Helsinki University of Technology
Laboratory of Computational Engineering
Espoo 2004

REPORT B37

MONTE CARLO SIMULATIONS IN THE STUDY OF SEMICONDUCTOR SURFACES

Laura Juvonen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Electrical and Communications Engineering, Helsinki University of Technology, for public examination and debate in Auditorium S1 at Helsinki University of Technology (Espoo, Finland) on the 4th of June, 2004, at 12 noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Computational Engineering

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Laskennallisen tekniikan laboratorio

Distribution:
Helsinki University of Technology
Laboratory of Computational Engineering
P. O. Box 9203
FIN-02015 HUT
FINLAND
Tel. +358-9-451 4826
Fax. +358-9-451 4830
http://www.lce.hut.fi

Online in PDF format: http://lib.hut.fi/Diss/2004/isbn9512271281/

E-mail: Laura.Juvonen@hut.fi

ISBN 951-22-7127-3 (printed)
ISBN 951-22-7128-1 (PDF)
ISSN 1455-0474
PicaSet Oy
Espoo 2004

Abstract

In this thesis, the Monte Carlo method is applied for the study of semiconductor surfaces. The focus is in the investigation of structural properties of epitaxially grown thin films on the Si(001) surface. Semiconductor surfaces are typically characterized by complicated energy landscapes, and the properties of these systems are often strongly influenced by long-range elastic effects. The Monte Carlo method is an attractive choice for large-scale relaxational problems because it is not bound to the true dynamical evolution of the system. This freedom can be utilized in designing new advanced algorithms which can significantly speed up the equilibration process. When combined with the computational efficiency of classical potentials, this approach can be used to reach even experimentally accessible time and length scales.

In this work, different silicon potentials are tested to evaluate their ability to describe the properties of Si(001), including the surface reconstruction and various defect structures. Significant differences are found in the performance of the tested models. Some of the potentials are poorly suited for finite-temperature surface simulations, while the Stillinger-Weber (SW) potential gives a fairly accurate overall description of Si(001).

The SW model is applied to study the effects of lattice-mismatch induced strain in the heteroepitaxial growth of Ge on Si(001). The surface undergoes a structural evolution in which the morphological changes are driven by a complex interplay between different temperature-dependent strain-relief mechanisms. The simulations provide a good overall explanation for experimental observations.

A new hybrid Monte Carlo - Molecular Dynamics algorithm is introduced for the study of relaxational problems involving large-scale configurational rearrangement. It is designed to circumvent the problem of getting trapped into deep metastable states in systems with complicated energy landscapes. The algorithm is used here to study islands and vacancy structures on the Si(001) surface, but in general, the same approach could be applied to study other semiconductor surfaces as well.

Preface

This thesis for the degree of Doctor of Technology has been prepared in the Laboratory of Computational Engineering at the Helsinki University of Technology during the years 2000-2004. My studies were mostly funded by the National Graduate School in Materials Physics and by the Academy of Finland under the Finnish Centre of Excellence Programme 2000-2005 (project on Computational Research of Semiconductor Materials, Project No. 1169043). I would also like to acknowledge the important support from the Finnish Cultural Foundation and the Finnish Academy of Science and Letters.

I would like to express my gratitude to several people who have participated in this work and helped me tremendously over the years. First of all, I am deeply grateful to my instructor Dr. Antti Kuronen for all his ideas and day-to-day guidance dating back to the summer of 1998 when I first began working in our lab as a summer trainee. I am also very thankful to my supervisor Academy Prof. Kimmo Kaski who has always been very supportive and given me wonderful opportunities, such as visiting international research groups and getting experience in teaching. I would like to express a special thanks to Prof. David P. Landau for showing such interest in my work and for giving me a lot of extra motivation throughout the years. He also gave me the valuable opportunity to collaborate with Dr. Francesca Tavazza, which became the most rewarding experience of my graduate student years. Using Francesca's own words, working with her was not only productive but real fun! I would also like to thank my roommates Maria Huhtala and Laura Laitinen for a pleasant and relaxed atmosphere.

Finally, I wish to thank my parents Markku and Tuula for their continuous support and encouragement, and also for the great example they have given me through their own research work. I also want to thank my sister Tuuli for many nice moments during these years. Last on the list, my deepest gratitude belongs to my husband Juha who is truly the greatest source of happiness in my life. Without his understanding, and those anything-but-physics conversations, this work would probably have never been completed.

Laura Juvonen

List of publications

This dissertation consists of an overview and the following publications:

I. A. Kuronen, L. Nurminen, K. Kaski, *Computer simulation of nucleation on patterned surfaces*, in Materials Issues and Modeling for Device Nanofabrication, volume 584 of Materials Research Society Symposium Proceedings, Eds. L. Merhari, L.T. Wille, K. Gonsalves, M.F. Gyure, S. Matsui, L.J. Whitman, pages 239-244, Materials Research Society, 2000.

II. L. Nurminen, A. Kuronen, K. Kaski, *Kinetic Monte Carlo simulation of nucleation on patterned substrates*, Physical Review B, **63** 035407 (2001).

III. L. Nurminen, A. Kuronen, K. Kaski, *Simulation of the Early Stages of Growth on a Patterned Substrate*, in Computer Simulation Studies in Condensed Matter Physics XIV, Eds. D.P.Landau, S.P.Lewis, H.-B. Shuttler, pages 62-66, Springer Verlag, 2001.

IV. L. Nurminen, F. Tavazza, D. P. Landau, A. Kuronen, K. Kaski, *Monte Carlo Simulation of the Surface Structure of Ge on Si(001)*, in Computer Simulation Studies in Condensed Matter Physics XV, Eds. D.P.Landau, S.P.Lewis, H.-B. Shuttler, pages 142-146, Springer Verlag, 2002.

V. L. Nurminen, F. Tavazza, D. P. Landau, A. Kuronen, K. Kaski, *Comparative study of Si(001) surface structure and interatomic potentials in finite-temperature simulations*, Physical Review B, **67** 035405 (2003).

VI. L. Nurminen, F. Tavazza, D. P. Landau, A. Kuronen, K. Kaski, *Reconstruction and intermixing in thin Ge layers on Si(001)*, Physical Review B, **68** 085326 (2003).

VII. L. Nurminen, F. Tavazza, D. P. Landau, A. Kuronen, K. Kaski, *Simulation of islands and vacancy structures for Si/Ge-covered Si(001) using a hybrid MC-MD algorithm*, in Computer Simulation Studies in Condensed Matter Physics XVII, Eds. D.P.Landau, S.P.Lewis, H.-B. Shuttler, Springer Verlag, 2004, in press.

VIII. F. Tavazza, L. Nurminen, D. P. Landau, A. Kuronen, K. Kaski, *A new hybrid MC-MD algorithm for the study of islands and step edges on semiconductor surfaces: Application to Si/Si(001)*, Physical Review E, accepted for publication.

In the overview, these publications are referred to by their roman numerals.

Author's contributions

This dissertation is a review of the author's work in the field of Computational Materials Physics, focusing on the properties of Semiconductor Surfaces. The author, Laura Juvonen (prev. Nurminen), has given a central contribution to all of the research work comprising this thesis. The selected publications have several contributing authors because the work is part of a large research project on semiconductor materials and has been carried out in collaboration with the Center for Simulational Physics, The University of Georgia, USA.

Laura Juvonen (in publications, L. Nurminen) is the principal author of publications II-VII and is fully responsible for all of the written material in these papers (valuable comments from the other contributing authors are acknowledged). She is the second author of the publications I and VIII and has also contributed to the writing of these papers.

In the following, the details of the author's contributions are summarized. The author has written the computer program and performed all of the numerical calculations reported in publications I, II and III. She also played a key role in developing the original ideas leading to these publications. The publications IV, V and VI are based on another computer program, also written by the author. The simulations and data analysis in these three papers were to a great extent designed by the author and she did almost all of the numerical calculations. The new simulation method presented in publications VII and VIII was initiated and designed by the author in collaboration with Dr. F. Tavazza. The author contributed to developing the simulation program and numerical tools for the data analysis. She has also implemented the new algorithm in her version of the simulation program and is responsible for producing part of the data and for the writing of publication VI. Publication VIII was mostly written by F. Tavazza who is also responsible for performing the numerical calculations for this paper. The author contributed to designing those calculations and has repeated some of the simulations as a check.

Discussions with Dr. Francesca Tavazza, Prof. David P. Landau, Dr. Antti Kuronen and Prof. Kimmo Kaski have been extremely helpful throughout the whole work in developing new ideas, designing simulations and analyzing the results.

Contents of publications

The overview part of this thesis is written in the style of a monograph in order to give a broader description of the frame of reference behind this work. The main results obtained in the eight publications of this thesis are described in the overview, but the primary aim of the text is not to give an extended summary of the publications. Therefore, the contents of each publication are briefly described here.

Publications I, II and III describe kinetic Monte Carlo simulations related to island nucleation during the epitaxial growth of crystalline materials. In publication I, we introduce two models (denoted as A and B) for describing an inhomogeneous growth substrate. The inhomogeneity is incorporated into the simulations by varying the activation energies for adatom diffusion, producing a surface with a nanoscale pattern. The first results reported in publication I indicate that a periodic inhomogeneity in the activation energy can have a strong confining effect on adatom nucleation. The effect is found to be strongly dependent on the temperature.

In publication II, this subject is investigated in more detail. We describe how the patterned substrate models are related to examples of real systems in which preferential nucleation at specific sites is encountered. Model A is related to vertically correlated quantum dot superlattices in which the energy surface seen by the diffusing adatom is spatially biased by buried islands. Model B is related to the heteroepitaxial growth of Ag on a Pt(111) surface covered by 2 ML of Ag. In this case, the substrate is patterned by a regular network of dislocations which act repulsively toward the diffusing adatoms. The simulation results reveal that both models lead to confined nucleation in a narrow temperature range which produces uniformly sized and spaced islands. We demonstrate that there exists an optimal set of growth conditions for uniform nucleation which is determined by the length scale of the substrate pattern. In addition, the two patterned substrate models are found to differ significantly when the stability of the nucleated islands is examined. Model B exhibits Ostwald ripening during post-deposition annealing, while for model A, the periodic arrangement of uniformly sized islands represents thermodynamic equilibrium.

In publication III, some additional details of the spatial ordering of the islands

are investigated. We examine how the patterned substrate affects the size distribution and density of islands as a function of temperature. These results can be directly compared to experimental measurements of the Ag/Pt(111) system. Such a comparison is made in Section 3.3 of the overview where the results of publication III are plotted with experimental data.

The main part of this thesis examines the properties of epitaxially grown Si- or Ge-layers on the Si(001) surface (publications IV-VIII). In publication IV, we briefly introduce the Monte Carlo method for conducting large-scale off-lattice simulations involving the Si(001) surface. The interactions between particles are modeled using empirical potentials of the Stillinger-Weber and Tersoff forms. We first examine the general properties of the potential models and identify some shortcomings related especially to surface simulations at finite temperatures. In order to further investigate the usefulness of the classical Monte Carlo approach, we use the SW potential to study the relaxation of small Ge islands on the Si(001) surface. We observe that the Ge islands show significant edge effects caused partly by the structure of the underlying substrate and to a lesser extent by lattice-mismatch induced strain in the islands.

Publication V is an extensive comparative study where the accuracy of the empirical potential models is examined in connection with finite-temperature simulations involving the Si(001) surface. We investigate how well the potentials can describe the main properties of Si(001), including the reconstruction of the surface and the formation of defect structures, such as antiphase boundaries and dimer vacancies. We find that good performance in the static limit does not ensure the reliability of results at finite temperatures. None of the empirical potentials can give a fully accurate description of Si(001) (e.g. effects of quantum mechanical origin, such as dimer buckling, cannot be reproduced by a classical model), but in our opinion, the Stillinger-Weber model is best suited for simulations conducted at finite temperatures.

In publication VI, the SW model is applied to study a physical application. We examine the complex interplay of two strain relief mechanisms, surface reconstruction and Si-Ge intermixing, in a system composed of a thin Ge layer on the Si(001) surface. We first investigate the structure and energetics of the $2 \times n$ reconstruction. In good agreement with experimental observations, the simulations show that this periodic vacancy-line pattern is stabilized by the Ge overlayer and the optimal periodicity decreases as the Ge layer becomes thicker. Si-Ge intermixing is studied using a so-called random-switch MC algorithm. This is a direct simulation method which produces the equilibrium distribution of Ge in the subsurface layers, including entropic effects. We find that strain-relief driven intermixing has a pronounced effect on the optimal periodicity of the $2 \times n$ reconstruction. In addition, the Ge atoms show a clear preference to certain lattice sites in subsurface layers which can be explained by the presence of local strain fields near surfaces.

In publications VII and VIII, we introduce a new hybrid MC-MD algorithm which has been designed for the study of semiconductor surfaces. These systems are typically characterized by complicated energy landscapes, and consequently, traditional Monte Carlo or Molecular Dynamics algorithms can easily get trapped into deep metastable states. In order to overcome these problems, we have developed a method which speeds up the equilibration of the system by utilizing collective moves of several particles. The version presented in publications VII and VIII (the so-called dimer-jump algorithm) is designed for the study of Si(001) surfaces, but the idea itself is rather general and could potentially be expanded to study other semiconductor surfaces as well. In publication VII, we give a brief description of the dimer-jump algorithm and its implementation. We then apply the method to study two physically interesting phenomena: the relaxation of islands and the formation of dimer-vacancy structures. Publication VIII gives a detailed description of the hybrid MC-MD method, along with a discussion of problems encountered in simulations with traditional algorithms. We explain how the collective MC trial moves are constructed and why it is necessary to include an MD-driven relaxation. In order to illustrate how the method works, the algorithm is applied to study the early stages of island formation, island stability and step-edge evolution on the Si(001) surface. In both publications, all the simulation results are found to be in good qualitative agreement with experimental observations.

Contents

Chapter 1

Introduction

Understanding the properties of semiconductor surfaces plays a key role in developing new materials for future electronic and optoelectronic devices. At the moment, silicon-based technology - with devices grown on semiconductor surfaces - dominates the field with an over 90% share of the total market. The properties of the Si(001) surface have been intensively investigated for several decades, not only due to its immense technological importance, but also because this surface exhibits a surprisingly rich variety of theoretically interesting phenomena. Currently, mixed SiGe systems are viewed as the most promising candidate for novel high-frequency nano-scale devices, and especially due to the difficulties encountered in manufacturing small-scale structures, understanding the properties and growth characteristics of these materials is extremely important.

There are various experimental methods for studying surfaces, developed to the point where real-space atomic-scale imaging of the surface structure has become possible. Using techniques such as the atom-tracking scanning tunneling microscopy (STM), experimentalists are able to even follow the motion of a single atom as it diffuses on the surface. Nevertheless, there are several problems that still remain outside the reach of even the most advanced experimental methods. For example, distinguishing between Si and Ge atoms is extremely difficult due to their similar atomic, electronic and chemical properties. This has prevented experimentalists from performing e.g. a quantitative analysis of the degree of intermixing in Si(001) systems covered with a thin layer of Ge.

These types of difficulties can often be overcome with the aid of computational methods which have become a vital tool for investigating the properties of semiconductor surfaces. On one hand, computational studies have allowed scientists to interpret experimentally observed features and to form a theoretical understanding of the underlying phenomena. On the other hand, computational methods can be used to address difficult problems that are beyond the limits of current experimental techniques. At their best, the results of simulations can be

used to predict the behavior of systems under conditions which have not or cannot be studied experimentally. Nevertheless, computational studies should not be considered as an alternative to experiments, but on the contrary, these two approaches should complement one another in forming the foundation for more complete theoretical understanding.

The Si(001) surface, like most semiconductor surfaces, is characterized by an extremely complicated energy landscape which results from the directional bonding and reconstruction of the surface. The behavior of such systems is often strongly influenced by the presence of elastic effects which have to be included in the computational method in order to reproduce the physics correctly. In addition, the interesting physical phenomena often involve time and length scales which propose extreme challenges to current computational techniques.

The objective of this thesis is to investigate the possibilities offered by large-scale Monte Carlo simulations. In this approach, we utilize the computational efficiency of classical interatomic potentials which allows us to reach sufficiently large system sizes while retaining an atomic-scale description of the system. In Chapter 2, we begin by describing the basic methodology related to Monte Carlo simulations, focusing especially on off-lattice simulations of surfaces. A brief description of the Molecular Dynamics method is also provided, and we describe the functional form and parametrization of two empirical silicon potentials which are used in the MC simulations. In Chapter 3, we discuss the general properties of epitaxial systems. In this part, we use a simplified lattice model and focus, in a very general sense, to the kinetic aspects of crystal growth.

In Chapter 4, we give an overall description of the Si(001) surface and discuss various properties related to surface defects, such as antiphase boundaries and dimer vacancies. These properties are also used as a testing ground for evaluating the suitability of the empirical potential models for use in finite-temperature simulations involving the Si(001) surface. In Chapter 5, we investigate the structural evolution encountered during the growth of Ge on Si(001). We begin by studying the energetics of the $2 \times n$ reconstruction, and then extend the study to include the possibility of Si-Ge intermixing. A so-called random-switch algorithm is implemented for the intermixing simulations, and it serves as a good example of the flexibility of the Monte Carlo method. In Chapter 6, we introduce a new hybrid MC-MD algorithm which we have developed in order to circumvent the problem of getting trapped in deep metastable states in systems with complicated energy landscapes. We describe how the method is constructed in the case of the Si(001) surface. As application examples, we study the formation and relaxation of Si islands and the formation of vacancy structures on Si or Ge-covered Si(001). Finally, Chapter 7 gives a brief summary and concluding remarks.

Chapter 2

Simulation methods

2.1 Metropolis Monte Carlo

2.1.1 Algorithm

In a general form, the term *Monte Carlo* (MC) corresponds to a wide range of stochastic methods which are based on the use of random numbers. In this section, we review the principles of the classical Metropolis algorithm [1] which was developed in 1953 for the calculation of thermal averages according to the principles of statistical mechanics in a given equilibrium ensemble.

In a Monte Carlo simulation, a set of random numbers is used to determine the sequence of states through which the system evolves. Consequently, the changes in the system do not occur in a predefined fashion, i.e. deterministically, but in a stochastic manner such that average quantities obtained from two independent runs agree within some statistical error. The time in connection with Metropolis Monte Carlo simulations is referred to as the "Monte Carlo time" and it is not directly related to the real physical time [2]. In contrast, the kinetic Monte Carlo method, which is discussed in section 2.2, has been specifically designed to describe the dynamical evolution of the system, and therefore it can be used to study time-dependent physical properties. In the following, the theoretical foundation behind the Monte Carlo method is reviewed and used to derive the basic form of the Metropolis algorithm.

The state space of a given system $S = \{C\}$ consists of all possible configurations C. The probability distribution of configurations at time t is denoted by $P(C,t)$. The time evolution of $P(C,t)$ is described by the *Master equation*

$$\frac{\partial P(C,t)}{\partial t} = -\sum_{C'} W(C \to C')P(C,t) + \sum_{C'} W(C' \to C)P(C',t), \qquad (2.1)$$

where $W(C \to C')$ is a transition rate from state C to state C'. In equilibrium, we obtain a unique distribution $P(C)$ which satisfies the following stationary condi-

tion: $\partial P(C,t)/\partial t = 0$. In practice, this can be replaced by a stronger condition called *detailed balance*:

$$W(C \to C')P(C,t) = W(C' \to C)P(C',t). \tag{2.2}$$

In a canonical ensemble, the probability distribution is given by

$$P(C,t) = \frac{1}{Z} e^{-E_C/k_\mathrm{B}T}, \tag{2.3}$$

where E_C is the energy of the configuration C, T is the temperature, k_B is the Boltzmann constant and Z is the partition function. The probabilities $P(C,t)$ are usually not known due to the denominator, but we can set up a Markov chain by generating a sequence of configurations that converges to the desired equilibrium distribution $P(C)$. If we calculate the ratio $P(C',t)/P(C,t)$, we notice that the two partition functions cancel out, which leads to the following form of the detailed balance condition

$$\frac{W(C \to C')}{W(C' \to C)} = \exp(-\Delta E/k_\mathrm{B}T), \tag{2.4}$$

where $\Delta E = E_{C'} - E_C$.

Any transition rate which satisfies the detailed balance is acceptable. The oldest and most common choice was introduced by Metropolis *et al.* [1] in 1953:

$$W(C \to C') = \begin{cases} \tau_0^{-1} \exp(-\Delta E/k_\mathrm{B}T), & \Delta E > 0 \\ \tau_0^{-1}, \quad \text{otherwise}, \end{cases} \tag{2.5}$$

where τ_0 is a unit time step between attempted moves (often τ_0 is set to unity). In some systems, the parameter τ_0 can be set equal to some characteristic time scale of the system, and consequently the MC time may be interpreted as a measure of the "effective physical time". Such an interpretation is well justified for example in simulations of processes which are random by nature (e.g. decay or diffusion). In general, the time step in MC simulations does not correspond to real physical time.

We will now describe the structure of a basic Metropolis Monte Carlo algorithm. As an example, let us consider a system of N interacting particles. The interactions between the particles are described by a potential $\mathcal{U}(\mathbf{r}^N)$, where $\mathbf{r}^N$ denotes the coordinates of all the particles. The Metropolis algorithm can be implemented by the following simple recipe:

1. Choose an initial state (possibly use a lattice that is suitable for the system in question).

2. Randomly choose a particle with label i.

3. Calculate a random displacement for the particle: $\mathbf{r}_\mathbf{i}' = \mathbf{r}_\mathbf{i} + \delta$, and calculate the corresponding energy change $\Delta\mathcal{U} = \mathcal{U}(\mathbf{r}'^N) - \mathcal{U}(\mathbf{r}^N)$.

4. If $\Delta\mathcal{U} < 0$, accept the move and go to 2.

5. If $\Delta\mathcal{U} > 0$, generate a random number r such that $0 < r < 1$. If $r < \exp(-\Delta\mathcal{U}/k_\mathrm{B}T)$, accept the move. Otherwise keep the old configuration. Go to 2.

In this algorithm, "Monte Carlo time" is measured in units of Monte Carlo steps (MCS) which corresponds to one trial move per each atom in the system.

For relaxational models (for which the time-dependent behavior is described by the Master equation, Eq. 2.1), this algorithm generates a sequence of states that converges to the desired equilibrium distribution. In a typical MC run, the system first passes through a transient stage during which it relaxes toward equilibrium. During the relaxation, the internal energy and other quantities are changing, usually with different characteristic time scales. Once equilibrium is established, the probability of visiting a particular state n is proportional to the Boltzmann factor $\exp(-E_n/k_\mathrm{B}T)$. Measured physical quantities show thermodynamic fluctuations but their average values remain constant. At this stage, we can begin to measure averages $\langle A\rangle = \sum_n P_n A_n$ by sampling over the set of generated states. Note that if the move is rejected, the old state is kept and counted again for the average.

In a more complex system, we might not know what the equilibrium state of the system is like. In some cases, the relaxation period might be extremely long or the system can even get trapped in a metastable state and never reach equilibrium. It is thus very important to be careful about checking the results. A simple check is to repeat the run by starting from different initial states to see if the same equilibrium distribution is reached. It is also worthwhile to make one very long run to see if anything unexpected appears at a later time.

2.1.2 Off-lattice models

In most textbooks, the Metropolis algorithm is described in connection with spin-lattice models (e.g. the Ising model). In this thesis, we concentrate mostly on simulations of atomic systems using off-lattice models with classical interaction potentials. In this approach, the particles are not restricted to certain fixed lattice sites but are allowed to occupy all possible positions in a three-dimensional space. The crystalline structure of materials, such as the diamond lattice of silicon, is stabilized by an interatomic potential $\mathcal{U}$, which is a function of the x, y and z

coordinates of all the atoms in the system. For an N-particle system, the particle positions are continuous variables, denoted by $\mathbf{r}^N$.

In the case of the canonical distribution, the probability of observing a configuration $\mathbf{r}^N$ is given by the Boltzmann weight

$$P(\mathbf{r}^N) = \frac{1}{Z} e^{-\mathcal{U}(\mathbf{r}^N)/k_\mathrm{B}T}, \tag{2.6}$$

where $\mathcal{U}(\mathbf{r}^N)$ is the total potential energy of the system. In some cases it is assumed that $\mathcal{U}(\mathbf{r}^N)$ is simply a sum of pairwise interactions $\mathcal{U}(\mathbf{r}_i - \mathbf{r}_j)$ over all interacting particle pairs in the system. But as we see later, such a model is clearly not suitable for accurate description of most materials. For example, it is necessary to include a three-body term into the potential in order to model the tetrahedral bonding in Si and Ge.

2.1.3 Boundary conditions

The aim of Monte Carlo simulations of atomic systems is to provide information about the properties of a macroscopic sample. The size of feasible simulation systems ranges, however, from a few hundred to a few million constituent particles of the system at best, which is far from the thermodynamic limit. Therefore, we cannot assume that the treatment of boundaries has a negligible effect on the outcome of the simulation.

In order to simulate systems in bulk-like conditions, it is important to treat the boundaries such that edge effects are effectively eliminated. This can be achieved by employing so-called *periodic boundary conditions* (pbc) where the volume containing the N particle system is considered to be the unit cell of an infinite periodic lattice (see Fig. 2.1 for illustration). Each particle now interacts not only with all other particles in the same cell but also with all other particles in the entire infinite lattice (including its own periodic image). If an accepted move takes a particle outside the volume of the unit cell, it is placed back inside the cell from the opposite "wall", thus conserving the total number of particles in the unit cell.

This procedure effectively eliminates effects due to the boundaries, but the system is still affected by the finite size of the unit cell. For example, one consequence of the periodicity of the model system is that the wavelength of allowed fluctuations is determined by the periodicity of the lattice. The maximum possible wavelength is the one that just fits the box: $\lambda = L$. For this reason, one should always check for effects caused by the finite system size; e.g. by performing several test runs using different system sizes.

For systems with continuous interatomic potentials, the potential energy calculation involves an infinite sum due to the periodic boundaries, unless the interactions are truncated beyond a certain cutoff distance. For short-range inter-

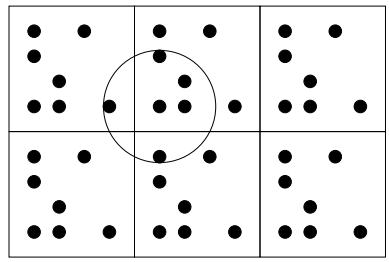

Figure 2.1: Illustration of the periodic boundary conditions. The circle represents a cutoff distance: the central atom interacts only with other atoms inside the circle.

actions such a truncation can usually be done without introducing errors into the simulation. Long-range interactions are considerably more difficult to handle since a truncation will seriously influence the physics of the system. In this thesis, we leave out the discussion of methods designed for the treatment of long-range forces, since classical potentials for semiconductor systems are usually sufficiently short ranged.

In the following, we will consider the treatment of short-range interactions in simulations with periodic boundary conditions. By short ranged, we mean that the largest contribution to the potential energy of a given particle comes from neighboring particles within some cutoff distance r_c, and therefore the effect of truncating the interactions will hopefully be small. For Monte Carlo simulations, a simple truncation of the potential at r_c is usually sufficient, whereas in Molecular Dynamics, the situation is somewhat more complex because the truncation can introduce a discontinuity into the force. In MD simulations, it is therefore more common to both truncate and *shift* the potential such that it vanishes at the cutoff. The choice of the cutoff distance depends on the potential and the particular system in question. In MC simulations of semiconductor systems, the cutoff distance r_c is usually clearly shorter than the size of the periodic box L, and thus for any particle i only interactions with the nearest image particles need to be considered (only one image of each particle j can contribute to the potential energy of particle i). This is called the *nearest periodic image convention*.

2.1.4 Slab geometry

In simulations of surfaces, the boundary conditions of the system are modified such that periodic boundaries are used in two directions (parallel to the surface) and free-edge boundary conditions are applied in the third direction (perpendicular to the surface). The unit cell is repeated periodically in two dimensions, thus creating two infinite surfaces. This is called *slab geometry*. In order to prevent the two surfaces from interacting with each other, it is important to use suffi-

ciently many atomic layers in the bulk section of the slab (in the z direction). This question should be given careful consideration because the finite thickness of the simulation slab may influence the outcome of the simulations in unexpected ways.

In publication V, we observed that the 2×1 reconstruction of the Si(001) surface induces anisotropic changes in the x and y dimensions of the simulation slab, which significantly influences the results of the simulations. Our investigation shows that the effects are caused by the finite thickness of the simulation slab, although not due to a direct interaction between the two surfaces. For example, twenty atomic layers is a sufficient thickness to prevent the two surfaces from interacting with each other through the bulk section of the slab (e.g. the average energy per particle assumes a constant value in a few atomic layers below the surface). However, in a 20-layer thick slab, the number of surface atoms is still a fair fraction of the total number of particles in the system, which means that the presence of the surfaces can distort the lattice structure of the whole simulation cell. In the case of Si(001), we observed that the average separation between atomic layers in the z direction shows clear oscillations through the whole bulk section of the slab.

Figure 2.2(a) shows our results for three different empirical silicon potentials (see Section 2.4 for description of the potentials). We notice that the amplitude of the oscillations stays constant throughout the bulk and does not decrease as a function of distance from the surface, as would be expected of a regular surface effect. These oscillations do not occur in bulk simulations (where periodic boundaries are applied in all three directions) which confirms that the lattice distortion is caused by the presence of the surface. When the simulations were repeated using varying slab thicknesses, we observed that the amplitude of the oscillations decreases when the slab becomes thicker. Figure 2.2(b) shows a comparison of 20- and 80-layer thick slabs. This reduction in the oscillation amplitude clearly indicates that this effect is caused by the finite slab thickness.

The reason for these oscillations is that the dimerization of the Si(001) surface produces an anisotropic change in the surface size of the simulation cell (for a description of the surface reconstruction, see Chapter 4). The system shrinks in the direction of the dimer bonds, while it expands in the other direction. In the bulk section of the slab, these changes are accommodated by an oscillating decrease/increase in the interlayer distances. Of course, this situation is unphysical since the purpose of the simulations is to study the properties of the surface on a much larger substrate. Our suggestion for improving the situation is to use an odd number of atomic layers in the slab which produces two surfaces with perpendicular dimer orientations. Using this odd-layer geometry, we observe no oscillations because the anisotropic effects due to the two surfaces cancel out. However, the problem cannot be fully solved this way because in many simulations the two surfaces are not identical in their atomic structure (see e.g. Chapter 6). In addition, in heteroepitaxial systems, such as Ge-covered Si(001), the lattice-mismatch in-

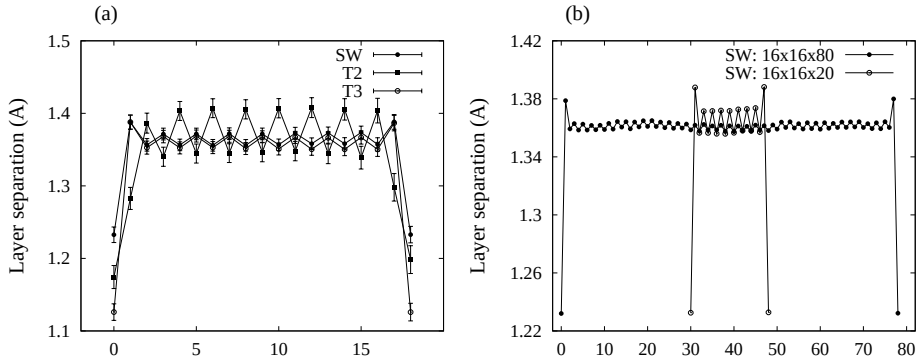

Figure 2.2: Separation between atomic layers in the direction perpendicular to the (001) plane. (a) Comparison of three different empirical silicon potentials: Stillinger-Weber (SW) and two parametrizations of the Tersoff potential (T2 and T3). The simulation slab is 20 atomic layers thick and the surface size is 16×16 atoms. (b) Effect of slab thickness. Data is shown for slab thicknesses of 20 and 80 atomic layers.

duced strain may cause other types of changes in the simulation cell (e.g. isotropic expansion) which cannot be eliminated by the odd-layer geometry. Therefore, this question should always be given careful thought when performing surface simulations.

2.1.5 Verlet lists and cell-linked lists

In large systems, it is computationally very inefficient to search through the entire system when looking for atoms within a cutoff distance $r_c \ll L$ of particle i. A simple method to reduce the amount of work is to construct a list of neighbors for each particle in the system. This so-called Verlet list contains the indices of all those neighboring particles which are within the distance r_c from the particle in question. After each accepted MC move, some of the neighboring particles may have moved out of the "interaction volume" and new ones may have entered. Thus we need to reconstruct the list after each accepted move. A way to avoid this recalculation after each accepted step is to include all particles within some distance $r_{\max} > r_c$ where $r_{\max} - r_c = n\delta_{\max}$ is large enough so that no particle may enter within r_c in n MC steps of maximum size $\delta_{\max}$. In this case, the Verlet list needs to be recalculated only after n MC steps.

An additional way to reduce the computational effort is to use a so-called cell list or linked-list method. The idea is to subdivide the system into smaller cells whose size is only slightly larger than the cutoff radius r_c. Each particle now interacts with other particles in the same cell or in neighboring cells, and so the remainder of the system need not be searched. This is a very fast method since the

allocation of a particle to a given cell is an operation that scales with N and the total number of cells that need to be searched for the calculation of the potential energy is independent of the system size. The linked-list method can be used either together with the Verlet list or independently.

2.1.6 Constant-pressure simulations

The isobaric-isothermal (NpT) ensemble is often used in off-lattice Monte Carlo simulations of both fluids and solids [2]. In many cases, using an extensive variable (like the volume V) as a control parameter of the simulation may constrain the system in an undesirable way. For example, in a constant-volume simulation, the system may be kept at a density where it would like to phase separate into two coexisting phases of different density but is prevented from doing so by finite-size effects. In studies of crystalline solids, we are often interested in the transformation of one crystalline structure to another or in the change of shape of the unit cell with temperature or applied stress. In such simulations, it is extremely important that the system has enough freedom to change shape without creating highly stressed configurations or defect structures.

We now describe the method which is used for constant-pressure simulations of crystalline solids. The derivation from statistical mechanics can be found e.g. in Ref. [3]. Since solids are intrinsically anisotropic, we need to consider anisotropic changes in the volume as a means of keeping the pressure constant at $P = 0$. In a Monte Carlo simulation, this is achieved by introducing an additional trial move, called a *volume variation*, into the Metropolis algorithm. In the basic algorithm, each MC step consists of one attempted random displacement of each individual particle. In the constant-pressure algorithm, we also attempt to change the linear sizes of the simulation cell, L_x, L_y and L_z. This is done by randomly choosing new values, L'_x, L'_y and L'_z, slightly altered from the previous ones. The positions of the atoms are scaled by the relative change in the linear size: $x' = xL'_x/L_x$, $y' = yL'_y/L_y$ and $z' = zL'_z/L_z$. The potential energy change, $\Delta\mathcal{U}$, associated with this global distortion of the system, is not the only quantity which now enters the Metropolis criterion for accepting the move. Instead, we have to add a term which describes the change in translational entropy of the system. The resulting change in the generalized Hamiltonian is given by [4]

$$\Delta H_{\text{eff}} = \Delta\mathcal{U} - Nk_{\text{B}}T \ln \frac{L'_x\, L'_y\, L'_z}{L_x\, L_y\, L_z}. \tag{2.7}$$

The frequency at which the volume changes should be tried depends on the efficiency with which the volume space is sampled. We can use the following criterion of efficiency [3]

$$\frac{\text{sum of squares of accepted volume changes}}{t_{\text{CPU}}}.$$

Thus the frequency of the attempted volume moves should depend on their cost. In general, a trial volume variation requires the calculation of all interactions in the system, and therefore it is comparable to carrying out N single-particle moves. In practice, it is common to choose that volume changes are attempted e.g. every 10 MCS.

An extension of the constant-pressure algorithm is available for simulations in the isotension-isothermal ensemble for cases where the system is under non-hydrostatic pressure (e.g. uniaxial stress). Description of the extended method can be found in Ref. [3].

2.1.7 Advanced techniques

The idea behind many advanced MC techniques is to utilize the fact that the Monte Carlo method is not limited by the natural dynamics of the system but we have a freedom to develop trial moves which can result in large jumps in phase space. The large jumps are usually designed to help the system to explore the phase space more rapidly by creating direct paths from one region to other regions which would be difficult to reach using only the single-particle jumps. For example, equilibration is often sufficiently fast at high temperatures, while at lower temperatures the system can get easily trapped in deep metastable states and thus the global free-energy minimum cannot be reached within a reasonable simulation time. Several advanced Monte Carlo techniques have been develop to tackle such problems. In the following, we review the method of parallel tempering and discuss the use of hybrid schemes. In Chapter 6, we give a detailed description of a new hybrid MC-MD algorithm which we have developed for simulations of semiconductor surfaces.

Parallel tempering is a method which has been developed to achieve good sampling of systems with complicated energy landscapes. It is also called the method of multiple Markov chains [5, 6] and exchange Monte Carlo [7] and it is similar to the method of simulated tempering [8, 9]. In the parallel tempering approach, we consider a set of n replicas of a given system. This set is simulated in a given ensemble, but the thermodynamic state is different in each replica of the system. In the simplest case, these states differ in temperature. Other choices can be intensive variables such as the chemical potential, but the formalism of the parallel tempering approach is also applicable to any parameter that appears in the effective Hamiltonian of the system [10]. For convenience, we illustrate the parallel tempering method for a system of replicas at different temperatures.

Consider a system which is characterized by multiple local minima in the potential energy landscape. The replicas of the system simulated at high temperatures can overcome the potential barriers, while the low-temperature replicas get trapped in the local minima. The idea of the parallel tempering algorithm is to introduce MC trial moves which attempt to swap replicas belonging to different

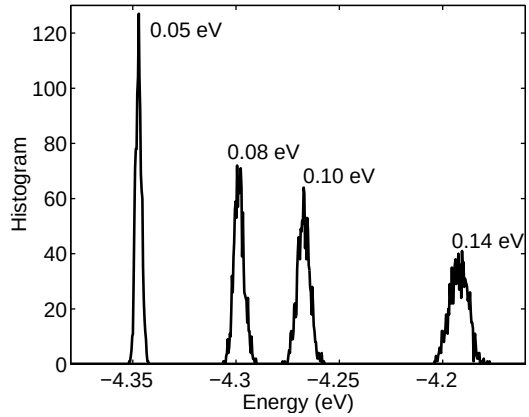

Figure 2.3: Examples of the energy distribution at four different temperatures (0.05, 0.08, 0.10 and 0.14 eV) for a system composed of a Ge island on the Si(001) surface. In the parallel tempering method, the energy distributions of neighboring replicas must have sufficient overlap. Thus at least one or two intermediate temperatures should be used between the systems at 0.08 and 0.10 eV and many more in the other intervals.

thermodynamic states. The trial move will have a very low acceptance probability if the two thermodynamic states are very different. Thus we should only attempt a swap between replicas with a small temperature difference.

Let us assume that the n replicas are arranged according to an increasing temperature scale, $T_1 < T_2 < \ldots < T_n$, where T_i is the temperature of replica i. In order to have a reasonable acceptance rate for swaps between neighboring replicas i and $i+1$, the energy distributions of the neighboring replicas should overlap. This is achieved by choosing the sequence of temperatures such that the temperature difference between neighboring replicas is sufficiently small. Figure 2.3 shows typical energy distributions obtained for an example system at four different temperatures. These distributions have no overlap between them, thus many more replicas should be used in the parallel tempering method if the whole temperature range from 0.05 eV to 0.14 eV were to be covered. From Fig. 2.3 we also notice that the energy distribution becomes highly-peaked and narrower as the temperature decreases. This means that if we are to cover a large range of temperatures, then we should not use a uniform distribution of the set of replicas, but smaller subintervals should be used between neighboring replicas at the low-temperature end of the range.

We are now ready to introduce an MC move which attempts to swap two neighboring ensembles. The acceptance probability for the swap can be derived from the condition of detailed balance (see e.g. Ref. [3]). Using the standard Metropolis scheme, the attempted swap is accepted with probability $\min(1, B)$ in

which B is given by the ratio of transition probabilities

$$B = \frac{w[(i,\beta_i),(j,\beta_j) \rightarrow (i,\beta_j),(j,\beta_i)]}{w[(i,\beta_j),(j,\beta_i) \rightarrow (i,\beta_i),(j,\beta_j)]} \tag{2.8}$$

$$= \frac{\exp[-\beta_i \mathcal{U}(j) - \beta_j \mathcal{U}(i)]}{\exp[-\beta_i \mathcal{U}(i) - \beta_j \mathcal{U}(j)]} \tag{2.9}$$

$$= \exp\{(\beta_j - \beta_i)[\mathcal{U}(i) - \mathcal{U}(j)]\} \tag{2.10}$$

Here β denotes the inverse temperature, (i, β_α) $(\alpha = i, j)$ configuration i at temperature β_α, and $\mathcal{U}(i)$ is the total potential energy of configuration i.

These swaps are computationally very inexpensive since the total energy of both configurations is known before the swap is attempted and the acceptance criterion does not require any further long calculations. The swap moves do not change the Boltzmann distribution corresponding to a particular ensemble (i.e. we can obtain ensemble averages of the n replicas in the same way as we do in a normal MC simulation). The true benefit of the method comes when it is applied to the study of systems with complicated energy landscapes. There the systems which are initially simulated at high temperatures can reach equilibrium fairly quickly. Through the swaps these equilibrated configurations are able to "travel" down to lower temperatures, which in turn enables us to get the low-temperature ensembles to jump over the potential barriers and eventually the system to reach equilibrium.

Let us now move on to the subject of hybrid Monte Carlo techniques. In conventional Monte Carlo algorithms, only the coordinates of a *single* atom are changed in a trial move. Consequently, such algorithms can be inefficient in representing collective moves of a larger set of atoms. In the hybrid Monte Carlo technique [11], the idea is to use MD to generate the trial MC moves. The advantage in doing this is that one can, for example, use a time step that would be too long for normal MD, but the resulting configuration may still constitute a good MC trial move. Such a move will not conserve energy, but this is acceptable in MC as long as the algorithm is time-reversible and area-preserving. A systematic method for constructing time-reversible, area-preserving MD algorithms is available [12]. In the trial move, the initial velocities of the particles are drawn from a Maxwell-Boltzmann distribution, and often the move consists of a sequence of MD steps. In the final step of the hybrid MC, the MD-generated trial configuration is either accepted or rejected using the Metropolis criterion.

The acceptance probability for the trial moves depends on the length of the MD time step and the number of steps used to generate a single move. For these reasons, the hybrid MC can be advantageous but it does not offer dramatic improvement over normal MD. In some cases, however, the efficiency of the hybrid MC method can be developed further. For example, a computationally demanding many-body potential can sometimes be replaced, to a first approximation, by

a less demanding pair potential. The pair potential could then be used in the MD steps to create the trial configuration, whereas the final Metropolis acceptance test would be performed using the many-body potential.

In another variation of the hybrid scheme, one utilizes the fact that trial moves in the MC method are not restricted by the real dynamics of the system. In Ref. [13], a hybrid MC algorithm is devised using fictitious dynamics which turns out to improve the sampling of polymer conformations compared to normal MD. In this example, the kinetic part of the Hamiltonian is adjusted to speed up conformational changes, but the potential energy part of the modified Hamiltonian is left unaltered.

In studies of semiconductor surfaces, the normal MC and MD algorithms can easily get trapped in a metastable state related to reconstruction of the surface. In Chapter 6, we describe a new algorithm which we have developed in order to overcome the problems related to the deep potential barriers on semiconductor surfaces. In this so-called *dimer-jump algorithm*, we introduce a trial MC move which consists of a large displacement of a pair of atoms (a dimer), followed by a local relaxation using MD. The MD part is essential for the algorithm because otherwise the acceptance rate for the long displacements would be extremely low. A full description of the algorithm is given in Chapter 6.

2.2 Kinetic Monte Carlo

Kinetic Monte Carlo (KMC) is a method for solving kinetic equations [14, 15]. The purpose of KMC is to reflect the time evolution of the system and to reproduce non-equilibrium processes. In contrast to the equilibrium nature of the Metropolis algorithm, the KMC method is well suited for studying time-dependent phenomena such as diffusion and growth processes. In the following, crystal growth is used as our example in order to describe how the KMC algorithm is constructed.

In a KMC simulation, the sequence of new configurations is created through moves that correspond to real physical events taking place in the system. The simulation model is constructed by first identifying all the possible processes which may occur in the system at any stage of the simulation. In crystal growth, typical processes include e.g. the deposition of new adatoms and different surface diffusion processes. Let N be the number of possible events which occur at a certain rate R_a ($a = 1, ..., N$) (probability per unit time) in a given configuration C. Both N and the rates $\{R_a\}$ depend on the configuration C. We now define the total rate by

$$\mathcal{Q} = \mathcal{Q}(C) = \sum_{a=1}^{N} R_a. \tag{2.11}$$

The transition probabilities are formally written as

$$W(C \to C') = \sum_{a=1}^{N} R_a V^a(C \to C'), \tag{2.12}$$

where the stochastic matrix $V^a(C \to C')$ specifies whether the transition $C \to C'$ can be realized by event a. The occurrence probability of event a is $R_a/\mathcal{Q}(C)$, which means that events are selected with probabilities proportional to their physical rates.

We will now describe the *BKL algorithm* [16] which is one of the earliest implementations of the KMC method and often used in simulations of crystal growth. The basic idea is that at each MC step one process is selected with its corresponding probability and then also realized. Because of this approach, the algorithm is not slowed down by unsuccessful attempts. The performance of the BKL algorithm can usually be further improved by applying it in a form known as the *N-fold way* which refers to grouping of the individual events according to their rates.

In order to illustrate this approach, consider having a system with n groups of different processes (e.g. deposition of an atom, diffusion jump occurring with a certain probability, etc.). The groups are labeled by $\alpha = 1, ..., n$, and their corresponding rates by ρ_α. In a given configuration C, a process α can be realized in n_α ways which is called the multiplicity of the process. For example, there might be a number of surface atoms with identical surroundings which have the same probability to diffuse and are thus members of the same group. Next, we assign a relative rate, $q_\alpha(C) = n_\alpha \rho_\alpha$, to each kind of process. The total transition rate in a configuration C is now $\mathcal{Q}(C) = \sum_{a=1}^{n} n_\alpha(C)\rho_\alpha$.

The algorithm for the kth step is as follows.

1. Choose a random number r_1 in the range $[0, \mathcal{Q}(C_k))$.

2. Decide which kind of process will take place by choosing the first index σ for which $\sum_{\alpha=1}^{\sigma} q_\alpha(C) \geq r_1$.

3. Select a realization of the process σ. Technically this can be done with the help of a list of coordinates for each kind of movement, and an integer random number r_2 in the range $[1, n_\sigma(C_k)]$. r_2 is generated and the corresponding member of the list is selected.

4. Perform the selected movement.

5. Update the multiplicities n_α, relative rates q_α, total rate Q and any data structure being used.

In principle, the KMC method will reproduce the correct dynamical evolution of the system given that we have accurate *a priori* information of the system and a complete list of all possible events can be constructed as an input to the simulation. In practice, this is a serious shortcoming because in many cases constructing such a list is an impossible task (e.g. transitions may involve concerted motion of multiple atoms and occur in completely unexpected ways).

In a recently developed off-lattice variation of the KMC method, the state-dependent rate catalog is not computed in advance but built on the fly [17]. In this self-learning scheme, the so-called Dimer Method of Henkelman *et al.* [18] is used to scan for possible diffusion mechanisms each time the system enters a new state. The corresponding activation barriers are then computed to construct the rate catalog for the current state, and the system is evolved in the same way as in the standard KMC. The on-the-fly KMC method offers substantial improvement with respect to the standard KMC, e.g. unexpected events can be found during the simulation. It can also reach time scales which are inaccessible even with the accelerated MD methods. As in all KMC simulations, the drawback is that it is difficult to know whether the list of possible diffusion mechanisms is complete for a given configuration or if some improbable events have been left out. Even extremely infrequent events may be significant for allowing the system to escape from a deep local superbasin of states [17].

2.3 Molecular Dynamics

Molecular Dynamics (MD) is a simulation method for studying equilibrium and transport properties of classical many-body systems. By classical it is meant that the nuclear motion of the constituent particles obeys the laws of Newtonian mechanics. The idea in an MD simulation is simply to set up the system and to solve the Newton's equations of motion for the collection of mutually interacting particles.

For simplicity, we begin by considering the NVE ensemble in which the number of particles N, the system volume V and the total energy E remain constant. The coupled equations of motion may be derived for example from Lagrange's or Hamilton's equations. In the former approach, Lagrange's equations for the system of N particles produce a set of $3N$ equations to be solved [2]:

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i = -\nabla_{r_i} \mathcal{U}(\mathbf{r}^N), \tag{2.13}$$

where m_i is the mass of particle i, $\mathbf{F}_i$ is the total force acting on it, and $\mathcal{U}$ is the appropriate interatomic potential. In the latter approach, where Hamilton's equations are used to derive the dynamics of the system, we obtain the following

set of $6N$ equations:

$$\begin{cases} \dot{\mathbf{r}}_i &= \mathbf{p}_i / m_i \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i , \end{cases} \tag{2.14}$$

where $\mathbf{p}_i$ is the momentum of the particle. In this approach, the energy of the system is invariant with time so that solving these equations would produce a sequence of states in the microcanonical ensemble. Either set of equations can be solved by finite difference methods using a time interval Δt which must be made sufficiently small for accurate results.

In contrast to the stochastic nature of Monte Carlo simulations, the particles in MD simulations follow the true trajectories dictated by the dynamics of the system. This means that we are able to address time-dependent properties of the system, such as diffusion of particles on a surface. Nevertheless, it is important to realize that the purpose of an MD simulation is not to predict precisely what will happen to a system that is initially in a precisely defined state. In fact, for almost all systems, the trajectory of the system through phase space is sensitively dependent on the initial conditions. This means that the trajectories of two systems which were initially very close to one another will diverge exponentially as time progresses. Therefore, the aim of the simulation is to predict the average behavior of the system in a statistical sense.

A basic Molecular Dynamics algorithm consists of the following steps:

1. *Initialization of the system*.
   In the beginning, we must assign initial positions and velocities to all particles in the system. The particle positions should be chosen such that they correspond sufficiently closely to the structure that is to be simulated. The initial velocities can be drawn directly from the Maxwell-Boltzmann distribution corresponding to the simulation temperature, or they can be assigned by first drawing each velocity component from a uniform distribution in the interval $[-0.5, 0.5]$, then shifting all velocities such that the total momentum is zero and finally scaling to adjust the mean kinetic energy to the desired value. The last step is based on the fact that, in thermal equilibrium, the following relation should hold:

   $$\langle v_{\alpha,i}^2 \rangle = k_\mathrm{B} T / m_i , \tag{2.15}$$

   where $v_{\alpha,i}$ is the velocity component α ($\alpha = x, y, z$) of particle i. This relation can be used to define an instantaneous temperature at time t:

   $$k_\mathrm{B} T(t) \equiv \sum_{i=1}^{N} \frac{m_i v_{\alpha,i}^2(t)}{N_f} \tag{2.16}$$

where N_f is the number of degrees of freedom ($N_f = 3N - 3$ for a system composed of N particles). The instantaneous temperature $T(t)$ can be adjusted to match the desired temperature T by scaling all velocities by a factor $\sqrt{T/T(t)}$. The initial velocities are, however, not particularly critical as the instantaneous temperature will fluctuate during the simulation. The relative fluctuations in the temperature will be of order $1/\sqrt{N_f}$, which means that in a 1000-particle system, the fluctuations are about 5-10%. An accurate estimate of the temperature can be obtained by averaging over many fluctuations.

2. *Computing the forces.*
   The most time consuming part in almost all MD simulations is the force calculation. In the most general case, we have to consider the contribution to the force on particle i due to all other particles in the system. Even in the case of simple additive, pairwise interactions with each particle interacting with only the nearest image of another particle, the force calculation involves evaluating $N \times (N-1)/2$ pair distances. Consequently, the time needed for the calculation scales as N^2, which, however, can be speeded up by using efficient techniques to get order N scaling (see Ref. [3] for details). In many cases, a cutoff distance r_c is used to limit the range of the interactions and the computational cost can be significantly reduced by using neighbor lists (see Section 2.1.5). The actual force calculation is performed for each particle i in the system. For example, in the case of simple pairwise interactions, the x component of the force on particle i due to a neighboring particle j is obtained by taking the partial derivative of the potential function with respect to x:

$$f_x(r) = -\frac{\partial \mathcal{U}(r)}{\partial x} \tag{2.17}$$

3. *Integrating the equations of motion.*
   Having computed all the forces, we can now proceed to integrating the Newton's equations of motion. The simplest numerical solution is obtained by making Taylor expansions of the positions and velocities about the current time t, i.e. [2]

$$\begin{cases} \mathbf{r}_i(t+\Delta t) & = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 + \cdots \\ \mathbf{v}_i(t+\Delta t) & = \mathbf{v}_i(t) + \mathbf{a}_i(t)\Delta t + \cdots . \end{cases} \tag{2.18}$$

   These equations are truncated after a small number of terms in order to make the calculation sufficiently simple, but consequently errors tend to build up quickly as more time steps are taken. A number of different

predictor-corrector methods have been developed to minimize the truncation errors, but often a better solution is to use the so-called *Verlet* algorithm. In this approach, the estimate of the new position is given by [3]

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \mathbf{a}_i(t)\Delta t^2, \tag{2.19}$$

where the error is of order Δt^4. The velocities can then be determined by taking numerical time derivatives of the position coordinates

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t}. \tag{2.20}$$

In this expression, the error is of order Δt^2. It is, however, possible to write the Verlet algorithm in a form that uses positions and velocities computed at equal times. This *velocity Verlet* algorithm consists of the following updates

$$\begin{cases} \mathbf{r}_i(t+\Delta t) & = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 \\ \mathbf{v}_i(t+\Delta t) & = \mathbf{v}_i(t) + \frac{1}{2}[\mathbf{a}_i(t+\Delta t) + \mathbf{a}_i(t)]\Delta t. \end{cases} \tag{2.21}$$

Although it is not immediately obvious, this scheme is equivalent to the original Verlet algorithm (see Ref. [3] for details).

4. *Measurement of average quantities*.
In an MD simulation, the state of the system is defined by the positions and momenta of the N constituent particles. The average value of a quantity A is calculated as a time average

$$\langle A_t \rangle = \frac{1}{t}\int_0^t A[\mathbf{x}(\tau)]d\tau, \tag{2.22}$$

where $\mathbf{x}(\tau)$ is the state vector of the system at time τ. In order to perform a measurement, we must be able to express the observable as a function of the positions and momenta of the particles. For example, we already saw how a convenient expression can be derived for the instantaneous temperature based on the theorem of equipartition of energy.

An appealing feature of the MD method is that it follows the actual dynamical evolution of the system and no predictions need to be made about the possible trajectories. Comparing to kinetic Monte Carlo, the MD method does not suffer from any of the problems related to forming a catalog of possible events or determining the correct rates for those events. However, the need to integrate the equations of motion dictates that the maximum time step is in the femtosecond range at best, so that problems which involve a much larger time scale cannot be addressed. Recently, several new "accelerated dynamics methods" have appeared which aim to circumvent the time scale problem. The common feature of this class of methods

is that they all try to speed up the natural dynamics of the system while retaining a sequence of states which is representative of the original dynamics. Further discussion is beyond the scope of this thesis, but an excellent review of this subject is given by Voter *et al.* [17] who have been actively involved in developing these methods; including e.g. hyperdynamics [19, 20], parallel replica dynamics [21] and temperature accelerated dynamics [22].

2.4 Empirical Si-Ge potentials

2.4.1 Introduction

In recent years, the role of computer simulations has become increasingly important in materials science. While it is desirable to use accurate first-principles quantum-mechanical methods, one often faces the reality that currently they are limited to static calculations involving only a few hundred atoms at best. In order to reach conditions comparable to experiments, it is necessary to use much larger systems composed of at least tens of thousands of atoms, and furthermore, the simulation method must include temperature dependence. For these reasons, classical empirical potentials are in many cases the only practical choice for conducting large-scale simulations. Although the classical potentials cannot capture all the details of the first-principles models, they can be well suitable for studying properties which involve larger length scales than those typical for quantum-mechanical effects. In addition, when combined with advanced Monte Carlo or Molecular Dynamics techniques, this approach can be used to study systems at nonzero temperatures where entropic effects may compete with the internal energy in determining what the state of lowest free energy is.

Due to its technological importance, silicon has served as the prototype material for developing empirical potential models for semiconductor systems. A number of such potentials has appeared in the literature (e.g. [23, 24, 25, 26, 27, 28, 29, 30]). Perhaps the two most widely used models are those by Stillinger and Weber (SW) [23] and Tersoff (T2 and T3) [24, 25]. The abbreviations T2 and T3 refer to two different parametrizations of the same functional form.

The similar properties of silicon and germanium have allowed many of the empirical silicon potentials to be reparametrized to model germanium and further generalized for multicomponent Si-Ge systems. For the SW potential, the original Ge parametrization is given in Ref. [31], while the model for mixed Si-Ge systems can be found in Ref. [32]. Similarly, the Si-Ge Tersoff potential was obtained by refitting the parameters for Ge and by generalizing the original form of the potential for a multicomponent system. The details and parameters can be found in Ref. [33].

In this section, we describe the functional form of the Stillinger-Weber and Tersoff models for multicomponent Si-Ge systems and give the corresponding

sets of potential parameters. In both cases, we describe the generalized form of the potential which includes the possibility of having three different types of interactions in the system (Si-Si, Si-Ge and Ge-Ge). When used to model single-component systems, the potentials reduce to the original SW and T3 models. The multicomponent Tersoff potential uses the T3 parametrization for Si atoms because this model provides a better description of the elastic properties than the earlier T2 model. The T2 parametrization can be found in Ref. [24].

2.4.2 Stillinger-Weber potential

The Stillinger-Weber potential models bonding with classical two- and three-body interactions. The structural energy of the system is given by the potential energy function which is written as a sum of two- and three-body terms:

$$E = \sum_{\langle i,j \rangle} V_2(r_{ij}) + \sum_{\langle i,j,k \rangle} V_3(r_{ij}, r_{ik}), \tag{2.23}$$

where the first sum is taken over all nearest-neighbor bonds $\langle i, j \rangle$ and the second sum over all triplets $\langle i, j, k \rangle$ with the vertex at site i (j and k are nearest neighbors of i).

In order to describe the different types of atoms in the system, we introduce a discrete variable S_i which describes the type of the atom such that $S_i = 0$ for a Si atom and $S_i = 1$ for a Ge atom. In addition, each atom is described by the three Cartesian coordinates $\mathbf{r}_i$. Next the functional form of the two- and three-body terms is described.

The two-body potential is given by

$$V_2(r_{ij}) = \epsilon(S_i, S_j) F_2\left(\frac{r_{ij}}{\sigma(S_i, S_j)}\right), \tag{2.24}$$

where ϵ and σ are the energy and length parameters which depend on the type of bond in question (Si-Si, Si-Ge or Ge-Ge), and r_{ij} is the distance between atoms i and j. The spatial dependence of the two-body interaction is introduced through the function F_2 as follows:

$$F_2(y) = \begin{cases} A(\frac{B}{y^p} - \frac{1}{y^q}) e^{\delta/(y-b)}, & \text{if } y < b, \\ 0, & \text{otherwise.} \end{cases} \tag{2.25}$$

Notice that $F_2(y)$ is a function of the rescaled bond length, $y = r_{ij}/\sigma(S_i, S_j)$, only, and therefore it does not depend on the type of bond in question. For this reason, the parameters of the function F_2 are identical for the three types of bonds. Another interesting feature of F_2 is that it vanishes without any discontinuity at $y = b$. This function exhibits a minimum equal to -1 at $y = 2^{1/6}$; i.e. the ideal bond length is given by $R_0(S_i, S_j) = 2^{1/6}\sigma(S_i, S_j)$.

The three-body term of the potential for the contribution from a triplet $\langle ijk \rangle$ with atom i at the vertex site is given by

$$V_3(r_{ij}, r_{ik}) = [\epsilon(S_i, S_j)\epsilon(S_i, S_k)]^{1/2}\, \mathcal{L}_{ijk} \times F_3\left(\frac{r_{ij}}{\sigma(S_i, S_j)}, \frac{r_{ik}}{\sigma(S_i, S_k)}\right)\left(\cos\theta_{ijk} + \frac{1}{3}\right)^2. \tag{2.26}$$

Notice that the total contribution from three neighboring atoms i, j and k consists of three such terms where each of the atoms is taken to be at the vertex site at a time. The cosine of the angle between r_{ij} and r_{ik} is obtained from the relation

$$\cos\theta_{ijk} = \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{ik}}{r_{ij} r_{ik}}. \tag{2.27}$$

The three-body function F_3 is given by

$$F_3(y_1, y_2) = \begin{cases} e^{\gamma/(y_1-b)+\gamma/(y_2-b)}, & \text{if } y_1, y_2 < b, \\ 0, & \text{otherwise,} \end{cases} \tag{2.28}$$

where y_1 and y_2 are the rescaled lengths of the bonds between the central atom i and its two nearest neighbors j and k. The function $\mathcal{L}_{ijk}$ in Eq. 2.26 is given by

$$\mathcal{L}_{ijk} = [\lambda(S_i)^2\lambda(S_j)\lambda(S_k)]. \tag{2.29}$$

Here the geometric mean of the interaction parameters λ for the pure substances has been used to obtain the corresponding value for the composite system.

The energy parameters ϵ correspond to the covalent binding energies. Originally the parameters of the SW model for pure Si were fitted to the Si lattice constant and the cohesive energy of the diamond structure with the additional requirement that the melting point and the structure of liquid silicon be described. The original parameters give the value -4.34 eV for the bulk cohesive energy of diamond silicon. In order to reproduce the experimental value -4.63 eV of the Si cohesive energy, the original energy parameters can be multiplied by a scale factor of 1.0668 [34]. Thus the energy parameters in the Si-Ge model are $\epsilon(0,0) = 2.315$ eV for Si and $\epsilon(1,1) = 1.93$ eV for Ge (from Ref. [31]). The Ge cohesive energy, -3.85 eV, is reproduced correctly by $\epsilon(1,1)$. The binding energy between Si and Ge was calculated by Laradji *et al.* in Ref. [32] using an unmixing enthalpy $\Delta H = 7.3$ meV. From mean-field considerations, one obtains that $\epsilon(0,1) = \frac{1}{2}[\epsilon(0,0) + \epsilon(1,1)] - \Delta H = 2.1152$ eV.

Slightly varying values have been used in the literature for the parameter R_0 which determines the ideal bond length in the diamond lattice. The reason is that once the value of R_0 is fixed, the SW model predicts linear thermal expansion for diamond silicon in the whole temperature range from 0 K to 1500 K

(see Ref. [32] for details). At temperatures above 250 K, the thermal expansion coefficient predicted by the SW model is in close agreement with experimental measurements [35], but at low temperatures ($T < 120$ K), a *negative* thermal expansion is observed experimentally. As a consequence, if the value of R_0 is fixed to give the experimental value at 0 K, then the average value of the lattice constant given by the SW model will deviate from the experimentally observed values at higher temperatures. Another option is to choose R_0 in such a way that the results from the SW model agree closely with experiments for temperatures higher than 250 K. Agreement at 0 K is obtained using $R_0(0,0) = 2.351670$ Å [23, 34], while choosing $R_0(0,0) = 2.34779$ Å [32] gives agreement at higher temperatures.

For Ge, the experimentally observed thermal expansion is linear, but the expansion coefficient predicted by the SW model deviates slightly from the experimentally measured value [32]. In the original Ge parametrization it was chosen that $R_0(1,1) = 2.448090$ Å [31]. This gives agreement with experiments only at 0 K. Better overall agreement is obtained by choosing $R_0(1,1) = 2.44589$ Å [32] (full agreement at 550 K).

The parameter $R(0,1)$ for the mixed system is obtained by assuming that Vegard's law is valid. Under this assumption, a mixed system with 50% composition should have a lattice constant which is the arithmetic mean of that of pure Si and pure Ge. This gives $R(0,1) = \frac{1}{2}[R(0,0) + R(1,1)]$. If we choose $R_0(0,0)$ and $R_0(1,1)$ to match the experimental data at higher temperatures, we obtain $R(0,1) = 2.396885$ Å.

The remaining parameters for the Si-Ge SW model are as follows:

$$A = 7.049\,556\,277,\ B = 0.602\,224\,558\,4,\ p = 4,\ q = 0,$$
$$\delta = 1,\ b = 1.80,\ \gamma = 1.20,\ \lambda(0) = 21.0,\ \lambda(1) = 31.0.$$

2.4.3 Tersoff potential

The original Tersoff potential for elemental Si has two parametrizations, T2 [24] and T3 [25]. The latter one is used to model Si in the Tersoff potential for the mixed Si-Ge system [33]. In the following, the functional form of the Si-Ge potential is presented and the corresponding parameter values are given in Table 2.1. The same potential can be used to model Si-C systems (for C parameters, see Ref. [33]).

Potentials of the Tersoff type are called cluster functionals [34]. Bonding is modeled with pairwise interactions where the attractive term f_A depends on the local environment of the atoms effectively including many-body interactions. The structural energy has the form

$$E = \frac{1}{2}{\sum_{i,j}}' f_c(r_{ij})[\, f_R(r_{ij}) + b_{ij}\, f_A(r_{ij})], \tag{2.30}$$

where the prime indicates that the summation is taken over distinct indices. The functions f_R and f_A are the repulsive and attractive pair-potentials, respectively, and f_c is a smooth cutoff function. The prefactor b_{ij} is a monotonically decreasing function of the effective coordination of atoms i and j, which represents a measure of *bond order* (i.e. the degree of bonding between two atoms relative to that of a single bond).

In this case, we use the notation that singly subscribed parameters, such as λ_i, depend only on the type of atom in question, whereas parameters with double subscripts, such as A_{ij}, depend on the type of bond in question (Si-Si, Si-Ge, Ge-Ge). Most of the parameters describing interactions between Si and Ge are obtained as mean values from the elemental Si and Ge parameters. One additional parameter, χ_{ij}, is introduced to the multicomponent potential to give correctly the cohesive energy -4.231 eV of the SiGe zinc-blende-structure compound.

The repulsive and attractive pairwise terms are given by

$$f_R(r_{ij}) = A_{ij} e^{-\lambda_{ij} r_{ij}} \tag{2.31}$$

$$f_A(r_{ij}) = -B_{ij} e^{-\mu_{ij} r_{ij}}, \tag{2.32}$$

and the cutoff function is given by

$$f_c(r_{ij}) = \begin{cases} 1, & \text{if } r_{ij} < R_{ij}, \\ \frac{1}{2} + \frac{1}{2}\cos[\pi(r_{ij} - R_{ij})/(S_{ij} - R_{ij})], & \text{if } R_{ij} < r_{ij} < S_{ij}, \\ 0, & \text{if } R_{ij} > S_{ij}. \end{cases} \tag{2.33}$$

The prefactor b_{ij} has the form

$$b_{ij} = \chi_{ij}(1 + \beta_i^{n_i} \xi_{ij}^{n_i})^{-1/(2n_i)}. \tag{2.34}$$

From this we see that the parameter χ_{ij} modifies the strength of the heteropolar bonds relative to the value obtained by simple interpolation. For bonds between two atoms of the same kind (Si-Si or Ge-Ge), $\chi_{ii} = 1$. One common value is used for mixed bonds (Si-Ge or Ge-Si), $\chi_{ij} = \chi_{ji}$. It is noted, however, that $b_{ij} \neq b_{ji}$. The effective coordination number ξ_{ij} is given by

$$\xi_{ij} = \sum_{k \neq i,j} f_c(r_{ik}) g(\theta_{ijk}) \exp[\gamma_{ij}^3 (r_{ij} - r_{ik})^3], \tag{2.35}$$

where the value of γ_{ij} can either be set to zero to simplify the potential (as in Ref. [33]), or to obtain the original T3 model, it can be set equal to μ_{ij} (as in Ref. [25]). Finally, the angle-dependent function g is given by

$$g(\theta_{ijk}) = 1 + c_i^2/d_i^2 - c_i^2/[d_i^2 + (h_i - \cos\theta_{ijk})^2], \tag{2.36}$$

where θ_{ijk} is the angle between $\mathbf{r}_{ij}$ and $\mathbf{r}_{ik}$.

The values of the compound parameters are obtained by taking simple arithmetic or geometric mean values of the corresponding single component parameters:

$$\lambda_{ij} = (\lambda_i + \lambda_j)/2,\ \mu_{ij} = (\mu_i + \mu_j)/2,$$
$$A_{ij} = (A_i A_j)^{1/2},\ B_{ij} = (B_i B_j)^{1/2},\ R_{ij} = (R_i R_j)^{1/2},\ S_{ij} = (S_i S_j)^{1/2}. \quad (2.37)$$

The parameters for the T3 pure Si model were originally fitted to a database consisting of cohesive energies of real and hypothetical structures of bulk silicon, along with the bulk modulus and bond length of the diamond structure. In addition, it was required that the three elastic constants of silicon are reproduced within 20%. For the multicomponent potential, the parameters for Ge were first determined independently using the same functional form as for Si. Then, by adding a *single* additional parameter (χ_{ij}), the potential was generalized to treat mixed Si-Ge systems. The parameter values for the Si-Ge Tersoff potential are summarized in Table 2.1.

Table 2.1: Parameters for the Si-Ge Tersoff potential, from [33] and [24].

	Si	Ge
A (eV)	1.8308×10^3	1.769×10^3
B (eV)	4.7118×10^2	4.1923×10^2
λ (Å^{-1})	2.4799	2.4451
μ (Å^{-1})	1.7322	1.7047
γ (Å^{-1})	0 or 1.7322	0 or 1.7047
β	1.1000×10^{-6}	9.0166×10^{-7}
n	7.8734×10^{-1}	7.5627×10^{-1}
c	1.0039×10^5	1.0643×10^5
d	1.6218×10^1	1.5652×10^1
h	-5.9826×10^{-1}	-4.3884×10^{-1}
R (Å)	2.7	2.8
S (Å)	3.0	3.1
χ_{Si-Ge}	1.00061	1.00061

Chapter 3

Generic growth models

3.1 Full-diffusion model

In this thesis, the kinetic Monte Carlo method is applied to the study of epitaxial crystal growth using a so-called full-diffusion model (FD) with a solid-on-solid (SOS) approximation. Epitaxial growth is a non-equilibrium process in which new material crystallizes on a substrate in a manner dictated by the lattice structure of the underlying solid. There are various fabrication techniques to achieve such growth, but in this study we concentrate on the case of Molecular Beam Epitaxy (MBE).

In the following, we introduce a generic KMC model which has been successfully applied to model the central features of MBE growth. It is outside the scope of this approach to aim at describing any particular material in great detail as this would require much more sophisticated models, including an off-lattice description of the system. Nevertheless, we have been able to demonstrate that the basic model can be useful in investigating general properties of epitaxial systems. We applied the method to study the initial stages of growth on non-uniform surfaces, and our results show that nanoscale patterning of the substrate can improve the spatial ordering and the size uniformity of small two-dimensional islands.

The KMC method is based on the solid-on-solid (SOS) [36] model of epitaxial growth which assumes a simple cubic lattice structure with neither vacancies nor overhangs. The basic processes included in the model are *deposition of adatoms* and subsequent *surface diffusion.* The process of desorption has been omitted from the model since it is negligible under usual MBE conditions. Thus, the fractional surface coverage is given by $\Theta = Ft$ where F is the constant deposition rate of atoms in monolayers per second (ML/s) and t is the physical time. The deposition of adatoms takes place onto an initially flat substrate. In the simulations, a deposition site is first selected at random, and then a search is carried out within a square of fixed linear size $2R_i + 1$, centered upon the selected site. The site with

a maximum number of lateral nearest neighbors is chosen as the deposition site.

The diffusion rate of a single adatom is defined as the probability of a diffusion jump per unit time. Based on Transition State Theory (TST) [37], it is given by the following Arrhenius-type expression:

$$k(E,T) = k_0 \exp(-E/k_{\mathrm{B}}T), \tag{3.1}$$

where E is the activation energy, T is the substrate temperature, k_{B} is the Boltzmann constant and the prefactor k_0 corresponds to the frequency of atomic vibrations. Assuming that each adatom has the same frequency and can be treated as a two-dimensional harmonic oscillator, we can use the equipartition theorem to derive a reasonable approximation for k_0 at high temperatures (corresponding to typical growth conditions) [38]: $k_0 = 2k_{\mathrm{B}}T/h$, where h is the Planck constant.

In the basic model, the activation energy E comprises a substrate term, E_{S}, and a contribution from each occupied lateral nearest neighbor atom, E_{N}:

$$E = E_{\mathrm{S}} + nE_{\mathrm{N}}, \tag{3.2}$$

where $n = 0-4$ is the number of occupied lateral nearest neighbors at the *initial* site. In this basic model of diffusion, the adatom lands with equal probability at any of the four neighboring sites.

3.2 Patterned substrate models

The patterned substrate is incorporated into the basic diffusion model by dividing the lattice into square-shaped domains of size ($l \times l$). The square geometry is adopted for simplicity and to speed up the computation. The energy barriers for diffusion are varied within the domain structure in two ways (see Fig. 3.1):

1. (Model A) The parameter E_{S} is let to vary piecewise linearly as a function of the lateral position of the adatom on the surface. The total barrier for diffusion is given by $E_{\mathrm{A}} = E_{\mathrm{S}}(x,y) + nE_{\mathrm{N}}$.

2. (Model B) An additional diffusion barrier E_{D} is introduced for the diffusion jumps directed towards the domain boundaries. The strength of the additional barrier is determined by the distance of the adatom from the boundary. The total barrier for diffusion is given by $E_{\mathrm{B}} = E_{\mathrm{S}} + nE_{\mathrm{N}} + E_{\mathrm{D}}$ where E_{D} depends on both the hop direction and the lateral position of the adatom on the surface.

The experimental motivation behind Model A comes from vertically correlated quantum dot superlattices. In these systems, the surface diffusion of adatoms and subsequent island nucleation are affected by the spatial bias that arises from

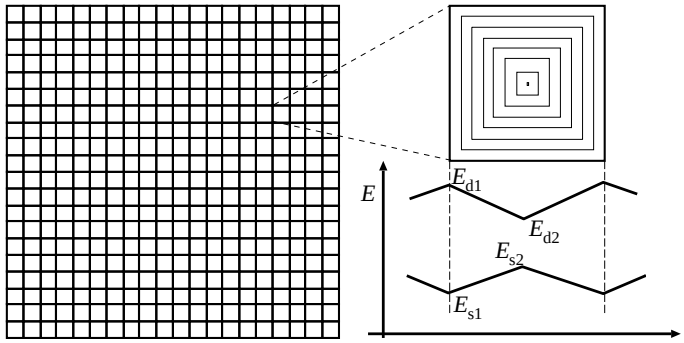

Figure 3.1: Domain structure in the patterned substrate simulation model. The right part shows the variation of the diffusion activation energies inside a single domain and the corresponding variations of the model parameters E_S (model A) and E_D (model B) when traversing through the cross-section of the domain.

the strain at the surface due to buried islands [39]. The idea in our work is to examine the origin of possible mechanisms that lead to spatially ordered nucleation through the variation of diffusion activation energies. Model A describes a situation in which the energy surface seen by the diffusing adatom is spatially biased by some underlying structure of the substrate (e.g. buried islands).

Motivation for Model B, in turn, is based on the experimental system of Ag / 2 ML Ag / Pt(111) heteroepitaxy. In this case, the substrate is patterned with a regular network of dislocations which act repulsively towards the diffusing adatoms [40]. In model B, we have included an additional, hop-direction dependent diffusion barrier, E_D, to describe the long-range repulsive adatom-dislocation interaction.

For model A, the variation of the substrate binding energy E_S inside each domain was chosen by performing some preliminary simulations. The values $E_{S1} = 0.65$ eV and $E_{S2} = 0.85$ eV (see Fig. 3.1) produced an effect which was clear but not artificially strong. In the case of model B, the magnitude of the additional diffusion barrier, E_D, does not need to be very large since even a small increase in the diffusion barrier leads to substantial changes in diffusivity. This is because adatoms have to make several jumps in an unfavorable direction to cross a dislocation. Based on this information, the values $E_{D1} = 0.02$ eV, $E_{D2} = 0$ eV were used for the maximum and minimum values of the additional barrier, respectively (see Fig. 3.1).

The deposition rate F was 0.0033 ML/s and the value $R_i = 1$ was used for the incorporation radius. The temperature range of interest was determined by the behavior of the system. We found that in the simulations for the chosen parameter set the effect of patterned substrate on nucleation was most pronounced for the temperature range $T = 360 - 420$K.

3.3 Results for nucleation on patterned surfaces

This section summarizes the main results obtained in publications I, II and III. We begin by first discussing the general features of patterned substrates which are common to both models A and B.

Figure 3.2 shows examples of island morphologies at three different temperatures for islands grown on a homogeneous and a patterned substrate. We see that the patterned substrate induces ordering into a regular arrangement of islands which displays the periodicity of the underlying pattern. Figures 3.2(d)-(f) all show that the nucleation probability is highest in the domain centers due to a net flow of adatoms into this area. However, the morphology of the islands is found to be strongly dependent on the temperature. At intermediate temperatures, the average diffusion length is of the same order as the lateral length of the substrate domains which means that the diffusing adatoms are able to diffuse across the whole domain before nucleating a new island or joining a pre-existing one. This leads to a situation where within a certain temperature range exactly one island nucleates within each unit cell of the substrate structure.

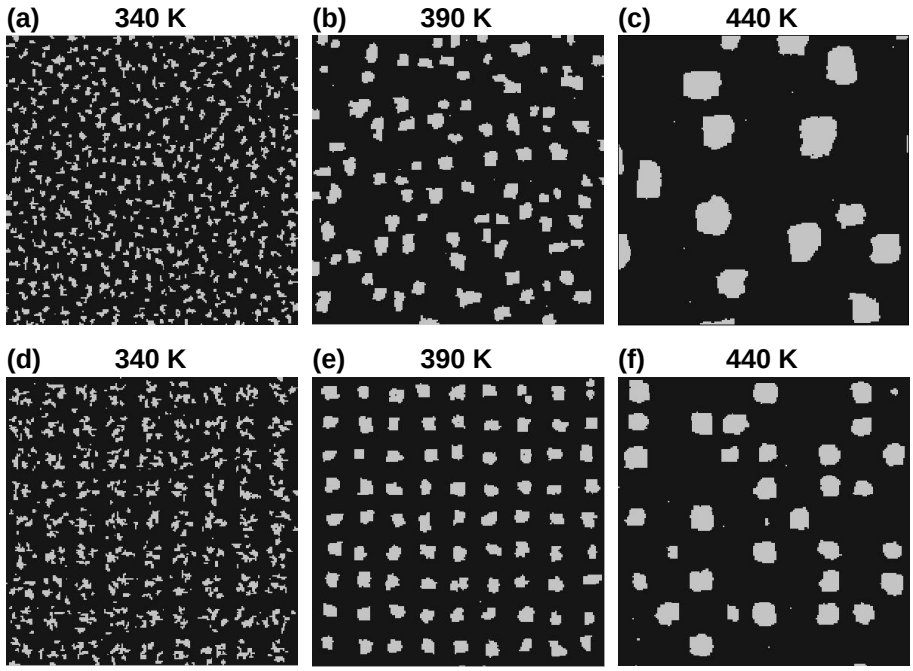

Figure 3.2: Island morphologies obtained at three different temperatures (from left, 340, 390 and 440 K) for a homogeneous substrate (a)-(c) and for a patterned substrate (model B) (d)-(f). Dark areas designate the substrate and light areas the first layer of adatoms. The coverage is 15%. 200×200 surface sections are shown.

In order to investigate these observations in more detail, let us consider the distribution of island sizes. Figure 3.3(a) shows the size distributions at $T = 390K$ for random nucleation on a homogeneous substrate and for ordered nucleation on a patterned substrate. The distributions are normalized according to scaling theory [41]. We notice that the spatial ordering leads to an enhanced size uniformity of the islands. The domains act as equally large capture areas for the islands since all material deposited into a certain domain stays confined within this domain due to the biased diffusion. This holds for temperatures at which the average diffusion length corresponds to the length scale of the substrate pattern. The simulation results are also compared to experimental data (shown as symbols) for Ag nucleation on a substrate consisting of 2 ML of Ag on Pt(111) (surface patterned with dislocation networks) and for Ag nucleation on clean Pt(111) (a homogeneous surface). We notice that KMC simulations agree well with the experimental measurements, which confirms that our model seems to be able to capture the essential features of ordered nucleation.

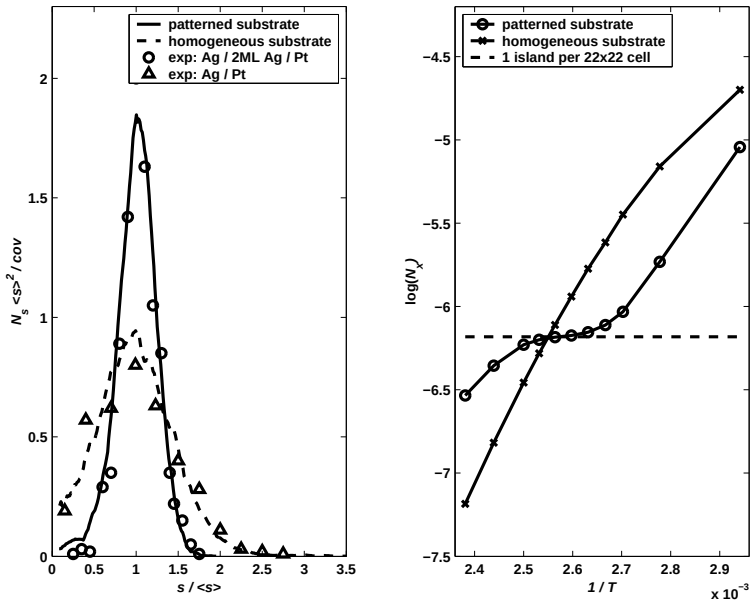

Figure 3.3: (a) Scaled island size distributions for random and ordered nucleation. The distributions are normalized according to scaling theory. s denotes the island size, $\langle s \rangle$ its average value, and N_s the areal density of islands composed of s atoms. The model parameters correspond to $D/F \approx 10^6$. The coverage is 15 %. The symbols, circles and triangles, are experimental data for the nucleation of Ag islands on 2ML of Ag on Pt(111) (substrate patterned with dislocation networks) and on clean Pt(111) (homogeneous substrate), respectively. (Experimental data from Ref. [40].) (b) Arrhenius plots of the total island density N_x for nucleation on a homogeneous and a patterned surface. All the data are for model A.

The temperature dependencies of the total island density, N_x, are shown in Fig. 3.3(b). In the patterned substrate case, the island density remains approximately constant for temperatures in the range 380-400K. This constant density corresponds to the case where each domain contains a single island. At lower temperatures, the island density increases, which means that several small islands nucleate within each domain due to the shorter diffusion length. At high temperatures, diffusion across the domain boundaries becomes activated. Some of the domains are left empty as the adatoms join islands in surrounding domains, thus decreasing the island density.

So far we have not considered the differences between models A and B. These become apparent when the total diffusion rate of an adatom in different positions on the surface is considered. In the case of model A, the adatoms are more tightly bound to the substrate the closer they are to the domain centers. Thus diffusion is fast near the boundaries and slow near the centers of the domains. The energetically most favorable configuration is a single island located at the center of each domain. In the case of model B, the hop-direction dependence of the diffusion barriers produces a net movement of adatoms toward the domain centers, but the change in the total diffusion rate is very small (maximally, 0.02 eV, in contrast to 0.2 eV for model A). Under favorable growth conditions, the islands first order into a regular array, but post-deposition relaxation of the structure leads to dissociation of small islands in favor of larger ones. This Ostwald ripening process does not occur in model A due to its different energetic structure. This behavior is illustrated in Fig. 3.4 which shows a series of three snapshots for both models of the post-deposition evolution of the surface morphology.

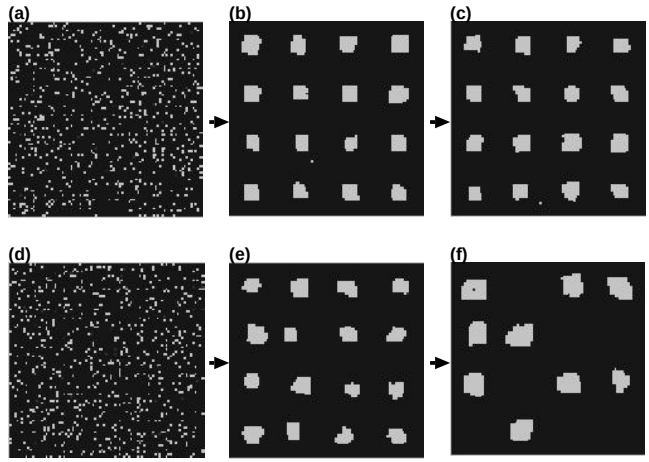

Figure 3.4: Post-deposition development of surface configurations for model A (a)-(c) and for model B (d)-(f). The initial configurations (a) and (d) were obtained by a random deposition of 10% coverage of adatoms. The temperature is 400 K.

Chapter 4

The Si(001) surface

4.1 2×1 reconstruction

In this chapter, we will review the basic properties of the Si(001) surface. The surface morphology is dominated by the 2×1 reconstruction which leads to strong direction-dependent effects such as the anisotropic distribution of strain and the existence of different types of step edges. The main defect structures on the Si(001) surface are also related to dimers which are the basic building blocks of the 2×1 reconstruction. We will discuss these various issues by comparing experimentally observed features to computational results and concentrating especially on the atomic-scale structure and energetics of the surface. We begin by describing the 2×1 reconstruction on a clean Si(001) surface and continue the discussion to stepped surfaces and defect structures.

The arrangement of atoms at surfaces often deviates significantly from the atomic structure of bulk material [42]. Especially semiconductor materials display a wide variety of atomic reconstructions at surfaces. The surface atoms rearrange in order to reduce the number of unsaturated bonds. This reduction of the dangling bonds results in an energy gain which often outweighs the energy cost of increasing the surface stress due to the atomic displacements.

When a silicon crystal is cut along the (001) plane, the surface atoms are left with two dangling bonds each. The Si(001) surface reconstructs by forming a structure with 2×1 symmetry consisting of rows of dimerized atoms. The reconstruction removes half of the dangling bonds, but also increases the strain energy due to bond distortion. According to *ab initio* calculations [43, 44], the total energy gain is approximately 1 eV per surface atom with respect to 1×1 ordering. The 2×1 reconstruction persists on Si(001) up to temperatures of at least 1475 K [45]. Figure 4.1 shows the structure of a bulk-terminated and a reconstructed Si(001) surface.

The bulk lattice constant of silicon is 5.43 Å (at 300 K) [35]. The corre-

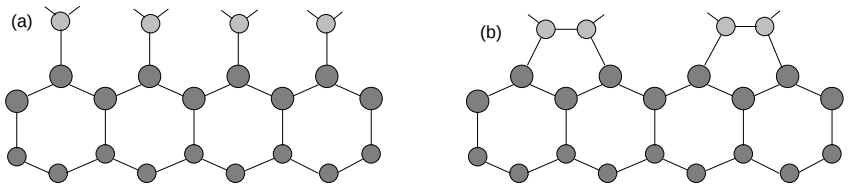

Figure 4.1: The 2×1 reconstruction of the Si(001) surface. (a) Unreconstructed (bulk-terminated) surface with two dangling bonds per atom. (b) Reconstructed (dimerized) surface with one dangling bond per atom.

sponding nearest-neighbor separation in the diamond lattice is 2.35 Å. On unreconstructed Si(001), the surface layer forms a square lattice with a separation of 3.84 Å between neighboring top-layer atoms. This symmetry is broken on the reconstructed surface as the distance between two adjacent atoms is reduced and a dimer is formed. The experimental value of the dimer bond length is very close to the nearest-neighbor separation in bulk (2.35 Å). The dimers are arranged in rows which are oriented perpendicular to the direction of the dimer bonds (see Fig. 4.2). The distance between two neighboring dimer rows is 7.84 Å.

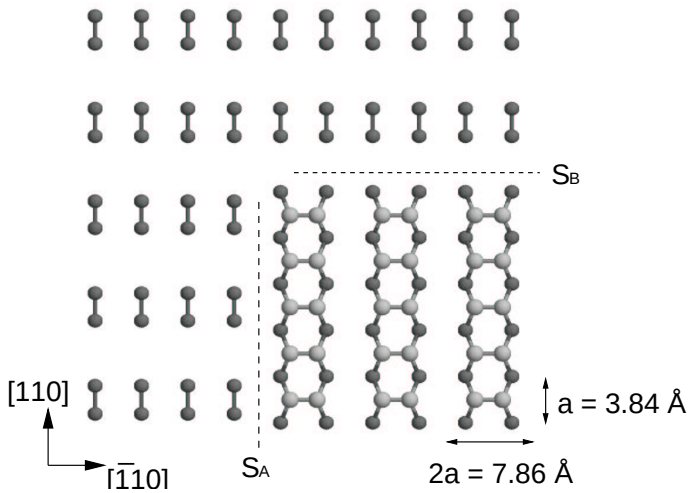

Figure 4.2: Schematic model of a stepped 2×1 Si(001) surface. Two different types of single-layer-height steps are shown: S_A (edge parallel to upper-terrace dimer rows) and S_B (edge perpendicular to upper-terrace dimer rows). In addition, the S_B step can have two different edges: nonrebonded (last upper-terrace dimers are not bonded to dimerized atoms on the lower terrace, shown in the figure) or rebonded (last dimers are bonded to dimerized atoms on the lower terrace). Due to the symmetry of the diamond lattice, the direction of the dimer bonds rotates by 90° on terraces separated by monatomic steps.

The 2×1 reconstruction establishes two characteristic directions on the surface: namely parallel and perpendicular to the dimer rows. The stress field introduced by the reconstruction is highly anisotropic such that the stress along the dimer bond is tensile, while in the perpendicular direction the surface is under compression [46]. Due to the tetrahedral bonding configuration of the diamond lattice, the direction of the dimer bonds rotates by 90° on terraces separated by monatomic steps. There are two different types of monatomic steps: the S_A step edge is parallel and the S_B step edge is perpendicular to the upper-terrace dimer rows. In addition, the S_B step has two alternative forms: rebonded (last atoms on the upper terrace are bonded to lower-terrace atoms which in turn are rebonded to other lower-terrace atoms) or nonrebonded (the lower terrace atoms are not rebonded). Figure 4.2 shows a schematic illustration of the dimer structure and different step types.

The different atomic structure of the S_A and S_B step edges has consequences for the step morphology on a larger length scale. The energy required to form kinks is much higher for an S_A than for an S_B step. In typical scanning tunneling microscopy (STM) images (see e.g. [47] and [45]), the S_A steps appear smooth and straight, while the S_B steps have a rough shape. In order to analyze these observations in more detail, a number of energy parameters need to be determined for a thermodynamic description of the different contributions to the free energy. These parameters are the step-edge formation energy, the kink-creation energy, and the energetic and entropic interactions between steps. The values of these parameters have been determined using both experimental [47, 48, 49] and theoretical [50, 51, 52, 53, 54] methods.

Isolated step edges on Si(001) can be characterized by three energy parameters: step energies ϵ_A and ϵ_B per atom corresponding to the S_A and S_B type steps, and by an effective corner energy E_C which is associated with the corners of a kink. In the following, we briefly review how the values of these parameters have been extracted both experimentally and computationally.

In a pioneering paper, Swartzentruber *et al.* [47] demonstrated that STM images with an equilibrium distribution of steps and kinks can be analyzed to obtain the distribution of kink separations and kink lengths. These distributions can consequently be used to determine the step and kink energies. In the simplest approach, it is assumed that the kink excitations are statistically independent. Under this assumption, the distribution of kink lengths, $N(n)$, is proportional to the Boltzmann factor; i.e. $N(n) \propto \exp[-E(n)/k_B T]$. Swartzentruber *et al.* measured the number of kinks of length n from a large set of STM images. The energy of a kink of length n can be obtained by plotting the data using the following relation

$$E(n) = -k_B T \ln\left(\frac{N(n)}{2N(0)}\right), \tag{4.1}$$

where $N(0)$ is the number of sites at which there is no kink and the factor 2

arises from the degeneracy of the excited states (i.e. the kinks can occur either into or out of the terrace). The STM data can be fitted by a functional form $E_{B,A}(n) = n\epsilon_{A,B} + E_C$ where $\epsilon_{A,B}$ is the energy of the S_A or S_B step without kinks. By carefully analyzing a set of experimental data obtained from different studies (Refs. [47, 49, 55]), Bowler *et al.* [53] were able to combine the data to extract the following experimental values : $\epsilon_A = 0.023 \pm 0.001$ eV/a, $\epsilon_B = 0.065 \pm 0.005$ eV/a and $E_C = 0.078 \pm 0.022$ eV.

The experimental values can be compared with computational values obtained from first-principles [54], tight-binding [50, 53] or empirical-potential [51, 52] calculations. The unit cell sizes used in most quantum-mechanical studies are very small, which means that the calculated values of the step-edge energy have not converged with respect to cell size. Comparison with experiments shows that results obtained in these studies are inconsistent with measured values. In a tight-binding study by Bowler *et al.* [53], the question of unit cell size was carefully addressed. Using sufficiently large terrace sizes in the unit cell to ensure convergence, they obtained the following values for the energy parameters: $\epsilon_A = 0.019 \pm 0.005$ eV/a, $\epsilon_B = 0.080 \pm 0.010$ eV/a and $E_C = 0.071 \pm 0.01$ eV. These are all in good agreement with the experimental values given above.

4.2 Antiphase boundaries

The structure of the 2×1 reconstruction has a strong influence on growth kinetics and morphology of growing films on Si(001). Diffusion of adatoms is much faster in the direction of the underlying dimer rows on the terrace than in the perpendicular direction. Due to the alternating direction of the dimer bonds on the vicinal Si(001) surface, there are two types of terraces. On the A-type terraces (T_A) the dimer rows run parallel to the edge of the down step (S_A). and on the B-type terraces (T_B) the dimer rows are perpendicular to the edge of the down step (S_B).

Due to anisotropic diffusion of adatoms and the alternating direction of the dimer rows on adjacent terraces, the growth mode is step flow on the T_B terraces (atoms diffuse quickly to one of the step edges), while on the T_A terraces islands are observed to nucleate (diffusion is fast in the direction parallel to the step edges) [42]. When the nucleated islands grow, they eventually coalesce with the upper terrace (S_B step). There is a 50 % chance that the 2×1 reconstruction of the joining island is in phase with the 2×1 reconstruction of the upper terrace. In this case, the island can merge into the upper terrace defectlessly. In contrast, if the 2×1 reconstruction of the island is shifted by one lattice unit a with respect to the position of the upper terrace dimer rows, an antiphase boundary (APB) is created. This is called an APB of type B because the boundary runs perpendicular to the dimer rows and is created when two out-of-phase S_B steps are joined. It is also possible that the dimer rows in two growing islands are not in register in the

direction parallel to the dimer rows. When such two islands meet, an extra row of substrate atoms is left in between the boundaries of the two islands. This A-type APB runs parallel to the dimer rows, thus separating two S_A steps. Figure 4.3 shows a schematic illustration of both types of APBs.

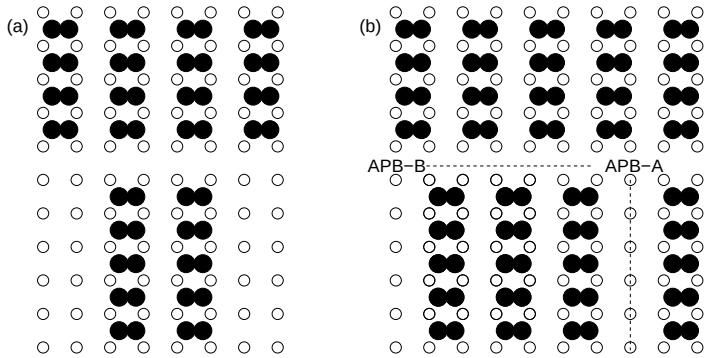

Figure 4.3: Two types of antiphase boundaries can be created upon coalescence of initially separate 2×1 reconstructed regions on the Si(001) surface. (a) Defect-free coalescence: dimer rows of the lower section are in phase with the upper section. (b) B-type APB (horizontal dashed line): dimer rows of the lower section are displaced by one lattice unit with respect to the upper section. A-type APB (vertical dashed line): parallel dimer rows are not in register which leaves an extra row of atoms in the underlying layer exposed.

STM experiments have shown that an APB-B acts as a preferential nucleation site for atoms landing on the upper terrace [56, 57, 58]. These new islands also have a 50 % chance to nucleate in such a way that another APB is formed upon coalescence. This leads to roughening of the growing thin film. The A-type APB, on the other hand, leads to the formation of long B-type double layer steps and narrow trenches [58]. Filling of these trenches is kinetically suppressed by slow diffusion in the direction perpendicular to the trenches, which also leads to roughening. Very similar APBs have also been observed to form during the growth Ge on Si(001) and on the Ge(001) surface [58, 59].

The formation of APBs serves as a good illustration of how the atomic scale reconstruction of the surface affects the morphology of the growth front on a much larger length scale. The smoothness of the growing film and abruptness of interfaces between epitaxial layers is often a crucial factor in manufacturing high-quality electronic devices. This leads to contradicting demands with respect to the growth conditions. On one hand, the growth temperature should be sufficiently high to avoid kinetic roughening, but on the other hand, the temperature should be sufficiently low to avoid undesirable intermixing and dopant segregation. It is therefore extremely important to understand the details of phenomena such as reconstruction, formation of APBs and intermixing, on an atomistic level.

In publication V, we compared the properties of the Stillinger-Weber (SW) and the Tersoff (T2 and T3) potentials and investigated how well these potentials are able to describe the energetics of antiphase boundaries. For the APB part of the study, we used a static energy-minimization method to calculate the excess surface energy of antiphase configurations with respect to a defect-free 2×1 reconstruction. The unit cell used in the calculations was $2n$ dimers long and one dimer wide with two antiphase boundaries separated by n dimers in the cell ($n = 1 - 100$). The bulk section of the cell was 25 atomic layers thick with the bottom layers fixed to diamond lattice positions.

Figure 4.4 shows the excess surface energy per APB as a function of the APB separation (expressed in number of dimers, n). For all three potentials, the energy increases as the APB separation becomes smaller, indicating that the interaction between two APBs is repulsive. Convergence to meV accuracy is reached at a separation of eight dimers. The shapes of the energy curves are very similar to the one obtained using TB calculations in Ref. [53]. The TB calculations give the value 0.162 eV for the converged value of the energy, which is of the same magnitude as the corresponding values 0.086 eV and 0.17 eV given by the SW and T2 potentials, respectively. In contrast, when the T3 potential is used, the excess surface energy is found to become clearly *negative* when the APBs are separated by at least two dimers. This indicates that the straight-row 2×1 reconstruction is a metastable state for the T3 potential, which in turn significantly limits the suitability of the T3 potential for finite temperature simulations, especially if the atoms are not restricted to maintain a preassigned surface configuration. Examples of such simulations are given later in this thesis (see Chapter 6).

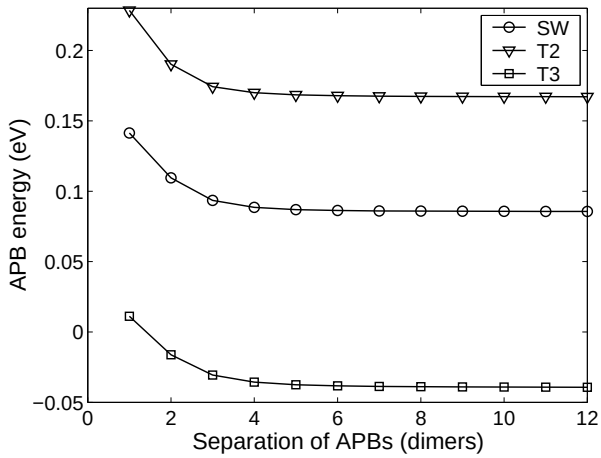

Figure 4.4: Excess surface energy per APB as a function of separation of APB's for a unit cell containing two equally sized antiphase domains. The reference point is a 2×1 reconstructed surface.

4.3 Vacancy structures

4.3.1 Background

There are various experimental techniques to study defect structures of solid systems. In the case of Si(001), STM experiments have produced consistent evidence that a high defect density is intrinsic to the surface [60, 61]. These defects have been identified to be missing dimers and their complexes, rather than missing individual atoms. An analysis of a typical STM image reveals that isolated dimer vacancies tend to congregate into small clusters and that these clusters form certain distinctive complexes which dominate over other possible configurations [62]. The structure and bonding of these and several other DV complexes were studied by Wang *et al.* [62] using *ab initio* total-energy calculations, and their results provide a good quantitative explanation for most of the features observed in the STM images. In the following, we review the main results obtained in Ref. [62] and compare them to our empirical potential calculations, pointing out the strengths and weaknesses of both approaches.

In the *ab initio* calculations of Ref. [62], the primitive 1-DV, 2-DV and 3-DV structures were modeled using a supercell with 2×5 periodicity, which means that in the direction of the dimer rows, the DVs in adjacent supercells are separated by only $5-n$ dimers ($n = 1, 2, 3$ for 1-DV, 2-DV and 3-DV structures, respectively). Moreover, the DVs occur in successive rows in the direction perpendicular to the dimer rows, i.e. the vacancies are not isolated but are arranged in a periodic structure where every 5th dimer from each row is missing. The main problem with using small supercells is that the amount of long-length-scale elastic relaxation is limited. In Ref. [62], additional calculations with the classical Keating potential [26] were performed to estimate the effect of lattice relaxation. After each full *ab initio* calculation, the resulting configuration was embedded in a large classical Keating supercell (up to 1708 atoms) and the surrounding material was allowed to relax completely. The resulting relaxation energy was then subtracted from the *ab initio* result.

The formation energy of each DV structure is obtained by comparing the total energy of the supercell containing the vacancy with the total energy of the corresponding supercell without the vacancy. The chemical potential is taken to be the energy of an atom in bulk silicon. The resulting formula for the formation energy per dimer is

$$E_f = \frac{1}{N_{DV}}(E_{tot} - E_{ref} + \Delta N \epsilon_b), \tag{4.2}$$

where E_{tot} and E_{ref} are the total energies of the system containing the vacancy and the reference system, respectively, ΔN is the difference in the number of atoms in the two systems, ϵ_b is the energy per atom in bulk silicon and N_{DV} is the number of dimer vacancies.

The problem with this approach is that there are different kinds of sites in a step edge to which excess atoms can attach [63]. The structure of a step edge should be taken into account when determining the formation energies. Keeping these questions in mind, we first discuss calculations where the formation energy is obtained by a direct comparison of the vacancy structure to a corresponding defectless surface with the bulk energy as the chemical potential. Then the discussion is extended to ways of improving the correspondence between numerical calculations and experimental conditions. For example, much larger supercells (composed of tens of thousands of atoms) can be used in calculations with empirical potentials. This allows a complete elastic relaxation of the configuration, and consequently the effects induced by the supercell size can be estimated. Some alternative ways of determining the formation energies are also discussed.

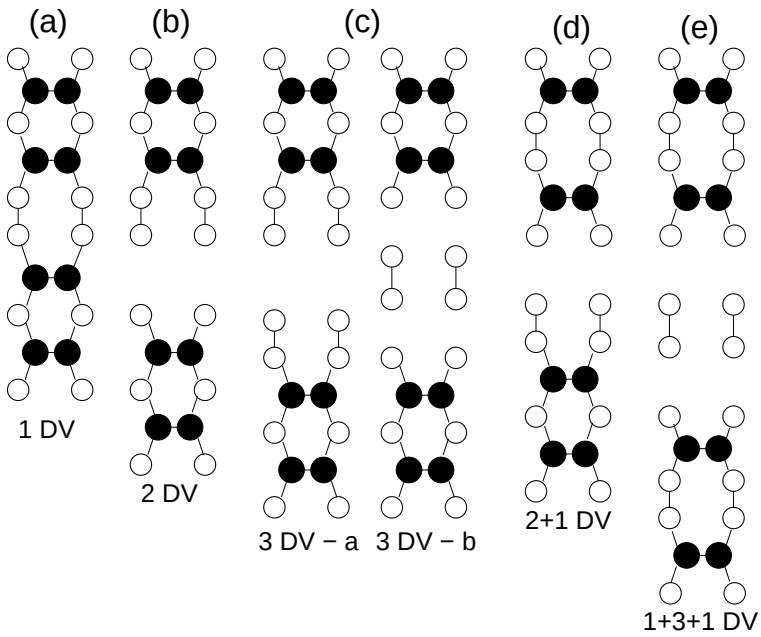

Figure 4.5: Illustration of the most common DV structures and their complexes. (a) The stable 1-DV configuration with rebonded second-layer atoms. (b) The stable 2-DV structure.The exposed second-layer atoms in the middle of the vacancy have a choice to rebond with the second-layer atoms on either side of the vacancy. The two possibilities are equivalent. (c) The stable (3 DV-a) and metastable (3 DV-b) 3-DV structures. (d) The stable 2+1-DV structure. (e) The stable 1+3+1-DV structure. Note that the rebonding of the triple vacancy in the 1+3+1-DV complex corresponds to the metastable 3-DV cluster. In all figures, the black circles represent dimerized surface atoms and the white circles atoms in the second layer.

4.3.2 1-DV structure

The basic 1-DV structure is formed by removing one dimer from a perfectly reconstructed 2×1 surface. Initially, the exposed second-layer atoms are pulled away from the vacancy, driven by the need to relax the backbonds to the surrounding dimers. This, however, is a metastable configuration. If the exposed second-layer atoms are brought closer together, bonds form between them, thus lowering the energy. This rebonded configuration is a weakly bonded but stable structure. It is depicted schematically in Fig. 4.5(a).

In Ref. [62], the formation energy of the rebonded structure was estimated to be 0.22 eV/dimer (including the Keating correction). The formation energy for the nonrebonded structure is not much higher, 0.64 eV/dimer. The small difference in energy is explained by the weakness of the bond between the exposed second-layer atoms: the bond length is 20 % longer than the bulk bond length. In an earlier *ab initio* calculation [43], a slightly larger value of 0.28 eV/dimer was obtained using a 2×4 supercell in the calculation. The difference between these two results is so small that based on these two values alone, we cannot draw any conclusions on the effect of the supercell size, and due to the large computational demands, this question has not been systematically studied using *ab initio* calculations.

Empirical potentials, on the other hand, are computationally much less intensive and have proved to be useful for investigating many of the mentioned questions related to supercell size, elastic relaxation and ways of calculating the formation energies. In publication V, we tested the SW, T2 and T3 potentials in their ability to model the dimer-vacancy structures. Of these three potentials, the Tersoff potentials were found to be unsatisfactory because T2 predicts that vacancy formation is exothermic (i.e. the formation energy per DV is negative) and T3 cannot produce spontaneous rebonding of the second layer atoms (the system remains in the metastable nonrebonded state even at high temperatures). In contrast, the SW potential was found to give results in fair agreement with corresponding *ab initio* data. The central features of dimer vacancies, the rebonding of the second-layer atoms and atomic displacements in the vicinity of the defect, are described with reasonable accuracy by SW. The formation energy for the stable, rebonded 1-DV structure was calculated to be 0.40 eV per dimer for a periodic 2×5 structure. This is somewhat higher than the corresponding *ab initio* value of 0.22 eV, but of the same order.

Our MC simulations with the SW potential were performed using a system composed of a 39-layers-thick slab with a surface area of at least 20×20 atoms, which is much larger than the supercell used in the *ab initio* calculations. The larger surface area allowed us to examine the effect of changing the periodicity of the vacancy structure; e.g. the 2×4, 2×5 and 2×10 structures can all be studied using the 20×20 surface area. For other periodicities, the surface area must be changed accordingly; e.g. the 24×24 surface can be used for the 2×6 structure.

The effect of changing the surface area does not influence the results which can be checked for example by doubling the surface area and repeating the calculation.

In publication V, we found that the formation energy of the 1-DV structure is dependent on the periodicity of the $2 \times n$ structure used for the calculation. To investigate this issue in detail, we use an approach first suggested by Oviedo *et al.* [64] for comparing the relative stability of surfaces with different stoichiometries. The main idea of this approach is that the energy change of forming a certain amount of monolayer (which may contain vacancies) cannot depend on whether this is accomplished by placing new atoms on a clean Si(001) surface or by bringing atoms to the boundary of a pre-existing piece of a monolayer. The relative stability of different $2 \times n$ structures is given by the quantity

$$\xi(n) = \frac{E_f(n) + E_p}{n - 1}, \tag{4.3}$$

where $E_f(n)$ is the energy difference per DV between the ideal 2×1 surface and the $2 \times n$ structure, and E_p is the monolayer formation energy for the ideal 2×1 surface. (See Chapter 5 for a detailed discussion.)

Figure 4.6 shows the value of ξ as a function of the periodicity of the vacancy structure. The energy of the $2 \times n$ structure decreases as the spacing between the vacancies becomes larger until the value $n = 10$ is reached. This indicates that two vacancies which are separated by less than 10 dimers are clearly interacting with each other due to the overlap of their surrounding strain fields. We can therefore, say that the 2×5 periodicity used in the *ab initio* calculations does not fully correspond to the experimental case where the concentration of vacancies is much less than 20%, but this approach is, nevertheless, useful for comparing the relative energies of the different DV clusters and their complexes.

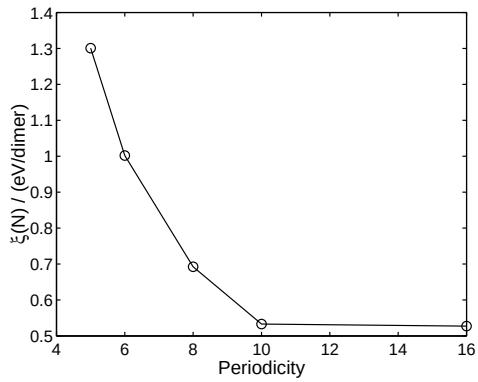

Figure 4.6: The relative energies of DV structures with different periodicities. The data are obtained using the SW potential.

4.3.3 2-DV and 3-DV structures

The 2-DV cluster is formed by removing two adjacent dimers from the 2×1 reconstructed Si(001) surface. In this case, rebonding between all three pairs of exposed second-layer atoms would lead to an unstable configuration. Instead, the pair in the middle of the vacancy rebonds with one of the other exposed second-layer atom pairs on either side. The resulting configuration is stable. This structure is illustrated in Fig. 4.5(b). The formation energy of the stable 2-DV structure is only 0.16 eV/dimer [62] which is clearly lower than the corresponding value for the stable 1-DV structure. This result is in agreement with the analysis of typical STM images where 2-DV clusters occur more frequently than 1-DVs. In addition, the strain of the 2-DV cluster is asymmetric, which significantly influences the way the 2-DV cluster interacts and binds with surrounding single vacancies.

The 3-DV cluster consists of three adjacent dimer vacancies. The stable 3-DV structure is a symmetric configuration where the exposed second-layer atoms are rebonded on both sides of the vacancy. This configuration is illustrated in the left part of Fig. 4.5(c) (labeled as 3-DV-a). The formation energy of the stable structure is 0.15 eV/dimer [62]. A metastable 3-DV structure can be formed by having rebonding between the middle pairs instead of having it on both sides (3-DV-b in Fig. 4.5(c)). This structure has more dangling bonds but it is also very stable with a formation energy of 0.29 eV/dimer [62], and it is further stabilized when combined with additional single vacancies in the 1+3+1-DV complex (see below for details). The strain field around both 3-DV configurations is symmetric.

4.3.4 DV cluster complexes

The strain fields surrounding the dimer vacancies induce interactions between them. The differences in the strain around the basic DV clusters determine how the basic DV clusters combine to form more complex structures. The 2-DV structure can link with one or two additional DVs. The resulting structures are labeled as 1+2-DV and 1+2+1-DV cluster complexes, respectively. Similarly, the 1+3-DV and 1+3+1-DV complexes consist of a triple vacancy linked with one or two single vacancies. Only two of these cluster complexes, the 1+2-DV and the 1+3+1-DV, occur in large numbers in STM images of Si(001). The reason for this is that these two combinations release the strain surrounding the 2-DV and 3-DV clusters.

The stable 1+2-DV complex consists of a rebonded 1-DV linked with the non-rebonded side of a 2-DV. The structure is illustrated in Fig. 4.5(d). The formation energy of the stable 1+2-DV complex is 0.42 eV, or 0.14 eV/dimer [62]. This is 0.13 eV lower than the sum of the formation energies of isolated 1-DV and 2-DV structures, which indicates that the combination is able to lower the strain at the junction of the two clusters. Linking the double vacancy from both sides to a single vacancy is energetically unfavorable due to rebonding on the other side of the

double vacancy.

The 1+3+1-DV complex is formed by linking a triple vacancy from both sides to a single, rebonded vacancy. In this case, the triple vacancy in the middle can be rebonded in two alternative ways as shown in Fig. 4.5(c). The energies of these two configurations are almost equal. In Ref. [62], the stable configuration was determined to be the one with rebonding between the exposed second-layer atoms in the middle of the 3-DV (shown in Fig. 4.5(e)). The formation energy of this 1+3+1-DV complex was calculated to be 0.73 eV, or 0.146 eV/dimer. This configuration is able to release the strain in both junctions.

The 1+2-DV cluster is the most frequently occurring structure in the typical STM images analyzed in Ref. [62]. Based on their *ab initio* calculations, the reason is that this is the only structure which minimizes the number of dangling bonds and, simultaneously, releases part of the strain at its junctions. As we will soon see, this structure has also been observed to dominate in ordered $2 \times n$ vacancy lines which have been observed to form on the Si(001) surface due to e.g. repeated quenching from high temperatures or exposure to contaminants such as Ni.

4.3.5 Ordered $2 \times n$ structures

All STM images of Si(001) show the presence of vacancies, but there has been discussion about their equilibrium distribution and ordering. We have already seen that the vacancies have a clear tendency to cluster into distinctive complexes due to strain-field mediated DV-DV interactions. The situation is, in fact, even more complicated. Several groups [65, 66, 67, 63, 68, 69, 70] have demonstrated that the dimer vacancy concentration can be manipulated, and once the concentration is sufficiently high, the vacancies form chain-like structures with long-scale periodicity. The orientation and atomic-scale structure of these vacancy lines depend on several factors, such as temperature, surface preparation technique and possible exposure to contaminants. At present, it is still partly unclear what role kinetic effects play in the formation of these ordered vacancy structures and which configuration represents the true equilibrium structure. In the following, we will review some of the questions related to the ordering of the dimer vacancies, especially because similar structures govern the morphology of the Ge-covered Si(001) surface.

The intrinsic dimer-vacancy concentration of the clean Si(001) surface can be increased by ion bombardment (Xe^+, Ar^+), etching (O_2, Br_2, I_2, etc.), or continued radiation quenching from high temperatures [67]. The activation energy for vacancy diffusion on the Si(001) surface is estimated to be less than 2 eV [71, 72], which allows the vacancies to become mobile at elevated temperatures (> 800 K). Their motion, in turn, leads to the formation of vacancy islands. When the concentration of vacancies is in the range $0.2-0.3$ ML, which is clearly more than the intrinsic defect density of Si(001) but still relatively low, the vacancy islands are

observed to have a highly anisotropic shape with their long axis aligned perpendicular to the direction of the dimer rows on the upper surface. These elongated islands then develop into long chains and eventually order into a periodic $2 \times n$ pattern formed by vacancy lines (VL) that extend for many thousands of Å [63]. At higher vacancy concentrations (above 0.3 ML), the orientation of the elongated vacancy islands is observed to rotate by 90° [67, 70]. These islands run parallel to the upper terrace dimer row and they are generally wider than the perpendicular vacancy lines. In agreement with these experimental observations, *ab initio* Molecular Dynamics calculations of Ref. [70] show that perpendicular lines are favored at low vacancy concentrations and the optimum width of the lines is two or three dimers. For higher vacancy concentrations, the calculations show a preference to parallel islands with an optimum width greater than three dimers.

The ordering of the dimer vacancies to a periodic line-structure is observed to be a strongly temperature-dependent process. Experimental results suggest that the ordered dimer-vacancy structures only exist in an intermediate temperature range [73]. The initial increase of the vacancy concentration using a technique such as ion sputtering leads to a random distribution of single vacancies and small vacancy clusters. An ordered line structure can be formed by annealing the surface at temperatures ranging from 600°C to 850°C. However, if the temperature is raised above 950°C, the random configuration can be completely restored. On the other hand, no ordering is observed at temperatures below 500°C. This implies that the transition temperature for ordering is between 500°C and 600°C.

The temperature-driven ordering of the dimer vacancies was investigated by Kim *et al.* [73] using tight-binding total-energy calculations. Their results suggest that dimer-vacancy clusters diffuse through a set of intermediate random configurations toward an ordered line structure. At low temperatures, the thermal energy is not sufficient to overcome the diffusion barrier for vacancies and therefore the initial random distribution remains unchanged. Using previously calculated values of 1.7-2.3 eV [71, 72] for the height of the diffusion barrier, the transition temperature was estimated to be about $380 - 610$°C which is in good agreement with the experimentally observed range of $500 - 600$°C. Above the transition temperature, thermal energy is sufficient for migration of the dimer vacancies and ordering will take place. However, if the temperature is further increased, the configurational entropy will begin to dominate the free energy of the system and consequently the ordered lines break into a random distribution of vacancies. This reasoning explains why the ordered vacancy lines exist only at intermediate temperatures.

In addition to the etching and sputtering techniques, vacancy lines also form spontaneously on the Si(001) surface if it is subject to Ni contamination [66]. Only a small amount of Ni, less than 1%, is sufficient to increase the vacancy density from below 2% to above 20% [68], and the intrinsic interaction between the dimer vacancies drives the ordering process. STM images [74, 68, 66] of the Ni-induced

$2 \times n$ vacancy-line structure show that the lines are mainly composed of double vacancies (2-DV) and 1+2-DV clusters. Some authors have suggested that the Ni-induced vacancy lines would be deeper than one atomic layer, but the high-resolution contour plot of Koo *et al.* [68] corresponds well to the atomic structure of the 1+2-DV cluster suggested by the *ab initio* study by Wang *et al.* The greater depth of the vacancy line is explained by the asymmetric rebonding of the 2-DV structure which leaves a deeper 'pitch' in the nonrebonded region. Fig. 4.7 shows a schematic illustration of the atomic structure of a 1+2-DV section of a vacancy line. In contrast to the DV lines formed by etching or sputtering, the Ni-induced lines cannot be annealed away even at high temperatures which implies that the $2 \times n$ structure is stabilized by Ni contamination [68, 66].

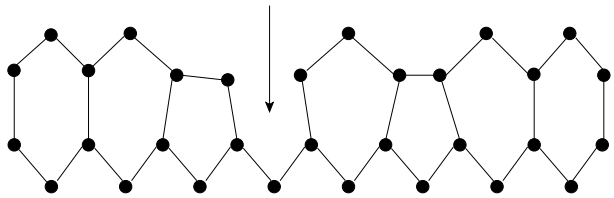

Figure 4.7: Illustration of the atomic structure of a 1+2-DV complex. The surface is viewed from the side. Four topmost atomic layers are shown. The arrow marks the position of the deepest pitch in the vacancy complex.

Men *et al.* [63] used the Stillinger-Weber potential to examine the interaction between adjacent vacancies in neighboring dimer rows. Their calculations show that a surface with two far-separated 2-DVs has an energy which is 0.08 eV higher than that of a surface where the vacancies are in neighboring dimer rows. For the 1+2-DV complex, the corresponding increase in energy is 0.28 eV. This excess energy represents an *attractive* interaction between two adjacent DVs. For two DVs in the same dimer row, the interaction is long-ranged and *repulsive*. This can be seen for example from Fig. 4.6 which shows the relative energies of the $2 \times n$ structures as a function of the vacancy-line separation (n). The long-range interaction decays as n^{-2} as a consequence of surface strain relief, which can also be theoretically justified by energetic considerations of the domain-wall structure [67]. Later in this thesis, we will see that dimer vacancies on the Ge-covered Si(001) surface behave very similarly, although the reason for their initial formation is different.

Chapter 5

Ge on Si(001)

5.1 Structural evolution

5.1.1 Experimental observations

The growth of Ge on the Si(001) surface is strongly influenced by the strain induced by the 4.2% lattice mismatch between Si and Ge. The initial stages of submonolayer growth of Ge are quite similar to Si homoepitaxy, but near the completion of the first monolayer both the kinetics and thermodynamics of further Ge growth change dramatically. As the amount of Ge increases, the surface undergoes a structural evolution driven by the need to relieve strain. The morphological changes are accompanied by complex effects on the step-edge roughness and surface stress field. In the following, we discuss experimental and theoretical results which contribute to our understanding of the underlying atomic-scale mechanisms that govern the growth of Ge on Si(001). The focus is mostly on quasi-equilibrium Ge films, corresponding to typical experimental growth conditions ($500-700^{\circ}$C), but attention is also given to important kinetic effects.

Initial strain relaxation in the growing Ge overlayer is achieved primarily by the formation of additional dimer vacancies. Several theoretical calculations (e.g. Refs. [75, 76, 64] and publication VI of this thesis) show that dimer vacancies have negative formation energies on the Ge-covered Si(001) surface with respect to a 2×1 reconstructed Ge overlayer. The kinetics of vacancy formation are not known, but experimental and theoretical evidence suggests that their formation is also kinetically favored on Ge-covered Si(001) because Si-Ge and Ge-Ge bonds are easier to break than Si-Si bonds [46].

The concentration of dimer vacancies increases as a function of Ge coverage, but initially the vacancies are randomly distributed on the surface. STM studies [77, 78, 79] show that a well-defined $2 \times n$ pattern appears when the Ge coverage reaches 0.8 ML. Further deposition leads to a decrease in the average value of n until saturation is reached at about 2 ML coverage. The distribution of

measured values of n alternates somewhat depending on the experimental conditions. For example, at 0.8 ML coverage, the measured average values of n range from $n = 17$ [79] to $n = 11$ [78]. Around 1.5 ML coverage, the n values have decreased to $n = 9-10$ [77, 78, 79], and convergence to $n \sim 8$ is reached around 2 ML coverage [79].

In addition, experiments show that the distribution of n is broad for low Ge coverages and becomes narrower as the thickness of the Ge overlayer increases [78]. The distribution of n is related to the number of kinks in the vacancy lines. When the vacancy lines first appear on the surface, they have a disordered structure with many kinks, and consequently, the average spacing between vacancies in the same dimer row has a wide distribution. The ordering of dimer vacancies into straight lines becomes energetically more favorable as the vacancy concentration increases, and it is reflected by the narrowing of the distribution of the average line separation.

The $2 \times n$ reconstruction becomes less effective in relieving the growing strain as more Ge is deposited on the surface. Experiments show that other structures begin to appear when the Ge film thickness exceeds 2ML. For example, a so-called $m \times n$ structure emerges where additional vacancy trenches (so-called dimer-row vacancies) form in the direction perpendicular to the vacancy lines in the $2 \times n$ reconstruction [42]. The resulting structure is a rectangular grid of intersecting vacancy lines. The separation m between the newly formed trenches is observed to decrease fast with increasing Ge coverage which implies that the dimer-row vacancies are less effective in relieving strain than the vacancy lines in the $2 \times n$ structure. Indeed, three-dimensional islands begin to form on the surface soon after the appearance of the $m \times n$ pattern. The critical thickness of the wetting layer in MBE growth is approximately 3 ML [42].

5.1.2 Surface energy calculations

Before discussing the results obtained in different calculations concerning the evolution of the Ge/Si(001) surface, we will address the question of how the energies of surfaces with different stoichiometries should be compared. In all theoretical calculations, we have to use a computational cell composed of a limited number of atoms. If we are studying a system such as the $2 \times n$ reconstructed Ge/Si(001) surface, we are forced to change the number of atoms in the surface layer depending on the periodicity of the reconstruction. For example, if the computational cell has a surface size of 20×20 atoms, then on a 2×1 reconstructed surface the topmost layer is composed of 400 atoms, while on a 2×10 reconstructed surface, it only contains 360 atoms. As a consequence, an appropriate chemical potential must be assigned for the removed atoms in order to compare the relative stabilities of surfaces with different stoichiometries.

This question was first addressed by Oviedo *et al.* [64] who introduced an

intelligent procedure for determining the chemical potential in such a way that close correspondence to experimental conditions is achieved. The idea is to use the ideal 2×1 reconstructed surface as a reference system and determine the relative stabilities of differently reconstructed surfaces with respect to the reference surface. The procedure for determining the relative stabilities does not use the bulk cohesive energy as the chemical potential, as is commonly used in surface energy calculations, but instead a redistribution of a fixed number of particles is considered. In the following, we first consider the theoretical idea behind the procedure, and then explain how the calculations are done in practice.

Let us consider a system with a fixed number of Ge atoms. Assuming that the surface is in thermal equilibrium, the probability of observing a particular configuration A is proportional to the Boltzmann factor $\exp(-E_A/k_BT)$ where E_A is the energy of configuration A. Corresponding to experiments, we are interested in the relative stability of different reconstructions obtained by distributing a fixed number of Ge atoms (i.e. a fixed amount of deposited material) in different ways. We take the 2×1 reconstruction as our reference point, and calculate the relative energies of different $2 \times n$ reconstructions with respect to the reference system. The same approach can easily be generalized for studies of different structures, e.g. the $m \times n$ pattern.

In order to understand the idea behind the computational procedure, consider a large section of the Si(001) surface covered with a 2×1 reconstructed layer of Ge. The shape of the area is unimportant, but the area is thought to be so large that edge effects are negligible. Let E_p denote the energy of a Ge dimer in the 2×1 reconstructed layer. Next an appropriate amount of Ge atoms is fetched from infinity and replaced at the borders of the Ge layer, which increases the energy of the system by E_p per additional Ge dimer. The $2 \times n$ reconstruction can now be formed by removing every nth dimer from each row and taking the removed dimers to infinity (the number of previously added atoms was chosen to equal the number of removed atoms). The energy cost of forming the dimer vacancies is denoted by E_f (per dimer). Combining the adding and removal of dimers, we obtain the energy change when going from the initial 2×1 reconstruction to the $2 \times n$ arrangement, denoted by $\xi(n)$, and given by

$$\xi(n) = \frac{E_f(n) + E_p}{n - 1}. \tag{5.1}$$

The relative stability of different $2 \times n$ reconstructions can now be determined by calculating the value of ξ as a function of n and finding the minimum. Positive values of ξ indicate that the configuration in question is less stable than the 2×1 reference surface.

The *computational* procedure for calculating ξ consists of determining the value of E_p and the values $E_f(n)$ for a range of n values. In practice, these are obtained from simulations as follows: For a given Ge coverage (e.g. 1 ML), (i)

E_p is given by the energy difference between a system with a 2×1 reconstructed surface and a system obtained by removing the top layer and relaxing the resulting surface with a 2×1 arrangement (e.g. using slab geometry with two surfaces in the system, if the reference system has 1 ML of Ge on both surfaces, the removal of the top layer yields 2×1 reconstructed Si-surfaces with the total number of atomic layers in the simulation system reduced by two). (ii) $E_f(n)$ is given by the energy difference per Ge dimer between a system with a $2 \times n$ reconstructed top layer and the reference surface with a 2×1 reconstruction. It is essential to notice that in both calculations, the total number of Ge atoms in the two systems is different. This should not be confused with the *conceptual* idea of comparing the relative energies of surfaces obtained by distributing a fixed number of Ge atoms in different ways.

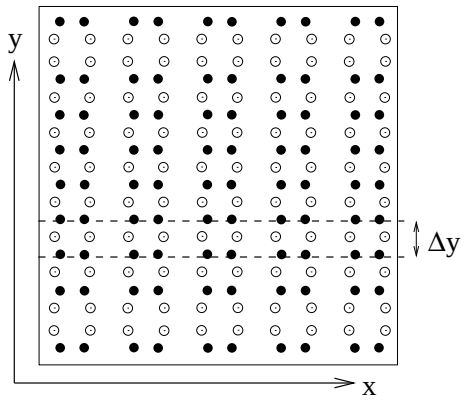

Figure 5.1: Illustration of a $2 \times n$ reconstructed surface ($n = 8$). Filled circles represent atoms in the top layer and open circles atoms in the layer beneath. Δy shows the direction in which the separation between rows of dimers is measured.

For full-monolayer coverages, the reference system consists of a predefined number of Ge layers (e.g. 1, 2 or 3 ML) with a 2×1 reconstructed surface. For submonolayer coverages, the surface layer consists of a random mixture of Ge-Ge, Si-Ge and Si-Si dimers with the percentage of Ge atoms given by the coverage; e.g. for 0.8 ML coverage, 80% of the top layer atoms are Ge. For fractional coverages exceeding 1 ML (e.g. 1.5 ML), the top layer is 100% Ge (due to the lower surface energy of Ge with respect to Si) and the layer beneath consists of a random Si-Ge mixture.

For all coverages, the $2 \times n$ reconstruction is obtained by removing every nth dimer from each dimer row in the surface layer. Figure 5.1 shows the structure of a $2 \times n$ reconstructed surface where the exposed second-layer atoms are rebonded. In simulations, top-layer atoms are usually dimerized in the initial configuration, but rebonding is not imposed initially (it is expected to occur spontaneously during

the relaxation). The surface area of the simulation cell must be chosen to match the periodicity of the $2\times n$ pattern; e.g. a 20×20 surface can be used for 2×4, 2×5 and 2×10 reconstructions. For details of how the relative stability of different $m \times n$ patterns can be compared, see Ref. [80].

5.1.3 $2 \times n$ reconstruction

In the following, we discuss the energetics and structure of the $2 \times n$ reconstruction. We compare results obtained by using *ab initio*, tight binding and empirical potential calculations and discuss how these results agree with experimental observations. In this section, we consider fully segregated Ge layers on the Si(001) surface, whereas the next section addresses the question of possible Si-Ge intermixing.

We begin by discussing the evolution of the optimal periodicity of the $2 \times n$ reconstruction as a function of the Ge coverage on the surface. The relative stability of the different $2 \times n$ reconstructions is determined in the manner described in the previous section. Figure 5.2 shows a plot of $\xi(n)$ as a function of n for Ge coverages of 0.5, 0.8, 1.0, 2.0 and 3.0 ML, obtained from Monte Carlo simulations using the Stillinger-Weber potential. The corresponding plot of $\xi(n)$ for a clean Si(001) surface was shown earlier in Fig. 4.6. We notice that the ξ values are all positive for the clean Si(001) surface, while for the Ge-covered surfaces the values are mostly negative. This indicates that the $2 \times n$ reconstruction is stabilized by the Ge layer, and the energy gain increases as the Ge layer becomes thicker.

The optimal periodicity for each coverage is given by the minimum of the $\xi(n)$ curve. For 0.5 ML coverage, the vacancy lines are energetically favorable only for very large values of n and there is no minimum value for $n \leq 16$. This indicates that while a low concentration of dimer vacancies is favorable at small submonolayer coverages of Ge, the vacancy-vacancy interactions are likely to be too weak for an organized line pattern to appear on the surface. As mentioned, experiments show that ordering of the vacancies is initiated around 0.8 ML coverage, and indeed, our SW calculations of $\xi(n)$ also show that a weak minimum appears at 0.8 ML coverage between $n = 12 - 14$. Further increase in the thickness in the Ge layer leads to a shift in the optimal value of n such that at 1.0 ML coverage the minimum is at $n = 10$, and at about 2 ML coverage a saturation point is reached after which the optimal value remains at $n = 8$.

The shift toward smaller n values is accompanied by an increasing energy gain until the saturation point is reached at 2 ML coverage. Further reduction of the periodicity is energetically unfavorable due to a repulsive interaction between neighboring vacancy lines. The repulsion is mediated by subsurface atoms in the vicinity of the vacancy line which are distorted from the usual diamond lattice positions. The interaction between two local strain fields results in an effective repulsion between the vacancy lines.

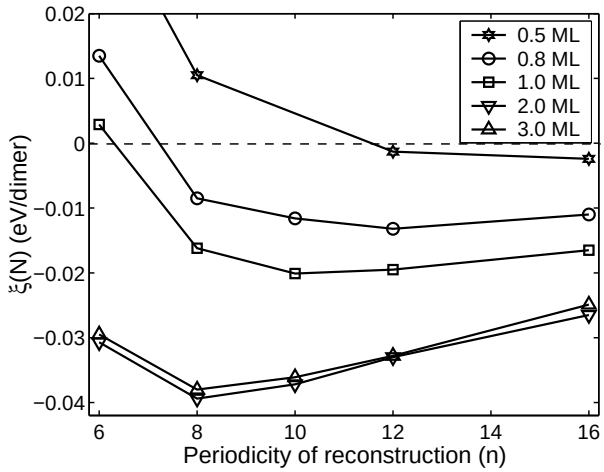

Figure 5.2: Energy change per Ge dimer when a $2 \times n$ reconstruction is formed from a 2×1 reconstructed reference surface. Negative values indicate that the $2\times n$ reconstruction is stable with respect to 2×1 ordering. The minimum of the curve gives the most favorable spacing between the dimer-vacancy lines. The results are from MC simulations using the Stillinger-Weber potential at 11 K.

These results were obtained from large-scale Monte Carlo simulations where typical system sizes were in the range from 15 000 to 35 000 atoms. The surface size of the simulation cell was kept relatively small (typical values were around 20×20), but the bulk section of the slab was thick (typically 59 atomic layers) in order to have close correspondence to experimental systems (i.e. to simulate a thin layer on a much larger substrate). We would like to point out that the constant-pressure algorithm (described in Section 2.1.6) allows the sides of the simulation cell to fluctuate independently which leads to anisotropic changes in the x and y dimensions of the system (due to the strain induced by the dimerization of the surface atoms). In order to eliminate this effect, we always used slabs where the two surfaces have perpendicular dimer orientations (i.e. an odd number of atomic layers in the z direction). This question is discussed in detail in Publication V.

Determination of the $\xi(n)$ values requires calculations involving extremely small energy differences which are then multiplied by much larger factors. Consequently, the simulations must yield very accurate energy values in order to be meaningful. In Fig. 5.2, the results are given with error bars smaller than or comparable to the size of the symbols. For details of the energy calculations, see publication VI.

Our SW results can be directly compared to those of two other studies which are based on first-principles [64] and tight-binding [80] calculations. Figure 5.3 shows a plot of $\xi(n)$ for 1, 2 and 3 ML Ge coverages obtained from tight-binding

calculations of Oviedo *et al.* [80]. The results for the 1 ML case agree fairly closely with corresponding DFT calculations of Ref. [64]. The common feature in all of the studies is that they predict that the $2 \times n$ reconstruction is stable on Ge-covered Si(001) surfaces with respect to the 2×1 ordering.

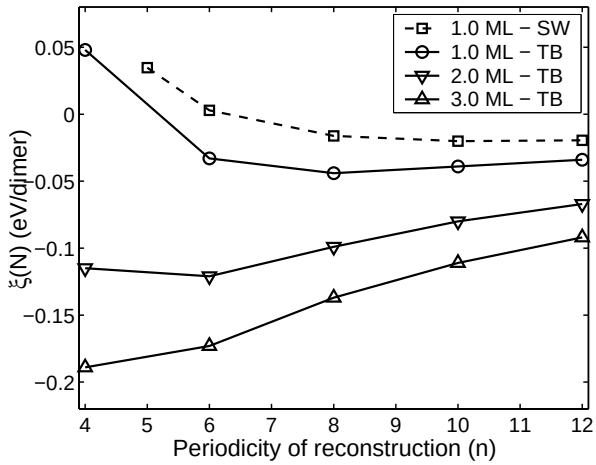

Figure 5.3: Results of tight-binding calculations showing the energy change per Ge dimer when a $2 \times n$ reconstruction is formed from a 2×1 reconstructed reference surface. For the 1ML coverage, Stillinger-Weber results are also given for comparison (the same data as shown in Fig. 5.2). The TB data are from Ref. [80].

In both the SW and TB calculations, the optimal periodicity of the vacancy-line structure is observed to decrease as a function of increasing Ge coverage. However, the results are found to disagree in their quantitative predictions of the optimal value of the vacancy line spacing. Table 5.1 shows a comparison between experimental data and results obtained from the SW and TB calculations. Surprisingly, the empirical SW potential gives a much closer agreement with experimental observations than do the TB calculations. The authors of Ref. [80] suggest that the discrepancy between their TB calculations and experiments may be explained by kinetic effects which may cause the experimentally observed structures to be out of thermal equilibrium. This explanation cannot be fully adequate because in our MC-SW simulations, the surface is in full thermal equilibrium, as in the TB calculations, and yet the SW results correspond closely to experimentally observed changes in the $2 \times n$ reconstruction of the surface. In addition, in the review of Liu *et al.* [46], it is explained that the structure and morphology of experimentally grown thin films is governed by thermodynamics due to the use of high temperatures ($500 - 700°$C) or long annealing times in typical experiments. Consequently, the experimental observations are made at a quasi-equilibrium state where the surface layers are very close to thermodynamic equilibrium but equilib-

rium between the surface and the bulk is not established (see the following section on Si-Ge intermixing for an extended discussion on this subject).

Table 5.1: Optimal periodicity of the $2 \times n$ reconstruction as a function of Ge coverage. The SW results are from publication VI and the TB results from Ref. [80]. The experimental data are collected from various STM measurements in Refs. [77, 78, 79].

θ_{Ge} (ML)	0.5	0.8	1.0	1.5	2.0	3.0
exp.	-	11-17	12	9-10	8	8
SW	> 16	12-14	10		8	8
TB			8		6	4

For comparison purposes, we repeated the Monte Carlo simulations using the empirical Tersoff potential T3 instead of the SW model. The calculations give the optimal periodicities of $n = 10, 6$ and 6 for Ge coverages of 1, 2 and 3 ML, respectively. These values are much closer to the corresponding results obtained using the SW potential ($n = 10, 8, 8$) than to those obtained from the TB calculations. The differences between the T3 and SW data result only from the use of two different potential models, whereas the TB results were obtained from static calculations of much smaller systems at constant volume (constant-pressure simulations were used for the SW and T3 calculations). It seems that the TB model does not fully predict the repulsive interaction between neighboring vacancy lines, possibly because the simulation cells were only one dimer wide and from four to twelve dimers long with a slab thickness of ten atomic layers [80]. This is much less than the systems used in the MC simulations where a typical system size was e.g. $20 \times 20 \times 59$. Consequently, the TB calculations may underestimate the energetic cost of introducing tightly-spaced vacancy lines on the surface.

Let us now consider some aspects related to the distribution of the average vacancy-line separation. As mentioned, experiments show that the distribution is broad for low Ge coverages and becomes narrower as the amount of Ge increases. The same trend can be seen in Fig. 5.2 where the relative energies for $n = 12 - 16$ are almost degenerate at 0.8 ML coverage, but as the Ge coverage increases, the potential well around the minimum in the $\xi(n)$ curve becomes deeper and narrower. This suggests that the straightening of the vacancy lines becomes energetically more favorable when the Ge film thickens.

On the other hand, the distribution of the average vacancy-line separation is directly related to kinks in the vacancy lines. In a tight-binding study [80], Li *et al.* calculated the energy cost of forming kinks in a line of missing dimers for Ge coverages of 1, 2 and 3 ML. The kink-formation energy is given by the energy change when a missing dimer is displaced from a straight-row configuration by a distance of one or two dimers in the direction perpendicular to the vacancy line.

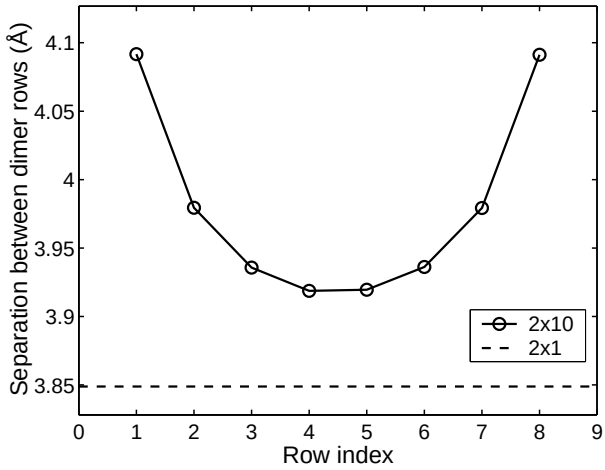

Figure 5.4: Average separation between rows of dimers on a 2×10 reconstructed surface with 1 ML Ge coverage. The average row separation on a 2×1 reconstructed surface is given for comparison. The row separation is measured in the direction perpendicular to the dimer bonds (shown as Δy in Fig. 5.1. The vacancy lines are situated at positions with indices 0 and 9. The results are from MC simulations using the Stillinger-Weber potential at 11 K.

The results of the tight-binding calculations show that a straight vacancy line is the energetically most favorable configuration for all Ge coverages. However, the formation energy of a kink of length 1 unit is only 0.036 eV for 1 ML Ge coverage, which implies that a fair amount of kinks will be present at typical experimental temperatures. The formation energy becomes larger as the Ge coverage increases: the values 0.069 and 0.089 eV were obtained for coverages of 2 and 3 ML, respectively. In agreement with experiments, the results suggest that the probability of kink formation decreases as the Ge coverage increases. The formation energy of 2-unit deep kinks is in all cases clearly larger than for kinks of depth 1.

The physical origin of this effective attraction is a subtle relaxation of the local atomic structure around the two neighboring vacancies in adjacent rows. Ordering of the vacancies in straight rows allows a better local relaxation of the atoms in the underlying layers due to the overlap and interference of the neighboring strain fields.

These results provide strong evidence that the dimer vacancy lines serve as a strain-relief mechanism in which the major contribution comes from large-scale elastic effects. The simplest explanation is that the vacancy lines provide room for the expansion of the Ge overlayer which effectively reduces the compressive strain in the system. Figure 5.4 shows the average separation between rows of dimers on a 2×10 reconstructed surface with 1 ML Ge coverage (see Fig. 5.1 for an illustration of how the distance is measured). For comparison, also the average

spacing on a 2×1 reconstructed surface is shown. Two vacancy lines are located at the edges of the surface area shown in the figure. We notice that the separation between the Ge atoms increases significantly in the whole region between the two vacancy lines when compared to the 2×1 surface. The relaxation is largest near the vacancy lines. Decreasing the periodicity provides more room for expansion, which is an effective way to reduce the strain energy in the system because the energetic cost of creating the dimer vacancies is relatively low due to rebonding of the exposed second-layer atoms. The optimal periodicity of the $2 \times n$ structure is determined by the balance between the energetic gain from the strain relaxation and the cost from forming the vacancies.

5.2 Si-Ge intermixing

Computational details

Despite the extensive efforts which have been dedicated to the morphology of Ge films on Si(001), there are still some gaps in our understanding of the complex mechanisms which influence the surface structure and stoichiometry. One of the fundamental issues is the possibility of Si-Ge intermixing which can change the stoichiometry of the surface layers and has profound consequences on the morphology, energetics and strain relaxation. STM and other experimental techniques are able to characterize the surface structure with atomic-scale resolution, but so far a direct measurement of the stoichiometry of the entire wetting layer has not been achieved because distinguishing between Si and Ge atoms is extremely difficult due to their similar atomic, electronic and chemical properties [46]. Contrary to earlier beliefs, recent experimental evidence suggest that a significant amount of intermixing may occur at typical growth temperatures. Computational methods have played an important role in building an understanding of the mostly indirect experimental information.

Intermixing is a complicated topic to study due to an interplay of different thermodynamic driving forces related to strain relaxation, effects of temperature and kinetic factors. Accurate first-principles studies can be used to obtain valuable information about the binding energies of individual Ge atoms at various lattice positions on the Si(001) surface, and they can be used to identify the mechanisms for place-exchange or subsurface diffusion processes. The main drawback of these methods is that the calculations are conducted in the zero-temperature limit after which the distribution at nonzero temperatures is estimated using e.g. the Boltzmann or the Fermi-Dirac distribution. In addition, effects due to coverage dependence (Ge-Ge correlations) and surface reconstruction are almost always neglected.

The Monte Carlo method, on the other hand, can be combined with a computationally effective classical potential, which allows one to perform simulations of

sufficiently large systems at any particular value of temperature with all correlation effects etc. intrinsically included in the simulation. In the MC approach, the idea is to obtain the equilibrium distribution for a given Ge coverage directly from the simulation where, in addition to the usual MC moves, randomly selected pairs of Si and Ge atoms can exchange positions. Each attempted exchange is accepted or rejected using the usual Metropolis criterion. In this so-called random-switch algorithm (publication VI), the total number of both types of atoms (Si and Ge) is conserved, and thus in a given simulation the Ge atoms are distributed over the system corresponding to the free-energy minimum.

The random-switch MC algorithm is a very fast method to achieve large configurational changes in the system, which would not be possible using regular single-particle moves in the MC algorithm or a method such as Molecular Dynamics where the particles must follow the trajectories dictated by the equations of motion. The MC method cannot be used to study the actual diffusion mechanism behind the intermixing processes but it is a useful tool for studying the thermodynamic driving forces behind the experimentally observed changes. Real systems are, of course, affected by kinetic limitations, but these can be estimated, at least to a certain extent, by investigating the system at two limiting cases: in the total absence of intermixing and in full thermodynamic equilibrium.

In addition, it is in some cases justifiable to consider a quasi-equilibrium case where intermixing is allowed to occur only in a region near the surface but the bulk region consists of pure Si. In real systems, bulk diffusion of Ge is practically negligible in usual growth conditions, but as we will soon see, several experiments indicate that near the surface Ge penetrates down to at least the fourth layer at temperatures as low as 400°C. The physical origin of the much faster diffusion rate near the surface is explained by differences in the diffusion mechanism. In bulk, Ge diffuses by two main mechanisms: interstitial exchange and vacancy diffusion. The DFT calculations of Uberuaga *et al.* [81] show that the interstitial-exchange process is clearly faster in the surface region because the formation energy of Si interstitials is significantly lower near the surface than in the bulk. Thus restricting the intermixing in simulations to a suitable region near the surface gives a better correspondence to the experiments than looking at the full thermodynamic equilibrium.

5.2.1 Place-exchange in the submonolayer regime

In a recent STM study by Qin *et al.* [82], empty-state imaging was successfully applied to directly identify atomic-scale intermixing sites on Si(001) at submonolayer Ge coverages. Such a differentiation between Si and Ge atoms had not been achieved prior to this study and it was commonly believed that it is not possible to distinguish Si and Ge on Si(001) using STM. In their study, Qin *et al.* were able to demonstrate that a measurable degree of place exchange occurs in the top

layer at temperatures as low as 330 K. In addition, the study shows that Ge/Si place exchange occurs randomly on the surface and steps or point-defects are not preferential intermixing sites. These results are in agreement with the findings of another advanced STM study by Lu *et al.* [83] where the different dynamic appearance of the mixed Si-Ge dimer was used to distinguish it from the Si-Si and Ge-Ge dimers. The abundance of mixed Si-Ge dimers indicates that the exchange process between deposited Ge atoms with Si atoms in the first layer involves a rather low energy barrier.

These STM measurements confirm the results of earlier indirect measurements [84, 85] which already suggested that below 1 ML Ge coverage, the surface layer consists of a mixture of Ge and Si atoms. In a displacive adsorption process, the deposited Ge atoms exchange places with Si atoms in the surface layer to form mixed Si-Ge dimers. The displaced Si atoms diffuse over the surface and are incorporated into steps. In the surface layer, the Ge content increases as more Ge is deposited until at 1 ML coverage the surface is completely covered by Ge. This displacive adsorption process is an essential part in understanding the gradual formation of the $2 \times n$ reconstruction in the submonolayer regime. If the Ge atoms would all adhere to step edges, then contiguous areas of Ge-covered surface would form from the very beginning of the deposition process, and consequently, a well-defined $2 \times n$ pattern should appear immediately in those areas. The surface would then be characterized by the coexistence of $2 \times n$ and 2×1 phases in the submonolayer regime. Instead, the ordered vacancy lines appear gradually with a poorly ordered $2 \times n$ pattern appearing around 0.8 ML coverage over the entire surface.

Several experiments [84, 82, 83] confirm that the Si-Ge place exchange occurs even at room temperature, whereas Ge incorporation to deeper layers involves a much higher energy barrier. In Ref. [84], measurements at 500°C show no evidence of incorporation into subsurface sites at 1 ML coverage, whereas in Ref. [85], annealing at submonolayer coverage to 600°C was reported to lead to an interdiffusion into deeper layers. More insight into the situation can be obtained by comparing these observations with results from recent computational studies.

At low temperatures, all Ge is expected to stay in the surface layer due to its lower surface energy with respect to Si. In all theoretical calculations, including both *ab initio* [81, 86, 87] and empirical potentials (our results, publication VI), the energy cost of substituting a Ge atom for Si is lowest for the topmost layer. Empirical potentials cannot produce dimer buckling, but according to the *ab initio* calculations, the Ge adsorption site corresponds to the up-atom position. The mechanism for the mixed Si-Ge dimer formation was investigated in detail by Ko *et al.* [87] using *ab initio* total-energy calculations. They concluded that the mixed Si-Ge dimer offers only a marginal energetic gain compared to a Ge-Ge growth mode. Furthermore, the displacive adsorption of individual Ge atoms was found to be unfavorable, both energetically and kinetically. Instead, the mixed Si-Ge

dimer formation was proposed to occur via an atomic exchange process during the dimerization of Ge atoms. The required activation energy was calculated to be 0.1 eV. This exchange-dimerization process may explain the predominant formation of mixed Si-Ge dimers at low Ge coverages. At the moment there is no direct experimental evidence of the microscopic process leading to the mixed dimer formation that would confirm the proposed mechanism of Ko *et al.*, and thus this question still remains partly open.

5.2.2 Intermixing in subsurface layers

At high temperatures, the probability of Ge diffusing to deeper layers increases. Several experiments [88, 89, 90, 91, 81] indicate significant subsurface Ge occupation even at submonolayer coverages at temperatures in the range $400-700^{\circ}$C. Varying subsurface distributions have been reported, but in all results, Ge is incorporated down to at least the fourth layer below the surface with most of the Ge still remaining in the first layer. Precise determination of the atomic-scale configuration in deeper layers is not possible using current experimental techniques, but computational studies have helped to identify the sites where the Ge atoms may be incorporated in the subsurface region.

The simplest approach is to calculate the energy cost of substituting a Ge atom in different sites near the Si(001) surface. Table 5.2 summarizes the results obtained using empirical potentials (Stillinger-Weber and Tersoff) (our results, publication VI) and DFT calculations (using the LDA approximation [86, 81] and the nonlocal PW91 exchange-correlation functional [81]). Figure 5.5 shows the indexing of the sites in layers $1-4$. All calculations show that the first layer is most favorable to Ge atoms with the up-atom site (1a) favored over the down-atom site (1b) in the DFT calculations. The compressively strained sites under the dimer rows (3a and 4a) are clearly higher in energy than are the tensile sites under the troughs (3b and 4b). For the second layer sites, the Tersoff potential gives an energy value which is clearly higher than the binding energy for both the compressive and tensile sites in layers 3 and 4. This result is in disagreement with all the other calculations where the relative order of the sites in increasing energy is 1, 2, 3b, 4b, 2, 4a, 3a.

In order to compare the computational results with existing experimental data, it is necessary to obtain the distribution of Ge atoms in subsurface layers at temperatures corresponding to the experimental conditions. In the DFT studies, the occupation probability of Ge atoms in various sites has been estimated using either the Fermi-Dirac distribution (corresponding to a situation where the number of Ge atoms in the unit cell is not fixed) or the Boltzmann distribution (corresponding to a fixed number of Ge atoms in the unit cell). In these calculations, the coverage dependence of the Ge binding energy is neglected (the occupation probability of a given site is assumed not to depend on the occupation of neigh-

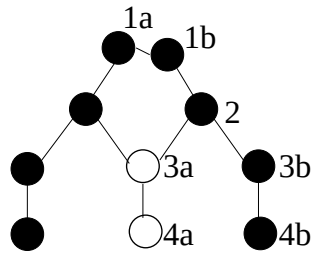

Figure 5.5: Illustration of the intermixing sites at the Ge/Si(001) surface. The reconstruction of the surface leads to the formation of two inequivalent sites in layers 3 and 4 (marked with a and b). In addition, the first layer sites are denoted as 1a (up-atom) and 1b (down-atom) corresponding to dimer buckling. The lattice sites which are under compression and thus unfavorable to Ge occupation are marked with open circles.

Table 5.2: Energy cost of substituting a Ge atom for Si in different sites near the Si(001) surface relative to the first layer (in eV). For the DFT results the difference between sites 1a (buckling-up) and 1b (buckling-down) is also given. The DFT calculations are from Refs. [86](LDA1) and [81] (LDA2 and PW91). The empirical potential calculations (SW and Tersoff) are our results from publication VI. The surface is 2×1 reconstructed in all cases.

Layer	SW	Tersoff	LDA1	LDA2	PW91
1a	0	0	0	0	0
1b	-	-	0.091	0.134	0.149
2	0.244	0.230	0.211	0.314	0.363
3a	0.289	0.133	0.249	0.361	0.415
3b	0.183	0.121	0.155	0.292	0.344
4a	0.274	0.173	0.237	0.344	0.419
4b	0.199	0.103	0.160	0.291	0.348

boring sites) and effects due to changes in the surface reconstruction (from 2×1 to $2 \times n$) are not considered. In contrast, in the MC method with the random-switch algorithm, all these factors can be directly included in the simulations, and as an outcome we obtain the equilibrium distribution for given values of Ge coverage and temperature.

Table 5.3 shows a comparison of the probability distribution of Ge in subsurface layers obtained using the MC-SW method (our results, publication VI) and DFT calculations [86, 81]. Two different cases are considered: 1.0 ML of Ge at 600 °C and 1.6 ML at 700 °C. The surface is 2×1 reconstructed in all calculations. The agreement is fairly good, especially when taking into account the differences in the techniques for obtaining the occupation probabilities. Fur-

thermore, in Ref. [81], the DFT results for the 1.6 ML case were compared with Auger electron diffraction (AED) measurements of a corresponding experimental system and the results were found to be in good qualitative agreement. It can be concluded that our SW-MC results are in similar agreement with the experiment since the differences between the SW-MC and DFT values are very small.

Table 5.3: Comparison of Ge occupation probabilities in subsurface layers obtained using the SW potential based MC method (publication VI) and DFT calculations from Refs. [86] (DFT1, LDA) and [81] (DFT2, PW91). The first set of results (DFT1, SW1) is for 1.0 ML Ge coverage at 600 °C and the second set (DFT2, SW2) is for 1.6 ML coverage at 700 °C. The SW distributions are a direct outcome from simulations, whereas the DFT results have been obtained by from the Ge binding energies using either the Boltzmann distribution (DFT1) or the Fermi-Dirac distribution (DFT2). The surface is 2×1 reconstructed in all cases. The statistical errors in the SW data are of the order 0.001.

	1.0 ML at 600 °C		1.6 ML at 700 °C	
Layer	DFT1	SW1	DFT2	SW2
1	0.648	0.694	0.893	0.836
2	0.131	0.095	0.257	0.251
3a	0.028	0.041	0.156	0.137
3b	0.082	0.168	0.303	0.367
4a	0.032	0.063	0.150	0.183
4b	0.078	0.149	0.293	0.338

Detailed quantitative comparison of the computational data to experimental measurements is not possible at the moment because such information cannot be produced using current experimental techniques (due to the difficulties in distinguishing between Si and Ge atoms). Indirect techniques have been used in several experimental studies, but the agreement between the different sets of measurements is not particularly good. The variations can be explained in part by the difficulties in measuring subsurface intermixing and in part by different experimental conditions. Due to these difficulties, computational studies have turned out to be extremely valuable in investigating questions which are inaccessible using current experimental techniques. In the following, we discuss how intermixing is connected to the reconstruction of the surface, and show that the computational results provide a reasonable explanation for qualitative features observed in experiments.

5.2.3 Site selectivity and strain relaxation

We now consider the energetic reasons for intermixing by investigating the system in the low-temperature limit where entropic effects do not contribute to the prob-

lem. We begin by discussing the site selectivity of intermixing which is caused by to main factors: dimerization of the surface and presence of dimer-vacancy lines. Then we consider how the relative stability of the different $2 \times n$ reconstructions is affected by the onset of intermixing.

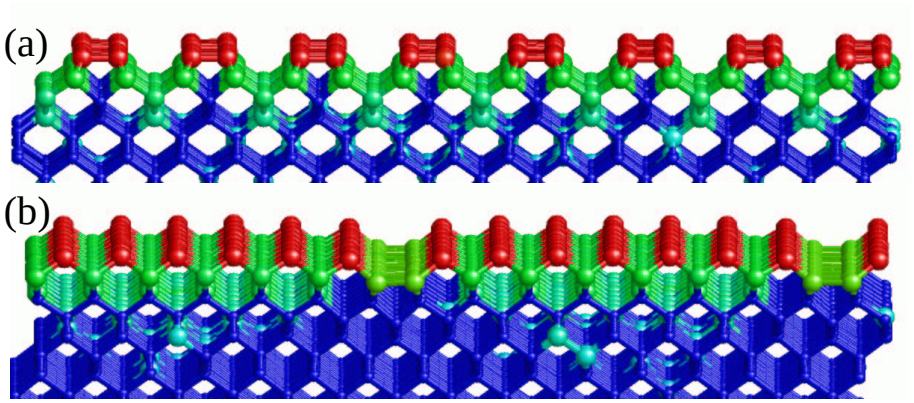

Figure 5.6: Snapshot of a typical configuration obtained from a Monte Carlo simulation of intermixing in a system with 3 ML of Ge on Si(001). (a) The surface is viewed from the side along the direction of the dimer rows (along the y axis in Fig. 5.1). (b) The surface is viewed from the other side along the direction of the vacancy lines (along the x axis in Fig. 5.1). Si atoms are shown in blue and Ge atoms are color coded according to their energy from red (high energy, dimers) to light blue (low energy).

Figure 5.6 shows a snapshot of a typical configuration obtained from a Monte Carlo simulation of intermixing in a system with 3 ML of Ge on Si(001). The simulation temperature is 0.001 eV (11 K), which means that in the equilibrium configuration, the Ge atoms will occupy the sites with lowest energy. In the low-temperature limit, it is not necessary to limit the intermixing region to a certain region near the surface because Ge segregates to the surface region when the overall fraction of Ge is sufficiently low (e.g. 3 ML or less while the entire simulation system consists of at least 40 atomic layers). For more details of the simulations, see publication VI.

In Fig. 5.6(a), the surface is viewed from the side along the direction of the dimer rows. The dimerization of the atoms in the surface layer leads to large distortions in the positions of the atoms in the underlying layers. Consequently, there is a strong preference in the third and fourth layers for the Ge atoms to occupy the sites under the troughs (3b and 4b), while the sites under the dimers (3a and 4a) are mainly occupied by Si atoms. This feature was predicted by both the DFT and empirical potential calculations of the binding energy of Ge atoms where the 3b and 4b sites were found to be clearly lower in energy than the 3a and 4a sites (see the previous section). The preference of Ge atoms to occupy

the tensile 3b and 4b sites has also been observed in experiments (e.g. in AED measurements in Ref. [81]). At typical growth temperatures, part of the Ge is driven to the compressively strained 3a and 4a sites due to entropy, but our MC simulations at elevated temperatures show that a majority of the Ge atoms remain in the tensile b sites even at 600°C. This leads to the formation of an ordered structure in subsurface layers as has also been observed in growth experiments of Si/Ge alloys on Si(001) [92].

In Fig. 5.6(b), the surface is viewed from the other side where vacancy lines are visible. The surface is 2×8 reconstructed which is an optimal periodicity for 3 ML coverage in the absence of intermixing. We observe that the presence of vacancy lines has a clear effect on the distribution of Ge atoms in the subsurface layers: the region under a vacancy line is clearly unfavorable for Ge occupation. This is due to the fact that the vacancy lines induce atomic displacements in their vicinity, which in turn leads to a local compressive strain in this region. From these observations, it becomes evident that there exists a complex interplay between the vacancy lines and subsurface distribution of Ge since both of these features contribute to strain relaxation on the Ge/Si(001) surface.

In order to address the question of strain relaxation in more detail, we return to the surface energy calculations which in section 5.1 were conducted in the absence of intermixing. We have seen that the optimal periodicity of the $2 \times n$ reconstruction decreases as the Ge layer becomes thicker due to an increasing need for strain relief in the system. Saturation is reached around 2 ML coverage after which the $2 \times n$ reconstruction becomes less effective in relieving the strain. We now propose a question: how are these results affected if intermixing can occur in the system? For example, it may be energetically more favorable to have less vacancy lines on the surface if part of the strain can be more effectively relieved by subsurface intermixing of Ge.

Figure 5.7 shows the effect of strain-relief driven intermixing on the relative stability of different $2 \times n$ reconstructions. For 2 ML coverage, the optimal value of n shifts from $n = 8$ to larger values when intermixing is allowed to occur in the system. STM measurements [46] show similar behavior: for 1.5 ML of Ge grown at a lower temperature, post-deposition annealing at 760°C changes the observed value of n from 9 to 14. The annealing enables the Ge atoms to overcome at least part of the kinetic barriers affecting their subsurface migration, and consequently, the periodicity of the vacancy-line structure changes as a response to the strain relaxation provided by the intermixing. In contrast, for the 3 ML coverage, intermixing leads to a large energy gain, but the optimal periodicity of the $2 \times n$ reconstruction remains at $n = 8$. The reason for this is that intermixing provides the additional strain relief which cannot be accomplished by further reduction of the vacancy-line periodicity. In experimental systems, the occurrence of intermixing is strongly dependent on the growth temperature. At low to moderate growth temperatures, the system is likely to be close to the limit of full surface segrega-

tion of Ge, whereas at high temperatures, intermixing could potentially change the surface morphology in the 2-3 ML regime. At least to our knowledge, this subject has not been investigated experimentally, but our simulation results suggest that the onset of intermixing could, for example, postpone the formation of the $m \times n$ pattern due to the reduced need for strain relaxation.

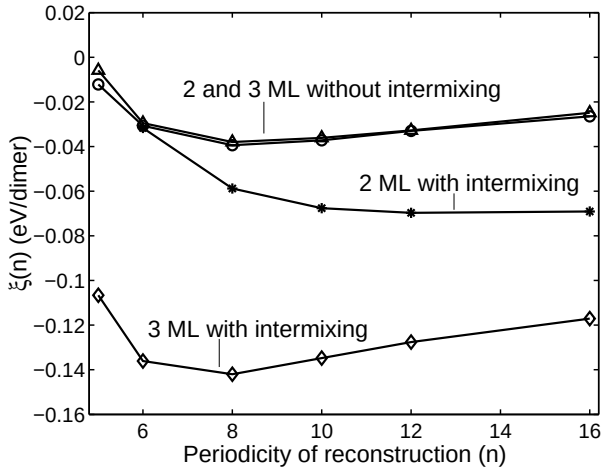

Figure 5.7: Effect of intermixing on the relative stability of different $2 \times n$ reconstructions. Results for 2 and 3 ML coverages are shown. The data for the case without intermixing is the same as shown in Fig. 5.2. Coverages up to 1 ML are not included in the figure because in the low-temperature limit all Ge remains in the top layer and the surface structure does not change.

Chapter 6

A new hybrid MC-MD algorithm

6.1 Background

In this last chapter, we present a new hybrid MC-MD algorithm which has been developed for the study of long time- or length-scale phenomena on semiconductor surfaces. The algorithm is targeted to circumventing the problem of getting trapped into metastable states in simulations of systems with complicated energy landscapes. We are not interested in the exact dynamical evolution of the system, but the aim is to achieve large-scale relaxation and to study the surface morphology at equilibrium or near equilibrium. For such studies, the Metropolis Monte Carlo is the method of choice because it is not constrained by the intrinsic dynamics of the system. The algorithm and simulation results presented here are based on publications VII and VIII.

The problem with conventional MC algorithms is that in simulations involving complicated energy landscapes, the escape from a metastable state would require large-scale rearrangement of the atomic configuration which is associated with very high energy barriers. In the case of the Si(001) surface, the formation of a dimer from two neighboring adatoms lowers their energy by approximately 1 eV/atom. On the contrary, the energy gained when two or more dimers align to form a straight dimer row is at least an order of magnitude smaller. Consequently, standard MC simulations typically produce configurations with an excessive amount of surface defects; e.g. undimerized atoms and antiphase boundaries (see publication VIII).

In order to overcome these problems related to large systems and long time scales, we have developed a new hybrid MC-MD algorithm which is based on introducing collective, large-scale moves to the standard Metropolis-MC method. The version presented in this thesis, the *dimer-jump algorithm*, is intended for the study of Si(001) systems, but the idea itself is quite general and could potentially be applied to other semiconductor surfaces. In the construction of the algorithm,

the identification of collective moves is the only part where knowledge of the particular physical system under study is required. In the case of the Si(001) surface, the 2×1 reconstruction of the surface plays the key role. Interesting phenomena such as the formation or relaxation of islands involve movement of the basic building blocks of the surface, namely diffusion of dimers. Using just the single atom moves of the standard MC algorithm, dimers are never observed to move distances that would be comparable to the lattice constant of silicon. An effective way to overcome the large energy barriers related to the movement of dimers is to consider the dimer as a unit rather than as two individual atoms.

In the following, we first give details of the implementation of the algorithm and then apply it to study the formation and relaxation of 2D Si islands on Si(001) and the formation of dimer-vacancy structures on Si- or Ge-covered Si(001).

6.2 Dimer-jump algorithm

The dimer-jump algorithm is based on the constant-pressure Metropolis Monte Carlo algorithm described in Chapter 2. In addition to the single-particle moves and volume variations, we introduce an additional type of trial move into the algorithm, namely the dimer jump. This move consists of two parts: the displacement of a pair of atoms (a dimer) which is followed by a local Molecular Dynamics relaxation. Fig. 6.1 shows a schematic illustration of a "dimer-jump" trial move.

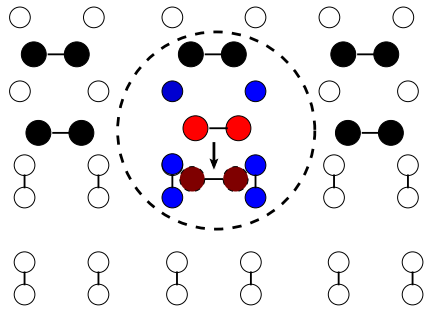

Figure 6.1: Schematic illustration of the "dimer-jump" trial move. The selected dimer (drawn in red) is first randomly displaced within the area indicated by the circle. Alternatively, a donut-shaped area can be used in order to exclude short jumps to the vicinity of the original position of the dimer. An example of a possible position after the jump is shown in the figure (drawn in darker red). Then the configuration is relaxed locally using MD (examples of atoms included in the relaxation are shown in blue). In the end, the relaxed configuration is either accepted or rejected using the Metropolis criterion. In the case of rejection, the algorithm returns to the original configuration before the dimer was displaced.

In the first part of the move, the dimer is displaced as a unit over a random distance which can extend up to 1.5 times the lattice constant of silicon. Typically the range of allowed jump lengths is chosen to coincide with the distance between two neighboring binding sites for the dimer. In order to explain the need for the subsequent MD relaxation, let us consider the arrangement of atoms in the underlying layers. Fig. 6.2 shows a snapshot of a typical step edge on Si(001). We see that the four nearest atoms under a dimer have relaxed to near-tetrahedral positions, whereas those atoms which are not directly under the step are themselves dimerized. If we now displace one of the upper-layer dimers by an amount corresponding to the average distance between binding sites, the dimer lands in a position where its nearest neighbors are either too close or too distant. Moreover, the configuration at the initial site has become unfavorable to those atoms which were previously under the dimer. Consequently, the dimer jump will result in a very high-energy configuration which will almost certainly be rejected in the Metropolis trial.

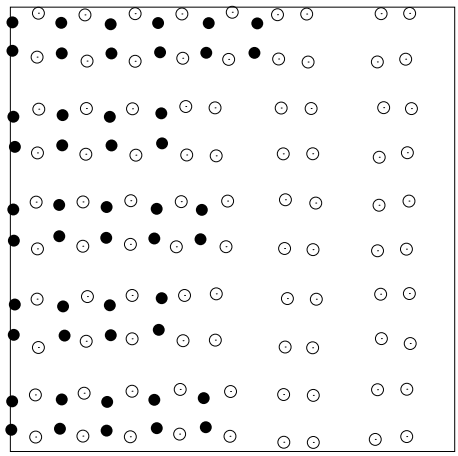

Figure 6.2: Snapshot of a typical S_B step edge on the Si(001) surface after relaxation at 700 K. The open and solid circles are lower and upper terrace atoms, respectively.

In order to achieve a much better acceptance rate, we have introduced a scheme in which the local environments around the initial and the landing site are relaxed prior to the acceptance test. The dimer itself is also included in this group of atoms. The relaxation is performed using standard MD with Velocity Verlet for integrating the equations of motion. Initial velocities are assigned from the Maxwell-Boltzmann distribution at the simulation temperature and constant-temperature conditions are maintained using velocity rescaling. In most cases, a relaxation of about 10 MD steps using a time step of 10^{-15} s is enough for achieving a good acceptance rate. After the MD relaxation, the resulting trial con-

figuration is tested for acceptance using the normal Metropolis criterion, and if it is rejected, the simulation returns to the original configuration before the dimer was displaced.

Further details about the algorithm are reported in publication VIII.

6.3 Results

6.3.1 Formation of Si islands

As the first application example, we consider the early stages of Si island formation on the Si(001) surface. The simulations are initiated by randomly depositing a small number of Si dimers onto the surface, and the system is allowed to evolve using the dimer-jump algorithm. The dimer jumps are performed at intervals of 5 MCS. Figure 6.3 shows a snapshot of a typical configuration obtained at $T = 700$ K after 8.5×10^4 MC steps. We notice that the dimers have moved on the surface to form elongated islands which are closely reminiscent to those observed in experiments (see e.g. [93]).

The experimentally observed shape anisotropy has generated a lot of theoretical interest. The formation of the almost one-dimensional islands is certainly linked to a strong anisotropy in the surface diffusion of adatoms and dimers. Movement in the direction of the underlying dimer rows is much faster than in the perpendicular direction. Experimentally, it is estimated to be about 1000 times faster for Si adatoms on Si(001) [94]. Similar anisotropy is observed for Si dimers, although this subject has not been addressed in such detail due to e.g. difficulties in identifying the possible mechanisms for dimer diffusion. In addition, the situation is influenced by a strong sticking anisotropy at the S_A and S_B step edges. The migrating adatoms or dimers prefer to stick to the ends of an existing dimer string but it is not clear how the atoms get transported to the ends if they first reach one of the long sides of the dimer string. Such events are not directly measurable even with the most advanced experimental techniques, and therefore this question remains partly open.

These experimentally observed features are qualitatively reproduced in our simulations, but we must emphasize that the dimer-jump algorithm is not a kinetic model and thus is not suitable for a quantitative comparison of time-dependent physical properties. Furthermore, the dimer-jump algorithm is not intended to mimic the real mechanism for dimer diffusion which is not even fully understood yet. Nevertheless, the directional anisotropy of dimer diffusion is effectively reproduced in our simulations. The dimers move predominantly in the direction parallel to the substrate dimer rows, and most of the jumps occur in the troughs between the dimer rows. We emphasize that the algorithm does not bias this direction in any way but jumps are tried at all directions with equal probability. Those jumps which are directed along the substrate dimer rows are more easily accepted

because they do not break the dimerization of the atoms in the underlying layers (the four nearest dimer pairs just move closer to the dimer to create a binding site). On the other hand, if we consider jumps in the perpendicular direction, the dimer must jump on top of the underlying dimer row which breaks up the dimer bonds of the four underlying atoms. Of course, infrequent jumps in the perpendicular direction must also occur in the process because otherwise long dimer strings would not form. Consequently, our simulations reproduce the essential features of anisotropic island formation on Si(001).

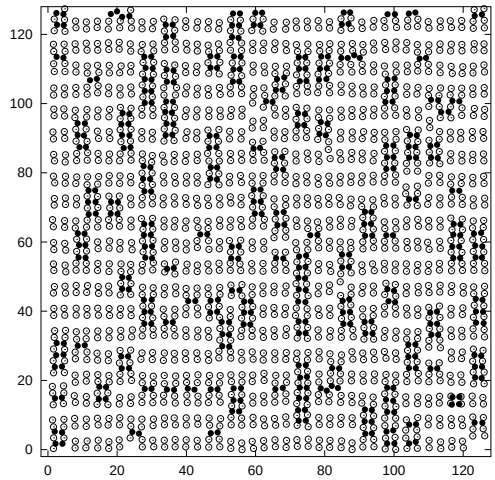

Figure 6.3: Snapshot of a Monte Carlo simulation showing the formation of elongated Si islands on Si(001). The configuration is shown after 8.8×10^4 MC steps at $T = 930$ K. The solid circles are adatoms and the open circles are atoms in the layer beneath. The 205 dimers were initially deposited at random positions. The surface size is 40×40 atoms and the entire simulation system is composed of 21610 atoms.

6.3.2 Relaxation of Si islands and step edges

The second problem addresses the stability and relaxation of preformed islands. In this case, the simulations are started with a square-shaped Si island on the Si(001) surface, which enables the simultaneous analysis of the S_A and S_B steps. The system is allowed to evolve using the dimer-jump algorithm until equilibrium is reached, and no further qualitative changes in the configuration are then observed.

Figure 6.4 shows typical snapshots of an island at two different temperatures. Looking first at the island on the left, we notice that the shape of the island has changed to a slightly anisotropic form. The S_A and S_B steps have evolved in a very different manner: the S_A steps are longer and have a smooth appearance, whereas

the S_B steps are clearly more rough and irregular. Furthermore, we observe that single and double vacancies have developed inside the island. The double vacancies form when two migrating single vacancies meet by chance, and the rebonding of the underlying substrate atoms stabilizes the double vacancies. This behavior is consistent with STM images of Si islands and step edges which all show very similar features as observed in our simulations (see e.g. [93, 55]).

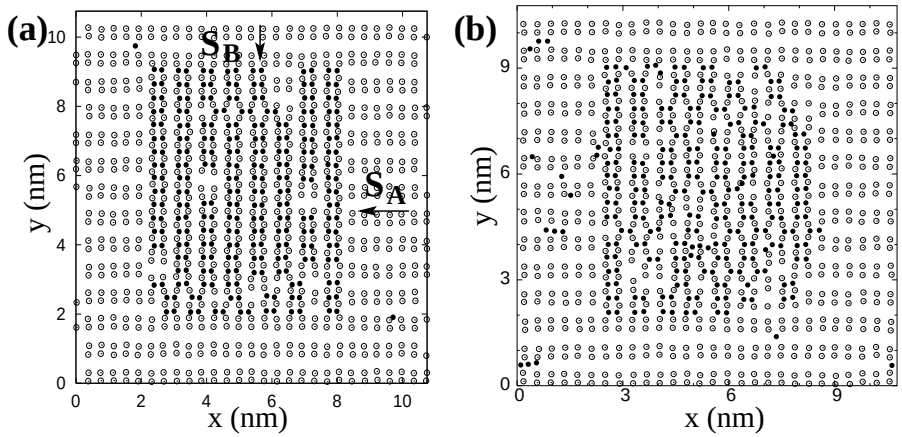

Figure 6.4: Snapshots of typical Si islands on Si(001). (a) $T = 700$ K, after 7×10^4 MCS. (b) $T = 930$ K, after 4×10^3 MCS. The solid circles are adatoms and the open circles are atoms in the first surface layer.

Comparison of the two figures at 700 K and 930 K shows that island stability is clearly dependent on temperature. At 700 K, the island changes its shape during equilibration but stays compact and well defined for very long simulation times. In contrast, when $T = 930$ K, the island dissociates very quickly. If we reduce the initial island size, we observe that dissociation takes place at much lower temperatures (see publication VIII). These observations indicate that there exists a temperature-dependent critical size beyond which islands are stable against dissociation and smaller islands will decay. It is not the aim of this work to address the subject in detail, but from these results it is evident that the algorithm is suitable for such an investigation.

6.3.3 Formation of Si and Ge vacancy structures

The investigation of the vacancy-line structures was initiated by the surface energy calculations described in Chapter 5. The results show that the $2 \times n$ pattern is stabilized by the Ge layer and increasing Ge coverage leads to a decrease in the optimal spacing between the vacancy lines. Our calculations are in excellent agreement with experimental observations, which confirms that our classical

model seems to capture the essential features of dimer vacancies in Si(001) surfaces. As the next step, we asked the question whether the ordering process can be simulated using the dimer-jump algorithm.

Figure 6.5 shows two snapshots from simulations using the dimer-jump algorithm for a pure Si(001) surface and for a surface covered with 2 ML of Ge. In both cases, we used the same initial configuration which consists of a random distribution of dimer vacancies. All dimers are allowed to move on the surface during the simulation, which in effect means that the vacancies migrate. We notice that on the clean Si(001) surface, the vacancies show a tendency to cluster together but the overall arrangement is random. Small segments of vertically lined-up dimers are sometimes observed during the simulations, but these structures dissociate in a relatively short time (within thousands of MC steps). In contrast, when the Si(001) surface is covered by a thin Ge layer, we observe that vacancy lines are clearly forming in the topmost Ge layer (in Fig. 6.5(b), one line passes through the whole system and two shorter segments have also developed). Once formed, the vacancy lines are very stable and do not dissociate.

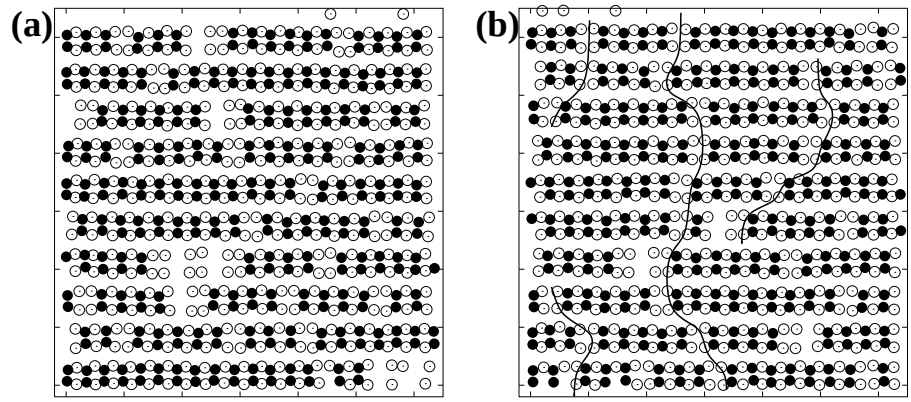

Figure 6.5: Formation of dimer vacancy structures at $T = 930$ K. (a) Clean Si(001), after 2×10^5 MCS and 250 accepted dimer jumps. (b) 2 ML of Ge on Si(001), after 5×10^5 MCS and 220 accepted dimer jumps. In both cases, the initial configuration was a random distribution of vacancies.

The drawback in these particular simulations is that we were forced to use a rather high temperature (930 K) in order to have sufficiently many dimer jumps accepted within a reasonable simulation time. However, the equilibrium shape of the vacancy lines is expected to be quite kinked and disordered at this temperature, which prevents the formation of a well-defined line structure. Therefore, an approach such as parallel tempering (see Chapter 2) could be useful in speeding up the equilibration at lower temperatures.

Nevertheless, these results are promising and suggest that the dimer-jump al-

gorithm gives the possibility to perform atomic-scale simulations of large systems such that experimentally accessible length scales are within reach. The algorithm is a useful tool for studying problems which involve large-scale restructuring of the system, and in particular, it is not necessary to know in advance what the equilibrium structure is, but the algorithm searches for the correct structure.

Chapter 7

Conclusions

The growth of thin semiconductor films is a crucial process in manufacturing various electronic and optoelectronic devices. The characteristic length scale of the components has decreased to the point where atomic-scale characteristics of the surface have become important. Consequently, detailed understanding of the underlying physical phenomena related to atomistic processes during thin film growth is required for the development of new solutions for future technological applications. Due to the advances made both in the experimental and the computational area of research, we are starting to approach a situation where the length- and time-scale gap between these two approaches has diminished and even some overlap can be achieved.

In this thesis, we have focused on the study of thin Si/Ge films on the Si(001) surface. The investigation has been carried out using large-scale Monte Carlo simulations in connection with classical interatomic potentials. This approach has allowed us to study large systems composed of tens of thousands of atoms, while preserving an off-lattice, atomic-scale description of the system. The objective in this work was to develop new, efficient simulation techniques for problems involving large-scale relaxation of systems with complicated energy landscapes. The Monte Carlo method is an ideal choice for such purposes because it is not bound to the real dynamics of the system but new algorithms can be designed to speed up the equilibration process. The Si(001) surface was chosen as our model system because it exhibits a surprisingly wide variety of interesting phenomena related to structural changes during growth.

We began the investigation by evaluating the suitability of different empirical potentials for use in finite-temperature simulations involving the Si(001) surface. The time spent for this study was well worthwhile because we found severe shortcomings in the performance of some widely-used potential models. None of the empirical potentials can give a completely accurate description of the Si(001) surface, but the Stillinger-Weber model was found to give the best overall perfor-

mance with no serious limitations. Based on these results, the SW potential was used for the majority of the simulations comprising this thesis.

The structure of the Si(001) surface is dominated by the formation of dimers, which leads to the well-known 2×1 reconstruction. Several important features of Si(001), such as the anisotropy of surface diffusion, existence of different types of step edges and strain characteristics, are related to the dimerization. In simulations of Si(001), the formation of dimers is predicted to be energetically highly favorable by almost all potential models, including *ab initio* and tight-binding calculations. Nevertheless, even the reproduction of the straight dimer rows in the 2×1 reconstruction is a surprisingly difficult task if the dimerization is not preimposed in the initial configuration. This is because on a bulk-terminated surface, the dimers initially form at random positions (i.e. each surface atom bonds equally likely with its left or right neighbor) and breaking up the dimer bond requires an energy of about 2 eV. Consequently, simulations typically produce a very disordered surface structure with an unphysically high amount of defects such as isolated atoms and antiphase boundaries. This situation is a good example of a deep metastable state which traps the system when only single-particle moves are used in the simulation algorithm (regardless whether we use standard Monte Carlo or Molecular Dynamics).

In order to tackle these problems, we began to work on developing more efficient algorithms for overcoming the potential barriers related to the structural rearrangement of the surface. For example, the 2×1 reconstruction can be reproduced in a Monte Carlo simulation by using a simple modification of the standard algorithm. In addition to the single-particle moves, we used coupled moves in which two neighboring atoms are moved simultaneously either closer together or further apart in the direction of the dimer bonds. This simple recipe allows the system to reorganize into a straight dimer-row structure without having to reduce the number of existing dimer bonds. A similar idea was used in the construction of the more complicated hybrid MC-MD scheme. In this case, the large dimer displacements allow the system to evolve toward the equilibrium structure because the algorithm enables the dimers to jump as a unit, effectively from one binding site to a neighboring one. The speed-up achieved by the dimer-jump algorithm is tremendous because the dimers are practically immobile in standard MC or MD simulations.

Apart from the work done on algorithmic development, this thesis also provides an overview and new information concerning the physical properties of the Si/Ge on Si(001) system. We have discussed the central features of the Si(001) surface and emphasized the role of computational studies in understanding and explaining experimental observations. The Ge/Si(001) has served as an ideal model system for studying the effects of lattice-mismatch induced strain in heteroepitaxial systems. We have illustrated how computational studies have played an important role in identifying strain relaxation as the main driving force for the ex-

perimentally observed structural changes which occur during the growth of Ge on Si(001). Our simulation results give a strong indication that much of the behavior is related to large-scale elastic effects because our classical interaction model produces consistent agreement with the experimental observations. We have also investigated some questions, such as the role of Si-Ge intermixing, that cannot be directly addressed using experimental techniques. This serves as a good example of how computational techniques can give extremely valuable information in situations which are difficult or even impossible to study experimentally.

In the future, the role of computational investigations will certainly continue to grow, not only as a result of an increase in the computing power, but also due to the work currently being done on algorithmic development. We are advancing toward a situation where simulations can reach conditions comparable to those in typical experiments. This will make computational methods an extremely valuable tool which, for example, can be used to perform a large number of computer 'experiments' with little cost compared to real experiments. Nevertheless, we conclude that the key to successful studies still remains in the ability to combine experimental and theoretical work closely together with the purpose of forming a profound understanding of the phenomena under investigation.

References

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[2] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge U. Press, Cambridge, 2000).

[3] D. Frenkel and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications.* (Academic Press, 1996).

[4] B. Dünweg and D. P. Landau, Phys. Rev. B **48**, 14182 (1993).

[5] C. Geyer and E. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).

[6] E. Orlandini, *Numerical Methods for Polymeric Systems* (Springer-Verlag, Berlin, 1998), p. 33.

[7] K. Hukushima, H. Takayama, and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[8] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).

[9] A. Lyubartsev, A. Martinovski, S. Shevkunov, and P. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).

[10] A. Bunker and B. Dünweg, Phys. Rev. E **63**, 016701 (2001).

[11] S. Duane, A. Kennedy, B. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[12] M. Tuckerman, B. Berne, and G. Martyna, J. Chem. Phys. **97**, 1990 (1992).

[13] B. Forrest and U. Suter, Mol. Phys. **82**, 393 (1994).

[14] M. Kotrla, Comp. Phys. Comm. **97**, 82 (1996).

[15] A. Levi and M. Kotrla, J. Cond. Matt. **9**, 299 (1997).

[16] A. B. Bortz, M. H. Kalos, and L. J. Lebowitz, J. Comput. Phys. **17**, 10 (1975).

[17] A. F. Voter, F. Montalenti, and T. C. Germann, Annu. Rev. Mat. Res. **32**, 321 (2002).

[18] G. Henkelman and H. Jónsson, J. Chem. Phys. **115**, 9657 (2001).

[19] A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).

[20] A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).

[21] A. F. Voter, Phys. Rev. B **57**, 13985 (1998).

[22] M. R. Sørensen and A. F. Voter, J. Chem. Phys. **112**, 9599 (2000).

[23] F. H. Stillinger and T. A. Weber, Phys. Rev. B **31**, 5262 (1985).

[24] J. Tersoff, Phys. Rev. B **37**, 6991 (1988).

[25] J. Tersoff, Phys. Rev. B **38**, 9902 (1988).

[26] P. N. Keating, Phys. Rev. **145**, 637 (1966).

[27] R. Biswas and D. R. Hamann, Phys. Rev. B **36**, 6434 (1987).

[28] B. W. Dodson, Phys. Rev. B **35**, 2795 (1987).

[29] K. E. Khor and S. D. Sarma, Phys. Rev. B **38**, 3318 (1988).

[30] J. Wang and A. Rockett, Phys. Rev. B **43**, 12571 (1991).

[31] K. Ding and H. C. Andersen, Phys. Rev. B **34**, 6987 (1986).

[32] M. Laradji, D. P. Landau, and B. Dünweg, Phys. Rev. B **51**, 4894 (1995).

[33] J. Tersoff, Phys. Rev. B **39**, 5566 (1988).

[34] H. Balamane, T. Halicioglu, and W. A. Tiller, Phys. Rev. B **46**, 2250 (1992).

[35] O. Madelung, editor, *Semiconductors, Group IV Elements and III-V Compounds* (Springer-Verlag, 1991).

[36] J. D. Weeks and G. H. Gilmer, Adv. Chem. Phys. **40**, 157 (1979).

[37] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1992).

[38] T. Shitara, D. D. Vvedensky, M. R. Wilby, J. Zhang, J. H. Neave, and B. A. Joyce, Phys. Rev. B **46**, 6815 (1992).

[39] M. Bartelt, A. Schmidt, J. Evans, and R. Hwang, Phys. Rev. Lett. **81**, 1901 (1998).

[40] H. Brune, M. Giovannini, K. Bromann, and K. Kern, Nature **394**, 451 (1998).

[41] H. Brune, Surf. Sci. Rep. **31**, 121 (1998).

[42] B. Voigtländer, Surf. Sci. Rep. **43**, 127 (2001).

[43] N. Roberts and R. J. Needs, Surf. Sci. **236**, 112 (1990).

[44] I. P. Batra, Phys. Rev. B **41**, 5048 (1990).

[45] H. J. W. Zandvliet, Rev. Mod. Phys. **72**, 593 (2000).

[46] F. Liu, F. Wu, and M. G. Lagally, Chem. Rev. **97**, 1045 (1997).

[47] B. S. Swartzentruber, Y.-W. Mo, R. Kariotis, M. G. Lagally, and M. B. Webb, Phys. Rev. Lett. **65**, 1913 (1990).

[48] N. C. Bartelt and R. M. Tromp, Phys. Rev. B **54**, 11731 (1996).

[49] H. J. W. Zandvliet, H. B. Elswijk, E. J. van Loenen, and D. Dijkkamp, Phys. Rev. B **45**, 5965 (1992).

[50] D. J. Chadi, Phys. Rev. Lett. **59**, 1691 (1987).

[51] T. Z. Poon, S. Yip, P. S. Ho, and F. F. Abraham, Phys. Rev. Lett. **65**, 2161 (1990).

[52] T. Z. Poon, S. Yip, P. S. Ho, and F. F. Abraham, Phys. Rev. B **45**, 3521 (1992).

[53] D. R. Bowler and C. M. Goringe, Phys. Rev. B **58**, 3937 (1998).

[54] A. Oshiyama, Phys. Rev. Lett. **74**, 130 (1995).

[55] N. C. Bartelt, R. M. Tromp, and E. D. Williams, Phys. Rev. Lett. **73**, 1656 (1994).

[56] R. J. Hamers, U. K. Köhler, and J. E. Demuth, J. Vacc. Sci. Technol. A **8**, 195 (1990).

[57] P. Bedrossian and E. Kaxiras, Phys. Rev. Lett. **70**, 2589 (1993).

[58] H. J. W. Zandvliet, E. Zoethout, W. Wulfhekel, and B. Poelsema, Surf. Sci. **482-485**, 391 (2001).

[59] E. Zoethout, P. W. van den Hoogenhof, H. J. W. Zandvliet, and B. Poelsema, J. Appl. Phys. **92**, 5785 (2003).

[60] R. M. Tromp, R. J. Hamers, and J. E. Demuth, Phys. Rev. Lett. **55**, 1303 (1985).

[61] O. L. Alerhand, N. A. Berker, . J. J. D, D. Vanderbilt, R. J. Hamers, and J. E. Demuth, Phys. Rev. Lett. **64**, 2406 (1990).

[62] J. Wang, T. A. Arias, and J. D. Joannopoulos, Phys. Rev. B **47**, 10497 (1993).

[63] F.-K. Men, A. R. Smith, K.-J. Chao, Z. Zhang, and C.-K. Shih, Phys. Rev. B **52**, 8650 (1995).

[64] J. Oviedo, D. R. Bowler, and M. J. Gillan, Surf. Sci. **515**, 483 (2002).

[65] P. Bedrossian and T. Klitsner, Phys. Rev. Lett. **68**, 646 (1992).

[66] H. J. W. Zandvliet, H. K. Louwsma, P. E. Hegeman, and B. Poelsema, Phys. Rev. Lett. **75**, 3890 (1995).

[67] H. J. W. Zandvliet, Surf. Sci. **377-379**, 1 (1997).

[68] J.-Y. Koo, J.-Y. Yi, C. Hwang, D.-H. Kim, S. Lee, and D.-H. Shin, Phys. Rev. B **52**, 17269 (1995).

[69] J.-Y. Koo, J.-Y. Yi, C. Hwang, D.-H. Kim, S. Lee, and J. Cho, Phys. Rev. B **54**, 10308 (1996).

[70] M.-H. Tsai, Y.-S. Tsai, C. S. Chang, Y. Wei, and I. S. T. Tsong, Phys. Rev. B **56**, 7435 (1997).

[71] N. Kitamura, M. G. Lagally, and M. B. Webb, Phys. Rev. Lett. **71**, 2082 (1993).

[72] H. Watanabe and M. Ichikawa, Phys. Rev. B **55**, 9699 (1997).

[73] E. Kim, C. Chen, T. Pang, and Y. H. Lee, Phys. Rev. B **60**, 8680 (1999).

[74] H. Niehus, U. K. Köhler, K. Copel, and J. E. Demuth, J. Micros. **152**, 735 (1988).

[75] B. D. Yu and A. Oshiyama, Phys. Rev. B **52**, 8337 (1995).

[76] F. Liu and M. G. Lagally, Phys. Rev. Lett. **76**, 3156 (1996).

[77] X. Chen, F. Wu, and M. G. Lagally, Phys. Rev. Lett. **73**, 850 (1994).

[78] F. Wu, X. Chen, , Z. Zhang, and M. G. Lagally, Phys. Rev. Lett. **74**, 574 (1995).

[79] B. Voigtländer and M. Kästner, Phys. Rev. B **60**, R5121 (1999).

[80] K. Li, D. R. Bowler, and M. J. Gillan, Surf. Sci. **526**, 356 (2003).

[81] B. P. Uberuaga, M. Leskovar, A. P. Smith, H. Jonsson, and M. Olmstead, Phys. Rev. Lett. **84**, 2441 (2000).

[82] X. R. Qin, B. S. Swartzentruber, and M. G. Lagally, Phys. Rev. Lett. **84**, 4645 (2000).

[83] Z.-Y. Lu, C.-Z. Wang, X. R. Qin, B. S. Swartzentruber, M. G. Laglly, and K.-M. Ho, Phys. Rev. Lett. **85**, 5603 (2000).

[84] R. M. Tromp, Phys. Rev. B **47**, 7125 (1993).

[85] L. Patthey, E. L. Bullock, T. Abukawa, S. Kono, and L. S. O. Johansson, Phys. Rev. Lett. **75**, 2538 (1995).

[86] J.-H. Cho and M.-H. Kang, Phys. Rev. B **61**, 1688 (2000).

[87] Y.-J. Ko, K.-H. Park, J. S. Ha, and W. S. Yun, Phys. Rev. B **60**, 8158 (1999).

[88] M. Sasaki, T. Abukawa, H. W. Yeom, M. Yamada, S. Suzuki, S. Sato, and S. Kono, Appl. Surf. Sci. **82-83**, 387 (1994).

[89] A. Ikeda, K. Sumitomo, T. Nishioka, T. Yasue, T. Koshikawa, and Y. Kido, Surf. Sci. **385**, 200 (1997).

[90] H. W. Yeom *et al.*, Surf. Sci. **381**, L533 (1997).

[91] K. Nakajima, A. Konishi, and K. Kimura, Nucl. Instr. and Meth. in Phys. Res. B **161-163**, 452 (2000).

[92] F. K. LeGoues, V. P. Kesan, S. S. Iyer, J. Tersoff, and R. Tromp, Phys.Rev. Lett. **64**, 2038 (1990).

[93] Y. W. Mo, B. S. Swartzentruber, R. Kariotis, M. B. Webb, and M. G. Lagally, Phys. Rev. Lett. **63**, 2393 (1989).

[94] Z. Zhang, F. Wu, and M. G. Lagally, Annu. Rev. Mater. Sci. **27**, 525 (1997).