Cellular Monte Carlo Simulation of Coupled Electron and Phonon Dynamics

by

Flavio F.M. Sabatti

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2018 by the
Graduate Supervisory Committee:

Saraniti Marco , Chair
David J. Smith
Robert Wang
Stephen M. Goodnick

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

A novel Monte Carlo rejection technique for solving the phonon and electron Boltzmann Transport Equation (BTE), including full many-particle interactions, is presented in this work. This technique has been developed to explicitly model population-dependent scattering within the full-band Cellular Monte Carlo (CMC) framework to simulate electro-thermal transport in semiconductors, while ensuring the conservation of energy and momentum for each scattering event. The scattering algorithm directly solves the many-body problem accounting for the instantaneous distribution of the phonons. The general approach presented is capable of simulating any non-equilibrium phase-space distribution of phonons using the full phonon dispersion without the need of the approximations commonly used in previous Monte Carlo simulations. In particular, anharmonic interactions require no assumptions regarding the dominant modes responsible for anharmonic decay, while Normal and Umklapp scattering are treated on the same footing.

This work discusses details of the algorithmic implementation of the three particle scattering for the treatment of the anharmonic interactions between phonons, as well as treating isotope and impurity scattering within the same framework. The approach is then extended with a technique based on the multivariable Hawkes point process that has been developed to model the emission and the absorption process of phonons by electrons.

The simulation code was validated by comparison with both analytical, numerical, and experimental results; in particular, simulation results show close agreement with a wide range of experimental data such as the thermal conductivity as function of the isotopic composition, the temperature and the thin-film thickness.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Figure Page

LIST OF SYMBOLS

$a(+\boldsymbol{q})$ Creation operator of quantum state (or particle) represented by $\boldsymbol{q}$

$a(-\boldsymbol{q})$ Annihilation operator of quantum state (or particle) represented by $\boldsymbol{q}$

BTE Boltzmann transport equation

BZ1 First Brillouin zone

CMC Particle-based full band cellular Monte Carlo

C_v Volumetric heat capacity

$D(\mathcal{E}, \Omega_{\boldsymbol{q}})$ Density of state in the cell $\Omega_{\boldsymbol{q}}$ at the energy $\mathcal{E}$

e Elementary charge

$\mathcal{E}$ Energy

$\mathbb{E}$ Expectation operator

EMC particle-based Ensemble Monte Carlo

exp Exponential function

f Distribution function

$\mathcal{F}$ Fermi integral

f Target distribution

$\boldsymbol{G}$ Reciprocal lattice vector

g Target distribution

G Optical counterpart of the Grüneisen parameter

$\hbar$ Reduced Planck constant

$\boldsymbol{H}$ Hessian matrix

$\mathcal{H}$ History, or filtration, of a point process

h Exciting function in a point process

H	Hamiltonian
I	Overlap integral
$\mathcal{J}$	Step function
$\boldsymbol{k}$	Crystal wavevector
k_B	Boltzmann constant
$\mathcal{M}$	Counting measure
M	Mass
N_a	Number of atoms in the crystal
$\mathcal{D}$	Dynamical matrix
Γ	Transitioning rate between eigenstates
$\boldsymbol{\epsilon}$	polarization vector
N	Number of simulated particles
n	Occupation number of a state
T	Temperature in Kelvin
U	Potential energy of the crystal
n_0	Bose-Einstein distribution
$\mathbf{P}$	momentum for a crystal ion
$\mathcal{V}$	Poisson probability distribution
$P(E)$	Probability for the event E
$p(E)$	Probability distribution function for the event E
pdf	Probability density function
PP	Point process
$\boldsymbol{q}$	Crystal wavevector
$\boldsymbol{r}$	Position vector

ϱ_u	Random number
R	Rate for one specific particle to transitioning between eigenstates
$\mathbf{R}_b$	equilibrium position of the Bravais lattice
σ	Thermal conductivity
t	Time
$\mathbf{u}$	deviation from equilibrium for a crystal ion
$U(0,1)$	Uniform probability distribution
$\boldsymbol{v}$	particle velocity vector
V_c	Volume of the crystal
W^c	Number of physical charges represented by one simulated particle
W^p	Number of physical phonons represented by one simulated particle
X,Y	Random variables
x,y	Real numbers
$\mathrm{Z_c}$	Concentration of ionized impurities
β_s	Screening parameter
χ	Volume element in the crystal phase space
η	Number of physical particles
γ	Grüneisen parameter
λ	Intensity or rate of a process
ω	Phonon frequency
$\Omega_{\boldsymbol{q}}$	Region (cell) in momentum space centered around the wavevector $\boldsymbol{q}$
ρ	Mass density
Ξ	Scattering parameter

Chapter 1

INTRODUCTION

Effective thermal management and controlled heat dissipation have always been considered key aspects of device engineering because of the role they play on both the device performance and the reliability. In addition, sophisticated heat management techniques are required due to the increasing density and operation frequency of active and passive components. The particle-based Monte Carlo approach has been previously used [1–12] to solve the heat transport problem by simulating phonons explicitly. However, most of the previous works treat the three-state interactions of simulated phonons as one-state by assuming either near-equilibrium or the relaxation time approximation. At the end of each time step, particles are removed from the simulation domain according to a scattering rate expressed in the form $\Gamma(\boldsymbol{q}, T)$. Then, a set of newly initialized phonons is introduced in the system according to a distribution function, aimed at restoring thermal equilibrium and conserve energy. The energy of the ensemble is statistically conserved, but momentum conservation is either not addressed or treated only partially [4, 12]. Furthermore, in particle-based Monte Carlo, the electron-phonon scattering is modeled either indirectly via macroscopic parameters [13–15] such as temperature and heat-generation rate, or via a one dimensional histogram [16] where a virtual phonon is added/deleted every time an emission/absorption occurs. In the first approach, the phonon distribution has some pre-defined shape, usually the Bose-Einstein. In the second one, the histogram is computed over one particular component of the wavevector [17] or via the magnitude of the phonon wavevector [16] (the phonon momentum space is assumed to be isotropic).

The present work proposes a Monte Carlo scattering algorithm designed to model non-equilibrium phonon distributions. The algorithm has been developed within the 3D Cellular Monte Carlo (CMC) full-band (dispersion) numerical framework [18], where transition probabilities between states before and after the scattering are precomputed and stored in a look-up table, which is loaded at the beginning of the simulation.

During the simulation, the precomputed table is used to determine the three states involved in the scattering, and a rejection technique is used to solve the scattering operator based on the local current phonon distribution without assumptions on its shape. All the states interact simultaneously, which allows enforcing the conservation of energy and momentum in each scattering event. Within this approach, all scattering processes are treated equally, and no formal distinction is required between Umklapp and Normal processes, while selection rules and decay paths are the results of momentum and energy conservation, rather than externally imposed hypotheses.

The population-dependent scattering technique introduced for phonons interactions is then extended to model electron-phonon scattering when both populations are treated as synchronous ensembles of simulated particles. In particular, the problem of modeling the interaction of simulated particles representing populations of largely different size is discussed. Two approaches are proposed to solve this problem: the first approach is simply based on modeling phonon emission and absorption as a Poisson distribution, while the second approach is based on the more sophisticated multivariable Hawkes point process.

Chapter 1 offers an overview on the BTE and the Monte Carlo method. Chapter 2 investigates the theoretical aspects of the rejection technique and presents the implementation within the CMC framework. The modeling of the electron-phonon interactions in terms of The Poisson distribution and Hawkes process is presented

in Chapter 3, electrons and phonons collision mechanisms are discussed in Chapters 4 and 5 respectively. Chapter 6 presents the models for the particle dynamics algorithm, while 7 presents the data structures employed in this work. Finally, in Chapter 8 the numerical results are presented.

Chapter 2

FULL BAND PARTICLE-BASED METHOD

This chapter examines the full band cellular Monte Carlo (CMC) approach for the particle-based simulation of electron devices. The first section discusses the history of the method and the assumptions leading to the formulation of Boltzmann Transport Equation (BTE) for the electrons and the analogous expression for the phonons. Section 2.2 offers an overview of the particles-based Monte Carlo as a technique to solve the BTE. Finally, the last two sections focus on the CMC particle-based simulation by discussing the basic implementation of the algorithm and the reasons behind this approach.

2.1 The Boltzmann Transport Equation

Within the semi-classical approach, charge transport in semiconductors can be described by the seven-dimensional distribution function $f(\mathbf{r}, \boldsymbol{q}, t)$, which carries information about the position $\boldsymbol{r}$ and the wavevector $\boldsymbol{q}$ of the particles at any time t. In other words, $f(\mathbf{r}, \boldsymbol{q}, t)$, provides a statistical description of the time evolution of the distribution of carriers in both momentum and position space, and it can be used to obtain various quantities of interest such as the mean energy of the carriers, their drift velocity, etc.

The time evolution of $f(\mathbf{r}, \boldsymbol{q}, t)$ is described by the Boltzmann Transport Equation [19]:

$$\frac{\partial f}{\partial t} + \frac{\partial \boldsymbol{r}}{\partial t} \cdot \nabla_{\boldsymbol{r}} f + \frac{\partial \boldsymbol{q}}{\partial t} \cdot \nabla_{\boldsymbol{q}} f = \left| \frac{\partial f}{\partial t} \right|_{Col}, \tag{2.1}$$

where the right side of the equation represent the collision integral describing the rate

of change of f because of the interactions of the carriers with their environment.

This equation was originally derived by Ludwig Boltzmann (1872) by merging mechanical concepts and statistical considerations to describe a gas of undeformable spheres. In particular, the BTE is a consequence of the Liouville theorem [20] and describes the effect of collisions on the carriers density in phase space under the following three assumptions: the system operate in the Boltzmann-Grad limit [21], that is the dimension of the particles tends to zero while the mean free path remains finite, the collisions are elastic, and two particles are uncorrelated before collisions (this is also known as the *Stosszahlansatz* or *Molecular Chaos* hypothesis). The first assumption requires particles to interact only via binary collisions, so interactions involving more than two particles are neglected; in addition, the collisions or scattering events must be localized in both space and time. The assumption of elastic collisions ensures the conservation of both momentum and energy. The last assumption allows the derivation of the time-asymmetric behavior of macroscopic systems from the chaotic, but time-symmetric, behavior of the microscopic dynamics. The breaking of the time-reversal symmetry ultimately ensures that the entropy increases only in the future. In other words, systems described within the framework of the Boltzmann equation obey the second law of thermodynamics and do indeed reach an equilibrium distribution.

Since a finite thermal conductivity requires accounting for three-particle (phonon) scattering (see Ch. 6), and the transport equation derived by Boltzmann assumes binary collisions, thermal transport in solids cannot be modeled by the BTE as described above. However, Rudolf Peierls [22], inspired by Boltzmann work, employed a similar statistical approach to model heat transport in terms of particles. His formulation allows for non-binary collisions, but it requires two additional assumptions: the occupation number contains all the information that is obtainable about the system,

and the phonon states are quantized. The result is an equation formally identical to the BTE which is referred to as the Peierls-BTE (PBTE), or the phonon BTE:

$$\frac{\partial f}{\partial t} + \frac{\partial \boldsymbol{r}(\boldsymbol{q}_l)}{\partial t} \cdot \nabla_{\boldsymbol{r}} f = \left| \frac{\partial f}{\partial t} \right|_{Col}, \tag{2.2}$$

where $\boldsymbol{q}_l$ is the momentum of the l-th phonon mode.

If we drop the mode index for the sake of clarity, the collision integral on the right side of Eq. 2.2 can be written as follows:

$$\left| \frac{\partial f}{\partial t} \right|_{Col} = \sum_{\boldsymbol{q} \subset BZ1} \sum_{m=1}^{M} \sum_{\boldsymbol{q}^1,...,\boldsymbol{q}^m} \Gamma^+(\boldsymbol{q}, \boldsymbol{q}^1, ..., \boldsymbol{q}^m, f) - \Gamma^-(\boldsymbol{q}, \boldsymbol{q}^1, ..., \boldsymbol{q}^m, f), \tag{2.3}$$

where $BZ1$ is the first Brillouin zone, m represents the number of states interacting with the initial state $\boldsymbol{q}$, M is the total number of phonon states, and $\Gamma^+(\boldsymbol{q}, .., f)$ and $\Gamma^-(\boldsymbol{q}, .., f)$ represent the rate of decrease and increase of phonons in the state $\boldsymbol{q}$, respectively.

In general, a close-form solution of this equation cannot be obtained without severe approximations being made on the nature of the collision integral (e.g. the relaxation time approximation [23]) or on the distribution function itself (e.g. the near-equilibrium approximation [24]). For this reason, both deterministic and stochastic integration schemes [3, 5, 9–11, 25–29] have been developed to determine numerically the distribution function. Deterministic approaches are often based on the lower order terms of the moment expansion of the distribution function [30], and approximations are always needed to obtain closure. In particular, the numerically calculated phonon dispersion cannot be included in such models.

Therefore, although these models have relatively low computational cost, they usually require reliable closure relations relating transport parameters [31, 32] and their validity domain is limited [33] by the accuracy of such parameters. Alternatively, the particle-based Ensemble Monte Carlo (EMC) method [34] provides a space-time solution of the full BTE which is exact [25] in statistical terms.

Within the EMC framework, a representative sample of the particle population in the system is directly simulated as individual particles. The simulated trajectories of the particles are tracked both in position and momentum space, while a Monte Carlo (stochastic) algorithm is employed to model the scattering processes. Finally, any physical information of interest (e.g. energy, particle concentration, diffusivity, etc.) can be easily extracted by averaging over the ensemble of simulated particles.

Compared to deterministic approaches, stochastic methods require less draconian approximations and offer an intuitive approach to modeling. However, these methods are based on complex and computationally intensive algorithms.

2.2 The Ensemble Monte Carlo Method

The EMC method has proven successful in modeling complicated phenomena in terms of elementary interactions [25], and by solving the BTE without the need of relying on the accuracy of macroscopic transport parameters, which are often an output of the EMC simulation itself. Moreover, EMC simulations allow for a valuable analysis of the microscopic individual phenomena responsible for the macroscopic behavior of the system as a whole.

Figure 2.1 shows the flowchart of a generic particle-based Monte Carlo algorithm. During the initialization step, the initial carrier distribution inside the computational domain representing the semiconductor device is computed according to some externally imposed conditions. These conditions correspond to imposing local charge neutrality and thermal equilibrium of the simulated particles. After the initialization, the main loop consists of four steps: data gathering, Cauchy problem solving, free flight, and scattering. The data gathering step collects, formats, and records the data to be used by other steps and for post-processing. Subsequently, one or more Cauchy problems [35] (typically field equations) are solved based on the system under study.

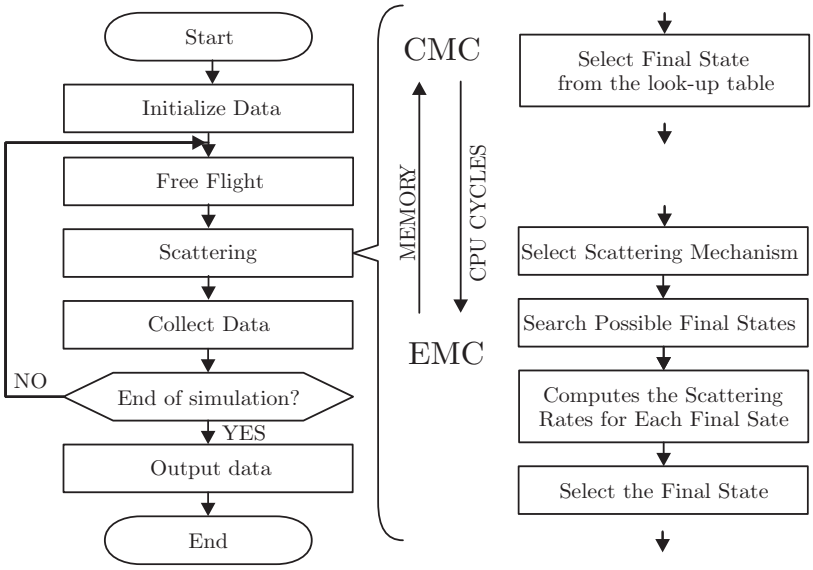

Figure 2.1: Flowchart of the particle-based Monte Carlo simulator. The right side expand on the differences between the EMC and the CMC approach. The EMC method has low memory requirements but is computationally intensive while the CMC requires more memory but is less computationally intensive.

This step allows accounting for the forces responsible for the carrier dynamics. For example, when simulating electrons, the charge distribution and the contact potentials are used to solve Poisson's equation [36] and to obtain the electric field distribution. Since carriers are moving, the carrier dynamics has to be periodically stopped and the field must be updated often enough so as to resolve plasma oscillations [37].

The free flight step calculates the carrier trajectories by employing a ray tracing algorithm to account for the device geometry, and a time integration scheme that accounts for the forces obtained in the previous step.

A stochastic Monte Carlo procedure is used to solve the collision integral by modeling the scattering as a Markovian process: the particle wavevector is changed instantaneously, with no memory of the state before the collision. The scattering rates in the collision integral are related by a recurrence relation, this reflects [38] the fact that carriers transition from one state to another in sequence. One should note that the rates of scattering are pre-calculated and tabulated according to material

characteristics such as the band structure, doping density, isotopic composition etc.

In the case of steady-state processes, the stability of the simulation allows to express the distribution function as the result of a convergent iterative process [38], where scattering is regarded as sequence of events causing the distribution to approach the steady state. The distribution function itself is formally expressed as the solution of a homogeneous integral equation. This allows to simplify the scattering algorithm by defining a new scattering process rate which leaves the state of a particle unchanged, this process is referred to as "self-scattering" [38], The self-scattering has no physical consequence: it modifies the probability functions while leaving the form of the recurrence relation unaltered.

A suitable [1] choice of the "self-scattering" rate allows equating each scattering to a time increment of the physical system. In other words, this approach allows the execution of the Monte Carlo algorithm for all particles at a constant time interval.

From the numerical point of view, an iterative solution of the integral-differential Boltzmann equation requires both differentiation and integration. In our approach, however, the iteration cycle involves only integration. Since errors in the numerical representation of a function tend to be accentuated by differentiation but reduced by integration, this approach is expected to have some numerical advantages over directly trying to solve iteratively the integro-differential BTE.

2.3 Cellular Monte Carlo Method

Traditional EMC simulators require a search over the whole discretized Brillouin zone after the occurrence of a scattering event in order to invert the energy/momentum relation, and find all the energy-conserving candidate cells in the

[1] The self scattering makes no net contribution to the rate of change of the distribution due to scattering and it could be arbitrarily defined. However, the self-scattering rate must be positive in order to model the distribution function as the result of a convergent iterative process

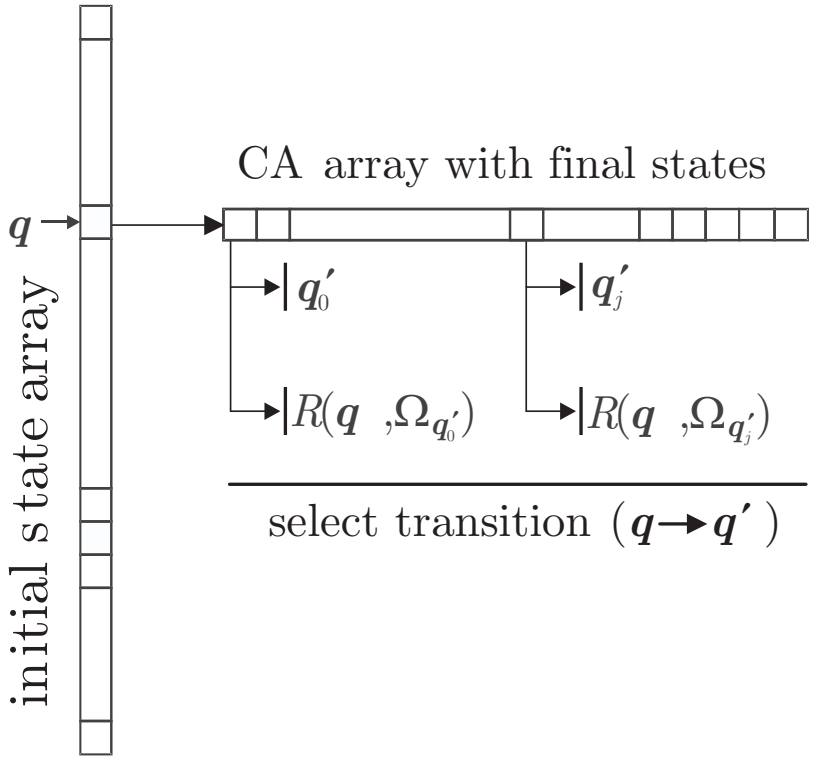

Figure 2.2: Scattering table for the particle-based Cellular Monte Carlo simulator. Each initial state state is connected to an ordered array with all of the possible final states and their scattering rates.

momentum space. Furthermore, the transition rate needs to be computed for each candidate in order to choose the final state. The Cellular Monte Carlo (CMC) algorithm [18] was developed to reduce the extremely high computational demand of the EMC scattering algorithm described above. To do so, the CMC algorithm employs a pre-computed transition look-up table for the probability of scattering, for each mechanism, from every initial state to every possible final state.

A schematic diagram of such a look-up table is shown in Fig. 2.2, where all of the possible initial states are represented in the vertical array. Each of those states is connected to an ordered list (horizontal array) of the possible final states and the corresponding scattering rate. When a scattering occurs, the carrier final momentum is chosen directly from candidates in the horizontal array simply by generating a

random number, which allows the efficient simulation of a large number of scattering events. Therefore, the CMC approach [18] trades the EMC intense computation burden with a large, but realistic, need of fast storage (RAM). However, the basic CMC approach achieves an efficient use of the memory by storing only the total rate; information is lost about the exact type of scattering process involved in the transition (e.g., the mode of the involved phonon, the temperature). This work expanded the basic approach by employing a rejection technique based on a new look-up table designed to retain all these information on the scatter events.

Finally, we note that this work employs the full-band representation of both the electronic band structure and the phonon spectra. In this approach, the energy dispersion for electrons and phonons are not analytically approximated but computed and tabulated across the whole first Brillouin zone (BZ1), which represents the Wigner–Seitz cell of the reciprocal lattice of the semiconductor crystal.

In particular, the full-band electronic structure is obtained with the non-local empirical pseudo-potential method [39, 40], while the phonon dispersion is computed by employing the 14 parameters valence shell model [41].

2.4 Format of the Transition Table

The CMC scattering algorithm employs a rate, R, representing the frequency of a specific particle transitioning between eigenstates. However, the result of the Fermi Golden Rule [20, 42] represents the total transition rate Γ between the initial and final eigenstates. In other words, while Γ describes the cumulative rate for all the carriers occupying the initial eigenstate, each particle transitions to the final eigenstate at a different rate, R. In the case of particles obeying the Pauli Exclusion Principle, such as electrons, Γ is equivalent to R. However, more than one phonon may, and in general do, occupy an initial eigenstate, and R must be obtained from Γ. This

important aspect will be further addressed in Chapter 6.

Moreover, in order to store all possible scattering rate in a look-up table of reasonable size, the entire (BZ1) of the crystal is discretized into an inhomogeneous tensor-product grid.

Formally, the scattering rate are expressed as the probability for a particles to transition from the initial state $\boldsymbol{q}$ to a small region (cell) $\Omega_{\boldsymbol{q}'}$ in momentum space centered around the point $\boldsymbol{q}'$:

$$R(\boldsymbol{q}, \Omega_{\boldsymbol{q}'}) = \int_{\Omega_{\boldsymbol{q}'}} d\boldsymbol{q}' R(\boldsymbol{q}, \boldsymbol{q}') \tag{2.4}$$

The cell $\Omega_{\boldsymbol{q}'}$ is assumed to be small enough for the transition rate to have small variations within $\Omega_{\boldsymbol{q}'}$, and the scattering rate can be approximated as:

$$R(\boldsymbol{q}, \Omega_{\boldsymbol{q}'}) \simeq R(\boldsymbol{q}, \boldsymbol{q}') D(\mathcal{E}(\boldsymbol{q}'), \Omega_{\boldsymbol{q}'}) \tag{2.5}$$

where $D\big(\mathcal{E}(\boldsymbol{q}'), \Omega_{\boldsymbol{q}'}\big)$ is the density of state in $\Omega_{\boldsymbol{q}'}$ at the expected final carrier energy $\mathcal{E}(\boldsymbol{q}')$.

Chapter 3

THE REJECTION TECHNIQUE

This chapter discusses an improved scattering algorithm used in the CMC method. The first section offers the motivations of the new approach, while sections 3.2 and 3.3 offer its theoretical framework. In particular, section 3.2 discusses the rejection technique in the context of a generic Monte Carlo method, while section 3.3 presents the rejection technique within the specific framework of the CMC method used in this work. Finally, the last section of this chapter discusses how to compute a crucial parameter for the implementation of the rejection technique, the occupation number.

3.1 Improving the CMC Scattering Algorithm

The algorithm that models scattering mechanisms within the CMC framework [18] is extremely efficient. Nevertheless, as it has been described in section 2.3, it cannot efficiently account for variations of local parameters, such as temperature and doping concentration, which affect the scattering probability. Indeed, different static look-up tables could be used to account for the different values of such parameters. However, such an approach would require an unrealistic number of rather large look-up tables, and corresponding fast memory devices (RAM), in order to treat all the possible combinations of these quantities. This work improves the modeling capability of the basic particle-based CMC scattering algorithm by employing a stochastic rejection technique to accept or reject scattering events based on parameters that are local in time and/or space. Such technique is based on an individual look-up table, and therefore comes with no overhead in terms of the required storage. The computational overhead results to be minimal.

3.2 Rejection Algorithm

The Monte Carlo rejection sampling is typically used to generate observations for a specific probability density function (pdf) $\mathrm{f}(x)$, from a different pdf $\mathrm{g}(x)$ by using a stochastic selection involving the ratio of the two pdf to accept or reject samples. The desired pdf $\mathrm{f}(x)$ is usually referred to as the target distribution, while the second pdf $\mathrm{g}(x)$ is usually referred to as the proposal distribution. In order to simplify the notation, the uppercase letters X and Y will be employed to represent random variables, while lowercase letters x, x', and ϱ_u will be employed to represent real numbers. Under this notation, the value $\mathrm{g}(x)dx$ represents the probability of the random variable X to fall within the interval $[x, x + dx]$.

The technique, also known as the acceptance-rejection method [43], is usually employed when the target pdf $\mathrm{f}(x)$ is difficult to sample but easy to evaluate, or when it is not known *a priori*. The proposal pdf $\mathrm{g}(x)$ is chosen to be easy to sample with the condition "$c\,\mathrm{g}(x) \geq \mathrm{f}(x)$", where c is a constant positive real number.

The flow-chart of the rejection algorithm is shown in Fig 3.1: first, a sample number x' is generated from the proposal pdf $\mathrm{g}(x)$, and a random number ϱ_u is independently extracted from the uniform distribution $U(0, 1)$. Then, the rejection probability is computed as $P_{rej} = \mathrm{f}(x')/\big(c\,\mathrm{g}(x')\big)$. Finally, if $\varrho_u < P_{rej}$ the sample is accepted, otherwise it is rejected.

Figure 3.2 provides a geometric interpretation of the rejection sampling algorithm for a simple one-dimensional case. The horizontal axis represents the possible values x for the random variable X. Each of the possible outcomes of the rejection algorithm can be described by the two values x and ϱ_u, these two values are then associate to a point of the plane of coordinate $(x, \varrho_u\, c\,\mathrm{g}(x))$. The dashed line identifies the points corresponding to the case of $\varrho_u = 1$, which correspond to the distribution $c\mathrm{g}(x)$. Since

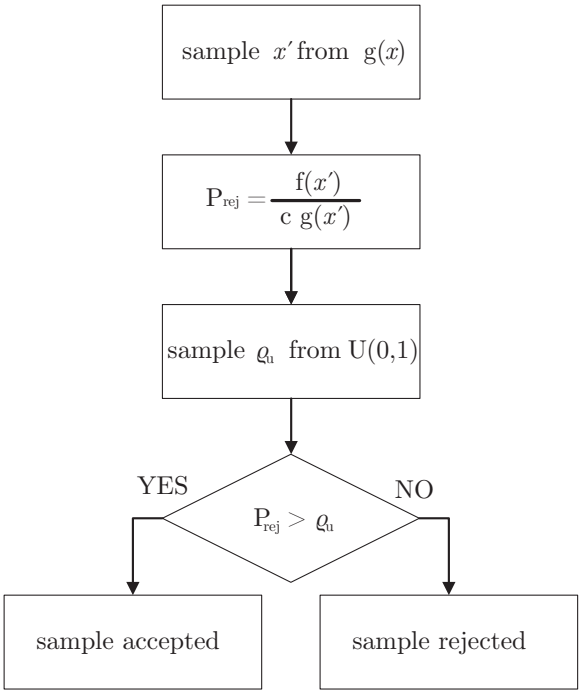

Figure 3.1: Flowchart of a generic acceptance/rejection algorithm.

the value of ϱ_u is in the interval $[0, 1]$, any point $(x, \varrho_u\, c\, g(x))$, corresponding to an iteration of the algorithm, must lay in the region between the horizontal axis and the dashed line.

Samples are rejected by the algorithm if $\varrho_u > P_{rej}$; a simple algebraic substitution shows that these occurrences correspond to points with ordinate larger than $P_{rej} c g(x) = f(x)$. In other words, all instances of the algorithm rejecting samples are represented by points which laying above the target distribution f(x) (solid line), while the accepted points lay under it. Intuitively, the algorithm works because the points corresponding to the accepted sample reproduce the area defined by the target distribution f(x).

More formally, the distribution resulting from the algorithm is the conditional probability density function of x given the occurrence of $\varrho_u < P_{rej}$. We prove now that $p(X = x|\varrho_u < P_{rej}) = f(x)$, where p denotes a pdf. The value of the pdf can be

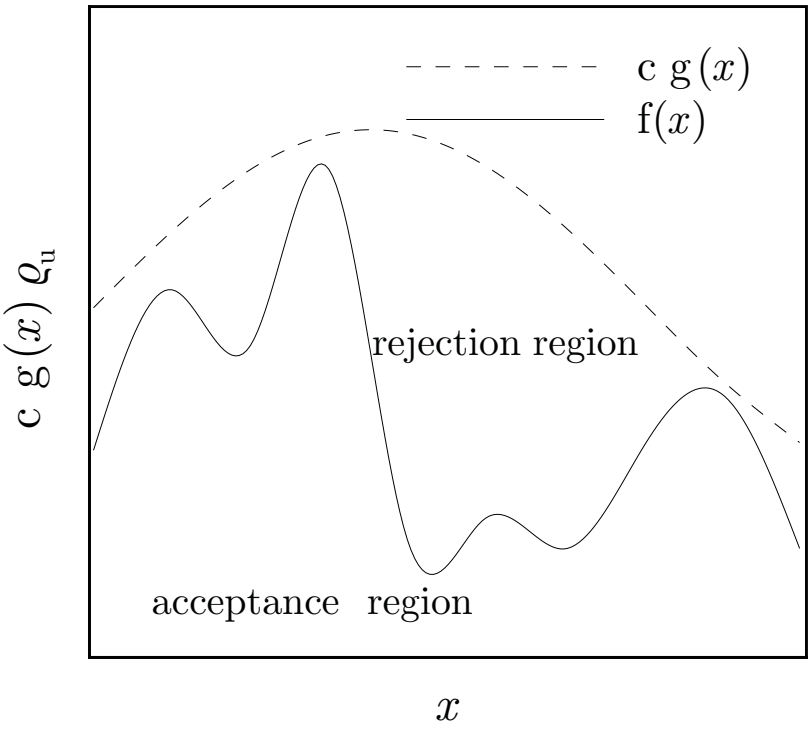

Figure 3.2: Graphical representation of the acceptance-rejection sampling in one dimension. The possible values of x and ϱ_u are associated to a point of coordinate $\big(x, \varrho_u\, c\, \mathrm{g}(x)\big)$. The curves for the proposal (dashed line) and target (solid line) distribution define the region for the accepted and rejected points. The algorithm samples uniformly under $c\,\mathrm{g}(x)$, keeping only the points that fall under $\mathrm{f}(x)$.

obtained from the definition of conditional probability [43]:

$$p(X = x|\varrho_u < P_{rej}) = \frac{p(\varrho_u < P_{rej}|X = x) * \mathrm{g}(x)}{p(\varrho_u < P_{rej})}. \tag{3.1}$$

The distribution function for an accepted sample after a specific x has been selected is P_{rej}:

$$p(\varrho_u < P_{rej}|X = x) = P_{rej} = \frac{\mathrm{f}(x)}{c\,\mathrm{g}(x)}. \tag{3.2}$$

The pdf of $p(\varrho_u < P_{rej})$ can therefore be computed as [1] :

$$p(R < P_{rej}) = \int dx p(\varrho_u < P_{rej}|X = x)\,\mathrm{g}(x) = \frac{1}{c}\int dx \mathrm{f}(x) = \frac{1}{c}. \tag{3.3}$$

[1]Given two random variable X, Y, and their joint distribution function $p_{X,Y}(x, y)$, the following equation are valid [43]: $p_X(y) = \int dx\, p_{X,Y}(x, y)$ and $p_{X,Y}(x, y) = p_X(x|y)p_Y(y)$.

Finally, the substitution of Eq. 3.2 and Eq. 3.3 in Eq. 3.1 supplies our proof:

$$p(X = x|\varrho_u < P_{rej}) = \frac{\mathrm{f}(x)}{c\mathrm{g}(x)} * \mathrm{g}(x)\frac{1}{1/c} = \mathrm{f}(x). \tag{3.4}$$

3.3 Rejection Sampling in the CMC Scattering Algorithm

This work models scattering by following an approach similar to the one described in [44]: the tabulated transition rate between two states is replaced by its maximum expected value for some parameter(s); when a transition occurs, the rejection technique is used to compare the maximum rate to its actual value as simply determined by the local value of the parameter(s). If the transition is rejected, the event is treated as a 'self-scattering' [38], in the sense that the final state is taken to be equal to the initial one, *i.e.* the particle state is unchanged.

More formally, employing the rejection sampling technique within the CMC frameworks requires defining in the space of possible final state $\boldsymbol{q}'$ the proposal $\mathrm{g}(\boldsymbol{q}')$ and the target distribution $\mathrm{f}(\boldsymbol{q}')$, alongside the constant c for an initial state $\boldsymbol{q}^{ini}$. The proposal pdf is obtained by normalizing values of the scattering rates $R_{max}(\boldsymbol{q}^{ini}, \boldsymbol{q}')$ stored in the look-up table and corresponding to the maximum value of some parameter(s):

$$\mathrm{g}(\boldsymbol{q}') = \frac{R_{max}(\boldsymbol{q}^{ini}, \boldsymbol{q}')}{\sum_{\boldsymbol{q}} R_{max}(\boldsymbol{q}^{ini}, \boldsymbol{q})}. \tag{3.5}$$

A similar approach is used to define the target distribution:

$$\mathrm{f}(\boldsymbol{q}') = \frac{R_{loc}(\boldsymbol{q}^{ini}, \boldsymbol{q}')}{\sum_{\boldsymbol{q}} R_{loc}(\boldsymbol{q}^{ini}, \boldsymbol{q})}, \tag{3.6}$$

where $R_{loc}(\boldsymbol{q}^{ini}, \boldsymbol{q}')$ is the transition rate from $\boldsymbol{q}^{ini}$ to $\boldsymbol{q}'$ given by the local values of the parameter(s).

In order to ensure a positive self-scattering rate, the value stored in the look-up table must by higher than the locally computed rate for each $\boldsymbol{q}^{ini}$ and $\boldsymbol{q}'$. Since

$R_{max}(\boldsymbol{q}^{ini}, \boldsymbol{q}') > R_{loc}(\boldsymbol{q}^{ini}, \boldsymbol{q}')$ for each $\boldsymbol{q}'$, the constant c can be simply set to:

$$c = \frac{\sum_{\boldsymbol{q}} R_{max}(\boldsymbol{q}^{ini}, \boldsymbol{q})}{\sum_{\boldsymbol{q}} R_{loc}(\boldsymbol{q}^{ini}, \boldsymbol{q})}. \tag{3.7}$$

This choice of c not only ensures that $c\,\mathrm{g}(\boldsymbol{q}') > \mathrm{f}(\boldsymbol{q}')$ for each $\boldsymbol{q}'$, but it also allows the computation of the rejection probability simply as the ratio of the target and the sampling probability:

$$P_{rej}(\boldsymbol{q}') = \frac{\mathrm{f}(\boldsymbol{q}')}{c\,\mathrm{g}(\boldsymbol{q}')} = \frac{R_{loc}(\boldsymbol{q}')}{R_{max}(\boldsymbol{q}')}. \tag{3.8}$$

A crucial simplification of the rejection probability due to the common terms between $R_{loc}(\boldsymbol{q}')$ and $R_{max}(\boldsymbol{q}')$ will be discussed in Chapter 5 and 6, for electron and phonon scattering, respectively.

As stated before, this technique requires that each transition rate $R_{max}(\boldsymbol{q}')$ stored in the look-up table is larger or equal to any rate $R_{loc}(\boldsymbol{q}')$ expected during the simulation. For example, in the case of phonon simulation, the scattering rate is maximized by assuming the values of parameters such as a temperature and doping concentration to be much larger than the realistically expected simulation values. Compared to the traditional CMC scattering, this technique has a computational overhead due to oversampling and to the computation of $P_{rej}(\boldsymbol{q}')$.

We note that the traditional CMC scattering algorithm [18] employs a look-up table optimized for memory requirement. This is achieved by employing a lossy compression method which merge together transition with the same $\boldsymbol{q}^{ini}$ and $\boldsymbol{q}'$ and discards information on the nature of the scattering mechanism causing each transition. However, the rejection technique requires all these information in order to compute a new transition rate based on the local values of the parameter(s). Therefore, a new modular look-up table able to retain information about the exact type of scattering process involved in the transition has been designed and implemented. Details on the structure and the compression algorithm employed by the new look-up table are

given in Chapter 8.

3.4 Occupation Number

In the case of scattering involving phonons, the rejection technique requires the value of the local and instantaneous occupation number $n(\boldsymbol{r}, \boldsymbol{q}, t)$ for the phase-space state involved in the scattering. A possible approach is to assume local thermal equilibrium and a consequent 'local' temperature T. The expected value of n can then be predicted by the Bose-Einstein distribution [20]:

$$n_0(\omega, T) = \left(\exp\left(\frac{\hbar\omega}{k_B T} \right) - 1 \right)^{-1}, \tag{3.9}$$

where $\hbar$ is the reduced Plank constant, ω is the frequency of the phonon, and k_B is the Boltzmann constant, and the phonon mode index has been omitted for the sake of clarity.

In a more general approach, n must be evaluated by sampling the population of simulated particles. This is achieved by approximating n with its average in the phase space volume identified by the volume $V_{\boldsymbol{r}}$ in position space and the volume $\Omega_{\boldsymbol{q}}$ in momentum space:

$$n(\boldsymbol{r}, \boldsymbol{q}, t) \approx \frac{\eta(\Omega_{\boldsymbol{q}}, V_{\boldsymbol{r}}, t)(8\pi^3)}{\Omega_{\boldsymbol{q}} V_{\boldsymbol{r}}}, \tag{3.10}$$

where $V_{\boldsymbol{r}}$ is centered on the position vector $\boldsymbol{r}$, $\Omega_{\boldsymbol{q}}$ is centered around the momentum $\boldsymbol{q}$, and $\eta(\Omega_{\boldsymbol{q}}, V_{\boldsymbol{r}}, t)$ is the number of phonons with momentum within $\Omega_{\boldsymbol{q}}$ located in the volume $V_{\boldsymbol{r}}$ at time t.

While $V_{\boldsymbol{r}}$ is obtained by simply defining a volume in the system geometry around the position $\boldsymbol{r}$, $\Omega_{\boldsymbol{q}}$ requires further consideration. First, $\Omega_{\boldsymbol{q}}$ must be defined accounting for the crystal periodicity. For example, a spherical $\Omega_{\boldsymbol{q}}$ of radius Δq can be defined as the collection of states $\boldsymbol{q}'$ in the first Brillouin zone BZ1, whose Euclidean distance

from $\boldsymbol{q}$ is smaller than Δq:

$$\Omega_{\boldsymbol{q}} = \left\{ \boldsymbol{q}' \subset BZ1 \colon |\boldsymbol{q}' - \boldsymbol{q} + \boldsymbol{G}| < \Delta q \right\} \tag{3.11}$$

where $\boldsymbol{G}$ is a reciprocal lattice vector and accounts for the periodicity of BZ1.

Moreover, $n(\boldsymbol{q}, \boldsymbol{r}, t)$ has an exponential dependence with the energy, and using a purely geometrical description of $\Omega_{\boldsymbol{q}}$ it may not be enough to ensure $n(\boldsymbol{r}, \boldsymbol{q}, t) \simeq \langle n \rangle$ unless $\Omega_{\boldsymbol{q}}$ is "small enough". This problem is approached by employing an additional constraint on the maximum energy span allowed in $\Omega_{\boldsymbol{q}}$.

Finally, an ellipsoid is used in place of a sphere to define $\Omega_{\boldsymbol{q}}$. This approach generalizes the spherical case and provides more flexibility when dealing with asymmetric Brillouin zones. In this case, $\Omega_{\boldsymbol{q}}$ is defined as:

$$\begin{aligned} \Omega_{\boldsymbol{q}} = \Big\{ & \boldsymbol{q}' \subset BZ1 \colon \ |\omega' - \omega| < \Delta\omega_{MAX} \\ & \wedge \frac{|q'_x - q_x + G_x|^2}{\Delta q_x^2} + \frac{\left|q'_y - q_y + G_y\right|^2}{\Delta q_y^2} + \frac{|q'_z - q_z + G_z|^2}{\Delta q_z^2} = 1 \Big\} \end{aligned}, \tag{3.12}$$

where ω' represents the energy of the state with momentum $\boldsymbol{q}'$, $\Delta\omega_{MAX}$ is the maximum allowed energy difference between two states, the subscript x, y, and z represent the vector projections on the main axes, and finally, Δq_x, Δq_y, and Δq_z are the length of the semi-principal axes that identify the ellipsoid in momentum space.

The bottleneck of this technique is the calculation of $\eta(\Omega_{\boldsymbol{q}}, V_{\boldsymbol{r}}, t)$. In particular, the algorithm must evaluate all the simulated particles and decide which one must be considered and which one can be discarded. In order to reduce the number of particles to be evaluated, each cell in position space has a linked list of all the carriers present in the cell itself. Moreover, each cell in position space has another list containing all the other cells within a distance d in position space. Regarding the reciprocal space, the data required to define $\Omega_{\boldsymbol{q}}$ is precomputed and saved in the band structure data. In particular, a hash function based on the memory location is used to associate each

cell to an integer number, this approach allows the use of an efficient integer-based search algorithm. For each $\boldsymbol{q}$ at the center of a reciprocal space cell an ordered list $S(\Omega_{\boldsymbol{q}})$ containing the hash number of all the cells intersecting $\Omega_{\boldsymbol{q}}$ is computed and stored alongside the value of $\Omega_{\boldsymbol{q}}$. In order reduce the computing time, the required controls are performed in sequence from the fastest to the slowest. The first control verifies if the particle is in the same phonon mode of $\boldsymbol{q}$, the second control ensures that the energy restriction on $\Omega_{\boldsymbol{q}}$ is respected. In the last step, a dichotomic search algorithm is employed to verify if the momentum of the simulated particle is included in the volume $\Omega_{\boldsymbol{q}}$. When the simulated phonon passes all these tests, the value of $\eta(\Omega_{\boldsymbol{q}}, V_{\boldsymbol{r}}, t)$ is incremented.

Chapter 4

ELECTRON-PHONON INTERACTION

This chapter discusses the modeling of the electron-phonon interactions when the two populations of particles are synchronous; both the theoretical and the algorithmic aspects of such simulations are described here. The first section of this chapter presents the challenges of simulating these interactions within a Monte Carlo framework, while in the following sections, two different stochastic techniques to address such challenges are offered. In particular, section 4.2 presents a technique of modeling the process of emission and absorption of phonons in terms of the Poisson probability distribution. While this technique is highly effective for systems in steady-state, it presents difficulties in simulating the transient regimes. Section 4.6 offers an alternate technique that is more suitable for simulation of transient, which is based on the more complex mathematical framework discussed in section 4.4 and 4.5.

4.1 Challenges

The interaction of electrons with the crystal lattice can be expressed in terms of absorption and emission of phonons. In the context of a thermo-electric simulation with heat modeled as particles, these processes correspond to the addition and removal of simulated phonons from the simulation domain. However, the process of emission or absorption cannot be simply modeled by creating or removing a simulated phonon as the result of the scattering with a simulated charged particle. This is due to the fact that there is not a one-to-one correspondence between the real particles and the simulated ones; in other words, each simulated particle represents an arbitrary number of real particles. In particular, let's say that a simulated charge carrier represents W^c

physical electrons, and a simulated phonon represents W^p physical phonons. In a simulated electron-phonon scattering event, the scattering represents an interaction of W^c electrons with W^p phonons. It appears obvious that, for this model to be statistically accurate, W^p should equal W^c. However, in a typical semiconductor, the density of phonons is orders of magnitude higher than the density of free charges. This means that the $W^p = W^c$ condition would require a number of simulated phonons orders of magnitude larger than the number of simulated charged particles, which is not realistic from a computational view point. A more sensible choice employs a similar number of simulated particles for both phonons and charged particles with the weight factor W^p orders of magnitude larger than W^c. Because of this difference, the processes of emission and absorption cannot be correctly simulated as one-to-one interactions between simulated phonons and simulated electrons. In order to address this problem, we developed two stochastic techniques that model the interactions between simulated particles with different weights while allowing for a statistically accurate representation of the carriers dynamics in the thermo-electric system.

4.2 Phonon Absorption and Emission in Steady-State

In a steady-state system, the number of phonons emitted in a given time interval can be assumed constant, therefore, the emission process can be described by the Poisson probability distribution [45, 46]:

$$\mathcal{V}\big(k, \lambda^e(\boldsymbol{q})\big) = \frac{\big(\lambda^e(\boldsymbol{q})\big)^k e^{-\lambda^e(\boldsymbol{q})}}{k!}, \tag{4.1}$$

where the parameter $\lambda^e(\boldsymbol{q})$, referred to as rate or intensity, is the expected (or average) number of phonons emitted during a time interval dt into the state $\boldsymbol{q}$, and the superscript e referring to the emission process. In particular, during dt, the intensity

is equal to the average number of emission scattering events per unit of time:

$$\lambda^e(\boldsymbol{q}) = \frac{\mathbb{E}\left[\eta^{ce}(\boldsymbol{q}, t+dt) - \eta^{ce}(\boldsymbol{q}, t)\right]}{dt} = \eta^c \Gamma^{ce}(\boldsymbol{q}), \tag{4.2}$$

where $\eta^{ce}(\boldsymbol{q}, t)$ is the total number of (real) charges emitting a phonons into the state $\boldsymbol{q}$ up to the time t, η^c represents the number of charged particle (either holes or electrons), and Γ^{ce} is the charge-phonon emission rate. Finally, the symbol $\mathbb{E}$ represents the expectation operator [45], such that the expected value of the random variable X, $\mathbb{E}[X]$, is the mean value of X. In the context of CMC simulations, the number of phonon emissions caused by simulated charged particles has a similar expression:

$$\lambda^{ce}(\boldsymbol{q}) = \frac{\mathbb{E}\left[N^{ce}(\boldsymbol{q}, t+dt) - N^{ce}(\boldsymbol{q}, t)\right]}{dt} = \frac{\eta_c}{W^c}\Gamma^{ce}(\boldsymbol{q}) = \frac{\lambda^e(\boldsymbol{q})}{W^c}, \tag{4.3}$$

where the capital letter N denotes the number of simulated particles, each one representing W^c charged particles. Dividing the expected number of emitted phonons, $\lambda^e(\boldsymbol{q})$, by the weight of the simulated phonon, W^p, gives the expected number of simulated phonons to be added to the simulation:

$$\lambda^{pe}(\boldsymbol{q}) = \frac{\lambda^e(\boldsymbol{q})}{W^p}. \tag{4.4}$$

The simple approach of generating a new simulated phonon with each scattering of a simulated charge would create $\lambda^{ce}(\boldsymbol{q})$ simulated phonons in average. However, it is clear from Eq. 4.4 and Eq. 4.3 that $\lambda^{ce}(\boldsymbol{q})$ differs from $\lambda^{pe}(\boldsymbol{q})$; in particular, $\lambda^{ce}(\boldsymbol{q}) > \lambda^{pe}(\boldsymbol{q})$ since $W^c < W^p$. Consequently, this simple approach cannot be used.

The intuitive solution is to create a new simulated phonon as result of only a fraction of the scattering events while preserving the Poisson nature of the phonon generation process. Indeed, a widely used variation on the rejection algorithm presented in Chapter 3 section 3.2, called thinning [47], allows to achieve our goal. The

algorithm generates events obeying a Poisson distribution, which will be referred to as the target distribution, by down-sampling events obeying a different Poisson distribution, which will be referred to as the sampling distribution. In other words, the algorithm reproduces the target distribution by randomly selecting a subset of events from a suitable sampling distribution. In particular, the events of the sampling distribution must be more frequent than the events of the target distribution; so that the sampling Poisson distribution is said to be stochastically dominating [48] the target distribution. The requirement of stochastic domination results in a lower bound for the Poisson intensity of the sampling distribution $\lambda^*(t)$:

$$\lambda^*(t) > \lambda(t) \ \ \forall t, \tag{4.5}$$

where $\lambda(t)$ is the intensity of the target distribution. The events are then retained with probability $\lambda(t)/\lambda^*(t)$ or rejected with probability $1-\lambda(t)/\lambda^*(t)$. Finally, $\lambda^*(t)$ is typically chosen to be a constant in order to simplify the algorithm which generates the sampling distribution. In the specific case of phonon emission, $\lambda^{ce}(\boldsymbol{q})$ is the intensity of the dominating distribution while $\lambda^{pe}(\boldsymbol{q})$ is the intensity of the target distribution.

In practice, after a simulated charged particle scatters, the thinning procedure employs the same rejection technique discussed in section 3.2 where the rejection probability is computed as:

$$P_{rej} = \frac{\lambda^{pe}(\boldsymbol{q})}{\lambda^{ce}(\boldsymbol{q})} = \frac{W^c}{W^p}. \tag{4.6}$$

When the process is accepted, a new simulated phonon is created in the same position of the simulated charged particle. The average number of simulated phonons created by the rejection algorithm is therefore given by:

$$\mathbb{E}\left[N^{pe}(\boldsymbol{q}, t+dt) - N^{pe}(\boldsymbol{q}, t)\middle| \varrho_u < P_{rej}\right] = \mathbb{E}\left[dN^{ce}(\boldsymbol{q})\right] P_{rej} = \frac{dt\, \lambda^e(\boldsymbol{q})}{W^p}, \tag{4.7}$$

Where $N^{pe}(\boldsymbol{q}, t)$ is the total number of simulated phonons in the state $\boldsymbol{q}$ created by an emission process up to the time t, ϱ_u is a number randomly extracted from the uniform distribution $U(0,1)$, $\big|$ is the conditional expectation operator[1][45], and $dN^{ce}(\boldsymbol{q}) = N^{ce}(t + dt, \boldsymbol{q}) - N^{ce}(t, \boldsymbol{q})$. The previous equation shows that the average number of simulated phonons created by the rejection algorithm corresponds to the expected value $\lambda^{pe}(\boldsymbol{q})$ calculated in Eq. 4.4. A similar treatment in the case of phonon absorption leads to a similar algorithm employing the same rejection probability in 4.6. In this case, a simulated phonon to be removed is chosen while calculating the local and instantaneous occupation number as described in section 3.4. In particular, the algorithm selects the simulated particle with the smallest deviation from the target value of $\hbar\omega_{\boldsymbol{q}}$. The rather limited capability of this approach to model a transient regime will be discussed in the following section.

4.3 The Limits of the Poisson Model

The limits of the technique presented in the previous section are shown when describing the expected rate at which energy is lost and gained by the two simulated populations.

In particular, the expected energy loss by simulated charged carriers in an infinitesimal time interval dt, $d\mathcal{E}^{ce}$, can be computed by adding together the contribution of each scattering:

$$\mathbb{E}\left[d\mathcal{E}^{ce}\right] = \mathbb{E}\left[\mathcal{E}^{ce}(t + dt) - \mathcal{E}^{ce}(t)\right] = \mathbb{E}\left[\sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^c dN^{ce}(t, \boldsymbol{q})\right], \tag{4.8}$$

where the symbol $\mathbb{E}$ represents the expectation operator [45], $\omega_{\boldsymbol{q}}$ is the frequency of the state $\boldsymbol{q}$, $dN^{ce}(t, \boldsymbol{q})$ has the same expression as in Eq. 4.7, and represents the

[1] $A\big|B$ is the expected value of A computed with respect to the conditional probability distribution $p(\mathrm{A}|\mathrm{B})$.

number of scattering between phonons in the state $\boldsymbol{q}$ and simulated charged particles in the interval $[t+dt, t]$, each one representing W^c charged particles. The linearity of the expectation operator [45] allows us to write the previous equation as:

$$\mathbb{E}\left[d\mathcal{E}^{ce}\right] = \sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^c \mathbb{E}\left[dN^{ce}(t,\boldsymbol{q})\right] = \sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^c \, dt \, \lambda^{ce}(\boldsymbol{q},t), \tag{4.9}$$

where the product $dt\,\lambda^{ce}(t,\boldsymbol{q})$ is the expectation, or the average value, of $dN^{ce}(t,\boldsymbol{q})$.

Similarly, the expected energy gained by the population of simulated phonons during dt, $\mathcal{E}^{pe}(t)$, is calculated by adding the contribution of each newly created particle:

$$\mathbb{E}\left[d\mathcal{E}^{pe}(t)\right] = \sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^p \mathbb{E}\left[dN^{pe}(t)\right] = \sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^p \, dt \, \lambda^{pe}(t,\boldsymbol{q}), \tag{4.10}$$

where $dN^{pe}(t,\boldsymbol{q}) = N^{pe}(t+dt,\boldsymbol{q}) - N^{pe}(t,\boldsymbol{q})$, $N^{pe}(t,\boldsymbol{q})$ is defined in Eq. 4.7, and W^p is the number of phonons represented by each simulated phonon. Finally, the product $dt\,\lambda^{pe}(t,\boldsymbol{q})$ is the expectation of $dN^{pe}(t,\boldsymbol{q})$.

Since the thinning algorithm imposes $\lambda^{pe}(t,\boldsymbol{q}) = \lambda^{ce}(t,\boldsymbol{q}) * W^c/W^p$, the average energy lost by the population of simulated charged carriers equals the energy gained by the population of simulated phonons. A similar approach can be applied to analyze the thinning algorithm applied to the emission process. In this case, the equations show that the energy gained by the population of simulated charged carriers via absorption, $d\mathcal{E}^{ca}$, equals the energy lost via removal of simulated phonons, $d\mathcal{E}^{pa}$.

The corresponding central variances for 4.8 and 4.10 are[2,3], respectively:

$$\begin{aligned}
\mathrm{Var}\left(d\mathcal{E}^{ce}(t)\right) &= \textstyle\sum_{\boldsymbol{q}\subset BZ1} \left(dt\,\hbar\omega_{\boldsymbol{q}} W^c\right)^2 \lambda^{ce}(\boldsymbol{q}) \\
\mathrm{Var}\left(d\mathcal{E}^{pe}(t)\right) &= \textstyle\sum_{\boldsymbol{q}\subset BZ1} \left(dt\,\hbar\omega_{\boldsymbol{q}} W^p\right)^2 \lambda^{pe}(\boldsymbol{q}) = \sum_{\boldsymbol{q}\subset BZ1} \left(dt\,\hbar\omega_{\boldsymbol{q}}\right)^2 W^p W^c \lambda^{ce}(\boldsymbol{q})
\end{aligned} \tag{4.11}$$

[2]If x is a random variable with a Poisson distribution of intensity λ and a is a real number, the random variable $y = ax$ has average $a\lambda$ and variance $a^2\lambda$.

[3]If x and y are independent random variables $\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y)$.

Since $W^p > W^c$ (typically, it is $W^p >> W^c$), the variance of the energy gained by the simulated phonon population is larger than the variance of the energy lost by the simulated charge population: $\mathrm{Var}(d\mathcal{E}^{pe}) > \mathrm{Var}(d\mathcal{E}^{ce})$. In the case of phonon absorption, the same argument gives in $\mathrm{Var}(d\mathcal{E}^{pa}) > \mathrm{Var}(d\mathcal{E}^{ca})$.

In the steady state regime, instantaneous differences between the energy gained by the simulated phonon population, $d\mathcal{E}^{pe}(t)$, and the energy lost by the simulated charged carriers, $d\mathcal{E}^{ce}(t)$, have no lasting effect since, from 4.7, both quantities have the same average value. The total energy of the charge-phonon system is therefore conserved on average, and the effect of the difference in variance is an increase in noise in the energy of the simulated phonon population as compared to the simulated charge population. Therefore, the relative difference between the energy of the two populations existing before the steady-state regime will remain unaltered. More explicitly, the conditions present at the end of a transient are preserved in the steady-state regime, as they should be. Following the same argument, any random fluctuation of the difference in energy $d\mathcal{E}^{pe}(t) - d\mathcal{E}^{ce}(t)$ present at the end of a transient will be preserved in the steady state regime. In other words, the statistical fluctuation of the transient regime carries over in the steady state regime. This creates an irremovable discrepancy between the energy gained by the simulated crystal and the energy lost by the simulated charge particles. The larger the fluctuation, the larger the error in the conservation of energy. This problem needs to be addressed by the algorithm.

It is possible to employ a book-keeping scheme to account for the error in the energy exchanged between the two populations. This approach allows compensating for any error by adding or removing the appropriate number of simulated phonons, when the energy discrepancy justifies it. However, this method requires somewhat arbitrary decisions on the state of the simulated phonon to be removed and on which state a new simulated phonon should be created. In order to avoid such arbitrary

decisions, a method based on the mathematical model called point process [49] is used in place of the model presented in section 4.2. This framework allows modeling the creation and removal of simulated phonons in a way that statistically enforces the conservation of the system energy. The mathematical basis for the modeling of absorption and emission point process will be offered in the next sections 4.4 and 4.5. The model employed in this work will be finally discussed in section 4.6.

4.4 Point Process

A point process [50], can be described as a sequence of points randomly located in some space. However, the only type of point process (PP) considered in this work will be defined on the one-dimensional time space, represented by the letter t. In this context, a PP can be simply thought of as being a collection of points (events) randomly placed on the time line. The specific location patterns and these points constitute the process itself. In this framework, the process will therefore denote a collection of events occurring in time, and the term *stochastic process* will identify a process that develops in time according to some probabilistic rule. A point process is typically described in terms of its counting measure [49, 51, 52], $\mathcal{M}(t)$, which represents the cumulative count of events in a process. This means that for any two times t_1 and t_0, such that $t_1 > t_0$, the value $\mathcal{M}(t_1) - \mathcal{M}(t_0)$ is a positive integer representing the number of events occurring in the time interval $[t_1, t_0]$. Moreover, $\mathcal{M}(t)$ is called a random counting measure when the counted events are parts of a stochastic process.

More formally, $\mathcal{M}(t)$ defined for time $t > 0$ is said to be a counting measure if [49, 51]:

- $\mathcal{M}(0) = 0$

- $\mathcal{M}(t)$ is a positive integer $\forall t > 0$

- $\mathcal{M}(t)$ is non-decreasing, right-continuous step function with jumps of finite size.

The counting function is typically expressed as a summation of step functions [45]:

$$\mathcal{M}(t) = \sum_{i=1}^{\infty} \mathcal{J}_i(t), \tag{4.12}$$

where the step function $\mathcal{J}_i(t)$ is defined as:

$$\mathcal{J}_i(t) = \begin{cases} 1 & t \geq t_i \\ 0 & t < t_i \end{cases}, \tag{4.13}$$

where t_i is the time of the *i-th* event. Figure 4.1 shows an graphical representation

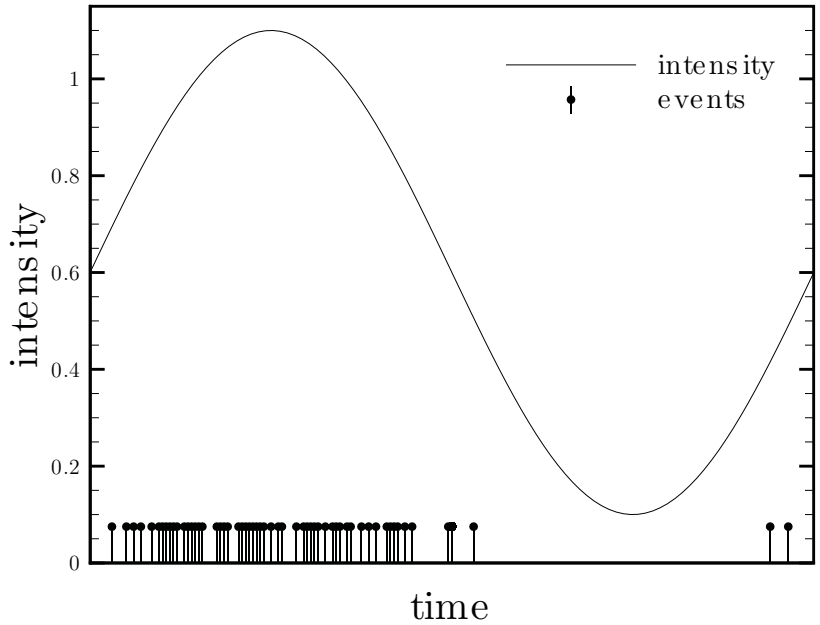

Figure 4.1: Example of a one dimensional PP. The solid line represent the corresponding intensity of the PP as a function of time, while each dot on the time line marks an event. A higher clustering of events corresponds to a higher intensity.

of a PP. The solid line represents the PP intensity as a function of time, while the dots on the time line mark the corresponding events. The average distance between events centered around a specific time t has an inverse relationship with the corresponding intensity, $\lambda(t)$. In other words, a higher intensity results in a higher density

of events while a lower intensity results in a more sparse distribution of events. In this particular case, the alternation between high and low intensity produces clustering effects centered around the peak of intensity.

A notable example of PP is obtained by defining the probability for new events to occur during dt as:

$$\lim_{dt\to 0+} P\left[\mathcal{M}(dt+t)-\mathcal{M}(t)=j\right]\begin{cases} o(dt) & j>1 \\ \lambda(t)\,dt+o(dt) & j=1 \\ 1-\lambda(t)\,dt+o(dt) & j=0 \end{cases}, \tag{4.14}$$

where the $P[A]$ is the probability of the event A, $\mathcal{M}$ is the random counting measure of the PP, $\lambda(t)$ is referred to as intensity, $o(dt)$ is the Bachmann-Landau (little o) notation [49], that explicitly indicates that the probability of two events happening at the same time tends to vanish. A point process with this property is said to be *simple* [49, 50]. Moreover, this property allows us to relate the probability, the intensity and the expected value [49, 52] of new events occurring in a time interval dt. In particular, for small values of dt, $o(dt)=0$ and the previous quantities are simply related by:

$$P\big[\mathcal{M}(t+dt)-\mathcal{M}(t)=1\big]=dt\,\lambda(t)=\mathbb{E}\left[\mathcal{M}(t+dt)-\mathcal{M}(t)\right]. \tag{4.15}$$

When $\lambda(t)$ is a deterministic function, the number of events in an interval dt obeys a Poisson distribution with intensity $\lambda(t)$. Therefore, the quantity $\mathcal{M}(t+dt)-\mathcal{M}(t)$ is a non-homogeneous Poisson random variable with intensity $\lambda(t)$ [45, 49]. In this general case, the PP is referred to as a non-homogeneous Poisson point process. When $\lambda(t)$ is simply a constant, the quantity $\mathcal{M}(t+dt)-\mathcal{M}(t)$ is a homogeneous Poisson random variable. In this case, the PP is referred to as a homogeneous Poisson point process. Moreover, this definition of Poisson PP in terms of counting measure allows the derivation of all the properties of the Poisson distribution [45, 46]. If $\lambda(t)$ is not

a deterministic function but another independent Poisson random variable, the PP is referred to as a Cox Doubly Stochastic Poisson Process [46, 53, 54], or simply Cox process. As such, the Cox process is usually considered a generalization of the Poisson process where the intensity itself is a random variable. For simplicity, this case can be thought of in terms of *local* behavior: in a small time interval, the Cox PP has a Poisson-like behavior with intensity $\lambda(t)$; however, at each time the value of the intensity is given by an independent Poisson random variable. The term Cox process is also commonly used in the literature to describe any PP with a Markovian [49] intensity $\lambda(t)$. In the case of a non-Markovian intensity, the PP is commonly referred to as a Hawkes process [46, 49, 53], named after the work of Alan G. Hawkes, who employed a PP with non-Markovian intensity to model clustering behavior [53]. In the same way, the Cox process can be considered a generalization of the Poisson process, the Hawkes process can be thought of as a non-Markovian extension of the Poisson PP. In a sense, the Hawkes PP is a Poisson PP which accounts for the effect of past events. For this reason, an approach based on the Hawkes PP seems a natural extension to the model based on the Poisson distribution presented in section 4.2. It may seems unusual to use a non-Markovian approach to model a Markovian systems. However, this method can be intuitively thought of as a sophisticated statistical form of book keeping.

4.5 Phonon Absorption and Emission as Hawkes Process

The point process (PP) used in this work to model the absorption and emission of simulated phonons is a Hawkes [53] point process. This type of PP models systems where each new event impacts the whole stochastic process. More explicitly, the chance of a new event depends on the sequence of events that preceded it. This type of process is typically employed to model systems where the observed events

tend to cluster in time. For example, seismic events present short-term clustering; an earthquake main-shock is both preceded (fore-shock) and followed (after-shock) by intermediate-size events [55]. Hawkes processes have been employed to model criminal activity such as property crime, gang violence, and terrorism. Each initial event increases the likelihood of more events, as a successful offender(s) tends to replicate the crime in nearby location in the following weeks [56]; or, in the case of gang violence, each violent event tends to ignite retaliation [57].

In the past years, this statistical model became popular in the fields of finance and economics [58] to model the dynamics of trades and prices, especially in order to predict price changes in the context of high frequency finance. This later application employed a multivariate form of the Hawks process. In this approach, several events interact with each other to determine the probability of an event occurring in the near future. For example, it is well known that in the financial market the trading activity tends to cluster in time. In particular, Hewlett [59] successfully fitted order arrivals data by employing a model where each new sell offer increases the expected number of both buy requests and sell offers for a short time, and vice versa. In all these examples, each event makes the occurrence of the following events more probable for some period of time, and this type of PP are referred to as Self-Exciting Point Processes. Formally, this behavior can be expressed in terms of counting measure as:

$$cov\left(\mathcal{M}(t_1) - \mathcal{M}(t_0), \mathcal{M}(t_2) - \mathcal{M}(t_1)\right) > 0, \tag{4.16}$$

for any $t_0 < t_1 < t_2$, where *cov* denotes the covariance of the two quantities. When the covariance is negative, the PP is called self-correcting or self-regulating; in that case each event makes the occurrence of the following events less probable for some period of time.

The Hawkes process is defined by relating the counting measure to the collection

of previously occurred events, which will be represented by the symbol $\mathcal{H}(t)$ and referred to as history[4]. In particular, the Hawkes process is defined by expressing the conditional probability of observing a particular counting function given the history:

$$\lim_{dt\to 0+} P\Big(\mathcal{M}(t+dt)-\mathcal{M}(t)=m\Big|\mathcal{H}(t)\Big)=\begin{cases} o(dt) & m>1 \\ \lambda\big(t,\mathcal{H}(t)\big)\,dt+o(dt) & m=1 \\ 1-\lambda\big(t,\mathcal{H}(t)\big)\,dt+o(dt) & m=0 \end{cases}, \tag{4.17}$$

where $\Big|$ represent the conditional probability operator[5] [45], and $\lambda\big(t,\mathcal{H}(t)\big)$ is referred to as the intensity of the process. For example, in the case studied by Hawkes [53] the intensity takes the form:

$$\lambda\big(t,\mathcal{H}(t)\big)=\lambda_0+\int_0^t ae^{b(t-s)}d\mathcal{M}(s), \tag{4.18}$$

where λ_0 is the background intensity of the process, a and b are positive real constants, and the term $d\mathcal{M}(s)$ is:

$$d\mathcal{M}(t)=\mathcal{M}'(t)\,dt=\sum_{i=1}^{\infty}\delta(t-t_i)\,dt, \tag{4.19}$$

where t_i is the time of the *i-th* event, $\mathcal{M}$ is replaced by the definition in in 4.12, and the delta function is the derivative of the step function. The previous equation allows us to rewrite 4.18 as:

$$\lambda(t,\mathcal{H}(t))=\lambda_0+\sum_{t_i<t} ae^{-b(t-t_i)} \tag{4.20}$$

It is now clear that each event instantaneously increases the arrival intensity by A, and this influence decays exponentially over time.

[4]In measure theory, $\mathcal{H}$ is called a filtration or history [49], that is an increasing sequence of σ-algebras on the measurable space.

[5]$P\Big(X<x_0\Big|Y<y_y\Big)$ is the probability of the event $X<x_0$ occurring given the event $Y<y_0$ has occurred.

In more general terms, the intensity of a non-linear Hawkes process can be written as [49, 60]:

$$\lambda\big(t, \mathcal{H}\big) = \vartheta\left(\zeta(t) + \int_{-\infty}^{t} \hbar\,(t, s)\, d\mathcal{M}(s)\right), \tag{4.21}$$

where the function ϑ is such that $\vartheta : \mathbb{R} \to \mathbb{R}^{+}$, and the function ζ accounts for the evolution of the stochastic process in absence of events, in other words, it provides a *background* intensity to the PP. The term $\hbar$ is referred to as the exciting function, and can be rewritten by using the definition of $d\mathcal{M}(s)$ in Eq 4.19 as:

$$\int_{0}^{t} \hbar\,(t, s)\, d\mathcal{M}(s) = \sum_{i=1}^{\infty} \int_{0}^{t} \hbar\,(t, s)\, \delta(s - t_i) ds. \tag{4.22}$$

where the summation and integration order has been reversed. The final integration can be split in two cases:

$$\begin{array}{ll} \int_{0}^{t} \hbar\,(t, s)\, \delta(s - t_i) ds = \hbar\,(t, t_i) & \forall t_i \leq t \\ \int_{0}^{t} \hbar\,(t, s)\, \delta(s - t_i) ds = 0 & \forall t_i \geq t \end{array}. \tag{4.23}$$

These last two equations allow us to rewrite Eq. 4.21 as:

$$\lambda\big(t, \mathcal{H}\big) = \vartheta\left(\zeta(t) + \sum_{t_i < t} \hbar\,(t, t_i)\right). \tag{4.24}$$

This expression clearly shows the relationship between the exciting function, past events and the intensity. More explicitly, each past event influences the future intensity $\lambda\big(t, \mathcal{H}\big)$ with the weight $\hbar$.

In the multivariate case [60], the effect of all the variates on the *i-th* one is represented in the exciting function:

$$\lambda^{i}\big(t, \mathcal{H}\big) = \vartheta\left(\zeta(t) + \sum_{j}^{N_v} \int \hbar^{ij}(t, s) d\mathcal{M}^{j}(s)\right), \tag{4.25}$$

where N_v is the number of variates, $\hbar^{ij}$ defines how the *j-th* variate affect the *i-th* variate, and $\mathcal{M}^{j}$ is the random counting measure of the *j-th* variate.

In this work, the process of creation and removal of simulated phonons is modeled in terms of a multivariate Hawkes PP in order to minimize the effects of the statistical fluctuations on the conservation of the system energy. In particular, the PP must ensure that the variation of energy in the phonon population closely mirrors the variation in the population of simulated charges, in spite of the difference between the weight of the particles in the two populations. In other words, in order to achieve high accuracy when modeling transients, the model must alter the rate of generation or removal of simulated phonons to adjust the instantaneous conservation of the system energy.

In this context, four type of variates allow representing all the energy exchange happening between the simulated populations. In particular, each phonon state with momentum $\boldsymbol{q}$ is connected to four variates:

- simulated electron emission scattering with phonons in $\boldsymbol{q}$, represented by *ce*
- simulated electron absorption scattering with a phonon in $\boldsymbol{q}$, represented by *ca*
- simulated phonon created with momentum $\boldsymbol{q}$ by emission scattering, represented by *pe*
- simulated phonon with momentum $\boldsymbol{q}$ removed from the simulation by absorption scattering, represented by *pa*.

These four types of variates allow the definition of the energy exchange between the two population in terms of an energy balance equation:

$$d\mathcal{E}(t) = d\mathcal{E}^{pe}(t) + d\mathcal{E}^{pa}(t) - d\mathcal{E}^{ce}(t) - d\mathcal{E}^{ca}(t), \tag{4.26}$$

where the notation $d\mathcal{E}(t)$ represents $E(t + dt) - E(t)$, and the superscript *ce*, *ca*, *pe*, and *pa* represent the four type of variate as previously described. In particular,

expressions for $d\mathcal{E}^{ce}$ and $d\mathcal{E}^{pe}$ have been explicitly defined in section 4.3 in the case of phonon emission. The quantities $d\mathcal{E}^{ca}(t)$ and $d\mathcal{E}^{pa}(t)$ are similarly defined for the case of phonon absorption:

$$\begin{aligned} d\mathcal{E}^{ca}(t) &= \textstyle\sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^c dN^{ca}(\boldsymbol{q},t) \\ d\mathcal{E}^{pa}(t) &= \textstyle\sum_{\boldsymbol{q}\subset BZ1} \hbar\omega_{\boldsymbol{q}} W^p dN^{pa}(\boldsymbol{q},t) \end{aligned} \tag{4.27}$$

The time integral of $d\mathcal{E}(t)$ is a measure of the error in the conservation of energy between the simulated populations.

The expected value, or expectation, of $d\mathcal{E}(t)$ when generation and removal of simulated phonon are modeled as Hawkes PP can be written as:

$$\begin{aligned} \mathbb{E}\left[d\mathcal{E}(t)|\mathcal{H}\right] = \sum_{\boldsymbol{q}} & \hbar\omega_{\boldsymbol{q}} W^p (\mathbb{E}\left[dN^{pa}(\boldsymbol{q},t)|\mathcal{H}\right] - \mathbb{E}\left[dN^{pe}(\boldsymbol{q},t)|\mathcal{H}\right]) \\ & - \hbar\omega_{\boldsymbol{q}} W^c (\mathbb{E}\left[dN^{ca}(\boldsymbol{q},t)|\mathcal{H}\right] - \mathbb{E}\left[dN^{ce}(\boldsymbol{q},t)|\mathcal{H}\right]) \end{aligned}, \tag{4.28}$$

where the sum has been separated because of the linearity of the expectation operator. The definition of PP intensity and 4.15 allows the rewriting of the previous equation in terms of the intensities:

$$\mathbb{E}\left[d\mathcal{E}(t)|\mathcal{H}\right] = dt \sum_{\boldsymbol{q}} \hbar\omega_{\boldsymbol{q}} W^p (\lambda^{pa}(\boldsymbol{q},t) - \lambda^{pe}(\boldsymbol{q},t)) - \hbar\omega_{\boldsymbol{q}} W^c (\lambda^{ca}(\boldsymbol{q},t) - \lambda^{ce}(\boldsymbol{q},t)), \tag{4.29}$$

where $\lambda^{ca}(\boldsymbol{q},t)$ and $\lambda^{ce}(\boldsymbol{q},t)$ are the intensities corresponding to the scattering of simulated charged particles as defined in section 4.3. $\lambda^{pa}(\boldsymbol{q},t)$ and $\lambda^{pe}(\boldsymbol{q},t)$ are the intensities corresponding to the removal and generation of simulated phonon respectively. A simple algebraic substitution shows that the thinning algorithm presented in section 4.2 produces in an expected value of zero for $d\mathcal{E}(t)$ at steady state.

When the creation and the removal of simulated phonons are modeled as Hawkes processes, the intensities $\lambda^{pe}(\boldsymbol{q},t)$ and $\lambda^{pa}(\boldsymbol{q},t)$ are designed to minimize the error in the conservation of energy between the simulated populations, which is represented by

the integral $\int d\mathcal{E}(t)\, dt$. In particular, the desired properties for the intensity $\lambda^{pa}(\boldsymbol{q}, t)$ are as follows:

- $\lambda^{pa}(\boldsymbol{q}, t)$ has a background intensity of $\lambda^{ca}(\boldsymbol{q}, t) * W^c/W^p$
- $\lambda^{pa}(\boldsymbol{q}, t)$ has an inverse dependence on $\int d\mathcal{E}(t)$
- $\lambda^{pa}(\boldsymbol{q}, t)$ has a maximum intensity of $\lambda^{ca}(\boldsymbol{q}, t)$.

The first property ensures that in absence of an error the process behaves like the Poisson model described in the previous section, the second property imposes a negative feedback on the value of $\int d\mathcal{E}(t)$, and the final property allows the simplification of the algorithm by ensuring that each simulated charged particle interacts with no more than one simulated phonon.

In a similar fashion, the desired properties for the intensity λ^{pe} are:

- $\lambda^{pa}(\boldsymbol{q}, t)$ has a background intensity of $\lambda^{ca}(\boldsymbol{q}, t) * W^c/W^p$
- $\lambda^{pa}(\boldsymbol{q}, t)$ has a forward dependence on $\int d\mathcal{E}(t)$
- $\lambda^{pa}(\boldsymbol{q}, t)$ has a maximum intensity of $\lambda^{ca}(\boldsymbol{q}, t)$.

There is a variety of definitions of λ^{pe} and λ^{pa} which can satisfy these properties, and each definition has its own merits and drawbacks. The most basic approach would employ a step-like intensity, for example:

$$\lambda^{pe}(\boldsymbol{q}, t) = \begin{cases} \lambda^{ce}(\boldsymbol{q}, t)\frac{W^c}{W^p} & \int d\mathcal{E}(t) \leq 0 \\ \lambda^{ce}(\boldsymbol{q}, t) & \int d\mathcal{E}(t) > 0 \end{cases} . \tag{4.30}$$

The implementation of this approach would result in an extreme burden on the algorithm by continuously imposing an unnecessary amount of removal and creation of simulated phonons. Moreover, the intensity reduces to the Poisson model presented

in section 4.2 only when $\int d\mathcal{E}(t) = 0$, which is not a likely event. This means that at steady state, the removal and creation of simulated phonons will largely differ from the expected Poisson distribution. This problem could be mitigated by using some constant $\mathcal{E}^*_{max}$ as a threshold:

$$\lambda^{pe}(\boldsymbol{q}, t) = \begin{cases} \lambda^{ce}(\boldsymbol{q}, t)\frac{W^c}{W^p} & \int d\mathcal{E}(t) < \mathcal{E}^*_{max} \\ \lambda^{ce}(\boldsymbol{q}, t) & \int d\mathcal{E}(t) \geq \mathcal{E}^*_{max} \end{cases} . \tag{4.31}$$

This approach does not burden the algorithm and it reproduces the desired Poisson-like distribution behavior for $\int d\mathcal{E}(t) < \mathcal{E}^*_{max}$. However, there is no effect on the intensity until $\int d\mathcal{E}(t) \geq \mathcal{E}^*_{max}$, and this can result in large delay between the algorithm response and the moment a statistical fluctuation causes the error in $\int d\mathcal{E}(t)$. This effect results in a statistical difference between the momentum distribution of the created phonons and the phonon states which created the error by scattering with the simulated charge carriers. In other words, the momentum distribution of the generated phonons does not correspond to the variation in the momentum of the scattered charged particles.

This issue can also be mitigated by employing some type of transition function between the maximum and minimum value of the intensity. The simplest approach is to employ a linear relationship in the definition of intensity:

$$\lambda^{pe}(\boldsymbol{q}, t) = \lambda^{ce}(\boldsymbol{q}, t)\frac{W^c}{W^p} + a \int d\mathcal{E}(t), \tag{4.32}$$

where the proportionality constant a determines the sensitivity of the intensity. Small values of a closely reproduce the Poisson-like presented in section 4.2 but allow for large values of the error $\int d\mathcal{E}(t)\, dt$. In contrast, while large values of a limit the magnitude of the error, the behavior of the intensity would largely differ from the Poisson case presented presented in section 4.2.

This work employs a log-linear model to balance the efficiency of the algorithm, the

need for an instantaneous response, and the need to limit the error $\int d\mathcal{E}(t)\,dt$. This approach provides an immediate and modulated response to the error represented by $\int d\mathcal{E}(t)\,dt$. In particular, the effect is minimal and the rate of removal and creation of simulated phonons does not diverge much from the Poisson model presented in section 4.2. As the error grows, the effect on the intensity becomes more pronounced ensuring the desired small delay between the insurgence of an error and the algorithm response.

4.6 Log Linear Hawkes Process

In general, a log-linear approach models functions whose logarithm equals a linear combination of parameters. In this case, it is the logarithm of the intensity that is proportional to a linear function. This is simply achieved by imposing ϑ in Eq. 4.24 to be an exponential function. In particular, $\lambda^{pe}(\boldsymbol{q},t)$ and $\lambda^{pa}(\boldsymbol{q},t)$ are respectively:

$$\begin{array}{l}\lambda^{pe}\big(\boldsymbol{q},t,\mathcal{H}\big) = \exp\Big(L^{pe}(\boldsymbol{q},t) + \sum_{\boldsymbol{q}}\int_{-\infty}^{t} h^{pe}(\boldsymbol{q},t,s)d\mathcal{M}(s)\Big) \\ \lambda^{pa}\big(\boldsymbol{q},t,\mathcal{H}\big) = \exp\Big(L^{pa}(\boldsymbol{q},t) + \sum_{\boldsymbol{q}}\int_{-\infty}^{t} h^{pa}(\boldsymbol{q},t,s)d\mathcal{M}(s)\Big)\end{array}. \tag{4.33}$$

In the absence of the exciting function $\hbar$, the intensity should reduce to the Poisson model presented in section 4.2. This allows to derive expressions for the terms $L^{pe}(\boldsymbol{q},t)$ and $L^{pe}(\boldsymbol{q},t)$ by imposing:

$$\begin{array}{lll}\frac{W^c}{W^p}\lambda^{ce} & = \lambda^{pe}\big(\boldsymbol{q},t,\mathcal{H}\big) & = \exp\left(L^{pe}(\boldsymbol{q},t)\right) \\ \frac{W^c}{W^p}\lambda^{ca} & = \lambda^{pa}\big(\boldsymbol{q},t,\mathcal{H}\big) & = \exp\left(L^{pa}(\boldsymbol{q},t)\right)\end{array}, \tag{4.34}$$

which lead us to rewrite Eq. 4.33 as:

$$\begin{array}{ll}\lambda^{pe}\big(\boldsymbol{q},t,\mathcal{H}\big) & = \frac{W^c}{W^p}\lambda^{ce}\exp\Big(\sum_{\boldsymbol{q}}\int_{-\infty}^{t} h^{pe}(\boldsymbol{q},t,s)d\mathcal{M}(s)\Big) \\ \lambda^{pa}\big(\boldsymbol{q},t,\mathcal{H}\big) & = \frac{W^c}{W^p}\lambda^{ca}\exp\Big(\sum_{\boldsymbol{q}}\int_{-\infty}^{t} h^{pa}(\boldsymbol{q},t,s)d\mathcal{M}(s)\Big)\end{array}. \tag{4.35}$$

The exciting functions must be proportional to the measure of error presented in Eq 4.8, and the two intensity should have opposite behavior:

$$h^{pe}(\boldsymbol{q},t,s) = -h^{pa}(\boldsymbol{q},t,s) = a_{\boldsymbol{q}}\big(\, d\mathcal{E}^{pa}(q,t) - d\mathcal{E}^{pe}(q,t) - d\mathcal{E}^{ca}(q,t) + d\mathcal{E}^{ce}(q,t)\big). \tag{4.36}$$

This equation allows us to explicitly write the arguments of the exponentials in Eq. 4.35:

$$\begin{array}{ll} \sum_{\boldsymbol{q}} \int_{-\infty}^{t} h^{pe}(\boldsymbol{q},t,s) d\mathcal{M}(s) & = +a \int_{-\infty}^{t} d\mathcal{E}(t) \\ \sum_{\boldsymbol{q}} \int_{-\infty}^{t} h^{pa}(\boldsymbol{q},t,s) d\mathcal{M}(s) & = -a \int_{-\infty}^{t} d\mathcal{E}(t) \end{array}, \tag{4.37}$$

where the factor $a_{\boldsymbol{q}} = a$ is assumed constant in $\boldsymbol{q}$, the summation and integration order have been reversed, and $d\mathcal{E}(t)$ is the same as in Eq. 4.26.

Finally, the intensities for the creation and removal of simulated phonons are, respectively:

$$\begin{array}{ll} \lambda^{pe}\big(\boldsymbol{q},t,\mathcal{H}\big) & = min(\frac{W^c}{W^p}\lambda^{ce} e^{+a\int d\mathcal{E}(t)}, \lambda^{ce}) \\ \lambda^{pa}\big(\boldsymbol{q},t,\mathcal{H}\big) & = min(\frac{W^c}{W^p}\lambda^{ca} e^{-a\int d\mathcal{E}(t)}, \lambda^{ca}) \end{array}, \tag{4.38}$$

where the minimum function accounts for the limit imposed to the maximum intensity. The factor a has the dimensions of the inverse of an energy, intuitively this factor determines the sensitivity of the intensity to changes in $\int d\mathcal{E}(t)$.

In the specific case of:

$$\begin{array}{ll} \lambda^{pa} & = \frac{W^c}{W^p}\lambda^{ca} e^{-a\int d\mathcal{E}(t)} \\ \lambda^{pe} & = \frac{W^c}{W^p}\lambda^{ce} e^{a\int d\mathcal{E}(t)} \end{array}, \tag{4.39}$$

the expectation in 4.29 can be simplified as:

$$\begin{aligned} \mathbb{E}\left[d\mathcal{E}(t)|\mathcal{H}\right] &= dt \sum_{q} \hbar\omega_q W^c \left(\lambda^{ca}(q,t)(e^{-a\int d\mathcal{E}(t)} - 1) - \lambda^{ce}(q,t)(1 - e^{a\int d\mathcal{E}(t)})\right) \\ &= dt(e^{-a\int d\mathcal{E}(t)} - 1)\mathcal{E}^{ca}(t) - dt(e^{a\int d\mathcal{E}(t)} - 1)\mathcal{E}^{ce}(t) \end{aligned}. \tag{4.40}$$

These equations allow us to write the expected values for the integral of $d\mathcal{E}(t)$ in terms of a recurrence relation [61]:

$$\mathbb{E}\left[\int_{-\infty}^{t_{i+1}} d\mathcal{E}(t)|\mathcal{H}\right] = E_{i+1} = E_i + (e^{-aE_i} - 1)\, dt\, \mathcal{E}^{ca} - (e^{aE_i} - 1)\, dt\, \mathcal{E}^{ce}, \qquad (4.41)$$

where t_i is the time of the *i-th* event. A simple substitution allows us to prove that $E_{i+1} = 0$ in the special case of $E_i = 0$. Since at the start of the simulation, the net energy exchange between the simulated populations is zero $E_0 = 0$. Therefore, $E_i = 0$ for each n, and the expected value for the integral of $d\mathcal{E}(t)$ must be zero for each t. Since $\int d\mathcal{E}(t)$ is a measure of the error in the conservation of energy in scattering between the simulated populations, the energy gained by the simulated phonon population must be equal to the energy lost by the population of simulated charged particles.

Finding the analytical expression for the variance of $\int d\mathcal{E}(t)$ is challenging and beyond the scope of this work. However, an upper bound to the central variances can be computed by employing the Popoviciu's inequality [62]. In particular, the variance of a random variable x is bound by:

$$\mathrm{Var}(x) \leq (max(x) - min(x))^2/4, , \qquad (4.42)$$

where $max(x)$ and $min(x)$ represents the maximum and minimum value of x respectively.

In order to compute the extreme of $\int d\mathcal{E}(t)$, we consider separately the cases of small and large values of a. For large values of a, the intensities defined in Eq. 4.38 oscillate between zero and the maximum value. In this case, the integral $\int d\mathcal{E}(t)$ is bounded by the discrete energy of the simulated phonons:

$$-W^p \hbar\omega_{max} < \int d\mathcal{E}(t) < +W^p \hbar\omega_{max}, \qquad (4.43)$$

where ω_{max} is the maximum phonon energy.

For small values of a, the extreme of $\int d\mathcal{E}(t)$ can be obtained from the definitions of intensity in Eq. 4.38. In particular, for positive values of $\int d\mathcal{E}(t)$ the upper bound is computed by imposing $\lambda^{pe} \leq \lambda^{ce}$:

$$\int d\mathcal{E}(t) < +\frac{1}{a} \ln \frac{W^p}{W^c}. \tag{4.44}$$

Similarly, for negative values of $\int d\mathcal{E}(t)$ the lower bound is computed by setting $\lambda^{pa} \geq \lambda^{ca}$:

$$\int d\mathcal{E}(t) > -\frac{1}{a} \ln \frac{W^p}{W^c}. \tag{4.45}$$

The inequalities in Eq. 4.43, 4.44, and 4.45 provide the final bound to $\int d\mathcal{E}(t)$:

$$\begin{aligned} \int d\mathcal{E}(t) < &\quad +\tfrac{1}{a} \ln \tfrac{W^p}{W^c} + W^p \hbar\omega_{max} \\ \int d\mathcal{E}(t) > &\quad -\tfrac{1}{a} \ln \tfrac{W^p}{W^c} - W^p \hbar\omega_{max} \end{aligned}, \tag{4.46}$$

Finally, the upper bound to the central variances is computed from Eq. 4.42 as:

$$\mathrm{Var}\left(\int d\mathcal{E}(t)\right) < \left(\frac{1}{a} \ln \left(\frac{W^p}{W^c}\right) + W^p \hbar\omega_{max}\right)^2. \tag{4.47}$$

This equation confirms the intuitive notion that the variance depends on the relationship between W^p and W^c. Moreover, the equation provides a practical approach for the selection of the parameter a by choosing the magnitude of the variance.

4.7 Simulation of Hawkes Process

The literature proposes various sampling algorithms [63–66] to generate samples of point processes. This work employs a technique based on Ogata's [65] thinning approach.

Ogata frames the thinning algorithm in terms of a target and a sampling PP when both PP s are *simple*, as defined in section 4.4. In this context, the rejection probability is expressed as a ratio of the PP intensity [49, 65]. If the intensities of both PP are deterministic, as in a Poisson PP, the thinning procedure follows

the basic thinning algorithm [47]. However, when at least one of the intensities is a random variable, as in a Hawkes processes, the algorithm must account directly [49] for the inequality in 4.5. That is, at any time the intensity of the sampling PP must be larger than the intensity of the target PP. Ogata achieves this goal by employing a sequential variant of the thinning algorithm. In this sequential approach, the inequality is ensured by recomputing the intensity of the sampling PP after each event alongside to the value of the intensity of the target PP.

In the case of the log-linear model presented in this work, the intensities of the Hawkes PP, as defined in Eq. 4.38, are bounded random variables. This condition allows to largely simplify the sampling algorithm by selecting an appropriate sampling PP which ensure the inequality in Eq. 4.5. The definition in Eq. 4.38 provides that the intensity for the generation of simulated phonons is always smaller than the intensity for the emission scattering of simulated charge carriers. Similarly, the intensity for the removal of simulated phonons is always smaller than the intensity for the absorption scattering of simulated charge carriers. These two relationships allow to implement the thinning procedure by simply following the procedure detailed in section 4.2 while employing the following rejection probabilities:

$$\begin{aligned} P^{a}_{rej} &= \frac{\lambda^{pa}\left(\boldsymbol{q},t,\mathcal{H}\right)}{\lambda^{ca}(\boldsymbol{q},t)} = \frac{W^{c}}{W^{p}}\, e^{-a\int d\mathcal{E}(t)} \\ P^{e}_{rej} &= \frac{\lambda^{pe}\left(\boldsymbol{q},t,\mathcal{H}\right)}{\lambda^{ce}(\boldsymbol{q},t)} = \frac{W^{c}}{W^{p}}\, e^{a\int d\mathcal{E}(t)} \end{aligned} \quad , \tag{4.48}$$

where P^{a}_{rej} and P^{e}_{rej} represent the probability of removing and creating a simulated phonon respectively, a is the same as in Eq. 4.38. and $\int d\mathcal{E}(t)$ is defined in Eq. 4.26.

Chapter 5

ELECTRON SCATTERING

This chapter offers an overview of the scattering mechanisms involving electrons and how the rejection technique is applied to model them. The first three sections discuss scattering between the electrons and their environment, while the last section analyze the modeling of electron-electron interactions. In all equations of this chapter the phonon mode index and the electron band index have been dropped for the sake of clarity.

5.1 Deformation Potential Scattering

Interactions between the Bloch electron and the deformations of the crystal periodicity caused by the thermal vibrations of the lattice atoms are commonly referred to as deformation potential scattering. This phenomenon can be expressed as either the absorption or the emission of an elementary quantum of thermal energy (phonon). In particular, an electron with wavevector $\boldsymbol{k}$ interacting with a phonon of wavevector $\boldsymbol{q}$ acquires a new wavevector $\boldsymbol{k'} = \boldsymbol{k} \pm \boldsymbol{q}$ where "+" and "-" refer to the absorption and the emission of the phonon, respectively. Moreover, the energy exchanged during the process is obviously: $\mathcal{E}(\boldsymbol{k'}) = \mathcal{E}(\boldsymbol{k}) \pm \mathcal{E}(\boldsymbol{q})$.

According to the approach discussed in section 2.4, the transition rate to be stored in the look-up table is expressed as the probability for a carrier with initial wavevector $\boldsymbol{k}$ to scatter to the cell $\Omega_{\boldsymbol{k'}}$ centered around $\boldsymbol{k'}$ [36]:

$$\Gamma^{def}(\boldsymbol{k}, \Omega_{\boldsymbol{k'}}) = \frac{\pi D(\mathcal{E}(\boldsymbol{k}) \pm \mathcal{E}(\boldsymbol{q})), \Omega_{\boldsymbol{k'}}}{\rho_{\mathcal{E}}(\boldsymbol{q})} \left|\Xi_{def}(\boldsymbol{k}, \boldsymbol{q}, \boldsymbol{k'})\right|^2 \left|I(\boldsymbol{k}, \boldsymbol{k'})\right|^2 \left(n + \frac{1}{2} \mp \frac{1}{2}\right), \tag{5.1}$$

where $D(\mathcal{E}, \Omega_{\boldsymbol{k}'})$ is the density of states with energy $\mathcal{E}$ in $\Omega_{\boldsymbol{k}'}$, ρ is the semiconductor density, $\Xi_{def}(\boldsymbol{k}, \boldsymbol{q}, \boldsymbol{k}')$ is the deformation potential parameter [36], $I(\boldsymbol{k}, \boldsymbol{k}')$ is the overlap integral [36] between the initial and final states, and n is the number of phonons with wavevector $\boldsymbol{q}$. The rate Γ^{def} can be written as the products of a constant term dependent only on the wavevectors $\boldsymbol{k}$, $\boldsymbol{k}'$, and $\boldsymbol{q}$, and a second term dependent on $n(\boldsymbol{q})$, which is the phonon occupation number. In the local equilibrium approximation, this second term is simply a function of the (local) crystal temperature.

In the case of absorption, an electron has the same chances of absorbing each of the n indistinguishable phonons in the state $\boldsymbol{q}$, as is reflected in the scattering rate that is linearly dependent on n. This obviously implies that the absorption process requires the presence of at least one phonon in the state $\boldsymbol{q}$.

An electron has also the same chances of interacting with each of the n indistinguishable phonons in the state $\boldsymbol{q}$. In this interaction, the charge will lose energy by emitting a phonon. Moreover, a phonon emission in the state $\boldsymbol{q}$ can happen even in the absence of phonons in that state. This two mechanisms are reflected in the total emission rate which is varying linearly with $n + 1$.

Since the rejection probability is expressed by the ratio between the locally computed scattering rate and the tabulated scattering rate, the constant terms are common and can be simplified.

The rejection probability employed by the scattering algorithm is therefore:

$$P_{rej}^{def} = \frac{\Gamma_{loc}^{def}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'})}{\Gamma_{max}^{def}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'})} = \frac{\left(n_{loc} + \frac{1}{2} \mp \frac{1}{2}\right)}{\left(n_{max} + \frac{1}{2} \mp \frac{1}{2}\right)} \tag{5.2}$$

where n_{loc} is the local number of phonons with wavevector $\boldsymbol{q}$, and n_{max} is the number of phonons with wavevector $\boldsymbol{q}$ computed at the maximum expected temperature T_{max} used to compute the look-up table. Importantly, this approach allows to employ the rejection technique without the need of the expansive calculation of the defor-

mation potential parameter, the overlap integral and the density of states after each scattering, so the numerical overhead of the rejection is almost negligible.

5.2 Polar Interactions

When phonons distort the lattice of a polar material, the local electric charge neutrality is destroyed. This effect causes an electric polarization and an associated macroscopic electric field to which the electron responds [67]. Similar to the deformation potential case, this interaction results in an exchange of energy and momentum between the Bloch electron and the crystal via the emission and absorption of phonons.

Polar interactions between electrons and optical phonons are referred to as polar optical scattering and can be expressed [37, 68–71] as:

$$\Gamma^{pol}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'}) = D(\mathcal{E}, \Omega_{\boldsymbol{k}'}) \frac{2\pi e^2}{\hbar |\boldsymbol{q}|^2} |\Xi_{pol}(\boldsymbol{k}, \boldsymbol{q}, \boldsymbol{k}')|^2 |I(\boldsymbol{k}, \boldsymbol{k}')|^2 \left(n + \frac{1}{2} \mp \frac{1}{2} \right), \tag{5.3}$$

where e is the elementary charge, $\boldsymbol{q}$ is the wavevector of the phonon, and $\Xi_{pol}(\boldsymbol{k}, \boldsymbol{q}, \boldsymbol{k}')$ is the polar scattering parameter. Since Γ^{pol} and Γ^{def} share a similar mathematical structure the considerations made for the deformation potential scattering also apply to the polar optical case.

In the case of polar optical scattering, the simplified rejection probability is:

$$P^{pol}_{rej} = \frac{\Gamma^{pol}_{loc}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'})}{\Gamma^{pol}_{max}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'})} = \frac{\left(n_{loc} + \frac{1}{2} \mp \frac{1}{2}\right)}{\left(n_{max} + \frac{1}{2} \mp \frac{1}{2}\right)}. \tag{5.4}$$

Polar disturbances of the electron dynamics related to an acoustic phonon are referred to as piezoelectric scattering. This scattering is usually treated as elastic because of the predominance of interactions with low energy phonons near the gamma valley. The scattering rate can be expressed as [67]:

$$\Gamma^{piezo}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'}) = D(\mathcal{E}(\boldsymbol{k}), \Omega_{\boldsymbol{k}'}) \frac{2\pi}{\hbar} K^2_{av} \frac{e^2 k_B T}{|\boldsymbol{q}|^2 \epsilon^*} |I(\boldsymbol{k}, \boldsymbol{k}')|^2, \tag{5.5}$$

where K_{av}^2 is the electro-mechanical coupling constant.

For this scattering mechanism, the rejection probability further simplifies to:

$$P_{rej}^{piezo} = \frac{\Gamma_{loc}^{piezo}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'})}{\Gamma_{max}^{piezo}(\boldsymbol{k}, \Omega_{\boldsymbol{k}'})} = \frac{T_{loc}}{T_{max}}, \tag{5.6}$$

where T_{loc} is the temperature at the scattering electron location, and T_{max} is the maximum expected temperature used to compute the look-up table.

5.3 Ionized Impurities

Ionized impurities in the lattice are also treated as a source of elastic scattering [34, 72], in the sense that only the direction of electrons' motion is changed by the scattering. Since this scattering mechanism is related to the concentration of impurities, as the impurity concentration increases, the carriers are subject to more scattering, consequently increasing the randomization of the carriers velocities.

While the scattering with ionized impurities hardly affects the saturation drift velocity [73] and the transport at high field, it is crucial in low field transport. Indeed, as the ionized impurities scattering increases, the mobility of carriers decreases, and a correct modeling of the mobility is important for accurate predictions of device performances. Moreover, the randomization of the carrier velocity via impurities increases the output current noise, so the doping concentration must be kept relatively low along the conduction path (channel), especially for low power analog applications.

Ridley's statistical screening model [74] allows modeling the ionized impurities scattering for a wide range of impurities concentration by combining the Brooks-Herring (BH) approach [75] with Conwell-Weisskopf (CW) model [76]. All these approaches assume that electrons interact with only one scattering center at a time.

In particular, the BH approach assumes that the distance between scattering centers is larger than the screening length and is valid when the screening effect is

strong. Under this assumption, the scattering can be treated as a two-body problem by using a screening parameter that accounts for the many-body effects [77]. The CW approach models the case of high impurity concentration or low temperature, when the screening effect is weak. This approach imposes a cut-off radius by assuming that the effects of all center lying further away than half the average distance between centers add up to zero.

Ridley's [74] approach to the problem was to quantify statistically the assumption that only one center is active. This approach effectively allows for a reconciliation [74] of the BH and the CW models as:

$$\Gamma^{ion}(\boldsymbol{k},\Omega_{\boldsymbol{k}'}) = D(\mathcal{E}(\boldsymbol{k}),\Omega_{\boldsymbol{k}'})\frac{|\boldsymbol{v}(\boldsymbol{k})|}{(2\pi\,\mathrm{Z_c})^{-1/3}}\Big[1-\exp\Big(\frac{\Gamma^{bh}(\boldsymbol{k},\boldsymbol{k}')(2\pi\,\mathrm{Z_c})^{-1/3}}{|\boldsymbol{v}(\boldsymbol{k})|}\Big)\Big]. \tag{5.7}$$

where $|\boldsymbol{v}(\boldsymbol{k})|$ is the carrier velocity, $\mathrm{Z_c}$ is the ionized impurity concentration, and $\Gamma^{bh}(\boldsymbol{k},\boldsymbol{k}')$ is [37]:

$$\Gamma^{bh}(\boldsymbol{k},\boldsymbol{k}') = \frac{Z_d^2\,\mathrm{Z_c}\,e^4}{4\pi^2\hbar\epsilon_0}\frac{|I(\boldsymbol{k},\boldsymbol{k}')|^2}{\left(\beta_s^2+|\boldsymbol{k}-\boldsymbol{k}'|^2\right)^2}\delta(\mathcal{E}(\boldsymbol{k})-\mathcal{E}(\boldsymbol{k}')), \tag{5.8}$$

where the quantity eZ_d is the charge of the dopant species, and β_s is the screening parameter depending on carriers concentration and distribution. Typically β_s is assumed to be equal to the Debye-Hückel [78] screening for non-degenerate systems.

In general, the ionized impurity concentration is a spatial parameter $N_i = N(\mathbf{r})$, while the screening depends both on time and position $\beta_s = \beta_s(\mathbf{r},t)$. The model for this scattering mechanism does not allow for a substantial simplification on of the rejection probability. However, the rejection algorithm still offers results similar to the standard EMC method while retaining an overwhelming computational advantage. Indeed, each scattering in the EMC approach requires building a small transition table by first selecting all the possible final state candidates, then computing the

corresponding rates for all the candidates, and the total partial sums, finally a final state can be statistically chosen. In contrast, the rejection algorithm requires computing only one transition rate for the preselected final state, then a simple comparison chooses between self scattering and scattering in the new state.

5.4 Carrier-Carrier Coulombic Interactions

Carrier-carrier coulombing interactions are particularly difficult to model efficiently, because the strength of the interactions depends on the local distribution function.

For this reason, the problem of modeling these phenomena is typically approached by treating separately interactions between carriers distant form each other (long range) and interactions between carriers in proximity to each other (short range).

The overall effect of long range interactions is a net force acting on the particles. This effect is accounted for throughout Hartree-Fock method by solving self-consistently Poisson's equation. The short-range Coulombic interactions are included throughout a screening length β_s by assuming a Debye-Hückel potential [79]:

$$\beta_s = \frac{\eta e^2}{V_c \epsilon k_B T_e} \frac{\mathcal{F}_{-1/2}(\mathcal{E}_f (k_B T_e)^- 1)}{\mathcal{F}_{1/2}(\mathcal{E}_f (k_B T_e)^{-1})} \tag{5.9}$$

where η is the number of carriers, V_c is the volume of the crystal, e is the elementary charge, k_B is the Boltzmann constant and T_e is the electron temperature, $\mathcal{F}$ is the Fermi integral, and $\mathcal{E}_f$ is the Fermi energy. In the case of non degenerate semiconductors, the carriers distribution is roughly Maxwellian, and the ratio of the Fermi integrals can be approximated to 1. As stated in previous sections, the screening length plays a crucial role in ionized impurities scattering that strongly affects the mobility at low electric field. For this reason, a local electron temperature $T_e(\mathbf{r}, t)$ is computed during the simulation by inverting the relationship between temperature

and average carrier energy $\mathbb{E}[\mathcal{E}]$ for a Maxwellian distribution. In particular, $<\mathcal{E}>$ is computed by employing a running exponential average, while the temperature is calculated as:

$$T_e = \frac{2\mathbb{E}[\mathcal{E}]}{3k_B}. \tag{5.10}$$

Carrier-carrier scattering does not dissipate either energy or momentum, but it influences the shape of the distribution function. These interactions can affect quantities which are more sensitive to the particular shape of the distribution function However, the literature [79] suggests electron-electron (or hole-hole) scattering has no first-order effect on the mobility. In particular, when the carrier concentration is low (as in high purity samples), the electron-electron scattering is infrequent, therefore, the effect on the mobility is not significant. At very large carrier concentrations, the conservation of energy and momenta near the Fermi surface implies that the distribution is practically unaltered by the electron-electron scattering, in this case the effect on the mobility is also expect to not be significant.

However, electron-electron scattering can play an important role in far from equilibrium situation, for example it can affect the hot-carrier degradation depending on the channel length and applied bias [80] and it may dominate the initial response of laser-excited carriers in bulk and quantum-well [44]. Since this process is fairly complex to model and not the focus of this work, electron-electron scattering is not included in the present code.

Chapter 6

PHONON MODELING

This chapter discusses the modeling of heat transport in solids, within the particle based framework discussed in the preceding Chapters. The first part reviews the theoretical aspects of heat transport modeling, while in the second part the modeling of crystal dynamics in the framework of particle-based CMC simulation is discussed.

In particular, the first section offers an introduction to the classic and the quantum theory of the harmonic crystal. The second section discusses the motivation and the history of the development of a phonon BTE. The third and fourth sections offer a formal treatment for the harmonic and anharmonic terms of the potential energy expansion, respectively. Finally, section 6.5 and 6.6 examine scattering mechanism involving phonons and how the rejection technique is applied to them. In all the equations of this chapter the phonon mode index has been dropped for the sake of clarity.

6.1 The Harmonic Crystal

In solids, thermal energy takes the form of collective atomic vibrations that are usually modeled by assuming both the adiabatic and the harmonic approximations [81, 82]. The adiabatic approximation is based on the observation that the electron velocity is typically orders of magnitude higher than the typical ion velocity. Since the particle dynamics operate on different timescales, one assumes that the ionic motion and the electron dynamics can be treated separately.

Within the harmonic approximation, the Taylor expansion of the crystal potential

energy, U^{tot}, is separated in three components:

$$U^{tot} = U^{eq} + U^{harm} + U^{anh}, \tag{6.1}$$

where U^{eq} is the equilibrium potential energy, which is constant, U^{harm} is the second order (harmonic) term of the potential expansion, U^{anh} contains the higher order (anharmonic) terms of the expansion.

The harmonic approximation postulates that the constant term, U^{eq}, has no effect on dynamical problems, and that the quadratic (harmonic) term dominates over higher order (anharmonic) terms.

The classic harmonic model based on these assumptions has two limitations: it cannot predict the relationship between crystal energy and temperature (heat capacity), and it cannot predict the relationship between energy flux and temperature gradient (thermal conductivity).

In the classic approach based on the equipartition theorem, the thermal energy density U^{tot} is given by the average of the possible ionic configuration weighted by $\exp(-\mathcal{E}/K_B T)$ [82]:

$$U^{tot} = \frac{\int d\chi \, \exp(\frac{H_c}{k_B T_c}) H_c}{\int d\chi \, \exp(\frac{H_c}{k_B T_c})}, \tag{6.2}$$

where T_c is the crystal temperature, k_B is the Boltzmann constant, and H_c is the crystal Hamiltonian.

The differential $d\chi$ represents for the volume element in crystal's phase space:

$$d\chi = \prod_{\mathbf{r}} \mathbf{u}_b \mathbf{P}(\mathbf{R}_b) \tag{6.3}$$

where $\mathbf{R}_b$ is an equilibrium position of the Bravais lattice, while $\mathbf{u}_b$ and $\mathbf{P}(\mathbf{R}_b)$ are the deviations from equilibrium and the momentum of the ion whose equilibrium position is $\mathbf{R}_b$, respectively. Equation 6.1 in the harmonic approximation reduces to:

$$U^{tot} = U^{eq} + \frac{3 N_a k_B T_c}{V_c}, \tag{6.4}$$

where V_c is the crystal volume, and N_a is the number of atoms in the crystal. The volumetric heat capacity, C_v, predicted by this model is constant:

$$C_v = \frac{\partial u}{V_c \partial T_c} = \frac{N_a k_B}{V_c}. \tag{6.5}$$

This result is in contrast with the experimental data, that show clearly that the heat capacity increases with the temperature. While such discrepancy at higher temperature could be explained as a limit of the harmonic approximation, this model fails even at lower temperatures, where the harmonic approximation should be reliable. The failure of the model is rooted in the underlying *continuum hypothesis* in Eq. 6.2. This hypothesis assumes that the states of the crystal momentum are part of a continuous space. Indeed, the correct behavior for the thermal capacity can be obtained via quantum mechanical modeling of the crystal momentum. In this case, heat is a collection of non-interacting quantum harmonic oscillators, therefore, the total thermal energy is calculated as:

$$U^{tot} = \sum_i \hbar\omega_i (n_i(T_c) + \frac{1}{2}), \tag{6.6}$$

where ω_i and n_i are the angular frequency and the (temperature dependent) excitation number of the $i-th$ normal mode [81, 82], respectively. The heat capacity therefore becomes:

$$C_v = \frac{1}{V_c} \sum_i \frac{\partial \hbar\omega_i n_i}{\partial T}, \tag{6.7}$$

which depends both on the frequency spectrum of the normal modes and on the temperature through the excitation number.

6.2 Thermal conductivity in the Harmonic Crystal

The description of heat based on non-interacting quantum harmonic oscillators allows to successfully predict the temperature dependence of the thermal capacity.

After this early success, some authors thought that this idealized model was also suitable to treat quasi-stationary problems. In particular, Schrödinger [83] and Kármán [84], attempted to prove that the assumption of quantum harmonic oscillator give rise to a finite thermal conductivity. However, this attempt has been rebutted by Ornstein and Zernjke [85], who proved that the conduction of heat is infinite within the harmonic model.

Their argument can be qualitatively explained by observing that energy exchanges between different waves are prevented by the conservation of momentum and energy. In other words, the harmonic model describes the motions of the atoms in terms of mutually independent sound waves. Since the vibrations are not interacting, any arbitrary initial energy distribution will be time independent and last forever. In this framework, there is not a thermal equilibrium which can be established. Indeed, the very concept of temperature loses its meaning since there cannot be an equilibrium distribution corresponding to the configuration of maximum entropy. Therefore, a thermal conductivity defined by the relationship between the net energy flux and the gradient of the temperature cannot be obtained. Moreover, this model also fails when modeling thermal transport directly in terms of energy. In this approach, the temperature is replaced by the energy density, and a quantity analogous to the thermal conductivity is defined as the ratio between the energy flux and the gradient of the energy density. In this framework, the absence of energy exchanges between states results in a stationary energy distribution with a net flux of energy but without energy gradient, which corresponds to an infinite value for the thermal conductivity analogous. In other words, the states in stationary distribution can provide a net transport of energy between two regions of space in equilibrium with each other, which is obviously unphysical.

Debye first observed that introducing the cubic terms of the potential expansion in

the harmonic sound waves model leads to normal modes that change with time. This model considered a wave-packet scattering against static deformations of the lattice which give rise to statistical (thermal) equilibrium between normal modes, allowing for the concept of temperature. In this model, imposing a stationary distribution results in a zero net energy flux and hence there cannot be any net transport of energy between two regions of space in equilibrium with each other. In other words, this model ensures that any net transport of energy requires the presence of a temperature gradient. Consequently, a finite thermal conductivity can always be obtained. In addition, the Debye model introduces the concept of lifetime and mean free path for an acoustic wave-packet.

While Debye noticed the importance of the anharmonic terms in modeling thermal transport phenomena, Ornstain and Zernick [85] observed that his assumption of static deformations of the lattice is not admissible. They argued that, since the deformations are created by propagating wave packet, the deformations themselves should propagate with the same speed. They proposed a model based on a one-dimensional elastic continuous with propagating deformation to show how the energy is exchanged between states.

Their approach results in a model where the energy flux remains unchanged when two waves interacts; this allows for the existence of stationary distributions with non-zero energy flux, thus infinite thermal conductivity. Indeed, Peierls [86] proved that the continuum assumption, which neglects the atomic structures, results in infinite thermal conductivity also in the three-dimensional case and even when including arbitrary high terms of the potential expansion. This can be intuitively explained by observing that the conservation of energy and momentum in a continuum momentum space allows for the existence of a *closed* group of states where the exchange of energy does not alter the collective flux. In other words, there is a group of states such that

each state can exchange energy only with other states within the group, and each exchange of energy conserves the total energy flux of the state involved.

In contrast, accounting for the atomic structure results in a quantized and periodic momentum space, which provides an additional path for the energy exchanges between vibrational states. Indeed, the energy transfer between states that conserve momentum within the BZ1 periodicity (Umklapp processes) is allowed. These type of processes ensure [86] that there cannot exist a group of states which exclusively exchange energy with other states within the same group, while simultaneously conserving the group energy flux with each exchange. For this reason, the literature usually differentiate between Umklapp processes and energy exchanges that conserve momentum explicitly (called Normal processes).

In addition, a quantized approach allows modeling the crystals vibrations as a linear combination of normal modes of vibration with quantized energy [81]. In this case, each elementary excitation can be associated to a quasi-particle called a phonon. Peierls noticed the similarity between the problem of phonons motion in solid and the problem of gas dynamics, and developed his theoretical framework by employing a procedure similar to the kinetic theory of gases. This approach accounts for the existence of an equilibrium phonon distribution and produces [22] a transport model for phonons formally similar to the BTE.

6.3 Normal modes of vibration

The total potential energy of the crystal can be expressed as the summation of the contribution of potential energy Φ between each pair of ions:

$$U^{tot} = \frac{1}{2} \sum_{b}^{N_a} \sum_{b'}^{N_a} \Phi(\mathbf{R}_b - \mathbf{R}_{b'} + \mathbf{u}_b - \mathbf{u}_{b'}) \tag{6.8}$$

where N_a is the number of ions in the crystal, $\mathbf{R}_b$ and $\mathbf{u}_b$ Are the equilibrium position and the deviation from the equilibrium position of the *b-th* ions, respectively. This potential energy can by expanded in a Taylor series for small values of $\mathbf{u}_b - \mathbf{u}_{b'}$. In this case, the equilibrium energy is:

$$U^{eq} = \frac{1}{2}\sum_{b}^{N_a}\sum_{b'}^{N_a}\Phi(\mathbf{R}_b - \mathbf{R}_{b'}). \tag{6.9}$$

The first order, or linear term of the Taylor expansion is simply:

$$\frac{1}{2}\sum_{b}^{N_a}\sum_{b'}^{N_a}[\mathbf{u}_b - \mathbf{u}_{b'}]\cdots\nabla\Phi\Big|_{\mathbf{R}_b-\mathbf{R}_{b'}}. \tag{6.10}$$

This term can be rearranged by separating the summations in:

$$\frac{1}{2}\sum_{b}^{N_a}\sum_{\mathbf{R}_b}\mathbf{u}_b\cdots\sum_{b'}^{N_a}\nabla\Phi\Big|_{\mathbf{R}_b-\mathbf{R}_{b'}} - \frac{1}{2}\sum_{b'}^{N_a}\mathbf{u}_{b'}\cdots\sum_{b}^{N_a}\nabla\Phi\Big|_{\mathbf{R}_b-\mathbf{R}_{b'}}. \tag{6.11}$$

Since Φ is a scalar potential, its gradient is a force. In particular, $\nabla\Phi\Big|_{\mathbf{R}_b-\mathbf{R}_{b'}}$ represents the force between the atoms in $\mathbf{R}_b$ and $\mathbf{R}_{b'}$. Consequently, the term $\sum_{b'}^{N_a}\nabla\Phi\Big|_{\mathbf{R}_b-\mathbf{R}_{b'}}$ is the total forces exerted on the *b-th* atom. There are no net forces on any ion in the equilibrium position, therefore, this linear term must vanish.

The second order, or the harmonic term of the expansion is:

$$U^{har} = \frac{1}{4}\sum_{b}^{N_a}\sum_{b'}^{N_a}(\mathbf{u}_b - \mathbf{u}_{b'})\boldsymbol{H}_{\Phi}\big|_{\mathbf{R}_b-\mathbf{R}_{b'}}(\mathbf{u}_b - \mathbf{u}(\mathbf{R}_{b'})^T) \tag{6.12}$$

where $\boldsymbol{H}_{\Phi}\big|_{\mathbf{R}_b-\mathbf{R}_{b'}}$ is the Hessian matrix of Φ evaluated at $\mathbf{R}_b - \mathbf{R}_{b'}$. This expression is usually rearranged as:

$$U^{har} = \frac{1}{2}\sum_{b}^{N_a}\sum_{b'}^{N_a}\mathbf{u}_b\mathbf{D}(\mathbf{R}_b - \mathbf{R}_{b'}))\mathbf{u}_{b'}^T \tag{6.13}$$

where the $\mathbf{D}$ is a symmetric matrix resulting from the algebraic rearrangement of the summations [82]. In this framework, the equation of motion of the ions in the crystal becomes:

$$M\ddot{\mathbf{u}}(\mathbf{R}_b) = -\sum_{b'}^{N_a}\mathcal{D}(\mathbf{R}_b - \mathbf{R}_{b'})\mathbf{u}_{b'}{}^T, \tag{6.14}$$

where $\mathbf{u}_b$ is the deviation from equilibrium of the *b-th* ions.

This equation of motion allows for solutions in the form of plane waves [82]:

$$\mathbf{u}_b = \boldsymbol{\epsilon} \exp(\boldsymbol{q} \cdot \mathbf{R}_b - \omega t), \tag{6.15}$$

where $\boldsymbol{\epsilon}$ is known as the polarization vector, $\boldsymbol{q}$ is the wavevector, and ω is the angular frequency of the plane wave. This expression for $\mathbf{u}_b$ can used in Eq. 6.14 to obtain:

$$\omega^2 \boldsymbol{\epsilon} = \mathcal{D}(\boldsymbol{q})\boldsymbol{\epsilon} \tag{6.16}$$

where $\mathcal{D}(\boldsymbol{q})$, known as the dynamical matrix, has the form:

$$\mathcal{D}(\boldsymbol{q}) = \frac{1}{2} \sum_{b}^{N_a} \mathbf{D}(\mathbf{R}_b) \exp(\boldsymbol{q} \cdot \mathbf{R}_b). \tag{6.17}$$

Finally, the application of the Born-von Kármán periodic boundary condition produces quantized normal modes of vibration by restricting the allowed wavevector $\boldsymbol{q}$ to the form:

$$\boldsymbol{q} = \frac{j_1}{J_1}\boldsymbol{b}_1 + \frac{j_2}{J_2}\boldsymbol{b}_2 + \frac{j_3}{J_3}\boldsymbol{b}_3, \tag{6.18}$$

where the $\boldsymbol{b}_i$ are reciprocal lattice vectors, lowercase j_i represent an integer, and the uppercase J_i represent three integers such that $J_1 + J_2 + J_3 = N_a$, which is the total number of atoms in the crystal.

In the general case of a crystal with p bases, Eq. 6.16 becomes a set of p equations. In particular, the *i-th* equation is:

$$\omega^2 \boldsymbol{\epsilon}_i = \sum_{j=1}^{p} \mathcal{D}_i(\boldsymbol{q}) \boldsymbol{\epsilon}_j^T. \tag{6.19}$$

It can be shown [82] that solving this eigenproblem leads to $3p$ vibration modes. In the three solutions with lower energy, the ions in the unit cell move in parallel with equal amplitude, and their frequencies vanish at $\boldsymbol{q} = 0$. These are called acoustic modes since their behavior is characteristic of the displacement in an elastic continuum that propagates sound waves.

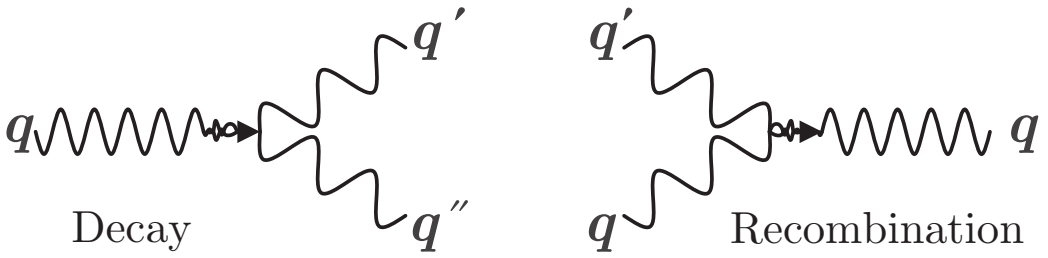

Figure 6.1: Diagram of three phonon scattering. On the left a phonon decays into two, on the right two phonons recombine into one.

In the remaining $3p-3$ solutions, the ions of the base vibrate with respect to each other even when the center of mass remains fixed at $\boldsymbol{q} = 0$. These modes of vibrations induce in the crystal a net fluctuating dipole moment that can interact with external electromagnetic fields. For this reason, these modes are usually referred to as optical modes.

6.4 Anharmonic Interaction

As discussed in the previous sections, the harmonic term of the potential energy expansion allows describing the crystal vibrations as a linear combination of normal modes of vibration with quantized energy. Consequently, the higher terms (anharmonic) can be treated as discrete energy exchanges between these modes. In this framework, each elementary excitation is associated to a quasi-particle called a phonon [81]. This approach allows us to model the interactions between the normal modes in terms of generation and annihilation of phonons. In particular, the anharmonic interactions are expressed in terms of phonons being annihilated in some state and new phonons been created in others. The use of perturbation theory [81, 87] in this phonons model shows that the *n-th* term of the potential expansion corresponds to a process involving n particles. As shown in section 6.2, the existence of these anharmonic terms give rise to the finite thermal conductivity of solids.

In practice, the tree-phonon scattering plays a predominant role in phonon dy-

namics while other anharmonic effects are expected to be negligible at room temperature [81]. For this reason, this work focuses on modeling the tree-phonon scattering, which will be simply referred to as anharmonic scattering.

The diagrams in Fig. 6.1 show the possible anharmonic interactions between a high energy state with momentum $\boldsymbol{q}$ and two lower energy states with momentum $\boldsymbol{q}'$ and $\boldsymbol{q}''$. In the left diagram, a phonon in state $\boldsymbol{q}$ is annihilated and phonons in state $\boldsymbol{q}'$ and $\boldsymbol{q}''$ are created in a process that here will be referred to as phonon decay. The right diagram represents a process reciprocal to the decay: a phonon in $\boldsymbol{q}'$ and one in $\boldsymbol{q}''$ disappear, while a new phonon is created in the state $\boldsymbol{q}$. This process will be referred to as recombination.

The Hamiltonian matrix element of the anharmonic scattering is written as the product of an anharmonic perturbation coefficient and the creation and annihilation operators [88, 89]:

$$H(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'') = \Xi_{anh}(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')a(\mp\boldsymbol{q})a(\pm\boldsymbol{q}')a(\pm\boldsymbol{q}''), \tag{6.20}$$

where the wave vector $\boldsymbol{q}$ is associated to the higher energy state, the notation $a(+\boldsymbol{q})$ and $a(-\boldsymbol{q})$ represent the creation and annihilation operators, respectively, and Ξ_{anh} represent the strength of the anharmonic interaction. The '$\pm$' and '$\mp$' signs are used to describe both the decay and the recombination process in one equation. In particular, when only the upper signs are considered, the equation describes a decay process, analogously, when only the lower signs are considered the equation describes a recombination processes. Klemens [88] derived an approximated expression for $\Xi^2_{anh}(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')$, which provides an accuracy within the order of magnitude in the case of long-wave acoustic phonons scattering:

$$\Xi^2_{anh}(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'') = (\frac{4\gamma^2}{3N_c})(\frac{M^2}{v^2})(\omega\omega'\omega'')^2, \tag{6.21}$$

where γ is the Grüneisen parameter [81], and N_c, M, and v represent the number

of elementary cell, the average atomic mass, and the sound speed in the crystal, respectively. Ridley, Gupta [90], and Ferry [91] obtained expressions for the optical phonon scattering rate that can be rewritten in a form similar to Eq. (6.21), where the Grüneisen parameter is replaced by an optical counterpart G. For the above stated reasons, this work employs Eq. (6.21), while adopting a constant generalized anharmonicity parameter G_l for each phonon mode l. The G_l parameters are chosen to fit the thermal conductivity at room temperature, and to ensure that in an isolated system at steady-state the correct equilibrium distribution is reached. It should be noted that the scattering algorithm implemented in this work has no requirement regarding the method used to compute the matrix element. This semi-empirical model has been chosen for its simplicity but more complex empirical or *ab-initio* approaches can be used.

As customary, the Fermi's Golden rule [42] is used to calculate the total transition rate:

$$\Gamma^{d,r}(\boldsymbol{q},\boldsymbol{q}',\boldsymbol{q}'') = \frac{2\hbar}{M^3}\frac{\Xi^2(\boldsymbol{q},\boldsymbol{q}',\boldsymbol{q}'')}{\omega\omega'\omega''}\delta(\omega-\omega'-\omega'')F^{d,r}(n,n',n'')\delta(\mathbf{q}-\boldsymbol{q}'-\boldsymbol{q}''+i\boldsymbol{G}), \quad (6.22)$$

where t is time; n, n', and n'' represent the occupation numbers of the states $\boldsymbol{q}$, $\boldsymbol{q}'$, and $\boldsymbol{q}''$, respectively; i is an integer, $\boldsymbol{G}$ is the reciprocal lattice, and the delta function, $\delta(\mathbf{q}-\boldsymbol{q}'-\boldsymbol{q}''+i\boldsymbol{G})$, enforces the momentum conservation within the Brillouin zone periodicity (Umklapp processes are allowed). Finally, $F(n,n',n'')$ is a population-dependent weighting factor due to the annihilation and creation operators and can have two forms:

$$F^d(n,n',n'') = n(n'+1)(n''+1) \quad (6.23)$$

$$F^r(n,n',n'') = (n+1)n'n'', \quad (6.24)$$

where F^d and F^r represent the case of decay and recombination respectively. The

latter descriptions are used to compute the CMC look-up tables, the corresponding rejection probabilities are presented in the next section.

6.5 Anharmonic Scattering Rate

The scattering rate R used in the CMC algorithm represents the probability of a specific particle transitioning between two eigenstates. In the case of anharmonic scattering, the conservation of energy and momentum allows describing the three-particle processes in terms of two-particle interactions. In other words, during the simulation the third state is recovered from the other two as a consequence of energy and momentum conservation. In particular, the recombination scattering rate R^r can be expressed, and tabulated as a function of only two states, as in $R^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}-\boldsymbol{q}') \equiv R^r(\boldsymbol{q}' \to \boldsymbol{q})$ and $R^r(\boldsymbol{q}, \boldsymbol{q}'', \boldsymbol{q}-\boldsymbol{q}'') \equiv R^r(\boldsymbol{q}'' \to \boldsymbol{q})$. Both rates are a representation of the same process and are simply related to the rate in Eq. 6.22 by:

$$\Gamma^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'') = n' R^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}-\boldsymbol{q}') + n'' R^r(\boldsymbol{q}, \boldsymbol{q}'', \boldsymbol{q}-\boldsymbol{q}'') \tag{6.25}$$

Consequently, the values stored in the CMC look-up tables are computed as:

$$R^r(\boldsymbol{q}' \to \boldsymbol{q}) \equiv R^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}-\boldsymbol{q}') = \quad \alpha \ \frac{\Gamma^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')}{n'} \tag{6.26}$$

$$R^r(\boldsymbol{q}'' \to \boldsymbol{q}) \equiv R^r(\boldsymbol{q}, \boldsymbol{q}'', \boldsymbol{q}-\boldsymbol{q}'') = \quad (1-\alpha)\frac{\Gamma^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')}{n''}, \tag{6.27}$$

where α is a number between 0 and 1.

The rejection probabilities for the two representations above can be written as:

$$P^r_{rej}(\boldsymbol{q}' \to \boldsymbol{q}) = \quad \frac{R^r_{loc}(\boldsymbol{q}' \to \boldsymbol{q})}{R^r_{max}(\boldsymbol{q}', \boldsymbol{q})} = \frac{(n_{loc}+1)n''_{loc}}{(n_{max}+1)n''_{max}} \tag{6.28}$$

$$P^r_{rej}(\boldsymbol{q}'' \to \boldsymbol{q}) = \quad \frac{R^r_{loc}(\boldsymbol{q}'' \to \boldsymbol{q})}{R^r_{max}(\boldsymbol{q}'', \boldsymbol{q})} = \frac{(n_{loc}+1)n'_{loc}}{(n_{max}+1)n'_{max}}, \tag{6.29}$$

where the subscript *loc* indicates the values evaluated at a specific position at the time of the scattering event, and the subscript *max* indicates the values stored in

the look-up table; finally, n, n', and n'' are the occupation number of the state $\boldsymbol{q}$, $\boldsymbol{q}'$, and $\boldsymbol{q}''$, respectively. During the simulation, a simulated particle with state $\boldsymbol{q}'$ at position $\boldsymbol{r}$ can start the scattering event, corresponding to $\Gamma^r(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q} - \boldsymbol{q}')$, with probability $R^r_{loc}(\boldsymbol{q}' \to \boldsymbol{q})$. The state $\boldsymbol{q}''$ is then recovered from energy and momentum conservation before proceeding with the scattering algorithm presented in Chapter 3. This algorithm will require the computation of n''; during this step, a candidate particle in $\boldsymbol{q}''$ is chosen. If the scattering is not statistically rejected based on P^r_{rej}, the simulated particle in $\boldsymbol{q}'$, which started the scattering event, and the candidate particle in $\boldsymbol{q}''$ are removed from the simulation, and a new particle is generated with momentum $\boldsymbol{q}$ at position $\boldsymbol{r}$. In an analogous way, this same scattering event can be started by a phonon in $\boldsymbol{q}''$ with probability $R^r_{loc}(\boldsymbol{q}'' \to \boldsymbol{q})$. In this instance, $\boldsymbol{q}'$ is recovered before proceeding with the scattering algorithm. While this two cases seem algorithmically different, they represent the same physical event.

A similar approach allows the calculation of the decay rate R^d to be stored in the two-states look-up table:

$$R^d(\boldsymbol{q} \to \boldsymbol{q}') \equiv R^d(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q} - \boldsymbol{q}') = \quad \beta \ \frac{\Gamma^d(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')}{n} \tag{6.30}$$

$$R^d(\boldsymbol{q} \to \boldsymbol{q}'') \equiv R^d(\boldsymbol{q}, \boldsymbol{q}'', \boldsymbol{q} - \boldsymbol{q}'') = \quad (1 - \beta)\frac{\Gamma^d(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')}{n}, \tag{6.31}$$

where β is a number between 0 and 1. In both descriptions the rejection probability is:

$$P^d_{rej}(\boldsymbol{q} \to \boldsymbol{q}') = P^d_{rej}(\boldsymbol{q} \to \boldsymbol{q}'') = \frac{(n'_{loc} + 1)(n''_{loc} + 1)}{(n'_{max} + 1)(n''_{max} + 1)} \tag{6.32}$$

The scattering algorithm is similar to the recombination case: a simulated particle with state $\boldsymbol{q}$ at position $\boldsymbol{r}$ can either start the scattering event corresponding to $\Gamma^d(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}'')$, with probability $R^d_{loc}(\boldsymbol{q} \to \boldsymbol{q}')$ or with probability $R^d_{loc}(\boldsymbol{q} \to \boldsymbol{q}'')$. Also here, both rates represent the same physical event. The state of the third phonon

is recovered during run-time before the rejection algorithm from energy and momentum conservation. If the scattering is not statistically rejected based on P_{rej}^{d}, the phonon that initiated the process is removed from the simulation. Then, new energy-conserving particles are generated in $\boldsymbol{q}'$ and $\boldsymbol{q}''$.

Finally, the values of α and β may be chosen arbitrary between 0 and 1, and we set them to 0.5 in order to simplify the rate calculation. It should be noted that, contrary to other approaches, this algorithm ensures momentum and energy conservation for each individual scattering process.

6.6 Isotope Scattering

An isotope is an atomic variant of an element with the nominal number of protons, but a different number of neutrons and hence a different atomic mass. The isotopes of an element show identical chemical behavior (except in the reaction speed) as well as both crystal and electronic structure. However, the different mass of isotopes disrupts the periodicity of the crystal enough to result in a large phonon scattering rate. The overall effect of this is seen in a large reduction of the thermal conductivity tensor (σ). The thermal conductivity of the natural crystal (σ^{nat}) and the isotopically pure crystal (σ^{iso}) can differ by up to one order of magnitude at low temperatures [92]. At room temperature, the effect is more ambiguous, and even for a material studied as extensively as silicon the reported σ^{iso} range is between 110% [92, 93] and 160% [94, 95] of the σ^{nat} value.or these reasons isotope scattering cannot be overlooked in the simulation of real semiconductor across different range of temperatures.

The matrix element for isotope scattering can be written as [81, 96]:

$$\Xi_{iso}(\boldsymbol{q},\boldsymbol{q}') = \frac{\hbar}{4\rho N_c}\sqrt{(}\omega,\omega')(\boldsymbol{\epsilon}^{*}\cdot\boldsymbol{\epsilon}'^{*})M_{\boldsymbol{q},\boldsymbol{q}'}a(-\boldsymbol{q})a(\boldsymbol{q}'), \tag{6.33}$$

where ρ is the material density, N_c is the number of elementary cells in the crystal, $\boldsymbol{\epsilon}$

are the polarization vectors, and $M_{\boldsymbol{q},\boldsymbol{q}'}$ is defined as

$$M_{\boldsymbol{q},\boldsymbol{q}'} = \sum_j (\overline{M} - M_j) exp[i(\mathbf{q} - \boldsymbol{q}') \cdot \mathbf{r}_j], \tag{6.34}$$

where M_j and $\mathbf{r}_j$ are the mass and the position of the $j-th$ atom respectively, and $\overline{M}$ is the average mass. Replacing the creation and annihilation operators by their equivalent matrix representations and using Fermi's Golden rule, the scattering rate can be derived as:

$$\Gamma^{iso}(\boldsymbol{q}, \boldsymbol{q}') = \frac{\pi K^{iso}}{2N_c} \omega\omega' \left(\boldsymbol{\epsilon}^* \cdot \boldsymbol{\epsilon}'^*\right)^2 \delta(\omega - \omega') n(n'+1), \tag{6.35}$$

where ρ is the material density, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}'$ are the polarization vectors of $\boldsymbol{q}$ and $\boldsymbol{q}'$, respectively. Finally, K^{iso} is [81, 96]:

$$K_{iso} = \sum_i f_i \frac{\left|M_i - \bar{M}\right|}{\bar{M}}, \tag{6.36}$$

where f_i and M_i are the fraction and the atomic mass of the $i-th$ isotopic specie, respectively, and $\bar{M}$ is the average mass. In this case, the scattering rate stored in the look-up table is computed as:

$$R^{iso}(\boldsymbol{q} \to \boldsymbol{q}') = \frac{\Gamma^{iso}(\boldsymbol{q}, \boldsymbol{q}')}{n}, \tag{6.37}$$

where n is the occupation number of the state $\boldsymbol{q}$. The corresponding rejection probability P^{rej}_{iso} is:

$$P^{iso}_{rej} = \frac{R^{iso}_{loc}(\boldsymbol{q}, \boldsymbol{q}')}{R^{iso}_{max}(\boldsymbol{q}, \boldsymbol{q}')} = \frac{(n'_{loc}+1)}{(n'_{max}+1)}, \tag{6.38}$$

where n'_{loc} and n'_{max} are the local number of phonons in $\boldsymbol{q}'$ when the scattering occurs, and the maximum value used to compute the rates stored in the look-up table, respectively.

Chapter 7

PARTICLE DYNAMICS

7.1 Initialization

The initialization procedure assumes local thermal equilibrium in each cell. Thus, the number of phonons present in the cell is given by:

$$\eta^{cell} = \frac{V_{cell}}{8\pi^3} \int_{BZ1} d\boldsymbol{q}\, n_0(\boldsymbol{q}, T_{cell}) D(\hbar\omega(\boldsymbol{q}), BZ1), \tag{7.1}$$

where V_{cell} is the cell volume, T_{cell} is the local temperature of the cell, and D is the phonon density of states, n_0 is the Bose-Einstein distribution as given by Eq. 3.9, $\hbar$ is the reduced Plank constant, and $\omega(\boldsymbol{q})$ is the frequency of the phonon in state $\boldsymbol{q}$.

Since the number of phonons and atoms in the device are comparable, the explicit simulation of every particle is usually impossible. For this reason, the actual number of phonons is divided by the number of (desired) simulated phonons N^* to obtain the weight factor W^p, which is assigned to each simulated particle.

Finally, the simulated particles are positioned initialized in each cell accordingly to a uniform random distribution, and their the momentum is chosen according to the Bose-Einstein distribution evaluated at the local temperature.

7.2 Contacts

Contacts are modeled as ideal reservoirs [97] of phonons with infinite thermal capacity and constant temperature. In other words, the energy density and particle distribution in the reservoir are unaffected by exchanges of particles between the reservoir and the device. Consequently, each scattering affecting the reservoir is also

assumed to thermalize instantaneously. This effect is modeled by forcing phonons in the reservoir to self-scattering any time there is a scattering event. Moreover, each simulated phonon crossing the contact boundary from the device is assumed to thermalize instantaneously, and removed from the simulation domain.

Figure 7.1 shows a schematic representation of the contacts. The dashed line represents the boundary between the reservoir contact,and the device. The reservoir acts as a source of thermalized carriers, which are injected into the device simply by crossing the surface boundary. The contact population itself is replenished by injecting new particles from the interface boundary (dashed line) into the reservoir itself. This approach results in the appropriate isotropic distribution in both position and momentum space within the reservoir. The injection algorithm itself is a velocity-weighted surface injection method [98]. This approach injects at the start of each time step, Δt, a number of simulated particles equal to:

$$N_{inj} = \frac{\sum_{\boldsymbol{q} \subset BZ1} n_0(\boldsymbol{q}, T) v_\perp(\boldsymbol{q}) * dt}{W^p} \tag{7.2}$$

where n_0 is the Bose-Einstein distribution as given in Eq. 3.9, and $v_\perp$ is the velocity component perpendicular to the surface. Each particle is randomly placed on the boundary surface and the momentum $\boldsymbol{q}'$ is chosen according to the probability distribution:

$$P_{inj}(\boldsymbol{q}') = \frac{n_0(\boldsymbol{q}', T) v_\perp(\boldsymbol{q}')}{\sum_{\boldsymbol{q} \subset BZ1} n_0(\boldsymbol{q}, T) v_\perp(\boldsymbol{q})}. \tag{7.3}$$

After the particle is initialized, a stochastic process selects a remaining free-flight time for the created particle. This final step emulates the injection of carriers at random times during the time step in order to prevent spatial clustering and spurious correlations between carriers.

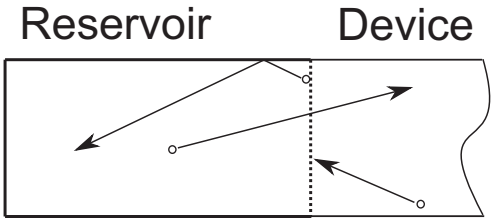

Figure 7.1: Schematic diagram of the reservoir type contact (dashed box). From the boundary between the device and the contact (vertical line at the center of the diagram) particles are injected in the reservoir by employing a velocity-weighted surface injection method [98]. On the right side a particle crossing into the reservoir from the device side is removed from the simulation domain, while particles are free to cross into the device from the reservoir.

7.3 Reflective Boundaries Conditions

The reflective surfaces are modeled [99, 100] by assuming that incident particles reflect either specularly or diffusively. Therefore, surfaces are defined by the fraction P_F of specularly reflected particles. In this approach, P_F is used in a stochastic process to discriminate between specular and diffusive reflection.

In the specular case, the incident and reflected angle are the same, therefore the initial and the final trajectories have maximum correlation. In the diffusive case, the reflected angle obeys the Lambert Cosine Law [101]. In this case, the particle loses memory of its previous state, hence the initial and the final trajectories are uncorrelated. Figure 7.2 offers a schematic representation of Lambert Cosine Law by comparing the incident and specularly reflected trajectories (dashed lines) with diffuse trajectory (solid line). The length of the diffuse trajectory vector is proportional to the reflection probability at that angle. In particular, the length of the vector is proportional to the cosine of the angle between the vector and the surface normal, hence the name cosine Law.

We implemented the Lambert Cosine Law by using the following probability dis-

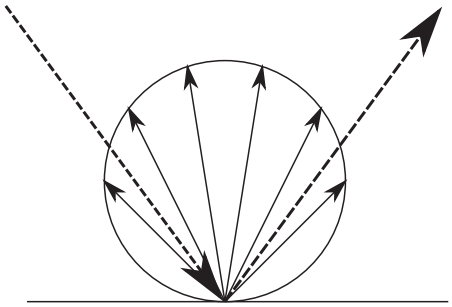

Figure 7.2: Comparison between the reflected trajectories in the case of specular reflection (dashed line) and that of diffusive reflection (solid line). In the case of diffusive reflection, the length of the vector is proportional to the reflection probability for that angle.

tribution for the final, reflected, wavevector $\boldsymbol{q}'$:

$$P_{dif}(\boldsymbol{q}') = \frac{v_{\perp}(\boldsymbol{q}')\delta(\omega(\boldsymbol{q}') - \omega)}{\sum_{\boldsymbol{q}\subset BZ1} v_{\perp}(\boldsymbol{q})\delta(\omega(\boldsymbol{q}) - \omega)}, \tag{7.4}$$

where $v_{\perp}$ is the velocity component perpendicular to the surface, ω is the particle frequency, and the term $\delta(\omega(\boldsymbol{q}') - \omega)$ ensures the conservation of the particle energy.

Chapter 8

DATA STRUCTURE

This chapter discusses the implementation of the rejection technique in the CMC framework from the point of view of data organization, and shows results of simulation in momentum space. The first two sections examine the structure of the look-up table employed in this work to store the scattering rate. Section 8.3 offers a validation of the look-up table approach presented in the two previous sections. Section 8.4 and 8.5 discuss the representation of the dispersion relation in momentum space and its impact on the scattering algorithm.

8.1 Scattering Table

The original CMC approach [18] tried to obtain for the most efficient use of computer memory and the consequent minimization of the transition table size. This was achieved by employing a lossy compression algorithm that discards non-essential information. However, some of that information is needed by the new scattering algorithm based on the rejection technique. In particular, information regarding the exact type of scattering process involved in the transition (e.g. the mode of the phonon involved, the temperature) is crucial for the new algorithm.

For this reason, a new three-level look-up table has been designed. This new structure allows storing all the information on the nature of each scattering event while keeping the performance advantages of the CMC approach over traditional EMC method. Figure 8.1 shows a graphical representation of the three level structure employed in the new look-up table. The left-most long vertical array represents the first level of the structure; this array represents all the possible particle states

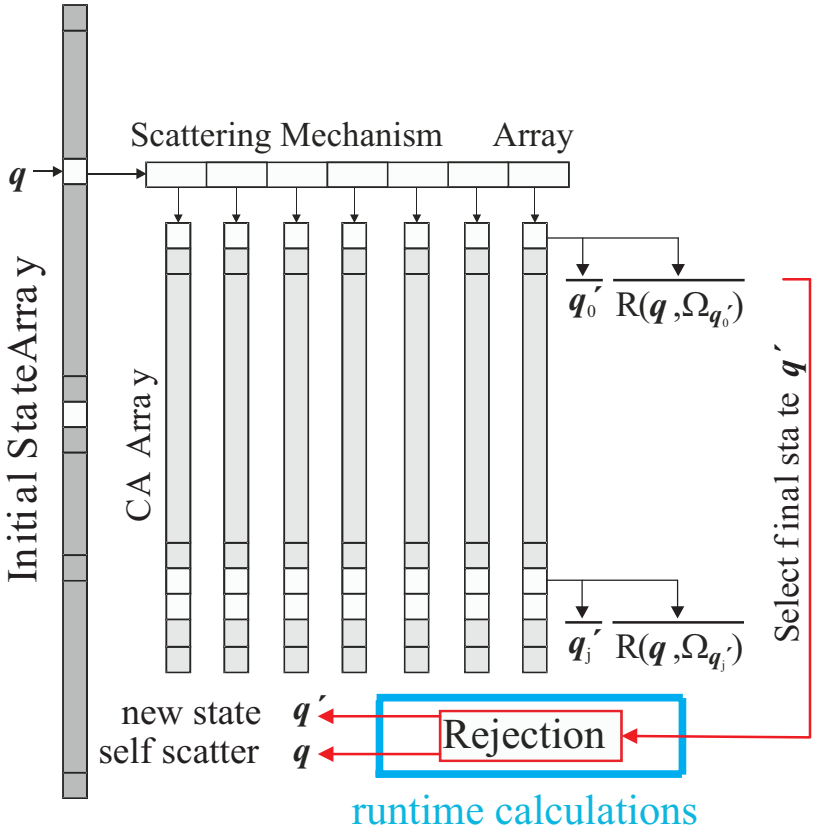

Figure 8.1: Graphical representation of the three-level data structure designed to be employed with the rejection technique.

$\boldsymbol{q}$ in momentum space and it will be referred to as the *initial state array*. Each element of the initial state array is linked to a *scattering mechanism array*, depicted horizontally. The scattering mechanism array represents all the possible scattering mechanism for the state $\boldsymbol{q}$. Indeed, the elements of the scattering mechanism array contain information used in a typical Ensemble Monte-Carlo code [34], such as the mechanism identification and the total probability of scattering out of $\boldsymbol{q}$. Finally, each element of the scattering mechanism array is linked to a *CA array*, depicted by the short vertical bars. The CA array represents all the possible final states, $\boldsymbol{q}'$ for a scattering particle in state $\boldsymbol{q}$. In particular, each element of the CA array contains the information on the final states $\boldsymbol{q}'$ and the corresponding probability (for that mechanism) to scatter from $\boldsymbol{q}$ to a small region (cell) $\Omega_{\boldsymbol{q}'}$ centered around the

final state $\boldsymbol{q}'$. This CA array constitutes the main difference from a traditional EMC table. When a particle scatters, an hash function efficiently associates the initial particle state $\boldsymbol{q}$ to one element of the initial state array. This element is linked to the scattering mechanism array that represents all the possible scattering mechanism for a particle in $\boldsymbol{q}$. The scattering algorithm for the new transition table is a four-step process:

1. A hash function efficiently associates the initial particle state to one element of the initial state array
2. A scattering mechanism is stochastically chosen from the corresponding scattering mechanism array
3. A final state is stochastically chosen from the corresponding CA array
4. The rejection algorithm establishes if the scatter occurs or is rejected

In case of three-particle scattering, like for phonons, the third state is retrieved after step 4 during run-time as consequence of the momentum and energy conservation.

Finally, we note that the basic CMC method could not provide accurate information on the energy of the particle after the scattering. Therefore, particles scattering from the initial state $\boldsymbol{q}$ into the region $\Omega_{\boldsymbol{q}'}$ were simply placed by that algorithm in the geometrical center of $\Omega_{\boldsymbol{q}'}$, which causes an error in the instantaneous conservation of energy and momentum. This error was mitigated by reducing the volume of the cell $\Omega_{\boldsymbol{q}'}$ at the expenses of larger look-up tables. The original CMC method obtained a balance between accuracy and computer memory requirement by employing a non-uniform discretization scheme of the BZ1 [18] with smaller cells in those areas that are more important for the transport phenomena.

However, the information about the scatter mechanism contained in the new table structure allows for a high accuracy in the determination of the final state energy. In particular, the increased accuracy is made possible by the detailed knowledge of the nature of the scattering mechanism, and its characteristic energy exchange. This means that the final state of a particles scattering from the $\boldsymbol{q}$ into the region $\Omega_{\boldsymbol{q}'}$ can be chosen by first inverting the dispersion relation near $\boldsymbol{q}'$ and then selecting a final momentum with the desired target energy.

8.2 Table Compression

As depicted in Figure 8.1, the size of the table is definitely large. Indeed, the number of elements in the look-up table is on the order of the square of the number of cells in the first Brillouin Zone (BZ1) times the number of bands, and a naive approach could results in look-up tables requiring hundreds of Gigabytes of RAM.

In this work, the storage issue is addressed by employing a compression algorithm which exploit physical, geometrical, and numerical properties of the model. First, the symmetry of the crystal is fully exploited so that the table is computed only for initial states located in the irreducible wedge of the BZ1 [82, 102]. From an algorithmic point of view, the initial state of the particle is rotated into the irreducible wedge prior to the scattering, and then the (stored) inverse rotation matrix is applied to the final state as selected with the CMC algorithm.

A data compression algorithm allows to efficiently store both the scattering rate and the information on the location of the final state cell $\Omega_{\boldsymbol{q}'}$ as one 4-Byte integer number. The most significant bits of the integer number will encode the scattering rate in the form of normalized partial sums, so that the whole number can be used in the stochastic selection process with no need of separating the compressed data in its two components. The least significant digits encode the information on the position

of the final state $\Omega_{\boldsymbol{q}'}$. This encoding is prepared in two steps: first a hash function is used to assign a unique integer value to the state depending on the cell and band, this hash code will be simply referred to as the 'address' of the cell $\Omega_{\boldsymbol{q}'}$. Then, the compression algorithm simply encodes the final cell address in term of relative distance from a reference cell that depends on the initial state of the carrier. Moreover, the address of the reference cell is numerically close to the address of final state cell. This offset-based encoding allows to minimize the number of bits required to represent the relative address, thus maximizing the numerical precision of the scattering rate. In this "compressed" format, the size of the fast memory bank used for each table is usually limited to a size in the order of magnitude of 10 GByte.

8.3 Model Validation and Performances

The look-up table structure introduced and described in section 8.1 has been implemented and validated against both experimental data and previous CMC simulations.

Figure 8.2 shows the drift velocity vs electric field curves for electrons in Si as obtained from momentum-space simulations performed assuming different temperature and different scattering algorithms. The reference simulations (lines) are computed by using the basic, fixed temperature, CMC look-up table and algorithm, therefore, simulations of different temperatures require different scattering tables computed separately. Test simulations (dots) employ the rejection algorithm using temperature as a driving mechanism and a scattering table build with the new structure discussed in section 8.1. In particular, these simulations employ only one look-up table computed by assuming a maximum crystal temperature of 400K. Since there is only one scattering table, the difference in simulation results are due to the locally imposed crystal temperatures and their effect on the scattering. At low values of the electric

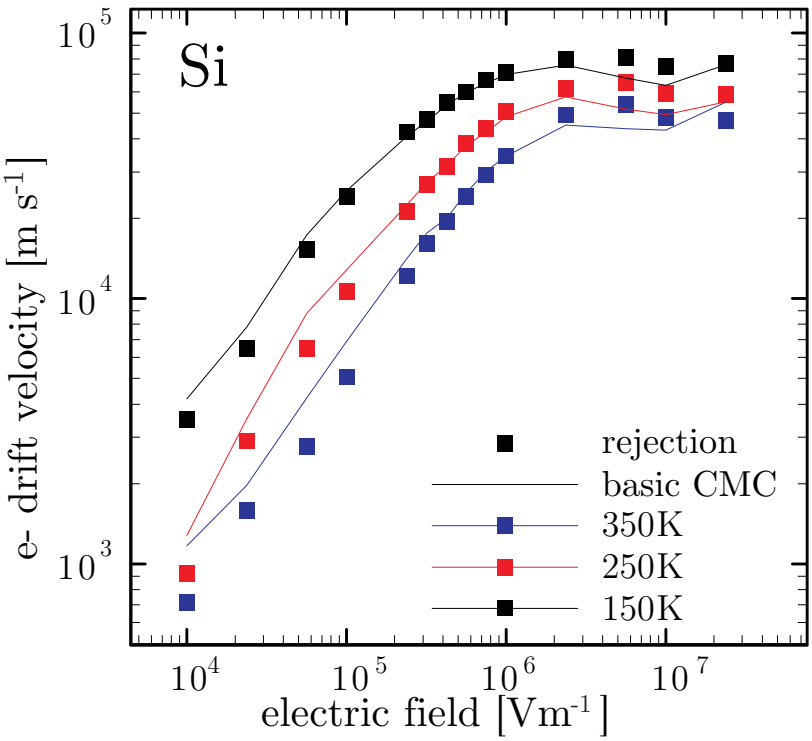

Figure 8.2: Si drift velocity as function of the electric field All the data is computed by using the same 400K table with the rejection technique (dots) compared to equivalent simulations computed with the basic CMC algorithm (lines) by using a different scattering table for each temperature. The two approaches show excellent agreement at low electric fields, the difference at high electric field is due to the effect of the coarse discretization of the BZ1 at high energy. Here the accurate energy conservation produces a flatter velocity saturation in the curves obtained with the rejection.

field, electrons are located close to the bottom of the conduction band. This is the most important region for the conduction properties of the materials and the momentum space discretization (cell size) is particularly fine. In this case, the drift velocity curves obtained with the two algorithm perfectly match.

At higher values of the electric field, the electrons start populating areas of the BZ1 which are usually less important for the conduction properties of the semiconductor. These areas have a coarser discretization in order to limit the table size. In this regime, the more accurate energy conservation obtained with the rejection technique results in a flatter velocity saturation corresponding to the correct physical behavior.

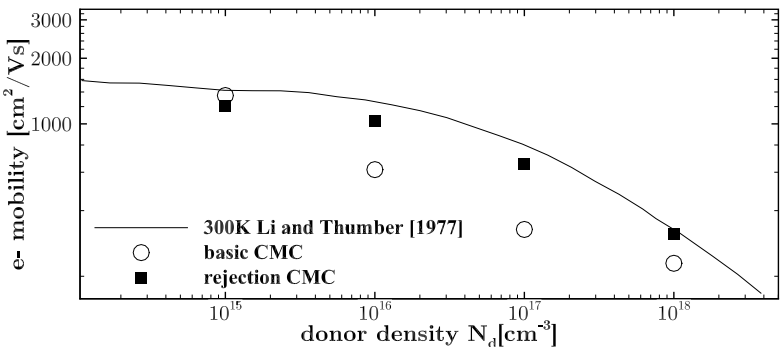

Figure 8.3: Si Mobility Computed with the Rejection Technique

Si mobility computed with the rejection technique (squares) by using the same 400K table and a concentration of 10^{19}[cm^{-3}] of n-doping. The results are compared with similar data computed with the basic CMC algorithm (circles) and experimental data (line) by Li and Thumber. At high concentrations, the energy conservation algorithm and the adaptive run-time screening allow for a more accurate modeling of the experimental mobility.

Figure 8.3 shows a comparison between the experimental data (solid line) for the Silicon mobility at 300K as a function of doping and the simulation results (symbols). The circles represent results computed by using the basic CMC approach with a different table for each doping concentration. These simulations show a good agreement with the experimental data at low doping, but underestimate the mobility at higher doping. The black squares represent results computed by using the rejection algorithm and a 350K table assuming a maximum doping of 10^{18}[cm^{-3}]. For this table, both temperature and doping concentration are employed by the rejection algorithm. These simulations show a very good agreement with the experimental data both at high and low concentration.

Finally, Fig. 8.4 shows more in detail the effects of the final state correction algorithm by comparing the computed average electron distributions at different doping concentrations, with no externally imposed electric field, and assuming a $300K$

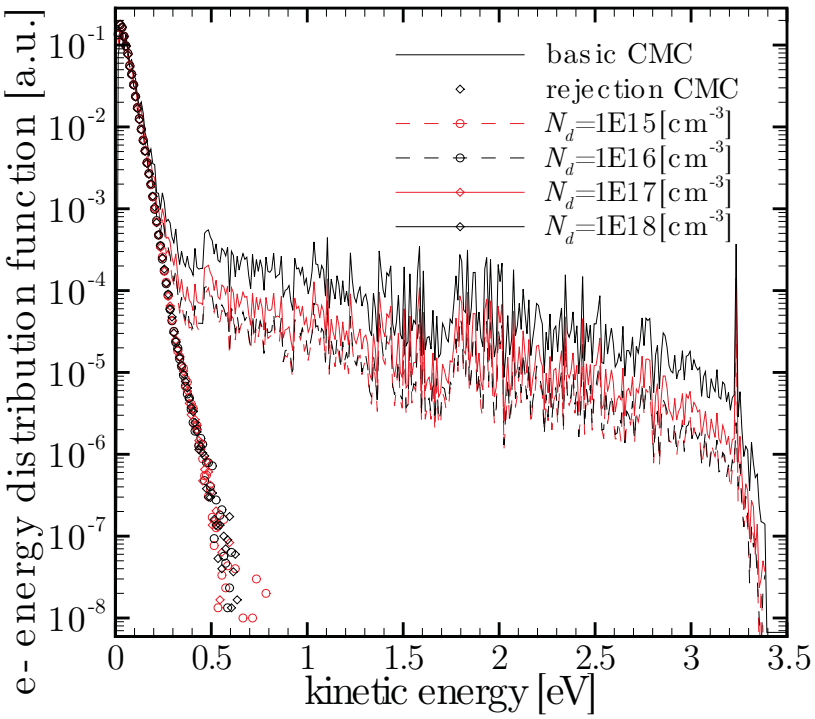

Figure 8.4: Low electric field electron distribution computed with the basic CMC algorithm (lines) compared to the result obtained with the CMC and rejection algorithm by using the same 400K table and a concentration of 10^{19}cm^{-3} of n-doping. At low energy, where the BZ1 has a fine discretization, the two algorithms show similar results. Above a kinetic energy of 0.5 eV, where the BZ1 has coarser discretization, the improved energy conservation allowed by the rejection algorithm provides the correct physical behavior independent on the donor density.

crystal temperature. Near the bottom of the conduction band, the BZ1 has a finer discretization and both the basic CMC (lines) and the CMC with rejection (symbols) simulations show similar results. However, the basic CMC produces spread in the distribution at high energy where the BZ1 discretization is coarser and the error on the energy conservation is more pronounced. The improved energy conservation provided by the rejection algorithm reproduces the expected behavior and compensate for the gridding issues.

8.4 Dispersion Relationship

In order to compromise between accuracy and simplicity [37], the full-band energy dispersion relationship is interpolated over the momentum space with a second-order scheme in the discretized space. In particular, the values for the second order Taylor expansion of the dispersion are pre-computed and stored only for the wavevectors $\boldsymbol{q}_i$ at the center of the *i-th* cell in the discretized momentum space. The dispersion in other position $\boldsymbol{q}$ is obtained by employing a trilinear interpolation scheme [103]. This approach requires first finding the cell containing $\boldsymbol{q}$ and the 7 nearest neighbors cells, then the energy $\mathcal{E}(\boldsymbol{q})$ is computed by summing up the energy contributions from each cell using the appropriate weights [37]:

$$\mathcal{E}(\boldsymbol{q}) = \sum_{l=1}^{8} w_i(\boldsymbol{q}) \cdot \mathcal{E}_i(\boldsymbol{q}), \tag{8.1}$$

where w_i is the weighting factor for the trilinear interpolation:

$$w_i(\boldsymbol{q}) = \left(1 - \frac{|q_x - q_{i,x}|}{L_{i,x}}\right)\left(1 - \frac{|q_y - q_{i,y}|}{L_{i,y}}\right)\left(1 - \frac{|q_z - q_{i,z}|}{L_{i,z}}\right), \tag{8.2}$$

where $L_{i,j}$ is the size of the *i-th* cell, $\Omega_{\boldsymbol{q}_i}$, along the j direction. Each $\mathcal{E}_i(\boldsymbol{q})$ is obtained from the quadratic expansion of the energy around the wavevector $\boldsymbol{q}_i$ at the center of the cell $\Omega_{\boldsymbol{q}_i}$:

$$\mathcal{E}_i(\boldsymbol{q}) \approx \mathcal{E}(\boldsymbol{q}_i) + (\boldsymbol{q} - \boldsymbol{q}_i)^T \cdot \nabla\mathcal{E}\big|_{\boldsymbol{q}_i} + \frac{1}{2}(\boldsymbol{q} - \boldsymbol{q}_i)^T \cdot \boldsymbol{H}(\mathcal{E})\big|_{\boldsymbol{q}_i} \cdot (\boldsymbol{q} - \boldsymbol{q}_i), \tag{8.3}$$

where the superscript T represent the vector transpose operator, and $\nabla\mathcal{E}\big|_{\boldsymbol{q}_i}$ is the gradient of the dispersion calculated at $\boldsymbol{q}_i$. The $\boldsymbol{H}(\mathcal{E})\big|_{\boldsymbol{q}_i}$ term in Eq. 8.3 represents the Hessian matrix of the energy calculated at $\boldsymbol{q}_i$:

$$\boldsymbol{H}(\mathcal{E})\big|_{\boldsymbol{q}_i} = \begin{bmatrix} \frac{\partial^2\mathcal{E}}{\partial q_x^2}\Big|_{\boldsymbol{q}_i} & \frac{\partial^2\mathcal{E}}{\partial q_x \partial q_y}\Big|_{\boldsymbol{q}_i} & \frac{\partial^2\mathcal{E}}{\partial q_x \partial q_z}\Big|_{\boldsymbol{q}_i} \\ \frac{\partial^2\mathcal{E}}{\partial q_x \partial q_y}\Big|_{\boldsymbol{q}_i} & \frac{\partial^2\mathcal{E}}{\partial q_y^2}\Big|_{\boldsymbol{q}_i} & \frac{\partial^2\mathcal{E}}{\partial q_y \partial q_z}\Big|_{\boldsymbol{q}_i} \\ \frac{\partial^2\mathcal{E}}{\partial q_x \partial q_z}\Big|_{\boldsymbol{q}_i} & \frac{\partial^2\mathcal{E}}{\partial q_y \partial q_z}\Big|_{\boldsymbol{q}_i} & \frac{\partial^2\mathcal{E}}{\partial q_z^2}\Big|_{\boldsymbol{q}_i} \end{bmatrix}, \tag{8.4}$$

where q_x, q_y, and q_z represent the directions of the orthogonal basis in the momentum space.

8.5 Energy Conservation Algorithm

When a scattering occurs, a final cell $\Omega_{\boldsymbol{q}_{fin}}$ is selected by the scattering algorithm and a new wavevector must be chosen within that cell. In principle, both energy and momentum must be conserved, however a perfect conservation of both quantities increases the algorithm complexity without offering any concrete advantage. Within the CMC framework, small differences in momentum simply correspond to small variations in the trajectories, however, this phenomenon is negligible compared to the effect of external forces, such as the electric field and scattering, especially at room temperature. Errors in the conservation of energy alter the population distribution. In turn, these differences affect the particle dynamics by altering screening factors, the tunneling probability and the population temperature. For these reasons, after a scattering event, the new particle wavevector is selected by maximizing the conservation of energy, while relaxing the constrains on momentum conservation.

The identification of wavevectors on the iso-energetic surface at the target energy $\mathcal{E}'$ within $\Omega_{\boldsymbol{q}fin}$ requires inverting the function in Eq. 8.1. Since there is no close-form for the inverse of this function, the new wavevector is chosen by employing a numerical algorithm based on non-iterative successive approximation to minimize the energy conservation error. The computations are executed in double precision, while the final value is stored in single precision.

At the start of the algorithm a first candidate wavevector, $\boldsymbol{q}_{c1}$ is chosen by and looking for potential wavevectors along the direction $\hat{\boldsymbol{q}}$ corresponding to the maximum

rate of change:

$$\hat{\boldsymbol{q}} = \frac{\nabla \mathcal{E}|_{\boldsymbol{q}_{fin}}}{\left|\nabla \mathcal{E}|_{\boldsymbol{q}_{fin}}\right|}. \tag{8.5}$$

In particular, the candidate must be a wavevector of the type:

$$\boldsymbol{q}_1(\alpha) = \boldsymbol{q}_{fin} + \alpha \hat{\boldsymbol{q}} \tag{8.6}$$

where $\boldsymbol{q}_{fin}$ is the wavevector at the center of $\Omega_{\boldsymbol{q}_{fin}}$ and α is a real number.

The energy at $\boldsymbol{q}_0(\alpha)$ can be written as a function of α by using Eq. 8.3:

$$\mathcal{E}(\boldsymbol{q}_1(\alpha)) = \mathcal{E}(\boldsymbol{q}_{fin}) + \alpha \hat{\boldsymbol{q}}^T \cdot \nabla \mathcal{E}|_{\boldsymbol{q}_{fin}} + \alpha^2 \hat{\boldsymbol{q}}^T \cdot \boldsymbol{H}(\mathcal{E})|_{\boldsymbol{q}_{fin}} \cdot \hat{\boldsymbol{q}}. \tag{8.7}$$

The previous equation allows writing a second order equation in α by imposing $\mathcal{E}(\boldsymbol{q}_1(\alpha)) - \mathcal{E}' = 0$:

$$\mathcal{E}(\boldsymbol{q}_{fin}) + \alpha \hat{\boldsymbol{q}}^T \cdot \nabla \mathcal{E}|_{\boldsymbol{q}_{fin}} + \alpha^2 \hat{\boldsymbol{q}}^T \cdot \boldsymbol{H}(\mathcal{E})|_{\boldsymbol{q}_{fin}} \cdot \hat{\boldsymbol{q}} - \mathcal{E}' = 0. \tag{8.8}$$

The solution in α of the quadratic formula allows computing the first candidate wavevector by using Eq. 8.6. In particular, we chose the root with the smallest amplitude to ensure that $\boldsymbol{q}_{c1}$ is close to $\boldsymbol{q}_{fin}$.

However, obtaining a reliable numerical solution for Eq.8.8 is not a trivial task. Indeed, the numerical evaluation of the quadratic equation $ax^2 + 2bx + c = 0$ requires accounting for the order of magnitude of its coefficients in order avoid unreliable results and instabilities caused by the limited precision of floating point arithmetics. Since only the root with the smaller amplitude is of interest, the solution of the quadratic equation can be divided in two cases. In the first case $b^2 >> ac$, this allow us to numerically approximate the smallest zero of the equation as [1] :

$$x \approx -\frac{b}{2c}(1 + \frac{ac}{b^2}). \tag{8.9}$$

[1] This result derives directly from the properties of the modulus and from the Taylor expansion $\sqrt{1-x} = 1 - x/2 - x^2/8 + O(x^3)$.

In the opposite case, the smaller root is computed without substantial loss of precision by using the Citardauq formula [103, 104]:

$$x = \frac{c}{-b - \mathrm{sgn}\,(b)\,\sqrt{b^2 - ac}}, \tag{8.10}$$

where sgn is the sign function. After the first candidate $\boldsymbol{q}_{c1}$ is selected, the algorithm can identify the 8 cells/wavevectors employed in the interpolation scheme of Eq. 8.1. In the next step of the algorithm, a search is made for a second candidate wavevector $\boldsymbol{q}_{c2}$ in close proximity of $\boldsymbol{q}_{c1}$:

$$\boldsymbol{q}_2(\alpha) = \boldsymbol{q}_1 + \alpha \cdot \hat{\boldsymbol{q}}. \tag{8.11}$$

In this case, $\boldsymbol{q}_{c2}$ is chosen by evaluating the energy at $\boldsymbol{q}_2(\alpha)$ resulting from the interpolation scheme by imposing $\mathcal{E}(\alpha) = \mathcal{E}(\boldsymbol{q}_1 + \alpha\hat{\boldsymbol{q}})$, which results in:

$$\begin{aligned}\mathcal{E}(\alpha) = \sum_{i=1}^{8} &\left(1 - \frac{\left|\Delta\boldsymbol{q}_{i,x} + \alpha \cdot \hat{\boldsymbol{q}}_x\right|}{L_{i,x}}\right)\left(1 - \frac{\left|\Delta\boldsymbol{q}_{i,y} + \alpha \cdot \hat{\boldsymbol{q}}_y\right|}{L_{i,y}}\right)\left(1 - \frac{\left|\Delta\boldsymbol{q}_{i,z} + \alpha \cdot \hat{\boldsymbol{q}}_z\right|}{L_{i,z}}\right) \\ &\left(\mathcal{E}(\boldsymbol{q}_i) + (\alpha\hat{\boldsymbol{q}} + \Delta\boldsymbol{q}_i)^T \cdot \nabla\mathcal{E}\big|_{\boldsymbol{q}_i} + \frac{1}{2}(\alpha\hat{\boldsymbol{q}} + \Delta\boldsymbol{q}_i)^T \cdot \boldsymbol{H}(\mathcal{E})\big|_{\boldsymbol{q}_i} \cdot (\alpha\hat{\boldsymbol{q}} + \Delta\boldsymbol{q}_i)\right),\end{aligned} \tag{8.12}$$

where $\Delta\boldsymbol{q}_i = \boldsymbol{q}_{c1} - \boldsymbol{q}_i$, and $\boldsymbol{H}(\mathcal{E})\big|_{\boldsymbol{q}_i}$ is the Hessian matrix as defined in Eq. 8.4.

In this case, the resulting polynomial is a fifth order equation in α:

$$\sum_{i=0}^{5} A_i \alpha^i, \tag{8.13}$$

where the coefficients A_i are given by the following expressions:

$$A_0 = \textstyle\sum_{i=1}^{8} \quad w_i(\boldsymbol{q}_c) \cdot c_i \tag{8.14}$$

$$A_1 = \textstyle\sum_{i=1}^{8} \quad w_i(\boldsymbol{q}_c) \cdot b_i + w_i^1 \cdot c_i \tag{8.15}$$

$$A_2 = \textstyle\sum_{i=1}^{8} \quad w_i(\boldsymbol{q}_c) \cdot a_i + w_i^1 \cdot b_i + w_i^2 \cdot c_i \tag{8.16}$$

$$A_3 = \textstyle\sum_{i=1}^{8} \quad w_i^1 \cdot a_i + w_i^2 \cdot b_i + w_i^3 \cdot c_i \tag{8.17}$$

$$A_4 = \textstyle\sum_{i=1}^{8} \quad w_i^2 \cdot a_i + w_i^3 \cdot b_i \tag{8.18}$$

$$A_4 = \textstyle\sum_{i=1}^{8} \quad w_i^3 \cdot a_i \tag{8.19}$$

where w_i are the weight as defined in Eq. 8.2, the factors a_i, b_i and c_i are obtained from the expansion of the energy dispersion:

$$a_i = \frac{1}{2}\hat{\boldsymbol{q}}^T \cdot \boldsymbol{H}(\mathcal{E})|_{\boldsymbol{q}_i} \cdot \hat{\boldsymbol{q}} \tag{8.20}$$

$$b_i = \hat{\boldsymbol{q}}^T \cdot \nabla\mathcal{E}|_{\boldsymbol{q}_i} + \hat{\boldsymbol{q}}^T \cdot \boldsymbol{H}(\mathcal{E})|_{\boldsymbol{q}_i} \cdot \Delta\boldsymbol{q}_i \tag{8.21}$$

$$c_i = \mathcal{E}_i(\boldsymbol{q}_c), \tag{8.22}$$

and the factors w_i^1 are given by:

$$w_i^1 = l_{i,x} \cdot w_{i,y} \cdot w_{i,z} + w_{i,x} \cdot l_{i,y} \cdot w_{i,z} + w_{i,x} \cdot w_{i,y} \cdot l_{i,z}, \tag{8.23}$$

where the values of $w_{i,j}$ and $l_{i,j}$ derive from the interpolation scheme in Eq.8.1 under the assumption that $\alpha\hat{q_j} < \Delta q_{i,j}$, ($j$ represents a direction, while i is the index of the neighbor cell):

$$1 - \frac{|q_j + \alpha\hat{\boldsymbol{q}} - q_{i,j}|}{L_{i,j}} = 1 - \frac{|q_j - q_{i,j}|}{L_{i,j}} + \operatorname{sgn}(\Delta q_{i,j})\frac{\alpha\,|\hat{q_{i,j}}|}{L_{i,j}} = w_{i,j} + \alpha l_{i,j}. \tag{8.24}$$

Finally, the values for w_i^2 and w_i^3 are:

$$w_i^2 = w_{i,x} \cdot l_{i,y} \cdot l_{i,z} + l_{i,x} \cdot w_{i,y} \cdot l_{i,z} + l_{i,x} \cdot l_{i,y} \cdot w_{i,z} \tag{8.25}$$

$$w_i^3 = l_{i,x} \cdot l_{i,y} \cdot l_{i,z} \tag{8.26}$$

$$\tag{8.27}$$

In general, fifth order polynomial equations have no algebraic solution [2] . Because of that the balance between accuracy and efficiency is achieved by employing a two-step algorithm.

The first step is to select a candidate value α_c based on the quadratic truncation of the polynomial:

$$A_2 \cdot \alpha^2 + A_1 \cdot \alpha + A_0 - \mathcal{E}' = 0. \tag{8.28}$$

[2]The Abel-Ruffini impossibility theorem [105] states that a polynomial equation of degree five or higher with arbitrary coefficients has no closed-form algebraic expression, i.e. the roots cannot be expressed in terms of its coefficients by using only addition, subtraction, multiplication, division, and exponentiation.

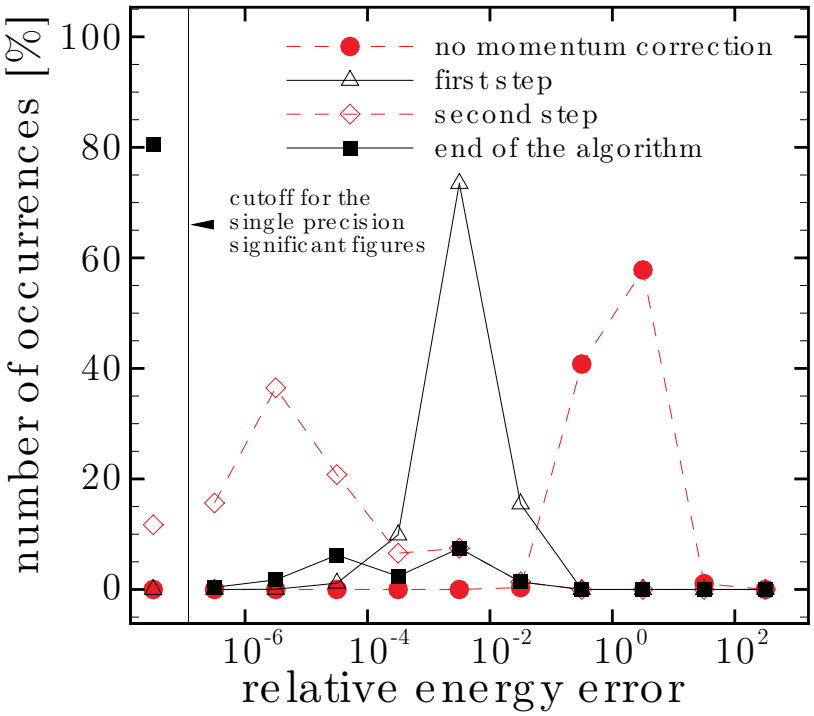

Figure 8.5: Error relative to the target energy at different points of the algorithm; the values are obtained from a low electric field bulk simulation. The vertical line is placed at single precision machine epsilon. The expected error is reduced at each step of the algorithm, and in the majority of cases the error at the end of the algorithm is smaller than the machine precision.

Eq 8.13 is then expanded around the candidate value α_c:

$$(\alpha-\alpha_c)^2\sum_{i=0}^{4}\left(\frac{\alpha_c^i A_{i+1}}{i+1}\right)+(\alpha-\alpha_c)\left(\sum_{i=0}^{3}\frac{\alpha_c^i A_{i+2}}{(i+1)(i+2)}\right)+\mathcal{E}(q(\alpha_c))-\mathcal{E}', \quad (8.29)$$

The smallest root of this equation α' is used to compute the final wavevector which maximize energy conservation:

$$\boldsymbol{q}' = \boldsymbol{q}_c + \alpha' \cdot \hat{\boldsymbol{q}}. \quad (8.30)$$

In order to evaluated the performance of the algorithm, we run low electric field simulations in momentum space with different selection schemes of the final energy of the simulated particles, and record the energy of the simulated carrier after the scattering, E_{fin} in terms of relative distance from the expected target energy, $\mathcal{E}'$. In

other words, with each scatter the recoded value is $E_{fin}/\mathcal{E}' - 1$. Figure 8.5 shows the histogram of the analyzed data, the horizontal axis represent the possible value of $E_{fin}/\mathcal{E}' - 1$ in logarithmic scale, the vertical axis represents the number of occurrences as a percentage. The vertical line is placed at single precision machine epsilon according to the ISO C standard, this value represents the relative distance between two consecutive numbers. In other words, this value represents the maximum precision achievable with the single precision arithmetic. The dashed line with dots represents the results of the basic CMC algorithm where the particles are placed at the center of Ω_{fin}. The solid line with the triangles represents the error produced by stopping the algorithm after the first step of the algorithm when a candidate wavevector $\boldsymbol{q}_{c1}$ is selected. In this approach the effect of the interpolation is not accounted for. The diamonds and the square represent the case where the interpolation is accounted for, and the dispersion relationship is represented by the fifth order polynomial in Eq. 8.13. In particular, the dashed line with diamonds represent the error produced by stopping at the second step of the algorithm after finding the value α_c is from Eq 8.28. Finally, the solid line with squares represents the relative error at the end of the algorithm.

The figure shows that in 80% of the cases the error at the end of the algorithm is smaller than the single precision machine epsilon. That is, the error is smaller than a difference between two consecutive numbers in single precision.

Chapter 9

SIMULATIONS

This chapter presents a series of simulations designed to test and validate the scattering algorithm [106] implemented in this work. In section 9.1 the capability of the technique to reach thermal equilibrium is investigated. Section 9.2 compares the simulated properties of Si at room temperature with available literature. Section 9.3 and 9.4 investigate the effects of isotope and boundary scattering, respectively. Section 9.5 offers a test of the scattering algorithm in transient simulations. Finally, section 9.6, 9.7 and 9.8 present simulations regarding the electron-phonon coupling.

9.1 Isolated System

In an isolated system at thermal equilibrium, the temperature T_c and the total energy E_c are related by:

$$E_c(T_c) = \sum_{\mathbf{q} \subset BZ} \left(n_0(\mathbf{q}, T_c) + 1/2\right) \hbar\omega, \tag{9.1}$$

where $n_0(\boldsymbol{q}, T_c)$ is the Bose-Einstein distribution as in Eq. 3.9. Equation 9.1 defines a bijective function between E_c and T_c. Consequently, phonons in an isolated system with energy $E_c(T_c)$ will reach the thermal equilibrium distribution given by $n_0(\boldsymbol{q}, T_c)$.

This behavior is tested for the purpose of validation by simulating an isolated cube of Si initialized with a non-equilibrium phonon distribution. In particular, the initial particle distribution is generated with the momentum and energy based on a 350K Bose-Einstein distribution, while limiting the number of particles in the system in order to achieve a total energy in the system corresponding to the 216K equilibrium (full) distribution. The phonon dynamics simulation kernel is then activated until

the system reaches a steady state solution. Figure 9.1 shows the results of this simulation: the analytical 216K equilibrium distribution (solid line), and the initial distribution (dots) obviously differ. However, after a transient of 400ps, the self-consistent solution (dashed line) is in excellent agreement with the expected 216K equilibrium distribution. The insert shows the time evolution of the deviation between the reference and the simulated distribution as measured by the Jensen-Shannon divergence [107]. The dot represents the difference between the reference distribution and the self-consistent distribution.

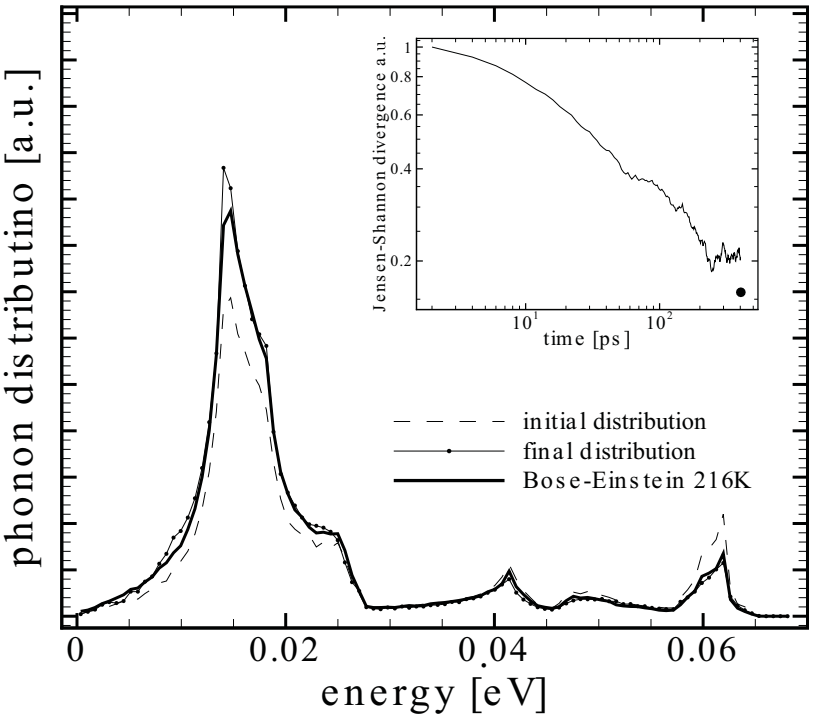

Figure 9.1: Phonon simulation of an isolated System. The CMC algorithm initialized with an out of equilibrium distribution (dots) reaches a steady state solution (dashed) comparable to a 216K equilibrium distribution (solid). The insert shows the time evolution of the Jensen-Shannon divergence [107] between the reference and the simulated distribution; the dot represents the divergence between the reference and the final distribution.

9.2 Bulk Simulation at Room Temperature

An isolated system with specular boundaries initialized at 300K is simulated for 35 ns by using a scattering table computed for a maximum temperature T_{max} = 400K. The resulting mean free path and relaxation time spectra are compared to data available in literature. Figure 9.2 shows the simulated mean free path compared with previous Monte Carlo [108] and molecular dynamics simulations [109]. All results have the same order of magnitude and show similar behavior. However, compared to the Monte Carlo simulations reported in ref. [108], the present work shows the small variations associated with the full-band dispersion. In particular, the energy corresponding to the peak density of states can be identified by the dip in the optical phonon mean free path.

Figure 9.3 shows the relaxation time spectra of longitudinal acoustic phonons compared with a first-principle study [110] and molecular dynamics [109] simulations. All results show the same order of magnitude and shows similar behavior, in particular this work and the molecular dynamics solutions are especially close. However, the comparison with the first-principle study suggests that there is an overestimation in the relaxation time of higher energy phonons which are close to the BZ1 boundaries.

9.3 Thermal Conductivity

The thermal conductivity for a Si cuboid of dimensions $2.00 \times 3.12 \times 20.5$ mm^3 is computed between 4K and 300K for both homogeneous isotopically enriched silicon, Si^{28} (14 protons and 14 neutrons), and silicon with naturally occurring isotopic composition, Si^{nat} (92.2% Si^{28}, 4.6% Si^{29}, 3.1% Si^{30}) and compared to the values measured by Inyushkin *et al.* [92]. The elements of the thermal conductivity tensor

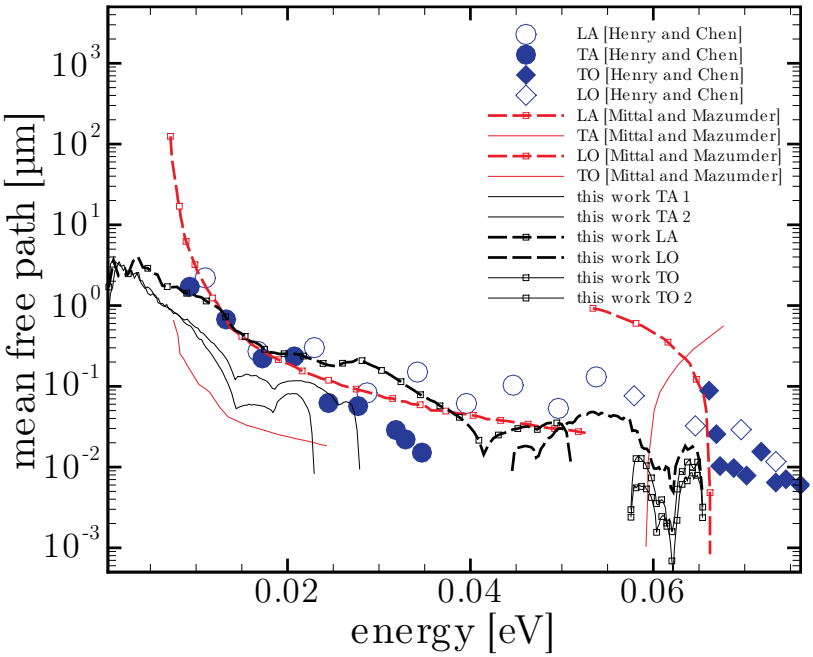

Figure 9.2: Bulk mean free path in Si computed at 300K (black lines) compared with previous Monte Carlo (Mitttal and Mazumder [108]) and molecular dynamics (Henry and Chen [109]) simulations. This work shows the spectral variations associated with a full-band dispersion

are extracted according to [88]:

$$\sigma_{k,j} = \sum_{q \subset BZ1} v_k v_j \tau(\mathbf{q}) \frac{(\hbar\omega)^2}{\mathrm{k}_B T} n(\mathbf{q},T)(n(\mathbf{q},T)+1), \tag{9.2}$$

where k and j represent a spatial direction, $\mathbf{q}$ is the phonon momentum, while v, τ, and ω represent the velocity, the average time between collisions, and the frequency of the phonon $\mathbf{q}$, respectively.

Figure 9.4 shows good agreement between the computed thermal conductivity (circles) and the experimental data (lines) for both Si^{nat} (open) and Si^{28} (solid).

The figure shows clearly the scattering-limited regime at temperatures above 30K, where the thermal conductivity is limited by anharmonic interactions, while a geometry-limited regime occurs at lower temperatures, when the mean free path is comparable with the device size and the thermal conductivity is limited by the interactions with the device boundaries.

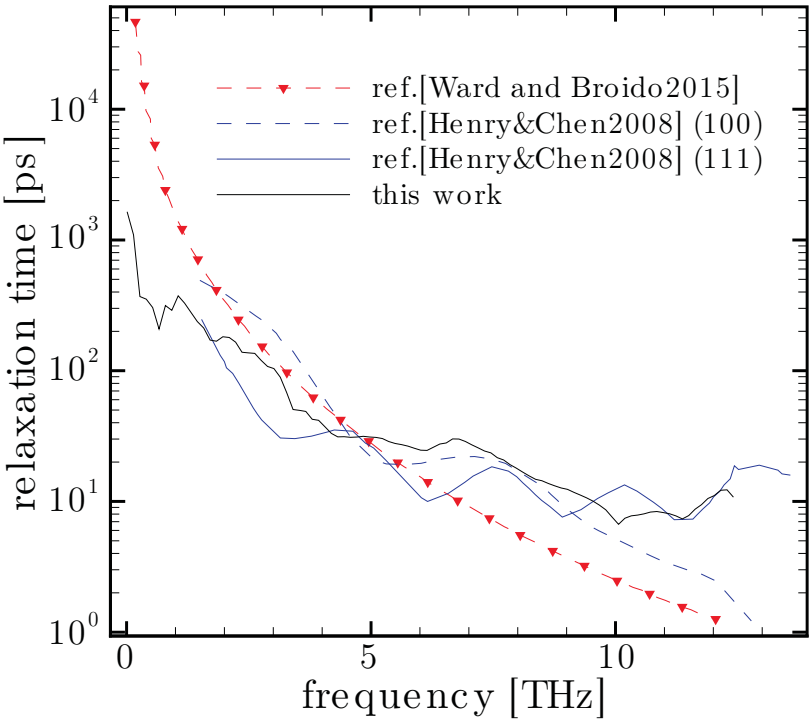

Figure 9.3: Comparison of the Si bulk relaxation time at 300K for longitudinal acoustic phonons between this work, a first-principles study (Ward and Broido [110]), and molecular dynamics simulations (Henry and Chen [109]). While this work shows close agreement to the molecular dynamic simulation, the comparison with the first-principle study shows an overestimation of the relaxation time for phonons close to the BZ1 boundaries.

9.4 Thin-Films

The 300K Si^{nat} thermal conductivity, σ_{tf}, is investigated as function thickness h_{tf} of a thin-film. As h_{tf} approaches the size of the phonon mean free path, the thermal conductivity is reduced by diffusive scattering with the boundaries. For Si^{nat}, the plot in fig. 9.2 suggests that this effect becomes visible for h_{tf} on the order of the micrometer. For this reason, the values of h_{tf} range between 5 nm and 1 μm. In all simulations the surfaces perpendicular to the thickness are assumed fully diffusive, while the surfaces of the other 4 sides are assumed perfectly specular. The thermal conductivity is computed by using Eq. (9.2), and similar to [99, 100] [111], τ is only affected by interactions which randomize the momentum, and so specular scattering does not contribute. In this approach, specular reflections have effects similar to periodic boundaries and do not limit the thermal conductivity. Figure 9.5 shows

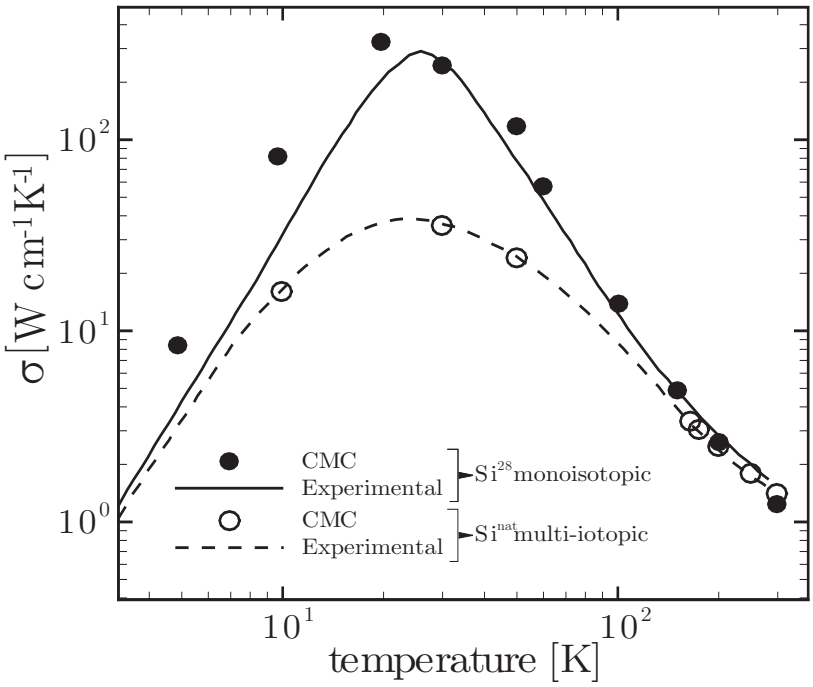

Figure 9.4: Isotope effect on the thermal conductivity as a function of the temperature: the natural silicon Si^{nat} (92.2% Si^{28}, 4.6% Si^{29}, 3.1% Si^{30}) is compared to a Si^{28} enriched sample. In the experimental results from Inyushkin *et al.* [92], the Si^{28} thermal conductivity (solid line) is up to ten times higher than Si^{nat} thermal conductivity (dashed line). The CMC simulations (symbols) reproduce the dependence on temperature of both thermal conductivities.

the predicted thermal conductivity (solid circles) compared to experimental results [112–115] (open symbols). The simulations show good agreement with the experimental results.

9.5 Transient Device Simulation

Our simulation code has been tested by comparing the simulation results of a simple 1D structure with an analytical solution.

A 2 μm long Si thermal resistor with initial temperature 290K has two side boundaries set to temperatures $T_H = 310$K and $T_L = 290$K as shown above the graphs in Fig. 9.6. The remaining boundaries are treated as specular boundaries. During the transient, an effective temperature as function of position is extracted from the local energy by inverting the energy-temperature relationship in Eq. (9.1). Here we simulate the time-dependent evolution of the resistor temperature profile for a step function change on the temperature

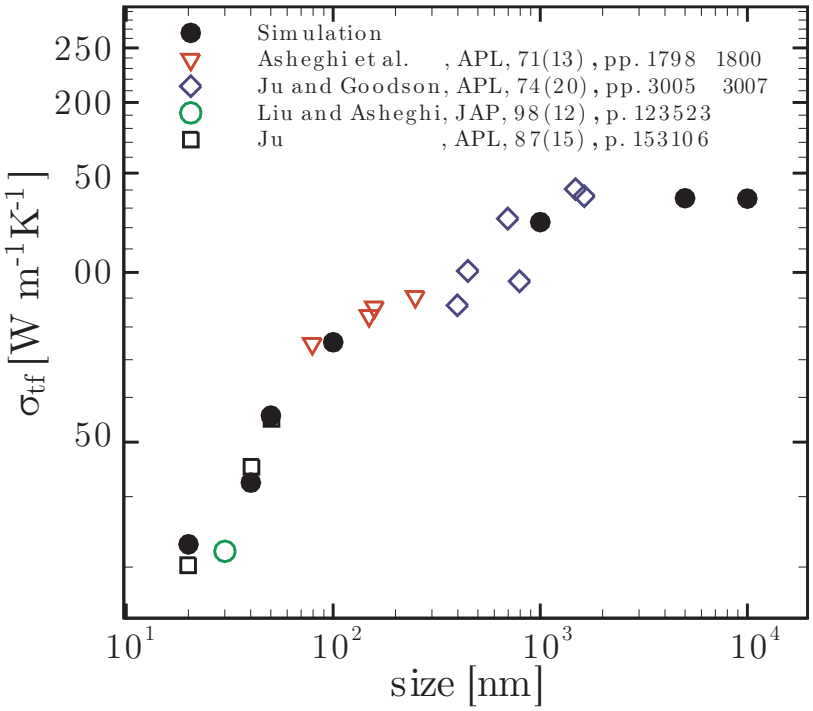

Figure 9.5: Comparison between the CMC results (solid circles) and experimental measurement[112–115] (open symbols) of the 300K thin-film thermal conductivity of Si for different film thickness.

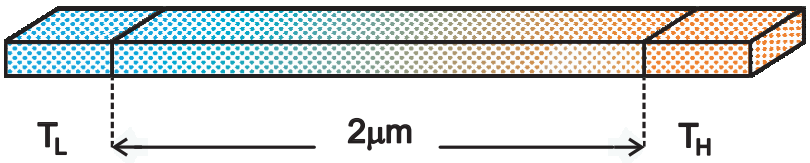

Figure 9.6: Schematic of the simulated Si thermal resistor, $T_H = 310$K and $T_L = 290$K, with the total length equal to 2.8 μm.

at the left contact.

The device operate on the limit for diffusive transport and some of the lower energy phonons have mean free path comparable with the device dimension as shown in Fig. 9.2, so the effects of ballistic phonons are visible in the steady-state solution, resulting in a temperature step close to the hot contact. For this one-dimensional heat transient problem, the evolution of the temperature distribution within the device can be determined analytically by solving Fourier's law of diffusion [116] by means of the Laplace transform [117]:

$$\frac{T(x,t)-T_L}{(T_H-T_L)} = \sum_{n=0}^{\infty} \text{erfc}\left(\frac{C_n - x}{2\sqrt{tk_D}}\right) - \text{erfc}\left(\frac{C_n + x}{2\sqrt{tk_D}}\right), \tag{9.3}$$

where $T(x,t)$ is the estimated temperature, k_D the thermal diffusivity ($8 \times 10^{-4}\mathrm{m}^2/\mathrm{s}$ for Si [118]), L the length of the device, $C_n = (2n+1)L$, and erfc is the complementary error function.

Figure 9.7 shows the Monte Carlo results averaged over over 50 simulations, each realized with a different seed of the pseudo-random number generator to improve the statistics.

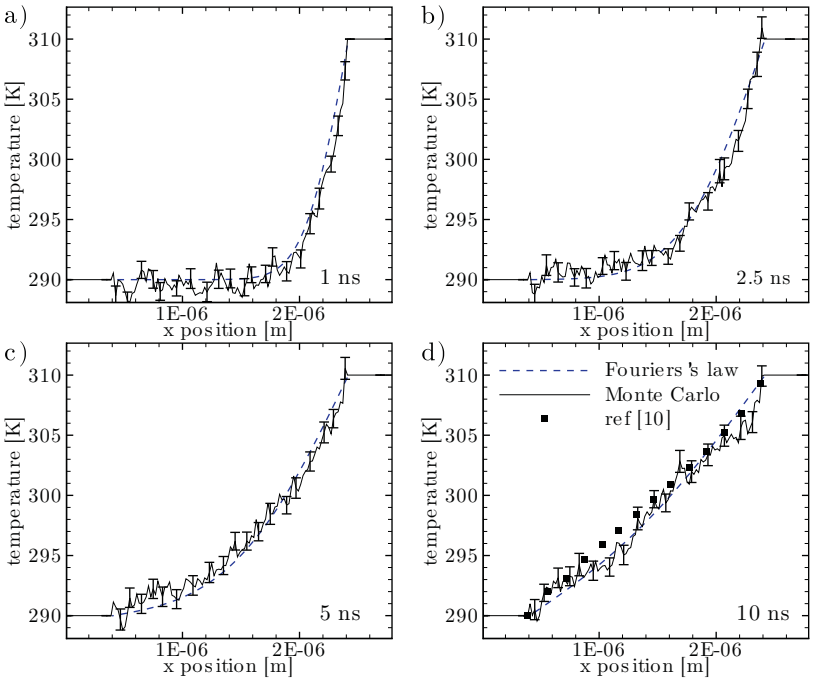

Figure 9.7: Comparison between the evolution of the temperature distribution obtained analytically by Eq. (9.3) (dashed lines), and the Monte Carlo simulation (solid lines). Each pane represents a snapshot of the temperature distribution at the time indicated on the bottom left corner. The last panel also shows a comparison with the steady state result of Lacroix *et al.* [5].

As it can be observed, the CMC simulations are in an excellent agreement with the analytical solution during the time transient for each time snapshot, and achieve the proper longtime limit of a linear temperature gradient as expected from Fourier's law, apart from the ballistic effects noted. The last panel shows a comparison with the steady-state result of Lacroix *et al.* [5], represented by solid squares.

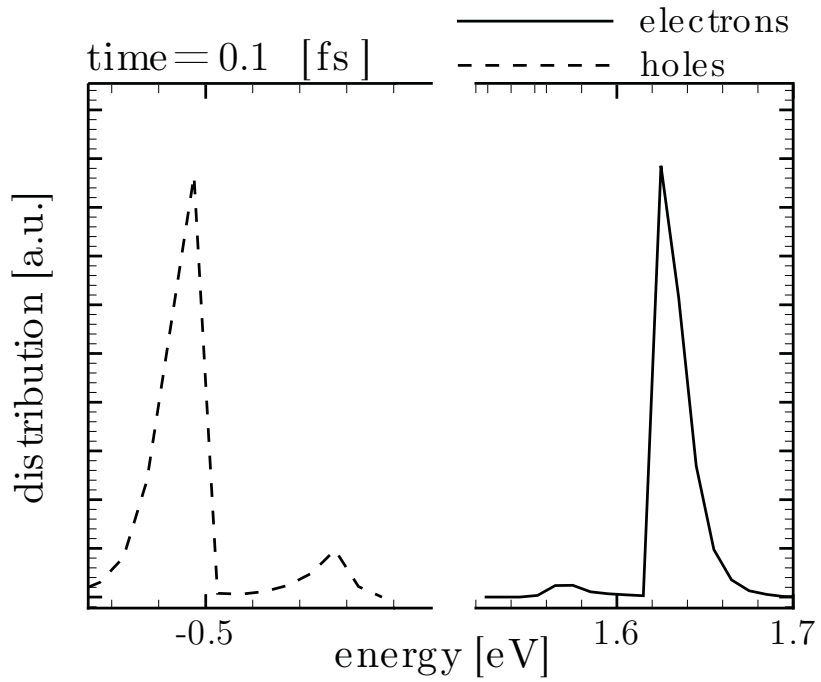

Figure 9.8: Distribution of holes and electrons in the isolated cube of Si at time $t = 0.1$ fs. The populations are concentrated at the initialization energy. Both charges and phonon are simulated as particles.

9.6 Electron Decay

The full algorithm for the coupled electron-phonon dynamics has been tested by simulating the thermalization process of electron-hole pairs in an isolated cube of Si that is initialized out of equilibrium. In particular, the simulated electrons are initialized with a kinetic energy of 0.6 eV, and the simulated phonons are initialized at 100 K, the system is left to relax. This simulation uses a scattering table computed for a maximum temperature of 400K. The focus of this first test is to observe the energy exchange between charges and phonons when modeling absorption and emission in terms of a Hawkes process, as discussed in section 4.6.

Figure 9.8 and 9.9 show the distribution of holes and electrons at time 0.1 fs and 30.1 fs, respectively for a ratio $W^c/W^p = 5.42 \times 10^{-3}$ (see Eq. 4.47 in Section 4.6). These plots represent the evolution of the charge population in the early transient period of the simulation. In Fig. 9.8, few carriers have scattered and the majority of the charges are still

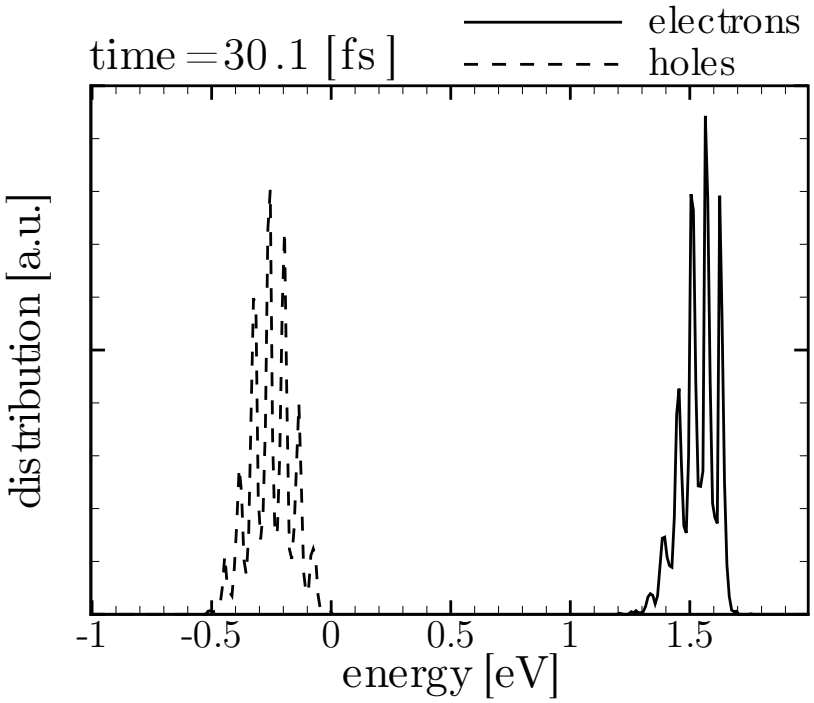

Figure 9.9: Distribution of holes and electrons in the isolated cube of Si at $t = 30.1$ fs with a ratio $W^c/W^p = 5.42 \times 10^{-3}$; the zero of the horizontal axis is placed at the top of the valence band. This picture clearly shows that the relaxation process of the charges results in regularly spaced peaks in the energy distribution. Fig 9.10 shows an enlarged representation of this electron distribution, where the distance between the peaks clearly corresponds to one optical phonon.

at the initial energy distribution.

The distributions in Fig. 9.9 show a clear pattern in the carrier decay. In particular, the picture shows evenly spaced density peaks for both electrons and holes. In order to better observe this behavior, Fig. 9.10 shows an enlarged picture of the electron distribution where the distance between the minor vertical grid lines is set to 0.06 eV, or roughly the average energy of an optical phonon. From this enlarged picture, it is clear that the density peaks are evenly spaced. Moreover, the distance between consecutive peaks corresponds to the energy of an optical phonon. Indeed, this behavior is expected since the emission of optical phonon is the dominant scattering mechanism, and hence the energy relaxation process, for high energy carriers out of equilibrium.

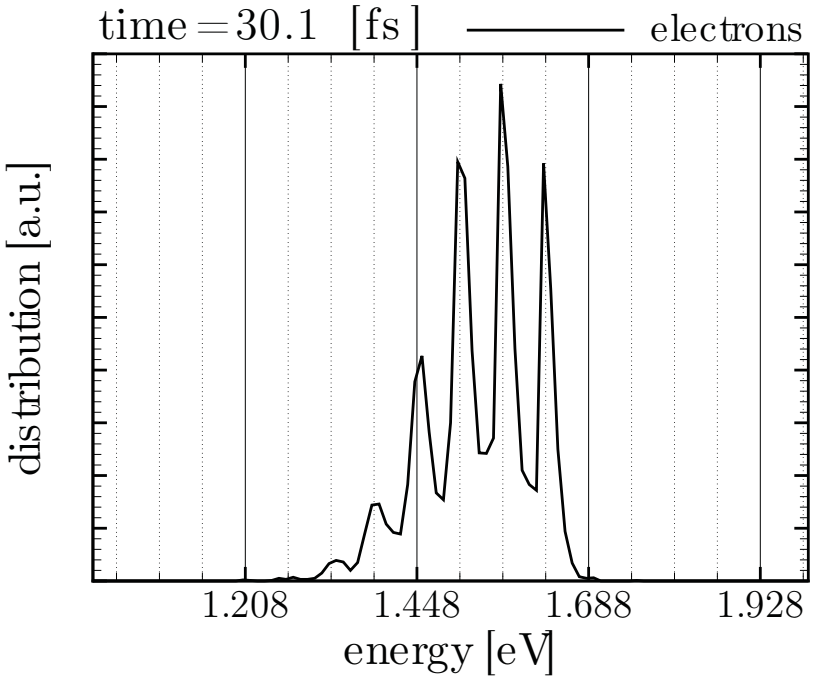

Figure 9.10: Enlarged picture of the electron population depicted in Fig 9.9. In this graph, the distance between the vertical grid lines has been set to 0.06 eV, which is roughly the average energy of an optical phonon. It is clear that optical emission is the dominant scattering mechanism since the distance between density peaks correspond to one optical phonon.

9.7 Conservation of Energy

A second group of simulations test the influence of the scattering algorithm and the weight of the simulated particles on the energy conservation of the system. The simulated system is similar to the previous case: an isolated cube of Si with a charge population initialized with a kinetic energy of 0.6 eV, and a phonon population initialized in thermal equilibrium at 100K. This simulations uses the same scattering table of the previous case. During the relaxation process, the instantaneous particles distribution is recorded at intervals of 0.1 fs. Figures 9.11, 9.12, and 9.13 show the time evolution of the system in terms of total energy lost of the charge population (dashed line) and the energy gained of the simulated phonons (solid line) in three different cases. In other words, the energy gained by the particles has positive sign for phonons and negative sign for charges. Since the scattering events conserve energy, the two lines should superimpose perfectly in each plot.

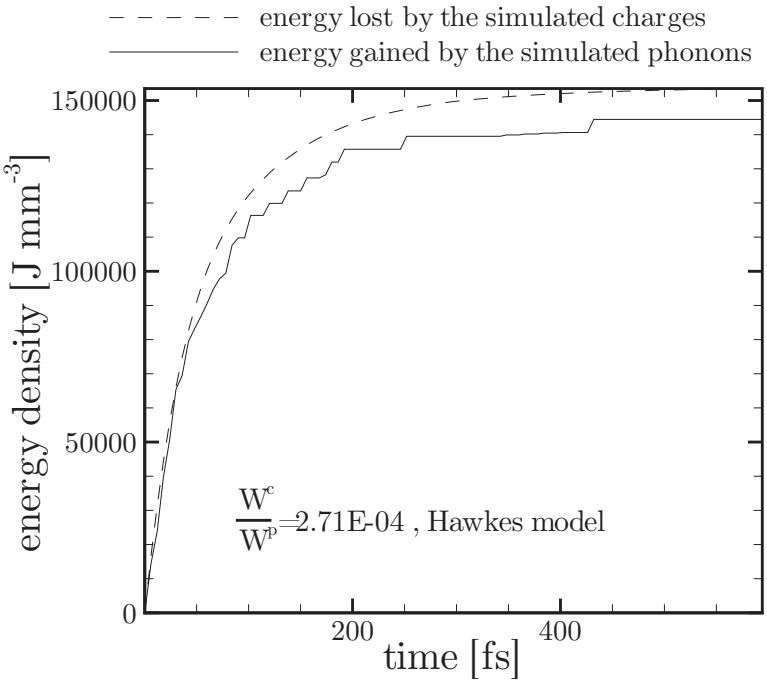

Figure 9.11: Evolution of the energy balance of charge and phonon population in the isolated cube of Si. The quantities are expressed in terms of total energy total lost by the simulated charges (dashed line) and the energy gained by simulated phonons (solid line). Here the phonon emission and absorption is treated as a Hawkes process.

In Fig. 9.11 and 9.12, the electron-phonon interaction is treated as a Hawkes process, as described in Ch. 4, with different W^c/W^p ratios. In particular, the simulation plotted in Fig. 9.11 has a ratio W^c/W^p half compared to the case presented in Fig. 9.12. The importance of this parameter is presented in Eq. 4.47 (Ch. 4), which suggests that the average distance between the two curves should scale with the inverse square of W^c/W^p. This means that simulations with a larger ratio W^c/W^p should perform better. Indeed, compared to the case in Fig. 9.11 the plot in Fig 9.12 shows a much closer agreement between the energy gained by the simulated phonons and the energy lost by the simulated charges

Figure 9.13 shows what happens when the charge-phonon interaction is simply treated in terms of Poisson probability distribution, as described in the first sections of Ch. 4. In this case the ratio W^c/W^p is much larger than in the previous two cases, however, there is

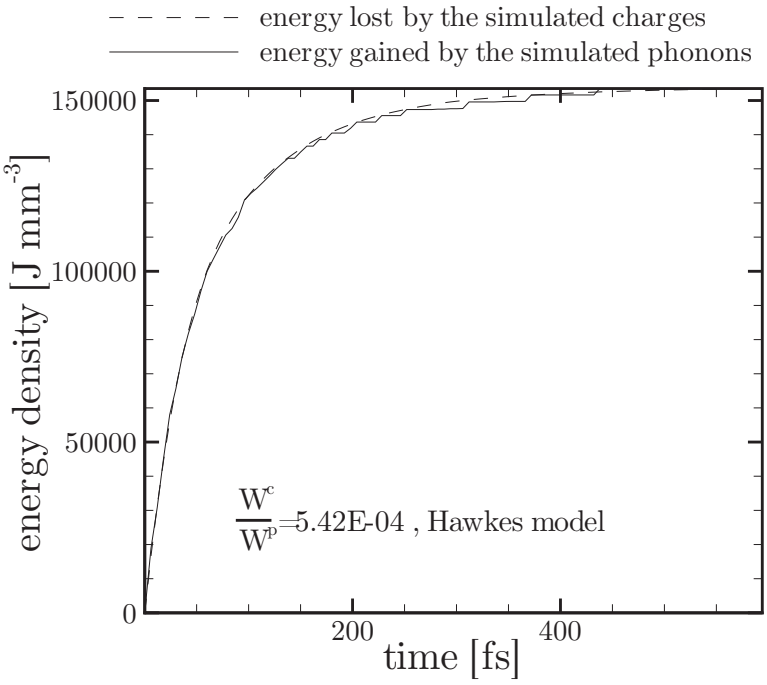

Figure 9.12: Evolution of the energy balance of charge and phonon population in the isolated cube of Si. The quantities are expressed in terms of total energy lost by the simulated charges (dashed line) and the energy gained by simulated phonons (solid line). Here the phonon emission and absorption is treated as a Hawkes process. In this case the two lines show excellent agreement.

a large difference between the energy gained by the simulated phonons and the energy lost by the simulated charges. Indeed, this is the expected behavior predicted in Ch. 4. These plots shows the importance of using an appropriate scattering algorithm and a sufficiently large ratio W^c/W^p.

9.8 Electron Thermalization

In another test, all the simulated electrons in the isolated Si cube are initialized with a kinetic energy of 20 meV and the simulated phonons are initialized at 300K. All these simulation use the same scattering table computed for a maximum temperature of 400K. This set-up tests the code's capability to correctly thermalize the simulated population of particles by both absorption and emission for different ratios of W^c/W^p. Figure 9.14 compares an equilibrium reference distribution of the simulated electron (circles) against distributions

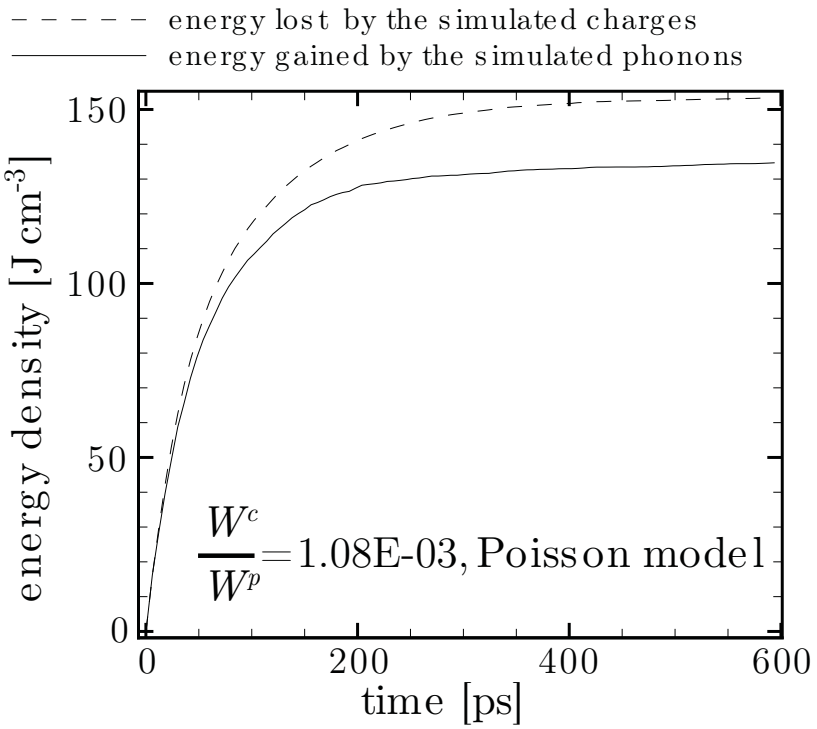

Figure 9.13: Evolution of the energy balance of charge and phonon population in the isolated cube of Si. The quantities are expressed in terms of total energy lost by the simulated charges (dashed line) and the energy gained by simulated phonons (solid line). Here the phonon emission and absorption is treated in terms of Poisson probability distribution. While the ratio W^c/W^p is close to 10^{-2}, this approach performs much worse than the case in Fig 9.11 where W^c/W^p 10^{-4} but emission and absorption are treated as Hawkes processes.

obtained at time 0.6 fs (dashed line) and 300 fs (solid line) with a ratio $W^c/W^p \approx 10^{-4}$. At 0.6 fs, few electrons have scattered and the majority of the charges are still at the initial energy distribution (dashed line). From the picture it is clear that the distribution at 300 fs corresponds to the reference distribution, while the initialized population is out of equilibrium. In section 9.6 optical phonons are the predominant scattering mechanism; however, in this simulation the acoustic phonon plays the crucial role in defining the final shape of the electron distribution function.

Fig. 9.15 compares the time evolution of the electron distribution for the simulation in Fig. 9.14 (dashed line) and for two other ratios of W^c/W^p corresponding to 2.71×10^{-3} (solid line) and 2.71×10^{-3} (dotted line). In this picture, the distributions are expressed

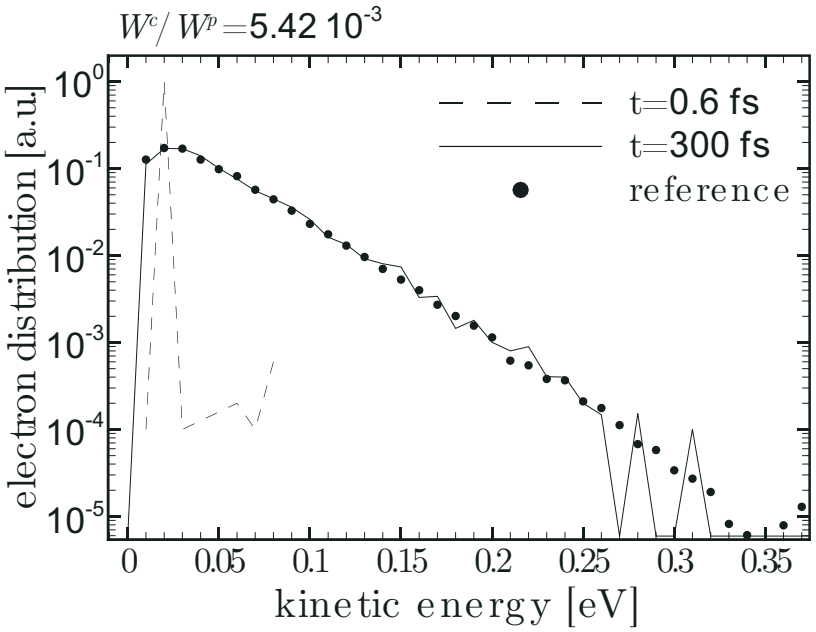

Figure 9.14: Thermalization of electrons in Si via electron-phonon scattering. The reference equilibrium distribution of the simulated electrons (circles) is compared with simulated distributions obtained at time 0.6 fs (dashed line) and 300 fs (solid line) with a ratio $W^c/W^p \approx 10^{-4}$. The 300 fs closely match the reference distribution.

in terms of their distance from the reference distribution by employing the Jensen-Shannon divergence. We can see in this figure small differences in the transient evolution: all three cases reach steady-state at the same time, around the 150 fs mark. After this point, the divergence value results from the statistical noise. The square indicates the divergence for the distribution at 300 fs from Fig. 9.14, this plot clearly shows that the system is indeed in steady state regime.

While all three cases reach steady-state at the same time, the solid and dashed lines are sensibly less noisy during the 150 fs transient. This effect is caused by the use of a momentum and time dependent occupation number (presented in Section 3.4) to dynamically evaluate the the electron scattering rate via the rejection technique. Indeed, the three simulations have the same W^c but different W^p. Since all simulation have the same starting temperature, larger W^p correspond to fewer simulated phonons. This causes a difference in the statistical noise when computing the occupation number via Eq. 3.10.

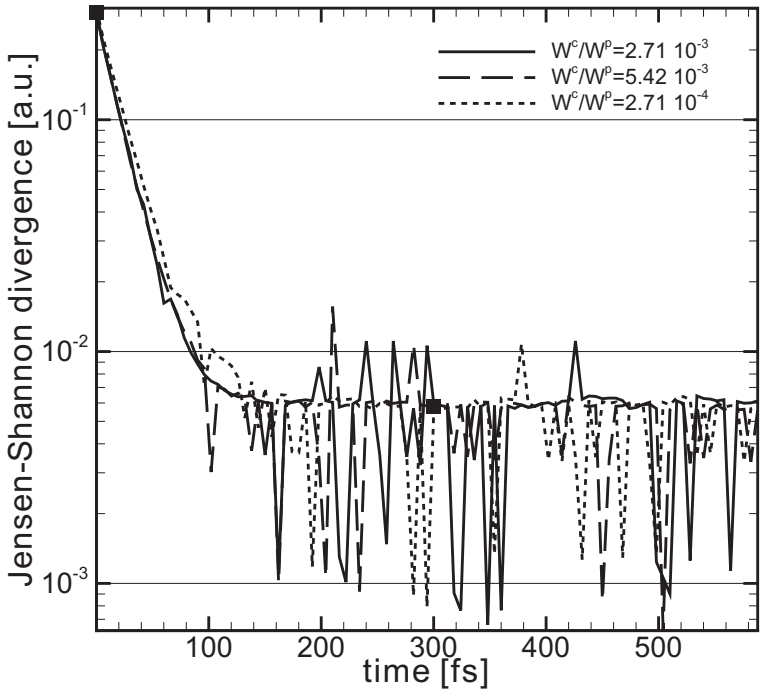

Figure 9.15: Time evolution of the charge distribution in the in the isolated Si cube. The quantities are represented in terms of their distance from the reference distribution. The vertical axis shows the logarithm of the Jensen-Shannon divergence. The three lines correspond to simulation with ratios W^c/W^p. All three cases reach steady-state around the 150 fs mark, then the divergence shows the magnitude of the statistical noise. The black square corresponds to the divergence for the case of $W^c/W^p = 5.42 \times 10^{-3}$ at 300 fs that is shown in Fig. 9.14.

Chapter 10

CONCLUSION

A rejection technique to explicitly model population-dependent scattering while ensuring the statistical conservation of energy and momentum has been presented. The technique has been implemented in a full-band Cellular Monte Carlo (CMC) framework to simulate phonon transport in semiconductors. The scattering algorithm explicitly solves the many-body problem allowing for the simulation of an arbitrary non-equilibrium phase space distribution of phonons with numerically calculated dispersion, without resorting to the linearization of the BTE itself. The details have been presented for the algorithmic implementation of both the three particle scattering for the treatment of the anharmonic interactions between phonons, and the two-state scattering process used to model scattering with isotope.

Challenges to the modeling electron-phonon scattering when both populations are treated as simulated particles have been investigated. Two possible families of solutions based on the population-dependent scattering technique have been presented. The first approach models phonon and emission absorption as a Poisson distribution and fails to reliably conserve energy during transient simulations. For this reason, a second approach based on the multivariable Hawkes point process has been successfully developed. This second method limit the maximum error in the conservation of the system energy by effectively acting as a statistical form of energy book-keeping.

The simulation results have been successfully compared to data available in literature. In particular, the numerical simulations show very good agreement with experimental results for thermal conductivity of silicon as function of temperature, isotopes composition and sample size.

Future work will continue testing the code by comparing electron-phonon simulations to available optical experiments of coupled electron phonon decay. The final goal of this

project is to use the code as a TCAD tool for early-stage design of real power electronic devices.

REFERENCES

[1] Klitsner, T., VanCleve, J., Fischer, H. E., and Pohl, R., 1988. "Phonon Radiative Heat Transfer and Surface Scattering". *Physical Review B,* **38**(11), p. 7576.

[2] Peterson, R. B., 1994. "Direct Simulation of Phonon-Mediated Heat Transfer in a Debye Crystal". *ASME J. Heat Transfer,* **116**(4), p. 815.

[3] Mazumder, S., and Majumdar, A., 2001. "Monte Carlo Study of Phonon Transport in Solid Thin Films Including Dispersion and Polarization". *ASME J Heat Transfer,* **123**(4), pp. 749–759.

[4] Chen, Y., Li, D., Lukes, J. R., and Majumdar, A., 2005. "Monte Carlo Simulation of Silicon Nanowire Thermal Conductivity". *ASME J Heat Transfer,* **127**(10), pp. 1129–1137.

[5] Lacroix, D., Joulain, K., and Lemonnier, D., 2005. "Monte Carlo Transient Phonon Transport in Silicon and Germanium at Nanoscales". *Physical Review B,* **72**(6), p. 064305.

[6] Péraud, J. M., and Hadjiconstantinou, N. G., 2011. "Efficient Simulation of Multidimensional Phonon Transport Using Energy-Based Variance-Reduced Monte Carlo Formulations". *Physical Review B,* **84**(20), p. 205331.

[7] Ramayya, E., Maurer, L., Davoody, A., and Knezevic, I., 2012. "Thermoelectric Properties of Ultrathin Silicon Nanowires". *Physical Review B,* **86**(11), p. 115328.

[8] McGaughey, A. J. H., and Jain, A., 2012. "Nanostructure Thermal Conductivity Prediction by Monte Carlo Sampling of Phonon Free Paths". *Applied Physics Letters,* **100**(6), p. 061911.

[9] Kukita, K., and Kamakura, Y., 2013. "Monte Carlo Simulation of Phonon Transport in Silicon Including a Realistic Dispersion Relation". *Journal of Applied Physics,* **114**(15), p. 154312.

[10] Mei, S., Maurer, L., Aksamija, Z., and Knezevic, I., 2014. "Full-Dispersion Monte Carlo Simulation of Phonon Transport in Micron-Sized Graphene Nanoribbons". *Journal of Applied Physics,* **116**(16), p. 164307.

[11] Landon, C. D., and Hadjiconstantinou, N. G., 2014. "Deviational Simulation of Phonon Transport in Graphene Ribbons With ab initio Scattering". *Journal of Applied Physics,* **116**(16), Oct., p. 163502.

[12] Mohamed, M., Aksamija, Z., Vitale, W., Hassan, F., Park, K. H., and Ravaioli, U., 2014. "A Conjoined Electron and Thermal Transport Study of Thermal Degradation Induced During Normal Operation of Multigate Transistors". *IEEE Transactions on Electron Devices,* **61**(4), Apr., pp. 976–983.

[13] Rowlette, J. A., and Goodson, K. E., 2008. "Fully coupled nonequilibrium electron-phonon transport in nanometer-scale silicon fets". *IEEE Transactions on Electron Devices,* **55**(1), Jan, pp. 220–232.

[14] Kamakura, Y., Mori, N., Taniguchi, K., Zushi, T., and Watanabe, T., 2010. "Coupled monte carlo simulation of transient electron-phonon transport in nanoscale devices". In 2010 International Conference on Simulation of Semiconductor Processes and Devices, pp. 89–92.

[15] Lai, J., and Majumdar, A., 1996. "Concurrent thermal and electrical modeling of sub-micrometer silicon devices". *Journal of Applied Physics,* **79**(9), pp. 7353–7361.

[16] Lugli, P., Bordone, P., Reggiani, L., Rieger, M., Kocevar, P., and Goodnick, S. M., 1989. "Monte carlo studies of nonequilibrium phonon effects in polar semiconductors and quantum wells. i. laser photoexcitation". *Phys. Rev. B,* **39**, Apr, pp. 7852–7865.

[17] Shi, Y. B., and Knezevic, I., 2014. "Nonequilibrium phonon effects in midinfrared quantum cascade lasers". *Journal of Applied Physics,* **116**(12), p. 123105.

[18] Saraniti, M., and Goodnick, S., 2000. "Hybrid full-band Cellular Automaton/Monte Carlo approach for fast simulation of charge transport in semiconductors". *IEEE Transactions on Electron Devices,* **47**(10), Oct., pp. 1909–1915.

[19] Hamaguchi, C., 2001. *Basic Semiconductor Physics.* Springer, Berlin ; New York.

[20] McQuarrie, D., 2000. *Statistical Mechanics.* University Science Books, Sausalito, CA.

[21] Cercignani, C., 1972. "On the boltzmann equation for rigid spheres". *Transport Theory and Statistical Physics,* **2**(3), pp. 211–225.

[22] Peierls, R., 1955. *Quantum Theory of Solids.* International series of monographs on physics. Clarendon Press, Oxford, UK.

[23] Grasser, T., Tang, T.-W., Kosina, H., and Selberherr, S., 2003. "A review of hydrodynamic and energy-transport models for semiconductor device simulation". *Proceedings of the IEEE,* **91**(2), pp. 251–274.

[24] SS, S., EE, P., RW, D., and KE., G., 2005. "Non-equilibrium phonon distributions in sub-100nm silicon transistors.". *ASME. J. Heat Transfer.,* **7**, pp. 638–647.

[25] Jacoboni, C., and Reggiani, L., 1983. "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials". *Reviews of Modern Physics,* **55**, July, pp. 645–705.

[26] Joshi, A., and Majumdar, A., 1993. "Transient Ballistic and Diffusive Phonon Heat Transport in Thin Films". *Journal of Applied Physics,* **74**(1), pp. 31–39.

[27] Chai, J. C., Lee, H. S., and Patankar, S. V., 1994. "Finite Volume Method for Radiation Heat Transfer". *J. Thermophys Heat Transfer,* **8**(3), pp. 419–425.

[28] Chen, G., 2002. "Ballistic-Diffusive Equations for Transient Heat Conduction From Nano to Macroscales". *ASME J Heat Transfer,* **124**(2), pp. 320–328.

[29] Sinha, S., Pop, E., Dutton, R., and Goodson, K., 2006. "Non-Equilibrium Phonon Distributions in Sub-100nm Silicon Transistors". *ASME J Heat Transfer,* **128**(7), pp. 638–647.

[30] Grasser, T., Tang, T.-W., Kosina, H., and Selberherr, S., 2003. "A review of hydrodynamic and energy-transport models for semiconductor device simulation". *Proc. IEEE,* **91**(2), Feb., pp. 251–274.

[31] Murthy, J. Y., Narumanchi, S. V., José A, P.-G., Wang, T., Ni, C., and Mathur, S. R., 2005. "Review of Multiscale Simulation in Submicron Heat Transfer". *Int. J. Mult. Comp. Eng,* **3**(1).

[32] Zhao, Y., Watling, J., Kaya, S., Asenov, A., and Barker, J., 2000. "Drift diffusion and hydrodynamic simulations of si/sige p-mosfets". *Materials Science and Engineering: B,* **72**(2), pp. 180–183.

[33] Rupp, K., Jungemann, C., Hong, S.-M., Bina, M., Grasser, T., and Jüngel, A., 2016. "A review of recent advances in the spherical harmonics expansion method for semiconductor device simulation". *Journal of Computational Electronics,* **15**(3), pp. 939–958.

[34] Jacoboni, C., and Lugli, P., 1989. *The Monte Carlo Method for Semiconductor Device Equations.* Springer–Verlag, Wien, NewYork.

[35] Hadamard, J., 1923. *Lectures on Cauchy's Problem in Linear Partial Differential Equations.* Yale University Press, New Haven.

[36] Ferry, D. K., 1991. *Semiconductors.* Macmillan ; Collier Macmillan Canada ; Maxwell Macmillan International, New York : Toronto : New York.

[37] Fischetti, M. V., and Laux, S. E., 1988. "Monte Carlo Analysis of Electron Transport in Small Semiconductor Devices Including Band-Structure and Space-Charge Effects". *Physical Review B,* **38**(14), p. 9721.

[38] Rees, H., 1969. "Calculation of distribution functions by exploiting the stability of the steady state". *Journal of Physics and Chemistry of Solids,* **30**(3), pp. 643–655.

[39] Phillips, J. C., 1958. "Energy-band interpolation scheme based on a pseudopotential". *Physical Review,* **112**, Nov., pp. 685–695.

[40] Phillips, J. C., and Kleinman, L., 1959. "New method for calculating wave functions in crystals and molecules". *Physical Review,* **116**, Oct., pp. 287–294.

[41] Kunc, K., and Nielsen, O. H., 1979. "Lattice dynamics of Zincblende structure compounds. II. Shell model". *Computer Physics Communications,* **17**, pp. 413–422.

[42] Dirac, P. A. M., 1927. "The Quantum Theory of the Emission and Absorption of Radiation". *Proc. R. Soc. London A,* **114**(767), pp. 243–265.

[43] Devolder, P., Janssen, J., and Manca, R., 2015. *Basic Stochastic Processes (mathematics and Statistics).* Wiley-ISTE, Hoboken, NJ.

[44] Goodnick, S., and Lugli, P., 1988. "Effect of Electron-Electron Scattering On Nonequilibrium Transport in Quantum-Well Systems". *Phys. Rev. B,* **37**(5), p. 2578.

[45] Ross, S. M., 1995. *Stochastic Processes.* John Wiley & Sons, New York.

[46] Streit, R. L., 2010. *The Poisson Point Process.* Springer, Boston, MA.

[47] Lewis, P. A. W., and Shedler, G. S., 1979. "Simulation of nonhomogeneous poisson processes by thinning". *Naval Research Logistics Quarterly,* **26**(3), pp. 403–413.

[48] Hadar, J., and Russell, W. R., 1969. "Rules for ordering uncertain prospects". *The American Economic Review,* **59**(1), pp. 25–34.

[49] Daley, D. J., and Vere-Jones, D., 2007. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods.* Springer, New York.

[50] Cox, D., and Isham, V., 1980. *Point Processes.* Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, London.

[51] Khinchin, A. Ã. A. Ã., 1949. *Mathematical Foundations of Statistical Mechanics.* [The Dover series in mathematics and physics]. Dover Publications, New York.

[52] Daley, D. J., and Vere-Jones, D., 2007. *An Introduction to the Theory of Point Processes: Volume Ii: General Theory and Structure.* Springer-Verlag, New York.

[53] Hawkes, A. G., 1971. "Spectra of some self-exciting and mutually exciting point processes". *Biometrika,* **58**(1), pp. 83–90.

[54] Cox, D. R., 1955. "Some statistical methods connected with series of events". *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 129–164.

[55] Ogata, Y., 1988. "Statistical models for earthquake occurrences and residual analysis for point processes". *Journal of the American Statistical Association,* **83**(401), pp. 9–27.

[56] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E., 2011. "Self-exciting point process modeling of crime". *Journal of the American Statistical Association,* **106**(493), Mar., pp. 100–108.

[57] Tita, G., and Ridgeway, G., 2007. "The impact of gang formation on local patterns of crime". *Journal of Research in Crime and Delinquency,* **44**(2), May, pp. 208–237.

[58] Omi, T., Hirata, Y., and Aihara, K., 2017. "Hawkes process model with a time-dependent background rate and its application to high-frequency financial data". *Phys. Rev. E,* **96**, July, p. 012303.

[59] Hewlett, P., 2006. "Clustering of order arrivals, price impact and trade path optimisation". In In Workshop on Financial Modeling with Jump processes, Ecole Polytechnique.

[60] Brémaud, P., and Massoulié, L., 1996. "Stability of nonlinear hawkes processes". *The Annals of Probability*, pp. 1563–1588.

[61] Bilotta, S., Pergola, E., Pinzani, R., and Rinaldi, S., 2015. *Recurrence Relations, Succession Rules, and the Positivity Problem.* Springer International Publishing, Cham, pp. 499–510.

[62] Sharma, R., Gupta, M., and Kapoor, G., 2010. "Some better bounds on the variance with applications". *Journal of Mathematical Inequalities*(3), pp. 355–363.

[63] Jensen, J. L., and Møller, J., 1991. "Pseudolikelihood for exponential family models of spatial point processes". *The Annals of Applied Probability*, pp. 445–461.

[64] Propp, J. G., and Wilson, D. B., 1996. "Exact sampling with coupled Markov chains and applications to statistical mechanics". *Random structures and Algorithms,* **9**(1-2), pp. 223–252.

[65] Ogata, Y., 1981. "On lewis' simulation method for point processes". *IEEE Transactions on Information Theory,* **27**(1), pp. 23–31.

[66] Billera, L. J., and Diaconis, P., 2001. "A geometric interpretation of the metropolis-hastings algorithm". *Statistical Science*, pp. 335–339.

[67] Ridley, B. K., 2000. *Quantum Processes in Semiconductors*, 4$^{\text{th}}$ ed. Oxford University Press, New York.

[68] Froehlich, H., 1937. "Theory of electrical breakdown in ionic crystals". *Proceedings of the Royal Society of London,* **A160**, p. 230.

[69] H. Fröhlich, 1954. "Electrons in lattice fields". *Advances in Physics,* **3**, pp. 325–361.

[70] Lee, B. C., Kim, K. W., Dutta, M., and Stroscio, M. A., 1997. "Electron optical-phonon scattering in Wurtzite crystals". *Physical Review B,* **56**, pp. 997–1000.

[71] Bulutay, C., Ridley, B. K., and Zakhleniuk, N. A., 2000. "Full-band polar optical phonon scattering analysis and negative differential conductivity in Wurtzite GaN". *Physical Review B,* **62**, pp. 15754–15763.

[72] Brennan, K. F., 1991. "Ensemble Monte Carlo investigation of nonlinear transport effects in semiconductor heterostructure devices". In *Monte Carlo Device Simulation: Full Band and Beyond*, K. Hess, ed., The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers Group, Dordrecht, The Netherlands, ch. 6, pp. 161–189.

[73] Moglestue, C., 1982. "Monte Carlo particle modelling of small semiconductor devices". *Computer Methods in Applied Mechanics and Engineering,* **30**(2), pp. 173–208.

[74] Ridley, B., 1977. "Reconciliation of the Conwell-Weisskopf and Brooks-Herring formulae for charged-impurity scattering in semiconductors: Third-body interference". *Journal of Physic C: Solid State Physics,* **10**, pp. 1589–1593.

[75] Brooks, H., and Herring, C., 1951. "Scattering by ionized impurities in semiconductors". *Physical Review,* **83**, p. 879.

[76] Conwell, E., and Weisskopf, V. F., 1950. "Theory of impurity scattering in semiconductors". *Physical Review,* **77**, Feb., pp. 388–390.

[77] Bude, J., 1991. "Scattering mechanism for semiconductor transport calculations". In *Monte Carlo Device Simulation: Full Band and Beyond*, K. Hess, ed., The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers Group, Dordrecht, The Netherlands, ch. 2, pp. 27–66.

[78] Yu, P., and Cardona, M., 1996. *Fundamental of Semiconductors - Physics and Material Properties*, first ed. Springer-Verlag, Berlin.

[79] Chattopadhyay, D., and Queisser, H. J., 1981. "Electron scattering by ionized impurities in semiconductors". *Reviews of Modern Physics,* **53**, Oct., pp. 745–768.

[80] Tyaginov, S., Bina, M., Franco, J., Wimmer, Y., Kaczer, B., and Grasser, T., 2015. "On the importance of electron–electron scattering for hot-carrier degradation". *Japanese Journal of Applied Physics,* **54**(4S), p. 04DC18.

[81] Srivastava, G. P., 1990. *The Physics of Phonons.* Taylor & Francis Group, New York, NY.

[82] Ashcroft, N. W., and Mermin, N. D., 1981. *Solid State Physics.* Holt–Sauders International Editions, Tokyo.

[83] Schrödinger, E., 1914. "Zur dynamik elastisch gekoppelter punktsysteme". *Annalen der Physik,* **349**(14), pp. 916–934.

[84] von Kármán Th., 1929. "Wärmeausgleich in Kristallen im Lichte der Quantenmechanik". *Naturwissenschaften,* **17**, pp. 385–387.

[85] L. S. Ornstein, F. Z., 1917. "The Unimpeded Spreading of Heat Even in Case of Deviations from Hooke's Law". *Proceedings of the Royal Netherlands Academy of Arts and Sciences,* **19**, pp. 1295–1304.

[86] Peierls, R., 1929. "Zur kinetischen Theorie der Wärmeleitung in Kristallen". *Annalen der Physik,* **395**(8), pp. 1055–1101.

[87] Klemens, P., 1966. "Anharmonic Decay of Optical Phonons". *Physical Review,* **148**(2), p. 845.

[88] Klemens, P., 1958. "Thermal Conductivity and Lattice Vibrational Modes". *Solid State Physics,* **7**, pp. 1–98.

[89] Han, Y.-J., and Klemens, P. G., 1993. "Anharmonic thermal resistivity of dielectric crystals at low temperatures". *Physical Review B,* **48**, Sept., pp. 6033–6042.

[90] Ridley, B., and Gupta, R., 1991. "Nonelectronic Scattering of Longitudinal-optical Phonons in Bulk Polar Semiconductors". *Physical Review B,* **43**(6), p. 4939.

[91] Ferry, D., 1974. "Decay of Polar-Optical Phonons in Semiconductors". *Physical Review B,* **9**(10), p. 4277.

[92] Inyushkin, A., Taldenkov, A., Gibin, A., Gusev, A., and Pohl, H., 2004. "On the Isotope Effect in Thermal Conductivity of Silicon". *Phys. Status Solidi (C),* **1**(11), pp. 2995–2998.

[93] Ruf, T., Henn, R., Asen-Palmer, M., Gmelin, E., Cardona, M., Pohl, H.-J., Devyatych, G., and Sennikov, P., 2003. "Erratum to "thermal conductivity of isotopically enriched silicon" [solid state communications, 115 (2000) 243–247]". *Solid State Communications,* **127**(3), p. 257.

[94] Burzo, M. G., Komarov, P. L., and Raad, P., 2004. "Non-Contact Thermal Conductivity Measurements of Gold Covered Natural and Isotopically-Pure Silicon and Their Respective Oxides". In 5^{th} Int. Conf. On Thermal, Mechanical and Thermo-Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems, Brussels, Belgium.

[95] Capinski, W. S., Maris, H. J., Bauser, E., Silier, I., Asen-Palmer, M., Ruf, T., Cardona, M., and Gmelin, E., 1997. "Thermal conductivity of isotopically enriched si". *Applied Physics Letters,* **71**(15), pp. 2109–2111.

[96] Klemens, P. G., 1967. "Decay of High-Frequency Longitudinal Phonons". *Journal of Applied Physics,* **38**, Nov., pp. 4573–4576.

[97] Adams, J. G., Tang, T.-W., and Kay, L. E., 1994. "Monte Carlo simulation of noise in gaas semiconductor devices". *IEEE Transactions on Electron Devices,* **41**(4), pp. 575–581.

[98] González, T., and D, P., 1996. "Physical Models of Ohmic Contact for Monte Carlo Device Simulation". *Solid-State Electronics,* **39**(4), pp. 555–562.

[99] Fuchs, K., 1938. "The Conductivity of Thin Metallic Films According to the Electron Theory of Metals". In Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 34, Cambridge Univ Press, pp. 100–108.

[100] Sondheimer, E. H., 1952. "The Mean Free Path of Electrons in Metals". *Advances in Physics,* **1**(1), pp. 1–42.

[101] Lambert, J., 1760. *Lamberts Photometrie : Photometria, sive De mensura et gradibus luminus, colorum et umbrae.* W. Engelmann, Leipzig,Ge.

[102] Kittel, C., 2005. *Introduction to Solid State Physics*, 8th ed. John Wiley and Sons Ltd, Hoboken, NJ.

[103] H.Press, W., A.Teukolsky, S., william T.Vetterling, and P.Flannery, B., 2007. *Numerical Recipes*, 3rd ed. Cambridge University Press, Cambridge, ENG.

[104] GOFF, G. K., 1976. "The citardauq formula". *The Mathematics Teacher,* **69**(7), pp. 550–551.

[105] Rosen, M. I., 1995. "Niels hendrik abel and equations of the fifth degree". *The American Mathematical Monthly,* **102**(6), pp. 495–505.

[106] Sabatti, F., Goodnick, S., and Saraniti, M., 2016. "Simulation of phonon transport in semiconductors using a population-dependent many-body cellular monte carlo approach". *Journal of Heat Transfer,* **139**(3), 3, pp. 032002–032002–10.

[107] Lin, J., 2006. "Divergence Measures Based on the Shannon Entropy". *IEEE Transactions on Information Theory,* **37**(1), Sept., pp. 145–151.

[108] Mittal, A., and Mazumder, S., 2010. "Monte Carlo Study of Phonon Heat Conduction in Silicon Thin Films Including Contributions of Optical Phonons". *ASME J Heat Transfer,* **132**(5), p. 052402.

[109] Henry, Asegun S.; Chen, G., 2008. "Spectral Phonon Transport Properties of Silicon Based on Molecular Dynamics Simulations and Lattice Dynamics". *Journal of Computational and Theoretical Nanoscience,* **5**(2), pp. 141–152.

[110] Ward, A., and Broido, D. A., 2010. "Intrinsic phonon Relaxation Times from First-Principles Studies of the Thermal Conductivities of Si and Ge". *Physical Review B,* **81**, Feb., p. 085205.

[111] Ziman, J., 1964. *Principles of the Theory of Solids.* Cambridge University Press, London, UK.

[112] Asheghi, M., Leung, Y. K., Wong, S. S., and Goodson, K. E., 1997. "Phonon-boundary Scattering in Thin Silicon Layers". *Appl Phys Lett,* **71**(13), pp. 1798–1800.

[113] Ju, Y. S., and Goodson, K. E., 1999. "Phonon Scattering in Silicon Films with Thickness of Order 100 nm". *Applied Physics Letters,* **74**(20), pp. 3005–3007.

[114] Liu, W., and Asheghi, M., 2005. "Thermal Conduction in Ultrathin Pure and Doped Single-Crystal Silicon Layers at High Temperatures". *Journal of Applied Physics,* **98**(12), p. 123523.

[115] Ju, Y., 2005. "Phonon Heat Transport in Silicon Nanostructures". *Applied Physics Letters,* **87**(15), p. 153106.

[116] Fourier, J. B. J., and Freeman, A., 1878. *The Analytical Theory of Heat, by Joseph Fourier. Translated, with Notes, by Alexander Freeman.* The University Press, Cambridge UK.

[117] Crank, J., 1975. *The Mathematics of Diffusion*, 2nd ed. Clarendon Press, Oxford [Eng].

[118] Shanks, H. R., Maycock, P. D., Sidles, P. H., and Danielson, G. C., 1963. "Thermal Conductivity of Silicon from 300 to 1400K". *Physical Review,* **130**, June, pp. 1743–1748.