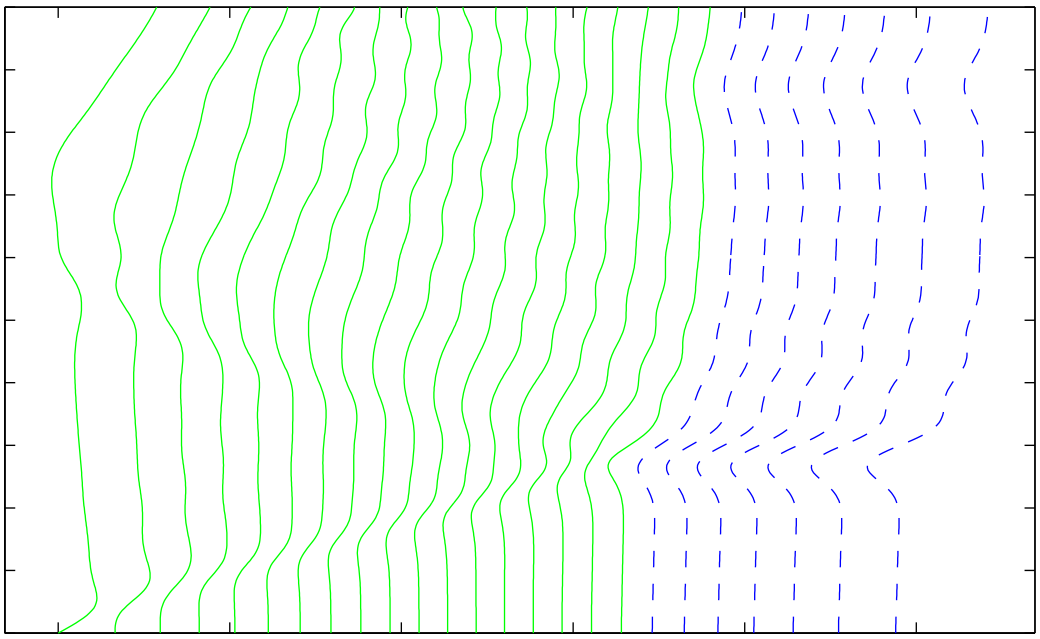
UNIVERSITY OF KAISERSLAUTERN

DEPARTMENT OF MATHEMATICS

Optimal Prediction in Molecular Dynamics



DIPLOMA THESIS

BENJAMIN SEIBOLD

Kaiserslautern, April 2003

Optimal Prediction in Molecular Dynamics

DIPLOMA THESIS

BENJAMIN SEIBOLD

Supervisors

Prof. Dr. Helmut Neunzert
Dr. Thomas Götz
Dr. Peter Klein

Department of Mathematics
University of Kaiserslautern
Gottlieb-Daimler-Straße
P.O. Box 3049, D-67653 Kaiserslautern
Germany

April 2003

Abstract

Optimal prediction is a method to approximate the average solution of a large system of ordinary differential equations by a smaller system. In this thesis we show how optimal prediction can be applied in the field of molecular dynamics in order to reduce the number of particles. We apply optimal prediction to a model problem describing a surface coating process and show how asymptotic methods can be used to approximate the original system by a smaller system. We consider the reduction of computational effort and analyze – analytically and by numerical experiments – under which conditions the optimal prediction system is a valid approximation to the original system.

Table of Contents

Table of Contents **1**

Introduction **3**

1 Optimal Prediction **5**

1.1 Mathematical and Physical Assumptions 5

1.2 Splitting the Variables 6

1.3 Conditional Expectation and Mean Solution 7

1.4 First Order Optimal Prediction 8

1.5 Mori-Zwanzig 9

2 Description of the Problem **13**

2.1 The Real Problem 14

2.2 The Model Problem 16

2.3 Physical Properties of the Model Problem 19

2.4 Numerical Simulation 20

3 Optimal Prediction Applied to the Model Problem **25**

3.1 Mathematical and Physical Setup 26

3.1.1 Appropriate Domains of Integration 26

3.1.2 The Optimal Prediction Hamiltonian 27

3.2 Low Temperature Asymptotics 28

3.2.1 Asymptotic Expansion of the Hamiltonian 29

3.2.2 An Example of Three Atoms 31

3.2.3 Properties of the Asymptotic Expansion 33

3.3 Zero Temperature Limit 36

3.3.1 Physical Interpretation 37
3.3.2 Equations of Motion 38
3.3.3 Equations of Motion in the Model Problem 40
3.3.4 A Boundary Layer Condition 41
3.4 Numerical Simulation 44
3.4.1 Examples 44
3.4.2 Computational Speed-Up 49

4 A Comparison by Numerical Experiments 53
4.1 Comparing Two Systems 54
4.1.1 Why Use Monte-Carlo Sampling? 55
4.1.2 Monte-Carlo Sampling in the Model Problem 57
4.1.3 Criteria of Comparison 57
4.2 A Model for the Copper Diffusion 59
4.2.1 Modeling Hopping as a Random Walk 59
4.2.2 Using the Model to Obtain Diffusion Parameters 63
4.3 An Experiment Containing Sonic Waves 65
4.3.1 Numerical Results 65
4.3.2 Sonic Waves 67
4.3.3 The Velocity of Sound 68
4.3.4 Sonic Waves in the Optimal Prediction System 69
4.4 An Experiment Without Sonic Waves 72
4.4.1 Estimating the Monte-Carlo Error 74
4.4.2 The Error in Distribution 76
4.4.3 Diffusion Parameters 77
4.4.4 The Number of Hopping Events 81
4.4.5 Energy Fluctuations 81

5 Conclusions and Outlook 85

Acknowledgements 87

List of Figures 88

Bibliography 90

Introduction

Computations in the field of molecular dynamics typically require a large computational effort due to two factors:

1. **Small time steps are required** in the numerical solver to resolve the fast atomic oscillations. Methods to remedy this problem are e.g. smoothing algorithms as in [38], which allow to take larger time steps by smoothing out the large oscillations and following the large scale evolution only.

2. **Large systems** due to the large amount of atoms which have to be computed. Methods to reduce the number of degrees of freedom are e.g. shrinking the support of pair potentials, or multipole methods [2, 20] in the context of long range particle interactions.

The method of optimal prediction as presented and analyzed in [3, 6, 7, 8, 9, 10, 11, 12, 13, 21, 23, 27, 35, 37] applies to "underresolved computation", in particular systems of ordinary differential equations, where only part of the initial data is known while the rest of the initial data is only known to be distributed according to some probability measure. Sought is the mean solution of the original system with respect to the unknown initial data being sampled from the underlying probability distribution. Optimal prediction yields a new system of ordinary differential equations, which is an approximation to the mean solution, but is smaller, and thus cheaper to compute, than the original system. In Chapter 1 we will derive the optimal prediction equations and their important properties.

In this thesis we show how the method of optimal prediction can be applied to problems in the field of molecular dynamics in order to reduce the number of degrees of freedom. So from the above two paths to reduce the computational effort, we walk the second one. (An example how to take the first path is [14].) In principle, the idea is to divide the particles into two groups: On the one hand the particles which are close to the region of interest and whose initial data is known, and on the other hand the particles outside the region of interest, which should be considered only in an averaged sense.

We present a one-dimensional model for a problem which arises in surface coating processes of copper onto a silicon crystal. The problem appears in a DFG project in the

Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM) in Kaiserslautern. Goal of this project is to use a molecular dynamics simulation to investigate which effects could cause bad material properties of the product. Of particular interest is the "hopping" of single copper atoms inside the silicon crystal. The complete coating process takes such a long time compared to the atomic time scales, that methods to reduce the computational effort are of great importance. In Chapter 2 we will describe the physical problem in detail and introduce the one-dimensional model problem. Information about the numerical implementation and computational properties are provided.

In Chapter 3 we apply the method of optimal prediction to the one dimensional model problem. This yields expressions involving high-dimensional integrals which in general cannot be evaluated explicitly. We thus derive an asymptotic approximation to the optimal prediction system, where we employ the fact that the characteristic temperature is low on a scale given by the bond energy between two atoms. This allows to evaluate the high dimensional integrals approximately and to obtain a smaller system which is an approximation to the optimal prediction system. Example computations show which aspects are important for the numerical implementation, and the aspect of numerical speed-up is considered.

In order to investigate if the so obtained smaller system is a valid approximation to the original system, criteria have to be defined how to compare the two systems. Since in molecular dynamics precise trajectories have no importance (unless for very short times), statistical quantities, such as diffusion constants and energy fluctuations, have to be considered. In Chapter 4 we compare the optimal prediction approximation to the original system by performing various numerical experiments in order to investigate whether and under which conditions important statistical quantities are preserved.

Chapter 1

Optimal Prediction

Optimal prediction was introduced in 1998 by Chorin, Kast and Kupferman [9] as a method for approximating the average solution of problems which are computationally too expensive or where not enough data is at hand, but prior statistical information is available. By the use of the Mori-Zwanzig formalism [33, 42] several ways have been derived how to approximate the average solution by a new system which is smaller than the original one.

In this thesis we are less interested in approximating the average solution, but rather in using the method of optimal prediction as a way to replace a large Hamiltonian system by a smaller one, which preserves the important properties of the original system. In particular the smaller system should be Hamiltonian again. In terms of molecular dynamics this means considering only a smaller number of particles and "averaging" the other ones away.

In the following we present the most important steps in deriving the optimal prediction equations and their connection with the Mori-Zwanzig formalism.

1.1 Mathematical and Physical Assumptions

Consider an n-dimensional system of ordinary differential equations

$$\begin{aligned} \frac{d}{dt}\varphi(x,t) &= R(\varphi(x,t)), \\ \varphi(x,0) &= x. \end{aligned} \tag{1.1}$$

Here $\varphi(x,t)$ denotes the solution (for later analysis interpreted as a phase flow) to the initial condition $x \in \mathbb{R}^n$. The right hand side R is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$, where the system's size n is typically a large number. Of particular interest in molecular dynamics

are Hamiltonian systems, i.e. $2n$-dimensional systems of the form

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q}, \end{aligned} \tag{1.2}$$

where $q = (q_1, \ldots, q_n)$ is the position vector, $p = (p_1, \ldots, p_n)$ is the momentum vector, and $H(p, q) = \frac{1}{2}p^2 + V(q)$ is the Hamiltonian function. Additionally, we consider a measure on the phase space $\mathbb{R}^{2n}$. In Hamiltonian molecular dynamics the canonical equilibrium measure

$$f(x) = Z^{-1} e^{-\beta H(x)} \tag{1.3}$$

is in many cases the appropriate choice. Here $x = (q_1, p_1, \ldots, q_n, p_n)$ is the position in state space, and $\beta = \frac{1}{k_B T}$ is a constant, where k_B is the Boltzmann constant and T is the temperature of the process. $Z = \int e^{-\beta H(x)} \, \mathrm{d}x$ is a normalization constant. The question, how to guarantee that Z is finite, will be treated in Subsection 3.1.1. In this thesis we will use only the above measure (1.3). Note that in the following the size of the systems may be denoted inconsistently. Typically, the size of a general system is n. However, if for a Hamiltonian system the vectors q and p are being considered, the system is assumed to be of size $2n$.

It has to be pointed out that the Hamiltonian system (1.2) is assumed to be in thermodynamical equilibrium, i.e. in particular that the temperature T is constant with respect to space and time. In Chapter 4 one important aspect for the applicability of optimal prediction will turn out to be the relevance of non-equilibrium effects.

1.2 Splitting the Variables

Assume now a separation of all degrees of freedom into two groups $\varphi = (\hat{\varphi}, \tilde{\varphi})$, where $\hat{\varphi} = (\varphi_1, \ldots, \varphi_m)$ represents the variables of interest, and $\tilde{\varphi} = (\varphi_{m+1}, \ldots, \varphi_n)$ are the variables which should be "averaged out". Typically, m is significantly smaller than n.

Now only part of the initial conditions $\hat{x} = (x_1, \ldots, x_m)$, namely the ones corresponding to the variables which are of interest $\hat{\varphi}$, are known, while the other components $\tilde{x} = (x_{m+1}, \ldots, x_n)$ are not known exactly. Instead, for each choice of $\hat{x}$ they are sampled from the conditioned measure

$$f_{\hat{x}}(\tilde{x}) = Z_{\hat{x}}^{-1} e^{-\beta H(\hat{x}, \tilde{x})}, \tag{1.4}$$

where $Z_{\hat{x}} = \int e^{-\beta H(\hat{x}, \tilde{x})} \, \mathrm{d}\tilde{x}$ is the appropriate normalization constant. Note that for Hamiltonian systems each pair (q_i, p_i) must belong either to the group $\hat{\varphi}$ or to $\tilde{\varphi}$.

1.3 Conditional Expectation and Mean Solution

As in [7] we define the conditional expectation of a function $f(\hat{x}, \tilde{x})$ as

$$\mathbb{E}[f|\hat{x}] = \frac{\int f(\hat{x}, \tilde{x}) e^{-\beta H(\hat{x}, \tilde{x})} d\tilde{x}}{\int e^{-\beta H(\hat{x}, \tilde{x})} d\tilde{x}}. \tag{1.5}$$

It is the orthogonal projection onto the space of functions $g(\hat{x})$ with respect to the inner product

$$(u, v) = \mathbb{E}[uv] = \frac{\int u(x) v(x) e^{-\beta H(x)} dx}{\int e^{-\beta H(x)} dx}, \tag{1.6}$$

so we denote

$$Pf = \mathbb{E}[f|\hat{x}]. \tag{1.7}$$

Note that in molecular dynamics the integrals in formula (1.5) are typically not defined over the whole space $\mathbb{R}^{n-m}$. The integrals are infinite, if the existence of (almost) free particles is possible, i.e. single particles far away from and not influenced by the other ones. We will deal with this issue in Subsection 3.1.1.

We are now interested in the mean solution of (1.1), where the initial conditions $\hat{x}$ are fixed and $\tilde{x}$ are sampled from (1.4). This mean solution is given by

$$P\varphi(x, t) = \mathbb{E}[\varphi(x, t)|\hat{x}] = \frac{\int \varphi((\hat{x}, \tilde{x}), t) e^{-\beta H(\hat{x}, \tilde{x})} d\tilde{x}}{\int e^{-\beta H(\hat{x}, \tilde{x})} d\tilde{x}}. \tag{1.8}$$

More precisely, we are only interested in the first m components of the mean solution $P\hat{\varphi}(x, t)$.

For the special case of linear Hamiltonian systems the mean solution (1.8) can be calculated explicitly by interchanging the linear solution operator e^{At} with the integral:

$$P\varphi(x, t) = \varphi(Px, t) = e^{At} \cdot \begin{pmatrix} \hat{x} \\ P\tilde{x} \end{pmatrix}. \tag{1.9}$$

This means in particular that for linear Hamiltonian systems the mean solution is Hamiltonian (with initial conditions which are the average over all possible initial conditions) and does hence not decay.

In molecular dynamics, however, the potentials between two particles are typically vanishing at infinity, hence the Hamiltonian system is nonlinear. In this case the conditional expectation P, involving the $(n - m)$-dimensional integral, can be approximated by Monte-Carlo sampling. The mean solution can be computed as follows:

- Fix $\hat{x} = (x_1, \ldots, x_m)$
- Sample $\tilde{x} = (x_{m+1}, \ldots, x_n)$ N times from the conditioned distribution given by (1.4)
- Solve N times (1.1) with initial values $(\hat{x}, \tilde{x})$
- Average over all solutions (arithmetic mean)

Obviously this is extremely expensive, since the original n-dimensional system has to be solved N times.

An important fact about nonlinear systems is that the mean solution decays. This is the case, even if the system itself is Hamiltonian and thus energy preserving. In terms of irreversible statistical mechanics this decay can be explained as a loss of information as the first m variables tend to the thermodynamical equilibrium, which is represented by the canonical measure (1.3). In [3] the authors give a deeper physical reasoning for the decay. A formal explanation for the decay of the mean solution yields the application of the Mori-Zwanzig formalism in Section 1.5.

1.4 First Order Optimal Prediction

The idea of first order optimal prediction is fairly simple: We want to approximate the first m components of the mean solution $P\hat{\varphi}(x,t)$ by an m-dimensional system. The right hand side R of the n-dimensional system is a function of n variables. Applying the conditional expectation to R

$$\mathfrak{R} = PR = \mathbb{E}[R|\hat{x}] \tag{1.10}$$

yields a function $\mathfrak{R}$ of just m variables. Then $\hat{\mathfrak{R}} = (\mathfrak{R}_1, \ldots, \mathfrak{R}_m)$ is a function from $\mathbb{R}^m$ to $\mathbb{R}^m$. So we define the optimal prediction system to be

$$\begin{aligned} \dot{y}(t) &= \hat{\mathfrak{R}}(y(t)), \\ y(0) &= \hat{x}, \end{aligned} \tag{1.11}$$

i.e. the m-dimensional optimal prediction system is obtained by applying the conditional expectation projection P to the right hand side R. An important result is the following

Theorem 1.1 (O.Hald [12])
If a system is Hamiltonian, then its optimal prediction system is also Hamiltonian with the Hamiltonian

$$\mathfrak{H}(\hat{q}, \hat{p}) = -\frac{1}{\beta} \log \left(\frac{1}{c} \int \int e^{-\beta H(\hat{q}, \hat{p}, \tilde{q}, \tilde{p})} \, \mathrm{d}\tilde{q} \, \mathrm{d}\tilde{p} \right). \tag{1.12}$$

Here c is a constant with unit $[c] = [\tilde{q}] \cdot [\tilde{p}] = \left(kg\frac{m^2}{s}\right)^{n-m}$ (for a $2n$-dimensional system). The exact value of c is of no importance for the dynamics.

Proof:

$$\begin{aligned} \dot{q}_i &= \mathbb{E}\left[\frac{\partial H}{\partial p_i}|\hat{q},\hat{p}\right] = \frac{\int\int \frac{\partial H}{\partial p_i}(q,p)e^{-\beta H(q,p)}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}}{\int\int e^{-\beta H(q,p)}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}} \\ &= \frac{\partial}{\partial p_i}\left(-\frac{1}{\beta}\log\left(\frac{1}{c}\int\int e^{-\beta H(q,p)}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}\right)\right) \end{aligned} \tag{1.13}$$

Analogously for p_i.

□

From this Theorem it automatically follows that for nonlinear problems first order optimal prediction will never be a good approximation for long times, since the mean solution decays, i.e. loses energy, while the first order optimal prediction solution is Hamiltonian, and thus energy preserving.

However, one can also look at the problem in another way:

We have an n-dimensional Hamiltonian system, but wish to compute only the first m components. We forget any knowledge about the last $n - m$ components and wish to obtain an m-dimensional system for the first m components. This system should approximate the behavior of the first m components accurately. However, this is not meant in the sense of trajectories, but the relevant statistical quantities should be the recovered. In particular this means that the m-dimensional system should be Hamiltonian again.

The mean solution may be the evolution which has the least error compared to the whole collection of possible solutions, but in molecular dynamics the error is no relevant quantity anyway – unless for very short times. Hence why not choose the first order optimal prediction system as the m-dimensional system? Of course this choice can only be reasoned, if it turns out in the end that the relevant statistical quantities are indeed preserved. This issue will be our task in Chapter 4.

1.5 Mori-Zwanzig

The application of the Mori-Zwanzig formalism to optimal prediction can be found in [6, 7, 8]. It gives insight why first order optimal prediction produces an error as an approximation to the mean solution and explains why the mean solution decays even for

Hamiltonian systems. However, we will not employ the final results of the Mori-Zwanzig formulation in this thesis, so we will just state the most important ideas.

Let

$$L = \sum_{j=1}^{n} R_j(x) \frac{\partial}{\partial x_j} \tag{1.14}$$

be the so-called *Liouville operator*. Then the Liouville equation

$$\begin{aligned} u_t(x,t) &= Lu(x,t), \\ u(x,0) &= g(x) \end{aligned} \tag{1.15}$$

is a linear partial differential equation with solution

$$u(x,t) = g(\varphi(x,t)), \tag{1.16}$$

where φ is the solution of (1.1). This can be proven by the method of characteristics (see e.g. [16]). Now define the semigroup operator e^{tL} via the relation

$$\left(e^{tL} g\right)(x,t) = u(x,t) = g(\varphi(x,t)). \tag{1.17}$$

Two important properties of this operator are

$$Le^{tL} = e^{tL} L \tag{1.18}$$

and

$$\frac{d}{dt} e^{tL} g = Le^{tL} g. \tag{1.19}$$

For the special choice of $g(x) = x_i$ one obtains

$$e^{tL} x_i = \varphi_i(x,t), \tag{1.20}$$

i.e. the ordinary differential equation (1.1) and the partial differential equation (1.15) are equivalent.

Recall that P is the conditional expectation operator (1.7), and define

$$Q = Id - P. \tag{1.21}$$

Note that both P and Q are orthogonal projections with respect to the inner product (1.6). In particular

$$PQ = 0. \tag{1.22}$$

Then

$$\begin{aligned} \frac{\partial}{\partial t} \varphi_i(x,t) &= \frac{\partial}{\partial t} e^{tL} x_i \overset{(1.19)}{=} Le^{tL} x_i \overset{(1.18)}{=} e^{tL} L x_i \\ &\overset{(1.21)}{=} e^{tL}(P+Q)Lx_i = e^{tL} PLx_i + e^{tL} QLx_i. \end{aligned} \tag{1.23}$$

The use of *Dyson's formula* [17]

$$e^{tL} = \int_0^t e^{(t-s)L} PLe^{sQL}\,\mathrm{d}s + e^{tQL} \tag{1.24}$$

yields the identity

$$\frac{\partial}{\partial t}\varphi_i(x,t) = e^{tL}PLx_i + \int_0^t e^{(t-s)L}PLe^{sQL}QLx_i\,\mathrm{d}s + e^{tQL}QLx_i. \tag{1.25}$$

The evolution operator e^{tQL} represents the so called *orthogonal dynamics*. The formal existence of this operator has been treated in [22]. In particular

$$QLe^{tQL} = e^{tQL}QL. \tag{1.26}$$

From these relations the following identities can be derived

$$Lx_i = \sum_{j=1}^n R_j(x)\frac{\partial x_i}{\partial x_j} = \sum_{j=1}^n R_j(x)\delta_{ij} = R_i(x), \tag{1.27}$$

$$e^{tL}PLx_i \overset{(1.27)}{=} e^{tL}PR_i(x) \overset{(1.10)}{=} e^{tL}\mathfrak{R}_i(x) = \mathfrak{R}_i(\hat{\varphi}(x,t)), \tag{1.28}$$

and

$$Pe^{tQL}QLx_i \overset{(1.26)}{=} PQLe^{tQL}x_i \overset{(1.22)}{=} 0. \tag{1.29}$$

Using these identities and applying P to (1.25) yields the relation

$$\frac{\partial}{\partial t}P\varphi_i(x,t) = P\mathfrak{R}_i(\hat{\varphi}(x,t)) + \int_0^t Pe^{(t-s)L}PLe^{sQL}QLx_i\,\mathrm{d}s. \tag{1.30}$$

In [7, 8] the authors interpret the first term on the right hand side as the Markovian self-interaction of the first m variables, and the integral term as a non-Markovian memory term, which comes from the interaction of the first m variables with the averaged variables.

Interchanging P and $\mathfrak{R}$ in the first term and dropping the memory term yields the approximate equation

$$\frac{\partial}{\partial t}P\varphi_i(x,t) = \mathfrak{R}_i(P\hat{\varphi}(x,t)). \tag{1.31}$$

A reasoning for these approximations is given in [8]. Recalling that $P\varphi_i(x,t)$ is the mean solution one can observe that this is exactly the first order optimal prediction equation (1.11), which is again Hamiltonian and does not decay.

In [7] the authors show how the integral term in (1.30) can be approximated by a memory kernel, yielding the approximate system

$$\frac{\partial}{\partial t}P\varphi_i(x,t) = \mathfrak{R}_i(P\hat{\varphi}(x,t)) + \int_0^t \sum_{j=1}^m K_{ij}(t-s)P\varphi_i(x,t)\,\mathrm{d}s, \tag{1.32}$$

which is an integro-differential equation. Here the matrix K is a so-called memory kernel, which can be obtained by Monte-Carlo sampling. Unlike first order optimal prediction, the solution to this system does decay, so this approximation is referred to as *higher order optimal prediction*. In several examples (as the one presented in [7]) it is a much better approximation to the mean solution than first order optimal prediction, especially for longer times. However, as reasoned previously, in this thesis we will use first order optimal prediction only, since we want a smaller system which is again Hamiltonian.

Chapter 2

Description of the Problem

Computations in the field of molecular dynamics are typically governed by a large computational effort, and methods are required for reducing the degrees of freedom. Our goal is to investigate if the method of optimal prediction can perform such a task in a practical application.

It has to be pointed out that optimal prediction as presented in Chapter 1 has some important assumptions, which the problem under consideration should satisfy:

1. The system must have a measure which makes physically sense. This should not be a problem, since many problems in the field of molecular dynamics are Hamiltonian and thus give rise to a canonical measure.

2. The system must be in thermodynamical equilibrium, i.e. the temperature is constant with respect to space and time. This is a rather restrictive condition, since most problems which require a numerical simulation are not in equilibrium.

3. The system must allow a clear separation of particles into ones of interest and ones to be averaged away, and this separation has to be valid over time. This assumption is reasonable for systems at low temperature, e.g. in crystalline structures. On the other hand, in liquids or gases, this assumption will be violated, since particles change their positions considerably over time.

A problem, which is treated in the *Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM)* in Kaiserslautern, satisfies these assumptions. In a DFG project the process of coating a thin surface of copper atoms onto a silicon crystal should be simulated by molecular dynamics. In Section 2.1 we will explain the practical problem, and in which sense it satisfies the above assumptions. Since the complete problem is too complex as a test for the applicability of optimal prediction, we introduce a simplified one-dimensional model problem in Section 2.2, which represents the important properties of the real problem well enough to allow to draw conclusions about the real problem from our results.

2.1 The Real Problem

In the production of micro chips a thin layer of copper has to be coated onto a silicon crystal. Technically this is carried out by sputtering copper onto the silicon crystal. Over time, the copper builds up a thin surface (a few monolayers thick) on top of the silicon crystal. The latter is constantly cooled, such that the hot copper does not heat up the crystal significantly. More information about the technical process is provided in [26].

The surface coating process should be simulated and analyzed numerically. A classical paper on this issue is [41], which uses Monte-Carlo methods for the simulation. In the DFG project in the *ITWM* the simulation is done by molecular dynamics. Technically, the simulation is similar to the one described in [18].

The silicon atoms inside the crystal oscillate on a time scale of 10^{-14} seconds. The copper atoms hit the crystal with a much lower frequency. A typical growth rate of a copper film in the real coating process is about 10 monolayers per second. For a typical size of the simulation crystal this means that the time between two copper atoms hitting the crystal surface is on a time scale of 10^{-4} seconds, i.e. 10 time scales slower than the atomic oscillations. Of course this value can vary depending on the size of the simulated crystal. In any case, the copper atoms hitting the crystal can be considered single events, compared to the time scales of the processes of interest. Atom by atom the copper then builds up a thin layer on top of the silicon crystal, such that in the end the copper atoms hit the copper layer rather than the silicon crystal.

When a single copper atom hits the surface, it has a rather high kinetic energy, which has three major effects. Firstly, the atom penetrates into the crystal, secondly, it produces a shock wave traveling into the crystal, and thirdly, it heats up the atoms at the top of the crystal. It is obvious that at this stage the process is far from thermodynamical equilibrium, i.e. optimal prediction as presented in Chapter 1 cannot be applied to this process.

However, the above effects of one copper event last only about 10^{-11} seconds, i.e. only 0.00001% of the time between two copper events the system is in non-equilibrium, while 99.99999% of the time, i.e. nearly all the time, it is in thermodynamical equilibrium.

The problem is, that also during the long time in equilibrium single copper atoms penetrate deeper into the silicon crystal. This is an unwanted effect, since single copper atoms inside the crystal possibly produce inner tensions and may thus significantly reduce the stability properties of the product. Unlike the event when a copper atom hits the surface and penetrates into the crystal, the movement of the copper atoms during the time in equilibrium is of diffusive nature. Most of the time the copper atoms are caught inside a crystal cell of silicon atoms, where “caught” has to be understood dynamically. Both the copper atom and the silicon atoms oscillate on a time scale of 10^{-14} seconds. (A copper atom has only about twice the mass of a silicon atom, so there is no separation of time scales here.) But once in a while the fast oscillations form a constellation which

provides the copper atom with enough energy to penetrate the potential barrier built up by the silicon atoms on one side of the cell, and the copper atom jumps – or hops – to the neighboring cell. Since these hopping events are due to the fast oscillations, they happen – although deterministic – completely unsystematically. The time between two hopping events of one copper atom inside the silicon crystal is on a time scale of 10^{-11} to 10^{-10} seconds, i.e. 3 to 4 time scales slower than the atomic oscillations, so about 10^6 to 10^7 hopping events happen between two copper atoms hitting the crystal, enough for single atoms to diffuse into the crystal. The problem of molecular hopping is treated e.g. in [4].

The goal of the DFG project in the *ITWM* is to simulate the process of a growing copper layer (approximately 5000 atoms) on top of a silicon crystal in three space dimensions. A typical size of the silicon crystal, which can be computed on standard computers, is $40 \times 40 \times 10$ atoms. Of course larger model crystals are desirable. The lower 40×40 surface is coupled to a bulk, which provides a constant outside temperature and pressure coupling, such that energy and sonic waves can be transported away. The copper layer grows on the upper 40×40 surface. On the other layers of the crystal periodic boundary conditions are imposed. The dynamics are Hamiltonian, but the potentials may have a rather complicated structure. Special focus should be on the hopping of copper atoms during the time of thermodynamical equilibrium. In order to visualize the computational effort, let us recall the time scales of the problem:

time scale	process
10^{-15} seconds	time step of the integrator
10^{-14} seconds	one atomic oscillation
10^{-11} to 10^{-10} seconds	time between two hopping events
10^{-4} seconds	time between two copper atoms hitting the crystal
10^{-1} seconds	time for a complete copper layer to grow

So a straightforward simulation of the complete coating process would require 10^{14} integration steps of a 10^5-dimensional system of equations. Obviously, methods are required to reduce the computational effort.

The hope is that optimal prediction can be used to reduce the degrees of freedom during the long time of thermodynamical equilibrium between two copper events. Silicon atoms far away from copper atoms can be "averaged out". Since silicon atoms normally do not perform hopping events, a clear separation between particles of interest and ones to be averaged away is possible. Finally, since the system is Hamiltonian, the canonical measure $Z^{-1}e^{-\beta H}$ should make sense in this application, although a complicated nonlinear structure in the potential energy might cause problems in applying the conditional expectations as in (1.10) efficiently (see Chapter 3).

Hence, the assumptions for optimal prediction are satisfied, and one can hope that the method can be applied successfully for the time in equilibrium. Since this is the case for

about 99.99999% of the whole computational time, a reduction of the number of atoms in each space dimension by just one half might result in a speed-up of factors of 10 and more.

2.2 The Model Problem

In order to investigate the question if optimal prediction can in principle be applied to the above problem, we set up a model problem which describes the hopping of a copper atom in a silicon crystal. It should reflect the important properties of the real problem, but it should be easier to handle with respect to computational effort and complexity of the expressions obtained when applying the method of optimal prediction.

In constructing the model problem, we assume two major simplifications:

- **Focus on a one-dimensional problem**, i.e. we consider n atoms lined up like beads on a cord, as shown in Figure 2.1.
- **The potential $V(q)$ depends only on the pairwise distances of the particles**, i.e.

$$V(q_1, \ldots, q_n) = \sum_{\substack{i,j=1 \\ i<j}}^{n} f_\alpha(q_i - q_j). \tag{2.1}$$

  Here $\alpha \in \{1, 2, 3\}$, where f_1 is the potential between two silicon atoms, f_2 is the potential between a silicon and a copper atom, and f_3 is the potential between two copper atoms. All three are even, analytic functions, which vanish at $\pm\infty$.

The pair potentials f_1, f_2 and f_3 look similar to the corresponding potentials in three space dimensions. They have a minimum at distance r_e. At a smaller distance, the two atoms will repel, while at greater distances, the two atoms will attract each other. Towards infinite distance, the force between the two atoms vanishes. As usual in this field we gauge the potentials to go to 0 at infinity.

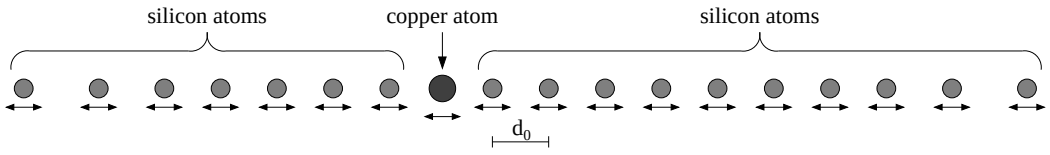


Figure 2.1: The one-dimensional silicon crystal with a copper atom

A property which can not be preserved from the three-dimensional problem is the behavior of the potentials towards distance 0. Since two particles classically cannot be at the same place, potentials – like the widely used Lennard-Jones potential (see e.g. [25]) – go to infinity as the distance goes to 0. However, in our one-dimensional problem, this would make hopping impossible. While in three space dimensions hopping of a copper atom means moving though the plane given by three silicon atoms forming the crystal structure, in one space dimension hopping can only mean a copper atom changing position with a silicon atom. Since classical particle paths are continuous, this requires the two atoms to have the same position at one time. Hence, in one space dimension, hopping between two particles is only possible if the potential between them is finite at distance 0.

In our modeling, we decide to allow hopping between copper and silicon atoms as well as hopping between two copper atoms, i.e. we make the potentials f_2 and f_3 finite at distance 0. Hopping between two silicon atoms, however, we disallow, i.e. the potential f_1 is infinite at the origin. This decision is motivated by technical reasons. In three space dimensions, silicon atoms can very well change positions, although such an event happens – due to the crystal structure and higher potential barriers – significantly less frequently than a hopping event of a copper atom. But since a hopping event of two silicon atoms through the (virtual) border between particles of interest and those to be averaged out would violate our third assumption for the applicability of optimal prediction, we disallow such a behavior.

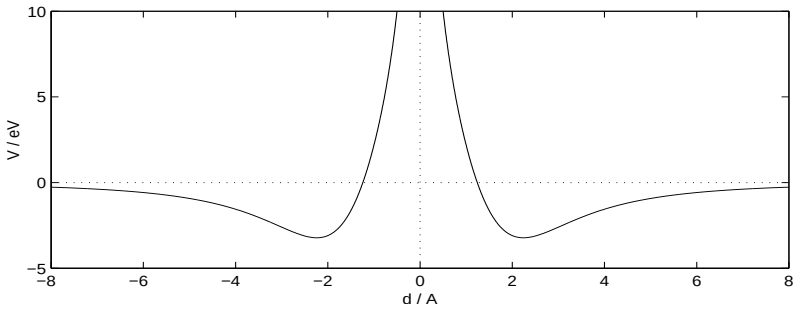


Figure 2.2: Potential f_1 between two silicon atoms

Figures 2.2 to 2.4 show the potentials used for our numerical simulations. The distance is given in Å ($1\text{Å} = 10^{-10}$m) and the energy in eV ($1\text{eV} = 1.6 \cdot 10^{-19}$J). Note the different scales of the y-axes.

The potentials are chosen to be close – at least qualitatively – to the correct three dimensional potential, which are similar to the Lennard-Jones potential. In particular, for f_1 the position of the minimum ($r_e = 2.24\text{Å}$) and the energy at the minimum ($E = -D$, where $D = 3.24\text{eV}$) are correct values. Furthermore, for f_3, $r_e = 2.21\text{Å}$ is the correct value. Apart from that the other characteristic quantities differ from their correct values, for technical reasons, e.g. making hopping more likely (see below). All numerical values are taken from [39].

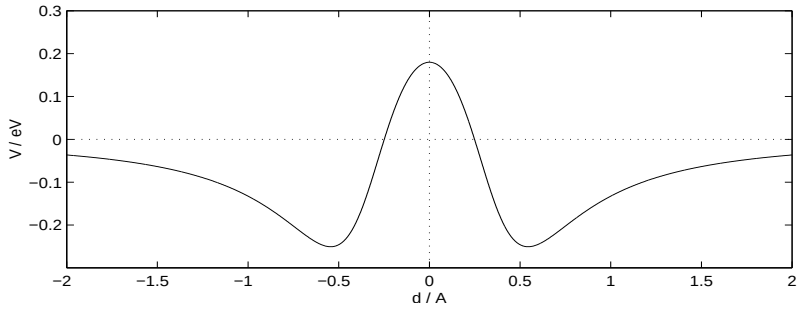


Figure 2.3: Potential f_2 between a copper and a silicon atom

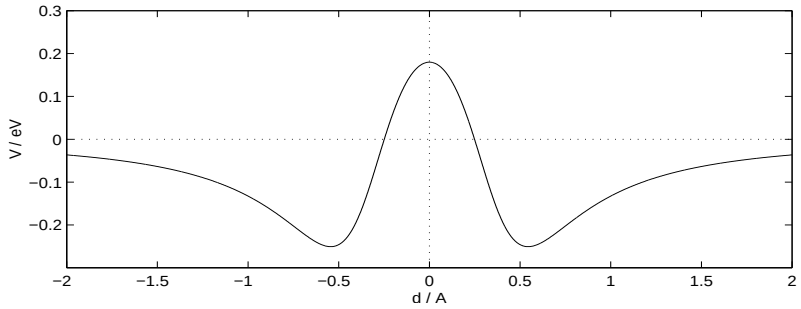


Figure 2.4: Potential f_3 between two copper atoms

Relevant for the frequency of hopping events is the value of the potential at the origin, since this potential barrier has to be exceeded for a hopping event to happen. For computational reasons we choose to lower the potential barrier. Additionally, as we will see in Section 2.3, we increase the temperature of the process by a relevant factor. Both modifications have the effect that hopping events happen much more frequently than in reality. On the other hand, they are still on a time scale much slower than the atomic oscillations. This helps us reducing the computational effort by a factor of 100, without changing the basic properties of the system. The computational factor becomes important in Chapter 4, when we have to perform extremely costly Monte-Carlo simulations in analyzing the diffusion of a copper atom due to hopping events. Note that in the particular examples in this thesis we will insert at most one copper atom into the crystal, hence the potential f_3 is not being used. However, since the ideas presented in the following are not restricted to just one copper atom, we include the potential f_3 into the model problem.

At this point one could ask the question: "If this is a model problem anyhow, why care about being close to correct values?" The answer to this question is: "Because it is important, which numerical value the expression $\frac{D}{k_B T}$ has." In Chapter 3 we will use asymptotic methods for approximating the conditional expectations (1.5), and for this approximation to be valid, $\frac{D}{k_B T}$ has to be a large number. In the following Section we will reason and estimate this constant.

2.3 Physical Properties of the Model Problem

Mathematically, the equations of motion, which completely describe our model problem, are already defined by the Hamiltonian function, including the potentials f_1, f_2 and f_3. However, since our system models a real process, some quantities have a physical meaning, which is important for later analysis in Chapter 3 and Chapter 4. In this section we provide numerical values of the important physical quantities and discuss how they relate to a numerical simulation of the model problem.

Temperature
As described in Section 2.1, the temperature of the real process is not very high. Indeed, the process will run at temperatures between 400K and 500K. The question if this temperature is a "low" temperature for this kind of process, can be quantified by the expression $\frac{D}{k_B T}$, where $-D$ is the minimum of a potential acting between two particles (which is typically negative). This number corresponds to the situation, when two atoms are at equilibrium distance, i.e. minimal potential energy. Recalling that this value for the potential between two silicon atoms is $D = 3.24$eV, we obtain for a temperature of 500K:

$$\frac{D}{k_B T} = 103.68. \tag{2.2}$$

Since $Z^{-1} e^{-\frac{H(q)}{k_B T}}$ is the probability density on the space of all states q, this means that two atoms are about $e^{103.68} \approx 10^{45}$ times more likely at equilibrium distance than at a distance corresponding to energy zero, which is the case e.g. far away from each other.

In the same manner one can reason why hopping events between copper and silicon are not very likely at these temperatures. The energy difference between equilibrium and zero distance for the potential between silicon and copper f_2 has the value $\Delta E = 0.43$eV, which yields

$$\frac{\Delta E}{k_B T} = 13.76. \tag{2.3}$$

Consequently, it is $e^{13.76} \approx 10^6$ times more likely for the copper atom to be at equilibrium distance than to perform a hopping event. Hence atomic hopping happens on a much slower time scale than the fast atomic oscillations.

In order to increase the number of hopping events, we increase the temperature for our simulations to $T_1 = 4000$K (as in Section 4.3) respectively $T_2 = 7000$K (as in Section 4.4). This yields a values of $\frac{D}{k_B T_1} = 12.96$ respectively $\frac{D}{k_B T_2} = 7.41$, which are for the above reasons still large enough numbers. Since later we will employ the fact that the temperature be low, we can expect the method to work even better at correct temperatures (400K to 500K) than it does for the temperatures of more than 4000K, which are used in our computations.

Pressure
As we consider a problem in thermodynamical equilibrium, we do not need to consider

pressure as an important quantity. The only point where pressure will come in implicitly, is in Section 4.3, when we estimate the velocity of sound in our silicon crystal.

Ergodicity
It is not our task to argue if or why our problem is (quasi-)ergodic or not.[1] We are content with the fact that our experiments yield results which reflect the predictions of statistical mechanics about ergodic system. Whenever in the following we employ a statement from statistical mechanics, we always assume that our system behaves like an ergodic system. In particular this means that we assume space averages over a number of atoms to be equal to time averages over a given time span.

Initial conditions
As assumed by optimal prediction, our system has to be in thermodynamical equilibrium, i.e. the average momentum of the atoms is constant (temperature is constant) in space and time, and potential and kinetic energy fluctuate around fixed values. Since nonlinear Hamiltonian systems always tend towards thermodynamical equilibrium, we are content with starting in some reasonable state on the energy surface and let molecular dynamics drive the system into equilibrium. We let our system start in the state of minimal potential energy, since this state can be computed directly by Newton iteration. It is the equilibrium state corresponding to temperature zero, i.e. the static equilibrium, where all forces cancel out. The momenta, however, are non-zero, hence they correspond to a positive temperature. We sample them independently from Gaussian distributions, subtract the mean momentum (to keep the center of mass constant in time), and normalize them to fit the chosen temperature $T = 4000$K, respectively $T = 7000$K, employing the relation

$$T = \frac{1}{k_B n} \sum_{i=1}^{n} \frac{p_i^2}{m_i}. \tag{2.4}$$

Obviously this initial state is not in the thermodynamical equilibrium for $T = 4000$K, respectively $T = 7000$K, but numerical experiments, as in Section 2.4, indicate that after a short time of about 10^{-13} seconds the system is in equilibrium. This fact can be seen (e.g. in Figure 2.5) by the kinetic energy not decaying anymore, but fluctuating around a constant value.

2.4 Numerical Simulation

The task of this thesis is to investigate if optimal prediction can in principle be applied to problems in molecular dynamics. It is not our task to consider highly expensive examples which require special methods or parallel computing. Instead we constructed our model problem in order to be computationally inexpensive. Two important changes compared to the real problem are

[1] After reading [19, pp. 24-27]

- Lowering the potential barrier between copper and silicon,
- Increasing the temperature from 500K to temperatures of 4000K and 7000K.

As described in Section 2.2, both changes make hopping events of copper atoms in the silicon crystal many times more frequent. Since in Chapter 4 we employ the hopping of a copper atom as a way to compare the optimal prediction system to the original system, a significant increase in the hopping frequency helps to reduce the computational effort.

The number of particles being computed should not be unnecessarily large, but at least large enough, such that the system still behaves as statistical mechanics predict. In particular, one needs a minimum amount of particles, such that statistical (averaged) quantities like temperature make sense and one can speak of a thermodynamical equilibrium. Additionally, enough atoms are required such that there is a reasonable interior region of the crystal, which is not affected by any boundary effects. In our numerical experiments we use 35 to 70 atoms. With more than 30 particles the above assumptions were satisfied satisfyingly well.

As reasoned in Section 2.3, we are reasonably close to thermodynamical equilibrium after a very short time, when starting at minimum potential energy and sampling the momenta from Gaussian distributions.

For simplicity, the integration is performed by the classical explicit fourth order Runge-Kutta method. Although usually symplectic (i.e. energy preserving) methods are used for Hamiltonian systems, we prefer the Runge-Kutta method for the following reasons:

- It is an explicit method and can thus be applied directly to any approximative right hand side which arises in optimal prediction, as derived in Chapter 3.
- It is very accurate. Usually in computational molecular dynamics accuracy is less important than energy conservation, since exact trajectories are not important anyhow. However, in our application, hopping of copper atoms is one major feature, and if a copper atom hops or not depends very sensitively on the positions and momenta of the atoms in the neighborhood of the copper atom. A brief numerical test during our experiments indicated that the first order accurate symplectic Euler method requires significantly smaller time steps than Runge-Kutta 4, to achieve that the hopping behavior is (almost) independent of the chosen time step.

The time step in the Runge-Kutta 4 method is chosen in such a way, that the total energy does not change too significantly. We choose equidistant time steps between $1.5 \cdot 10^{-15}$s and $2.5 \cdot 10^{-15}$s. For the example computation shown in Figure 2.5, which had 1000 time steps, the total energy increased by 0.04%. More physically motivated is to express the change in energy in units of $k_B T$. In the example here, we obtain

$$\frac{H(t = 2 \cdot 10^{-12}\text{s}) - H(t = 0\text{s})}{k_B T} = -0.16, \tag{2.5}$$

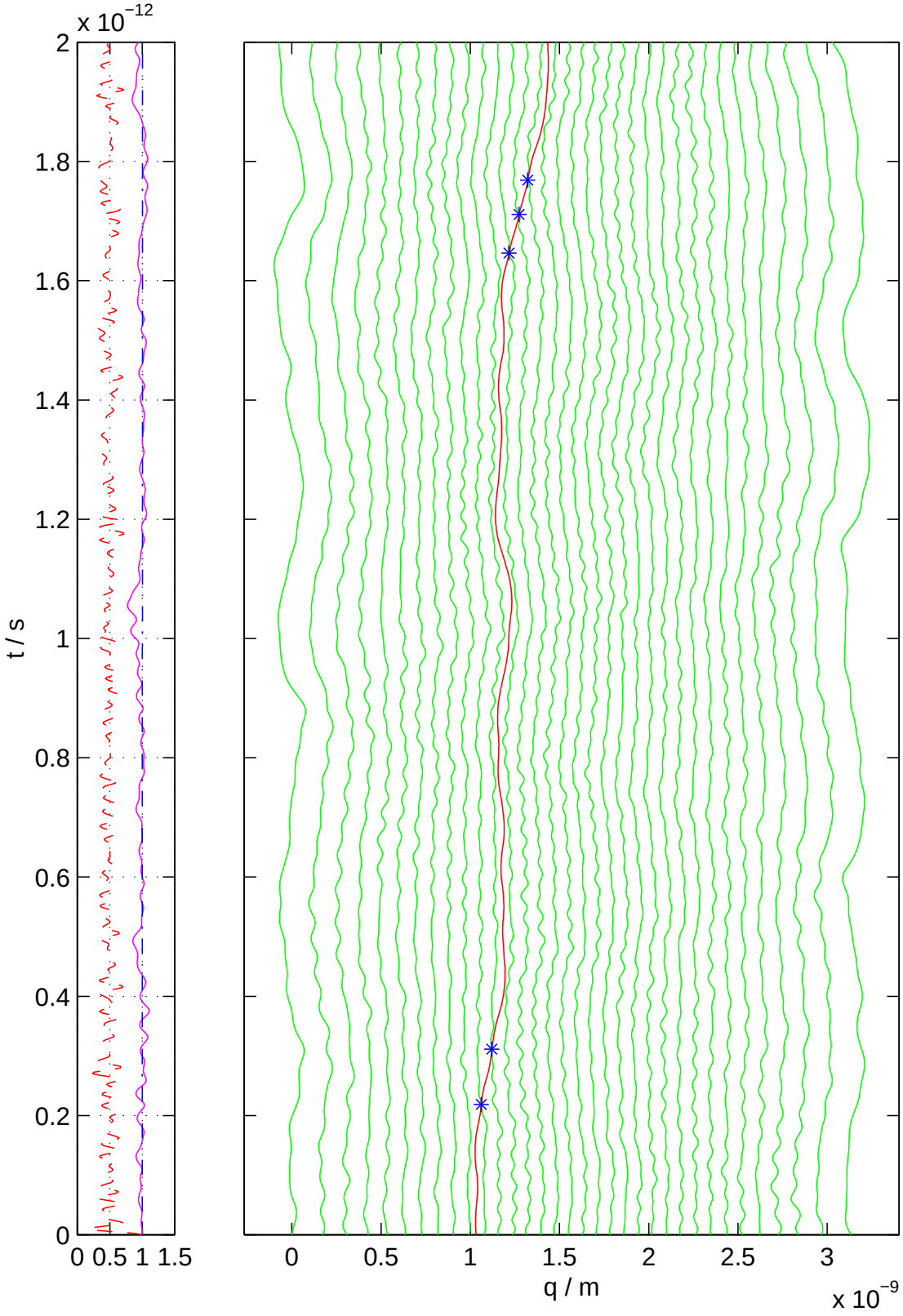


Figure 2.5: A simulation of the model problem

which is still reasonable for our task.

The structure of the potential energy (2.1) makes the effort for one evaluation of the equation's right hand side of order n^2, where n is the number of particles, since each atom is affected by any other atom. However, as potentials between two atoms typically decay quite fast for increasing distances (e.g. like $1/r^6$ for the classical Lennard-Jones potential), one can often assume that the potentials have a compact support, i.e. range over a distance of only k equilibrium distances. Under this assumption, the effort for one right hand side evaluation is only of order n. However, depending on the specific potentials and the accuracy required, k might have to be quite large. Another important requirement is that the approximation errors must not stack into one direction. For the original system the errors in principle cancel out due to the spatial symmetry of the geometry. For the optimal prediction approximation, as presented in Chapter 3, we will not have such a symmetry anymore. As it can be seen in Section 3.4, a choice for k which works perfectly fine for the original system can be too small for the optimal prediction approximation to yield accurate results. Of course, the above approximation in cutting off the potentials is technically a very simple one. With respect to the real, three-dimensional problem, one can think of more accurate and efficient methods, as smoothing out the cut off potentials, or – for long ranged potentials – using the fast multipole method [20]. The point, that care has to be taken when using other approximations together with optimal prediction, stays valid and will also be dealt with in Section 3.4.

Figure 2.5 shows the numerical results for an example computation of the model problem. A crystal of 34 silicon and one copper atom (i.e. $n = 35$) is being computed over a time of $2.0 \cdot 10^{-13}$s. The time step of the Runge-Kutta 4 integrator is $2.0 \cdot 10^{-15}$s. The temperature is $T = 7000$K.

In the right subplot the positions $q_i(t)$ of the 35 atoms are plotted over time. Here the horizontal axis is space and the vertical axis is time. One can observe that the silicon atoms oscillate non-periodically around their starting positions on a fast time scale. The copper atom oscillates between two silicon atoms on a time scale irrelevantly slower due to its larger mass. Important are the hopping events, i.e. the copper atom hops over a silicon atoms and thus changes its position inside the crystal. Each hopping event is emphasized by a blue asterisk in the plot. Observe further the surface effect on both ends of the crystal. As for the outermost silicon atoms other atoms are missing on one side, the equilibrium distance between two atoms is significantly larger than in the middle of the crystal. Since the potential between two silicon atoms reaches over several equilibrium distances (see Figure 2.2), this surface effect affects not only the outermost two atoms, but a whole layer. In many contexts, the above effect is called *surface tension*. Obviously one has to take care about the surface tension when dealing with quantities which depend on the equilibrium distance between two silicon atoms, e.g. the velocity of sound, as used in Chapter 4. Also, the diffusion of a copper atom by hopping events may change its behavior when coming too close to the crystal's surface.

In the left subplot three important quantities are plotted over time:

1. The total energy over the initial energy

$$\frac{H(q(t),p(t))}{H(q(0),p(0))}, \tag{2.6}$$

where

$$H(q,p) = \frac{1}{2}\sum_{i=1}^{n}\frac{p_i^2}{m_i} + \sum_{\substack{i,j=1\\ i<j}}^{n} f_\alpha(q_i - q_j). \tag{2.7}$$

It is the dash-dotted blue curve. Observe that it is constant 1 in the plot. Numerically, its value is 1.0002 at time $t = 2.0 \cdot 10^{-12}$s.

2. The kinetic energy over the initial kinetic energy

$$\frac{T(p(t))}{T(p(0))}, \tag{2.8}$$

where

$$T(p) = \frac{1}{2}\sum_{i=1}^{n}\frac{p_i^2}{m_i}. \tag{2.9}$$

It is denoted by the dashed red curve. This function starts at 1, and decays very fast to a value of 0.5, around which it oscillates from then on. The reason for this behavior is that we start our system in the static equilibrium, i.e. with minimal potential energy, and hence with maximal kinetic energy. Since for ergodic Hamiltonian systems in equilibrium the total energy is – averaged over time – distributed equally in kinetic and potential energy, the kinetic energy decays to half of its initial value as the system goes into thermodynamical equilibrium. Hence the function $\frac{T(p(t))}{T(p(0))}$ is an important indicator for the time the system takes to reach equilibrium. The fact that this time is very small in this case (about $2.0 \cdot 10^{-14}$s) shows that we can be content with our strategy of choosing the initial values.

3. The total energy of the left $m = 25$ atoms over its initial value

$$\frac{E_{\text{left}}(q(t),p(t))}{E_{\text{left}}(q(0),p(0))}, \tag{2.10}$$

where

$$E_{\text{left}}(q,p) = \frac{1}{2}\sum_{i=1}^{m}\frac{p_i^2}{m_i} + \sum_{\substack{i,j=1\\ i<j}}^{m} f_\alpha(q_i - q_j). \tag{2.11}$$

It is the magenta curve in the plot, which fluctuates around 1. Note that it is stretched by a factor of 5 around the value 1, in order to make the fluctuations visible. The value of (2.11) is close to 1, but not exactly 1, since energy is constantly exchanged between the left m and the right $n - m$ atoms. The amplitude and frequency of this function are measures for the rate of exchange of energy between silicon atoms. Both are statistical quantities, which should be preserved by the smaller optimal prediction system. In Chapter 4 we will make use of these quantities in order to compare the optimal prediction to the original system.

Chapter 3

Optimal Prediction Applied to the Model Problem

In this chapter we apply the method of first order optimal prediction as defined in Section 1.4 to the one-dimensional model problem described in Section 2.2. As reasoned in Section 1.4, we choose the first order optimal prediction system as an approximation to our n-dimensional Hamiltonian system, because it is Hamiltonian itself. The mean solution, which first order optimal prediction is constructed to approximate, is of no interest for us, since it decays and is thus not energy preserving for the nonlinear problem here.

Since in one space dimension we have a clear ordering of atoms, the choice which atoms to average away is fairly obvious. Hence the formal application of optimal prediction is straightforward. The main problem are the high-dimensional integrals appearing in the conditional expectation projection (1.5), respectively in the new Hamiltonian function (1.12) given by Theorem 1.1. The first part of this problem is to define the correct subsets of $\mathbb{R}^{n-m}$, which to integrate the q-variables in (1.5) over. We will deal with this issue in Subsection 3.1.1. The second part of this problem is how to evaluate the $(n-m)$-dimensional integrals. Since a direct evaluation is in general not easily possible and a Monte-Carlo approximation for a single right hand side evaluation is far too expensive, we employ an asymptotic expansion for low temperatures in order to obtain an approximate expression, which can be evaluated directly. This will be done in Subsection 3.2.1.

In Section 3.3 we will use the zero temperature limit of the asymptotic expansion, derive the corresponding equations of motion (Subsection 3.3.2), and present approximative equations, which describe a boundary layer condition which simulates an infinitely continued crystal (Subsection 3.3.4). In Section 3.4 we present numerical examples and consider the question of computational speed-up.

3.1 Mathematical and Physical Setup

In this section we show how to set up the problem, such that the physically motivated measure $Z^{-1}e^{-\beta H(q,p)}$ makes sense mathematically, as well as how to separate the optimal prediction Hamiltonian (1.12) into kinetic and potential energy.

3.1.1 Appropriate Domains of Integration

We assume that for our Hamiltonian system the canonical measure

$$f(q,p) = Z^{-1}e^{-\beta H(q,p)} \tag{3.1}$$

is the "correct" measure, where "correct" is meant in the sense of statistical mechanics. In statistical mechanics this measure can be derived from the principle of maximum entropy, as in [19]. Typically in such derivations special properties about the potentials are assumed, or one restricts the possible positions of the particles. The reason is that, once free particles are possible, the expression

$$\int_{\mathbb{R}^n} e^{-\beta H(x)}\,\mathrm{d}x \tag{3.2}$$

is not finite, i.e. $Z^{-1}e^{-\beta H(q)}$ does not make sense as a probability distribution. The easiest example is one free particle, i.e. no potential acting. Since in our problem the potentials are chosen to vanish at infinity, we are in the same problematic situation.

What is often done in text books on statistical mechanics, is to put the whole system, i.e. the positions $q_1, \ldots, q_n$, into a box of finite volume. This makes the integral in (3.2) finite, but is often still unsatisfying, since the measure depends on the size of the box. (The integral over the momenta $p_1, \ldots, p_n$ can of course always be taken over the whole $\mathbb{R}^n$.)

In the following , we will force the system into a domain of finite volume, such that $Z^{-1}e^{-\beta H(q)}$ can be interpreted as a measure. Unlike most text books on statistical mechanics, we do not choose simply a box $[-L, L]^n$, but a simplex-shaped subset of this box (for the silicon atoms). In Subsection 3.2.1 this domain can be extended to infinity again, since the asymptotic expansion in this section will yield an expression which is integrable over the whole space $\mathbb{R}^{n-m}$.

The reason why we restrict the shape of the domain of integration to a small part of a box is the fact that our silicon atoms are ordered. We know their initial order, and since their potential is infinite at zero distance, this initial order is preserved for all times. This is an information, which we do not want to average out. We will see in Subsections 3.2.1 and 3.3.1 why it is indeed the correct choice to preserve the information about the ordering of

the silicon atoms. Assume for a moment that no copper atoms are present. Then we can restrict the domain of integration from the whole $\mathbb{R}^n$ to

$$M^\infty = \{(q_1, q_2, \ldots, q_n) | q_1 < q_2 < \cdots < q_n\} \subset \mathbb{R}^n. \tag{3.3}$$

Analogously, for the conditional expectations, $\mathbb{R}^{n-m}$ is replaced by

$$M^\infty_{\hat{q}} = \{(q_{m+1}, \ldots, q_n) | q_m < q_{m+1} < \cdots < q_n\} \subset \mathbb{R}^{n-m}. \tag{3.4}$$

Here only such $\hat{q}$ are considered, which satisfy

$$\hat{q} \in \{(q_1, \ldots, q_m) | q_1 < \cdots < q_m\}. \tag{3.5}$$

This choice is appropriate for a crystal consisting only of silicon atoms. Since copper atoms can change positions inside the silicon crystal by hopping events, their position cannot be restricted in the above manner. Instead, for any copper atom, we must allow for the position that $q_i \in [-L, L]$. However, as we will reason in Subsection 4.1.1, we can choose the computation times so short, that we can restrict to situations where no copper atom comes close to the virtual border between particles computed in optimal prediction $(q_1, \ldots, q_m)$ and the ones averaged away $(q_{m+1}, \ldots, q_n)$. We can afford such short computation time by obtaining important statistical quantities by Monte-Carlo sampling instead. Under this assumption, the presence of copper atoms would influence the set M^∞, but the set $M^\infty_{q_1,\ldots,q_m}$, which is the actual domain of integration, remains unchanged.

Note that expression

$$\int_{\mathbb{R}^n} \int_{M^\infty} e^{-\beta H(q)} \, \mathrm{d}q \, \mathrm{d}p \tag{3.6}$$

is still infinite, so at this stage the appropriate sets of integration are

$$M^L = \{(q_1, \ldots, q_n) \in [-L, L]^n | q_1 < \cdots < q_n\} \tag{3.7}$$

respectively

$$M^L_{\hat{q}} = \{(q_{m+1}, \ldots, q_n) \in [-L, L]^{n-m} | q_m < q_{m+1} < \cdots < q_n\}. \tag{3.8}$$

In both cases L should be large enough for the whole crystal to fit into the box $[-L, L]^n$. As previously stated, we will be able to omit the restriction to finite domains in Subsection 3.2.1.

3.1.2 The Optimal Prediction Hamiltonian

In following derivation let $l = n - m$ denote the number of particles to be averaged out. As shown in Chapter 1, we can circumvent the application of the conditional expectation

(1.5) to the system's right hand side by computing the Hamiltonian function of the first order optimal prediction system (1.12):

$$\mathfrak{H}(\hat{q},\hat{p}) = -\frac{1}{\beta}\log\left(\frac{1}{c}\int_{\mathbb{R}^l}\int_{M_{\hat{q}}^L} e^{-\beta H(\hat{q},\hat{p},\tilde{q},\tilde{p})}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}\right). \tag{3.9}$$

One can easily check that the positions q and momenta p separate

$$\begin{aligned}\mathfrak{H}(\hat{q},\hat{p}) &= -\frac{1}{\beta}\log\left(\frac{1}{c}\int_{\mathbb{R}^l}\int_{M_{\hat{q}}^L} e^{-\beta(T(\hat{p},\tilde{p})+V(\hat{q},\tilde{q}))}\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}\right)\\ &= \underbrace{-\frac{1}{\beta}\log\left(\frac{1}{c_p^l}\int_{\mathbb{R}^l} e^{-\beta T(\hat{p},\tilde{p})}\,\mathrm{d}\tilde{p}\right)}_{=\mathfrak{T}(\hat{p})}\underbrace{-\frac{1}{\beta}\log\left(\frac{1}{c_q^l}\int_{M_{\hat{q}}^L} e^{-\beta V(\hat{q},\tilde{q})}\,\mathrm{d}\tilde{q}\right)}_{=\mathfrak{V}(\hat{q})}.\end{aligned} \tag{3.10}$$

Here c_p and c_q are constants with units $[c_p] = \mathrm{kg}\frac{\mathrm{m}}{\mathrm{s}}$ and $[c_q] = \mathrm{m}$. Since $T = \frac{1}{2}\sum_{i=1}^n \frac{p_i^2}{m_i}$, the first term of (3.10) can be computed directly as

$$\mathfrak{T}(\hat{p}) = \frac{1}{2}\sum_{i=1}^m \frac{p_i^2}{m_i} + C, \tag{3.11}$$

where the constant $C = -\frac{1}{2\beta}\sum_{i=m+1}^n \log\left(\frac{2\pi m_i}{\beta c_p^2}\right)$ is of no relevance for the dynamics. Hence, with respect to the momenta applying first order optimal prediction is just omitting the momenta $p_{m+1},\ldots,p_n$.

However, for the potential V, life is far from being as easy as for the kinetic energy T. Typically, in molecular dynamics the q-variables do not separate and are no quadratic functions, not even polynomials. Hence, an analytic calculation of the first order optimal prediction potential

$$\mathfrak{V}(\hat{q}) = -\frac{1}{\beta}\log\left(\frac{1}{c_q^l}\int_{M_{\hat{q}}^L} e^{-\beta V(\hat{q},\tilde{q})}\,\mathrm{d}\tilde{q}\right) \tag{3.12}$$

is in general impossible, or at least too complicated to be of any use. In the following section, we will obtain an asymptotic approximation to (3.12), which we can then continue to work with.

3.2 Low Temperature Asymptotics

In Section 2.3 we found the dimensionless quantity

$$\varepsilon = D\beta = \frac{D}{k_B T} \tag{3.13}$$

to carry the information if the process is at low or high temperature. Here $-D$ is the minimum potential energy between two silicon atoms, which is typically negative. The distance of the two atoms corresponding to the minimum energy is called equilibrium distance. In a system in thermodynamical equilibrium it is e^{ε} times more likely to find two atoms at equilibrium distance than at a distance corresponding to energy zero, e.g. far away from each other. Consequently, if ε is large, the atoms will form an equidistant grid at equilibrium distance, i.e. a solid body, while for low values of ε one will have a fluid or even a gas. Here ε being large means $\varepsilon > 7$, since even for $\varepsilon = 7$ two atoms having minimum energy is $e^7 \approx 10^3$ times more likely than having energy zero. As computed previously, in our model problem we have temperatures of $T = 4000$K and $T = 7000$K, i.e. $\varepsilon = 13.0$, respectively $\varepsilon = 7.4$. In the real problem, the value of ε is more than 100, so one can expect the following approximations to yield much better results when applied to the real problem.

3.2.1 Asymptotic Expansion of the Hamiltonian

We can now express (3.12) in terms of the dimensionless quantity ε:

$$\mathfrak{V}(\hat{q}) = -\frac{D}{\varepsilon} \log \left(\frac{1}{c_q^l} \int_{M_{\tilde{q}}^L} e^{-\varepsilon \tilde{V}(\hat{q},\tilde{q})} \, \mathrm{d}\tilde{q} \right), \tag{3.14}$$

where $\tilde{V}(\hat{q},\tilde{q}) = \frac{1}{D} V(\hat{q},\tilde{q})$ is the potential normalized in such a way, that the potential of two atoms at equilibrium distance has the value -1.

Using *Laplace's method for integrals of real variables*[1] [34, 36], we can find an asymptotic approximation to (3.14) for ε large. The idea is, for any given $\hat{q}$, to approximate $\tilde{V}(\hat{q},\tilde{q})$ by a quadratic function located at the minimum.

Assume for the moment that for a fixed choice of $\hat{q}$ the function $\tilde{V}(\hat{q},\tilde{q})$ has a unique global minimizer $r(\hat{q}) \in \mathbb{R}^l$ with respect to $\tilde{q}$, and that the Hessian at this point $\frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ is regular. Then Laplace's method yields the following asymptotic approximation for $\varepsilon \to \infty$:

$$\begin{aligned} \int_{M_{\tilde{q}}^L} e^{-\varepsilon \tilde{V}(\hat{q},\tilde{q})} \, \mathrm{d}\tilde{q} &\approx \int_{M_{\tilde{q}}^L} e^{-\varepsilon \left(\tilde{V}(\hat{q},r(\hat{q})) + \frac{1}{2}(\tilde{q}-r(\hat{q}))^T \cdot \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q},r(\hat{q})) \cdot (\tilde{q}-r(\hat{q})) \right)} \, \mathrm{d}\tilde{q} \\ &= e^{-\varepsilon \tilde{V}(\hat{q},r(\hat{q}))} \cdot \int_{M_{\tilde{q}}^L} e^{-\frac{\varepsilon}{2} \left((\tilde{q}-r(\hat{q}))^T \cdot \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q},r(\hat{q})) \cdot (\tilde{q}-r(\hat{q})) \right)} \, \mathrm{d}\tilde{q} \\ &\approx e^{-\varepsilon \tilde{V}(\hat{q},r(\hat{q}))} \cdot \int_{\mathbb{R}^l} e^{-\frac{\varepsilon}{2} \left((\tilde{q}-r(\hat{q}))^T \cdot \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q},r(\hat{q})) \cdot (\tilde{q}-r(\hat{q})) \right)} \, \mathrm{d}\tilde{q}. \end{aligned} \tag{3.15}$$

[1] In some textbooks this method is also referred to as *Watson Lemma*.

Extending the set of integration to the whole $\mathbb{R}^l$ is valid, since the minimum is always in the interior of $M_{\hat{q}}^L$, provided L is large enough (see [36]). Since $\frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ is assumed to be regular, one obtains by applying the transformation rule that

$$\frac{1}{c_q^l} \int_{\mathbb{R}^l} e^{-\frac{\varepsilon}{2}(\tilde{q}-r(\hat{q}))^T \cdot \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \cdot (\tilde{q}-r(\hat{q}))} \, \mathrm{d}\tilde{q} = \sqrt{\frac{(2\pi)^l}{\varepsilon^l \left| \det c_q^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right|}}. \tag{3.16}$$

Given $\tilde{V}$ is of class C^2 (which is the case in our model problem), the complete asymptotic expansion including the error term is

$$\begin{aligned} \frac{1}{c_q^l} \int_{M_{\hat{q}}^L} e^{-\varepsilon \tilde{V}(\hat{q}, \tilde{q})} \, \mathrm{d}\tilde{q} = \sqrt{\frac{(2\pi)^l}{\varepsilon^l \left| \det c_q^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right|}} \cdot e^{-\varepsilon \tilde{V}(\hat{q}, r(\hat{q}))} \\ + e^{-\varepsilon \tilde{V}(\hat{q}, r(\hat{q}))} \cdot O\left(\varepsilon^{-\frac{l}{2}-1}\right), \end{aligned} \tag{3.17}$$

which follows directly from the one-dimensional case as shown in [34, pp. 33-34]. Substituting (3.17) into (3.14) yields

$$\begin{aligned} \mathfrak{V}(\hat{q}) &= -\frac{D}{\varepsilon} \log \left(e^{-\varepsilon \frac{1}{D} V(\hat{q}, r(\hat{q}))} \cdot \left(\varepsilon^{-\frac{l}{2}} \sqrt{\frac{(2\pi)^l}{\left| \det c_q^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right|}} + O\left(\varepsilon^{-\frac{l}{2}-1}\right) \right) \right) \\ &= V(\hat{q}, r(\hat{q})) + \frac{Dl}{2\varepsilon} \log\left(\frac{\varepsilon}{2\pi}\right) - \frac{D}{\varepsilon} \log \left(\frac{1}{\sqrt{\left| \det c_q^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right|}} + O\left(\frac{1}{\varepsilon}\right) \right) \\ &= V(\hat{q}, r(\hat{q})) + \frac{Dl}{2\varepsilon} \log\left(\frac{\varepsilon}{2\pi}\right) + \frac{D}{2\varepsilon} \log \left| \det c_q^2 \frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right| + O\left(\frac{1}{\varepsilon^2}\right). \end{aligned} \tag{3.18}$$

Here the last equality holds due to the fact that $\log(1+x) \sim x$ as $x \to 0$. In (3.18), the second term is just a constant with respect to $\hat{q}$, and thus of no relevance for the dynamics. Hence we found a zeroth order asymptotic expansion (up to constants) for $\mathfrak{V}$

$$\mathfrak{V}_0(\hat{q}) = V(\hat{q}, r(\hat{q})), \tag{3.19}$$

as well as a first order asymptotic expansion in ε

$$\mathfrak{V}_1(\hat{q}) = V(\hat{q}, r(\hat{q})) + \varepsilon^{-1} \cdot \frac{D}{2} \log \left| \det \frac{c_q^2}{D} \frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right|, \tag{3.20}$$

such that

$$\nabla_{\hat{q}} \mathfrak{V}(\hat{q}) \sim \nabla_{\hat{q}} \mathfrak{V}_0(\hat{q}) \tag{3.21}$$

and

$$\nabla_{\hat{q}}\mathfrak{V}(\hat{q}) \sim \nabla_{\hat{q}}\mathfrak{V}_1(\hat{q}). \tag{3.22}$$

Note that in the above derivation we neglected constants, hence $\mathfrak{V}_0$ respectively $\mathfrak{V}_1$ approximate $\mathfrak{V}$ only up to constants, which are irrelevant for the acting forces. If in the following we speak of $\mathfrak{V}_i$ approximating $\mathfrak{V}$, we always mean: "up to constants".

The above results mean in particular that the conditional expectation $\mathfrak{V}$ converges pointwise to $\mathfrak{V}_0$ as $\varepsilon \to \infty$. The question, whether the convergence is also uniform in $\hat{q}$, is covered in Subsection 3.2.3.

3.2.2 An Example of Three Atoms

In order to get an impression about the accuracy of the above asymptotic expansions, we consider the simple case of three silicon atoms, i.e. $n = 3$, where one is being averaged out, i.e. $m = 2$. Since the problem is translation invariant, we can without loss of generality set $q_1 = 0$. For this example the potential energy takes the form

$$V(q_2, q_3) = f(q_2) + f(q_3) + f(q_2 - q_3). \tag{3.23}$$

The optimal prediction potential (3.14) is (up to constants)

$$\mathfrak{V}(q_2) = f(q_2) - \frac{D}{\varepsilon} \log \left(\int_{q_3=q_2}^{L} e^{-\frac{\varepsilon}{D}(f(q_3)+f(q_2-q_3))} \, \mathrm{d}q_3 \right). \tag{3.24}$$

The zeroth order approximation is

$$\begin{aligned} \mathfrak{V}_0(q_2) &= f(q_2) + \min_{q_3 > q_2} \left(f(q_3) + f(q_2 - q_3) \right) \\ &= f(q_2) + \left(f(r_3(q_2)) + f(q_2 - r_3(q_2)) \right), \end{aligned} \tag{3.25}$$

where $r_3(q_2)$ is the minimizer of $(f(q_3) + f(q_2 - q_3))$. Finally, the first order approximation is

$$\begin{aligned} \mathfrak{V}_1(q_2) = f(q_2) &+ \left(f(r_3(q_2)) + f(q_2 - r_3(q_2)) \right) \\ &+ \frac{D}{2\varepsilon} \log \left(f''(r_3(q_2)) + f''(q_2 - r_3(q_2)) \right). \end{aligned} \tag{3.26}$$

All expressions $\mathfrak{V}$, $\mathfrak{V}_0$ and $\mathfrak{V}_1$ can be computed numerically for different values of ε. Note that the conditional expectation $\mathfrak{V}$ and the first order approximation $\mathfrak{V}_1$ depend on ε, while the zeroth order approximation $\mathfrak{V}_0$ is temperature-independent.

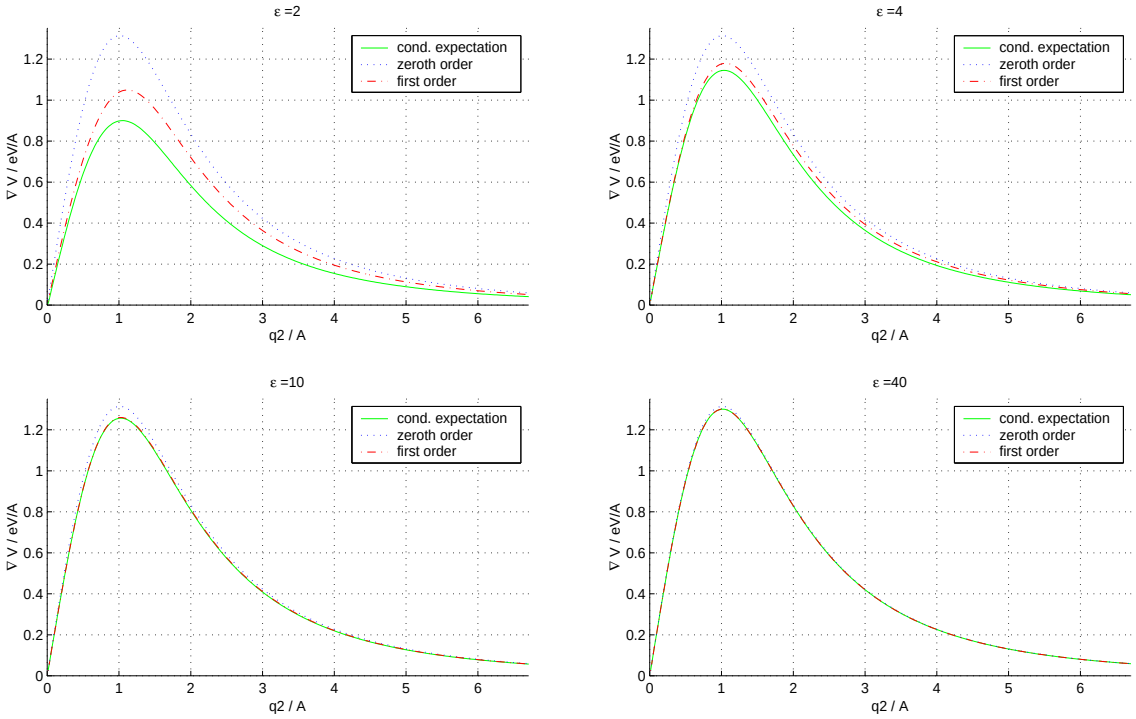


Figure 3.1: The asymptotic approximation for different ε

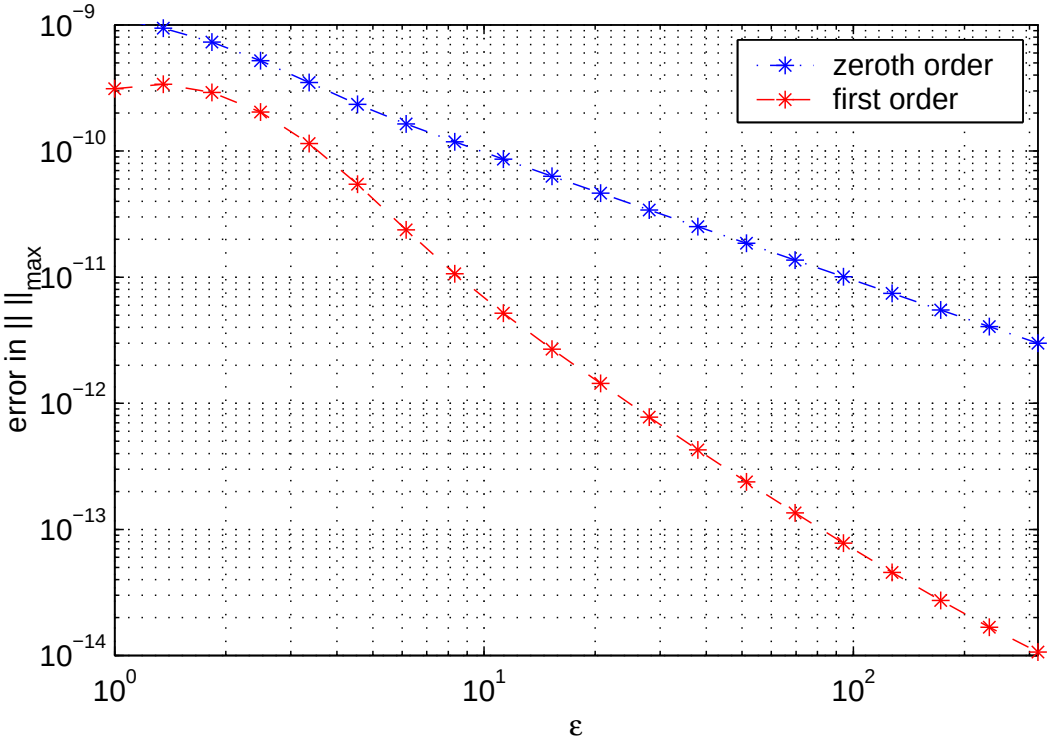


Figure 3.2: The decay of the approximation error depending on ε

Since the non-averaged term $f(q_2)$ is preserved in each expression, we consider the averaged contribution only

$$g(q_2) = \mathfrak{V}(q_2) - f(q_2) = -\frac{D}{\varepsilon}\log\left(\int_{q_3=q_2}^{L} e^{-\frac{\varepsilon}{D}(f(q_3)+f(q_2-q_3))}\,\mathrm{d}q_3\right) \tag{3.27}$$

$$g_0(q_2) = \mathfrak{V}_0(q_2) - f(q_2) = (f(r_3(q_2)) + f(q_2 - r_3(q_2)) \tag{3.28}$$

$$g_1(q_2) = \mathfrak{V}_1(q_2) - f(q_2) = g_0(q_2) + \frac{D}{2\varepsilon}\log\left(f''(r_3(q_2)) + f''(q_2 - r_3(q_2))\right. \tag{3.29}$$

respectively their derivatives $g'(q_2)$, $g_0'(q_2)$ and $g_1'(q_2)$.

Figure 3.1 shows the functions $g'(q_2)$, $g_0'(q_2)$ and $g_1'(q_2)$ for $\varepsilon \in \{2, 4, 10, 40\}$. Observe that already for $\varepsilon = 4$ the error in the first order approximation is fairly small, and for $\varepsilon = 40$ also the zeroth order approximation is almost identical to the conditional expectation. When considering that in the real problem one is confronted with values of $\varepsilon > 100$, these results give rise to the assumption that also for larger systems the error in the asymptotic approximations is negligibly small.

Figure 3.2 shows the order of convergence of the asymptotic approximations with respect to ε. Plotted is the error in the maximum norm on a log-log scale. The numerical results coincide perfectly with the analytical results: The error in the zeroth order approximation is of order $O(\varepsilon^{-1})$, and the error in the first order approximation is of order $O(\varepsilon^{-2})$.

In this thesis, we consider only the zeroth order approximation (3.19), since the Hessian $\frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ could not be included without problems into the equations which we will derive in the following sections. However, as the results in Figure 3.2 indicate, the first order approximation is by a significant factor more accurate. Hence, a next step in a further investigation should be to include the $O(\varepsilon^{-1})$-term into the formalism.

3.2.3 Properties of the Asymptotic Expansion

Some important remarks have to be given about the results of the derivation in Subsection 3.2.1:

1. The derivation in Subsection 3.2.1 was performed under the assumptions, that $\tilde{V}(\hat{q}, \tilde{q})$ has a unique global minimizer $r(\hat{q})$ with respect to $\tilde{q}$, and that the Hessian at this point $\frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ is regular. Both assumptions can be relaxed for the zeroth order expansion:

   - In the case of several minimizers $r_1(\hat{q}), \ldots, r_k(\hat{q})$, the first order asymptotic expansion changes to

$$\mathfrak{V}(\hat{q}) \sim V(\hat{q}, r_i(\hat{q})) + \varepsilon^{-1} \cdot D \cdot \log \sum_{i=1}^{k} \left| \det \frac{c_q^2}{D} \frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r_i(\hat{q})) \right|^{-\frac{1}{2}}. \tag{3.30}$$

Note that is is not important which minimizer $r_i(\hat{q})$ is inserted into $V(\hat{q}, r_i(\hat{q}))$.

- In the case of a singular Hessian, higher order derivatives have to be considered. For analytic potentials (as in our model problem) nonzero higher derivatives always exist. As in the above derivation, they also affect only the first order term in the asymptotic expansion (3.20).

2. In Subsection 3.2.1 we assumed the global minimizer inside the domain $M^L_{\hat{q}}$ to be unique. In the whole $\mathbb{R}^{n-m}$ the global minimizer is not unique, due to the possibility to exchange particles. As an example consider the case $n = 3$ and $m = 1$. The first atom is fixed at position $q_1 = 0$. Altogether, there are 6 global minimizers for (q_2, q_3), corresponding to the 6 possibilities to order 3 distinct numbers. Figure 3.3 shows the contour plot of the corresponding potential energy landscape. The 6 distinct minimizers are marked by red stars. The domain $M^L_{q_1}$ is the interior of the black triangle in the upper right corner. Note that inside this domain the minimizer is unique. This fact holds in a similar manner for higher dimensions for most "physical" potentials. We will explain in the following, what "physical" means in this context.

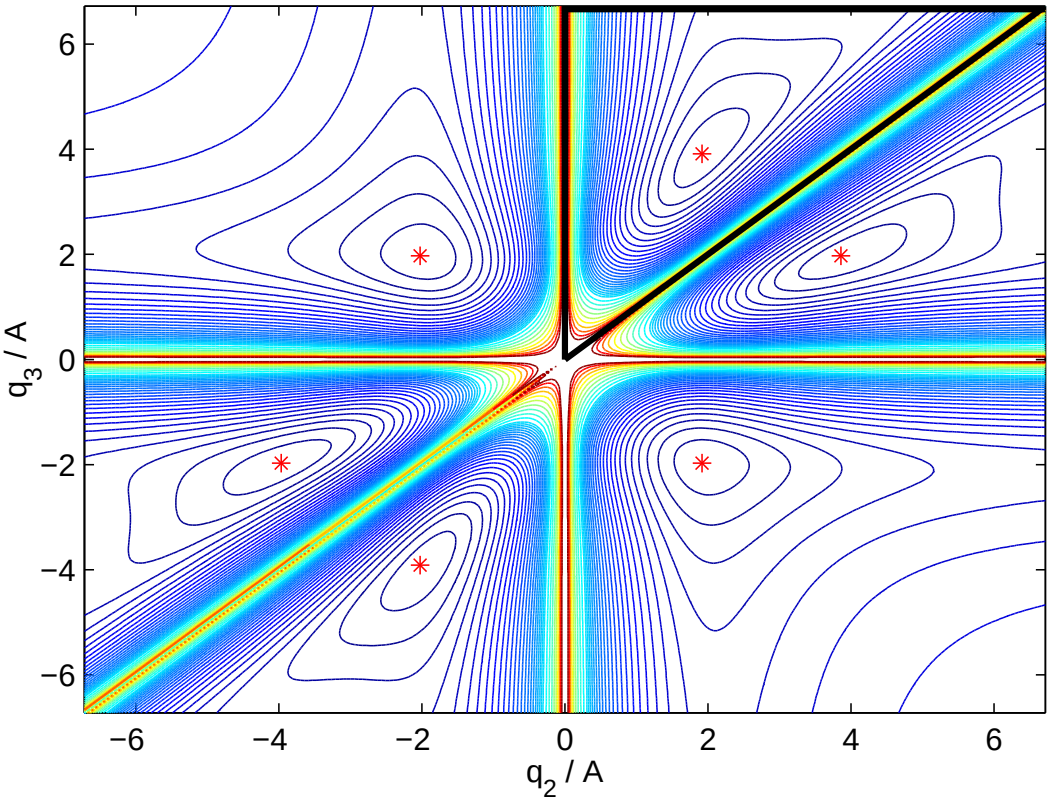


Figure 3.3: Six global minimizers for (q_2, q_3)

3. Note that one can construct potential functions which yield non-unique minimizers even inside the sets $M_{\hat{q}}^L$, but these will be "unphysical", i.e. they have features untypical for potentials which arise in molecular dynamics. Figure 3.4 shows an example of a potential which does not yield unique minimizers. Similarly, one can construct potentials, for which the Hessian of the minimizer inside the domain $M_{\hat{q}}^L$ is singular. However, for our model problem the Hessian turned out to be always regular.

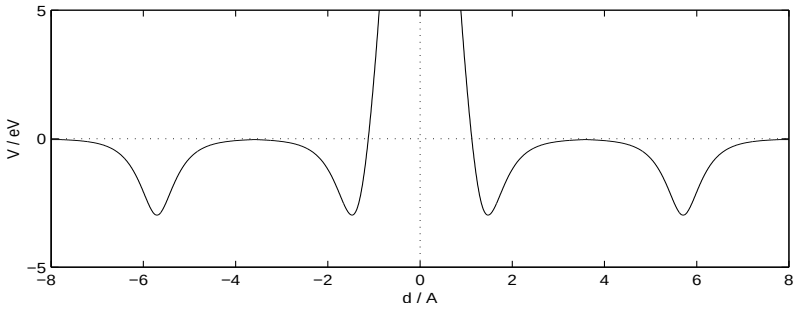


Figure 3.4: An "unphysical" potential which does not yield unique minimizers

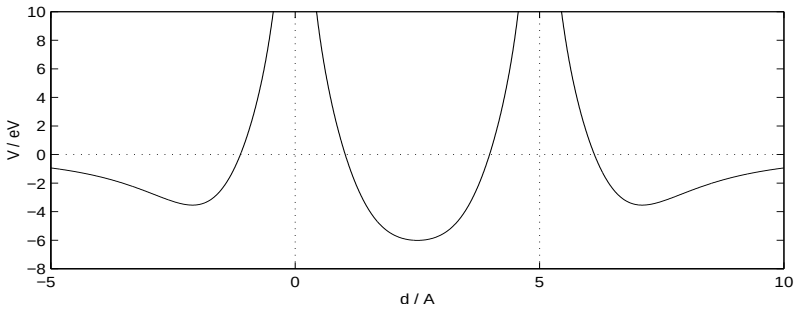


Figure 3.5: The minimizer inside $M_{\hat{q}}$ need not be the global minimizer

Note further that even if there is always a unique global minimizer inside the domain $M_{\hat{q}}^L$, this point will in general not be a global minimizer inside the whole $\mathbb{R}^{n-m}$. Consider the example from Subsection 3.2.2, i.e. the case $n = 3$ and $m = 2$. The first two atoms are at positions $q_1 = 0$Å and $q_2 = 5$Å. The effective potential of these two atoms if shown in Figure 3.5. The third atom q_3 has to be placed in the minimum. By definition, the domain $M_{\hat{q}}^L$ is the interval $(q_2, L]$, i.e. $q_3 > 5$Å. Note that inside the domain $M_{\hat{q}}^L$ the minimizer is indeed unique. However, it is not the global minimizer, which is located at position 2.5Å, i.e. outside of the domain $M_{\hat{q}}^L$. In Subsection 3.3.1 we will explain why it makes physically sense to choose the minimizer inside the domain $M_{\hat{q}}^L$, even if it is not a global minimizer.

4. Expressions (3.19) and (3.20) are correct asymptotic expansions to $\mathfrak{V}(\hat{q})$. But they were derived only as pointwise approximations with respect to $\hat{q}$. In particular

(3.19) was shown to be the pointwise limit of $\mathfrak{V}(\hat{q})$ as $\varepsilon \to \infty$

$$\forall \hat{q} \in \{(q_1, \ldots, q_m) | q_1 < \cdots < q_m\} : \lim_{\varepsilon \to \infty} \mathfrak{V}(\hat{q}) = \min_{\tilde{q} \in M^L_{\hat{q}}} V(\hat{q}, \tilde{q}). \tag{3.31}$$

Note that in general the convergence is not uniform in $\hat{q}$. Although it is not necessary to have a uniform convergence for the analysis presented in the following, it may very well be of interest for further work to know under which conditions $\mathfrak{V}$ converges uniformly to (3.19). In Subsection 3.2.1 we derived that up to constants the following relation holds as $\varepsilon \to \infty$:

$$\mathfrak{V}(\hat{q}) = V(\hat{q}, r(\hat{q})) + O\left(\varepsilon^{-1}\right). \tag{3.32}$$

Hence, for each $\hat{q} \in \hat{M}^L$, where

$$\hat{M}^L = \{(q_1, \ldots, q_m) \in [-L, L]^m | q_1 < \cdots < q_m\}, \tag{3.33}$$

one can find a $C^L(\hat{q})$, such that

$$|\mathfrak{V}(\hat{q}) - V(\hat{q}, r(\hat{q}))| \leq C^L(\hat{q}) \cdot \varepsilon^{-1}. \tag{3.34}$$

Given V is smooth enough and assuming the Hessian $\frac{\partial^2 \tilde{V}}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ is regular for any $\hat{q}$, one can find constants $C^L(\hat{q})$, which depend continuously on $\hat{q}$ (as in [36]). For each L the domain $\hat{M}^L$ is compact, hence the maximum

$$C^L = \max_{\hat{q} \in \hat{M}^L} C^L(\hat{q}) \tag{3.35}$$

exists. If the potential energy between two atoms $f(d)$ vanishes as $d \to \infty$, then the maxima C^L stay bounded as $L \to \infty$. Hence, under the given assumptions, the convergence of $\mathfrak{V}$ to (3.19) can be expected to be even uniform.

In other words, the zeroth order asymptotic expansion (3.19) is the zero temperature limit to the optimal prediction Hamiltonian (3.14). In the following section we will provide a physical interpretation for this limit and derive a new system of equations from it.

3.3 Zero Temperature Limit

We found that the zeroth order asymptotic expansion to $\mathfrak{V}(\hat{q})$

$$\mathfrak{V}_0(\hat{q}) = \min_{\tilde{q} \in M^L_{\hat{q}}} V(\hat{q}, \tilde{q}), \tag{3.36}$$

is the uniform limit for $\varepsilon \to \infty$, i.e. $T \to 0$. Hence we call $\mathfrak{V}_0$ the *zero temperature limit potential*. Of course we are not interested in molecular dynamics at temperature zero,

since this would mean all atoms standing still in their potential minimum. However, as reasoned before, our dynamics is taking place at at low temperatures, so one can expect the correct optimal prediction potential function

$$\mathfrak{V}(\hat{q}) = -\frac{1}{\beta} \log \left(\frac{1}{c_{\hat{q}}^{n-m}} \int_{M_{\hat{q}}^L} e^{-\beta V(\hat{q},\tilde{q})} \, \mathrm{d}\tilde{q} \right) \tag{3.37}$$

to be close to the zero temperature limit potential $\mathfrak{V}_0(\hat{q})$ for small but positive temperatures (the results in Subsection 3.2.2 support this expectation). Hence the idea is to run the low temperature optimal prediction dynamics, which would be correctly described by $\mathfrak{V}$, with the zero temperature limit dynamics given by $\mathfrak{V}_0$.

3.3.1 Physical Interpretation

By going over from $\mathfrak{V}$ to $\mathfrak{V}_0$ we have formally replaced an $(n-m)$-dimensional integration by an $(n-m)$-dimensional minimization problem. At first glance this is no real improvement, since high-dimensional minimization is also computationally expensive, especially since we are searching for a global minimum of a function, which may eventually have many local minima. A method for solving such problems is e.g. *simulated annealing* [28], but it is in general still too expensive to yield a speed-up compared to solving the original system.

To remedy this problem one can use a physical interpretation of (3.37):

> Given m atoms arbitrarily. Now place $n-m$ further atoms in such a way, such that the total potential energy is minimized.

Since $\mathfrak{V}_0$ is formally just m-dimensional, we call the $n-m$ atoms from the above physical interpretation *virtual particles*.

Now, one can see why it was important to restrict the domains of integration to the sets $M_{\hat{q}}^L$, as we did in Subsection 3.1.1. Having the whole $\mathbb{R}^{n-m}$ as domain for $\tilde{q}$ would mean that the $n-m$ virtual particles can be left, right and in between the m real particles (the real particles would "swim" in a sea of virtual particles). But this would violate the third assumption about the applicability of optimal prediction given in the introduction to Chapter 2, namely that the system must allow a clear separation of particles into ones of interest and ones to be averaged away, and this separation must stay valid over time. On the other hand, having $\tilde{q}$ restricted to the set $M_{\hat{q}}^L$, forces the virtual particles to be only *right* of the m real particles (without loss of generality we always assume the real particles to be on the left and the virtual particles to be on the right), which exactly satisfies the above assumption. Additionally, by giving the silicon particles – both real and virtual

– an order, we made the global minimizer unique for most "physical" potentials, so the high-dimensional minimization problem becomes less threatening.

In the following section we will derive a differential equation for the minimizer, which allows to follow its position over time, instead of computing the minimum over and over again. Furthermore we will see, that in general one does not have to compute $n - m$ virtual particles. Instead one has to follow just the first 5 to 10 of them. This will yield a *boundary layer condition* for the optimal prediction dynamics.

3.3.2 Equations of Motion

The zero temperature limit potential for the optimal prediction potential is

$$\mathfrak{V}_0(\hat{q}) = \min_{\tilde{q} \in M^L_{\hat{q}}} V(\hat{q}, \tilde{q}) = V(\hat{q}, r(\hat{q})), \tag{3.38}$$

where $r(\hat{q})$ is the minimizer, which we assume to be unique. In terms of the physical interpretation from the previous section, $r(\hat{q})$ is the position vector of the $n - m$ virtual particles, which sit in the potential minimum. In the following we assume $V(\hat{q}, r)$ and $r(\hat{q})$ to be of class C^1. This allows to compute the equations of motion corresponding to the zero temperature limit Hamiltonian

$$\mathfrak{H}_0(\hat{q}, \hat{p}) = \mathfrak{T}_0(\hat{p}) + \mathfrak{V}_0(\hat{q}) = \frac{1}{2} \sum_{i=1}^{m} \frac{p_i^2}{m_i} + V(\hat{q}, r(\hat{q})). \tag{3.39}$$

The change of position is simply

$$\frac{\partial \mathfrak{H}_0}{\partial p_i}(\hat{q}, \hat{p}) = \frac{\partial \mathfrak{T}_0}{\partial p_i}(\hat{p}) = \frac{p_i}{m_i} \qquad \forall i = 1, \ldots, m. \tag{3.40}$$

The change of momentum, i.e. the acting force, is given by

$$\frac{\partial \mathfrak{H}_0}{\partial \hat{q}}(\hat{q}, \hat{p}) = \frac{\partial \mathfrak{V}_0}{\partial \hat{q}}(\hat{q}) = \frac{\partial V}{\partial \hat{q}}(\hat{q}, r(\hat{q})) + \underbrace{\frac{\partial V}{\partial r}(\hat{q}, r(\hat{q}))}_{=0} \cdot \frac{dr}{d\hat{q}}(\hat{q}) = \frac{\partial V}{\partial \hat{q}}(\hat{q}, r(\hat{q})), \tag{3.41}$$

i.e. the force acting on one of real particles is just the sum of the forces from the other $m - 1$ real and the $n - m$ virtual particles. Note that $\frac{\partial V}{\partial r}(\hat{q}, r(\hat{q}))$ is zero, since $r(\hat{q})$ is the minimizer of $V(\hat{q}, r(\hat{q}))$.

Formally, the equations of motion (3.40) and (3.41) describe a $2m$-dimensional system, instead of a $2n$-dimensional system. This was our goal right from the start, but with the aim to speed up our computation. Solving an $(n - m)$-dimensional minimization problem (in order to place the virtual particles) is too expensive and thus not going to help in achieving this aim. So in the following we derive equations of motion for the

virtual particles, too. Note that since the virtual particles always sit in the minimum of the potential energy, they do not have inertia as the m real particles have. Hence, no momenta are required to describe the virtual particles' movement.

To obtain an evolution equation for r we first have to find a way to compute $\frac{\partial r}{\partial \hat{q}}(\hat{q})$, i.e. how the *virtual* particles have to be moved to stay in the minimum, once the first m particles are being moved. We define

$$v(\hat{q}) := \frac{\partial V}{\partial \tilde{q}}(\hat{q}, r(\hat{q})). \tag{3.42}$$

Since $r(\hat{q})$ is always chosen to minimize V, we have that $v(\hat{q}) = 0 \; \forall \hat{q}$. Thus

$$0 = \frac{\partial v}{\partial \hat{q}}(\hat{q}) = \frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q})) + \frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \cdot \frac{\partial r}{\partial \hat{q}}(\hat{q}). \tag{3.43}$$

Here the Hessian $\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ is an $(n-m) \times (n-m)$ matrix, and $\frac{\partial r}{\partial \hat{q}}(\hat{q})$ and $\frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q}))$ are both $(n-m) \times m$ matrices. Solving for $\frac{\partial r}{\partial \hat{q}}(\hat{q})$ yields

$$\frac{\partial r}{\partial \hat{q}}(\hat{q}) = -\left(\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right)^{-1} \cdot \frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q})). \tag{3.44}$$

Now we can set up a relation for the time derivative of the minimizer $r(\hat{q})$

$$\frac{d}{dt} r(\hat{q}) = \frac{\partial r}{\partial \hat{q}}(\hat{q}) \cdot \frac{d\hat{q}}{dt} = -\left(\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right)^{-1} \cdot \frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q})) \cdot \mathfrak{M}^{-1} \cdot \hat{p} \tag{3.45}$$

The matrix $\mathfrak{M}$ in (3.45) is a diagonal matrix containing the masses m_i of the atoms

$$(\mathfrak{M})_{ii} = m_i. \tag{3.46}$$

Altogether the zero temperature limit optimal prediction system of differential equations is given by

$$\begin{aligned} \frac{d}{dt}\hat{q} &= \mathfrak{M}^{-1} \cdot \hat{p} \\ \frac{d}{dt}\hat{p} &= -\frac{\partial V}{\partial \hat{q}}(\hat{q}, r(\hat{q})) \\ \frac{d}{dt} r &= -\left(\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \right)^{-1} \cdot \frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q})) \cdot \mathfrak{M}^{-1} \cdot \hat{p}, \end{aligned} \tag{3.47}$$

with initial conditions

$$\begin{aligned} \hat{q}_i(0) &= q_i(0) \qquad \forall i = 1, \ldots, m \\ \hat{p}_i(0) &= p_i(0) \qquad \forall i = 1, \ldots, m \\ r(0) &\qquad \text{s.t. } V(\hat{q}(0), r(0)) \text{ is minimal.} \end{aligned} \tag{3.48}$$

The minimization problem to find the initial value $r(0)$ can be solved by a few Newton steps

$$r^{(k+1)}(\hat{q}) = r^{(k)}(\hat{q}) - \left(\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r^{(k)}(\hat{q}))\right)^{-1} \cdot \frac{\partial V}{\partial \tilde{q}}(\hat{q}, r^{(k)}(\hat{q})). \tag{3.49}$$

As initial value for the integration one can choose an equidistant alignment of particles at equilibrium distance d_0. Numerically, it may pay off to solve the minimization problem once in a while during the computation, too, to eliminate integration and round-off errors.

Since the virtual particles have no momentum, system (3.47) is just $(n+m)$-dimensional, instead of $2n$-dimensional. It is a closed system of ordinary differential equations, and the right hand side requires no integration or minimization problem to be solved. Still, there is an $(n-m)$-dimensional linear system of equations to be solved, which is also expensive, unless the matrix $\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ has a special structure or the number of virtual particles can be reduced. Note that so far we have not made any use of the special structure of the potential energy (2.1) in our problem. We will employ this special structure in the following section and show that it allows further reduction of the size of the system and the computational effort.

3.3.3 Equations of Motion in the Model Problem

For the potential energy in the model problem (2.1), the matrices in (3.47) have a special structure

$$\forall i : \left(\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))\right)_{(i,i)} = \sum_{j=1}^{m} f_\alpha''(r_i - q_j) + \sum_{\substack{j=m+1 \\ j\neq i}}^{n} f_1''(r_i - r_j) \tag{3.50}$$

$$\forall i \neq j : \left(\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))\right)_{(i,j)} = -f_1''(r_i - r_j) \tag{3.51}$$

and

$$\frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q})) = \begin{pmatrix} f_\alpha''(r_{m+1} - q_1) & \dots & f_\alpha''(r_{m+1} - q_m) \\ \vdots & \ddots & \vdots \\ f_\alpha''(r_n - q_1) & \dots & f_\alpha''(r_n - q_m) \end{pmatrix}. \tag{3.52}$$

Here $\alpha = 1$, if the particle q_j is a silicon atom, and $\alpha = 2$, if it is a copper atom. As in Subsection 3.1.1 we assume to have the copper atom(s) far away from the border between real and virtual particles, and we further assume to compute only for so long, as this assumption is still satisfied. In this context, "far away" means that the potential $f(q_i - q_j)$ between the two atoms is negligibly small. Typically this is the case at a distance greater than 5 to 10 times the equilibrium distance. So the virtual particles will "see" only silicon atoms.

Since at low temperatures the silicon atoms more or less just oscillate around their equilibrium positions in the crystal structure, one can assume the silicon atoms in the inside

of the crystal to be aligned almost equidistantly. Employing the fact that k equilibrium distances away from each other the potential is nearly zero, the matrices in (3.47) have a special structure:

- The Hessian $\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ is almost a diagonal band matrix with band width k.
- The matrix $\frac{\partial^2 V}{\partial \hat{q} \partial \tilde{q}}(\hat{q}, r(\hat{q}))$ is almost an upper triangular matrix with width k.

Substituting the special structure of these matrices into the equation of motion for the positions of the virtual particles r yields that only the first k virtual particles $r_{m+1}, \ldots, r_{m+k}$ have to be considered. The others will align equidistantly right to the first k ones. This assumption is valid if effects at the right boundary of the crystal can be neglected. This is the case if $n - m$ is large ($n - m > 10$), which is exactly the case which we are interested in, since then shrinking the system from $2n$ to $2m$ dimensions promises a great computational speed-up.

3.3.4 A Boundary Layer Condition

Assume that the potential between two silicon atoms f_1 is negligibly small at a distance $k \cdot d_0$, where d_0 is the equilibrium distance of silicon atoms inside the crystal. In this context "equilibrium" means that all forces from either side cancel out, thus d_0 is given by the relation

$$\sum_{i=1}^{\infty} f_1'(i \cdot d_0) = 0. \tag{3.53}$$

Typically one is interested in cases, where the number of virtual particles $n - m$ is much larger than k, even larger than $2k$. As reasoned in the previous section, in such a case only the first k virtual particles have to be computed, while the others will align equidistantly right to the first k ones. Since the potential reaches only over k atoms, it is sufficient to consider only $2k$ virtual particles, where the last k ones are aligned equidistantly. Hence, the position vector for the $2k$ virtual particles is

$$r = \begin{pmatrix} r_{m+1} \\ \vdots \\ r_{m+k} \\ r_{m+k} + d_0 \\ \vdots \\ r_{m+k} + kd_0 \end{pmatrix}, \tag{3.54}$$

and its time derivative

$$\frac{d}{dt}r = \begin{pmatrix} \dot{r}_{m+1} \\ \vdots \\ \dot{r}_{m+k} \\ \dot{r}_{m+k} \\ \vdots \\ \dot{r}_{m+k} \end{pmatrix}. \tag{3.55}$$

Also the matrices $\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ and $\frac{\partial^2 V}{\partial \hat{q}\partial \tilde{q}}(\hat{q}, r(\hat{q}))$ can be approximated by matrices with a special structure, as described in the previous section. We obtain

$$\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q})) \approx \left(\begin{array}{ccc|ccc} a_{1,1} & \dots & a_{1,k} & a_{1,k+1} & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ a_{k,1} & \dots & a_{k,k} & a_{k,k+1} & \dots & a_{k,2k} \\ \hline a_{k+1,1} & \dots & a_{k+1,k} & a_{k+1,k+1} & \dots & a_{k+1,2k} \\ 0 & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{2k,k} & a_{2k,k+1} & \dots & a_{2k,2k} \end{array}\right), \tag{3.56}$$

where each of the four blocks is of size $k \times k$. Similarly

$$\frac{\partial^2 V}{\partial \hat{q}\partial \tilde{q}}(\hat{q}, r(\hat{q})) \approx \left(\begin{array}{ccc|ccc} \multicolumn{3}{c|}{\overbrace{}^{m-k}} & \multicolumn{3}{c}{\overbrace{\phantom{b_{1,m-k+1}\ \dots\ b_{1,m}}}^{k}} \\ 0 & \dots & 0 & b_{1,m-k+1} & \dots & b_{1,m} \\ \vdots & \ddots & \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & b_{k,m} \\ \hline 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{array}\right) \begin{array}{l} \left.\phantom{\begin{array}{c}0\\0\\0\end{array}}\right\}_{k} \\ \left.\phantom{\begin{array}{c}0\\0\\0\end{array}}\right\}_{n-m-k} \end{array} \tag{3.57}$$

The sizes of the blocks can be seen in (3.57). The zero entries in both approximate matrices (3.56) and (3.57) are not exactly zero in $\frac{\partial^2 V}{\partial \tilde{q}^2}(\hat{q}, r(\hat{q}))$ and $\frac{\partial^2 V}{\partial \hat{q}\partial \tilde{q}}(\hat{q}, r(\hat{q}))$, but negligibly small. Hence, we can expect to have a good approximation.

Substituting these special vectors and matrices into equation (3.45) yields the relation

$$\left(\begin{array}{ccc|ccc} a_{1,1} & \dots & a_{1,k} & a_{1,k+1} & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ a_{k,1} & \dots & a_{k,k} & a_{k,k+1} & \dots & a_{k,2k} \\ \hline a_{k+1,1} & \dots & a_{k+1,k} & a_{k+1,k+1} & \dots & a_{k+1,2k} \\ 0 & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{2k,k} & a_{2k,k+1} & \dots & a_{2k,2k} \end{array}\right) \cdot \left(\begin{array}{c} \dot{r}_{m+1} \\ \vdots \\ \dot{r}_{m+k} \\ \hline \dot{r}_{m+k} \\ \vdots \\ \dot{r}_{m+k} \end{array}\right)$$
$$= \left(\begin{array}{ccc|ccc} 0 & \dots & 0 & b_{1,m-k+1} & \dots & b_{1,m} \\ \vdots & \ddots & \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & b_{k,m} \\ \hline 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{array}\right) \cdot \left(\begin{array}{c} p_1 \\ \vdots \\ p_{m-k} \\ \hline p_{m-k+1} \\ \vdots \\ p_m \end{array}\right), \tag{3.58}$$

which implies the following k-dimensional system for the first k virtual particles

$$\left(\begin{array}{ccc} a_{1,1} & \dots & a_{1,k-1} & \sum_{j=k}^{k+1} a_{1,j} \\ \vdots & \ddots & \vdots & \vdots \\ a_{k,1} & \dots & a_{k,k-1} & \sum_{j=k}^{2k} a_{k,j} \end{array}\right) \cdot \left(\begin{array}{c} \dot{r}_{m+1} \\ \vdots \\ \dot{r}_{m+k} \end{array}\right)$$
$$= \left(\begin{array}{ccc} b_{1,m-k+1} & \dots & b_{1,m} \\ 0 & \ddots & \vdots \\ 0 & 0 & b_{k,m} \end{array}\right) \cdot \left(\begin{array}{c} p_{m-k+1} \\ \vdots \\ p_m \end{array}\right). \tag{3.59}$$

This relation can be interpreted as a boundary layer condition which acts as if the crystal of silicon atoms was continued to infinity, although it is actually cut off after the m-th particle.

With the above modifications the zero temperature limit optimal prediction system (3.47) becomes a $(2m + k)$-dimensional system of equations. Since for typical potentials k is about 5 to 10, solving the linear system (3.59) should be of minor effort. Hence it is reasonable to expect that system (3.47) with the modified relation for the virtual particles is significantly cheaper to compute than the original, where the speed-up factor of course depends on the values n, m and k. In Subsection 3.4.2 we will compare the two systems with respect to computational effort.

The main question, however, is if the system we derived above yields the same dynamics as the original system. This is not clear at all, since several approximations have been done in achieving the smaller system, and each approximation had specific assumptions about the system, which also have to be checked:

	system	dimension	assumptions
	original system	$2n$	system in equilibrium
$\rightsquigarrow$	optimal prediction at $T > 0$	$2m \left[+ (n-m) \right]$	low temperature
$\rightsquigarrow$	zero temperature limit	$n + m$	potentials short ranged
$\rightsquigarrow$	approximative system	$2m + k$	

Since trajectories have no meaning in statistical mechanics (as the initial positions and momenta are not known anyhow), the new system has to be compared to the original system by other means. These are statistical quantities, e.g. the diffusion constant of one copper atom inside the silicon crystal, or fluctuations of the energy of the first m atoms. In Chapter 4 we will compare the new system to the original system and investigate under which conditions the new system reflects the "correct" dynamics, and under which conditions it does not.

It has to be pointed out that the ideas, which led from the conditional expectations to reasonable (approximate) equations of motions, apply in principle to much more general cases than the specific model problem here. In particular, the given asymptotic approximation for low temperatures may find an application in various problems which are in low temperature equilibrium. Also we have only used the zeroth order asymptotic expansion in the derivation of equations of motion. Using the first order expansion may yield even better results, and should be one important aspect in future investigation on this subject.

3.4 Numerical Simulation

In this section we provide example computations of the optimal prediction system and show a warning example, in which optimal prediction cannot be used successfully together with another approximation, which works satisfyingly well for the original system (cutting off the potentials). In Subsection 3.4.2 we investigate to which extent optimal prediction yields a computational speed-up.

3.4.1 Examples

The optimal prediction system (3.47), with or without the approximate equation for the virtual particles (3.59), is solved under the same parameters and using the same method as the original system, as it was described in Section 2.4, i.e. the integration is done over a time of $1.0 \cdot 10^{-11}$s using Runge-Kutta 4 with time step $2.0 \cdot 10^{-15}$s, and the temperature is $T = 7000$K.

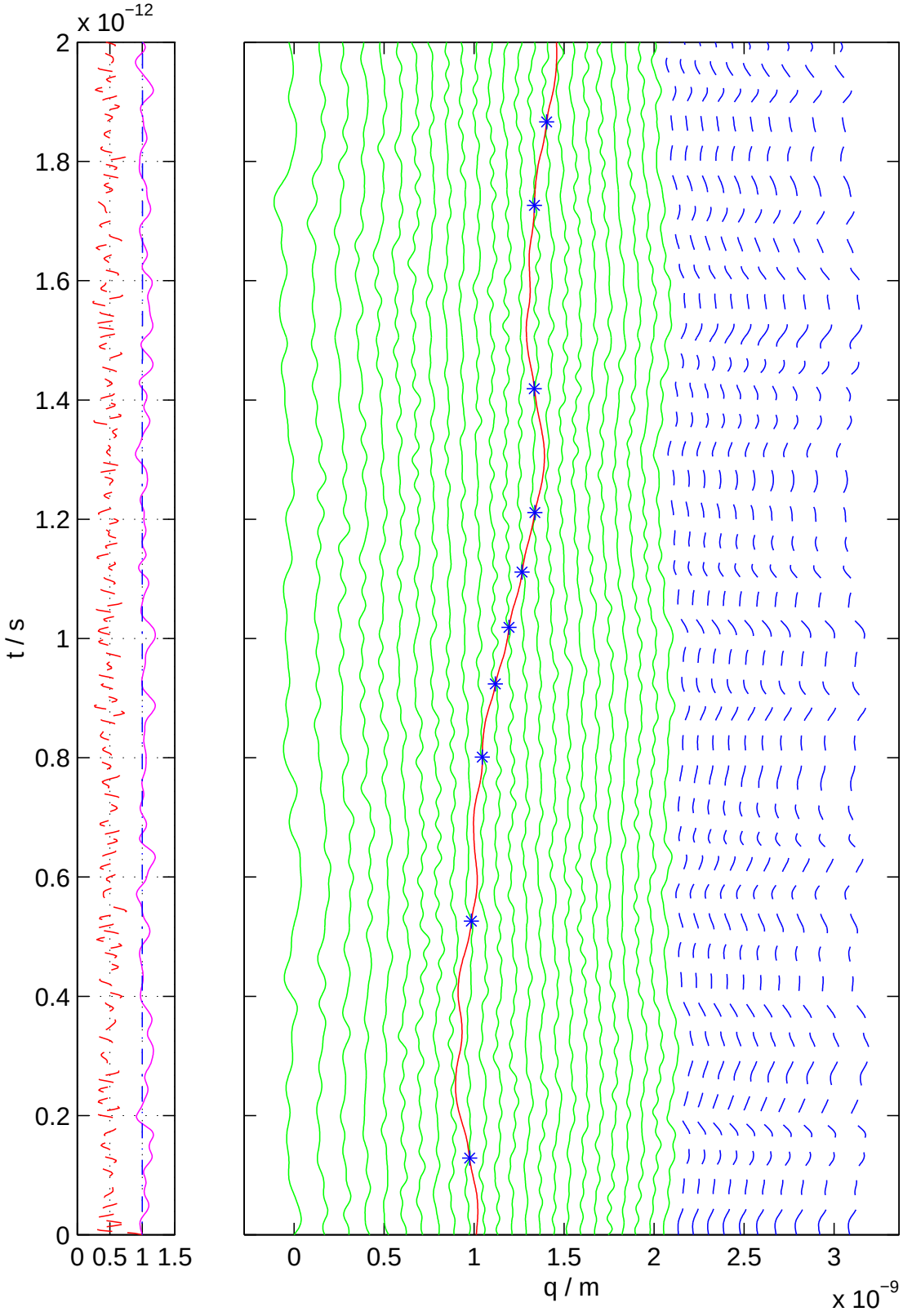


Figure 3.6: A simulation of the optimal prediction system

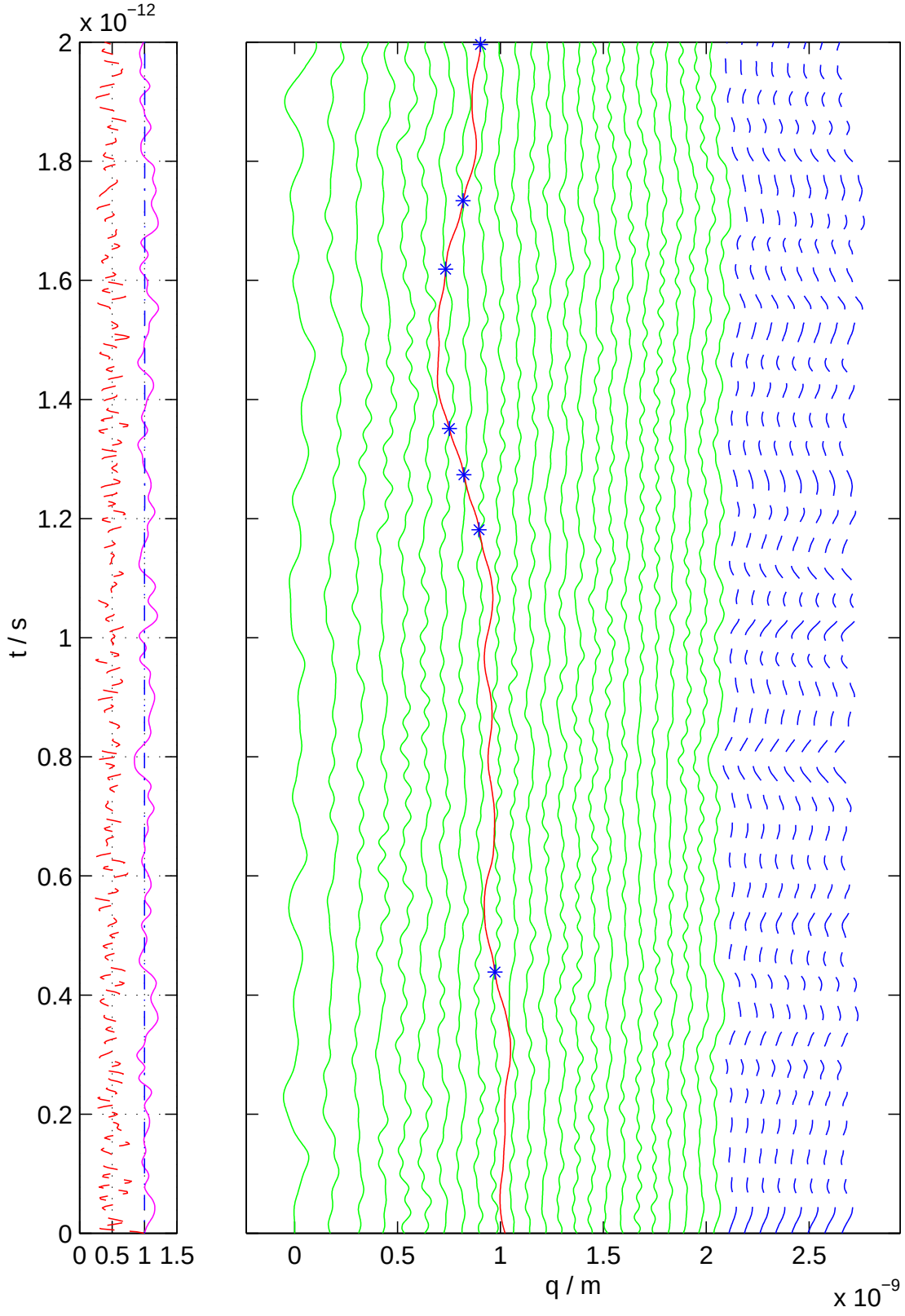


Figure 3.7: A simulation of the optimal prediction approximation for an infinite crystal

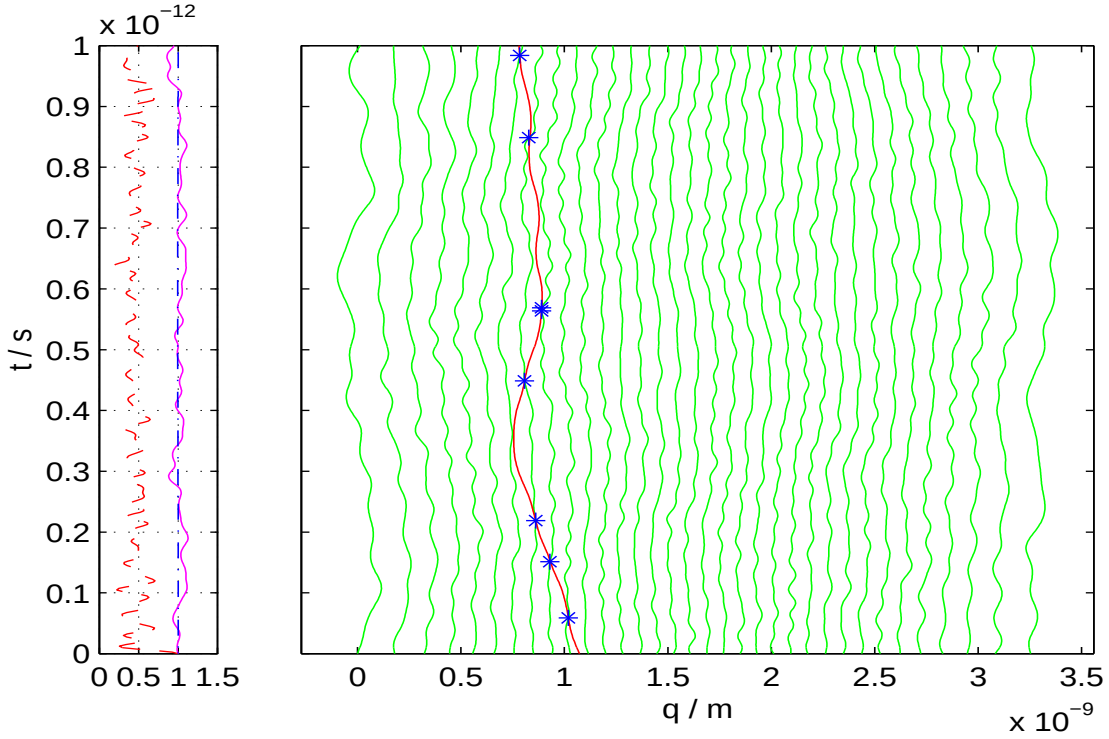


Figure 3.8: The original problem with cut-off potentials

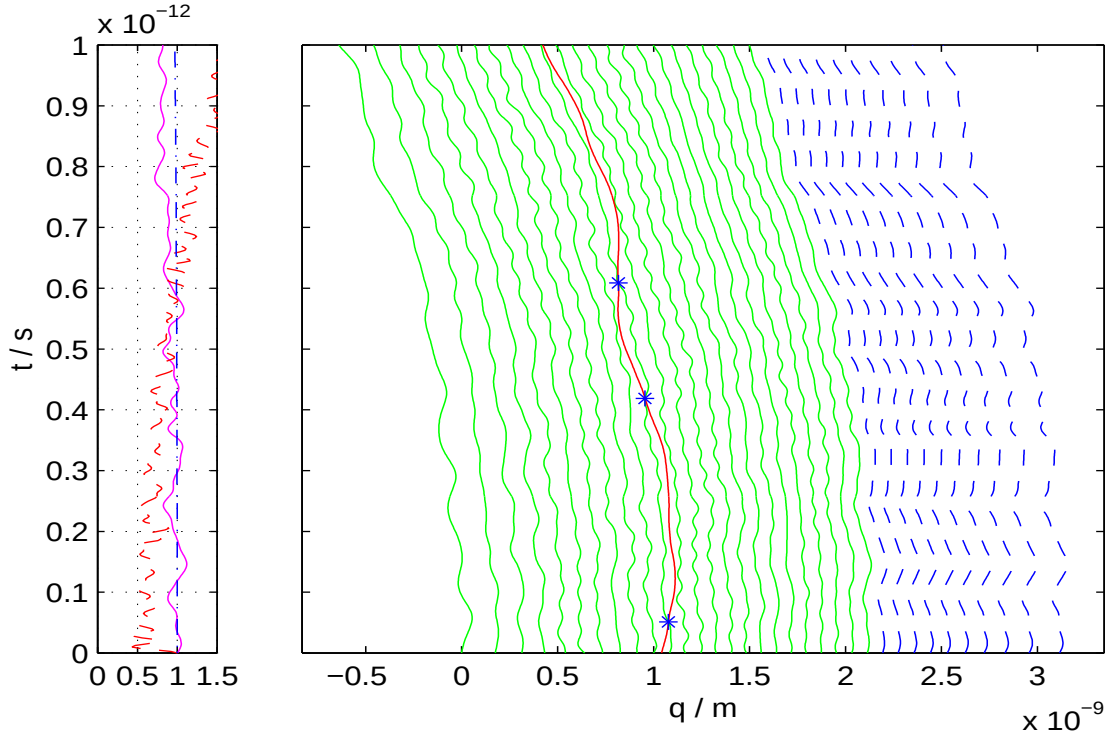


Figure 3.9: The optimal prediction approximation with cut-off potentials

Figure 3.6 shows the numerical results for an example computation of the zero temperature limit optimal prediction system (3.47) for 24 silicon and one copper atom (i.e. $m = 25$). Since the original system contained $n = 35$ atoms, 10 virtual silicon atoms appear in the computation.

In the right subplot the positions $q_i(t)$ of the 25 atoms and the 10 virtual atoms are plotted over time. One can observe that the real atoms oscillate freely, while the virtual atoms more or less follow the motion of the rightermost real atoms. An important aspect is that the surface tension effect on the right side of the crystal affects now the virtual atoms, while the rightermost real atoms are aligned equidistantly.

Obviously in computations with the optimal prediction system care has to be taken that no copper atom reaches into the virtual atoms, as this would destroy the third assumption about the applicability of optimal prediction, which is given in the Introduction to Chapter 2.

As in Section 2.4 for the original system, the three quantities

1. total energy over initial total energy (dash-dotted blue curve)
2. kinetic energy of the first $m = 25$ atoms $T_{\text{left}}(p) = \frac{1}{2}\sum_{i=1}^{m} \frac{p_i^2}{m_i}$ over its initial value (dashed red curve)
3. total energy of the $m = 25$ real atoms (given by Formula (2.11)) over its initial value, stretched by a factor of 5 (solid magenta curve)

for the optimal prediction approximation are shown in the left subplot. All three functions show the same features as observed in Section 2.4 for the original system. As the total energy of the $m = 25$ real atoms neglects the contribution of the virtual atoms to the potential energy, it is not constant over time also for the optimal prediction system.

Figure 3.7 shows the numerical results for an example computation of the zero temperature limit optimal prediction system, but with the boundary layer condition corresponding to an infinitely extended silicon crystal, as described in Subsection 3.3.4. In other words, the effect of surface tension in the virtual atoms is neglected. We solve system (3.59) for 8 virtual atoms. Also this simulation reflects the behavior of the original system with respect to the three quantities shown in the left subplot.

Since computational effort is not a problem for a single computation with 35 particles, we do not cut-off the potentials between two atoms (as described in Section 2.4) in the numerical simulations in this section. For larger systems, however, such an approximation may be necessary. It has to be pointed out that care has to be taken, when using other approximations together with optimal prediction.

Figure 3.8 and Figure 3.9 show the numerical solution to the problem described in Section 2.4 up to time $1.0 \cdot 10^{-12}$s. Unlike the computations shown in Figure 2.5 and Figure 3.6, the potentials between two atoms were cut off, such that they reach only over 10 atoms in each direction. Figure 3.8 shows a solution to the original problem, while Figure 3.9 shows a solution to the optimal prediction system with 10 virtual atoms. One can observe that the simplification in the potentials works fine for the original problem, as errors cancel out due to the spatial symmetry of the crystal. For the optimal prediction approximation, however, the simplification is inappropriate, as the whole crystal is apparently being accelerated in the direction away from the virtual particles. The reason for this effect lies in the asymmetry of the optimal prediction system. Cutting off the potentials yields a small additional force acting on each of the real particles. In the original system, these small forces cancel out, but in the optimal prediction system, the virtual atoms "push" the whole crystal, as they always align in the potential minimum.

This observation may serve as a warning example when using other approximative methods together with the optimal prediction approximation. In particular, optimal prediction could in principle be used together with averaging methods or multipole methods, as mentioned in the Introduction. This example indicates that a careful investigation of any approximation errors involved may be necessary.

Of course the example computations in this section only give a hint that the optimal prediction approximation may reflect the dynamics of the original system well. A rigorous investigation of this issue will be done in Chapter 4, when we compare statistical quantities as diffusion constants and energy fluctuations numerically by Monte-Carlo sampling.

3.4.2 Computational Speed-Up

Although the result that optimal prediction yields a boundary layer condition, which simulates a crystal continued to infinity, is also of theoretical interest, the actual intention was to use optimal prediction as a method to reduce the computational effort. In Section 3.3, we derived two versions of optimal prediction for a system of n atoms, which shall be reduced to m atoms:

- The original version, as described in Subsection 3.3.2, which considers $n-m$ virtual atoms.
- The boundary layer condition version, as described in Subsection 3.3.4, which considers only k virtual atoms.

In principle, one can achieve arbitrarily high speed-up factors with the boundary layer condition system by keeping m and k fixed and increasing n. But in most cases such a consideration does not make much sense, since the particular values for n and m typically

depend on the context in which the problem arises. Usually the original size of crystal n is given a priori, and the first step in applying optimal prediction is to choose an appropriate value for m. The decision, which value for m works best for a particular problem, is influenced by two factors:

- The required accuracy, i.e. the question, how well the important properties of the system are preserved. In general one has: The larger m, the better.
- The reduction of computational effort. Here one has in general: The smaller m, the better.

The question of accuracy will be considered in Chapter 4, in particular will criteria be provided to investigate how well important properties are preserved. The question of computational effort will be dealt with in the following.

We investigate the computational speed-up by comparing the CPU time for computations of the optimal prediction system with the CPU time for the corresponding computations of the two versions of the original system. The comparison is carried out in dependence of the sizes of the systems, given by the numbers n and m. Since the original version of optimal prediction does only replace real particles by virtual ones, while on the other hand the boundary layer condition version allows to really omit atoms, a real speed-up can only be expected from the boundary layer condition version.

The computations were done on a network of AMD Athlon-6 1.4 GHz computers. The CPU times are given in seconds, but the absolute values are not important, since we are only interested in speed-up factors. The computations with the boundary layer condition version of optimal prediction were done with $k = 10$ virtual atoms.

Figure 3.10 shows the CPU times for the original system and the optimal prediction system, first in the original version, then in the boundary layer condition version, all in dependence of n and m, where of course only $m < n$ makes sense. Note that for a fixed n the CPU times for the original system are of course independent of m, while in the boundary layer condition version the computational effort is independent of n. The results show that optimal prediction in its original version is *never* cheaper to solve than the original system. Apparently, for our model problem setting up and solving the linear system (3.45) is more expensive than computing the full system of equations, especially for $n - m$ large. On the other hand, for the boundary layer condition version the effort decays significantly with m.

Figure 3.11 shows the actual speed-up factor with respect to the original system for the two versions of optimal prediction. While the original version of optimal prediction does not yield any acceleration (the speed-up factors are less than 1), the boundary layer condition version yields high speed-up factors for small m. While this looks promising, one must not forget that it depends on the particular context, how small m can be chosen.

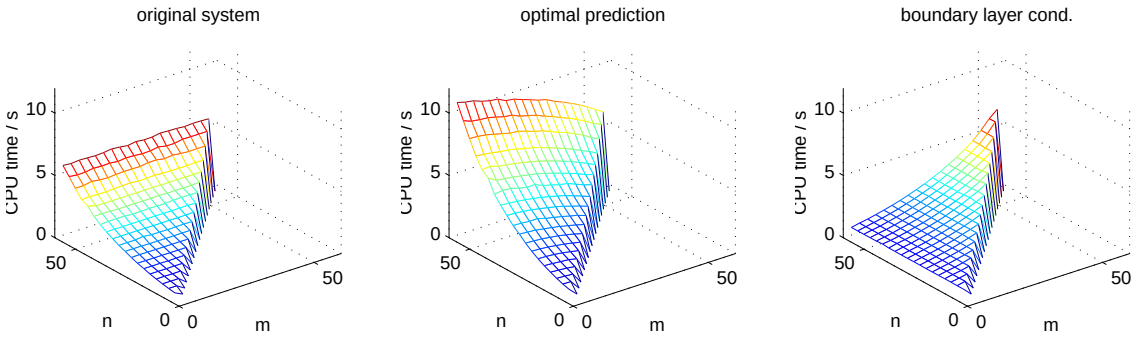


Figure 3.10: CPU times for the three systems depending on n and m

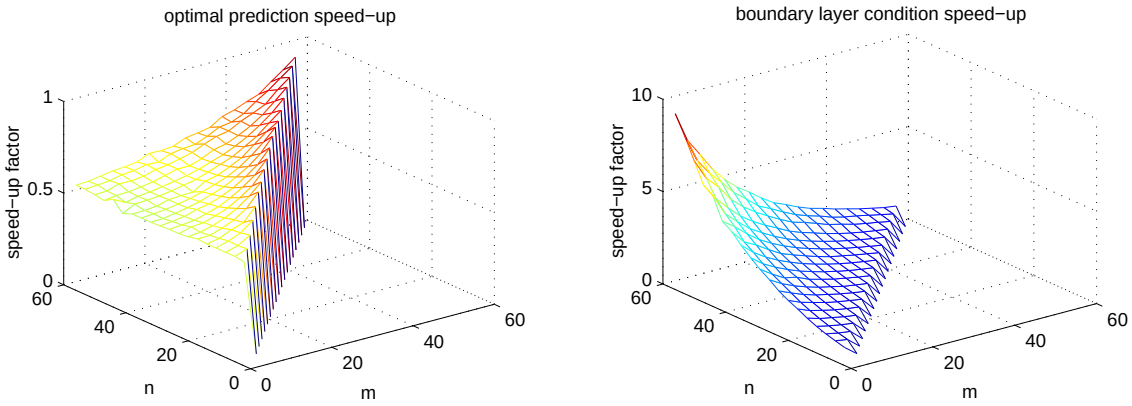


Figure 3.11: Speed-up factors depending on n and m.

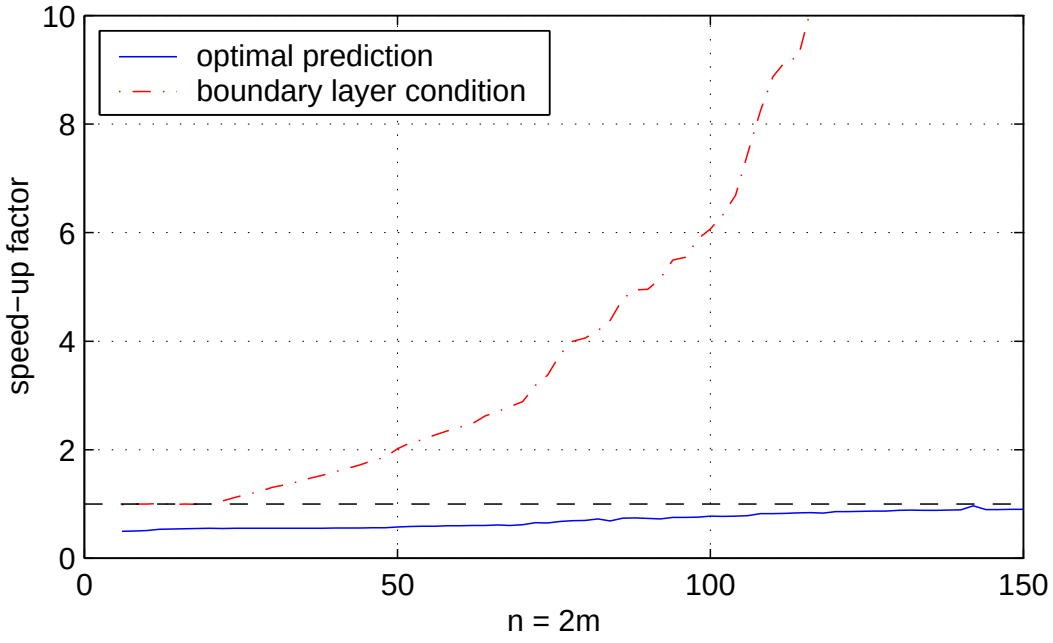


Figure 3.12: Speed-up factors for $n = 2m$

Replacing a system of $n = 60$ atoms by a system of $m = 5$ atoms is unlikely to preserve the important properties, e.g. the hopping behavior of a copper atom.

A choice of m which is reasonable in many cases, is cutting the whole crystal of n atoms into two halves, i.e. $m = \lfloor \frac{n}{2} \rfloor$. The speed-up factors for the two versions of optimal prediction are shown in Figure 3.12. While the original version does not reduce the computational effort, the speed-up for the boundary layer condition version increases significantly for n increasing.

While these results are interesting for our model problem, one should not assume that they are the case for optimal prediction in general. In our model problem the potential functions are fairly easy and cheap to evaluate. In more complicated systems it could very well pay off to solve the linear system (3.45) instead, or even to solve the minimization problem directly. On the other hand the boundary layer condition version of optimal prediction could possibly fail to work in other applications, e.g. in three space dimensions. Such questions will have to be investigated when applying optimal prediction to a new problem.

In any case, the results for the boundary layer condition version look promising with respect to computational speed-up. Of course they are only of any use, if optimal prediction turns out to preserve the important properties of the original system. This issue will be our task in the following chapter.

Chapter 4

A Comparison by Numerical Experiments

In Chapter 3 we showed that for a $2n$-dimensional Hamiltonian system, which describes n atoms, the evolution of $m < n$ atoms can be approximated by an $(n + m)$-dimensional system of ordinary differential equations. Under the assumption that the influence of one atom ranges over just k other atoms, the system's dimension could be further reduces to just $2m + k$ dimensions. Since k is typically much smaller than $n - m$, the approximate system is significantly smaller than the original $2n$-dimensional system.

The smaller system was derived by first applying the optimal prediction procedure as presented in Chapter 1, then calculating an asymptotic limit of the optimal prediction Hamiltonian for temperature zero, and finally deriving a system of equations which approximates this asymptotic limit. It is unclear, how well the various assumptions, under which the above approximations are valid, are satisfied in our particular problem.

- Optimal prediction assumes the validity of the canonical measure $Z^{-1}e^{-\beta H(q,p)}$ and the system being in thermodynamical equilibrium. It is unclear how small disturbances to the equilibrium, caused e.g. by hopping events of copper atoms, affect the result.
- The asymptotic limiting process assumes a low temperature. Here it depends on the point of view which temperatures can still be called "low".
- Finally, the $(2m + k)$-dimensional approximate system of equations employs the fact that the potential between two atoms is negligibly small at some distance. Again, it is unclear, whether these small errors are uncorrelated enough to stay negligible in the final result. The numerical results shown in Figure 3.9 may stand as a warning example.

In this chapter we try to find answers to these unclear points by comparing the approximate optimal prediction system with the original system. Since we are dealing with high-dimensional nonlinear differential equations, an analytic analysis would be far too complicated. Instead we compare the two systems by numerical experiments, as presented in Section 2.4 and Section 3.4.

In Section 4.1 we show by which criteria and using which methods we can compare the two systems. The diffusive behavior of the hopping of a copper atom will be the most important criterion. Hence, we set up an analytical model for the nature of hopping in Section 4.2 and show how it can be used in order to obtain diffusion parameters. In Section 4.3 we will numerically investigate a first example. It will turn out that sonic waves, which travel through the silicon crystal, cause disturbances, which would make a comparison by statistical quantities untrustworthy. On the other hand, the sonic waves will enlighten an interesting aspect about the nature of optimal prediction. In Section 4.4 we will present a second numerical experiment, in which sonic waves have a very small influence on the relevant quantities. The numerical results are used for a comparison of the original system with the optimal prediction approximation by statistical quantities.

4.1 Comparing Two Systems

In molecular dynamics "comparing" two systems does not mean to analyze two trajectories in phase space and plot their difference over time. Since initial positions and momenta can never be known exactly, and molecular dynamics is typically chaotic, trajectories in high-dimensional phase space are no appropriate means to compare two systems. Instead, "comparing" means to test whether both systems have similar dynamics, i.e. show the same statistical behavior. This is represented by statistical quantities, such as pressure, time correlation functions, diffusion constants, fluctuations of energy, etc. In the following, we will consider two statistical processes in order to compare the two systems:

- **The diffusion of a single copper atom in the silicon crystal due to hopping events.**
  Here we consider two quantities in order to compare the two systems:
  - The distribution of the position the copper atom
  - The distribution of the number of hopping events up to a fixed time
- **The fluctuation of energy of the first m atoms.**
  Here we use the variance of the energy over a fixed time interval as a criterion to compare the two systems.

We will obtain these quantities numerically by Monte-Carlo sampling. In the following we explain why this is a reasonable thing to do in this context.

4.1.1 Why Use Monte-Carlo Sampling?

Whenever, in molecular dynamics, statistical quantities are sought, statistical mechanics gives two canonical ways to obtain them:

- Running molecular dynamics
- Monte-Carlo sampling

In this case "molecular dynamics" often means running a single computation for a long time and using limiting processes (e.g. the Einstein relation (4.3)) to approximate statistical quantities (e.g. diffusion constants). Monte-Carlo methods, on the other hand, in principle perform the same computation over and over again, but the initial conditions are sampled from some given measure, in our case $Z^{-1}e^{-\beta H(q,p)}$. In other words: Statistical quantities can only be obtained by some kind of averaging. Molecular dynamics achieves the averaging by observing one experiment over a long time. Monte-Carlo methods achieve the averaging by observing many samples.

One important statistical quantity, which we are interested in here, is the diffusion "constant" of a diffusion process. Since so far we cannot be sure, whether the diffusion is time-independent, we rather speak of a "diffusion parameter" in the following. In many applications, diffusion parameters are statistical quantities of theoretical and practical importance. In most examples, the considered diffusion process is self-diffusion, as e.g. in [4]. In this thesis, however, we are not interested in self-diffusion (which does not happen at all in our model problem, since two silicon atoms cannot swap their positions), but rather in the diffusion of a single copper atom inside a crystal of silicon atoms. The diffusion happens due to hopping events, i.e. the copper atom changes position with silicon atoms. Since the hopping of a copper atom has been the effect of interest in the first place (as introduced in Chapter 2), it is a natural decision to use the diffusion of a copper atom as a criterion to compare the optimal prediction system to the original system.

In textbooks and papers on statistical mechanics (as in [25, 31]) diffusion constants are often calculated by the *Green-Kubo formula*

$$\kappa = \int_{t=0}^{\infty} \langle v_i(t) \cdot v_i(0) \rangle \, \mathrm{d}t. \tag{4.1}$$

This formula assumes the system to be in thermodynamical equilibrium and ergodic. Here $v_i(t)$ is the velocity of particle i at time t. The function

$$a(t) = \langle v_i(t) \cdot v_i(0) \rangle \tag{4.2}$$

is called *velocity autocorrelation function*. The parentheses $\langle \cdot \rangle$ denote ensemble average, which is equivalent to space and time averaging, since the system is ergodic.

Another statement from statistical mechanics is the *Einstein relation*

$$\kappa = \lim_{t\to\infty} \frac{1}{2t} \left\langle |r_i(t) - r_i(0)|^2 \right\rangle, \tag{4.3}$$

where $r_i(t)$ is position of particle i at time t. Again, the system is assumed to be in equilibrium and ergodic. Both relations (4.1) and (4.3) are presented for the one-dimensional case. In three space dimensions they have to be multiplied by the constant $\frac{1}{3}$. The equality of (4.1) and (4.3) is proven e.g. in [25].

In our computations, however, we do not use any of these relations. Instead we obtain the diffusion constant by Monte-Carlo sampling, i.e. we solve the same system over and over again, but sample the initial conditions from the canonical distribution $\tilde{Z}^{-1}e^{-T(p)}$. The reason, why we prefer this over using the Green-Kubo formula or the Einstein relation, which require the system to be solved only once, is that for our model problem (and also for many real problems) the expressions

$$\int_{t=0}^{T} \langle v_i(t) \cdot v_i(0) \rangle \, \mathrm{d}t \tag{4.4}$$

and

$$\frac{1}{2T} \left\langle |r_i(T) - r_i(0)|^2 \right\rangle \tag{4.5}$$

converge very slowly to (4.1) respectively (4.3). This is problematic, since in long time computations the copper atom often travels towards the boundaries of the silicon crystal, which should be avoided for the following reasons:

- Boundary effects appear due to the larger equilibrium distance between two silicon atoms (surface tension).
- The copper atom may hop over the outermost silicon atom and completely leave the crystal.
- In the case of the optimal prediction system, the copper atom could hop too close to (even into) the area of virtual particles, which would violate the assumption of a clear separation between particles.

In solving the system repeatedly, we can compute up to much shorter times, such that the above effects are extremely unlikely to happen. And if they do still happen, their influence will not be relevant in the average.

Since in Monte-Carlo sampling the expected error decays with the square root of the number of samples, on the one hand the quantity of interest can in principle be approximated arbitrarily well, on the other hand a very high accuracy would require an inefficiently large number of samples. For our model problem, between 5000 and 25000 samples turned out to yield completely satisfying results, in the sense that the error due to the Monte-Carlo sampling is significantly smaller than the difference in the two quantities in comparison, as shown by the numerical results in Subsection 4.4.1.

4.1.2 Monte-Carlo Sampling in the Model Problem

Monte-Carlo sampling in our numerical experiments means solving the same system, as shown in Figure 2.5, Figure 3.6, and Figure 3.7, N times. The initial conditions, i.e. positions q_i and momenta p_i should be sampled from the canonical distribution $Z^{-1}e^{-\beta H(q,p)}$. Due to the simple structure of the kinetic energy, sampling the momenta p_i is just sampling independently from Gaussian distributions. Sampling the positions from $\tilde{Z}^{-1}e^{-\beta V(q)}$, on the other hand, would require high-dimensional sampling from a nonlinear distribution, which additionally for low temperatures has very high peaks. For such a distribution the standard methods as *acceptance-rejection methods* or *Metropolis sampling* (as described e.g. in [15, 24]), would be far from easy to apply efficiently. We circumvent such problems by sampling only the momenta p_i from the canonical distribution, and keeping the positions q_i fixed at the potential minimum (as described in Section 2.4).

Obviously sampling only the momenta is not exactly the same as sampling both the positions and the momenta from $Z^{-1}e^{-\beta H(q,p)}$. But in the same manner as the kinetic energy decays very quickly to its equilibrium value (as observed in Section 2.4), we can assume both the positions and the momenta to be distributed according to the canonical distribution after a similarly short time. Additionally, keeping the positions fixed in the initial conditions, automatically guarantees that we remain in the correct domain M^L, as introduced in Subsection 3.1.1.

4.1.3 Criteria of Comparison

As reasoned in the previous subsection, we will compute statistical quantities about the original system and the optimal prediction approximation by Monte-Carlo sampling. The so obtained quantities hold as criteria for a comparison of the two systems. In the following sections, we will focus on the following three statistical quantities:

1. **The distribution of the position of one copper atom inside the silicon crystal.** A single copper atom, which is initially located always at the same position, will describe very different paths over time, given its momentum and the positions and momenta of the silicon atoms are randomly sampled from the canonical distribution $Z^{-1}e^{-\beta H(q,p)}$, respectively sampled in the way described in Subsection 4.1.2. The copper atom will change its position inside the crystal by hopping over silicon atoms. Since hopping events depend very sensitively on the positions and momenta at a given time, the way the copper atom moves through the crystal will be very different in each experiment. If the initial conditions are randomly sampled from the canonical distribution, the position of the one copper atom can be described by a diffusion process, i.e. the distribution function for the position of the copper atom might be described by a diffusion equation. We approximate this distribution function by Monte-Carlo sampling.

Once there are distribution functions for the position of the copper atom for the original system and the optimal prediction approximation, these distribution functions have to be compared. One way to compare them, which we will do in Subsection 4.4.2, is simply evaluating the difference between them, respectively plotting the relative error in the maximum norm in space. Another way, which is more physically motivated, is to construct a model for the diffusion process, which can be solved analytically. This model can then be used to obtain diffusion parameters, which hold as statistical quantities for comparing the two systems. In Section 4.2, we will model the copper diffusion as a random walk and provide its analytical treatment. The numerical application will then follow in Subsection 4.4.3.

2. **The distribution of the number of hopping events up to a fixed time.** Our task is to check if optimal prediction preserves the nature of hopping. One criterion is the number of hopping events up to a fixed time. As we are performing N experiments of the same system with different initial data, we obtain a distribution for the number of hopping events, i.e. each nonnegative number is mapped to a number of experiments which had exactly the given number of hopping events. This distribution should be the same for the original system and the optimal prediction approximation, if the nature of hopping is preserved. We will investigate this issue in Subsection 4.4.4.

3. **The fluctuation of energy of the first $\mathbf{m}$ atoms.** For the original system and for the optimal prediction approximation, we compute the energy of the first m atoms, which is given by Formula (2.11). While for the original system the total energy of all n atoms is constant over time (the system is Hamiltonian), the energy of the first m atoms fluctuates around some fixed average, given the system is in thermodynamical equilibrium. For the approximate optimal prediction system, the energy of the m real particles is computed without considering the virtual particles. Hence also this energy will fluctuate around some mean. In Subsection 4.4.5 we consider the variance of this fluctuating energy in order to compare the two systems. This test is particularly interesting, since we compare the exchange of energy among real particles (for the original system) with the exchange of energy between real and virtual particles (for the optimal prediction approximation).

Since the diffusion of the copper atom is the process we wanted to study in the first place, the diffusion constants for the diffusion of the copper atom are of particular interest. In the following section we will investigate the nature of the given diffusion process. In particular, we construct a mathematical model which describes the hopping of the copper atom as a special random walk.

4.2 A Model for the Copper Diffusion

As reasoned previously, the diffusion process of a copper atom due to atomic hopping is the main criterion for comparing the optimal prediction approximation with the original system. In this section, we present an analytical model for the copper diffusion and show how this model can be used to compute the time-dependent diffusion parameter of a diffusion process which is obtained by numerical experiments.

4.2.1 Modeling Hopping as a Random Walk

When speaking of the diffusion of the copper atom we mean diffusion due to hopping events, i.e. small displacements from its initial condition due to oscillations between the same two silicon atoms should not be taken into account, but changes in the position inside the crystal due to hopping events should. Hence the appropriate diffusion process to describe the copper diffusion is discrete in space, with a spatial discretization d_0, which is the equilibrium distance between two silicon atoms. Unless the copper atom comes too close to the boundaries of the crystal, it is valid to assume d_0 to be constant.

Let us assume for the moment, that the diffusion process is linear. If it was continuous in space, it would be described by the heat equation

$$\begin{aligned} u_t(x,t) &= \kappa u_{xx}(x,t), && (4.6)\\ u(x,0) &= \delta_0(x), && (4.7) \end{aligned}$$

where κ is the diffusion constant. The heat equation is in many cases a valid approximation to discrete diffusion processes, as often the atomic distance d_0 is small compared to the relevant length scales. Hence for a large number of atoms, especially in liquids and gases, the heat equation is typically used to describe atomic diffusion, e.g. in [25]. However, in our model problem and for the short times we consider, the atomic distance d_0 is the relevant length scale for the diffusion due to hopping. Hence, we need a diffusion process which is discrete in space (on the spatial grid $\{-\mathfrak{m}d_0, \ldots, (\mathfrak{m}-1)d_0, \mathfrak{m}d_0\}$). Linear diffusion processes which are discrete in space are described by so called *compartment models*:

$$\begin{aligned} \dot{u}(t) &= A \cdot u(t),\\ u(0) &= u_0, \end{aligned} \qquad (4.8)$$

where $A \in \mathbb{R}^{(2\mathfrak{m}+1)\times(2\mathfrak{m}+1)}$ is a matrix with column sums equal to zero (to ensure mass

conservation, i.e. $\frac{d}{dt}\sum_i u_i(t) = 0$), $u(t) \in \mathbb{R}^{2\mathfrak{m}+1}$, and

$$u_0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{2\mathfrak{m}+1}. \tag{4.9}$$

Here the initial conditions (4.7), respectively (4.9) represent the fact that the copper atom always starts at the same position, in particular between the same two silicon atoms.

The analytical solution to (4.8) is

$$u(t) = \exp(tA) \cdot u_0. \tag{4.10}$$

A simple example for a diffusion process which is discrete in space is described by the tridiagonal matrix (which is Toeplitz up to the boundary entries a_{11} and a_{nn})

$$A = \frac{\kappa}{d_0^2} \cdot \begin{pmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & 1 & -2 & 1 & \\ & & & 1 & -1 & \end{pmatrix}, \tag{4.11}$$

which is the standard finite difference approximation to the heat equation (4.6). Hence, the compartment model (4.8) with the matrix (4.11) converges to the heat equation (4.6) (with Neumann boundary conditions in this case) as $d_0 \to 0$. But as there are many possible finite difference approximations to u_{xx}, there are many possible matrices A which yield a diffusion process converging to the heat equation for $d_0 \to 0$. It has to be pointed out that two different matrices A_1 and A_2, which both yield the heat equation in the limit $d_0 \to 0$, will in general describe different *discrete* diffusion processes.

To investigate which diffusion process describes the copper diffusion most accurately and to obtain diffusion constants (given the process is linear), we model the evolution of the copper atom's position as a specific random walk. The experiment in Figure 2.5 shows the typical behavior of the copper atom. It oscillates for some time between two silicon atoms, and then suddenly jumps to a neighboring position inside the crystal. Sometimes one hopping event is followed immediately by a further hopping event, resulting in a jump over two or more silicon atoms. The latter event is likely to happen, since the copper atom requires a certain amount of kinetic energy when performing a hopping event, in order to overcome the potential barrier between silicon and copper. Since this kinetic energy is not lost in the hopping event, the chances are good that the consecutive silicon atom is being hopped over, too. In other words:

> Hopping events are *correlated*, in the sense that given the copper atom has just hopped to the left (right), it is much more likely than normally that a second hopping event to the left (right) follows.

This important aspect is ignored in many papers on this issue. E.g. in [4] the authors assume that "The velocities of the particles before the jump is understood to be uncorrelated from the velocities after the jump", which is not a reasonable assumption in the case of atomic hopping in a crystal at low temperatures.

The correlation between hopping events makes the modeling as a random walk much more complicated, since a random walk with correlated jumps will have memory, i.e. be non-Markovian. However, we can describe the hopping process as a Markovian random walk by using the following trick:

- Hopping events are assumed to be uncorrelated.
- In a single hopping event the copper atom can also hop over two or more silicon atoms at once.

These two assumptions lead to a model, which is typical in the context of stochastic processes (e.g. in [30]). Let X_t denote the position of the copper atom at time t. The random walk is given by the following properties:

1. Let $(T_n)_{n\in\mathbb{N}}$ be the hopping times of the process. Then the times between two hopping events $\Delta T_n = T_n - T_{n-1}$ are independent and have the distribution

$$P\left(\Delta T_n \in [s, s+\,\mathrm{d}s)\right) = \alpha \cdot \exp(-\alpha s)\,\mathrm{d}s, \tag{4.12}$$

   where α is a parameter controlling the hopping rate. We will investigate its connection to the diffusion constant κ later. From the derivations in [30, pp. 53-58] and [29, pp. 222-227] it follows that for a Markovian process which is discrete in space and continuous in time, the hopping times T_n must be a Poisson process and satisfy (4.12).

2. A hopping event is a change of position of the copper atom inside the silicon crystal. Let $\Delta_n \in \mathbb{Z}\setminus\{0\}$ denote the number of silicon atoms which the copper atoms hops over to the right in the n-th hopping event. Consequently $\Delta_n < 0$ means hopping to the left. Δ_n is assumed to be independent and distributed according to

$$P\left(|\Delta_n| = i\right) = p_i, \tag{4.13}$$

   where $(p_i)_{i\in\mathbb{N}}$ is a non-negative sequence, i.e. $(p) : \mathbb{N} \to \mathbb{R}_0^+$, which satisfies $\sum_{i\in\mathbb{N}} p_i = \frac{1}{2}$ and $\sum_{i\in\mathbb{N}} i^2 p_i < \infty$. For an easy analysis one can assume that only finitely p_i are non-zero, i.e. the hopping behavior is described by a vector

$p = (p_1, \ldots, p_k)$. By assuming that $P(\Delta_n = i) = P(\Delta_n = -i)$ we consider only symmetric random walks. While this restriction is in principle not necessary, it simplifies some of the following analysis. Assumed that the copper atom does not come close to the boundaries of the silicon crystal, our random walk is indeed symmetric, as the numerical experiment in Subsection 4.4.3 will show.

An analysis of the described random walk (along the lines of the analysis in [30, pp. 53-58]) yields that its variance at time t is

$$\text{Var}(X_t) = \alpha t \cdot 2 \sum_{i=1}^{k} (d_0 i)^2 p_i. \tag{4.14}$$

In comparison equals the variance of the heat equation (4.6) and the discrete diffusion process given by (4.11)

$$\text{Var}(Y_t) = 2\kappa t \tag{4.15}$$

at time t. As the described random walk should also be described by the heat equation in the limit $d_0 \to 0$, we obtain the relation

$$\kappa = \alpha \cdot \sum_{i=1}^{k} (d_0 i)^2 p_i, \tag{4.16}$$

which allows us to speak also for the discrete diffusion process (4.8) of a diffusion constant κ, which is consistent with the diffusion constant κ in the continuous diffusion process (4.6).

Following the derivation in [30, pp. 56-57] one obtains that the probability distribution of the random walk X_t is described by the differential equation (4.8), where the matrix A is given by

$$A = \alpha \cdot \begin{pmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & & & \\ & p_k & \ldots & p_1 & -1 & p_1 & \ldots & p_k & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & p_k & \ldots & p_1 & -1 & p_1 & \ldots & p_k & \\ & & & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4.17}$$

In the context of random walks, the matrix (4.17) is called *infinitesimal generator*. Of course, as a matrix is finite dimensional, this k-diagonal structure has to be changed at the upper and lower rows. In the termini of differential equations this means that *boundary conditions* have to be specified. In our case we can assume that in no Monte-Carlo experiment the copper atom leaves the crystal or comes close to the boundaries, which is reasonable since the computational time is comparably short. In this case the appropriate boundary conditions are *Neumann boundary conditions*, i.e. no flux through the boundaries, and thus energy conservation, i.e.

$$\frac{d}{dt} \sum_i u_i(t) = 0. \tag{4.18}$$

This assumption is satisfied, if the matrix A has zero column sums. The best results are achieved, if A is chosen at least so large, that for the whole computation time the values of $u(t)$ close to the boundaries will be negligibly small anyhow. Note that depending on the type of boundary conditions the matrix A need not be symmetric anymore.

An important example for the above kind of process is the case

$$\begin{cases} p_1 = \frac{1}{2} \\ p_i = 0 \quad \forall i \geq 2, \end{cases} \tag{4.19}$$

which corresponds to a symmetric hopping over only single silicon atoms, where the single hopping events are completely uncorrelated. The generating matrix would be exactly the finite difference approximation (4.11). Since in our model problem single hopping events are correlated, we have at least $p_2 > 0$.

The question, which values for $p_1, \ldots, p_k$ are to be chosen to describe the real process in the best way, can be answered by Monte-Carlo sampling, which we will do in Subsection 4.4.3. While one could in principle also approximate the constant α in (4.12) using Monte-Carlo sampling, we leave it as parameter and instead compute the diffusion parameter κ by the method described in Subsection 4.2.2, since it turned out to be more stable and additionally allows to include a time dependence in κ, i.e. to consider $\kappa(t)$.

4.2.2 Using the Model to Obtain Diffusion Parameters

The random walk model, as presented above, can be used to obtain diffusion parameters of a distribution which is obtained by Monte-Carlo sampling of a numerical simulation. Let $v(x,t)$ denote the numerically obtained probability distribution for the position of the copper atom. Here x stands for the position of the copper atom (i.e. $x \in \mathbb{Z}$), and t denotes time. In the argumentation below, t is always assumed to be continuous. Of course, $v(x,t)$ is actually sampled only at discrete times t_i, where $t_{i+1} - t_i$ is the time step of the integrator. At time $t = 0$ we have

$$v(x,0) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0. \end{cases} \tag{4.20}$$

Since it is unclear, if the diffusion parameter κ is constant in time, we let it be a time dependent parameter $\kappa(t)$. Over a comparably short time interval $\tau \in [t, t+\Delta t)$, however, we can assume the diffusion process (4.8) to be linear with a constant diffusion parameter $\kappa(t)$.

Let A be the generating matrix (4.17) as constructed in Subsection 4.2.1. Motivated by

relation (4.16) we define

$$\tilde{A} = \frac{1}{\sum_{i=1}^{k} (d_0 i)^2 p_i} \cdot \begin{pmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & & & \\ & p_k & \dots & p_1 & -1 & p_1 & \dots & p_k & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & p_k & \dots & p_1 & -1 & p_1 & \dots & p_k & \\ & & & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{4.21}$$

such that

$$A = \kappa \tilde{A}. \tag{4.22}$$

Since in Subsection 4.4.3 we approximate the values $p_1, \dots, p_k$ by Monte-Carlo sampling, the matrix $\tilde{A}$ can be completely determined.

On the interval $\tau \in [t, t + \Delta t)$ we define the error functional

$$F(\kappa) = \int_t^{t+\Delta t} \| e^{(\tau - t)\kappa \tilde{A}} \cdot v(t) - v(\tau) \|_2^2 \, \mathrm{d}\tau, \tag{4.23}$$

i.e. we use the $\mathbb{L}^2$ norm on the domain

$$(x, t) \in \{-\mathfrak{m} d_0, -(\mathfrak{m} - 1) d_0, \dots, (\mathfrak{m} - 1) d_0, \mathfrak{m} d_0\} \times [t, t + \Delta t). \tag{4.24}$$

We choose this norm, since all points in the domain (4.24) are equally taken into account. Hence it is particularly stable with respect to errors in v due to the Monte-Carlo sampling.

On the domain (4.24) we wish to chose the particular κ which minimizes the error (4.23) between the real evolution $v(x, \tau)$ and the evolution which the analytical model (4.8) with initial values $v(x, t)$ would yield

$$F(\kappa) \to \min. \tag{4.25}$$

A necessary condition for (4.25) is

$$F'(\kappa) = 0. \tag{4.26}$$

This implicit relation can be solved by Newton iteration

$$\kappa^{(i+1)} = \kappa^{(i)} - \frac{F'(\kappa^{(i)})}{F''(\kappa^{(i)})}. \tag{4.27}$$

The functions F' and F'', which are required for (4.27), can be obtained by some analysis as

$$F'(\kappa) = v(t)^T \cdot \tilde{A} \cdot \int_t^{t+\Delta t} (\tau - t) e^{(\tau - t)\kappa \tilde{A}} \cdot \left(e^{(\tau - t)\kappa \tilde{A}} \cdot v(t) - v(\tau) \right) \mathrm{d}\tau, \tag{4.28}$$

and

$$F''(\kappa) = v(t)^T \cdot \tilde{A}^2 \cdot \int_t^{t+\Delta t} (\tau - t)^2 e^{(\tau-t)\kappa\tilde{A}} \cdot \left(2e^{(\tau-t)\kappa\tilde{A}} \cdot v(t) - v(\tau)\right) \, \mathrm{d}\tau. \quad (4.29)$$

In both expressions the integrals can be numerically approximated using the trapezoidal rule on the time steps t_i. In this way the diffusion parameter $\kappa(t)$ belonging to the time interval $[t, t + \Delta t)$ can be computed for all time steps t_i. Hence, the value $\kappa(t_{i-1})$ is a good starting value for the Newton iteration for $\kappa(t_i)$.

The length Δt of the considered time interval must on the one hand be large enough to have already some diffusion taken place, on the other hand – given the diffusion turns out to be time-dependent – it must be small enough to justify the approximation that the diffusion parameter $\kappa(t)$ is constant on the interval $[t, t+\Delta t)$. In Subsection 4.4.3 we will compute diffusion parameters for the numerical data obtained for our model problem.

4.3 An Experiment Containing Sonic Waves

In this section we present the results of a first Monte-Carlo experiment. We solve the original system, consisting of $n = 35$ atoms, where the one copper atom starts at the 11^{th} position from the left, $N = 5000$ times for a time of $t^* = 2.0 \cdot 10^{-12}$s. The temperature of the experiments in this section is $T = 4000$K.

4.3.1 Numerical Results

The distribution function over time for the position of the copper atom in the silicon crystal is plotted in Figures 4.1 and 4.2. The initial position of the copper atom is position 11. Since for each experiment the copper atom starts at the same position, at time $t = 0$ the distribution is 1 at $x = 11$ and 0 for all other positions.

As it can be observed in Figure 4.1, the copper atom describes some diffusion process on a long time scale, as the variance of the distribution is significantly larger at time $t = 2.0 \cdot 10^{-12}$s than at time $t = 0$s. But obviously this process is no linear diffusion with a constant diffusion parameter. In particular, the distribution does not satisfy a maximum principle, since the maximum increases around the times $t_1 = 0.5 \cdot 10^{-12}$s, $t_2 = 1.0 \cdot 10^{-12}$s, $t_3 = 1.5 \cdot 10^{-12}$s, and $t_4 = 2.0 \cdot 10^{-12}$s. This behavior is too systematic and too large in effect to be only due to errors in the Monte-Carlo sampling (as estimates as described in Subsection 4.4.1 indicate). It must be a feature of the process, and this feature is not caught in the random walk model in Subsection 4.2.1.

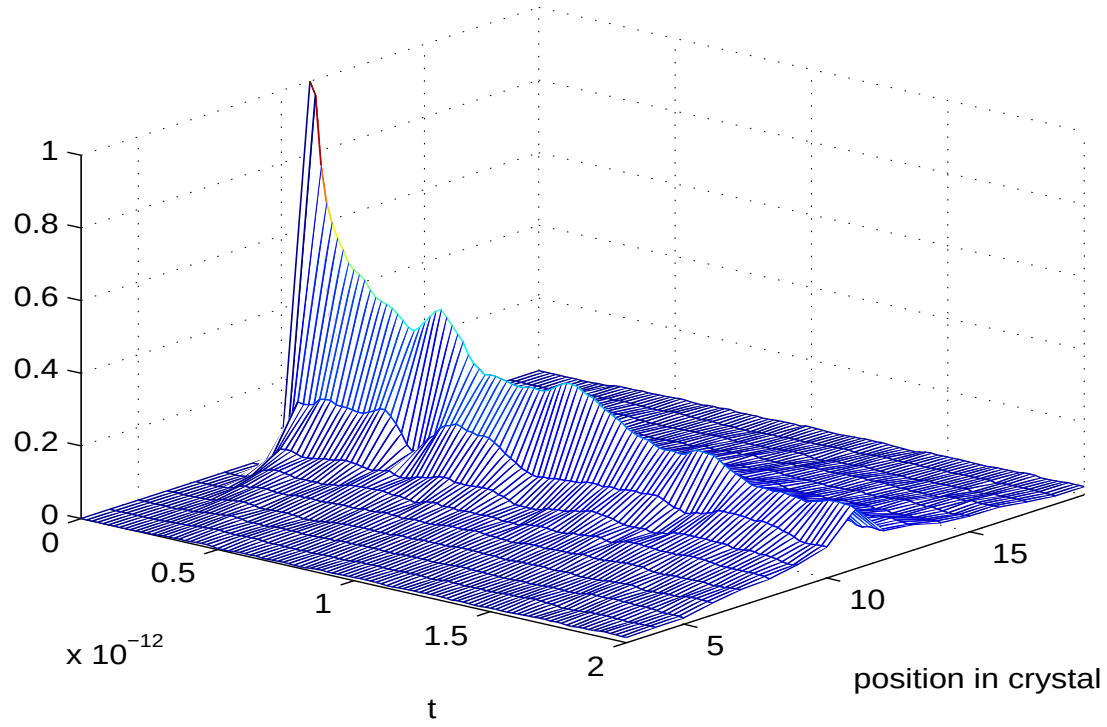


Figure 4.1: Distribution of the copper atom's position (viewed from the side)

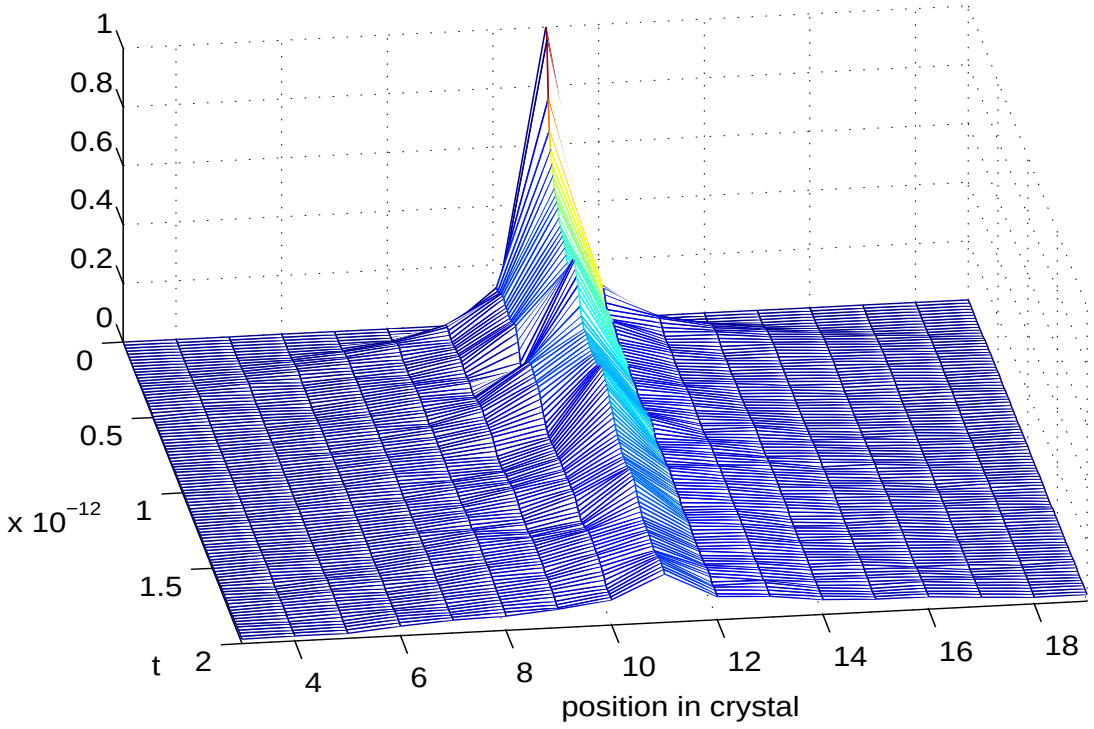


Figure 4.2: Distribution of the copper atom's position (viewed from the top)

4.3.2 Sonic Waves

The fact that the distribution's maximum increases for a short time each $5 \cdot 10^{-13}$ seconds means that at these times for more Monte-Carlo experiments than before the copper atom is "pushed" back to its initial position. A process which can possibly "push" the copper atom back to its initial position is a sonic wave, caused by a hopping event (obviously a non-equilibrium effect) or by the starting configuration (remember that we do not start exactly in thermodynamical equilibrium, as explained in Section 2.3). Such a wave is normally small in effect and not to recognize by looking at the plot. But it can certainly be strong enough to influence the hopping behavior when it hits the copper atom again after being reflected.

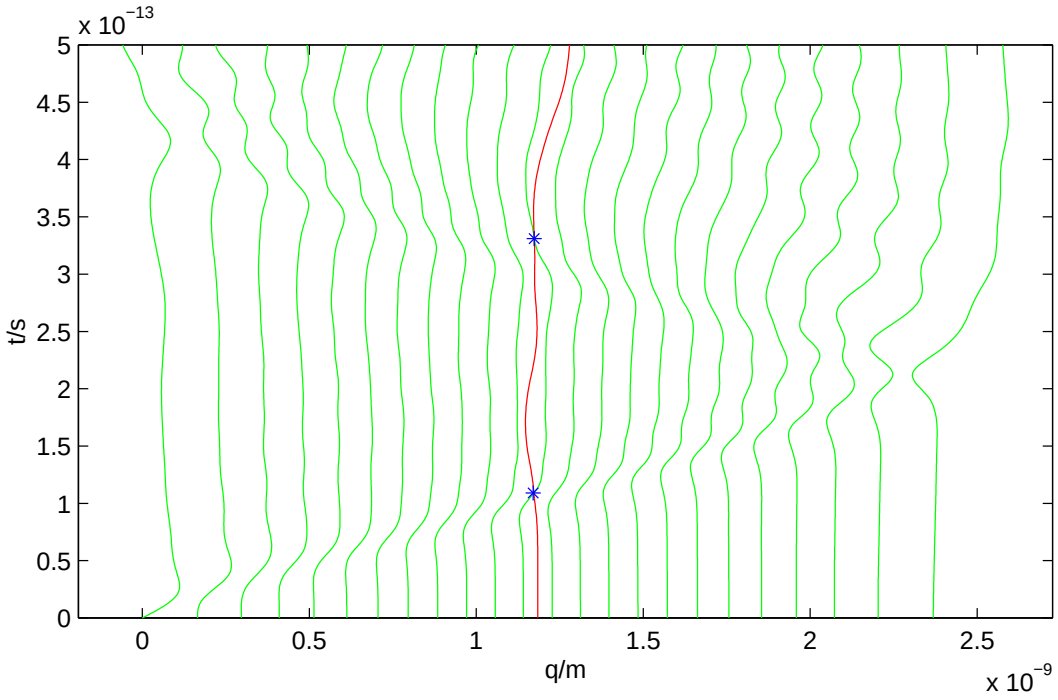


Figure 4.3: A sonic wave traveling through the crystal

Figure 4.3 shows an exaggerated example of a sonic wave. The wave is started at the left side of the crystal, travels through to the right, is being reflected and travels back to the left. Since the system is non-linear, the wave loses structure while traveling, and especially when being reflected. This behavior reflects the fact that the system is going towards thermodynamical equilibrium. In this example the wave indeed produces hopping events of the copper atom, when running over its position. It has to be pointed out that not every wave which hits the copper atom yields a hopping event. On the contrary, the waves in a "real" simulation are much weaker and seldom yield a hopping event. It just happens a little bit more often than normally, which is enough to increase the maximum of the distribution in Figures 4.1 and 4.2

In order to ensure whether sonic waves can be the reason for the short time intervals of anti-diffusion in distribution in Figures 4.1 and 4.2, we need to compute, or at least estimate, the velocity of sound inside the silicon crystal. The example shown in Figure 4.3 may already give a hint, but as it is strongly exaggerated, we wish to estimate the velocity of sound analytically.

4.3.3 The Velocity of Sound

In the following we assume that the equilibrium distance between two silicon atoms d_0 is constant value. Obviously this in an incorrect assumption close to the boundaries. But for a rough estimate for the velocity of sound we can neglect the boundary effects. The equilibrium distance is then given implicitly by the relation

$$\sum_{i=1}^{\infty} f'(d_0 \cdot i) = 0. \tag{4.30}$$

This relation can be solved numerically by Newton iteration. For our model problem we obtain the value

$$d_0 = 1.87 \cdot 10^{-10}\text{m}. \tag{4.31}$$

As derived in [25] we need to compute the *Young modulus*

$$Y = d_0 \cdot \frac{\partial^2 H_{\text{loc}}}{\partial d_0^2}(d_0). \tag{4.32}$$

Here H_{loc} is the potential energy of a single silicon atom (without loss of generality we can assume its position to be 0)

$$H_{\text{loc}} = \sum_{\substack{i=-\infty \\ i\neq 0}}^{\infty} f(0 - q_i) = \sum_{\substack{i=-\infty \\ i\neq 0}}^{\infty} f(d_0 \cdot i) = 2 \cdot \sum_{i=1}^{\infty} f(d_0 \cdot i). \tag{4.33}$$

In three space dimensions, H_{loc} is differentiated with respect to volume, which in one dimension is differentiation with respect to d_0. Substituting (4.33) into (4.32) yields

$$Y = 2d_0 \cdot \sum_{i=1}^{\infty} i^2 \cdot f''(d_0 \cdot i) \tag{4.34}$$

In our model problem the numerical value is

$$Y = 5.05 \cdot 10^{-9} \tfrac{\text{kg}\cdot\text{m}}{\text{s}^2}. \tag{4.35}$$

The velocity of sound is then obtained by the relation

$$c = \sqrt{\frac{Y}{\rho}} = \sqrt{\frac{Y \cdot d_0}{m_{Si}}}, \tag{4.36}$$

which yields for our model problem

$$c = 4.50 \cdot 10^3 \tfrac{\mathrm{m}}{\mathrm{s}}. \tag{4.37}$$

Further important for our estimate are the distances

$$s_{\text{left}} = q_{11} - q_1 = 1.04 \cdot 10^{-9}\mathrm{m}, \tag{4.38}$$

$$s_{\text{right}} = q_{35} - q_{11} = 2.12 \cdot 10^{-9}\mathrm{m}, \tag{4.39}$$

which are the distances of the copper atom to the left respectively right boundary of the crystal. Assuming that the velocity of sound is constant inside the crystal, we can now estimate the time the sonic waves take to come back to the copper atom.

$$t^*_{\text{left}} = \frac{2 \cdot s_{\text{left}}}{c} = \frac{2.08 \cdot 10^{-9}\mathrm{m}}{4.50 \cdot 10^3 \frac{\mathrm{m}}{\mathrm{s}}} = 4.64 \cdot 10^{-13}\mathrm{s}, \tag{4.40}$$

$$t^*_{\text{right}} = \frac{2 \cdot s_{\text{right}}}{c} = \frac{4.24 \cdot 10^{-9}\mathrm{m}}{4.50 \cdot 10^3 \frac{\mathrm{m}}{\mathrm{s}}} = 9.43 \cdot 10^{-13}\mathrm{s}. \tag{4.41}$$

So here we have the special case that the wave traveling to right takes twice as long to hit the copper atom again as the wave traveling to the left. Considering the strong simplifications in the above derivation we obtain the rough estimate, that the wave to the left influences the hopping behavior every $0.5 \cdot 10^{-12}$ seconds, and the wave to the right influences the hopping behavior every $1.0 \cdot 10^{-12}$ seconds. These values coincide perfectly with the times observed in Figures 4.1 and 4.2, which brings us to the conclusion that indeed sonic waves are the reason for the short intervals of antidiffusion which could be in Figures 4.1 and 4.2.

Figure 4.4 shows a sketch of how sonic waves travel inside the crystal for the original system (as presented in Figures 4.1 and 4.2), given a non-equilibrium events starts them at $t = 0$. Note that in a real experiment waves lose structure while traveling and when being reflected due to the nonlinearity of the system. Hence the wave to right has not the same influence when coming back to the copper atom as the wave to the left. Also do the waves in reality not hit the copper atom at exactly the same time.

Observe further (this aspect can be seen in Figure 4.2) that at times $t_1 = 0.5 \cdot 10^{-12}$s and $t_3 = 1.5 \cdot 10^{-12}$s the distribution is a bit asymmetric, while at times $t_2 = 1.0 \cdot 10^{-12}$s and $t_4 = 2.0 \cdot 10^{-12}$s it is symmetric again. This can be explained by the fact that at times t_1 and t_3 only a wave from the left hits the copper atom, while at times t_2 and t_4 the wave coming from the right eliminates this asymmetry.

4.3.4 Sonic Waves in the Optimal Prediction System

So sonic waves, which travel through the crystal, disturb the pure diffusive nature of the hopping of the copper atom. This is an important observation, since sonic waves are

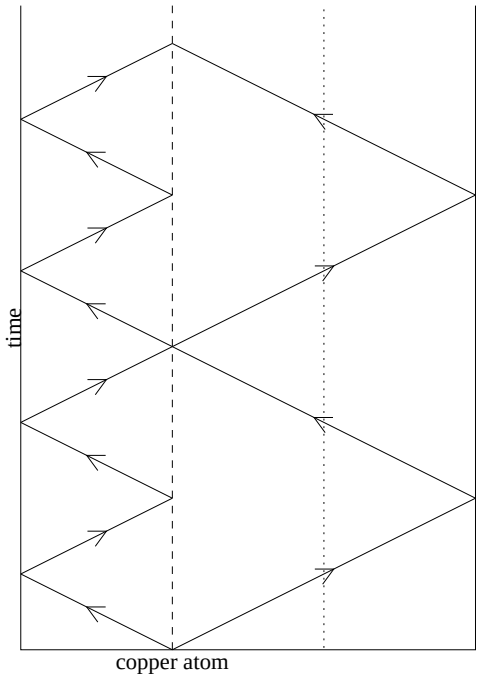


Figure 4.4: Behavior of sonic waves in the original system

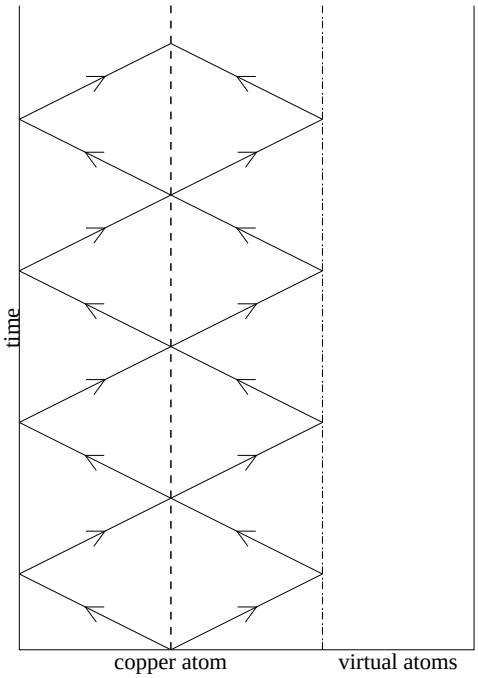


Figure 4.5: Behavior of sonic waves in the optimal prediction approximation

non-equilibrium effects, but optimal prediction in this context assumes the system to be in perfect thermodynamical equilibrium. Which leads us to the question, in which way sonic waves are reproduced by the optimal prediction system.

Figure 4.6 shows an exaggeratedly strong sonic wave in the optimal prediction system (without any copper atoms). Note that this wave is being *reflected* by the block of virtual particles. This is understandable, since optimal prediction assumes the system to be in equilibrium, while a sonic wave is a non-equilibrium effect. But this observation means that optimal prediction in the form presented in Chapters 1 and 3 does not reproduce the correct behavior of non-equilibrium effects like sonic waves.

Figure 4.7 shows the probability distribution for the position of the copper atom in Monte-Carlo computations of the optimal prediction system. While in principle the long scale diffusive behavior is very similar, the non-equilibrium antidiffusion effects are not visible. This can be explained by computing the distance from the copper atom to the first virtual particle

$$\tilde{s}_{\text{right}} = q_{25} - q_{11} = 1.07 \cdot 10^{-9}\text{m}. \tag{4.42}$$

Note that $\tilde{s}_{\text{right}}$ is almost equal to s_{left}. This yields a behavior as shown in Figure 4.5. Sonic waves travel almost symmetrically to both sides, hence their influences cancel each other

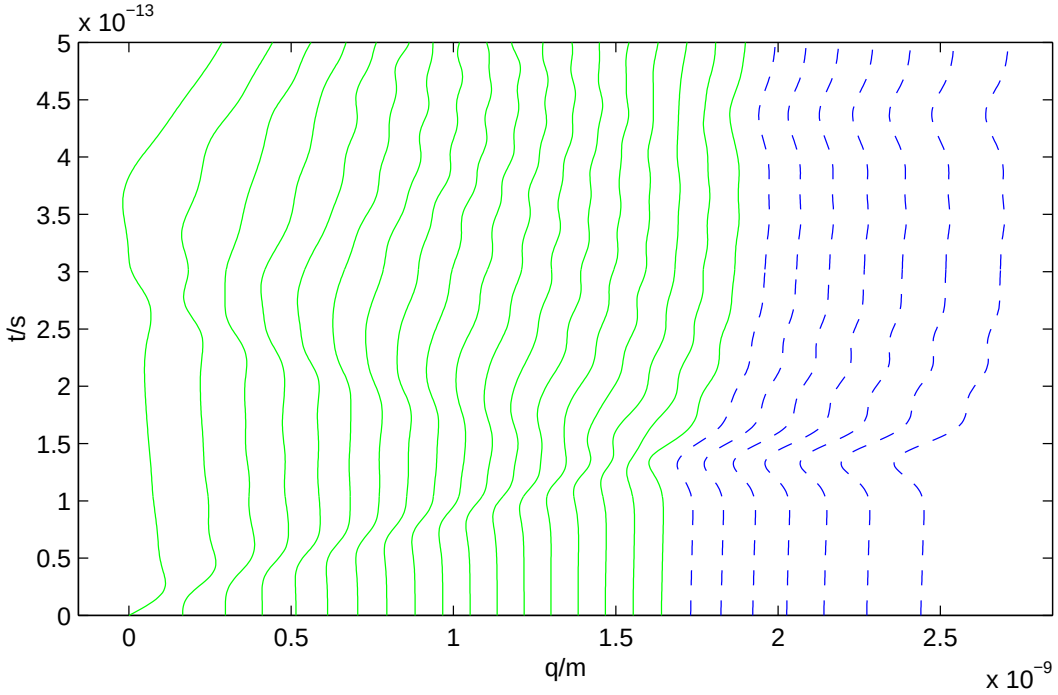


Figure 4.6: A sonic wave in the optimal prediction system

out. This was not the case in the original system, since the wave to the right lost a lot of its structure while traveling twice the distance.

Due to this difference in the way the optimal prediction approximation treats sonic waves compared to the original system, a comparison between the original system and the optimal prediction approximation becomes quite complicated in the presence of sonic waves. Therefore we will show in the following section how one can eliminate the influence of sonic waves.

On the other hand it has to be pointed out that the effect of sonic waves is particularly strong in a one-dimensional problem. In three space dimensions waves travel to all directions and hence are negligibly small already after a short distance of traveling. Since waves caused by hopping events or by the initial conditions are small in effect anyhow, optimal prediction is still a promising candidate for computations in three space dimensions. As an important result we can conclude:

> Optimal prediction in the form presented in Chapters 1 and 3 promises to be a good approximation to the original system, if the influence of non-equilibrium effects is small.

The question if the method of optimal prediction can be generalized to a molecular dynamics problem which is not in equilibrium (e.g. with a temperature gradient present)

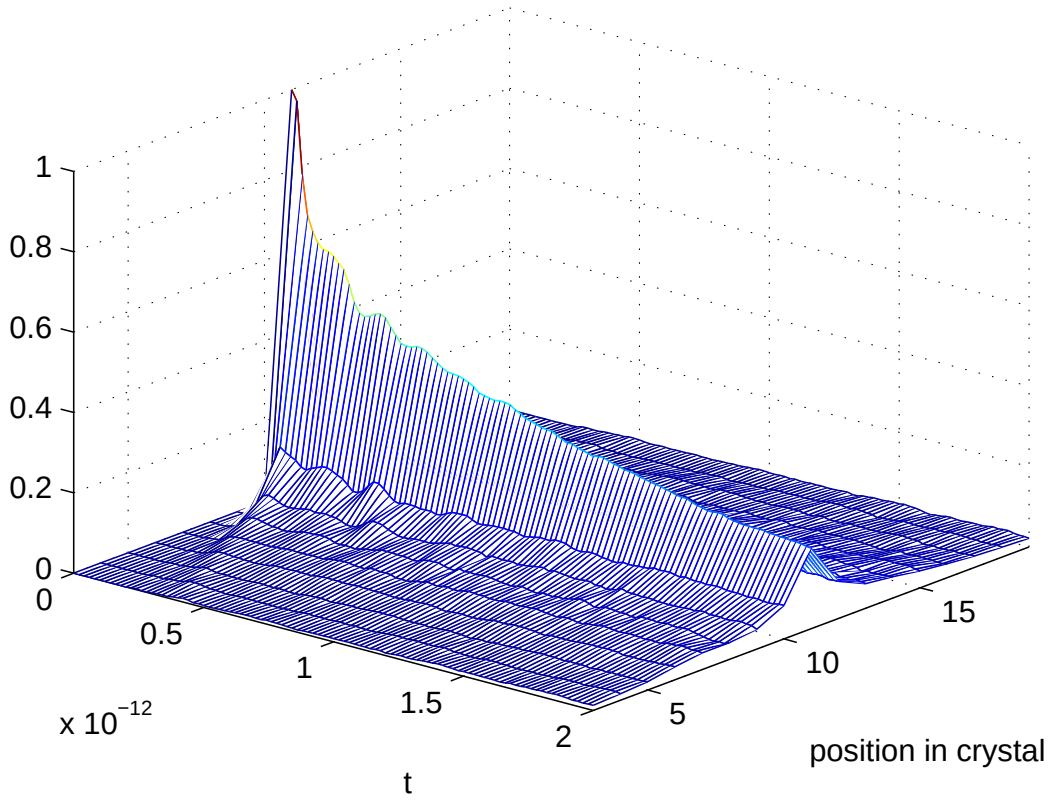


Figure 4.7: Distribution of the copper atom's position for the optimal prediction system

could be an interesting challenge for further work.

Also we have only investigated so far, under which circumstance optimal prediction does *not* work. The question, whether it holds as a valid approximation if non-equilibrium effects are small, will be answered in the following section.

4.4 An Experiment Without Sonic Waves

In Section 4.3 we showed that sonic waves are disturbing the pure diffusive behavior of the hopping of the copper atom. Since the diffusion should hold as an important criterion to compare the original system with the optimal prediction approximation, we have to eliminate the sonic waves' influence. In order to achieve this, we change the following parameters:

- Increase the number of atoms to $n = 70$, and place the copper atom the 22^{nd} position. For the optimal prediction system, we choose $m = 50$. This increases the

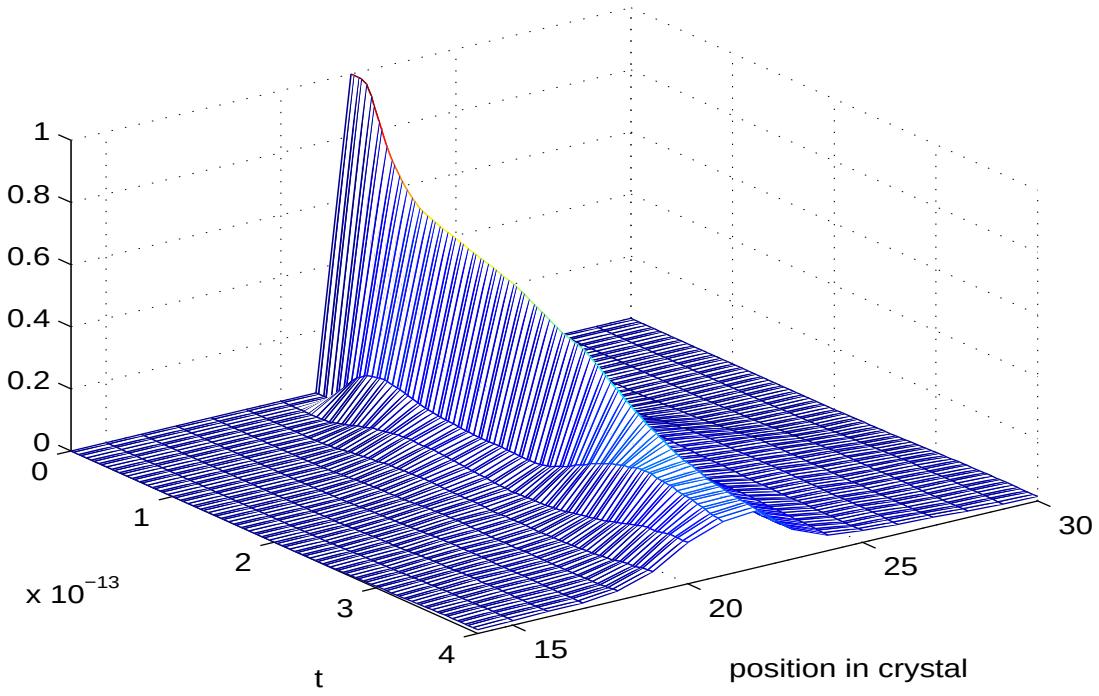


Figure 4.8: Distribution of the copper atom's position for the original system with $n = 70$

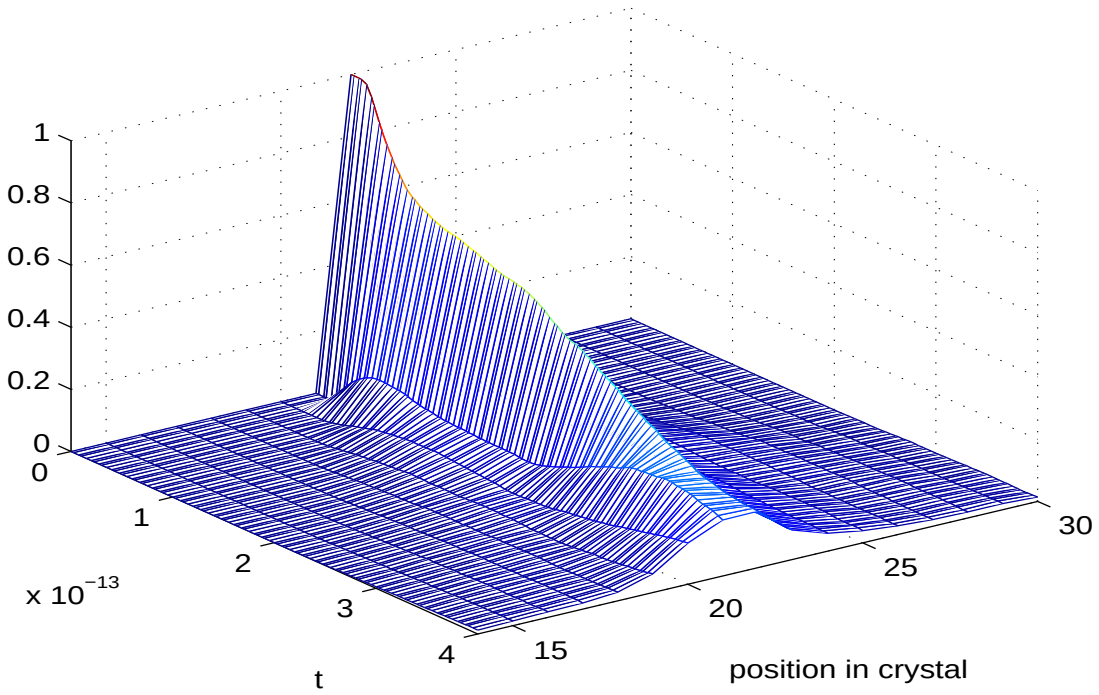


Figure 4.9: Distribution of the copper atom's position for optimal prediction approximation with $n = 70$, $m = 50$

distances s_{left} and s_{right} to

$$s_{\text{left}} = q_{22} - q_1 = 2.01 \cdot 10^{-9}\text{m}, \tag{4.43}$$

$$s_{\text{right}} = q_{70} - q_{22} = 3.64 \cdot 10^{-9}\text{m}. \tag{4.44}$$

Of importance for the optimal prediction system is the distance

$$\tilde{s}_{\text{right}} = q_{50} - q_{22} = 1.90 \cdot 10^{-9}\text{m}. \tag{4.45}$$

- Decrease the computation time to $t^* = 4.0 \cdot 10^{-13}$s.
- Increase the temperature to $T = 7000$K, in order to increase the diffusion.

These changes have the effect that the computation time t is longer than the time a sonic wave takes to travel from the copper atom to left boundary of the crystal, respectively to the first virtual particle, and vice versa, i.e.

$$ct^* = 1.8 \cdot 10^{-9}\text{m} < \min\left(s_{\text{left}}, s_{\text{right}}, \tilde{s}_{\text{right}}\right). \tag{4.46}$$

Hence, we can expect to observe no disturbances due to sonic waves, as we did in Section 4.3. We can afford to decrease the computation time such significantly, since we can re-obtain the required accuracy by taking more Monte-Carlo samples. For the following results, we used $N = 25000$ Monte-Carlo samples. As a further change, we increase the time step to $2.5 \cdot 10^{-15}$s, which we can afford to do, because the computation time t^* is so short.

Figure 4.8 shows the distribution of the position of the copper atom in the silicon crystal when solving the original system. In Figure 4.9 the analogous distribution can be seen for the optimal prediction system. Unlike the numerical results for the first experiment in Section 4.3, these two distributions look quite similar, and the features from the first experiment due to sonic waves cannot be observed here.

Still, the distributions show unexpected aspects. One of these aspects is the time evolution of the maximum, which is obviously not equal to the behavior of the maximum for the heat equation. Another aspect is the asymmetry in the distribution, which shows up after $t = 3.0 \cdot 10^{-13}$s. While we cannot provide explanations for these effects, we will observe connections in the following analysis of the results.

4.4.1 Estimating the Monte-Carlo Error

We can only trust the numerical results in this chapter, if the error due to Monte-Carlo sampling is significantly smaller than the difference between the results of the original and the optimal prediction system. Hence, we need to estimate the Monte-Carlo error. We

will do this estimate for the distribution of the original system, as shown in Figure 4.8. Similar results are obtained for the computations with the optimal prediction system.

Let $v_N(x,t)$ denote the distribution which is obtained numerically using N Monte-Carlo samples. Let further $u(x,t)$ denote the truth, which we wish to approximate, and let $e_N = \|v_N(x,t) - u(x,t)\|_{\max}$ denote the error. Typically, in Monte-Carlo sampling, the error decays like

$$e_N = O\left(N^{-\frac{1}{2}}\right). \tag{4.47}$$

Consider $N^* = 25000$ to be the number of samples in our experiment. In order to estimate the error e_{N^*}, consider experiments with $N_1, \ldots, N_k$ samples, where $N_i \ll N^* \ \forall i = 1, \ldots, k$. The differences with respect to $v_{N^*}(x,t)$ may be denoted by $\tilde{e}_{N_i} = \|v_{N_i}(x,t) - v_{N^*}(x,t)\|_{\max}$. Under these circumstances we can assume that

$$\tilde{e}_{N_i} \approx e_{N_i} = O\left(N_i^{-\frac{1}{2}}\right). \tag{4.48}$$

The values $\tilde{e}_k$ can be computed numerically. Plotted on a log-log scale they should lie on a line with slope $-\frac{1}{2}$. Extrapolating this line to the point N^* should yield an estimate for the error due to Monte-Carlo sampling for N^* samples.

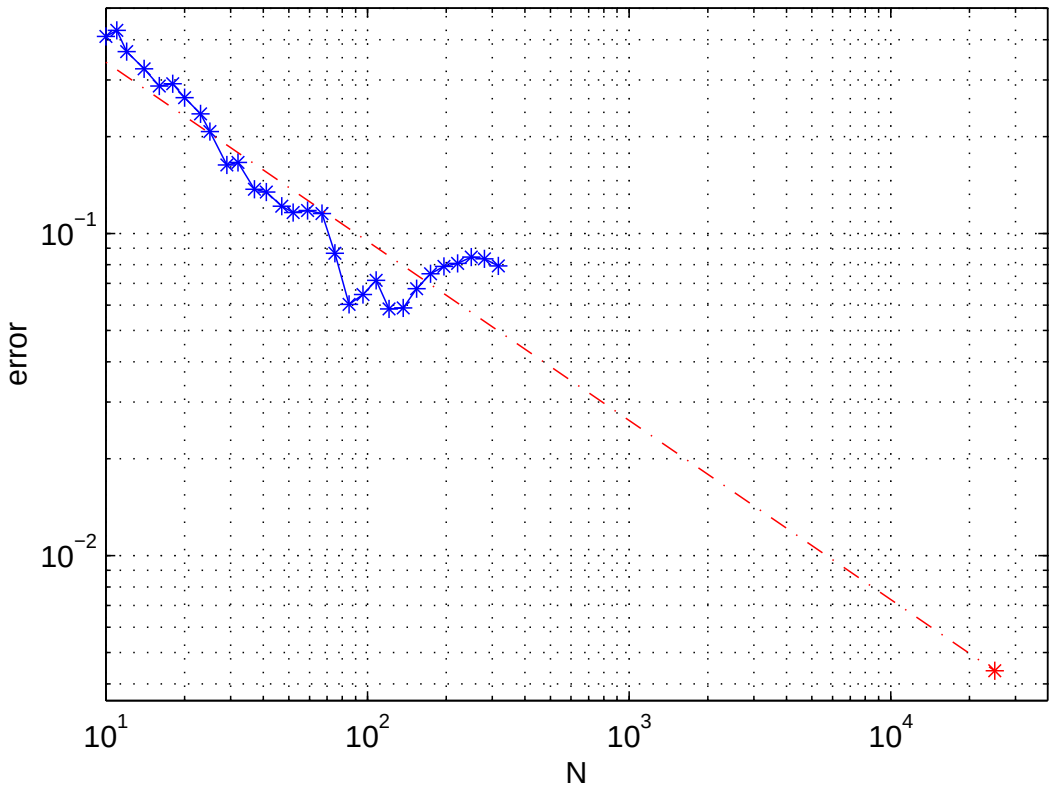


Figure 4.10: Estimating the error due to Monte-Carlo sampling

Figure 4.10 shows this idea applied to our current problem. We obtain an error of about $e_{N^*} \approx 4.4 \cdot 10^{-3}$. Compared to the distance to the optimal prediction plot, as shown in the following subsection, the error due to Monte-Carlo sampling turns out to be indeed negligibly small.

4.4.2 The Error in Distribution

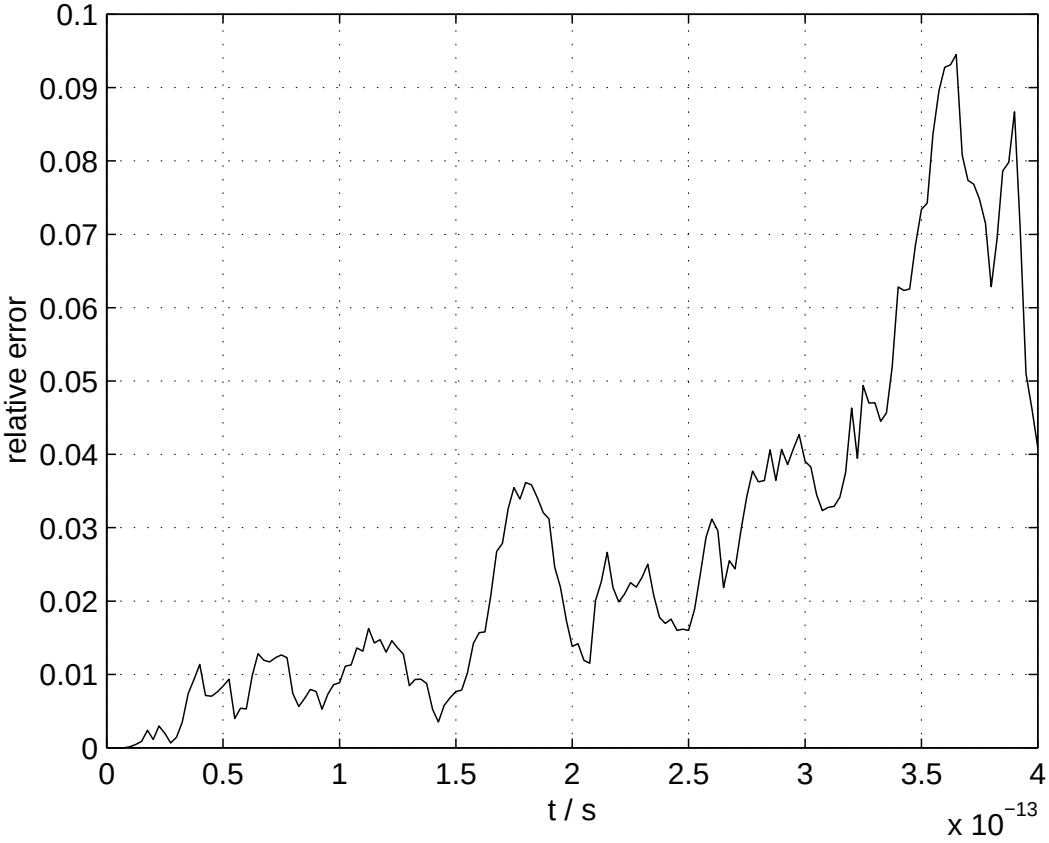


Figure 4.11: Relative error in the two distributions

Let $v(x,t)$ denote the numerically obtained distribution for the original system and $\tilde{v}(x,t)$ the numerically obtained distribution for the optimal prediction approximation. As a first comparison of these two distributions, we consider the difference between them, respectively, the error which is made in $\tilde{v}(x,t)$ with respect to $v(x,t)$. However, we do not want to consider the absolute error

$$|v(x,t) - \tilde{v}(x,t)|, \tag{4.49}$$

since the same difference is less relevant where the distribution $v(x,t)$ is large, and more relevant where the distribution $v(x,t)$ is small. Therefore we wish to consider the relative error. As the distribution is almost zero for large x, dividing the difference by $v(x,t)$ at

each point will make the relative error go to infinity for large x. Hence we consider the expression

$$e(t) = \frac{\max_x |v(x,t) - \tilde{v}(x,t)|}{\max_x |v(x,t)|} \tag{4.50}$$

as the relative error. By taking for each t the maximum over all x we eliminate the effects described above.

Figure 4.11 shows the relative error $e(t)$ over time. Note that up to the time $t = 3.0 \cdot 10^{-13}$s the relative error is less than 4.5%. After this time the relative error increases to 9.4%. Apart from errors simply due to the Monte-Carlo sampling, this increase could very well be connected to the asymmetries, which could be observed in the plots in Figure 4.8 and Figure 4.9. As a first result in comparing the two system we can conclude that a relative error of less than 10% (less than 5% up to time $t = 3.0 \cdot 10^{-13}$s), is not overwhelmingly small, but indicates very well that the distributions are quite similar. The following analysis of diffusion parameters should provide a more physically motivated comparison.

4.4.3 Diffusion Parameters

In the following we wish to compute the diffusion parameters $\kappa(t)$ for the two systems as shown in Figure 4.8 and Figure 4.9, using the method described in Subsection 4.2.2. For applying this method, we first need to obtain the parameters of the random walk model as described in Subsection 4.2.1, i.e. the sequence $(p_i)_{i\in\mathbb{N}}$, where p_i is the probability that in a hopping event the copper atom hops over exactly i silicon atoms. Since we are dealing with finite crystal, we will consider a vector $(p_1, \ldots, p_k)$, where $\sum_{i=1}^{k} p_i = \frac{1}{2}$. Note that due to the importance of the variance (4.14), the question which value for k is reasonable depends on the decay of $(i^2 p_i)_{i\in\mathbb{N}}$, rather than the decay of $(p_i)_{i\in\mathbb{N}}$ itself.

The values $p_1, \ldots, p_k$ can be approximated by Monte-Carlo sampling. We use the experiments for the original system, which we used to obtain the distribution shown in Figure 4.8. Since in each of the $N = 25000$ Monte-Carlo experiments we saved the position of the copper atom over time, we can precisely locate any hopping event. As the copper atom does not reach too large velocities, in the data the copper atom always hops over exactly one silicon atom. Since we wish to interpret two or more directly consecutive hopping events into the same direction as one large hopping event, we fix a time ($\Delta t_1 = 6.0 \cdot 10^{-14}$s worked well for our data) in which consecutive hopping events are clustered to a single one.

Before this evaluation can be done, a technical problem has to be solved, namely the elimination of double-hopping events into different directions.

Figure 4.12 and Figure 4.13 show two typical examples of such cases. In Figure 4.12 two "hopping" events follow very closely, one to the right followed by one to the left. But obviously this is not what we would consider a real hopping event, since it is due to the

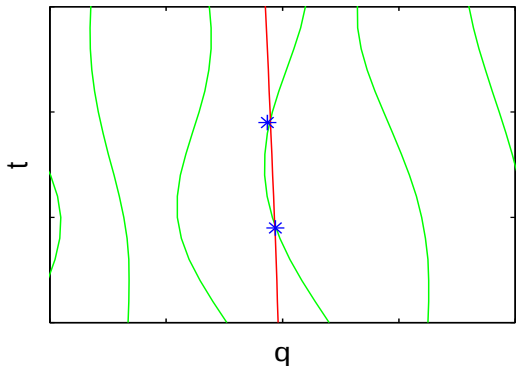


Figure 4.12: A hop hop-back event

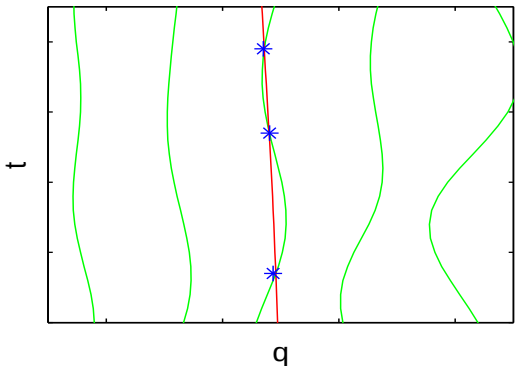


Figure 4.13: A hop hop-back hop event

momentum of the neighboring silicon atom, and does effectively not affect the position of the copper atom. Consequently, the event shown in Figure 4.13 should be counted as only one hopping event to the left. We achieve to treat such cases correctly by eliminating all events as shown in Figure 4.12, given they happen in a short time interval of length $\Delta t_2 = 2.0 \cdot 10^{-14}$s, before the clustering of hopping events is done.

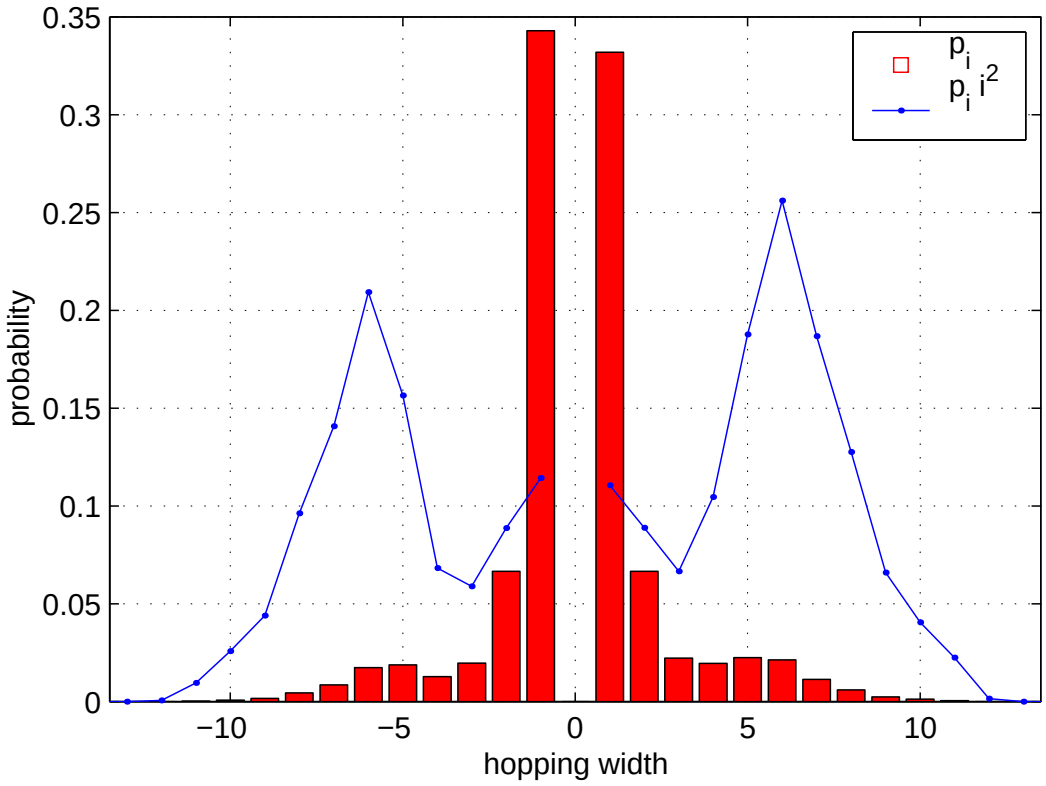


Figure 4.14: Probabilities of hopping over more than a single atom

The numerical results of the process described above are shown in Figure 4.14. The hopping probabilities are given by the red box histogram. Note that the hopping probabilities to the left are not exactly the same as the probabilities to hop to the right. This is partly due to Monte-Carlo errors, partly due to boundary effects from those experiments where the copper atom comes close to a boundary of the crystal. Since we wish to consider a symmetric random walk, we choose as values $p_1, \ldots, p_k$ the average probabilities of hopping to the left and hopping to the right from the results shown in Figure 4.14.

As derived in Subsection 4.2.1, the variance of the above distribution $\sum_{i=1}^{k} i^2 p_i$ is proportional to the diffusion constant κ in the random-walk model (see equation (4.16)). In Figure 4.14 the values $i^2 p_i$ are denoted by the blue curve (scaled to fit into the same plot).

As Figure 4.14 indicates, only $p_1, \ldots, p_{11}$ are greater than zero. Hence, we choose $k = 11$ for the computation of the diffusion parameters.

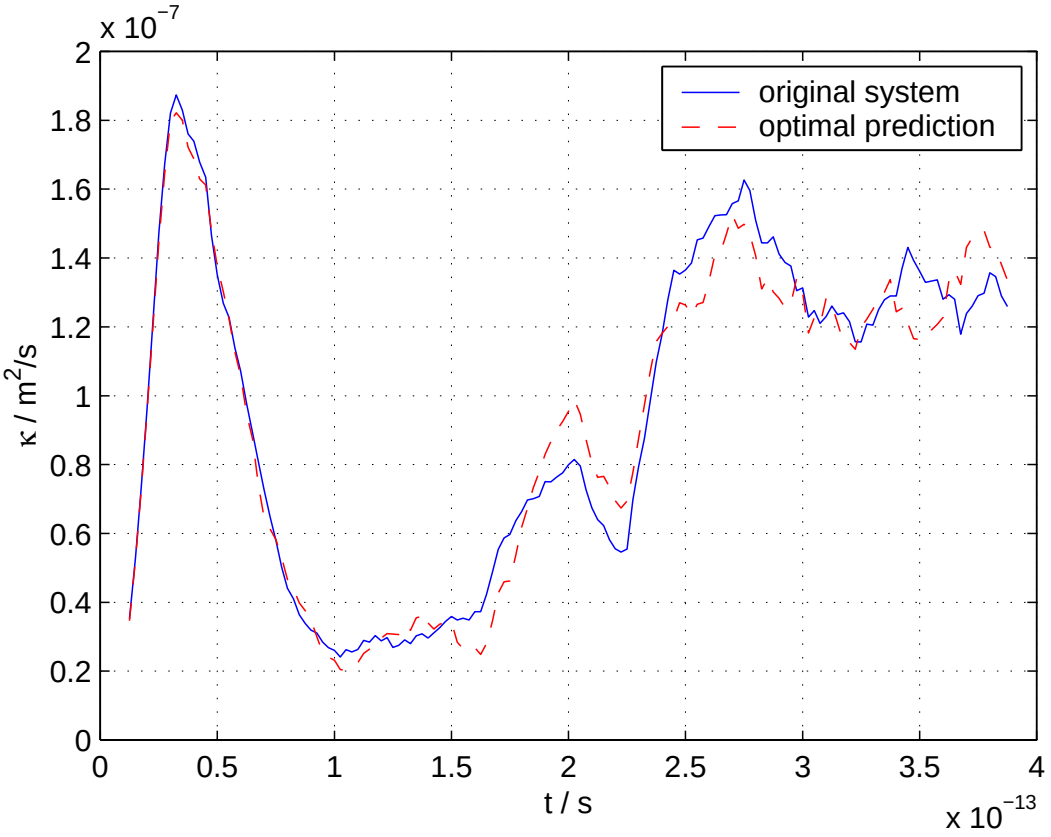


Figure 4.15: Diffusion parameters over time

Once the values $p_1, \ldots, p_{11}$ are obtained, the time-dependent diffusion parameters can be computed using the method described in Subsection 4.2.2. Figure 4.15 shows the so obtained diffusion parameters $\kappa(t)$ for the original system and the optimal prediction approximation. In both cases we chose $\Delta t = 2.5 \cdot 10^{-14}$s. Note that in the figure the

value $\kappa(t)$, which corresponds to the diffusion on the interval $[t, t+\Delta t)$, is plotted at time $t + \frac{\Delta t}{2}$. Important observations and estimates are:

- The diffusion parameters $\kappa(t)$ are far from being constant. Instead, they start close to 0, have a strong peak at $t = 3.0 \cdot 10^{-14}$s, after which they rapidly decay to a small value again, followed by a slow increase to a moderate value, around which they fluctuate from then on. Since sonic waves cannot be the reason for this behavior, it can be due to the following aspects:
  - The random walk model as described in Subsection 4.2.1 is just a simplified model, which need not represent the actual hopping behavior accurately enough. An investigation of this aspect and finding better models for the hopping process would be first steps in further work on this issue.
  - Due to the choice of initial values the initial configuration is not exactly in thermodynamical equilibrium. Hence, the strong changes over time in $\kappa(t)$ may express the system approaching its equilibrium.
- Considering that $\kappa(t)$ shows such pronounced behavior, the two functions $\kappa(t)$ for the original system, and $\tilde{\kappa}(t)$ for the optimal prediction approximation are remarkably close to each other. Up to time $t = 1.4 \cdot 10^{-13}$s the the distance between the two curves is quite small. After that time the error becomes larger, but shows the same features, in particular they fluctuate around the same value after time $t = 3.0 \cdot 10^{-13}$s.
- In order to roughly estimate if the obtained diffusion parameter is reasonable, we compare it with a correct diffusion constant measured in reality. In [5] the diffusion constant of copper in a silicon crystal at a temperature of $T_{\mathrm{r}} = 1273$K is given as $\kappa_{\mathrm{r}} = 4.4 \cdot 10^{-10} \frac{m^2}{s}$. In [25] the dependence of the diffusion constant on temperature is given by the relation
$$\kappa = \kappa_0 \cdot \exp\left(-\frac{\Delta E}{k_B T}\right). \tag{4.51}$$
Since the process of hopping is technically more complicated in three dimensions, no precise and unique value for the energy barrier ΔE, which the copper atom has to overcome, can be given. A reasonable value is $\Delta E_{\mathrm{r}} \approx 2$eV. This yields for the copper diffusion in reality
$$\exp\left(-\frac{\Delta E_{\mathrm{r}}}{k_B T_{\mathrm{r}}}\right) \approx 6.3. \tag{4.52}$$
In our model problem we had $T_{\mathrm{m}} = 7000$K and $\Delta E_{\mathrm{m}} = 0.43$eV, which yields:
$$\exp\left(-\frac{\Delta E_{\mathrm{m}}}{k_B T_{\mathrm{m}}}\right) \approx 1.0. \tag{4.53}$$
Employing relation (4.51) we obtain that
$$\kappa_{\mathrm{m}} \approx \kappa_{\mathrm{r}} \cdot e^{6.3-1.0} \approx 200 \cdot \kappa_{\mathrm{r}} \approx 9 \cdot 10^{-8} \frac{m^2}{s}. \tag{4.54}$$

A look at Figure 4.15 shows that our experimentally obtained diffusion parameters are indeed in this region.

Summarizing the results from Subsection 4.4.2 and Subsection 4.4.3 one can conclude that the optimal prediction approximation reproduces the diffusion process due to hopping of the copper atom quite accurately.

4.4.4 The Number of Hopping Events

In the previous Subsections we have seen that the diffusive behavior of the hopping of the copper atom is preserved by optimal prediction. To be a valid approximation, optimal prediction should preserve the nature of hopping itself, in particular the number of hopping events which happen up to a given time t^*.

Figure 4.16 shows the number of hopping events for the computation described in Section 4.4. The solid blue line represents the original system, and the dashed red line stands for the optimal prediction system. Plotted are the histograms, i.e. the ratio of all experiments which had exactly N hopping events over the number of hopping events N. Obviously the two graphs are close enough to each other, that one can speak of the same hopping behavior.

It should be remarked that for the experiment which is described in Section 4.3 the corresponding curves, as shown in Figure 4.17, show the same behavior, but are not such close as the ones shown in Figure 4.16. Obviously the sonic waves in the first experiment, which were not reproduced correctly by the optimal prediction system, disturbed also the nature of hopping in principle. In the absence of sonic waves, however, optimal prediction preserves the nature of hopping. Note that the results shown in Figure 4.16 are obtained at a temperature of $T = 7000$K, while the results shown in Figure 4.17 are obtained at a temperature of $T = 4000$K. Hence the plots in Figure 4.17 have their maximum at zero.

4.4.5 Energy Fluctuations

While the total energy is constant for the original system as well as for the optimal prediction approximation, the energy of the first m atoms

$$E_{\text{left}}(t) = \frac{1}{2}\sum_{i=1}^{m}\frac{p_i^2(t)}{m_i} + \sum_{\substack{i,j=1\\ i<j}}^{m} f_\alpha(q_i(t) - q_j(t)) \tag{4.55}$$

fluctuates over time. This expression fluctuates also for the optimal prediction, since the influence of the virtual atoms is neglected in the potential energy. The fluctuations in

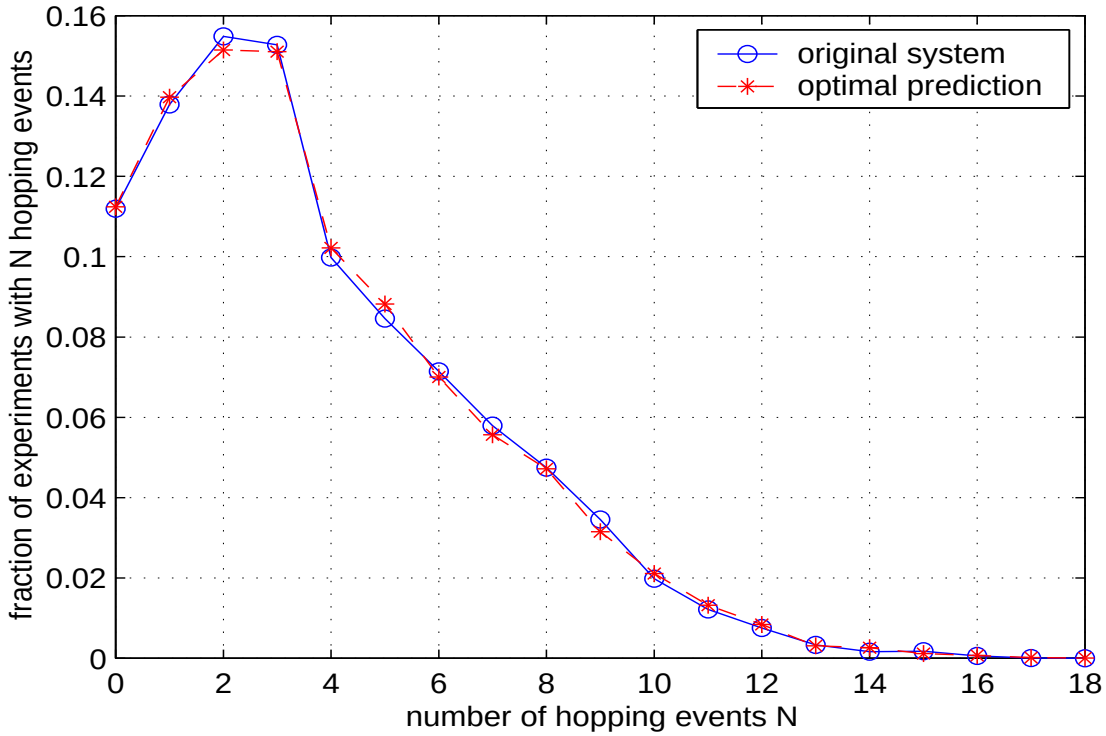


Figure 4.16: Number of hopping events for $n = 70$ atoms at $T = 7000$K

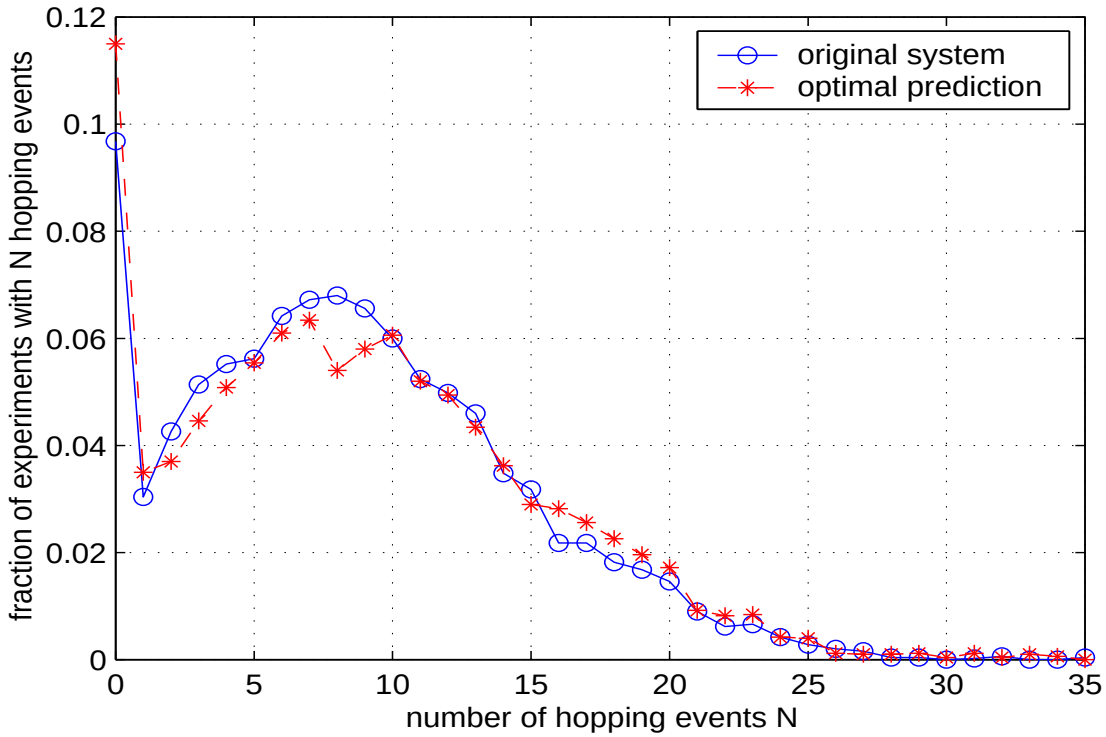


Figure 4.17: Number of hopping events for $n = 35$ atoms at $T = 4000$K

(4.55) represent the exchange of energy between atoms, which is a quantity that should be preserved by optimal prediction. Since for optimal prediction we consider the energy of exactly the real particles, this test enlightens the exchange of energy between real and virtual atoms.

For each of the $N = 25000$ experiments we consider the variance of (4.55) over time

$$V = \int_{t=0}^{t^*} (E_{\text{left}}(t) - E_{\text{left}}(0))^2 \, \mathrm{d}t, \tag{4.56}$$

which measures the impact of fluctuation. Hence we obtain N values $V_1, \ldots, V_N$ for the original system and for the optimal prediction approximation.

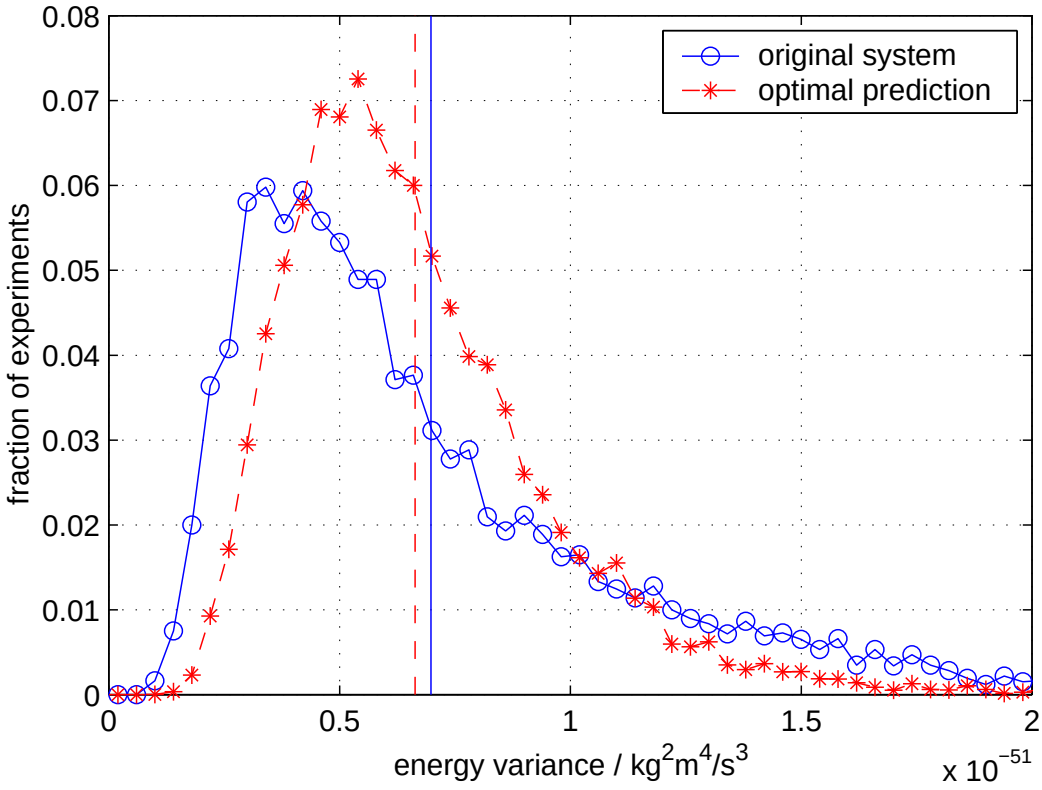


Figure 4.18: Histogram of total energy fluctuation for the two systems

Figure 4.18 shows the histogram for the variances V_i for the two systems. The solid blue line stands for the original system, and the dashed red line corresponds to optimal prediction. The average energy fluctuations for both systems are denoted by the blue and red vertical lines. While the average fluctuations are close for the two systems, and the two distributions look similar in principle, they obviously do not coincide. For optimal prediction most energy fluctuations are stronger than for the original system. On the other hand, particularly high energy fluctuations happen more frequently in the original system.

Possible physical reasons for this behavior could be:

- The fact, that in optimal prediction most fluctuations are stronger than in the original system, can be due to the additional fluctuative "Langevin" terms in expression (1.25).
- The high fluctuations in the original system can happen, since around the m^{th} atom free energy exchange can take place among a whole cluster of atoms. In optimal prediction, on the hand, there is no free energy exchange between the virtual atoms, since they have no momentum, but instead always follow the potential minimum.

Apart from these physical explanations, the difference in the distributions in Figure 4.18 can of course also be due to the error in the asymptotic approximation in Subsection 3.2.1. Remember that we only used the zeroth order approximation, which places the virtual particles always at the potential minimum. This could be too inflexible to correctly simulate the energy exchange. On the other hand could the first order approximation model the rate of exchange of energy more accurately. Also remember that our computations were done at strongly increased temperatures ($T = 7000$K). At "correct" temperatures ($T = 500$K) the energy fluctuations in optimal prediction might be much closer to the truth.

Chapter 5

Conclusions and Outlook

In this thesis, we could successfully apply the method of optimal prediction to a model problem in the context of molecular dynamics. The model problem was chosen to simulate the important effects which arise in a surface coating process between copper and silicon, in particular the effect of atomic *hopping*.

Under the assumption of a low temperature, asymptotic methods were applied to resolve the conditional expectations, which arise in optimal prediction. The zeroth order asymptotic expansion was used to derive a new and smaller system of equations. As a direct result of the asymptotic method we obtained a system which was formally smaller than the original system, but did not efficiently decrease the computational effort. Employing the property that potentials in molecular dynamics range only over short distances, we could derive a new system of equations, which gave rise to an interpretation as a *boundary layer condition*. This system yielded an obvious computational speed-up for our model problem.

In order to investigate, if the approximate optimal prediction system preserves the important properties of the original system, we provided various criteria, which could be checked by numerical experiments. In particular, we modeled and investigated the diffusion process of a single copper atom due to atomic hopping. While on the one hand the diffusion process turned out to be of a complicated nature, on the other hand the optimal prediction approximation showed to be very close to the original system in its diffusive behavior. Other statistical quantities, as the number of hopping events, were also preserved by the optimal prediction system.

On the other hand various examples showed under which conditions optimal prediction fails. In particular, sonic waves and other non-equilibrium effects caused a very different behavior between the original system and optimal prediction, which assumes the system to be in a perfect equilibrium.

A natural next step in research on this topic would be to apply the method to a more complex problem, possibly in two or three space dimensions, and with focus on other effects

than atomic hopping. Many of the results shown in this thesis should be transferable to three space dimensions in a straightforward manner, but various other aspects will become more problematic. On the other hand one can expect that non-equilibrium effects, as sonic waves, be smaller in effect in three space dimensions, and the fraction of atoms, which can be averaged out, be larger. Also many of the statistical quantities, which we considered at strongly increased temperatures, should be approximated much better at the physically correct temperatures.

Another obvious step in a further investigation is to analyze the first order asymptotic expansion, which was derived in Chapter 3, and to use it for a derivation of an improved system, which preserves physical quantities, as e.g. the exchange of energy between atoms, more accurately.

A further question, which is to be investigated, is how to generalize the methods presented herein to problems not in equilibrium, in particular how to introduce non-equilibrium effects into the equations of motion derived in Chapter 3.

Summarizing the results, we have shown that optimal prediction can in principle be applied to problems in molecular dynamics which take place at comparably low temperatures. Although the method turned out to react sensitively to non-equilibrium effects and application together with other approximations, it showed to approximate the truth well in equilibrium. Using the boundary layer condition version we could achieve an obvious speed-up. While the speed-up by optimal prediction itself cannot compete with the speed-up gained by other methods (e.g. the fast multipole method in a similar field), the emphasis lies more on the physical interpretation and on the possibility to apply it in principle together with other methods. Altogether, the method of optimal prediction can be called “promising” in the context of molecular dynamics.

Acknowledgements

In the end I wish to thank ...

- Prof. Dr. Helmut Neunzert for his effort in finding this interesting problem, which ideas both from Kaiserslautern and from Berkeley could be applied to. Furthermore I wish to thank for his encouragement and him always providing useful ideas in which direction else to look and go.
- Dr. Thomas Götz for many hours of interesting and helpful discussions. He always took time for my questions and problems, and always helped me to judge my own mathematical point of view, even in the deepest physical discussions.
- Dr. Peter Klein for providing the physical counterpart to Dr. Thomas Götz. He truly explained the "deeper secrets" of statistical mechanics to someone with previously no background in this field: me.
- Prof. Alexandre J. Chorin and Prof. Ole H. Hald for teaching me the ideas of optimal prediction in lots of interesting conversations and discussions, in a friendly and helpful, and yet challenging, atmosphere at the Lawrence Berkeley Laboratory and the UC Berkeley. I also wish to thank Prof. Chorin for reading the draft of my thesis and giving interesting and useful comments.
- Dr. Michael Junk for useful advice in Monte-Carlo sampling and Dr. Jochen Voß for his advice on stochastic processes.
- Mr. Lutz Justen for reading the draft of this thesis and giving interesting comments.
- Ms. Theresa Chow, Mr. Eugene Ingerman, Mr. Kevin Lin, Mr. Pavel Okunev and Ms. Helen Shvets for helpful discussions and comments during my work on optimal prediction and for helping me having a great time in Berkeley.
- Mr. Otto A. Wipprecht and the people from his foundation, whose scholarship made my two semesters in Berkeley possible.
- My parents and family for continuous support in just all possible fields.

- Ms. Diana Lemke for her support during my work on this theses and for bearing my periods of shifted day-night-cycles while typing.
- Mr. Lutz Justen, Mr. Claus Massion, Ms. Steffi Müller, Mr. Niko Ruf, the Frisbee players from the "Fabulous Ultimate Club Kaiserslautern", and many, many others who helped me not *only* to do mathematics during my work on this thesis.

List of Figures

2.1 The one-dimensional silicon crystal with a copper atom 16
2.2 Potential f_1 between two silicon atoms 17
2.3 Potential f_2 between a copper and a silicon atom 18
2.4 Potential f_3 between two copper atoms 18
2.5 A simulation of the model problem 22

3.1 The asymptotic approximation for different ε 32
3.2 The decay of the approximation error depending on ε 32
3.3 Six global minimizers for (q_2, q_3) 34
3.4 An "unphysical" potential which does not yield unique minimizers 35
3.5 The minimizer inside $M_{\hat{q}}$ need not be the global minimizer 35
3.6 A simulation of the optimal prediction system 45
3.7 A simulation of the optimal prediction approximation for an infinite crystal 46
3.8 The original problem with cut-off potentials 47
3.9 The optimal prediction approximation with cut-off potentials 47
3.10 CPU times for the three systems depending on n and m 51
3.11 Speed-up factors depending on n and m. 51
3.12 Speed-up factors for $n = 2m$ 51

4.1 Distribution of the copper atom's position (viewed from the side) 66
4.2 Distribution of the copper atom's position (viewed from the top) 66

4.3 A sonic wave traveling through the crystal 67

4.4 Behavior of sonic waves in the original system 70

4.5 Behavior of sonic waves in the optimal prediction approximation 70

4.6 A sonic wave in the optimal prediction system 71

4.7 Distribution of the copper atom's position for the optimal prediction system 72

4.8 Distribution of the copper atom's position for the original system with $n = 70$ 73

4.9 Distribution of the copper atom's position for optimal prediction approximation with $n = 70$, $m = 50$ 73

4.10 Estimating the error due to Monte-Carlo sampling 75

4.11 Relative error in the two distributions 76

4.12 A hop hop-back event 78

4.13 A hop hop-back hop event 78

4.14 Probabilities of hopping over more than a single atom 78

4.15 Diffusion parameters over time 79

4.16 Number of hopping events for $n = 70$ atoms at $T = 7000$K 82

4.17 Number of hopping events for $n = 35$ atoms at $T = 4000$K 82

4.18 Histogram of total energy fluctuation for the two systems 83

Bibliography

[1] M.P. Allen, D.J. Tildesley, *Computer simulation of liquids*, Oxford Science Publications, 1987.

[2] J. Barnes, P. Hut, *A hierarchical $o(n \log n)$ force-calculation algorithm*, Nature (London), 324:446-449, 1986.

[3] J. Bell, A.J. Chorin and W. Crutchfield, *Stochastic optimal prediction with application to averaged Euler equations*, Proc. 7th Nat. Conf. Comput. Fluid Mech., C.A. Lin (ed), Pingtung, Taiwan, 2000, pp. 1-13.

[4] B.J. Berne, J.D. Gezelter, E. Rabani, *Calculating the hopping rate for self-diffusion on rough potential energy surfaces: Cage correlations*, J. Chem. Phys. 107, No. 17, 1997.

[5] L.J. Chen, S.L. Cheng, H.H. Lin, *The Failure Mechanisms and Phase Formation For Ni, Co and Cu Contacts on Ion Implanted (001)Si under high current stress*, Nucl. Instr. and Meth.in Phys. Res. B 169, pp. 161-165, 2000.

[6] A.J. Chorin, O. Hald and R. Kupferman, *Optimal prediction and the Mori-Zwanzig representation of irreversible processes*, Proc. Nat. Acad. Sc. USA, 97, 2000, pp. 2968-2973.

[7] A.J. Chorin, O. Hald and R. Kupferman, *Non-markovian optimal prediction*, Monte Carlo Methods and Applications, Vol. 7, No. 1-2, pp. 99-109, 2001.

[8] A.J. Chorin, O. Hald and R. Kupferman, *Optimal prediction with memory*, Physica D, 2002.

[9] A.J. Chorin, A. Kast and R. Kupferman, *Optimal prediction of underresolved dynamics*, Proc. Nat. Acad. Sc. USA, 95, 1998, pp. 4094-4098.

[10] A.J. Chorin, A. Kast and R. Kupferman, *Unresolved computation and optimal prediction*, Comm. Pure Appl. Math., 52, 1999, pp. 1231-1254.

[11] A.J. Chorin, A. Kast and R. Kupferman, *On the prediction of large-scale dynamics using unresolved computations*, Contemp. Math., 238, 1999, pp. 53-75.

[12] A.J. Chorin, *Probability, Mechanics, and Irreversibility*, Lecture notes, UC Berkeley Math. Dept., 2000.

[13] A.J. Chorin, R. Kupferman and D. Levy, *Optimal prediction for Hamiltonian partial differential equations*, J. Comput. Phys., 162, 2000, pp. 267-297.

[14] P. Deuflhard, *From molecular dynamics to conformational dynamics in drug design*, ZIB-Report 02-20, 2002.

[15] L. Devroye, *Non-uniform random variate generation*, Springer, 1986.

[16] L.C. Evans, *Partial differential equations*, American Mathematical Society, 1998.

[17] D. Evans, G. Morriss, *Statistical mechanics of non-equilibrium liquids*, Academic Press, London, 1990.

[18] T. Fraunheim, H. Hensel, P. Klein, H.M. Urbassek, *Comparison of classical and tight-binding molecular dynamics for silicon growth*, Physical Review B, 1996.

[19] W.T. Grandy, Jr., *Foundations of statistical mechanics, Vol. I*, D. Reidel Publishing Company, 1987.

[20] L. Greengard, V. Rokhlin, *A fast algorithm for particle simulations*, J. Comp. Physics, 73:pp. 325-348, 1987.

[21] O. Hald, *Optimal prediction and the Klein-Gordon equation*, Proc. Nat. Acad. Sc. USA, 96, 1999, pp. 4774-4779.

[22] O. Hald and R. Kupferman, *Existence of orthogonal dynamics*, preprint, 2001.

[23] O. Hald and R. Kupferman, *Convergence of optimal prediction for nonlinear Hamiltonian systems*, submitted for publication, 2000.

[24] J.M. Hammersley, D.C. Handscomb, *Monte Carlo methods*, Chapman and Hall, 1979.

[25] J. P. Hansen, I. R. McDonald, *Theory of Simple Liquids*, Academic Press, London, 1976.

[26] S. Hasegawa, S. Ino, Z.H. Zhang, *Epitaxial growth of Cu onto Si(111) surfaces at low temperatures*, Surface Science, 1998.

[27] A. Kast, *Optimal prediction of stiff oscillatory mechanics*, Proc. Nat. Acad. Sci. USA, 97, (2000), in press.

[28] S. Kirkpatrick, C.D. Gelatt Jr, M.P. Vecchi, *Optimization by Simulated Annealing*, Science, V. 220, No. 4598, pp. 671-680, 1983.

[29] U. Krengel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg, 1991.

[30] G.F. Lawler, *Introduction to Stochastic Processes*, Chapman and Hall, London, 1995.

[31] M.H. Lee, *Fick's law, Green-Kubo formula, and Heisenberg's equation of motion*, Physical Review Letters, Col. 85, No. 12, 2000.

[32] A. Martin-Löf, *Statistical mechanics and the foundations of thermodynamics*, Lecture Notes in Physics, Springer, 1979.

[33] H. Mori, *Transport, collective motion, and Brownian motion*, Prog. Theor. Phys., 33, pp. 423-455, 1965.

[34] J.D. Murray, *Asymptotic analysis*, Springer, 1984.

[35] P. Okunev, *On comparative performance of three different algorithms for the non-markovian optimal prediction applied to the Hald system*, LBNL-Report, 2001.

[36] F.W.J. Olver, *Asymptotics and special functions*, Academic Press, 1974.

[37] B. Seibold, *Non-markovian optimal prediction with integro-differential equations*, LBNL-Report, 2001.

[38] Ch. Schütte, *A quasiresonant smoothing algorithm for solving large highly oscillatory differential equations from quantum chemistry*, Dissertation, FU Berlin, 1994.

[39] B.M. Smirnov, A.S. Yatsenko, *Properties of dimers*, Physics – Uspekhi 39, No. 3, 1996.

[40] A.M. Stuart, J.O. Warren, *Analysis and experiments for a computational model of a heat bath*, to appear in J. Stat. Phys. SCCM Technical Report SCCM-98-17, 1998.

[41] J.A. Venables, *Nucleation calculations in a pair-binding model*, Physical Review B, 1986.

[42] R. Zwanzig, *Problems in nonlinear transport theory, Systems far from equilibrium*, L. Garrido, ed., Springer, New York, pp. 198-225, 1980.