Multiscale Simulation of Phase Change Memory

Jie Liu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

M. P. Anantram, Chair

Lucien Brush

Michael Hochberg

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Multiscale Simulation of Phase Change Memory

Jie Liu

Chair of the Supervisory Committee:
Professor M. P. Anantram
Department of Electrical Engineering

This PhD thesis investigates the phase change material (PCM) based memory technology using multiscale theoretical modeling and numerical simulation. The multiscale analysis consists of three hierarchical parts: (i) the macroscopic simulation ($\approx 10^1 - 10^2$ nm and beyond), which studies the electro-thermal transport by using analytic derivation and time-dependent finite element method (TD-FEM); (ii) the mesoscopic simulation ($\approx 10^0 - 10^1$ nm), which investigates the electron transport properties by using the non-equilibrium Green's function (NEGF) method; and (iii) the microscopic simulation ($\approx 10^{-1} - 10^0$ nm), which studies the electronic structure, amorphization and crystallization processes by using density functional theory (DFT) and ab-initio molecular dynamics (AIMD).

Keywords: phase change memory (PCM), multiscale simulation, electro-thermal transport, thermal boundary resistance (TBR), electron transport, time dependent finite element method (TD-FEM), non-equilibrium Green's function (NEGF), density functional theory (DFT), and ab initio molecular dynamics (AIMD).

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

GLOSSARY

AIMD: ab initio molecular dynamics.

CG: conjugate gradient.

DFT: density functional theory.

MPI: message passing interface.

NEGF: non-equilibrium Green's function.

NW: nanowire.

PAO: pseudo-atomic orbital.

PCM: phase change material.

PL: principal layer.

RGF: recursive Green's function.

SCF: self-consistent field.

SGF: surface Green's function.

TBR: thermal boundary resistance.

TD-FEM: time-dependent finite element method.

ACKNOWLEDGMENTS

I want to express sincere appreciation to my PhD advisor Prof. M.P. Anantram for his help during my PhD research. Also, I appreciate Prof. Scott Dunham, Prof. Lucien Brush, Prof. Michael Hochberg and Prof. Xiaodong Xu for their precious time to serve in the committee.

Chapter 1

INTRODUCTION

The chalcogenide phase change material (PCM) based technologies are promising to offer the next-generation ultra-fast and ultra-dense non-volatile information storage solutions. This thesis focuses on the theoretical modeling and numerical simulation of the PCM technology, in order to provide more knowledge of the device physics, the scaling scenario, and the ultimate scaling limit of the PCM technology.

In this Chapter, the field of phase change memory is overviewed. In Section 1.1 and Section 1.2, the operation mechanisms and the history of the PCM devices are discussed, respectively. In Section 1.3, the current research status of the PCM device field is briefly reviewed. The Section 1.4 discusses some of the prominent but unsolved problems in the PCM device field, which will be investigated in this thesis. In Section 1.5, the methodology of multiscale simulation and modeling used in this thesis is introduced.

1.1 Background

The amorphous phase and the crystalline phase are the two structural phases of the PCM. The material property, e.g. optical reflectivity and electrical conductivity, can be orders of magnitude different in these two phases. This contrast can be used to store information.

1.1.1 Categorization

There are two categories of PCM devices – the optical PCM devices and the electrical PCM devices. The optical PCM devices take advantage of the optical reflectivity contrast between the two structural phases to store data. The electrical PCM devices take advantage of the electrical conductivity contrast of the two structural phases to store data. In this thesis, the electrical PCM devices are investigated.

1.1.2 Operation Mechanism

During the PCM device operations, electrical current pulses can be used to reversibly switch the PCM between its crystalline phase and amorphous phase. The resistivity of the amorphous phase is several orders of magnitude higher than that of the crystalline phase. By taking advantage of the resistance contrast, the binary bits (0 and 1) can be stored in a non-volatile manner.

One of the most widely applied PCM cell geometry is the thin film mushroom structure, as shown in Figure 1.1. During RESET and SET operations, the electrical current pulses are driven accross the chalcogenide thin film to switch it between crystalline and amorphous states.

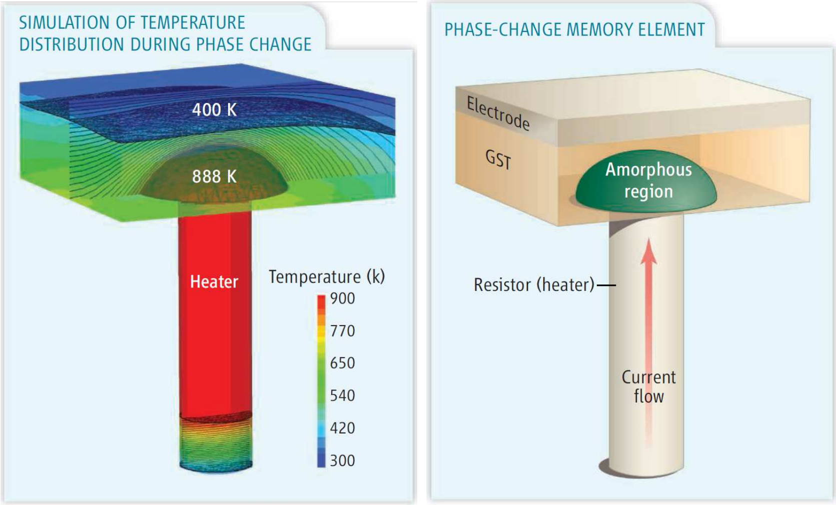


Figure 1.1: Thin film mushroom PCM cell structure [8]

To perform the RESET operation, a short current pulse of high current amplitude (pulse duration is about 10^0 ns and pulse amplitude is about $10^{-1} - 10^0$ mA) is applied to melt the crystalline PCM into liquid. After turning off the current pulse, the liquid will be quenched into amorphous phase.

To perform the SET operation, a long current pulse of low current amplitude (pulse duration is about $10^1 - 10^2$ ns and pulse amplitude is about 10^{-2} mA) is applied to anneal the amorphous PCM into crystalline. The SET operation electrical current pulse can generate moderate Joule heat, such that the temperature of the PCM active region is higher than the crystallization temperature but lower than the melting temperature. Before the SET pulse is turned off, the amorphous chalcogenide has been recrystallized.

1.2 History

The contemporary prosperity of the PCM technologies roots deeply into the wealth of PCM knowledge which has been and is being continuously developed by generations of PCM researchers in the recent one century. This section guides the reader through this period of interesting history.

1.2.1 Discovery of PCM

The history of the PCM science and technology can be traced back to the 1910s. In A.T. Waterman's papers [116, 117], it was found that the conductivity of the chalcogenide MoS_2 can be altered progressively. A.T. Waterman pointed out that the chalcogenide MoS_2 may have two phases, a high resistance phase and a low resistance phase. Furthermore, some other important traits of PCM, like the negative differential resistance and the snap-back current-voltage characteristics, were observed (Figure 1.2).

However, without the help of modern lattice structure analysis methodologies, A.T. Waterman failed to unveil the microscopic lattice structural change. Also, due to the lack of rapid melt-quench technologies, the conductivity increase could not be reversed and had to be kept permanent.

Despite these limitations, A.T. Waterman's pioneering work hallmarked the birth of a new scientific branch – phase change materials. Unfortunately, partially due to the limitation of irreversibility and the consequent lack of application, A.T. Waterman's research received little attention at that time. And the computer scientists chose to use other technologies, e.g. the mechanical punched cards (1930s to 1940s), the magnetic core memory (1950s), the magnetic hard disk, and the electronic floating gate transistors, as non-volatile memory in various computing devices.

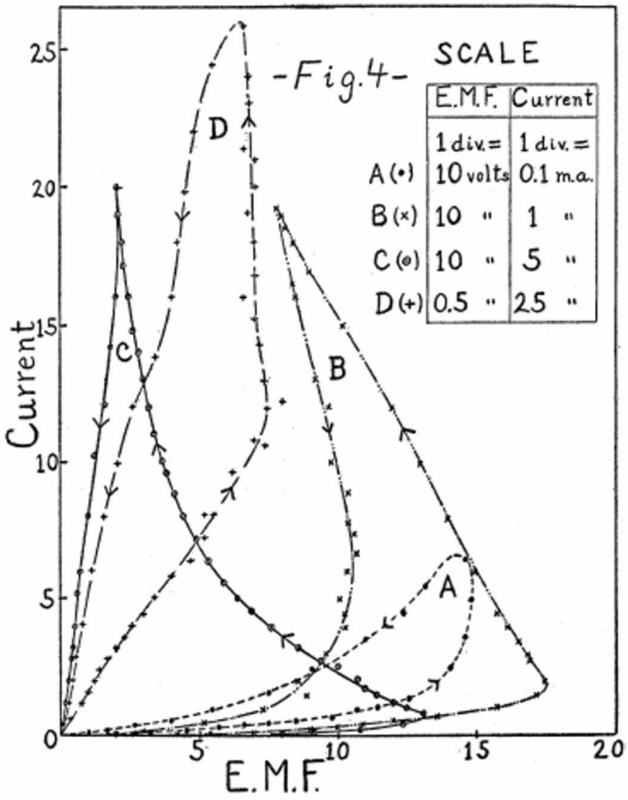


Figure 1.2: Conductance contrast and negative differential resistance of MoS_2 [116].

1.2.2 Pioneers of PCM memory technology

The bottleneck of irreversibility is broken by S. Ovshinsky in the 1960s. In S. Ovshinsky's pioneering paper [86], he described the methodology to reversibly switch the chalcogenide with 10% Ge, 12% Si, 30% As, and 48% Te between the amorphous and crystalline phases. Besides reversibility, S. Ovshinsky also showed the durability of the PCM device by cycled RESET/SET operations.

These breakthroughs in the late 1960s ignited enthusiasm for the development of PCM memory devices for the computers in the early 1970s. The most noteworthy work was the 256 bit PCM memory device fabricated by R.G. Neale, D.L. Nelson, and G.E. Moore [81]. The PCM memory device they fabricated could operate reversibly and repeatedly. Furthermore, the information could be stored in a non-volatile manner. However, the SET operation took 5 μs with 200 mA operation current and 25 V operation voltage; and the RESET operation took 10 ms with 5 mA operation current and 25 V operation voltage. These operation conditions made the newly-born PCM memory devices extremely energy hungry. Even worse, it imposed impractically high voltage and current requirements on the external circuits. As a consequence, the PCM devices were outperformed by its non-volatile memory competitors, like the magnetic hard disk and the electronic floating gate transistors (also known as Flash memory). Gradually, the ignited enthusiasm of PCM technology extinguished and the PCM memory device field received little attention in the next three decades.

1.2.3 Revival and Commercialization

The status of hibernation came to an end in the late 1990s and early 2000s, not in a sudden way, but in a gradual manner. Interestingly, the revival of the PCM memory technology was not triggered by the evolution or revolution of the PCM technolgy per se. Instead, it was largely caused by the gradual and continuous development of the integrated circuit fabrication technologies, which is exemplified by the Moore's Law.

The PCM devices back in the 1970s were energy hungry, simply because they were fabricated using μm technology. Since the device size was large, the energy consumption, the operation current, and the operation voltage were high and the operation speed was slow (see Chapter 2 for the device physics analysis that explains the underlying reasons).

The development of semiconductor fabrication technologies enables the downscaling of the PCM devices, which leads to the significant reduction of operation current, voltage, and energy. Nowadays, with the help of the state-of-art nanoscale fabrication technologies, the PCM devices can finish one RESET (SET) operation using about hundreds of μA current [123]; the energy consumption per operation can be reduced to several femto Joule [122]; the operation voltage can be reduced below 1 V [122]; and the operation time can be reduced to 0.5 ns [69].

The substantial device performance improvements have not only re-ignited intensive PCM research enthusiasm, but also paved a way for the commercialization of the PCM memory technologies. In 1999, Ovonyx was formed and the Ovonic Unified Memory (OUM) was licensed. Within several years, many semiconductor companies (Intel, Samsung, Micron, STMicroelectronics, Elpida which is now part of Micron, Macronix, Toshiba, Hitachi, Sony, etc.) launched their projects to commercialize the PCM memory technology.

1.3 Current Research Status

The commercialization of the PCM technology in the industry is facilitated by, and in turn fosters, the research efforts in both academia and industry. This section attempts to provide an overview of the current status of the PCM research. It worth noting that the overview is by no means comprehensive or exhaustive, because the information contained in the PCM literature is massive. The overview only reflects the author's partial knowledge of this field, as well as the relevance of the overviewed materials to this thesis.

In Section 1.3.1, the three types of most widely used PCM devices are briefly introduced. In Section 1.3.2, the scaling research pertaining to the PCM technology scalability and the PCM device performance scaling scenario is reviewed. In Section 1.3.3, some of the established device physics research methodologies and results are discussed.

1.3.1 Diverse Device Design

Different types of PCM devices have been designed, for both research purposes and application purposes. These different designs are briefly reviewed in this section.

1.3.1.1 Thin Film Device

The thin film PCM device as schematically shown in Figure 1.1 is one of the most popular PCM device structures, due to its simplicity. In thin film PCM devices, the RESET operation generates Joule heat to melt the active PCM thin film region, in order to transform the PCM from crystalline phase to amorphous phase. The shape of the amorphous region resembles the mushroom, so it is also referred to as mushroom PCM devices.

1.3.1.2 Confined Structure Device

The simple thin film PCM devices suffer from high energy consumption, since they cannot confine the heat in an effective way. In order to improve the device efficiency, a large number of confined PCM devices have been proposed. An example of the confined PCM device structure is shown in Figure 1.3 [14]. It can be seen that the PCM active region is surrounded by thermal insulating material to encapsulate thermal energy.

Most of these confined PCM devices are purposefully designed to lower the energy consumption by confining either the active PCM region [20, 60, 61] or the electrodes [121]. To achieve the goal, they use various device structures like cross-spacers [113], edges [124], trenches [99], pores [12], ring-shape contacts [121], and dash-type confinements [29], etc.

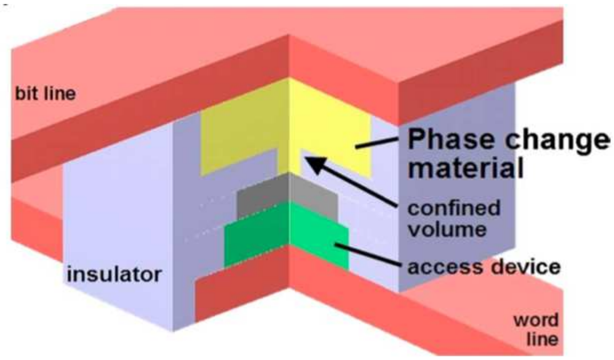


Figure 1.3: Confined PCM cell structure [14]

1.3.1.3 Nanowire Device

Different from the two dimentional thin film structure, another type of PCM device uses the one dimensional PCM nanowire to fabricate PCM devices, as shown in Figure 1.4[108, 127, 107, 105, 128]. When the PCM nanowire is exposed to vacuum, the interfacial energy can decrease the melting temperature, leading to possible reduction of the operation energy (see Chapter 2 for explanation of the underlying physics).

Also, the PCM nanowires have been used to investigate the microscopic structural physics [79, 43, 26, 38, 44, 41, 64, 63, 10, 42, 62], because the one dimensional geometry facilitates the microscope observation.

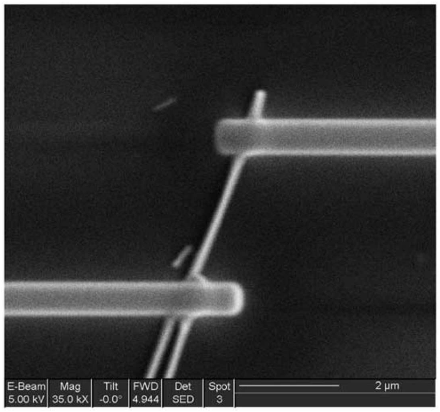


Figure 1.4: Exposed Nanowire PCM cell structure [128]

1.3.2 Scaling Research

Good scalability is one of the most attractive merits of the PCM memory technology. Downscaling not only helps increase the data density, but also helps improve the device performance.

In order to uncover the possible obstacles of PCM device downscaling, a lot of research bas been done. Basically, there are two categories: (1) the PCM scalability research (discussed in Section 1.3.2.1) and (2) the PCM scaling scenario research (i.e. discussed in Section 1.3.2.2). The first category of research studies how small the PCM devices can be made. The second category of research studies how will the device performance change if the device is downscaled.

1.3.2.1 Scalability Research

The scalability of the PCM technology hinges on three factors: (1) the ability to perform write operations, (2) the ability to perform read operations, and (3) the ability to maintain amorphous phase in the PCM nanostructures over a certain period of time (say several years).

As discussed in Section 1.1, the PCM devices rely on repeated reversible switching between the amorphous phase and the crystalline phase of the PCM to perform write operations. While the capabilities of repeated reversible phase switching is not a problem in the bulk PCM, it might be a problem in PCM nanostructures. To address this issue, many scaling experiments have been carried out [101, 14, 39, 17, 16, 120, 119, 92, 95, 93, 23]. It was found that, in sub-10 nm scale, some material properties become size-dependent, which might hinder aggressive scaling. For example, the crystallization temperature rises, sometimes even approaching the melting temperature, when the PCM nanostructure is scaled to sub-10 nm [95, 96, 17]. However, this problem has been solved by the minimizing stress, which can help to maintain the crystallization temperature of the 2 nm PCM nanostructures at the bulk value [102]. Now, it is known that the phase change properties can be maintained in ultra-scaled sub-2 nm PCM nanostructures, demonstrating extremely attractive scaling scenario of the so-called ultra-scaled PCM technology.

Besides keeping phase change properties, the ultra-scaled PCM nanostructures also should be able to retain device endurance and sufficient large electrical conductance contrast, in order to enable reliable read operations. To study these issues, device scaling experiments have been performed, which show that the PCM devices can operate endurably when the device size is reduced to the sub-10 nm scale [49]. It is also shown that 6 nm PCM nanostructures still can retain a two order of magnitude ON/OFF ratio, enabling reliably read operations [49].

The ability to keep amorphous PCM endurable is largely determined by the scaling scenario of the SET operation, which has been investigated by using nanowire PCM devices [129] and bridge PCM devices [54, 55, 24, 23]. It has been found that, though the threshold switching of the bulk amorphous PCM (about tens of nm or larger) obeys a constant field scaling law [54, 55, 24, 23], the threshold switching of amorphous PCM obeys a constant voltage scaling law in the sub-10 nm scale [129]. This change is found to be indispensible to enable the ultra-scaled sub-10 nm PCM devices by ensuring the stablity of the ultra-scaled amorphous PCM nanostructures [129].

In a word, the state-of-art scaling experiments have unveiled extremely promising scalability of the PCM technology. Now it is well known that the PCM devices can be downscaled to sub-10 nm.

1.3.2.2 Scaling Performance Research

Besides the scalability, researchers are also interested in understanding how the device performance changes during scaling. The pioneering experimental research on this topic unveiled the fact that the downscaling of the PCM devices helps reduce operation energy, operation current, operation voltage, and operation time [90].

Though the quantitative scaling laws of various device performance metrics have been experimentally determined in 2003 [90], they were not deeply understood at that time. In 2007, the underlying device physics of these quantitative laws was explained by using analytic and numerical analysis of the governing electro-thermal transport partial differential equations [48].

Driven by the promising device performance gain in scaling, a lot of efforts have been devoted in order to break the PCM device performance record. To mention a few prominent examples: by using nanoscale carbon nanotubes as electrodes, it has been shown that the PCM device energy consumption can be reduced to several femto Joule per operation [122]; by agressively downscaling the PCM nanostructure, it has been shown that both the operation speed and the device endurance can be boosted significantly [115]; and by applying pre-annealing techniques, the nanoscale PCM device operation time can be reduced to 0.5 ns [69].

In summary, it is now well known that downscaling can significantly improve various device performance metrics of PCM devices; and state-of-art PCM device performance is very promising for large scale application and commercialization.

1.3.3 Device Physics Research

The underlying physics of the PCM devices has attracted considerable research attention, especially in the past decade. Various experimental and theoretical approaches have been applied, in order to understand the underlying physical mechanisms that govern the PCM device operation.

Two of most fundamental physical processes in the PCM devices are the armophization of crystalline PCM (RESET operation) and the crystallization of amorphous PCM (SET operation). To achieve amorphization and crystallization, one has to rely on electrical and thermal transport to deliver operation energy. Furthermore, the read operation hinges on the eletrical transport properties of the PCM.

Therefore, amorphization, crystallization, and electrothermal transport are the three most important physical processes for the the operation of PCM devices. Since the efforts to understand these three key processes are central topics of PCM device physics research, the current status of modeling and simulation research used to understand them are reviewed in Section 1.3.3.1, Section 1.3.3.2, and Section 1.3.3.3, respectively.

1.3.3.1 Amorphization

During the RESET operation of the PCM devices, an electrical pulse of large magnitude and short duration is applied to generate Joule heat and melt the crystalline PCM. After the pulse is turned off, the liquid PCM is quenched into amorphous state. This melt-quench amorphization process can be modeled at the atomistic scale by using ab initio molecular dynamics (AIMD) simulations, which is based on density functional theory (DFT).

DFT and AIMD simulation of the PCM melt-quench process was pioneered by the research of three groups (Z.M. Sun and R. Ahuja et. al. [109]; J. Akola and R. Jones et. al. [2]; and S.R. Elliot et. al. [31]) in 2007 and 2008. With the assistance of powerful microscopic modeling techniques, there has been a significant increase of tomistic scale knowledge of the amorphization process and the amorphous phase of the PCM [53, 32, 6, 4, 5, 3, 111, 73, 50].

By using DFT and AIMD simulations, it has been found that amorphization can be achieved by destroying week bonds in the crystalline PCM [50]. As a consequence, though the amorphous PCM lacks long range order, its local short range lattice contains a lot of basic building blocks (like four-fold rings and eight-atom cubes) of the crystalline phase [32, 6, 4, 5, 3, 50]. The presence of these blocks facilitates the recrystallization of the amorphous PCM.

In the DFT and AIMD simulations, periodic boundary conditions (PBC) have to be applied at the boundaries of the simulation supercells, due to the requirements of the simulation algorithm. The application of PBC means that there is artifical long range order in the amorphous simulation system. Interestingly, despite this drawback, the DFT and AIMD simulations still can yield decent physical insight and good agreement with the experimental measurements.

1.3.3.2 Crystallization

Compared to the melt-quench AIMD simulations, the AIMD simulations of the PCM crystallization process is much more computationally demanding, since the crystallization time (nanosecond level) is much longer than the amorphization time (about $10^1 - 10^2$ ps).

To simulate the crystallization process using the AIMD, the melt-quench AIMD simulations are first performed to obtain the amorphous PCM models. Then, the annealing AIMD simulations are run in a brute-force manner for a long time until the crystallization is done. The brute-force AIMD simulation research [69, 65, 32, 31, 45, 110] has provided rich atomistic scale knowledge about the recrystallization process of amorphous PCM. It has been found that the presence of the four-fold rings, eight-atom cubes, and small crystalline seeds facilitates the fast recrystallization of amorphous PCM.

Besides offering scientific understanding, the brute-force anneal AIMD also helps the development of new technology and the improvement of PCM device performance. One of the most prominent examples is the use of AIMD simulations for the development of ultra-fast PCM devices by applying pre-annealing pulses [69].

1.3.3.3 Electrothermal Transport

As the fundamental governing physical process of both the write operations and the read operations, the electrothermal transport in the PCM devices has attacted intensive research attention. A lot of modeling and simulation methodologies have been applied to investigate eletrothermal transport. To mention a few prominent examples: (1) finite element method (FEM) numerical simulations [22, 48, 47] have been applied to investigate electrothermal transport under the diffusive transport assumption, in order to quantitatively obtain PCM device performance metrics and to investigate PCM device scaling scenario; (2) Monto Carlo simulations [15] have been used to understand the peculiar negative differential resistance and S-shape snap-back current voltage curve [34]; (3) analytic analysis and modeling have been applied to explain the electron transport properties of amorphous PCM [58, 35, 36, 37].

1.4 Prominent Unsolved Problems

Though the PCM device technology has attracted intensive research attention and efforts, some prominent problems remain unsolved, which will be investigated in this thesis:

1. The existing PCM device operation schemes exhibit diversity. To perform RESET/SET operations, some research uses constant current pulses [57, 128, 62, 63, 107, 105, 41, 46, 125, 70]; some research uses constant voltage pulses [130, 127, 75, 64, 106, 44, 129]; and some research use constant field pulses [57, 129]. However, a unified understanding of the these operation schemes and their different scaling scenarios is still missing. In order to better design PCM devices and to better choose the most appropriate operation scheme in different situations, it is important to investigate the the underlying device physics of these operation schemes. Chapter 2 is devoted to investigate these problems by using both analytic derivation and time-dependent finite element method (TD-FEM) simulations.

2. While the crystalline PCM exhibits ohmic electron transport behavior, the amorphous PCM shows fairly interesting sub-threshold electron transport properties. For instance, when the bias is small, the current-voltage curve of the amorphous PCM is largely linear; when the bias is large, the shape of the current-voltage curve of amorphous PCM becomes exponential. While these electron transport properties were experimentally discovered a long time ago, an understanding of the underlying atomistic scale physics is still missing, especially for ultra-scaled PCM devices. In order to better understand the underlying physics of these phenomena, Chapter 3 investigates the sub-threshold electron transport properties by using the non-equilibrium Green's function (NEGF) simulations.

3. It is known that the PCM device technology has an extremely promising scaling scenario. State-of-art PCM device scaling experiments have shown that the PCM devices scaled down to 6 nm still can operate stably with roughly a two orders of magnitude ON/OFF ratio [49]. Interestingly, though the ON/OFF ratio of the ultra-scaled PCM devices is much smaller than the bulk PCM ON/OFF ratio (about four to five orders of magnitude), it still can support reliable read operations. Furthermore, while the PCM devices scaled down to 6 nm still show a large ON/OFF ratio, the ultimate scaling limit of the PCM technology is still an open question. To explore the ultimate scalability of the PCM technology, it is important to find out what is the smallest feature size of a PCM device that can still support a sufficiently large ON/OFF ratio to allow reliable read operations. In Chapter 3, the ultimate scaling limit of the PCM technology and the underlying device physics are investigated.

4. It is known that the scattering of electrons by phonons plays an important role in determining the electron transport properties in the PCM devices. Though phenomenological explanations have been published [15, 58, 35, 36, 37], the quantitative atomistic scale analysis is still missing. In Chapter 3, the influence of the electron-phonon scattering is investigated by using the NEGF simulations, to offer deeper knowledge of this underlying physical process.

5. The PCM scaling experiments have shown that the ultra-scaled sub-2 nm PCM nanostructures still preserve bulk phase change properties [95]. The underlying reason is still not clearly understood. Density functional theory (DFT), ab initio molecular dynamics (AIMD) simulations and thermodynamic theory are applied in Chapter 4, in order to develop an understanding of the physical reasons for the superb scalability of the PCM.

6. Currently, the different roles of the various physical quantities (Gibbs free energy density, interfacial energy density, elastic energy, specific heat capacity, critical formation energy, critical nuclei radius, etc.) that determine the crystallization properties of amorphous PCM are still not clear. Also, the temperature dependence of these phys-

ical quantities is still unknown. They are calculated and quantitatively analyzed in Chapter 4.

1.5 Multiscale Simulation

The governing physical laws of the RESET and SET operations span multiple spatial scales. To capture the multiscale nature of PCM device operations, the hierarchical multiscale simulation methodology used in this thesis is summarized in Table 1.1.

Table 1.1: Multiscale simulation hierarchy

	Microscopic	**Mesoscopic**	**Macroscopic**
spatial scale	$\approx 10^{-1} - 10^{0}$ nm	$\approx 10^{0} - 10^{1}$ nm	$\approx 10^{1} - 10^{2}$ nm
temporal scale	$\approx 10^{-2} - 10^{0}$ ns	$\approx 10^{0} - 10^{1}$ ns	$\approx 10^{1} - 10^{2}$ ns
physics described	electronic structure, phonon, bond forming and breaking	deep nm-scale electron transport	electro-thermal transport
governing equation	Schrödinger equation, Newton equation	Green's function	heat transfer equation, Laplace equation
methodology	DFT, AIMD	NEGF	TD-FEM
tools	SIESTA, Matlab	C, MPI, Matlab	FreeFEM, Matlab

1.5.1 Macroscopic Simulation

The macroscopic analysis ($\approx 10^{1} - 10^{2}$ nm and beyond) is presented in Chapter 2. It aims at analyzing the device-level intra-cell and inter-cell electro-thermal transport properties at spatial dimensions larger than tens of nm. The governing heat transfer equation and Laplace equation are solved by TD-FEM to obtain the crucial physical quantities (e.g., operation voltage bias, electrical current, and electrical field) that determine device performance. Based on the numerical TD-FEM results and analytical derivation, the PCM scaling scenarios under three operation schemes (constant field, constant voltage, and constant current) are analyzed and compared. Analysis of a cylindrical pore-like geometry PCM is used as an example. The widely-used thin film PCM and novel nanowire PCM are briefly compared. The macroscopic analysis indicates that more than 90% of RESET energy is wasted as thermal flux into the substrate and the electrode in the current PCM geometry design.

1.5.2 Mesoscopic Simulation

The mesoscopic analysis ($\approx 10^0 - 10^1$ nm) is presented in Chapter 3. It is appropriate at the intra-cell sub-10 nm scale for the study of mesoscopic transport properties and uses NEGF simulations. In macroscopic simulations, diffusive electro-thermal transport theory is applied. However, in the sub-10 nm, the carrier transport is largely elastic. So, Green's function method is applied to analyze the electron transport properties of the ultrascaled PCM nanostructure. The system used in the analysis is a GeTe ultrathin film sandwiched between two TiN electrodes. It is found that the 40 Åthick (13 atomic layer) GeTe ultrathin films can still show a two order of magnitude ON/OFF ratio. However, if the ultrathin film thickness is further reduced, the ON/OFF ratio drops rapidly. When the film is thinner than 33 Å, the ON/OFF ratio becomes smaller than 10, making it difficult to reliably perform read operation. This sets up an ultimate scaling limit for the PCM technology. It is found that the metal induced gap states (MIGS) inside the amorphous GeTe is responsible for the reduction of the ON/OFF ratio.

1.5.3 Microscopic Simulation

The microscopic analysis ($\approx 10^{-1} - 10^0$ nm) is presented in Chapter 4. The crystallization properties of the prototypical binary PCM GeTe system are investigated by using thermodynamic theory, DFT, and AIMD simulations. The temperature dependence of the physical quantities that determine the crystallization properties is calculated and analyzed. It is shown that the critical nuclei radius of the crystalline cluster is smaller than 1.4 nm when the annealing temperature is lower than 600 K, indicating an extremely promising scaling scenario. The analysis reveals that the elastic energy, which is largely ignored in the existing PCM crystallization studies, plays an important role in determining various crystallization properties and the ultimate scaling limit of the PCM. By omitting the influence of elastic energy, the critical formation energy (critical nuclei radius) will be underestimated by 41.7% (22.4%) and the nucleation rate will be overestimated by 74.2% when the annealing temperature is 600 K. Besides, the microscopic DFT and AIMD simulations generate ab initio Hamiltonian and overlap matrices, which are used as inputs in mesoscopic electron transport simulations.

Chapter 2

MACROSCOPIC SIMULATION

In this chapter, the macroscopic electro-thermal transport processes in PCM devices are investigated, by using analytical derivation and time-dependent finite element method (TD-FEM) numerical simulation [66, 68].

2.1 Device Geometry

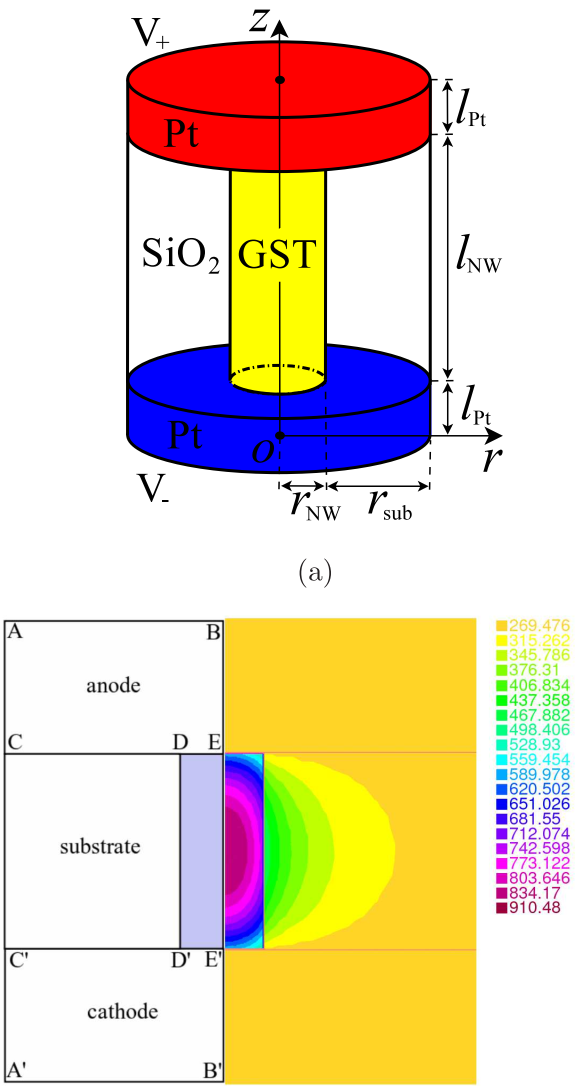
Figure 2.1 shows the PCM device model geometry used in the macroscopic simulation. The model can be considered as a PCM nanowire (NW) since it is axisymmetric where the horizontal (vertical) direction is the radial (axial) direction. The PCM active region (region DEE′D′) is wrapped by the substrate (region CDD′C′). At the two ends of PCM, there are two electrodes (region ABEC and A′B′E′C′). During the RESET operation, an electrical pulse is driven through the PCM by applying a bias between the two electrodes. The Joule heat generated by the pulse could melt the crystalline PCM. After switching off the pulse, the melted PCM will be quenched into amorphous state. The size values in Figure 2.1 are listed in Table 2.1.

Table 2.1: problem geometry

	DE	EE′	CD	BE and B′E′
meaning	NW radius	NW length	substrate thickness	electrode length
value (nm)	r (variable)	l (variable)	400	200

2.2 Assumptions and Approximations

In the macroscopic analysis, the following assumptions and approximations are made:



(a)

(b)

Figure 2.1: (a) Schematic axisymmetric cylindrical problem geometry and (b) snapshot of the simulated temperature distribution, where the region DEE′D′ is the phase change material (GeTe), the region CDD′C′ is the substrate (SiO_2), and the region ABEC and A′B′E′C′ are the electrodes (Pt).

1. The PCM devices analyzed are about tens of nm or even larger. In this spatial scale, the device size is much larger than the carrier mean free path (MFP) in typical PCM

[59, 80]. So the electro-thermal transport is assumed to be diffusive.

2. The phenomenological coefficients (i.e. C_v, κ, and ρ, etc.) in the governing electro-thermal transport equations are assumd to be temperature independent. This is justified since the temperature dependence of them does not have dominant influence on the analysis results [47].

3. The phenomenological coefficients (i.e. C_v, κ, and ρ, etc.) in the governing electro-thermal transport equations are assumed to be fixed during scaling. In the spatial scale of our interest (about tens of nm or larger), the diffusive nature of the transport justifies this assumption [80, 129].

Since the transport is assumed be diffusive in the macroscopic simulation, the results and conclusions made in this chapter are not applicable for the aggressively scaled sub-10 nm PCM devices, since the transport is no longer diffusive in this scale [129]. In Chapter 3, NEGF is applied to investigate the transport in agressively scaled PCM devices.

2.3 Governing Equations

As mentioned in Section 2.2, the carrier MFP in typical PCM is much smaller than the device size of our interest here, so the diffusive electro-thermal transport equations can be used to describe the PCM device physics. Mathematically, the thermal transport process is governed by the heat transfer partial differential equation

$$C_v \frac{\partial T}{\partial t} = \nabla \cdot (\kappa \nabla T) + \rho_e |\mathbf{J}_e|^2 \tag{2.1}$$

and the electrical transport process is governed by the Laplace equation

$$\nabla \cdot (\sigma \nabla \phi) = 0 \tag{2.2}$$

where the current density

$$\mathbf{J}_e = -\sigma \nabla \phi \tag{2.3}$$

T is the temperature, $\mathbf{J}_e$ is the electrical current density, and ϕ is the electrical potential. The C_v, κ, ρ_e, and σ are material specific parameters, as listed in Table 2.2. On the right hand side of Equation 2.1, the first term means the thermal diffusion and the second term means the Joule's heating.

Table 2.2: Values of physical parameters

	heat capacity c_v (J/m^3·K)	thermal conductivity κ (W/m·K)	electrical resistivity ρ_e (Ω·m)	latent heat L(J/m^3)	melting temperature T_m(K)
GeTe	1.60×10^6	4.4	4.78×10^{-4}	1.45×10^9	998
$Ge_2Sb_2Te_5$	1.30×10^6	0.46	4.16×10^{-4}	1.121×10^9	905
SiO_2	1.94×10^6	1.4	1×10^{16}	–	–
Pt	2.84×10^6	71.6	1×10^{-7}	–	–

2.4 Analytic Scaling Analysis

In this section, the isotropic scaling (all three dimensions of the NW PCM are scaled by the same factor k ($k > 1$)) is analyzed. This means that both the NW radius and NW length are reduced by k times simultaneously.

2.4.1 Analysis Methodology

The isotropic scaling changes the cylindrical spatio-temporal coordinate from (r, z, t) to (r_s, z_s, t_s)[1], where

$$r_s = r/k \qquad z_s = z/k \qquad t_s = \alpha t \tag{2.4}$$

and coefficient α is to be determined. Due to the Ohm's law, the NW electrical resistance scales as

$$R_s = Rk \tag{2.5}$$

---

[1] Hereafter, the variables with (without) subscript s denote physical quantities after (before) scaling.

The geometric change due to the isotropic scaling will exert impacts on some physical quantities (e.g., T, ϕ, and $\mathbf{J}_e$). So, the device performance will be changed by scaling. The relation between the device performance and the isotropic scaling can be studied by analyzing the governing PDEs (i.e., Equation 2.1 and Equation 2.3). To start the analysis, Equation 2.1 and Equation 2.3 are expressed as

$$C_v \frac{\partial T}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(\kappa r \frac{\partial T}{\partial r} \right) + \frac{\partial}{\partial z} \left(\kappa \frac{\partial T}{\partial z} \right) + \rho J^2 \tag{2.6}$$

and

$$-\frac{\partial \phi}{\partial r} = \rho J_r \qquad -\frac{\partial \phi}{\partial z} = \rho J_z \tag{2.7}$$

in the cylindrical coordinates[2], where

$$J = |\mathbf{J}_e| = \sqrt{J_r^2 + J_z^2} \tag{2.8}$$

and J_r (J_z) is the radial (axial) component of $\mathbf{J}_e$ (for convenience, the subscript e is suppressed).

The derivations of the isotropic scaling properties of NW PCM's RESET performance in the three device operation schemes (constant field, constant voltage, and constant current) are presented below.

2.4.1.1 Constant Field Scaling

If the device operation electric field is kept constant during device scaling, one could have

$$E_{rs}(r_s, z_s) = E_r(r, z), \quad E_{zs}(r_s, z_s) = E_z(r, z) \tag{2.9}$$

So,

$$J_{rs}(r_s, z_s) = J_r(r, z), \quad J_{zs}(r_s, z_s) = J_z(r, z) \tag{2.10}$$

should be valid since the Ohm's law

$$E = \rho J \tag{2.11}$$

---

[2] Hereafter, the variables with subsript r (z) denote radial (axial) components of the corresponding physical quantities.

is valid both before and after scaling. As the cross-section scaling factor is $1/k^2$, the current scales as

$$I_s = I/k^2 \tag{2.12}$$

Because

$$-\frac{\partial \phi}{\partial r} = E_r, \quad -\frac{\partial \phi}{\partial z} = E_z \tag{2.13}$$

should be valid both before and after scaling, one could have

$$-\frac{\partial \phi_s(r_s, z_s)}{\partial r_s} = E_{rs}(r_s, z_s) = E_r(r, z) = -\frac{\partial \phi(r, z)}{\partial r} = -\frac{\partial \phi(r, z)}{\partial r_s}\frac{1}{k} \tag{2.14}$$

$$-\frac{\partial \phi_s(r_s, z_s)}{\partial z_s} = E_{zs}(r_s, z_s) = E_z(r, z) = -\frac{\partial \phi(r, z)}{\partial z} = -\frac{\partial \phi(r, z)}{\partial z_s}\frac{1}{k} \tag{2.15}$$

Therefore,

$$\phi_s(r_s, z_s) = \phi(r, z)/k \tag{2.16}$$

should be valid.

Due to Equation 2.10 and the fact that Equation 2.6 holds both before and after scaling, one can obtain

$$\begin{aligned}
&C_v \frac{\partial T_s(r_s, z_s, t_s)}{\partial t_s} - \frac{1}{r_s}\frac{\partial}{\partial r_s}\left(\kappa r_s \frac{\partial T_s(r_s, z_s, t_s)}{\partial r_s}\right) - \frac{\partial}{\partial z_s}\left(\kappa \frac{\partial T_s(r_s, z_s, t_s)}{\partial z_s}\right) - \rho J_s^2(r_s, z_s) \\
=&C_v \frac{\partial T(r, z, t)}{\partial t} - \frac{1}{r}\frac{\partial}{\partial r}\left(\kappa r \frac{\partial T(r, z, t)}{\partial r}\right) - \frac{\partial}{\partial z}\left(\kappa \frac{\partial T(r, z, t)}{\partial z}\right) - \rho J^2(r, z) \\
=&C_v \frac{\partial T(r, z, t)}{\partial (t_s/\alpha)} - \frac{1}{(kr_s)}\frac{\partial}{\partial (kr_s)}\left(\kappa (kr_s) \frac{\partial T(r, z, t)}{\partial (kr_s)}\right) - \frac{\partial}{\partial (kz_s)}\left(\kappa \frac{\partial T(r, z, t)}{\partial (kz_s)}\right) - \rho J_s^2(r, z) \\
=&C_v \frac{\partial T(r, z, t)}{\partial t_s}\alpha - \frac{1}{r_s}\frac{\partial}{\partial r_s}\left(\kappa r_s \frac{\partial T(r, z, t)}{\partial r_s}\right)\frac{1}{k^2} - \frac{\partial}{\partial z_s}\left(\kappa \frac{\partial T(r, z, t)}{\partial z_s}\right)\frac{1}{k^2} - \rho J_s^2(r, z) \\
=&0
\end{aligned} \tag{2.17}$$

Thus, equalities

$$T_s(r_s, z_s, t_s) = T(r, z, t)/k^2 \tag{2.18}$$

and $\alpha = 1/k^2$, i.e.

$$t_s = t/k^2 \tag{2.19}$$

must hold.

The scaling factors of I, ϕ, and t are $1/k^2$, $1/k$, and $1/k^2$, respectively. Therefore, the power

$$P = I\Delta\phi \tag{2.20}$$

and the energy

$$Q = I\Delta\phi\Delta t \tag{2.21}$$

scale as

$$P_s = P/k^3 \tag{2.22}$$

and

$$Q_s = Q/k^5 \tag{2.23}$$

2.4.1.2 Constant Voltage Scaling

If the device operation electrical voltage is kept constant during scaling, one could have

$$\phi_s(r_s, z_s) = \phi(r, z) \tag{2.24}$$

Then,

$$\rho J_{rs}(r_s, z_s) = -\frac{\partial \phi_s(r_s, z_s)}{\partial r_s} = -\frac{\partial \phi(r, z)}{\partial r_s} = -\frac{\partial \phi(r, z)}{\partial r}k = \rho J_r(r, z)k \tag{2.25}$$

$$\rho J_{zs}(r_s, z_s) = -\frac{\partial \phi_s(r_s, z_s)}{\partial z_s} = -\frac{\partial \phi(r, z)}{\partial z_s} = -\frac{\partial \phi(r, z)}{\partial z}k = \rho J_z(r, z)k \tag{2.26}$$

Therefore,

$$J_{rs}(r_s, z_s) = J_r(r, z)k \quad J_{zs}(r_s, z_s) = J_z(r, z)k \tag{2.27}$$

So, the current scales as

$$I_s = I/k \tag{2.28}$$

as the cross-section scaling factor is $1/k^2$; and the electric field scales as

$$E_{rs}(r_s, z_s) = E_r(r, z)k \quad E_{zs}(r_s, z_s) = E_z(r, z)k \tag{2.29}$$

due to Equation 2.11.

Furthermore, due to the fact that Equation 2.6 holds both before and after scaling, one can obtain

$$\begin{aligned}
&C_v \frac{\partial T_s(r_s, z_s, t_s)}{\partial t_s} - \frac{1}{r_s}\frac{\partial}{\partial r_s}\left(\kappa r_s \frac{\partial T_s(r_s, z_s, t_s)}{\partial r_s}\right) - \frac{\partial}{\partial z_s}\left(\kappa \frac{\partial T_s(r_s, z_s, t_s)}{\partial z_s}\right) - \rho J_s^2(r_s, z_s) \\
=&C_v \frac{\partial T(r, z, t)}{\partial t} - \frac{1}{r}\frac{\partial}{\partial r}\left(\kappa r \frac{\partial T(r, z, t)}{\partial r}\right) - \frac{\partial}{\partial z}\left(\kappa \frac{\partial T(r, z, t)}{\partial z}\right) - \rho J^2(r, z) \\
=&C_v \frac{\partial T(r, z, t)}{\partial (t_s/\alpha)} - \frac{1}{(kr_s)}\frac{\partial}{\partial (kr_s)}\left(\kappa (kr_s) \frac{\partial T(r, z, t)}{\partial (kr_s)}\right) - \frac{\partial}{\partial (kz_s)}\left(\kappa \frac{\partial T(r, z, t)}{\partial (kz_s)}\right) - \rho J_s^2(r, z)\frac{1}{k^2} \\
=&C_v \frac{\partial T(r, z, t)}{\partial t_s}\alpha - \frac{1}{r_s}\frac{\partial}{\partial r_s}\left(\kappa r_s \frac{\partial T(r, z, t)}{\partial r_s}\right)\frac{1}{k^2} - \frac{\partial}{\partial z_s}\left(\kappa \frac{\partial T(r, z, t)}{\partial z_s}\right)\frac{1}{k^2} - \rho J_s^2(r, z)\frac{1}{k^2} \\
=&0
\end{aligned} \tag{2.30}$$

Thus, equalities

$$T_s(r_s, z_s, t_s) = T(r, z, t) \tag{2.31}$$

and $\alpha = 1/k^2$, i.e.

$$t_s = t/k^2 \tag{2.32}$$

must hold.

Due to Equation 2.20, Equation 2.21, and the fact that the scaling factors of I, ϕ, and t are $1/k$, 1, and $1/k^2$, respectively, P and Q scale as

$$P_s = P/k \tag{2.33}$$

and

$$Q_s = Q/k^3 \tag{2.34}$$

2.4.1.3 Constant Current Scaling

If the device operation electrical current is kept constant during scaling, the current density scaling obeys

$$J_{rs}(r_s, z_s) = J(r, z)k^2 \quad J_{zs}(r_s, z_s) = J(r, z)k^2 \tag{2.35}$$

as the PCM cross-section is scaled by factor $1/k^2$ when the PCM radius is scaled by factor $1/k$.

Thus, one can get

$$I_s = I \tag{2.36}$$

as the PCM cross-section scaling factor is $1/k^2$.

Due to Equation 2.11, one could have

$$E_{rs}(r_s, z_s) = E(r, z)k^2 \quad E_{zs}(r_s, z_s) = E(r, z)k^2 \tag{2.37}$$

and

$$-\frac{\partial \phi_s(r_s, z_s)}{\partial r_s} = \rho J_{rs}(r_s, z_s) = \rho J_r(r, z)k^2 = -\frac{\partial \phi(r, z)}{\partial r}k^2 = -\frac{\partial \phi(r, z)}{\partial r_s}k \tag{2.38}$$

$$-\frac{\partial \phi_s(r_s, z_s)}{\partial z_s} = \rho J_{zs}(r_s, z_s) = \rho J_z(r, z)k^2 = -\frac{\partial \phi(r, z)}{\partial z}k^2 = -\frac{\partial \phi(r, z)}{\partial z_s}k \tag{2.39}$$

Therefore,

$$\phi_s(r_s, z_s) = \phi(r, z)k \tag{2.40}$$

must be valid.

Due to the fact that Equation 2.6 holds both before and after scaling, one can obtain

$$\begin{aligned}
&C_v \frac{\partial T_s(r_s, z_s, t_s)}{\partial t_s} - \frac{1}{r_s}\frac{\partial}{\partial r_s}\left(\kappa r_s \frac{\partial T_s(r_s, z_s, t_s)}{\partial r_s}\right) - \frac{\partial}{\partial z_s}\left(\kappa \frac{\partial T_s(r_s, z_s, t_s)}{\partial z_s}\right) - \rho J_s^2(r_s, z_s) \\
=&C_v \frac{\partial T(r, z, t)}{\partial t} - \frac{1}{r}\frac{\partial}{\partial r}\left(\kappa r \frac{\partial T(r, z, t)}{\partial r}\right) - \frac{\partial}{\partial z}\left(\kappa \frac{\partial T(r, z, t)}{\partial z}\right) - \rho J^2(r, z) \\
=&C_v \frac{\partial T(r, z, t)}{\partial (t_s/\alpha)} - \frac{1}{(kr_s)}\frac{\partial}{\partial (kr_s)}\left(\kappa (kr_s) \frac{\partial T(r, z, t)}{\partial (kr_s)}\right) - \frac{\partial}{\partial (kz_s)}\left(\kappa \frac{\partial T(r, z, t)}{\partial (kz_s)}\right) - \rho J_s^2(r, z)\frac{1}{k^4} \\
=&C_v \frac{\partial T(r, z, t)}{\partial t_s}\alpha - \frac{1}{r_s}\frac{\partial}{\partial r_s}\left(\kappa r_s \frac{\partial T(r, z, t)}{\partial r_s}\right)\frac{1}{k^2} - \frac{\partial}{\partial z_s}\left(\kappa \frac{\partial T(r, z, t)}{\partial z_s}\right)\frac{1}{k^2} - \rho J_s^2(r, z)\frac{1}{k^4} \\
=&0
\end{aligned} \tag{2.41}$$

Thus, equalities

$$T_s(r_s, z_s, t_s) = T(r, z, t)/k^2 \tag{2.42}$$

and $\alpha = 1/k^2$, i.e.

$$t_s = t/k^2 \tag{2.43}$$

must hold.

Due to Equation 2.20, Equation 2.21, and that I, ϕ, and t is scaled by1, k, and $1/k^2$, respectively, P and Q scale as

$$P_s = Pk \tag{2.44}$$

and

$$Q_s = Q/k \tag{2.45}$$

2.4.2 Analysis Results

2.4.2.1 Scaling Scenarios

The PCM device operation schemes exhibit diversity. To perform RESET operations, the existing PCM research uses either constant current pulses [57, 128, 62, 63, 107, 105, 41, 46, 125, 70] or constant voltage pulses [130, 127, 75, 64, 106, 44, 129]. Also, previous studies have shown that the threshold switching of PCM SET operation demonstrates constant field scaling properties [57, 129]. While these existing studies aim at researching the PCM properties in various experimental conditions, the focus here is providing a general theoretical analysis and comparison of these operation schemes during scaling.

The scaling factors of the physical quantities derived in Section 2.4.1 are summarized in Table 2.3.

From Table 2.3, one can conclude:

1. The scaling factors for time under three operation schemes all are $1/k^2$, indicating that the temperature changing rate in time is k^2 times faster in the scaled PCM device than in the original one [48]. Therefore, downscaling helps improving the PCM device speed [114, 13].

2. The energy scaling factors for the three operation schemes indicate that the energy needed to RESET/SET PCM device can be significantly reduced by isotropic volumetric scaling. This agrees with the PCM experiment that the RESET current and energy can be significantly decreased due to the reduction of the PCM programmable volume [128].

Table 2.3: Scaling factors of three PCM operation schemes

	Quantity	Constant E		Constant V		Constant I	
r	radius	$1/k$	Equation 2.4	$1/k$	Equation 2.4	$1/k$	Equation 2.4
l	length	$1/k$	Equation 2.4	$1/k$	Equation 2.4	$1/k$	Equation 2.4
R	resistance	k	Equation 2.5	k	Equation 2.5	k	Equation 2.5
E	electric field	1	Equation 2.9	k	Equation 2.29	k^2	Equation 2.37
ϕ	voltage	$1/k$	Equation 2.16	1	Equation 2.24	k	Equation 2.40
I	current	$1/k^2$	Equation 2.12	$1/k$	Equation 2.28	1	Equation 2.36
J	current density	1	Equation 2.10	k	Equation 2.27	k^2	Equation 2.35
T	temperature	$1/k^2$	Equation 2.18	1	Equation 2.31	k^2	Equation 2.42
t	time	$1/k^2$	Equation 2.19	$1/k^2$	Equation 2.32	$1/k^2$	Equation 2.43
P	power	$1/k^3$	Equation 2.22	$1/k$	Equation 2.33	k	Equation 2.44
Q	energy	$1/k^5$	Equation 2.23	$1/k^3$	Equation 2.34	$1/k$	Equation 2.45

3. In constant field (voltage) operation scheme, the operation current can be reduced by k^2 (k) times. This is particularly favorable to reduce the large RESET current, which is the most severe constraint of the PCM development. Since the large RESET current requires large size cell selector, which mainly determines the upper limit of the PCM array density [14], this is also beneficial to increase the data density.

4. In constant field operation scheme, temperature scaling factor is $1/k^2$. This means that the temperature profile is lowered during scaling, leading to reduced thermal proximity disturbance. In constant voltage operation scheme, temperature scaling factor is 1. This means that the temperature profile is kept the same during scaling, leading to unchanged thermal proximity disturbance [48]. So the thermal proximity disturbance between the adjacent PCM device cells is not a limiting factor of PCM device scaling.

5. In constant voltage (current) operation scheme, the electric field and current density

increase linearly (quadratically) in scaling. This might cause electro-migration and intrinsic device failure. For example, it has been observed that the current density as large as 3×10^{11} A/m^2 causes $Ge_2Sb_2Te_5$ atom separation and device failure [46]. And the RESET current for 60×60 nm cross-section $Ge_2Sb_2Te_5$ PCM is measured to be about 0.25 mA (i.e. 7×10^{10} A/m^2) [63]. As these two current densities are already close to each other, special attention should be paid to choose more stable material to enable reliable aggressive PCM device scaling.

6. The constant voltage (field) operation scheme requires the same (reduced) potential bias during PCM device scaling. The constant current operation scheme requires increased voltage supply from the external circuit.

Based on these conclusions, the author recommends constant voltage RESET operation scheme and constant field SET operation scheme in PCM device scaling. Also, it is worth mentioning that the PCM threshold switching in the SET operation exhibits constant voltage scaling properties in the sub-10 nm scale [129]. So the SET operation of PCM devices should also adopt constant voltage operation scheme during aggressive scaling.

2.4.2.2 Impact of TBR

The analytical isotropic scaling results, as summarized in Table 2.3, reveal the huge potential that scaling can offer to aggressively reduce the operation energy consumption and to significantly increase device switching speed and data storage density. Once the assumptions made in Section 2.2 are valid, these scaling factors can be used to predict the device scaling behavior.

But there is one factor deserves attention – the impact of thermal boundary resistance (TBR) during scaling. To make the presentation concise, only the axial TBR is analyzed here. The radial TBR can be analyzed in the same manner by substituting the subscript z with r. The axial TBR is defined as

$$\frac{\Delta T}{\mathrm{TBR}} = \frac{T_-(z_0-) - T_+(z_0+)}{\mathrm{TBR}} = -\kappa_- \frac{\partial T_-(z_0-)}{\partial z} = -\kappa_+ \frac{\partial T_+(z_0+)}{\partial z} \tag{2.46}$$

if the TBR is at the interface $z = z_0$, where the variables with subscript $+$ $(-)$ mean physical quantities at $z > z_0$ $(z < z_0)$, as shown in Figure 2.2.

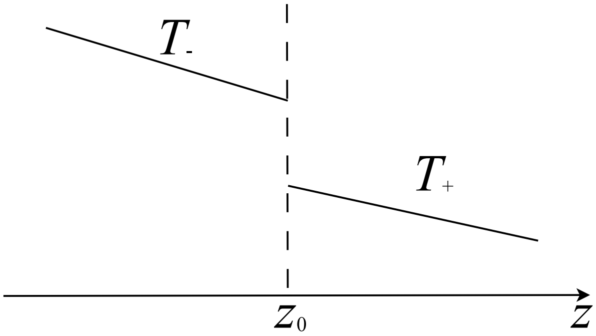


Figure 2.2: Schematic figure of the impact of TBR on temperature distribution at material interface.

In constant field operation scheme, the scaling factor of temperature T is $1/k^2$, and the scaling factor of coordinate z is $1/k$. So

$$\mathrm{TBR}_s = \frac{\Delta T_s}{-\kappa \frac{\partial T_s}{\partial z_s}} = \frac{\Delta\left(T/k^2\right)}{-\kappa \frac{\partial (T/k^2)}{\partial (z/k)}} = \frac{\Delta T}{-\kappa \frac{\partial T}{\partial z}} \frac{1}{k} = \mathrm{TBR}/k \tag{2.47}$$

In constant voltage operation scheme, the scaling factor of temperature T is 1, and the scaling factor of coordinate z is $1/k$. So

$$\mathrm{TBR}_s = \frac{\Delta T_s}{-\kappa \frac{\partial T_s}{\partial z_s}} = \frac{\Delta T}{-\kappa \frac{\partial T}{\partial (z/k)}} = \frac{\Delta T}{-\kappa \frac{\partial T}{\partial z}} \frac{1}{k} = \mathrm{TBR}/k \tag{2.48}$$

In constant field operation scheme, the scaling factor of temperature T is k^2, and the scaling factor of coordinate z is $1/k$. So

$$\mathrm{TBR}_s = \frac{\Delta T_s}{-\kappa \frac{\partial T_s}{\partial z_s}} = \frac{\Delta\left(Tk^2\right)}{-\kappa \frac{\partial (Tk^2)}{\partial (z/k)}} = \frac{\Delta T}{-\kappa \frac{\partial T}{\partial z}} \frac{1}{k} = \mathrm{TBR}/k \tag{2.49}$$

The Equation 2.47, Equation 2.48, and Equation 2.49 indicate that the TBR is scaled artificially by factor $1/k$ in the analytic isotropic scaling analysis. Actually, the physical TBR should be larger than this artificially reduced value. So the actual scaled device performance would be better than that shown in Table 2.3.

2.4.2.3 Exposed Nanowire

In the recent years, a novel PCM NW cell structure has been proposed. As shown in Figure 1.4, the PCM NW is not wrapped by thermal insulators. Instead it is exposed to vaccum. This exposed NW cell structure can suppress the melting point, and hence reduce the RESET energy consumption.

Here, it is shown that the NW melting point is lower than that of the bulk material [128]. As shown in Figure 2.3, during the RESET operation of NW PCM, the energy conservation relation

$$LdV - T_{NW}dSdV - \gamma dA = 0 \tag{2.50}$$

is valid when the radius of soid-liquid interface r_i is changed by dr_i (r_i is the NW radius at the beginning of melting). During this process, the volume of the solid region is changed by $dV = 2\pi r_i dr_i l$ (m^3); the area of solid-liquid interface is changed by $dA = 2\pi dr_i l$ is (m^2). In Equation 2.50, L is the latent heat of fusion (J/m^3); T_{NW} is the NW melting point (K); T_{bulk} is bulk material melting point (K); and γ is the interfacial energy at the solid-liquid interface (J/m^2).

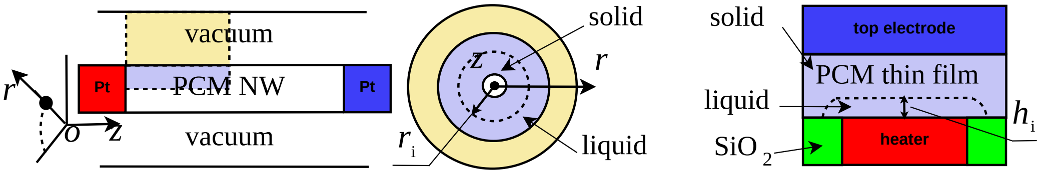


Figure 2.3: Schematic figure of NW PCM (left), NW PCM cross-section (middle), and the conventional thin film PCM structure (right). Here, the solid region is inside the liquid region. It deserves emphasizing that this figure is for illustrative purpose. In the real PCM device, the melt region will not be annular. To accomplish the RESET operation, one entire cross-section of the PCM NW should be amorphized.

If $\gamma = 0$, Equation 2.50 becomes

$$L - T_{bulk}dS = 0 \tag{2.51}$$

so, the entropy density change is $dS = L/T_{bulk}$ (J/m^{3}K). Therefore,

$$\begin{aligned} T_{NW} &= \frac{LdV - \gamma dA}{dSdV} \\ &= \frac{L\left(2\pi r_i dr_i l\right) - \gamma\left(2\pi dr_i l\right)}{\left(L/T_{bulk}\right)\left(2\pi dr_i l\right)} \\ &= T_{bulk}\left(1 - \frac{\gamma}{Lr_i}\right) \end{aligned} \tag{2.52}$$

In Equation 2.52, γ and L are unkowns. Here their values from NW are determined from experiments. The GeTe NW with diameter 40-80 nm is measured to melt at 390 °C, which is lower than the bulk GeTe melting point 725°C [107, 128]. So, $r_i = 30$ nm, $T_{NW} = 663$ K, and $T_{bulk} = 998$ K are used estimate that

$$\frac{\gamma}{L} = r_i\left(1 - \frac{T_{NW}}{T_{bulk}}\right) \approx 10.07 \text{ nm} \tag{2.53}$$

Combining Equation 2.52 and Equation 2.53, the dependence of GeTe NW melting point on NW radius can be obtained, as shown in Figure 2.4.

In contrast, thin film PCM can not take advantage of γ to reduce melting point. As shown in Figure 2.3, if the solid-liquid interface position h_i is changed by dh_i, dA in Equation 2.50 is approximately zero. So for thin film PCM, melting point is T_{bulk}.

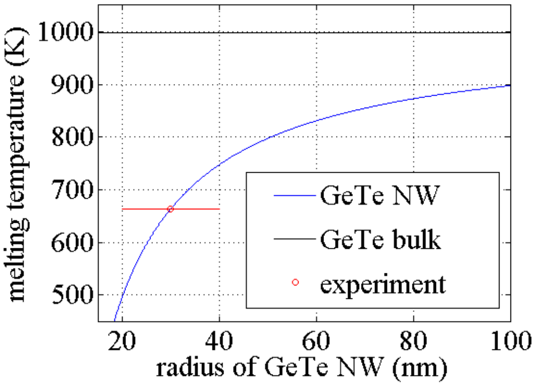


Figure 2.4: The dependence of GeTe NW melting point on NW radius

The derivation in Section 2.4.2.3 follows [51]. It deserves emphasizing that, the purpose of Section 2.4.2.3 is merely to provide an illustrative, yet not physically rigorous, analysis to show that the melting temperature can be influenced by geometry and material parameters of PCM nanostructures. Before applying the analysis and conclusions in Section 2.4.2.3, one should carefully note:

1. The geometry dependence of the melting temperature is not included in the rest part of the thesis. In the multiscale analysis out of this section, the melting temperature is chosen as the bulk value. In macroscopic analysis, the simulations do not include the size and curvature induced effect in any of the simulated quantities.

2. Figure 2.3 is for illustrative purpose only. In real physical systems, the melting onset region and the melting region geometry can be very different from what is schematically shown in Figure 2.3. For example, it can be shown via theoretical derivation that, if the liquid region is inslide the solid region at the beginning of melting, melting temperature can be higher than the bulk value. Furthermore, the actual shape of the liquid region (assumed to be annular here for illustrative purpose) can be very different from what is ideally shown in Figure 2.3. These effects are not addressed in this thesis. They deserve future research.

3. It has been experimentally shown that melting temperature of nanowire and nanoparticles are smaller than bulk value [128, 51]. It is interesting to theoretically understand these measured phenomena. This section attempts to provide a first-order understanding. To provide more complete understanding, more physics should be included in the model.

4. The geometry dependence of melting temperature shown in Equation 2.52, Equation 2.53, and Figure 2.4 are not applicable for small radius nanowire, since γ/L is fitted from measured melting temperature of nanowire with radius 30 nm. The value of γ/L can be very different if the nanowire radius is very small.

2.5 Numerical Scaling Analysis

In the analytical scaling analysis, the TBR is artificially scaled (Section 2.4.2.2) and the latent heat is not considered. In this section, the numerical simulation is applied to account for those missed influencing factors.

2.5.1 Analysis Methodology

2.5.1.1 Finite Element Method

The electro-thermal transport process during the RESET operation is simulated in two steps. First, the Equation 2.2 is solved by using the FEM to obtain $\phi(r, z)$, and $\mathbf{J}_e(r, z)$ is obtained by using Equation 2.3. Then, the TD-FEM is used to obtain $T(r, z, t)$ by solving the Equation 2.1 using implicit time stepping.

2.5.1.2 Isotropic Scaling

Batch simulations are performed to investigate the isotropic scaling properties of PCM device RESET performance. In the simulations, the magnitude of the voltage bias between AB and A′B′ in Figure 2.1 is chosen to be a value such that the maximum temperature at DD′ is equal to the PCM melting point (the pulse duration is 10 ns). This means that one entire cross-section of the PCM active region will be melted during a 10 ns pulse. After the pulse is switched off, the melted region ($T > T_m$) will quench into amorphous state to achieve RESET operation.

2.5.2 Simulation Parameters

In the TD-FEM simulation, a lot of quantities (mesh quantity, time step, TBR layer thickness, and effective heat capacity implementation of latent heat) should be controlled such that there is no artifical numerical impact on the simulation results. This section aims at choosing appropriate values for these quantities to guarantee simulation accuracy.

The results in this section are obtained by using PCM cell with radius 20 nm and length 100 nm. The voltage bias is set as 0.16 V. The RESET pulse width is 10 ns. Both the axial

TBR and the radial TBR are 20 m^2K/GW.

2.5.2.1 *Mesh Quality*

The finer mesh leads to more accurate result at the cost of longer computational time and larger memory consumption. Here, the mesh controlling factor N_{mesh} (larger value indicates finer mesh, see code for details) is tuned to investigate the influence of mesh quality on the calculation results, as shown in Figure 2.6.

The mesh's influence on the temperature simulation results is visualized in the Figure 2.5. It can be seen that the mesh quality has no significant influence on the result if N_{mesh} is larger than 20. In the TD-FEM simulations, the N_{mesh} =30 is chosen.

2.5.2.2 *Time Step*

According Figure 2.7, the time step should be chosen smaller than 0.1 ns. In the TD-FEM simulations, the time step is chosen as 0.05 ns.

2.5.2.3 *TBR region thickness*

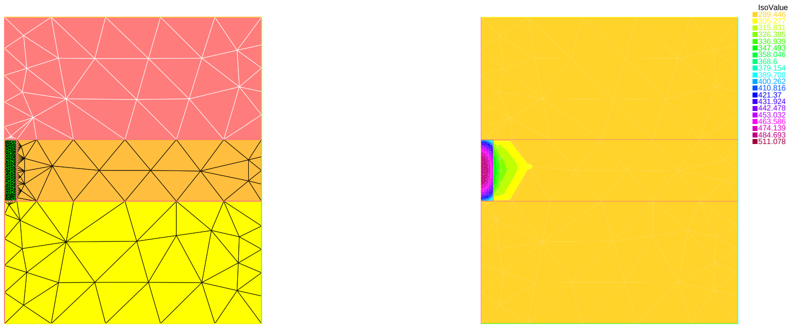
The thermal boundary resistance (TBR, also called Kapitza resistance) is included by inserting a thin artificial layer with thermal conductivity $\kappa_{\text{TBR}} = L_{\text{TBR}}/R_{\text{TBR}}$, where L_{TBR} is the TBR thickness and R_{TBR} is the TBR value. Dereasing the L_{TBR} is desirable to make the geometry less artificial. But if L_{TBR} is too small, the mesh will become more singular and undesirable. Therefore, striking a balance to choose the L_{TBR} is necessary.

In Figure 2.8, it can be seen that the temperature error can be controlled to be less than 1K, by chosing L_{TBR} <0.2 nm. So, in the TD-FEM simulations L_{TBR} =0.1 nm is chosen.

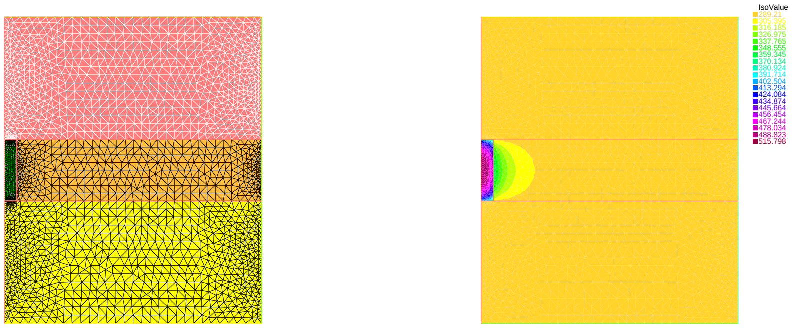
2.5.2.4 *Latent Heat Simulation*

In the TD-FEM simulation, the latent heat L is included via changing the volumetric specific heat capacity C_v into the effective volumetric specific heat capacity

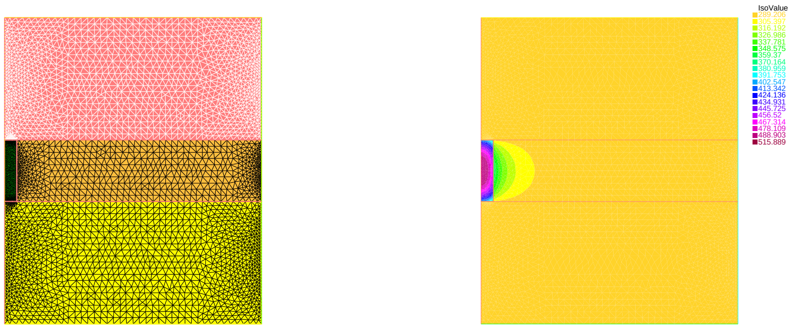
$$C_v^L = \begin{cases} C_v + \frac{L}{T_2 - T_1}, & \text{if } T_1 < T < T_2 \\ C_v, & \text{elsewise} \end{cases} \tag{2.54}$$



(a) N_{mesh}=5



(b) N_{mesh}=30



(c) N_{mesh}=40

Figure 2.5: Mesh (left) and temperature distribution (right) obtained by using different values of N_{mesh}.

where T_1 and T_2 are two slightly different values, both of which are close to the melting point of phase change material.

The $T_2 - T_1$ in Equation 2.54 is artificially introduced in order to capture the influence of latent heat during phase change. It's influence on simulation results must be tested.

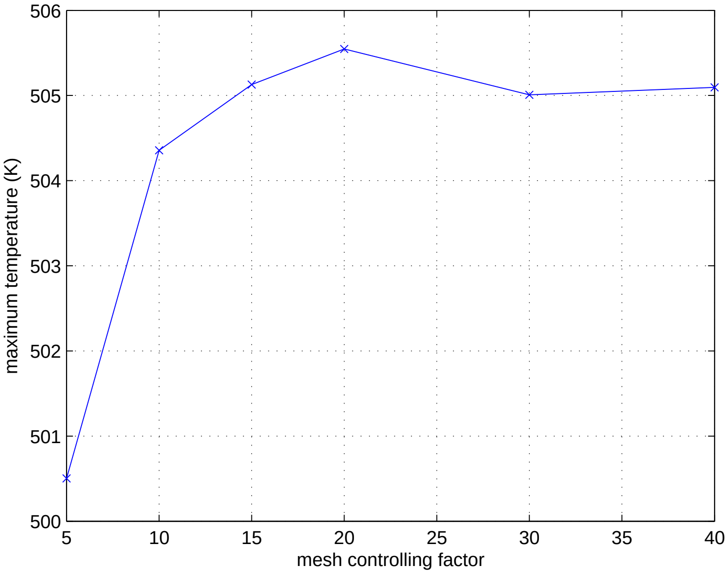


Figure 2.6: Influence of mesh quality on the calculation result.

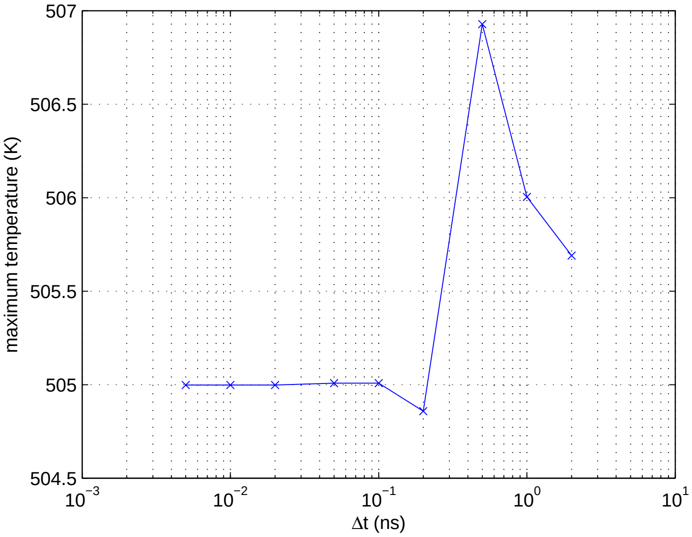


Figure 2.7: influence of TD-FEM timestep on the calculation result.

It can be seen from Figure 2.9 that if $T_2 - T_1$ is chosen to be smaller than 1% of T_m (melting temperature), the change of $T_2 - T_1$ has little influence on the simulation results.

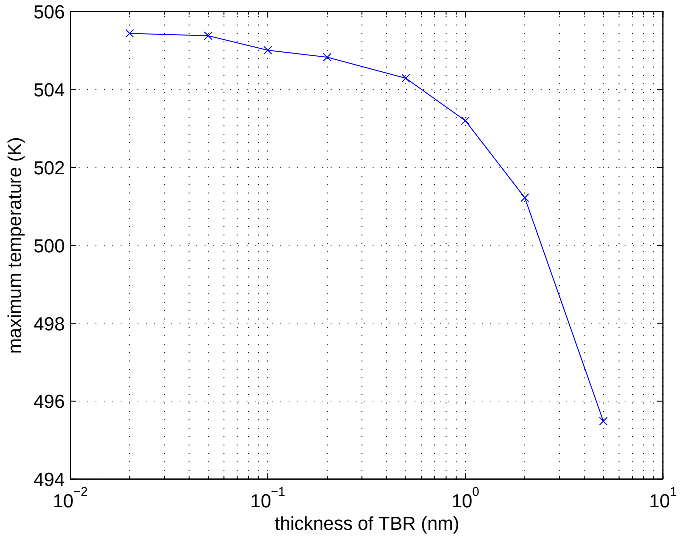


Figure 2.8: Influence of TBR region thickness on the calculation result.

In the TD-FEM simulations, $T_2 - T_1$ is set to be 1% of T_m.

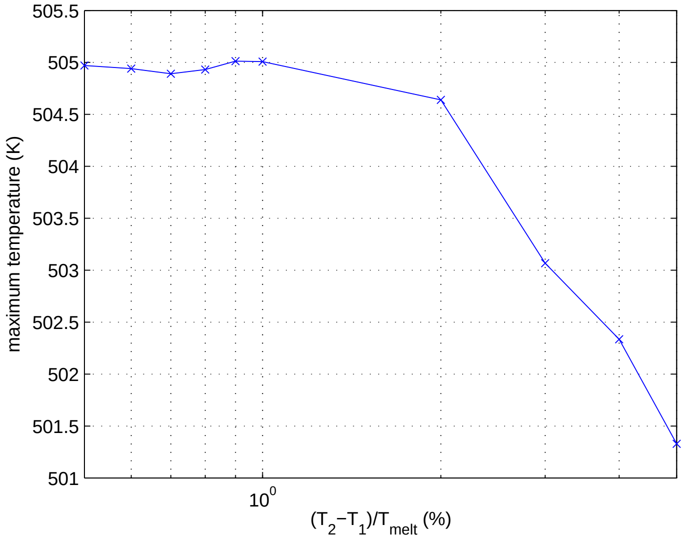


Figure 2.9: influence of $T_2 - T_1$ on the calculation result.

2.5.3 Analysis Results

2.5.3.1 Scaling Scenarios

The quantitative isotropic scaling scenarios are summarized in Table 2.4. In Figure 2.10, Figure 2.11, and Figure 2.12, the isotropic scaling properties of PCM devices with small, medium, and large AR are shown, respectively. From Figure 2.10, Figure 2.11, Figure 2.12, and Table 2.4, one can conclude that:

1. Isotropic scaling can significantly improve the RESET performance of PCM devices. Quantitatively speaking, if the device is scaled by a factor of $1/k = 1/2$, the RESET current, energy, and voltage can be reduced to less than 50.0%, 12.5%, and 100.0% of their original values, respectively. This scaling relation is valid for PCM devices with various aspect ratios.

2. Isotropic scaling cannot reduce the internal electric field effectively. So the electromigration problem (atom seperation when the electric field is large) will not be eliminated by isotropic scaling.

3. During the isotropic scaling, the relative fraction of four energy consumption components (energy consumed due to axial heat loss, radial heat loss, latent heat, and specific heat capacity) remains largely unchanged. The ralative fraction is largely determined by the aspect ratio. The larger (smaller) the aspect ratio, the larger the contribution of the radial (axial) direction heat loss.

4. In the four energy components, latent heat and specific heat consumptions are not important ($< 10\%$). What is important is the radial and axial heat loss ($> 90\%$). This relation does not change during isotropic scaling. So for PCM devices with any aspect ratio and with any size, the most important task to optimize the PCM device design is to more effectively confine heat. For example, for PCM devices with radius 20 nm (leftmost points of Figure 2.10 for AR = 1), $Q \approx$0.64 pJ, which consists of four components: heat loss due to radial direction thermal flux $Q_{rad} \approx$0.27 pJ, heat loss

due to axial direction thermal flux Q_{axi} ≈0.32 pJ, energy consumed due to specific heat capacity Q_{cap} ≈0.04 pJ, and energy consumed due to latent heat Q_{lat} ≈0.01 pJ. Only $Q_{cap} + Q_{lat}$ ≈50 fJ is necessary to achieve the RESET operations. The wasted Q_{rad} and Q_{axi} consumes 92% of the RESET energy. Therefore, confining the heat has the potential of reducing the RESET energy consumption by one order of magnitude[3].

5. To more clearly reveal the quantitative scaling relations, the numerical scaling factors are collected from the scaling results shown in Figure 2.10, Figure 2.11, and Figure 2.12, and listed in Table 2.4. From Figure 2.10, Figure 2.11, Figure 2.12, and Table 2.4, one can see that RESET operation approximately follows the constant voltage scaling scenario.

6. The numerical scaling factors of E, V, I, J, P, and Q are smaller than the corresponding analytic scaling factors. The reason is that the analytic scaling factors were obtained when the TBR is artificially scaled. If the TBR is kept constant in scaling, as implemented in our TD-FEM code, the scaling can offer more desirable impacts than those summarized in Table 2.3. So, one can use Table 2.3 as the lower limit estimation of device performance in scaling. For example, if the device is scaled by a factor of $1/k$, the RESET energy (current) will be reduced by $> k^3$ ($> k$) times.

7. Constant voltage scaling yields $T_s(r_s, z_s, t_s) = T(r, z, t)$, as shown in Table 2.3, leading to spatially squeezed temperature profile without change of its magnitude. This reveals that the cross-cell thermal proximity effect is unchanged in scaling. If the TBR is not scaled artificially, the thermal proximity effect is alleviated by scaling. This counter-intuitive impact offers an extra scaling benefit.

2.5.3.2 Impact of TBR

In Section 2.4.2.2, the analytic derivation qualitatively shows that the TBR helps alleviating cross-cell thermal proximity effect. Here, the numerical simulation is applied to quantita-

---

[3] Here, Q_{axi} is calculated by numerically integrating the thermal flux at the GST-electrode interface; and Q_{rad} is calculated by numerically integrating the thermal flux at the GST-substrate interface.

Table 2.4: Isotropic scaling scenarios of PCM device with different aspect ratios (AR)

	TD-FEM results						Const. V scaling in Table 2.3
	small AR		medium AR		large AR		
	AR=1	AR=2	AR=5	AR=6	AR=10	AR=12	
r	0.500	0.500	0.500	0.500	0.500	0.500	0.500
l	0.500	0.500	0.500	0.500	0.500	0.500	0.500
R	2.000	2.000	2.000	2.000	2.000	2.000	2.000
E	1.539	1.627	1.722	1.741	1.774	1.778	2.000
V	0.770	0.814	0.861	0.871	0.887	0.889	1.000
I	0.384	0.406	0.429	0.434	0.442	0.443	0.500
J	1.539	1.627	1.722	1.741	1.774	1.778	2.000
T	1.000	1.000	1.000	1.000	1.000	1.000	1.000
t	0.250	0.250	0.250	0.250	0.250	0.250	0.250
P	0.296	0.331	0.370	0.379	0.393	0.395	0.500
Q	0.074	0.083	0.093	0.095	0.098	0.099	0.125

tively investigate this impact.

The model shown in Figure 2.1 is used in the simulation. The radius of PCM device is set as 20 nm and the aspect ratio is set as 1. The PCM device cell isotropically downscaled by a factor of $1/k$ (k >1) and the RESET operation of the scaled device is simulated. The temperature $T\left(r, z_0, t_0\right)$ is observed during scaling, where z_0=0 nm and $t_0 = 5/k^2$ ns (RESET pulse duration). The TBR is fixed as 20 m^{2}K/GW. The results are shown in Figure 2.13. It is obvious that the thermal proximity effect is alleviated during downscaling, due to the presence of the TBR.

2.5.3.3 Impact of Aspect Ratio

The Figure 2.14 demonstrates the impact of AR on the RESET performance of the PCM devices. It is shown that decreasing the AR reduces the RESET energy, but at the cost of increasing the RESET current. As the large RESET current is a limiting factor of the PCM cell selector size, which largely determines the PCM device integration density, one need to choose an appropriate AR value to strike a balance between the data density and energy efficiency.

2.5.3.4 Impact of Interfacial Energy

In Section 2.4.2.3, the impact of the radius of exposed NW on the the PCM melting point is analyzed. As shown in Equation 2.52, $a \equiv \gamma/L$ influences the melting point and hence the PCM device performance. Here, γ is the interfacial energy which is up to the material properties of PCM the substrate material; L is the latent heat of the PCM. Generally speaking, the dependence of γ and L on the nanowire radius is not available.

The values of γ and L can change when different materials are chosen to fabricate the PCM device. In Section 2.4 and Section 2.5, it is assumed that the melting point of PCM are the same as the bulk melting point (i.e. $\gamma/L = 0$), in order to focus on the influence of the size scaling of PCM devices. But at real interfaces, γ/L will not be zero. According to Equation 2.52, the melting point will no longer be equal to bulk melting poing. So, the RESET energy and current will be changed correspondingly. In this section, the impact of $a = \gamma/L$ on the device performance is investigated by using numerical simulation, as shown in the Figure 2.15. In the simulation, the PCM device with AR=5 and $r_0 = 20$ nm is used.

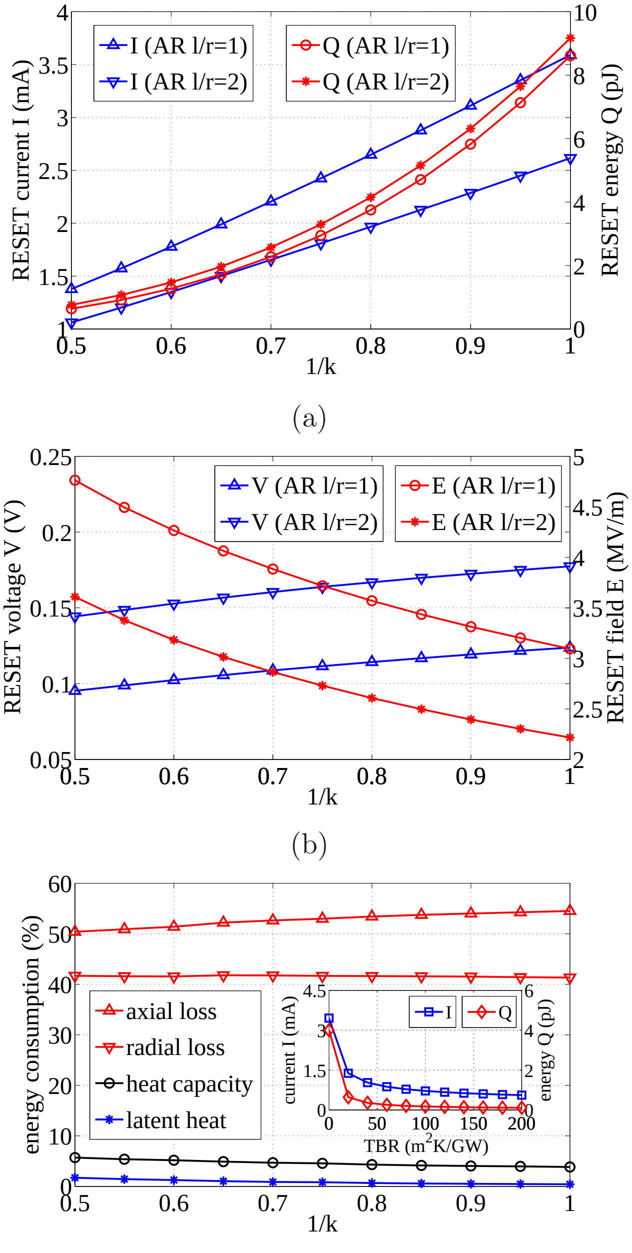
It can be seen from Figure 2.15 that increasing $a = \gamma/L$ can significantly reduce the RESET current and energy. Since L is fixed once the PCM is chosen, it is important to select appropriate materials to wrap the NW, in order to increase γ and therefore to decrease RESET energy and current.

It worth emphasizing that, the PCM NW shown in Figure 2.3 is exposed to vacuum in the radial direction [107, 128]. As a consequence, the interior region of NW is solid and the exterior region of NW is liquid, leading to decrease of melting point. It deserves noting that, if the liquid is inside the solid, the melting point will be increased.

2.5.3.5 Impact of Radius

According to Equation 2.52, both the γ/L and the radius r of exposed NW influence the melting point and hence the RESET energy and current. The impact of the γ/L is investigated in Section 2.5.3.4. Here, the influence of the radius r is studied.

As shown in Figure 2.16, both the RESET energy and RESET current can be significantly reduced when the radius of nanowire is decreased. This conclusion is valid for nanowire with different length.



(a)

(b)

(c)

Figure 2.10: The impact of isotropic scaling on PCM RESET performance. Note: scaling factor k is defined in Equation 2.4; in (a) and (b), r =40/k nm and l =40/k nm for AR=1, r =40/k nm and l =80/k nm for AR=2; in (c), AR l/r = 1; in (a)-(c), radial and axial TBR are 20 m^{2}K/GW.

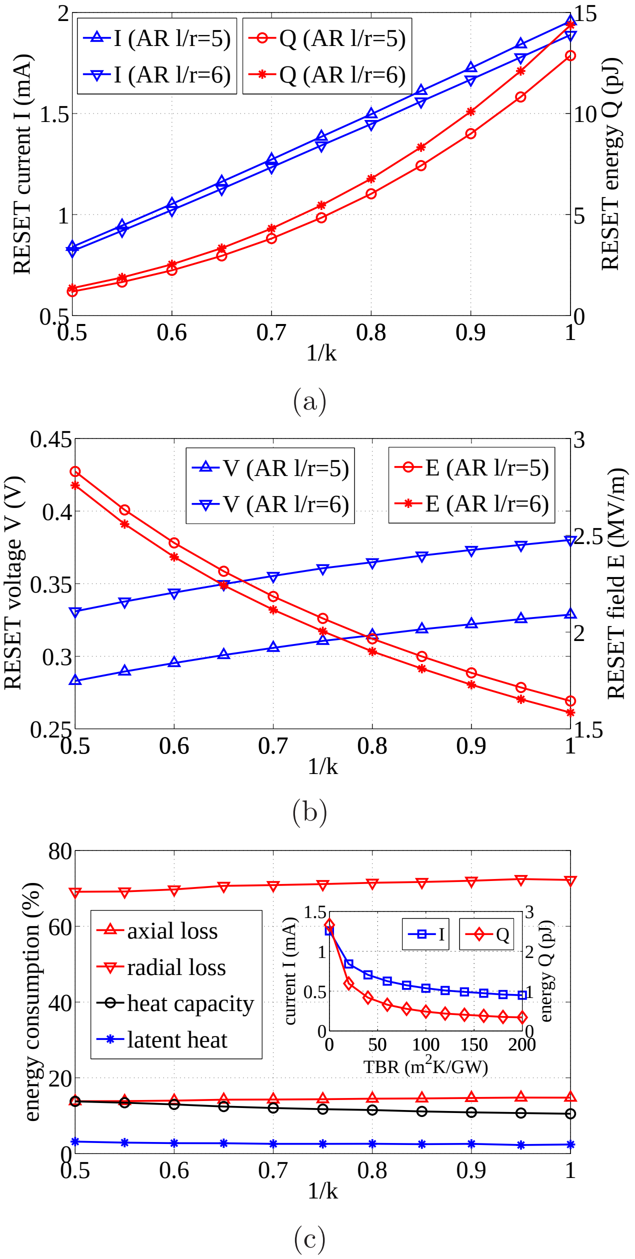
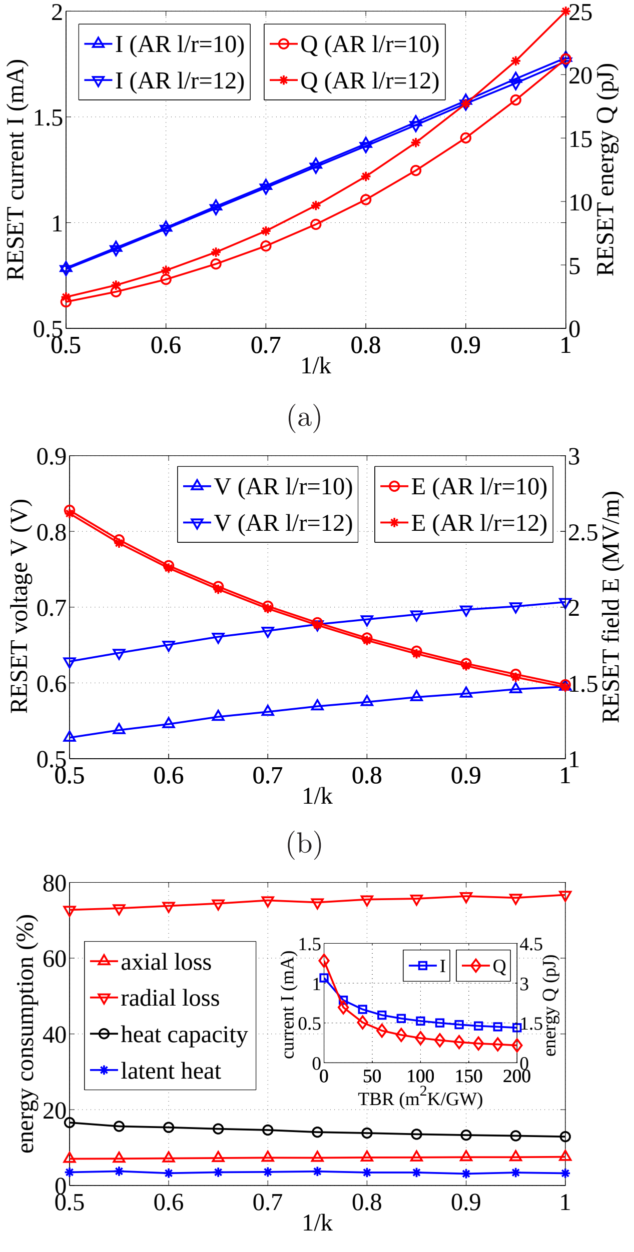


Figure 2.11: The impact of isotropic scaling on PCM RESET performance. Note: scaling factor k is defined in Equation 2.4; in (a) and (b), r =40/k nm and l =200/k nm for AR=5, r =40/k nm and l =240/k nm for AR=6; in (c), AR $l/r = 5$; in (a)-(c), radial and axial TBR are 20 m^{2}K/GW.



(a)

(b)

(c)

Figure 2.12: The impact of isotropic scaling on PCM RESET performance. Note: scaling factor k is defined in Equation 2.4; in (a) and (b), r =40/k nm and l =400/k nm for AR=10, r =40/k nm and l =480/k nm for AR=12; in (c), AR l/r = 10; in (a)-(c), radial and axial TBR are 20 m^{2}K/GW.

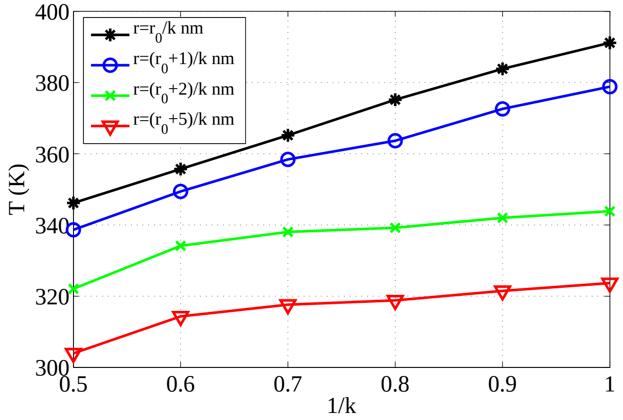


Figure 2.13: Influence of TBR on thermal proximity effect during scaling.

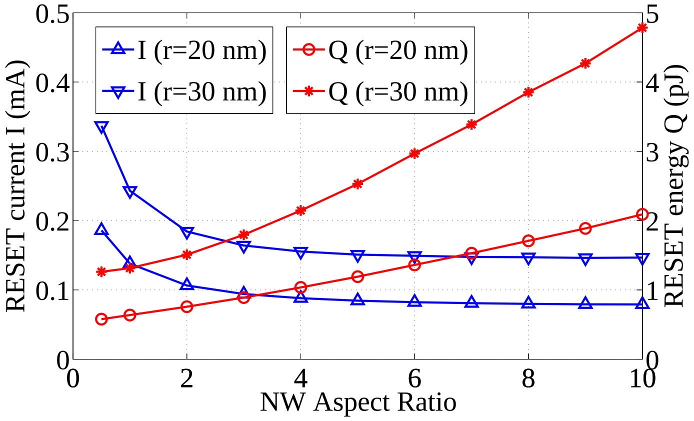


Figure 2.14: Impact of NW AR on RESET performance.

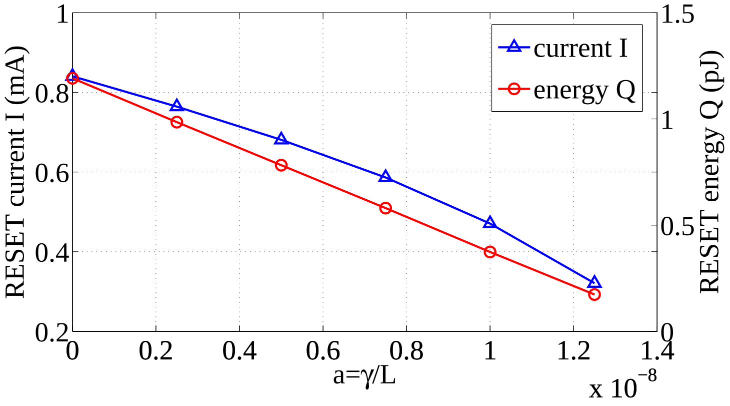


Figure 2.15: Impact of $a = \gamma/L$ on the NW PCM device performance.

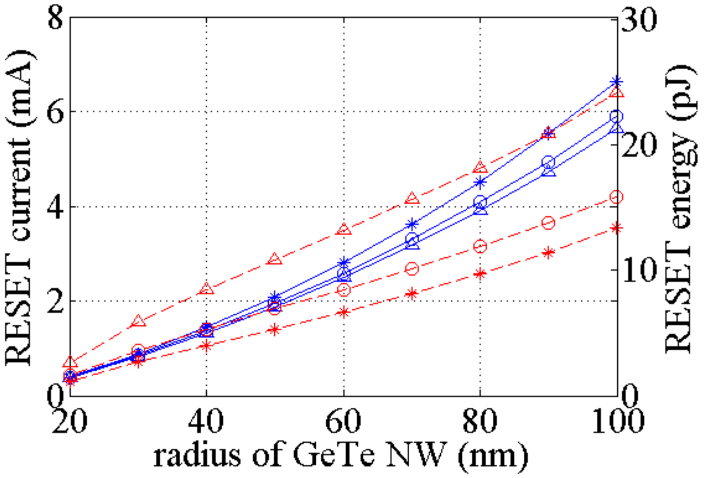


Figure 2.16: The impact of NW radius (Δ: l_0=200 nm; o: l_0=300 nm; *: l_0=500 nm where l_0 is the length of nanowire) on the NW PCM RESET energy (red dash lines) and current (blue solid lines).

Chapter 3

MESOSCOPIC SIMULATION

The macroscopic simulation discussed in Chapter 2 is able to offer some important qualitative and quantitative insight[1], when PCM device cell size is larger than tens of nanometers. Also, the macroscopic simulation is comparatively much less computationally expensive, so it is appropriate to simulate large-size systems. However, macroscopic simulation has its inherent limitations:

1. The governing electro-thermal transport equations are phenomenological, so they are not able to provide knowledge of the governing physical processes in PCM device;
2. In macroscopic simulation, it is assumed that the PCM device feature size is much larger than the carrier mean free path. This assumption will cease to be valid when the PCM device is aggressively downscaled;
3. Macroscopic simulation is not capable of predicting the physical scaling limit of PCM device technology.

Therefore, mesoscopic non-equilibrium Green's function (NEGF) simulations ($\approx 10^0 - 10^1$ nm) are applied in this chapter to investigate the electron transport properties in the sub-10 nm scale [67]. After introducing the model geometry in Section 3.1, the simulation algorithms are described in Section 3.2. The details of efficient implementation of the NEGF algorithm are presented in Section 3.4. The code is verified by comparing its simulation results against experiments. In Section 3.5, the simulation results are shown.

---

[1] e.g. scaling scenarios in different operation schemes, impact of PCM device cell aspect ratio and interfacial energy on device performance, and influence of thermal boundary resistance.

3.1 Simulation Model

As shown in Figure 3.1, the germanium telluride (GeTe) ultrathin films are sandwiched by titanium nitride (TiN), which is the most widely-used electrode material in PCM devices due to its excellent thermal and mechanical properties. Here the material to be studied is the prototypical binary PCM GeTe, instead of the most popular ternary $Ge_2Sb_2Te_5$, because binary PCM can better retain phase change properties in the ultra-scaled nanostructures probably due to its simpler stoichiometry and smaller crystalline unit cell.

To create the electrode-PCM-electrode sandwich structures, the *c*-GeTe and *a*-GeTe models are created by the microscopic simulations (Chapter 4). Then, the *c*-GeTe and *a*-GeTe models are sandwiched by TiN electrodes. The positions of Ti, N, Ge, and Te atoms in the 10 Å thick regions near each TiN-GeTe interface and the distances between TiN and GeTe are relaxed using the CG method. The CG convergence criteria is 40 meV/Å.

After obtaining the transport simulation model, the DFT simulation (Section 4.1.1) is applied to obtain Hamiltonian matrix **H** and overlap matrix **S**, which contain the ab initio description of the model's electronic strucutre.

3.2 Numerical Algorithm

3.2.1 Simulation Input

The mesoscopic NEGF simulator takes the output of the microscopic AIMD and DFT simulators as input.

As discussed in Section 3.1, the AIMD simulations generate the atomic coordinates of the NEGF simulation model; and the DFT simulations generate Hamiltonian matrices (Figure 3.2, Figure 3.3) and overlap matrices, which contain the ab initio description of the electronic structure of the systems shown in Figure 3.1.

3.2.2 Non-Equilibrium Green's Function

The Green's function solver takes the Hamiltonian matrix **H** and overlap matrix **S** from the microscopic simulation (see Chapter 4 for details) as input, and outputs mesoscopic transport properties of the nanoscale device [27, 28]. The Green's function solver begins

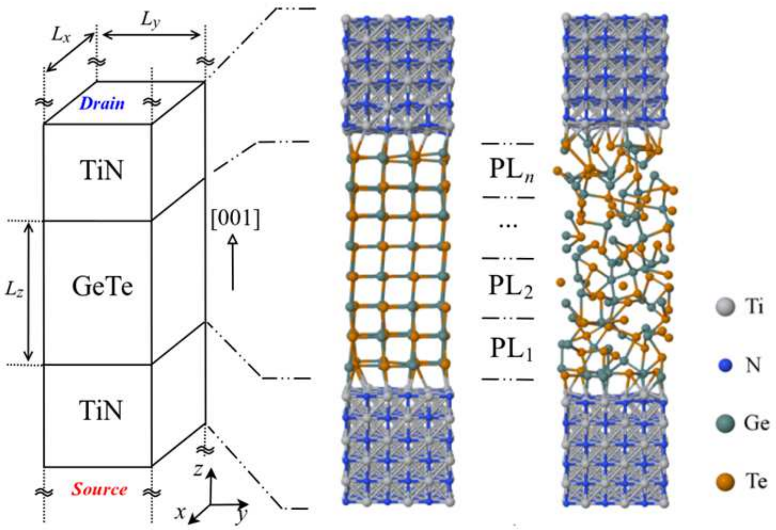


Figure 3.1: Ab initio simulation models, in which c-GeTe and a-GeTe ultrathin films with different thicknesses (L_z) are sandwiched by TiN electrodes. At supercell boundaries in x and y directions, the periodic boundary conditions (PBC) are applied. Thus, the models represent ultrathin films by infinitely repeating the supercells in x and y directions. Each principal layer (PL) contains 32 GeTe atoms in the supercell (i.e., 2 crystalline atomic layers).

with

$$\bar{\mathbf{H}}\mathbf{G} = \mathbf{I} \tag{3.1}$$

where $\bar{\mathbf{H}} = \epsilon\mathbf{S} - \mathbf{H}$, ϵ is the energy, $\mathbf{G}$ is the Green's function, and $\mathbf{I}$ is the identity matrix.

For a sandwich structure device whose central scattering region (O) is sandwiched by the source electrode (S) and drain electrode (D) (Figure 3.1), Equation 3.1 can be written as

$$\begin{bmatrix} \bar{\mathbf{H}}_S^\infty & \bar{\mathbf{H}}_{SO}^\infty & 0 \\ \bar{\mathbf{H}}_{SO}^{\infty\dagger} & \bar{\mathbf{H}}_O & \bar{\mathbf{H}}_{OD}^\infty \\ 0 & \bar{\mathbf{H}}_{OD}^{\infty\dagger} & \bar{\mathbf{H}}_D^\infty \end{bmatrix} \begin{bmatrix} \mathbf{G}_S^\infty & \mathbf{G}_{SO}^\infty & \mathbf{G}_{SD}^\infty \\ \mathbf{G}_{SO}^{\infty\dagger} & \mathbf{G}_O & \mathbf{G}_{OD}^\infty \\ \mathbf{G}_{SD}^{\infty\dagger} & \mathbf{G}_{OD}^{\infty\dagger} & \mathbf{G}_D^\infty \end{bmatrix} = \begin{bmatrix} \mathbf{I}^\infty & & \\ & \mathbf{I} & \\ & & \mathbf{I}^\infty \end{bmatrix} \tag{3.2}$$

where $\bar{\mathbf{H}}_i^{(\infty)} = \epsilon\mathbf{S}_i - \mathbf{H}_i$ ($i = S, O, D$) represent the source electrode region, scattering region (active phase change material region), and drain electrode region; $\bar{\mathbf{H}}_{SO}^\infty$ and $\bar{\mathbf{H}}_{SO}^{\infty\dagger}$ ($\bar{\mathbf{H}}_{OD}^\infty$ and

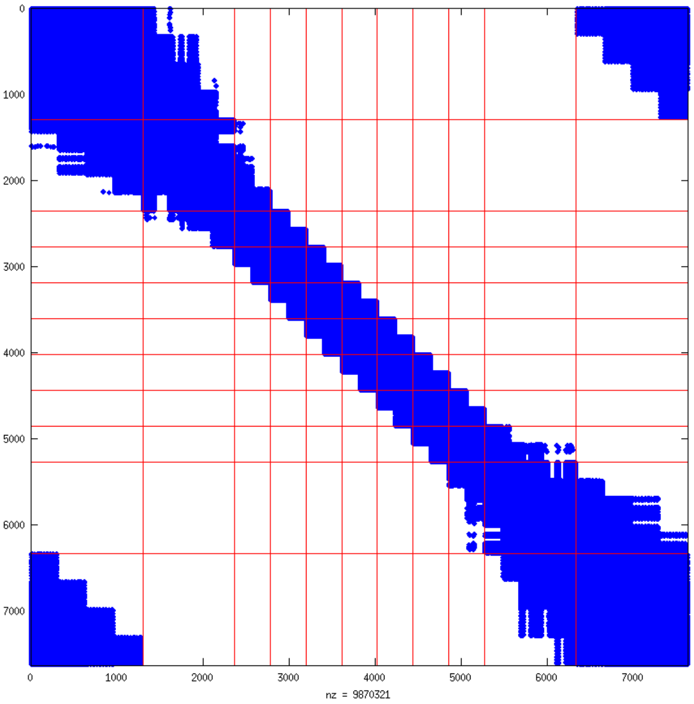


Figure 3.2: Hamiltonian of *c*-GeTe sandwiched by TiN.

$\bar{\mathbf{H}}_{OD}^{\infty\dagger}$) represent the coupling between the S (D) region and the scattering region. In order to eliminate the infinite terms (those with superscript ∞), Equation 3.2 is converted into

$$\mathbf{G}_O^r = [\epsilon \mathbf{S}_O - \mathbf{H}_O - \mathbf{\Sigma}_S^r - \mathbf{\Sigma}_D^r]^{-1} \tag{3.3}$$

where the two self energy matrices due to left and right contacts are

$$\mathbf{\Sigma}_S^r = \bar{\mathbf{H}}_{SO}^{\infty\dagger} \left(\bar{\mathbf{H}}_S^{\infty}\right)^{-1} \bar{\mathbf{H}}_{SO}^{\infty} \tag{3.4}$$

and

$$\mathbf{\Sigma}_D^r = \bar{\mathbf{H}}_{OD}^{\infty} \left(\bar{\mathbf{H}}_D^{\infty}\right)^{-1} \bar{\mathbf{H}}_{OD}^{\infty\dagger} \tag{3.5}$$

respectively.

To include the influence of electron-phonon scattering, the electron-phonon self energy matrix $\mathbf{\Sigma}_{ph}^r$ is added into Equation 3.3. The retarded Green's function becomes (hereafter,

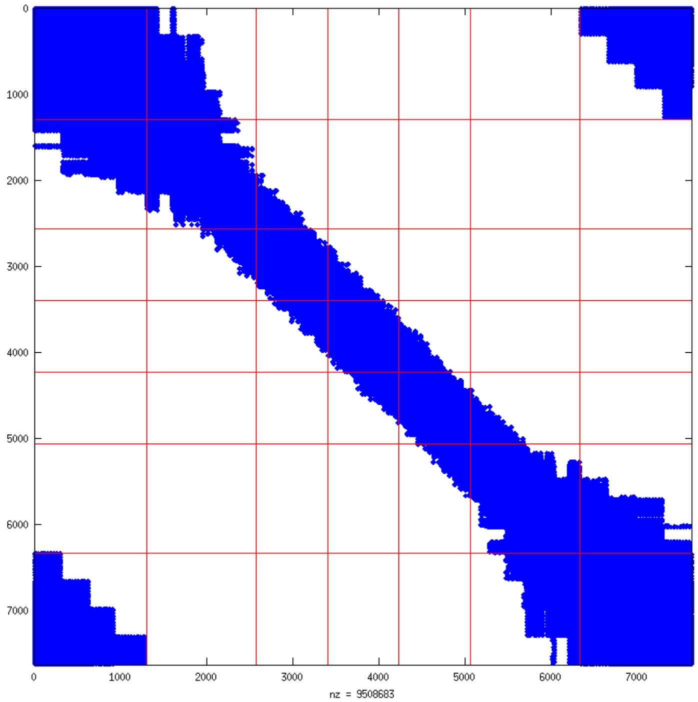


Figure 3.3: Hamiltonian of *a*-GeTe sandwiched by TiN.

the subscript O is suppressed for convenience)

$$\mathbf{G}^r = \mathbf{A}^{-1} \tag{3.6}$$

where

$$\mathbf{A} = \left(\epsilon\mathbf{S} - \mathbf{H} - \boldsymbol{\Sigma}_S^r - \boldsymbol{\Sigma}_D^r - \boldsymbol{\Sigma}_{ph}^r\right) \tag{3.7}$$

The electron Green's function $\mathbf{G}^n$ and the hole Green's function $\mathbf{G}^p$ defined in

$$\mathbf{A}\mathbf{G}^n = \boldsymbol{\Sigma}^{in}\mathbf{G}^{r\dagger} \tag{3.8}$$

and

$$\mathbf{A}\mathbf{G}^p = \boldsymbol{\Sigma}^{out}\mathbf{G}^{r\dagger} \tag{3.9}$$

respectively. $\mathbf{G}^r$, $\mathbf{G}^n$, and $\mathbf{G}^p$ are evaluated using the recursive Green's functon (RGF) algorithm.

Physically, the elements of retarded Green's function matrix $\mathbf{G}^{r}_{\mu\nu}$ defines the correlation between ϕ_μ and ϕ_ν; the elements of the electron Green's function $\mathbf{G}^{n}_{\mu\nu}$ defines electron density related to the overlap of ϕ_μ and ϕ_ν (see Equation 3.24 for mathematical definition); and the elements of the hole Green's function $\mathbf{G}^{p}_{\mu\nu}$ defines hole density related to the overlap of ϕ_μ and ϕ_ν (see Equation 3.25 for mathematical definition). Here, ϕ_μ means the μ^{th} pseudo atomic orbital.

The total in-scattering self-energy $\mathbf{\Sigma}^{in}$ and the total out-scattering self-energy $\mathbf{\Sigma}^{out}$ are defined as

$$\mathbf{\Sigma}^{in} = \mathbf{\Sigma}^{in}_{S} + \mathbf{\Sigma}^{in}_{D} + \mathbf{\Sigma}^{in}_{ph} \tag{3.10}$$

and

$$\mathbf{\Sigma}^{out} = \mathbf{\Sigma}^{out}_{S} + \mathbf{\Sigma}^{out}_{D} + \mathbf{\Sigma}^{out}_{ph} \tag{3.11}$$

respectively. Physically, the source and drain in-scattering matrices $\mathbf{\Sigma}^{in}_{S}$ and $\mathbf{\Sigma}^{in}_{D}$ describe the injection of electrons from souce and drain into the channel; the source and drain out-scattering matrices $\mathbf{\Sigma}^{out}_{S}$ and $\mathbf{\Sigma}^{out}_{D}$ describe the outflow of electrons from the channel into the souce and drain.

Here, the in-scattering self-energy and the out-scattering self-energy due to the source are

$$\mathbf{\Sigma}^{in}_{S} = \mathbf{\Gamma}_{S} f_{S} = i\left(\mathbf{\Sigma}^{r}_{S} - \mathbf{\Sigma}^{r\dagger}_{S}\right) f_{S} = -2\mathrm{Im}\left(\mathbf{\Sigma}^{r}_{S}\right) f_{S} \tag{3.12}$$

and

$$\mathbf{\Sigma}^{out}_{S} = \mathbf{\Gamma}_{S}\left(1 - f_{S}\right) = i\left(\mathbf{\Sigma}^{r}_{S} - \mathbf{\Sigma}^{r\dagger}_{S}\right)\left(1 - f_{S}\right) = -2\mathrm{Im}\left(\mathbf{\Sigma}^{r}_{S}\right)\left(1 - f_{S}\right) \tag{3.13}$$

respectively. The in-scattering self-energy and the out-scattering self-energy due to the drain are

$$\mathbf{\Sigma}^{in}_{D} = \mathbf{\Gamma}_{D} f_{D} = i\left(\mathbf{\Sigma}^{r}_{D} - \mathbf{\Sigma}^{r\dagger}_{D}\right) f_{D} = -2\mathrm{Im}\left(\mathbf{\Sigma}^{r}_{D}\right) f_{D} \tag{3.14}$$

and

$$\mathbf{\Sigma}^{out}_{D} = \mathbf{\Gamma}_{D}\left(1 - f_{D}\right) = i\left(\mathbf{\Sigma}^{r}_{D} - \mathbf{\Sigma}^{r\dagger}_{D}\right)\left(1 - f_{D}\right) = -2\mathrm{Im}\left(\mathbf{\Sigma}^{r}_{D}\right)\left(1 - f_{D}\right) \tag{3.15}$$

respectively. The in-scattering self-energy and the out-scattering self-energy due to inelastic electron-phonon scattering are

$$\mathbf{\Sigma}^{in}_{ph,inel} = D_{inel}\left\{n_B\left(\hbar\omega\right)\mathbf{G}^{n}\left(\epsilon - \hbar\omega\right) + \left[n_B\left(\hbar\omega\right) + 1\right]\mathbf{G}^{n}\left(\epsilon + \hbar\omega\right)\right\} \tag{3.16}$$

and

$$\mathbf{\Sigma}^{out}_{ph,inel} = D_{inel}\left\{\left[n_B\left(\hbar\omega\right)+1\right]\mathbf{G}^p\left(\epsilon-\hbar\omega\right)+n_B\left(\hbar\omega\right)\mathbf{G}^p\left(\epsilon+\hbar\omega\right)\right\} \tag{3.17}$$

respectively. The imaginary part of $\mathbf{\Sigma}^{r}_{ph,inel}$ is calculated by using

$$\mathrm{Im}\left(\mathbf{\Sigma}^{r}_{ph,inel}\right) = -\frac{1}{2}\mathrm{Re}\left(\mathbf{\Sigma}^{in}_{ph,inel}+\mathbf{\Sigma}^{out}_{ph,inel}\right) \tag{3.18}$$

and the real part of $\mathbf{\Sigma}^{r}_{ph,inel}$ is ignored.

Physically, the inelastic electron-phonon in-scattering matrix $\Sigma^{in}_{ph,inel}(\epsilon)$ describes the in-scattering of electrons into energy ϵ from other energies by the phonons; the inelastic electron-phonon out-scattering matrix $\Sigma^{out}_{ph,inel}(\epsilon)$ describes the out-scattering of electrons from energy ϵ into other energies by the phonons.

The

$$f_S = \frac{1}{1+e^{\frac{\epsilon-\mu_S}{k_BT}}} \tag{3.19}$$

and

$$f_D = \frac{1}{1+e^{\frac{\epsilon-\mu_D}{k_BT}}} \tag{3.20}$$

mean the Fermi function of the source and drain electrodes, where ϵ is the energy of electron;

$$\mu_S = \epsilon_F + \frac{q_e}{2}v \tag{3.21}$$

and

$$\mu_D = \epsilon_F - \frac{q_e}{2}v \tag{3.22}$$

are the electrochemical potential of the source and drain; k_B is the Boltzmann constant; T is the temperature; q_e is the elementary charge; ϵ_F is the Fermi level; and v is the external bias.

3.2.3 Self-Consistent Field Solution

The schematic flowchart of the general self-consistent field (SCF) NEGF simulation algorithm is shown in Figure 3.4. The functional blocks shown in Figure 3.4 are described below:

1. The module 1. calculates the contact self energy matrix due to source (Equation 3.4) and drain (Equation 3.5). As shown in Figure 3.1, Figure 3.2, and Figure 3.3, the system and the Hamiltonian are finite in the ab initio DFT simulations. The calculation of the contact self-energy matrices starts from the finite Hamiltonian and includes the source and drain electrodes as infinite electron reserviors. The details about the mathematical algorithm and the efficient implementation are introduced in Section 3.4.1.2.

2. The module 2. sets the initial guess of the potential profile for the SCF calculation. The initial guess of the potential profile should be as close as possible to the final converged potential profile, in order to minize the iterative computational cost. Typically, the initial guess of potential profile is set as a linear drop from drain to source. In this case, an extra term is added to the DFT Hamiltonian $\mathbf{H}$

   $$\mathbf{H}_{\mu\nu} \leftarrow \mathbf{H}_{\mu\nu} + \mathbf{S}_{\mu\nu}\frac{v_\mu + v_\nu}{2} \tag{3.23}$$

   in order to approximately account for the linear potential drop, where v_μ is the potential at the atom to which the pseuto-atomic orbital (PAO) μ belongs.

3. The module 3. takes the potential profile as input and calculates the retarded Green's function (Equation 3.6), the electron Green's function (Equation 3.8), and the hole Green's function (Equation 3.9) are computed using the recursive Green's function algorithm. The derivation and implementation details are presented in Section 3.4.1.3.

4. The module 4. takes the Green's functions as input and calculates the electron-phonon in-scattering self-energy matrix (Equation 3.16) and the electron-phonon out-scattering self-energy matrix (Equation 3.17).

5. The module 5. checks whether or not the SCF calculation of the electron-phonon scattering between the module 3. and and the module 4. has converged. The convergence is considered as achieved if the criteria for electron

   $$\left|\frac{n_A^{(k)} - n_A^{(k-1)}}{n_A^{(k-1)}}\right| \leq \eta \tag{3.24}$$

and for hole

$$\left| \frac{p_A^{(k)} - p_A^{(k-1)}}{p_A^{(k-1)}} \right| \leq \eta \tag{3.25}$$

are satisfied for every atom A, where the superscript (k) means the k^{th} iteration of the electron-phonon SCF calculation; η is the pre-defined tolerance of convergence; the electron number and the hole number are

$$n_A = \frac{1}{2\pi} \sum_{\mu \in A} \sum_{\nu} \int_{-\infty}^{\infty} d\epsilon \mathbf{G}_{\mu\nu}^{n}(\epsilon) \mathbf{S}_{\mu\nu} \tag{3.26}$$

and

$$p_A = \frac{1}{2\pi} \sum_{\mu \in A} \sum_{\nu} \int_{-\infty}^{\infty} d\epsilon \mathbf{G}_{\mu\nu}^{p}(\epsilon) \mathbf{S}_{\mu\nu} \tag{3.27}$$

respectively.

6. The module 6. takes the Green's functions as input and calculates the potential profile. The charge density can be computed from the electron Green's function and the hole Green's function. Then the potential profile can be calculated from the charge density.

7. The module 7. checks whether or not the convergence of the potential profile has achieved.

8. The module 8. performs the post-process calculations. It computes various physical quantities, like current density, transmission coefficients, and local density of states, etc. The details are introduced in Section 3.2.4.

3.2.4 Post-Process Calculations

After the SCF calculation as shown in Figure 3.4 is converged, the electron transport properties of the PCM (Figure 3.1) can be calculated in the post-process module.

3.2.4.1 Transmission Coefficient

When the inelastic scattering is ignored, the transmission coefficient

$$T(\epsilon) = \mathrm{Tr}[\mathbf{\Gamma}_S(\epsilon) \mathbf{G}^r(\epsilon) \mathbf{\Gamma}_D(\epsilon) \mathbf{G}^{r\dagger}(\epsilon)] \tag{3.28}$$

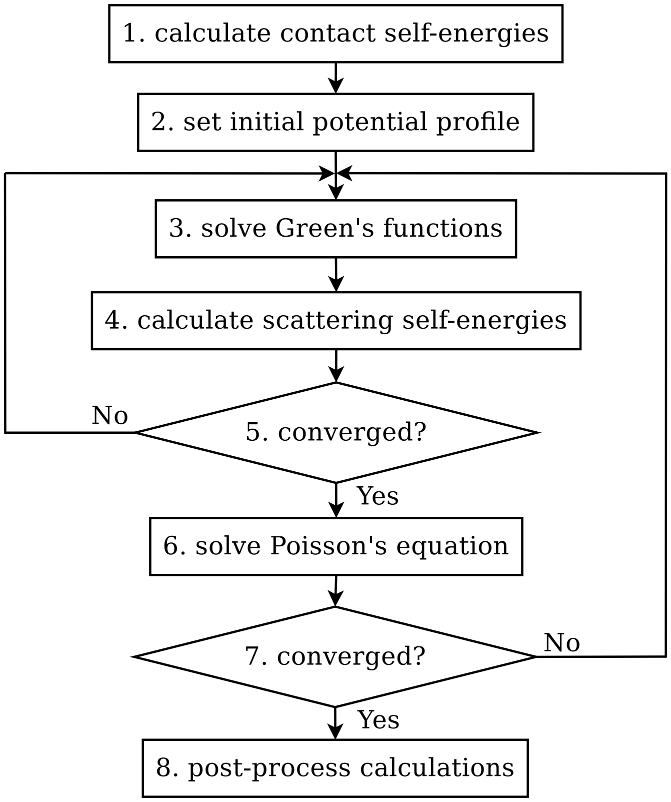


Figure 3.4: Schematic flowchart of the NEGF SCF simulation algorithm.

is computed from the retarded Green's function

$$\mathbf{G}^r(\epsilon) = [\epsilon\mathbf{S} - \mathbf{H} - \mathbf{\Sigma}_S^r(\epsilon) - \mathbf{\Sigma}_D^r(\epsilon)]^{-1} \tag{3.29}$$

using the recursive Green's function algorithm (Section 3.4.1.3). Here, **H** and **S** are the Hamiltonian and overlap matrices obtained using DFT; $\mathbf{\Sigma}_{S/D}^r(\epsilon)$ is the self-energy of the source/drain, which are computed using the iterative surface Green's function algorithm (Section 3.4.1.2); and

$$\mathbf{\Gamma}_{S/D}(\epsilon) = i \times [\mathbf{\Sigma}_{S/D}^r(\epsilon) - \mathbf{\Sigma}_{S/D}^{r\dagger}(\epsilon)] \tag{3.30}$$

is the broadening matrix of source/drain. Physically, the broadening matrices of the source/drain $\mathbf{\Gamma}_{S/D}$ describes the broadening of the electron energy levels due to the presence of source/drain; and the transmission $T(\epsilon)$ defines the transmission probability of electrons at energy ϵ, summed over all bands.

3.2.4.2 Low-Bias Conductance

Since thickness of the ultrathin films of our interest here is much smaller than the electron mean free path in GeTe, the electrical conductance of the *c*-GeTe (*a*-GeTe) ultrathin films G_c (G_a) can be approximately calculated by using[2]

$$G_{c/a} = \frac{2q_e^2}{\hbar} \int g(\epsilon) d\epsilon \tag{3.31}$$

where

$$g(\epsilon) = -T(\epsilon) \partial f(\epsilon) / \partial \epsilon \tag{3.32}$$

in the sub-threshold voltage range, in which low bias is applied to read out the stored data. Here, ϵ is the electron energy; $f(\epsilon)$ is the Fermi function; q_e is the elementary charge; $\hbar$ is the reduced Planck constant. It deserves mentioning that in Equation 3.31, the inelastic scattering is ignored.

3.2.4.3 Local Density of States

The local density of states (LDOS) of atom i is computed by using

$$\text{LDOS}(\epsilon, i) = -\frac{1}{\pi} \text{Im} \left\{ \text{Tr} \left[(\mathbf{G}^r \cdot \mathbf{S})_{i,i} \right] \right\} \tag{3.33}$$

where $\mathbf{G}^r$ is the retarded Green's function and $\mathbf{S}$ is the overlap matrix in the SIESTA PAO representation.

3.2.4.4 Current Density

The current density from the k^{th} principal layer to the $(k+1)^{\text{th}}$ principal layer is computed by using

$$J_{k \to k+1} = \frac{2q_e}{\hbar L_x L_y} \int_{-\infty}^{+\infty} d\epsilon \frac{1}{2\pi} \text{Tr} \left[\mathbf{T}_{k,k+1} \mathbf{G}^n_{k+1,k} - \mathbf{G}^n_{k,k+1} \mathbf{T}_{k+1,k} \right] \tag{3.34}$$

where

$$\mathbf{T}_{\mu\nu} = \epsilon \mathbf{S}_{\mu\nu} - \mathbf{H}_{\mu\nu} \tag{3.35}$$

---

[2] It worth noting that in Equation 3.31, the inelastic scattering of the electrons, say by phonons, is ignored. So, in Equation 3.31, the transport is elastic. The role of the inelastic electron-phonon scattering is investigated in Section 3.5.3.

L_x and L_y are the supercell size perpendicular to the transport direction, as shown in Figure 3.1.

When the electron-phonon scattering is turned off in the simulation, the transport being modelled will be elastic. In this case, Equation 3.34 can be simplified into

$$J = \frac{2q_e}{\hbar L_x L_y} \int_{-\infty}^{+\infty} d\epsilon \frac{1}{2\pi} T(\epsilon)[f_S(\epsilon) - f_D(\epsilon)] \tag{3.36}$$

where the subscript k is suppressed, as J is no longer dependent on k in the elastic transport. Here, T is tranmission coefficient (Equation 3.28); and $f_{S/D}$ is the Fermi function of source/drain (Equation 3.19 and Equation 3.20).

3.2.4.5 Electron Mean Free Path

In order to estimate the electron-phonon scattering strength (D_{inel} in Equation 3.16 and Equation 3.17), the electron mean free path (MFP) of the *c*-GeTe needs to be calculated and fitted to the measured value.

The MFP is calculated by using

$$\text{MFP} = \sum_{\epsilon_i > E_c} v(\epsilon_i)\tau(\epsilon_i) \times \left\{ \frac{f(\epsilon_i)\text{DOS}(\epsilon_i)}{\sum_{\epsilon_i > E_c} f(\epsilon_i)\text{DOS}(\epsilon_i)} \right\} \tag{3.37}$$

where ϵ_i is the eigenvalue of the Schrödinger equation $H\psi_i = \epsilon_i\psi_i$; E_c is the bottom of the conduction band;

$$v(\epsilon_i) = \sqrt{\frac{2(\epsilon_i - E_c)}{m^*}} \tag{3.38}$$

is the electron velocity; m^* is the effective mass; f is the Fermi function; the density of states (DOS) can be computed from the retarded Green's function and the overlap matrix (Equation 3.33);

$$\tau^{-1}(\epsilon_i) = \sum_{\mu} \frac{-2\text{Im}\left[\mathbf{\Sigma}^r_{ph,inel,\mu\mu}(\epsilon_i)\right]}{\hbar} \times \left\{ \frac{|c_{i\mu}|^2}{\sum_\mu |c_{i\mu}|^2} \right\} \tag{3.39}$$

is the life time; $\mathbf{\Sigma}^r_{ph,inel}$ is the retarded inelastic electron-phonon scattering self-energy matrix as discussed in Section 3.2.2; $c_{i\mu}$ is the coefficient used to express the eigenstates $\psi_i = \sum_\mu c_{i\mu}\phi_\mu$ as a linear combination of the pseudo-atomic orbitals (PAO) ϕ_μ.

Since the DFT calculations in SIESTA are performed in the PAO representation, the MFP should be calculated by performing a two-level summation. In the first level summation (Equation 3.39), the contribution of all the PAOs ϕ_μ to the scattering rate corresponding to the eigenstate ψ_i should be summed by using the PAO expansion coefficients $c_{i\mu}$ (term in the curly bracket) as the weighting coefficients. These weighting coefficients are normalized and represent the electron population contribution to the eigenstate ψ_i from the PAO ϕ_μ. Since $-2\text{Im}\left[\mathbf{\Sigma}^r_{ph,inel,\mu\mu}(\epsilon_i)\right]/\hbar$ represents the contribution from the PAO ϕ_μ to the electron scattering rate corresponding to the eigenstate ψ_i, Equation 3.39 physically means that the more heavily populated the PAO ϕ_μ, the larger its contribution to the scattering rate. The same principle can be applied to the second level summation (Equation 3.37). Since $f(\epsilon_i)\text{DOS}(\epsilon_i)$ means the electron population corresponding to eigenstate ψ_i, the contribution to the total MFP should be summed using the coefficients shown in the curly brackets.

It worth noting that in Equation 3.39, the contributions to the scattering rate from the μ^{th} PAO are summed by using the normalized weighting coefficients shown in the curly brackets, which represent the electron population of this PAO. In Equation 3.37, the summation should consider the degeneracy. If the ϵ_i is degenerate with a degree of k, then the summation should include the contribution from the ϵ_i term for k times.

3.3 Assumptions and Approximations

The mesoscopic simulation analysis presented in Chapter 3 are based on the following assumptions and approximations:

1. During the creation of the simulation model (Section 3.1), the metal-PCM interfaces are created by using CG relaxation, instead of the brute-force AIMD. The later is apparently more physically appropriate. However, the brute-force AIMD treatment of the interfaces requires including the metal layers in the computationally demanding AIMD calculations. Since the AIMD simulations without the interfaces are already very computationally expensive, the interfaces are created by using CG.

2. In the mesoscopic electron tranport simulations, the atom positions are fixed. In

reality, the thermal vibration of lattice can affect the transport properties, which is approximated by using the inelastic electron-phonon in-scattering and out-scattering self energy matrices (Equation 3.16 and Equation 3.17). This means that the phonon distribution is assumed to be constant and not affected by the electron tranport.

3. The *c*-PCM used in the crystalline tranport model is assumed to be pure. In real PCM devices, there will be defects and vacancies, which are ignored here. Furthermore, only the tranport along the [100] direction is simulated.

4. The *a*-PCM transport model is assumed to be periodic in the xy plane, which is perpendicular to the tranport direction (z direction). This assumption is made in order to enable the tranport simulations. In reality, there is no long range order in the xy plane. However, in the simulation, the amorphous nature along the transport direction is taken into consideration.

The above assumptions and approximations are made to facilitate the calculation of the systems of the practical scale. Without thes approximations, it would be practically difficult to perform the simulation and analysis. The influence of these assumptions deserves future research efforts.

3.4 Algorithm Implementation

In this section, the details of the NEGF code implementation are introduced. The code implementation is verified by comparing the simulation results against experiments and the existing simulations.

3.4.1 Efficient Implementation

Since the NEGF simulations are very computationally demanding, it is very important to improve the simulation efficiency by implementing fast algorithms and parallelization. In Section 3.4.1.1, the parallelization methodology implemented in the NEGF simulator is briefly introduced. The Section 3.4.1.2, Section 3.4.1.3, and Section 3.4.1.4 introduce the efficient algorithms implemented in the NEGF simulator to compute the contact self energy, retarded Green's function, and electron and hole Green's functions, respectively. In Section 3.4.3, the running time and it's components of the NEGF code are analyzed, in order to provider a general overview of the simulation time of the simulator.

3.4.1.1 MPI Parallelization

As shown in Equation 3.31, the Green's function calculation should run through many energy points. This makes the calculation very computationally demanding. Since the calculations of these energy points are independent of each other, which is ideal for massive parallelization, the simulation code parallelizes the calculations via message passing interface (MPI).

3.4.1.2 Contact Self Energy

The calculation of Equation 3.4 and Equation 3.5 to include the contacts' influence should run through many energy values. So its efficiency should be optimized to save computational time. In the NEGF code, the efficient iterative algorithm proposed in the seminal paper [97] is implemented, as discussed in this section. This efficient algorithm can achieve exponential convergence speed, compared to the other algorithms which could only converge in a linear speed.

In Equation 3.4, typically only the top right block (bottom left block) of $\bar{\mathbf{H}}_{LC}^{\infty\dagger}$ ($\bar{\mathbf{H}}_{LC}^{\infty}$) is nonzero[3]. So it only need the bottom right block of $\left(\bar{\mathbf{H}}_{L}^{\infty}\right)^{-1}$ to calculate the left contact self energy. The inverse of $\bar{\mathbf{H}}_{L}^{\infty}$ is defined as $\mathbf{G}_{L}^{\infty}$, which obeys

$$\bar{\mathbf{H}}_{L}^{\infty}\mathbf{G}_{L}^{\infty} = \mathbf{I}^{\infty} \tag{3.40}$$

---

[3] In Equation 3.4, the source (drain) is assumed to be the left (right) contact. Since the bias polarity can be reversed, the subscripts S, O, D is replaced by L (left), C (center), and R (right) in this section.

whose block version is

$$\begin{bmatrix} \ddots & \ddots & & \\ \ddots & \bar{\mathbf{H}}_L & \bar{\mathbf{H}}_{LL} & \\ & \bar{\mathbf{H}}_{LL}^\dagger & \bar{\mathbf{H}}_L & \bar{\mathbf{H}}_{LL} \\ & & \bar{\mathbf{H}}_{LL}^\dagger & \bar{\mathbf{H}}_L \end{bmatrix} \begin{bmatrix} \ddots & \vdots & \vdots & \vdots \\ \dots & \mathbf{G}_{n-2,n-2} & \mathbf{G}_{n-2,n-1} & \mathbf{G}_{n-2,n} \\ \dots & \mathbf{G}_{n-1,n-2} & \mathbf{G}_{n-1,n-1} & \mathbf{G}_{n-1,n} \\ \dots & \mathbf{G}_{n,n-2} & \mathbf{G}_{n,n-1} & \mathbf{G}_{n,n} \end{bmatrix} = \begin{bmatrix} \ddots & & & \\ & \mathbf{I} & & \\ & & \mathbf{I} & \\ & & & \mathbf{I} \end{bmatrix} \tag{3.41}$$

Physically speaking, Equation 3.41 describe the interactions among the atom layer n and $n - 2^0$ of the left contact. At this point, only $2^0 + 1 = 2$ layer of the left contact exists.

For $i < n$, Equation 3.41 gives

$$\bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{i-4,n} + \bar{\mathbf{H}}_L \mathbf{G}_{i-3,n} + \bar{\mathbf{H}}_{LL} \mathbf{G}_{i-2,n} = 0 \tag{3.42}$$

$$\bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{i-3,n} + \bar{\mathbf{H}}_L \mathbf{G}_{i-2,n} + \bar{\mathbf{H}}_{LL} \mathbf{G}_{i-1,n} = 0 \tag{3.43}$$

$$\bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{i-2,n} + \bar{\mathbf{H}}_L \mathbf{G}_{i-1,n} + \bar{\mathbf{H}}_{LL} \mathbf{G}_{i,n} = 0 \tag{3.44}$$

From Equation 3.42 and Equation 3.44, one could have

$$\mathbf{G}_{i-3,n} = -\bar{\mathbf{H}}_L^{-1} \left(\bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{i-2,n} + \bar{\mathbf{H}}_{LL} \mathbf{G}_{i,n} \right) \tag{3.45}$$

$$\mathbf{G}_{i-1,n} = -\bar{\mathbf{H}}_L^{-1} \left(\bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{i,n} + \bar{\mathbf{H}}_{LL} \mathbf{G}_{i+2,n} \right) \tag{3.46}$$

which are inserted into Equation 3.43 to obtain

$$-\bar{\mathbf{H}}_{LL}^\dagger \bar{\mathbf{H}}_L^{-1} \bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{i-4,n} + \left(\bar{\mathbf{H}}_L - \bar{\mathbf{H}}_{LL}^\dagger \bar{\mathbf{H}}_L^{-1} \bar{\mathbf{H}}_{LL} - \bar{\mathbf{H}}_{LL} \bar{\mathbf{H}}_L^{-1} \bar{\mathbf{H}}_{LL}^\dagger \right) \mathbf{G}_{i-2,n} - \bar{\mathbf{H}}_{LL} \bar{\mathbf{H}}_L^{-1} \bar{\mathbf{H}}_{LL} \mathbf{G}_{i,n} = 0 \tag{3.47}$$

To simplify the derivation, Equation 3.47 is written in compact form as

$$-\mathbf{A}_{22} \mathbf{G}_{i-4,n} + \mathbf{B} \mathbf{G}_{i-2,n} - \mathbf{A}_{11} \mathbf{G}_{i,n} = 0 \quad (i < n) \tag{3.48}$$

From Equation 3.41, one could have

$$\bar{\mathbf{H}}_{LL}^\dagger \mathbf{G}_{n-2,n} + \bar{\mathbf{H}}_L \mathbf{G}_{n-1,n} + \bar{\mathbf{H}}_{LL} \mathbf{G}_{n,n} = 0 \tag{3.49}$$

i.e.

$$\mathbf{G}_{n-1,n} = -\bar{\mathbf{H}}_L^{-1}\left(\bar{\mathbf{H}}_{LL}^{\dagger}\mathbf{G}_{n-2,n} - \bar{\mathbf{H}}_{LL}\mathbf{G}_{n,n}\right) \tag{3.50}$$

and

$$\bar{\mathbf{H}}_{LL}^{\dagger}\mathbf{G}_{n-1,n} + \bar{\mathbf{H}}_L\mathbf{G}_{n,n} = \mathbf{I} \tag{3.51}$$

Therefore,

$$-\bar{\mathbf{H}}_{LL}^{\dagger}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL}^{\dagger}\mathbf{G}_{n-2,n} + \left(\bar{\mathbf{H}}_L - \bar{\mathbf{H}}_{LL}^{\dagger}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL}\right)\mathbf{G}_{n,n} = \mathbf{I} \tag{3.52}$$

whose compact form is

$$-\mathbf{A}_{22}\mathbf{G}_{n-2,n} + \mathbf{C}\mathbf{G}_{n,n} = \mathbf{I} \tag{3.53}$$

In Equation 3.48 and Equation 3.53, the coefficients are

$$\begin{cases} \mathbf{A}_{11} & = \bar{\mathbf{H}}_{LL}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL} \\ \mathbf{A}_{22} & = \bar{\mathbf{H}}_{LL}^{\dagger}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL}^{\dagger} \\ \mathbf{B} & = \bar{\mathbf{H}}_L - \bar{\mathbf{H}}_{LL}^{\dagger}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL} - \bar{\mathbf{H}}_{LL}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL}^{\dagger} \\ \mathbf{C} & = \bar{\mathbf{H}}_L - \bar{\mathbf{H}}_{LL}^{\dagger}\bar{\mathbf{H}}_L^{-1}\bar{\mathbf{H}}_{LL} \end{cases} \tag{3.54}$$

Physically speaking, Equation 3.53 describes the interactions between the layer n and $n-2^1$ of the left contact. Now, $2^1+1=3$ layers of atoms of the left contact have been included.

By substituting subscript i with $i-4$, $i-2$, and i in Equation 3.48, one can obtain

$$-\mathbf{A}_{22}\mathbf{G}_{i-8,n} + \mathbf{B}\mathbf{G}_{i-6,n} - \mathbf{A}_{11}\mathbf{G}_{i-4,n} = 0 \tag{3.55}$$

$$-\mathbf{A}_{22}\mathbf{G}_{i-6,n} + \mathbf{B}\mathbf{G}_{i-4,n} - \mathbf{A}_{11}\mathbf{G}_{i-2,n} = 0 \tag{3.56}$$

$$-\mathbf{A}_{22}\mathbf{G}_{i-4,n} + \mathbf{B}\mathbf{G}_{i-2,n} - \mathbf{A}_{11}\mathbf{G}_{i,n} = 0 \tag{3.57}$$

respectively, where $i < n$. From Equation 3.55 and Equation 3.57, one could have

$$\mathbf{G}_{i-6,n} = \mathbf{B}^{-1}\left(\mathbf{A}_{22}\mathbf{G}_{i-8,n} + \mathbf{A}_{11}\mathbf{G}_{i-4,n}\right) \tag{3.58}$$

$$\mathbf{G}_{i-2,n} = \mathbf{B}^{-1}\left(\mathbf{A}_{22}\mathbf{G}_{i-4,n} + \mathbf{A}_{11}\mathbf{G}_{i,n}\right) \tag{3.59}$$

which can be inserted into Equation 3.56 to obtain

$$-\mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{22}\mathbf{G}_{i-8,n}+\left(\mathbf{B}-\mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{11}-\mathbf{A}_{11}\mathbf{B}^{-1}\mathbf{A}_{22}\right)\mathbf{G}_{i-4,n}-\mathbf{A}_{11}\mathbf{B}^{-1}\mathbf{A}_{11}\mathbf{G}_{i,n}=0 \quad (i<n) \tag{3.60}$$

Setting $i=n$ in Equation 3.48,

$$-\mathbf{A}_{22}\mathbf{G}_{n-4,n}+\mathbf{B}\mathbf{G}_{n-2,n}-\mathbf{A}_{11}\mathbf{G}_{n,n}=0 \tag{3.61}$$

i.e.

$$\mathbf{G}_{n-2,n}=\mathbf{B}^{-1}\left(\mathbf{A}_{22}\mathbf{G}_{n-4,n}+\mathbf{A}_{11}\mathbf{G}_{n,n}\right) \tag{3.62}$$

can be obtained. Together with Equation 3.53, one could have

$$-\mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{22}\mathbf{G}_{n-4,n}+\left(\mathbf{C}-\mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{11}\right)\mathbf{G}_{n,n}=\mathbf{I} \tag{3.63}$$

If one substitutes the coefficients of Equation 3.60 and Equation 3.63 by following

$$\begin{cases}\mathbf{A}_{11} & := \mathbf{A}_{11}\mathbf{B}^{-1}\mathbf{A}_{11}\\ \mathbf{A}_{22} & := \mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{22}\\ \mathbf{B} & := \mathbf{B}-\mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{11}-\mathbf{A}_{11}\mathbf{B}^{-1}\mathbf{A}_{22}\\ \mathbf{C} & := \mathbf{C}-\mathbf{A}_{22}\mathbf{B}^{-1}\mathbf{A}_{11}\end{cases} \tag{3.64}$$

the compact form

$$-\mathbf{A}_{22}\mathbf{G}_{i-8,n}+\mathbf{B}\mathbf{G}_{i-4,n}-\mathbf{A}_{11}\mathbf{G}_{i,n}=0 \quad (i<n) \tag{3.65}$$

and

$$-\mathbf{A}_{22}\mathbf{G}_{n-4,n}+\mathbf{C}\mathbf{G}_{n,n}=\mathbf{I} \tag{3.66}$$

can be obtained. Physically speaking, Equation 3.66 describes the interactions between the atom layer n and $n-2^2$ of the left contact. Now, $2^2+1=5$ layers of the left contact have been included.

The derivation can start over by using Equation 3.65 and Equation 3.66. In the i^{th} calculation, 2^i+1 layers of atoms of the left contact will included. It is obvious that this iterative algorithm can achieve exponential convergence speed while calculating the surface

```
1. Set $\mathbf{A} := \begin{bmatrix} \bar{\mathbf{H}}_{LL} & 0 \\ 0 & \bar{\mathbf{H}}_{LL}^{\dagger} \end{bmatrix}$, $\mathbf{B} := \bar{\mathbf{H}}_L$, and $\mathbf{C} := \bar{\mathbf{H}}_L$ as initial condition
2. do:
     $\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{22} \end{bmatrix} \mathbf{B}^{-1} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{22} \end{bmatrix}$, $\mathbf{B} := \mathbf{B} - \mathbf{A}_{21} - \mathbf{A}_{12}$, and $\mathbf{C} := \mathbf{C} - \mathbf{A}_{21}$
   while $\max_{i,j} \left| (\mathbf{A}_{21})_{ij} \right| < \delta$
3. the left contact surface Green's function is $\mathbf{C}^{-1}$
```

Algorithm 1: Iterative algorithm to compute the left contact surface Green's function.

Green's functions. From the above derivation, one can summarize the iterative algorithm to efficiently compute the left contact suface Green's function, as shown in Algorithm 1.

Similarly, the iterative algorithm to calculate the right contact surface Green's function can be derived, as shown in Algorithm 2.

```
1. Set $\mathbf{A} := \begin{bmatrix} \bar{\mathbf{H}}_{RR} & 0 \\ 0 & \bar{\mathbf{H}}_{RR}^{\dagger} \end{bmatrix}$, $\mathbf{B} := \bar{\mathbf{H}}_R$, and $\mathbf{C} := \bar{\mathbf{H}}_R$ as initial condition
2. do:
     $\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{22} \end{bmatrix} \mathbf{B}^{-1} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{22} \end{bmatrix}$, $\mathbf{B} := \mathbf{B} - \mathbf{A}_{12} - \mathbf{A}_{21}$, and $\mathbf{C} := \mathbf{C} - \mathbf{A}_{12}$
3. while $\max_{i,j} \left| (\mathbf{A}_{12})_{ij} \right| < \delta$
4. the right contact surface Green's function is $\mathbf{C}^{-1}$
```

Algorithm 2: Iterative algorithm to compute the right contact surface Green's function.

3.4.1.3 Retarded Green's Function

The calculation of Equation 3.3 to obtain the retarded Green's function of the central scattering region is also very computationally demanding. Here, the recursive Green's function method is implemented to reduce the computational cost from $O\left(N_x^3 N_y^3\right)$ to $O\left(N_x^3 N_y\right)$, where N_x is the degree of freedom in one principal layer and N_y is the number of principal layers [56]. The algorithm details are discussed below.

If the scattering region have only one principal layer, Equation 3.6 can be written as

$$\mathbf{A}_{11}\mathbf{G}_{11}^{r} = \mathbf{I} \tag{3.67}$$

So the left connected Green's function of this layer is

$$\mathbf{g}_{11}^{rL} = \mathbf{G}_{11}^{r} \tag{3.68}$$

Now the second layer of the central region is included by using

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{G}_{11}^{r} & \mathbf{G}_{12}^{r} \\ \mathbf{G}_{21}^{r} & \mathbf{G}_{22}^{r} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \tag{3.69}$$

According to the matrix inversion operation

$$\left[\begin{array}{cc|cc} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{I} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} & 0 & \mathbf{I} \end{array}\right]$$

$$\sim \left[\begin{array}{cc|cc} \mathbf{I} & \mathbf{g}_{11}^{rL}\mathbf{A}_{12} & \mathbf{g}_{11}^{rL} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} & 0 & \mathbf{I} \end{array}\right] \qquad \left(\mathbf{g}_{11}^{rL} = \mathbf{A}_{11}^{-1}\right) \tag{3.70}$$

$$\sim \left[\begin{array}{cc|cc} \mathbf{I} & \mathbf{g}_{11}^{rL}\mathbf{A}_{12} & \mathbf{g}_{11}^{rL} & 0 \\ 0 & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{g}_{11}^{rL}\mathbf{A}_{12} & -\mathbf{A}_{21}\mathbf{g}_{11}^{rL} & \mathbf{I} \end{array}\right] \tag{3.71}$$

$$\sim \left[\begin{array}{cc|cc} \mathbf{I} & \mathbf{g}_{11}^{rL}\mathbf{A}_{12} & \mathbf{g}_{11}^{rL} & 0 \\ 0 & \mathbf{I} & -\mathbf{G}_{22}^{r}\mathbf{A}_{21}\mathbf{g}_{11}^{rL} & \mathbf{G}_{22}^{r} \end{array}\right] \qquad \left(\mathbf{G}_{22}^{r} = \left[\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{g}_{11}^{rL}\mathbf{A}_{12}\right]^{-1}\right) \tag{3.72}$$

$$\sim \left[\begin{array}{cc|cc} \mathbf{I} & 0 & \mathbf{g}_{11}^{rL} - \mathbf{g}_{11}^{rL}\mathbf{A}_{12}\mathbf{G}_{21}^{r} & -\mathbf{g}_{11}^{rL}\mathbf{A}_{12}\mathbf{G}_{22}^{r} \\ 0 & \mathbf{I} & \mathbf{G}_{21}^{r} & \mathbf{G}_{22}^{r} \end{array}\right] \qquad \left(\mathbf{G}_{21}^{r} = -\mathbf{G}_{22}^{r}\mathbf{A}_{21}\mathbf{g}_{11}^{rL}\right) \tag{3.73}$$

$$\sim \left[\begin{array}{cc|cc} \mathbf{I} & 0 & \mathbf{G}_{11}^{r} & \mathbf{G}_{12}^{r} \\ 0 & \mathbf{I} & \mathbf{G}_{21}^{r} & \mathbf{G}_{22}^{r} \end{array}\right] \qquad \left(\mathbf{G}_{12}^{r} = -\mathbf{g}_{11}^{rL}\mathbf{A}_{12}\mathbf{G}_{22}^{r};\ \mathbf{G}_{11}^{r} = \mathbf{g}_{11}^{rL} - \mathbf{g}_{11}^{rL}\mathbf{A}_{12}\mathbf{G}_{21}^{r}\right) \tag{3.74}$$

one could obtain $\mathbf{G}^r_{22}$, $\mathbf{G}^r_{21}$, $\mathbf{G}^r_{12}$, and $\mathbf{G}^r_{11}$.

Similarly, the third layer can be added and so on. To sum up, the recursive Green's function method to calculate the retarded Green's function of a N-layer central region is shown in Algorithm 3.

```
1. $\mathbf{g}^{rL}_{11} := \mathbf{A}^{-1}_{11}$

2. for $k = 2, 3, \cdots, N$, calculate
   $\mathbf{g}^{rL}_{k,k} := \left[\mathbf{A}_{k,k} - \mathbf{A}_{k,k-1}\mathbf{g}^{rL}_{k-1,k-1}\mathbf{A}_{k-1,k}\right]^{-1}$

3. $\mathbf{G}^r_{N,N} := \mathbf{g}^{rL}_{N,N}$

4. for $k = N-1, N-2, \cdots, 1$, calculate
   $\mathbf{G}^r_{k+1,k} := -\mathbf{G}^r_{k+1,k+1}\mathbf{A}_{k+1,k}\mathbf{g}^{rL}_{k,k}$
   $\mathbf{G}^r_{k,k} := \mathbf{g}^{rL}_{k,k} - \mathbf{g}^{rL}_{k,k}\mathbf{G}^r_{k,k+1}\mathbf{A}_{k+1,k}$
   $\mathbf{G}^r_{k,k+1} := (\mathbf{G}^r_{k+1,k})^\dagger$
```

Algorithm 3: Recursive algorithm to calculate the retarded Green's function

3.4.1.4 Electron and Hole Green's Functions

Once the retarded Green's function calculation is done, the electron Green's function defined in Equation 3.8) and the hole Green's function defined in Equation 3.9) can be calculated using the recursive Green's function algorithm, as shown in Algorithm 4.

3.4.2 Validation of the Code

To verify the implementation of the NEGF simulator, two examples are calculated. The results obtained by using the NEGF simulator agree well with the published experimental and simulation results. Here, since the model only contains a few atoms and the transport is largely elastic, the validation simulations are done without including the inelastic electron-phonon scattering.

```
1. $\mathbf{g}_{11}^{inL} := \mathbf{g}_{11}^{rL} \mathbf{\Sigma}_{11}^{in} (\mathbf{g}_{11}^{rL})^\dagger$

2. for $k = 2, 3, \cdots, N$, calculate
   $\mathbf{g}_{k,k}^{inL} := \mathbf{g}_{k,k}^{rL} \left[ \mathbf{\Sigma}_{k,k}^{in} + \mathbf{A}_{k,k-1} \mathbf{g}_{k-1,k-1}^{inL} (\mathbf{A}_{k,k-1})^\dagger \right] (\mathbf{g}_{k,k}^{rL})^\dagger$

3. $\mathbf{G}_{N,N}^{n} := \mathbf{g}_{N,N}^{inL}$

4. for $k = N-1, N-2, \cdots, 1$, calculate
   $\mathbf{G}_{k+1,k}^{n} := -\mathbf{G}_{k+1,k+1}^{r} \mathbf{A}_{k+1,k} \mathbf{g}_{k,k}^{inL} - \mathbf{G}_{k+1,k+1}^{n} (\mathbf{A}_{k,k+1})^\dagger (\mathbf{g}_{k,k}^{rL})^\dagger$
   $\mathbf{G}_{k,k}^{n} := \mathbf{g}_{k,k}^{inL} + \Theta - \Xi$
   $\mathbf{G}_{k,k+1}^{n} := (\mathbf{G}_{k+,k}^{n})^\dagger$
   where $\Theta = \mathbf{g}_{k,k}^{rL} \mathbf{A}_{k,k+1} \mathbf{G}_{k+1,k+1}^{n} (\mathbf{A}_{k,k+1})^\dagger (\mathbf{g}_{k,k}^{rL})^\dagger$
   and $\Xi = \mathbf{g}_{k,k}^{inL} (\mathbf{A}_{k+1,k})^\dagger (\mathbf{G}_{k,k}^{r})^\dagger + \mathbf{G}_{k,k+1}^{r} \mathbf{A}_{k+1,k} \mathbf{g}_{k,k}^{inL}$

5. for $k = 1, 2, \cdots, N$, calculate
   $\mathbf{G}_{k+1,k}^{p} := -\mathbf{G}_{k+1,k}^{n} + i \times [\mathbf{G}_{k+1,k}^{r} - (\mathbf{G}_{k,k+1}^{r})^\dagger]$
   $\mathbf{G}_{k,k}^{p} := -\mathbf{G}_{k,k}^{n} + i \times [\mathbf{G}_{k,k}^{r} - (\mathbf{G}_{k,k}^{r})^\dagger]$
   $\mathbf{G}_{k,k+1}^{p} := -\mathbf{G}_{k,k+1}^{n} + i \times [\mathbf{G}_{k,k+1}^{r} - (\mathbf{G}_{k+1,k}^{r})^\dagger]$
```

Algorithm 4: Recursive algorithm to calculate the electron and hole Green's functions

3.4.2.1 Example 1. 1D Gold Chain.

The transmission coefficient of 1D gold chain as shown in Figure 3.5 is calculated. The result is demonstrated in Figure 3.6, which means the conductance is

$$G = G_0 T(\epsilon_F) = G_0 \tag{3.75}$$

as $T(\epsilon_F) = 1$. This agrees well with the literature [82, 18, 104].

The I-V curve of the gold chain is calculated by using the Green's function code, as shown in Figure 3.7. The results in Figure 3.7 are listed in Table 3.1. It is obvious that the results of the code agree well with the existing literature.

Figure 3.5: Model of the 1D gold chain.

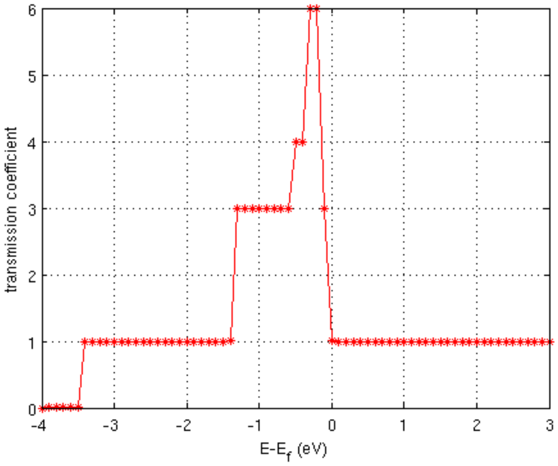


Figure 3.6: Transmission coefficient of 1D gold chain.

Table 3.1: Conductance of the gold chain.

bias v (V)	current i (μA)	NEGF resistance $R = v/i$ ($k\Omega$)	ref. resistance $R_{\text{ref}}(k\Omega)$	relative error $\frac{R-R_{\text{ref}}}{R_{\text{ref}}}$
0.01	0.7739	12.922	12.9	+0.1698 %
0.02	1.5485	12.916	12.9	+0.1227 %
0.03	2.3230	12.914	12.9	+0.1099 %
0.04	3.0985	12.909	12.9	+0.0722 %

3.4.2.2 Example 2. Au-CO system

The I-V curves of the gold chain (Figure 3.5) and the gold chain with a CO molecule attached (Figure 3.9) are calculated and shown in Figure 3.8. It is obvious that the CO molecule saliently reduces the conductance of the gold chain. This tendency agrees well with the existing literature.

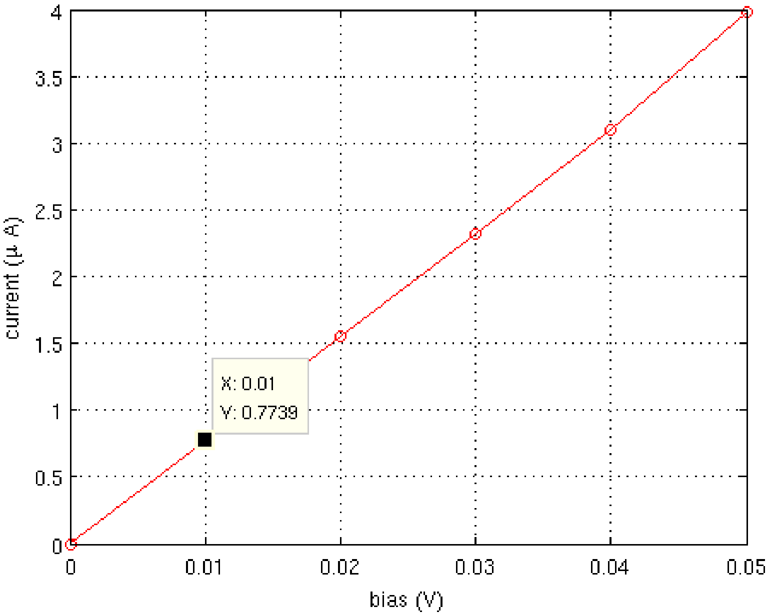


Figure 3.7: I-V curve of the gold chain.

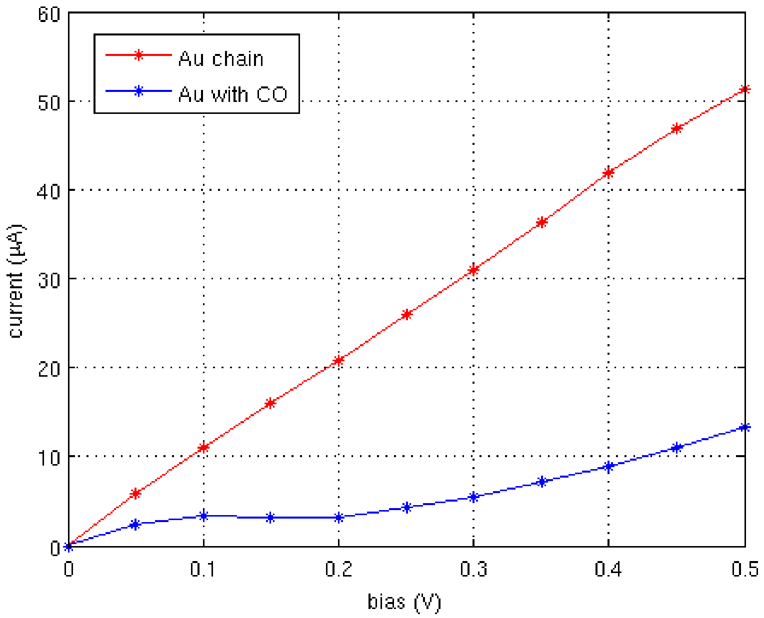


Figure 3.8: I-V curves of the two test systems.

The transmission coefficient of the gold chain with a CO molecule attached is also calculated. The result is shown in Figure 3.10, which indicates that the conductance is dramatically reduced to $G \approx 0$. This calculation reveals that the CO molecule prohibits the delocalization of gold chain electron wave function. So the conductance is significantly suppressed [82, 18].

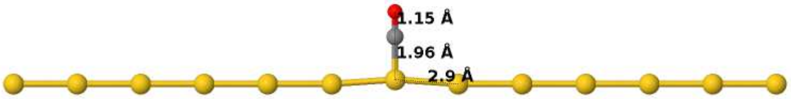


Figure 3.9: Model of the 1D gold chain with CO molecule attached.

3.4.3 Running Time Analysis

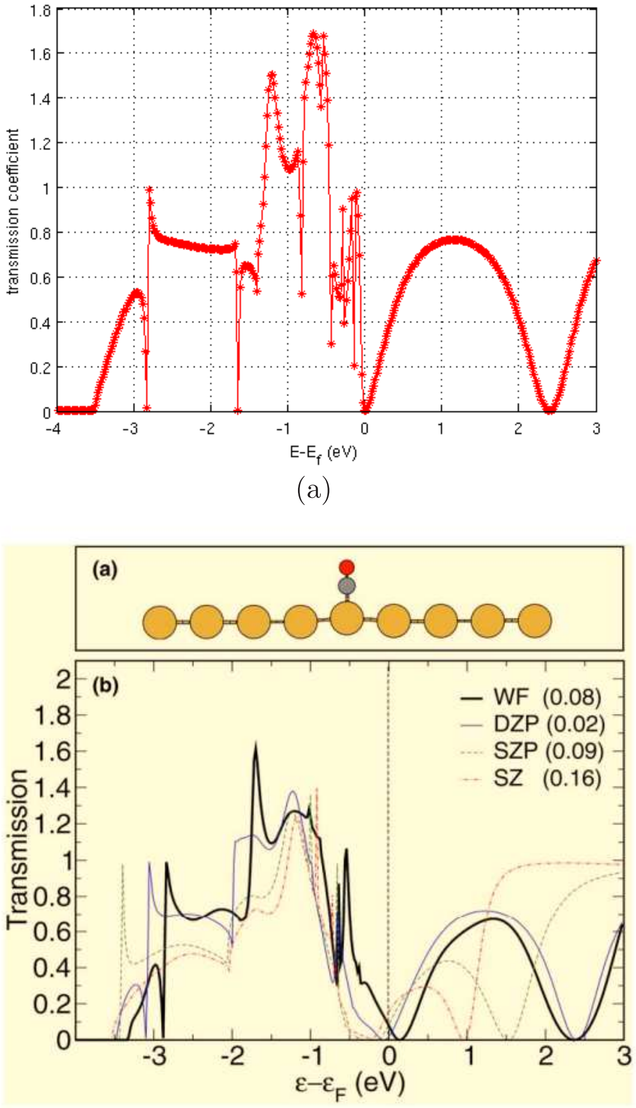
The NEGF simulations are very computationally demanding. In order to offer more quantitative insights of the total simulation time and its components, the timing analysis is performed in this section.

3.4.3.1 Test System

In the timing analysis, the TiN-GeTe-TiN sandwich electron transport simulation models as shown in Figure 3.1 are calculated using the NEGF code.

During the simulations, the single-zeta plus polarization (SZP) pseudo-atomic orbitals (PAO) are used for the Ti and N atoms. Since the $3d^2 4s^2$ electrons are included for Ti atoms and the $2s^2 2p^3$ electrons are included for N atoms, there are 9 PAO per each Ti or N atom. The double-zeta plus polarization (DZP) PAO are used for the Ge and Te atoms. Since the $4s^2 4p^2$ electrons are included for Ge atoms, the $5s^2 5p^4$ electrons are included for Te atoms, there are 13 PAO per each Ge or Te atom.

In the simulations, the number of TiN atoms is fixed at 512, which means there are 216 TiN atoms (6 atomic layers) in the source and 216 TiN atoms (6 atomic layers) in the drain, respectively. Six systems are simulated, in which 128, 160, 192, 224, 256, 288 GeTe atoms are contained between source and drain respectively. In these six models, the number of



(a)

(b)

Figure 3.10: Transmission coefficient of 1D gold chain with CO molecule attached. (a): result obtained by using the Green's function code with DZP basis Hamiltonian and overlap matrices from SIESTA; (b): benchmark calculation from [104].

GeTe atomic layers are 8, 10, 12, 14, 16, 18, respectively. These GeTe channel lengths of these models vary from about 2.5 nm to about 6 nm.

Since the number of the GeTe atoms vary, the size of the Hamiltonians are 5552, 5968, 6384, 6800, 7216, and 7632, respectively. The Hamiltonian of the 18 GeTe atomic layer model is shown in Figure 3.2. When computing the contact surface Green's functions, 4 TiN atomic layers is treated as a principal layer, which means the size of the contact self energy matrix $\mathbf{\Sigma}^r_{S/D}$ is fixed at 1296 for all of the six models.

As shown in Figure 3.2, the system is partitioned into principal layers, in order to accelerate the calculation by taking advantage of the tridiagonal matrix structure of the Hamiltonian.

3.4.3.2 Simulation Time Components

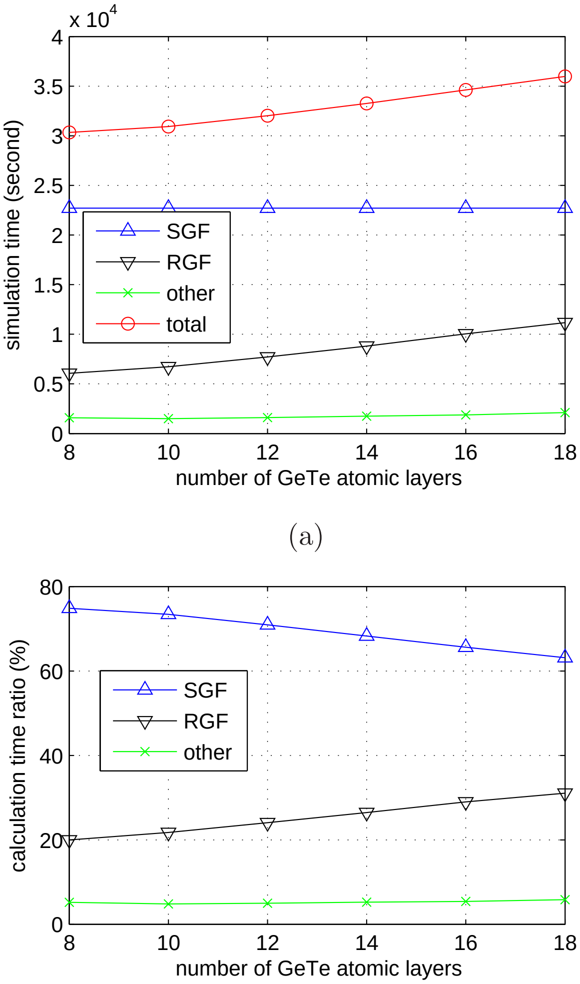
It is well known that the surface Green's function (SGF) calculation (Section 3.4.1.2) and the recursive Green's function (RGF) calculation (Section 3.4.1.3 and Section 3.4.1.4) are the most computationally expensive parts of the NEGF code.

As shown in Figure 3.11, the SGF calculation time is fixed and consumes most of the simulation time. This is due to the facts that the size of the contact self energy matrices are fixed in all six models; and the size of the Hamiltonian blocks that represents the TiN contacts is much larger than those represent the GeTe channel (Figure 3.2).

It can be seen from Figure 3.11 that the RGF calculation time depends linearly on the GeTe channel length, which is a manifestation of the $O(N_{xy}^3 N_z)$ computational complexity of the RGF algorithms. Here, N_z represents the number of principal layers along the transport direction; and N_{xy} represents the number of PAO inside one principal layer. In Figure 3.11, the "other" represents the simulation time spent to read input information, pre-process data, post-process data, and to output information.

3.4.3.3 Discussion

Regarding the total computational time and its components of the NEGF simulator, it worth noting that:



(a)

(b)

Figure 3.11: Running time analysis of the NEGF code (201 energy points are calculated): (a) the total simulation time and its components; and (b) the ratio of these components in the total simulation time. On the x axis, $N_L = 8$ means $N_H = 5552$ and $N_A = 560$; $N_L = 10$ means $N_H = 5968$ and $N_A = 592$; $N_L = 12$ means $N_H = 6384$ and $N_A = 624$; $N_L = 14$ means $N_H = 6800$ and $N_A = 656$; $N_L = 16$ means $N_H = 7216$ and $N_A = 688$; and $N_L = 18$ means $N_H = 7632$ and $N_A = 720$, Here, N_L is number of GeTe atomic layers, N_H is dimension of the Hamiltonian; and N_A is number of atoms in the simulation model.

1. The computational complexity of the SGF algorithm is $O(N_s^3)$, where N_s is the dimension of the contact self energy matrix. If the N_s is chosen to be large, the SGF calculation will consume a significant fraction of the total calculation time.

2. The computational complexity of the RGF algorithm is $O(N_{xy}^3 N_z)$, which is much faster than the direct matrix inversion ($O(N_{xy}^3 N_z^3)$). Since the RGF algorithm is about N_z^2 faster than the direct matrix inversion, the system should be partitioned into as many principal layers as possible, to accelerate the RGF calculation.

3. In Figure 3.11, the SGF and the RGF are calculated for only one time at each energy point, since electron-phonon scattering is turned off. As a consequence, the SGF calculation consumes much more time than the RGF calculation. However, in the NEGF simulations that include the scattering, e.g. electron-phonon scattering, the SGF calculation needs to be run only once but the RGF calculation needs to be run many times (typically 10^1 times) to achieve convergence (Figure 3.4). In this case, the RGF calculation will consume much more time than the SGF calculation.

3.5 Simulation Results

In this section, the simulation results of mesoscopic NEGF simulations are presented. The results under equilibrium low-bias operation condition and non-equilibrium operation condition are shown in Section 3.5.1 and Section 3.5.2, respectively. In Section 3.5.3, the results pertaining to electron-phonon scattering are presented.

3.5.1 Equilibrium Low-Bias Operation

In this section, the equilibrium low-bias electron transport properties are simulated.

3.5.1.1 Assumptions and Approximations

The following assumptions are made in Section 3.5.1:

1. Since the operation bias is small, the device is assumed to be in equilibrium condition. This means that the solultion of the Poisson equation (module 6) and the SCF loop

of potential in Figure 3.4 are omitted.

2. Since the focus of this section is the transport properties of the ultra-thin PCM film, whose thickness is smaller than the mean free path (MFP) of electrons, the electron-phonon scattering is ignored. This means that the module 4 and the SCF loop of electron-phonon scattering in Figure 3.4 are ignored.

3.5.1.2 Low-Bias Conductance

The L_z-dependent G_c and G_a computed using Equation 3.31, together with the ON/OFF ratio G_c/G_a, are shown in Figure 3.12. It worth noting that the thickest film (roughly 6 nm thick) in our ab initio simulations shows two orders of magnitude ON/OFF ratio. Interestingly, this agrees with the state-of-art scaling experiments, in which the thinnest $Ge_2Sb_2Te_5$ film (6 nm thick) sandwiched by TiN electrodes also shows two orders of magnitude ON/OFF ratio, though the PCM stoichiometry differs.

While the experimental ON/OFF ratio data for sub-6 nm PCM ultrathin films is not available, it is computed here as shown in Figure 3.12. It can be seen that the two orders of magnitude ON/OFF ratio can be retained even if the ultrathin film thickness is reduced down to about 40 (13 atomic layers). However, if the ultrathin film thickness is further reduced, the ON/OFF ratio drops rapidly. When the thickness is reduced below about 33 (11 atomic layers), the ON/OFF ratio becomes smaller than 10, which is an ON/OFF ratio value typically needed for reliable read operation in useful devices.

It worth noting that sub-20 Å PCM nanostructures still keep phase change properties, making write operations possible in ultrascaled PCM devices. Therefore, here it is pointed out, for the first time according to our knowledge, that the aggressive scaling might be limited by the loss of adequate ON/OFF ratio, which makes it difficult to reliably perform read operation.

3.5.1.3 Transmission Coefficient

Before simulating the transmission coefficient, the impact of the k-point sampling on the calculation accuracy is investigated. To test the influence of k-point sampling, the $n_x \times$

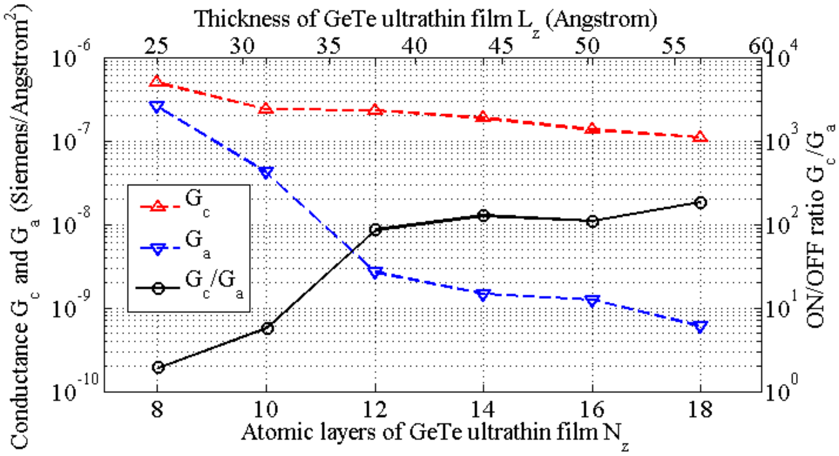


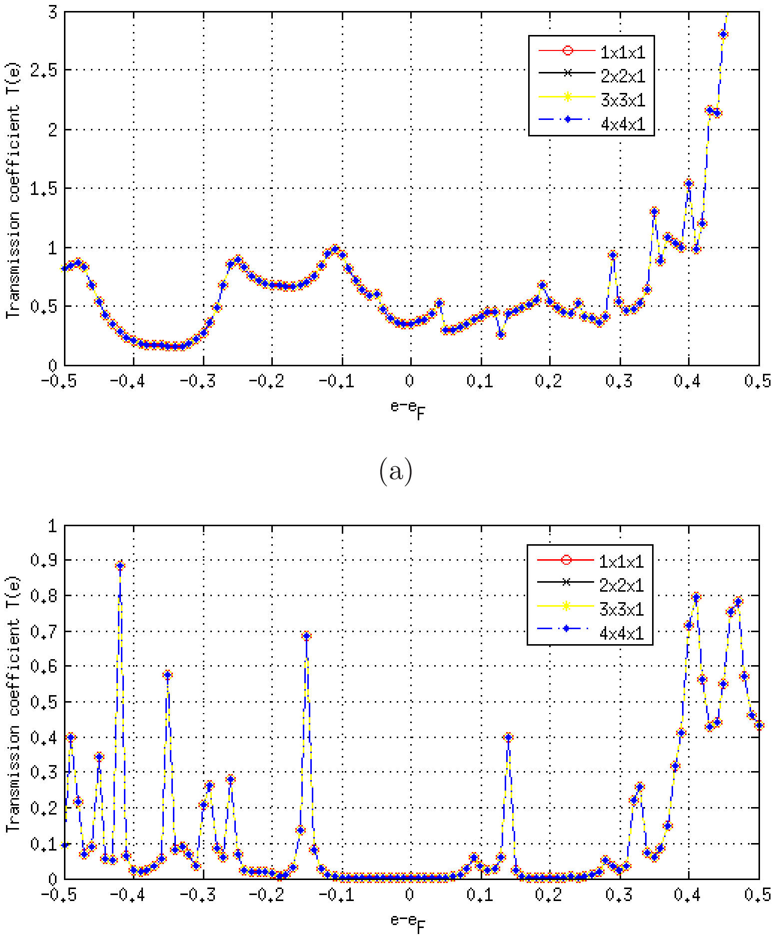
Figure 3.12: Conductance of *c*-GeTe (G_c) and *a*-GeTe (G_a) ultrathin films, and the ON/OFF ratio G_c/G_a.

$n_y \times n_z$ Monkhorst-Pack k-points sampling algorithm [77] is used to compute transmission coefficient

$$T(\epsilon) = \frac{1}{\Omega} \int_{\mathrm{1BZ}} T(k;\epsilon) dk \approx \sum_i w_i \times T(k_i;\epsilon) \tag{3.76}$$

where Ω is the area of the first Brillouin zone and w_i is the numerical quadrature weighting coefficient corresponding to the Monkhorst-Pack sampling point k_i. Among all models used in the simulation, the supercell length in z direction ranges from 49.21 Å (crystalline system when N_z=8 atomic layers) to 84.75 Å (amorphous system when N_z = 18 atomic layers). As all supercell lengths in the z direction are very large, n_z is constantly set as 1. Supercell length in x and y directions are comparatively smaller (L_x=L_y=12.06 Å), so we choose n_x=n_y=1,2,3,4, in order to test the Brillouin zone sampling precision. When n_x=n_y=1, the $n_x \times n_x \times n_z$ Monkhorst-Pack k-points sampling is reduced to Γ-point sampling. The calculation results are shown in Figure 3.13. It can be seen that Γ-point sampling has satisfactory precision compared against refined $n_x \times n_y \times 1$ Monkhorst-Pack sampling results. Its known that the Γ-point sampling result is more reliable for larger supercells. Figure 3.13 indicates that the supercell size in x and y directions is large enough to justify the choice of Γ-point sampling.

Its known that the Γ-point sampling result is more reliable for larger supercells. Figure 3.13



(a)

(b)

Figure 3.13: Transmission coefficients of crystalline (a) and amorphous (b) TiN-GeTe-TiN systems (N_z=12 atomic layers) calculated using $n_x \times n_y \times n_z$ Monkhorst-Pack k-point sampling (n_x=n_y=1,2,3,4 and n_z=1).

indicates that the supercell size in x and y directions (L_x=L_y=12.06 Å) is large enough to justify the choice of Γ-point sampling. so, the Γ-point sampling is used in the simulation to alleviate CPU intensity.

In order to understand the underlying device physics that causes the sharp reduction of ON/OFF ratio shown in Figure 3.12, the $T(\epsilon)$ and the $g(\epsilon)$ of Equation 3.31 in Figure 3.14 are investigated. The $T(\epsilon)$ and $g(\epsilon)$ provide energy resolved insight – the former (latter) unravels the electron transmission probabilities (contribution to the total conductance) as a function of the electron energy ϵ. Here, only the $T(\epsilon)$ and the $g(\epsilon)$ results for three thickness values (N_z=10,12,14) are presented, since the sharp ON/OFF ratio reduction mainly occurs in this critical thickness range. Physically, the *c*-GeTe and *a*-GeTe ultrathin films, which are sandwiched by metal TiN electrode as shown in Figure 3.1, form potential barriers for electron transport during read operation. When the lattice structure is changed between the amorphous phase and the crystalline phase, the electronic structure of the barrier is altered, causing the change of the electron transmission behavior as shown in Figure 3.14(a)-(c).

It is obvious that near the Fermi level ϵ_F, the amorphous phase $T(\epsilon)$ is always smaller than the crystalline phase $T(\epsilon)$, making the conductance of the amorphous phase G_a smaller than that of the crystalline phase G_c (Figure 3.12). If the ultrathin films are scaled down to 12 atomic layers, the amorphous phase still exhibits clear band gaps (Figure 3.14(b)-(c)) and the electrical conductance is mainly due to the electron transport via intra-gap donor-like and acceptor-like states (Figure 3.14(e)-(f)). However, if the ultrathin film is further scaled, the amorphous phase band gap disappears (Figure 3.14(a)), leading to both quantitatively significant increase and qualitatively sharp change of $g(\epsilon)$ (Figure 3.14(d)). Now the conductance is no longer determined by the distinct donor-like states and acceptor-like states, but dominated by transport near the Fermi level of the metal electrodes.

In contrast to the significant change of electron transport behavior when the amorphous phase film is aggressively scaled, one could observe only a slight quantitative increase of electron transmission probabilities when the crystalline phase ultrathin film is scaled to the same dimensions (Figure 3.14(a)-(c)). The crystalline films in the size scale of our interest here show no band gap, and the conductance is always dominated by electron transport near the Fermi level. As a consequence, when the ultrathin films are aggressively scaled, G_a increases significantly but G_c increases only slightly (Figure 3.12). Therefore, the loss of adequate ON/OFF ratio G_c/G_a is mainly caused by the rapid increase of G_a with decreasing film thickness.

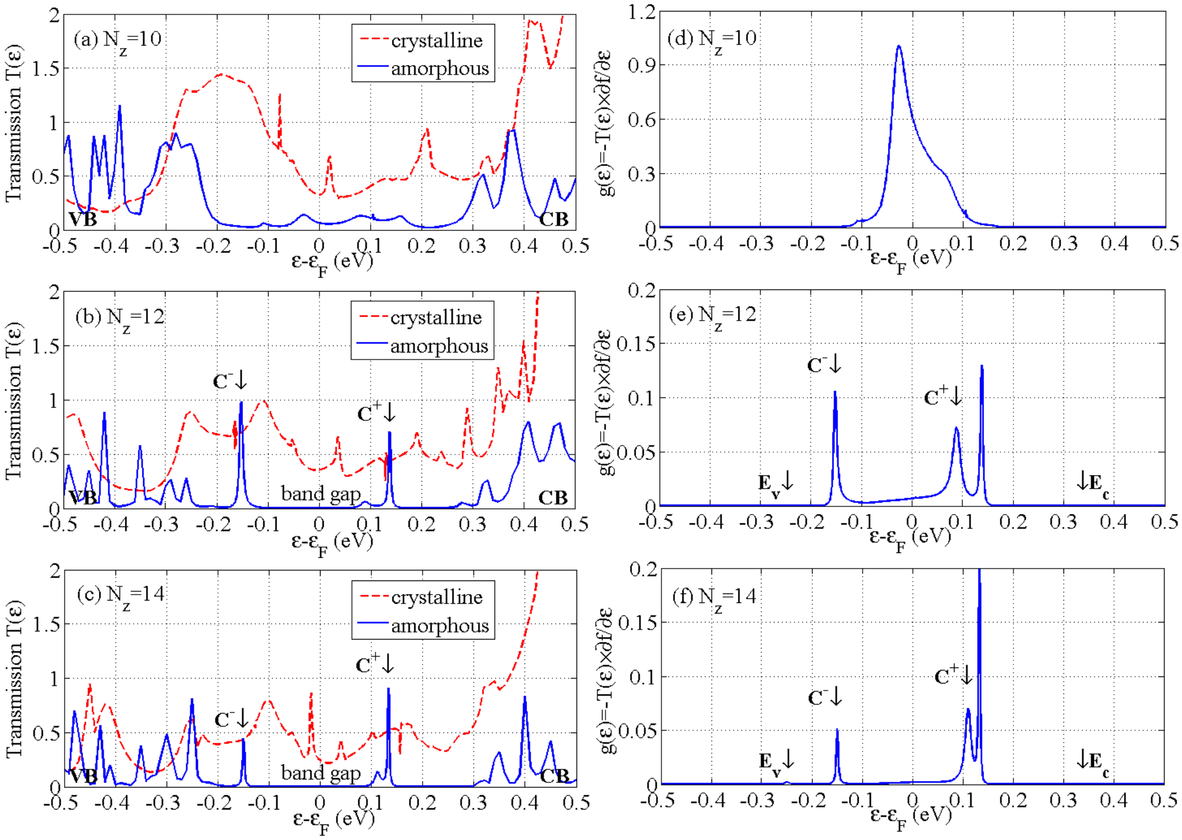


Figure 3.14: The transmission coefficient $T(\epsilon)$ and contribution to conductance as a function of electron energy $g(\epsilon) = -T(\epsilon) \times \frac{\partial f(\epsilon)}{\partial \epsilon}$ of the ultrathin films. The donor-like states (C^+) and acceptor-like states (C^-) are clearly seen in plots of $g(\epsilon)$ and local density of states in Figure 3.15. N_z: number of atomic layers in the ultrathin films; CB: conduction band; VB: valence band; E_c: conduction band edge; E_v: valence band edge; and ϵ_F: Fermi level.

3.5.1.4 Local Density of States

The difference in scaling behavior between *c*-GeTe and *a*-GeTe ultrathin films evokes our interest to inspect their electronic structure difference, in order to further understand the governing device physics. From Figure 3.12, it can be seen that the ON/OFF ration has a turning point at N_z=12. For $N_z \geq 12$, there is a roughly two orders of magnitude ON/OFF ration; but for $N_z < 12$, the ON/OFF ratio quickly drops. Here, the LDOS at this critical thickness point (N_z=12) is presented.

In the discussion here, a principal layer (PL) is defined as 32 Ge and Te atoms, as shown in Figure 3.1 (for crystalline phase, one PL means two atomic layers of *c*-GeTe). The ultrathin film with N_z=12 consists of six PLs stacking in the z-direction from source to drain. Each of the six PLs is labeled PLn where n varies from 1 to 6. It can be seen from Figure 3.15 that in the middle region of *a*-GeTe ultrathin films, which is far away from the metal electrodes (PL_3 and PL_4), the LDOS is very small near the Fermi level, and there is a clear band gap with pronounced intra-gap donor-like and acceptor-like states (Figure 3.15(c)-(d)). These states are responsible for the conduction in Figure 3.14(b). In the *a*-GeTe region that is spatially close to TiN electrodes (PL_1, PL_2, PL_5, and PL_6), the band gap disappears and the LDOS is large near the Fermi level (Figure 3.15(a)-(b) and Figure 3.15(e)-(f)) due to the hybridization with the electronic states inside the metal electrodes. These metal induced gap states (MIGS) evanescently penetrate two PLs into the *a*-GeTe (Figure 3.15(a)-(b) and Figure 3.15(e)-(f)) and decay in the deeper *a*-GeTe region (Figure 3.15(c)-(d)).

In comparison, the MIGS penetrate all the way across the ultrathin film in *c*-GeTe thin films (Figure 3.15). As a result, near the Fermi level, the *c*-GeTe LDOS is much larger than the *a*-GeTe LDOS in the region far away from the TiN electrodes (Figure 3.15(c)-(d)). This LDOS magnitude difference of the two phases explains why the$T(\epsilon)$ of *c*-GeTe is larger than that of *a*-GeTe, and why G_c is always larger than G_a (Figure 3.12).

3.5.1.5 Role of metal induced gap states

It worth mentioning that, in the two regions (PL_1 and PL_2; PL_5 and PL_6) close to the TiN electrodes, the *a*-GeTe LDOS is in general comparable to, though slightly smaller than, the *c*-GeTe LDOS near the Fermi level (Figure 3.15(a)-(b) and Figure 3.15(e)-(f)). If the ultrathin films are thick enough ($N_z > 12$), these two MIGS-controlled high-LDOS conductive regions are spatially separated by a MIGS-free low-LDOS resistive region (PL_3 and PL_4) when the PCM state is amorphous, making the amorphous phase much more resistive than the crystalline phase. But if the ultrathin films are aggressively scaled ($N_z < 10$), these two regions become closer, bridging a path for electrons to transport through the amorphous

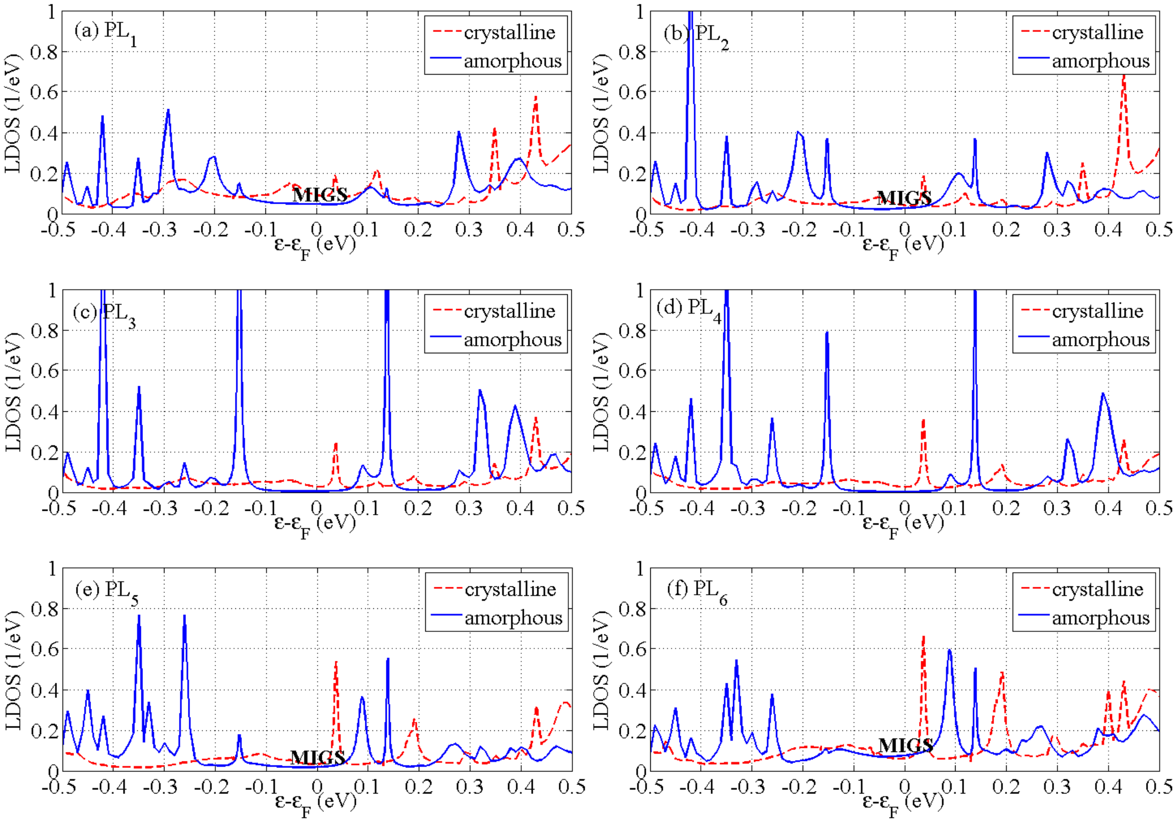


Figure 3.15: Local Density of State (LDOS) of different principal layers (PL$_n$, as defined in Figure 3.1) in the GeTe ultrathin film (N_z=12). MIGS: metal induced gap states in amorphous phase.

phase. As a consequence, in the amorphous phase transmission probability (Figure 3.14(a)), tunneling via MIGS near the Fermi level (Figure 3.14(d)), and conductance G_a Figure 3.12 increases significantly, leading to the loss of adequate ON/OFF ratio.

As mentioned above, it is observed that the MIGS penetration depth in c-GeTe is larger than that in a-GeTe. This phenomenon can be attributed to the well known fact that c-GeTe band gap is smaller than a-GeTe band gap and in general MIGS penetrate deeper into the semiconductors with smaller band gap at the metal-semiconductor interface.

3.5.1.6 Analysis and Discussion

The above analysis on the role of MIGS during scaling indicates that if the PCM with larger amorphous phase band gap is used, the MIGS penetration depth can be reduced and the two MIGS-controlled regions can be better separated. Therefore, it might be promising to push the ultimate scaling limit of PCM to even smaller dimension by using PCM which has a larger amorphous phase band gap, or by applying other techniques that can effectively reduce the MIGS penetration depth. It also worth noting that the larger amorphous band gap will lead to increased threshold voltage required for set operation. This is a typical example where there is a need to strike a balance between two competing device performance metrics. Although it has been shown that ultrathin films can be scaled to at least 33 Angstroms from the viewpoint of keeping adequate ON/OFF ratio for reliable read operation, the scaling scenario can also be shaped by many other technical issues and device performance metrics, e.g. crystallization temperature, operation energy, device endurance, and fabrication cost. Their roles in determining the PCM scaling scenario deserve future research efforts.

3.5.2 Non-Equilibrium Sub-Threshold Operation

In this section, the sub-threshold electron transport properties of *a*-GeTe ultra-thin films are investigated by using NEGF simulations.

The simulation results reproduce the trends in measured electron transport properties, e.g. current-voltage curve, intra-bandgap donor-like and acceptor-like defect states, and *p*-type conductivity. The underlying physical mechanism of electron transport in ultra-scaled *a*-PCM is unraveled. It is found that, though the current-voltage curve of the ultra-scaled *a*-PCM resembles that of the bulk *a*-PCM, their physical origins are different. Unlike the electron transport in bulk *a*-PCM, which is governed by the Poole-Frenkel effect, the electron transport in ultra-scaled *a*-PCM is largely dominated by tunneling transport via intra-bandgap donor-like and acceptor-like defect states. It is found that the current-voltage characteristics in the sub-threshold region have linear and exponential regions.

3.5.2.1 Assumptions and Approximations

The following assumptions and approximations are made in Section 3.5.2:

1. Since the thickness of the PCM ultra-thin film of our interets here is smaller than the electron MFP, the electron-phonon scattering is ignored. This means that the module 4 and the electron-phonon scattering SCF loop in Figure 3.4 are ignored.

2. Since the sub-threshold operation condition is the focus and the distance between source and drain is small, the potential is assumed to vary linearly from srouce to drain. This means that the module 6 and the potential SCF loop in Figure 3.4 are ignored in the calculation.

3.5.2.2 Sub-threshold Current-Voltage Curve

The current-voltage curves, conductance, and differential conductance results obtained from the NEGF simulations are shown in Figure 3.16.

It can be found that the simulation results yield the linear region (when v is smaller than about 0.05 V, see insert of Figure 3.16 (a) and Figure 3.16 (b)) and the exponential region (when v is larger than about 0.05 V, see Figure 3.16), the trends of which are similar to measured a-PCM current-voltage curves. The linear and exponential dependence of current on bias can be seen more clearly from the conductance and the differential conductance results (Figure 3.16 (b)).

3.5.2.3 Transmission Coefficient

In order to study the underlying physical mechanism behind the current-voltage curve shape shown in Figure 3.16 (a), the contribution of electrons with energy ϵ to the total current is investigated, as shown in Figure 3.17. It can be seen that:

1. There is a clear bandgap near the Fermi level, indicating the semiconducting nature of a-GeTe. Figure 3.17 (c) shows that the transmission is not significantly dependent on the applied bias at small voltages but Figure 3.17 (d) shows that the transmission changes at larger applied biases.

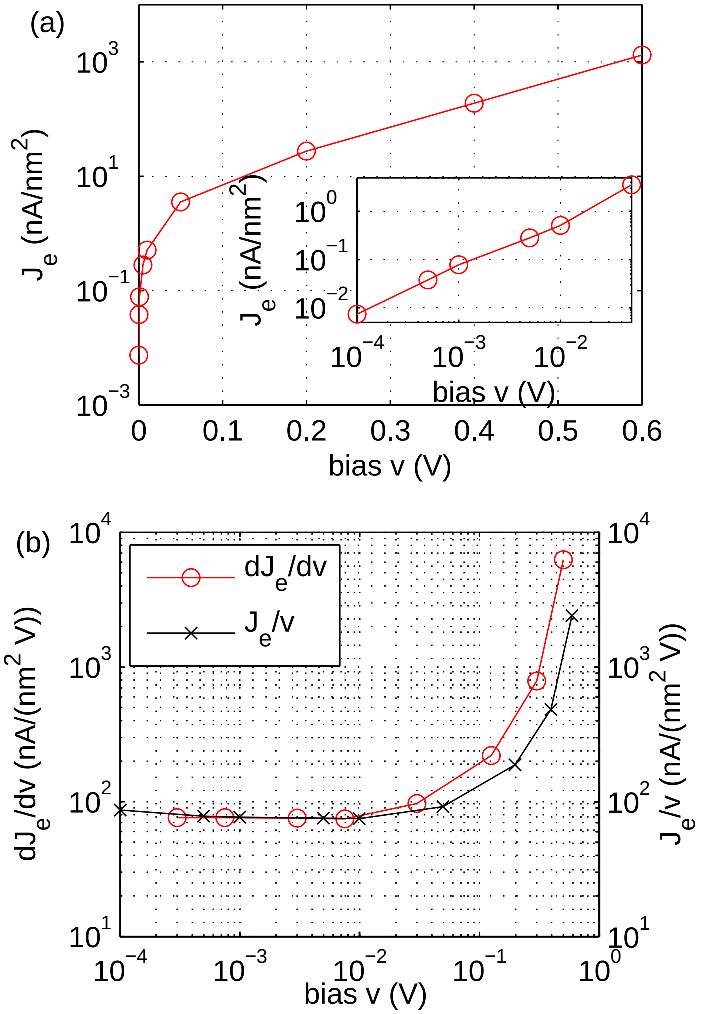


Figure 3.16: The sub-threshold transport properties of a-GeTe ultra-thin film. (a) Current density (J_e) versus applied bias (v); insert: zoom of small bias region. (b) Differential conductance (dJ_e/dv) and conductance J_e/v calculated from (a).

2. The electron transport across the a-GeTe ultra-thin film is dominated by the intra-bandgap donor-like and acceptor-like defect states. This is seen both in the plots of transmission (Figure 3.17 (c) and (d)) and current flowing per unit energy in Figure 3.17 (a) and (b). It follows from our calculations that the electron transport via acceptor-like states is more significant than via donor-like states (Figure 3.17

(a) and (b)), indicating the p-type conductivity of a-GeTe. The trends in our results for the transport properties agree well with experiments [94].

3.5.2.4 Influence of Temperature

In this section, the influence of the temperature on electron transport is analyzed. The current density is computed using Equation 3.36, where the transport is assumed to be elastic. Figure 3.18 shows the simulation results of the temperature dependent current-voltage curve of 6 nm thick a-GeTe ultrathin film sandwiched by TiN electrodes.

From Figure 3.18, it can be seen that the change of temperature exerts little influence on the current density when the voltage is large (say larger than 0.4 V). Comparatively speaking, the impact of temperature on the current density is larger when the voltage is small (say 0.1 to 0.2 V). This phenonmenon can be explained by using the underlying transport mechanism of the ultra-scaled PCM nanostructures. As shown in Figure 3.17, the conduction is dominated by the transport via the intra-bandgap defect states, which locate at 0.1-0.2 eV away from the Fermi level. So, according to Equation 3.36, the magnitude of the current density is determined by how the bias window $(f_S - f_D)$ includes the transmission peaks at these defect states. When the voltage is large, the transmission is very small at the energies near $\epsilon_F \pm v/2$, where $(f_S - f_D)$ is significantly altered by temperature (v is bias and ϵ_F is Fermi level). When the voltage is small, both the trasmission and the change of $(f_S - f_D)$ by temperature are large near $\epsilon_F \pm v/2$, leading to large dependence of current density on the temperature.

3.5.2.5 Large-Bias Results

The elastic transport simulation is performed for bias that is larger than the measured threshold voltage (for 6 nm PCM ultra-thin film, the threshold voltage is measured to be about 0.6 V [49]). As shown in Figure 3.19, the simulation fails to capture the threshold switching and the snap-back of current-voltage curve. The reason is that the possible underlying mechanisms of the threshold switching, e.g. microscopic crystallization and/or charge redistribution, etc., are not included in the simulation model. Including these effects

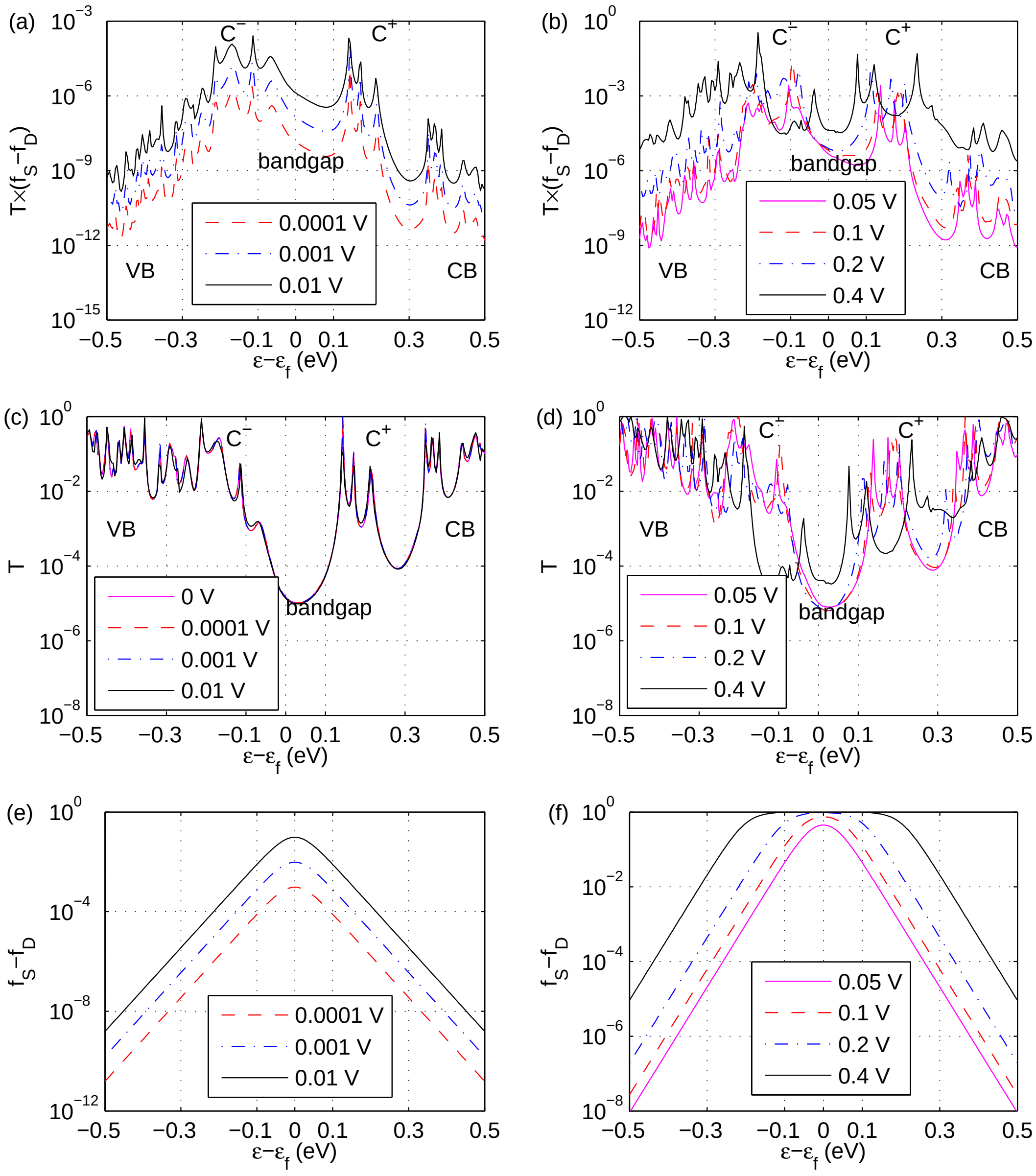


Figure 3.17: Energy resolved transport properties of a-GeTe ultrathin film. (a) and (b): Current flowing per unit energy in units of $2q_e/\hbar L_x L_y$, which is $T(\epsilon)(f_S(\epsilon) - f_D(\epsilon))$; (c) and (d): Transmission coefficient, which is $T(\epsilon)$; and (d) and (e): Fermi functions difference between the source and drain contacts, which is $f_S(\epsilon) - f_D(\epsilon)$. Here ϵ is the electron energy and ϵ_f is the equilibrium Fermi energy.

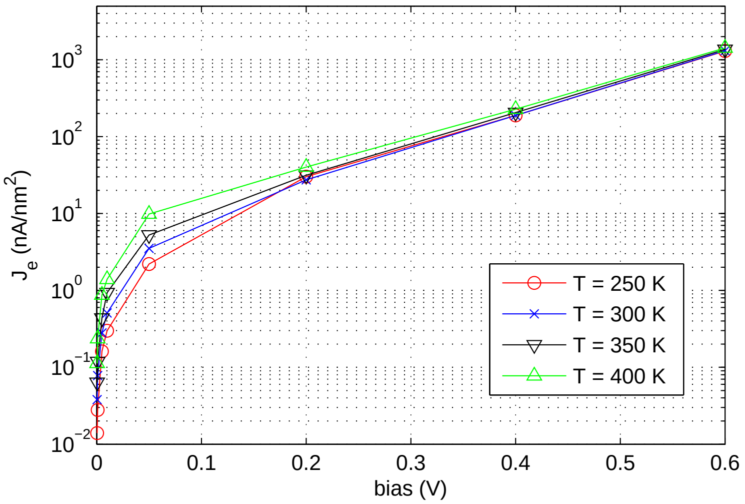


Figure 3.18: Temperature dependence of sub-threshold current of *a*-GeTe.

needs more sophisticated models and more computationally expensive calculations, which deserve future research efforts.

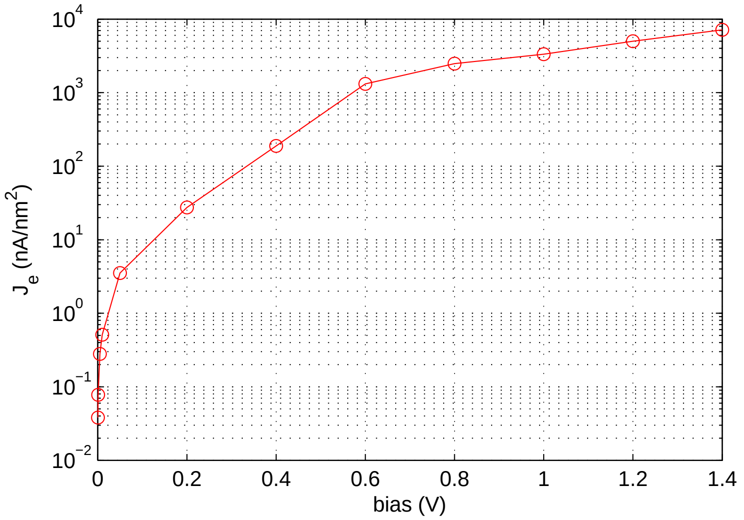


Figure 3.19: Simulated large-bias current-voltage curve of *a*-GeTe, which failed to capture the measured snap-back characteristics.

3.5.2.6 Analysis and Discussion

When the bias is small, the simulations show that the current-voltage curve is linear, in agreement with experiments. The microscopic reason is that the transmission coefficient T is roughly independent of bias (Figure 3.17 (c)). The linear region of the current-voltage curve at small voltages (Figure 3.16) is mainly due to the linear dependence of the difference in the Fermi functions $f_S - f_D$ with energy as shown in Figure 3.17 (e). It also can be found that the electrons carry maximum current at energies corresponding to the acceptor-like defect states ϵ_{C-} and donor-like defect states ϵ_{C+}. This is most clearly demonstrated in the plot of current density per unit energy, $T(\epsilon)(f_S(\epsilon) - f_D(\epsilon))$, as shown in Figure 3.17 (a). The Figure 3.17 (c) and (d) show that the transmission is large outside of the bandgap and smaller at the location of the defect energies. However, as a result of the exponential dependence of the Fermi functions on energy, the current is primarily carried at energies corresponding to the intra-bandgap defect states (Figure 3.17 (a) and (b)).

When the applied bias is increased beyond about 0.05 V, the current depends exponentially on the applied bias (Figure 3.16). The physical origin of the exponential dependence is now discussed. In the exponential region, the calculations show that there are significant shifts in the defect state energy levels with bias, which modifies the location of the peaks of the transmission coefficient (Figure 3.17 (d)) and hence the current density (Figure 3.17 (b)). It can be found that the exponential dependence of current with applied bias results from:

1. the bias induced change of the transmission coefficient (Figure 3.17 (d));
2. the exponential dependence of the term $f_S - f_D$ at defect energies contributing to current (Figure 3.17 (f)).

In conclusion, it is found that the underlying governing mechanism of transport properties in ultra-scaled *a*-GeTe thin films is different from that of the bulk *a*-GeTe in the sub-threshold region. In the deep nanometer scale, the transport is no longer governed by thermal emission of trapped electrons over potential barriers. Instead, the tunneling transport mediated by the intra-bandgap donor-like and acceptor-like defect states explain the experimental trends of linear and exponential regions of the current versus voltage curve in the sub-threshold region of the ultra-scaled PCM devices. The linear dependence at low voltages arises mainly from the dependence of the Fermi functions on the bias. The exponential dependence at higher voltages arises from an interesting interplay between the change in transmission peaks with applied bias and the accompanying large changes in the values of the Fermi functions at these transmission peaks.

3.5.3 Electron-Phonon Scattering

Compared to the simple Ohmic transport behavior of crystalline PCM, the amorphous PCM shows very sophisticated and interesting transport properties (e.g. linear/exponential current-voltage curve shapes under low/large bias in the sub-threshold region; and the snap-back S-shape current-voltage curve after threshold switching), which has spawned intensive research to understand the underlying governing physics. Though many models have been proposed to explain these interesting phenomena [34], one of the most popular and well accepted models is based on the Poole-Frenkel Law [58, 35, 36, 37], in which the inelastic scattering of electrons (e.g. by phonons) plays a pivotal role in shaping the electron transport properties of the bulk PCM (tens of nm or larger) .

Despite the importance of electron-phonon scattering in the PCM devices, the quantum mechanical level and atomistic scale understanding of this process is still missing. In this section, the impact of inelastic electron-phonon scattering on the electron transport properties of the PCM is investigated by using the NEGF simulations, in order to offer more knowledge of this important physical process.

3.5.3.1 Assumptions and Approximations

In the NEGF simulations of the electron-phonon scattering presented in Section 3.5.3, the following assumptions and approximations are made:

1. Since the channel length of the electron transport simulation model used here is very small (6 nm), the potential distribution, which is largely governed by the source and drain, is assumed to vary linearly between the two electrodes. As a consequence, the

solution of the Poisson equation and the potential convergence SCF loop are omitted in Figure 3.4.

2. In all the electron-phonon scattering calculations, the phonon energy is set as 20 meV by following the experimental measurements and numerical simulations of the GeTe, which is the PCM to be investigated here [30, 100].

3. The inelastic in-scattering (out-scattering) of the electrons due to the scattering by the phonons is accounted for by using Equation 3.16 (Equation 3.17). This assumes that the electron-phonon coupling is treated in a uniform manner (by using a phenomenological constant D_{inel}), regardless of the different properties of different phonons.

4. The phonons are approximately treated as an infinite reservior where its distribtuion is governed by the Bose-Einstein distribtution law. This assumes that the phonons are not influenced by the electrons.

Since the NEGF simulations with the electron-phonon scattering are very computationally expensive and the experimental data for the phonon-dependent electron-phonon coupling is not available, the above assumptions and approximations are made to enable the modeling and analysis.

3.5.3.2 Estimation of Scattering Strength

The objective of this section is to determine the electron-phonon scattering strength, which is represented by the phenomenological constant D_{inel} in Equation 3.16 and Equation 3.17, to be used in the following PCM NEGF simulations.

The rhombohedral crystalline GeTe model [21] (Section 4.3.2.1), which contains 192 GeTe atoms is used to estimate the D_{inel}. Along the transport direction (z direction), there are 12 atomic layers in the model. In each atomic layer in the xy plane, which is perpendicular to the transport direction, there are 4×4 atoms. First, the DFT simulation (Section 4.1.1) is applied to obtain the Hamiltonian matrix (as shown in Figure 3.20) and the overlap matrix. Then, the NEGF simulation (Section 3.2.2) is applied to obtain various tranport properties. As shown in Figure 3.20, the model can be partitioned into 6 principal layers, where within each principal layer there are 2 atomic layers (i.e. 32 atoms). Each of the two electrodes consists of two atomic layers (i.e. 32 atoms). The contact self energy matrices are evaluated using the SGF algorithm; and the retarded Green's function, the electron Green's function, and the hole Green's function are evaluated using the RGF algorithm Section 3.4.1.

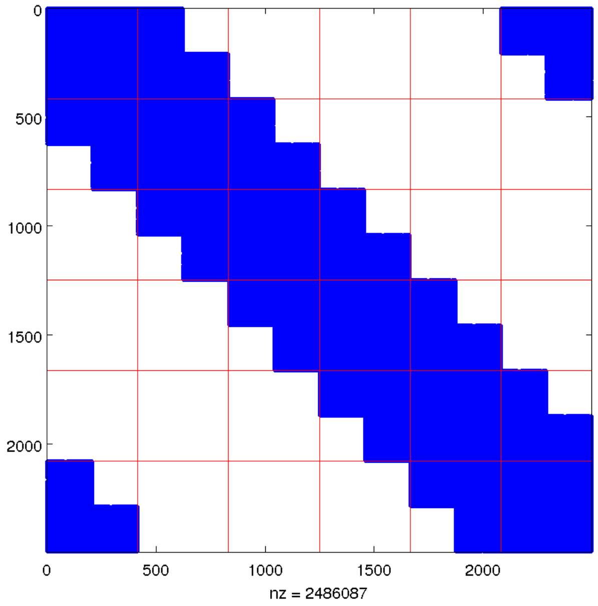


Figure 3.20: Hamiltonian of *c*-GeTe for MFP calculation.

Before using this *c*-GeTe model to estimate the D_{inel}, its validity is verified by computing the transmission coefficient and the density of states (DOS), as shown in Figure 3.21 and Figure 3.22, respectively. It can be seen that the *c*-GeTe model exhibits a 0.24 eV band gap near the Fermi level. This is in agreement with the experimentally measured value 0.20 eV [19], showing that the ab initio *c*-GeTe simulation model used here can fairly reproduce the *c*-GeTe electronic strucuture.

When using the *c*-GeTe model to estimate the electron MFP (Equation 3.37), zero bias is applied; the effective mass (Equation 3.38) is set as $0.8 \times m_e$ (which is taken from measurements [9]), where m_e is the free electron mass; the conduction band minimum E_c

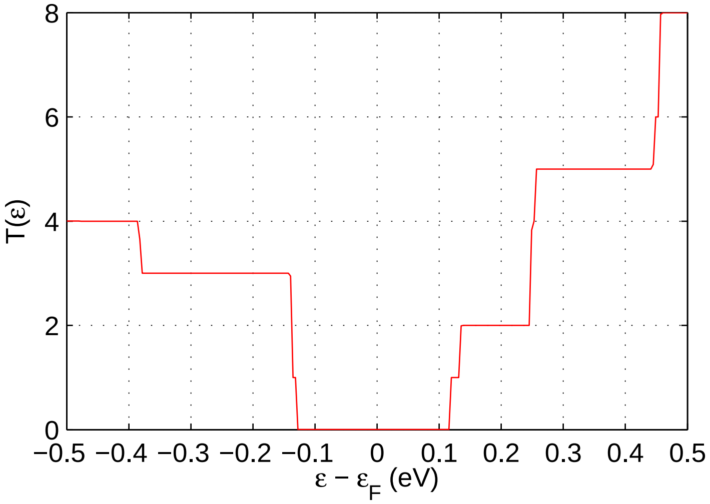


Figure 3.21: Transmission of the *c*-GeTe.

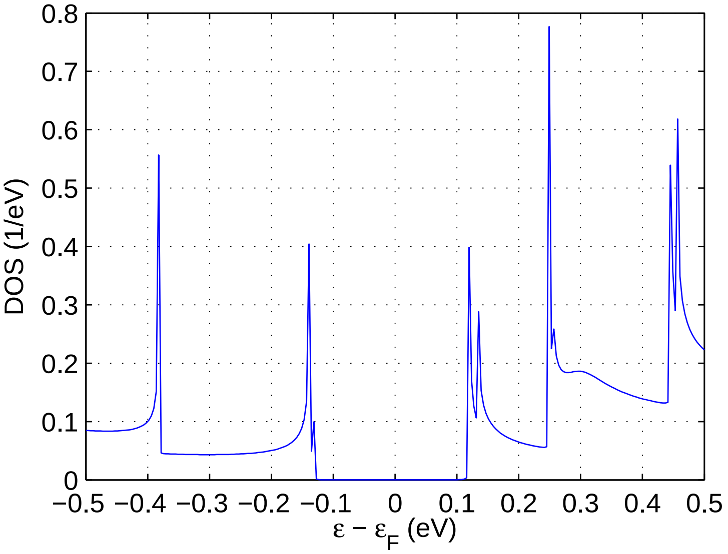


Figure 3.22: Density of States of the *c*-GeTe.

(Equation 3.38) is taken from the transmission simulation results (Figure 3.21); the DOS (Equation 3.39) is obtained from the ab initio simulations (Figure 3.22); and the tranport properties (with different D_{inel} values set in the simulation) along the z direction are investigated using NEGF.

Physically, D_{inel} controls the scattering strength and hence the electron mean free path

(MFP). So, each D_{inel} results in an electron MFP, which is computed using the methodology introduced in Section 3.2.4.5. Typically, the value of D_{inel} is around 10^{-2} eV^2 [52, 126]. Here, the D_{inel} is tuned around the neighborhood of this value. As shown in Figure 3.23, when the D_{inel} is increased from 10^{-3} eV^2 to 10^{-1} eV^2, the electron MFP is decreased monotonously from about 3 μm to about 2 Å. As a matter of fact, the electron mean free path (MFP) in the PCM is estimated to be about tens of nm by using experimental measurements [129, 94]. From Figure 3.23, it can be seen that when $D_{inel} \approx 0.6 \times 10^{-2}$ eV^2, the electron MFP is about tens of nm.

From the analysis above, it can be seen that the value $D_{inel} = 0.6 \times 10^{-2}$ eV^2 can phenomenologically represent the electron-phonon coupling strength in terms of the electron MFP. Therefore, this value is adopted in the following electron-phonon calculations.

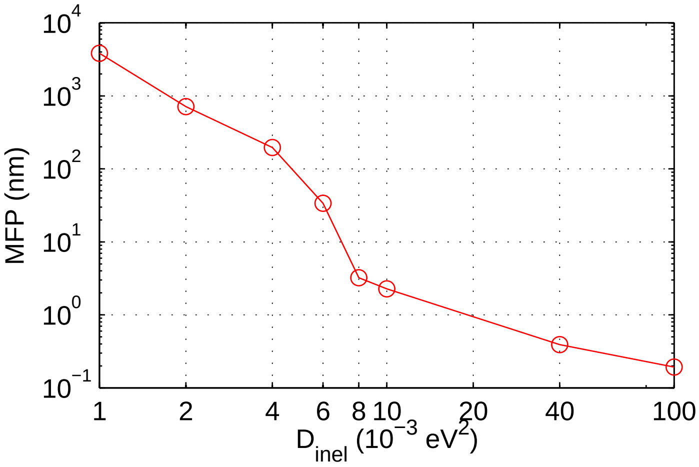


Figure 3.23: MFP of the *c*-GeTe under different D_{inel} values.

3.5.3.3 Scattering Rate

The energy dependent scattering rate that is evaluated using Equation 3.39 is plotted in Figure 3.24. It can be observed directly that the scattering rate is increased monotonously as the phennomenological parameter D_{inel} is increased. It also can be seen that when the $D_{inel} \approx 0.006$ eV^2 is chosen to fit the electron MFP to the experimental measurements, the corresponding scattering rate is about 10^{12} s^{-1}. This provides a rough estimation of the scattering time (about 1 ps or so).

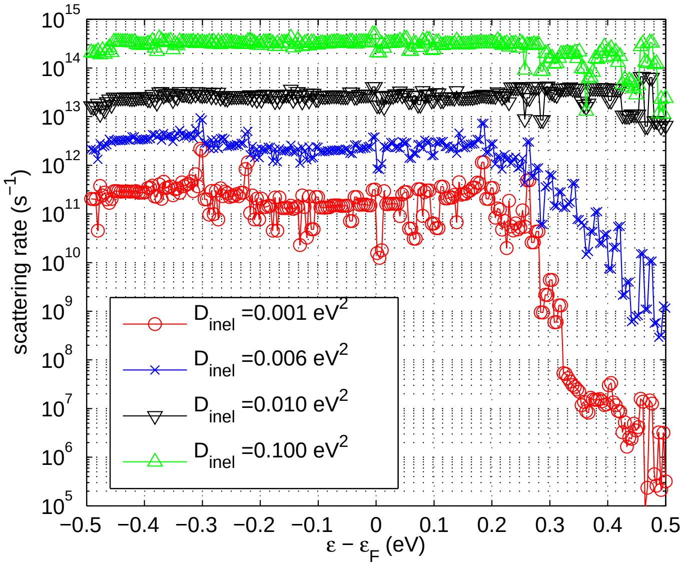


Figure 3.24: Energy-dependent electron-phonon scattering rate of *c*-GeTe.

3.5.3.4 Sub-threshold Current-Voltage Curve

As mentioned above, the quantum mechanical and atomistic scale understanding of the electron-phonon scattering on the *a*-GeTe electron transport properties is still missing, especially for the ultra-scaled PCM nanostructures. In order to investigate this issue, the sub-threshold current-voltage curves of the *a*-GeTe is simulated, as shown in Figure 3.25.

It can be seen that the electron-phonon interaction exerts little impact on the shape of the 6 nm thick *a*-GeTe current-voltage curve. The current-voltage curve still keeps the linear shape in the low bias region (insert of Figure 3.25) and the exponential shape in the large bias region (Figure 3.25). Furthermore, compared to the purely elastic tranport ($D_{inel} = 0$ eV^2), the current values are kept nearly the same after turning on the inelastic electron-phonon scattering ($D_{inel} = 0.006$ eV^2). This implies that the trapped electrons are not effectively excited by the scattering of phonons to participate in the tranport, which is physically reasonable since the electron MFP (tens of nm) is much larger than the channel length of the ultra-scaled PCM devices (6 nm).

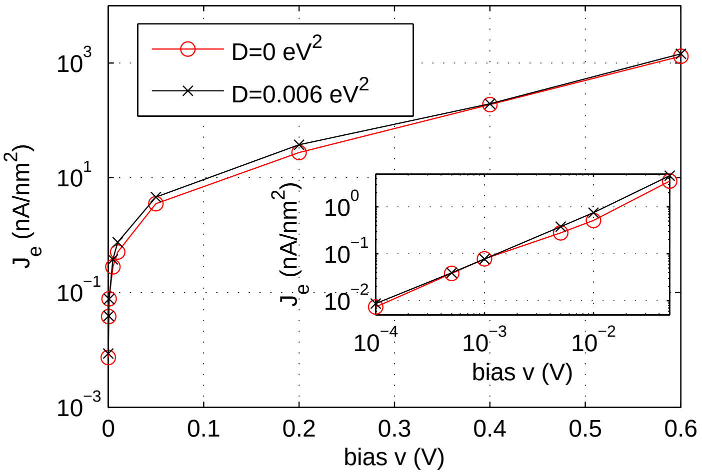


Figure 3.25: Current-voltage curves of the 6 nm a-GeTe (Figure 3.1) without ($D_{inel} = 0$ eV^2) and with ($D_{inel} = 0.006$ eV^2) inelastic electron-phonon scattering.

3.5.3.5 Electron Energy Relaxation

In order to investigate the impact of the inelastic electron-phonon scattering on the electron energy relaxation during the transport, the energy-resolved and position-dependent current density (integrand of Equation 3.34) is plotted in Figure 3.26 (bias $v = 0.2$ V) and Figure 3.27 (bias $v = 0.6$ V). It can be seen from Figure 3.26 and Figure 3.27 that the inelastic electron-phonon scattering effectively broadens the current density peaks, due to the broadening of the density of states. However, the energy distribution of the current remains nearly the same in different positions in the channel, indicating very week energy relaxation due to inelastic electron-phonon scattering in the 6 nm a-GeTe.

3.5.3.6 Analysis and Discussion

From the electorn-phonon scattering analysis presented above, it can be seen that, though the inelastic electron-phonon scattering plays an important role in bulk PCM electron tranport process, its exerts very small impact on the electron tranport properties of the ultra-

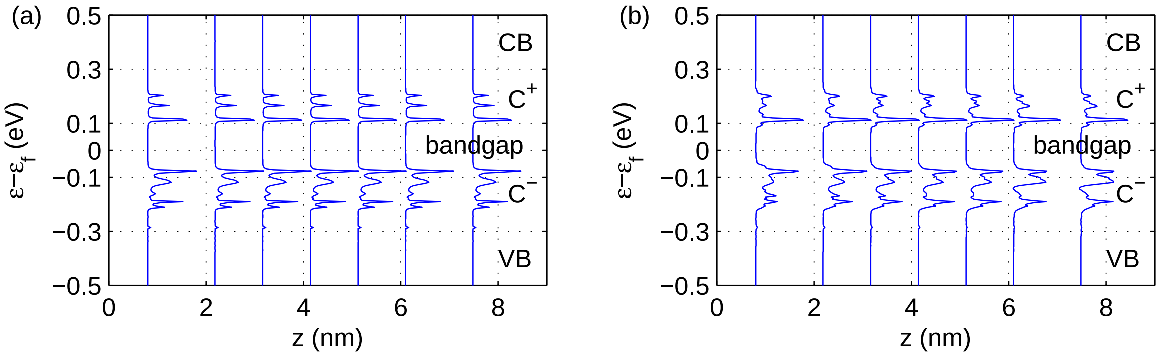


Figure 3.26: Current density as a function of position z (Figure 3.1) and electron energy ϵ, when the bias is 0.2 V. (a) $D_{inel} = 0$ eV^2; and (b) $D_{inel} = 0.006$ eV^2. VB: valence band; CB: conduction band; C^+: donor-like states; and C^-: acceptor-like states.

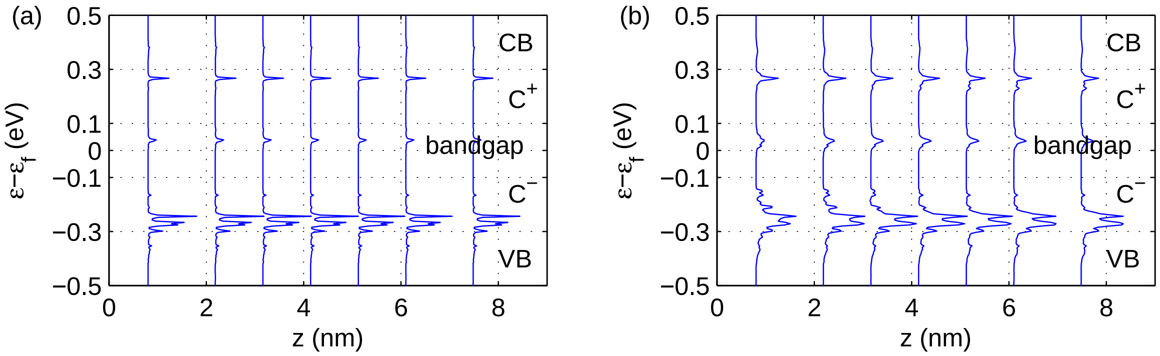


Figure 3.27: Current density as a function of position z (Figure 3.1) and electron energy ϵ, when the bias is 0.6 V. (a) $D_{inel} = 0$ eV^2; and (b) $D_{inel} = 0.006$ eV^2. VB: valence band; CB: conduction band; C^+: donor-like states; and C^-: acceptor-like states.

scaled PCM nanostructures. The reason is that, in the ultra-scaled PCM nanostructures, the electorn MFP is much larger than the channel length. So, the electrons transport through the PCM nanostructures largely in an elastic way.

Physically, this means that the Ohm's law and Joule heating will cease to be valid in the deep nanoscale PCM nanostructures. As a consequence, the Joule heating cannot be generated inside the PCM active region of ultra-scaled PCM devices.

Chapter 4

MICROSCOPIC SIMULATION

In microscopic simulations, both amorphization and crystallization processes are simulated from first principle, in order to investigate RESET and SET operations of PCM device.

In Section 4.1, the numerical algorithms are briefly introduced. In Section 4.2, the assumptions and approximations of the microscopic analysis are described. In Section 4.3 and Section 4.4, the amorphization and crystallization processes are studied.

4.1 Numerical Algorithm

Ab initio density functional theory (DFT), DFT-based ab initio molecular dynamics (AIMD), and DFT-based conjugate gradient (CG) relaxation, which are implemented in the DFT package SIESTA, are used to simulate the amorphization and crystallization processes.

4.1.1 Density Functional Theory

4.1.1.1 Single-Particle Hamiltonian

In the DFT calculations, the Kohn-Sham single-particle Schrödinger equation

$$\hat{H}\left(\mathbf{r}\right)\psi_i\left(\mathbf{r}\right) = \epsilon_i \psi_i\left(\mathbf{r}\right) \tag{4.1}$$

is solved, where ϵ_i and ψ_i are eigenvalues and eigenfunctions [87]. Here,

$$\hat{H} = \hat{T} + \hat{V}_{PS} + \hat{V}_H + \hat{V}_{xc} \tag{4.2}$$

where $\hat{T}$ represents the kinetic energy, $\hat{V}_{PS}$ is the pseudopotential which represents the interaction between the electron and the pseudoatoms (nuclei and the core electrons), $\hat{V}_H$ represents the Coulombic interaction between the valence electrons, $\hat{V}_{xc}$ represents the exchange-correlation effects.

4.1.1.2 Pseudo-Atomic Orbital Representation

The wave function $\psi_i(\mathbf{r})$ is expressed as a linear combination of the pseudo-atomic orbitals (PAO) $\{\phi_\nu(\mathbf{r})\}$, i.e.,

$$\psi_i(\mathbf{r}) = \sum_\nu c_{i\nu}\phi_\nu(\mathbf{r}) \tag{4.3}$$

Inserting equation 4.3 into equation 4.1, one can obtain

$$\sum_\nu c_{i\nu}\hat{H}(\mathbf{r})\phi_\nu(\mathbf{r}) = \epsilon_i \sum_\nu c_{i\nu}\phi_\nu(\mathbf{r}) \tag{4.4}$$

and

$$\sum_\nu c_{i\nu} \underbrace{\int d^3\mathbf{r}\phi^*_\mu(\mathbf{r})\hat{H}(\mathbf{r})\phi_\nu(\mathbf{r})}_{\mathbf{H}_{\mu\nu}} = \epsilon_i \sum_\nu c_{i\nu} \underbrace{\int d^3\mathbf{r}\phi^*_\mu(\mathbf{r})\phi_\nu(\mathbf{r})}_{\mathbf{S}_{\mu\nu}} \tag{4.5}$$

i.e.,

$$(\epsilon_i\mathbf{S} - \mathbf{H})\boldsymbol{\Psi}_i = 0 \tag{4.6}$$

where the column vector $\boldsymbol{\Psi}_i = \{c_{i\nu}\}$ is the wave function ψ_i in the PAO representation. The ab initio description of the system is contained in $\mathbf{H}$ and $\mathbf{S}$.

4.1.2 Ab Initio Molecular Dynamics

4.1.2.1 Born-Oppenheimer Approximation

The AIMD algorithm used in this thesis is based on the DFT and the Born-Oppenheimer approximation. The AIMD simulation consists of two steps:

1. Given the positions of the atomic nuclei, the electronic Schrödinger equation is solved by using the DFT to obtain the total energy and the inter-atomic forces.

2. Given the inter-atomic forces, the atomic nuclei positions are updated by integrating the Newton's equation

$$M_A\frac{d\mathbf{R}_A}{dt} = \nabla E \tag{4.7}$$

   where M_A and $\mathbf{R}_A$ are the mass and position of atom A, and E is the energy obtained by using DFT. [72].

These two steps iterates until a predefined number of steps is done.

4.1.2.2 Thermoestat and Barostat

The AIMD algorithm introduced in Section 4.1.2.1 is able to simulate microcanonical (NVE) ensemble. In the AIMD simulations of PCM, the system is either canonical (NVT) ensemble or isothermal-isobaric (NPT) ensemble. To simulate the NVT ensemble and the NPT ensemble, the temperature and pressure need to be controlled.

To control the temperature of the simulation system, the Nosé-Hoover thermostat is applied in the AIMD simulations [84, 33]. To control the internal pressure of the simulation system, the Parrinello-Rahman barostat is applied in the AIMD simulations [88].

4.1.3 Conjugate Gradient Relaxation

The conjugate gradient (CG) algorithm is used to relax the atomic lattice structure, in order to find the most energetically favorable lattice configuration of the PCM systems [40].

4.1.4 Simulation Parameters

In the AIMD, CG, and DFT simulations, Γ-point is used to sample the Brillouin Zone; the plane wave cutoff is chosen to be 100 Ry; the generalized gradient approximation (GGA) of Perdew, Burke, and Ernzerhof (PBE) [89] is used to approximate the exchange-correlation energy; the double-zeta (single-zeta) plus polarization PAO is used for Ge and Te (Ti and N) atoms; and the CG relaxation convergence criteria is set as 40 meV/Å. The AIMD timestep is chosen as 5 fs.

4.2 Assumptions and Approximations

The simulation and analysis presented in Chapter 4 are based on the following assumptions and approximations:

1. In the DFT and AIMD simulations, the many body problem is approximately described by using single particle Kohn-Sham equation and the PBE general gradient approximation (GGA) is used to approximate the exchange-correlation effects [89].

2. In the DFT and AIMD simulations, only the valence electrons are simulated. The inner shell core electrons are approximated by using the pseudopotential theory.
3. In the DFT and AIMD simulations, the periodic boundary conditions are applied. For crystalline PCM, there is no problem to do so. For amorphous PCM, the application of the periodic boundary conditions will impose artificial long range order on the system.
4. In the DFT and AIMD simulations of the material interfaces, only the [100] interface is investigated. The interfaces of other lattice orientations deserves future research efforts.

4.3 Amorphization Simulation

It is important to study the amorphization process of PCM, since it is the underlying mechanism of the RESET operation of the PCM-based devices. This section investigates the thermally activated amorphization by using the melt-quench AIMD simulations. The atomic coordinates of the *a*-PCM obtained in this section are used to create the mesoscopic transport simulation models, as discussed in Section 3.1.

4.3.1 Modeling Methodology

4.3.1.1 Melt Simulation

The amorphization process of the PCM is calculated by using melt-quench AIMD simulation, which consists of three consecutive steps. In the first step, the crystalline GeTe is heated up from room temperature to 1100K in 10 ps, to obtain the liquid GeTe. The temperature profile of the melt AIMD simulation is shown in Figure 4.1.

4.3.1.2 Quench Simulation

In the second step, the liquid GeTe is quenched from 1100K to room temperature in 15 ps, to obtain the amorphous GeTe. In the third step, the amorphous GeTe is relaxed using conjugate gradient (CG) algorithm, to obtain the final amorphous GeTe atomic coordinates. The temperature profile of the quench AIMD simulation is shown in Figure 4.2

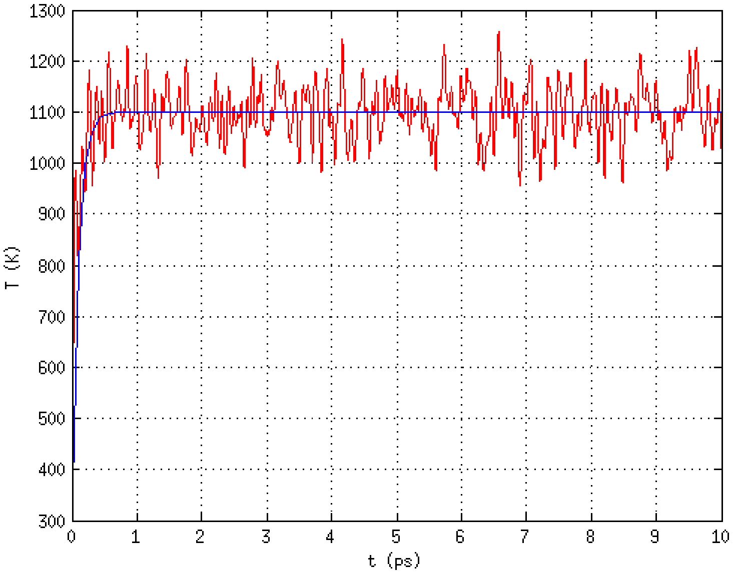


Figure 4.1: Temperature profile of the melt AIMD simulation. Blue line: target temperature; red line: actual AIMD termperature.

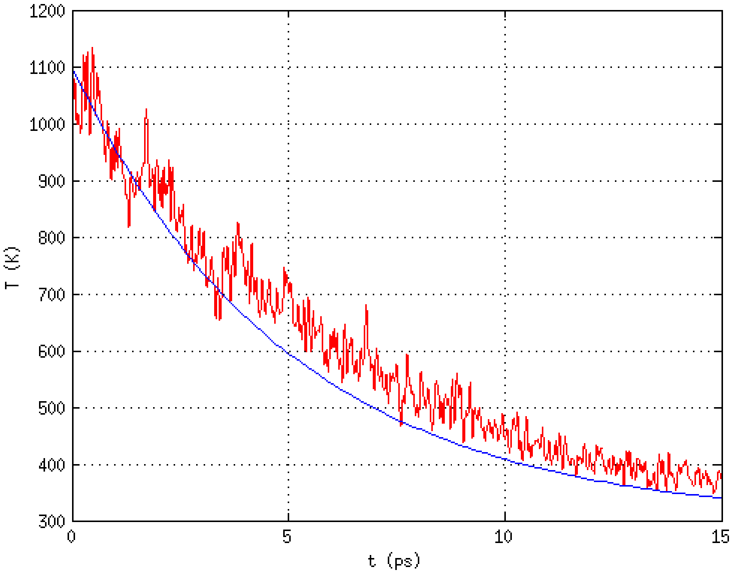
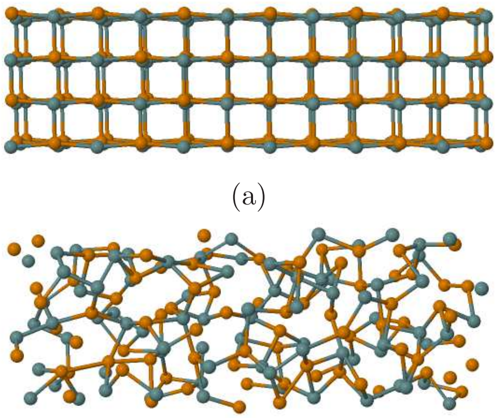


Figure 4.2: Temperature profile of the quench AIMD simulation. Blue line: target temperature; red line: actual AIMD termperature.

4.3.2 Analysis Results

4.3.2.1 Atomic Coordinates

The melt-quench AIMD simulations start from the c-GeTe model (Figure 4.3 (a)), and end up with the atomic coordinates of the a-GeTe (Figure 4.3 (b)).



(a)

(b)

Figure 4.3: Crystalline (a) and amorphous (b) GeTe AIMD simulation models.

4.3.2.2 Pair Distribution Function

To show the validity of the melt-quench AIMD simulation discussed above, the pair distribution function (PDF)

$$g(r) = \frac{1}{\rho^2}\left\langle \sum_{i} \sum_{i \neq j} \delta(r_i)\delta(r_j - r) \right\rangle \tag{4.8}$$

of the final amorphous atomic coordinates is computed using the methodology introduced in [2], where ρ is the density and r_i is the position of atom i. The calculation results are compared against experimental measurements, as shown in Figure 4.4.

Since the amorphous materials have no regular lattice structure, the PDF is one of the most effective ways to describe the inter-atomic distance in the amorphous materials. Physically, the PDF describes the distances between pairs of atoms in the system. Mathematically, the PDF denoted by $g_{ij}(r)$ defines the probability of finding atom j at the distance of r from atom i. The PDF shown in Figure 4.4 is computed by averaging all GeTe atoms in the ab initio simulation supercell.

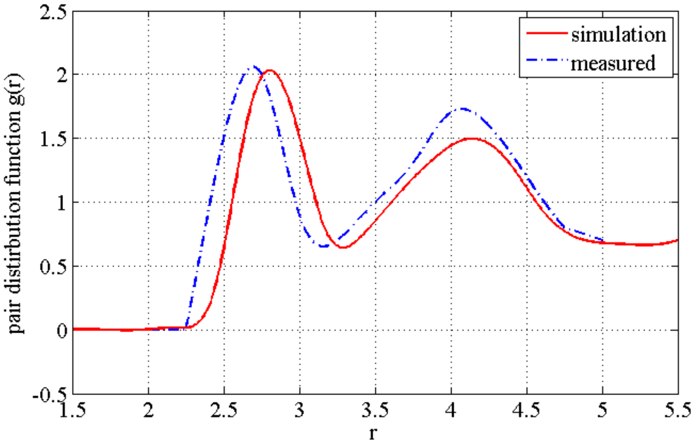


Figure 4.4: Pair distribution function of amorphous GeTe.

4.3.2.3 Nearest Neighbor

As shown in Figure 4.4, the first (second) nearest neighbor distance is calculated to be 2.8 Å (4.1 Å), which agrees well with the measured value 2.7 Å (4.2 Å). The match between the simulated and measured nearest neighbor distances shows that the ab initio simulations fairly captured the microscopic nature of the *a*-GeTe.

4.3.2.4 Bond Energy Hierarchy

It is well known that one of the most serious bottlenecks of the PCM technology is its large RESET energy consumption, which inherently originates from the thermally-induced melting-based amorphization of the crystalline PCM. To break this bottleneck and to offer greener nonvolatile data storage solutions, it is meaningful to investigate the possible non-melting amorphization mechanisms to achieve the RESET operations, by taking advantage of the bond energy hierarchy in the crystalline PCM [50].

In this section, the crystalline GeTe is analyzed by using DFT simulations, in order to quantitatively investigate the bond energy hierarchy. The bond energy can also be called the contributing bond energy, which physically represents the contribution that the bond makes to the total atomization energy of the system [98]. The bond energy can be computed by decomposing the DFT total energy (the mathematical derivation and physical explanation are presented in Appendix C).

As shown in Figure 4.5, the Ge-Te bond lengths differ in the rhombohedral crystalline GeTe. By using the bond energy analysis methodology introduced in Appendix C, it is calculated here that the short Ge-Te bond (about 2.9 Å) has 2.93 eV bond energy, while the long Ge-Te bond (about 3.2 Å) has 1.42 eV bond energy.

This bond energy hierarchy can possibly be used to achieve the non-melting amorphization, by destroying the crytalline lattice via selectively breaking the weak bonds. Though this could be achieved by using laser and/or coherent phonon excitation, the development of the practical non-melting amorphization methodology that can be used in commodity PCM devices is still ongoing.

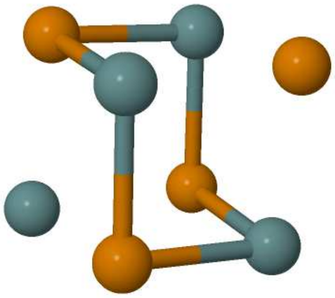


Figure 4.5: Short and long bonds in crystalline GeTe. Short bonds (about 2.9 Å) are visualized as bars; while long bonds (about 3.2 Å) are not.

4.4 Crystallization Simulation

It has been found that the chalcogenide PCM possess a lot of interesting and useful physical properties like the nanosecond-level fast phase change speed and the nanometer-level high scalability, which are rarely found in other materials. These findings have spurred intensive interests to theoretically understand the microscopic origin of the PCMs unique physical properties. Also, they have invoked extensive efforts of using PCM to implement the next-generation ultra-fast and ultra-dense non-volatile information storage devices.

The fast phase change speed and the high scalability of PCM largely hinge on the unique crystallization properties of *a*-PCM. So it is important to investigate the crystallization behaviors of *a*-PCM, in order to offer deeper scientific understanding, and to search better material stoichiometry to enable faster and more scalable information storage technologies. Although the crystallization of *a*-PCM has been extensive explored, the detailed knowledge of the roles of different physical mechanisms that determines the crystallization properties is still missing.

In this section, the temperature dependent crystallization properties of the prototypical binary PCM germanium telluride (GeTe) is investigated by combining the thermodynamic theory and the ab initio simulations. The physical mechanisms (atom diffusion, Gibbs free energy, elastic energy, and interfacial energy) that govern the crystallization process and their temperature dependence are quantitatively analyzed. In Section 4.4.1, the ab initio simulation procedure and the thermodynamic analysis methodology are introduced. In Section 4.4.2, the simulation results are presented and analyzed. In Section 4.4.3, the technological implications and the limitations in the simulation methodology are discussed.

4.4.1 Modeling Methodology

4.4.1.1 Thermodynamic Theory

It has been experimentally observed that the crystallization of *a*-GeTe is a two-stage process: (1) activated by thermal energy, the homogeneous nucleation occurs inside the a-GeTe to form small crystalline clusters; (2) then the growth of crystalline clusters happens at the interface between *a*-GeTe and *c*-GeTe.

According to the thermodynamic theory of crystallization, the nucleation of *a*-GeTe can be considered as the following three-step imaginary operation: (1) remove a small volume from the *a*-GeTe, and allow it to undergo an unconstrained transformation into *c*-GeTe; (2) apply surface traction to the *c*-GeTe and stretch it to its original size when it was *a*-GeTe, then insert it back into the hole of the *a*-GeTe again (the mass density of the *c*-GeTe is larger than that of the a-GeTe); (3) weld the small volume of c-GeTe with the a-GeTe. The energy penalties of these three steps are the three components of the crystalline cluster formation energy

$$\Delta G = -\frac{4}{3}\pi r^3 \Delta G_{ac} + \frac{4}{3}\pi r^3 E_e + 4\pi r^2 \sigma \tag{4.9}$$

where r is the effective radius of the crystalline cluster seed of c-GeTe; $\Delta G_{ac} = G_a - G_c$ and G_a (G_c) is the Gibbs free energy density of the c-GeTe (a-GeTe); E_e is the elastic energy density of the c-GeTe; σ is the interfacial energy density between a-GeTe and $c-$GeTe. As schematically shown in Figure 4.6(a), the system has to conquer the critical formation energy barrier ΔG_c to enable nucleation. The growth could begin when the effective radius of the c-GeTe cluster is larger than the critical nuclei radius r_c, as shown in Figure 4.6(a). The growth of c-GeTe is a consequence of two counter-balancing processes occurring at the interface: 1) amorphous to crystalline transition; and 2) crystalline to amorphous transition. Their energy barriers are E_1 and $E_1 + E_2$ respectively, as shown in Figure 4.6(b). Since the crystallization of a-GeTe is a thermally activated and diffusion limited process, $E_1 = E_D$ and $E_2 = \Delta G_{ac}$ where E_D is the activation energy of atom diffusion in a-GeTe.

According to the Arrhenius law, the nucleation rate I

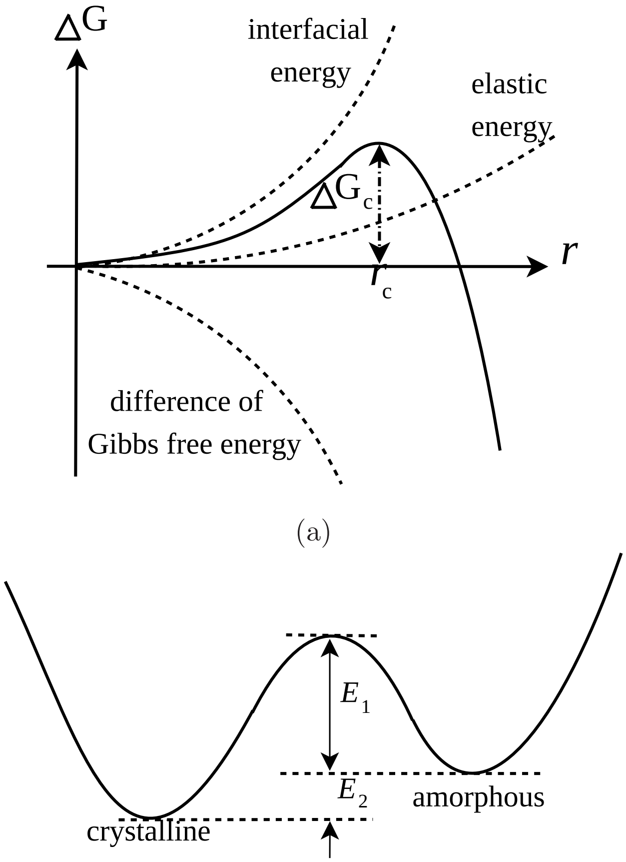
$$I = I_0 e^{-\frac{E_D + \Delta G_c}{kT}} \tag{4.10}$$

and the growth speed u

$$u = u_0 e^{-\frac{E_D}{kT}} (1 - e^{-\frac{\Delta G_{ac}}{kT}}) \tag{4.11}$$

where I_0 and u_0 are the prefactors; k is the Boltzmann constant; and T is the temperature [25].

Physically, the Equation 4.10 and Equation 4.11 quantitatively define the dependence of the I and u on the characteristic energies, e.g. E_D, G_c, G_{ac}, and the temperature T. The E_D, G_c, and G_{ac} can be computed from practical-scale ab initio molecular dynamics simulations. However, it worth noting that the prefactors I_0 and u_0 physically represent the characteristic attempt frequency, which are more difficult to compute. To calculate the numerical values of I_0 and u_0, one can run brute-force AIMD crystallization simulation. However, the precise numerical values of I_0 and u_0 are not needed to unveil the different roles of the various physical quantities (Gibbs free energy density, interfacial energy density, elastic energy, specific heat capacity, critical formation energy, critical nuclei radius, etc.) during the crystallization process. Therefore, the values of the prefactors I_0 and u_0 are not calculated and the I/I_0 and I/u_0 are used instead in the analysis.



(a)

(b)

Figure 4.6: Schematic energy barriers for the nucleation (a) and the growth (b) (ΔG: formation energy; r: effective radius of crystalline nuclei; ΔG_c: critical formation energy; r_c: critical radius; $E_1 = E_D$: atom diffusion activation energy; $E_2 = \Delta G_{ac}$: Gibbs free energy difference between a-GeTe and c-GeTe).

4.4.1.2 Simulation Models

In order to obtain the quantitative values pertaining to the crystallization process of *a*-PCM, three models are created using canonical ensemble (NVT) ab initio simulations:

1. The *c*-GeTe model contains 192 atoms. The supercell volume $V_c = L_x \times L_y \times L_z^c$ (upper case V and L mean fixed value) is chosen such that the mass density is the same as the measured value 6.06 g/cm^3 [83]. The lattice structure of the measured rhombohedral *c*-GeTe [21] is used to generate the atomic coordinates, which are relaxed using conjugate gradient (CG) algorithm.

2. The *a*-GeTe model, which also contains 192 atoms, is obtained from the melt-quench ab initio molecular dynamics (AIMD) simulations model [67]. First, the *c*-GeTe model is melted at 2000 K for 7 ps, and then brought down to 1000 K for 3 ps, in order to destroy the crystalline lattice structure and generate liquid (*l*-) GeTe. Then, the *l*-GeTe is quenched to 300 K in 20 ps. After the CG relaxation, the *a*-GeTe model can be obtained. In the melt-quench AIMD simulations, the supercell volume is chosen as $V_a = L_x \times L_y \times L_z^a$ such that the mass density is the same as the measured value of *a*-GeTe 5.60 g/cm^3 [83].

3. The interface model, which contains 384 atoms, is created by putting the *c*-GeTe model and the *a*-GeTe model together in the supercell with volume $V_ac = L_x \times L_y \times (L_z^a + L_z^c)$, followed by the CG relaxation. These three models will be used as the initial atomic coordinates of the following AIMD simulations.

4.4.1.3 Simulation Procedure

The isothermal-isobaric ensemble (NPT) AIMD simulations are performed on the three simulation models (Section 4.4.1.2).

In the AIMD simulations, the temperature T is controlled to be the annealing temperature by using the Nosé thermostat (Section 4.1.2.2). Since the simulation supercell size is very small, the AIMD annealing temperature is set as uniform in the system. The

Parrinello-Rahman method (Section 4.1.2.2) is used to alter the supercell volume on the fly, such that the pressure P is controlled to be roughly zero, in order to capture the temperature dependence of the mass density. The supercell volumes of the c-GeTe model, the a-GeTe model, and the interface model are $v_c = l_x \times l_y \times l_z^c$, $v_a = l_x \times l_y \times l_z^a$, and $v_{ac} = l_x \times l_y \times l_z^{ac}$, respectively (lower case v and l mean variable values). The AIMD simulations run for 8 ps, and the time-averaged temperature dependent internal energies are shown in Figure 4.7. Here, the simulation time is chosen to be long enough to average out the energy fluctuation as a function of time; but short enough to prevent the phase change from a-GeTe to c-GeTe, which needs more than $10^2 - 10^3$ ps [94].

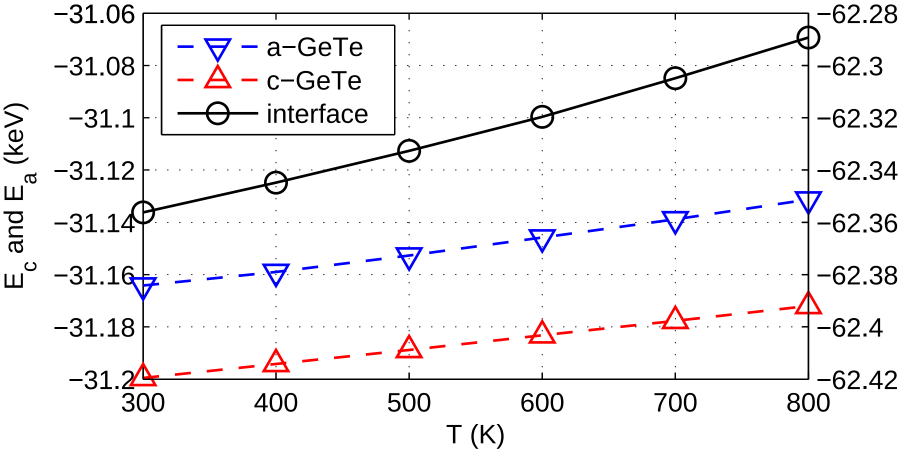


Figure 4.7: Time-averaged temperature dependent internal energies of the c-GeTe model (E_c), the a-GeTe model (E_a), and the interface model (E_{ac}).

4.4.2 Analysis Results

In this section, the results obtained from ab initio simulations are presented. First it is shown that the results of mass density and specific heat capacity obtained from the simulations can agree well with experimental measurements. Then the atomic diffusion in a-GeTe is investigated. Based on the simulation results of the three components of Equation 4.9 (Gibbs free energy, elastic energy, interfacial energy), the nucleation and growth properties and their temperature dependence are analyzed.

4.4.2.1 Mass Density

In the NPT AIMD simulations, the volume of the supercell is allowed to change such that the internal pressure approximately vanishes. This enables the calculation of the temperature dependent mass densities of *a*-GeTe (ρ_a) and *c*-GeTe (ρ_c), as shown in Figure 4.8. It can be seen that the simulated ρ_a agrees well the measured value 5.60 g/cm^3; the simulated ρ_c/ρ_a values could reproduce the experimental measurements that the mass density of the *c*-GeTe is about 8% higher than that of the *a*-GeTe [94, 83].

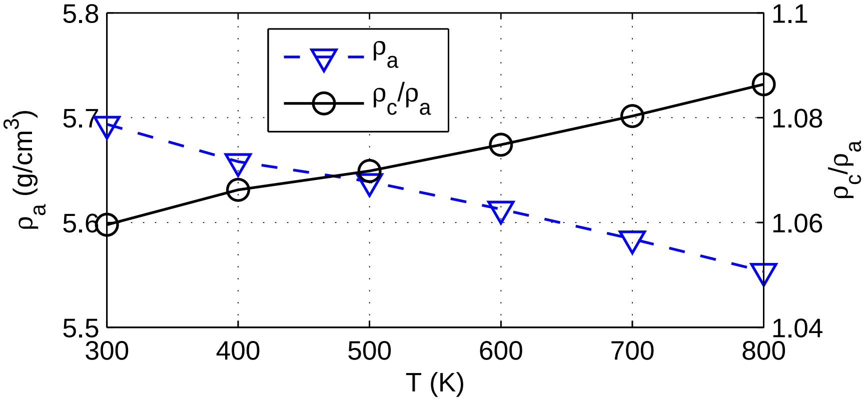


Figure 4.8: Temperature dependent mass density.

4.4.2.2 Specific Heat Capacity

The specific heat capacity of a-GeTe (c-GeTe) can be computed using

$$c_{a(c)} = (\partial U_a(c)/\partial T)_P \tag{4.12}$$

where T is the annealing temperature; $U_{a(c)} = E_{a(c)}/v_{a(c)}$ is the internal energy density; $E_{a(c)}$ is the internal energy (as shown in Figure 4.7); and $v_{a(c)}$ is the supercell volume of the *a*-GeTe (*c*-GeTe) model.

The simulation shows that $c_c = a_c + b_c T$, where $a_c = 1.45 \times 10^6$ J/Km^3 and $b_c = 292$ J/m^3, which are close to the experimentally measured values $a_c = 1.35 \times 10^6$ J/Km^3 and $b_c = 351$ J/m^3 [76]. This means that when $T = 600$ K, the simulation result is $c_c = 1.62 \times 10^6$

J/Km^3, which is in decent agreement with the measured value 1.56×10^6 J/Km^3 [76]. While the measured value of ca is not available, the simulations yield $c_a = a_a + b_a T$ ($a_a = 1.11 \times 10^6$ J/Km^3 and $b_a = 1.51 \times 10^3$ J/m^3) using ab initio simulations, as shown in Figure 4.9.

The specific heat capacity depends on the number of the degree of freedom (DOF) that is available in the material to store energy. It can be seen from Figure 4.9 that at low temperature, the c_a is slightly larger but comparable to c_c. As temperature increases, the c_a becomes much larger than c_c because b_a is larger than b_c. This reveals that by increasing the annealing temperature, the *a*-GeTe can enable more DOF than the *c*-GeTe due to the difference of lattice structure and chemical bonding.

It deserves mentioning that while c_c can be measured, it would be difficult to experimentally measure c_a, due to the nanosecond-level fast phase change speed from *a*-GeTe to *c*-GeTe. Fortunately, the picosecond-level AIMD simulations can be used to obtain c_a. The specific heat capacity can influence the crystallization of *a*-GeTe via entropy difference, which will be calculated in the following sections.

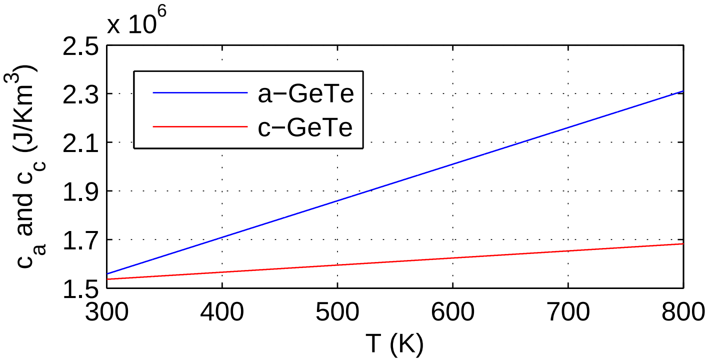


Figure 4.9: Temperature dependent specific heat capacity of *a*-GeTe (c_a) and *c*-GeTe (c_c).

4.4.2.3 Diffusivity

The diffusivity of atoms in *a*-GeTe can be calculated using

$$D \approx \frac{1}{6t} < |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 > \tag{4.13}$$

wher $\mathbf{r}_i(t)$ is the position of atom i (at time t) obtained from the AIMD simulations. The results of the mean square displacement (MSD) $< |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 >$ for $T = 800$ K are shown in Figure 4.10. The temperature dependence of diffusivities is shown in Figure 4.11. Since the diffusion is a thermally activated process, the diffusivity obeys the Arrhenius law

$$D = D_0 e^{-\frac{E_D}{KT}} \tag{4.14}$$

where D_0 is prefactor; and E_D is activation energy. Using Equation 4.14 and D values shown in Figure 4.11, it can computed that $E_D = 0.30$ eV for Ge atoms, $E_D = 0.32$ eV for Te atoms, and $E_D = 0.31$ eV for a-GeTe.

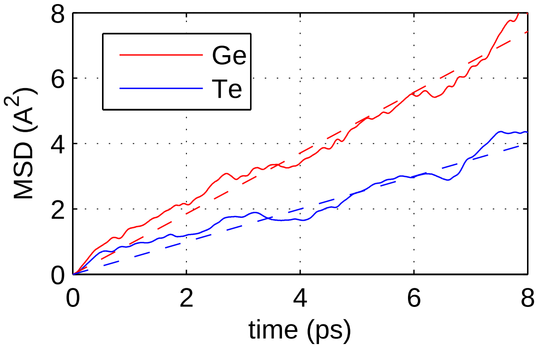


Figure 4.10: The mean square displacement (MSD) of atoms in a-GeTe (T=800 K).

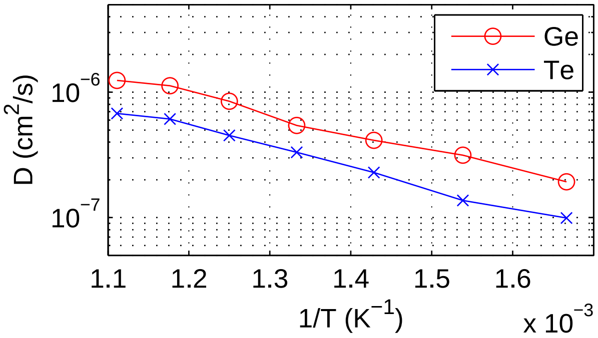


Figure 4.11: Temperature dependent diffusivity (D) of atoms in a-GeTe.

4.4.2.4 Gibbs Free Energy

As schematically shown in Figure 4.6(b), the a-GeTe is a metastable structure, whose Gibbs free energy density (G_a) is higher than that of the c-GeTe (G_c). The difference between G_a and G_c can be calculated using [1]

$$\Delta G_{ac} = \Delta U_{ac} - T\left(\frac{\Delta H_m}{T} + \int_{T_m-}^{T} \frac{c_a - c_c}{T} dT\right) \tag{4.15}$$

where $\Delta U_{ac} = U_a - U_c$; $U_{a(c)}$ is the internal energy density of a-GeTe (c-GeTe); T_m = 998 K is the melting temperature of GeTe [94]; ΔH_m=1.45 GJ/m^3 is the latent heat of fusion [11]; $c_{a(c)}$ is the specific heat capacity of a-GeTe (c-GeTe) as shown Figure 4.9. In Equation 4.15, the terms inside the parenthesis represents $\Delta S_{ac} = S_a - S_c$, where $S_{a(c)}$ is the entropy density of the a-GeTe (c-GeTe). The simulation results of ΔG_{ac} are presented in Figure 4.12.

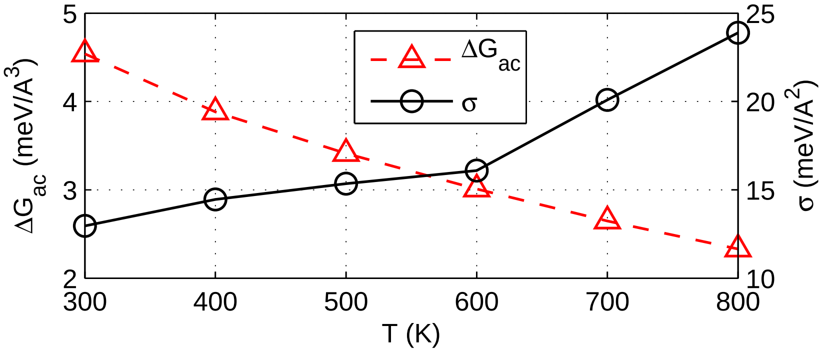


Figure 4.12: Gibbs free energy density difference between a-GeTe and c-GeTe (ΔG_{ac}) and interface energy density (σ).

4.4.2.5 Interfacial Energy

The interfacial energy of the interface between a-GeTe and c-GeTe is computed using [2]

$$\sigma = \frac{E_{ac} - E_a - E_c}{2A} \tag{4.16}$$

---

[1] See Section B.1 for details of the derivation.

[2] See Section B.2 for details of the derivation.

where $A = l_x \times l_y$ is the area of the interface between a-GeTe and c-GeTe in the interface model; E_a, E_c, and E_{ac} are the internal energies of the a-GeTe model, the c-GeTe model, and the interface model, as shown in Figure 4.7. The results of σ are shown in Figure 4.12.

4.4.2.6 Elastic Energy

The elastic energy density Ee of the c-GeTe can be calculated using [3]

$$E_e = \frac{B_0}{B_0'}\left[\frac{1}{B_0'-1}\left(\frac{V_0}{V}\right)^{B_0'}+1\right]-\frac{B_0}{B_0'-1}\frac{V_0}{V} \tag{4.17}$$

where B_0 (B_0') is the (pressure derivative of) bulk modulus; and V (V_0) is the (most energetically favorable) supercell volume [78]. The parameters in Equation 4.17 are computed using ab initio simulations. First the c-GeTe model is isotropically stretched to have different supercell volumes (V). Then the atoms are relaxed using CG to obtain the total energy of the system with the supercell volume V. Finally, the total energy as a function of V is fitted into the Murnaghan equation of state Figure 4.13.

The fitting shows that:

1. The most energetically favorable lattice constant of the c-GeTe is 6.04 Å, which is in decent agreement with the measured value 6.0 Å [21];

2. The bulk modulus is calculated to be 50.0 GPa, which is in agreement with the values 48.9 GPa [85] and 51.0 GPa [91] in the previous studies;

3. The pressure derivative of the bulk modulus is calculated to be 4.62.

Using the simulation results of B_0, B_0', V, and V_0, the elastic energy density is computed using Murnaghan equation of state (Figure 4.13). As shown in Figure 4.6 and Equation 4.9, the elastic energy density (E_e) counterbalances the influence of Gibbs free energy density difference (ΔG_{ac}), leading to raised nucleation formation energy barrier. As shown in Figure 4.12 and Figure 4.13, the E_e is comparable to ΔG_{ac}. Thus, qualitatively speaking, if the contribution of the elastic energy is ignored, the nucleation will be estimated to be

---

[3] See Section B.3 for details of the derivation.

easier than the actual situation. The quantitative analysis of this tendency will be presented in the following sections.

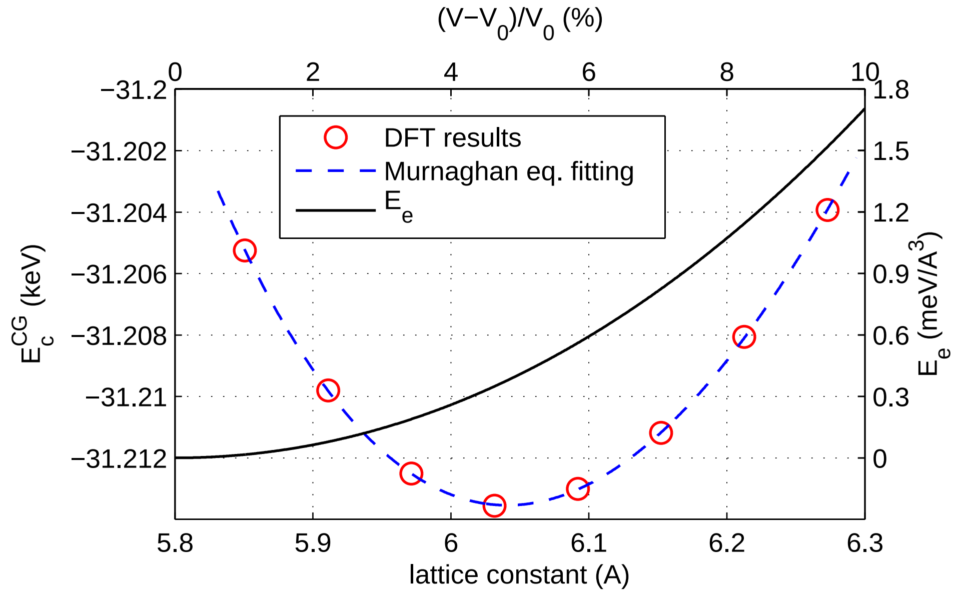


Figure 4.13: Total energy of the *c*-GeTe model (E_c^{CG}) under strain (corresponding to different lattice constants) and the fitting to the Murnaghan equation of states; and the elastic energy density (E_e) as a function of $(V - V_0)/V_0$ where V is the *c*-GeTe supercell volume and V_0 is the most energetically favorable supercell volume of the *c*-GeTe model.

4.4.2.7 Critical Nuclei Radius and Formation Energy

Using the simulation results of ΔG_{ac}, σ, and E_e, the nucleation formation energy ΔG can be obtained, as shown in Figure 4.14. The x and y coordinates of the peak of the ΔG curves represent the critical nuclei radius r_c (Figure 4.15) and the critical formation energy ΔG_c (Figure 4.16), respectively.

It can be seen from Figure 4.14, Figure 4.15, and Figure 4.16 that:

1. both ΔG_c and r_c increase monotonously as temperature rises, since the increased thermal vibration tends to break the chemical bonds and make nucleation more difficult to happen;

2. E_e plays an important role in determining ΔG, ΔG_c, and r_c. For $T = 400$, 600, 800 K, ignoring E_e leads to 27.2%, 41.7%, and 63.4% underestimation of ΔG_c; and 14.8%, 23.4%, and 39.5% underestimation of r_c.

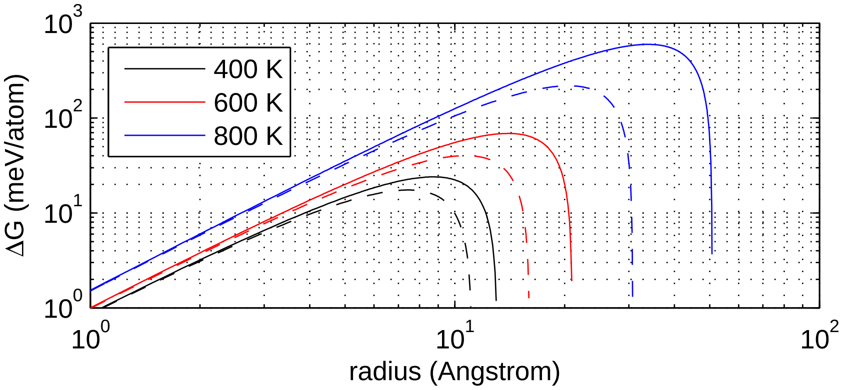


Figure 4.14: Formation energy (ΔG) when elastic energy is included (solid lines) and ignored (dashed lines).

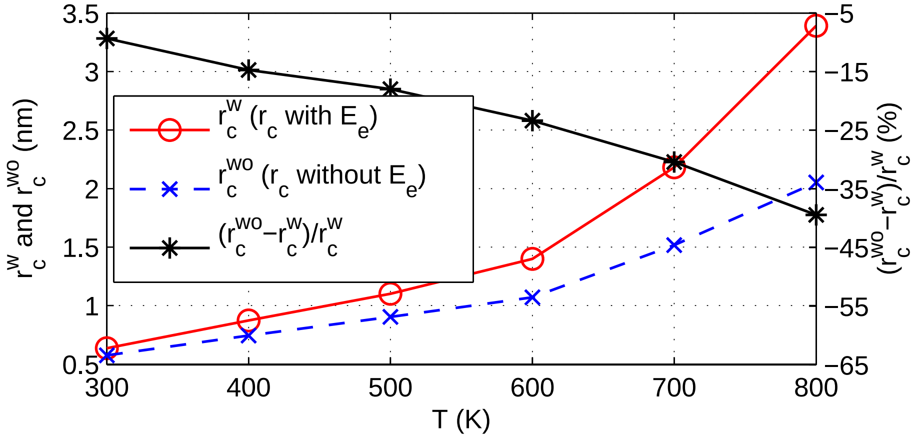


Figure 4.15: Critical nuclei radius with elastic energy E_e included (r_c^w) and without E_e (r_c^{wo}).

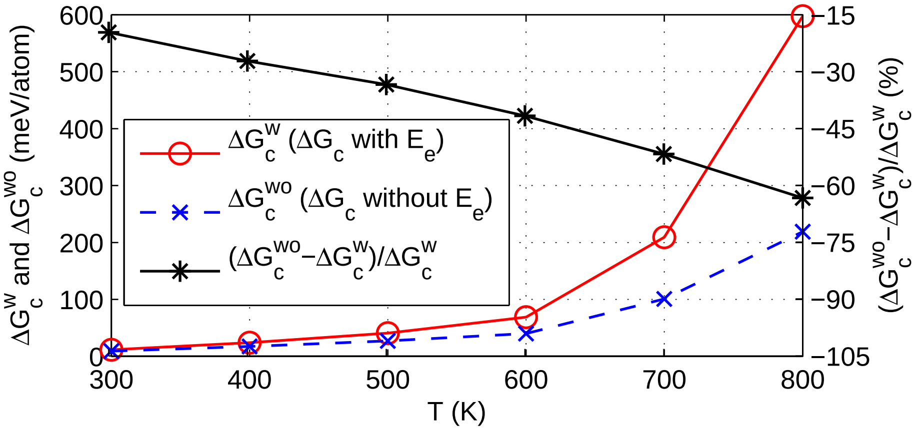


Figure 4.16: Critical formation energy with elastic energy E_e included (ΔG_c^w) and without E_e (ΔG_c^{wo}).

4.4.2.8 Nucleation Rate and Growth Speed

Using the values of E_D, ΔG_{ac}, and ΔG_c, the nucleation rate I and the growth rate u can be computed by using Equation 4.10 and Equation 4.11, as shown in Figure 4.17. Here the focus is the role of elastic energy and the temperature dependence of I and u, instead of the absolute value of I and u. So I/I_0 and u/u_0 are plotted. To obtain the values of I_0 and u_0, the brute-force AIMD simulation of crystallization is needed, which requires much longer simulation time (about $10^2 - 10^3$ ps).

Figure 4.17 shows that the nucleation rate (I/I_0 with E_e) peaks at around 600 K. This is caused by the fact that the nucleation is determined by both the atom diffusion and the formation energy, as shown in Equation 4.10. When the temperature is increased, atom diffusivity D increases monotonously (Figure 4.11), which tends to increase the nucleation rate I. However, this tendency is counterbalanced by the increasing critical formation energy barrier ΔG_c (Figure 4.14). Below 600 K, the increase of ΔG_c is slow so the I is largely dominated by D. Above 600 K, the ΔG_c increases rapidly, leading to decrease of I. It is obvious from Figure 4.17 that if the elastic energy is ignored, I will be significantly overestimated (by 20.8%, 74.2%, and 240.2% when T = 400, 600, and 800 K, respectively). This originates from the underestimation of the critical formation energy barrier (Figure 4.16).

In the temperature range of our interests here, the growth speed u increases as temperature increases, as shown in Figure 4.17. According to Equation 4.11, u is determined by both the diffusivity D and the Gibbs free energy density difference ΔG_{ac}. When T is low, ΔG_{ac} is large, so u is largely determined by D which increases as T increases. When T approaches the melting temperature $T_m = 998$ K, ΔG_{ac} will decrease to a very small value (Figure 4.12), which will lead to the decrease of u.

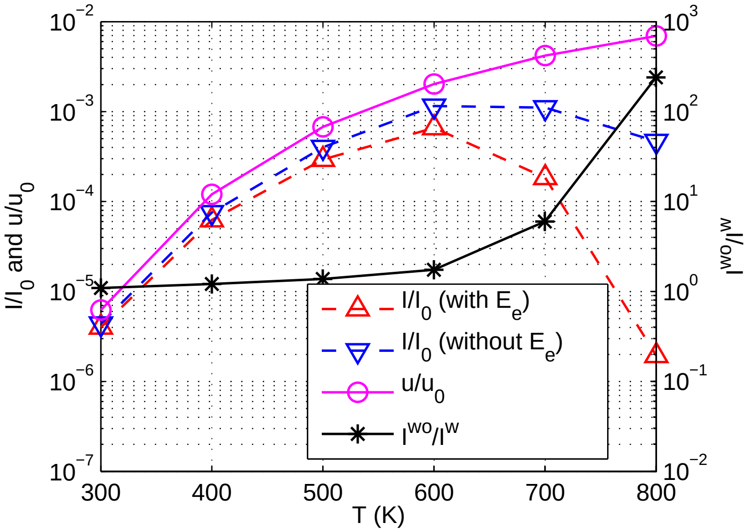


Figure 4.17: Nucleation rate I and growth speed u (the superscript w/wo means the calculation of I using Equation 4.10 is performed with/without elastic energy E_e).

4.4.2.9 Brute Force Annealing Simulation

In this section, the of brute force AIMD annealing simulation results are presented. The simulation starts from the pristine crystalline GeTe (Figure 4.18 (a)). The melt-quench AIMD simulation as introduced in Section 4.3 is applied to obtain the atomic coordinates of amorphous GeTe (Figure 4.18 (b)). Then, the amorphous GeTe is annealed at a constant temperature (650 K) to obtain the "recrystallized" GeTe (Figure 4.18 (c))[4].

---

[4] Though the annealed GeTe model shown in Figure 4.18 (c) appears to be more ordered than the amorphous GeTe, it deserves more rigorous structural analysis to determine whether or not it is fully recrystallized. So, it is referred to as "recrystallized" GeTe hereafter.

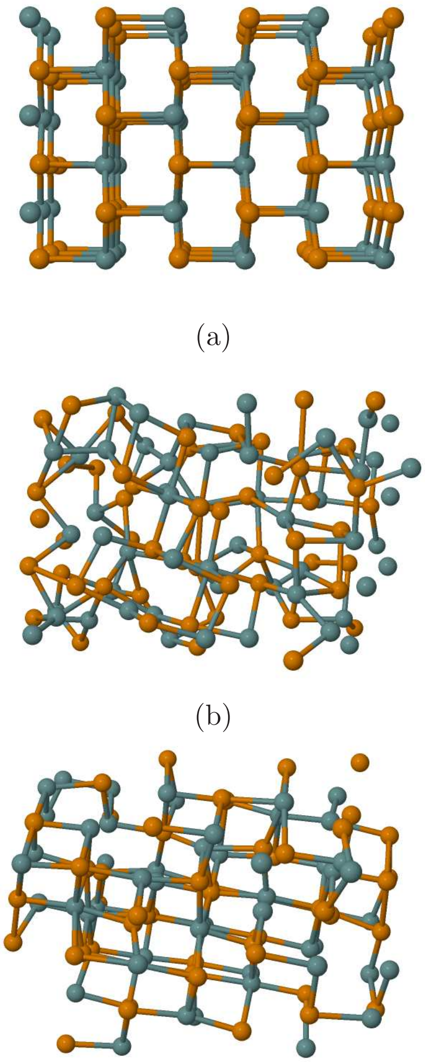
In the rhombohedral crystalline GeTe lattice, the basic building blocks are the correct bonds (Ge-Te), the 4-fold rings, and the 8-atom cubes. During the annealing simulation, it can be observed that the the number of 8-atom cubes and the number of 4-fold rings are increased gradually (Figure 4.19 and Figure 4.20). This is accompanied by the gradual decrease of the number of wrong bonds (Ge-Ge and Te-Te), as shonw in Figure 4.21 [65].

In Figure 4.22, the pair distribution function (PDF) of "recrystallized" GeTe (Figure 4.18 (c)) is demonstrated. As a comparison, the simulated PDF of amorphous GeTe (Figure 4.4) is also shown. It can be seen that the "recrystallized" GeTe has more pronounced PDF peaks than the amorphous GeTe. This indicates that the "recrystallized" GeTe has more ordered lattice structure.

4.4.3 Discussion

Though crystallization of *a*-PCM is very fast, it is slow compared to the amorphization of *c*-PCM. So the SET operation (crystallization) is the speed-limiting factor of the phase change memory devices [94]. As a two-step process (nucleation and growth), the speed of crystallization is largely limited by nucleation, since the growth speed is very fast (about $10^0 - 10^1$ nm/ns) [65]. The analysis shows that the nucleation rate has the largest value around 600 K. In the design of PCM-based devices, in order to shorten SET operation time, the SET operation current pulse shape should be optimized such that the PCM active region is controlled at this optimal temperature.

The analysis shows that the elastic energy exerts considerable impacts on the crystallization properties of *a*-PCM by raising the nucleation formation energy barrier. This unfavorable effect deteriorates the recrystallization speed, which is the bottleneck of the speed of PCM-based devices. Predictably, if the elastic energy can be reduced, say by optimizing the chalcogenide stoichiometry and/or doping, the speed of crystallization and the SET operation of the PCM-based devices can be increased. The simulation results show that, when the annealing temperature is 600 K, the critical nuclei radius is roughly 1.4 nm. This means that the crystallization might be able to happen in ultrascaled nanostructures, indicating extremely promising scaling scenario of the PCM-based information



(a)

(b)

(c)

Figure 4.18: Brute force AIMD simulation of GeTe phase change. (a) pristine crystalline GeTe; (b) amorphous GeTe; and (c) "recrystallized" GeTe.

storage device technologies.

It worth noting that, due to the expensive computational cost of ab initio simulations, this paper focuses on the homogeneous and isotropic crystallization. Of course, in real PCM-based devices, some complicated processes (e.g. sympathetic nucleation, undercool-

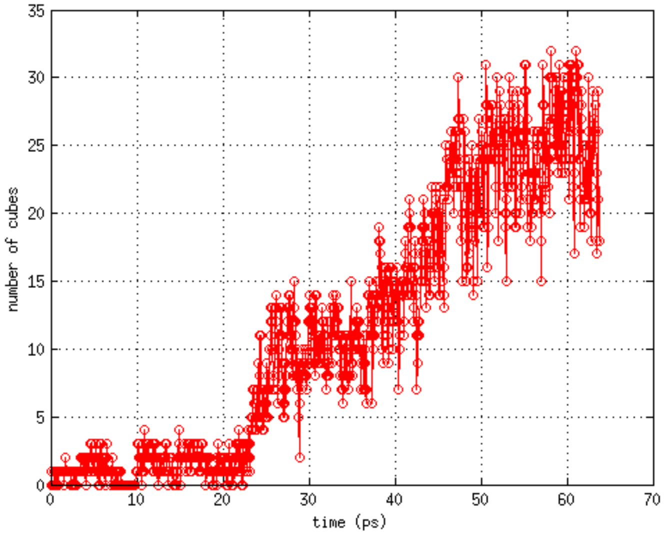


Figure 4.19: Number of 8-atom crystalline GeTe cubes in annealing AIMD simulation.

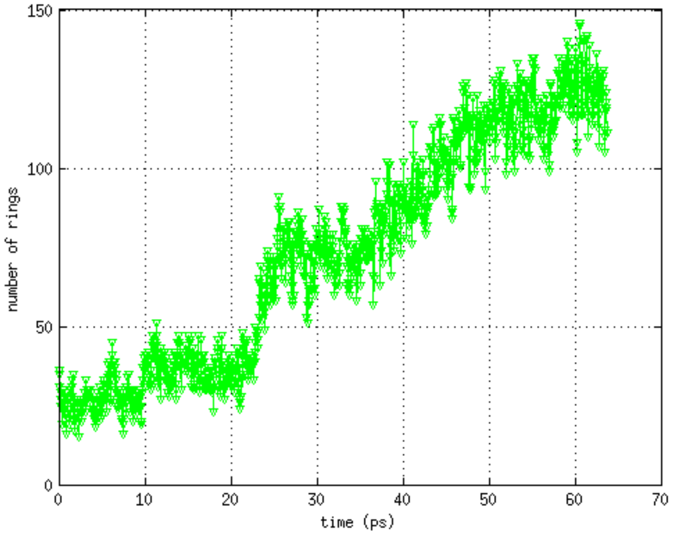


Figure 4.20: Number of 4-fold crystalline GeTe rings in annealing AIMD simulation.

ing, dislocation, grain boundary, and interface between PCM and encapsulating materials) [25, 1], which are not accounted for here, can influence the crystallization properties. Their impacts on the PCM crystallization deserve future research efforts.

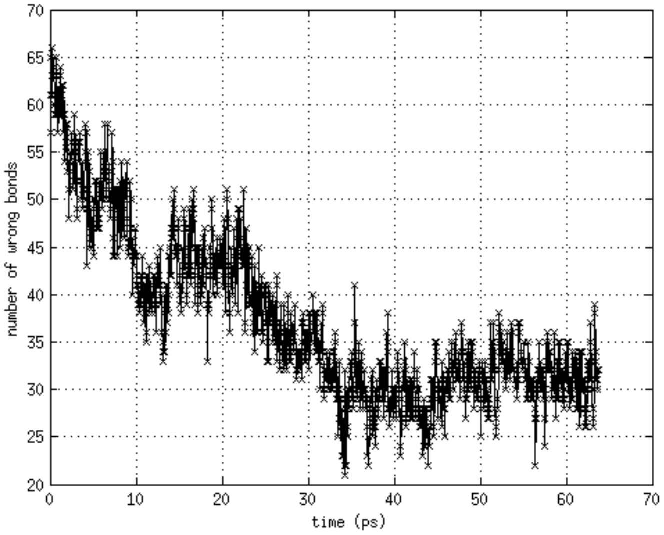


Figure 4.21: Number of wrong bonds in annealing AIMD simulation.

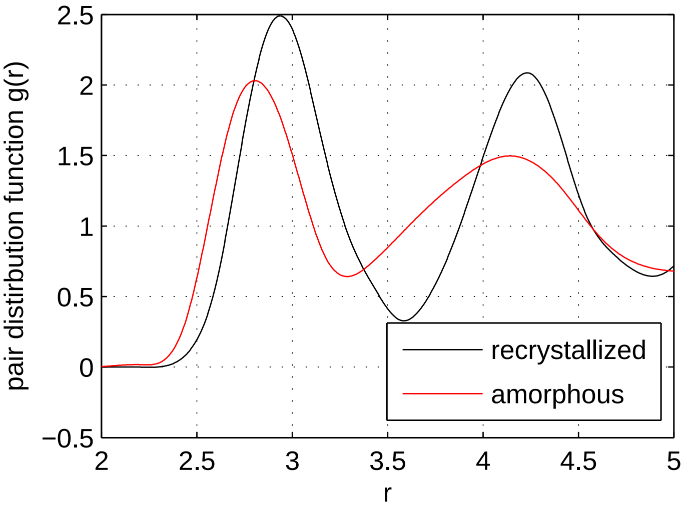


Figure 4.22: Pair distribution function of “recrystallized” GeTe.

Chapter 5

CONCLUSION

In this chapter, a brief conclusion of the thesis is presented. In Section 5.1, the modeling and simulation used in the thesis is briefly reviewed. In Section 5.2, the main results and conclusions are briefly summarized. In Section 5.3, the author's projection of the future PCM research direction and the PCM technology development is tentatively discussed.

5.1 Methodology

This thesis investigates the PCM device technology by using multiscale simulations. In the macroscopic scale ($\approx 10^1 - 10^2$ nm and beyond), the analytical derivation and TD-FEM are used to analyze the governing eletro-thermal transport equation and Laplace equation, to investigate the scaling scenario of PCM devices. In the mesoscopic scale ($\approx 10^{-1} - 10^0$ nm), the NEGF is applied to investigate the electron transport properties in deep nanoscale and the ultimate scaling limit of the PCM devices. In the microscopic scale ($\approx 10^0 - 10^1$ nm), the ab initio DFT and AIMD are used to investigate the atomistic level amorphization and crystallization processes.

5.2 Main Results

5.2.1 Macroscopic Simulation

The macroscopic simulation analyzes the scaling property of PCM based devices using analytic and numerical methods. The scaling scenarios of the three widely used PCM operation schemes (i.e., constant electric field, constant voltage, and constant current) are studied and compared. It is shown that if the device size is downscaled by a factor of $1/k$ ($k > 1$), the operation energy (current) will be reduced by more than k^3 (k) times, and the operation speed will be increased by k^2 times. It is also shown that more than 90% of operation energy is wasted as thermal flux into substrate and electrodes. We predict that, if the wasted thermal flux is effectively reduced by heat confinement technologies, the energy consumed per RESET operation can be decreased from about 1 pJ to less than 100 fJ. It is shown that reducing the aspect ratio of the PCM cell helps decreasing PCM energy consumption. It is revealed that cross-cell thermal proximity disturbance is counter-intuitively alleviated by scaling, leading to a desirable scaling scenario.

5.2.2 Mesoscopic Simulation

In mesoscopic simulation, the sub-threshold electron transport properties of amorphous (*a*-) germanium telluride (GeTe) PCM ultra-thin films, which is sandwiched by titanium nitride (TiN) electrodes, are investigated by using the non-equilibrium Greens function (NEGF) simulations. The simulation results successfully reproduce the measured electron transport properties, e.g. current-voltage (IV) curve, intra-bandgap donor-like and acceptor-like defect states, and *p*-type conductivity. The underlying physical mechanism of the electron transport in ultra-scaled *a*-PCM is unraveled. We found that, though the IV curve of the ultra-scaled *a*-PCM highly resembles that of the bulk *a*-PCM, their physical origins are totally different. Unlike the electron transport in bulk *a*-PCM, which is governed by the Poole-Frenkel effect, the electron transport in ultra-scaled *a*-PCM is largely ballistic and dominated by transport via intra-bandgap donor-like and acceptor-like defect states. It is found that *a*-GeTe ultrathin films scaled down to about 38 Å (12 atomic layers) still show a band gap. If the ultrathin films are further scaled, the *a*-GeTe band gap disappears due to overlap of the two metal induced gap states (MIGS) regions near the TiN electrodes, leading to sharp increase of *a*-GeTe conductance and significant decrease of *c*-GeTe/*a*-GeTe conductance ratio. The *c*-GeTe/*a*-GeTe conductance ratio drops below one order of magnitude if the ultrathin films are scaled below about 33 Å, making it difficult to reliably perform read operations. This overlap of the MIGS regions sets up the ultimate scaling limit of PCM device technology. The simulation results suggest that the ultimate scaling limit can be pushed to even smaller size, by using PCM with larger amorphous phase band gap than *a*-GeTe. The NEGF simulation shows that the electron-phonon scattering exerts little influence on the electron transport properties of the ultra-scaled PCM nanostructures.

Since the channel length of the ultra-scaled PCM devices is shorter than the electron mean free path, the electron transport is largely elastic.

5.2.3 Microscopic Simulation

In microscopic simulation, both the amorphization (RESET operation) and crystallization (SET operation) are simulated using the density functional theory (DFT) and the ab initio molecular dynamics (AIMD) simulations, respectively. The atomic coordinates generated by using melt-quench AIMD simulation and the Hamiltonian and overlap matrices obtained by using DFT simulation are used as input of mesoscopic simulations. The crystallization properties of the prototypical binary PCM GeTe are investigated by using thermodynamic theory and the annealing AIMD simulations. The temperature dependence of the physical quantities that determine the crystallization properties is calculated and analyzed. It is shown that the critical nuclei radius of the crystalline cluster is smaller than 1.4 nm when the annealing temperature is lower than 600 K, indicating extremely promising scaling scenario. The analysis reveals that the elastic energy, which is largely ignored in the existing PCM crystallization studies, plays an important role in determining various crystallization properties and the ultimate scaling limit of PCM. By omitting the influence of elastic energy, the critical formation energy (critical nuclei radius) will be underestimated by 41.7% (22.4%) and the nucleation rate will be overestimated by 74.2% when the annealing temperature is 600 K.

5.3 Future Research

The PCM technology is undergoing rapid development. In order to enable greener, denser, and faster non-volatile information storage solutions, a lot of research need to be done in the future.

5.3.1 Energy Consumption

It is known that one of the most serious bottlenecks of the large-scale application of the PCM technology is the large RESET operation energy and current of the contemporary

PCM devices. There are several ways to reduce RESET energy and current, to offer greener data storage devices and to elongate the electronic device battery life. Firstly, the RESET energy and current can be significantly reduced by more effectively confining thermal energy. As discussed in Chapter 2, more than 90% of the operation energy is wasted as thermal flux into substrate and electrodes in the contemporary PCM device design. Therefore, one of the promising research directions would be to develop thermal confinement techniques, say by (1) optimizing the device geometry, (2) selecting better thermal insulating materials to fabricate substrate and electrode, (3) purposefully designing the material interface to increase thermal boundary resistance, and (4) taking advantage of both the material and interfacial thermoelectric effects, etc.

Secondly, the RESET energy and current can be aggressively reduced, if the amorphization can be achieved via non-melting physical processes in commodity PCM devices. As discussed in Chapter 4, the the bond lengths differ in chalcogenide crystalline PCM, leading to bond energy hierarchy. In the contemporary PCM devices, the RESET operation is achieved via thermally-activated and melting-induced destruction of the crystalline lattice, which destroys not only the weak bonds but also the strong bonds without discrimination. This is the fundamental reason of the large energy consumption. However, the crystalline lattice can possibly be destroyed to achieve amorphization, by selectively destroying the weak bonds using, say coherent phonon excitation or purposefully designed energy pulses.

It worth mentioning that experiments have found that RESET energy can be reduced to at least single fJ per operation [122], while the RESET energy of the contemporary PCM devices is at the order of pJ. There is plenty room for improvements.

5.3.2 Device Speed

Another bottleneck of the PCM technolgy is its long SET operation time of the contemporary PCM devices. This problem can possibly be solved by (1) applying low-amplitude pre-annealing pulses or (2) taking advantage of the material interfaces as the epitaxial growth template.

It has been shown that the SET operation can be finished in only 0.5 ns in laboratory prototype PCM devices [69], which is much faster than the $10^1 - 10^3$ ns SET time in contemporary PCM devices. This difference indicates huge potential for enhancing the PCM device speed.

5.3.3 Data Capacity

Nowadays, the capacity of the commodity PCM devices are still very small (at the order of Mega-Byte) compared to the more mature technologies like DRAM and flash (Giga-Byte or Tera-Byte). The downscaling and multi-level device development are the two promising directions that will lead to significant increase of data density and capacity.

Since it is known that PCM devices can be scaled to single-digit nanometer scale [95, 49, 67], which is much more scalably than DRAM and flash, the PCM technology is very promising to become the next-generation mainstream memory technology to offer large-capacity non-volatile data storage solutions.

Appendix A

IMPLEMENTATION OF SIMULATORS

In order to perform the multiscale simulations of PCM devices, several numerical simulators are implemented. They are used to generate a part of the results presented in this thesis. In this Appendix, the implementation details and the instrucstions to use them are presented.

A.1 PCM Scaling Scenario TD-FEM Simulator

A.1.1 Introduction

The PCM scaling scenario TD-FEM simulator is implemented using the FreeFEM package [1]. It is used to generate the simulation results shown in Section 2.5.3.

A.1.2 Underlying Algorithm

The TD-FEM simulator solves the Laplace equation (Equation 2.2) and the heat transfer equation (Equation 2.1) to obtain the potential distribution and the temperature distribution, respectively.

The Dirichlet boundary conditions are applied at the top and bottom boundaries of the electrodes (AB and A'B' of Figure 2.1). The homogeneous Neumann boundary condition is applied at the outmost simulation domain boundary (AA' of Figure 2.1).

A.1.3 Code Description

The commented code is stored in the directory "./code/1-TDFEM-FreeFEM/commented". The ".edp" files are the input files, which are explained below:

1. "main.edp" is the entry of the code. It controls the batch simulation to be performed.

---

[1] http://www.freefem.org

2. “material.edp” defines the material properties (Table 2.2) used in the simulation.

3. “bisection.edp” performs bisection optimization to search the appropriate bias value, such that the temperature of one cross-section of the PCM active regionperpendicular to the electrical current flow direction is higher than the melting temperature (to ensure that one cross-section of the crystalline PCM can be melted and then quenched into amorphous during RESET operation).

4. “calc.edp” defines the weak form of the governing equations (Section 2.3), and performs the calculations to solve the Laplace equation and the heat transfer equation.

5. “preproc.edp” defines the problem geometry; performs 2-D meshing of the calculation domain; and initializes the parameters for the following calculations.

6. “postproc.edp” performs post-process calculations, in order to obtain current density (Equation 2.3), operation energy (Equation 2.21), and contributions of the four energy consumption components (Figure 2.10, Figure 2.11, and Figure 2.12), etc.

7. “loopbegin.edp” outputs the input information, like device geometry and TBR value in the radial and axial directions, etc.

8. “loopend.edp” outputs the simulation results, like temperature, potential, operation energy, operation current, operation electric field, contribution of the energy consumption components, etc.

A.1.4 Instruction for Users

To run the code, please:

1. install the FreeFEM package by following the manual;

2. copy the code from “./code/1-TDFEM-FreeFEM/commented” to the working directory;

3. type “$ nohup FreeFem++ -v 0 main.edp > output &”;
4. read the “output” file to obtain the PCM device performance results.

Here are several examples to generate the plots shown in Chapter 2:

1. To generate Figure 2.10, go to ./code/1-TDFEM-FreeFEM/isotropic/ratio-1/ and ./code/1-TDFEM-FreeFEM/isotropic/ratio-2/, type “$ nohup FreeFem++ -v 0 main.edp > output &”. Then run Matlab script ./code/1-TDFEM-FreeFEM/isotropic/read1.m to read output file to generate the figures.

2. To generate Figure 2.11: go to ./code/1-TDFEM-FreeFEM/isotropic/ratio-5/ and ./code/1-TDFEM-FreeFEM/isotropic/ratio-6/, type “$ nohup FreeFem++ -v 0 main.edp > output &”. Then run Matlab script ./code/1-TDFEM-FreeFEM/isotropic/read2.m to read output file to generate the figures.

3. To generate Figure 2.12: go to ./code/1-TDFEM-FreeFEM/isotropic/ratio-10/ and ./code/1-TDFEM-FreeFEM/isotropic/ratio-12/, type “$ nohup FreeFem++ -v 0 main.edp > output &”. Then run Matlab script ./code/1-TDFEM-FreeFEM/isotropic/read3.m to read output file to generate the figures.

4. To generate Figure 2.6: modify the integer “n” (mesh controlling factor N_{mesh}) in the “preproc.edp”, then run the FreeFEM by typing “FreeFem++ -v 0 main.edp”. All FreeFEM files are stored in ./code/1-TDFEM-FreeFEM/test-mesh/freefem/. The Matlab script to do plot is stored in ./code/1-TDFEM-FreeFEM/test-mesh/matlab/.

5. To generate Figure 2.7: modify the real variable “dt” (time step for TD-FEM) in the “preproc.edp”, then run the FreeFEM by typing “FreeFem++ -v 0 main.edp”. All FreeFEM files are stored in ./code/1-TDFEM-FreeFEM/test-dt/freefem/. The Matlab script to do plot is stored in ./code/1-TDFEM-FreeFEM/test-dt/matlab/.

6. To generate Figure 2.8: modify the real variable “xtbr” (TBR layer thickness) in the “preproc.edp”, then run the FreeFEM by typing “FreeFem++ -v 0 main.edp”.

All FreeFEM files are stored in ./code/1-TDFEM-FreeFEM/test-TBR-thick/freefem/. The Matlab script to do plot is stored in ./code/1-TDFEM-FreeFEM/test-TBR-thick/matlab/.

7. To generate Figure 2.9: modify the real variable “Lratio” ($\frac{T_2 - T_1}{T_{melt}}$) in the “preproc.edp”, then run the FreeFEM by typing “FreeFem++ -v 0 main.edp”. All FreeFEM files are stored in ./code/1-TDFEM-FreeFEM/test-dTL/freefem/. The Matlab script to do plot is stored in ./code/1-TDFEM-FreeFEM/test-dTL/matlab/.

A.2 PCM Ultimate Scalability MATLAB Simulator

A.2.1 Introduction

The PCM ultimate scalability MATLAB simulator is implemented using the MATLAB platform. The simulator is based on the Green’s function algorithms and simulates the equilibrium elastic electron transport properties. It is used to generate the simulation results shown in Section 3.5.1.

A.2.2 Underlying Algorithm

The code calculates the elastic electron transport properties under equilibrium condition using the Green’s function algorithms. As discussed in Section 3.5.1, both the electron-phonon scattering SCF loop and the potential SCF loop in Figure 3.4 are ignored in the calculation; and the potential is assumed to vary linearly from source to drain using Equation 3.23.

A.2.3 Code Description

The functionalities of the files of this code are described below. For more details, please read the code and the comments in these files, which are stored in the folder “./code/2-GF-MATLAB”.

1. “main.m” is the entry of the code. The user needs to specify various input information (PAO, PL partition, full paths of the Hamiltonian and overlap matrices, and Fermi level, etc.) in this file.

2. “paodefine.m” performs pre-process calculations. It prepares the data pertaining to PAO. Also, it reads the atomic species and 3D atomic coordinates from the SIESTA output.

3. “readdmhs.m” reads the “DMHS.nc” file dumped by the SIESTA and prepare the MATLAB matrices that contain the Hamiltonian and the overlap matrices.

4. “cutmatrix.m” takes the Hamiltonian and overlap matrices as input and partition them into principal layers along the transport direction (z-direction in Figure 3.1), in order to prepare input data structure for block version RGF algorithms (Section 3.4.1.3).

5. “calcsigma.m” calculates the contact self-energy matrices (Equation 3.4 and Equation 3.5) using the algorithm introduced in Section 3.4.1.2.

6. “calctrans.m” calculates the retarded Green’s function (Equation 3.29) and the transmission coefficients (Equation 3.28).

7. “calcldos.m” calculates the local density of states (Equation 3.33) from the retarded Green’s function and the overlap matrices.

A.2.4 Instruction for Users

Before running the code, the user should:

1. copy the “*.m” files introduced in Section A.2.3 into the working directory;

2. generate the Hamiltonian matrix and the overlap matrix (contained in file “DMHS.nc”) using the DFT package SIESTA and specify the full path of the “DMHS.nc” file as the first input variable of the function “readdmhs” in the “main.m” file;

3. calculate the Fermi level (contained in file “scat.EIG”) using the DFT package SIESTA; and specify the full path of the “scat.EIG” file in the “read Fermi level” part of the “main.m” file;

4. output the atomic species and the 3D atomic coordinates of the system under investigation (contained in the "scat.xyz" file) using the DFT package SIESTA; and specify the full path of the "scat.xyz" file in the file "paodefine.m".

5. download the NetCDF-MATLAB interface package to enable MATLAB to fetch data from NetCDF files [2]; and specify the full path of this interface package at the first line of the "main.m" file.

6. open MATLAB and run the "main.m" file to perform the simulations.

A.3 PCM Transport MPI Parallel NEGF C Simulator

A.3.1 Introduction

The PCM transport MPI parallel NEGF C simulator is implemented using the C language. The simulator is based on the NEGF algorithms and simulates both elastic and inelastic electron tranport properties in non-equilibrium operation conditions. The calculation of the Green's functions are parallelized via MPI. It is used to generate the simulation results shown in Section 3.5.2 and Section 3.5.3.

A.3.2 Underlying Algorithm

The code calculates non-equilibrium electron transport properties using NEGF algorithms (Section 3.2.2). The inelastic electron-phonon scattering can be included (Equation 3.16 and Equation 3.17).

In the calculation, the electron-phonon SCF loop in Figure 3.4 is implemented. The self-consistent calculation continues until the convergence criteria (Equation 3.26 and Equation 3.27) are satisfied. The potential SCF loop in Figure 3.4 is ignored. The potential is assumed to vary linearly from source to drain using Equation 3.23

---

[2] http://www.unidata.ucar.edu/downloads/netcdf/netcdf-java-4/index.jsp

A.3.3 Code Description

The source code is stored in the folder "./code/3-NEGF-C-MPI/src". The files are explained below.

1. "Makefile" defines the rules to facilitate the compilation of this code by following the GNU make [103]. This code uses BLAS and LAPACK packages [7] and their efficient implementations (e.g. Intel MKL, ATLAS [118], etc.) to perform the computationally expensive matrix operations. The paths of the libraries of these packages are specified in the "Makefile". The folder "arch" contains several examples "Makefile.machine", which are directly usable on those machines [3].

2. "main.c" is the entry of the code. It initializes the MPI calculations; reads the input data; outputs the system information to screen; drives the electron transport properties simulations ("calciv" function); and finalizes the MPI calculations.

3. "calciv.c" is the most important nexus of the code. It contains only one function "calciv" which calls various functions (intorduced below) to calculate the electron transport properties. The calculation procedure follows the Figure 3.4 and the description in Section A.3.2.

4. "preproc.c" contains the "preproc" function, which is called by the "calciv" function, in order to perform pre-process calculations and preparations for the following simulations. The preparations include: (1) read the Fermi level from SIESTA output file; (2) read the Hamiltonian and overlap matrices output by SIESTA package; (3) initalize data structure for contact self-energy calculations; (4) initalize data structure for computing the Green's functions; (5) read atomic coordinates of the transport simulation system; and (6) prepare data structures for PAO and other system information; etc.

5. "calcdv.c" defines the linear variation of the potential from source to drain.

---

[3] For example, to compile this code on the PNNL Chinook supercomputers, one can use the "arch/Makefile.PNNL" file to replace "Makefile" file and execute command "make" to generate executables

6. “defepoint.c” prepares the various data structures pertaining to the electron energy. These preparations include: (1) define energy spectrum of interests in the NEGF simulations; (2) allocate the energy points to different MPI processes; and (3) initialize the energy dependent data structures (like integrand of current density, contact self-energies, electron-phonon scattering self-energies, Green’s functions, and LDOS, etc.) which are local at different MPI processes.

7. “calcsgf.c” and “postsgf.c” compute the contact self-energies due to the presence of source and drain (Section 3.4.1.2).

8. “preprgf.c”, “calcrgf.c”, and “postrgf.c” compute the Green’s functions using the recursive Green’s function algorithm (Section 3.4.1.3)

9. “calcphscat.c” calculates the electron-phonon in-scattering and out-scattering self-energies from the Greens functions.

10. “postproc.c” performs post-process calculations, to obtain physical quantities of our interests. These quantities inlude: (1) current density; (2) electron and hole densities; (3) transmission; and (4) LDOS, etc.

11. “memfree.c” free the memory that is used to store the data structure of the NEGF simulator.

12. “utility.c” contains some small auxiliary functions, e.g. matrix multiplication solver, linear equations solver, and SCF convergence checking function, etc.

13. “headers.h” defines the headers of the C functions, and some global variables.

A.3.4 Instruction for Users

To use the code, one should [4]:

---

[4] In this thesis, the DFT Hamiltonian is used as input of the NEGF simulator. This NEGF simulator can also be used to analyze electron transport properties from the Hamiltonian obtained from other methods, like tight-binding and effective mass approximation.

1. Install the prerequisite packages (MPI, BLAS, LAPACK).

2. Copy the source code files under the folder "code/3-NEGF-C-MPI/src" into the compilation directory; specify the location of the libraries of the prerequisite packages, e.g., MPI, LAPACK, and BLAS in "Makefile"; type "make" to generate the executable "negf-core".

3. Prepare the data files for contact self-energies calculations. These files include: (1) "ndim" file which defines the dimension of the contact self-energy matrix; (2) "sl", "sd", and "su" files which store the overlap matrix blocks used to compute contact self-energy matrices [5]; (3) "hl", "hd", and "hu" files which store the Hamiltonian matrix blocks used to compute contact self-energy matrices . To generate these files, one can do a seperate single-point DFT calculation of the electrode material using SIESTA. The Hamiltonian matrix and the overlap matrix are partitioned into blocks which represent principal layers. Then the diagonal blocks are "hd" and "sd"; and the lower (upper) off-diagonal blocks are "hl" and "sl" ("hu" and "su"). The full path of the folder that contains these files should be specified in the second line of the "in.log" file.

4. Prepare the data files for the Green's functions calculations. These files include: (1) "hd-i" and "sd-i" ($i = 1, 2, \cdots, N_{PL}$ where N_{PL} is the number of principal layers along the transport direction) which are the diagonal blocks of the Hamiltonian matrix and overlap matrix; (2) "hl-i" and "sl-i" ($i = 1, 2, \cdots, N_{PL} - 1$) which are the lower off-diagonal blocks of the Hamiltonian matrix and overlap matrix; (3) "hu-i" and "su-i" ($i = 1, 2, \cdots, N_{PL} - 1$) which are the upper off-diagonal blocks of the Hamiltonian matrix and overlap matrix; and (4) "ndim" which specifies the dimension of the matrices (the first line is "N_{PL}" and the second line contains "N_{PL}" integers which are dimensions of the square matrices of the diagonal blocks). The full path of the folder that contains these files should be specified in the third line of the "in.log" file.

---

[5] All matrices read and output by this simulator are stored in a column-major manner.

5. Run SIESTA DFT calculations to generate three files “DMHS.nc” (Hamiltonian and overlap matrices), “scat.xyz” (atomic species and coordinates) and “scat.EIG” (Fermi energy). The full path of the folder that contains these files should be specified in the fourth line of the “in.log” file.

6. Write the input file “in.log” for the NEGF simulator. The first line of “in.log” can be any digit which is the label of the system being simulated. The second, third, and fourth lines have been explained before. The fifth line has one number which is the external bias v between source and drain (V). The sixth line is the electron energy spectrum resolution (eV). The seventh line specifies the ϵ_{ext} which controls the range of the electron energy spectrum $[\epsilon_{\mathrm{F}}-v/2-\epsilon_{\mathrm{ext}}, \epsilon_{\mathrm{F}}+v/2+\epsilon_{\mathrm{ext}}]$. The eighth line defines the elastic electron-phonon scattering coupling strength (set as zero when computing the inelastic electron-phonon scattering). The nineth and tenth lines defines the inelastic electron-phonon coupling strength (D_{inel} in Equation 3.16 and Equation 3.17 in the unit of eV^2) and the phonon energy ($\hbar\omega$ in Equation 3.16 and Equation 3.17 in the unit of eV). The eleventh line defines the convergence tolerance (η in Equation 3.24 and Equation 3.25).

7. In the folder that contains the file “in.log”, type “mpirun -np N negf-core | tee output”, where N is the number of CPU cores used to perform the calculation. Here, the locations of mpirun and negf-core executables should be defined in the enviramental variables before running the command in shell.

Appendix B

DERIVATION OF THERMODYNAMIC EQUATIONS

This appendix derives the thermodynamic equations that are used in the microscopic analysis (Chapter 4).

B.1 Derivation of Gibbs free energy difference

The internal energy U (eV) of a thermodynamic system

$$U = U\left(S, V, N_i\right) \tag{B.1}$$

is a function of entropy S (eV/K), volume (Å^3), and number of atom species i ($i = 1, 2, \cdots, N$) N_i. Taking the total differentiation of internal energy,

$$dU = \left(\frac{\partial U}{\partial S}\right)_{V,N_i} dS + \left(\frac{\partial U}{\partial V}\right)_{S,N_i} dV + \sum_i \left(\frac{\partial U}{\partial N_i}\right)_{S,V} dN_i \tag{B.2}$$

If we define temperature T (K), pressure P (Pa), and chemical potential of atom species i μ_i (eV) as

$$T \equiv \left(\frac{\partial U}{\partial S}\right)_{V,N_i}, \; P \equiv -\left(\frac{\partial U}{\partial V}\right)_{S,N_i}, \; \mu_i \equiv \left(\frac{\partial U}{\partial N_i}\right)_{S,V}, \tag{B.3}$$

respectively, we have

$$dU = TdS - PdV + \sum_i \mu_i dN_i \tag{B.4}$$

The enthalpy H (eV) of a thermodynamic system is defined as

$$H = U + PV \tag{B.5}$$

For the NPT system, the total differentiation of enthalpy

$$\begin{aligned} dH &= dU + PdV \\ &= TdS + \sum_i \mu_i dN_i \\ &= TdS \end{aligned} \tag{B.6}$$

gives

$$dS\left(T\right)=\frac{dH\left(T\right)}{T} \tag{B.7}$$

By taking integration, we have

$$\begin{aligned} S\left(T\right) &= S\left(T_m-\right)+\int_{T_m-}^{T}\frac{dH\left(T\right)}{T} \\ &= S\left(T_m-\right)+\int_{T_m-}^{T}\frac{cV}{T}dT \end{aligned} \tag{B.8}$$

where c is the specific heat capacity at constant pressure (eV/K$\mathring{\mathrm{A}}^3$), and T_m is the melting temperature (K). Therefore,

$$s\left(T\right)=s\left(T_m-\right)+\int_{T_m-}^{T}\frac{c}{T}dT \tag{B.9}$$

where s is the entropy density (eV/K$\mathring{\mathrm{A}}^3$), which is written as S hereafter for consistency with the main context. For amorphous GeTe,

$$S_a\left(T\right)=S_a\left(T_m-\right)+\int_{T_m-}^{T}\frac{c_a}{T}dT \tag{B.10}$$

where S_a is the amorphous GeTe entropy density (eV/K$\mathring{\mathrm{A}}^3$) and c_a is the amorphous GeTe specific heat capacity at constant pressure (eV/K$\mathring{\mathrm{A}}^3$). For crystalline GeTe,

$$S_c\left(T\right)=S_c\left(T_m-\right)+\int_{T_m-}^{T}\frac{c_c}{T}dT \tag{B.11}$$

where S_c is the crystalline GeTe entropy density (eV/K$\mathring{\mathrm{A}}^3$) and c_c is the crystalline GeTe specific heat capacity at constant pressure (eV/K$\mathring{\mathrm{A}}^3$). Therefore,

$$\begin{aligned} \Delta S_{ac}\left(T\right) &= S_a\left(T\right)-S_c\left(T\right) \\ &= S_a\left(T_m-\right)-S_c\left(T_m-\right)+\int_{T_m-}^{T}\frac{c_a-c_c}{T}dT \\ &= S_l\left(T_m+\right)-S_c\left(T_m-\right)+\int_{T_m-}^{T}\frac{c_a-c_c}{T}dT \\ &= \frac{\Delta H_m}{T_m}+\int_{T_m-}^{T}\frac{c_a-c_c}{T}dT \end{aligned} \tag{B.12}$$

where S_l is the liquid GeTe entropy density (eV/K$\mathring{\mathrm{A}}^3$) and ΔH_m is the latent heat of fusion of GeTe (eV/$\mathring{\mathrm{A}}^3$).

The Gibbs free energy of a system is defined as

$$G = U + PV - TS \tag{B.13}$$

For the NPT system in an unconstrained state (P=0),

$$G = U - TS \tag{B.14}$$

Therefore,

$$\begin{aligned} \Delta G_{ac}(T) &= G_a(T) - G_c(T) \\ &= [U_a(T) - U_c(T)] - T[S_a(T) - S_c(T)] \\ &= \Delta U_{ac}(T) - T\Delta S_{ac}(T) \\ &= \Delta U_{ac}(T) - T\left[\frac{\Delta H_m}{T_m} + \int_{T_m-}^{T} \frac{c_a - c_c}{T} dT\right] \end{aligned} \tag{B.15}$$

which is the Equation 4.15.

B.2 Derivation of the interfacial energy density

To obtain the interfacial energy density $\sigma(T)$ (eV/Å^2), the NPT AIMD is used to simulate three systems with periodic boundary conditions applied at the supercell boundaries:

1. amorphous GeTe supercell with size $L_x \times L_y \times L_z^a$.
2. crystalline GeTe supercell with size $L_x \times L_y \times L_z^c$.
3. welded supercell with size $L_x \times L_y \times (L_z^a + L_z^c)$ containing both amorphous and crystalline GeTe supercells.

During the NPT simulation, L_x, L_y, L_z^a, and L_z^c are allowed to change to control pressure $P = 0$. The internal energies of the three systems are U_a (eV), U_c (eV), and U_{ac} (eV), respectively.

For the NPT (P=0) systems, the Gibbs free energy (eV)

$$\begin{aligned} G &= U + PV - TS \\ &= U - TS \end{aligned} \tag{B.16}$$

where U is the internal energy (eV), P is the pressure (Pa), V is the volume (Å^3), T is the temperature (K), and S is the entropy (eV/K).

Since the NPT AIMD only last for several ps and no phase change will occur and entropy describes how disordered the system is, the welded system entropy (S_{ac}) will be approximately the same as the sum of crystalline system entropy (S_c) and amorphous system entropy (S_a). Thus,

$$\begin{aligned} G_{ac} - (G_a + G_c) &= U_{ac} - (U_a + U_c) - T\left[S_{ac} - (S_a + S_c)\right] \\ &\approx U_{ac} - (U_a + U_c) \end{aligned} \tag{B.17}$$

Therefore,

$$\sigma(T) = \frac{U_{ac} - (U_a + U_c)}{2A} \tag{B.18}$$

where A is the interface area (Å^2).

B.3 Derivation of the Murnaghan equation of state

The bulk modulus is defined as

$$B = -V\left(\frac{\partial P}{\partial V}\right)_T \tag{B.19}$$

and the pressure derivative of bulk modulus is

$$B' = \left(\frac{\partial B}{\partial P}\right)_T \tag{B.20}$$

where V is the volume (Å^3), T is the temperature (K), and P is the pressure (Pa).

Since B' is experimentally found to change little with pressure, $B' = B'_0$ (the subscript 0 means $P = 0$, i.e. the system has the most energetically favorable configuration). Thus,

$$\begin{aligned} B &= B_0 + \int_0^P B' dP \\ &= B_0 + B'_0 P \\ &= -V\left(\frac{\partial P}{\partial V}\right)_T \end{aligned} \tag{B.21}$$

Therefore,

$$\frac{dV}{V} = \frac{-dP}{B_0 + B'_0 P} = \frac{-d\left(P + \frac{B_0}{B'_0}\right)}{B'_0\left(P + \frac{B_0}{B'_0}\right)} \tag{B.22}$$

Taking the integration,

$$\begin{aligned} \ln \frac{V}{V_0} &= \left[-\frac{1}{B_0} \ln \left(P + \frac{B_0}{B_0'}\right)\right]_0^P \\ &= \ln \frac{\left(P + \frac{B_0}{B_0'}\right)^{-\frac{1}{B_0}}}{\left(\frac{B_0}{B_0'}\right)^{-\frac{1}{B_0}}} \\ &= \ln \left(1 + \frac{B_0'}{B_0} P\right)^{-\frac{1}{B_0}} \end{aligned} \tag{B.23}$$

So,

$$V(P) = V_0 \left(1 + \frac{B_0'}{B_0} P\right)^{-\frac{1}{B_0}} \tag{B.24}$$

i.e.,

$$P(V) = \frac{B_0}{B_0'} \left[\left(\frac{V_0}{V}\right)^{B_0'} - 1\right] \tag{B.25}$$

According to Equation B.4

$$dU = TdS - PdV + \sum_i \mu_i dN_i \tag{B.26}$$

the internal energy of crystalline GeTe is changed by

$$dU \approx -PdV \tag{B.27}$$

when the volume is change by dV. Here, we take advantage of that the entropy S is a description of how disordered the system is, thus $-PdV$ is much more important than TdS and it dominates dU.

By taking the integration,

$$\begin{aligned} U(V) &= U(V_0) - \int_{V_0}^{V} PdV \\ &= U(V_0) - \int_{V_0}^{V} \frac{B_0}{B_0'} \left(\frac{V_0}{V}\right)^{B_0'} dV + \int_{V_0}^{V} \frac{B_0}{B_0'} dV \\ &= U(V_0) - \frac{B_0 V_0^{B_0'}}{B_0'(1 - B_0')} \left(V^{1-B_0'} - V_0^{1-B_0'}\right) + \frac{B_0}{B_0'}(V - V_0) \\ &= U(V_0) + \frac{B_0 V}{B_0'} \left[\frac{1}{B_0' - 1} \left(\frac{V_0}{V}\right)^{B_0'} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \end{aligned} \tag{B.28}$$

So, the elastic energy density (eV/Å^3) is

$$E_e(T) = \frac{U(V) - U(V_0)}{V} \tag{B.29}$$
$$= \frac{B_0}{B_0'}\left[\frac{1}{B_0'-1}\left(\frac{V_0}{V}\right)^{B_0'} + 1\right] - \frac{B_0}{B_0'-1}\frac{V_0}{V}$$

i.e., Equation 4.17. Since $V_0 = V_0(T)$, $B_0 = B_0(T)$, and $B_0' = B_0'(T)$, E_e is temperature dependent too.

Appendix C

BOND ENERGY ANALYSIS

The code of open source density functional theory (DFT) package SIESTA [40] is modified, in order top perform bond energy analysis (Section 4.3.2.4). The modified SIESTA code is called SIESTA-bond hereafter.

This appendix describes the underlying analysis methodology, numerical implementation, and tutorials to use SIESTA-Bond. In Section C.1, the numerical algorithm that underlies the SIESTA-Bond is described. Numerical implementation details are introduced in Section C.2 and Section C.3. In Section C.4, the test cases of SIESTA-Bond are presented to show the simulation validity. Section C.5 contains the hands-on tutorials to install and use SIESTA-Bonds.

C.1 Simulation Methodology

The SIESTA-Bond implements the numerical scheme proposed in [74, 112], to compute the bond energy. In this Section, the simulation methodology is introduced.

The DFT Kohn-Sham total energy (Section C.1.1) consists of five components – kinetic energy of electrons, ion-electron energy, electron-electron Hartree energy, exchange-correlation energy, and ion-ion energy. The decomposition of the ion-ion energy is introduced in Section C.1.2. The decomposition of the first four components is introduced in Section C.1.3.

C.1.1 DFT Kohn-Sham Energy

The DFT Kohn-Sham total energy E^{KS} can be written as [1]

---

[1] Please read SIESTA method paper [40] for explanations of the variables.

$$\begin{aligned} E^{KS} &= \sum_{\mu\nu} H_{\mu\nu}\rho_{\mu\nu} - \frac{1}{2}\int V^{H}\left(\mathbf{r}\right)\rho\left(\mathbf{r}\right)d^{3}\mathbf{r} \\ &\quad + \int\left(\epsilon^{xc}\left(\mathbf{r}\right) - V^{xc}\left(\mathbf{r}\right)\right)\rho\left(\mathbf{r}\right)d^{3}\mathbf{r} + \sum_{I<J}\frac{Z_I Z_J}{R_{IJ}} \end{aligned} \tag{C.1}$$

By transforming the density operator from orbital representation into Cartesian representation

$$\rho\left(\mathbf{r}\right) = \sum_{\mu\nu}\rho_{\mu\nu}\phi_{\mu}\left(\mathbf{r}\right)\phi_{\nu}\left(\mathbf{r}\right) \tag{C.2}$$

we have

$$\begin{aligned} E^{KS} &= \sum_{\mu\nu} H_{\mu\nu}\rho_{\mu\nu} - \frac{1}{2}\sum_{\mu\nu}\rho_{\mu\nu}\int V^{H}\left(\mathbf{r}\right)\phi_{\mu}\left(\mathbf{r}\right)\phi_{\nu}\left(\mathbf{r}\right)d^{3}\mathbf{r} \\ &\quad + \sum_{\mu\nu}\rho_{\mu\nu}\int\left(\epsilon^{xc}\left(\mathbf{r}\right) - V^{xc}\left(\mathbf{r}\right)\right)\phi_{\mu}\left(\mathbf{r}\right)\phi_{\nu}\left(\mathbf{r}\right)d^{3}\mathbf{r} + \sum_{I<J}\frac{Z_I Z_J}{R_{IJ}} \end{aligned}$$

i.e.,

$$\begin{aligned} E^{KS} &= \sum_{\mu\nu}\rho_{\mu\nu}H_{\mu\nu} - \sum_{\mu\nu}\rho_{\mu\nu}\frac{1}{2} < \phi_{\mu}|V^{H}|\phi_{\nu} > \\ &\quad + \sum_{\mu\nu}\rho_{\mu\nu} < \phi_{\mu}|\left(\epsilon^{xc} - V^{xc}\right)|\phi_{\nu} > + \sum_{I<J}\frac{Z_I Z_J}{R_{IJ}} \end{aligned} \tag{C.3}$$

Equation C.3 can be written as

$$E^{KS} = \sum_{IJ} M_{IJ} + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{R_{IJ}} \tag{C.4}$$

with

$$M_{IJ} = \sum_{\mu\in I,\nu\in J}\rho_{\mu\nu}\left[H_{\mu\nu} - \frac{1}{2}\left(M^{(1)}_{\mu\nu} + M^{(2)}_{\mu\nu}\right) + M^{(3)}_{\mu\nu}\right] \tag{C.5}$$

where, $\mu \in I$ means that orbital μ is centered at ion I, the matrices

$$M^{(1)}_{\mu\nu} =< \phi_{\mu}|V^{H}_{\rho_{atom}}|\phi_{\nu} > \tag{C.6}$$

$$M^{(2)}_{\mu\nu} =< \phi_{\mu}|V^{H}_{\delta\rho}|\phi_{\nu} > \tag{C.7}$$

$$M^{(3)}_{\mu\nu} =< \phi_\mu | \left(\epsilon^{xc} - V^{xc}\right) |\phi_\nu > \tag{C.8}$$

are dumped by modifying SIESTA. Here, V^H consists of two components

$$V^H = V^H_{\rho_{atom}} + V^H_{\delta\rho} \tag{C.9}$$

which are calculated by solving the Poisson equation, from source ρ_{atom} and $\delta\rho$, respectively. Their sum

$$\rho = \rho_{atom} + \delta\rho \tag{C.10}$$

is the electron density.

C.1.2 Decomposition of Ion-Ion Energy Component

The decomposition of the ion-ion energy (last term in Equation C.4) is straightforward. The ion-ion energy between ion I and ion J is simply $\frac{1}{2}\frac{Z_I Z_J}{R_{IJ}}$.

C.1.3 Decomposition of other DFT Energy Components

We can take advantage of the localized nature of SIESTA pseudo atomic orbitals, to decompose the kinetic energy, ion-electron energy, electron-electron Hartree energy, and exchange-correlation energy. As shown in Equation C.4, the contribution from ion pair IJ is M_{IJ}.

C.2 Modification of SIESTA

As shown in Equation C.5, five matrices ($\rho_{\mu\nu}$, $H_{\mu\nu}$, $M^{(1)}_{\mu\nu}$, $M^{(2)}_{\mu\nu}$, $M^{(3)}_{\mu\nu}$) are needed for bond energy calculation. $\rho_{\mu\nu}$ and $H_{\mu\nu}$ can be dumped by using unmodified SIESTA by turning on the flag “WriteDMHS.NetCDF” in the input.fdf file.

However, to obtain $M^{(1)}_{\mu\nu}$, $M^{(2)}_{\mu\nu}$, and $M^{(3)}_{\mu\nu}$, the SIESTA source code should be modified. The modifications made to dump these three matrices are introduced in Section C.2.1, Section C.2.2, and Section C.2.3, respectively. They are implemented in the SIESTA source code file “dhscf.F”.

C.2.1 Hartree Contribution due to ρ_{atom}

There are 3 steps to obtain the matrix $M^{(1)}_{\mu\nu}$.

(1) Solution of the Poisson equation to obtain $V^{H}_{\rho_{atom}}$ from ρ_{atom}:

```
C Solve Poisson’s equation to find Hartree energy of rhoatm
        call reord( rhoatm, rhoatm, nml, nsm, +1)
        call poison( cell, ntml(1), ntml(2), ntml(3), ntm, rhoatm,
     .               Uharrs, Vscf, DStres, nsm )
        call reord( rhoatm, rhoatm, nml, nsm, -1)
```

(2) Evaluation of the matrix elements:

```
!       matrix of <phi_i|v_rhoatm|phi_j>
        mat = 0.0_dp
        if (spiral) then
           call vmatsp( norb, nmpl, dvol, nspin, v_rhoatm, maxnd,
     .          numd, listdptr, listd, mat, nuo,
     .          nuotot, iaorb, iphorb, isa, qspiral )
        else
           call vmat( norb, nmpl, dvol, nspin, v_rhoatm, maxnd,
     .          numd, listdptr, listd, mat, nuo,
     .          nuotot, iaorb, iphorb, isa )
        endif
```

(3) Store the matrix:

```
!       Store matrix
        flagfile = 1
        call setup_mat_netcdf_file( maxnh, no_l, nspin,
     $       no_s, indxuo,
     $       numh, listhptr, listh, flagfile)
        call write_mat_netcdf( no_l, maxnh, nspin, mat,
     $       overwrite=.true. )
```

C.2.2 Hartree Contribution due to $\delta\rho$

Similar to the procedure introduced in Section C.2.1, the matrix $M^{(2)}_{\mu\nu}$ is also obtained using 3 steps. The difference is that ρ_{atom} is changed to $\delta\rho$ here.

C.2.3 Exchange Correlation Contribution

There are three steps to dump the matrix $M^{(3)}_{\mu\nu}$.

(1) ϵ_{xc} and V_{xc} are evaluated:

```
      call cellxcnew(0, 0, cell, ntml, ntml, ntpl, 0, aux3, nspin, DRho,
     .             Ex, Ec, DEx, DEc, Vscf, e_xc, DVxcdn_dummy, stressl)
```

```
!     xc potential
      v_xc = Vscf
```

(2) the matrix elements are computed:

```
!     matrix of <phi_i|e_xc-v_xc|phi_j>
      mat = 0.0_dp    if (spiral) then
         call vmatsp( norb, nmpl, dvol, nspin, e_xc-v_xc, maxnd,
     .        numd, listdptr, listd, mat, nuo,
     .        nuotot, iaorb, iphorb, isa, qspiral )
      else
         call vmat( norb, nmpl, dvol, nspin, e_xc-v_xc, maxnd,
     .        numd, listdptr, listd, mat, nuo,
     .        nuotot, iaorb, iphorb, isa )
      endif
```

(3) the matrix is stored:

```
!     Store matrix
      flagfile = 3
      call setup_mat_netcdf_file( maxnh, no_l, nspin,
```

```
     $     no_s, indxuo,
     $     numh, listhptr, listh, flagfile)
      call write_mat_netcdf( no_l, maxnh, nspin, mat,
     $     overwrite=.true. )
```

C.3 Ewald Summation

The Ewald summation technique [71] is applied to compute the ion-ion interaction energy per supercell under periodic boundary condition

$$\frac{1}{2}\sum_{IJ} Z_I Z_J \left[\frac{4\pi}{\Omega} \sideset{}{'}\sum_{\mathbf{G}\neq 0} \frac{1}{|\mathbf{G}|^2} e^{-\frac{|\mathbf{G}|^2}{4\eta^2}} \cos\left(\mathbf{G}\cdot\mathbf{r}_{IJ}\right) - \frac{2\eta}{\sqrt{\pi}}\delta_{IJ} + \sideset{}{'}\sum_{\mathbf{T}} \frac{\operatorname{erfc}\left(\eta|\mathbf{r}_{IJ}-\mathbf{T}|\right)}{|\mathbf{r}_{IJ}-\mathbf{T}|} \right] \tag{C.11}$$

where Z_I and Z_J are the valence charges, Ω is the volume of supercell, $\mathbf{G}$ is the translation vector in the reciprocal space, $\mathbf{T}$ is the translation vector in the direct space, $\mathbf{r}_{IJ}$ is the vector between atom I and atom J, and η is a constant.

In Equation C.11, the first term can be evaluated in the reciprocal space efficiently; the last term can be evaluated in the direct space efficiently. The implementation is done using MATLAB (calcewald.m file in the source directory).

C.4 Test Cases

To test the validity of the SIESTA-Bond package, the bond energy of crystalline silicon, germanium, and tin (diamond lattice structures) are computed. During the calculation, PBE GGA is used as exchange-correlation functional. In the diamond lattice structure supercell, 64 atoms are included. Γ point is used to sample the Brillouin zone. DZP is used as the pseudo atomic orbital basis.

As shown in Table C.1, the SIESTA-Bond results can correctly predict the relative bond strength ($E_{Si} > E_{Ge} > E_{Sn}$). Compared against experimental measurement, the SIESTA-Bond underpredicts the bond energy by 0.2-0.4 eV.

It worth noting that the precision of energy decomposition calculation depends on a lot of factors, like DFT parameters and bond strength. The developer of SIESTA-Bond does not guarantee the simulation validity – the users themselves are responsible to check the results carefully.

Table C.1: Test cases of SIESTA-Bond simulation

	SIESTA-Bond E_{sim} (eV)	experiment E_{exp} (eV)	$\lvert E_{sim} - E_{exp} \rvert$ (eV)
Si	1.53	1.82	-0.29
Ge	1.38	1.65	-0.27
Sn	1.12	1.51	-0.39

C.5 Hands-on Tutorial

C.5.1 Installation

The installation and parallelization of SIESTA-Bond are identical to that of SIESTA. The users are recommended to first install and parallelize SIESTA on their computers. Then, they can use the precisely the same procedure to install and parallelize SIESTA-Bond.

To enable SIESTA-Bond simulation, the compilation must be done with NetCDF enabled, since the five matrices in Equation C.5 needed to perform bond energy analysis are in NetCDF format. The users are refered to SIESTA manual for technical details.

C.5.2 Example Run

The folder “example” contains four subfolders:

1. “coord” contains python script to generate coordinates for the example crystalline silicon calculation.
2. “dft” contains input files of SIESTA-Bond.
3. “epart-system” contains MATLAB scripts to compute bond energy.
4. “netcdf” contains NetCDF interface to read “*.nc” files dumped by SIESTA-Bond.

C.5.2.1 DFT Calculation

The first step is to go into the “dft/system” directory, and run

```
$MPI/bin/mpirun -np 2 $SIESTA/Obj/siesta < input.fdf | tee output &
```

in the terminal. After the DFT calculation is completed, the user should see “DMHS.nc”, “mat-1.nc”, “mat-2.nc”, and “mat-3.nc” files in the “dft/system” directory. If not, it's recommended to check whether or not the installation is correct, especially the NetCDF interface in SIESTA installation process.

C.5.2.2 MATLAB Postprocess

Then, the user can go to “epart-system” directory in the MATLAB. The “main.m” file should be modified according to the system to be calculated. For example, if the system contains two species “Ge” and “Te”, which have 4 and 6 valence electrons respectively, the “main.m” file should look like:

```
in.ele = {'Ge','Te'};
in.valence = [4,6];
in.ndirect = 20;
in.nreciprocal = 30;
```

Here, “in.ndirect” (“in.nreciprocal”) controls the number of repeated supercells in direct (reciprocal) space. The larger these two values, the more accurate and CPU-intensive the calculations. The user can make a test and strike a balance between accuracy and efficiency.

After the parameters in the “main.m” are correctly set, one can type “main” in the MATLAB command window to run the simulation. The calculation should end up with a matrix “bondm03” in the workspace. The value of the (I, J) element of this matrix times 2 is the bond energy between atom pair $I - J$. The positive value means repulsive energy, and the negative value means attractive energy.

BIBLIOGRAPHY

[1] H.I. Aaronson, G. Spanos, R.A. Masamura, R.G. Vardiman, D.W. Moon, E.S.K. Menon, and M.G. Hall. Sympathetic nucleation: an overview. *Materials Science and Engineering: B*, 32(3):107–123, July 1995.

[2] J. Akola and R. O. Jones. Structural phase transitions on the nanoscale: The crucial pattern in the phase-change materials ge2sb2te5 and gete. *Phys. Rev. B*, 76(23):235201, December 2007.

[3] J Akola and R O Jones. Density functional study of amorphous, liquid and crystalline ge2sb2te5: homopolar bonds and/or ab alternation. *Journal of Physics: Condensed Matter*, 20(46):465103, 2008.

[4] J. Akola and R. O. Jones. Structure of amorphous ge8sb2te11: Gete-sb2te3 alloys and optical storage. *Phys. Rev. B*, 79(13):134118, April 2009.

[5] J. Akola, R. O. Jones, S. Kohara, S. Kimura, K. Kobayashi, M. Takata, T. Matsunaga, R. Kojima, and N. Yamada. Experimentally constrained density-functional calculations of the amorphous structure of the prototypical phase-change material ge2 sb2 te5. *Phys. Rev. B*, 80(2):020201, July 2009.

[6] J. Akola, J. Larrucea, and R. O. Jones. Polymorphism in phase-change materials: melt-quenched and as-deposited amorphous structures in ge2sb2te5 from density functional calculations. *Phys. Rev. B*, 83(9):094113, March 2011.

[7] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999.

[8] Greg Atwood. Engineering: Phase-change materials for electronic memories. *Science*, 321(5886):210–211, 2008.

[9] S. K. Bahl and K. L. Chopra. Amorphous versus crystalline gete films. iii. electrical properties and band structure. *Journal of Applied Physics*, 41(5):2196–2212, 1970.

[10] D. A. Baker, M. A. Paesler, G. Lucovsky, S. C. Agarwal, and P. C. Taylor. Application of bond constraint theory to the switchable optical memory material ge2sb2te5. *Phys. Rev. Lett.*, 96(25):255501, June 2006.

[11] Robert F. Brebrick. Partial pressures in equilibrium with group iv tellurides. iii. germanium telluride. *J. Chem. Phys.*, 41(4):1140–1146, August 1964.

[12] M. Breitwisch, T. Nirschl, C.F. Chen, Y. Zhu, M.H. Lee, M. Lamorey, G.W. Burr, E. Joseph, A. Schrott, J.B. Philipp, R. Cheek, T.D. Happ, S.H. Chen, S. Zaidr, P. Flaitz, J. Bruley, R. Dasaka, B. Rajendran, S. Rossnage, M. Yang, Y.C. Chen, R. Bergmann, H.L. Lung, and C. Lam. Novel lithography-independent pore phase change memory. In *VLSI Technology, 2007 IEEE Symposium on DOI - 10.1109/VLSIT.2007.4339743*, pages 100–101, 2007.

[13] G. Bruns, P. Merkelbach, C. Schlockermann, M. Salinga, M. Wuttig, T. D. Happ, J. B. Philipp, and M. Kund. Nanosecond switching in gete phase change memory cells. *Appl. Phys. Lett.*, 95(4):043108–3, July 2009.

[14] Geoffrey W. Burr, Matthew J. Breitwisch, Michele Franceschini, Davide Garetto, Kailash Gopalakrishnan, Bryan Jackson, Bulent Kurdi, Chung Lam, Luis A. Lastras, Alvaro Padilla, Bipin Rajendran, Simone Raoux, and Rohit S. Shenoy. Phase change memory technology. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 28(2):223–262, 2010.

[15] Fabrizio Buscemi, Enrico Piccinini, Rossella Brunetti, Massimo Rudan, and Carlo Jacoboni. Monte carlo simulation of charge transport in amorphous chalcogenides. *J. Appl. Phys.*, 106(10):103706–6, November 2009.

[16] Marissa A. Caldwell, Rakesh Gnana David Jeyasingh, H.-S. Philip Wong, and Delia J. Milliron. Nanoscale phase change memory materials. *Nanoscale*, pages –, 2012.

[17] Marissa A. Caldwell, Simone Raoux, Robert Y. Wang, H.-S. Philip Wong, and Delia J. Milliron. Synthesis and size-dependent crystallization of colloidal germanium telluride nanoparticles. *J. Mater. Chem.*, 20(7):1285–1291, 2010.

[18] Arrigo Calzolari, Carlo Cavazzoni, and Marco Buongiorno Nardelli. Electronic and transport properties of artificial gold chains. *Phys. Rev. Lett.*, 93(9):096404–, August 2004.

[19] L. L. Chang, P. J. Stiles, and L. Esaki. Electron barriers in al-al2o3-snte and al-al2o3-gete tunnel junctions. *IBM Journal of Research and Development*, 10(6):484–486, 1966.

[20] Der-Sheng Chao, Yi-Chan Chen, F. Chen, Ming-Jung Chen, P.H. Yen, Chain-Ming Lee, Wei-Su Chen, Chenhsin Lien, Ming-Jer Kao, and Ming-Jinn Tsai. Enhanced thermal efficiency in phase-change memory cell by double gst thermally confined structure. *Electron Device Letters, IEEE DOI - 10.1109/LED.2007.906084*, 28(10):871–873, 2007.

[21] T Chattopadhyay, J X Boucherle, and H G vonSchnering. Neutron diffraction study on the structural phase transition in gete. *Journal of Physics C: Solid State Physics*, 20(10):1431–1440, 1987.

[22] I-Ru Chen and E. Pop. Compact thermal model for vertical nanowire phase-change memory cells. *Electron Devices, IEEE Transactions on DOI - 10.1109/TED.2009.2021364*, 56(7):1523–1528, 2009.

[23] Y.C. Chen, C.T. Rettner, S. Raoux, G.W. Burr, S.H. Chen, R.M. Shelby, M. Salinga, W.P. Risk, T.D. Happ, G.M. McClelland, M. Breitwisch, A. Schrott, J.B. Philipp, M.H. Lee, R. Cheek, T. Nirschl, M. Lamorey, C.F. Chen, E. Joseph, S. Zaidi, B. Yee, H.L. Lung, R. Bergmann, and C. Lam. Ultra-thin phase-change bridge memory device using gesb. In *Electron Devices Meeting, 2006. IEDM '06. International DOI - 10.1109/IEDM.2006.346910*, pages 1–4, 2006.

[24] Yi-Chou Chen, Yuyu Lin, Shih-Hung Chen, Huai-Yu Cheng, Hsiang-Lan Lung, S. Raoux, C.T. Rettner, G.W. Burr, and C.H. Lam. Characteristics of a highly scalable bridge phase change memory. In *Solid-State and Integrated-Circuit Technology, 2008. ICSICT 2008. 9th International Conference on DOI - 10.1109/ICSICT.2008.4734693*, pages 909–912, 2008.

[25] J. W. Christian. *The Theory of Transformations in Metals and Alloys.* Elsevier Science Ltd., 2002.

[26] Hee-Suk Chung, Yeonwoong Jung, Seul Cham Kim, Do Hyun Kim, Kyu Hwan Oh, and Ritesh Agarwal. Epitaxial growth and ordering of gete nanowires on microcrystals determined by surface energy minimization. *Nano Letters*, 9(6):2395–2401, June 2009.

[27] Datta. *Electronic Transport in Mesoscopic Systems.* 1995.

[28] Supriyo Datta. *Quantum transport atom to transistor.* Cambridge University Press, Cambridge, UK, 2005.

[29] S.L.Cho-H.G.An D. H. Kim I. S. Kim H. Park D. H. Ahn H. Horii S. O. Park U-I. Chung D.H.Im, J.I.Lee and B J. T. Moon. A unified 7.5 nm dash-type confined cell for high performance pram device. *Proc. IEEE Int. Electron Devices Meeting*, 2008.

[30] M. Forst, T. Dekorsy, C. Trappe, M. Laurenzis, H. Kurz, and B. Bechevet. Phase change in ge[sub 2]sb[sub 2]te[sub 5] films investigated by coherent phonon spectroscopy. *Appl. Phys. Lett.*, 77(13):1964–1966, September 2000.

[31] J. Hegedus and S. R. Elliott. Microscopic origin of the fast crystallization ability of ge-sb-te phase-change memory materials. *Nat Mater*, 7(5):399–405, May 2008.

[32] J. Hegedus and S. R. Elliott. Computer-simulation design of new phase-change memory materials. *physica status solidi (a)*, 207(3):510–515, 2010.

[33] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, March 1985.

[34] Stephen Hudgens. Progress in understanding the ovshinsky effect: Threshold switching in chalcogenide amorphous semiconductors. *Phys. Status Solidi B*, 249(10):1951–1955, 2012.

[35] Daniele Ielmini. Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses. *Phys. Rev. B*, 78(3):035308, July 2008.

[36] Daniele Ielmini and Yuegang Zhang. Analytical model for subthreshold conduction and threshold switching in chalcogenide-based memory devices. *J. Appl. Phys.*, 102(5):054517, September 2007.

[37] Daniele Ielmini and Yuegang Zhang. Evidence for trap-limited transport in the subthreshold conduction regime of chalcogenide glasses. *Appl. Phys. Lett.*, 90(19):192102, May 2007.

[38] Andrew T. Jennings, Yeonwoong Jung, Johanna Engel, and Ritesh Agarwal. Diameter controlled synthesis of phase change germanium telluride nanowires via the vapor-liquid-solid mechanism. *The Journal of Physical Chemistry C*, 113(17):6898–6901, April 2009.

[39] Rakesh G. D. Jeyasingh, Marissa A. Caldwell, Jiale Liang, Chiyui Ahn, and H.-S. Philip Wong. Methodologies to study the scalability and physics of phase change memory devices.

[40] Julian D Gale Alberto Garca Javier Junquera Pablo Ordejon Daniel Sanchez-Portal Jose M Soler, Emilio Artacho. The siesta method for ab initio order- n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745–2779, 2002.

[41] Yeonwoong Jung, Se-Ho Lee, Andrew T. Jennings, and Ritesh Agarwal. Core-shell heterostructured phase change nanowire multistate memory. *Nano Letters*, 8(7):2056–2062, July 2008.

[42] Yeonwoong Jung, Se-Ho Lee, Dong-Kyun Ko, and Ritesh Agarwal. Synthesis and characterization of ge2sb2te5 nanowires with memory switching effect. *Journal of the American Chemical Society*, 128(43):14026–14027, November 2006.

[43] Yeonwoong Jung, Sung-Wook Nam, and Ritesh Agarwal. High-resolution transmission electron microscopy study of electrically-driven reversible phase change in ge2sb2te5 nanowires. *Nano Lett.*, 11(3):1364–1368, January 2011.

[44] Yeonwoong Jung, Chung-Ying Yang, Se-Ho Lee, and Ritesh Agarwal. Phase-change gesb nanowires: Synthesis, memory switching, and phase instability. *Nano Letters*, 9(5):2103–2108, May 2009.

[45] J. Kalikka, J. Akola, J. Larrucea, and R. O. Jones. Nucleus-driven crystallization of amorphous ge2sb2te5: A density functional study. *Phys. Rev. B*, 86(14):144113, October 2012.

[46] Cheolkyu Kim, Dongmin Kang, Tae-Yon Lee, Kijoon H. P. Kim, Youn-Seon Kang, Junho Lee, Sung-Wook Nam, Ki-Bum Kim, and Yoonho Khang. Direct evidence of phase separation in ge[sub 2]sb[sub 2]te[sub 5] in phase change memory devices. *Appl. Phys. Lett.*, 94(19):193504–3, May 2009.

[47] Dae-Hwang Kim, Florian Merget, Michael Forst, and Heinrich Kurz. Three-dimensional simulation model of switching dynamics in phase change random access memory cells. *J. Appl. Phys.*, 101(6):064512–12, March 2007.

[48] S. Kim and H.-S.P. Wong. Analysis of temperature in phase change memory scaling. *Electron Device Letters, IEEE DOI - 10.1109/LED.2007.901347*, 28(8):697–699, 2007.

[49] SangBum Kim, Byoung-Jae Bae, Yuan Zhang, R.G.D. Jeyasingh, Youngkuk Kim, In-Gyu Baek, Soonoh Park, Seok-Woo Nam, and H.-S.P. Wong. One-dimensional thickness scaling study of phase change material(ge2sb2te5) using a pseudo 3-terminal device. *Electron Devices, IEEE Transactions on*, 58(5):1483–1489, 2011.

[50] KolobovA.V., KrbalM., FonsP., TominagaJ., and UrugaT. Distortion-triggered loss of long-range order in solids with bonding energy hierarchy. *Nat Chem*, 3(4):311–316, April 2011.

[51] Olga Koper and Slawomir Winecki. *Specific Heats and Melting Points of Nanocrystalline Materials*. John Wiley & Sons, Inc., New York, 2001.

[52] S.O. Koswatta, S. Hasan, M.S. LUNDSTROM, M. P. Anantram, and D.E. Nikonov. Nonequilibrium green's function treatment of phonon scattering in carbon-nanotube transistors. *Electron Devices, IEEE Transactions on*, 54(9):2339–2351, 2007.

[53] M. Krbal, A. V. Kolobov, P. Fons, J. Tominaga, S. R. Elliott, J. Hegedus, and T. Uruga. Intrinsic complexity of the melt-quenched amorphous ge2sb2te5 memory alloy. *Phys. Rev. B*, 83(5):054203, February 2011.

[54] Daniel Krebs, Simone Raoux, Charles T. Rettner, Geoffrey W. Burr, Martin Salinga, and Matthias Wuttig. Threshold field of phase change memory materials measured using phase change bridge devices. *Appl. Phys. Lett.*, 95(8):082101–3, August 2009.

[55] Daniel Krebs, Simone Raoux, Charles T. Rettner, Geoffrey W. Burr, Robert M. Shelby, Martin Salinga, C. Michael Jefferson, and Matthias Wuttig IBM/Macronix PCRAM Joint Project. Characterization of phase change memory materials using phase change bridge devices. *J. Appl. Phys.*, 106(5):054308–7, September 2009.

[56] Roger Lake, Gerhard Klimeck, R. Chris Bowen, and Dejan Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, 81(12):7845–7869, June 1997.

[57] Martijn H. R. Lankhorst, Bas W. S. M. M. Ketelaars, and R. A. M. Wolters. Low-cost and nanoscale non-volatile memory concept for future silicon chips. *Nat Mater*, 4(4):347–352, April 2005.

[58] S. Lavizzari, D. Ielmini, and A.L. Lacaita. Transient simulation of delay and switching effects in phase-change memories. *Electron Devices, IEEE Transactions on DOI - 10.1109/TED.2010.2078822*, 57(12):3257–3264, 2010.

[59] Jaeho Lee, Zijian Li, John P. Reifenberg, Sangchul Lee, Robert Sinclair, Mehdi Asheghi, and Kenneth E. Goodson. Thermal conductivity anisotropy and grain structure in ge[sub 2]sb[sub 2]te[sub 5] films. *J. Appl. Phys.*, 109(8):084902–6, April 2011.

[60] J.I. Lee, H. Park, S.L. Cho, Y.L. Park, B.J. Bae, J.H. Park, J.S. Park, H.G. An, J.S. Bae, D.H. Ahn, Y.T. Kim, H. Horii, S.A. Song, J.C. Shin, S.O. Park, H.S. Kim, U-In Chung, J.T. Moon, and B.I. Ryu. Highly scalable phase change memory with cvd gesbte for sub 50nm generation. In *VLSI Technology, 2007 IEEE Symposium on DOI - 10.1109/VLSIT.2007.4339744*, pages 102–103, 2007.

[61] Jinil Lee, Sunglae Cho, Dongho Ahn, Mansug Kang, Seokwoo Nam, Ho-Kyu Kang, and Chilhee Chung. Scalable high-performance phase-change memory employing cvd gebite. *Electron Device Letters, IEEE DOI - 10.1109/LED.2011.2157075*, 32(8):1113–1115, 2011.

[62] Minhyea Lee, Liliana Viciu, Lu Li, Yayu Wang, M. L. Foo, S. Watauchi, R. A. Pascal Jr, R. J. Cava, and N. P. Ong. Large enhancement of the thermopower in naxcoo2 at high na doping. *Nat Mater*, 5(7):537–540, July 2006.

[63] Se-Ho Lee, Yeonwoong Jung, and Ritesh Agarwal. Highly scalable non-volatile and ultra-low-power phase-change nanowire memory. *Nat Nano*, 2(10):626–630, October 2007.

[64] Se-Ho Lee, Yeonwoong Jung, and Ritesh Agarwal. Size-dependent surface-induced heterogeneous nucleation driven phase-change in ge2sb2te5 nanowires. *Nano Letters*, 8(10):3303–3309, October 2008.

[65] T. H. Lee and S. R. Elliott. Abinitio computer simulation of the early stages of crystallization: Application to ge2sb2te5 phase-change materials. *Phys. Rev. Lett.*, 107(14):145702, September 2011.

[66] J. Liu, B. Yu, and M. P. Anantram. Scaling analysis of nanowire phase-change memory. *Electron Device Letters, IEEE DOI - 10.1109/LED.2011.2162390*, 32(10):1340–1342, 2011.

[67] Jie Liu and M. P. Anantram. Low-bias electron transport properties of germanium telluride ultrathin films. *J. Appl. Phys.*, 113(6):063711, February 2013.

[68] Jie Liu, Bin Yu, and M.P. Anantram. Isotropic and anisotropic scaling analysis of nanowire phase change memory. In *Nanotechnology (IEEE-NANO), 2011 11th IEEE Conference on DOI - 10.1109/NANO.2011.6144493*, pages 1343–1347, 2011.

[69] D. Loke, T. H. Lee, W. J. Wang, L. P. Shi, R. Zhao, Y. C. Yeo, T. C. Chong, and S. R. Elliott. Breaking the speed limits of phase-change memory. *Science*, 336(6088):1566–1569, 2012.

[70] M. Longo, C. Wiemer, O. Salicio, M. Fanciulli, L. Lazzarini, and E. Rotunno. Au-catalyzed self assembly of gete nanowires by mocvd. *Journal of Crystal Growth*, In Press, Corrected Proof:–, 2010.

[71] Richard M. Martin. *Electronic structure basic theory and practical method*. Cambridge University Press, Cambridge, UK, 2004.

[72] Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics basic theory and advanced methods*. Cambridge University Press, Cambridge, UK, 2009.

[73] Toshiyuki Matsunaga, Jaakko Akola, Shinji Kohara, Tetsuo Honma, Keisuke Kobayashi, Eiji Ikenaga, Robert O. Jones, Noboru Yamada, Masaki Takata, and Rie Kojima. From local structure to nanosecond recrystallization dynamics in aginsbtephase-changematerials. *Nat Mater*, 10(2):129–134, February 2011.

[74] I. Mayer. An exact chemical decomposition scheme for the molecular energy. *Chemical Physics Letters*, 382:265–269, December 2003.

[75] Stefan Meister, David T. Schoen, Mark A. Topinka, Andrew M. Minor, and Yi Cui. Void formation induced electrical switching in phase-change nanowires. *Nano Letters*, 8(12):4562–4567, December 2008.

[76] K. C. Mills. *Thermodynamic Data for Inorganic Sulphides, Selenides and Tellurides.* Butterworth-Heinemann Ltd. London, UK, 1974.

[77] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, June 1976.

[78] F. D. Murnaghan. The compressibility of media under extreme pressures. *Proceedings of the National Academy of Sciences*, 30(9):244–247, 1944.

[79] Sung-Wook Nam, Hee-Suk Chung, Yu Chieh Lo, Liang Qi, Ju Li, Ye Lu, A.T. Charlie Johnson, Yeonwoong Jung, Pavan Nukala, and Ritesh Agarwal. Electrical wind force-driven and dislocation-templated amorphization in phase-change nanowires. *Science*, 336(6088):1561–1566, 2012.

[80] Prem Nath and K. L. Chopra. Thermal conductivity of amorphous and crystalline ge and gete films. *Phys. Rev. B*, 10(8):3412–3418, October 1974.

[81] R. G. Neale, D. L. Nelson, and G. E. Moore. Nonvolatile and reprogrammable, the read mostly memory is here. *Electronics*, pages 56–60, 1970.

[82] N. Nilius, T. M. Wallis, and W. Ho. Localized molecular constraint on electron delocalization in a metallic chain. *Phys. Rev. Lett.*, 90(18):186102–, May 2003.

[83] Toshihisa Nonaka, Gentaro Ohbayashi, Yoshiharu Toriumi, Yuji Mori, and Hideki Hashimoto. Crystal structure of gete and ge2sb2te5 meta-stable phase. *Thin Solid Films*, 370(1-2):258–261, July 2000.

[84] Shuichi Nose. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, July 1984.

[85] C M I Okoye. Electronic and optical properties of snte and gete. *Journal of Physics: Condensed Matter*, 14(36):8625–, 2002.

[86] Stanford R. Ovshinsky. Reversible electrical switching phenomena in disordered structures. *Phys. Rev. Lett.*, 21(20):1450, November 1968.

[87] Parr. *Density Functional Theory of Atoms and Molecules*. 1989.

[88] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, December 1981.

[89] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77(18):3865–3868, October 1996.

[90] A. Pirovano, A.L. Lacaita, A. Benvenuti, F. Pellizzer, S. Hudgens, and R. Bez. Scaling analysis of phase-change memory technology. In *Technical Digest of IEEE International Electron Devices Meeting*, pages 29.6.1–29.6.4, 2003.

[91] K. M. Rabe and J. D. Joannopoulos. Structural properties of gete at t=0. *Phys. Rev. B*, 36(6):3319–3324, August 1987.

[92] S. Raoux, G. W. Burr, M. J. Breitwisch, C. T. Rettner, Y.-C. Chen, R. M. Shelby, M. Salinga, D. Krebs, S.-H. Chen, H.-L. Lung, and C. H. Lam. Phase-change random access memory: A scalable technology. *IBM Journal of Research and Development DOI - 10.1147/rd.524.0465*, 52(4.5):465–479, 2008.

[93] S. Raoux, C.T. Rettner, Yi-Chou Chen, J. Jordan-Sweet, Yuan Zhang, M. Caldwell, H.-S.P. Wong, D. Milliron, and J. Cha. Scaling properties of phase change materials. In *Non-Volatile Memory Technology Symposium, 2007. NVMTS '07 DOI -*, pages 30–35, 2007.

[94] Simone Raoux. *Phase Change Materials Science and Applications.* Springer Science and Business Media LLC, New York, USA, 2009.

[95] Simone Raoux, Jean L. Jordan-Sweet, and Andrew J. Kellock. Crystallization properties of ultrathin phase change films. *J. Appl. Phys.*, 103(11):114310, June 2008.

[96] Simone Raoux, Robert M. Shelby, Jean Jordan-Sweet, Becky Munoz, Martin Salinga, Yi-Chou Chen, Yen-Hao Shih, Erh-Kun Lai, and Ming-Hsiu Lee. Phase change materials and their application to random access memory technology. *Microelectronic Engineering*, 85(12):2330–2333, December 2008.

[97] M P Lopez Sancho, J M Lopez Sancho, and J Rubio. Highly convergent schemes for the calculation of bulk and surface green functions. *Journal of Physics F: Metal Physics*, 15(4):851–858, 1985.

[98] R. T. Sanderson. Interrelation of bond dissociation energies and contributing bond energies. *J. Am. Chem. Soc.*, 97(6):1367–1372, March 1975.

[99] G. Servalli. A 45nm generation phase change memory technology. In *Electron Devices Meeting (IEDM), 2009 IEEE International*, pages 1–4, 2009.

[100] R. Shaltaf, X. Gonze, M. Cardona, R. K. Kremer, and G. Siegle. Lattice dynamics and specific heat of alpha -gete : Theoretical and experimental study. *Phys. Rev. B*, 79(7):075204, February 2009.

[101] Jean L. Jordan-Sweet Vaughn R. Deline Jan Boris Philipp Simone Raoux, Charles T. Rettner and Hsiang-Lan Lung. Scaling properties of phase change nanostructures and thin films. *Proceedings of the European Symposium on Phase Change and Ovonic Science, Grenoble, France, May 2006*, page 127134.

[102] R. E. Simpson, M. Krbal, P. Fons, A. V. Kolobov, J. Tominaga, T. Uruga, and H. Tanida. Toward the ultimate limit of phase change in ge2sb2te5. *Nano Letters*, 10(2):414–419, February 2010.

[103] Stallman. *GNU Make*. Free Software Foundation, 2010.

[104] M. Strange, I. S. Kristensen, K. S. Thygesen, and K. W. Jacobsen. Benchmark density functional theory calculations for nanoscale conductance. *J. Chem. Phys.*, 128(11):114714–8, March 2008.

[105] Xuhui Sun, Bin Yu, and M. Meyyappan. Synthesis and nanoscale thermal encoding of phase-change nanowires. *Appl. Phys. Lett.*, 90(18):183116–3, April 2007.

[106] Xuhui Sun, Bin Yu, G. Ng, M. Meyyappan, Sanghyun Ju, and D.B. Janes. Germanium antimonide phase-change nanowires for memory applications. *Electron Devices, IEEE Transactions on DOI - 10.1109/TED.2008.2005160*, 55(11):3131–3135, 2008.

[107] Xuhui Sun, Bin Yu, Garrick Ng, and M. Meyyappan. One-dimensional phase-change nanostructure: Germanium telluride nanowire. *The Journal of Physical Chemistry C*, 111(6):2421–2425, February 2007.

[108] Xuhui Sun, Bin Yu, Garrick Ng, Thuc Dinh Nguyen, and M. Meyyappan. Iii-vi compound semiconductor indium selenide (in[sub 2]se[sub 3]) nanowires: Synthesis and characterization. *Applied Physics Letters*, 89(23):233121, 2006.

[109] Zhimei Sun, Jian Zhou, and Rajeev Ahuja. Unique melting behavior in phase-change materials for rewritable data storage. *Phys. Rev. Lett.*, 98(5):055505, February 2007.

[110] Zhimei Sun, Jian Zhou, Andreas Blomqvist, Borje Johansson, and Rajeev Ahuja. Fast crystallization of chalcogenide glass for rewritable memories. *Appl. Phys. Lett.*, 93(6):061913–3, August 2008.

[111] Zhimei Sun, Jian Zhou, Andreas Blomqvist, Börje Johansson, and Rajeev Ahuja. Formation of large voids in the amorphous phase-change memory ge2sb2te5 alloy. *Phys. Rev. Lett.*, 102(7):075504, February 2009.

[112] Sergei F. Vyboishchikov, Pedro Salvador, and Miquel Duran. Density functional energy decomposition into one- and two-atom contributions. *J. Chem. Phys.*, 122(24):244110–13, June 2005.

[113] D. S. Chao-Y. C. Chen F. Chen C. W. Chen-R. Yen M. J. Chen-W. H. Wang T. C. Hsiao J. T. Yeh S. H. Chiou M. Y. Liu T. C. Wang L. L. Chein C. Huang N. T. Shih L. S. Tu D. Huang T. H. Yu M. J. Kao W. S. Chen, C. Lee and B M.-J. Tsai. A novel cross-spacer phase change memory with ultra-small lithography independent contact area. page 319322, 2007.

[114] J.-S. Wang, J. Wang, and J. T. L. Quantum thermal transport in nanostructures. *The European Physical Journal B - Condensed Matter and Complex Systems*, 62(4):381–404, 2008-04-01.

[115] Weijie Wang, Desmond Loke, Luping Shi, Rong Zhao, Hongxin Yang, Leong-Tat Law, Lung-Tat Ng, Kian-Guan Lim, Yee-Chia Yeo, Tow-Chong Chong, and Andrea L. Lacaita. Enabling universal memory by overcoming the contradictory speed and stability nature of phase-change materials. *Sci. Rep.*, 2:1–6, April 2012.

[116] A. T. Waterman. Positive ionisation of certain hot salts, together with some observations on the electrical properties of molybdenite at high temperatures. *Phil. Mag.*, 33:225, 1917.

[117] A. T. Waterman. The electrical conductivity of molydenite. *Phys. Rev.*, 21:540–549, 1923.

[118] R. Clint Whaley and Antoine Petitet. Minimizing development and maintenance costs in supporting persistently optimized BLAS. *Software: Practice and Experience*, 35(2):101–121, February 2005. `http://www.cs.utsa.edu/~whaley/papers/spercw04.ps`.

[119] H.-S.P. Wong, SangBum Kim, Byoungil Lee, M.A. Caldwell, Jiale Liang, Yi Wu, R.G.D. Jeyasingh, and Shimeng Yu. Recent progress of phase change memory (pcm) and resistive switching random access memory (rram). In *Solid-State and Integrated Circuit Technology (ICSICT), 2010 10th IEEE International Conference on DOI - 10.1109/ICSICT.2010.5667542*, pages 1055–1060, 2010.

[120] H.P. Wong, S. Raoux, S. Kim, J. Liang, J.P. Reifenberg, B. Rajendran, M. Asheghi, and K.E. Goodson. Phase change memory. *Proceedings of the IEEE DOI - 10.1109/JPROC.2010.2070050*, 98(12):2201–2227, 2010.

[121] J.Y. Wu, M. Breitwisch, S. Kim, T.H. Hsu, R. Cheek, P.Y. Du, J. Li, E.K. Lai, Y. Zhu, T.Y. Wang, H.Y. Cheng, A. Schrott, E.A. Joseph, R. Dasaka, S. Raoux, M.H. Lee, H.L. Lung, and C. Lam. A low power phase change memory using thermally confined tan/tin bottom electrode. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, pages 3.2.1–3.2.4, 2011.

[122] Feng Xiong, Albert Liao, David Estrada, and Eric Pop. Low-power switching of phase-change materials with carbon nanotube electrodes. *Science*, 332:568–570.

[123] Cheng Xu, Zhitang Song, Bo Liu, Songlin Feng, and Bomy Chen. Lower current operation of phase change memory cell with a thin tio2 layer. *Applied Physics Letters*, 92(6):062103, 2008.

[124] H. Horii-J. H. Park S. H. Joo S. O. Park U.-In. Chung Y. H. Ha, J. H. Yi and J. T. Moon. An edge contact type cell for phase change ram featuring very low power consumption. *Proc. Symp. Very Large Scale Integr. (VLSI) Technol.*, page 175176, 2003.

[125] Joanne W. L. Yim, Bin Xiang, and Junqiao Wu. Sublimation of gete nanowires and evidence of its size effect studied by in situ tem. *Journal of the American Chemical Society*, 131(40):14526–14530, October 2009.

[126] Youngki Yoon, Dmitri E. Nikonov, and Sayeef Salahuddin. Role of phonon scattering in graphene nanoribbon transistors: Nonequilibrium green's function method with real space approach. *Appl. Phys. Lett.*, 98(20):203503–3, May 2011.

[127] Bin Yu, Sanghyun Ju, Xuhui Sun, Garrick Ng, Thuc Dinh Nguyen, M. Meyyappan, and David B. Janes. Indium selenide nanowire phase-change memory. *Appl. Phys. Lett.*, 91(13):133119–3, September 2007.

[128] Bin Yu, Xuhui Sun, Sanghyun Ju, D.B. Janes, and M. Meyyappan. Chalcogenide-nanowire-based phase change memory. *Nanotechnology, IEEE Transactions on DOI - 10.1109/TNANO.2008.926374*, 7(4):496–502, 2008.

[129] Dong Yu, Sarah Brittman, Jin Seok Lee, Abram L. Falk, and Hongkun Park. Minimum voltage for threshold switching in nanoscale phase-change memory. *Nano Letters*, 8(10):3429–3433, October 2008.

[130] Dong Yu, Junqiao Wu, Qian Gu, and Hongkun Park. Germanium telluride nanowires and nanohelices with memory-switching behavior. *Journal of the American Chemical Society*, 128(25):8148–8149, June 2006.