COLLECTOR CURRENT TRANSPORT MECHANISMS IN SIGE HBTS OPERATING AT CRYOGENIC TEMPERATURES

A Dissertation
Presented to
The Academic Faculty

By

Hanbin (Victor) Ying

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Electrical and Computer Engineering

Georgia Institute of Technology

May 2019

COLLECTOR CURRENT TRANSPORT MECHANISMS IN SIGE HBTS OPERATING AT CRYOGENIC TEMPERATURES

Approved by:

Dr. John D Cressler, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Asif Islam Khan
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Dragomir Davidovic
School of Physics
Georgia Institute of Technology

Date Approved: April 12, 2019

This thesis is dedicated to my parents, who always encourage me to pursue my dreams, and to my fiancee, Qianqi, who is always supportive when I'm faced with difficulties.

ACKNOWLEDGEMENTS

Along the journey of this research, there are so many people that gave me big helps. Without them, this work will not be possible. Although there is not enough room to properly thank every single one of them, I would still like to include a few of such people. First, I would like to thank my family. They are always there, giving me unconditional love, embracing my shortcomings, and gently leading me out of my lowest moment. Then I would like to thank my advisor, Dr. Cressler, whose decision to let me in the team in 2015 completely changed my life trajectory. Apart from technical knowledge, his wisdom really brings out my potential not only in research, but in all facets of life. I would also like to thank Dr. Khan and Dr. Davidovic, who are willing to serve in my committee and answer my infinite number of questions. I would like to thank Dr. Yoder, who spent quite a long time answering my questions. I want to express my gratitude to my colleagues including Dr. Mourigal and Luwei (kind enough to share the lab with me and teach me how to use the PPMS measurement system), Jason (who makes the tedious packaging time a lot more fun), Brian, Uppili, Anup, Nelson, Adrian, and Zach, all of whom spent a tremendous amount of time teaching me device physics and circuit theory, checking grammars in my writing, debugging TCAD or Cadence problems, and answering my questions. I would also like to thank everyone else in the SiGe Device and Circuit Team. It is simply a wonder how much I learned from these people. I'm also indebted to the administrative staff at TSRB for their support, including but not limited to Scott, Carolyn, and Alison, without whom I will not be able to purchase the right tools and carry out the research plans in a timely manner.

I would like to formally thank people at Sandia Corporation, including T. England, M. Holmes, and S. Carr. With their financial sponsorship, I am able to push through the boundary of science and engineering. Last but not least, I would like to thank A. Joseph, V. Jain, N. Cahoun, T. Lamouthe, D. Harame and the GlobalFoundries SiGe team for their support on wafer fabrications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Silicon germanium heterojunction bipolar transistors (SiGe HBTs) have recently gained attention due to their potential for use in quantum computing readout circuits. To serve such applications, which requires low noise, low to medium speed, and low power, it is crucial to understand the transport physics of SiGe HBTs at cryogenic temperatures. This thesis aims to extend the existing transport theories for collector current through theoretical analysis, experimental data, and TCAD simulation. A novel transport mechanism, namely the direct tunneling mechanism, is found to account for a portion of collector current at cryogenic temperatures. A method is proposed to differentiate between direct tunneling and quasi-ballistic transport. With the understanding of collector current and its physics, the impact of technology scaling on future SiGe HBTs are evaluated. Portions of this work have been published in [1, 2, 3].

CHAPTER 1
INTRODUCTION AND BACKGROUND

1.1 Background

SiGe HBTs have long been recognized as a viable candidate for cryogenic temperature applications [4, 5, 6, 7]. When cooled, SiGe HBTs naturally exhibit improved frequency response (f_T and f_{MAX}, as shown in Fig. 1.1), current gain, transconductance (shown in Fig. 1.2), noise, bandwidth, and output conductance[8, 9]. Additionally, the compatibility with CMOS technology enables highly-integrated solutions to satisfy a wide variety of applications [10, 11, 12, 13, 14, 15].

An emerging field with interesting application opportunities for cryogenically-operated SiGe HBTs is readout pre-amplification in quantum computing. Traditionally, qubit (quantum bit) is placed at deep cryogenic temperatures (< 4 K) to maintain coherence, and the information of qubit is read out by quantum amplifiers such as superconducting quantum interference devices (SQUIDs) or single electron transistors (SETs). This information is sent through cables to room temperature electronics for the amplification and processing [17, 18]. Due to the low signal-to-noise ratio (SNR), low bandwidth, and size inefficiency associated with this "indirect" readout method, there are growing efforts to amplify the output from quantum amplifiers directly at cryogenic temperatures with conventional electronics before sending it to room temperature[1] [19]. These kinds of electronics are frequently referred to as pre-amplifiers or interface circuits. Previously, CMOS devices, HEMTs, and MESFETs have successfully demonstrated for such cryogenic preamplification tasks [20, 21, 22, 23, 24]. SiGe HBTs, however, are largely absent from this picture until the recent

[1] IBM used such scheme in its quantum computer. The infographic showed a pre-amplifier at 4 K stage, followed by another amplifying stage at room temperature (source: https://www.research.ibm.com/ibm-q/learn/what-is-quantum-computing/).

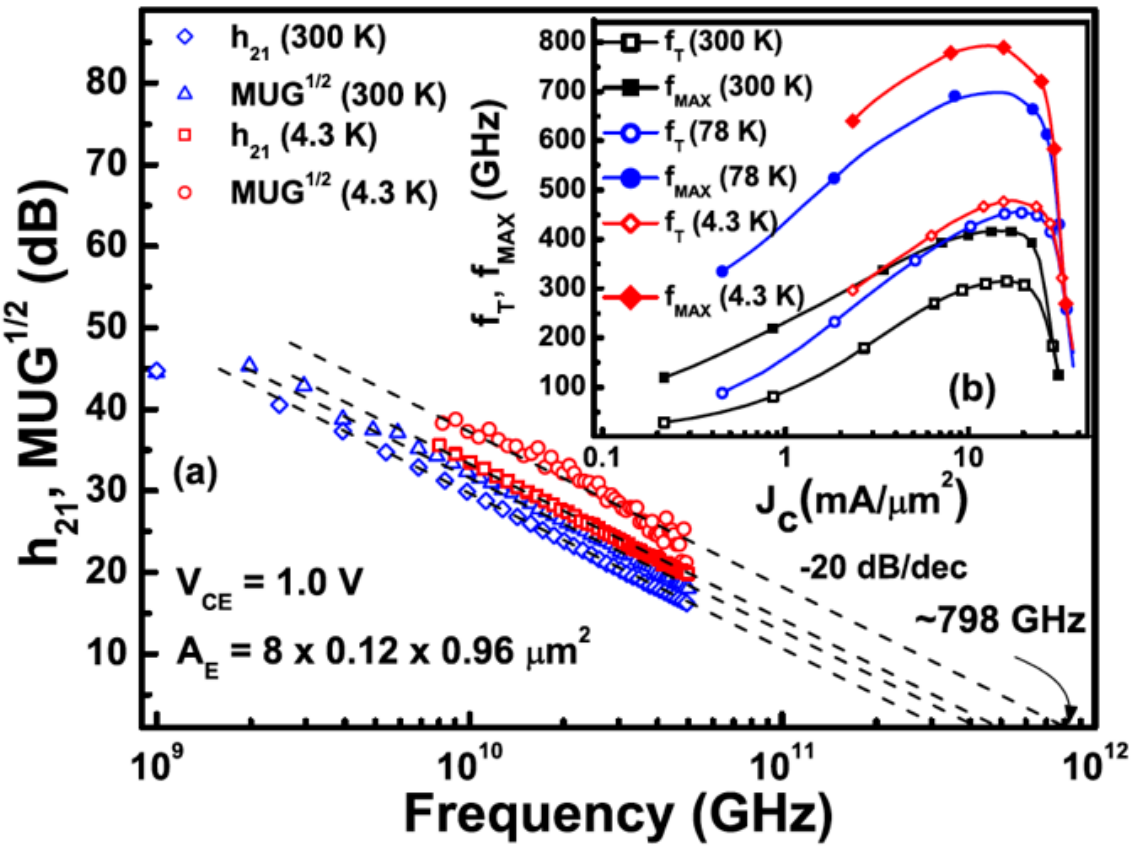

Figure 1.1: The h_{21} parameter of SiGe HBTs across frequency at 300 K and 4.3 K. Inset: f_T/f_{MAX} of SiGe HBTs at 300 K, 78 K, and 4.3 K (after [16]).

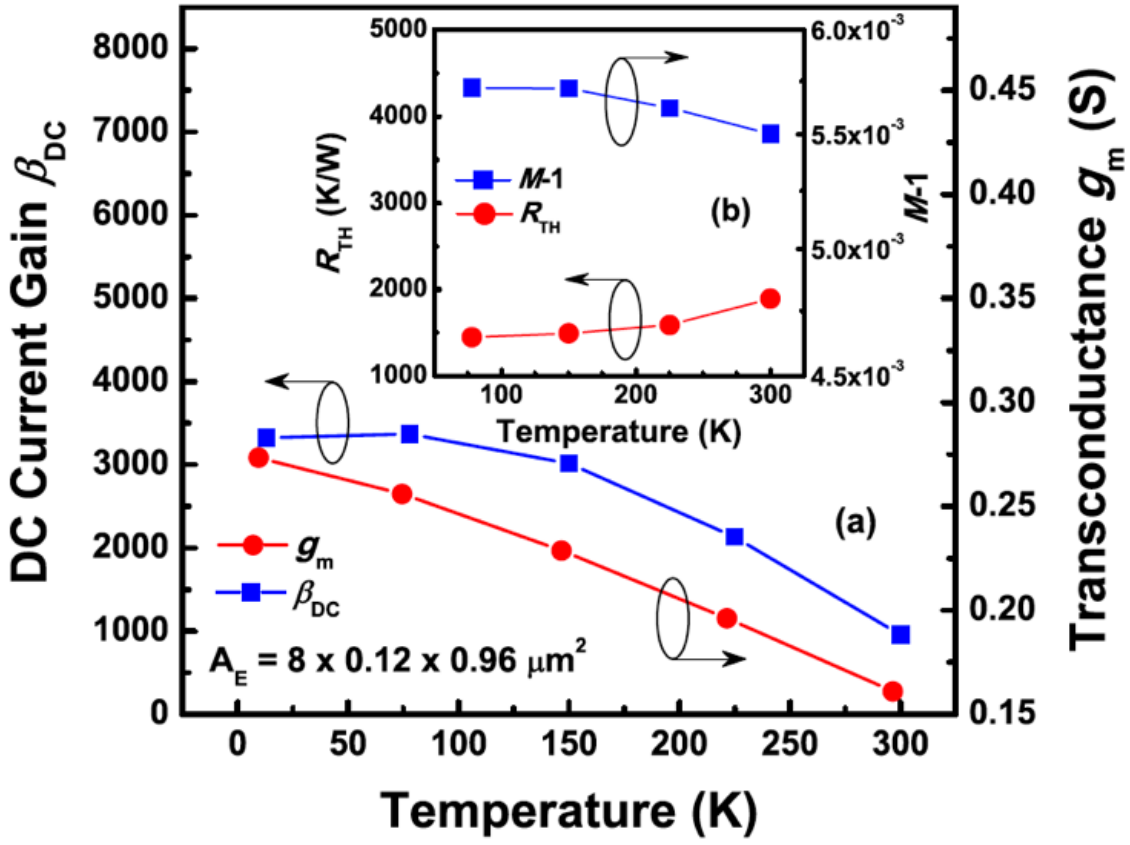

Figure 1.2: The DC current gain (β) and transconductance (g_m) of SiGe HBTs across temperatures. Inset: Thermal resistance and avalanche multiplication (M-1) of SiGe HBTs across temperatures (after [16]).

demonstration [18]. Compared with CMOS devices, SiGe HBTs have lower noise (1/f and broadband), higher gain, and higher transconductance (g_m), which is clearly desired for interface circuit.

Apart from these, SiGe HBTs also provide the scalability needed for quantum computing in an inexpensive way compared with III-V compound semiconductor devices. This is especially the case since the push[2] for large-scale quantum computers has been recently boosted by the advent of quantum algorithms and, in particular, the invention of quantum error correction codes because such methods require a large number of physical qubits [25, 26]. In a large-scale quantum computer, if the information of each qubit is sent out through a physical cable, thousands of wires (DC and AC) are fed out of the refrigerator, which clearly is not feasible [19]. To solve this problem, integrated circuits can be employed to amplifier the signals, combine the data streams, and implement local control and qubit initializations at cryogenic temperatures, dramatically reducing the number of communication lines needed between the refrigerator chambers and the room temperature instrumentations. Among various common choices of integrated circuits, silicon-based technology such as BiCMOS technology (SiGe HBTs and CMOS on the same chip) is clearly a strong candidate with its high yield and relatively low cost. In an integrated communication link composed from BiCMOS technology, SiGe HBTs can provide the low-noise front end for readout amplification, while the CMOS circuits provide the compact form of data processing. Therefore, it is meaningful to qualify SiGe HBTs at deep cryogenic temperatures for such potential applications.

1.2 Methodology

In the present work, multiple technology generations of SiGe HBTs were characterized at cryogenic temperatures. All SiGe HBTs were obtained from commercial product lines of GlobalFoundries through MOSIS multi-project wafers (MPW). There are two reasons

[2]Such push has already seen results. IBM drew attention with its 50-qubit quantum computer in 2017. Google announced the 72-qubit quantum computers in 2018

to study multiple commercial technology generations instead of any particular batch of devices or in-house fabrications.

First, the mature SiGe technology means that almost all SiGe HBTs available today come from commercial semiconductor foundries. Therefore, the details of profiles and device structures are always classified information and unavailable for the public. However, from literatures (including those from foundries themselves), the scaling rules for technology are known [27]. Such scaling rules have been observed from the early generation (0.5 μm node) to the advanced generation (90 nm node) [28]. By studying the change in device characteristics across multiple generations and assuming a known scaling rule, we can obtain correlations between doping profiles, device structures, and transport physics with less uncertainty. On the other hand, SiGe technology has so far been driven by the high frequency applications at room temperature, such as automative long-range radar, Tera-Hz imaging, and high speed wireline applications [29]. These applications are most likely following a different scaling rule than what the cryogenic amplification demands [14, 16]. Therefore, it is meaningful to evaluate how the existing scaling rules affects the operation of future SiGe HBTs at cryogenic temperatures.

Second, early technology generations of SiGe HBTs exhibited non-ideal characteristics at cryogenic temperatures. In particular, quasi-ballistic (non-equilibrium) transport and trap-assisted tunneling mechanisms appear as temperature decreases [30, 11]. Despite the same underlying operating principles, the present technologies are in many shapes and forms different from the early generations of devices. A simple comparison of f_{MAX} reveals the drastic difference between devices used in [18], which have a f_{MAX} of 110 GHz, and those in [31], which have a f_{MAX} as high as 720 GHz, both at 300 K. Since the market is dominated by the new technology generations nowadays, which have not seen much characterization at extreme temperatures, it is of interest to see if the previously observed non-ideality still exists in the advanced technology generations. A comparison between generations also yields useful information on whether the effect of any undesirable non-

ideality has been enhanced or diminished with the technology scaling.

In the present work, the measurement of one batch of devices down to 70 mK reveal a novel non-ideality in transistor terminal currents. This gives evidence to support the presence of a new transport mechanism, namely, the direct tunneling mechanism. In light of this newly discovered mechanism, the present work proposes a simple method to differentiate direct tunneling mechanisms from other transport mechanisms based on simple DC measurements. The measurement of various technology generations also enables the empirical observation of how technology scaling impacts the characteristics of SiGe HBTs at cryogenic temperatures. In addition, SiGe HBTs of multiple generations were simulated in TCAD at cryogenic temperatures in order to examine the sensitivity of transport mechanisms to process parameters and to provide insight for future profile designs. The transport mechanisms for collector current will be the major focus, due to the available literatures on cryogenic collector transport mechanisms. The transport mechanisms operative in the base current, though clearly important, will be investigated in a future work.

The remaining part of this thesis is organized into five chapters. Chap. 2 explains the measurement setup and device technologies used in this thesis. Chap. 3 summarizes the existing transport theories and presents a qualitative picture of direct tunneling mechanism in the context of other known mechanisms. Chap. 4 presents the measurement data at 70 mK temperature together with the theoretical analysis that shows the plausibility of direct tunneling mechanism for collector current. Chap. 5 studies the impact of technology scaling on collector transport based on empirical observations of experimental data and calibrated TCAD simulations. Lastly, Chap. 6 concludes the thesis.

CHAPTER 2

MEASUREMENT SETUP AND DEVICE TECHNOLOGY

2.1 Measurement Setup

Cryogenic measurements were made using a Quantum Design Physical Property Measurement System (PPMS) DynaCool® system. It includes a cryogenic refrigerator capable of cooling the base (chamber) down to 1.8 K. A superconducting magnetic is placed in the base and is capable of generating a static magnetic field of $\pm$14 T. A dilution refrigerator (DR) module with a DR insert (which will be inserted into the base) can provide the additional cooling capability to reach a temperature of 50 mK at the tip of the DR insert. The system is shown in Fig. 2.1.

The PPMS offers two configurations for measurement. The first configuration is with a DR insert, which achieves a minimum sample temperature of 50 mK while allowing a maximum of 25 μW power dissipation. This configuration is shown in Fig. 2.2(a). The second configuration is without a dilution refrigerator insert. In other words, the sample is placed directly inside the base of the chamber. This achieves a minimum temperature of 1.8 K but allows a maximum heat dissipation of 5 mW. Both configurations can achieve any temperature between room temperature and the lowest achievable temperature, by controling a heater in the base chamber with a PID controller.

Test samples were die-attached on custom gold packages (also called "pucks") using indium solder to ensure good thermal conductivity, as shown in Fig. 2.2(b). Electrical connections between the die and package were made by gold wirebonds of 25.4 μm diameter. The package is placed in the DR or in the base, and all electrical connections inside the PPMS system are provided via low-loss superconducting aluminum twisted pairs. No cryogenic filtering was applied. Outside the system, the connections were adapted to tri-

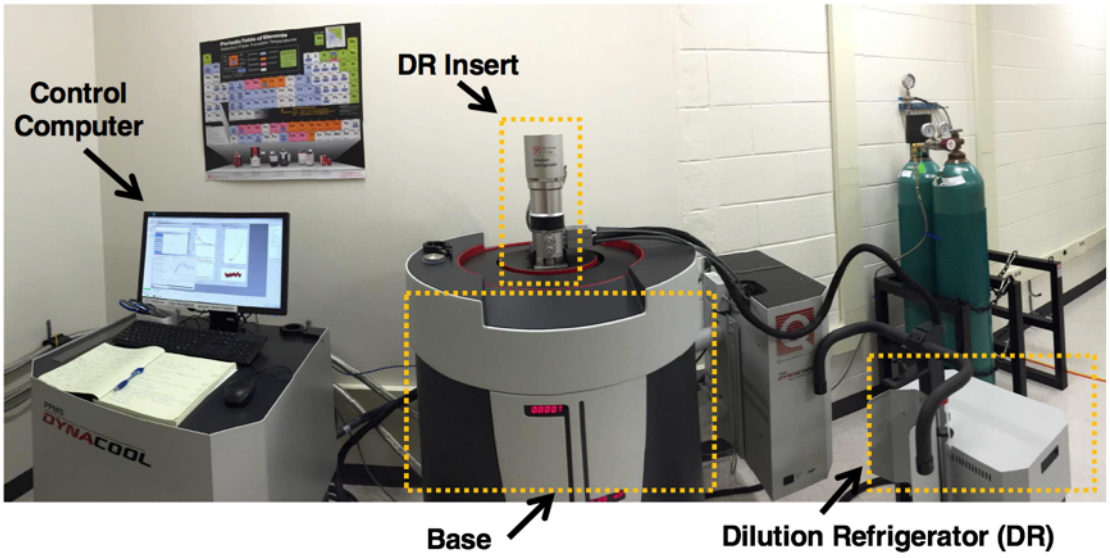

Figure 2.1: The first configuration of the PPMS system, showing the base chamber, dilution refrigerator (DR) insert, dilution refrigerator module, and the control computer.

axial cabling to minimize residual noise and fed into the Agilent 4156C Semiconductor Parameter Analyzer for characterizations. The exterior and interior design of the breakout box is shown in Fig. 2.2(c) and Fig. 2.2(d). During operation, the samples (either in the DR insert or in the base) were placed in the bore of a superconducting cryomagnet in the base that is capable of generating $\pm$14 T of magnetic field.

The package temperature was constantly monitored to ensure minimal temperature fluctuations across all measurements. One limitation in the first configuration of measurement (with the DR insert) is the limited cooling power (0.25 μW at 50 mK). All measurements at 4 K and below were limited to a maximum collector current of 1 μA in order to maintain a stable DUT temperature. To obtain the higher current, a quick DC sweep (with small integration time setting in the analyzer) was performed up to 0.1 mA collector current. This caused a transient temperature rise up to 50 mK above ambient (the ambient temperature was noticed to begin rapidly changing at currents $>$ 1 μA). Higher current sweeps were avoided in the first configuration (i.e., when DR insert is used) because the rapid heating can potentially damage the measurement system. If the device is measured in the second configuration, no such limitation exist. As will be shown in this thesis, the DC operation of

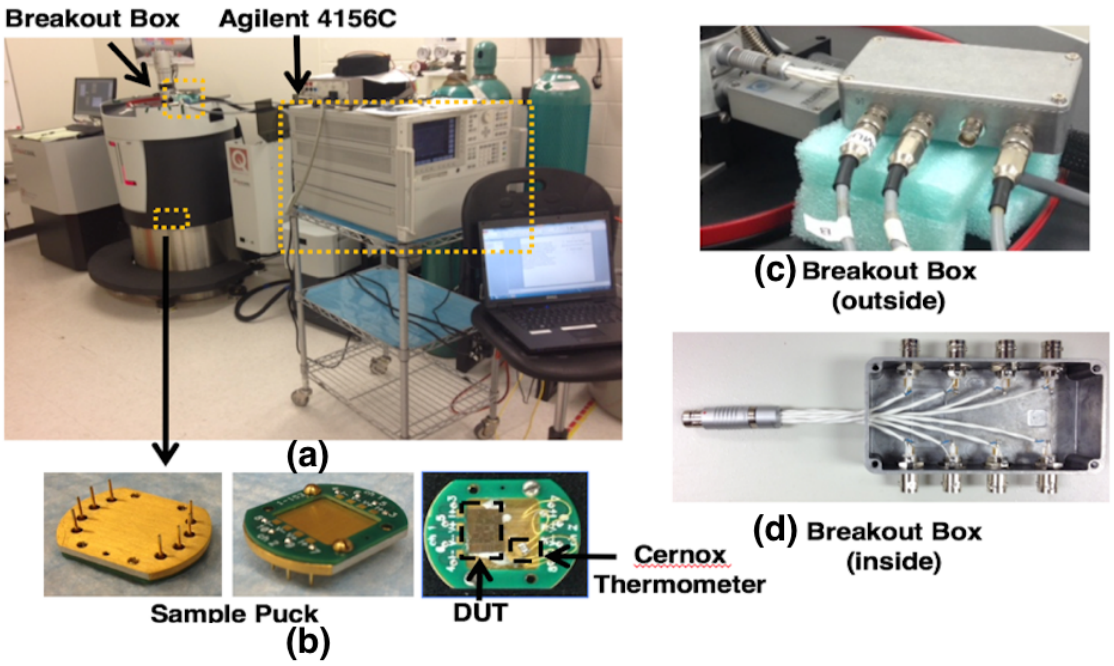

Figure 2.2: Details of the experimental setup. (a) The measurement setup of PPMS system together with 4156C, showing the location of the sample puck and the breakout box. (b) The empty sample puck (first two figures) and the puck with a sample DUT and a Cernox thermometer. (c) The connector (Triaxial) at the exterior of the breakout box. (d) The interior look of the breakout box that converts the custom Fisher connectors to standard Triaxial connectors using twisted pairs.

SiGe HBTs at 16 K and below is mostly identical, except in a handful of bias range. Therefore, after the initial characterization at mK temperatures, most subsequent measurements were obtained with the second configuration of PPMS system (with higher cooling power), and long integration time was used to obtain high resolution data.

2.2 Device Technology

The devices investigated in this work are from GlobalFoundries BiCMOS 5AM[1] (0.5 μm), BiCMOS 8HP (130 nm), and BiCMOS 9HP (90 nm) technologies, and are referred to as the first [32], third [33], and fourth [28] generation devices. The categorization of generation is mainly based on the peak f_T/f_{MAX} value of each technology. As shown in Fig. 2.3, the high-performance (HP) devices from the first, the third, and the fourth generations have peak f_T/f_{MAX} values of 50/90, 210/270, and 300/350 GHz, respectively, at 300 K. All devices presented are high-performance (HP) device variants from each generation, meaning that their collector profiles were optimized for maximum speed by increasing the doping of selectively implanted collector (SIC). For simplicity, only NPN SiGe HBTs transistors will be discussed because all processes investigated here only include high-performance NPN devices. The lateral PNP devices are of considerably lower performance in a typical BiCMOS process, unless it is specifically designed as a complementary platform. C-B-E-B-C layout structures are used, which means the devices have an emitter contact in the center of the device, followed by two base and two collector contacts on each side of the emitter contact. The emitter width is fixed by the process node, at 0.5 μm, 130 nm, and 90 nm for the first, third, and the fourth generation, respectively. The emitter length in the range of 0.76 - 10 μm were measured for each generation and the presented results were verified to be both reproducible and consistent across samples of multiple sizes.

[1]The process is officially called 5HP, though it is often referred to as 5AM because later versions of 5HP has the improved back end of the line (BEOL) dubbed "AM".

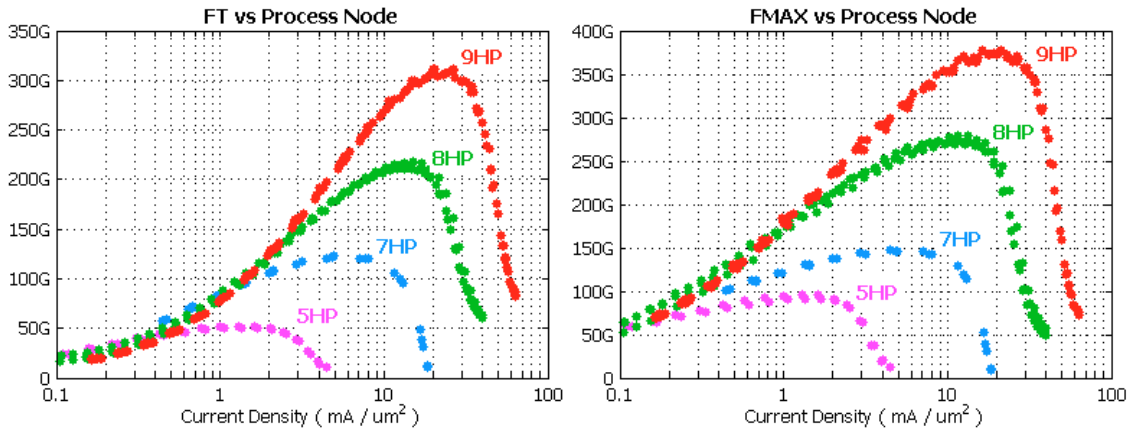

Figure 2.3: The f_T/f_{MAX} of SiGe HBTs from Globalfoundries 5HP, 7HP, 8HP, and 9HP, categorized as the first, second, third, and fourth generation of the technology, respectively (after [28]).

2.2.1 TCAD Cross Sections and TEMs

Although the exact doping profile and device structures are classified information, the foundries do reveal limited information on profiles and device structures in published literatures. Such device structure and profile are considered to be qualitative, though it does showcase key features of the device reasonably well. The cross section of the third generation device in Technology Computer-Aided Design (TCAD) simulation is shown in Fig. 2.4 [34, 35]. The pink portions are the bottom of the tungsten plug. The contacts for all terminals are ohmic contacts and there are no Schottky contact throughout the device, in all generations. The poly-silicon emitter (red in Fig. 2.4) is doped with arsenic (As) at a level close to solid solubility to reduce the emitter resistance. The small brown region between the poly-emitter pedestal (red) and the extrinsic base is the E-B spacer oxide. The intrinsic base (the blue region right between emitter and collector) is SiGe layer grown epitaxially on top of the silicon collector. The extrinsic base (the blue region connecting the intrinsic base to the base contact) is doped heavily with Boron (B) to reduce the base resistance. The SiGe layer has a varying Ge mole fraction and is doped in-situ with boron during with the epitaxial process. The collector is formed on top of the heavily doped sub-collector region to reduce the collector contact resistance. Both the collector and the sub-collector

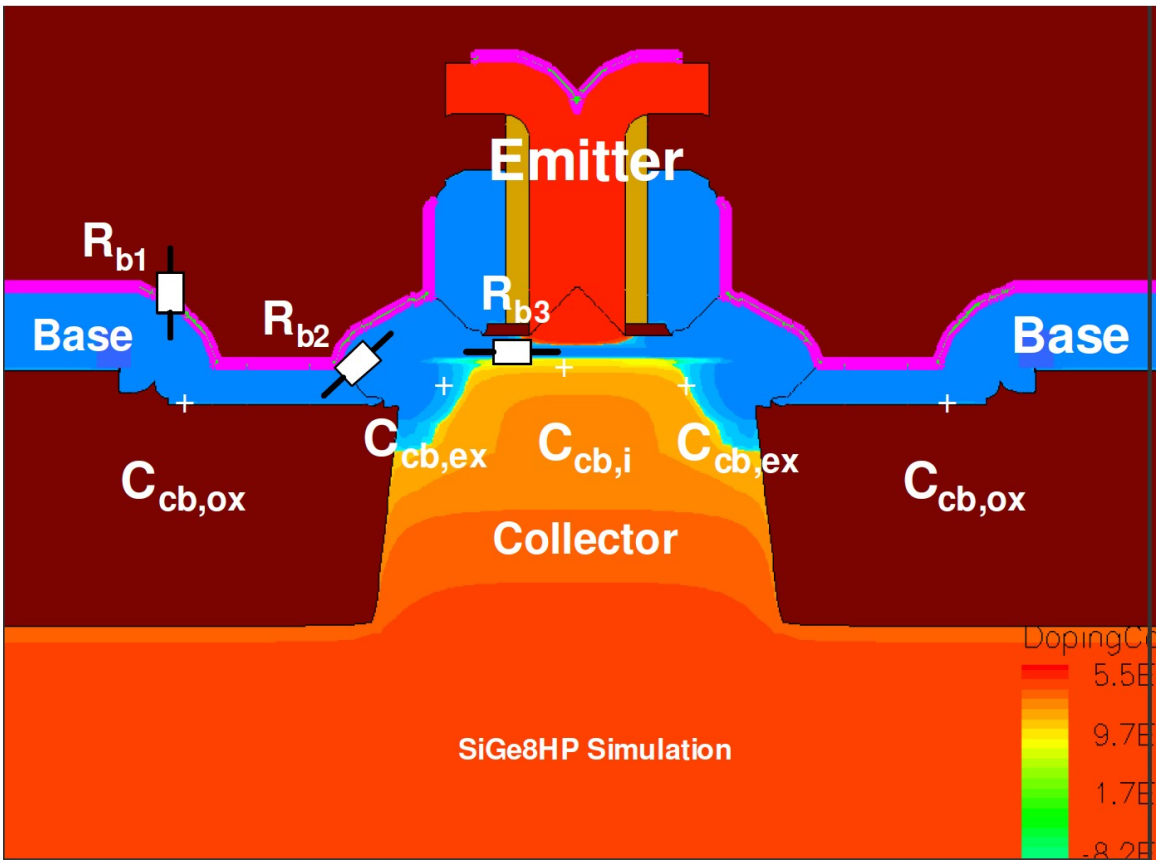

Figure 2.4: The cross section of a third generation SiGe HBT device in Sentaurus TCAD, showing the device structure and a qualitative view of the doping profile (after [34, 35]).

are doped with phosphorus (P). In high performance (HP) devices, a pocket of collector is doped (the so-called SIC) by ion implantation to a higher level than the background collector to suppress the Kirk effect. The collector opening is defined by the distance between the two shallow trench isolation (STI). A scanning electron micrograph of the device including the first metallization layer is shown in Fig. 2.5.

The cross section of the fourth generation device in TCAD simulation is shown in Fig. 2.6 [28]. The detailed doping profile is not revealed. However, compared to the third generation device, poly-silicon emitter doping stays the same (at solid solubility), while the base and collector doping increase. The emitter window (i.e., the spacing between the two E-B oxides) is decreased to reduce the length of the lateral linkage between the intrinsic and the extrinsic base, thereby decreasing base resistance. Similarly, collector window is

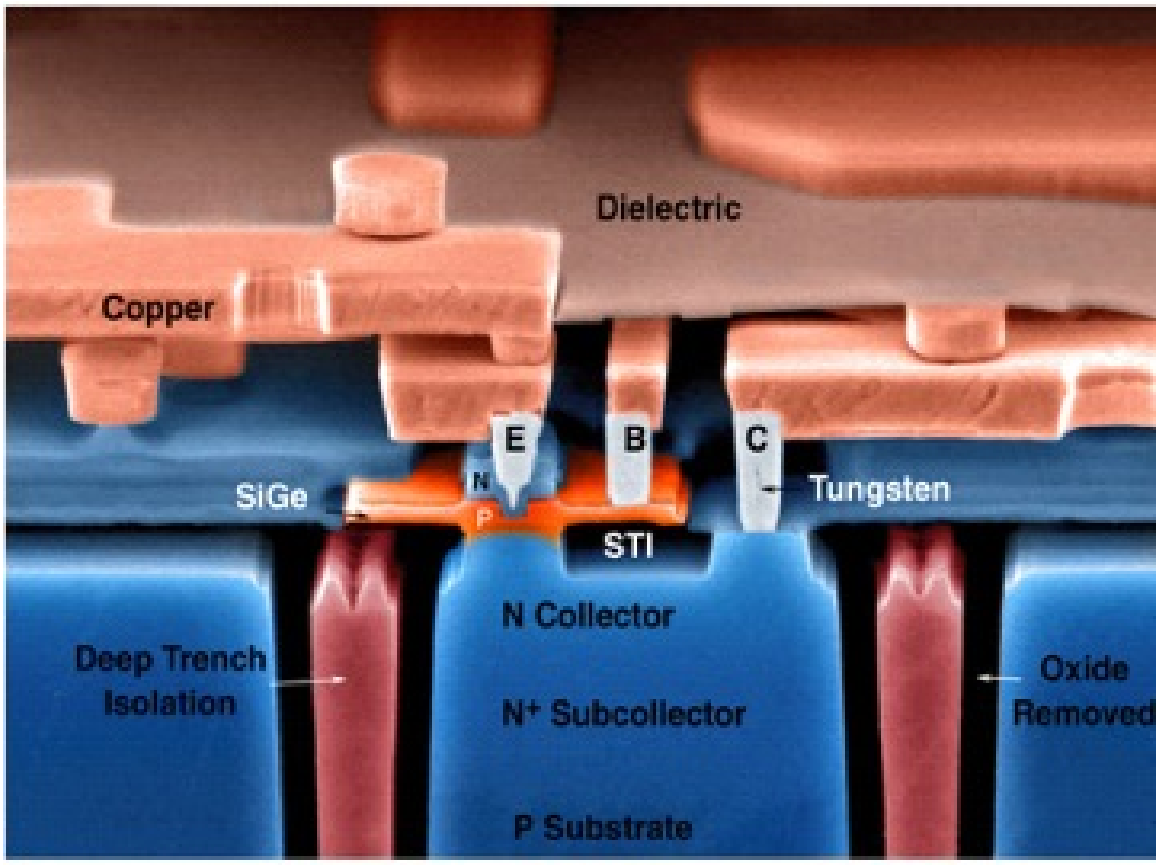

Figure 2.5: The scanning electron micrograph (SEM) of a third generation SiGe HBT device, showing the vertical stack from the substrate up to the first metal layer (after [2]).

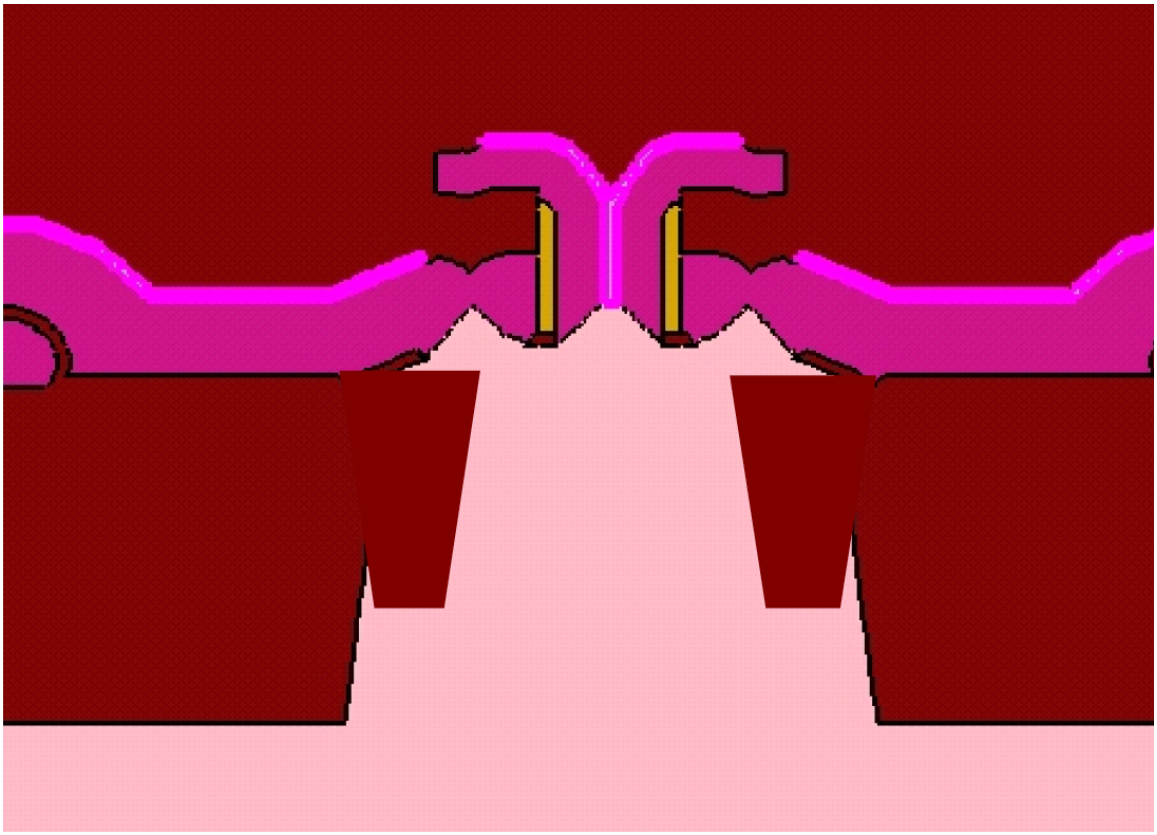

Figure 2.6: TCAD cross section of the fourth generation SiGe HBT (after [28]).

reduced by reducing the distance between the shallow trench isolation (STI), which reduces the capacitive coupling between the highly doped extrinsic base and the intrinsic collector and increases the peak speed. From the TEM picture of the B-E junction in the fourth generation device, as shown in Fig. 2.7, it can be seen that the actual device structure is similar to the published TCAD cross section. Therefore, it is assumed that the published TCAD cross section captures the essence of device structure reasonably well.

The Secondary Ion Mass Spectrometry (SIMS) profile of the first generation device is shown in Fig. 2.8. Compared to the more advanced generations, the base and collector doping in the first generation is lowered, and the base width and collector depth is wider. The peak Ge mole fraction is also lowered compared to close to 30% in the fourth generation. Despite the lowered doping, even the first generation of SiGe HBTs does not suffer much from the freeze-out effect. Freeze-out occurs when the lack of thermal energy at low

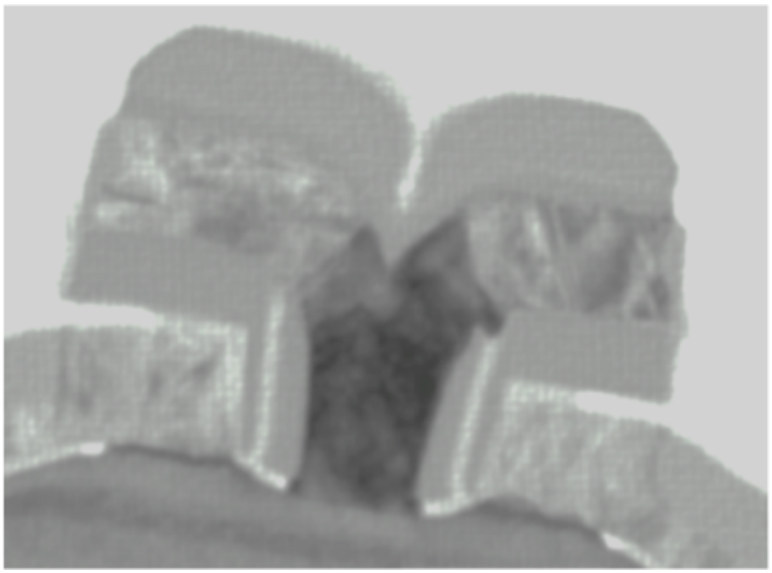

Figure 2.7: The TEM micrograph of a fourth generation SiGe HBT (after [28]).

temperatures cause carriers to be bound to their donor (or acceptor) atoms and therefore can not contribute to the electrical conduction. Experimentally, this is equivalent to having a large resistance in the terminal that has the freeze-out effect. When the doping density is sufficiently high (above Mott-transition threshold), the impurity band merges with the conduction band and the dopants are considered activated regardless of temperatures [36]. From the experimental data and theoretical analysis, the threshold doping level for Mott-transition is around $3\times10^{18}/cm^3$ for Si:P, $4\times10^{18}/cm^3$ for Si:As, and $1.7\times10^{18}/cm^3$ for Si:B, above which the dopants are considered to be activated across all temperature [37, 38]. This also explained the almost constant conductivity of highly doped Si samples across temperatures [39].

2.2.2 SiGe Band Structure

The base of the SiGe HBTs are graded SiGe alloy, where the Ge mole fraction is mostly below 30%, due to the stability limit of the thin-films. The graded alloy is engineered such that the final band structure of the intrinsic device does not contain any discontinuity. Since the SiGe base is pseudomorphically grown on the Si-100 surface of the n-type collector,

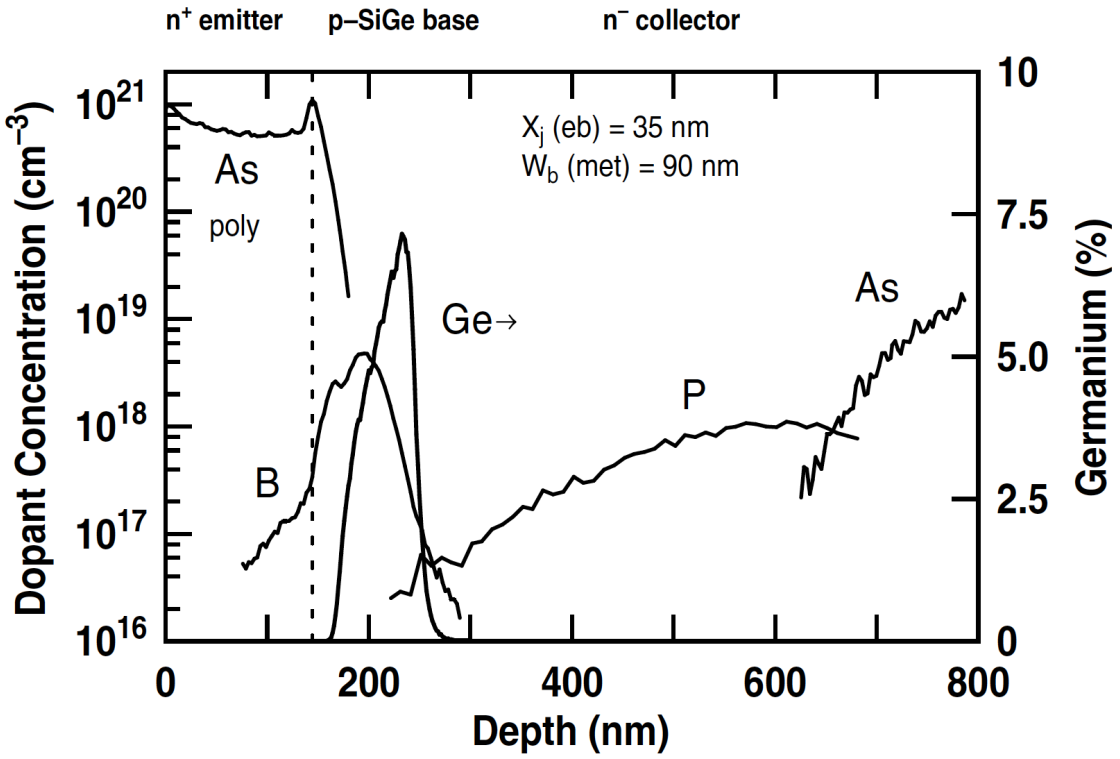

Figure 2.8: SIMS profile of a first generation SiGe HBT, showing the doping concentration of the respective species (left axis) and the Ge mole fraction (right axis) (after [14]).

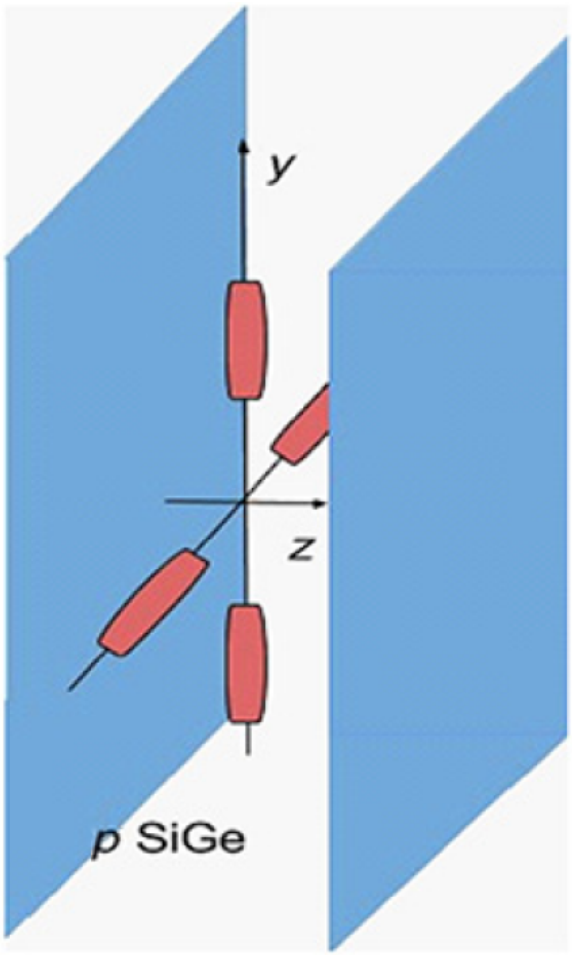

Figure 2.9: The illustration of four-fold valley degeneracy of SiGe alloy with respect to the direction of transport z. The 6-fold valley degeneracy is lifted due to the compressive strain.

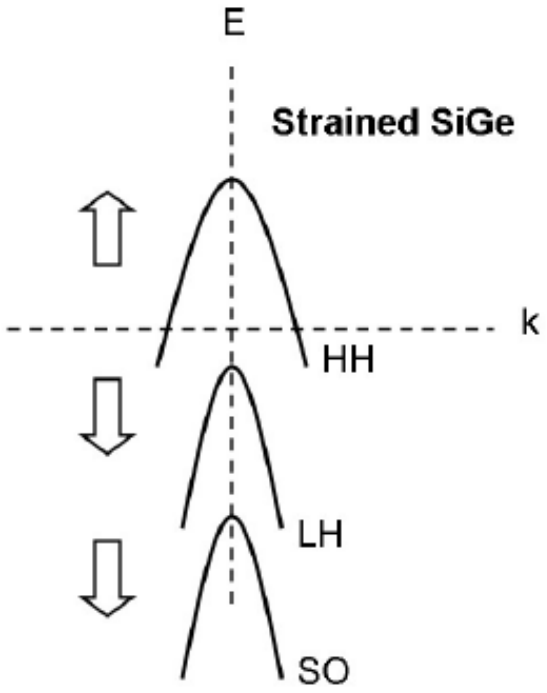

Figure 2.10: The effect of strain on the valence band of SiGe alloy (after [14]).

the sixfold valley degeneracy of pure Si is broken due to the compressive strain, and the bottom of the conduction band has a fourfold valley degeneracy, as shown in Fig. 2.9. Due to the compressive strain, the degeneracy of heavy hole, light hole, and split-off band also shifts correspondingly, as shown in Fig. 2.10. In particular, the top of the heavy hole band shifts closer to the conduction band, resulting in the bandgap energy reduction.

2.3 TCAD Models Used in This Work and Calibrations

2.3.1 TCAD Models

Since the aforementioned TCAD models from the foundry are unaccessible, similar TCAD models for the first, third, and the fourth generation were used in this thesis. The doping profile of the fourth generation device is shown in Fig. 2.11. The doping profiles for the first and the third generation device are qualitatively similar to that of the fourth generation, except the peak doping concentration in the base and collector and the peak Ge content is

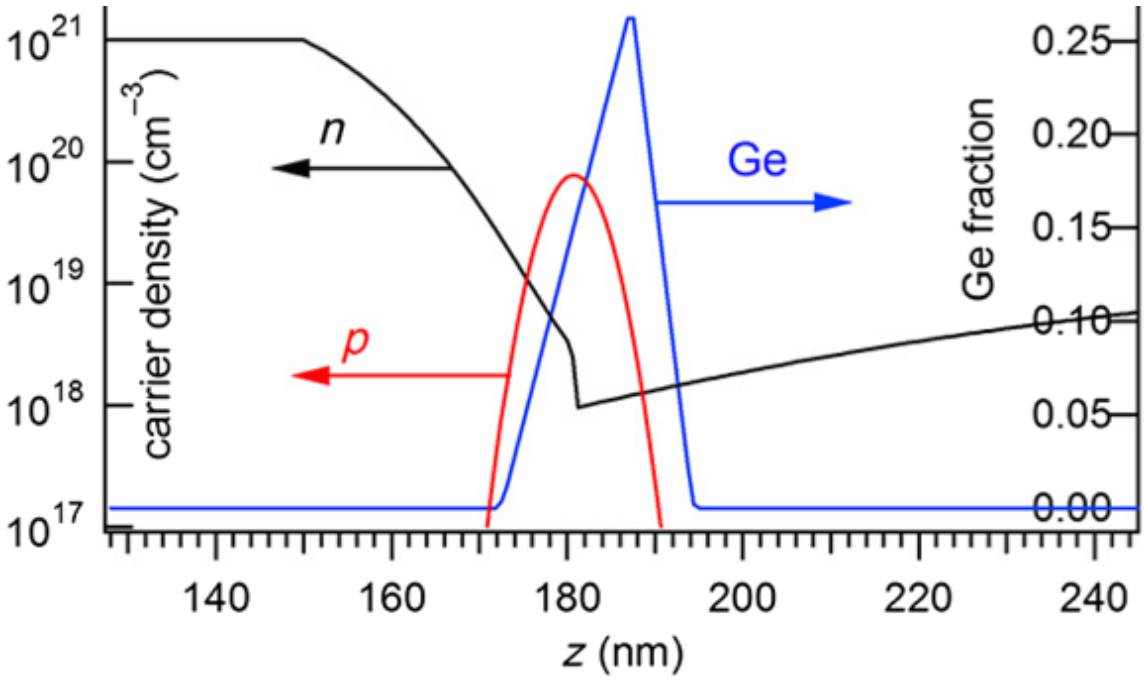

Figure 2.11: The net doping profile and the Ge mole fraction of the fourth generation SiGe HBT used in this thesis. The doping species are identical to those in Fig. 2.8.

lower, and the vertical distance for each region is correspondingly longer. More information on the simulation models will be given in Chap. 4 and Chap. 5 when the TCAD models are discussed.

2.3.2 Calibrations

The TCAD model was calibrated with the actual measurement data to make sure the profile is correctly modeling the actual device. Since the number of process parameters for calibration is large, a systematic approach was adopted to limit the possible range of profiles that match the actual device.

First, the lateral dimensions of the device is calibrated against the mask set in the process design kit (PDK). These include the emitter width, the distance between emitter and extrinsic base (the width of emitter-base oxide), the extrinsic base location and width, the collector reach-through location, the shallow trench isolation (STI) thickness, and the selectively implanted collector (SIC) opening, etc. Second, device parameters related to the

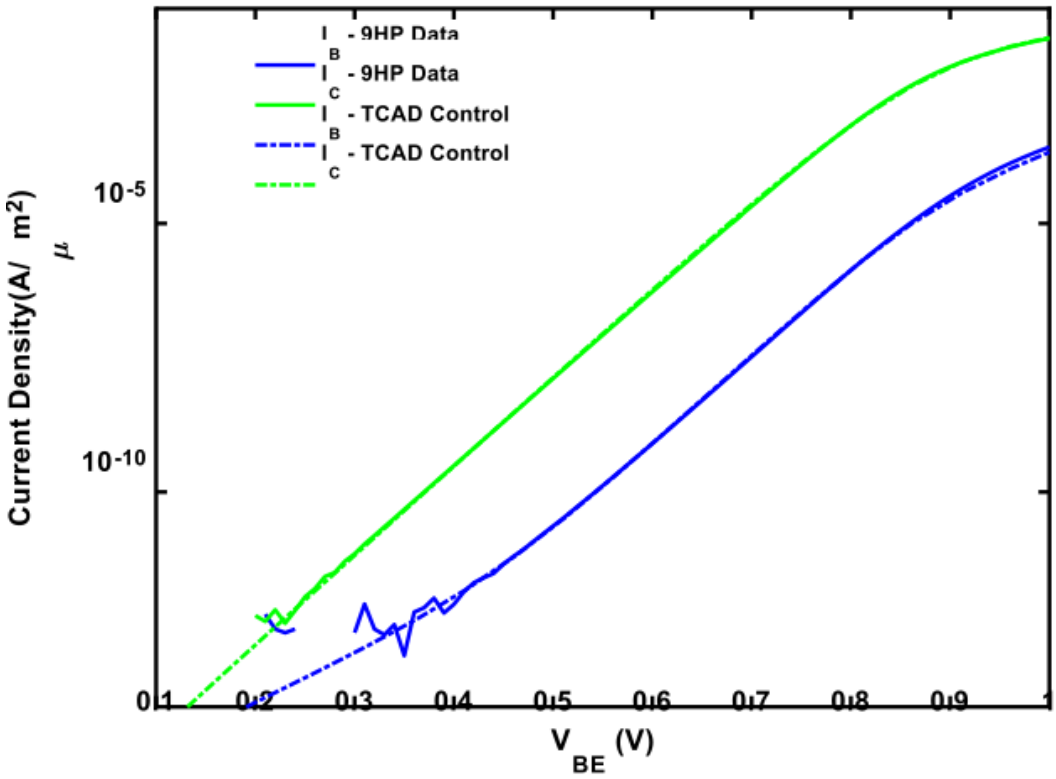

Figure 2.12: The calibration of Gummel characteristics of the profile.

vertical transport is calibrated, such as the emitter-base intercept Ge values, peak Ge percentage, emitter decay length, base peak doping and decay length, and SIC location and shape. The calibration is done by comparing TCAD simulation to the Gummel and f_T measurement data. Third, the vertical profile in the extrinsic region of the device is tuned to match the f_{Max} data. In all calibration steps, the decay length of each doping is evaluated to ensure that the profile is compatible with the processing technology limitations, such as the out diffusion from thermal rapid annealing (RTA). It is assumed with good faith that only a very limited set of profile will match the measurement data with all of the above constraints.

The Gummel calibration is shown in Fig. 2.12. The DC current gain, β, is calibrated in Fig. 2.13. The f_T/f_{Max} calibration is shown in Fig. 2.14.

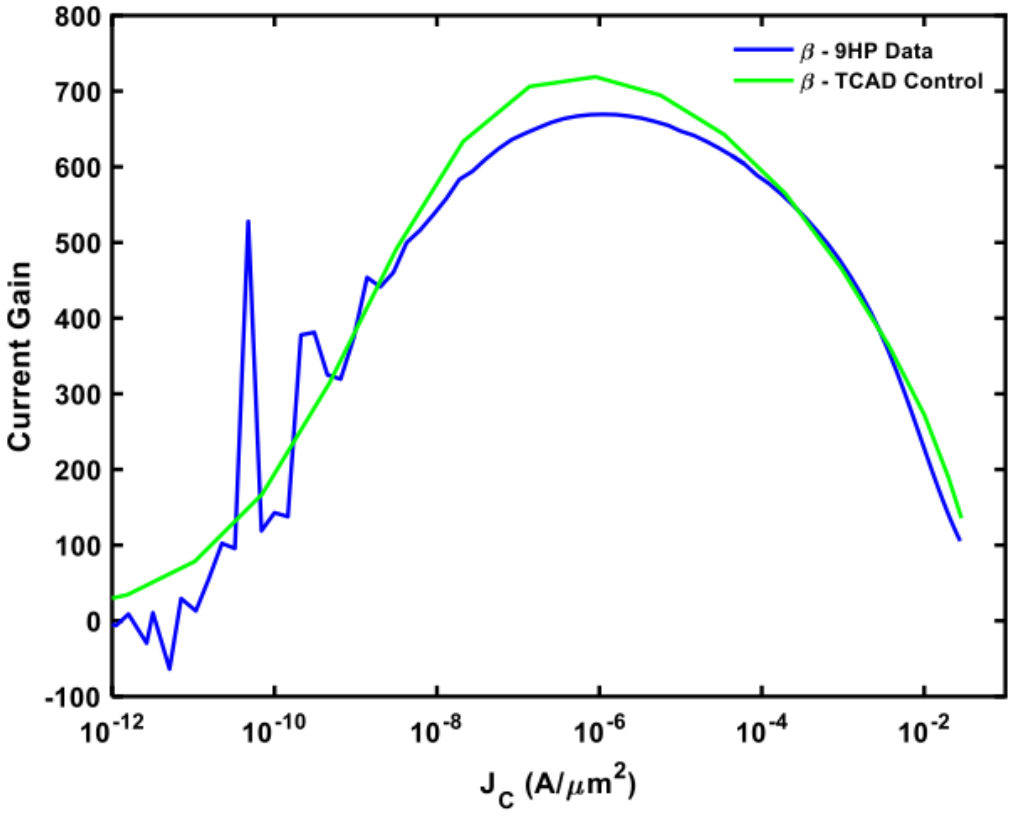

Figure 2.13: Calibration of DC current gain, β.

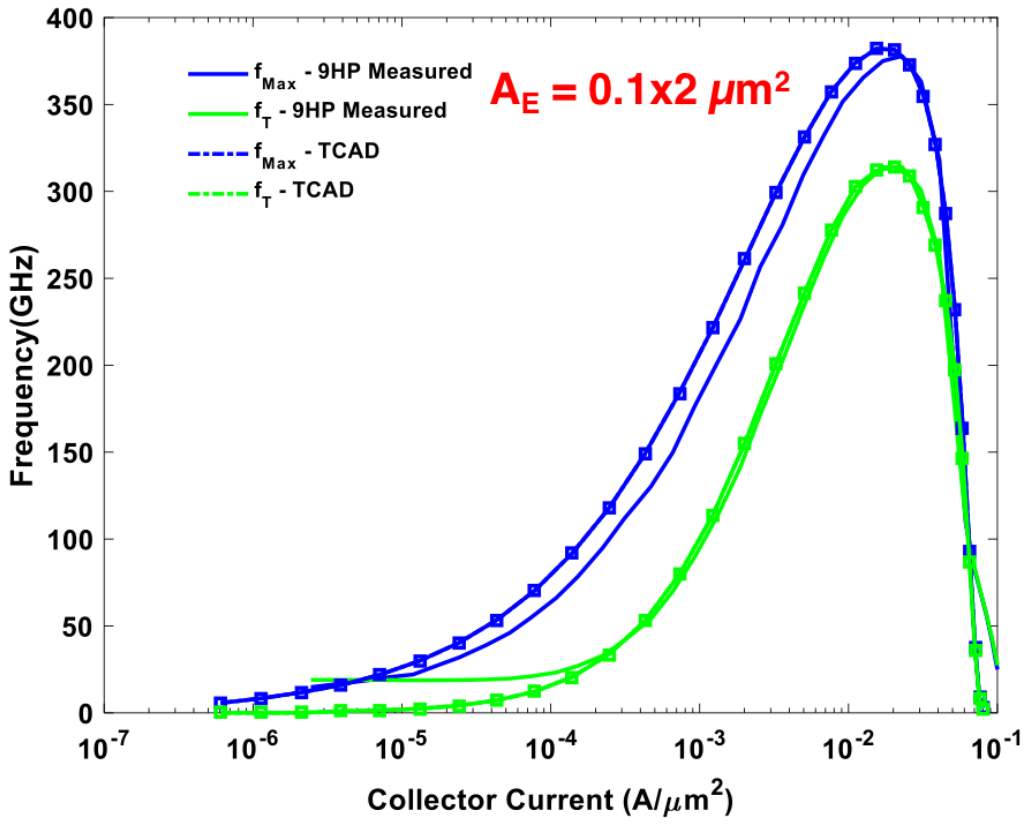

Figure 2.14: The calibration of AC performance of the profile through f_T/f_{Max} comparison with the measurement data.

CHAPTER 3
EXISTING TRANSPORT MECHANISMS

Two transport mechanisms have been inferred in literatures to account for the collector current of SiGe HBTs operating at cryogenic temperatures: quasi-ballistic transport [30] and trap-assisted tunneling [11]. This chapter summarizes the two mechanisms briefly in order to explain qualitatively how the direct tunneling could fit in the context of other transport mechanisms. Chap. 4 will explain the quantitative results of direct tunneling in detail.

3.1 Quasi-ballistic Transport

Quasi-ballistic transport was inferred to account for the collector current of SiGe HBTs at low temperatures [30]. Although there is no concrete proof for its presence, no evidence has thus far been raised against it, either. Therefore, we assume in this thesis that such mechanism is present. We do need to emphasize that the validity of Sec. 4.6 needs to be re-examined shall the mechanism be proved wrong in the future, and that other sections of the thesis do not depend upon the presence of quasi-ballistic transport.

To understand the quasi-ballistic transport, we shall first introduce the semiclassical transport mechanism in general. Charge carriers can contribute to the collector current through two classical transport mechanisms. Here, the word "classical" simply denotes that carriers can be described semi-classically as particles. This is different from the quantum mechanical description of carriers as propagating waves (e.g., in tunneling). Classical transport includes drift-diffusion (in a thick base) and quasi-ballistic transport (in a thin base). The word "thick" versus "thin" refers to how the base width compares to the mean free path of carriers that contribute to the collector current. In a transistor with a thick base, the transport of carriers through the base is mostly by diffusion, where carriers encounter

numerous collisions and the transport can be characterized by "average" parameters such as mobility. For an extremely thin base, however, most carriers traverse the base without scattering and the transport is called ballistic transport [40, 41, 42]. For an intermediate base width, both forms of transport exist and the transport is called quasi-ballistic transport.

Quasi-ballistic transport was used to explain the saturation of transconductance (slope of collector at zero V_{CB}) at low temperatures [30]. At cryogenic temperatures, the lack of phonons results in less energy exchange between the electron ensemble and the lattice, which translates to a reduced cooling power on the electrons. This shortage of phonon cooling can cause the electron temperature, even in the steady state, to be higher than the lattice (i.e., ambient) temperature. Therefore, previous work has modeled the collector current phenomenologically as drift-diffusion current, but with an electron temperature higher than the ambient temperature [30, 43].

To model the quasi-ballistic transport, we can start with the drift-diffusion model. In the drift-diffusion model, the collector current consists of emitter electrons diffusing through the emitter-base depletion region and subsequently traversing the base to reach the collector. Since the base is short, and the base-collector built-in field is strong, it can be assumed that the majority of the electrons reaching the emitter side of the base will reach collector. In other words, the collector current is proportional to the number of minority electrons at the emitter side of the base (diffusing downstream from emitter to base). If the emitter electrons, due to reduced phonon scattering at low temperatures, have a higher energy when they reach the base, we can associate the transport with a higher effective temperature. Therefore, the collector current will be proportional to

$$I_C \propto e^{\frac{qV_{BE}}{k_B T_e}} \tag{3.1}$$

where I_C is the collector current, V_{BE} is the base-emitter voltage, and T_e is the electron temperature that is higher than ambient temperature.

Generally, it is observed that the temperature scaling of collector current initially tracks

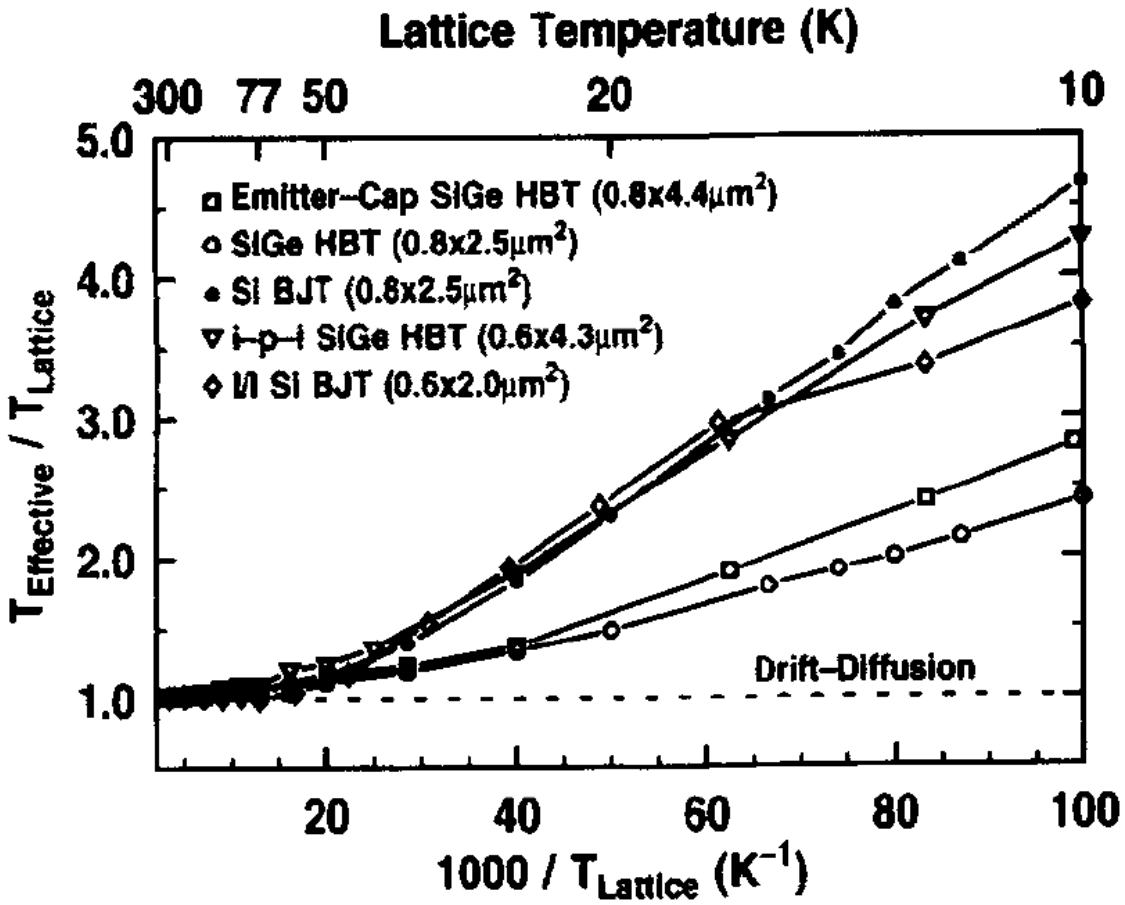

Figure 3.1: Effective temperature (electron temperature) versus the ambient temperature for various kinds of SiGe HBT devices (after [30]).

the ambient temperature until the temperature is "low", at which point the slope saturates to a constant value [43]. Since the slope of current implies a particular electron temperature as in Equ. 3.1, the saturation of slope implies that electron temperature saturates at low temperature. The ratio of effective electron temperature versus the ambient temperature for multiple types of devices is plotted in Fig. 3.1. As illustrated, the electron temperature starts to diverge from the ambient temperature at around 50 K and below. For this reason, in the present work, the deviation of the collector current slope from the ambient temperature with cooling is assumed to come from the quasi-ballistic transport.

As mentioned above, the proof or disproof of quasi-ballistic transport is still an open research question, and we do not intend to discuss such subject in this thesis.

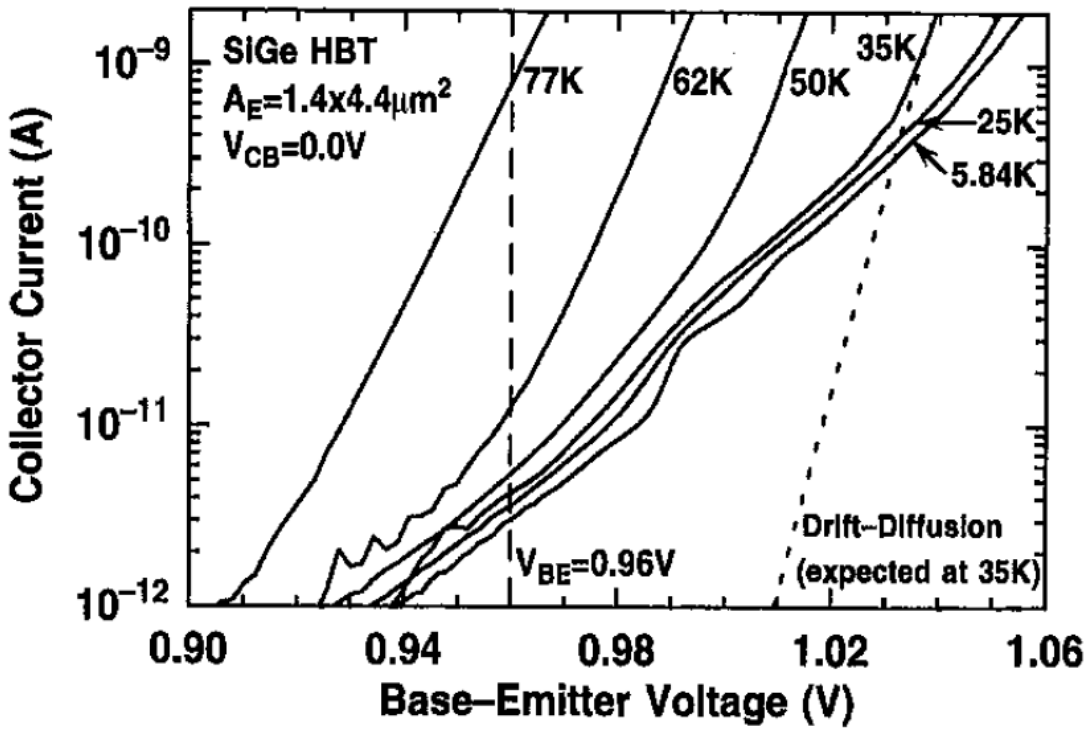

Figure 3.2: Collector current versus base-emitter voltage of SiGe HBTs from 77 K down to 5.84 K as the leakage current below 1 nA becomes more pronounced as the temperature decreases (after [11]).

3.2 Trap-assisted Tunneling

Another form of transport mechanism in collector current is trap-assisted tunneling. Carriers can tunnel quantum-mechanically from the emitter to the collector through trap states in the base [11]. Such mechanism exists when there is a finite distribution of traps in the bandgap of base region and the base width is small. Trap-assisted tunneling was observed to be prevalent in bipolar junction transistors (BJTs) of various base profiles and device structures at cryogenic temperature [11]. As shown in Fig. 3.2, the signature of trap-assisted tunneling is a progressively noticeable leakage current at low injection as the temperature is reduced.

Trap-assisted tunneling is weakly temperature dependent, because the temperature dependence mostly comes from the dependence of the potential profile on temperature [11].

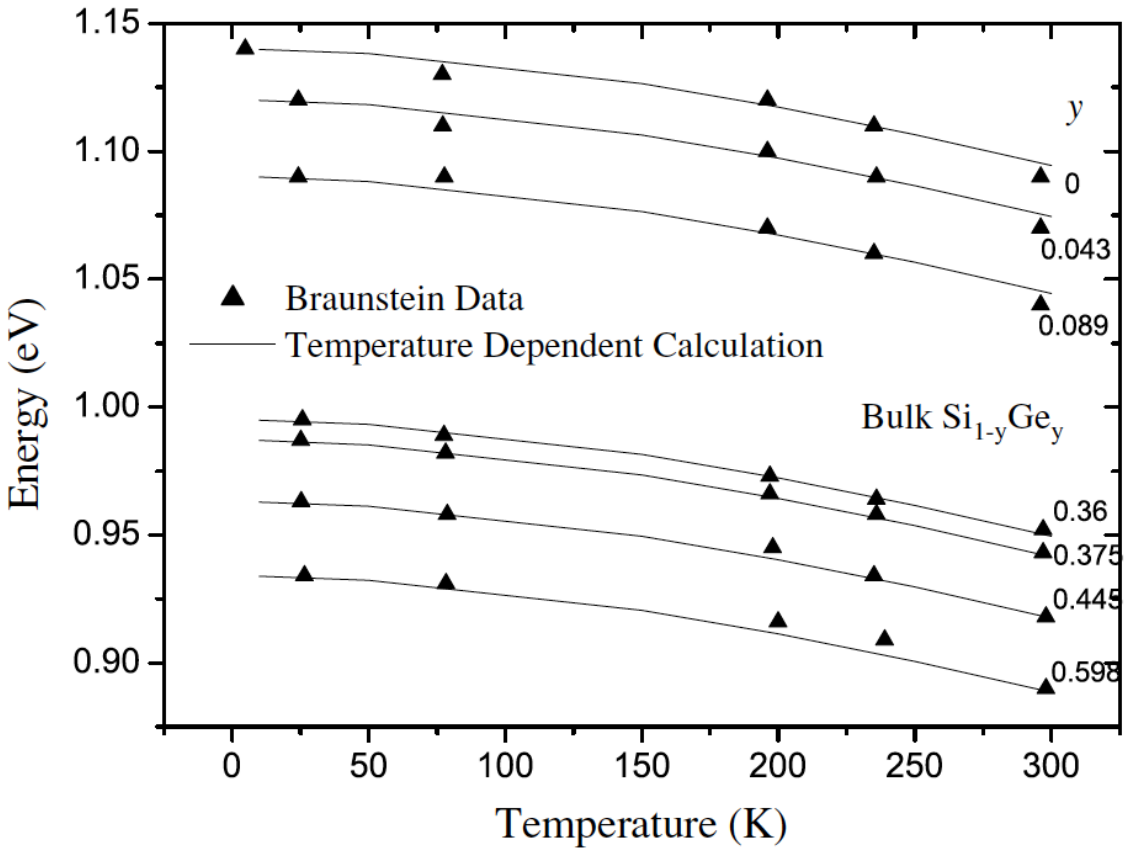

Figure 3.3: The bandgap energy (simulated and measured data) of the Si and SiGe alloy with different Ge mole fractions across temperature (after [44]).

For SiGe HBTs that do not have the freeze-out effect, the potential profile depends on the charge carrier density and the inherent band structures (bandgap). On one hand, the charge profile of the device is constant across temperatures, because at low injection, the charge profile is determined by the background doping which are constant across temperature (no freeze-out effect as discussed in Chap. 2). On the other hand, the bandgap of Si and SiGe alloy both saturate at low temperatures, as shown in Fig. 3.3. Therefore, at temperatures where trap-assisted tunneling is observed, the band structure is weakly dependent on temperature, which results in a weak dependence on temperature for trap-assisted tunneling.

3.3 Overall Picture

A qualitative illustration of the transport mechanisms with respect to bias is shown in Fig. 3.4. Here we show the conduction band edge of a typical SiGe HBT.

When the base-emitter voltage (V_{BE}) is small, no conduction is possible because carriers do not have sufficient energy to surmount the base potential barrier, and the base width is too large for any tunneling process.

As V_{BE} is increased, however, the conduction band energy in the emitter is raised relative to the base. For a positive base-emitter bias, we can define the base barrier width as the distance between two points of the same energy in the conduction band, one to the left and the other to the right of the neutral base. With increasing V_{BE}, the quasi-Fermi level of emitter (i.e. energy of electrons in the emitter) is increased, for which energy the base barrier width is reduced, as can be visually seen in Fig. 3.4. The distance reduction mainly comes from the collector side, where the conduction band edge is moving closer towards base with higher energy. The barrier height for electrons in the emitter is reduced as well, as the quasi-Fermi level of emitter is raised closer to the peak of base conduction band. If the base barrier width is small and trap levels exist in the base, electrons can tunnel from the emitter into the trap states in the base and subsequently tunnel into the collector to generate transport current. With increasing V_{BE}, base width is reduced even further.

Direct tunneling may become possible when the base width is so small that electrons can tunnel directly (instead of via traps) from emitter to collector. At higher V_{BE}, electrons are brought close to the top of the base conduction band, and classical conduction occurs.

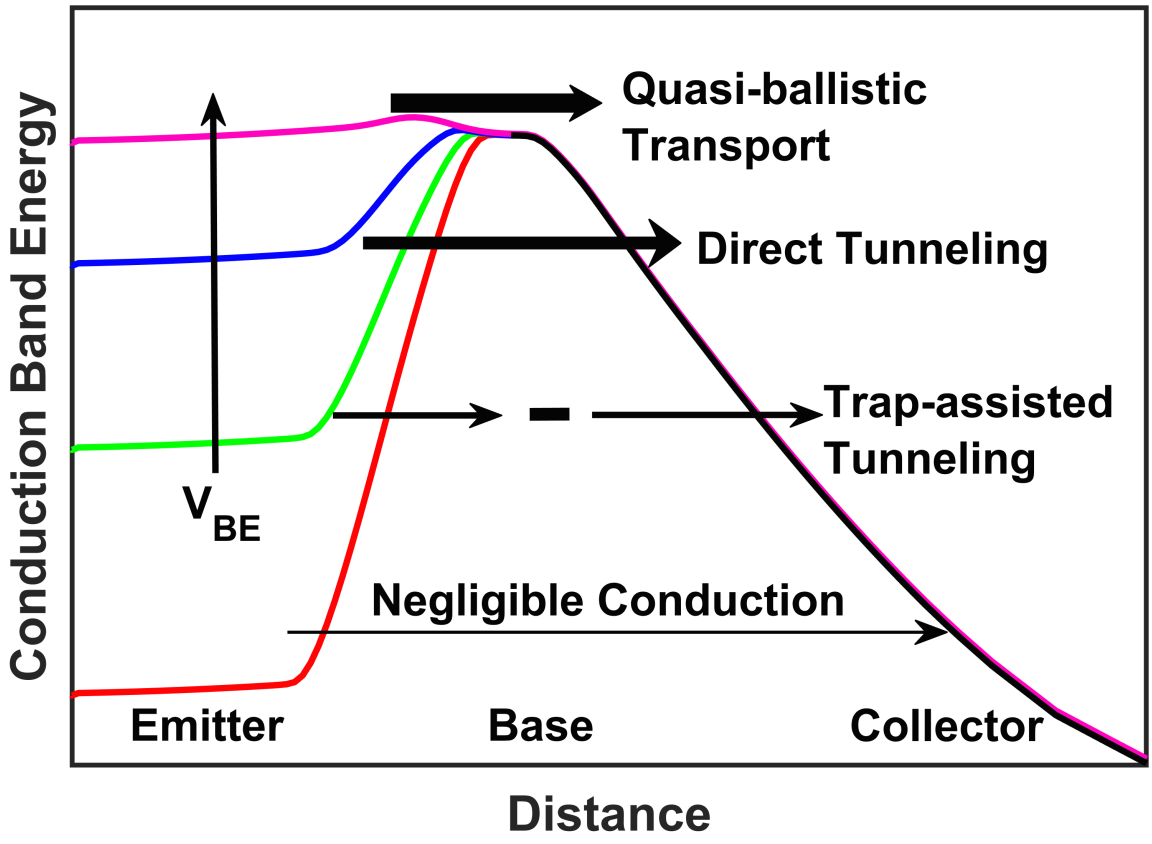

Figure 3.4: Qualitative illustration of conduction band of SiGe HBTs under various V_{BE} values, showing the dominant transport mechanisms in different bias regimes.

CHAPTER 4

NEW TRANSPORT MECHANISM-DIRECT TUNNELING

The transport theories from Chap. 3 are able to explain the experimental data well, until we characterized some early versions of the fourth generation devices. In Sec. 4.1.1, we present the data that can be explained by existing transport theories. In Sec. 4.1.2, we point out a "step" non-ideality whose temperature dependence raises doubts about existing theories. Due to the similarity of "step" to the single-charge tunneling, we introduce the basics of tunneling in Sec. 4.3 and conducted a quantitative analysis on the non-ideality. The conclusion is that a different transport mechanism, likely direct tunneling, exists in the regime previously considered to be quasi-ballistic transport. Based on the quantitative analysis in Sec. 4.4, it is very plausible that direct tunneling is present. Based on this, we compare between multiple technology generations and find the characteristics of this mechanism is congruent with that of direct tunneling, which gives further evidence that direct tunneling is present. In Sec. 4.6, a method is proposed to differentiate between the quasi-ballistic transport and direct tunneling, which have very similar I-V characteristics.

4.1 Experimental Data

4.1.1 Typical Characteristics

In this section, we present the measurement data of cryogenically-operated SiGe HBTs. The intention is to familiarize readers with the common features of SiGe HBTs data at cryogenic temperatures, such as the loss of temperature dependence.

A prototype version of the fourth generation SiGe HBT devices is characterized from 300 K to 70 mK. This is the lowest temperature any SiGe HBTs have ever been characterized to be operable [1]. The Gummel characteristics, or transfer characteristics (base

current I_B and collector current I_C versus base-emitter voltage V_{BE}[1]), of this device are shown in Fig. 4.1. As illustrated, the collector and the base current is temperature dependent down to 16.7 K, but below 16.7 K, and all the way down to 70 mK, the curves (both I_C and I_B) essentially overlap, even though the difference in reciprocal temperature is enormous (1000/T = 60 at 16.7 K vs. 1000/T = 14,286 at 70 mK). Since the drift-diffusion transport is exponentially dependent on the reciprocal temperature, the measured current is clearly different from the thermally-activated drift-diffusion mechanism.

As a closer examination of the non-ideal temperature dependence, the transconductance (g_m) and DC current gain (β) were extracted from the Gummel characteristics at a fixed collector current density and plotted versus reciprocal temperature (1/T) in Fig. 4.2. As can be seen, from 300 K to 40 K, the curve is linear, which means $log(g_m)$ is linear with 1/T. This is congruent with the drift-diffusion theory. Below 40 K, however, g_m flattens, indicating the slope of the collector is no longer steepening. Similarly, the current gain flattens below 40 K, indicating the slope of base current also stops changing together with that of the collector current. This is consistent with the lack of temperature scaling between 16.7 K and 70 mK seen in Fig. 4.1. The temperature independent collector and base current was observed in two other measured samples. Experimentally, this characteristic is common among cryogenically-operated SiGe HBTs and frequently observed [30, 43, 45]. In [30], the quasi-ballistic transport is used to explain the loss of temperature dependence, as explained in Chap. 3, and is widely accepted [46]. However, the presence of another non-ideality raised our attention.

[1]The V_{BE} is assumed to be applied on the actual junction until the emitter current starts to flatten from the exponential curve (slope becomes smaller) due to the parasitic resistance. When the flattening occurs in the emitter current, it also shows in collect and base current (which combine to give the emitter current). The effect of collector resistance versus emitter resistance is not easily separated. However, in the measurement presented in this thesis, the measurement is clipped much below the high-injection region where the resistance effect is visible.

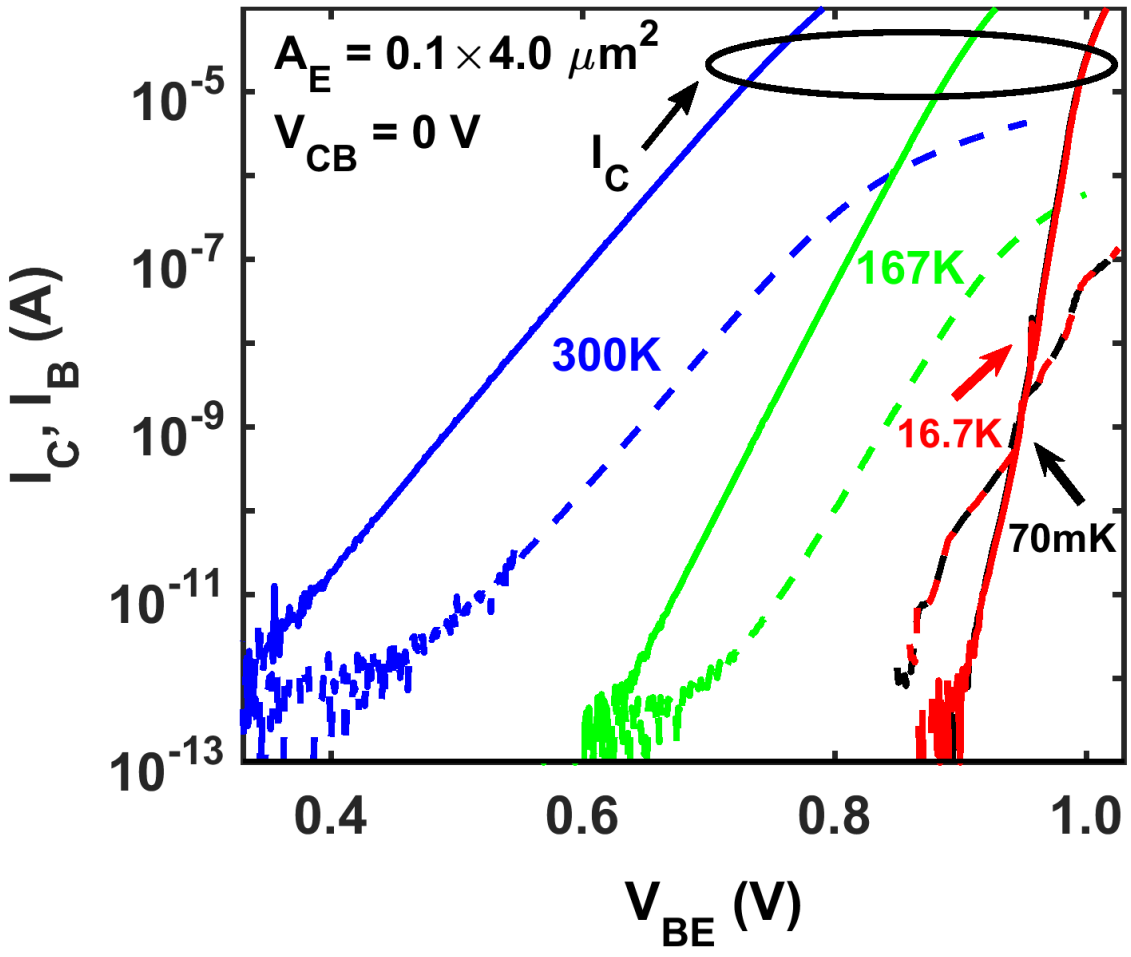

Figure 4.1: The Gummel characteristics (collector and base current versus base-emitter voltage) of SiGe HBTs at 300 K, 167 K, 16.7 K, and 70 mK. The curves at 16.7 K and 70 mK overlaps. The black arrow at 70 mK points to the "step" in base current (dashed line).

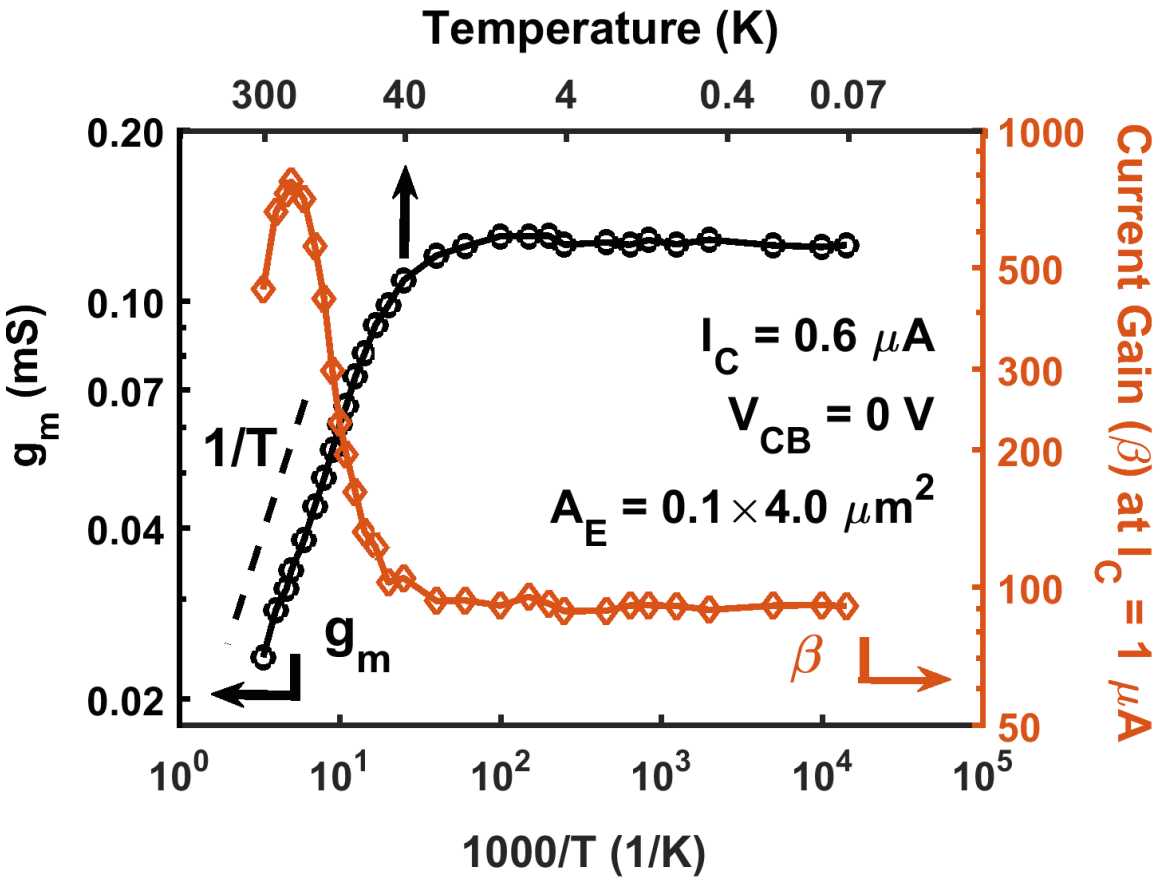

Figure 4.2: The transconductance (g_m) and DC current gain (β) of a SiGe HBT versus 1/T showing the loss of temperature dependence for these parameters below around 40 K. Above 40 K, the expected 1/T dependence of transconductance from the drift-diffusion transport is shown in dashed line.

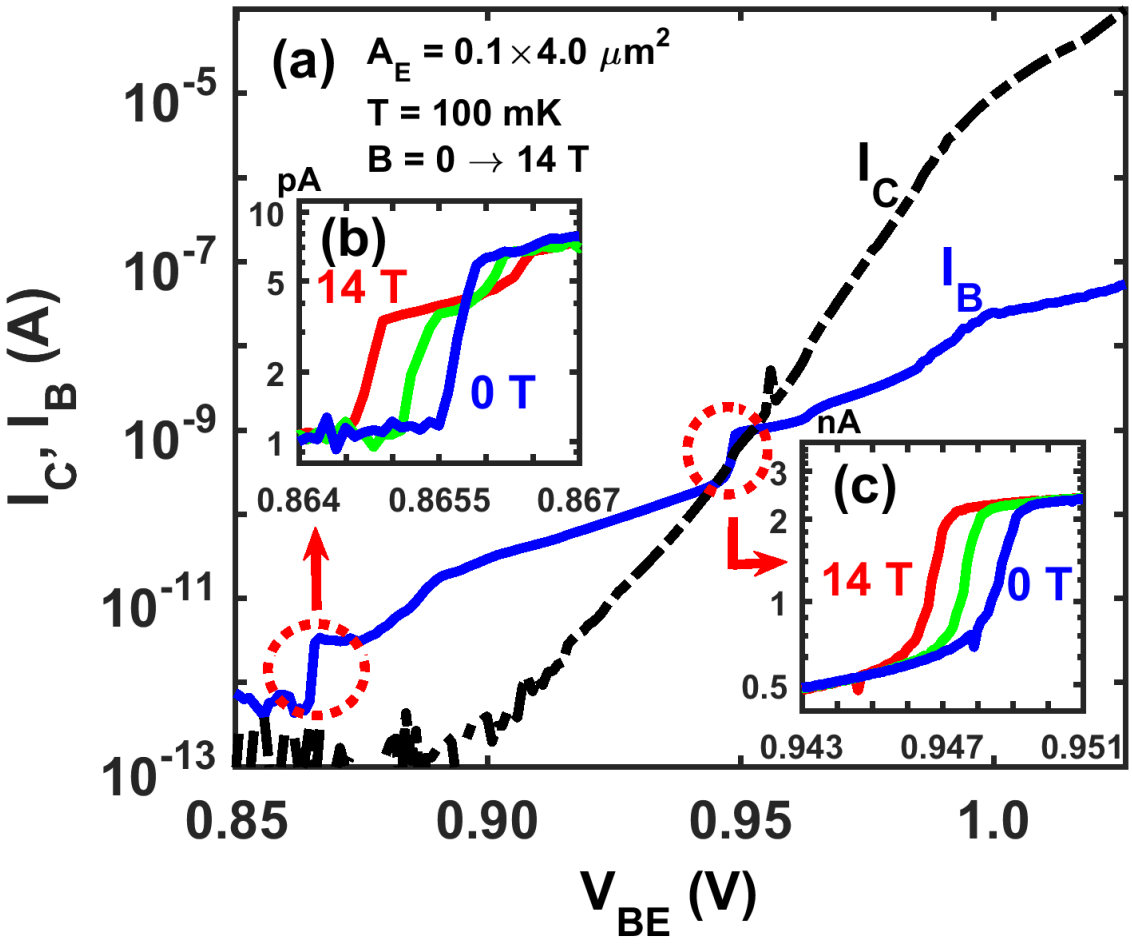

Figure 4.3: (a) The zoom-in view of the Gummel characteristics of SiGe HBTs at 100‘ mK. (b) and (c): zoom-in view of the base current under a magnetic field of 0 T, 7 T, and 14 T. The same axis and unit is used in (a)-(c).

4.1.2 “Step” Discontinuity

This section presents the “step” non-ideality never seen before in SiGe HBTs at cryogenic temperatures, which stimulates our interest to study it further.

The Gummel measurement at 16.7 K and 70 mK reveals a novel non-ideality. In Fig. 4.1, the base current exhibits a sharp transition (“step”) around the region where collector and base current crosses. A zoomed-in view of Gummel (i.e., V_{BE} = 0.85 V to 1.02 V) at 100 mK is provided in Fig. 4.3. As can be seen, the base current has two “step” discontinuities that have very steep slopes.

This "step" non-ideality is present in the other two SiGe HBTs from the same batch of fabrication. It is present in the inverse Gummel measurement as well, where we biased the transistor in reverse active mode (i.e. treating the physical emitter as collector and physical collector as emitter). This configuration is typically not useful for circuits, but does provide insights into the device physics. An interesting inferrence can be made based on the comparison of number of "steps" in forward versus inverse mode, as discussed in Sec. 4.3.2. A total of 3 samples were measured (including the aforementioned sample 1) and showed similar behaviors except the "steps" occur at different V_{BE} values. Both forward and inverse Gummel measurement of all three samples at 70 mK under different V_{CB} (for forward Gummel) or V_{EB} (for inverse Gummel) are shown in Fig. 4.4. The variation of V_{CB} and V_{EB} reveals some interesting behavior, which we will discuss in Sec. 4.6.

More insight on the "steps" is obtained through further experiment. A tunable static magnetic field (-14 T to +14 T) was applied on the device at 100 mK. The reason for applying a magnetic field is two-fold. First, the temperature-independent Gummel characteristics imply that the transport mechanism should not have a strong temperature dependence. One possible mechanism which is consistent with this hypothesis is tunneling through the base region. Second, the "step" in the base current resembles the characteristic I-V curve of electrons tunneling through a discrete energy level under an electric field [47]. According to the Zeeman effect, a degenerate energy level will split into two sub-levels under a static magnetic field. Therefore, if the transport is initiated by tunneling, and the "steps" are caused by discrete trap levels, an applied magnetic field should affect this transport mechanism and reveal itself in the magnetic field dependent Gummel characteristics.

Thus, a static and uniform magnetic field parallel to the transport direction (i.e., from emitter to collector) was applied to sample 1. The Gummel characteristics around the "step" region under different magnetic fields are shown in Fig. 4.3(b) and Fig. 4.3(c), for the lower and upper step, respectively. At 14 T, the lower step splits into two steps. As

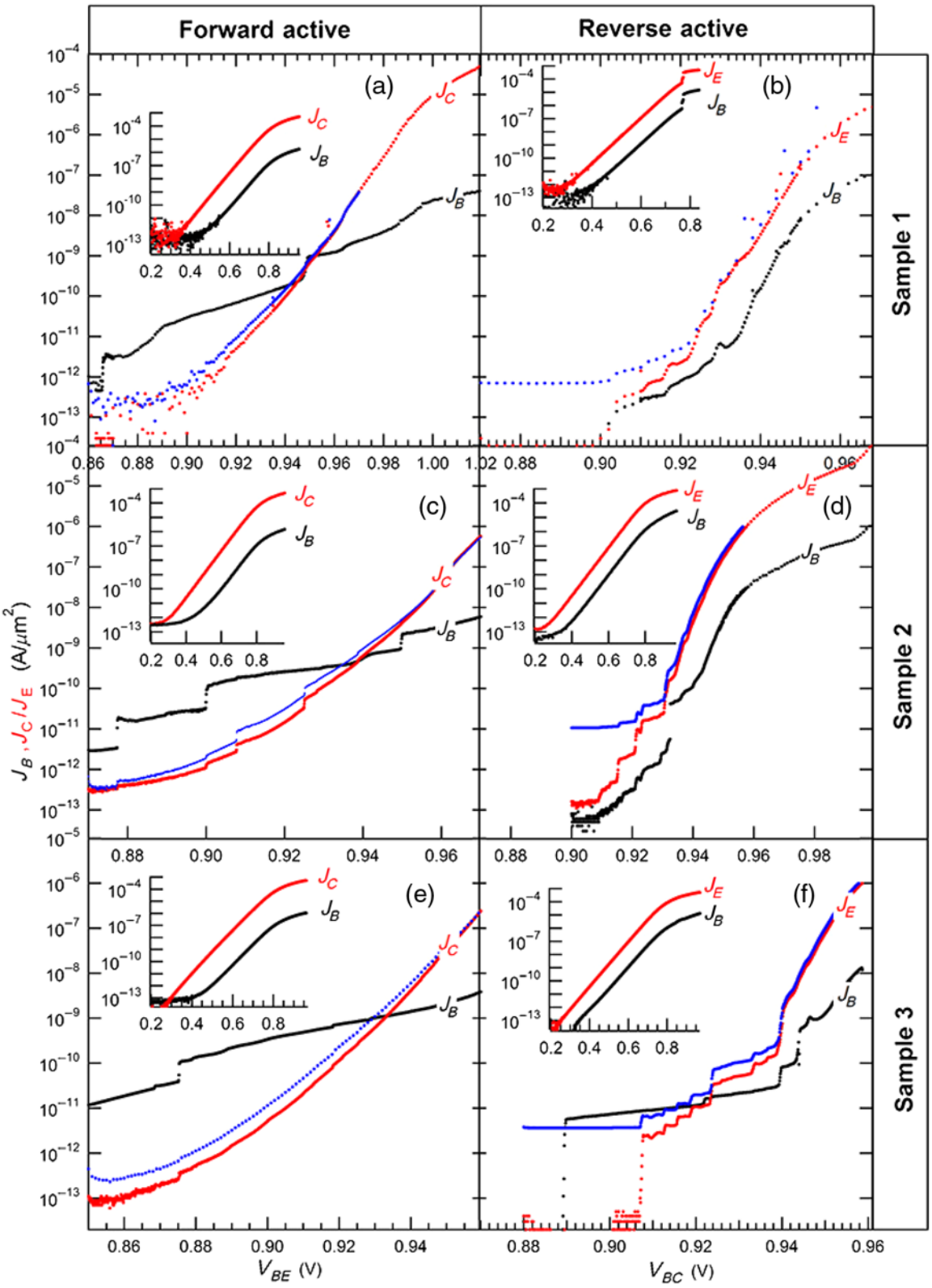

Figure 4.4: (a), (c), (e): forward Gummel characteristics of sample 1-3 at 70 mK. (b), (d), (f): inverse Gummel characteristics of sample 1-3 at 70 mK. Red curves indicate collector current density at V_{CB} (for forward Gummel) or V_{EB} (for inverse Gummel) of 0 V while blue curves are V_{CB} (for forward Gummel) or V_{EB} (for inverse Gummel) of 0.2 V. Inset: Gummel characteristics at 300 K.

shown in the red curve in Fig. 4.3(b), the lower portion of the current shifts to the left while the upper portion of the current shifts to the right, resulting in two discrete steps. This infers that for this "step", the tunneling is occurring through a discrete energy level, where the level splits into two non-degenerate levels under an applied magnetic field. At 7 T, the current lies between the 0 and 14 T curves, denoting that the current splitting has a linear dependence on the applied magnetic field, which further confirms the presence of the Zeeman effect. Interestingly, the upper step does not show such a splitting, but rather a uniform shift in current under magnetic field. It is likely that when the discrete energy level splits, the upper sub-level is suppressed due to asymmetric transport rates [48]. Therefore, the discrete level essentially has a lower energy. This translates to current conduction at a lower base-emitter voltage, as shown in Fig. 4.3(c).

The visual inspection reveals that "steps" have steep slopes, similar to the observed steps in current when a single charge tunnels into a discrete energy state. To examine it further, we adopted the transport spectroscopy methods frequently used in the field of single electron transistors (SET) and tunnel junctions [49, 50, 51, 52]. The method is well established to determine the electron temperature from the temperature dependence of full width at half maximum (FWHM) of current steps. It has been used recently to analyze tunneling through an individual donor in a semiconductor device [53]. Before we embark on that journey, however, we will take a detour now and introduce some important assumptions and simplifications that will later be referred to during the derivation or calculation.

4.2 Assumptions and Simplifications

The problem of calculating tunneling current in a SiGe HBT comes down to finding the tunneling rate of electrons. Though the exact solution is found from one electron Schrodinger equation in Equ. 4.1, it is often unnecessary due to its complexity in computation. Rather, the effective mass equation, as in Equ. 4.2, is used to solve for an envelope function instead

of the electron wave function.

$$E(k)\psi_0(r) = (-\frac{\hbar^2}{2m_0}\nabla^2 + U_L(r) + U_E(r))\psi_0(r) \tag{4.1}$$

$$E\psi(r) = -\frac{\hbar^2}{2m^*}\nabla^2\psi(r) + E_C(r)\psi(r) \tag{4.2}$$

As mentioned in Appendix A, the effective mass equation comes with an assumption. To make the analysis in Sec. 4.4 easier, we also make additional assumptions. In total, we adopt the following three assumptions and discuss them individually in this section: 1) all electrons in the transport can be described by the same constant effective mass, 2) the self-periodic part of the basis function $\mu_{v,k}$ is slowly varying in k space and is of similar nature between the initial and final state, and 3) the potential E_C is slowly varying within a unit cell.

4.2.1 Assumption 1

Assumption 1: all electrons in the transport can be described by the same constant effective mass.

For Assumption 1) to be valid, electrons have to occupy energy ranges near the bottom of the conduction band. It is known that the conduction band is parabolic near the band minimum, where electrons have constant effective mass. There are two potential concerns for this 1) at cryogenic temperatures, electrons could freeze out to the impurity band, and no free electrons are available in the conduction band, not to mention contributing to any conduction, and 2) the tunneling calculation spans across multiple regions of the device with different doping and Ge alloy mole fraction, and the effective mass in those regions may be different. Fortunately, the two concerns are alleviated based on experimental data.

For concern 1), electron spin resonance data on degenerately and non-degenerately doped Si samples show that instead of having two types of electrons (from the impurity

band and the conduction band), degenerately doped samples only one type of electrons as conduction band electrons [54]. The reason is that in a heavily doped Si sample, the impurity band and the conduction band overlaps in energy, and the interaction between the two bands delocalizes electrons. In other words, the nature of the electrons participating in the transport from the impurity band is the same as those from the conduction band. An interesting fact that results from this is the similar values of effective mass obtained from different types of measurement, such as hall mobility and cyclotron resonance measurement [55, 56, 57, 58]. This is the case even with the presence of non-significant compensation doping (i.e. in the emitter-base depletion region where the the net doping is low but either species is doped heavily) [56]. In addition, the free electrons only populate the bottom of the conduction band at zero temperature ($T = 0\ K$), when the sharp Fermi-Dirac distribution prevents electrons from occupying higher energy portion of the band.

For concern 2), a closer examination shows that for SiGe HBTs under consideration, the emitter and collector are heavily doped n-type silicon while the base is heavily doped p-type SiGe alloy (Ge mole fraction $< 30\%$). In any case, we assume a constant effective mass for all regions in this thesis. This approximation is justified in two aspects. First, the effective mass of electron is weakly dependent on the Ge mole fraction or on the induced strain by Ge until the Ge mole fraction is larger than 85% [59]. This is also showcased from the band structure calculation in Appendix C. The reason is that the addition of Ge only affects the curvature of the higher conduction bands and not the lowest one that we are concerned with [60]. Second, the effective mass is mostly independent from the As doping concentration until the doping concentration is above $1 \times 10^{21}/cm^3$ [61]. From this, we assume that effective mass is mostly independent from B and P doping concentration as well, due to their similarity. In all regions, the doping concentration is less than $1 \times 10^{21}/cm^3$, so it is justified to use the same effecive mass in all regions. Experimentally, the transverse effective mass is around $0.18m_0$ at 4.2 K [62]. Since the effective mass is weakly dependent

on temperature, we assume at 40 K and below the effective mass is all $0.18m_0$ [63].

The fact of constant effective mass in all regions alleviates a less noticeable concern, which is the boundary condition for inhomogenous material. The boundary condition for Equ. 4.2 will be a condition for the envelope function, $\phi(r)$, and not the original wave function $\psi_0(r,t)$. In the limiting case of an abrupt junction of inhomogenous material on either side, the current-conserving boundary condition between different regions along x direction is

$$\frac{1}{m_L}\frac{\partial\psi(r)}{\partial x}|_{x\to 0_L} = \frac{1}{m_R}\frac{\partial\psi(r)}{\partial x}|_{x\to 0_R} \tag{4.3}$$

where m_L and m_R are the effective mass to the left and to the right of the boundary, and the derivative terms are the envelope function evaluated at either side of the boundary. Since the effective mass in all regions is the same, the treatment on boundary conditions is simplified.

4.2.2 Assumption 2

Assumption 2: the self-periodic part of the basis function $\mu_{v,k}$ is slowly varying in k space and is of similar nature between the initial and final state.

As detailed in Appendix A, the absence of $\mu_{v,k}$ (replaced by the envelope wave function) in the construction of the effective mass equation comes with approximations, the Assumption 2 being one of them. To prove this assumption for SiGe HBTs under question, we use the 15-band $k\cdot p$ method from [64, 65] and compare the eigenvectors at wave vectors of interest. In particular, we compare wave vector k whose energy lies within $3kT$ of the conduction band minimum, for different Ge mole fraction and doping. One approximation that we take here is that doping level between $4\times 10^{18}/cm^3$ and $5\times 10^{20}/cm^3$ affects the band structure in a similar way. As mentioned in Assumption 1, the effective mass does not change much in this range of doping concentration, and the curvature of the band is weakly dependent on the doping level. To get an exact evaluation, a Schrodinger-Poisson

solver needs to be constructed to include the self-consistent calculation of band structure due to doping concentration, which is beyond the scope of this thesis. Therefore, the task of evaluating wave vector k across doping and Ge mole fraction simplifies to the evaluation across Ge mole fraction only.

The details of the Hamiltonian matrix from the $k \cdot p$ construction can be found in [64]. The direction of interest is [001] for the SiGe HBTs device under investigation. The 15-band construction denotes the wave function is constructed from 15 basis functions, which are the wave functions of 2S, 2Pz, 2Py, 2Px, 3S, 3Pz, 3Py, 3Px, 4S, 3Dx, 3Dz, 4Pz, 4Py, 4Px, and 5S orbitals. Therefore, the solution to the Hamiltonian is the actual wave function, which is represented as an eigenvector whose values are the coefficients for each basis function, and the linear combination of all basis function with the coefficient constructs the actual wave function. The eigenvalue is the energy of this wave function. In essence, this is an eigenvalue problem. The eigenvalues and eigenvectors are solved by eig function in *MATLAB R2017b*. The numbers shown in Table C.1 in Appendix C are the coefficient in front of each basis function.

SiGe alloy is similar to Si when the Ge mole fraction is less than 85%. In such alloy, the indirect bandgap occurs at around $k = 0.88 \times 2\pi/a$ along [100] as in Si. The range of k for energy within $3kT$ ($\approx 3.4meV$ for a T of $40K$) of the conduction band minimum is $0.86 - 0.9 \times 2\pi/a$. In Table C.2 and Table C.3 in the Appendix C, we list the eigenvectors at this range of k for Ge mole fraction of $0 - 30\%$.

The first part of the assumption that $\mu_{v,k}$ is slowly varying in k space is now easy to justify. From Table C.2 and C.3, the eigenvectors vary within 7% when k is close to the bottom of the conduction band, and the approximation of them being the same is not outrageous. The second part of the assumption can be justified from Table C.1 in Appendix C. The maximum variation of eigenvectors is 12% when the Ge mole fraction is varied from 0% to 30%. Therefore, for initial and final states close to the bottom of the conduction band in SiGe HBTs, the error due to approximating the overlap integral as unity is at most 12%.

4.2.3 Assumption 3

Assumption 3: the potential E_C is slowly varying within a unit cell.

Since this assumption needs to be justified for the Wentzel-Kramers-Brillouin (WKB) approximation we will adopt shortly, a brief introduction to WKB approximation is given here.

To obtain the transmission probability and therefore tunneling rate, we need to solve the envelope wave function, ψ, from the effective mass equation as in Equ. 4.2. For complicated shapes of potential profile, WKB approximation can be applied to solve Equ. 4.4 for tunneling calculation. In addition, it gives an analytical result that is easy to understand and is also incorporated in the device simulator.

$$-\frac{\hbar}{2m}\frac{\partial^2\psi}{\partial x^2} + U(x)\psi = E\psi \tag{4.4}$$

The derivation and requirement for applying WKB approximation is well documented in literatures and is briefly summarized in Appendix B [66]. Repeating the same result from the Appendix B, we have (same as Equ. B.7)

$$\frac{\partial U(x)}{\partial x} \ll 2\sqrt{\frac{2m}{\hbar^2}}(|U(x) - E|)^{3/2} \tag{4.5}$$

where U is the potential energy along the tunneling direction x, m is the effective mass of electron, and E is the energy of the electron.

Equ. 4.5 essentially states that the potential energy should vary slowly compared to the wavelength of the tunneling electrons, because the right-hand side has the dimensionality of $(U - E)/\lambda$, or energy over the wavelength. This is a slightly more stringent requirement than the Assumption 3, because the electron wavelength tends to be comparable or larger than the lattice constant. Therefore, if we can satisfy Equ. 4.5, we can satisfy the Assumption 3 as well.

In a SiGe HBT, $U(x) = E_C(x)$, where $E_C(x)$ is the conduction band edge, and $m =$

m_e, where we use the effective mass of electron $m_e = 0.18m_0$ in the transport direction from emitter to collector. To assess whether the WKB condition is met in SiGe HBTs, we estimate the left- and right-hand side of Equ. 4.5 at the extreme case. At the right-hand side, consider an electron encountering a 0.1 eV barrier ($U - E = 0.1\ eV$). This corresponds to an electron in the emitter when emitter-base junction is forward biased at around 0.95 V and the SiGe base has a bandgap of around 1.05 eV. In this case, $2\sqrt{2m_e/\hbar^2}(E_C - E)^{3/2} \approx 2.2 \times 10^{-11}$ J/m. For the left-hand side, we obtain it from the TCAD model of a fourth generation SiGe HBT whose doping profile was presented Sec. 2.2.1. This represents the worst scenario compared to the TCAD models of the first and third generation devices, because in the fourth generation device, the vertical profile is reduced and the built-in field is stronger in the B-E junction, which translates to a faster varying potential. Since the SiGe alloy in the base is carefully engineered to prevent any abrupt heterojunction interface, the derivative of E_C will never approach infinity. From TCAD simulations (details will be given below), the maximum slope of conduction band edge E_C with respect to the position x occurs in the emitter-base depletion region, where $\frac{\partial U}{\partial x} = \frac{\partial E_C}{\partial x} \approx 4.8\times10^{-12} < 2.2\times10^{-11}$ J/m. This means that Equ. 4.5 is reasonably satisfied but cautions should be exercised for potential barriers less than $0.1\ eV$, where the WKB approximation starts to fail.

4.3 Analysis of "Steps"

Coming back from the detour in the previous section, we should now march towards the goal and use the transport spectroscopy method to analyze the "steps" region and obtain insight on the transport of collector current.

Qualitatively, the nature of the defect states which we are going to analyze is as follows. We assume they are individual defect states in the neutral base region, and that they are both small in number and spaced apart in energy compared to k_BT. This assumption is re-examined in Sec 4.3.2. The defects need not be spatially spaced apart. Actually they could be spatially close as long as the energy difference between them is considerably larger than

k_BT. Also, because the distance of the EB space charge region is on the order of 2-3 nm, electrons can easily tunnel across the barrier if there are another states available in the neutral base. With these in mind, we can draw a qualitative picture of the single charge tunneling as follows.

As the quasi-Fermi energy of electrons in the emitter is raised with increasing V_{BE}, the emitter electron energy will screen through those discrete defect levels in the neutral base. When the defect level aligns with the quasi-Fermi energy, conduction band electrons [2] in the emitter can tunnel through the potential barrier in the EB space charge region and reach the defect state in the base. The electrons subsequently recombine in the base due to the abundance of holes. The electrons could also tunnel into the collector (the neutral base is only about 5 nm thick in these V_{BE} bias levels), but the process is more complicated and is not considered here. When the Fermi distribution in the emitter is a step function (i.e., zero temperature limit where all electrons sit at Fermi level), the change in current will be abrupt as the quasi-Fermi level aligns with the discrete level. When the temperature is higher, the emitter Fermi distribution will be smoother, and the current step will correspondingly be smoother. Therefore, it is intuitive to see why the sharpness of the current step is related to the emitter electron temperature (in the Fermi-Dirac distribution), which we will extract after the following derivation.

An illustration of single charge tunneling is shown in Fig. 4.5. The current from single-charge tunneling, as derived extensively in [48], is

$$I(V) = I_+ f(E_1(V) - E_2 - k_B T ln(\frac{2\gamma_r + \gamma_l}{\gamma_r + \gamma_l})) \tag{4.6}$$

where $I_+ = 2e\frac{\gamma_r\gamma_l}{2\gamma_r+\gamma_l}$, $f(x)$ is the Fermi function in Equ. 4.7, γ_l and γ_r are the tunneling

[2] Note that there are no distinction between impurity band versus conduction band electron, as mentioned in Sec. 4.2.

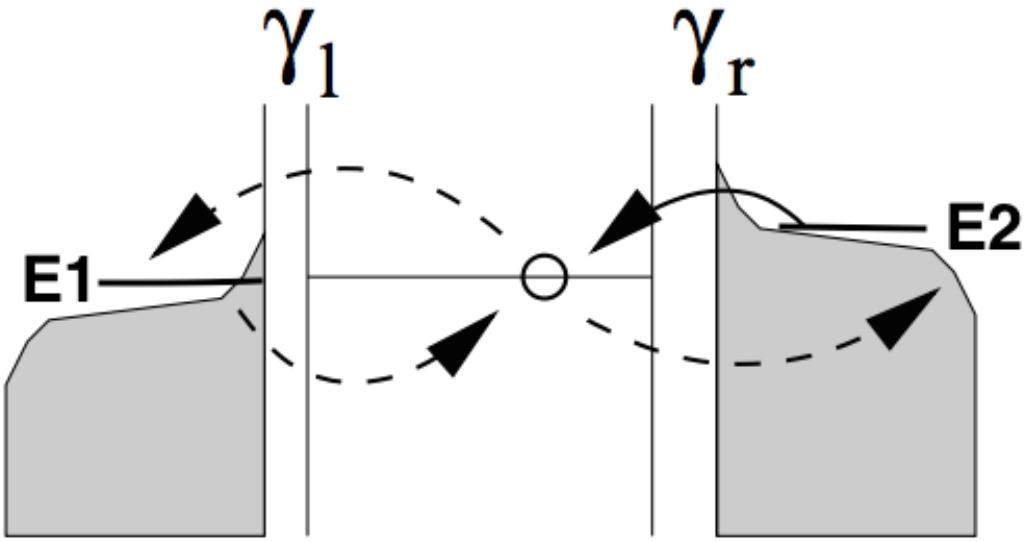

Figure 4.5: Illustration of tunneling rate and energy levels.

rate from left and right contact to the energy level, as illustrated in Fig. 4.5.

$$f(x) = \frac{1}{1 + e^{\frac{x}{k_B T}}} \tag{4.7}$$

Therefore, the output current will have the shape of Fermi function. Equ. 4.6 is slightly different from the standard Fermi function due to the last term in the argument. This term causes a temperature dependent shift in center energy. However, at low temperatures that we are considering, such shift is barely noticeable, and we can regard the function as the standard Fermi-Dirac function. The most prominent effect of temperature on the current is the steepening slope as the temperature decreases. The steepness of the slope is directly controlled by T. Therefore, by extracting the steepness of the slope, we can obtain the temperature of the lead from which the electrons tunnel into the discrete state. Usually, this approximation is checked by comparing the extracted FWHM versus temperature to the theoretical FWHM of Fermi-Dirac distribution ($3.526 k_B T$ from simple algebraic calculations). This will reveal whether the current indeed can be approximated as Fermi-Dirac function.

4.3.1 Electron Temperature Extraction

To obtain the electron temperature, we will extract the "sharpness" of the step by fitting the tunneling current expression to the measurement. The goal is to obtain the derivative of this fitting function versus voltage, from which we can obtain the FWHM for each temperature. From the previous section, plotting this FWHM versus temperature yields a line whose slope indicates whether the current has the shape of Fermi-Dirac function.

First, we chose from sample 2 a couple of "steps" that only show up in the base current (instead of in both base and collector current). Having only two terminal simplifies the analysis while still yields sufficient information for electron temperature. However, directly fitting Equ. 4.7 to the data is not practical, because there are other transport mechanisms causing a leakage current while the single charge tunneling takes place. Those background leakage current causes the current to slowly increase. Since we are ultimately concerned about the steepness of the slope, we fit the data to a modified function as in Equ. 4.8.

$$f(V_{BE}) = \frac{a}{1 + e^{(b - V_{BE})/k_B T}} + cV_{BE} + d \tag{4.8}$$

In Equ. 4.8, a, b, c, d, and T are fitting parameters, where a scales the function vertically, b translates the function laterally along V_{BE}, T determines the sharpness of the distribution at the transition, and c and d translates the curves vertically. The linear term cV_{BE} and d are placed here only to fit the gradual increase of the current in the relatively flat region of the "step" (i.e. on either side the sharp transition), because the base current consists of not only the "step" but also the classical diffusion and recombination current. These linear terms do not affect the following analysis in any way, because we only need to calculate the conductance (dI/dV), which is the derivative of the fitted Fermi function with respect to V_{BE}. Therefore, the offset term d will disappear, and the linear terms cV_{BE} will show up only as a constant offset in the derivative of Equ. 4.8. As one can easily see, the fitting is unique because each fitting parameter controls one and only one feature of the shape.

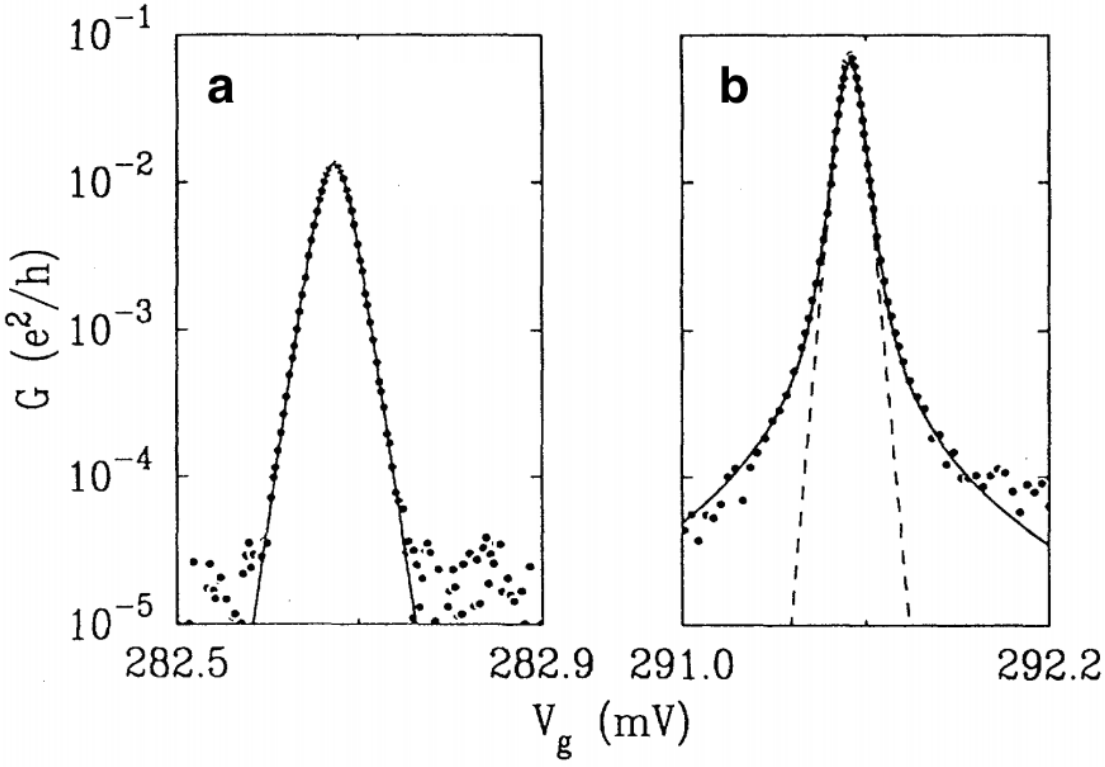

Figure 4.6: (a) Conductance peak fitted by Fermi function. (b) Conductance peak fitted by a broadened Lorentzian (solid line). The dashed line shows the fitting in (a) (after [67]).

The data and the fitted Fermi function for the two "steps" are shown in Fig. 4.7(a) and Fig. 4.7(b), respectively. The fitting is obtained by the built-in *fit* function with least absolute residual robust (LAR) method in *MATLAB R2017a*. The parameters obtained from the fitting and the coefficient of determination (R^2) are given in Appendix D. From the derivative of the fitted Fermi function (i.e., the conductance of the "step" current), the FWHM (in units of voltage) is extracted at each temperature and plotted in Fig. 4.7(c). Visually, the trend looks exactly like the FWHM extracted from a typical current spike in a SET (e.g., see inset of Fig. 2 in [68]). A linear fit is applied to the two curves in Fig. 4.7(c) from 0.5 K to 4 K. The obtained best fit is $k_B(3.48T+0.197)/e$ and $k_B(3.52T+1.69)/e$ for step 1 and 2, respectively. The closeness of the slope to $3.52k_B$ shows that approximating the single charge tunneling as Fermi-Dirac function is not a bad approximation, and that the distribution function of the lead (i.e., emitter) is indeed close to Fermi-Dirac distribution. Such argument is valid only up to a current density of around 2 nA/um^2, above which no "step" are distinctively visible to verify the electron temperature. It is likely that the device is quickly heating up as the bias current increases, or the current from other transport process dominates over the single charge tunneling. It should be noted that the above analysis does not rely on the type of functions used for fitting, as long as the fitting extracts the steepness of the slope. Therefore, we obtain approximately the same FWHM using erf function and arctan function in the fitting. The fitting will be valid for any function that is proportional to x in the limit of x approaching zero.

Next, we would like to obtain the electron temperature from a more acccurate method. In single charge tunneling through a quantum dot, the conductance peak is usually accounted for by Fermi distribution, as shown in Fig. 4.6(a), while the Lorentzian broadening also occurs at the tail of the conductance peak, as shown in Fig. 4.6(b). Therefore, to obtain the electron temperature, it is necessary to remove the effect of Lorentzian broadening on the line width of FWHM. In order to obtain the Lorentzian broadening, the FWHM is calculated as the convolution of Fermi distribution with a Lorentzian [67]. Straightforward

calculations of the convolution shows that $FWHM_E = 0.675\gamma + \sqrt{0.11\gamma^2 + (3.52k_BT)^2}$, where the linewidth $\gamma = \gamma_c + \gamma_d + \gamma_n$ is the charge and discharge transfer rates between the trap and transistor terminals and the coupling of the trap to a classical noise source. Here, $FWHM_E$ is in units of energy instead of voltage. For the case of $k_BT \gg \gamma$, $FWHM_E = 0.675\gamma + 3.52k_BT$, indicating that the y-axis crossing of the previous FWHM in voltage is directly related to the γ. Using the y-axis crossing from the linear fit in Fig. 4.7(c), we calculated the γ for step 1 and 2 as 0.025 and 0.22, respectively. Then we solved for electron temperature from $FWHM_E$ and plot them in Fig. 4.7(d). As shown, the electron temperature at 4 K and below tracks the ambient temperature fairly well, until hitting the limit of broadening at around 0.5 K. The same analysis yielded similarly low electron temperature, as expected from the sharp "steps" present in the collector in sample 2. The exact values of electron temperature will be less accurate for collector current because the tunneling process is more complicated (can occur between the trap and the base or between the trap and the collector). However, the electron temperature in the collector is still well below 40 K.

The fact that the extracted electron temperature is well below 40 K has an important implication. As discussed before, the loss of temperature dependence in collector current below 40 K was attributed to quasi-ballistic transport. The quasi-ballistic transport should be initiated by "hot" electrons (much higher than ambient temperature). However, the electron temperature of these "steps" tracks the ambient temperature well below 40 K. Since all transport mechanisms initiated from emitter should occur in parallel, it is very unlikely that one transport mechanism will have a much higher electron temperature than another simultaneous mechanism from the source. This suggests that some portions of current previously considered to be quasi-ballistic are explanable even with "cold" electrons. Such mechanism is likely in the form of tunneling. It is known that trap-assisted tunneling assumes a different slope compared to quasi-ballistic transport [43, 11]. Therefore, we want to evaluate whether direct tunneling is plausible because as explained in Chap. 3 the effective base

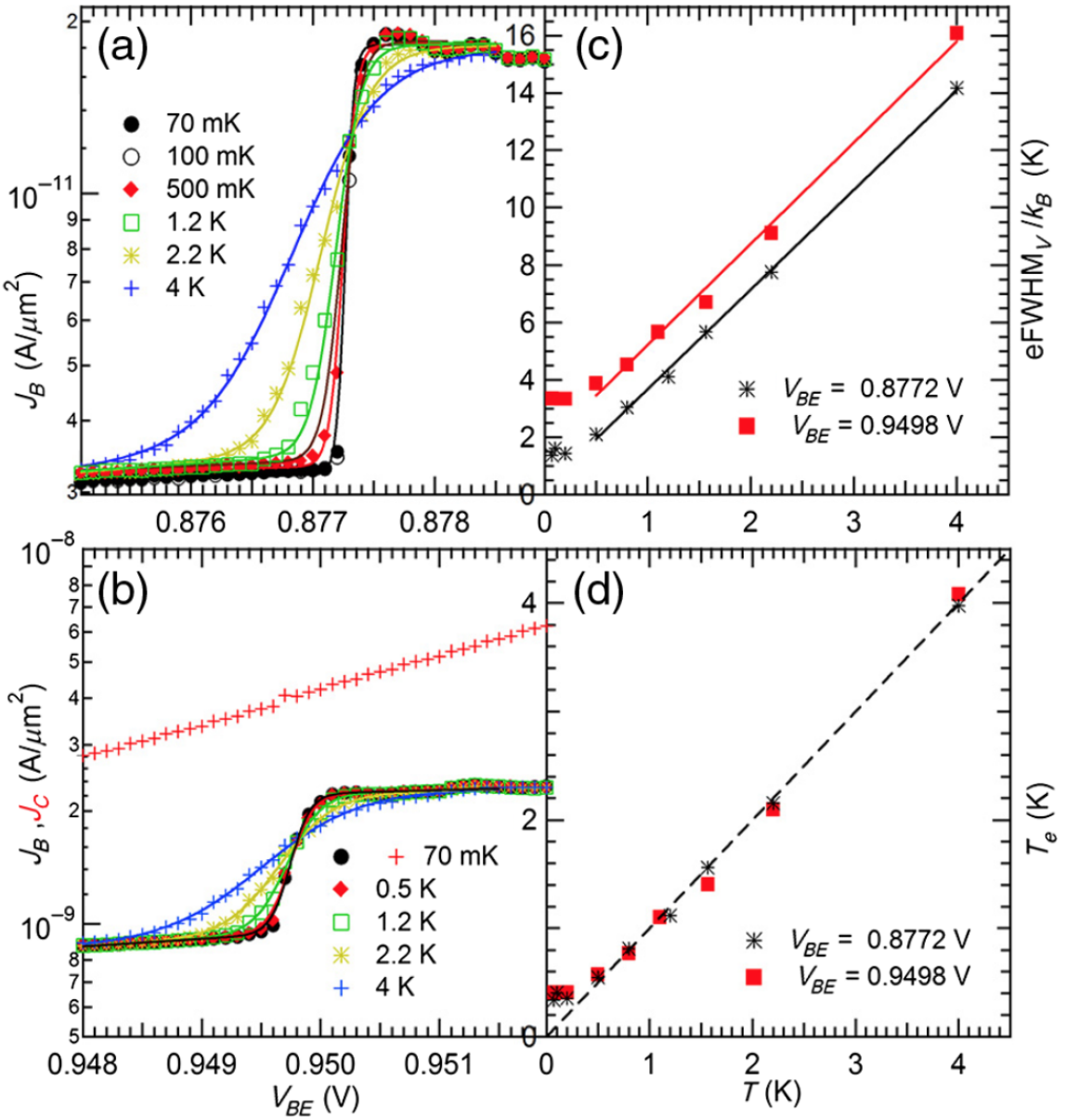

Figure 4.7: (a)-(b) The Fermi function fitted to the base current at $V_{BE} \approx 0.87\ V$ and $V_{BE} \approx 0.95\ V$ in sample 2. (c) Extracted FWHM of the conductance peak versus temperature, showing the linear relationship of FHWM and temperature until below about 0.5 K. The red curve is from (a), or step 1, while the black curve is from (b), or step 2. (d) The same data from (c) but corrected for broadening.

barrier width is reduced at high V_{BE}, and the probability of transmission could become non-negligible. In Sec. 4.4, we will conduct quantitative analysis to analyze whether direct tunneling is present.

4.3.2 Discussions of "Steps"

There are a couple of notes before we go to the analysis of direct tunneling. First, these "steps" are only observed in one early version of the fourth generation devices. The later version does not show any discontinuity, but a smooth I-V curve. However, all versions of devices show characteristics of tunneling, as explained in Sec. 4.6. Due to the variations in process technology between the early and the later versions (including a change of epitaxy tool, improved trap density control, etc), it is difficult to conclude what was causing the discrete steps. Similar "steps" were frequently observed in the transport spectroscopy of tunneling through individual donors in silicon transistors [53]. Therefore, some steps could be due to the unintentionally diffused donor into the neutral base. The exact nature of the "steps", however, require further study. Despite the difference in process tools, the device structure and doping profile from early to later versions is largely similar. Therefore, the potential profile used in the following quantitative analysis is applicable to both the early and the later versions of the fourth generation devices. In addition, from the perspective of circuit design, all steps except one observed in this experiment (in all 3 samples) are below unity current gain (β). Typically, circuits require a β of 100 to be practical. Therefore, the steps lie outside the bias range of practical circuits.

Now that we have analyzed the step, it is interesting to infer the type of defects that could cause such "steps". From the energy perspective, a Gummel sweep essentially uses the quasi-Fermi energy of emitter to screen through defect levels in the base. Therefore, we can relate the energy level to the base-emitter voltage. This is only valid at low temperature, when the Fermi distribution is quite narrow, and the density of states at the band edge can be assumed to be constant. From the Gummel voltage at which the "steps" occur, we can

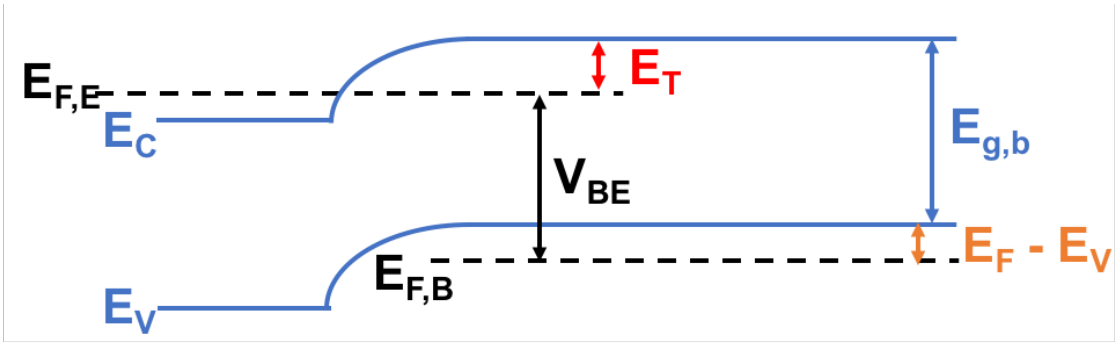

Figure 4.8: Band diagram illustration of defect energy level with respect to quasi-Fermi energy of emitter, quasi-Fermi energy of base, and band gap.

calculate the defect energy level relative to the conduction band in the base as shown in Fig. 4.8. Clearly, defect energy with respect to base conduction band is related to bandgap and V_{BE} as

$$E_C - E_T = E_{g,b} + (E_F - E_V) - eV_{BE}$$

For $E_{g,b}$, we need to consider the bandgap narrowing due to the heavy doping and Ge incorporation. The bandgap for $1 \times 10^{20} cm^{-3}$ of p-type doping is about 1.14 eV at 20 K [69]. With an average of 15% Ge in the base, the bandgap narrowing is about 0.11 eV [14]. The Fermi level with respect to base conduction band can be calculated from Joyce and Dixon approximation, which gives about 50 meV for $1 \times 10^{20} cm^{-3}$ doping [70]. Therefore, the defect energy can be calculated as

$$E_C - E_T = E_{g,b} + (E_F - E_V) - eV_{BE} = 1.14 + 0.05 - eV_{BE}$$

Note that such calculation of energy level is a crude estimate to the actual defect level, because the Ge mole fraction and the doping in the base is nonuniform, which results in a position-dependent values of $E_{g,b}$ and E_F. Therefore, cautions should be taken not to over-interpret the data.

We can count the number of "steps" in each Gummel sweep and obtain a histogram of defect energy (calculated above). Shown in Fig. 4.9 and Fig. 4.10 are the histogram

of "steps" that we obtained from the *findpeaks* function in *MATLAB*. The routine extracts the local maximum according the the prominence and width of the peak in conductance. With the small sample size, the result of the algorithm was checked manually to make sure no spurious "steps" are found. As shown, the defect levels for forward and inverse mode mostly centers at 150-160 meV below the conduction band, which means they are shallow traps. Given that the base Ge mole fraction could be in the range of $0 - 27\%$, the defect level obtained from the histogram has a large uncertainty, and could be in the range of $0 - 300\ meV$.

This is quite a large range that covers many types of impurity or defect states. It could be due to impurities like oxygen, nitrogen [71], copper [72], tantalum, iron [73], lead, tellurium [74], erbium [75], among others. In addition, many defects have more than one state of incorporation, meaning that they can have different defect energy levels with different concentrations. For example, iron could show up as both donor and acceptor with a wide range of possible energies [73].

To gain further information on the defect, we can estimate the volume density of the defect as the number of "step" occurrence in the Gummel divided by the active device volume. The volume is estimated as the emitter area times 10 nm of thickness. The choice of thickness is based on the knowledge that the rate of direct tunneling decreases significantly if the barrier width is more than 10 nm. Since each Gummel sweep from each Gummel predicts a particular density, we plot in Fig. 4.11 the density obtained from all samples in both forward and inverse mode. Surprisingly, the density is in a similar range around $3 \times 10^{15}/cm^3$. This is a reasonably low concentration compared to intentional doping (even in collector the doping is above $1 \times 10^{18}/cm^3$, but higher than the background impurity in a bare silicon wafer. Therefore, it is likely that in this particular batch of fabrication, wafers are contaminated during the processing but does not have high enough leakage current at room temperature to warrant a quality rejection.

The forward and inverse Gummel comparison shows that inverse mode in general has

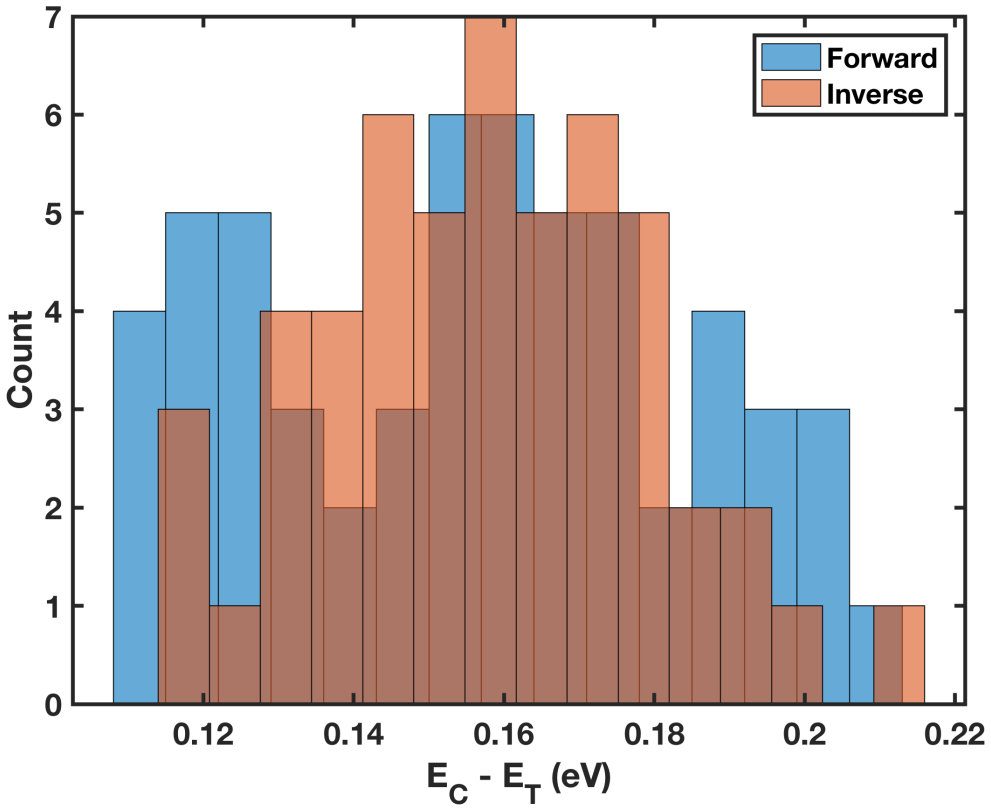

Figure 4.9: Histogram of trap levels versus the energy inferred from base current of forward and inverse Gummel.

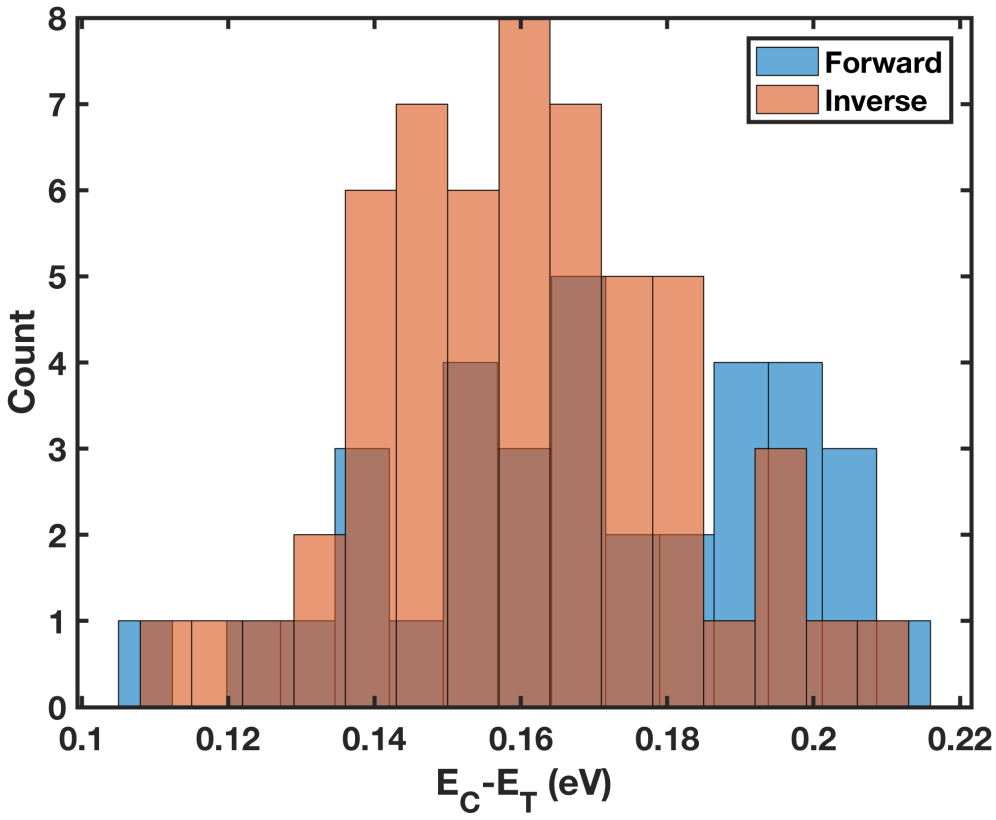

Figure 4.10: Histogram of trap levels versus the energy inferred from collector current of forward and emitter current of inverse Gummel.

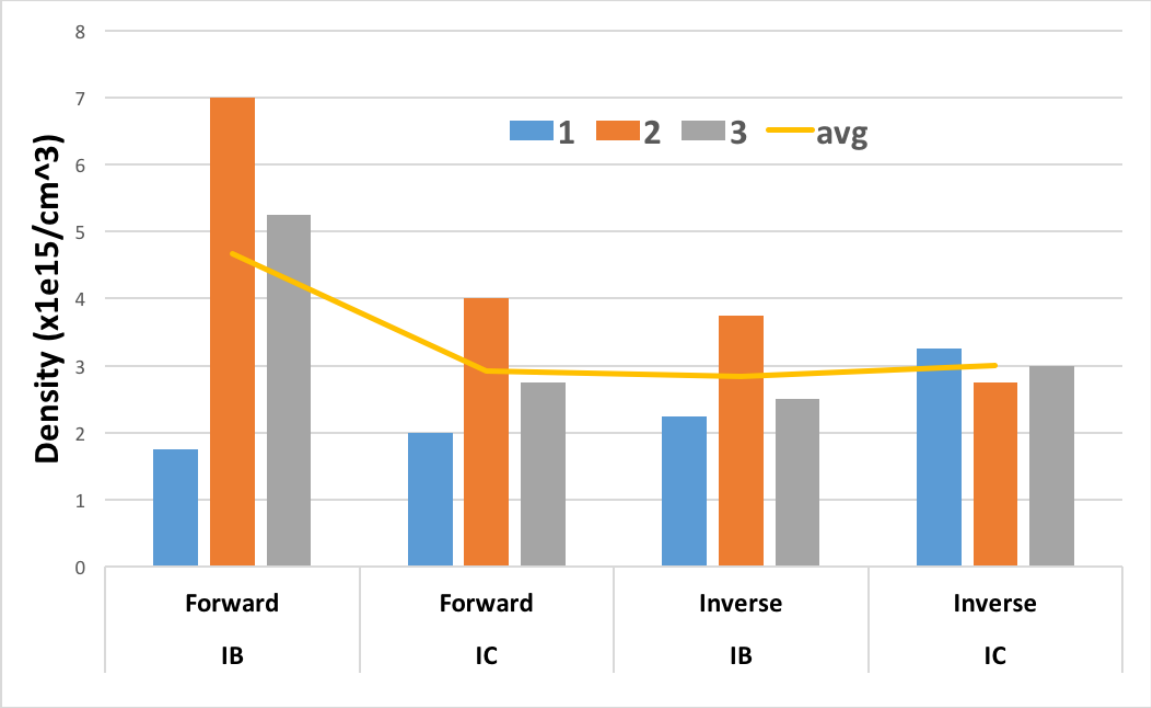

Figure 4.11: Histogram of volume defect density.

more "steps", which means the defects are located closer to the collector than the emitter. This could indicate that the defects originate from the epitaxial growth process in the base. However, as emphasized before, cautions should be taken in interpreting these inference because of the uncertainty from relating "steps" to defect energy.

The "step" non-ideality is very stable against thermal cycling and bias. It can be reproduced at the same V_{BE} even after thermal cycling the device from 70 mK to 300 K and back to 70 mK. This is different from the defects caused by poor interface passivation, which is very unstable and could easily be affected by current or temperature cycling.

4.4 Direct Tunneling in Collector

In this section, we analyze the direct tunneling quantitatively and suggest that it is plausible to have direct tunneling in the collector transport.

In quantum mechanical tunneling, electrons (or more generally, charge carriers) have a finite probability to go through a potential barrier even if the energy (E) of the electrons is

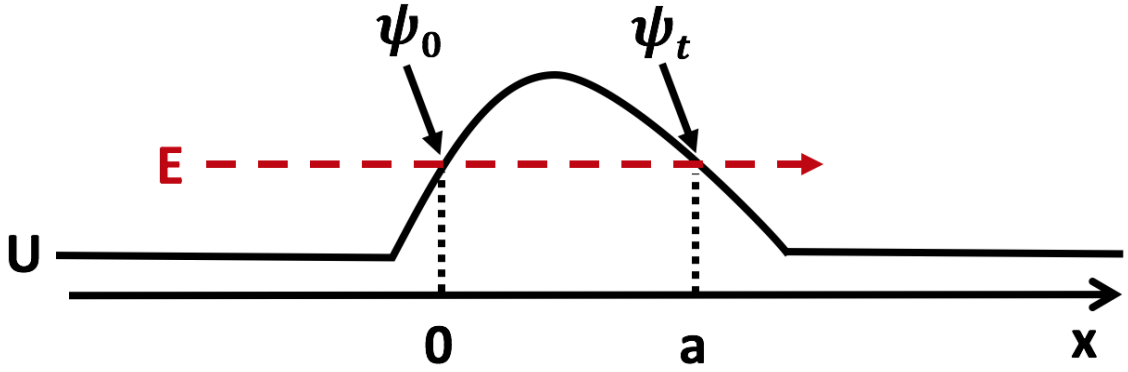

Figure 4.12: Qualitative illustration of an electron with energy E tunneling through a potential barrier U. The distance of the tunneling path under the barrier is a, and the wave function in the incidental and the output side is ψ_0 and ψ_t, respectively.

smaller than the potential energy (U).

Classically, this is forbidden because electrons do not have sufficient kinetic energy to overcome the barrier. Quantum mechanically, however, electrons behave like waves. The wave function decays exponentially in the region where the energy of the electron is smaller than the potential barrier. Such penetration of the wave function into the potential barrier denotes that electrons have a finite probability to be found within the classically forbidden region. If the potential barrier is thin, the exponential decay of the wave function may still result in a nonzero value coming out of the other end of the potential barrier, which means the electrons can go, or "tunnel", through the potential barrier.

We will calculate the tunneling probability and the tunneling current in the following sections. It should be emphasized early on that the calculation of tunneling current is not trivial. As a result, we use many approximations along the way to aid the analysis and the simulator implementation. We acknowledge that some assumptions, though reasonable, does not have experimental proof. In those cases, we try to state them explicitly so that the theory can be examined in the future when the experimental data are available.

4.4.1 Tunneling Probability and Current

From the WKB approximation, the tunneling current is proportional to the transmission probability $T = \frac{|\psi_t|^2}{|\psi_0|^2}$, where ψ_t and ψ_0 is the envelope wave function at the right and the left side of the barrier, as shown in Fig. 4.12. The wave function ψ on either side of the barrier is solved to be

$$\psi_t = \psi_0 e^{-\int_0^a \sqrt{2m(U(x)-E)/\hbar^2}dx} \tag{4.9}$$

Therefore, the transmission probability is proportional to

$$e^{-2\int_0^a \sqrt{2m(U(x)-E)/\hbar^2}dx} \tag{4.10}$$

To calculate the transmission probability, the band profile of a realistic SiGe HBT was obtained from the TCAD simulation at 300 K. The potential barrier (E_C) was calculated self-consistently as a function of the base, emitter, and collector chemical potentials and local current density from hydrodynamic model. Based on the result of measurement data, as in Sec. 4.6, tunneling current only occurs at low injection, at which point the band profile is not affected by the current. To calculate the current density, simulations were set up in the Synopsys Sentaurus TCAD suite. The device structure was similar to those given in Chap. 2 and the doping profile was given in Sec. 2.2.1. The transistors are biased such that the base terminal is grounded, the collector voltage is fixed, and the emitter voltage is swept, identical to the biasing scheme used in the actual measurement.

The TCAD simulation to obtain the potential profile was performed at 300 K. The band diagram in the emitter and base was found to vary minimally from 300 K to 40 K, as expected, because the doping in the emitter and base are above the Mott transition, not subject to freeze-out effect [37].

4.5 Tunneling Model and Measurement

Two implementations of the tunneling model is discussed in this thesis. In this section, calculations will be performed on the internal 1-D transistor to assess the likelihood of direct tunneling in SiGe HBTs. The calculation shows that direct tunneling is very likely to exist. Based on this conclusion, Chap. 5 introduces a non-local tunneling model set up in a 2-D TCAD model. Since both methods used the potential profile obtained from the TCAD simulation and tunneling is mostly dependent on the potential profile, they showed remarkable agreement.

The form of tunneling current density (A/m^2) in a position dependent barrier is derived by Floyd and Walmsley [76] as

$$J(V) = \frac{4\pi me}{h^3} \int_0^{E_F - eV} W_1 P(E_z, V) dE_z + \frac{4\pi me}{h^3} \int_{E_F - eV}^{E_F} W_2 P(E_z, V) dE_z \qquad (4.11)$$

where E_F is the Fermi energy at the left electrode, $E_F - eV$ is the Fermi energy at the right electrode under bias, and E_z is the energy of the electron. During the derivation, Fermi distribution is assumed. At $T = 0\ K$, the Fermi distribution becomes a step function, and $W_1 = eV$ and $W_2 = E_F - E_z$ in Equ. 4.11. At deep cryogenic temperatures of a few K, such approximation is valid. The two terms in Equ. 4.11 correspond to the two summations for the transverse momentum, as illustrated in Fig. 4.13. The spherical shell lying between E_F and $E_F - eV$ is the filled states from which electrons in the left electrode may tunnel to unfilled states in the final electrode when a bias V is applied. Depending on the magnitude of the incidental momentum k_z (perpendicular to the tunneling interface), the summation of the transverse momentum k_t (parallel to the tunneling interface) can be an annulus (Fig. 4.13(a)) or a disk (Fig. 4.13(b)).

In SiGe HBTs under consideration, the emitter is the aforementioned left electrode while the collector is the right electrode. In a Gummel measurement, base and collector

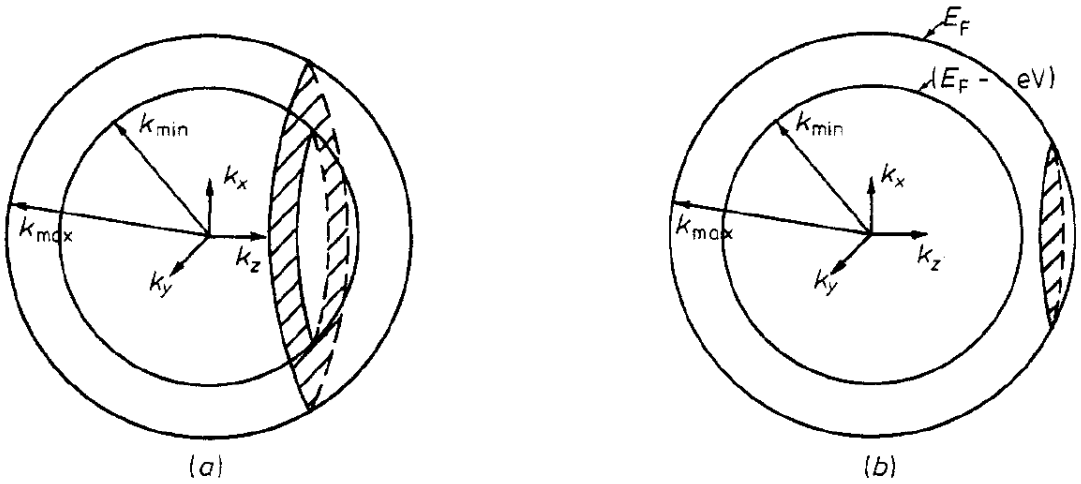

Figure 4.13: Graphical illustration of two cases of $\sum_{k_t}$, the summation of all momentum k transverse to the tunneling direction z (after [76]). The summation can be an annulus (a), or a disk (b), depending on the magnitude of the wave vector k_z.

are tied to the same potential while the emitter-base junction is forward biased at V_{BE}. Therefore, E_F in Equ. 4.11 is the emitter quasi-Fermi energy (for electrons in NPN transistor) while $E_F - eV = E_F - eV_{BE}$ is the collector quasi-Fermi energy for electrons. In heavily doped n-type Si, E_F is comparable to the bandgap energy of silicon, or 1.1 eV. For $V_{BE} \approx 1\ V$, $E_F - eV_{BE} \approx 0.1\ eV$ is much smaller than E_F. Therefore, the inner sphere is much smaller than the outer one in Fig. 4.13 and the first term in Equ. 4.11 can be neglected compared to the second term. If $T \neq 0\ K$, $W_2 = k_B T ln(1 + e^{(E_F - E_z)/k_B T})$ is the Fermi-Dirac integral for energy states transverse to the direction of tunneling. Based on the above approximations, the equation becomes

$$J(V_{BE}) = \frac{4\pi me}{h^3} \int_{E_F - eV_{BE}}^{E_F} W_2 P(E_z, V_{BE}) dE_z \tag{4.12}$$

A pre-factor of c is added to account for the effect of band structure on the density of states electron mass. $c = 4\sqrt{ab}$, where $a = 0.98$ and $b = 0.18$ are the ratios of the longitudinal and transverse effective electron masses over the electron mass, respectively. The prefactor 4 accounts for valley degeneracy. It is supposed that the higher-energy valleys will not contribute to the tunneling current. In addition, the energy of electrons is typically

referenced to the conduction band edge E_C. Since there are no energy state E_z within the bandgap of emitter, the range of integration from $E_F - eV_{BE}$ to E_F is actually only valid from E_C to E_F. Here, we assume a sharp conduction band and the effect of band tails on the transport is negligible, because the electrons contributing to the tunneling only come from the vicinity of the Fermi energy [77] . By using the notation E'_F, E'_z to denote the energy relative to the conduction band edge, we can change the limit of the integral to

$$J(V_{BE}) = \frac{4c\pi me}{h^3} \int_0^{E'_F} W_2(E') P(E'_z, V_{BE}) dE_z \tag{4.13}$$

where $E'_z = \hbar^2 k_z^2 / 2m$, m is the free-electron mass, and $P(E'_z, V_{BE})$ is the tunneling probability.

Using the WKB approximation for the transmission probability as in Equ. 4.10, we find

$$P(E'_z, V_{BE}) = e^{-2\int \sqrt{(2mb/\hbar^2)[E_C(z) - eV_{BE} + E'_F - E'_z}} dz \tag{4.14}$$

The relationship between $E_C(z)$, eV_{BE}, E_F, E'_F, E_z and E'_z are shown graphically in Fig. 4.14. The integral is taken over the region where the square root is real.

E_C was obtained from TCAD simulations of a fourth generation SiGe HBT device, and the tunneling current was calculated and compared to the measurement. The comparison is plotted in Fig. 4.15(a). Although the calculation looks different than the data in terms of slope, the magnitude of the current is actually remarkably close given that no fitting parameters are used and only reasonably approximations have been made to obtain the tunneling current with a realistic SiGe HBT profile. Due to the exponential dependence of tunneling current on the detailed shape of potential profile, a slight deviation of the potential profile from that in the actual SiGe HBT sample will change the current exponentially. Inspecting Fig. 4.15(a) in a different way, we found that for a given collector current density, the difference between the measured and calculated base-emitter voltage varies less than 40 mV, within 5% of the actual V_{BE}. In other words, the barrier height could be

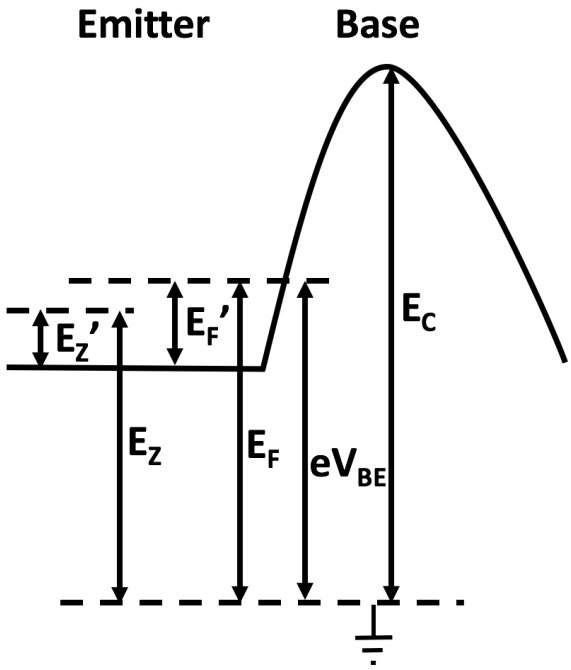

Figure 4.14: Graphical illustration of $E_C(z)$, eV_{BE}, E_F, and E_z used in Equ. 4.13 and Equ. 4.14 for the tunneling probability P in a SiGe HBT.

only 5% off despite the seeming discrepancy. It was found that the change of base width can change the simulation result dramatically, as shown in Fig. 4.15(b). For a particular bias voltage, a larger base width increases the slope and decreases the magnitude of the current. Considering the TCAD model is not exactly identical to the real device, such a small variation proves that direct tunneling is very likely the actual mechanism. The same mechanism was used to explain the measurement data in a different SiGe technology, suggesting the mechanism is prevalent in advanced SiGe HBTs [78].

4.6 Direct Tunneling versus Quasi-ballistic Transport

So far, it has been shown that a portion of collector current previously thought of as quasi-ballistic transport is due to a different mechanism with "cold" electrons. Tunneling is very likely to be this mechanism. However, it would be useful to provide more experimental evidence of tunneling. In particular, it would be useful to show that direct tunneling and

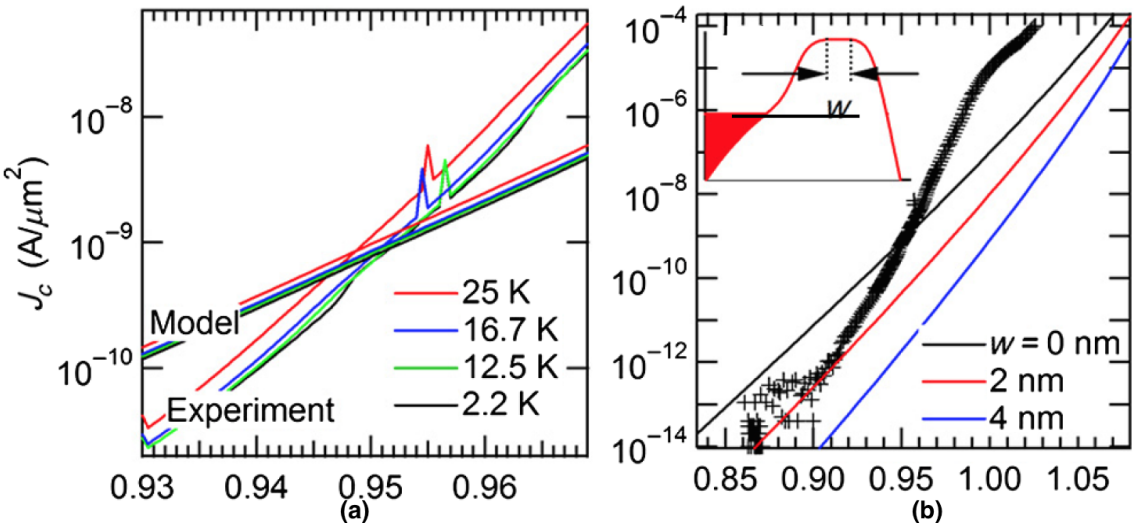

Figure 4.15: (a) The calculated direct tunneling current density (J_C) and measured J_C versus base-emitter voltage (V_{BE}) for the transistor operating in the forward active mode. The spikes in the measurement are single-point anomaly, likely due to the ranging circuit of the instrumentation. (b) The calculated tunneling current for various base widths. A larger base width decreases the tunneling current but increases the slope of the current.

quasi-ballistic transport can be differentiated in the measurement. To achieve this, we measured three generations of SiGe HBTs and compared their characteristics at cryogenic temperatures. Since the data of devices from the same generation are similar, only one representative device from each generation is showcased here.

The collector current density (J_C) versus V_{BE} of SiGe HBTs from three technology generations across temperature is shown in Fig. 4.16. At high temperatures, the current is linear on a log scale (exponential) as drift-diffusion transport dictates. At both 1.8 K and 18 K, a non-ideal current at low injection is observed for all generations. In the 1^{st} generation device, the non-ideal current below 10 nA/μm^2 was shown to be driven by trap-assisted tunneling (TAT) [11]. Since a similar slope and location of the non-ideal current is observed in the 3^{rd} and the 4^{th} generation, it is plausible that the same TAT mechanism is present in the more advanced generations. In the present investigation, the region with a smaller slope, circled in Fig. 4.16, is assumed to be due to a TAT mechanism. Chap. 5 will discuss the effects of scaling more in depth. For now, the main question to address is under which conditions the other mechanisms, namely direct tunneling and

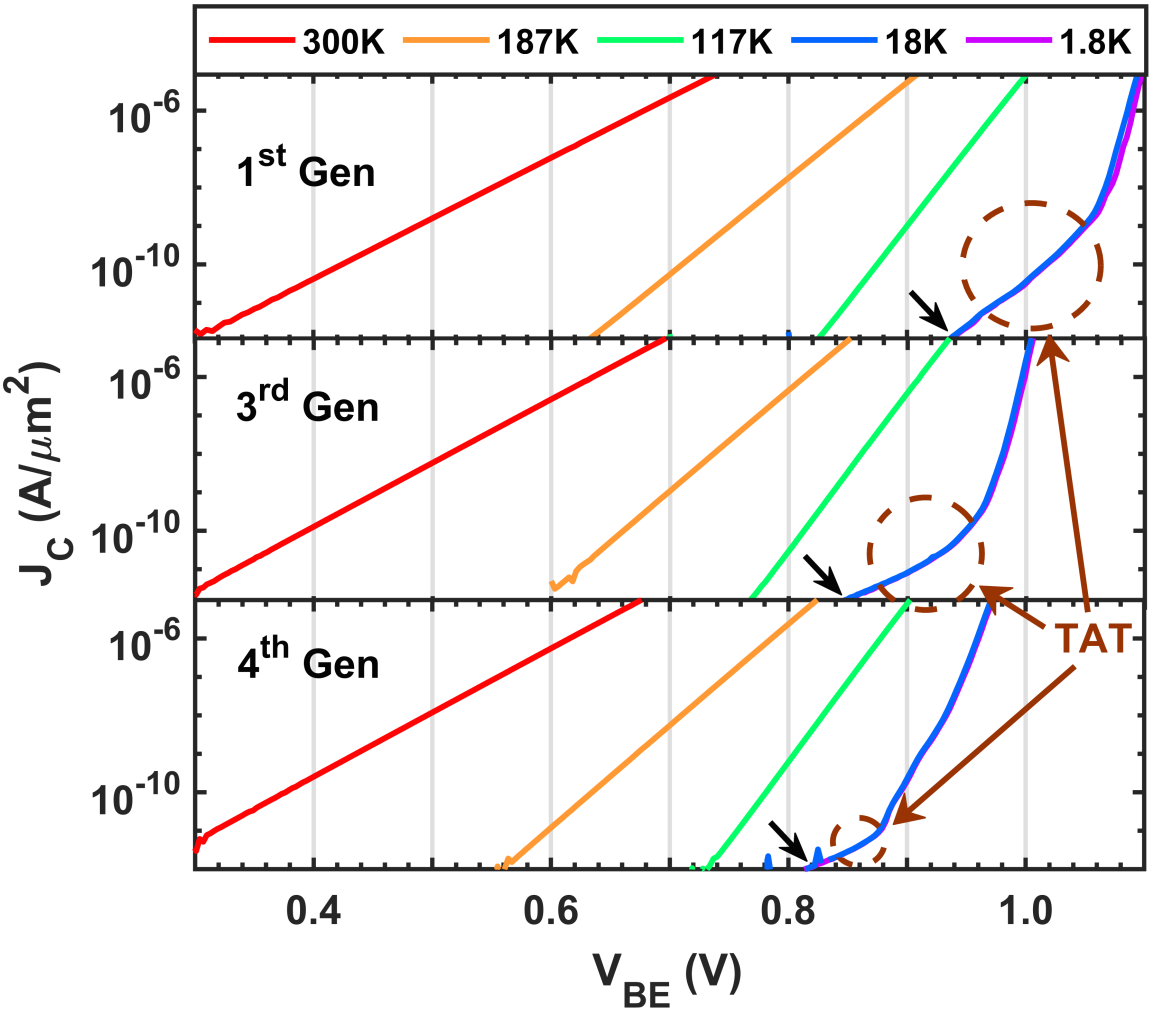

Figure 4.16: Measured collector current density versus V_{BE} of all three generations from 300 K to 1.8 K. Below 18 K, the curves overlap. The trap-assisted tunneling region is circled, with its onset marked with arrows.

quasi-ballistic transport, exist, and more importantly, how to differentiate between them in the measurement.

To distinguish between the two, it is recognized that quasi-ballistic transport should be only weakly dependent on the base width. In quasi-ballistic transport, the majority of carriers travel across the base without scattering. Therefore, a small change in base width does not change the already small scattering rate for carriers. In other words, the collector current component from quasi-ballistic transport should be invariant to changes in the base width. On the other hand, direct tunneling current is proportional to the transmission

probability P for a rectangular base barrier, according to [66]

$$P \propto e^{-2W\sqrt{2m(U-E)/\hbar^2}} \tag{4.15}$$

where m is the effective mass of the electron, U is the potential energy, E is the energy of the electron ($E < U$ inside the potential barrier), and W is the barrier width. Although the base barrier is not rectangular, the exponential dependence still applies, and a change in barrier width (base width) is expected to result in an exponential change in tunneling current [2, 78]. Therefore, the difference between direct tunneling versus the quasi-ballistic transport can be revealed by measuring collector current for different base widths.

To vary the base width without changing the other characteristics of devices, collector-base voltage V_{CB} is used as a tuning parameter. As shown in Fig. 4.17, increasing V_{CB} shifts the collector conduction band energy (E_C) down, effectively reducing the tunneling barrier width. Similarly, decreasing V_{CB} increases the barrier width. Experimentally, the Gummel characteristics were measured at $V_{CB} = -0.3V$ and $V_{CB} = 0.5V$. As shown in Fig. 4.18, apart from the increase of current at low injection due to band-to-band tunneling in the collector-base junction, the collector current of the 1^{st} generation device is invariant to changes in V_{CB}. However, a portion of the collector current in the 3^{rd} and the 4^{th} generation devices changes with V_{CB}.

To quantify this change in current, the normalized collector current at multiple V_{CB} values is plotted versus collector current density (J_C) in Fig. 4.19. Clearly, in the 3^{rd} and the 4^{th} generation devices, collector current density from 10^{-11} to 10^{-7} $A/\mu m^2$ is very sensitive to V_{CB} and this sensitivity disappears towards higher J_C. In the 1^{st} generation, however, the sensitivity is much smaller throughout the J_C range. The strong sensitivity can be explained by the strong dependence of direct tunneling current on the base width (i.e., V_{CB}), while the weak sensitivity indicates quasi-ballistic transport dominates, which is independent of the base width. In other words, quasi-ballistic transport is present in all generations, but direct tunneling is present only in the 3^{rd} and the 4^{th} generation devices.

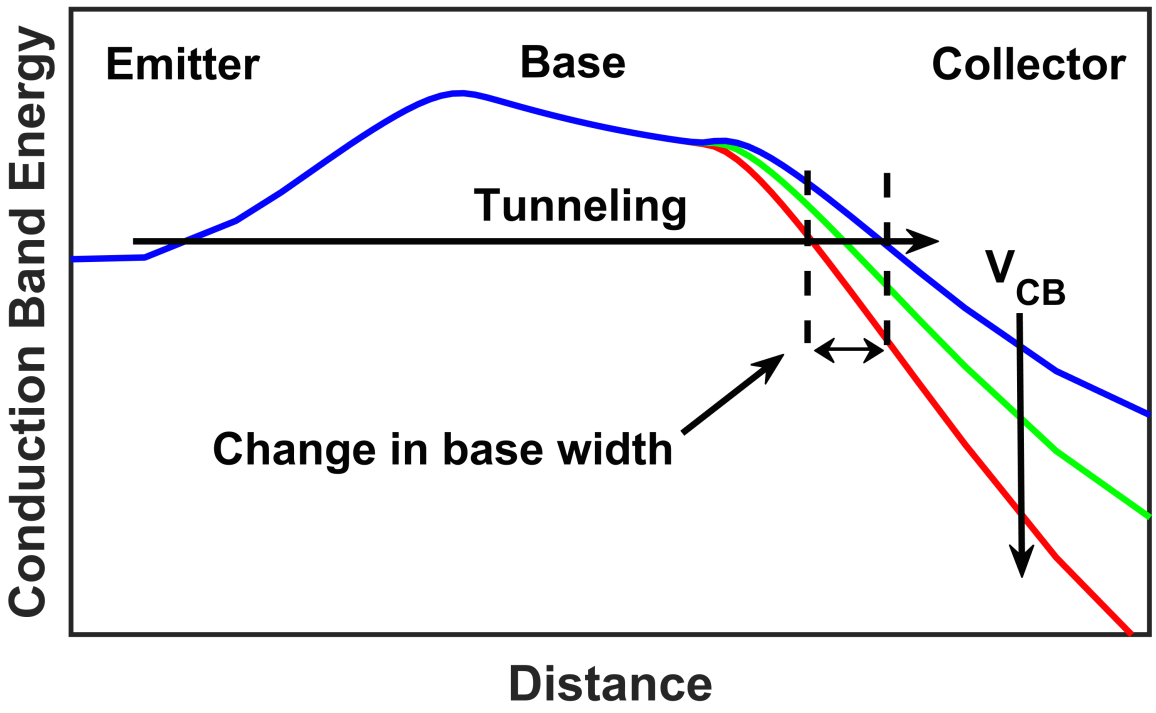

Figure 4.17: Qualitative illustration of the conduction band diagram of SiGe HBTs showing the variation of base barrier width under multiple V_{CB} values.

This makes intuitive sense.

There are two interesting observations to be made. First, it may seem surprising that the ratio of current is larger in the 3^{rd} than the 4^{th} generation device, although the 3^{rd} generation device should have less tunneling due to its larger base width. This can be understood, since the ratio of the current is proportional to the ratio of the tunneling probability, P. This ratio, from (4.15), is proportional to $e^{\Delta W}$, where ΔW is the change in base width. In other words, the ratio is proportional to the absolute change of base width, not its percentage change. Compared to the 4^{th} generation, the collector and base doping are lower in the 3^{rd} generation, causing a larger change in depletion width for the same change in V_{CB}. This causes the effective base width to change more significantly, resulting in a larger change in tunneling current for the 3^{rd} generation, as can be seen in Fig. 4.19. The tunneling current, however, is still larger in the 4^{th} generation if we compare them at a fixed V_{BE} (barrier height), because of its smaller base width. The second observation is that direct tunneling can potentially degrade the device transconductance, g_m, as shown by the smaller slope

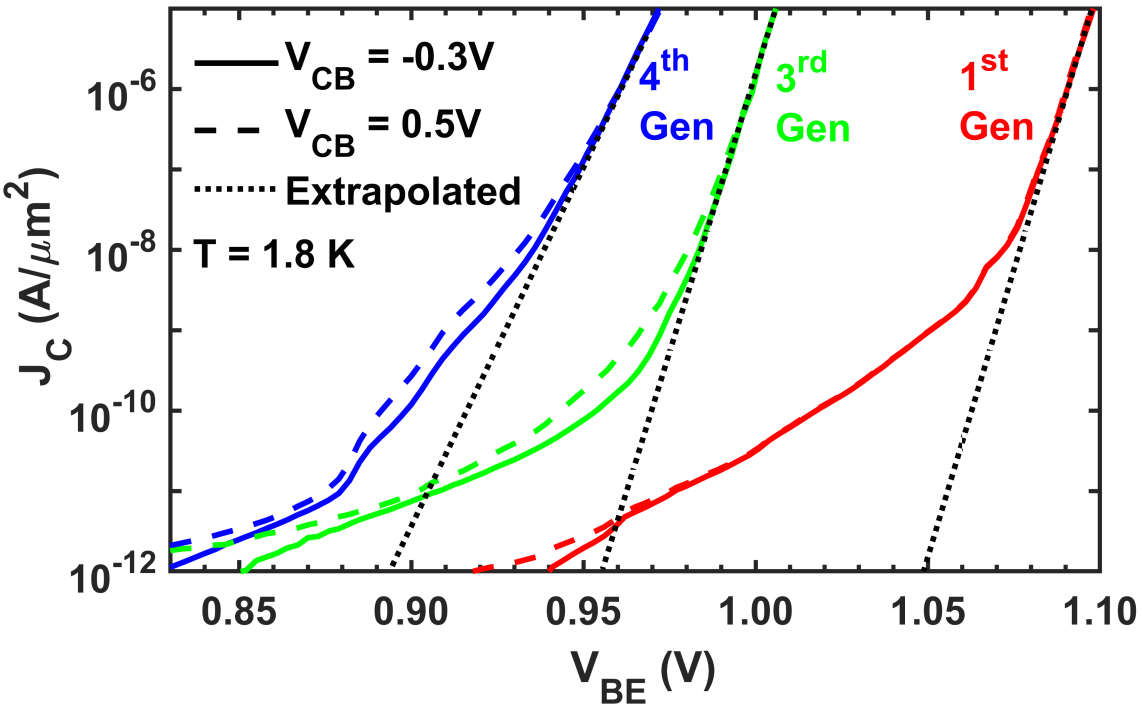

Figure 4.18: Collector current from the Gummel characteristics at 1.8 K for $V_{CB} = -0.3V$ and $0.5V$. Extrapolations of quasi-ballistic current (dotted lines) estimates the turn-on voltage for this mechanism.

of current in Fig. 4.18. Therefore, the presence of direct tunneling requires more careful considerations for cryogenic circuit designs.

To gain more insight into how direct tunneling becomes significant as temperature decreases, the ratio of collector current between $V_{CB} = 0.5V$ and $-0.3V$ is plotted across temperature in Fig. 4.20. At 82 K, the ratio is mostly constant for all generations, indicating that direct tunneling is negligible. As the temperature is lowered, a "hump" is progressively observable in the 3^{rd} and the 4^{th} generation devices, but not in the 1^{st} generation device. Since at a fixed V_{BE}, drift diffusion current decreases with temperature while the tunneling current remains roughly constant with temperature, tunneling becomes the dominant mechanism at low temperatures, thereby increasing the ratio. In particular, the hump due to direct tunneling first shows up in low injection and then slowly encroaches towards high injection as the temperature is lowered. When the drift diffusion current is replaced by quasi-ballistic current, the percentage of tunneling also stops changing, as shown in the 4^{th}

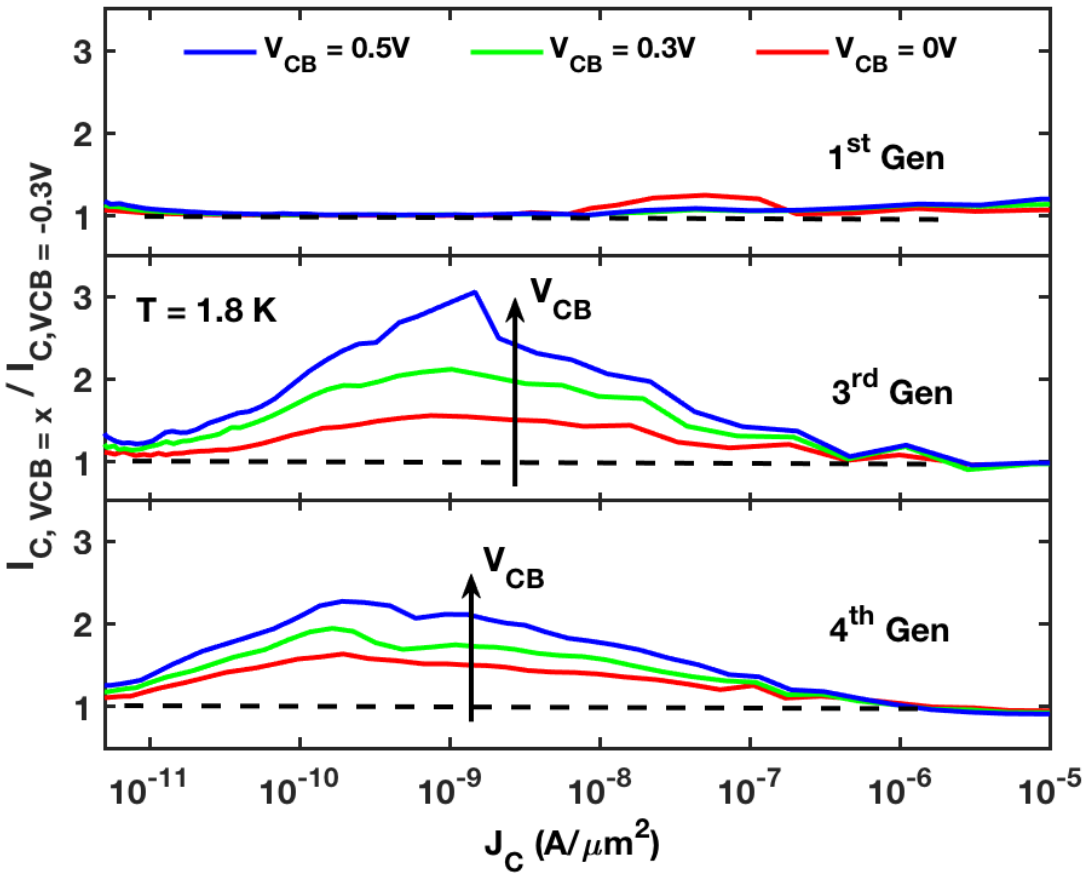

Figure 4.19: The ratio of collector current at various V_{CB} values to the current at $V_{CB} = -0.3V$ at 1.8 K in the first, third, and fourth generation SiGe HBT devices.

generation device. For the 3^{rd} and the 4^{th} generation devices, the hump at 1.8 K corresponds to a $3\times$ and $2\times$ increase, respectively, which is very significant. In comparison, the ratio is constant in the 1^{st} generation, as the base width is too large for any direct tunneling. In short, direct tunneling can be observed in scaled technologies over a finite range of collector currents by simply varying V_{CB}.

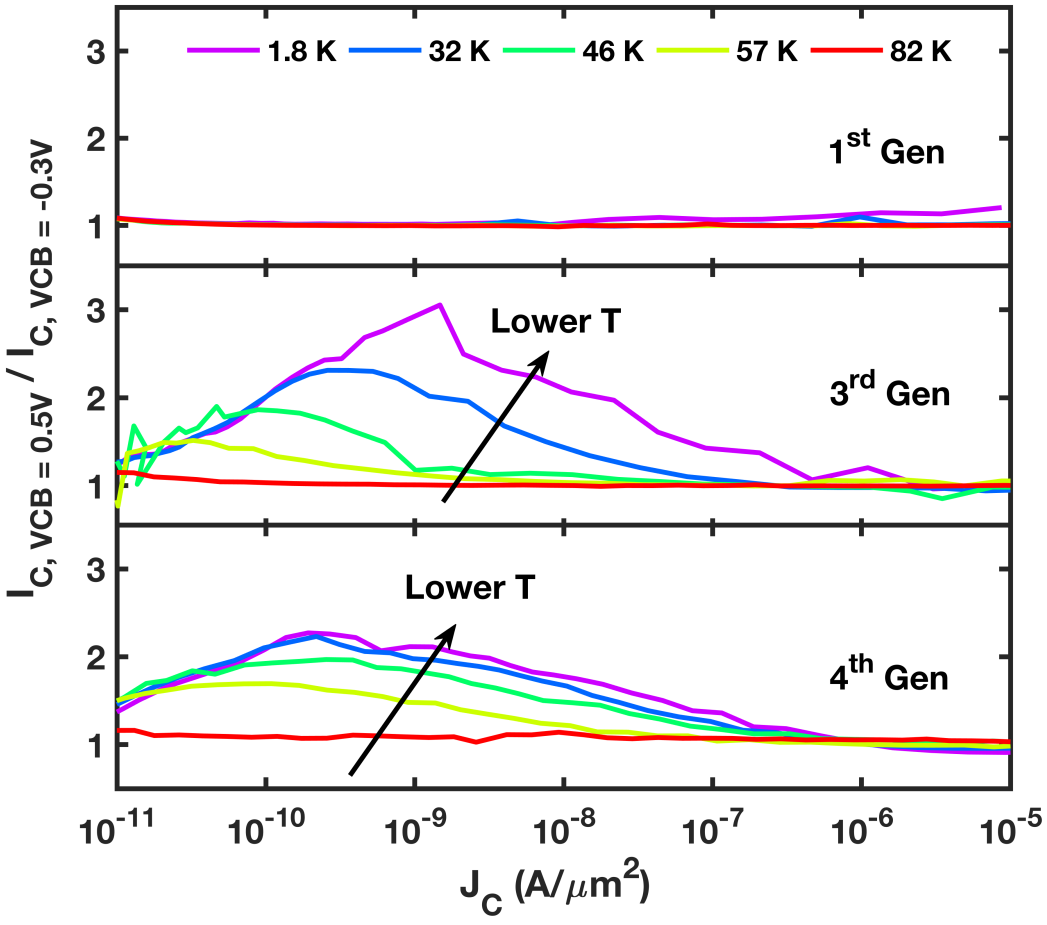

Figure 4.20: The ratio of collector current under $V_{CB} = 0.5V$ and $V_{CB} = -0.3V$ from 82 K to 1.8 K in each generation.

CHAPTER 5
EFFECT OF SCALING ON TRANSPORT MECHANISMS

5.1 Qualitative Discussions

As SiGe HBT technology scales, there is a consequent decrease in base width, increase in collector doping, increase in Ge mole fraction, and increase in base doping. Aside from the increase in base doping, which raises the base potential barrier, all the other modifications will lower the base potential barrier. A lower barrier enhances the direct tunneling mechanism, which means if quasi-ballistic transport is not increased at the same rate, direct tunneling will become significant.

Quasi-ballistic transport is mostly dependent on the base barrier height. This can be understood by recognizing that most electrons do not have enough energy to overcome the base barrier at cryogenic temperatures until the barrier height is close to zero, which is when the emitter quasi-Fermi level is similar in height to the base conduction band edge. The base barrier height is directly dependent on the integrated Ge content, which is indirectly related to the peak Ge content and the width of the Ge profile. As seen from the extrapolated line in Fig. 4.18, the onset of the quasi-ballistic transport shifts towards smaller V_{BE} as technology scales, because the peak Ge content increases in the more advanced generations. To decrease the turn-on voltage of the quasi-ballistic transport, a larger Ge mole fraction is required. However, a larger Ge mole fraction requires a thinner base to maintain Ge film stability [14]. Additionally, technology scaling targets improved high frequency operation at room temperature, and typically does not exceed a 30% peak Ge content. Instead one must also shrink the base width using decreased thermal cycles to achieve the improved performance [27]. Therefore, the onset of quasi-ballistic transport is likely fixed if room temperature scaling rules are followed. The reduced base width, however, will increase the

direct tunneling current relative to quasi-ballistic current. Such increase will be visible until quasi-ballistic current eventually rises above the direct tunneling current. In other words, direct tunneling is expected to dominate the collector current up to a higher V_{BE} in more scaled technology generations.

It is difficult to predict how the trap-assisted tunneling will change with technology scaling, since TAT, which relies on the presence of traps, depends on technology sensitive process steps, such as epitaxial growth conditions and tooling. For example, in [78], no TAT region is observed, which indicates that the specifics of the device structure design and processing conditions can play a major role in eliminating the TAT. In particular, a perimeter versus area (P/A) analysis can give more insight into the physical location of the traps in play. However, a P/A analysis is not meaningful in the present case because the standard device sizes in these highly scaled technologies allow for only a small range of P/A ratio (their emitter stripe width is fixed at minimum geometry and cannot be altered). That said, we can still estimate the energy of the trap levels based on the onset of the TAT from the Gummel characteristics. As illustrated in Fig. 3.4, quasi-ballistic transport occurs when the emitter quasi-Fermi level is close to the base conduction band edge. Since the onset of trap-assisted tunneling occurs earlier than the onset of the quasi-ballistic transport, the trap states must be located below the base conduction band. From Fig. 4.16, the onset of trap-assisted tunneling is 0.87 V, 0.82 V, and 0.81 V for the 1^{st}, 3^{rd}, and 4^{th} generation devices, respectively. From Fig. 4.18, the extrapolated onset of quasi-ballistic transport is about 1.05 V, 0.96 V, and 0.90 V. Therefore, TAT occurs about 100-200 mV before the onset of quasi-ballistic transport, indicating that the traps are likely located within 100-200 meV from the base conduction band edge. The good news here is that, as shown in Fig. 4.18, in the 4^{th} generation devices, the TAT leakage current extends only to about 10 pA/um^2, much lower than in the 1^{st} generation devices, where it extends to 10 nA/um^2. This may be the result of both improved process technology and film tooling (yielding reduced trap density) and, equally importantly, an earlier onset of direct tunneling (as discussed

Table 5.1: Summary of Transport Mechanisms

	Trap-assisted Tunneling	Direct Tunneling	Drift-diffusion/Quasi-ballistic
Key Process Dependence	base trap/defect density	base width	Ge mole fraction
Region of Significance	low injection	low-medium injection	high-injection
g_m	small	medium	large

above). From the data, the onset of direct tunneling current appears to shift to a smaller V_{BE} from the 1^{st} to the 4^{th} generations, thereby masking the trap-assisted tunneling. It is expected that, with scaling, such trends will continue and the direct tunneling will become more significant, causing the trap-assisted tunneling current to only appear at extremely low injection.

A table that summarizes the characteristics of each transport process is given in Table. 5.1.

5.2 TCAD Simulations

5.2.1 TCAD Simulation Setup

Sentaurus TCAD models (2D) of three SiGe HBT generations were used to investigate how scaling impacts the direct tunneling current. The doping profiles and geometries were calibrated to match on-wafer DC and small-signal AC measurements of the respective measurement. Hydrodynamic transport model was used with the parameter sets calibrated to 300 K measurements for Slotboom bandgap narrowing and Phillips unified mobility models [79, 80].

In addition, the nonlocal tunneling model was enabled to augment the hydrodynamic model [81]. The model was enabled only for electrons in the conduction band because the tunneling in the valence band is negligible. The non-locality of tunneling was incorporated

as a local generation/recombination rate $G(r)$ at location r, as in Equ. 5.1.

$$\begin{aligned} G(r) &= \frac{1}{q}\nabla \cdot J_{Tunnel} \\ &= \frac{1}{q}\frac{dJ_{Tunnel}}{d\phi} \cdot \nabla\phi \\ &= \frac{dJ_{Tunnel}}{-d\varepsilon}(-\xi) \\ &= \frac{dJ_{Tunnel}}{d\varepsilon}\xi \end{aligned} \tag{5.1}$$

$G(r)$ is dependent on the local electrostatic potential ϕ, energy ε, and electric field ξ, respectively [81]. To obtain the tunneling paths, a special-purpose mesh was defined to cover the emitter-base junction, the neutral base, and part of the base-collector junction. The tunneling paths were along the paths in the special-purpose mesh, and are the geometrically shortest line connecting vertices within specified distance to the interface. For this simulation, the distance for tunneling calculation is set as 15 nm to each side of the interface at the base and emitter junction (30 nm in total). The transmission probability $\gamma(u, l, \varepsilon)$ for particles from location l to $u > l$ with energy ε is written as in Equ. 4.10, with the range of integral replaced as l to u. The local recombination rate rate $R(u, l, \varepsilon)$ minus the generation rate $G(u, l, \varepsilon)$ at point u from $l < u$ is given, based on Equ. 5.1, as [82]

$$\begin{aligned} R(u,l,\varepsilon) - G(u,l,\varepsilon) = \\ \frac{A}{qk_B}\vartheta[\varepsilon - E_C(u), -\frac{dE_C}{dx}(u)]\vartheta[\varepsilon - E_C(l), \frac{dE_C}{dx}(l)]\gamma(u,l,\varepsilon) \\ \times [T(u)ln(1 + e^{(E_F(u)-\varepsilon)/k_BT(u)}) - T(l)ln(1 + e^{(E_F(l)-\varepsilon)/k_BT(l)})] \end{aligned} \tag{5.2}$$

where $\vartheta(x, y) = \delta(x)|y|\Theta(y)$. In Equ. 5.2, $A = 4\pi m_e kq/h^3$ is the Richardson constant, $T(r)$ is the electron temperature at location r. In the definition of $\vartheta(x, y)$, $\delta(x)$ is the Kronecker delta function, $|y|$ comes from $\nabla\phi$ in Equ. 5.1, and Θ is the heavy-side step function restricting the shape of the potential barrier to be an incline in the incidental side

and a decline in the output side.

The tunneling current density at point l from $u > l$ is the integral of the generation/recombination rate over energy and distance, as

$$\frac{dj_{Tunnel}}{dl}(l) = -q \sum_{v} \int_{l}^{\infty} \int_{-\infty}^{\infty} [R(u,l,\varepsilon) - G(u,l,\varepsilon)] d\varepsilon du \tag{5.3}$$

The potential profile, the quasi-Fermi level, and the total current density (including the contribution from tunneling) are calculated self-consistently.

Quasi-ballistic transport was not set up in the TCAD due to the lack of physics-based models. For simulations below 50 K, recombination in the base was found to not affect the simulated collector current (as expected for direct tunneling).

5.2.2 TCAD Simulation Studies

Two investigations were conducted using TCAD. The first evaluated the amount of tunneling versus conventional drift-diffusion present in the transport current as the device scales. To evaluate this, the device was simulated with either hydrodynamic (HD) and tunneling models engaged, or with HD model alone. The ratio of collector current with and without tunneling at V_{BE}, corresponding to 10 nA of current from HD model alone, is plotted in Fig. 5.1. As expected, the ratio of current with or without tunneling is constant throughout the temperature range in the 1^{st} generation device because the base width is too large for tunneling processes. For the 3^{rd} and the 4^{th} generation devices, however, the effects of direct tunneling on collector current begin to appear at as high as 180 K in the 4^{th} generation and 130 K in the 3^{rd} generation. Initially, the tunneling is simply a fractional increase compared to the hydrodynamic current. At lower temperatures, however, the hydrodynamic current becomes smaller compared to the direct tunneling current, and the ratio between the two increases dramatically. This can be visually seen in Fig. 5.2, where the collector current from the simulation with and without the tunneling model, as well as the measured data, are plotted across temperature for the 4^{th} generation device. At high temperatures,

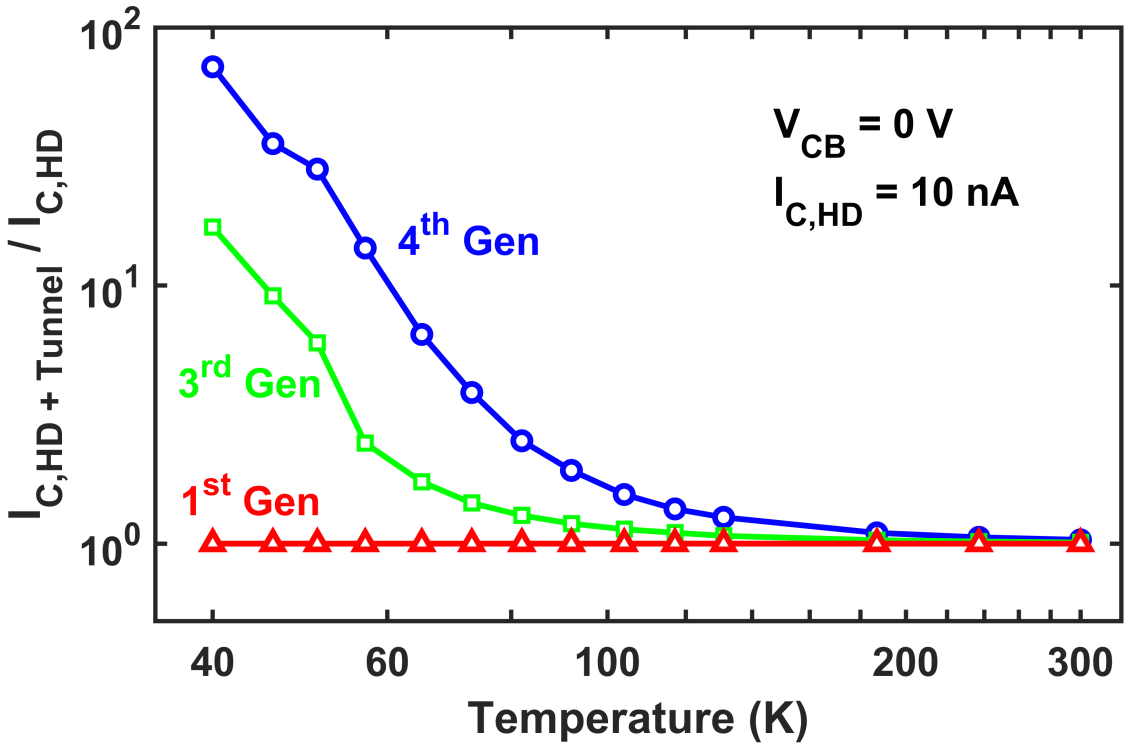

Figure 5.1: The ratio of collector current from hydrodynamic (HD) model with and without the tunneling model. The collector current is extracted from forward Gummel simulation at the V_{BE} corresponding to $I_C = 10\ nA$ in the HD model

the tunneling does not affect the collector current, and both simulations overlap with the data. At low temperatures, however, the simulated current with only the HD model can be as much as 2 to 4 orders of magnitude smaller than measured data! The HD model with direct tunneling, on the other hand, still models the data fairly well until a divergence above $1\ \mu A/um^2$ at low temperatures, as shown in Fig. 5.2 for 46 K. The inaccuracies at high currents is due to the absence of a quasi-ballistic transport model, which is necessary to explain the continuous exponential increase in the measured current.

The second investigation evaluates the sensitivity of tunneling to the technology process parameters that can modify the base barrier shape. Four process parameters were chosen; namely, peak base doping, peak doping of the selectively implanted collector (SIC), peak Ge, and base width. In addition, a scenario where the base width is reduced while the base doping is increased by the same factor (to keep constant integrated base charge) was also

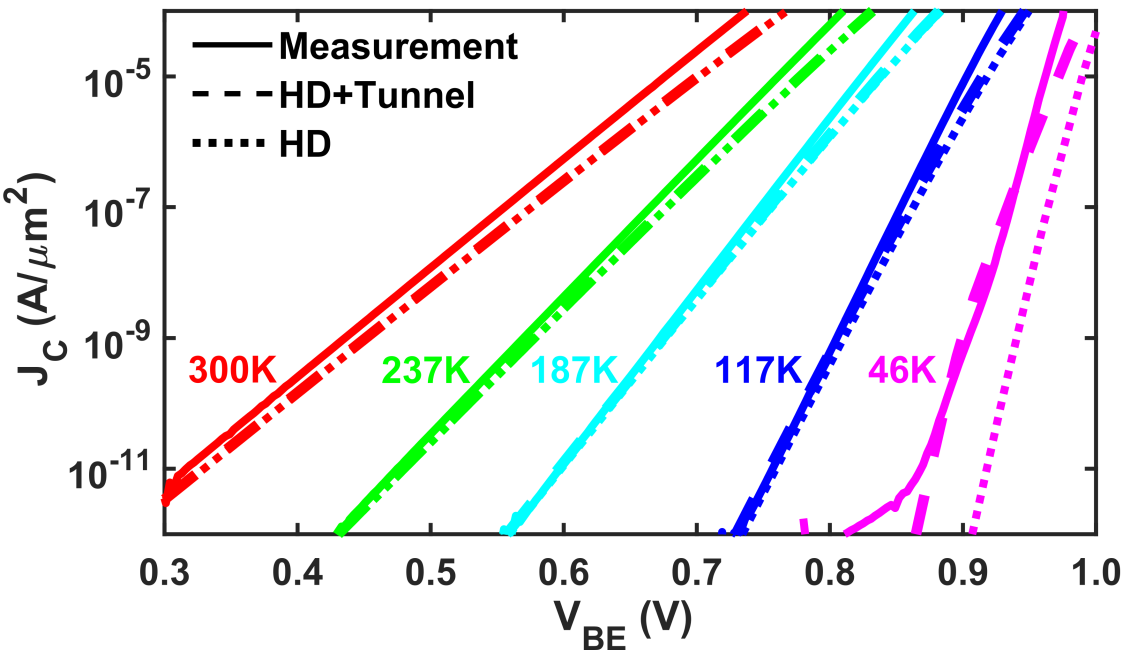

Figure 5.2: Collector current of the fourth generation device from Gummel characteristics and TCAD simulations. TCAD simulations are in hydrodynamic model with or without the tunneling model.

simulated. For comparison purposes, the base width is defined as the distance between the E-B and C-B metallurgical junctions. Though clearly multiple parameters will be scaled simultaneously in the real world (e.g., vertical reduction of base/collector profiles, reduction of emitter cap layer thickness, and modification of Ge profile, etc.), it is difficult to assess the contribution of individual parameters and make a fair comparison. Instead, the sensitivity to the various individual parameters was examined through the use of TCAD simulations. The base and SIC doping profile are assumed to be Gaussian shaped, and defined by the peak and standard deviation. The base width was adjusted by varying the standard deviation of the Gaussian boron profile, which changes the E-B and C-B doping intercepts. All simulations were performed using the calibrated 4^{th} generation TCAD model deck. The doping profile across the cutline in the intrinsic device is published in [2].

Shown in Fig. 5.3 is the normalized collector current at a fixed V_{BE} versus the various process parameters, all simulated at 57 K. The process parameters were normalized to the values given in Table 5.2. Observe that scaling Ge affects the current density significantly,

Table 5.2: Summary of Parameters in TCAD Simulation

Parameters	**Values**
Peak Base Doping	$7.8 \times 10^{19}\ cm^{-3}$
Peak SIC Doping	$9.0 \times 10^{18}\ cm^{-3}$
Peak Ge Fraction	0.27
Base Width	$16\ nm$

as even a slight reduction in Ge increases the base barrier width and height and thereby decreases the tunneling current exponentially. A separate simulation (not shown) reveals that at low temperatures the Ge content at the BE junction or the Ge grading in the base alone are much less important to the magnitude of tunneling than the total integrated Ge. To understand why, we can look at a typical Ge profile, which ramps up before the EB intercept and ramps down after the CB intercept. Compared to a graded Ge profile, a box Ge profile reduces the base barrier height not only within the neutral base, but also in the EB and CB depletion regions, which effectively reduces the barrier width and exponentially increases the tunneling. Therefore, if a large collector current is desired (e.g., for a large current gain β), or if the turn-on voltage needs to be reduced, a larger Ge mole fraction (close to a box profile shape) throughout the base is preferred. There should be less concern over using a box profile in scaled technologies as the Ge film stability requirement would be more relaxed due to inherently smaller base widths. The base width reduction decreases the tunneling barrier width and allows significantly more collector current to flow. Even if the base doping is increased (higher barrier) while the base width is reduced, the current still increases significantly because the effect of base width dominates over the effect of base doping. The effect of collector doping is also small, which means that a higher collector doping to suppress Kirk and heterojunction barrier effects for the room temperature operation will not have much impact on cryogenic operation.

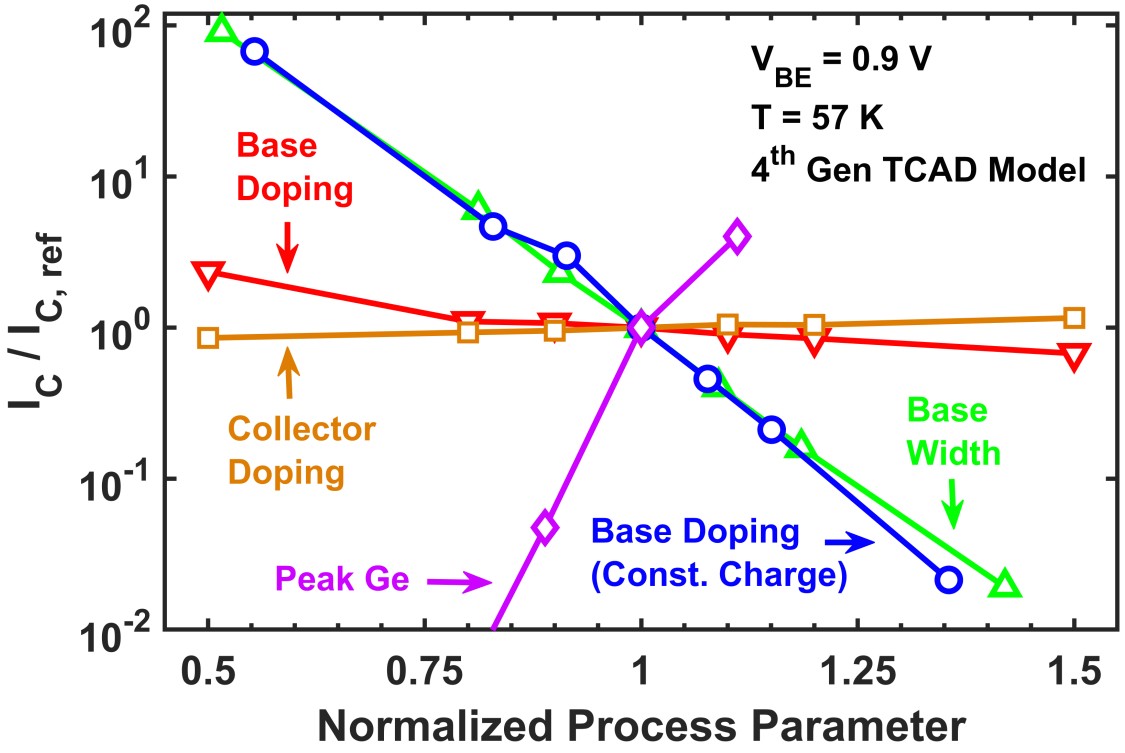

Figure 5.3: The simulated collector current versus process parameters that are individually varied in TCAD. Both the current and process parameters are normalized to the initial values.

CHAPTER 6
CONCLUSION

There are many useful information that was obtained from this thesis. First and foremost, the DC behavior of multiple technology generations of SiGe HBTs have, for the first time, been characterized at deep cryogenic temperatures as low as 70 mK and subjected to strong magnetic field as large as 14 T. The devices are fully operating and usable, even in the first generation device where the freeze-out effect could potentially has detrimental effect. The characteristics of all terminal currents and DC parameters (g_m and β) saturates at low temperatures. Such loss of temperature dependence has been accounted for by existing theories and a new transport mechanism-direct tunneling. The loss of temperature dependence can be considered desirable from a cryogenic circuit perspective, since circuit designers can set an universal operating point at around 40 K and expect it to remain fixed down to 70 mK, provided self-heating is minimized. This also greatly simplifies the effort of characterizing SiGe HBTs for low temperature applications, because the devices only need to be cooled down to 40 K instead of tens of mK. This reduces the complexity of experimental apparatus by large and shortens the length of characterization time.

This thesis also provides insight into cryogenic collector transport in SiGe HBTs. A unified picture of transport is summarized for three technology generations, where direct tunneling can be distinguished from quasi-ballistic transport through a simple experimental method. Among all transport mechanisms, trap-assisted tunneling could be potentially eliminated using optimized process technology, device structures, and film deposition tooling with lower epi trap densities. At least, the industry scaling trend so far suggests that the effects of trap-assisted tunneling is diminishing in more advanced generations. Direct tunneling and quasi-ballistic transport are expected to become more dominant when the base width is scaled down, and the competition between the two transport mechanisms

centers around the integrated Ge profile (width and Ge profile shape). Without an increase in Ge content, direct tunneling may dominate over quasi-ballistic transport to a higher collector current level. Through a process parameter sensitivity analysis using TCAD, total integrated Ge content and base width are determined to be the most important factors for optimizing cryogenic collector current in SiGe HBTs.

APPENDIX A

DERIVATION OF EFFECTIVE MASS EQUATION

In this section, we derive the effective mass equation from the one-electron Schrodinger equation. Quantum mechanically, the electron motion is a complex many-body problem, where electrons and phonons can interact with one another. Normally, one assumes that each electron feels an average force due to the presence of other electrons and phonons. This allows one to write the one-electron Schrodinger equation

$$i\hbar\frac{\partial}{\partial t}\psi_0(r,t) = H\psi_0(r,t) + U_{S,E}(r,t)\psi_0(r,t) \tag{A.1}$$

where

$$H = -\frac{\hbar^2}{2m_0}\nabla^2 + U_L(r)$$

Here, $\psi_0(r,t)$ is the wave function of the electron at location r and time t, $U_L(r)$ is the time-independent potential energy of the lattice, m_0 is the rest mass of electron (different from the effective mass below!) and $U_{S,E}(r,t)$ is the sum of time-dependent potential energy due to scattering (U_S) and electric field (U_E).

For the discussion in this thesis, we assume $U_S = 0$, since the phonon scattering is minimal at low temperatures, and we ignore impurity and alloy scattering. In addition, with a DC voltage, $U_E(r,t) = U_E(r)$. Therefore, we only need to solve the time-independent Schrodinger equation

$$E\psi_0(r) = (-\frac{\hbar^2}{2m_0}\nabla^2 + U_L(r) + U_E(r))\psi_0(r) \tag{A.2}$$

As $U_L(r)$ and $U_E(r)$ approach zero (i.e., free electron), the solutions approach plane waves e^{ikr}. For $U_E(r) = 0$, only the lattice potential $U_L(r)$ is present, the problem becomes

the simple infinite lattice problem. The solutions are Bloch waves

$$\psi_{v,k}(r) = \mu_{v,k}(r)e^{ikr}$$

, and the time-evolution is of the simple form $\psi_{v,k}(r,t) = \mu_{v,k}(r)e^{ikr}e^{-iE(k)t/\hbar}$. In solid state physics, it often takes a major effort to find the appropriate $U_L(r)$ in order to calculate $E(k)$. Therefore, it is often more convenient to use the effective-mass equation where U_L is implicitly formulated into $E(k)$, as in Equ. A.3. Here, U_L is absorbed into the energy term, $E_v(k)$ with k replaced by $-i\nabla$. $\psi(r)$ is the so-called envelope wave function, different than the original wave function $\psi_0(r)$ solved from Equ. A.2. Since $E_v(k)$ values are readily available for most semiconductors as a function of wave number k, the effective mass equation is used more often.

$$E\psi(r) = E_v(-i\nabla)\psi(r) + U_E(r)\psi(r) \tag{A.3}$$

Since the transport problem in this thesis centers around the bottom of the conduction band, the parabolic band approximation can be used. In other words, $E_v(k) = \frac{\hbar^2}{2m^*}k^2 + E_{C0}$, where m^* is the effective mass related to the curvature of the band. Therefore, $E_v(-i\nabla) = -\frac{\hbar^2}{2m^*}\nabla^2 + E_{C0}$, and we obtain the familiar

$$E\psi(r) = -\frac{\hbar^2}{2m^*}\nabla^2\psi(r) + E_C(r)\psi(r) \tag{A.4}$$

where $E_C(r) = E_{C0} + U_E(r)$ is the conduction band energy frequently calculated in device simulators.

The readers are referred to [83] for detailed steps of the above derivations. The only approximation involved in the derivation is

$$< v, k|U_{S,E}|v', k' >= \delta_{v,v'} < k|U_{S,E}|k' >$$

where $|v, k>$ is the Dirac notation for the wave function at a wave vector of k in the energy band v. Since we assume $U_S = 0$, and E_{C0} is a constant, the approximation becomes

$$< v, k|E_C|v', k' >= \delta_{v,v'} < k|E_C|k' >$$

In other words, the matrix elements of potential between states from different bands are negligible.

APPENDIX B

WKB APPROXIMATION DERIVATION

We will derive the WKB approximation in the context of SiGe HBTs here. To begin, we have the effective mass equation in 1D as

$$E\psi(x) = -\frac{\hbar^2}{2m}\nabla^2\psi(x) + U(x)\psi(x) \tag{B.1}$$

The solution for a constant box potential U assumes the form

$$e^{\pm ik(x)x}$$

where

$$k(x) = \begin{cases} i\sqrt{2m(U(x)-E)/\hbar^2}, & U(x) > E. \\ \sqrt{2m(E-U(x))/\hbar^2}, & E > U(x). \end{cases} \tag{B.2}$$

If, instead, the potential U is slowly varying (the exact meaning of "slowly varying" is discussed below), the solution can still be assumed to have a similar form as $e^{\pm iu(x)}$. Plugging this ansatz into Equ. B.1 and using the definition in Equ. B.2, we obtain the Schrodinger equation as

$$i\frac{\partial^2 u(x)}{\partial x^2} - (\frac{\partial u(x)}{\partial x})^2 + [k(x)]^2 = 0 \tag{B.3}$$

Equ. B.3 can be written in an equivalent form by rearranging terms and integrating with respect to x, which becomes (omitting the argument x from now on for clarity)

$$u = \pm\int\sqrt{i\frac{\partial^2 u}{\partial x^2} + k^2}dx + C \tag{B.4}$$

where C is an integration constant. To solve it, the WKB approximation assumes

$$|\frac{\partial^2 u}{\partial x^2}| \ll |k|^2 \tag{B.5}$$

This is reasonable because $\frac{\partial^2 u}{\partial x^2}$ is exactly zero for a free electron. Setting $\frac{\partial^2 u}{\partial x^2}$ to zero, Equ. B.4 yields

$$u = \pm \int \sqrt{k^2} dx + C = \pm \int |k| dx + C$$

Taking a second derivative with respect to x on both sides, we get

$$\frac{\partial^2 u}{\partial x^2} = \pm \frac{\partial k}{\partial x} \tag{B.6}$$

Plugging Equ. B.6 into the requirement for WKB approximation as in Equ. B.5, we have

$$|\frac{\partial k}{\partial x}| \ll |k|^2$$

Replacing k with its definition in Equ. B.2, we obtain Equ. B.7

$$\frac{\partial U(x)}{\partial x} \ll 2\sqrt{\frac{2m}{\hbar^2}}(|U(x) - E|)^{3/2} \tag{B.7}$$

APPENDIX C

BAND STRUCTURE CALCULATION

Table C.1: Eigenvectors for Different Ge Fraction at $k = 0.88 \times 2\pi/a$

Ge fraction	**2S**	**2Pz**	**2Py**	**2Px**	**3S**	**3Pz**	**3Py**	**3Px**	**4S**	**3Dx**	**3Dz**	**4Pz**	**4Py**	**4Px**	**5S**
0	0	0.408	0	0	0	0.492	0	0	0	-0.686	0	0	0	-0.343	-0.057
0.05	0	-0.411	0	0	0	0.494	0	0	0	-0.686	0	0	0	-0.335	0.058
0.1	0	-0.415	0	0	0	-0.496	0	0	0	0.686	0	0	0	0.328	-0.059
0.15	0	-0.419	0	0	0	0.498	0	0	0	0.686	0	0	0	-0.322	-0.060
0.2	0	0.422	0	0	0	-0.500	0	0	0	0.685	0	0	0	-0.315	-0.060
0.25	0	-0.426	0	0	0	-0.502	0	0	0	0.684	0	0	0	-0.309	0.061
0.3	0	0.429	0	0	0	0.504	0	0	0	0.683	0	0	0	-0.302	0.062

Table C.2: Eigenvectors for Si (0% Ge Fraction) at $k = 0.86 - 0.9 \times 2\pi/a$

k (2π/a)	**2S**	**2Pz**	**2Py**	**2Px**	**3S**	**3Pz**	**3Py**	**3Px**	**4S**	**3Dx**	**3Dz**	**4Pz**	**4Py**	**4Px**	**5S**
0.860	0	-0.411	0	0	0	0.489	0	0	0	-0.7	0	0	0	0.331	-0.055
0.866	0	0.410	0	0	0	-0.490	0	0	0	0.7	0	0	0	-0.335	-0.056
0.871	0	0.409	0	0	0	-0.491	0	0	0	0.7	0	0	0	0.338	0.056
0.877	0	0.408	0	0	0	-0.491	0	0	0	0.7	0	0	0	-0.341	-0.057
0.883	0	0.407	0	0	0	0.492	0	0	0	0.7	0	0	0	-0.344	0.057
0.889	0	-0.406	0	0	0	0.493	0	0	0	-0.7	0	0	0	-0.347	0.057
0.894	0	-0.406	0	0	0	-0.494	0	0	0	-0.7	0	0	0	-0.351	-0.058
0.900	0	0.405	0	0	0	0.495	0	0	0	-0.7	0	0	0	-0.354	-0.058

Table C.3: Eigenvectors for 30% Ge Fraction at $k = 0.86 - 0.9 \times 2\pi/a$

k (2π/a)	**2S**	**2Pz**	**2Py**	**2Px**	**3S**	**3Pz**	**3Py**	**3Px**	**4S**	**3Dx**	**3Dz**	**4Pz**	**4Py**	**4Px**	**5S**
0.860	0	0.433	0	0	0	0.502	0	0	-0.687	0	0	0	0	-0.292	-0.060
0.866	0	0.432	0	0	0	-0.502	0	0	0.686	0	0	0	0	-0.295	-0.061
0.871	0	0.431	0	0	0	-0.503	0	0	0.685	0	0	0	0	-0.298	-0.061
0.877	0	0.430	0	0	0	0.503	0	0	0.684	0	0	0	0	-0.301	0.062
0.883	0	0.429	0	0	0	0.504	0	0	0.683	0	0	0	0	0.304	0.062
0.889	0	0.428	0	0	0	-0.505	0	0	0.681	0	0	0	0	-0.307	-0.062
0.894	0	-0.427	0	0	0	-0.505	0	0	-0.680	0	0	0	0	-0.309	0.063
0.900	0	0.426	0	0	0	-0.506	0	0	-0.679	0	0	0	0	0.312	-0.063

APPENDIX D

VALUES USED IN FITTING FUNCTIONS

Below are values used in the fitting functions of Equ. 4.8 for Fig. 4.7.

Table D.1: Parameters for Fitting ($V_{BE} = 0.8772V$)

Temperature (K)	**a**	**b**	**T**	**c**	**d**	R^2
4.000	12.397	0.877	3.645	678.79	-590.73	0.99331
2.200	14.399	0.877	2.267	63.18	-52.05	0.99892
1.565	14.400	0.877	1.592	101.65	-85.72	0.99882
1.200	14.485	0.877	1.216	71.31	-59.15	0.99874
0.800	14.441	0.877	0.867	82.90	-69.29	0.99874
0.500	14.480	0.877	0.445	72.37	-60.09	0.99870
0.200	14.412	0.877	0.289	69.86	-58.01	0.99880
0.100	14.425	0.877	0.353	72.49	-60.30	0.99884
0.070	14.399	0.877	0.264	71.76	-59.65	0.99884

Table D.2: Parameters for Fitting ($V_{BE} = 0.9498$)

Temperature (K)	**a**	**b**	**T**	**c**	**d**	R^2
4.000	0.116	0.950	4.079	8.00	-7.50	0.99919
2.200	0.118	0.950	2.313	7.41	-6.94	0.99882
1.565	0.122	0.950	1.776	6.34	-5.92	0.99867
1.200	0.126	0.950	1.479	4.67	-4.34	0.99840
0.800	0.127	0.950	1.181	4.45	-4.14	0.99840
0.500	0.128	0.950	1.056	4.21	-3.90	0.99841
0.200	0.128	0.950	1.025	4.10	-3.80	0.99950
0.100	0.128	0.950	0.998	4.46	-4.14	0.99951
0.070	0.129	0.950	1.008	4.14	-3.83	0.99950

REFERENCES

[1] H. Ying, B. R. Wier, J. Dark, N. E. Lourenco, L. Ge, A. P. Omprakash, M. Mourigal, D. Davidovic, and J. D. Cressler, "Operation of sige hbts down to 70 mk," *IEEE Electron Device Letters*, vol. 38, no. 1, pp. 12–15, 2017. DOI: 10.1109/LED.2016.2633465.

[2] D. Davidovi ć, H. Ying, J. Dark, B. R. Wier, L. Ge, N. E. Lourenco, A. P. Omprakash, M. Mourigal, and J. D. Cressler, "Tunneling, current gain, and transconductance in silicon-germanium heterojunction bipolar transistors operating at millikelvin temperatures," *Phys. Rev. Applied*, vol. 8, p. 024 015, 2 2017. DOI: 10.1103/PhysRevApplied.8.024015.

[3] H. Ying, J. Dark, A. P. Omprakash, B. R. Wier, L. Ge, U. Raghunathan, N. E. Lourenco, Z. E. Fleetwood, M. Mourigal, D. Davidovic, and J. D. Cressler, "Collector transport in sige hbts operating at cryogenic temperatures," *IEEE Transactions on Electron Devices*, vol. 65, no. 9, pp. 3697–3703, 2018. DOI: 10.1109/TED.2018.2854288.

[4] S. Finn, J. Yuan, R. Krithivasan, L. Najafizadeh, P. Cheng, and J. D. Cressler, "A 10 mbps sige bicmos transceiver for operation down to cryogenic temperatures," in *Proceedings of 2007 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2007, pp. 115–118. DOI: 10.1109/BIPOL.2007.4351849.

[5] Y. Yao, F. Dai, R. C. Jaeger, and J. D. Cressler, "A 12-bit cryogenic and radiation-tolerant digital-to-analog converter for aerospace extreme environment applications," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 7, pp. 2810–2819, 2008. DOI: 10.1109/TIE.2008.924174.

[6] J. C. Bardin, P. Ravindran, S. W. Chang, R. Kumar, J. A. Stern, M. D. Shaw, D. S. Russell, and W. H. Farr, "A high-speed cryogenic sige channel combiner ic for large photon-starved snspd arrays," in *Proceedings of 2013 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2013, pp. 215–218. DOI: 10.1109/BCTM.2013.6798179.

[7] C. Ulaganathan, N. Nambiar, B. Prothro, R. Greenwell, S. Chen, B. J. Blalock, C. L. Britton, M. N. Ericson, H. Hoang, R. Broughton, K. Cornett, G. Fu, H. A. Mantooth, J. D. Cressler, and R. W. Berger, "A sige bicmos instrumentation channel for extreme environment applications," in *Proceedings of the 51st Midwest Symposium on Circuits and Systems*, 2008, pp. 217–220. DOI: 10.1109/MWSCAS.2008.4616775.

[8] J. D. Cressler and A. Mantooth, *Extreme Environment Electronics*. CRC Press, 2012.

[9] J. Yuan, J. D. Cressler, R. Krithivasan, T. Thrivikraman, M. H. Khater, D. C. Ahlgren, A. J. Joseph, and J. S. Rieh, "On the performance limits of cryogenically operated sige hbts and its relation to scaling for terahertz speeds," *IEEE Transactions on Electron Devices*, vol. 56, no. 5, pp. 1007–1019, 2009. DOI: 10.1109/TED.2009.2016017.

[10] J. D. Cressler, "On the potential of sige hbts for extreme environment electronics," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1559–1582, 2005. DOI: 10.1109/JPROC.2005.852225.

[11] A. J. Joseph, J. D. Cressler, and D. M. Richey, "Operation of sige heterojunction bipolar transistors in the liquid-helium temperature regime," *IEEE Electron Device Letters*, vol. 16, no. 6, pp. 268–270, 1995. DOI: 10.1109/55.790731.

[12] J. D. Cressler, "Sige-base bipolar transistors for cryogenic bicmos applications," in *Proceedings of the 22nd European Solid State Device Research Conference (ESSDERC)*, 1992, pp. 841–848. DOI: 10.1016/0167-9317(92)90556-7.

[13] L. Najafizadeh, J. S. Adams, S. D. Phillips, K. A. Moen, J. D. Cressler, S. Aslam, T. R. Stevenson, and R. M. Meloy, "Sub-1-k operation of sige transistors and circuits," *IEEE Electron Device Letters*, vol. 30, no. 5, pp. 508–510, 2009. DOI: 10.1109/LED.2009.2016767.

[14] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Boston, MA: Artech House, 2002.

[15] J. D. Cressler, *The Silicon Heterostructure Handbook: Materials, Fabrication, Devices, Circuits, and Applications of SiGe and Si Strained-Layer Epitaxy*. New York, NY: CRC Press, 2005.

[16] P. S. Chakraborty, A. S. Cardoso, B. R. Wier, A. P. Omprakash, J. D. Cressler, M. Kaynak, and B. Tillack, "A 0.8 thz sige hbt operating at 4.3 k," *IEEE Electron Device Letters*, vol. 35, no. 2, pp. 151–153, 2014. DOI: 10.1109/LED.2013.2295214.

[17] H. Tanaka, Y. Sekine, S. Saito, and H. Takayanagi, "Dc-squid readout for qubit," *Physica C: Superconductivity*, vol. 368, no. 1, pp. 300 –304, 2002. DOI: https://doi.org/10.1016/S0921-4534(01)01185-6.

[18] M. J. Curry, T. D. England, N. C. Bishop, G. Ten-Eyck, J. R. Wendt, T. Pluym, M. P. Lilly, S. M. Carr, and M. S. Carroll, "Cryogenic preamplification of a single-electron-transistor using a silicon-germanium heterojunction-bipolar-transistor," *Applied Physics Letters*, vol. 106, no. 20, p. 203 505, 2015. DOI: 10.1063/1.4921308. eprint: https://doi.org/10.1063/1.4921308.

[19] C. K. Andersen, J. Kerckhoff, K. W. Lehnert, B. J. Chapman, and K. Mølmer, "Closing a quantum feedback loop inside a cryostat: autonomous state preparation and long-time memory of a superconducting qubit," *Phys. Rev. A*, vol. 93, p. 012 346, 1 2016. DOI: 10.1103/PhysRevA.93.012346.

[20] A. A. Prager, H. C. George, A. O. Orlov, and G. L. Snider, "Experimental demonstration of hybrid cmos-single electron transistor circuits," *Journal of Vacuum Science & Technology B*, vol. 29, no. 4, p. 041 004, 2011.

[21] K. Das and T. Lehmann, "A cryogenic single electron transistor readout circuit: practical issues and measurement considerations," in *2012 IEEE International Symposium on Circuits and Systems*, 2012, pp. 1359–1362.

[22] A. Robinson and V. Talyanskii, "Cryogenic amplifier for 1 mhz with a high input impedance using a commercial pseudomorphic high electron mobility transistor," *Review of Scientific Instruments*, vol. 75, pp. 3169–3176, 2004.

[23] I. Vink, T Nooitgedagt, R. Schouten, L. Vandersypen, and W. Wegscheider, "Cryogenic amplifier for fast real-time detection of single-electron tunneling," *Applied Physics Letters*, vol. 91, no. 12, p. 123 512, 2007.

[24] A. T. Lee, "Broadband cryogenic preamplifiers incorporating gaas mesfets for use with low-temperature particle detectors," *Review of Scientific Instruments*, vol. 60, no. 10, pp. 3315–3322, 1989.

[25] S. J. Devitt, W. J. Munro, and K. Nemoto, "Quantum error correction for beginners," *Reports on Progress in Physics*, vol. 76, no. 7, p. 076 001, 2013.

[26] A. Montanaro, "Quantum algorithms: an overview," *npj Quantum Information*, vol. 2, p. 15 023, 2016.

[27] M. Schroter, G. Wedel, B. Heinemann, C. Jungemann, J. Krause, P. Chevalier, and A. Chantre, "Physical and electrical performance limits of high-speed sigec hbts part i: vertical scaling," *IEEE Transactions on Electron Devices*, vol. 58, no. 11, pp. 3687–3696, 2011. DOI: 10.1109/TED.2011.2163722.

[28] J. J. Pekarik, J. Adkisson, P. Gray, Q. Liu, R. Camillo-Castillo, M. Khater, V. Jain, B. Zetterlund, A. DiVergilio, X. Tian, A. Vallett, J. Ellis-Monaghan, B. J. Gross, P. Cheng, V. Kaushal, Z. He, J. Lukaitis, K. Newton, M. Kerbaugh, N. Cahoon, L. Vera, Y. Zhao, J. R. Long, A. Valdes-Garcia, S. Reynolds, W. Lee, B. Sadhu, and D. Harame, "A 90nm sige bicmos technology for mm-wave and high-performance analog applications," in *Proceedings of 2014 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2014, pp. 92–95. DOI: 10.1109/BCTM.2014.6981293.

[29] P. Chevalier, T. F. Meister, B. Heinemann, S. V. Huylenbroeck, W. Liebl, A. Fox, A. Sibaja-Hernandez, and A. Chantre, "Towards thz sige hbts," in *2011 IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, 2011, pp. 57–65. DOI: 10.1109/BCTM.2011.6082749.

[30] D. M. Richey, A. J. Joseph, J. D. Cressler, and R. C. Jaeger, "Evidence for non-equilibrium base transport in si and sige bipolar transistors at cryogenic temperatures," *Solid-State Electronics*, vol. 39, no. 6, pp. 785–789, 1996.

[31] B. Heinemann, H. Rcker, R. Barth, F. Brwolf, J. Drews, G. G. Fischer, A. Fox, O. Fursenko, T. Grabolla, F. Herzel, J. Katzer, J. Korn, A. Krger, P. Kulse, T. Lenke, M. Lisker, S. Marschmeyer, A. Scheit, D. Schmidt, J. Schmidt, M. A. Schubert, A. Trusch, C. Wipf, and D. Wolansky, "Sige hbt with fx/fmax of 505 ghz/720 ghz," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 3.1.1–3.1.4. DOI: 10.1109/IEDM.2016.7838335.

[32] D. C. Ahlgren, G. Freeman, S. Subbanna, R. Groves, D. Greenberg, J. Malinowski, D. Nguyen-Ngoc, S. J. Jeng, K. Stein, K. Schonenberg, D. Kiesling, B. Martin, S. Wu, D. Harame, and B. Meyerson, "A sige hbt bicmos technology for mixed signal rf applications," in *Proceedings of the 1997 Bipolar/BiCMOS Circuits and Technology Meeting*, 1997, pp. 195–197. DOI: 10.1109/BIPOL.1997.647434.

[33] B. Jagannathan, M. Khater, F. Pagette, J. S. Rieh, D. Angell, H. Chen, J. Florkey, F. Golan, D. R. Greenberg, R. Groves, S. J. Jeng, J. Johnson, E. Mengistu, K. T. Schonenberg, C. M. Schnabel, P. Smith, A. Stricker, D. Ahlgren, G. Freeman, K. Stein, and S. Subbanna, "Self-aligned sige npn transistors with 285 ghz fmax and 207 ghz ft in a manufacturable technology," *IEEE Electron Device Letters*, vol. 23, no. 5, pp. 258–260, 2002. DOI: 10.1109/55.998869.

[34] R. A. Camillo-Castillo, Q. Z. Liu, J. W. Adkisson, M. H. Khater, P. B. Gray, V. Jain, R. K. Leidy, J. J. Pekarik, J. P. Gambino, B. Zetterlund, C. Willets, C. Parrish, S. U. Engelmann, A. M. Pyzyna, P. Cheng, and D. L. Harame, "Sige hbts in 90nm bicmos technology demonstrating 300ghz/420ghz ft/fmax through reduced rb and ccb parasitics," in *2013 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2013, pp. 227–230. DOI: 10.1109/BCTM.2013.6798182.

[35] Q. Z. Liu, J. W. Adkisson, V. Jain, R. A. Camillo-Castillo, M. H. Khater, P. B. Gray, J. J. Pekarik, B. Zetterlund, A. W. Divergilio, M. L. Kerbaugh, and D. L. Harame, "Sige hbts in 90nm bicmos technology demonstrating ft/fmax 285ghz/475ghz through simultaneous reduction of base resistance and extrinsic collector capacitance," *ECS Transactions*, vol. 64, no. 6, pp. 285–294, 2014. DOI: 10.1149/06406.0285ecst. eprint: http://ecst.ecsdl.org/content/64/6/285.full.pdf+html.

[36] S. S. Li and W. R. Thurber, "The dopant density and temperature dependence of electron mobility and resistivity in n-type silicon," *Solid-State Electronics*, vol. 20, no. 7, pp. 609–616, 1977.

[37] P. P. Altermatt, A. Schenk, and G. Heiser, "A simulation model for the density of states and for incomplete ionization in crystalline silicon. i. establishing the model in si: p," *Journal of Applied Physics*, vol. 100, no. 11, p. 113 715, 2006.

[38] A. F. Da Silva, "Metal-insulator transitions in doped silicon and germanium," *Physical Review B*, vol. 37, no. 9, p. 4799, 1988.

[39] F. Morin and J. Maita, "Electrical properties of silicon containing arsenic and boron," *Physical Review*, vol. 96, no. 1, p. 28, 1954.

[40] S. Tanaka and M. S. Lundstrom, "A flux-based study of carrier transport in thin-base diodes and transistors," *IEEE Transactions on Electron Devices*, vol. 42, no. 10, pp. 1806–1815, 1995. DOI: 10.1109/16.464415.

[41] M. A. Stettler and M. S. Lundstrom, "A microscopic study of transport in thin base silicon bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 41, no. 6, pp. 1027–1033, 1994. DOI: 10.1109/16.293317.

[42] A. A. Grinberg and S. Luryi, "Diffusion in a short base," *Solid-State Electronics*, vol. 35, no. 9, pp. 1299 –1309, 1992. DOI: https://doi.org/10.1016/0038-1101(92)90165-9.

[43] J. C. Bardin, "Silicon-germanium heterojunction bipolar transistors for extremely low-noise applications," PhD thesis, California Institute of Technology, 2009.

[44] L. Yang, J. R. Watling, R. C. Wilkins, M. Boriçi, J. R. Barker, A. Asenov, and S. Roy, "Si/sige heterostructure parameters for device simulations," *Semiconductor Science and Technology*, vol. 19, no. 10, p. 1174, 2004.

[45] Z. Xu, G. Niu, L. Luo, and J. Cressler, "A physics-based trap-assisted tunneling current model for cryogenic temperature compact modeling of sige hbts," *ECS Transactions*, vol. 33, no. 6, pp. 301–310, 2010.

[46] J. Bardin, S Montazeri, and S.-W. Chang, "Silicon germanium cryogenic low noise amplifiers," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 834, 2017, p. 012 007.

[47] D. C. Ralph, C. T. Black, and M. Tinkham, "Spectroscopic measurements of discrete electronic states in single metal particles," *Physical Review Letters*, vol. 74, no. 16, pp. 3241–3244, 1995. DOI: 10.1103/PhysRevLett.74.3241.

[48] E. Bonet, M. M. Deshmukh, and D. C. Ralph, "Solving rate equations for electron tunneling via discrete quantum states," *Physical Review B*, vol. 65, no. 4, 2002. DOI: 10.1103/PhysRevB.65.045317.

[49] G.-L. Ingold and Y. V. Nazarov, "Charge tunneling rates in ultrasmall junctions," in *Single charge tunneling*, Springer, 1992, pp. 21–107.

[50] U Meirav and E. Foxman, "Single-electron phenomena in semiconductors," *Semiconductor Science and Technology*, vol. 11, no. 3, p. 255, 1996.

[51] H. Grabert and M. H. Devoret, *Single charge tunneling: Coulomb blockade phenomena in nanostructures*. Springer Science & Business Media, 2013, vol. 294.

[52] D. Mitra *et al.*, *Advanced Semiconductor Heterostructures: Novel Devices, Potential Device Applications and Basic Properties*. World Scientific, 2003, vol. 28.

[53] D. Moraru, A. Samanta, T. Mizuno, H. Mizuta, M. Tabe, *et al.*, "Transport spectroscopy of coupled donors in silicon nano-transistors," *Scientific reports*, vol. 4, p. 6219, 2014.

[54] R. C. Fletcher, W. A. Yager, G. L. Pearson, and F. R. Merritt, "Hyperfine splitting in spin resonance of group v donors in silicon," *Phys. Rev.*, vol. 95, pp. 844–845, 3 1954. DOI: 10.1103/PhysRev.95.844.

[55] G. Dresselhaus, A. F. Kip, and C. Kittel, "Cyclotron resonance of electrons and holes in silicon and germanium crystals," *Phys. Rev.*, vol. 98, pp. 368–384, 2 1955. DOI: 10.1103/PhysRev.98.368.

[56] H Fritzsche, "Resistivity and hall coefficient of antimony-doped germanium at low temperatures," *Journal of Physics and Chemistry of Solids*, vol. 6, no. 1, pp. 69–80, 1958.

[57] A. Miller and E. Abrahams, "Impurity conduction at low concentrations," *Physical Review*, vol. 120, no. 3, p. 745, 1960.

[58] E. M. Conwell, "Impurity band conduction in germanium and silicon," *Physical Review*, vol. 103, no. 1, p. 51, 1956.

[59] M. M. Rieger and P Vogl, "Electronic-band parameters in strained si 1- x ge x alloys on si 1- y ge y substrates," *Physical Review B*, vol. 48, no. 19, p. 14 276, 1993.

[60] W. Li, X. Qian, and J. Li, "Envelope function method for electrons in slowly-varying inhomogeneously deformed crystals," *Journal of Physics: Condensed Matter*, vol. 26, no. 45, p. 455 801, 2014.

[61] A Slaoui and P Siffert, "Determination of the electron effective mass and relaxation time in heavily doped silicon," *physica status solidi (a)*, vol. 89, no. 2, pp. 617–622, 1985.

[62] J. Hensel, H Hasegawa, and M Nakayama, "Cyclotron resonance in uniaxially stressed silicon. ii. nature of the covalent bond," *Physical Review*, vol. 138, no. 1A, A225, 1965.

[63] F. L. Madarasz, J. E. Lang, and P. M. Hemeger, "Effective masses for nonparabolic bands in p-type silicon," *Journal of Applied Physics*, vol. 52, no. 7, pp. 4646–4648, 1981.

[64] M. Cardona and F. H. Pollak, "Energy-band structure of germanium and silicon: the k· p method," *Physical Review*, vol. 142, no. 2, p. 530, 1966.

[65] R. M. Ward, "Modelling of silicon-germanium alloy heterostructures using double group formulation of k. p theory," 2012.

[66] E Merzbacher, *Quantum mechanics, 1998*.

[67] E. Foxman, P. McEuen, U Meirav, N. S. Wingreen, Y. Meir, P. A. Belk, N. Belk, M. Kastner, and S. Wind, "Effects of quantum levels on transport through a coulomb island," *Physical Review B*, vol. 47, no. 15, p. 10 020, 1993.

[68] D Goldhaber-Gordon, J. Göres, M. Kastner, H. Shtrikman, D Mahalu, and U Meirav, "From the kondo regime to the mixed-valence regime in a single-electron transistor," *Physical Review Letters*, vol. 81, no. 23, p. 5225, 1998.

[69] J. Wagner and J. A. del Alamo, "Band-gap narrowing in heavily doped silicon: a comparison of optical and electrical data," *Journal of applied physics*, vol. 63, no. 2, pp. 425–429, 1988.

[70] W. Joyce and R. Dixon, "Analytic approximations for the fermi energy of an ideal fermi gas," *Applied Physics Letters*, vol. 31, no. 5, pp. 354–356, 1977.

[71] M. Suezawa, K. Sumino, H. Harada, and T. Abe, "Nitrogen-oxygen complexes as shallow donors in silicon crystals," *Japanese journal of applied physics*, vol. 25, no. 10A, p. L859, 1986.

[72] M Kittler, C Ulhaq-Bouillet, and V Higgs, "Influence of copper contamination on recombination activity of misfit dislocations in sige/si epilayers: temperature dependence of activity as a marker characterizing the contamination level," *Journal of applied physics*, vol. 78, no. 7, pp. 4573–4583, 1995.

[73] K Graff and H Pieper, "The properties of iron in silicon," *Journal of the Electrochemical Society*, vol. 128, no. 3, pp. 669–674, 1981.

[74] M Schulz, "Determination of deep trap levels in silicon using ion-implantation and cv-measurements," *Applied physics*, vol. 4, no. 3, pp. 225–236, 1974.

[75] J. Benton, J Michel, L. Kimerling, D. Jacobson, Y.-H. Xie, D. Eaglesham, E. Fitzgerald, and J. Poate, "The electrical and defect properties of erbium-implanted silicon," *Journal of applied physics*, vol. 70, no. 5, pp. 2667–2671, 1991.

[76] R. Floyd and D. Walmsley, "Tunnelling conductance of clean and doped al-i-pb junctions," *Journal of Physics C: Solid State Physics*, vol. 11, no. 22, p. 4601, 1978.

[77] E. O. Kane, "Band tails in semiconductors," *Solid-State Electronics*, vol. 28, no. 1-2, pp. 3–10, 1985.

[78] H. Rücker, J. Korn, and J. Schmidt, "Operation of sige hbts at cryogenic temperatures," in *2017 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2017, pp. 17–20. DOI: 10.1109/BCTM.2017.8112902.

[79] D. Klaassen, J. Slotboom, and H. de Graaff, "Unified apparent bandgap narrowing in n- and p-type silicon," *Solid-State Electronics*, vol. 35, no. 2, pp. 125 –129, 1992. DOI: https://doi.org/10.1016/0038-1101(92)90051-D.

[80] D. Klaassen, "A unified mobility model for device simulationi. model equations and concentration dependence," *Solid-State Electronics*, vol. 35, no. 7, pp. 953–959, 1992. DOI: https://doi.org/10.1016/0038-1101(92)90325-7.

[81] M. Ieong, P. M. Solomon, S. Laux, H.-S. Wong, and D. Chidambarrao, "Comparison of raised and schottky source/drain mosfets using a novel tunneling contact model," in *Electron Devices Meeting, 1998. IEDM'98. Technical Digest., International*, IEEE, 1998, pp. 733–736.

[82] *Sentaurus device user guide*, version Version N-2017.09, Synopsys Inc., September, 2017, 1546 pp.

[83] S. Datta, *Quantum phenomena*. Addison-Wesley Longman, 1989, vol. 8.