is accurate.

We now further define the state matrix as

$$A_{\mathbf{u}} = (I + \Delta t \sum_{i=1}^{m} k_i(\mathbf{u}) N_i). \tag{4.19}$$

When $\mathbf{u}$ is held constant, $A_{\mathbf{u}}$ is also constant and may be identified using purely linear computation. In particular, we construct the generalized observability matrix

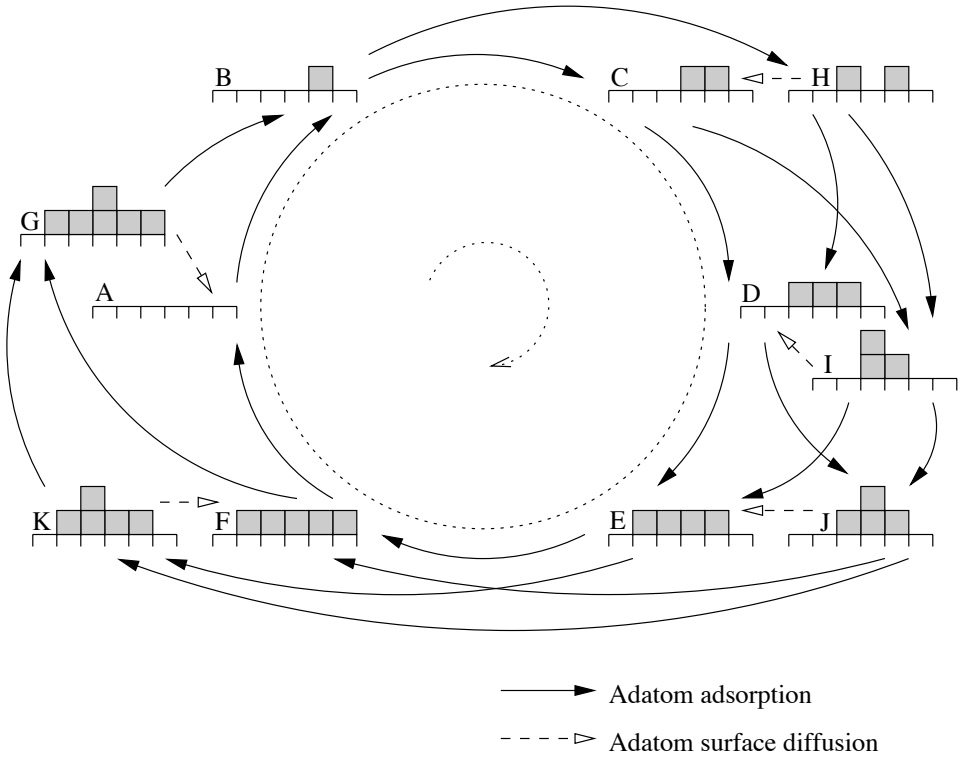


Figure 4.10: Configurations and transitions for the eleven-state reduced master equation, for transition mechanisms of adsorption and adatom surface diffusion.

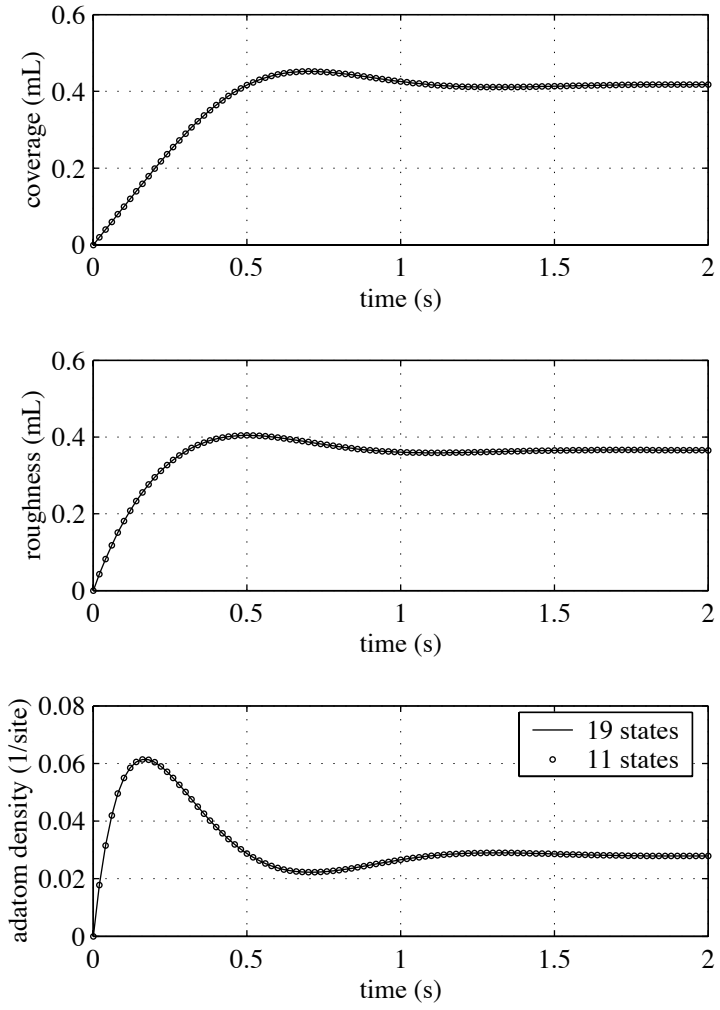


Figure 4.11: Comparison of the reduced eleven-state master equation to the original nineteen-state master equation, for $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{ads}} = 1 \times 10^4\ s^{-1}$.

$\mathcal{O}_{\mathbf{u}}$, which is defined as

$$\mathcal{O}_{\mathbf{u}} = \begin{bmatrix} C \\ CA_{\mathbf{u}} \\ CA_{\mathbf{u}}^2 \\ CA_{\mathbf{u}}^3 \\ \vdots \end{bmatrix} \tag{4.20}$$

for our discrete-time master equation. The observability matrix is prominent in linear systems theory, system identification [29], and model reduction [39], and is used to describe the relationship between the state of a system and its output.

Again holding $\mathbf{u}$ fixed, we observe that $\mathcal{O}_{\mathbf{u}}$ may be constructed from simulations of equations (4.17) and (4.18). The j^{th} column of $\mathcal{O}_{\mathbf{u}}$ is equal to the output column vector $\{\mathbf{y}_d[0], \mathbf{y_d}[1], \mathbf{y_d}[2], ...\}$, generated from initial condition $\mathbf{x}_d[0]$ having probability one in the j^{th} configuration. The observability matrix may alternatively be generated using kinetic Monte Carlo simulations. Once again, the j^{th} column of $\mathcal{O}_{\mathbf{u}}$ is equal to the expected values of outputs, after initiating the simulation in the j^{th} configuration.

Once $\mathcal{O}_{\mathbf{u}}$ has been constructed, the output matrix C can be extracted as the first block of $\mathcal{O}$, while $A_{\mathbf{u}}$ may be determined in a straightforward linear least squares computation, exploiting the shift property of $\mathcal{O}_{\mathbf{u}}$:

$$\begin{bmatrix} C \\ CA_{\mathbf{u}} \\ CA_{\mathbf{u}}^2 \\ \vdots \\ CA_{\mathbf{u}}^{n_t-1} \end{bmatrix} A_{\mathbf{u}} = \begin{bmatrix} CA_{\mathbf{u}} \\ CA_{\mathbf{u}}^2 \\ CA_{\mathbf{u}}^3 \\ \vdots \\ CA_{\mathbf{u}}^{n_t} \end{bmatrix}, \tag{4.21}$$

where n_t is the number of time steps performed in the simulations.

The linear least squares computation is guaranteed to give the globally optimal solution for $A_{\mathbf{u}}$, in which optimal means that the two-norm of the residual is minimized. However, this may not be the optimal solution for our application. In particular, we know that the continuous-time state matrices N_i are stochastic matrices, in which the columns sum to zero, the diagonal elements are non-positive, the off-diagonal elements are nonnegative, and the eigenvalues never have positive real part. These properties guarantee the conservation of probability, and may be enforced through linear equality and inequality constraints in a *constrained* linear least squares solution to equation (4.21). In particular, we constrain each element of $A_{\mathbf{u}}$ to be nonnegative, and constrain the columns to sum to 1, the analogous properties of a discrete-time stochastic matrix. When Δt is small, this also enforces the stochastic properties of the N_i in continuous time. We further note that an element of $A_{\mathbf{u}}(p, q)$ may be set to zero using additional equality constraints, to eliminate the possibility of a transition from configuration p to configuration q. Such a constraint would typically be justified based on physical arguments.

The identification algorithm developed in this section produces $A_{\mathbf{u}}$, but does not independently yield the state matrices N_i. However, because $A_{\mathbf{u}}$ is linear in

N_i, several observability matrices may be constructed for different (but constant) transition rates, which are then assembled into a single constrained linear-least squares problem, yielding $N_i, i = 1, ...m$.

There may or may not be a unique solution to the constrained linear least squares computation. This computation is guaranteed to yield a solution with the minimum two-norm of the residual, but if the problem is underconstrained, this solution may not be unique. The condition for unique determination of $A_{\mathbf{u}}$ depends not only on the rank of $\mathcal{O}_{\mathbf{u}}$, but also on the equality constraints imposed. We suggest only that if the problem is underdetermined, one could either used more data in the identification, select additional outputs to provide more information about the state, or perhaps reduce the number of configurations, if some are redundant or are not contributing to the output. Alternatively, if the problem is underdetermined, then the connectivity implied by the least squares solution to $A_{\mathbf{u}}$ may be sufficient to describe the output of interest.

The system identification algorithm is demonstrated using the six-site system considered earlier in this chapter. The two state matrices (for adsorption and surface diffusion) and the output matrix are computed using KMC simulation data with initial conditions in each of the eleven configurations of the reduced system, for 0.5 s with data saved at 0.1 s intervals, and with two sets of transition rates: $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{dif},0} = 1 \times 10^4\ s^{-1}$, and $k_{\text{ads}} = 1\ s^{-1}$ and $k_{\text{dif},0} = 1 \times 10^2\ s^{-1}$. At each set of conditions, 1000 realizations are performed and averaged. This data is then used to compute the state matrices with and without constraints. The results are shown in Figure 4.12 with adsorption rate varied randomly at 0.1 s intervals between 0 and 1 s^{-1}, and diffusion rate varied between 0 and $2{\times}10^4\ s^{-1}$. The constrained system compares much better with the original master equation than does the system obtained with the unconstrained computation. In the latter case, the eigenvalues of the state matrices have positive real parts, making the system unstable. Because it is straightforward to include linear constraints in the optimization, we will always include them in computations in the remainder of this work, which will guarantee that the identified system is also a master equation.

4.3 Alternate formulation of the master equation

A high-dimensional master equation is the mathematical representation underlying the morphology evolution of KMC simulations like those pictured in Figure 2.3. However, there may be circumstances in which the *expected* value of the film measures predicted by the master equation differs from *typical* behaviors observed in individual KMC realizations. We observe this effect in Figure 4.6, in which individual realizations exhibit oscillations correlated with the deposition of a monolayer, but in which the expected values exhibit little oscillation. The disparity between the expected and typical behaviors is due to a dephasing effect, stemming from randomness in the adsorption times. The variation in phase is particularly great in this small, six-site system, since each adsorption event results in the deposition of a large fractional coverage. In the opposite limit, as the number of sites grows

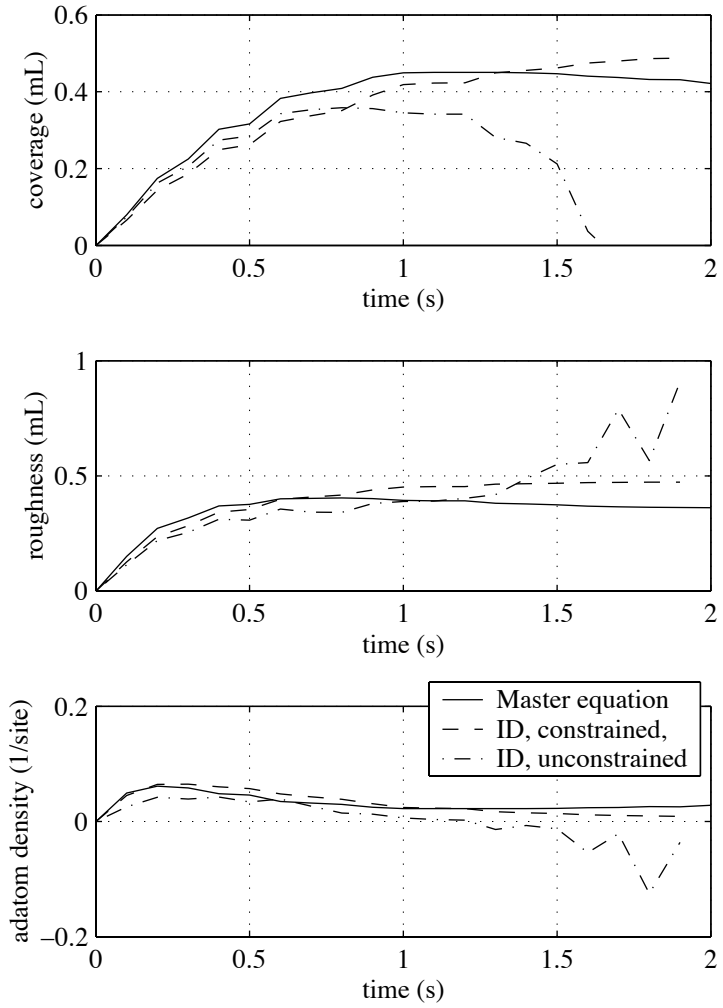


Figure 4.12: System identification of the six-site lattice model. The identification procedure is applied with and without constraints, and both systems are compared to a direct integration of the master equation. The inputs are randomly varying at 0.1 s intervals, with an adsorption rate between 0 and 1 s^{-1} and a diffusion rate between 0 and 2×10^4 s^{-1}.

large, the dephasing effect diminishes.

Ultimately, we would like to model the behavior seen in large KMC simulations with low-order representations, and thus seek a mathematical framework that is capable of capturing the types of behavior seen in large lattice systems. We thus reformulate the lattice model and master equation to eliminate the dephasing effect due to the random adsorption. We simply take the total number of adsorption events as the dependent variable, instead of time. Because the adsorption mechanism is site-independent, the mean time between adsorption events is known *a priori*, as well as the distribution of times, which is a Poisson distribution [15]. Additionally, over many events the time can be correlated with the total number of adsorption events using the adsorption rate k_{ads}.

The mean time between adsorption events is $1/(k_{\text{ads}}N)$, where N is the number of surface sites. Note that this is the mean time between adsorption events anywhere on the surface, and that each adsorption event may occur with equal probability at any site on the surface. The distribution of adsorption times τ follows the Poisson distribution:

$$P(\tau) = k_{\text{ads}} N \exp\left(-k_{\text{ads}} N \tau\right). \tag{4.22}$$

We develop an alternate, discrete formulation of the master equation by computing the probability distribution immediately following each adsorption event. Between these events, the system evolution is governed by the remaining transition mechanisms.

The adsorption event is represented by a discrete-time stochastic matrix, which we call A^d_{ads}. To guarantee the conservation of probability, each element must be nonnegative, and each column must sum to one. An adsorption event is then executed by multiplying the probability distribution by A^d_{ads}. Column j of A^d_{ads} is constructed by considering the probabilities of being in each of the n configurations, after beginning in configuration j and executing an adsorption event. This idea is illustrated by Figure 4.13, with corresponding matrices,

$$\begin{aligned} N_1 &= \begin{bmatrix} -b & 0 & 0 & 0 \\ b & -c-d & 0 & e \\ 0 & c & 0 & 0 \\ 0 & d & 0 & -e \end{bmatrix} \\ A^d_{\text{ads}} &= \begin{bmatrix} \frac{a}{a+b} & 0 & 0 & 0 \\ \frac{b}{a+b} & 0 & 0 & 0 \\ 0 & \frac{c}{c+d} & 0 & 1 \\ 0 & \frac{d}{c+d} & 0 & 0 \end{bmatrix}, \end{aligned}$$

where adsorption is considered to be mechanism 1, and N_1 is the continuous-time state matrix for adsorption. Both N_1 and A^d_{ads} reflect the adsorption transitions pictured by the connectivity graph of Figure 4.13, but serve different purposes. Note that N_1 does not contain the transition from 1 to itself, while A^d_{ads} incorporates this null transition since there is a nonzero probability that no configuration

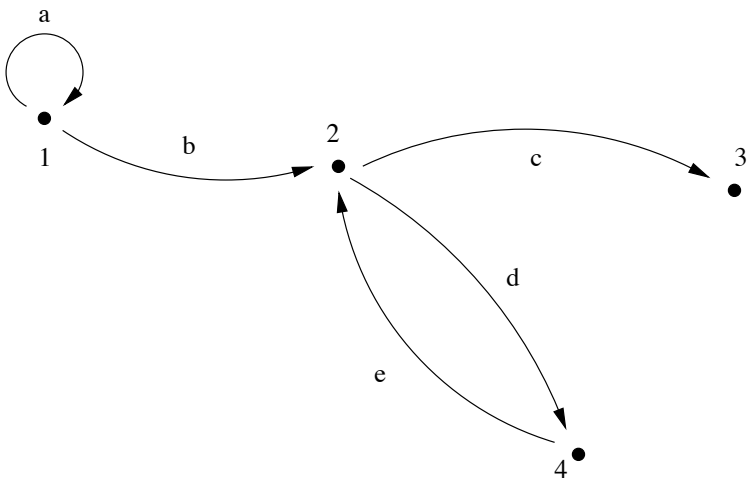


Figure 4.13: Sample graph for adsorption, to illustrate the construction of a discrete-time adsorption matrix.

change will occur as the result of the adsorption event. Conversely, the coefficient e for transition from 4 to 2 is not specified explicitly in A^d_{ads}, since it is the only transition out of configuration, and therefore must occur when an adatom adsorbs on configuration 4.

Adsorption events are executed by multiplication of A^d_{ads} with the probability distribution at adsorption times specified by the Poisson distribution. Between these points in time, the probability distribution is integrated continuously over the remaining $i = 2, ...m$ mechanisms. If we assume that the input is constant over each of these intervals, we obtain

$$\begin{aligned}\mathbf{x}_s(t) &= \exp\left(\sum_{i=2}^{m} k_i(\mathbf{u}_r)N_i(t-\sum_{q=1}^{r}\tau_q)\right)A^d_{\text{ads}}\exp\left(\sum_{i=2}^{m} k_i(\mathbf{u}_{r-1})N_i\tau_{r-1}\right)A^d_{\text{ads}}\cdots \\ &\quad A^d_{\text{ads}}\exp\left(\sum_{i=2}^{m} k_i(\mathbf{u}_2)N_i\tau_2\right)A^d_{\text{ads}}\exp\left(\sum_{i=2}^{m} k_i(\mathbf{u}_1)N_i\tau_1\right)\mathbf{x}_s(0),\end{aligned}$$

where $\mathbf{x}_s(t) \in \mathbb{R}^n$ is a stochastic realization of the probability distribution, τ_j is the time between adsorption events $j-1$ and j, $\mathbf{u}_j$ is the corresponding input for the j^{th} interval, and $t > \sum_{q=1}^{r}\tau_q$ and $t \leq \sum_{q=1}^{r+1}\tau_q$. Note that $\{\tau_j\}$ may be precomputed, since it is subject only to the Poisson distribution, and is not dependent on the probability distribution.

We may not need to know the probability distribution at all times, but instead would be satisfied with the probability distribution at the discrete times following each adsorption event. In this case we formulate a discrete master equation

$$\mathbf{x}^d_s[j+1] = A^d_{\text{ads}}\exp\left(\sum_{i=2}^{m} k_i(\mathbf{u}_{j+1})N_i\tau_{j+1}\right)\mathbf{x}^d_s[j], \tag{4.23}$$

where $\mathbf{x}_s^d[j] \in \mathbb{R}^n$ is the probability vector after j adsorption events.

A simulation of equation (4.23) is a stochastic realization, since it depends on the Poisson-distributed adsorption times $\{\tau_j\}$. However, we may instead compute the expected probability after j adsorption events:

$$\begin{aligned}
\left\langle \mathbf{x}_s^d[j] \right\rangle &= \int_0^\infty \cdots \int_0^\infty \int_0^\infty A_{\text{ads}}^d \exp\left(\sum_{i=2}^m k_i(\mathbf{u}_j) N_i \tau_j\right) \cdots \\
&\quad A_{\text{ads}}^d \exp\left(\sum_{i=2}^m k_i(\mathbf{u}_j) N_i \tau_2\right) A_{\text{ads}}^d \exp\left(\sum_{i=2}^m k_i(\mathbf{u}_j) N_i \tau_1\right) \mathbf{x}_s^d[0] \\
&\quad k_1(\mathbf{u}_j) N \exp\left(-k_1(\mathbf{u}_j) N \tau_j\right) \cdots \\
&\quad k_1(\mathbf{u}_2) N \exp\left(-k_1(\mathbf{u}_2) N \tau_2\right)\, k_1(\mathbf{u}_1) N \exp\left(-k_1(\mathbf{u}_1) N \tau_1\right) d\tau_1 d\tau_2 \cdots d\tau_j \\
&= \prod_{q=1}^j A_{\text{ads}}^d \left\langle \exp\left(\sum_{i=2}^m k_i(\mathbf{u}_q) N_i \tau_q\right) \right\rangle \mathbf{x}_o,
\end{aligned}$$

by integrating over the Poisson distribution for each of the uncorrelated adsorption times, assuming a known initial probability distribution of $\mathbf{x}_s^d[0] = \mathbf{x}_o$, and integrating sequentially, beginning with τ_1. Defining

$$\hat{A}(\mathbf{u}_j) \equiv A_{\text{ads}}^d \left\langle \exp\left(\sum_{i=2}^m k_i(\mathbf{u}_j) N_i \tau_j\right) \right\rangle \tag{4.24}$$

according to the Poisson distribution, we finally obtain the discrete evolution equation

$$\mathbf{x}^d[j+1] = \hat{A}(\mathbf{u}_j) \mathbf{x}^d[j], \tag{4.25}$$

where $\mathbf{x}^d[j] \in \mathbb{R}^n$ is the probability distribution after j adsorption events. The matrix $\hat{A}(\mathbf{u}_j)$ is a discrete-time stochastic matrix, such that the columns sum to one and the elements are positive, as required for a probabilistic system. However, there is now no direct map back through a matrix exponential to continuous-time state matrices. $\hat{A}(\mathbf{u}_j)$ is now a nonlinear function of $k_i(\mathbf{u})$ and N_i, although the linearity in the probability vector has been preserved.

Synchronizing changes in the input vector with each adsorption event would pose many practical challenges. Of course, if the input is held constant, there is no difficulty. At this point in the development we are not actually interested in implementing a control strategy that would be synchronized with individual adsorption events, but are more concerned with identifying a suitable mathematical structure to represent the behavior of large lattice systems. In a system with many surface sites, the time between adsorption events is very small, and if one is interested in changing the inputs at larger times, one may approximate the number of adsorption events with the mean time to deposit them. Before we move on to large surfaces and their KMC simulations, we demonstrate the discrete model on another small surface.

4.3.1 Example: adsorption-desorption

The discrete formulation of the lattice model is demonstrated with a six-site model evolving via adsorption and adatom desorption. Kinetic Monte Carlo simulations are performed for an adsorption rate of $k_{\text{ads}} = 1\ s^{-1}$ and adatom desorption rates of $k_{\text{des},0} = 100$ and $1000\ s^{-1}$. Each set of rates is simulated 1000 times, with output measures computed and saved not at discrete intervals of time, but instead immediately after each adsorption event. The simulations run until 12 adsorption events have occurred.

A corresponding master equation is also constructed, by first selecting an appropriate finite set of configurations. We take here a more data-driven approach to the identification of configurations, by adsorbing up to seven atoms at random locations on the surface, allowing no desorption. We then eliminate configurations considered to be improbable, in which two or more adatoms are stacked on top of each other, or in which four or more layers are occupied with fractional coverages. The resulting configuration set consists of 65 configurations. The continuous-time state matrices are constructed based on this configuration set, and are then used to compute the discrete maps of equation (4.24), one for each of the two desorption rates considered.

Configurations that evolve in fixed ratios were determined through simulations of the master equation. The time-dependent probability of each configuration was compared against every other configuration, and those that evolve together were grouped together through a coordinate transformation, yielding a reduced state dimension of 24.

Figure 4.14 shows the simulation results for both desorption rates considered, and compares the KMC simulations, the 65-state discrete master equation, and the reduced discrete master equation. All three simulation methods compare well within the error bars of the KMC simulations, which denote 95% confidence intervals. The predictions of thickness begins to diverge at the end of the low desorption rate simulations, which could be improved if additional rougher configurations were added to the configuration set.

The system identification procedure is now applied to this discrete model. The algorithm proceeds similarly to that developed for continuous-time master equations. KMC simulation data is used to generate output data, with simulations performed at both desorption rates and with initial conditions at each of the 24 configurations in the reduced set. Ensembles of 1000 simulations are used to compute the expected output, and an observability matrix is constructed for each set of transition rates considered.

In the identification of the continuous-time system, the individual state matrices N_i are computed in the constrained linear least squares computation. However, in the discrete model the N_i do not appear linearly, and instead a single stochastic matrix is generated for each set of transition rates, corresponding to $\hat{A}(\mathbf{u}_j)$ in equation (4.25). Thus, we exploit the linearity of the state in the identification process, but can no longer exploit the linearity of the N_i. Recall also our earlier reference to a singular perturbation approach to configuration reduction. We rejected

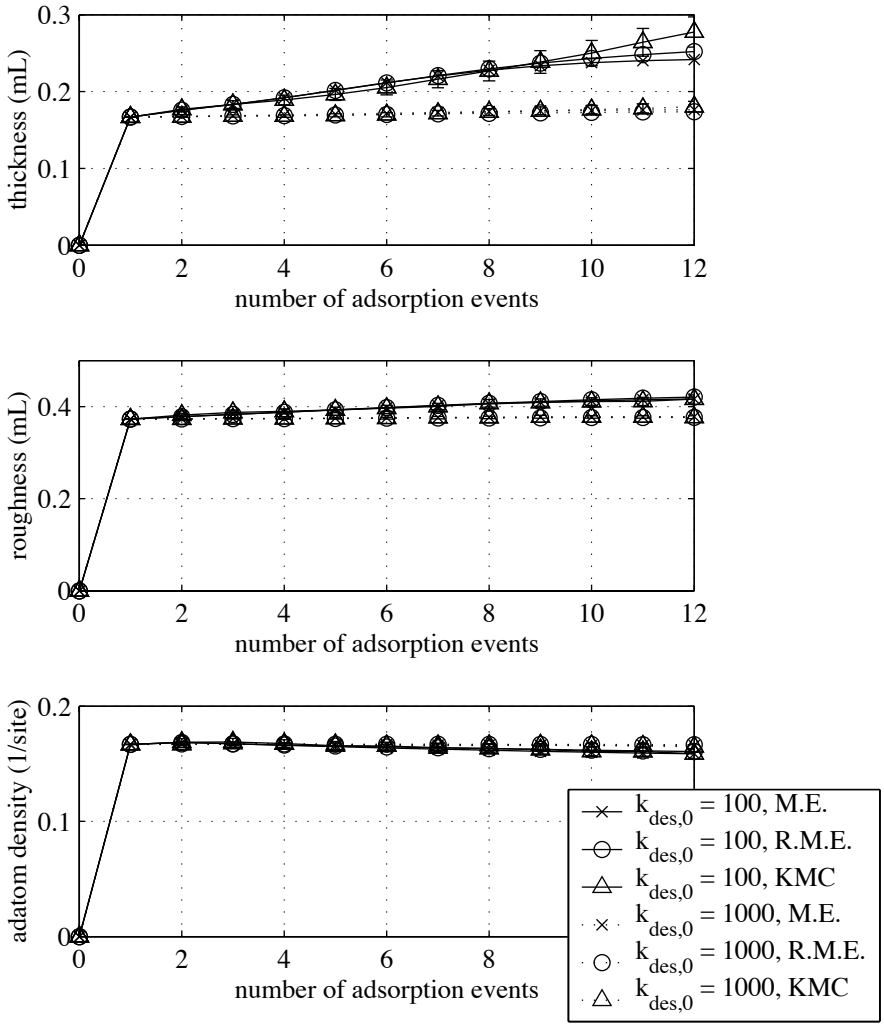


Figure 4.14: Comparison of the evolution predicted by KMC simulations, a 65-state master equation, and a reduced-order model, with $k_{\text{ads}} = 1\ s^{-1}$ and two values of the desorption rate.

this approach because it did not preserve linearity in transition rates, although in the discrete model the linearity is lost for other reasons. The connectivities determined through the system identification procedure likely include effects like setting probabilities of additional configurations to their equilibrium values.

The evolution of the identified system for inputs not used in the identification is not explicitly captured, and we use linearly interpolation to obtain the state matrix at intermediate values of the inputs. While this is only an approximation, a linear interpolation between stochastic matrices produces another stochastic matrix, so stability is guaranteed. Additionally, the continuity of the matrix exponential of equation (4.24) guarantees that the state matrices change continuously with the transition rates.

We demonstrate the identified model in Figure 4.15, comparing an ensemble of KMC simulations to the predictions of the identified model. The desorption rate is held at 100 s^{-1} during the first six events, after which it is raised to 200 s^{-1}. The state matrix for $k_{\mathrm{des},0} = 200\ s^{-1}$ is computed via a linear interpolation of the matrices for 100 and 1000 s^{-1}. Notice that the surface follows a roughening trajectory during the first half, but then levels out after the desorption rate is raised.

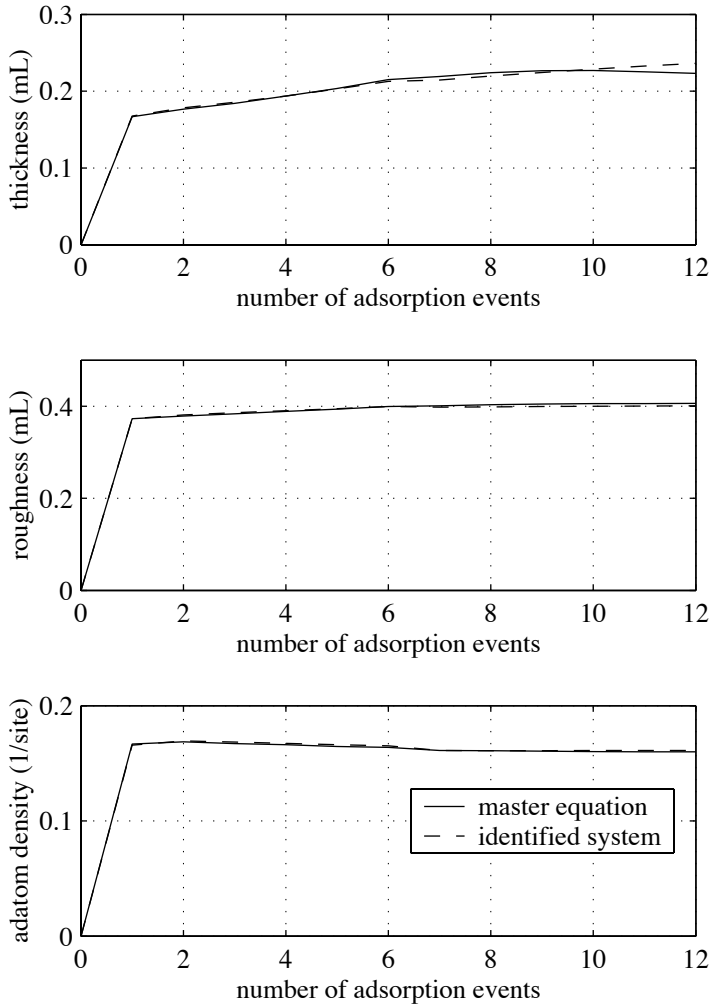


Figure 4.15: Comparison between an ensemble of KMC simulations and the reduced-order model obtained through system identification. The adatom desorption rate is 100 s^{-1} during the first six events, and 200 s^{-1} during the remainder, while the adsorption rate is held at 1 s^{-1}.

Chapter 5

Application to Large Lattices

In the previous chapter, we demonstrated model reduction and system identification techniques on small lattice systems. Based on this investigation, we observe that the number of configurations and corresponding state dimension may be larger than is necessary to capture the film metrics of interest. The number of configurations may be reduced either because some configurations are not occupied, because they only provide fast transition pathways, or because redundant configurations, or sets of configurations, exist.

In the examples of the previous chapter, we considered one-dimensional surfaces with a small number of surface sites. In such systems, the configurations could be enumerated explicitly to construct the state and output matrices. The master equation was then formally reduced, either by modifying the state matrices, or by performing a linear coordinate transformation.

Ultimately, we are interested in the evolution of large surfaces, in which the coordinated interaction of a large number of atoms leads to the large-scale features seen in the simulations of Figure 2.3. In such cases, the large number of possible atomic arrangements makes the enumeration process impractical, even for the simplest behaviors like two-dimensional nucleation and growth. Instead, we observe that there are many configurations that have the same film metrics (roughness, step density, etc.) and that these groups of configurations tend to evolve similarly. Because it is not feasible to list all configurations and then group them through a coordinate transformation, we postulate that such groups exist, and, as in the small systems considered previously, select one microscopic configuration to represent each group. When considering a particular growth mode, such as spiral growth or two-dimensional nucleation, one would expect that some groups of similar configurations that will be accessed frequently, while others will not.

Selection of the configuration set requires an understanding of the physics, and may require iteration on the part of the user. These configurations may be constructed directly, or may be extracted from KMC simulations. The states of the reduced master equation are then the probabilities of each of the configuration groups. Once the states of the master equation have been established, it is straightforward to apply the system identification algorithm of Chapter 4 to compute the associated state and output matrices—the observability matrices are constructed

using KMC simulations with each of the representative configurations as an initial condition. In this chapter, a master equation determined by this process is referred to as a reduced-order model (ROM).

It would be desirable to have a systematic way of determining the states of the master equation. Data-driven methods have been used to identify a reduced number of modes in high-dimensional systems, with particular success in complex fluid flows [26]. Linear combinations of the states are selected as the new states, which are orthogonal and which capture the maximum amount of energy, as measured by the two-norm. The full equation is then projected onto the modes, yielding a reduced-order model.

A similar idea has been applied to surface evolution in film growth [18, 46], including work by the author that is reported in Appendix B. In both of these studies, spatial modes were identified via KMC simulation. However, it is not clear how one can formulate an analogous evolution equation using spatial coordinates as the states, since the states in the master equation are probabilities of configurations. No reduced-order model was developed in either study, as there was no original equation on which to project the spatial modes. One might instead envision selecting linear orthogonal modes based on the original probability coordinates of the master equation. In fact, this is the idea underlying each of our reduced coordinates, which are interpreted as the sum over the probabilities of all configurations in the group, and are therefore orthogonal. We restrict ourselves to this type of coordinate, because we wish our reduced-order model to also be a master equation. One could consider developing a systematic algorithm to identify the best modes of this form. We also note that, in general, the task of identifying modes through KMC simulations is complicated by the fact that the states (probability distribution) are not available directly, even in simulation, since the simulations are stochastic realizations, and the number of possible configurations is extremely large.

In this work, we postulate the configuration sets, and then identify corresponding reduced-order models, leaving the systematic identification of coordinates as future work. We consider two physical scenarios in which time-varying process parameters produce altered surface morphology. In both cases, KMC simulations are first used to identify a model. The model is then used to generate optimal process parameters that minimize various cost functions.

5.1 Example 1: transition from smooth to rough growth

We first consider the evolution of a surface through adsorption and adatom surface diffusion. The adsorption rate is fixed, while the temperature, and thus the diffusion rate, may vary within a limited operating range. At the maximum temperature, the surface approaches the limiting behavior of two-dimensional nucleation and coalescence, while at the minimum temperature, the surface demonstrates three-dimensional roughening after only a few layers of deposition. The application of periodic inputs usually leads to a roughening of the surface when compared

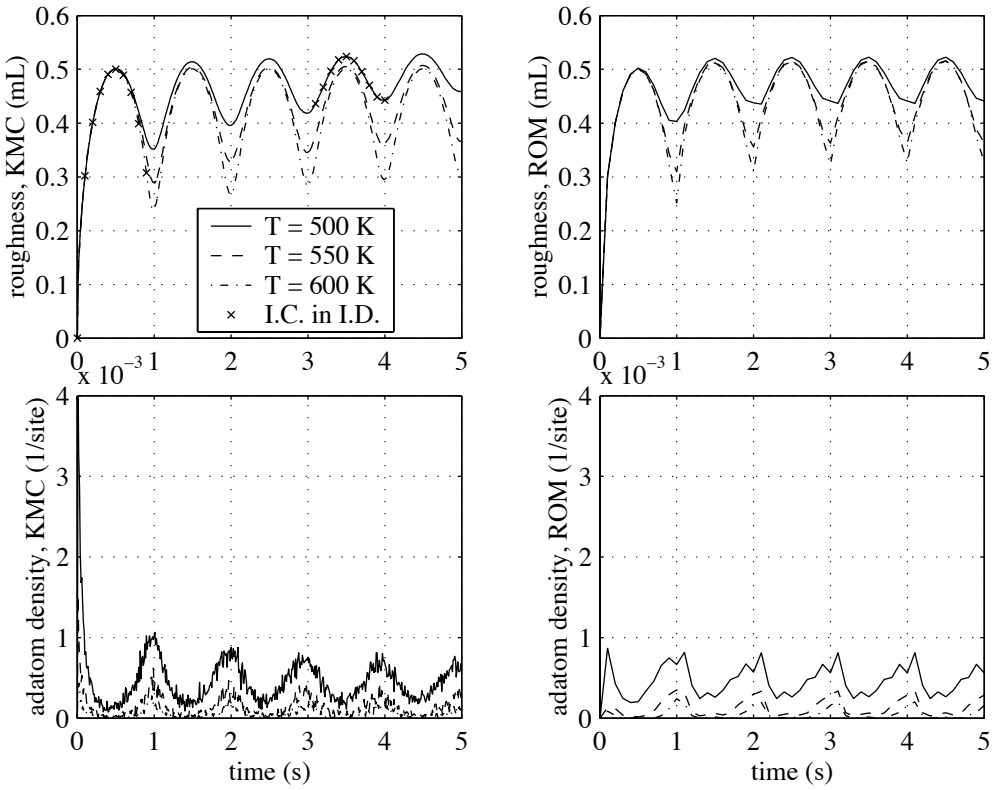


Figure 5.1: Comparison of roughness and adatom density predicted by kinetic Monte Carlo simulations and a reduced-order model. KMC simulations are performed on a 300×300 lattice at temperatures of 500, 550, and 600 K. Configurations used as states in the reduced model are denoted by 'x's and are chosen to represent limiting behaviors of the system.

to growth at the mean values of the process parameters. This roughening is attributed to the creation of compact rough features when the instantaneous adatom density is high [58]. These features do not decay fully when the adatom density is lowered, ultimately yielding a rougher surface. We demonstrate this effect in KMC simulations, identify a reduced-order model using KMC simulations, and then show that it also predicts the roughening effect of periodic inputs. The model is then used to generate optimal temperature profiles.

The KMC simulations are performed on a 300×300 domain, with $k_{\text{ads}} = 1\ s^{-1}$, $k_{\text{dif},0} = 10^{13} \exp\left(E_{\text{dif},0}/k_b/T\right)\ s^{-1}$, and $E_{\text{dif},0}/k_b = 10000\ K$. The temperature is allowed to vary between 500 K and 600 K. The plots on the left side of Figure 5.1 show KMC simulations at three constant temperatures. Oscillations in both roughness and adatom density are indicative of smooth island nucleation and coalescence behavior, which decays faster at lower temperature. The 'x's on the plot mark representative points—and corresponding configurations—that we consider important in the evolution. In particular, ten smooth configurations at uniform intervals of surface coverage are extracted from the first layer of growth at the

highest temperature, while corresponding rougher configurations are taken from the simulations at the lowest temperature after several layers of deposition. After identifying these 20 configurations, we construct two configurations from each: one with all adatoms removed, and one with adatoms added randomly to produce a density of 10^{-3} site^{-1}.

The configuration set consists of these 40 configurations, yielding 40 corresponding states in the reduced-order model. The observability matrices for the identification are constructed by performing 120 KMC simulations, using the 40 initial configurations, and performing simulations at 500, 550, and 600 K for 0.1 s. Only a single realization is performed for each combination of initial condition and temperature. The simulation domain is sufficiently large such that the fluctuations, as seen in Figure 5.1, are small. However, the fluctuations do introduce noise into the observability matrix, which ultimately tests the robustness of the identification technique. We do generally expect the algorithm to be robust, since it is based on a linear least squares computation.

The discrete identification algorithm is used to compute a single state matrix at each of the three temperatures used in the simulation, as well as to generate the output matrix. We choose to identify a discrete equation, and not a continuous master equation, because we wish to capture the oscillatory behavior associated with island nucleation and coalescence, as observed in the KMC simulations, and to do so using a small number of configurations. In a continuous-time master equation, we would observe the dephasing effect seen in Figure 4.6, so we choose to index the probability distribution by the number of adsorption events that have occurred.

Because we are dealing with a dependent variable that measures adsorption events, and because the configurations exist only at intervals of 0.1 mL, the probability distribution is actually only computed after a series of adsorption events have occurred, resulting in the deposition of 0.1 mL. The state matrix may thus be viewed as a map from one coverage level to the next. This picture is manifested in the constraints on the linear least squares computation. From one timestep to the next, a configuration is only allowed to transition to a configuration with a coverage of 0.1 mL greater than its own. All other transitions are set to zero using equality constraints. The state matrices are additionally constrained to be stochastic, with all elements nonnegative, and columns summing to one.

The evolution of the reduced, identified model is pictured by the plots on the right side of Figure 5.1 for the constant temperatures used in the identification. Note that the dependent variable is the time, not the number of adsorption events. In the limit of a large domain, the time to deposit 0.1 mL fluctuates little from the mean time, so we substitute back to time, despite our use of the discrete model.

The qualitative comparison between the KMC simulations and model predictions is good. A notable exception is the initial adatom density, whose initial spike is not captured by the model. First note that the reduced model is discrete, and in fact does not resolve the fast initial transient of the KMC simulation. However, the model prediction at 0.1 s does compare well with the KMC simulation. Additionally, the reduced model is not capable of predicting an adatom density

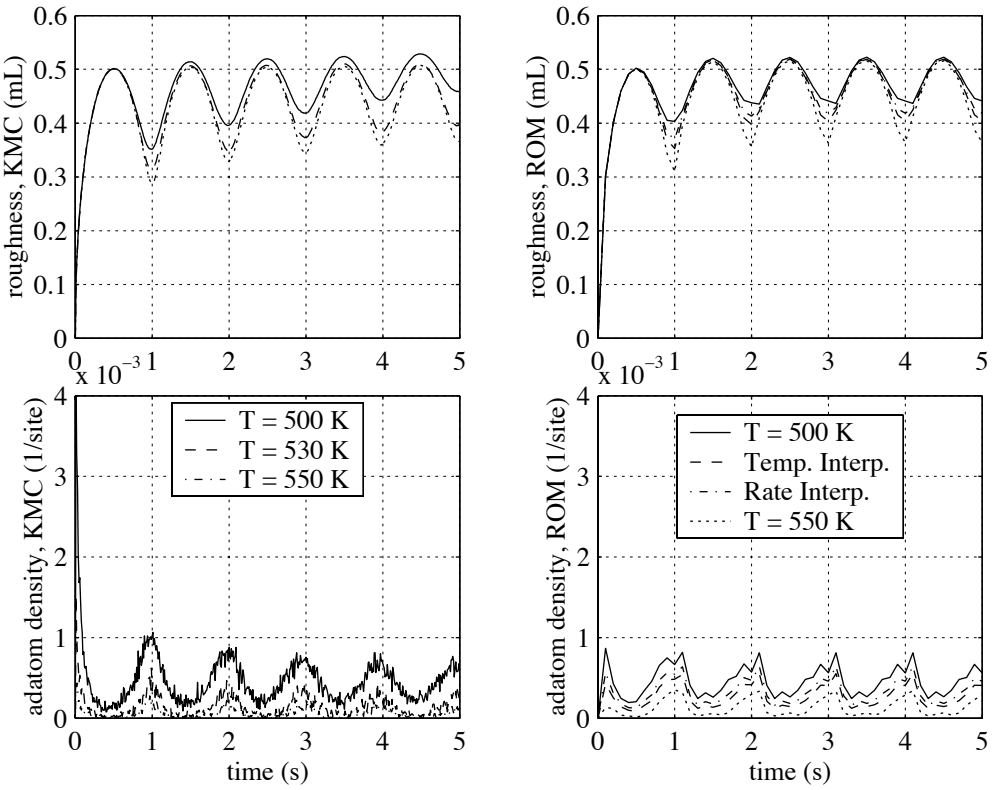


Figure 5.2: Comparison between KMC simulations and the reduced-order model at a temperature not used in the identification. To simulate 530 K, the state matrices for 500 and 550 K are linearly interpolated, either by temperature or by adatom diffusion rate.

of 4×10^{-3} site^{-1}, since all the states either have adatom densities of zero or 1×10^{-3} site^{-1}. The output predicted by the reduced model is always a convex combination of the outputs of each configuration, since the state is a probability vector that sums to one. Thus, if we had wanted to capture a higher adatom density, we would have had to include at least one configuration with this density as a state. As a general point, the output predicted by the model is restricted to the convex hull of the outputs for each configuration, which must be kept in mind when selecting the configuration set, i.e., the configuration set should include the *limiting*, or extreme, configuration that are to be predicted by the reduced model.

The simulations of Figure 5.1 were conducted at the same temperatures as those used in the identification. A better test of the model would to simulate an intermediate temperature. We simulate a temperature of 530 K in Figure 5.2, by linearly interpolating between the state matrices at 500 K and 550 K, using alternately the temperature and the diffusion rate for the interpolation. The roughness and the adatom density for 530 K lie at intermediate values between those for 500 K and 550 K, as we expect based on the physics. Because the differences between the outputs at 500 K and 550 K are not large, we do not see significant

differences between the interpolation methods. This is the desired situation—neither interpolation method provides the correct state matrix for 530 K, so we must identify matrices at intervals of temperature in which the growth behavior does not change dramatically. We also make the general comment that because the state matrices is stochastic, a linear interpolation yields another stochastic matrix, which thus preserves the stability and probabilistic interpretation of the states. This is of critical importance and makes possible an interpolation between the state matrices.

In the previous simulations of the reduced-order model, the initial configuration was the atomically flat surface, and yielded an initial condition of probability one for the corresponding state vector. This configuration was one of the representative configurations used to generate the model, but we now consider an initial condition not in the configuration set. An initial condition is generated with a KMC simulation at 500 K for 1 s (and 1 mL). The roughness and adatom density for this system are not consistent with any single state in the reduced model, so we instead computed a linear combination of probabilities that yields the same output. This probability distribution was then used as the initial condition for the reduced model. A simulation at a different temperature, 600 K, is shown in Figure 5.3, comparing the reduced model to the analogous KMC simulation. The quantitative comparison between the roughness is not perfect, but reduced model does capture the transition to the smoother growth at 600 K, versus the decaying oscillations of 500 K seen in Figure 5.1. We do not expect perfect quantitative comparison between the simulations, but rather strive for reasonable quantitative comparison, and good prediction of the trends associated with changes in the process parameters.

With the goal of predicting the effects of time-varying process parameters, we now consider a periodic temperature profile in which the temperature switches between 500 K and 600 K at 0.2 s intervals. Simulations of the Monte Carlo and reduced-order model are plotted in Figure 5.4, along with simulations at a continuous temperature of 577 K, whose a diffusion rate equals the mean diffusion rate of 500 K and 600 K. Note that in both models, the roughness is lower under the constant temperature, and that the maximum instantaneous adatom density under periodic temperature is much greater than under continuous growth. The reduced-order model predicts the same trends as the Monte Carlo simulations: the periodic temperature leads to an elevated instantaneous adatom density, and ultimately produces a rougher surface, as compared to growth with the mean diffusion rate.

We now continue on to the computation of optimal temperature profiles. The reduced-order model is well suited to this application, due to its low computational demands. The cost function to be minimized is

$$Cost = W(N_{\text{steps}}) + \alpha \sum_{j=1}^{N_{\text{steps}}} |T_j - \min T| + \beta \sum_{j=1}^{N_{\text{steps}}-1} |T_j - T_{j-1}|^2, \tag{5.1}$$

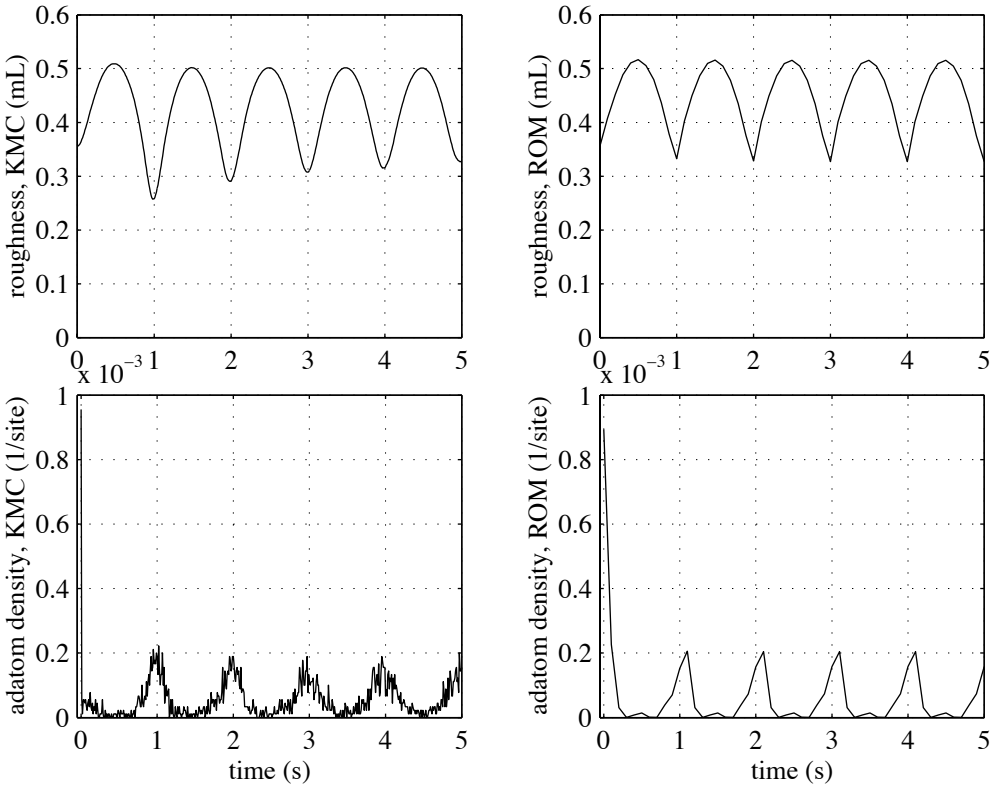


Figure 5.3: Evolution of reduced-order model, and comparison with KMC, beginning with an initial condition that does not correspond to a single state in the reduced model. To generate the initial condition for the identified model, a constrained least linear squares problem was solved to find a probability distribution that yielded the appropriate outputs.

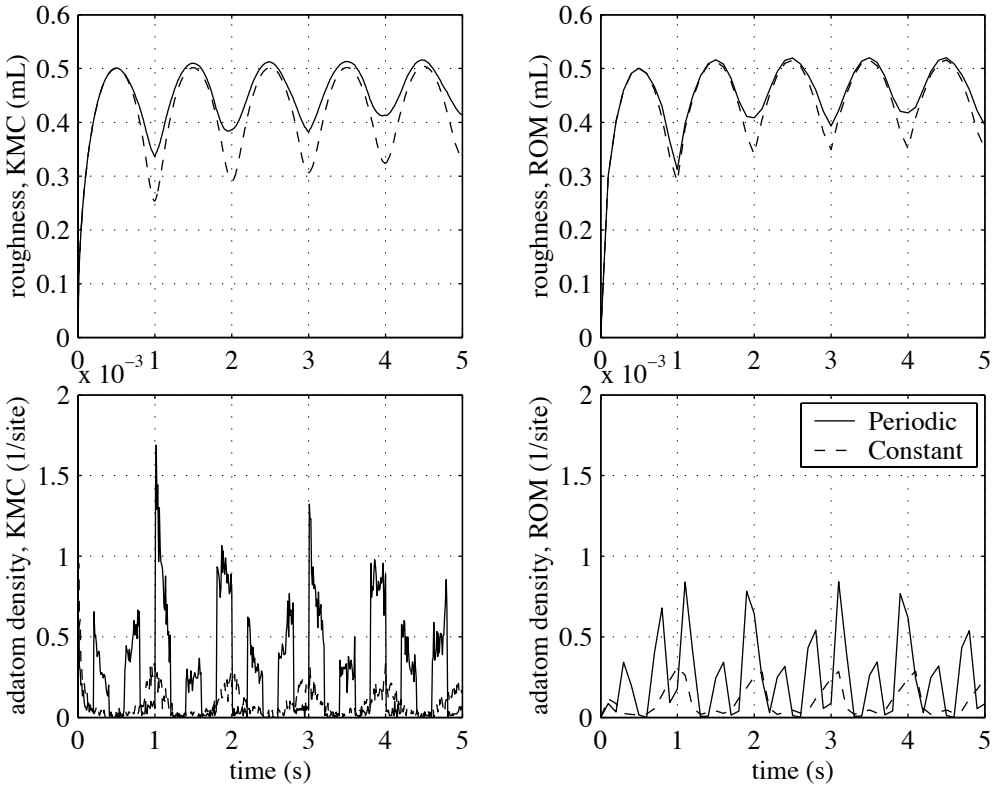


Figure 5.4: KMC and ROM simulations of growth with temperature switching between 500 and 600 K every 0.2 s. This evolution is compared to growth at a constant temperature of 577 K, which yields a diffusion rate equal to the mean of the diffusion rates at 500 and 600 K.

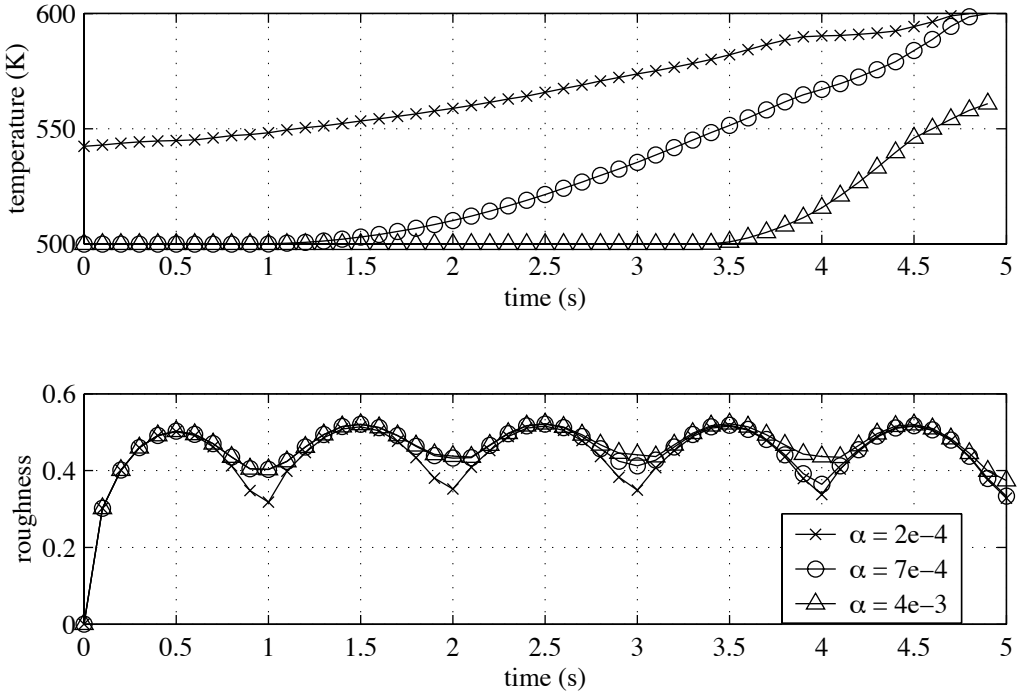


Figure 5.5: Optimal temperature profiles computed using the identified model and the cost function of equation (5.1), with $\beta = 2 \times 10^{-3}$ and with three values of α.

where N_{steps} is the number of time steps considered in the optimization, $W(N_{\text{steps}})$ is the final roughness, and α and β are the coefficients that reflect the cost associated with a high temperature and with a fast temperature change, respectively. We fix β at 2×10^{-3}, and perform the optimization for three values of α: 2×10^{-4}, 7×10^{-4}, and 4×10^{-3}. This cost function reflects not only the final surface properties, but also operating costs and reactor constraints associated with changing the temperature input.

Optimal temperature profiles and roughness trajectories are plotted in Figure 5.5. In general, a high temperature leads to a smooth surface, but as α is raised, a high temperature is increasingly penalized in the cost function, resulting in a lower overall temperature. Note also that raising the temperature at the end of growth is more desirable than raising it at the beginning, since the cost function only reflects the final roughness.

Table 5.1 lists the value of the cost function for the three values of α, and for different temperature trajectories. The cost for the optimal trajectory is compared to the cost for constant growth at T = 500 K and T = 600 K, and for the periodic growth strategy of Figure 5.4. The cost for the periodic temperature is high, due to the instantaneous switching of temperature between its minimum and maximum values, and therefore the cost is not strongly dependent on α. The difference in cost between $T = 600\ K$ and the optimal temperature is quite small for the lowest value of α, since a high temperature is not penalized much, but the costs diverges

	$\alpha = 2 \times 10^{-4}$	$\alpha = 7 \times 10^{-4}$	$\alpha = 4 \times 10^{-3}$
$T = 500\ K$	0.441	0.441	0.441
$T = 600\ K$	0.347	0.397	0.727
periodic T	10.0	10.0	10.2
optimal T	0.346	0.368	0.420

Table 5.1: Values of the cost function, equation (5.1), for various temperature profiles and values of α, with $\beta = 2 \times 10^{-3}$.

for increasing values of α. In the opposite limiting case of large α, a constant temperature of $T = 500$ will become optimal.

This type of optimization study would be impractical to perform with KMC simulations, due to their inherent noise and high computational demands. (See Section 5.3 for an analysis of the computational costs of direct optimization using KMC simulations.) However, by extracting the essential dynamics of the KMC simulations into a reduced-order model, we may apply tools like optimization to study the dynamics of surface morphology evolution.

5.2 Example 2: synchronized nucleation

We now consider a second example of surface morphology evolution in which time-varying temperature produces altered morphology. In contrast to the previous example, time-varying temperature may be used here to produce a surface with desirable properties that are unattainable using a constant temperature profile. In particular, we consider a temperature trajectory that is low during the beginning of a layer, and high near the end [37].

When the surface evolves under a constant high temperature, the adatom density on the surface is low, leading to the nucleation of a relatively small number of islands. As these islands grow, they produce large terraces on which new islands may nucleate. In contrast, at a low temperature, the adatom density is higher, and many islands are nucleated. Because the adatoms are less mobile at lower temperatures, new islands also nucleate on top of existing islands prior to the complete coalescence of the first layer. Both scenarios lead to three-dimensional growth and roughening of the surface. In general, as the temperature is raised, a smoother surface is produced, but a periodic growth strategy can sometimes produce an even smoother surface. If the temperature is lowered at the beginning of each layer, a large number of islands are nucleated. If the temperature is then raised following nucleation, the adatoms will be less likely to nucleate on top of existing islands, since the existing islands are smaller.

We produce this effect in KMC simulation, and then identify a reduced-order model that also captures this behavior. The KMC simulations are again performed on a 300×300 lattice, with $k_{\text{ads}} = 1\ s^{-1}$, $k_{\text{dif},i} = 10^{13} \exp\left((E_{\text{dif},0} + i\Delta E)/k_b/T\right)\ s^{-1}$, $E_{\text{dif},0}/k_b = 9000\ K$, $\Delta E/k_b = 2500\ K$, and i is the number of side bonds. The temperature is restricted to the range $400 \leq T \leq 500\ K$.

KMC simulations of this model are pictured on the left side of Figure 5.6. In addition to the roughness and adatom density of the previous example, the step density is also included now, since we wish to distinguish between configurations with equivalent coverage, but with differing island densities. At the highest temperature, persistent oscillations are observed over 10 layers, although at the end of each monolayer, the roughness does not go back to an atomically flat surface. In contrast, at the lowest temperature, the surface becomes much rougher, with highly damped oscillations.

As in the previous example, we select representative configurations from the KMC simulations to use as states in the reduced-order model. The 'x's in Figure 5.6 denote these configurations, which represent a range of fractional coverages, roughnesses, and step densities. Each configuration is then further simulated to produce surface with no adatoms, and with an adatom density of 4×10^{-3} site^{-1}, yielding a final configuration and state dimension of 80.

These configurations are then each simulated with KMC for 0.1 s at temperatures of 400 K, 450 K, and 500 K. A state matrix is generated for each of these three temperatures through the identification process, as described in the previous example. The right side of Figure 5.6 displays the evolution predicted by the model at these three temperatures. The comparison in Figure 5.6 is good, both qualitatively and quantitatively. Notice that the adatom density predicted by the model is noisy, since the original KMC prediction is also noisy.

We now simulate a periodic temperature strategy, in which the temperature is 400 K during the first half of each monolayer and 500 K during the second half. This strategy is compared to growth at a constant temperature of 500 K. The roughness at the end of each layer is lower under the periodic temperature profile, due to enhanced layer completion prior to nucleation of the next layer. Figure 5.7 shows Monte Carlo simulations, and simulations with the reduced-order model, using this periodic temperature trajectory. Both models predict that the roughness at the end of each layer is lower with the periodic temperature profile.

This periodic temperature strategy may produce smoother interfaces than constant temperature, but is not necessarily the best strategy. As in the previous example, we perform an optimization study, using the cost function of equation (5.1). In this study, we set $\alpha = 0$, with no penalty on the magnitude of the temperature, but instead vary β, the penalty on temperature *changes*. These results are shown in Figure 5.8, for three values of β: 10^{-3}, 10^{-4}, and 10^{-5}. When the penalty on temperature changes is high, it is best to hold the temperature near its maximum values, but as the penalty is reduced, a reduction in the cost is produced by lowering the temperature near the beginning of the second layer. Little benefit is obtained by lowering the temperature during the first layer, although for the lower values of β, the temperature is lowered slightly from its maximum value.

Table 5.2 summarizes the costs for the three values of β considered, and for constant, periodic, and optimal temperature trajectories. Notice that as the penalty on temperature changes is lowered, the periodic strategy considered nears the optimal cost.

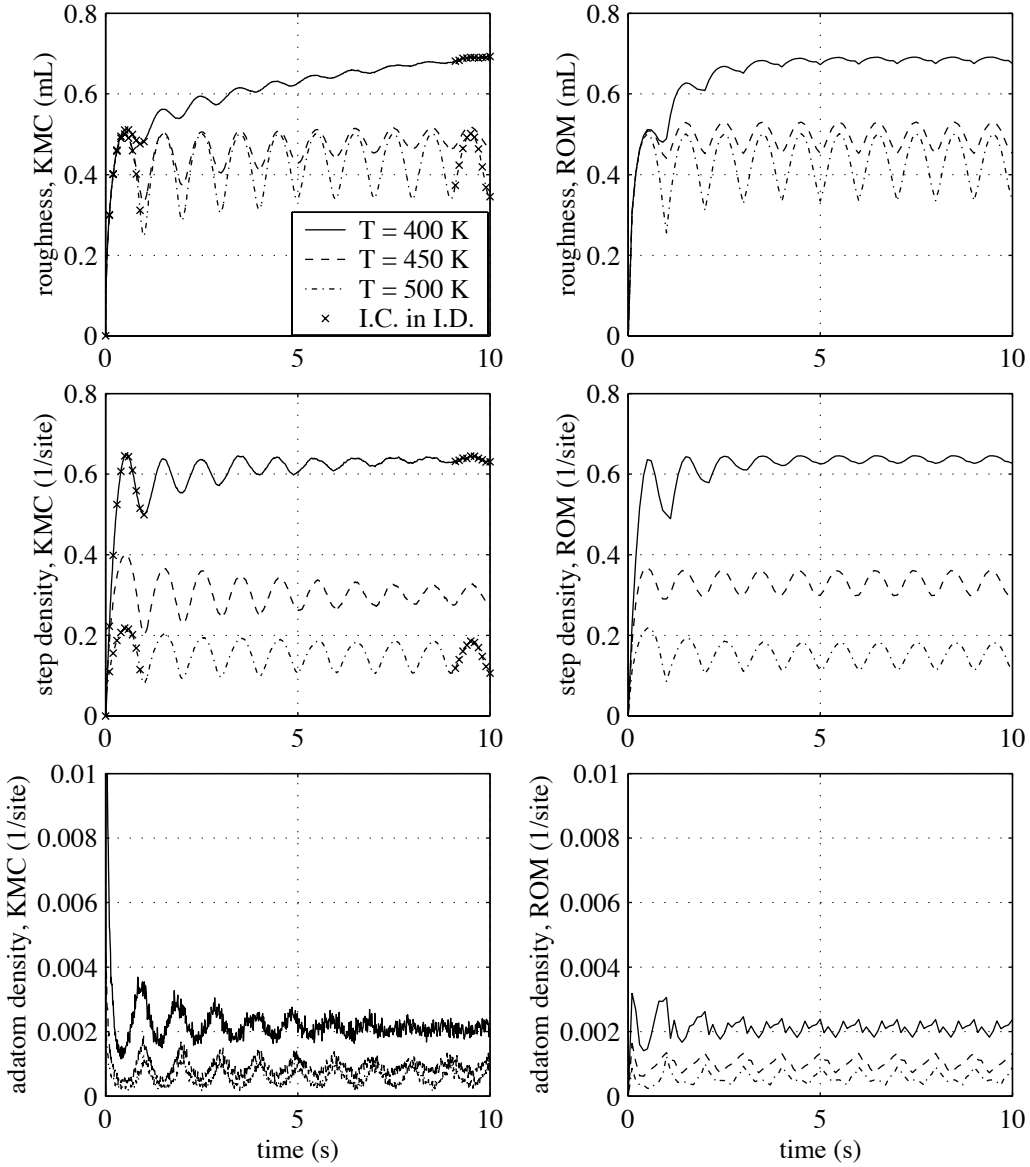


Figure 5.6: Comparison of roughness, step density, and adatom density predicted by KMC and ROM simulations. A 300×300 lattice is used for the KMC simulations. Configurations used as states in the reduced-order model are denoted by 'x's and are chosen to represent limiting behaviors of the system.

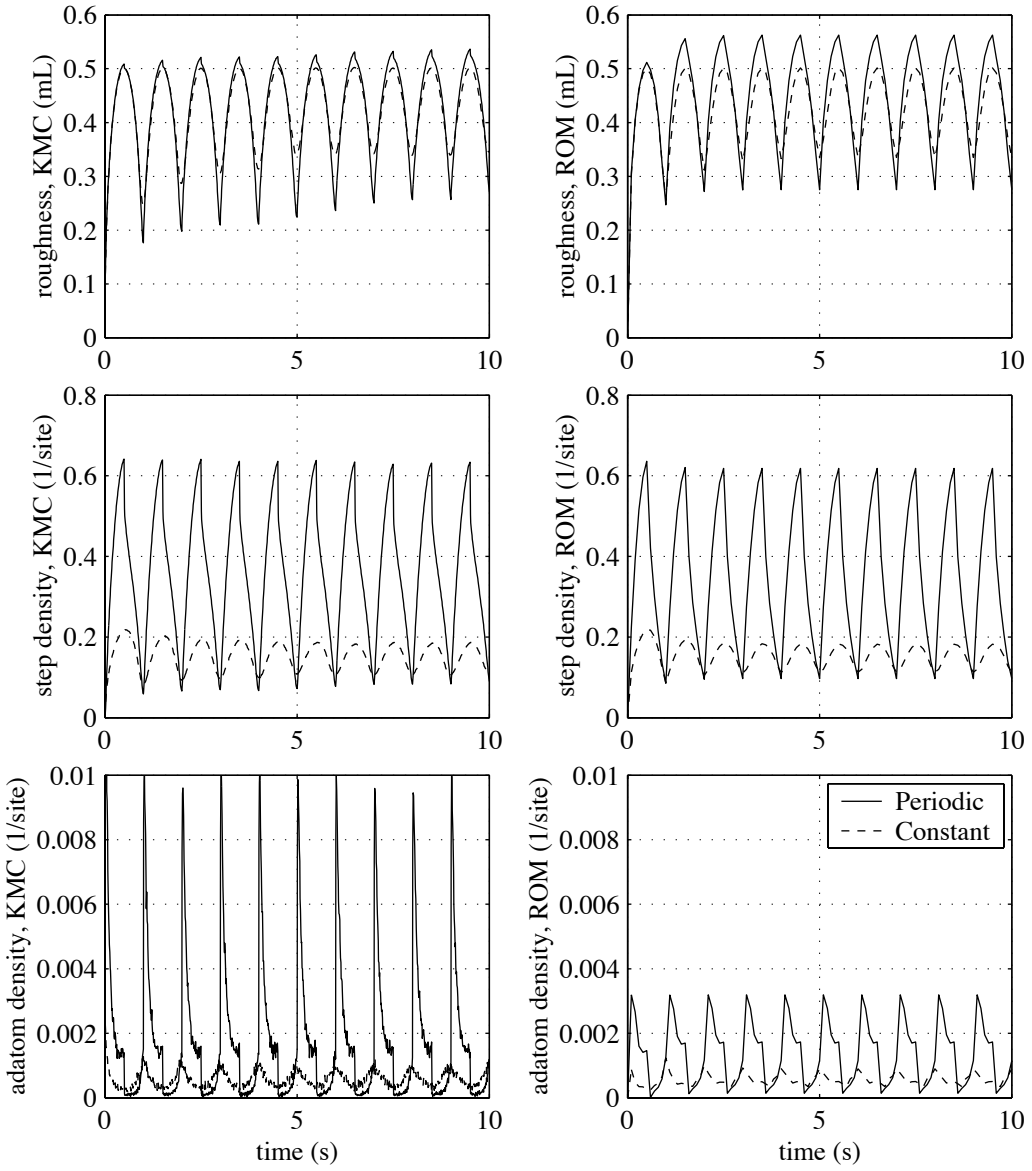


Figure 5.7: Simulations of growth at 500 K, and with a temperature of 400 K during the first half of each layer, and 500 K during the second half. The predictions of kinetic Monte Carlo simulations and the reduced-order model are compared.

	$\beta = 10^{-3}$	$\beta = 10^{-4}$	$\beta = 10^{-5}$
$T = 400\ K$	0.608	0.608	0.608
$T = 500\ K$	0.311	0.311	0.311
periodic T	1.77	0.422	0.287
optimal T	0.302	0.282	0.261

Table 5.2: Values of the cost function, equation (5.1), for various temperature profiles and values of β, with $\alpha = 0$.

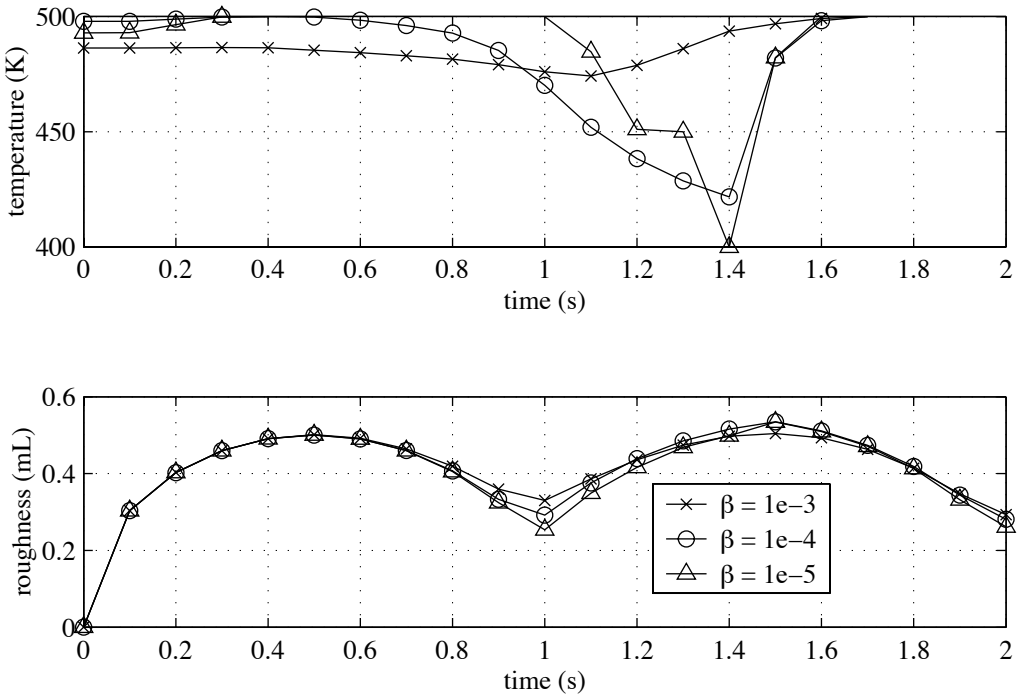


Figure 5.8: Optimal temperature profiles computed using the identified model and the cost function of equation (5.1), with $\alpha = 0$ and for three values of β.

5.3 Computational costs

In the previous sections we computed trajectories for the process parameters by first computing a low-order model and then applying a gradient-based optimization algorithm to the low-order model. However, this optimization algorithm does not exploit any specific feature of the reduced-order model—only the input-output map between process conditions and film properties is used. One could instead imagine using the KMC simulations in the optimization, eliminating the need for the low-order model. Even if the computational effort for KMC simulation is greater than for the reduced model, the cost of the system identification might be sufficiently high to make direct KMC optimization less expensive.

To compare the two approaches, we define C_{ROM} to be the total computational cost associated with identification of and optimization on a reduced-order model, and similarly define C_{KMC} to be the computational cost of the optimization using KMC simulation data. We do not take into account the inherent noisiness of KMC simulations, and instead assume that the optimization may be performed using individual realizations (not ensembles), which might yield an underprediction of C_{KMC}.

We further define c_{ROM} to be the typical computational cost of one second of simulation of the reduced-order model, c_{KMC} to be the typical cost of one second of KMC simulation, τ_{opt} to be the length of time in seconds over which

the optimization is to be performed, τ_{ID} to be the length of time in seconds for each simulation used in the identification, n_{opt} to be the number of simulations required for the optimization, and n_{ID} to be the number of simulations required for the identification, i.e., the produce of the state dimension and the number of different inputs used.

The total costs may now be expressed approximately in terms of these new variables:

$$C_{\mathrm{ROM}} = n_{\mathrm{opt}}\, \tau_{\mathrm{opt}}\, c_{\mathrm{ROM}} + n_{\mathrm{ID}}\, \tau_{\mathrm{ID}}\, c_{\mathrm{KMC}} \tag{5.2}$$

$$C_{\mathrm{KMC}} = n_{\mathrm{opt}}\, \tau_{\mathrm{opt}}\, c_{\mathrm{KMC}}. \tag{5.3}$$

This is only an approximation, since the computational cost per second of simulation time is not constant, and depends on the process conditions. The purpose of these expressions is to aid in understanding the factors that contribute to the total computational cost. Also realize that this comparison assumes that the reduced model is only used for a single optimization, even though it may be used repeatedly once identified.

We now consider the example of Section 5.2. A model with 80 states is identified, using three different inputs settings, yielding $n_{\mathrm{ID}} = 240$. Each simulation runs for $\tau_{\mathrm{ID}} = 0.1$ s. The value for c_{KMC} is taken from the fastest simulations, at 400 K, using chronological time in seconds to measure to the computational cost, such that $c_{\mathrm{KMC}} = 67$. The cost of the identification alone is thus 1600 s.

The temperature is optimized over $\tau_{\mathrm{opt}} = 2\ s$ of growth, which takes approximately $n_{\mathrm{opt}} = 2000$ simulations to converge. The computational demand of the reduced model is approximated based on the length of time required for the total optimization, such that $c_{\mathrm{ROM}} = 0.025$. These estimates yield computational costs for the optimization algorithm of 100 s and 270,000 s for the reduced model and KMC simulations, respectively. Adding in the cost of the identification, we arrive at $C_{\mathrm{ROM}} = 1700$ s and $C_{\mathrm{KMC}} = 270{,}000$ s. These costs are only estimates, and are intended to represent the order of magnitude of the computational cost. In this case the cost of optimizing on the KMC simulations is two orders of magnitude greater than optimizing on the reduced model.

Note first that the cost of the optimization on the reduced model is an order of magnitude less than the cost to identify the model. This is true primarily because $c_{\mathrm{ROM}} \ll c_{\mathrm{KMC}}$. In this limit, the ratio of the total costs is strongly dependent on the ratios of τ_{opt} and τ_{ID}, and n_{opt} and n_{ID}. When the cost of identifying the reduced-order model is much less than the cost of optimizing on the KMC simulations, we may view the reduced model as an efficient way of encoding the gradients observed in the KMC simulations, instead of computing them from KMC during every step of the optimization. As a final note we also mention that the issue of noise in the KMC realizations might actually be the biggest huddle to applying a gradient-based optimization algorithm, and that the linear least squares computation used in the identification is an effective method for determining gradients in the presence of noisy data.

Chapter 6

Model System: Germanium Homoepitaxy

During thin film deposition, the surface is randomly bombarded by atoms, which then attach to the surface. The stochastic nature of this incoming flux leads to a roughening of the surface. At high growth temperatures, the smoothening influence of surface diffusion counteracts the roughening, resulting in an atomically flat surface. Flat surfaces are often required in the manufacture of layered devices, but low growth temperatures may also be required to prevent dopant diffusion, to prevent degradation of a previously deposited layer, or to minimize manufacturing cost. It would be useful to be able to deposit smooth surfaces at low temperatures.

The conditions under which a thin film is deposited are generally held fixed throughout deposition. However, time-varying conditions may sometimes produce smoother surfaces than continuous growth at the average, minimum, and maximum conditions. This intentional modulation of temperature or flux can produce smoother surfaces when synchronized with the monolayer growth time [51]. Faster pulsing, as in a pulsed laser deposition (PLD) process, can also provide smoother surfaces than continuous growth by molecular beam epitaxy (MBE). These two processes were compared in experiment and simulation by Taylor and Atwater [58]. Simulations predict that the energetic effects of PLD contribute to smoothing, and that the time-varying flux alone would actually roughen the surface.

To separate the effects of pulsed flux from those of energetic flux, we deposit germanium in a molecular beam epitaxy process, generating pulsed flux by periodically blocking the wafer with a shutter. The chamber is equipped with a reflection high-energy electron-diffraction (RHEED) gun, which provides continuous real-time information throughout growth. We compare the RHEED signals associated with pulsed growth over a range of temperatures, growth rates, and pulse durations, and interpret the data in the context of atomistic surface models, existing STM studies, and various physical models of RHEED. Germanium was selected as our material system because it is relatively isotropic and has a very weak Schwoebel barrier, which simplifies our interpretation and comparison to models of surface evolution. We deposit germanium on highly oriented germanium substrates that are aligned with the (001) direction to minimize the number of steps due to substrate miscut.

6.1 Previous work

Scanning tunneling microscopy

Studies using scanning tunneling microscopy (STM) provide images of atomic arrangement [11, 42, 43]. The STM measurements cannot be performed during growth, but are used in "quench-and-look" experiments, in which growth is interrupted and the surface quickly cooled to obtain images associated with various points in time during the growth process. Particular studies have focused on island nucleation during low temperature growth [42], and surface smoothening during annealing [11].

Reflection high-energy electron diffraction

RHEED can be used as a real-time sensor to obtain information about the surface morphology as the film evolves. RHEED has been used to study the growth of Ge(001) over a range of growth parameters. In growth on highly oriented surfaces, persistent oscillations have been observed in the RHEED pattern that are correlated with the growth of individual atomic monolayers [1, 2, 7, 56]. On surfaces with higher miscut, oscillations are not observed, but decay in the RHEED pattern is associated with a roughening of the surface [9, 10].

RHEED interpretation

The diffraction of electrons from a crystal surface requires consideration of quantum mechanical phenomena, and is particularly dependent on the distribution of electrons in the crystal. However, in certain cases a kinematical approximation is valid, in which the incoming electrons are assumed to only scatter once before leaving the crystal surface [35]. In this approximation, the RHEED pattern is simply the Fourier transform of the autocorrelation function of the surface height. This interpretation makes comparison to simulated surfaces straightforward [41]. The kinematical approximation has been used in the past to interpret RHEED patterns of Ge(001) [9]. We continue to explore the validity of the kinematical approximation during growth of Ge and during periods in which the Ge flux is blocked and the surface is allowed to reorganize through diffusive processes.

Depending on the angle of the incident electron beam, electrons scattering from adjacent layers in the crystal may interfere constructively or destructively. Within the kinematical approximation, the dependence of the RHEED pattern on surface morphology has a particularly simple interpretation for the latter "out-of-phase" diffraction condition. The integrated intensity of the specular spot depends on the relative numbers of surface sites on even and odd layers. Scattering from adjacent layers interferes destructively, while sites differing by two levels add constructively. When the distribution of heights is known, surface coverage Θ may be directly tied to the intensity of the specular spot I. For example, in two-level growth $I(t)/I(0) = (1-2\Theta(t))^2$ [12], while on a statistical surface $I(t)/I(0) = \exp(-4\Theta(t))$ [9].

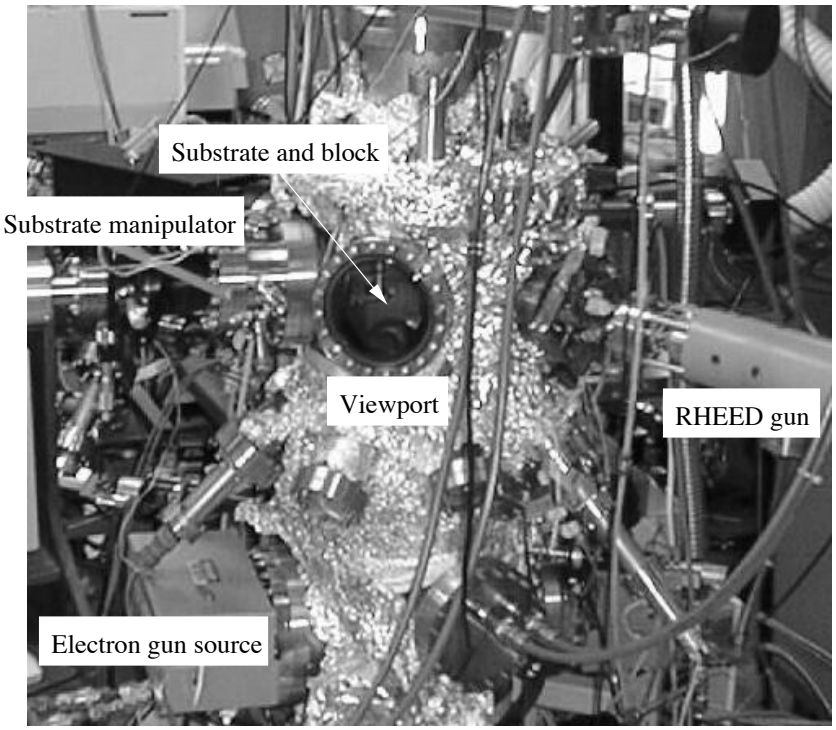


Figure 6.1: Photograph of the molecular beam epitaxy chamber.

6.2 Experimental apparatus

6.2.1 MBE chamber

Germanium deposition was performed in a molecular beam epitaxy (MBE) chamber, which is pictured in Figure 6.1. Key features relevant to the experiments include an electron-beam evaporative germanium source, a cryopump to obtain ultra high vacuum conditions, a tungsten filament sample heater, a quartz crystal thickness sensor, and a RHEED gun and phosphor screen for real-time analysis of the growing film. The germanium source is located in the base of the chamber, and is separated from the sample by a shroud and a shutter—when retracted, the molecular beam of germanium impinges on the substrate. A schematic of the interior of the growth chamber is shown in Figure 6.2. Figure 6.3 is a temperature calibration of the tungsten filament heater. A type-K thermocouple-instrumented silicon wafer was used in the calibration. Further detail on the MBE chamber may be found in [23].

6.2.2 RHEED

RHEED is the primary diagnostic in this work. The electron gun is directed at the growing surface at a glancing angle; the diffraction pattern from the surface is then visualized on a phosphor screen, which fluoresces when impacted by electrons.

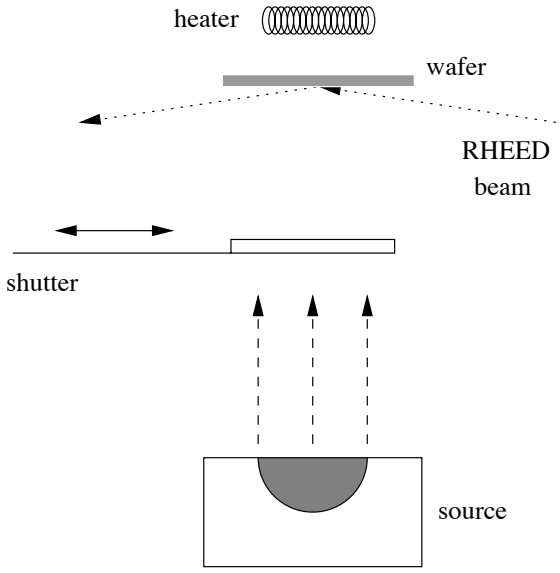


Figure 6.2: Schematic of the molecular beam epitaxy chamber.

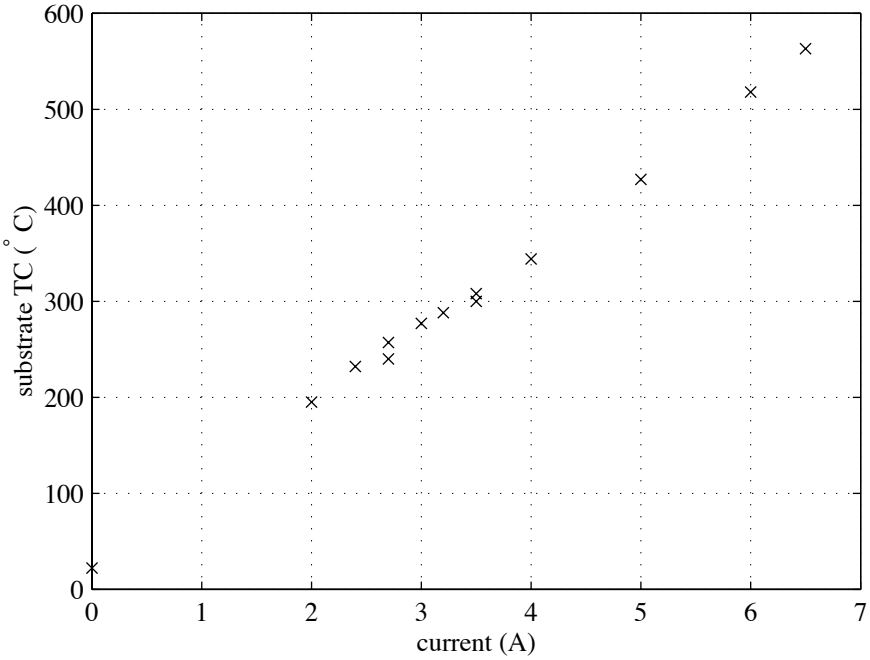


Figure 6.3: Temperature calibration of tungsten filament heater with a thermocouple-instrumented silicon wafer.

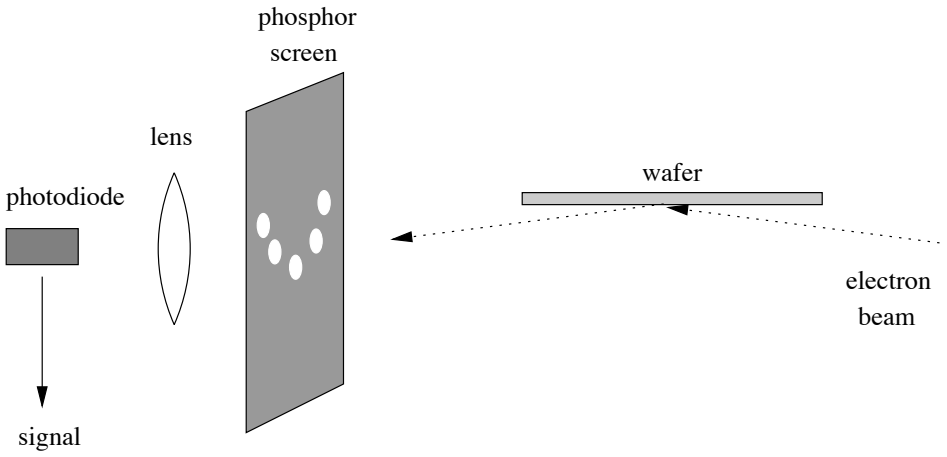


Figure 6.4: Layout and key features of the data acquisition system for RHEED. The diffraction pattern on the phosphor screen is collected here with a lens and photodiode, but may alternatively be captured with a camera.

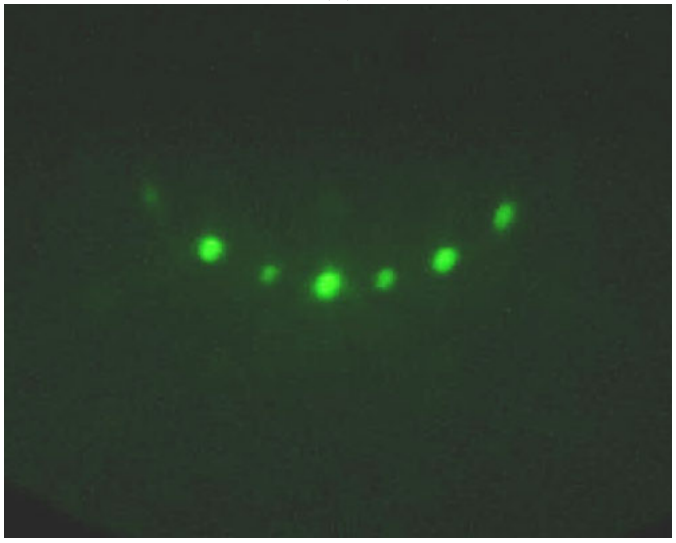
The image on the phosphor screen is collected by one of two methods. The first uses a CCD camera, which is linked to a framegrabber in a personal computer. We use the software package k-space from KSA to acquire and processes these images. With this method we may independently analyze the intensity of multiple features in the diffraction pattern. However, the rate of image acquisition is limited by the camera and framegrabber. When a higher acquisition rate is needed, we instead focus a particular diffraction feature onto a photodiode, and collect the output voltage directly into a personal computer. Figure 6.4 shows the key features in the sensing and data acquisition. The screen is mounted in a viewport on the wall of the chamber, and the photodiode and lens are located outside the chamber.

Typical RHEED patterns seen during our Ge growth are pictured in Figure 6.5. In Figure 6.5(a), the RHEED beam is oriented along the (110) crystallographic direction. In this configuration, we obtain a symmetric pattern in which the spots correspond to diffraction from atomic planes parallel to the surface. In our analysis, we are particularly interested in the central, or specular, spot, which results from reflection of the RHEED beam. Figure 6.5(b) is the RHEED pattern obtained after an in-plane substrate rotation of 3°. The specular spot is still visible in the center, but the other diffraction spots are now weaker and have been rotated down relative to the specular spot. This particular orientation minimizes multiple scattering of the electrons on the surface, which make RHEED interpretation more difficult.

6.2.3 Experimental procedure

Our germanium wafers are from Eagle Picher. They are lightly doped (n-type), and are specified to be oriented $0.05° \pm 0.02°$ from the (001) crystallographic direction. The wafers were originally 2" in diameter. However, we cleaved the wafers and

(a)



(b)

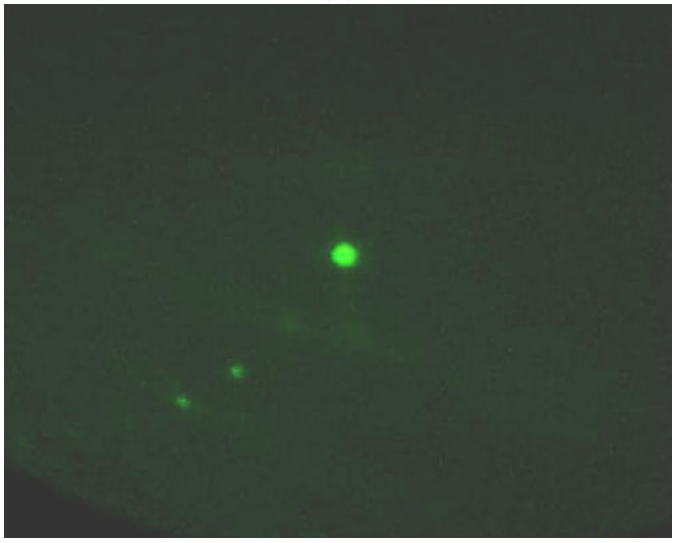


Figure 6.5: Typical RHEED pattern for Ge growth on a Ge(001) highly oriented wafer: (a) 0° off (110) azimuth; (b) $\sim 3°$ off (110) azimuth. The angle of incidence is 5° and the electron energy is 17 keV.

grew on smaller pieces.

Each sample was cleaned by sonicating it sequentially in acetone, methanol, and filtered water. The sample was then dried and placed in a UV-ozone cleaner for 10 minutes to oxidize the surface. Final cleaning steps include further sonication in water and a dip in 5% solution of hydrofluoric acid to make the surface hydrophobic. The sample is then blown dry and attached to the sample block. For more background on surface preparation for germanium, see [8].

The block is loaded into the MBE chamber through a load lock, and is baked at 250°C for several hours. A typical base pressure after this bake is 1×10^{-10} torr. The substrate is then heated to 450°C to drive off remaining impurities and to induce the (2×1) surface reconstruction, as seen by Bragg rods in the RHEED pattern. As a final step in our surface preparation, a 5000 Å buffer layer of Ge is deposited at temperatures of 500°C and a rate of 1 Å/s. At this point, we achieve a RHEED pattern like that shown in Figure 6.5.

A typical growth pressure is 3×10^{-9} torr. We deposit films at rates from 0.05 to 1 Å/s at temperatures from ambient to 560°C. We are primarily interested in low-temperature growth between 200 and 300°C and deposit tens of Angstroms at these temperatures. Between each low temperature deposition, we raise the temperature to 500°C and grow a 1000 Å buffer layer at 1 Å/s to return to a smooth surface, as indicated by RHEED. The typical electron energy for RHEED is 17 keV, with an azimuthal angle from (110) of 3° and an angle in the out-of-phase diffraction conditions near 5°.

6.3 Experimental results

6.3.1 Submonolayer deposition

To determine how the surface will respond to continuous and pulsed flux, we investigate surface evolution during the deposition of half a monolayer, and its subsequent reorganization following deposition. For this purpose five temperatures (230, 260, 280, 290, 305°C) and two growth rates (0.05 and 0.4 Å/s) are used. Growth proceeds for either 2 or 16 s, depending on the growth rate, followed by 40 s during which the flux is blocked by a shutter. Plots of intensity versus time have the same general features as the data in Figure 6.6. Instead of showing all of the curves, we plot in Figure 6.7 the intensity at the end of growth and at the end of the 40 s recovery period.

Notice first that the relationship between the four curves at each temperature is nearly independent of temperature. The lowest intensity is associated with growth at the higher growth rate at the end of the deposition period. After these surfaces are allowed to recover for 40 s, the intensity increases substantially. In fact, across the range of temperatures, the intensity after recovery for the higher growth rate is similar to the intensity of the lower growth rate at the end of the deposition period. This collection of data is consistent with an island nucleation and coarsening interpretation. At a fixed temperature and at the higher growth rate, the island density should be greater, resulting in a larger initial intensity

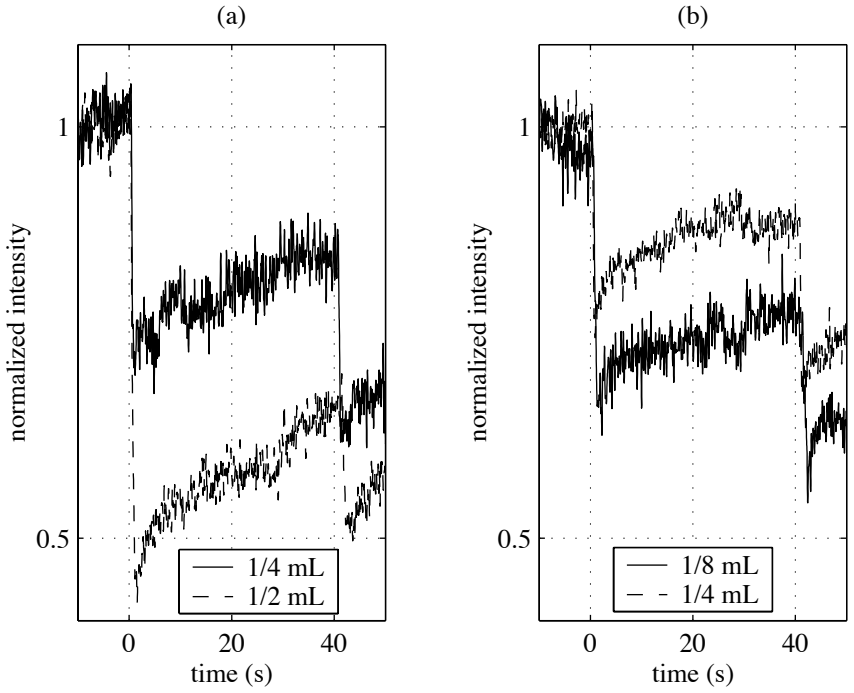


Figure 6.6: Normalized intensity of the specular spot during growth and recovery at 290°C for various submonolayer doses and growth rates: (a) 0.8 Å/s, (b) 0.4 Å/s.

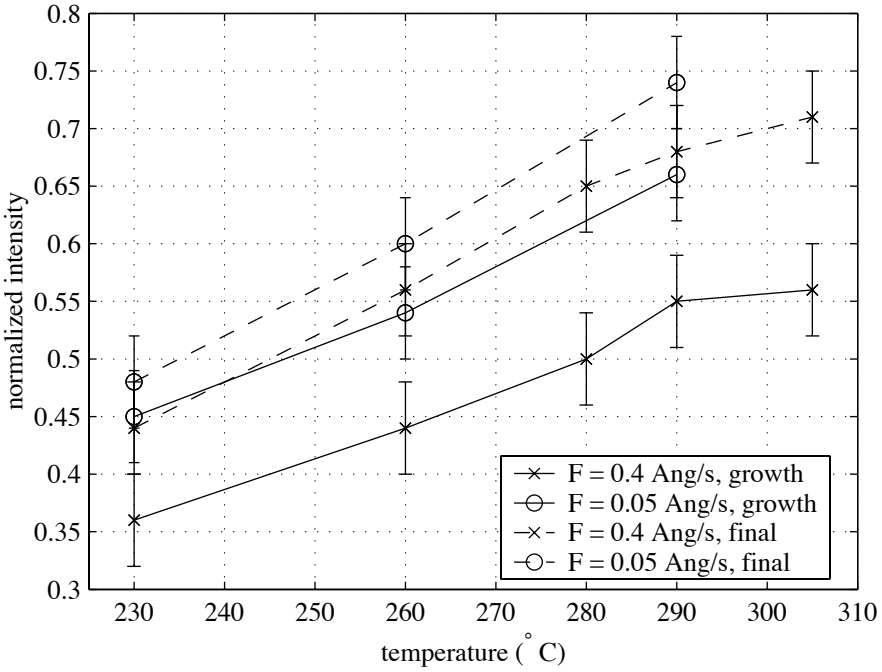


Figure 6.7: Intensity of the specular spot at the end of 1/2 mL of growth, and after 40 s of recovery.

drop. However, the coarsening rate during the recovery period is also higher since the islands are small and close together.

Error in the intensity measurements is attributed to uncertainties in the temperature and the amount of material deposited per pulse. The temperature calibration in Figure 6.3 shows a linear relationship between heater current and temperature, with some variation visible in the range of temperatures considered in the parameter study. The temperature variations are attributed to changes in contact resistence due to thermal expansion, which we expect to be greater in the instrumented wafer because it is larger and thicker than the germanium wafers used during growth. Nevertheless, we observe a temperature variation of $\pm 10°$C near 250°C, which we take as our temperature error. We translate this into intensity using the data in Figure 6.7, observing a typical slope in intensity versus temperature of $0.003/°$C, or ± 0.03 in normalized intensity for $\pm 10°$C.

The pulse time is precisely regulated by a computer-controlled shutter, so error in the coverage per pulse stems from the the electron-gun source. This variation is assessed using the data presented later in Figure 6.10 of Section 6.3.3, in which intensity oscillations indicate the growth of an individual monolayer. An analysis of three runs at identical conditions shows a standard deviation of 0.06 mL per monolayer of deposition, or 0.03 mL for our 1/2 mL pulses. We estimate the sensitivity of intensity to coverage using the data shown in Figure 6.6 to obtain an intensity uncertainty of 0.03 for the 1/2 mL pulses. Combining the two sources of error, our standard deviation for the intensity measurements is 0.04, as reflected in the error bars of Figure 6.7.

6.3.2 Multilayer growth

The study of individual submonolayer pulses suggests that a slower instantaneous flux rate provides a smoother surface after a single pulse. We next present RHEED data for a series of pulses, resulting in the growth of multiple layers. Figure 6.8 shows the RHEED intensity during pulsed growth at 305°C with different pulse lengths and growth rates, for a total of 10 pulses. In (a) the flux is 0.2 Å/s. The shutter is opened each cycle for 2 s, and then closed for 40 s. After the completion of 10 cycles, the quartz crystal monitor records a total deposition of 4 Å. In Figure 6.18(b), the flux remains unchanged, but the pulse length is increased to 4 s, resulting in 10 Å of deposition. Finally, in Figure 6.18(c), the flux is increased to 0.8 Å/s, with pulse length 0.5 s and a total deposition of 3 Å. During the first pulse, the intensity seems consistent with the cubic lattice model and our previous interpretation of RHEED. The initial decay of (a) is less than (b) or (c), due to larger coverage in (b) and growth rate in (c). However, after multiple cycles, the intensity approaches a steady state value that is not sensitive to growth rate or coverage in the range of conditions accessible to us. Instead, at the end of 10 cycles, the intensity is near 0.75 in all cases.

The steady-state intensity does not exhibit a strong dependence on growth rate or coverage, but is dependent on the growth temperature. Figure 6.9 show RHEED intensity for growth at 230°C. At this lower temperature, the steady state intensity

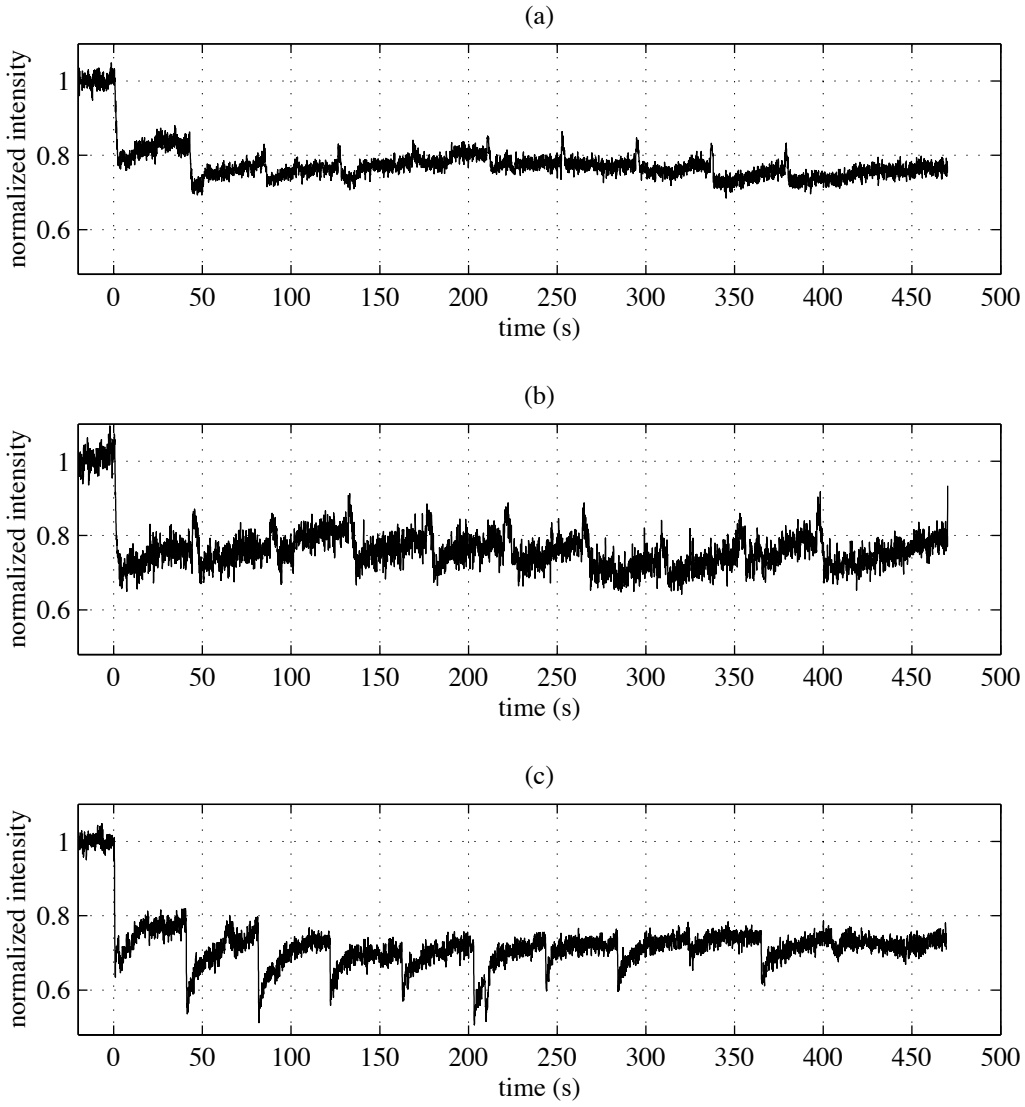


Figure 6.8: Integrated intensity of the spectral spot for pulsed growth at 305°C. In all three cases, 10 pulsing cycles are executed: (a) 4 Å are deposited in pulses of 2 s on, 40 s off at 0.2 Å/s; (b) 10 Å in pulses of 4 s on, 40 s off at 0.2 Å/s; (c) 3 Å in pulses of 0.5 s on, 40 s off at 0.8 Å/s.

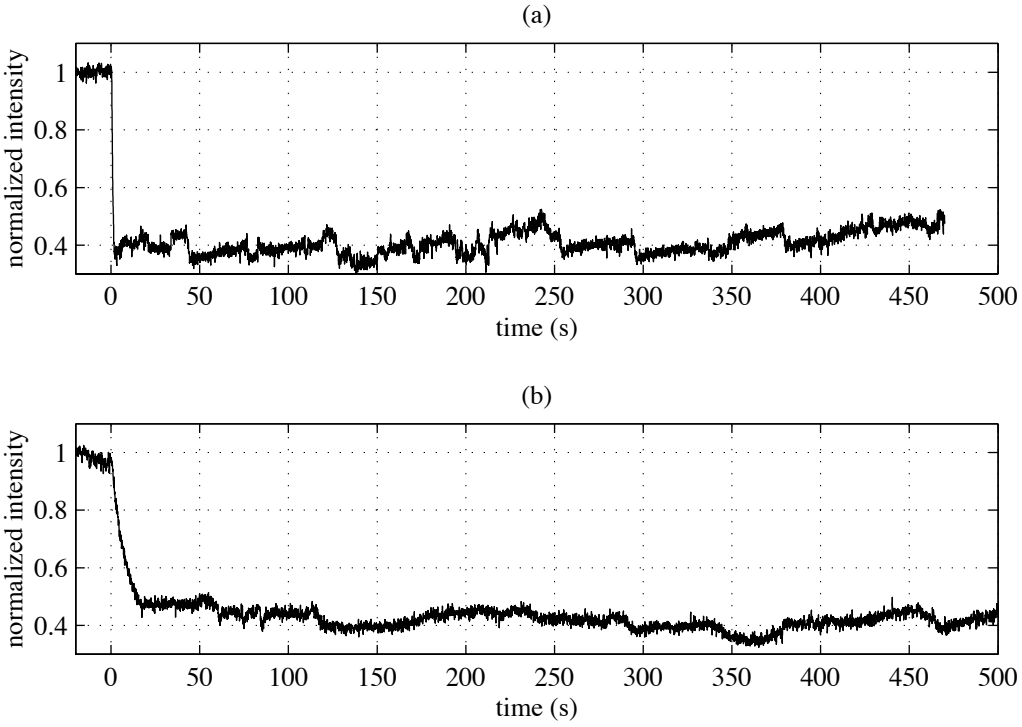


Figure 6.9: Integrated intensity of the spectral spot for pulsed growth at 230°C. In both cases, 10 pulsing cycles are executed: (a) 12 Å are deposited in pulses of 2 s on, 40 s off at 0.4 Å/s; (b) 7 Å in pulses of 16 s on, 40 s off at 0.05 Å/s.

Run	Description	α	β
1	Constant growth	0.0126	0.0138
2	Short recovery	0.0113	0.0142
3	Constant growth	0.0094	0.0128
4	Long recovery	0.0108	0.0163
5	Constant growth	0.0143	0.0178

Table 6.1: Exponential fit to $\Delta I = \alpha e^{\beta t}$.

has decreased near 0.4–0.5. In Figure 6.9 (a), 12 Å are grown in 10 cycles of 2 s on and 40 s off at a flux of 0.4 Å/s, while in (b), the flux is lowered to 0.05 Å/s, the growth time is raised to 16 s, for a total of 7 Å.

The intensity shows a strong dependence on the growth temperature, but appears to quickly reach a steady-state value that is independent of growth rate or thickness. While the intensity for submonolayer coverage is dependent on growth rate and coverage, this dependence is not observed in multi-layer growth.

6.3.3 Synchronized pulsing

As reported previously in the literature [1, 2, 7, 56], we have observed oscillations in the spectral spot intensity at glancing angles of incidence. The period of the oscillations is similar to the monolayer frequency, when correlated with the film thickness measurement obtained with a quartz crystal monitor. The oscillations are most pronounced at glancing angles less than 1°, far from 5° angle at which we study submonolayer pulsed growth. It is difficult to imagine that these well-defined persistent oscillations are occurring due to a periodic mechanism that is not the monolayer frequency. However, as observed by Aarts [2] with photoemission measurements, the occurrence of persistent oscillations in RHEED do not necessarily indicate smooth layer-by-layer growth. Berrie et al. [7] observe bimodal oscillations, which they attribute to a bilayer growth mechanism. However, bimodal oscillations are also reported to exist due to dynamical scattering [38].

We show in Figures 6.10 and 6.11 five growth runs performed sequentially, separated by only by the growth of a 1500 Å buffer layer at 450°C. Growth occurs at 270°C at a rate of 0.3 Å/s. The runs are labeled as Run 1–Run 5 in the order in which they were performed. In all runs, growth proceeds for 70 s, but in Runs 2 and 4, growth is interrupted at the end of each of the first five layers (as measured real-time by the RHEED oscillations) for periods of 5 s and 60 s, respectively. Substantial recovery in the intensity can be seen during these pauses, particularly during the long pauses of Run 4. We look for changes in the signal following this pulsing by comparing the oscillations after the pulsing sequence to the oscillations of Runs 1, 3, and 5. The amplitudes of the oscillation are plotted in Figure 6.12 as a function of the *growth* time. We also fit the decay in the amplitude to an exponential of the form $\Delta I = \alpha e^{\beta t}$ for each of the five runs. The values of the parameters are given in Table 6.1. The outlying values for both parameters are associated with the continuous runs, such that the decay

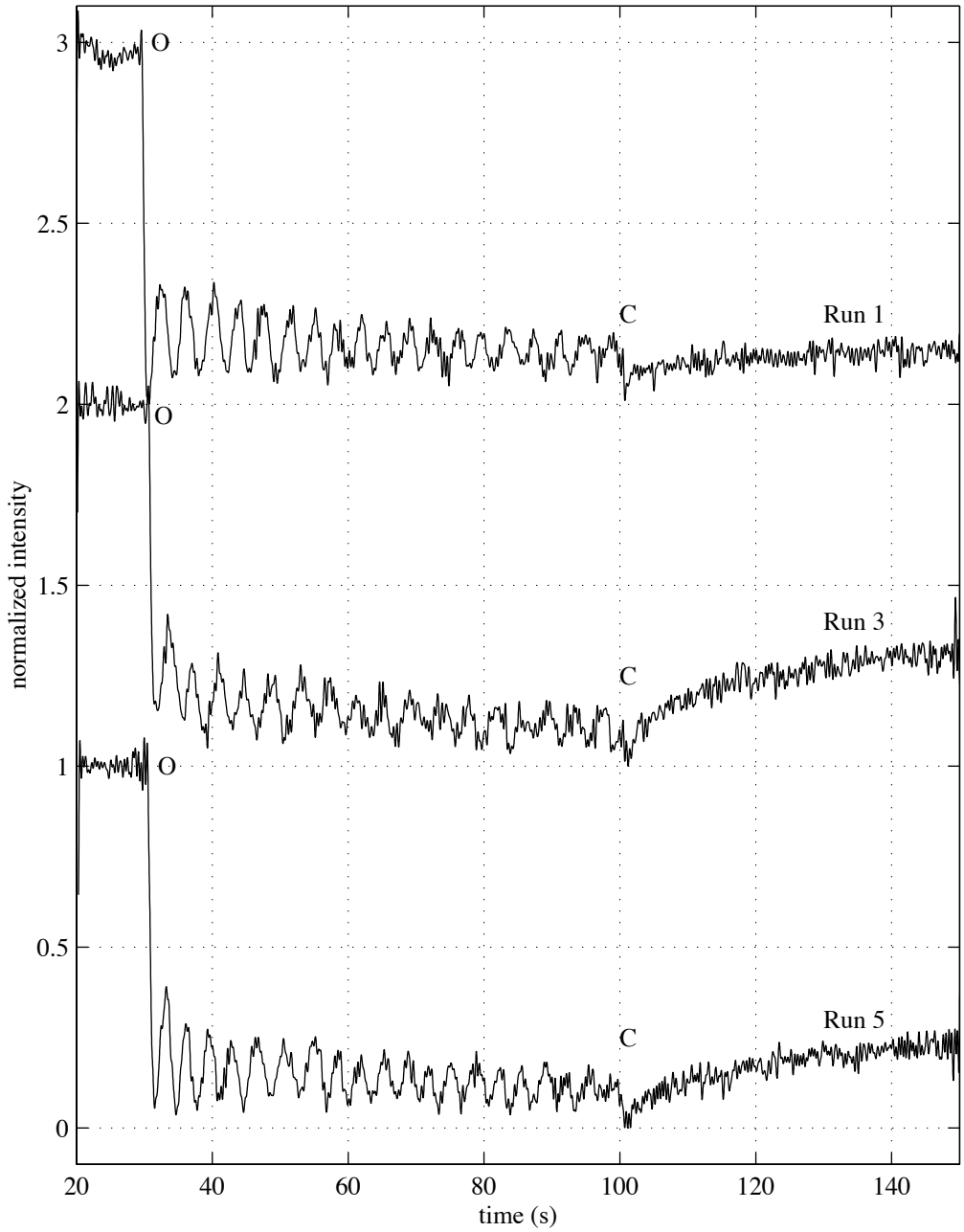


Figure 6.10: Germanium growth at 0.3 Å/s at 270°C. The 'O' denotes an opening of the shutter, and the 'C' denotes closing. Runs 1, 3, and 5 were performed under identifical conditions. Displacements in the intensity curves are arbitrary.

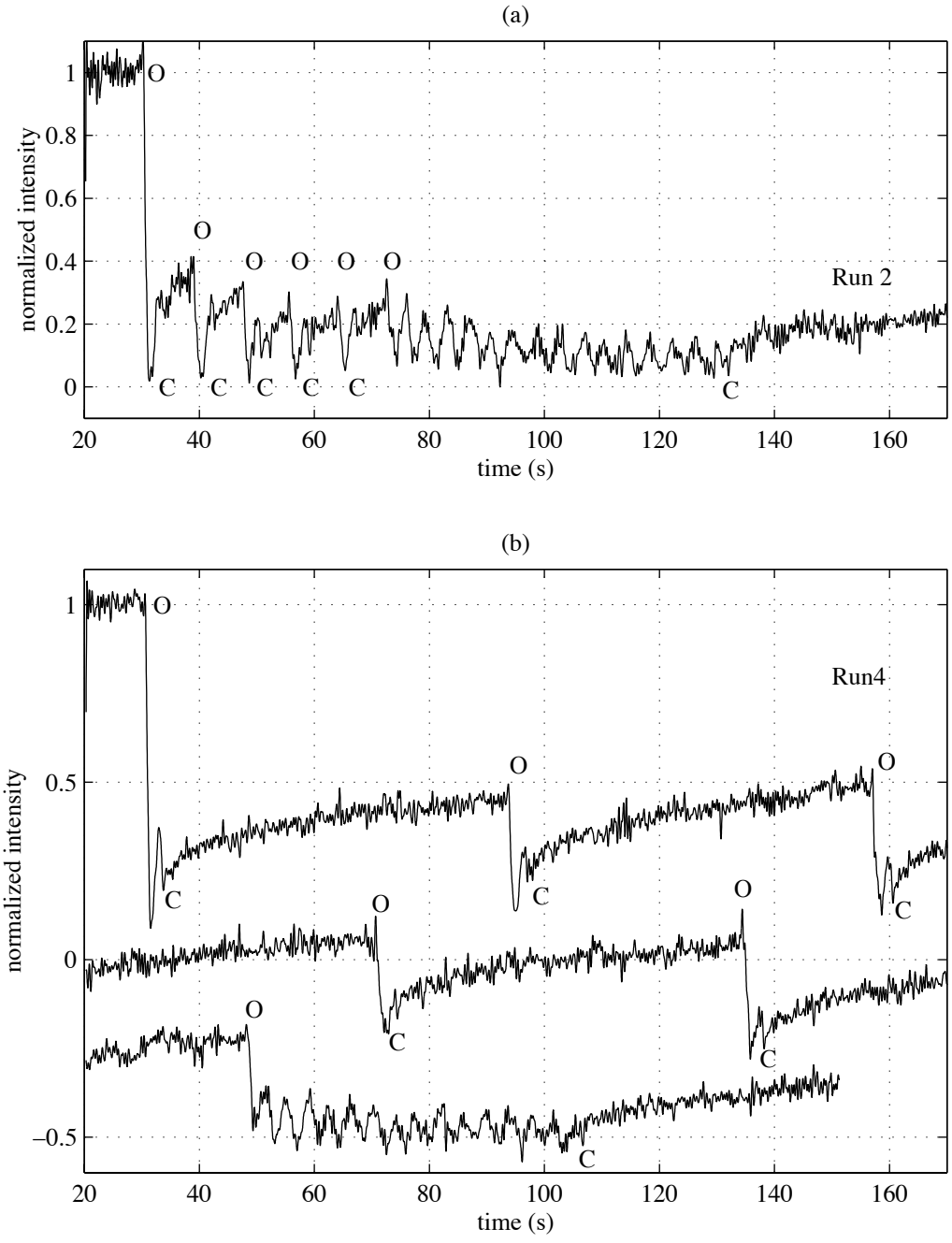


Figure 6.11: Germanium growth at 0.3 Å/s at 270°C. Growth is interrupted at the conclusions of each of the first five layers, for (a) 5 s; (b) 60 s. The 'O' denotes an opening of the shutter, and the 'C' denotes closing. Displacements in the intensity curves are arbitrary, and the curve for Run 4 in (b) is broken up into 3 segments, with the initial segment plotted at the top.

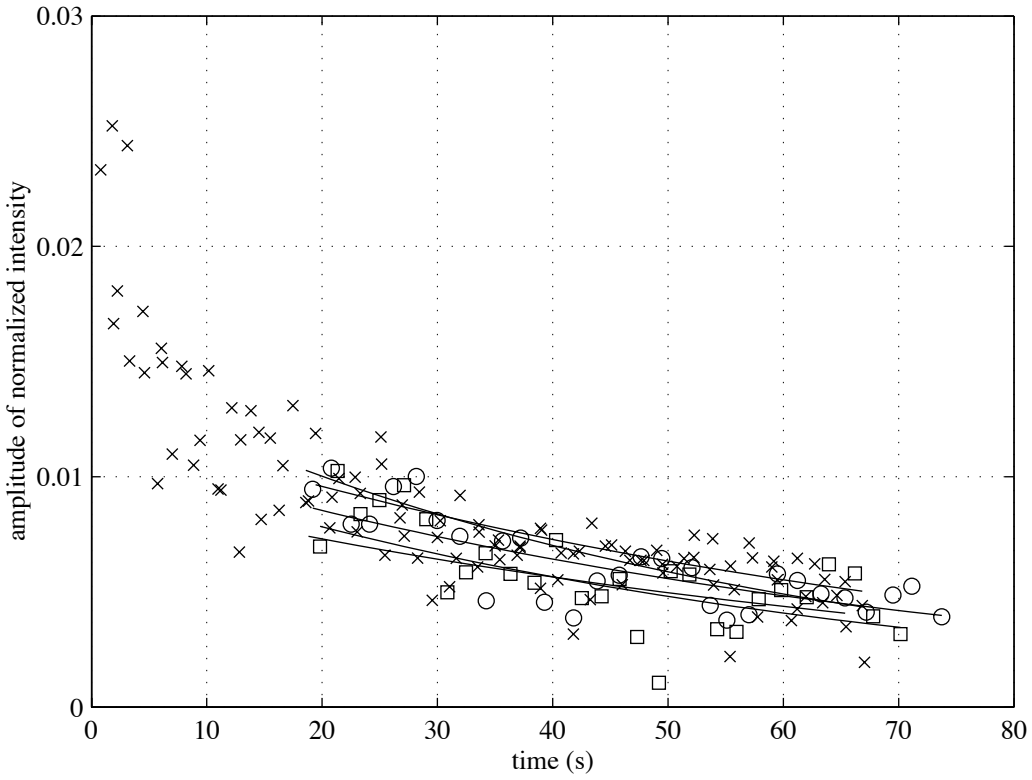


Figure 6.12: Amplitude of the intensity oscillations of Figures 6.10 and 6.11. The 'x's denote the three continuous growth runs, the 'o's are associated with the short pauses of Run 2, and the squares with the long pauses of Run 4. The decay associated with pulsed growth is indistinguishable from continuous growth. The parameters for the exponential fits are given in Table 6.1.

	α	β
mean	0.0121	0.0148
standard deviation	0.0025	0.0026

Table 6.2: Statistics for the exponentials fit to Runs 1, 3, and 5.

associated with the pulsed growth cannot be distinguished from the continuous growth. This is also reflected in the statistical analysis shown in Table 6.2, in which the parameters for the pulsed growth are contained within one standard deviation of the parameters for continuous growth. The pulsing seems to have little effect on the subsequent evolution of the surface, or at least cannot be seen in the intensity of the RHEED specular spot. This is consistent with Aarts' earlier statement that RHEED oscillations may be a better measure of film thickness than of surface morphology [2].

6.4 Comparison with simulation

6.4.1 Interpretation of RHEED signal

We have examined the validity of the kinematical approximation in Ge growth at low temperatures, using the camera and framegrabber to simultaneously analyze various features in the diffraction pattern. The decay in the intensity of the specular spot during low temperature growth is shown in Figure 6.13. As the temperature is lowered, the decay approaches the value associated with a statistical surface. We expect this limit at low temperature, since surface diffusion becomes negligible and the random deposition of atoms dominates the evolution. This exponentially decay has previously been observed by Chason et al. [9]. Also notice that after an initial decay, the RHEED intensity reaches a steady state value. We see this steady-state intensity over the range of temperatures studied, and observe that it is not consistent with the two-level or stochastic interpretations mapping intensity to coverage.

In our interpretation of the specular spot intensity, we must be sure that other diffraction features are not impinging on the specular spot and artificially contributing to the measured intensity. We investigate this issue by collecting line scans through the specular spot during growth at 125°C in the same run shown in Figure 6.13. A horizontal line scan reveals the contribution of the background to the measured intensity, while a vertical line scan shows the relative intensity of the specular spot and the Bragg rods, which might impinge upon the specular spot during growth. Line scans at the beginning and end of growth are given in Figure 6.14. In both scans, the intensity of the specular spot is large, while the surrounding intensity is within a few bits of zero. We conclude that even at the lowest temperature growth considered, the intensity of the specular spot is not corrupted by other diffraction features. However, in our photodiode measurements of the spectral spot intensity, we continue to measure and subtract off the background intensity for each growth run.

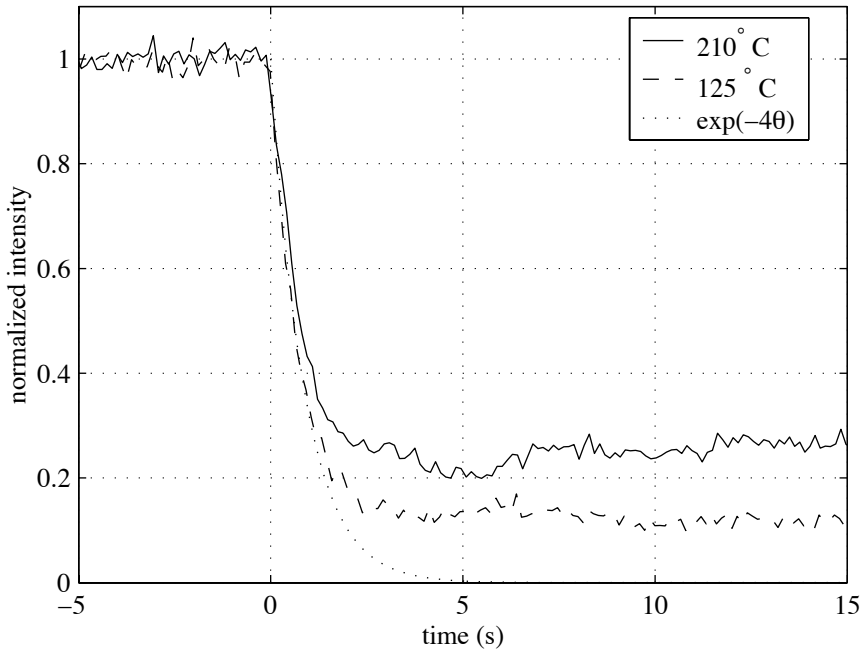


Figure 6.13: Decay of the integrated intensity of the RHEED specular spot during growth at 125°C and 210°C at a rate of 0.4 Å/s. Growth proceeds for 10 s, after which the shutter is closed. The decay is compared to the decay for stochastic growth for an equivalent monolayer coverage Θ, using 1.4 Å = 1 mL. As the temperature is lowered, the decay approaches the stochastic limit, but retains a small nonzero steady-state component for the temperatures considered.

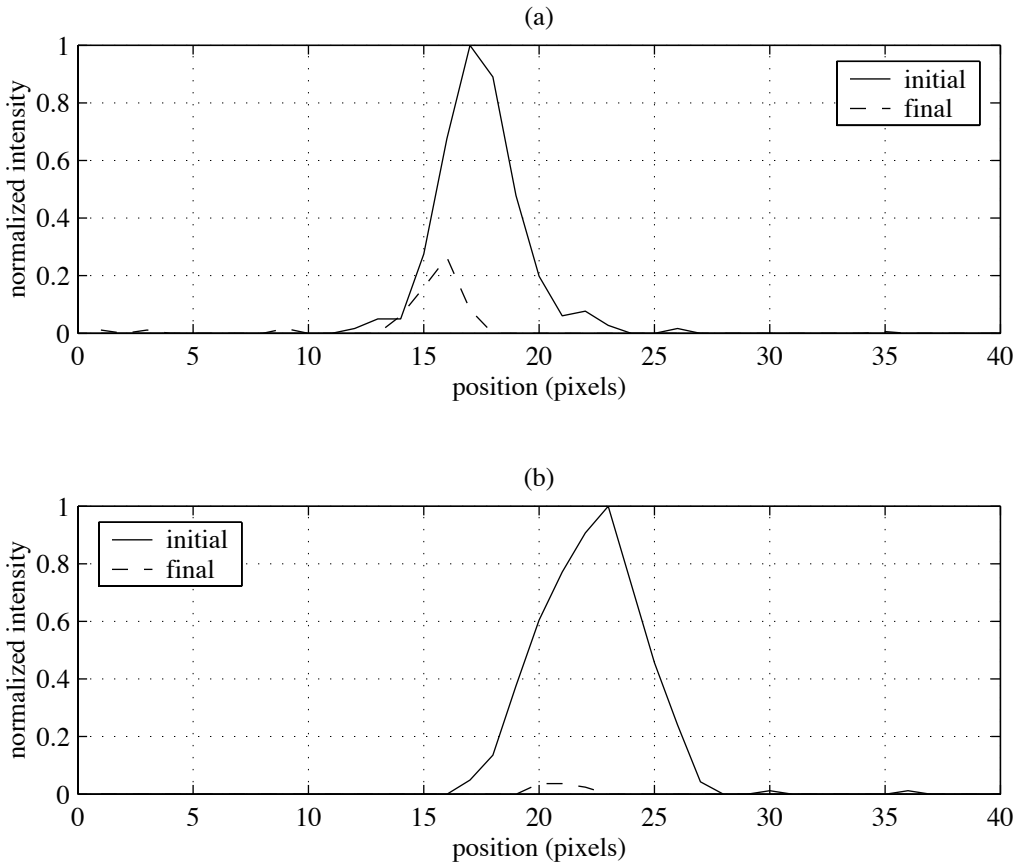


Figure 6.14: Line scans through the specular spot for the growth at 125°C shown in Figure 6.13: (a) line oriented vertically, along the Bragg rods; (b) line oriented horizontally, perpendicular to the Bragg rods. The Bragg rod and background do not contribute significant intensity to the measured specular spot intensity.

We now continue on with the kinematical interpretation, and consider the evolution of the specular spot first during growth, and then during subsequent recovery, as pictured in Figure 6.6. The main features in this plot are the decay in intensity during growth and the subsequent signal recovery. This recovery is not consistent with a straightforward mapping between surface coverage and signal intensity. Post-growth, the coverage is not changing, but the intensity is increasing. During this recovery phase, we expect the surface atoms to rearrange through adatom diffusion and detachment from islands, leading to island coarsening.

In the kinematical approximation, the RHEED pattern is ideally the Fourier transform of the surface autocorrelation function. However, imperfections in the measurement system result in a smoothening of the features, primarily via divergence of the electron beam. In some situations, this loss of spatial coherence is modeled by convolving the RHEED pattern with a Gaussian instrument response function. The width of the Gaussian corresponds to a critical distance on the surface at which the electron beam is no longer coherent [24]. The effect of convolution with the Gaussian eliminates contributions to the RHEED pattern that originate from atoms at distances greater than the transfer width. However, in our interpretation, we consider a nominal intensity resulting from pure reflection, which is diminished when electron waves interfere destructively from layers differing by an atomic layer. In this interpretation, we do not wish to eliminate the contribution from sites that are far apart. Instead, we should do exactly the opposite and subtract out destructive interference only from sites that are *within* the transfer width. This interpretation is clearly tied to a very different view of the RHEED specular spot intensity, which is that the intensity decay is proportional to the density of steps on the surface [54]. If we consider destructive interference only within some region surrounding the step edges, we obtain the same result. A consistent picture includes two limits: when steps are close together, the interference model is valid, but when terraces between steps are large, the step density model is appropriate. In the intermediate regime, we expect intensity to be inversely correlated with step density.

6.4.2 Simulation of experimental conditions

Submonolayer deposition

We now compare the experimental data shown in Figure 6.7 with kinetic Monte Carlo simulations of surface evolution. We assume a cubic lattice with unit cell 1.4 Å and with activation energies for diffusion and detachment based on a nearest-neighbor bond-counting scheme, with adsorption rate $k_{\text{ads}} = F$, diffusion rate $k_{\text{dif},i} = \nu \exp\left(-(E_{\text{dif},0} + i\Delta E)/(k_b T)\right)$, where F is the Ge flux, ν is a vibrational frequency, $E_{\text{dif},0}$ is the activation energy for adatom diffusion, i is the number of nearest-neighbor side bonds, ΔE is the additional activation energy associated with each side bond, k_b is Boltzmann's constant, and T is the substrate temperature. In the simulation study, we use the standard vibrational frequency of 10^{13} s^{-1}, and consider various values of $E_{\text{dif},0}$ and ΔE.

We also note that the crystal lattice of germanium is a diamond lattice, not a cubic lattice. However, cubic lattice simulations capture many important features of surface evolution, such as island nucleation and step propagation. In the cubic lattice model, $k_{\mathrm{dif},0}$ is the rate of adatom diffusion, $k_{\mathrm{dif},1}$ is the rate of dimer dissociation, and $k_{\mathrm{dif},2}$ is the rate of atom detachment from compact clusters. Consequently, one can expect to capture the main features of Ge(001) surface evolution that are described by the relative rates of adsorption, surface diffusion, and detachment. Altered rates of surface diffusion for adatoms diffusing up or down steps are not included in the simulations of germanium. Ehrlich-Schwoebel barriers at steps may eventually cause mound formation in very thick germanium films [11], but are unlikely to significantly impact surface evolution for films of only a few atomic layers [42]—the additional energy barrier at steps has been reported as only 0.024 eV.

In the kinetic Monte Carlo simulations we use step density for comparison with experimental intensity measurements. As discussed in Section 6.4.1, the step density is assumed to be monotonically decreasing with increasing intensity, for equivalent coverage and temperature. Furthermore, Figure 6.6 is used to estimate the magnitude of step densities that should be resolvable with RHEED. In Figure 6.6, coverages varying by a factor of two have been deposited under otherwise identical conditions. We thus assume that the island density between each pair is fixed, and therefore that a ratio of coverages of 2 yields a step ratio of $\sqrt{2} \sim 1.4$. For the similar growth conditions of the parameter study, step densities differing by this ratio are deemed to be visible in the RHEED intensity.

We now present the results of the kinetic Monte Carlo simulations for the experimental conditions of the data in Figure 6.7. Values of $E_{\mathrm{dif},0} = 0.60$, 0.65, and 0.70 eV and ΔE of 0.20, 0.25, and 0.30 eV are used in this study. While additional parameter values were explored, these values of ΔE provided the best agreement, while $E_{\mathrm{dif},0}$ was centered around a previously reported value [42]. Results for two temperatures are shown in Figures 6.15 and 6.16.

At both temperatures shown we observe a similar relationship among the four curves. At the highest bond energy, the step edge density after growth at 0.4 Å/s does not approach the value for growth at 0.05 Å/s, while for $\Delta E = 0.25$ eV, greater recovery is observed. At this intermediate bond energy, we near the previously identified ratio of 1.4 between the curves, so we are less confident that the intensities would not be the same. However, we get the best agreement between the experimental results of Figure 6.7 and the simulation results for $\Delta E = 0.20$ eV, regardless of the adatom diffusion rate specified by $E_{\mathrm{dif},0}$. We conclude that the recovery measurements are highly sensitive to ΔE, are most consistent with a value of 0.20 eV, and are not very sensitive to $E_{\mathrm{dif},0}$. At smaller values of ΔE, the simulations exhibited extreme fluctuations at the temperatures of interest, because the detachment rate is extremely high and islands do not remain compact. Consequently, we take 0.20 ± 0.05 eV to be an upper bound on ΔE, based on the parameter study.

An activation energy for adatom migration has been determined by a previous study by VanNostrand et al. [42] to be 0.65 eV. This value was extracted from

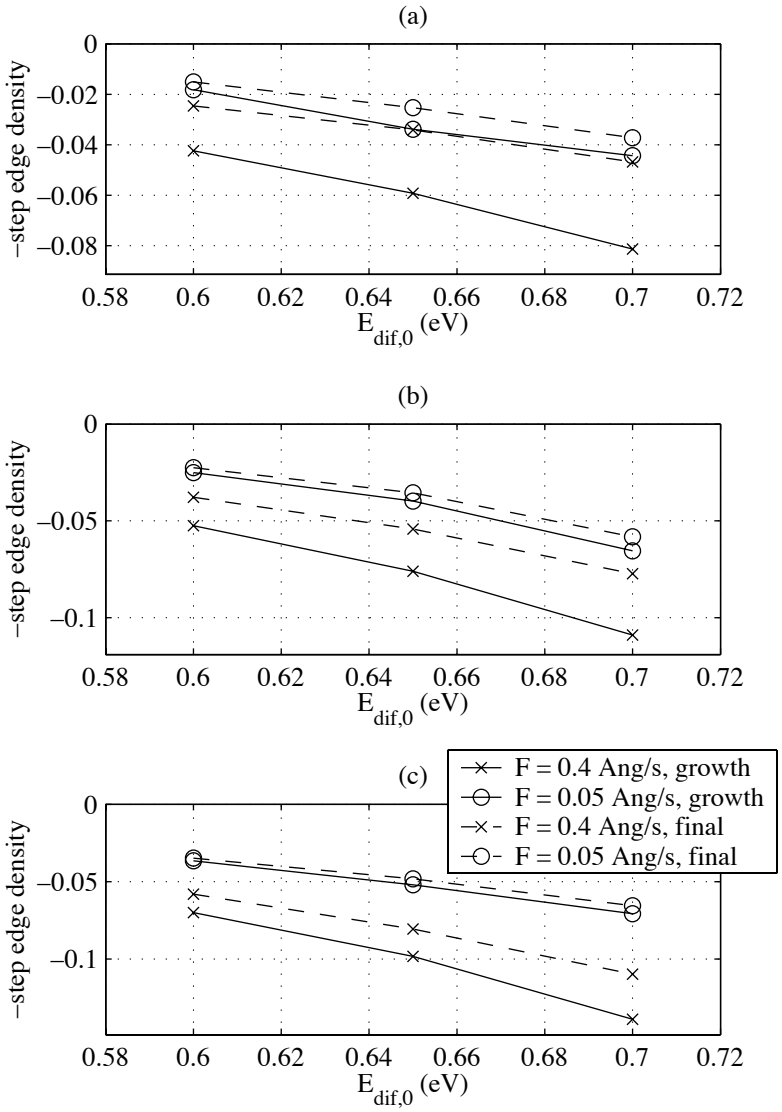


Figure 6.15: Simulated step edge density at 230°C immediately following growth, and after 40 s of recovery: (a) $\Delta E = 0.20$ eV, (b) $\Delta E = 0.25$ eV, and (c) $\Delta E = 0.30$ eV.

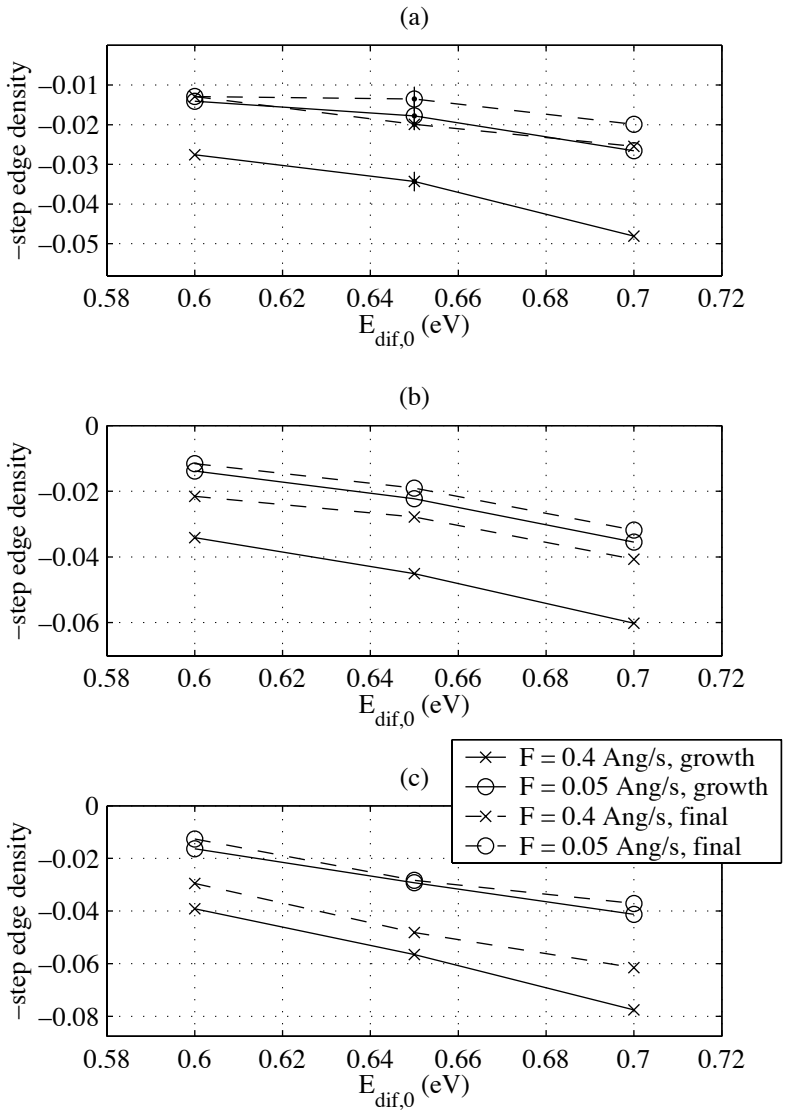


Figure 6.16: Simulated step edge density at 280°C immediately following growth, and after 40 s of recovery: (a) $\Delta E = 0.20$ eV, (b) $\Delta E = 0.25$ eV, and (c) $\Delta E = 0.30$ eV. Error bars for standard deviation over 6 KMC runs are shown in (a) for $E_{\text{dif},0}$=0.65 eV.

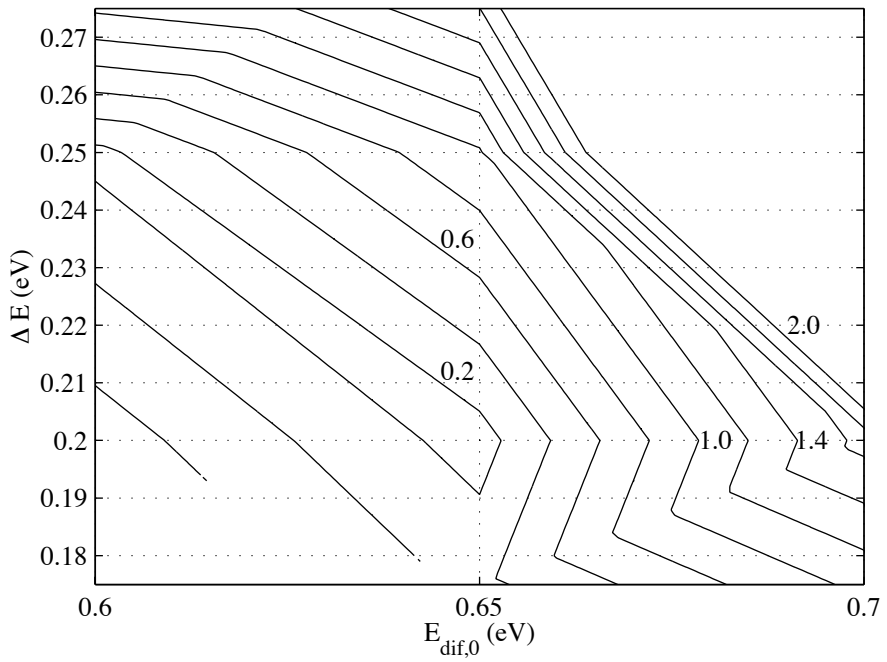


Figure 6.17: Error in island densities between KMC simulations and the experiments of [42]. Error is measured by the two-norm of the percent error at each temperatures, with contours shown at increments of 0.2.

low temperature STM studies of island density, using only the lowest temperature portion of the total data set and assuming no detachment. We used our Monte Carlo model to compare simulated island densities to those reported in [42] for temperatures of 155°C and 230°C and for the range of parameters used in the previous set of simulations. Figure 6.17 is a contour plot showing the error between the island densities of the KMC simulations, and those reported in [42]. The error is defined as the two-norm of the percent error at each of the two temperatures considered. The best agreement is obtained for parameter values of $\{E_{\text{dif},0}, \Delta E\} = \{0.65, 0.20\}$ eV and for $\{E_{\text{dif},0}, \Delta E\} = \{0.60, 0.25\}$ eV. Because the first set of parameters agrees best with our estimate of ΔE and with the value of $E_{\text{dif},0}$ reported previously [42], we ultimately arrive at parameter values of $E_{\text{dif},0} = 0.65 \pm 0.05$ eV and $\Delta E = 0.20 \pm 0.05$ eV.

We further compare our activation energies to an annealing study on Ge(001) [11], in which the activation energy for smoothening of large surface features is reported as 1.9±0.25 eV. This energy is argued to be the sum of the formation and diffusion energies for the mobile species. In our microscopic model, the formation energy for an adatom from a kink site is 1.05 eV, while its activation energy for surface diffusion is 0.65 eV, for a total of 1.70 eV—consistent with the annealing study.

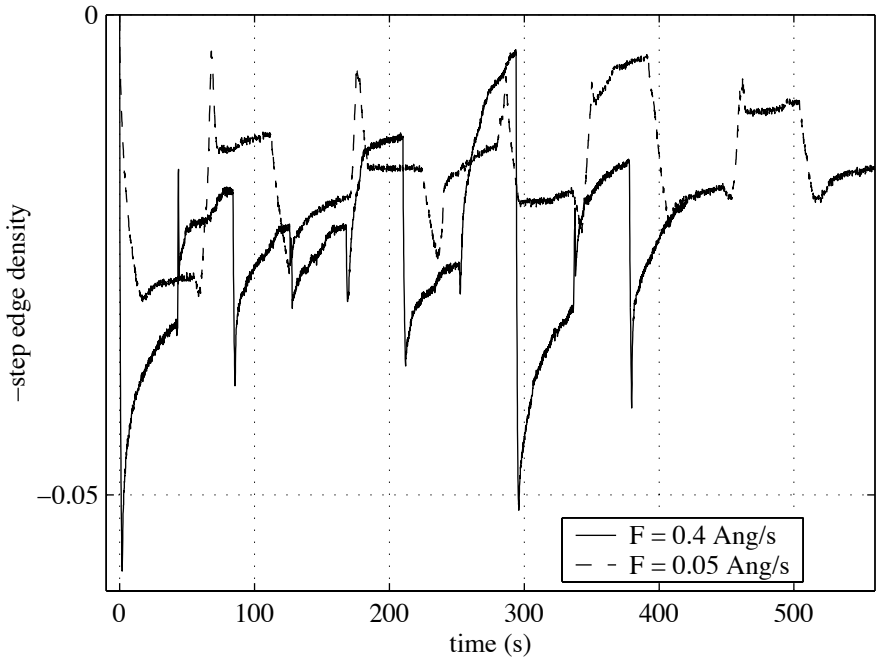


Figure 6.18: Kinetic Monte Carlo simulations of multilayer pulsed growth. The physical parameters are those determine previously for Ge(001): $\nu = 10^{13}s^{-1}$, $E_{\text{dif},0} = 0.65$ eV, and $\Delta E = 0.20$ eV. The growth conditions correspond to experimental conditions of T = 230°C and F = 0.4 and 0.05 Å/s. Pulse times of 2 s and 16 s, respectively, result in the deposition of approximately half a monolayer per pulse. Ten pulsing cycles are simulated, with 40 s pauses between the pulses. The faster growth rate results in a slightly higher step density.

Multilayer growth

We now use the parameters previously obtained for the cubic lattice model to make predictions about growth strategies based on time-varying conditions. First the conditions of the two of runs of Figure 6.9 are simulated, with corresponding step densities plotted in Figure 6.18. In both cases, the step density decays during growth and recovers when the source is shuttered. The decay of the higher growth rate is greater, but in both cases, the surfaces remain smooth and do not show any overall decay or recovery over multiple pulses.

We also simulate growth under the conditions shown in Figures 6.10 and 6.11 in which persistent RHEED oscillations are seen. Step edge density is plotted in Figure 6.19 for growth under continuous flux, and with one-monolayer pulses separated by 5 s pauses. In the simulations, the surface remains smooth, as measured by the oscillations, during both strategies, so one would not expect the pauses in growth to substantially alter evolution—in the simulations, both growth strategies result in growth by two-dimensional island nucleation and coalescence.

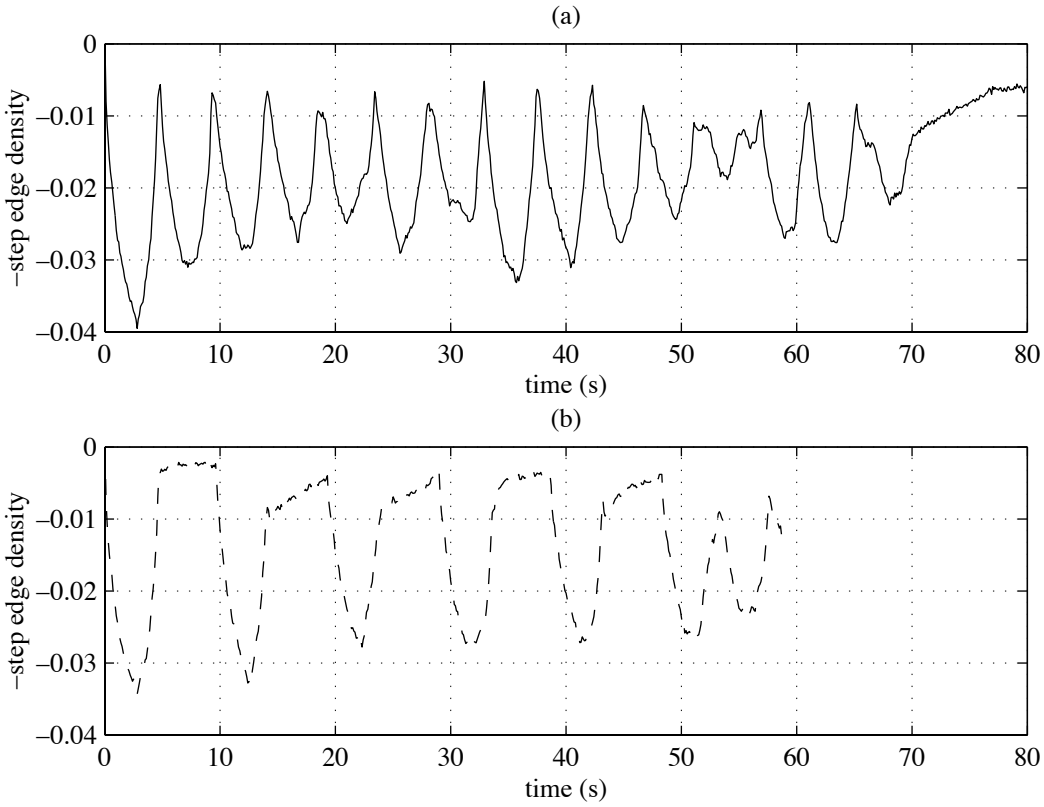


Figure 6.19: Kinetic Monte Carlo simulation of roughness evolution at 0.3 Å/s and 270°C, with $E_{\mathrm{dif},0} = 0.65$ eV, $\Delta E = 0.20$ eV, and $\nu = 10^{13} s^{-1}$. The growth conditions match those of the experiments of Figures 6.10 and 6.11, while the physical parameters are those determined for Ge(001) from the submonolayer experiments: (a) continuous flux (b) 5 s pauses after each of the first five monolayers.

6.4.3 Simulation of alternative growth strategies

Further simulations were performed to contrast the experimental conditions with alternative growth strategies. Figure 6.20 shows simulation results for lower growth temperatures, in which the growth temperature is restricted to the range of 75–150°C. We first simulate growth under constant conditions, with a flux of 1 mL/s and a range of temperatures, as shown in Figure 6.20(a). At the highest temperature, roughness oscillations are seen, which indicate a smooth two-dimensional surface. These oscillations decay as growth proceeds. However, note that the maximum remains near 0.5, indicating that at a coverage of 0.5, the surface is still two-dimensional. We next compare continuous growth to a pulsing strategy like those used in the experiments, in which a dose of one monolayer is deposited, followed by a pause in which the flux is zero. A comparion is made in Figure 6.20(b) for growth at 150°C between a continuous flux of 1 mL/s, and a pulsed flux with 1 mL deposited in 0.2 s, follow by a pause of 0.8 s. There is no substantial change in the final roughness after 10 layers are deposited, indicating that the evolution under pulsed flux is not significantly different than under continuous flux.

We next make a comparison between continuous growth and growth under a different pulsing strategy, in which either temperature is lowered at the beginning of each layer, or alternatively the flux is raised. Either method can lead to smoothing by increasing island density, thereby reducing three-dimensional nucleation. Figure 6.20(c) demonstrates this effect. We compare continuous growth at 150°C to growth in which the temperature is lowered to 75°C during the first 0.2 s of each layer, after which it is raised to 150°C. The final thickness under this synchronized nucleation strategy is substantially lower than under continous growth. However, in both cases the roughness at 1/2 mL coverage is 0.5, so at half coverage, in both cases the surface is two-dimensional. The benefit of the synchronization strategy is to reduce island density and delay the initiation of a new layer before the active one has been completed, providing a surface closer to an atomically flat one after the deposition of an integer number of monolayers.

A final comparison is made between continuous flux and a periodic flux profile reminiscent of pulsed laser deposition. One motivation for this study was to test the hypothesis [58] that the time-varying flux is generally roughening, and that the smoothening effect observed in PLD is the result of energetic effects. Based on the experiments, and the simulations of the experiments, we see no evidence that pulsed MBE creates a significant smoothening *or* roughening effect. However, simulations with the higher fluxes seen in PLD do generate roughening. The results of this simulation study are shown in Figure 6.21. Simulations are performed for the germanium parameters determined earlier in the chapter, and for the silicon parameters used in [58].

We use a pulse time of 5 μs, as in the PLD simulations of Taylor and Atwater [58]. The number of pulses per second is varied, with the instantaneous flux adjusted to maintain a mean growth rate of 1 mL/s. With low pulse rate and high instantaneous flux, a significant roughening effect is seen, which is diminished as the pulse rate is increased. When the pulse rate becomes large, we expect to

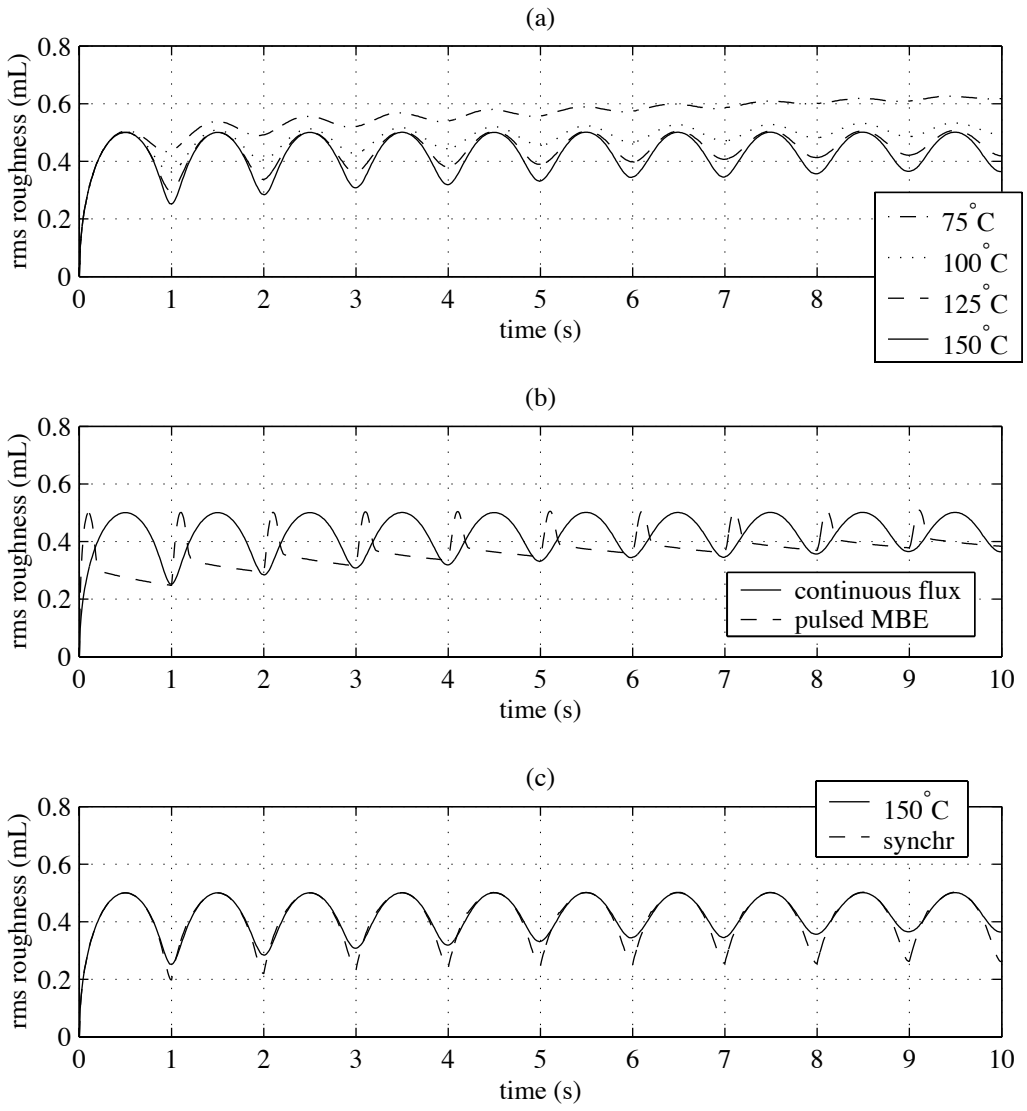


Figure 6.20: Kinetic Monte Carlo simulations of roughness evolution at a mean flux of 1 mL/s and temperatures between 75 and 150°C. The physical parameters are those identified for Ge(001): $E_{\mathrm{dif},0} = 0.65$ eV, $\Delta E = 0.20$ eV, and $\nu = 10^{13} s^{-1}$. In (a), the temperature is held constant throughout growth, while in (b) continuous growth is compared to a pulsed-MBE strategy in which 1 mL is deposited during the first 0.2 s, with the remaining 0.8 s having no flux. In (c) constant temperature growth is compared to a temperature synchronization strategy, in which the temperature is lowered to 75°C during the first 0.2 s of each layer, after which it is raised to 150°C.

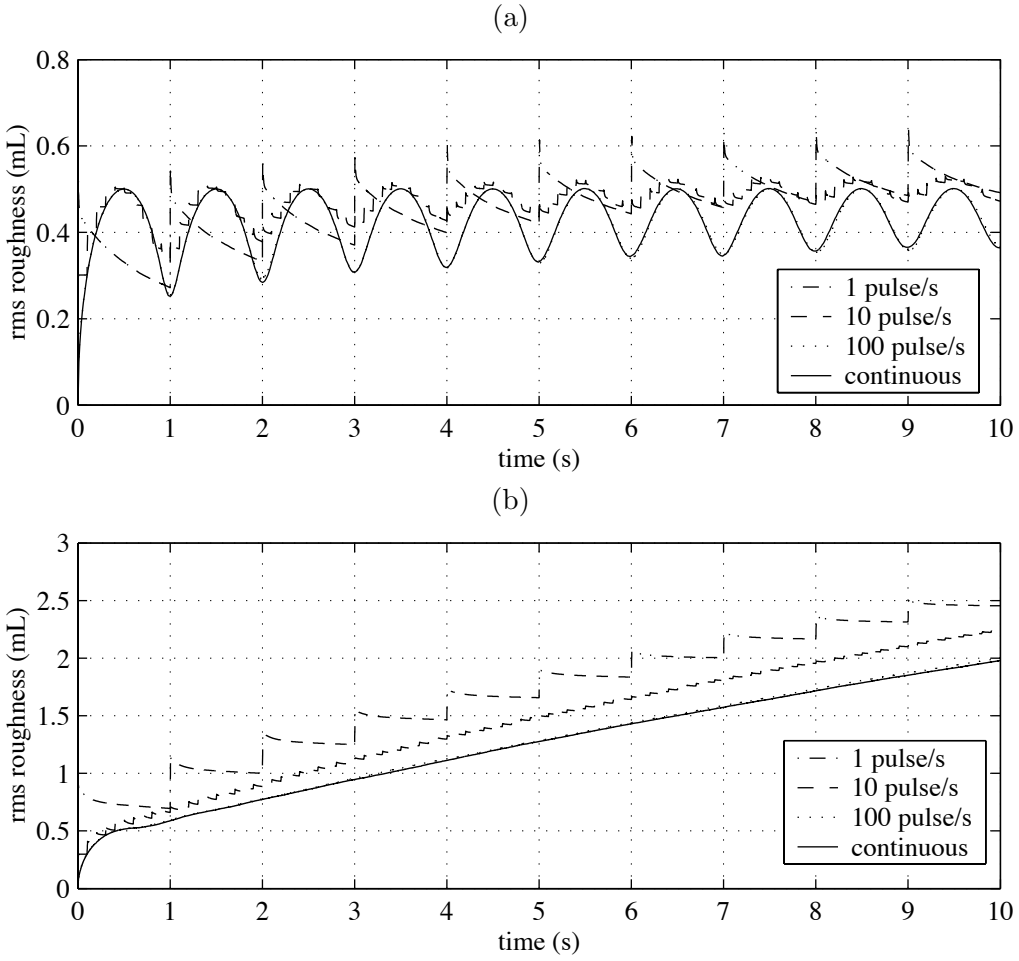


Figure 6.21: Kinetic Monte Carlo simulations of pulsed growth. Growth in pulses of 5 μs is compared to growth with continuous flux. In all cases the mean flux is 1 mL/s. (a) Physical parameters are those for Ge, with a growth temperature is 150°C. (b) Physical parameters, including step-edge barriers, are taken from Taylor and Atwater [58] to represent silicon. The growth temperature is 400°C.

recover the continuous flux evolution, and do for a pulse rate of 100 pulses/mL, with corresponding instantaneous flux only twice that of continous growth. Our lattice model and simulations are thus consistent with the predictions in [58] that an intense pulsed flux without energetic effects is roughening. We further conclude that at typical temperatures and growth rates of MBE, the maximum instantaneous flux is not high enough to create significant roughening.

6.5 Discussion and conclusions

We have used RHEED as a real-time diagnostic to examine the evolution of germanium surface morphology under pulsed flux. The intensity of the specular spot is dependent upon growth rate, temperature, and coverage for submonolayer deposition. However, for multilayer growth, the intensity is only dependent upon temperature within the range of growth rates accessible with our electron gun source. One might alternatively attribute the loss of sensitivity to a change in growth mode to rough three-dimensional growth. However, our RHEED pattern throughout growth indicates an atomically smooth surface—faint Bragg rods are consistent with the presence of atomic-height steps. Additionally, another study has reported that even at our lowest growth temperature, we should expect two-dimensional growth up to at least 50 monolayers, before three-dimensional mounds appear [42] (our experiments consist of less than 10 monolayers).

We conclude that a straightforward application of the kinematical approximation for RHEED is not warranted based on our data. Instead the analysis of the RHEED data is limited to the submonolayer regime, where the specular spot intensity is dependent on both the growth rate and coverage. The intensity is compared for surfaces with equal temperature and equal coverage, and it is observed that the intensity decay is greater when the growth rate is higher. We thus assume only a monotonic dependence between step density and intensity. This enables an estimation of bond energy for the cubic lattice KMC simulations, based only on the conjecture that when temperature and coverage are equal, equal intensity implies equal step density. The kinetic Monte Carlo simulations predict a bond energy of 0.20 eV, using a vibrational frequency of $\nu = 10^{13}\ s^{-1}$ and an adatom diffusion energy of 0.65 eV. This model is consistent with our data, with previously reported activation energies [11, 42], and with a comparison between our simulations and previously reported data [42].

In the multilayer growth experiments, differences in the RHEED intensity due to changes in the instantaneous flux are not observed over multiple layers of growth, and the interpretation of the RHEED pattern is uncertain in this regime. Simulations of the same experimental conditions were performed, which predict that faster instantaneous growth rates increase the step density slightly, and in all cases two-dimensional nucleation and coalescence is the dominant growth mode. The experiments and simulation are consistent with the explanation that pulsed-MBE does not result in a significant effect.

We also simulated growth under the synchronized nucleation strategy, and

under intense flux pulsing characteristic of pulsed laser deposition. A smoother interface is obtained using the synchronization strategy, while intense flux pulsing was roughening. While the pulsed-MBE strategy may not be an effective at generating altered morphology, the simulations do predict that other strategies using time-varying flux and temperature would produce unique surface properties.

At the conclusion of this study, several promising directions for future investigation are suggested. For the purpose of producing smoother films at low temperatures, our simulations suggest that the synchronized nucleation strategy could be beneficial for Ge(001) at low growth temperatures. Previous studies on Ge(111) [37] show that the synchronization can prolong the existence of RHEED oscillations. Connections between this growth strategy and final surface morphology could be of great utility in germanium, and particularly in silicon, and should be investigated experimentally.

It is also clear from the experimental work that while RHEED is sensitive to surface morphology, and provides real-time information during growth, a straightforward interpretation based on the kinematical approximation is not justified. Further development of RHEED models based on multiple scattering would be beneficial, particularly ones that can be run in real-time during growth to enable feedback control. However, in the absence of such models, calibration of the RHEED signal to STM measurements could also be tremendously useful. Our chamber is not equipped with an STM, and we instead relied on another STM study to complement our RHEED data [42]. It is clear that RHEED and an STM provide different types of information each with its own advantages, and that in a surface study such as ours, the combination of the two would be particularly useful.

Chapter 7

Conclusions and Future Work

7.1 Thesis summary and contributions

This thesis brings together theoretical, simulation, and experimental work in modeling the evolution of surface morphology during thin film deposition. The dynamics may be expressed through local interactions within a large collection of discrete atoms. The difficulty with such a high-dimensional representation arises when the computational demands of simulation become high. Although it may be feasible to make predictions, computing gradients or inverting the input-output map may be impractical.

The goal of this work has been to identify simpler descriptions and interpretations of the input-output behavior observed in a high-dimensional lattice model. In Chapter 3, the use of fast periodic processing conditions was examined, exploiting the mathematical structure of the master equation, and in particular the nonlinear map from macroscopic process parameters to microscopic transition rates. Application of the averaging theorem to the master equation suggests that when the period of oscillation is faster than the response time of the surface, the surface evolves according to constant transition rates that are the average of the instantaneous values applied. Because the maps from process parameters to inputs may be nonlinear, these effective transition rate may not be attainable with constant reactor conditions. This analysis requires consideration only of the transition rates, despite the extremely high dimension of the evolution equation. The set of all possible effective rates is the convex hull of the set of instantaneous rates, so linear programming techniques may be employed to compute the input trajectories.

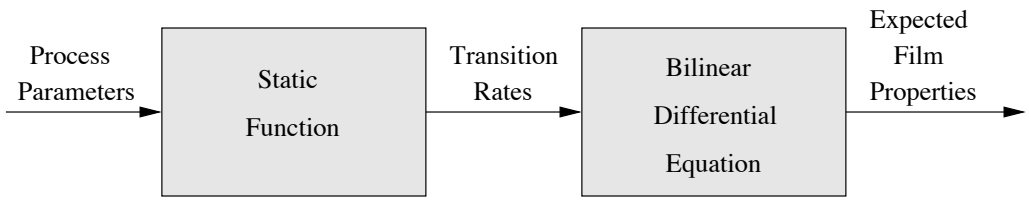


Figure 7.1: Block diagram of the lattice model.

The averaging analysis opens the space of possible transition rate inputs into the second dynamic block. These rates act as inputs for the bilinear differential equation that describes the probability distribution. Although fast periodic processing parameters yield new constant effective inputs, one also needs to understand the effect of inputs that vary more slowly. Because it is not possible to explicitly formulate the differential equation, due to its high dimension, the known linearity of the state was combined with the KMC simulations to identify low-order master equations that capture the transfer of probability between groups of similar microscopic configurations. This model was demonstrated to be compatible with a gradient-based optimization algorithm, through which optimal time-varying process parameters were computed for various cost functions.

The effect of time-varying process parameters was also studied in the model system of germanium, through experiment and simulation. Germanium was deposited onto germanium substrates in an ultra high vacuum process, using reflection high-energy electron diffraction as a morphology sensor during growth. The experimental data were used to determine parameters for the lattice model and KMC simulations, and as well as to test periodic growth strategies. The identified model predicts that within the range of inputs available in the molecular beam epitaxy process, a periodic growth strategy would not produce significantly different morphology over constant conditions, which was consistent with our experimental observations. The simulations also predict that more intense flux pulsing or temperature modulation can produce altered morphology.

In this thesis, specific problems in surface morphology control have been studied, such as the synchronized nucleation strategy of Chapter 5, and the particular material system of germanium in Chapter 6. However, at this point in time the greater contribution seems to be the generality of the approaches developed here. Both the averaging method and the KMC model reduction may be applied to any system described by a master equation. At the outset of this work, the master equation was viewed as an abstract representation, without much utility due to its uncountable dimension. However, the model reduction approach and examples have demonstrated that behavior seen in KMC simulations is characteristic of master equation dynamics, and that the master equation is a good mathematical structure on which to base the system identification.

7.2 Future directions

Because the dimension of the master equation is extremely large for all but the simplest systems, we do not attempt to compute state matrices for physically realistic surfaces. Instead, we borrow the mathematical structure of the underlying master equation, and then use Monte Carlo simulations for model identification and for demonstration of our ideas. This is a strategy common to both the averaging analysis and the model reduction. In the former case, we suggest that the timescales of surface evolution are slower than the input period, which in a finite dimensional system could be supported by the study of the eigenvalues of the

state matrices. However, in the lattice model, we not only have an infinite number of eigenvalues, but also have an uncountable number of configurations, for which there are few mathematical results. Similarly, in the model reduction work error bounds were proven for finite-dimensional systems, after which the ideas, but not the bounds, were extended to larger systems.

One of the strengths of the control theory is its attention to rigor and its quantification of performance via norm bounds. In this work, we have extended ideas that are quantifiable to uncountably infinite spaces, which work quite well in demonstration. However, further study of these approaches and their associated bounds is warranted, and might be accomplished either through the development of better theoretical constructs for uncountable spaces, or by formally reducing the system to a more tractable form.

Another approach presented here that could benefit from future study is the identification of a reduced configuration set that forms a finite-dimensional state space for the identified models. Methods have been developed to compute reduced coordinates in high-dimensional systems, and in particular to generate these modes from data. However, we have chosen to restrict ourselves only to coordinates representing probabilities, so that we preserve the stability of the associated stochastic matrices. It would be useful to have a method that would produce optimal coordinates within this constraint, or alternatively to develop a less constrained method that also guaranteed stability.

In this thesis we integrated theory, simulation, modeling, and experiment to investigate the effects of time-varying process parameters, through periodic forcing and open loop optimization. We placed less focus on the issue of feedback control, which ultimately would also be part of any practical implementation. We hope that eventually the overall approach advocated in this thesis will be applied in experiment, from first-principles modeling and KMC model reduction to computation of optimal inputs to sensor integration and finally to demonstration of optimal trajectories coupled with sensor feedback. At the initiation of the research in this thesis, it seemed that the biggest barrier to implementation of a coherent control strategy was the lack of a suitable model for both computation of optimal open-loop trajectories and for design of feedback controllers. However, at the conclusion of this thesis, a path has been identified to reduce large lattice models. We now see no single barrier to the implentation of control strategy based on physical modeling, but instead see many practical challenges. One of the greatest may be the difficulty in sensing. While many *in-situ* exist, they typically do not provide direct state information, and may require extensive modeling for interpretation. We view the sensor development issue as distinct from the dynamics of the evolving system, but note that both must be addressed in order to implement control in practice.

The modeling results presented in this thesis were developed in the context of thin film deposition and, in particular, the evolution of surface morphology. The lattice model with its associated master equation captures many important features, so we took the lattice model to be the actual physics, and then subsequently searched for simpler representations of it. Many other physical effects,

beyond surface morphology, may be captured by the lattice model, including multispecies effects, faceting, grain growth, and surface reconstruction. Additionally, lattice models capture bulk effects including magnetization and species interdiffusion. Furthermore, master equations provide probabilistic representations of many systems, and are particularly important in small systems where fluctuations are important, like quantum computation and intracellular chemical reactions. The methods developed in this thesis are not specific to surface morphology, film growth, or even lattice models. Many potential applications exist in other fields and should be actively pursued.

Appendix A

Computation of periodic process conditions

```
% Matlab code  9-21-01 Martha Gallivan

keff = [1.0; 500.005; 505];             % Desired effective transition rates.

% Define physical parameters.
m = 3;                                  % Three transition mechanisms.
Fmin = 0;                               % Define the bounds on the process
Fmax = 1;                               %   parameters.
Tmin = 2/3;
Tmax = 1;
vdif = 1e13;                            % Define the vibrational frequencies
vdes = 1e7;                             %   and activation energies.
Edif = 23.0258509299405;
Edes = 9.21034037197618;

% Discretize process parameters and compute transition rates.
F = [Fmin:(Fmax-Fmin)/10:Fmax];         % Discretize the process parameters.
T = 100*log([exp(Tmin/100):(exp(Tmax/100)-exp(Tmin/100))/10:exp(Tmax/100)]);
np = length(F)*length(T);               % Number of parameter settings.
kconst = zeros(m,np);                   % Compute the transition rates at
for i = 1:length(F)                     %   each setting.
     for j = 1:length(T)
       kconst(:,i) = [F(i); vdif*exp(-Edif/T(j)); F(i)*vdes*exp(-Edes/T(j))];
     end
end

% Solve for the fraction of time spent at each setting.
v = [ones(1,np); kconst];
Aeq = v;
beq = [1; keff];
LB = zeros(np,1);
```

```
UB = Inf*ones(np,1);
lambda = linprog([],[],[],Aeq,beq,LB,UB);
[v*lambda [1; keff]]

% Reduce the number of settings to m+1.
for j = np:-1:m+2
  n = null(v);
  mu = n(:,1);
  L = max(lambda)/min(abs(mu));
  for k = 1:j
    if (mu(k) > 0)
      L = min(L,lambda(k)/mu(k));
    end
  end
  lambda2 = lambda - L*mu;
  [Y,I] = min(lambda2);
  v = [v(:,1:I-1) v(:,I+1:j)];
  lambda = [lambda2(1:I-1); lambda2(I+1:j)];
end
[v*lambda [1; keff]]

% The vector lambda contains the fractions of the period spent at each
% of the kc rates.
kc = v(2:m+1,:);

% Solve for the process parameters from the transition rates.
F = kc(1,:);
T = Edif./(-log(kc(2,:)./vdif));
```

Appendix B

Spatial modes

In this study, we explore the dynamics of film growth on a vicinal surface using a kinetic Monte Carlo (KMC) model, over a temperature range spanning the transition from step flow growth at high temperature to island growth and coalescence on terraces between rough steps at low temperature. We use the technique of proper orthogonal decomposition (POD) to identify spatial modes that represent the major features of the surface profiles.

B.1 Kinetic Monte Carlo simulations

We studied the surface morphology of a growing crystal with a kinetic Monte Carlo model. We consider a single-species material on a cubic lattice, and we increment time as in Fichthorn and Weinberg [15] to achieve a physically based time. Vacancies in the crystal are prohibited. We define transition rates for adsorption, desorption, and surface diffusion based on a nearest-neighbor bond-counting model:

$$k_{\text{ads}} = \gamma P_j \sqrt{\frac{1}{2\pi m k_b T}} \tag{B.1}$$

$$k_{\text{des},i} = \frac{k_b T}{h} \exp\left(-\frac{E_{\text{des},0} + i\Delta E}{k_b T}\right) \tag{B.2}$$

$$k_{\text{dif},i} = \frac{k_b T}{h} \exp\left(-\frac{E_{\text{dif},0} + i\Delta E}{k_b T}\right), \tag{B.3}$$

where i, ranging from 0 to 4, is the number of adjacent side neighbors, k_{ads} is the adsorption rate, $k_{\text{des},i}$ is the desorption rate for a surface site with i nearest neighbors, and $k_{\text{dif},i}$ is the diffusion rate for a surface site with i nearest neighbors. The Boltzmann constant is denoted with k_b, Planck's constant is h, the sticking coefficient is γ, the mass of the particle is m, temperature is T, and the precursor partial pressure is P_j. The chemistry model has four free parameters: three activation energies and a constant in the adsorption rate. The activations energies $E_{\text{des},0}$, $E_{\text{dif},0}$, and ΔE are the depths of the potential energy wells associated with the occurrence of a surface event. Specifically, $E_{\text{des},0}$ is the energy for the desorp-

tion of an atom with no side neighbors, $E_{\mathrm{dif},0}$ is the energy for the diffusion of an atom with no side neighbors, and ΔE is the additional energy barrier associated with a single side neighbor.

We performed simulations on a 256×256 domain and deposited 4 layers of atoms. Eight equally spaced steps were inserted into the initialized lattice, and periodic boundary conditions were used to simulate an infinite train of steps. The transition rate parameters were $\gamma(2\pi m k_b)^{-0.5} = 5\sqrt{\mathrm{K}}\mathrm{Pas}^{-1}$, $E_{\mathrm{des},0} = 2.64 \times 10^{-18}J$, $E_{\mathrm{dif},0} = 3.02 \times 10^{-19}J$, and $\Delta E = 7.59 \times 10^{-20}J$. The activation energy for desorption is sufficiently high such that desorption is negligible in the simulations. We considered a nominal partial pressure $P_{j,o} = 1$ Pa and nominal temperatures T_o of 950 K, 1050 K, and 1150 K.

The temperature range was selected to span the transition from step flow growth at high temperature to growth primarily by island nucleation at low temperature. At intermediate temperatures, both processes contribute, as shown in Figure B.1. This figure shows that the islands are not distributed randomly on the terrace. Since the steps are sinks for adatoms, the adatom density is highest away from the steps, and it is here that islands preferentially nucleate.

Layer-by-layer growth via island nucleation and step growth results in a periodicity at the monolayer growth frequency that is usually detected as oscillations in the measured RHEED signal. This behavior is seen in plots of the step edge density and of the root-mean-square roughness. These measures are plotted in Figure B.2 for the simulation pictured in Figure B.1. The oscillations decay due to roughening of the step edges. Note in Figure B.1 that the step edge is not straight at the completion of one layer of growth.

The oscillations in surface properties are influenced by the spatial non-uniformities imposed by the steps. Because the islands are concentrated near the original center of the terrace, the collision of steps with islands happens suddenly. Once the steps have collided with the islands, the surface becomes smoother. Notice that in Figure B.1(b) the steps have not yet collided with most of the islands, while in Figure B.1(c), the step has merged with the islands, creating a minimum in roughness and step edge density.

To better understand the spatial distribution of islands, we examine the shape of the surface height profile for a range of growth conditions, from step-dominated growth at 1150 K to island-dominated growth at 950 K. We also consider growth under sinusoidally varying temperature and precursor partial pressure in an attempt to excite different spatial modes. To obtain spatial profiles across the terrace, we average the surface height over the direction parallel to the steps. We also average over the eight steps to yield the surface height profile for a single step as a function of the distance along the terrace. Spatial profiles throughout the growth of four layers are shown in Figure B.3. At low temperature, islands grow between the steps and dominate the growth. As time advances, the surface becomes more disordered, which may be seen in the smoothening of the surface profile. At high temperatures, growth occurs as adatoms attach to the step edge. The propagation of the step is the dominant spatial feature in this limit. Notice that the vertical scale is greatly expanded, so that the initial stepped surface appears to be a

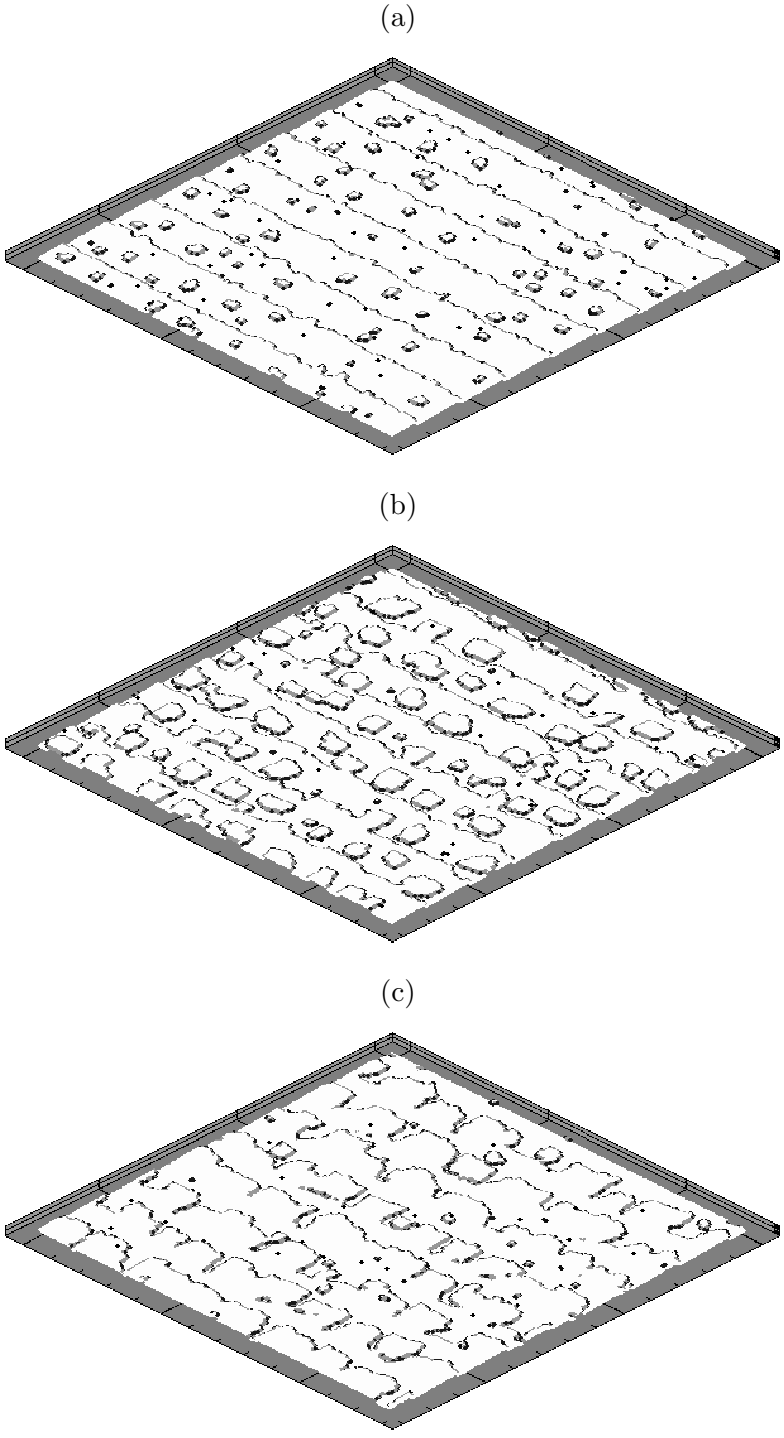


Figure B.1: Monte Carlo simulation at 1050 K at various stages of growth: (a) 0.16 layers, (b) 0.54 layers, (c) 1.0 layers.

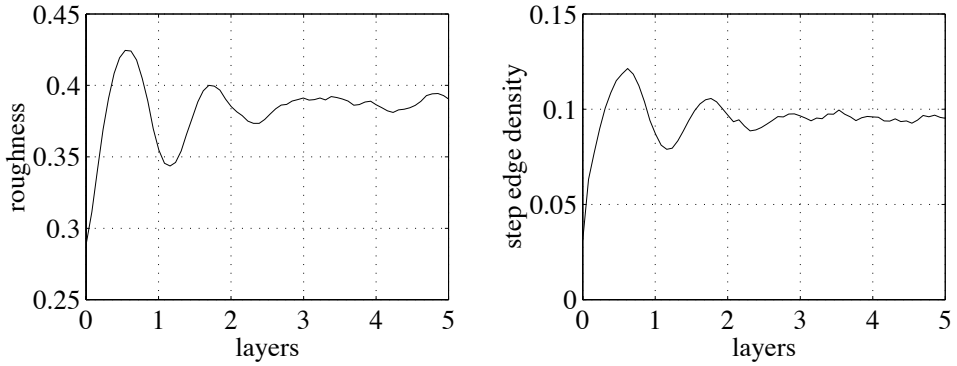


Figure B.2: Roughness and step edge density from a KMC simulation at T = 1050 K.

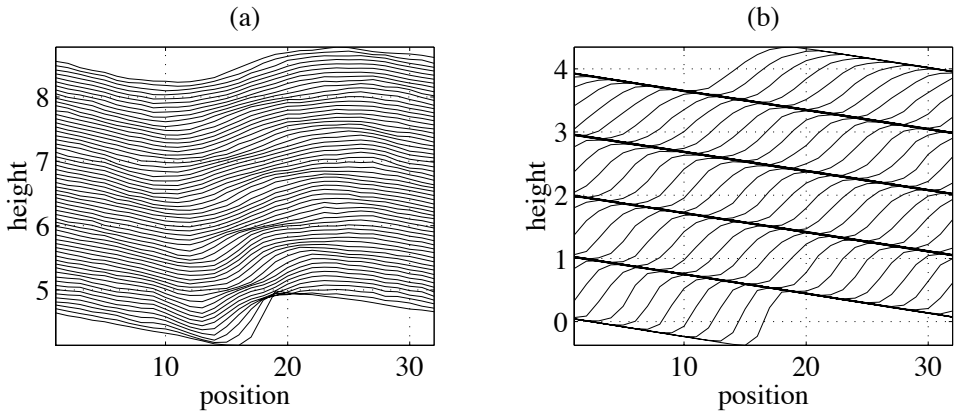


Figure B.3: Surface height as a function of position in the direction perpendicular to the step, for KMC simulations at (a) 950 K, (b) 1150 K.

sawtooth wave.

B.2 Proper orthogonal decomposition

We use a method called proper orthogonal decomposition (POD) to extract the typical shape of the surface height profile [26]. The input to POD is a collection of "snapshots" of data. As output, the POD method returns a series of orthogonal spatial modes, which are ordered in decreasing importance. The root-mean-square error between the snapshots and their projections onto the first n modes possesses the minimum root-mean-square error that any n orthogonal modes could (where n is an integer from 1 to the number of snapshots). Thus, the POD modes are considered to be optimal. If low-order behavior exists in the spatial profile, the first few POD modes will capture a large percentage of the total spatial profile.

In this study we take as our snapshots all of the spatial profiles plotted in Figure

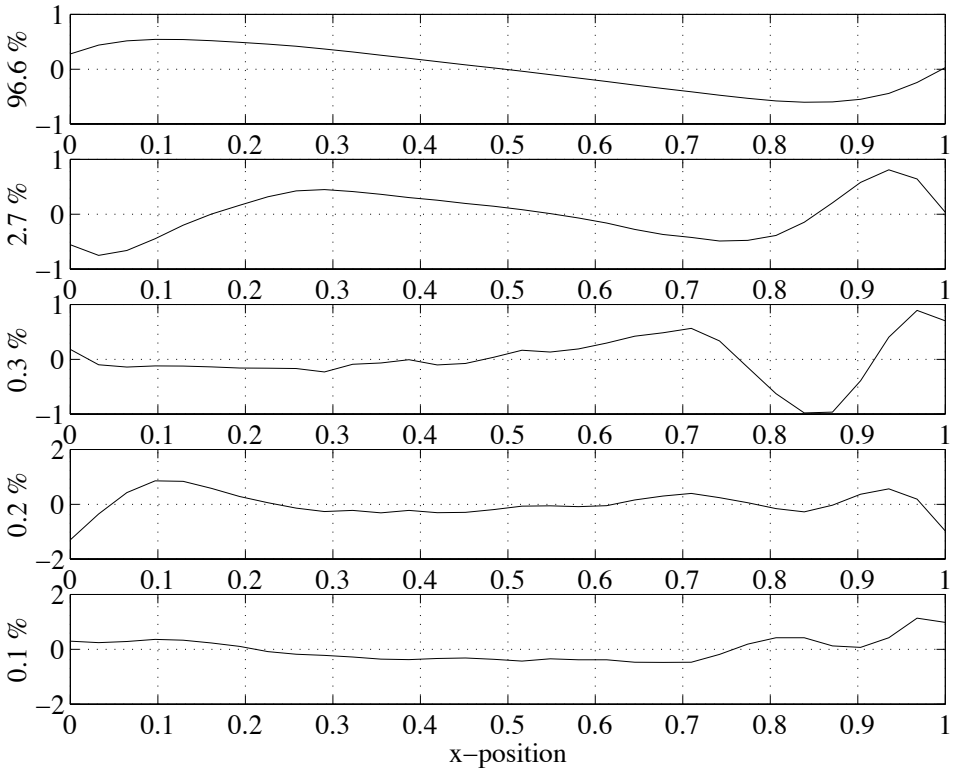


Figure B.4: The first five spatial modes of surface height, obtained by proper orthogonal decomposition. The percentages denote the amount of energy captured by each mode.

B.3. In Figure B.3(b) the surface maintains a typical shape which propagates in time. To capture more energy in fewer POD modes, we preprocess our snapshots with a procedure called template fitting [52], in which the spatial profiles are shifted to line up with each other.

The first five spatial modes determined by POD are plotted in Figure B.4. The first mode captures 96.6% of the energy, meaning that the root-mean-square error between the KMC surface profiles and the projections onto the first mode is only 3.4%. The shape of the first mode represents the overall shape of the original step in the KMC simulations. Its corners are rounded, consistent with the smoothening of the profiles seen in Figure B.3. The second mode contains 2.7% of the energy, totaling 99.3% for the first two modes combined. This mode has features on its ends which, when added to the first mode, can sharpen or smooth the overall shape. The feature in the center of the second mode represents island growth on the terrace. The third, fourth, and fifth modes are also shown in Figure B.4. They contain little energy. The most dominant features of these modes are on the right side. As in the second mode, the features on the right side resemble sinusoids in various phases with each other. These modes can be combined to reconstruct a traveling wave.

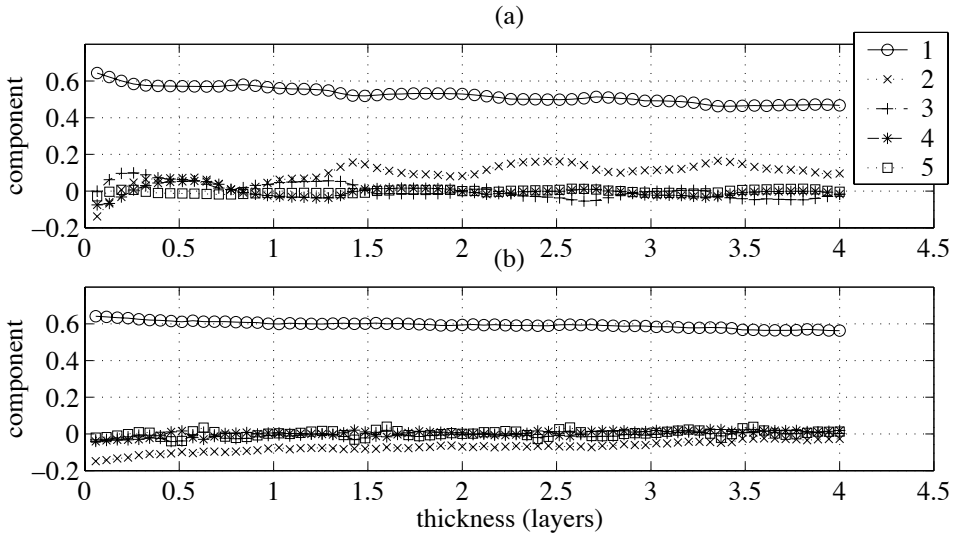


Figure B.5: The size of the projection of the first five POD modes onto the snapshots (a) 950 K, (b) 1150 K (shown in Figure B.3).

The projection size of each snapshot onto each mode is shown in Figure B.5. At the low temperature, the first mode decreases during growth while the second mode increases. This represents an overall roughening of the surface. In addition, oscillations in the second mode occur once per monolayer, signaling that the spatial height profile is oscillating near the monolayer frequency. The higher modes do not appear to contribute significantly to the spatial distribution. At the high temperature, when step flow dominates, oscillations are not evident. The sharp sawtooth-like shape of the step becomes smoother as the step becomes more disordered, as seen in a decrease in the first mode and an increase in the second mode.

B.3 Comparison to other snapshots

We have demonstrated that the POD method generates modes in which the first few modes capture the majority of the KMC spatial height profile. Next we test these modes on the spatial profiles of a KMC simulation associated with an intermediate temperature, 1050 K, at which both islands and steps grow. The new snapshots are shown in Figure B.6(a). In addition, we vary the temperature and partial pressure sinusoidally near the monolayer frequency, with amplitudes of 25 K and 0.95 Pa. We might expect that different spatial modes would be present in such a case. By projecting these profiles onto the previously computed modes, we test whether or not the previously computed modes represent overall spatial features of film growth on steps.

We project the spatial height profiles of the new simulation onto the previously computed POD modes. The energy captured by the first five modes is 94.7%.

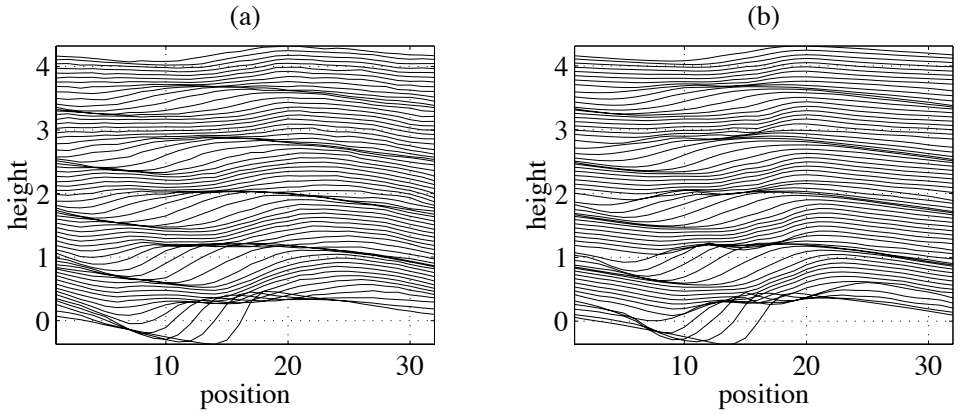


Figure B.6: Surface height as a function of position perpendicular to the step, from a KMC simulation at 1050 K with oscillating growth conditions: (a) original snapshots, (b) reconstructed surface height.

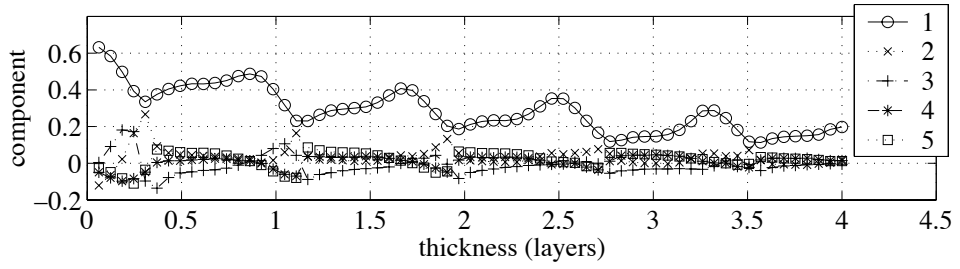


Figure B.7: The size of the projection of the first five POD modes onto the snapshots in Figure B.6(a).

While it is not as good as the 99.8% captured for the original data set, most of the energy is captured. This is quite remarkable considering that the spatial profiles contain different features. Because both islands and steps contribute to growth, there is a stationary feature, the islands, and a traveling feature, the steps, in the spatial profile. Template fitting cannot align both features, so a traveling wave is present in the final snapshots. During the early stages of growth, the first, second, third, and fifth modes all appear significantly in the response, as shown in Figure B.7. The second, third, and fifth modes contain spatial oscillations near the right side which when combined represent the traveling wave. Even though these modes did not play a large role in the spatial signals from which they were derived, they were captured and play a significant role for other growth conditions. The reconstructed spatial profiles are shown in Figure B.6(b). They appear very similar to the original spatial profiles.

B.4 Conclusions

We studied the surface morphology of film growth on a vicinal surface with a kinetic Monte Carlo simulation. To better understand the role of spatial non-uniformities, we searched for characteristic shapes of the surface height. The proper orthogonal decomposition technique produced spatial modes from the kinetic Monte Carlo data. These modes captured 99% of the spatial height profiles for the data used to generate the modes. For simulations performed at another growth condition, the modes captured 95% of the profile.

Bibliography

[1] J. Aarts, W. M. Gerits, and P. K. Larsen. Observations on intensity oscillations in reflection high-energy electron diffraction during epitaxial growth of Si(001) and Ge(001). *Applied Physics Letters*, 48(14):931–933, 1986.

[2] J. Aarts, A. J. Hoeven, and P. K. Larsen. Crystal growth of Ge studied by reflection high-energy electron diffraction and photoemission. *Journal of Vacuum Science and Technology A*, 6(3):607–610, 1988.

[3] O. Auciello and A. R. Krauss, editors. *In Situ Real-Time Characterization of Thin Films*. John Wiley and Sons, Inc., New York, NY, 2001.

[4] C. Baggio, R. Vardavas, and D. D. Vvedensky. Fokker-Planck equation for lattice deposition models. *Physical Review E*, 64(4):art. no. 045103 Part 2, 2001.

[5] J. Baillieul. Energy methods for stability of bilinear systems with oscillatory inputs. *International Journal of Robust and Nonlinear Control*, 5:285–301, 1995.

[6] C. C. Battaile and D. J. Srolovitz. A kinetic Monte Carlo method for the atomic-scale simulation of chemical vapor deposition: application to diamond. *Journal of Applied Physics*, 82:6293–6300, 1997.

[7] C. L. Berrie and S. R. Leone. Observation of monolayer and bilayer period RHEED oscillations during epitaxial growth of Ge on Ge(100). *Journal of Crystal Growth*, 216(1-4):159–170, 2000.

[8] L. H. Chan, E. I. Altman, and Y. Liang. Development of procedures for obtaining clean, low-defect-density Ge(100) surfaces. *Journal of Vacuum Science and Technology A*, 19(3):976–981, 2001.

[9] E. Chason, J. Y. Tsao, K. M. Horn, and S. T. Picraux. Dynamics of growth roughening and smoothening of Ge(001). *Journal of Vacuum Science and Technology A*, 7(2):332–336, 1989.

[10] E. Chason, J. Y. Tsao, K. M. Horn, S. T. Picraux, and H. A. Atwater. Surface roughening of Ge(001) during 200 eV Xe ion bombardment and Ge molecular beam epitaxy. *Journal of Vacuum Science and Technology A*, 8(3):2507–2511, 1990.

[11] S. J. Chey, J. E. Van Nostrand, and D. G. Cahill. Dynamics of rough Ge(001) surfaces at low temperatures. *Physical Review Letters*, 76(21), 1996.

[12] P. I. Cohen, P. R. Pukite, J. M. Van Hove, and C. S. Lent. Reflection high energy electron diffraction studies of epitaxial growth on semiconductor surfaces. *Journal of Vacuum Science and Technology A*, 4(3):1251–1258, 1986.

[13] S. Das Sarma and P. Tamborenea. A new universality class for kinetic growth: one-dimensional molebular-beam epitaxy. *Physical Review Letters*, 66:325–328, 1991.

[14] K. Eberl, W. Wegscheider, and G. Abstreiter. Group IV element (Si, Ge, and α-Sn) superlattices — low temperature MBE. *Journal of Crystal Growth*, 111:882–888, 1991.

[15] K. A. Fichthorn and W. H. Weinberg. Theoretical foundations of dynamical Monte Carlo simulations. *Journal of Chemical Physics*, 95:1090–1096, 1991.

[16] M. G. Flynn, R. Smith, R. Abraham, and S. DenBaars. Control of a III-V MOCVD process using ultraviolet absorption and ultrasonic concentration monitoring. *IEEE Transactions on Control Systems Technology*, 9:728–740, 2001.

[17] M. A. Gallivan, D. G. Goodwin, and R. M. Murray. A design study for thermal control of a CVD reactor for YBCO. In *Proceedings of the 1998 IEEE International Conference on Control Applications*, pages 1194–1199, 1998.

[18] M. A. Gallivan, R. M. Murray, and D. G. Goodwin. The dynamics of thin film growth: a modeling study. In M. D. Allendorf and M. L. Hitchman, editors, *CVD XV: Proceedings of the Fifteenth Symposium on Chemical Vapor Deposition*, volume 616, pages 168–175. The Electrochemical Society, 2000.

[19] M. A. Gallivan, R. M. Murray, and D. G. Goodwin. Kinetic Monte Carlo simulations of dynamic phenomena in thin film growth. In H. N. G. Wadley, G. H. Gilmer, and W. G. Barker, editors, *New Methods, Mechanisms, and Models of Vapor Deposition*, volume 2000-13, pages 61–66. Materials Research Society, 2000.

[20] D. T. Gillespie. The chemical Langevin equation. *Journal of Chemical Physics*, 113(1), 2000.

[21] G. H. Gilmer and P. Bennema. Simulation of crystal growth with surface diffusion. *Journal of Applied Physics*, 43(4):1347–1360, 1972.

[22] M. F. Gyure, C. Ratsch, B. Merriman, R. E. Caflisch, S. Osher, and J. J. Zinck. Level-set methods for the simulation of epitaxial phenomena. *Physical Review E*, 58(6):R6927–R6930, 1998.

[23] Gang Hang. *Novel Group IV alloy semiconductor materials*. Ph.D. thesis, California Institute of Technology, Pasadena, CA 91125, 1997.

[24] M. Henzler. Defects at semiconductor surfaces. *Surface Science*, 152/153:963–976, 1985.

[25] I. P. Herman. *Optical Diagnostics for Thin Film Processing*. Academic Press Inc., San Diego, CA, 1996.

[26] P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, coherent structures, dynamical systems, and symmetry*. Cambridge University Press, Cambridge, UK, 1996.

[27] S. Imai, S. Takagi, O. Sugiura, and M. Matsumura. A novel atomic layer epitaxy method of silicon. *Japanese Journal of Applied Physics*, 30:3646–3651, 1991.

[28] B. A. Joyce, D. D. Vvedensky, G. R. Bell, J. G. Belk, M. Itoh, and T. S. Jones. Nucleation and growth mechanisms during MBE of III-V compounds. *Materials Science and Engineering B*, 67:7–16, 1999.

[29] J.-N. Juang and R. S. Pappa. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics*, 8(5):620–627, 1985.

[30] M. Kardar, G. Parisi, and Y.-C. Zhang. Dynamic scaling of growing interfaces. *Physical Review Letters*, 56(9):889–892, 1986.

[31] H. Karloff. *Linear Programming*. Birkhäuser, Boston, 1991.

[32] P. V. Kokotovic, R. E. O'Malley Jr., and P. Sannuti. Singular perturbations and order reduction in control theory—an overview. *Automatica*, 12:123–132, 1976.

[33] S. Lall, J. E. Marsden, and S. Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear systems. *International Journal of Robust and Nonlinear Control*, 12(6):519–535, 2002.

[34] O. Lange. *Price Flexibility and Employment*, pages 103–109. The Principia Press, Bloomington, Indiana, 1952.

[35] C. S. Lent and P. I. Cohen. Diffraction from stepped surfaces. I. Reversible surfaces. *Surface Science*, 139:121–154, 1984.

[36] F. Liu, A. H. Li, and M. G. Lagally. Self-assembly of two-dimensional islands via strain-mediated coarsening. *Physical Review Letters*, 87(12):art. no. 126103, 2001.

[37] V. A. Markov, O. P. Pchelyakov, L. V. Sokolov, S. I. Stenin, and S. Stoyanov. Molecular beam epitaxy with synchronization of nucleation. *Surface Science*, 250:229–234, 1991.

[38] Z. Mitura. RHEED from epitaxially grown thin films. *Surface Review and Letters*, 6(3–4):497–516, 1999.

[39] B. C. Moore. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.

[40] W. W. Mullins. Theory of thermal grooving. *Journal of Applied Physics*, 28(3):333–339, 1957.

[41] A. K. Myers-Beaghton. Two-dimensional kinematic diffraction patterns from simulations of epitaxial growth. *Surface Science*, 241:439–453, 1991.

[42] J. E. Van Nostrand, S. J. Chey, and D. G. Cahill. Low temperature growth morphology of singular and vicinal Ge(001). *Physical Review B*, 57(19):12536–12543, 1998.

[43] J. E. Van Nostrand, S. J. Chey, M.-A. Hasan, D. G. Cahill, and J. E. Greene. Surface morphology during multilayer epitaxial growth of Ge(001). *Physical Review Letters*, 74(7):1127–1130, 1995.

[44] D. S. Olson, M. A. Kelly, S. Kapoor, and S. B. Hagstrom. A mechanism of CVD diamond film growth deduced from the sequential deposition from sputtered carbon and atomic hydrogen. *Journal of Materials Research*, 9:1546–1551, 1994.

[45] R. G. Phillips and P. V. Kokotovic. A singular perturbation approach to modeling and control of Markov chains. *IEEE Transactions on Automatic Control*, 26(5):1087–1094, 1981.

[46] S. Raimondeau and D. G. Vlachos. Low-dimensional approximations of multiscale epitaxial growth models for microstructure control of materials. *Journal of Computational Physics*, 160:564–576, 2000.

[47] L. L. Raja, R. J. Kee, R. Serban, and L. R. Petzold. Computational algorithm for dynamic optimization of chemical vapor deposition processes in stagnation flow reactors. *Journal of the Electrochemical Society*, 147:2718–2726, 2000.

[48] D. W. Robinson and J. W. Rogers. Low temperature atomic layer growth of aluminum nitride on Si(100) using dimethylethylamine alane and 1,1-dimethylhydrazine. *Thin Solid Films*, 372:10–24, 2000.

[49] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

[50] F. Roozeboom and N. Parekh. Rapid thermal processing systems: a review with emphasis on temperature control. *Journal of Vacuum Science and Technology B*, 8(6):1249–1259, 1990.

[51] G. Rosenfeld, N. N. Lipkin, W. Wulfhekel, J. Kliewer, K. Morgenstern, B. Poelsema, and G. Comsa. New concepts for controlled homoepitaxy. *Applied Physics A*, 61:455–466, 1995.

[52] C. W. Rowley and J. E. Marsden. Reconstruction equations and the Karhunen-Loeve expansion for systems with symmetry. *Physica D*, 142:1–19, 2000.

[53] K. Seshan, editor. *Handbook of Thin-Film Deposition Processes and Techniques*. Noyes Publications, Norwich, NY, 2002.

[54] T. Shitara, D.D. Vvedensky, M.R. Wilby, J. Zhang, J.H. Neave, and B.A. Joyce. Morphological model of reflection high-energy electron-diffraction intensity oscillations during epitaxial-growth on GaAs(001). *Applied Physics Letters*, 60(12):1504–1506, 1992.

[55] H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29(2):111–138, 1961.

[56] L. V. Sokolov, M. A. Lamin, V. A. Markov, V. I. Mashanov, O. P. Pchelyakov, and S. I. Stenin. Oscillations in the optical characteristics of the growth surface on Ge films during molecular beam epitaxy. *JETP Letters*, 44(6):357–359, 1986.

[57] Y. Suda. Migration-assisted Si subatomic-layer epitaxy from Si_2H_6. *Journal of Vacuum of Science and Technology A*, 15:2463–2468, 1997.

[58] M. E. Taylor and H. A. Atwater. Monte Carlo simulations of epitaxial growth: comparison of pulsed laser deposition and molecular beam epitaxy. *Applied Surface Science*, 129:159–163, 1998.

[59] J. A. Venables. Rate equation approaches to thin film nucleation kinetics. *Philosophical Magazine*, 27:697–738, 1973.

[60] D. D. Vvedensky, A. Zangwill, C. N. Luse, and M. R. Wilby. Stochastic-equations of motion for epitaxial-growth. *Physical Review E*, 48:852–862, 1993.

[61] S. C. Warnick and M. A. Dahleh. Feedback control of MOCVD growth of submicron compound semiconductor films. *IEEE Transactions on Control Systems Technology*, 6(1):62–71, 1998.

[62] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer-Verlag, New York, 1990.

[63] C. Wu, C. H. Crouch, L. Zhao, J. E. Carey, R. Younkin, J. A. Levinson, E. Mazur, R. M. Farrell, P. Gothoskar, and A. Karger. Near-unity below-band-gap absorption by microstructured silicon. *Applied Physics Letters*, 78:1850–1852, 2001.

[64] A. Zangwill and E. Kaxiras. Submonolayer island growth with adatom exchange. *Surface Science*, 326:L483–L488, 1995.