MODELING AND DESIGN OF SILICON-GERMANIUM HETEROJUNCTION BIPOLAR TRANSISTORS FOR RELIABILITY-AWARE CIRCUIT DESIGN

Approved by:

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

Professor Shyh-Chiang Shen
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Professor Asif Islam Khan
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Professor Dragomir Davidovic
School of Physics
*Georgia Institute of Technology*

Dr. Nelson E. Lourenco
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Date Approved: November 30, 2018
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr. Cressler for all his guidance, support, and encouragement he has provided over the almost seven full years I have worked as a student in his group. He has been a truly excellent support in fostering an interest in me for research and providing me the needed push and support when facing adversity in my research over the years. He has been an exceptional mentor, especially in preaching the need to always take a step back to appreciate the bigger picture not only in research but especially in life.

I would like to thank Dr. Shyh-Chiang Shen and Dr. Paul Douglas Yoder for taking the time to serve on my PhD proposal committee and provide invaluable feedback as well as Dr. Asif Khan, Dr. Dragomir Davidovic, and Dr. Nelson Lourenco for their contributions as members of my defense committee.

I am deeply greatly to all current and past members of the SiGe Devices and Circuits Group for their support over the years. I would like to give my greatest thanks to Uppili who has served as a close collaborator for my entire tenure and who has always been an endless source of new ideas, and I would also like to give special thanks also go to Partha, Michael, Rafael, and Victor, who I have enjoyed working and learning with over the years. Additionally, I would like to thank Adrian, Zach, Chris, Nelson, Peter, and Rob, all lab members who I never worked closely with but truly helped me survive over the years whether it be just hanging out or having a few lab beers,

I would like to thank my family, especially my parents Charles and Jane, who have provided constant support over the years and fostered my intellectual curiosity from a young age. I would also like to my friends in the broader Georgia Tech community
and beyond who have allowed me to make them so many sandwiches over the years and always eaten with joy (there hasn’t been a bad one yet!).

Finally, I would never have wanted to go through these past several years without the love and support of my wonderful partner Emma, who has made every day a joy, even when struggle to see it. Seeing the passion and determination you face your work with every day has always provided me with a drive to seek success and strive to always find joy in what I do.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................. iii

LIST OF FIGURES ......................................................... vii

SUMMARY ................................................................. xiii

1 INTRODUCTION ............................................................ 1
   1.1 Origin and History of the Problem ............................. 1
   1.2 Organization ....................................................... 2

2 SILICON-GERMANIUM BICMOS TECHNOLOGIES .............. 4
   2.1 Technology Overview ........................................... 4
       2.1.1 Silicon Germanium Heterojunction Bipolar Transistors ... 4
       2.1.2 SiGe BiCMOS History and Development ................. 7
   2.2 Collector Scaling and Breakdown Mechanics in SiGe HBTs ... 10
   2.3 Reliability Effects in SiGe HBTs ............................. 12

3 VERTICAL SUPERJUNCTION COLLECTORS FOR ENHANCED
   BREAKDOWN PERFORMANCE IN SIGE HBTS ..................... 15
   3.1 Motivation ....................................................... 15
   3.2 Breakdown and Superjunction Mechanics ................... 17
   3.3 Vertical Superjunction Collector Design Parameterization ... 22
       3.3.1 Doping Level ............................................. 22
       3.3.2 Placement ............................................... 27
       3.3.3 Shape .................................................... 28
   3.4 Practical Superjunction Fabrication ....................... 29
   3.5 Superjunction Measurement Results ....................... 32
   3.6 Improved Superjunctions ..................................... 38
       3.6.1 Epitaxial Superjunctions ............................. 39
       3.6.2 Implanted Superjunctions ............................. 41
   3.7 Summary ....................................................... 44
4 HOT CARRIER AGING IN SIGE HBTS ........................................ 46
  4.1 Motivation ............................................................... 46
  4.2 Hot-Carrier Physics .................................................... 48
  4.3 Device Stress Measurements ......................................... 53
    4.3.1 Mixed-Mode Stress .............................................. 54
    4.3.2 High-Current Stress ............................................ 57
  4.4 High-Current Simulations ........................................... 63
  4.5 Summary ............................................................... 66

5 COMPACT AGING MODELING FOR SIGE HBTS ......................... 67
  5.1 Motivation ............................................................... 67
  5.2 Model Formulation ..................................................... 68
  5.3 Stress Measurement and Model Calibration ......................... 75
  5.4 Aging Simulations ..................................................... 79
  5.5 Circuit Aging ............................................................ 82
  5.6 Summary ............................................................... 84

6 AGING MODELING OF SIGE HBTS FOR RELIABILITY-AWARE CIRCUIT DESIGN ................................................................. 86
  6.1 Motivation ............................................................... 86
  6.2 Testing Setup ........................................................... 88
  6.3 Model Formulation ..................................................... 89
  6.4 RF Device Stressing ................................................... 94
  6.5 Full Aging Simulations ............................................... 97
  6.6 Summary ............................................................... 99

7 CONCLUSION ................................................................. 101
  7.1 Summary of Contributions ........................................... 101
  7.2 Future Work ............................................................ 102

REFERENCES ................................................................. 104

VITA ................................................................. 123
LIST OF FIGURES

2.1 (a) Schematic cross-section and (b) measured SIMS profile of a representative first-generation SiGe HBT (after [13]). 5

2.2 Energy band diagram for an npn Si BJT and SiGe HBT showing the effects of Ge induced band offsets and band gap grading (after [13]). 6

2.3 Schematic view of a SiGe HBT cross-section used for scaling (after [14]). 8

2.4 Pairs of $f_T/BV_{CEO}$ values for SiGe BiCMOS platforms from major foundries. Constant $f_T \times BV_{CEO}$ product lines are shown for reference. 9

2.5 Schematic view of doping and germanium profile changes associated with typical vertical scaling in a SiGe HBT. 10

2.6 Avalanche multiplication process in a bipolar device (after [13]). 12

2.7 Example mixed-mode stress response showing (a) base current degradation in the Gummel characteristic and (b) current gain $\beta$ degradation. 13

3.1 Simulated electric field and electron temperature in the collector-base (CB) space charge region (SCR) for both standard collector device demonstrating the dead space region as the separation in the peak electric field and carrier temperature. 18

3.2 A lateral superjunction, with multiple n/p stripes displaced in the third dimension of the drift region (after [68]). 19

3.3 Schematic drawing of a two-layer superjunction collector structure within a typical SiGe HBT. Abrupt $p/n$ stripes are stacked some distance from the CB metallurgical junction. Doping levels ($N_{shallow}/N_{deep}$) and layer widths ($W_{shallow}/W_{deep}$) may differ for the shallow and deep pairs of $p/n$ stripes. 20

3.4 Simulated electric field and electron temperature in the CB SCR for both standard collector (dashed line) and superjunction collector (solid line) designs demonstrating the dead space region. 21

3.5 Simulated electron temperature and electron ionization coefficient in the CB SCR for both standard collector (dashed line) and superjunction collector (solid line) designs. 21

3.6 Simulated $BV_{CEO}$ and peak $f_T$ values as a function of peak superjunction layer doping with fixed layer thickness and position. $f_T$ values for $V_{CB} = 0.5 \text{ V}$ and $4 \text{ V}$ correspond to undepleted and fully-depleted superjunction layers, respectively. 23
3.7 Simulated electron temperature (solid lines) and electron ionization coefficient $\alpha_n$ (dashed lines) in the CB SCR for a range of superjunction doping values showing the development of a secondary ionization peak with increasing doping. .................................................. 24

3.8 Simulated (a) electric field profiles showing an early collapse and (b) electron density profiles for standard collector and superjunction collector designs for $V_{BE}$ ranging 0.8 – 0.85 V, where conductivity modulation effects set in. The field collapse and subsequent accumulation of carriers at the superjunction edge leads to quasi-saturation at low $V_{CB}$ bias and the start of roll-off in $f_T/f_{max}$. The location of the $p^+$ superjunction layers is marked by dotted lines. ........................................... 25

3.9 Accumulated transit time $\tau_{ec}$ through standard and superjunction collector devices for $V_{CB} = 0 – 1$ V. The field collapse and carrier accumulation due to conductivity modulation cause a sharp increase in $\tau_{ec}$ between the CB metallurgical junction and the superjunction layers, marked by the dotted lines, which improves with increasing $V_{CB}$ . . . . 26

3.10 Simulated values for open-base collector-emitter breakdown voltage ($BV_{CEO}$) and peak unity-gain cutoff frequency ($f_T$) at $V_{CB} = 0.5/4$ V as the distance from the CB metallurgical junction to the leading edge of the most shallow superjunction layer is swept. ......................... 27

3.11 Simulated values of $BV_{CEO}$ for varying doping ratio $N_{deep}/N_{shallow}$ between the deep and shallow superjunction layers. With the shallow layer width ($W_{shallow}$) held constant, variation of the width of the deep layer ($W_{deep}$) shows an increase in $BV_{CEO}$ with increasing width. . . . 29

3.12 Device cross-section and schematic example profiles with integrated compensating dopant for the (top) modified epi method and (bottom) low-energy implant method. ................................. 30

3.13 Measured Gummel characteristics showing matched collector and base current densities for the control and superjunction collector devices. . 32

3.14 Measured fixed-$I_B$ output characteristics for the control and superjunction profile designs. Small signal output resistance $r_o$ for $J_B = 0.1$ $\mu A/\mu m^2$ is shown in the inset and demonstrates the effects of conductivity modulation at low $V_{CE}$ due to quasi-saturation effects. ............................ 33

3.15 Measured $BV_{CEO}$ vs input $J_E$ curves under fixed-$I_E$ operation for the control, superjunction, and commercial 5PAE device designs. $BV_{CEO}$ was extracted from measurement via the base current reversal point. . 34
3.16 Measured \((M - 1)\) curves extracted from fixed-\(I_E\) operation for the control, superjunction, and commercial 5PAE device designs. With \(\beta \approx 100\) for each device, the voltage shifts for \((M - 1) = 0.01\) match the differences in \(BV_{CEO}\). 35

3.17 Measured \(f_T\) (dashed lines) and \(f_{MAX}\) (solid lines) at \(V_{CB} = 1\) V for the control, superjunction, and modified EB superjunction devices. 36

3.18 Measured \(f_T\) (dashed lines) and \(f_{MAX}\) (solid lines) at \(V_{CB} = 3\) V for the control, superjunction, and modified EB superjunction devices. 36

3.19 Figure of merit (Peak \(f_{MAX}\) × (min. \(BV_{CEO}\)) plotted across applied \(V_{CB}\) for the control, superjunction, modified emitter-base (EB) superjunction, and commercial 5PAE device designs. The inset plot shows a Johnson curve plot for the values at \(V_{CB} = 4\) V with constant product lines at 500/700 GHz-V. 37

3.20 Measured \(f_T/f_{MAX}\) curves at \(V_{CB} = 0.5\) V for the control and epitaxial superjunction devices showing matched \(f_T\) and a degraded \(f_{MAX}\) for the superjunction. 39

3.21 Measured inverse mode Gummel characteristic for the control and epitaxial superjunction devices. 40

3.22 Measured \(BV_{CEO}\) vs input \(J_E\) curves under fixed-\(I_E\) operation for the revised control and implanted superjunction device designs found via the base current reversal point. 41

3.23 Measured fixed-\(I_B\) output characteristics for the improved control and superjunction profile designs. In comparison to Fig. 3.14, conductivity modulation effects at low-\(V_{CE}\) have been corrected. 42

3.24 Measured \((f_T/f_{MAX})\) for revised control and implanted superjunction devices at (a) \(V_{CB} = 1\) V and (b) \(V_{CB} = 3\) V. 43

3.25 Figure of merit (Peak \(f_{MAX}\) × (min. \(BV_{CEO}\)) plotted across applied \(V_{CB}\) for the revised control, revised superjunction, original superjunction, and commercial 5PAE device designs. 44

4.1 Cross-section of a 2-D device model annotated with the basic process of the mixed-mode degradation mechanism (after [83]). 50

4.2 Schematic representation of a e-e-h Auger recombination process, where an e-h pair recombine and donate energy to a third carrier. 52

4.3 Measurement setup for hot carrier stress characterization with a fixed emitter current density \(J_E\) and collector-base voltage \(V_{CB}\). 54
4.4 Mixed-mode stress evolution over 10,000 s for the stress condition \( J_E = 400 \, \mu A/\mu m^2 \) and \( V_{CB} = 7 \, V \). (a) Forward-mode and (b) inverse-mode Gummel characteristics show the effects of interface traps at the EB spacer oxide and STI oxide, respectively. ........................................... 55

4.5 Mixed-mode stress evolution over 10,000 s for the stress condition \( J_E = 400 \, \mu A/\mu m^2 \) and \( V_{CB} = 7 \, V \). (a) Forward-mode and (b) inverse-mode Gummel characteristics show the effects of interface traps at the EB spacer oxide and STI oxide, respectively. ........................................... 56

4.6 High-current stress evolution over 10,000 s for the stress condition \( J_E = 20 \, mA/\mu m^2 \) and \( V_{CB} = 1 \, V \). (a) Forward-mode and (b) inverse-mode Gummel characteristics show the effects of interface traps at the EB spacer oxide and STI oxide, respectively. ........................................... 58

4.7 Time evolution of base and collector current for the stress condition \( J_E = 20 \, mA/\mu m^2 \) and \( V_{CB} = 1 \, V \). Non-ideal base current shifts are evident at low \( V_{BE} \), and shifts in the ideal base current and \( R_E \) are evident at higher \( V_{BE} \). ........................................... 59

4.8 Base current degradation at \( V_{BE} = 0.5 \, V \) from forward and inverse mode Gummel characteristics after 10,000 s of stress at \( V_{CB} = 1 \, V \) and range of emitter current densities. ........................................... 60

4.9 Base current degradation at \( V_{BE} = 0.5 \, V \) from forward and inverse mode Gummel characteristics after 10,000 s of stress at \( J_E = 20 \, mA/\mu m^2 \) and range of collector-base voltages. ........................................... 61

4.10 Base current degradation extracted at \( I_B = 1 \, pA \) from forward-mode Gummel characteristics after 10,000 s of the stress condition at \( J_E = 20 \, mA/\mu m^2 \) and \( V_{CB} = 2 \, V \) at both 25 °C and 100 °C. ........................................... 62

4.11 TCAD simulated peak Auger recombination rate within the neutral base region showing the threshold behavior and increase in high current hot carrier degradation........................................... 64

4.12 TCAD simulated conduction band energy within the neutral base region for a range of current densities near the onset of base push-out. The shape of the germanium profile is overlaid to help illustrate the formation of the pseudo-potential well. ........................................... 65

4.13 TCAD simulated electric field in the collector region for \( V_{CB} = 4 \, V \) for a range of emitter current densities demonstrating the compression of the electric field at the sub-collector under heavy base push-out conditions. In addition, the field at \( J_E = 20 \, mA/\mu m^2 \) and \( V_{CB} = 1 \, V \) is shown for comparison. The sub-collector location is shown by magenta dashed line. ........................................... 65
5.1 Gummel characteristic evolution of a SiGe HBT biased under typical mixed-mode stress conditions for 10,000 s.

5.2 TCAD generated hot carrier impingement rates for a 1st generation SiGe HBT technology across a wide range of stress conditions.

5.3 Plot of $\Delta I_b$ at each time step of mixed-mode stress, showing the extrapolated saturation current resulting from linear regression.

5.4 Evolution of $\Delta I_{n,sat}$ over time for a mixed-mode stress condition. Also shown is a fitted aging line used to extract the aging rate.

5.5 Measurement extracted and model simulated age rates for a range of mixed-mode stress bias conditions. The simulated (line) and measured (line) values show good agreement across the full data range.

5.6 Simulation flow diagram for the aging process within Spectre.

5.7 Simulated Gummel characteristics showing degradation of a SiGe HBT under mixed-mode stress conditions.

5.8 Simulated degradation map for a SiGe HBT under mixed-mode stress conditions showing the normalized change in $I_{reps}$, the HICUM peripheral B-E recombination saturation current.

5.9 Schematic of current mirror stress setup in both simulation and measurement. A constant current density $J_{in} = 1 \text{ mA}/\mu\text{m}^2$ is supplied at the input device, and a constant voltage $V_{out} = 4 \text{ V}$ is applied on the output device.

5.10 Simulated and measured current mirror ratio ($I_{out}/I_{in}$) over input current density for pre-stress and post-stress circuits.

6.1 Simulated load line of a device in a driver amplifier showing schematic representation of the device aging regions, SOA, and failure boundaries.

6.2 Calculated aging rate due to mixed-mode stress at room temperature with (a) a surface showing a full stress output plane and (b) cut-lines showing model-to-data fit at $J_E = 1 \text{ mA}/\mu\text{m}^2$ and at $V_{CB} = 3 \text{ V}$.

6.3 Hot carrier aging rate measured as a function of stress emitter current density at $V_{CB} = 0 \text{ V}$. A power law current density is observed as expected from the Auger-hot-carrier mechanism.

6.4 Combined hot carrier aging rate for the mixed-mode and high-current stress mechanisms at room temperature. Mixed-mode dominates at high voltages and moderate current, while high-current becomes apparent beyond peak-$f_T$ current density.

6.5 Normalized hot carrier aging rate showing a linear dependence of device aging on emitter perimeter.
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.6</td>
<td>Measurement setup for RF stress characterization.</td>
</tr>
<tr>
<td>6.7</td>
<td>Simulated RF stress showing how instantaneous age rate changes with the device bias $V_B$ and $V_C$.</td>
</tr>
<tr>
<td>6.8</td>
<td>Normalized hot carrier aging rate showing a linear dependence of device aging on emitter perimeter.</td>
</tr>
<tr>
<td>6.9</td>
<td>Simplified schematic of driver circuit used for aging simulation.</td>
</tr>
<tr>
<td>6.10</td>
<td>Circuit response degradation over time due to stress of a 20 GHz CW input signal with $P_{in} = 10 \text{ dBm}$. The individual bias shifts over time for devices $Q_0$, $Q_2$, and $Q_4$ are shown in the inset.</td>
</tr>
</tbody>
</table>
SUMMARY

The objective of this work is to analyze the mechanisms of breakdown and aging in silicon-germanium (SiGe) heterojunction bipolar transistor (HBT) devices to help maximize the deliverable, reliable performance of high-frequency SiGe HBT circuits and systems. Conventional circuit design methodologies follow conservative and overly simplistic guidelines when considering reliability, but with a more detailed understanding of device reliability, these conventional design limitations may be pushed to enhance performance. This broad objective is explored primarily in three separate approaches, namely the investigation of SiGe HBT aging mechanisms, the development of predictive aging models, and the design of novel SiGe HBT devices with enhanced breakdown performance. It is the intention of the author that device engineers may obtain an enhanced understanding of the physics of aging in SiGe HBTs and that circuit designers may obtain an appreciation of the benefits of “reliability-aware” circuit design methodologies. Building a link between these two groups to leverage an understanding of physics to enhance circuit and system design methodologies is the overall goal of this work. These goals will be addressed in the manner enumerated below.

1. An introduction to SiGe HBT technologies detailing the concepts necessary to understand the work is established, including a general overview of the use, fabrication, and physics of SiGe HBTs.

2. The design of novel SiGe HBT profiles targeting improved breakdown for enhanced high-voltage and reliability performance. This work was first published in the IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)
3. Exploration of the mechanics and various sources of long-term aging effects observed in SiGe HBTs is presented. Hot-carriers created under both high-current and high-field conditions create defects at oxide interfaces in SiGe HBTs. This work was first published in the International Semiconductor Device Research Symposium (ISDRS) © 2013 [3] and later extended in TED © 2015 [4].

4. Development of a physics-based compact model and simulation framework for incorporating predictive aging capabilities in a circuit simulation environment. This work was first published in TED © 2016 [5].

5. Extension of the physics-based compact model to allow comprehensive simulation of base current degradation in SiGe HBTs including analysis and validation for using the model for RF circuit design. This work was first published in the IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS) © 2018 [6].

Copyright Statement

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Georgia Institute of Technology’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

Portions of the enclosed text and many of the figures may become IEEE copyrighted material in the future.
CHAPTER 1

INTRODUCTION

1.1 Origin and History of the Problem

The use of advanced silicon-germanium (SiGe) bipolar CMOS (BiCMOS) platforms for high-speed radio-frequency (RF) and mm-wave applications has driven a rapid development of SiGe heterojunction bipolar transistor (HBT) technologies as an alternative option for traditionally III-V transistor dominated applications, offering the superior integrability, yield, and cost benefits associated with silicon manufacturing. The vertical scaling required to achieve improved high-frequency performance unfortunately comes at the cost of breakdown voltage and operating current density. Current state-of-the-art SiGe BiCMOS platforms have been pushed to achieve unity-gain cutoff frequencies ($f_T$) above 500 GHz and unity-power-gain frequencies ($f_{MAX}$) as high as 720 GHz at room temperature using conservative lithography [7], and additional studies exploring the performance limits of SiGe HBTs show the capability of SiGe HBT platforms for pushing towards terahertz (THz) performance limits [8,9]. These performance advancements will enable the growth and expansion of low-cost silicon IC platforms into new applications such as mm-wave radar, THz imaging, and multi-Gb/s wireless networks.

Given that breakdown voltage must be sacrificed to achieve such high-frequency performance, the practical application of these advanced SiGe technologies are often limited by low breakdown voltages that compromise their RF power generation capabilities. In SiGe HBT technologies, conventional collector design methodology calls for increased doping in the selectively implanted collector (SIC) region to boost $f_T$ by delaying the onset of high-current effects (i.e., base push-out, Kirk effect, and
heterojunction barrier effects); however, this leads to increased impact ionization in the collector-base (CB) junction and degraded breakdown voltages. This fundamental trade-off leads to the well-known “Johnson limit” [10].

As maximum operation voltages decrease, circuit designers are driven to push against classical foundry defined safe operating area (SOA) boundaries. The SOA will define a region of biases for which stable device performance can be expected for a defined lifetime (most commonly 10 years). These SOA boundaries, however, are defined at the single-device level and based on dc measurements, presenting a fairly simplistic view of device reliability, and abstracted to the circuit level, such rules are applied uniformly for each transistor in a circuit regardless of target application. When considering the space between SOA boundaries and device electro-thermal failures limits, circuit performance is potentially being thrown away for the sake of over-conservative reliability concerns.

Two approaches present themselves for tackling the issues this speed/power trade-off presents to designers at the circuit level. First, novel device structures that improve and help decouple this trade-off must be investigated, loosening the constricting SOA boundaries designers must observe. Second, a more thorough understanding and modeling of the mechanisms determining device lifetime must be achieved. This would allow circuit designers to evaluate reliability on a case-by-case basis and selectively violate foundry defined SOA boundaries and rules in ways that do not effect long-term circuit reliability.

1.2 Organization

The goal of this work is to investigate both design and modeling techniques at the device level to address the concerns of a shrinking SOA in SiGe BiCMOS technologies. Chapter 2 provides an introduction to SiGe BiCMOS technologies including their development, scaling, and basic reliability concerns. Chapter 3 explores the
design, simulation, and measurement of a novel superjunction collector SiGe HBT for enhanced breakdown performance. Chapter 4 investigates the aging mechanisms driving the long-term reliability of SiGe HBTs. Chapter 5 develops a modeling framework for the simulation of aging effects in SiGe HBTs. Chapter 6 expands upon the use of aging models for comprehensive simulation of aging effects in SiGe HBTs and their use in a circuit design environment.
CHAPTER 2

SILICON-GERMANIUM BICMOS TECHNOLOGIES

For much of the history of solid-state electronics, silicon has been the semiconductor of choice due to the large number of benefits it presents relative to other materials including but not limited to its abundance, ease of producing high-purity crystals, moderate band gap contributing to low leakage currents, and presence of a stable and easily grown oxide. Bulk silicon technologies have since enjoyed the rapid exponential growth and scaling as predicted by Gordon Moore back in 1965 [11], and while the primary result of continuous scaling has been the increased density of digital logic in CMOS technologies, reduced dimensions leading to decreased transit times have enabled the use of silicon technologies in RF applications. When targeting high-frequency RF and mm-wave applications, however, CMOS technologies have been shown to be limited relative to bipolar technologies by back end of line (BEOL) parasitics [12]. Coupled with the intrinsic advantages bipolar technologies offer compared to CMOS such as noise performance and output conductance [13], bipolar devices are an appealing candidate for high-frequency applications, and this role is one that SiGe HBTs have grown into over the past 30 years.

2.1 Technology Overview

2.1.1 Silicon Germanium Heterojunction Bipolar Transistors

At a basic level, a SiGe HBT is a standard Si homojunction BJT with a graded Ge profile incorporated into the base region. A schematic cross-section and measured SIMS profile of a representative SiGe HBT showing the location of the introduced Ge is shown in Fig. 2.1. The introduction of this Ge effectively lowers the band gap in the base region, and through clever band gap engineering via the introduced Ge
Figure 2.1: (a) Schematic cross-section and (b) measured SIMS profile of a representative first-generation SiGe HBT (after [13]).

profile, the SiGe HBT can overcome the limitations of a standard Si BJT.

We can consider the primary ways in which this band gap engineering affects device performance by analyzing the effect of the Ge on the band diagram, as seen in Fig. 2.2. We take $\Delta E_{g,Ge}(0)$ and $\Delta E_{g,Ge}(W_b)$ to be the Ge-induced band gap offsets at the emitter-base (EB) and collector-base (CB) edges of the quasi-neutral base, respectively, and define

$$\Delta E_{g,Ge}(\text{grade}) = \Delta E_{g,Ge}(W_b) - \Delta E_{g,Ge}(0)$$  \hspace{1cm} (2.1)$$
as the change in band gap across the base due to the graded Ge profile. In an $nnp$ device, this change in the band gap materializes in the conduction band within the base, and consequently, the energy barrier for electrons at the EB junction is lowered. This results in an enhancement of the collector current relative to a Si BJT for a fixed voltage, and as the EB valence band barrier is unaffected, the base current for the Si BJT and SiGe HBT should be comparable, additionally providing an enhanced current gain $\beta$. Coupling a spatially dependent band gap into the the generalized Moll-Ross relation, we can define a SiGe current gain enhancement factor as
Figure 2.2: Energy band diagram for an npn Si BJT and SiGe HBT showing the effects of Ge induced band offsets and band gap grading (after [13]).

\[
\frac{\beta_{\text{SiGe}}}{\beta_{\text{Si}}} \bigg|_{V_{BE}} = \frac{\tilde{\gamma} \tilde{\eta} \Delta E_{g,Ge(\text{grade})}/kT e^{\Delta E_{g,Ge(0)}/kT}}{1 - e^{-\Delta E_{g,Ge(\text{grade})}/kT}}
\]  

(2.2)

where \(\tilde{\gamma}\) and \(\tilde{\eta}\) are the position averaged enhancement factors for the effective density of states and diffusivity in the base region, respectively [13]. Additionally, with a graded Ge profile, a field is effectively generated across the neutral base in equilibrium conditions due to the changing band gap across the base. This field naturally weights the intrinsic free carrier density towards the CB junction, which allows a graded-base SiGe HBT to resist back depletion of base relative to a Si BJT. This results in an enhancement in the early voltage \(V_A\) given in [13] by

\[
\frac{V_{A,\text{SiGe}}}{V_{A,\text{Si}}} \bigg|_{V_{BE}} = e^{\Delta E_{g,Ge(\text{grade})}/kT} \left[ \frac{1 - e^{-\Delta E_{g,Ge(\text{grade})}/kT}}{\Delta E_{g,Ge(\text{grade})}/kT} \right].
\]  

(2.3)

One of the key benefits to these enhancements that a Ge profile provides is that key dc performance metrics of the SiGe HBT can be effectively decoupled from the
design of the base region. This allows for the use of high base doping to reduce the intrinsic base resistance, providing valuable improvements to device speed and noise. Additionally, the field created across the base region will encourage drift transport in favor of diffusion across the neutral base, providing an enhancement in the base transit time $\tau_b$ given in [13] as

$$\tau_{b,\text{SiGe}} \frac{\tau_{b,\text{Si}}}{2} \frac{kT}{\eta} \Delta E_{g,\text{Ge}}(\text{grade}) \left\{ 1 - \frac{kT}{\Delta E_{g,\text{Ge}}(\text{grade})} \left[ 1 - e^{-\Delta E_{g,\text{Ge}}(\text{grade})/kT} \right] \right\}. \quad (2.4)$$

Such improvements to device performance through the use of band gap engineering in the SiGe HBT make it a promising candidate for high-frequency applications.

2.1.2 SiGe BiCMOS History and Development

The challenge in pushing SiGe HBT technology development for use in RF and mm-wave applications lies not only in process development, namely learning how to reliably create devices with increasingly smaller feature sizes and tighter tolerances, but also process integration into a bipolar CMOS (BiCMOS) technology, where standard CMOS and SiGe HBT devices are fabricated simultaneously. Such practice is standard for SiGe HBT technology platforms, as the integration advantage of being able to place both digital and RF components side by side on the same die presents clear advantages over similar performance III-V technologies, thereby allowing for more compact and lower cost designs. For a modern, highly scaled SiGe HBT structure such as in Fig. 2.3 (after [14]), technologists create road maps for future performance improvements as well as general scaling rules for the relative change in each parameter needed to achieve a desired performance level [9, 14, 15]. As scaling of a device structure advances, limiting factors for device performance shift as device dimensions and doping levels change by orders of magnitude, and new structures are developed to counteract these limiting factors as a result.
Due to the maturity of silicon processing technology, generational growth of SiGe BiCMOS platforms has been rapid since the advent of successful growth methods of SiGe films such as ultra-high vacuum chemical vapor deposition (UHV/CVD). First-generation platforms boasting performance of 50 GHz $f_T$ and 3.3 V $BV_{CEO}$ [16] quickly gave way to second- ($f_T > 100$ GHz and 2.5 V $BV_{CEO}$ [17]) and third-generation ($f_T > 200$ GHz and 1.7 V $BV_{CEO}$ [18]) platforms over the course of a few years, though each technology generation still finds use in a variety of applications. More recent fourth-generation technologies boast $f_T/f_{MAX}$ values of 300/500 GHz [19, 20] and emerging technologies are quickly approaching THz performance [7], all while maintaining $BV_{CEO}$ values above 1.5 V.

With each new technology generation in SiGe, new process innovations have been necessary to keep up with performance demands. First generation platforms and the transition to the second generation saw the development of the double-polysilicon self-aligned (DPSA) structure for high-performance bipolar s [21], moving from "base-during-gate" to "base-after-gate" BiCMOS integration for managing thermal cycle
Figure 2.4: Pairs of $f_T/BV_{CEO}$ values for SiGe BiCMOS platforms from major foundries. Constant $f_T \times BV_{CEO}$ product lines are shown for reference.

issues [22,23], and the integration of carbon doping for the suppression of base-dopant out-diffusion [24, 25]. As scaling advanced, base profiles were required to have high doping and narrower widths, and the implantation of extrinsic device regions became problematic due to the associated introduction of interstitials that enhance boron diffusion. The emergence of the raised extrinsic base structure [18] came as a solution to this problem and is a notable feature of any modern SiGe BiCMOS technology. Furthermore, the raised extrinsic base allows for greater physical separation of the extrinsic base and SIC regions, reducing the CB overlap capacitance, a major concern in highly scaled devices [26].

With the rapid development of SiGe BiCMOS technologies, many of the major foundries have adopted it as a platform for high-speed mixed-signal applications. In Fig. 2.4, a wide array of $f_T/BV_{CEO}$ pairs for SiGe BiCMOS platforms from GLOBALFOUNDRIES [17–19,27], IHP [7,20,28–30], TowerJazz [31,32], STMicroelectronics [33–36], and Texas Instruments [37,38] show off the level of versatility and proliferation of SiGe BiCMOS technologies.
Figure 2.5: Schematic view of doping and germanium profile changes associated with typical vertical scaling in a SiGe HBT.

2.2 Collector Scaling and Breakdown Mechanics in SiGe HBTs

The primary motivation for vertical scaling in a SiGe HBT is typically tuning for enhanced high-frequency performance, generally characterized by $f_T$ and $f_{MAX}$, and we can reduce this problem of increasing $f_T$ to one of simply decreasing transport distances and transit times. When we decrease these distances, such as by decreasing the thickness of the intrinsic emitter and collector layers as in Fig. 2.5, junction capacitances will inevitably increase. To maintain high-$f_T$ performance, operating current density $J_C$ must increase in tandem to minimize the charging time effects of depletion and parasitic capacitances that reduce RF capabilities. The opponent to increasing $J_C$, though, is the onset of high-current effects such as the Kirk effect when injected carrier densities exceed the local ion density in the CB space charge region (SCR) [39]. The onset of high-current effects will collapse the CB field and increase the neutral base width, significantly degrading both $dc$ and $ac$ device characteristics. With base push out in action, additional mechanics such as heterojunction barrier
effects can also occur [40]. The push for higher $f_T$ values then requires increasing the selectively implanted collector (SIC) doping to delay onset of the Kirk effect by requiring a higher $J_{C,Kirk}$ to begin base push-out effects.

As a result of increasing SIC doping, the CB SCR width will decrease. This will not only lead to reduced CB transit time but also decreased breakdown voltage due to a larger electric field strength. Shown well in Fig. 2.4, this fundamental trade-off with breakdown leads to the well known “Johnson limit” [10, 41, 42]. With simple approximations of velocity saturation at $v_{sat}$ and breakdown at a critical electric field strength $E_{crit}$, values for $f_T$ and $BV_{CEO}$ must scale inversely to one another with the CB SCR width, as the CB transit time and $V_{CB,crit}$ decrease and increase, respectively. Of note, however, in Fig. 2.4 is the divergence from the classical “Johnson limit” for highly scaled nodes with $f_T > 100$ GHz, where length scales are in a regime such that non-equilibrium transport effects must be considered. In particular, breakdown in a highly scaled device cannot be considered solely as a field-based effect, and carrier energy becomes the primary consideration for breakdown. In these highly scaled SiGe HBT devices, more complex breakdown behavior must account for non-local impact ionization dependent on carrier kinetic energy [43] and velocity overshoot behavior [44]. Simulation of these effects is done either by advanced hydrodynamic energy-transport models that consider the average energy of the carrier ensemble [45] or by direct solution of the Boltzmann transport equation (BTE) [46].

The process of bipolar breakdown typically entails the coupling of the current gain and impact ionization mechanisms. Illustrated in Fig. 2.6 (after [13]), impact ionization will generate an amount of carriers $(M-1)I_{n,in}$ in the CB junction, where $M$ is the avalanche multiplication factor, defined as the ratio of the current entering and exiting the CB space-charge region. During operation, the CB electric field will sweep these generated carriers across the base, where $\beta$ times more carriers will be injected from the emitter. These secondary carriers can subsequently induce additional impact
ionization events. When this feedback mechanism progresses such that $\beta \times (M - 1) > 1$, avalanche behavior will begin and the device will enter breakdown. When the device is in open base operation, all impact ionization generated carriers will be multiplied, yielding the worst case open-base collector-emitter breakdown voltage $BV_{CEO}$, and in resistive operation where some secondary carriers can exit via the base a higher value $BV_{CER}$ is achieved [47].

2.3 Reliability Effects in SiGe HBTs

To be proven “reliable,” any device technology (SiGe and others) must show that under typical operating conditions, devices, circuits, and systems constructed using the technology will not degrade to a point such that they fail during the lifetime of the system. To qualify the reliability of a SiGe BiCMOS system, the building-block components encompassing passive elements, interconnects and metalization, and both CMOS devices and SiGe HBTs must each be tested individually.

Qualifying the reliability of a SiGe HBT is typically performed by a series of accelerated stress tests using over-stress conditions, or operation of a device in extreme measurement conditions outside of a typical SOA such as high-current/voltage bias, high/low temperatures, or some combination thereof. Afterwards, results are extrapolated to typical operating conditions, and an SOA is defined. While breakdown and instability are a concern for a device operated outside a traditional SOA
Figure 2.7: Example mixed-mode stress response showing (a) base current degradation in the Gummel characteristic and (b) current gain $\beta$ degradation.

definition, device operation at or near the boundaries of this SOA can lead to device aging effects. Aging effects, sometimes referred to as “soft” breakdown, are a result of physical degradation of the device structure over time, which can lead to changes in a device’s performance characteristics, as in Fig. 2.7.

Conventional ways of characterizing aging mechanisms has typically fallen into the categories of high forward-current density stress studies [48–54], reverse-EB stress studies [23, 55–58], and more recently, mixed-mode stress studies [59–61]. Many of these studies, particularly those concerned with high-current density stress, have investigated electromigration-induced degradation of the emitter contact, which will alter the emitter resistance and affect the device operation at high-current bias [51, 52, 54]. The primary result of device aging, however, is the degradation of oxide interfaces over time due to the de-passivation of dangling bonds and the freeing of hydrogen, which typically follows a power law time dependence [62]. The de-passivation of these dangling bonds is done through interaction with a high-energy or “hot” electron or hole. These dangling bonds then serve as a trap state and a center for Shockley-Read-Hall (SRH) recombination. The increase in SRH recombination can lead to the development of a significant leakage current, sharply degrading the forward current gain $\beta_F$ over time, particularly at low bias levels. A representative
mixed-mode stress response showing the degradation of the base current and current gain is provided in Fig. 2.7. With the shrinking of breakdown voltages with device scaling, aging effects due to operation at and beyond standard SOA boundaries becomes more significant, and the need to develop an understanding of the mechanisms of aging becomes clear.
CHAPTER 3

VERTICAL SUPERJUNCTION COLLECTORS FOR ENHANCED BREAKDOWN PERFORMANCE IN SIGe HBTS

This chapter reviews original research to introduce and develop the concept of a vertical superjunction collector structure in a SiGe HBT platform. This analysis resulted in two published works [1,2] and provides a look at the design and integration of a superjunction collector in a SiGe BiCMOS technology.

3.1 Motivation

The use of advanced SiGe BiCMOS platforms for high-speed RF and mm-wave applications has driven a rapid development of SiGe HBT technologies as an alternative in traditionally III-V transistor dominated applications, offering the superior integrability, yield, and cost benefits associated with silicon manufacturing. These benefits have helped pushed SiGe BiCMOS to become a major player in applications such mm-wave radar, THz imaging, and multi-Gb/s wireless networks, with state-of-the-art SiGe technologies now achieving cutoff frequencies $f_T$ above 500 GHz and maximum oscillation frequencies $f_{MAX}$ as high as 720 GHz at room temperature for conservative lithology [7]. Given that breakdown voltage must be sacrificed to achieve such high-frequency performance, the practical application of these advanced SiGe technologies are often limited by low breakdown voltages, which compromises their RF power handling capabilities.

In SiGe HBT technologies, increased doping in the SIC region is used to boost $f_T$ by delaying the onset of high-current effects such as base push-out (i.e., Kirk effect);
however, this leads to increased impact ionization in the CB junction and degraded breakdown voltages. Finding a way to decouple the $f_T/BV_{CEO}$ trade-off (i.e., the Johnson limit) in SiGe HBTs to enable the development of further advanced, but widely usable, technologies remains an important question. Taking a closer look at the traditional understanding of the Johnson limit, the maximum allowable breakdown voltage and cutoff frequency pair for a given geometry is determined by the saturation velocity $v_{sat}$ and breakdown field $E_{breakdown}$ of the material. When considering a silicon-based platform, this figure-of-merit (FOM) provides an approximate limit of $300 \text{ GHz \cdot V}$, given by

$$f_T \times BV \approx v_{sat} \times E_{breakdown} = 10^6 \frac{\text{cm}}{s} \times 0.3 \frac{\text{MV}}{\text{cm}} = 300 \text{ GHz \cdot V}. \quad (3.1)$$

In a highly scaled device however, assumptions of saturation velocities may no longer be generally valid due to device dimensions approaching the mean-free-path length (i.e., $W_B \gg \lambda$). Additionally the non-local energy dependence of impact ionization events partially invalidates the use of $E_{breakdown}$. These deficiencies in the traditional FOM give promise that there is room for improvement over the classically understood limits.

Previous studies have proposed and shown simulated results for the use of a “superjunction collector” to enhance the breakdown performance of SiGe HBTs [63–65]. The superjunction collector design uses a series of alternating $pn$-junctions embedded in the CB SCR to manipulate the carrier temperature profile. In this study, a practical superjunction collector device is developed based on a commercially available SiGe HBT platform intended for wireless power amplifier (PA) design [27]. Both TCAD simulations and measurements are used to show the design parameterization of a superjunction collector and demonstrate the practical limits of the technique.
3.2 Breakdown and Superjunction Mechanics

Breakdown of the CB junction in a SiGe HBT under normal operating conditions is traditionally caused by runaway carrier generation via impact ionization, or avalanche generation, in the CB junction, which can combine with the intrinsic gain mechanism at the EB junction to form a feedback loop and thereby create breakdown conditions. Manipulating the impact ionization behavior through clever design of the collector should therefore allow for increasing the device breakdown voltage. As current gain is a useful feature of SiGe HBT operation and couples to many other device performance metrics, changing the breakdown voltage by decreasing current gain can lead to undesirable effects and is thus not viable as an optimization approach.

Impact ionization results from the acceleration via the electric field of carriers to high energies, which can then ionize the lattice and create additional electron-hole pairs. These secondarily generated carriers can induce additional impact ionization, leading to an avalanche process and causing the device to enter breakdown. To delay the onset of breakdown, the probability of carrier ionization must be reduced by modifying the carrier energy profiles in the CB SCR. Using a calibrated TCAD model of an $nnp$ SiGe HBT, hydrodynamic transport simulations can be used to analyze the energy profile of carriers in the CB junction via the carrier temperature, as shown in Fig. 3.1. As carriers must travel some distance in the electric field to reach the energy threshold for impact ionization, the location of peak electron temperature does not match the location of peak electric field, creating a so-called “dead space” region [66]. Through analysis of this non-local effect, insight can be given as to how to exploit the dead space region to improve avalanche breakdown behavior. The electron ionization coefficient $\alpha_n$ calculated using the Okuto-Crowell model in these simulations is given by [67] as

$$\alpha_n(E_{\text{eff}}) = a \cdot (1 + c(T_n - T_0)) E_{\text{eff}}^n \cdot e^{-\left(\frac{b(1+d(T_n-T_0))}{E_{\text{eff}}^n}\right)^\delta} \quad (3.2)$$
where $a$, $b$, $c$, $d$, and $δ$ are fitting parameters. The non-local dead space effects are accounted for in the use of the electron temperature $T_n$ to calculate an effective electric field $E_{n eff}$ [67]. Because the peak in carrier temperature is caused by electrons traveling through the dead space region, lowering the electric field near the CB metallurgical junction will decrease the peak $T_n$. As $α_n$ has an exponential dependence on $T_n$, a small decrease in $T_n$ can lead to a significant decrease in $α_n$. Additionally if the electric field profile can be re-shaped without modifying the CB SCR width, the breakdown voltage can be increased with minimal effect on $ac$ performance, since the CB transit time is largely determined by the SCR width when electric fields are sufficiently high to reach velocity saturation.

Superjunction techniques have historically been applied to semiconductor power devices, as the superjunction structure enables a high doping level for reduced on-state resistance $R_{on}$ and the sustainment of an electric field across a widened SCR under high bias, improving upon the traditional $R_{on}$/breakdown voltage tradeoff [68–70]. By
replacing a uniformly, lowly doped drift region in a device with a series of alternating 
$p/n$ stripes or a superjunction, as in Fig. 3.2, the depletion edges of each stripe will 
merge laterally to fully deplete the drift region when a sufficient voltage is applied. 
The unique structure of these stripes will cause the potential to drop uniformly across 
the drift region, enabling the sustainment of a large electric field. A similar method 
using alternating $p^+/n^+$ layers has been used to manipulate the electric field and take 
advantage of velocity overshoot behavior in AlGaAs/GaAs HBTs to create ballistic 
collection transistors [71]. This methodology can be further modified to place a 
superjunction structure vertically in CB SCR of a SiGe HBT in order to “flatten” 
the electric field profile, improving breakdown performance [63,64]. 

In an $n pn$ device, the placement of a thin $p^+$ layer inside the CB SCR will re-
distribute the electric field towards the newly introduced dopants, and an equivalent 
compensating $n^+$ layer placed adjacent to the $p^+$ layer will maintain the SCR width,
as a net zero charge is introduced. By using a series of superjunction layers as in Fig. 3.3, the CB electric field profile can be reshaped to minimize avalanche generation while maintaining a constant SCR width. As used in traditional superjunction devices for high-power applications, the ideal field distribution for maximizing breakdown is uniform, distributing the potential drop across the entire SCR. The field and electron temperature profiles of an example device using Gaussian-shaped superjunction layers are compared against a standard collector design in Fig. 3.4. The addition of the superjunction layers leads to about a 20% decrease in both the peak electric field and peak electron temperature. Due to the exponential dependence of the ionization coefficient $\alpha_n$ on the carrier temperature in (3.2), a significant decrease in $\alpha_n$ can be seen in Fig. 3.5, with more than an order of magnitude drop for the simulated conditions.

The gains in breakdown performance due to the vertical superjunction collector
Figure 3.4: Simulated electric field and electron temperature in the CB SCR for both standard collector (dashed line) and superjunction collector (solid line) designs demonstrating the dead space region.

Figure 3.5: Simulated electron temperature and electron ionization coefficient in the CB SCR for both standard collector (dashed line) and superjunction collector (solid line) designs.
are determined by the degree to which the electric field distribution is reshaped and made to uniform, which is in turn set by the amount and location of charge implanted in the CB SCR via the superjunction layers. The limits of performance gains and restrictions to superjunction design are explored through a parameterized study in the following section.

3.3 Vertical Superjunction Collector Design Parameterization

The exploration of superjunction design presented here follows the influence of various superjunction layer parameters on the key device performance metrics $BV_{CEO}$ and $f_T/f_{MAX}$. Simulation data presented in this section uses a model based on a first-generation SiGe HBT device with a $BV_{CEO}$ of 6 V and $f_T/f_{MAX}$ of 35/80 GHz [27]. Using this model, two sets of rectangular $p/n$ layers were placed within the collector region as in Fig. 3.3 and the layer doping levels, location, and shape (i.e. width and relative doping levels) were modified to observe the effects of each parameter on overall device performance. A standard design used in each parametric sweep has a superjunction structure where the deep (farthest from the CB junction) and the shallow (closest to the CB junction) layers are matched, with a peak doping level of $1.25 \times 10^{17} \text{ cm}^{-3}$ and equal width, and are located 0.18 $\mu$m from the CB metallurgical junction.

3.3.1 Doping Level

Since the primary driver of electric field redistribution is the volume of charge implanted within the depletion region, the doping level of superjunction layers will logically have the most direct effect on breakdown enhancement in a superjunction collector. As seen in Fig. 3.6, two clear behaviors of the superjunction become apparent. For fixed layer thickness and position, $BV_{CEO}$ will increase with increasing
doping until a point where $BV_{CEO}$ will start to degrade. The initial increase is expected as the CB electric field is more strongly redistributed with increasing doping, and the decrease in $BV_{CEO}$ is caused by the creation of a secondary peak in carrier temperature and impact ionization rates, as seen in Fig. 3.7. Additionally, peak $f_T$ values decrease monotonically with increasing doping level. The primary driver of this decrease is a premature onset of the Kirk effect and conductivity modulation effects caused by the superjunction layers.

These conductivity modulation effects are a major concern for design at low-$V_{CB}$. Conductivity modulation occurs when a high level of injected carriers leads to a quasi-saturation regime where the CB junction is effectively forward biased and is essentially the low-$V_{CE}$ equivalent of the Kirk effect, where carriers are not moving at $v_{sat}$ [72]. At low $V_{CB}$ bias, the $p^+$ layers will still alter the magnitude of the field at the CB metallurgical junction even if they are undepleted. This allows for the CB field to collapse at much lower current densities, as seen in Fig. 3.8(a), and carriers
Figure 3.7: Simulated electron temperature (solid lines) and electron ionization coefficient $\alpha_n$ (dashed lines) in the CB SCR for a range of superjunction doping values showing the development of a secondary ionization peak with increasing doping.

will begin to accumulate at the front edge of the superjunction layer closest to the base, as seen in Fig. 3.8(b). Due to this behavior, quasi-saturation behavior will begin at a lower current density within the device, limiting $f_T/f_{\text{MAX}}$.

The severity of the conductivity modulation effect can be assessed through transit time analysis focusing on the CB depletion region. Transit time analysis can be used to monitor the contribution of individual regions within the device towards the overall transit time $\tau_{ec}$ \cite{73}. The transit time is found by integrating over the charge transport path using the equation

$$\tau_{ec} = \frac{1}{2\pi f_T} = \frac{q}{2\pi f_T} \int_S \frac{\Delta n(s)}{|\Delta j_n(s)|} \bigg|_{V_{CE}} \cdot ds$$ \hspace{1cm} (3.3)

where $\Delta n$ is the small-signal differential in electron density, $\Delta j_n$ is the small-signal differential in electron current density, and $S$ is the carrier transport path. A plot
Figure 3.8: Simulated (a) electric field profiles showing an early collapse and
(b) electron density profiles for standard collector and superjunction collector
designs for $V_{BE}$ ranging 0.8–0.85 V, where conductivity modulation effects set in.
The field collapse and subsequent accumulation of carriers at the superjunction
dge leads to quasi-saturation at low $V_{CB}$ bias and the start of roll-off in $f_T/f_{max}$.
The location of the $p^+$ superjunction layers is marked by dotted lines.
Figure 3.9: Accumulated transit time $\tau_{ec}$ through standard and superjunction collector devices for $V_{CB} = 0 - 1$ V. The field collapse and carrier accumulation due to conductivity modulation cause a sharp increase in $\tau_{ec}$ between the CB metallurgical junction and the superjunction layers, marked by the dotted lines, which improves with increasing $V_{CB}$.

Comparing the accumulated transit time of standard and superjunction collector devices is shown in Fig. 3.9. At higher applied $V_{CB}$ values where the $p^+$ layers are fully depleted and the peak CB field at the metallurgical junction, this quasi-saturation behavior is mitigated, but the altered electric field distribution within the CB SCR still leads to a minor accumulation of carriers compared to a standard collector design. This causes early onset of the Kirk effect and a slight decrease in peak $f_T/f_{\text{MAX}}$ values. As seen in Fig. 3.6, the relative change of peak $f_T$ values with collector-base voltage ($V_{CB}$) increases with increased $p^+$ layer doping due to this increased accumulation. Mitigating the effects of this enhanced quasi-saturation behavior while maximizing breakdown performance mandates a careful balance between the doping level and placement of the superjunction layers within the collector. By considering the $f_T \times BV_{CEO}$ product to explore the tradeoff, we can find a maximum value and ideal superjunction design at a doping value backed off of the peak $BV_{CEO}$ point.
3.3.2 Placement

Due to the degraded $ac$ performance presented by an undepleted superjunction structure, its placement within the collector can have a large impact on overall device performance. As the superjunction layers are moved closer to the CB metallurgical junction and the peak electric field region, the superjunction will become fully depleted at lower values of $V_{CB}$. Ideally, the entire superjunction structure should be depleted at zero $V_{CB}$ bias to minimize quasi-saturation behavior. As the superjunction is moved closer to the CB junction, the amount of superjunction layer charge the depletion region can support without developing a secondary ionization peak is reduced. If the superjunction layers are placed too close to the natural position of peak electric field and carrier temperature, those peaks will instead be enhanced and degrade breakdown behavior, but at the same time, the accumulation of carriers on the front side of the superjunction decreases as it moves closer to the CB metallurgical junction. This tradeoff between quasi-saturation degradation and breakdown enhancement can be seen clearly in Fig. 3.10 where the location of the superjunction...
layers relative to the CB metallurgical junction is swept. Again, at sufficiently high $V_{CB}$ bias where the superjunction is fully depleted, quasi-saturation is not a concern and $f_T/f_{max}$ values show minimal change with layer placement, but at low $V_{CB}$, a clear inflection point can be seen around a distance of 0.2 $\mu m$ where the superjunction layers begin to move entirely into the zero-bias CB depletion region.

### 3.3.3 Shape

In traditional superjunction power devices, uniformly doped, symmetric $p/n$ stripes are created parallel to the direction of current flow to evenly distribute the electric field across the drift region and maximize breakdown. In the context of the vertical processing flow of a modern SiGe HBT collector, this method of depletion region manipulation cannot be achieved with superjunction layers perpendicular to the direction of current flow. It is possible, however, to use a series on non-uniformly doped, asymmetric superjunction layers to linearize the potential drop across the SCR and make the electric field more uniform. In our two-layer superjunction collector structure, the ratio of peak doping level between the deep and shallow layers ($N_{deep}/N_{shallow}$ in Fig. 3.3) was varied while keeping the total integrated charge volume constant. $BV_{CEO}$ will increase as a larger portion of the net integrated charge in the superjunction layers is shifted to the deeper layer as seen in Fig. 3.11, though a limit exists where the highly doped deeper layer can cause an excessive peak in the electric field and lead to a decrease in breakdown voltage. This behavior is evident in the black curve in Fig. 3.11 for equal width superjunction layers. However, distributing the integrated charge over a larger area, as evidenced by the increase of $BV_{CEO}$ with increasing asymmetry of layer width in Fig. 3.11, leads to more gradual changes in the electric field and smaller secondary ionization peaks. Changes in $ac$ performance metrics with the integrated charge distribution are negligible when considering a fully depleted superjunction structure, but changes in quasi-saturation behavior must be
Figure 3.11: Simulated values of $BV_{CEO}$ for varying doping ratio $N_{deep}/N_{shallow}$ between the deep and shallow superjunction layers. With the shallow layer width ($W_{shallow}$) held constant, variation of the width of the deep layer ($W_{deep}$) shows an increase in $BV_{CEO}$ with increasing width.

considered to prevent strong accumulation behavior triggered by too highly-doped or too wide superjunction layers.

### 3.4 Practical Superjunction Fabrication

While the vertical superjunction collector offers an intriguing method to boost and partially decouple the breakdown and high-frequency performance characteristics of a SiGe HBT (effectively “violating” the classical Johnson limit), the practicality of inserting a superjunction structure without disrupting the process flow of a modern SiGe BiCMOS technology platform remains a question for commercial implementation. Simple modification of pre-existing processing steps without the need for additional masksets would be the ideal method for superjunction integration, and with that objective, two separate methods for exploring the practicality of integration are apparent. In both methods, the charge compensating dopant species is integrated into the background SIC doping level to decrease the number of layers instead of using
alternating $p$-$n$ regions (i.e. increased dose of a phosphorus SIC implant in an $n pn$ device). In addition to simplifying the structure, this removes the need to alternate between $n$- and $p$-type dopant species, which can be problematic. This simplified superjunction structure has been previously shown through simulation to have negligible performance difference compared to an alternating $pn$-junction structure [64]. Schematic examples of each method to be described below are shown in Fig. 3.12.

In a typical $npn$ SiGe HBT, the SiGe base layer is grown by low temperature epitaxy (LTE) on top of a $n$-type epi-layer. The base region is formed by depositing a thin boron layer simultaneous to the SiGe alloy growth, but if a series of boron layers are included during the growth of the buffer layer and Ge retrograde, an abrupt superjunction can be carefully placed within the CB SCR. Charge introduced in these $p^+$ regions can be compensated for by modifying a shallow SIC implant that covers the intrinsic buffer region.
A second method for superjunction fabrication is the use of low-energy boron implants before growing the SiGe epi-layer. For a device with SIC implants applied before SiGe epi-growth, these superjunction implants can be performed with the same blockout mask. The effectiveness of this method is limited by the thickness of the SiGe epi-layer as control over the superjunction location is restricted, but no modifications to the SiGe epi-layer growth process are necessary. Using the simplified superjunction improves the manufacturability of this design particularly, due to limitations in stacking low-energy narrow implants from a single interface. Interrupting collector-epi growth and placing superjunction implants from multiple growth interfaces could allow for greater freedom in superjunction shaping but may be undesirable in a commercial BiCMOS process flow.

The GLOBALFOUNDRIES 5PAE platform, a 0.35 µm SiGe HBT platform with a $BV_{CEO}$ of 6 V and $f_T/f_{MAX}$ of 35/80 GHz, was chosen as the model technology for this set of superjunction experiments. In the initial design of these experimental devices, a control device was fabricated to have similar performance characteristics of the process-of-record (POR) 5PAE device while maintaining a process flow conducive to experimenting with and modeling superjunction parameters. The superjunction device fabricated during the first round of experiments uses the low-energy implant method due to its simplicity in integration. The control device has a thick SiGe epi-layer, so the superjunction does not sit within the zero-bias CB SCR. This heavily degrades the frequency response at low $V_{CB}$ due to an enhancement of quasi-saturation behavior. This deep placement though does provide the opportunity to use an aggressive superjunction structure with a high level of integrated dopant due to the large area the CB SCR spreads over at high $V_{CB}$ bias. Such a deep structure allows for significant gains in breakdown performance, and as these SiGe HBTs are intended for use in PA design, any gains in voltage-swing capability and linearity are noteworthy.
Figure 3.13: Measured Gummel characteristics showing matched collector and base current densities for the control and superjunction collector devices.

### 3.5 Superjunction Measurement Results

Measurements presented here are from the 5PAE platform, the separate experimental control profile, and a superjunction collector design utilizing low-energy boron implants at the SiGe epi-growth interface. All measurements presented here were performed on a multi-finger device with emitter area $A_E = 0.8 \times 20 \, \mu m^2 \times 3$. This device geometry was chosen based on availability and the importance of multi-finger devices in compact PA design. Comparisons to the commercially-available 5PAE device are given where informative to allow for envisioning the benefits of a fully optimized superjunction design.

The Gummel characteristics of both the control and superjunction devices are ideal as shown in Fig. 3.13. As expected, the addition of the superjunction structure has negligible effect on the current drive and current gain characteristics of the device. This supports the notion that, to first order, the presence of the superjunction has no effect on operation of the EB junction. The fixed-$I_B$ output characteristics for
Figure 3.14: Measured fixed-$I_B$ output characteristics for the control and superjunction profile designs. Small signal output resistance $r_o$ for $J_B = 0.1 \, \mu\text{A}/\mu\text{m}^2$ is shown in the inset and demonstrates the effects of conductivity modulation at low $V_{CE}$ due to quasi-saturation effects.

The control and superjunction profiles, measured in Fig. 3.14, also show near ideal behavior, with the two devices having matched characteristics in the forward-active regime. The superjunction profile demonstrates quasi-saturation behavior at higher current levels with degraded output conductance at low-$V_{CE}$ due to the $p^+$ region outside the zero-bias SCR, but the difference in avalanche behavior due to breakdown at high-$V_{CE}$ is readily apparent in the “flatness” of the curves, which should present a gain to linearity performance.

To directly measure the effects of the superjunction on breakdown behavior, $BV_{CEO}$ has been measured across a range of input emitter current densities via finding the base current reversal point under fixed-$I_E$ operation. Avalanche breakdown occurs at the point when the avalanche multiplication factor $(M - 1)$ and $\beta$ product equals 1 representing a positive feedback process, and this condition can be directly
Figure 3.15: Measured $BV_{CEO}$ vs input $J_E$ curves under fixed-$I_E$ operation for the control, superjunction, and commercial 5PAE device designs. $BV_{CEO}$ was extracted from measurement via the base current reversal point. $BV_{CEO}$ was measured outside of open-base conditions by finding the base current reversal point, where the avalanche contribution to $I_B$ is equivalent to the $V_{CB} = 0$ value resulting in $I_B = 0$ [13]. As shown in Fig. 3.15, the superjunction collector design provides an extraordinary 41% (2.30 V) increase in the minimum $BV_{CEO}$ value over the standard collector design. The control device sees worsened breakdown characteristics compared to the 5PAE device, but with an optimized superjunction design applied to a commercial platform, a similar increase in $BV_{CEO}$ could be expected. As the current gain $\beta$ for each of these devices is about 100, breakdown should occur the point where $(M - 1) \approx 0.01$, and as expected, the differences in voltage at that point, shown in Fig. 3.16, matches the differences in minimum $BV_{CEO}$ in Fig. 3.15. This confirms that the gains in $BV_{CEO}$ provided by the superjunction design are due to reduced levels of avalanche generation.
Figure 3.16: Measured \((M - 1)\) curves extracted from fixed-\(I_E\) operation for the control, superjunction, and commercial 5PAE device designs. With \(\beta \approx 100\) for each device, the voltage shifts for \((M - 1) = 0.01\) match the differences in \(BV_{CEO}\).

Moving to characterization of the \(ac\) performance by the superjunction structure, the measured \(f_T\) and \(f_{MAX}\) at low \(V_{CB}\), shown in Fig. 3.17, shows an expected degradation of frequency response. This drop in measured \(f_T/f_{MAX}\) is due to the (unoptimized) deep superjunction inducing quasi-saturation within the device. The frequency response for higher \(V_{CB}\), shown in Fig. 3.18, shows an expected recovery of \(f_T\) and \(f_{MAX}\) as the superjunction layers become fully depleted and quasi-saturation is no longer a concern. Further secondary-ion mass spectrometry (SIMS) analysis of the superjunction device revealed a more aggressive than intended superjunction structure with an overly wide \(p^+\) region. While this can contribute to further \(BV_{CEO}\) enhancement, the negative effects of this design on quasi-saturation behavior and early triggering of the Kirk effect is apparent in Fig. 3.17 and Fig. 3.18.

Also shown are the measured results from a device design incorporating both a superjunction collector and a modified EB junction. A separate experiment was run
Figure 3.17: Measured $f_T$ (dashed lines) and $f_{\text{MAX}}$ (solid lines) at $V_{CB} = 1\,\text{V}$ for the control, superjunction, and modified EB superjunction devices.

Figure 3.18: Measured $f_T$ (dashed lines) and $f_{\text{MAX}}$ (solid lines) at $V_{CB} = 3\,\text{V}$ for the control, superjunction, and modified EB superjunction devices.
Figure 3.19: Figure of merit (Peak$f_{MAX}$) × (min.$BV_{CEO}$) plotted across applied $V_{CB}$ for the control, superjunction, modified EB superjunction, and commercial 5PAE device designs. The inset plot shows a Johnson curve plot for the values at $V_{CB} = 4$ V with constant product lines at 500/700 GHz-V.

aimed at reducing the EB junction capacitance while maintaining matched $dc$ performance, and the two designs were combined to test the independence of their contributions to device performance. As can be seen, the modified EB structure achieves the goal of enhanced $f_T/f_{MAX}$. Likewise, the high-current roll-off behavior matches the superjunction structure. Additionally, the $dc$ performance of the hybrid design matches the standard superjunction design, with identical Gummel characteristics, output characteristics, and breakdown.

The tradeoff between $f_T$ and breakdown voltage represents a fundamental challenge in the development of high-performance SiGe HBT technologies [10, 74]. The “superjunction” collector design detailed here presents an interesting solution for partially decoupling these two important performance metrics. Benchmarking via the product of $f_T$ or $f_{MAX}$ and $BV_{CEO}$ gives an understanding of the relative performance of these devices, as shown in Fig. 3.19. The superjunction profile suffers at
low $V_{CB}$ but shows an improved figure-of-merit at higher bias, even surpassing the 5PAE device at high $V_{CB}$. Additionally the enhanced breakdown performance, seen particularly in the “flatness” of $I_C$ in Fig. 3.14, potentially represents a significant gain for large-signal performance in PA applications.

### 3.6 Improved Superjunctions

In a second round of experiments, the control device was redesigned and additional superjunction variants using a less aggressive design were implemented. Having shown proof of concept in the first design round of the capability of the superjunction collector to enhance breakdown performance, care was taken in the second run to design a more practical superjunction that still achieves breakdown enhancement with minimal degradation of $ac$ performance. As epitaxially grown superjunction layers are able to be more tightly controlled, both implanted and epitaxial superjunction structures were designed to be fabricated. In the second design run, the control was redesigned to account for deficiencies in the design of the base region in the original run. An additional experiment performed in conjunction with the superjunction design sought to explore optimization of the EB junction for improved $f_T/f_{MAX}$. Due to tool related issues, the base region was not fabricated as desired in the first run, and the control was redesigned for a second design run to account for these issues. As the superjunction structure solely manipulates the characteristics of the CB junction, this control redesign largely did not affect the superjunction device performance, as evidenced by the apparent independence of the performance enhancements provided by the EB modification and the superjunction in the first round of experiments.

As an additional note, TCAD simulations of ideal 2-D analytical structures approximating both epitaxially grown and implanted superjunction structures were performed. The two variants showed negligible difference in regards to device breakdown and small-signal performance for equivalent superjunction layer parameters. This
means that the choice in the method for superjunction integration with SiGe HBT processing steps becomes a tradeoff between the need for precise control over the structure with an epitaxial superjunction and an easily integrated structure with an implanted superjunction.

3.6.1 Epitaxial Superjunctions

In measurement of the epitaxially grown superjunction collector devices, however, unforeseen behavior was observed indicating an issue with the design. The control device showed a measured $BV_{CEO}$ of about 5.3 V and the epitaxially grown superjunction device showed no change in the breakdown performance, while the chosen design exhibited a $BV_{CEO}$ about 1 V higher than the control device in simulations. When measuring $ac$ performance as well, the epitaxial superjunction showed a near identical $f_T$ curve to the control device, shown in Fig. 3.20, which matches the design simulations. A drop of about 10% in peak $f_{MAX}$ is seen however, which for equivalent
Figure 3.21: Measured inverse mode Gummel characteristic for the control and epitaxial superjunction devices.

$f_T$ indicates a shift in $r_b$ and/or $C_{bc}$ [13]. When observing the inverse mode Gummel characteristics of the epitaxial superjunction compared to the control device, shown in Fig. 3.21, two things of note appear. First, both the base and collector current are suppressed at high injection. Coupled with the fact that the inverse mode collector current is identical at lower injection, this change indicates a likely increase in the base resistance for the superjunction device. Second, the low injection base current shows a non-exponential behavior, likely indicating the presence of parasitic transport mechanisms such as tunneling. The control device also shows a highly non-ideal base current, but it has a pure $2kT/q$ recombination slope, likely indicating a poorly designed collector. These non-idealities only appear during inverse mode operation and were present in the first run of devices as well.

The likely cause for these degradations in the superjunction device is the superjunction layers interfering with the operation of the extrinsic base. These devices were process simulated with a 1-D simulator only, and the 2-D structure used in design
Figure 3.22: Measured $BV_{CEO}$ vs input $J_E$ curves under fixed-$I_E$ operation for the revised control and implanted superjunction device designs found via the base current reversal point.

was an idealized structure, likely inaccurate in the extrinsic regions of the device. In fabrication, it is possible that the superjunction layers could combine with the out-diffused extrinsic base. This could potentially provide an unintended resistive path for base current as well as broadening the extrinsic CB overlap capacitance, which could describe the unexpected performance degradations.

### 3.6.2 Implanted Superjunctions

The revised implanted superjunction device, however, showed excellent improvement, resolving many of the issues the overly aggressive design of the first experiment. By moving to a less aggressive superjunction design, $BV_{CEO}$ improvements were lessened, but as shown in Fig. 3.22, an approximately 2 V gain was still achieved. $BV_{CBO}$ was additionally measured for these devices. Across a full wafer, the control device measured $BV_{CBO} = 19.2 \pm 0.2$ V and the superjunction device measured $BV_{CBO} = 19.8 \pm 0.2$ V. Without a larger multi-wafer sample size, it would be difficult to
determine if the difference in $BV_{CBO}$ is due to process variation or the superjunction, but it can be stated with confidence that the superjunction inclusion does not degrade $BV_{CBO}$ in this technology.

Considering the forced-$I_B$ output characteristics, seen in Fig. 3.23, the harmful conductivity modulation effects at low-$V_{CE}$ have been corrected for, and the control and superjunction devices nearly match, with the breakdown improvement clearly evident at high $V_{CE}$. A small degradation in the knee voltage is observed however, but the cause of this change is unknown at this point in time.

Looking at the small-signal $ac$ response in Fig. 3.24, $f_T/f_{MAX}$ look much improved over the first superjunction design. There remains a reduction in peak ($f_T/f_{MAX}$) due to early onset of the Kirk effect, but the reduction remains in the $10−20\%$ range, and severe degradation due to conductivity modulation is not seen.
Figure 3.24: Measured $f_T/f_{MAX}$ for revised control and implanted superjunction devices at (a) $V_{CB} = 1 \text{ V}$ and (b) $V_{CB} = 3 \text{ V}$. 
As a final comparison to evaluate the overall effectiveness of the revised superjunction designs, the peak $f_{\text{MAX}} \times BV_{CEO}$ product is shown in Fig. 3.25 and should be compared to Fig. 3.19. The revised superjunction device maintains a FOM exceeding the control device for all bias conditions and even matches the commercial 5PAE device across bias. Given the lack of extensive process simulation and optimization taken in this design process, it would be reasonable to expect that the superjunction device can exceed the commercial device with further optimization.

3.7 Summary

This work investigated the processing methods to create a SiGe HBT with a superjunction collector for enhanced breakdown performance using layers of alternating $pn$-junctions embedded in the CB SCR. The limiting factors on superjunction performance were investigated and the trade space for successful superjunction design was
defined. Measurements of a superjunction collector designs show an impressive improvements in $BV_{CEO}$ and an expected reduction in $f_T/f_{MAX}$ due to enhanced quasi-saturation behavior. While overly aggressive superjunction designs showed severely degraded performance at low $V_{CB}$, revised designs showed superjunction devices with performance exceeding the control device and even matching the commercially optimized 5PAE device across bias.

The implanted superjunction structure was shown to have the potential to improve device performance across the board. Results for the epitaxially grown superjunction were mixed with unexpected behavior having been measured, likely attributed to a parasitic interaction between the superjunction layers and the extrinsic BC junction. The performance metrics intrinsic to the internal device such as the forward-mode Gummel characteristics and $f_T$ showed little change, which provides hope that the epitaxial superjunction could be properly implemented after careful design using full process simulation.
CHAPTER 4

HOT CARRIER AGING IN SÏGÉ HBTS

This chapter introduces in detail the concepts of hot carrier aging in SiGe HBTs and reviews original research investigating and differentiating the various mechanisms driving hot carrier degradation. Portions of this analysis have resulted in two publications [3, 4]. Provided in this study is an exploration of the different damage mechanisms present during mixed-mode and high-current stress and a comparison of the different characteristics they present. Evidence to support Auger generation as the driving mechanism during high-current stress is provided including both a range of measurements and TCAD simulations.

4.1 Motivation

A consequence of continuous device scaling driven by the desire for greater performance is the shrinking of classical safe operating area (SOA) boundaries, due to the drop in breakdown associated with enhancing high-frequency performance. As circuit designers seek to maximize the circuit performance afforded by a given SiGe technology platform, operating conditions are inevitably pushed nearer and nearer to the DC-defined SOA boundaries. Because of this, an understanding of the physics that determine the time-to-failure (TTF) for a circuit is necessary. The main concern with end-of-life performance of a circuit is how circuit parameters shift due to transistor degradation over time. The primary result of aging in a SiGe is the generation of interface traps at the emitter-base (EB) spacer oxide and the shallow trench isolation (STI) oxide by “hot,” or high-energy, charge carriers, which causes the base current to increase over time (i.e., degrade). In reverse-EB stress, the high electric field in the EB junction will accelerate carriers from the base to emitter, but the degradation will
be dominated by electrons tunneling from the base valence band to the emitter [75].

Under forward-active mode biasing, where most devices will typically operate, hot carrier degradation is generally caused by the mixed-mode degradation mechanism [59, 60], which can be traced to the generation of hot carriers by large electric fields in the CB junction under simultaneous high $V_{CB}$ and moderate $J_C$ biasing conditions. In particular, the main source of hot carriers in mixed-mode stress operation is the acceleration of secondary carriers generated via impact ionization (e.g., holes generated in the CB junction of an npn device) [76]. Interestingly, similar hot carrier driven interface trap generation has also been seen under simultaneous low-voltage and high-current stress conditions, where electric fields are not large enough to produce hot carriers. The most commonly cited explanation for hot carrier generation is a non-radiative Auger generation process that becomes a dominant mode of recombination at high-current density [48–50].

Other modes of degradation have also been reported on, particularly additional structural degradations associated with high-current density stress. The most commonly cited change has been a shift in the ideal base current at medium-to-high values of $V_{BE}$, with both increases and decreases in the current having been reported. This has been attributed to degradation of emitter poly/mono-crystalline Si interface [49, 54, 77, 78]. Variance in behavior can likely be attributed to the quality of the interface and the presence of trace oxide along the poly-Si boundary, with degradation leading to either an increase in tunneling centers at the interface [50], an increase in the recombination velocity at the interface [49, 51], or a change in the local mobility due to the creation of trap states along polysilicon grain boundaries in the emitter near the EB interface [78].
4.2 Hot-Carrier Physics

In order to understand the differences between mixed-mode and high-current stress degradation mechanisms, an understanding of hot carrier driven trap creation is necessary. Hot-carrier degradation relies on two major processes: 1) the generation and transport of hot carriers, and 2) the formation and annealing of traps at oxide interfaces [79]. These processes occur under traditional mixed-mode stress conditions when minority carriers from the base enter the CB depletion region and are accelerated by the high electric field. These carriers can gain sufficient energy to undergo impact ionization, creating additional energetic carrier pairs. These newly generated hot carriers can then proceed to participate in additional impact-ionization events, creating an avalanche effect. Throughout this process, energetic carriers can impact and interact with oxide interfaces, which are layered with a variety of dangling bond states. During fabrication, hydrogen is introduced to the system to passivate these dangling bonds, as they serve as harmful mid-gap trap states if left electrically active. If hot carriers reach the oxide interfaces with sufficient energy, they can de-passivate dangling bonds at the interface to reactivate these trap states [79]. Traps will generally be created across many regions of oxide in the device, but the particular regions of interest are at the EB spacer oxide and STI oxide interfaces. When trap states are created along these oxide interfaces, any generated traps present in the EB or CB space charge region will lead to an increase in the non-ideal base current in forward-mode and inverse-mode (electrical E and C swapped/reverse-active mode), respectively, due to increased Shockley-Read-Hall (SRH) recombination. As the effects on the forward-mode operation are more relevant to typical circuit operation, the traps generated at the EB spacer oxide are of greater concern for normal device use, but measuring the traps along the STI interface serves as a useful tool for reliability characterization.

The generation and transport of hot carriers has commonly been modeled using the “lucky-electron model” (LEM), which applies individual probabilities to each
step in the hot carrier degradation process. The lucky-electron model was originally formulated as a method for predicting channel hot-electron injection for gate leakage current in MOSFETs [80]. The model provides probabilities for several different components of hot carrier injection, namely 1) the probability that a carrier will gain sufficient energy to overcome the energetic barrier presented by the gate, 2) the probability that the hot carrier will be redirected from the channel towards the gate, and 3) the probability that the hot carrier will traverse to the location of the energy barrier peak with sufficient energy to overcome it. Variations on this model have been formulated over the years to deal the complexities of sub-micron devices such as non-constant electric fields and a need to consider carrier energy in lieu of the electric field [81]. Given its success in predicting degradation in MOSFETs, it has since been adapted for modeling mixed-mode stress and the acceleration of carriers via the CB electric field in a SiGe HBT [82, 83]. A note to consider with the LEM, however, is that its physical accuracy is limited in describing carrier energy distributions and the propagation of hot carriers over a distance. The LEM, however, still remains an invaluable tool for providing a high level understanding of the processes behind hot carrier degradation.

A discussion of the various elements of the LEM adapted to a SiGe HBT environment following [83] is provided here. A schematic view of the hot carrier degradation process is provided in Fig. 4.1. The rate at which interface states are created is directly related to the rate at which hot carriers impact the oxide interface. Following the LEM, this rate is given by

$$r_{e/h} = \frac{|J_{n,p}|}{q} \cdot P_1 \cdot P_2$$  \hspace{1cm} (4.1)

where $|J_{n,p}|$ is the electron or hole current density, $P_1$ is the probability that a carrier will gain the threshold energy $\phi_{hot}$ to de-passivate a dangling bond and will be redirected toward the oxide interface, and $P_2$ is the probability that hot carriers will
Figure 4.1: Cross-section of a 2-D device model annotated with the basic process of the mixed-mode degradation mechanism (after [83]).

traverse the distance to the oxide with sufficient energy remaining to create a trap state. The hot carrier redirection probability $P_{1,e/h}$ is defined in the LEM as

$$P_{1,e/h} = \int_{\phi_{hot}}^{\infty} P_{\text{hot},e/h}(\epsilon) P_{\text{red}}(\epsilon)$$  \hspace{1cm} (4.2)$$

where $P_{\text{hot},e/h}$ is the probability that a carrier gains energy $\epsilon$, given by

$$P_{\text{hot},e/h}(\epsilon) = \frac{1}{\lambda F_{\text{eff}}(x,y)} e^{-\epsilon/\lambda F_{\text{eff}}(x,y)} d\epsilon$$  \hspace{1cm} (4.3)$$

and $P_{\text{red}}(\epsilon)$ is the probability that that carrier will undergo a redirecting collision while maintaining sufficient energy, with

$$P_{\text{red}}(\epsilon) = \frac{1}{2\lambda_r} \left( 1 - \sqrt{\frac{\phi_{hot}}{\epsilon}} \right)$$  \hspace{1cm} (4.4)$$

In an $npn$ device under mixed-mode conditions, this redirection probability for electrons being redirected towards the EB interface remains low given the natural
transport path from emitter to collector; however, holes originating as secondary carriers generated by impact ionization events have a relatively high probability of being directed towards the EB oxide while also becoming hot [76].

Once at an oxide interface, trap formation by hot carriers is caused by de-passivation of silicon dangling bonds along the interface. The reaction-diffusion (R-D) framework has long been used to model trap state creation in MOSFETs that lead to bias temperature instability (BTI) [62], but has also been adapted successfully to explain hot carrier effects in bipolar devices [83]. The evolution of the interface state density $N_{it}$ can be computed from the R-D formalism considering both forward (damaging) and reverse (annealing) reactions, given by

$$\frac{\partial N_{it}}{\partial t} = K_F(N_0 - N_{it}) - K_R N_{it} H_2$$  \hspace{1cm} (4.5)$$

where $N_0$ is the total dangling bond density, $N_H$ is the density of hydrogen at the interface, $K_F$ is the forward reaction rate, and $K_R$ is the reverse reaction rate. Under normal conditions, the time dependence of the actual passivation and de-passivation processes are taken to be near instantaneous, which leads to primary time dependence of the reaction being driven by the rate at which freed hydrogen released from the dangling bonds is able to diffuse away. Under the assumption that the process is far from saturation, an approximate solution of (4.5) can be given as

$$N_{it} = \sqrt{\frac{K_F N_0}{2K_R}} \cdot (D_H \cdot t)^{\alpha}$$ \hspace{1cm} (4.6)$$

where $D_H$ is the hydrogen diffusion rate in the oxide and $\alpha$ sets the power-law time dependence [62]. For a diffusion limited reaction, the value of $\alpha$ will vary depending on the species of diffusing hydrogen: 1/6 for $H_2$, 1/4 for $H^0$, and 1/2 for $H^+$, and in a reaction limited process, meaning hydrogen diffusion is fast relative to the rate of trap creation or nonexistent, the time exponent will be 1 [84]. This situation
is most prominent during early stress times. In SiGe HBTs, stress measurements have typically shown an evolution with a time exponent in the range 0.2-0.3 [82,83], indicating a process limited by the diffusion of atomic hydrogen.

When examining degradation under high-current stress, we again see an increase in the non-ideal base current and a similar time dependence. This suggests that high-current degradation physics has a framework similar to mixed-mode degradation physics. Under low-voltage, high-current stress, the large electric field needed to produce hot carriers is missing. However if another mechanism is capable of producing hot carriers, then it is reasonable to assume that the proceeding processes of re-direction, travel to an oxide interface, and de-passivation of a dangling bond are the same. A prevalent process and likely source of hot carriers under high-current operation is Auger recombination, wherein an electron-hole pair recombines and a third free carrier is subsequently accelerated to higher energy, shown schematically in Fig. 4.2. In a SiGe HBT, Auger generation should not typically be able to provide...
the energy required to de-passivate a silicon dangling bond at the oxide interface as
the band gap of silicon is less than the 2.3 eV needed to initiate de-passivation, but
with a sufficient carrier density, repeated recombination processes or many-carrier
interactions should be able to populate the energy distribution function (EDF) at
higher energy with a sufficiently large thermal tail, which shows a Maxwellian shape
[85].

Similar to the avalanching impact-ionization process that plays a large role in
mixed-mode degradation, a possible avalanching of the Auger process, which is es-
esentially the inverse process to impact ionization, would provide sufficient energy for
hot carrier degradation. Studies of hot carrier injection in MOSFETs have shown
that Auger recombination and the similar electron-electron scattering process can
contribute to hot carrier degradation under certain conditions [86,87]. Though it has
reduced probability of creating hot carriers with sufficient energy, the Auger genera-
tion degradation process is enhanced relative to mixed-mode degradation due to the
physical location where these processes occur within the device. The highest rates
of Auger generation occur in the EB depletion region or within the neutral base,
as opposed to the CB depletion region for high fields in mixed-mode stress. Given
the increased proximity to the EB spacer oxide of Auger generated hot carriers, the
probability of energy-robbing collisions is reduced exponentially because hot carriers
have to travel a shorter distance.

4.3 Device Stress Measurements

The devices used in this investigation were first-generation SiGe HBTs from a 0.3 µm
commercial complementary SiGe HBT technology with a 50 GHz peak $f_T$ at $J_C$
of about 1 – 2 mA/µm², and a minimum $BV_{CEO}$ of 3.3 V. Measurements were
performed on-wafer using an automated DC probing system equipped with a hot-
chuck for over-temperature measurements. Transistors were stressed using an Agilent
Figure 4.3: Measurement setup for hot carrier stress characterization with a fixed emitter current density $J_E$ and collector-base voltage $V_{CB}$.

4155C Semiconductor Parameter Analyzer by pulling a set current density through the emitter and applying a set reverse-bias voltage across the CB junction with the base grounded for a pre-determined time interval. The basic stress setup is shown in Fig. 4.3. Device degradation was monitored by periodically interrupting the stress and measuring forward-mode and inverse-mode Gummel characteristics. Each stress condition was measured on a fresh unstressed device.

### 4.3.1 Mixed-Mode Stress

In this study, mixed-mode stress data is briefly examined to set a baseline for comparison with the high-current stress results. A stress condition representative of peak avalanche generation in mixed-mode stress is shown in Fig. 4.4, which shows the time evolution of a device’s forward and inverse Gummel characteristics over 10,000 s, for a stress bias of $J_E = 400 \, \mu A/\mu m^2$ and $V_{CB} = 7 \, V$. As expected for a high-voltage stress condition, the non-ideal base current in both the forward-mode (Fig. 4.4(a)) and inverse-mode (Fig. 4.4(b)) Gummel characteristics increases, which indicates the formation of interface traps at the EB spacer oxide and at the STI oxide edge, respectively.

The degradation of these devices was measured across a large set of bias points in the mixed-mode stress region, and as seen the stress map shown in Fig. 4.5(a), mixed-mode stress follows the same bias dependency as avalanche generation, shown in Fig. 4.5(b). That is, base current degradation increases with voltage due to increased
Figure 4.4: Mixed-mode stress evolution over 10,000 s for the stress condition $J_E = 400 \ \mu A/\mu m^2$ and $V_{CB} = 7 \ V$. (a) Forward-mode and (b) inverse-mode Gummel characteristics show the effects of interface traps at the EB spacer oxide and STI oxide, respectively.
Figure 4.5: Mixed-mode stress evolution over 10,000 s for the stress condition $J_E = 400 \ \mu A/\mu m^2$ and $V_{CB} = 7 \ \text{V}$. (a) Forward-mode and (b) inverse-mode Gummel characteristics show the effects of interface traps at the EB spacer oxide and STI oxide, respectively.
impact ionization rates and hot carrier probability, and the degradation increases with current due to the increase in available carriers to participate in damage physics. At high currents, however, the degradation decreases rapidly due to the Kirk effect, which collapses the CB electric field and pushes the hot carrier generation location away from the EB spacer oxide. Trap creation along the STI oxide, however, will continue to increase with current density despite the Kirk effect because the hot carrier generation location moves along, not away from, the interface. A more detailed explanation on the bias dependencies of mixed-mode stress in the framework of the LEM can be found in [83].

4.3.2 High-Current Stress

Using the same setup as utilized during mixed-mode stress, devices were stressed at current densities larger than $J_C$ at peak $f_T$. In order isolate the effects of high-current damage from and mixed-mode effects, stress voltages were kept small to minimize the contribution of field-generated hot carriers to trap generation. The evolution of the Gummel characteristics over 10,000 s, for a representative high-current stress bias of $J_E = 20 \text{ mA/µm}^2$ and $V_{CB} = 1 \text{ V}$, is shown in Fig. 4.6. An increase in the non-ideal base current is apparent in both forward-mode (Fig. 4.6(a)) and inverse-mode (Fig. 4.6(b)) operation, which indicates interface trap creation at the EB spacer and STI oxides.

Upon closer inspection of the Gummel characteristics, additional current shifts are visible. As shown in Fig. 4.7, there is a small initial decrease in the ideal base current as well as an eventual increase in both the collector and base current for high values of $V_{BE}$. This has previously been attributed to an annealing process of interface traps at the poly-silicon/crystalline-silicon emitter interface that reduces recombination at the interface and leads to a decrease in the emitter resistance $R_E$ [49,50]. Carroll et al. [49] and Rieh, et al. [50] reported associated current shifts as
Figure 4.6: High-current stress evolution over 10,000 s for the stress condition $J_E = 20 \text{mA}/\mu\text{m}^2$ and $V_{CB} = 1 \text{V}$. (a) Forward-mode and (b) inverse-mode Gummel characteristics show the effects of interface traps at the EB spacer oxide and STI oxide, respectively.
Figure 4.7: Time evolution of base and collector current for the stress condition $J_E = 20 \text{ mA/µm}^2$ and $V_{CB} = 1 \text{ V}$. Non-ideal base current shifts are evident at low $V_{BE}$, and shifts in the ideal base current and $R_E$ are evident at higher $V_{BE}$.

high as 30% in a silicon BiCMOS platform and 15 – 20% in a higher performance SiGe HBT platform, respectively. A more recent study looking at similar effects in a highly scaled SiGe BiCMOS platform has attributed such performance shifts at high injection to the creation and annealing of trap states along polysilicon grain boundaries within the emitter, which degrades mobility and thus back-injection from the base and resistivity [78]. Sacrificial or unintentional oxides at the poly-Si/c-Si are uncommon in highly scaled SiGe BiCMOS platforms, which also tend to have larger grain size (more crystalline) polysilicon emitters. The exact nature of these effects has been seen to vary significantly from technology-to-technology, from which we can infer these degradation processes are highly sensitive to the quality of the interface and the availability of atomic hydrogen, and the end result will be some combination of interfacial oxide and polysilicon degradation. Additionally, an effect involving the Germanium profile that leads to increased damage, explored in more detail later in this study, may also contribute to the difference compared to the silicon BiCMOS
Figure 4.8: Base current degradation at $V_{BE} = 0.5 \, V$ from forward and inverse mode Gummel characteristics after 10,000 s of stress at $V_{CB} = 1 \, V$ and range of emitter current densities.

platform presented in [49]. As these devices show minimal annealing behavior, the changes in the non-ideal base current due to trap creation will overtake the effects of the annealing over time.

Looking at the effects of high-current stress over a wider range of stress conditions, a few different behaviors become apparent. When examining the effect over a wider range of current densities, an interesting threshold appears around a stress current density of 10 mA/µm², as shown in Fig. 4.8. For a stress below 10 mA/µm², the base current indicates both damage and annealing behaviors over time, represented by a base current that oscillates up down but remains near the unstressed value. Referring to the R-D formalism, we can infer that there is a range of current values where the forward and reverse reaction rates, $K_F$ and $K_R$, are balanced. The reverse reaction process is driven by the availability of hydrogen species in proximity of the oxide interface and is greatly affected by temperature due to change in hydrogen diffusivity.
Figure 4.9: Base current degradation at $V_{BE} = 0.5 \, V$ from forward and inverse mode Gummel characteristics after 10,000 s of stress at $J_E = 20 \, mA/\mu m^2$ and range of collector-base voltages.

[62]. At high current densities, self-heating within the device can become significant, leading to an increased reverse reaction; however, it is clear that above a certain current density, the forward reaction process begins to dominate. The cause of this behavior is discussed in the following section. Fig. 4.8 also shows that degradation of the inverse mode is significantly less than forward mode degradation, which is due to the STI oxide interface being located farther away from the region of peak Auger regeneration in the EB depletion region and the neutral base.

As the stress voltage is increased while maintaining a very high current density, one would expect the forward mode damage to increase as the mixed-mode mechanism begins to contribute; however, minimal change in the damage at the EB oxide occurs, as seen in the forward mode data in Fig. 4.9. Considering the Kirk effect is strongly engaged, causing the maximum electric field to decrease and push deep into the device away from the EB oxide interface, the observed negligible damage increase is reasonable. Additionally, Auger recombination is largely dependent on the
carrier densities and relatively independent of the electric field. If we instead focus on the inverse-mode operation, damage does in fact increase consistently with voltage. This behavior requires further investigation through the use of TCAD because this increased damage occurs for stress voltages as low as $4 \text{ V}$, for mixed-mode damage at the EB oxide is typically small in this technology.

With an increase in operating temperature, the effects of mixed-mode degradation will decrease because an increase in the phonon scattering probability will reduce the mean free path length $\lambda$ and the probability of impact ionization. Because the scattering and re-direction probabilities for Auger hot carriers are similar to those for mixed-mode hot carriers, a similar temperature dependence would be expected; however, increased Auger hot carrier degradation is observed with increasing temperature. This can be seen in Fig. 4.10, where the degradation for a single stress condition of $J_E = 20 \text{ mA}/\mu\text{m}^2$ and $V_{CB} = 2 \text{ V}$, measured at both room temperature (25 $^\circ\text{C}$) and...
100 °C, is shown. The higher temperature stress condition clearly develops damage at a faster rate. The comparison here is made by looking at the relative change in the Gummel characteristics from each temperature sampled at the $V_{BE}$ corresponding to a pre-stress base current of $I_B = 1$ pA, depicted in the inset of Fig. 4.10, which allows for isolating the degradation from the effects of the difference in temperatures. This current value was chosen to sample an operating point at which interface traps contribute significantly to the total base current. There are two factors that account for this temperature dependence with Auger hot carriers: the Auger recombination rate and the hot carrier EDF. Previous studies show that Auger recombination rates will increase by about $20 - 50\%$ depending on the carrier density over this temperature range [88,89]. An increase in recombination rates means an increase in the number of hot carriers generated. Additionally, due to the thermalized energy tail of the Auger recombination EDF, increased temperatures will increase the probability that charge carriers excited by the Auger process have sufficient energy to create trap states.

4.4 **High-Current Simulations**

To more thoroughly investigate the physical origins of the peculiarities seen during high current stress measurements, simulations of high current device operation were performed using a well-calibrated 2-D device profile in the Synopsys Sentaurus TCAD suite. The particular effects under investigation are 1) the threshold behavior around 10 mA/$\mu$m$^2$ seen in the results in Fig. 4.8 and 2) the effect of stress voltage during high-current operation on inverse-mode degradation seen in the results in Fig. 4.9.

By comparing Auger recombination rates within the device across current density, the underlying reason for the threshold behavior becomes evident. From 1 – 10 mA/$\mu$m$^2$, Auger recombination within the neutral base increases several orders of magnitude, and with hot carriers being created closer to the EB spacer oxide, the trap formation rate will increase. Shown in Fig. 4.11, the peak Auger recombination
rate within the neutral base increases sharply until $J_E \approx 6\ \text{mA}/\mu\text{m}^2$. The cause of this increase is the exposure of the triangular Germanium profile utilized in the base. At the onset of the Kirk effect, the neutral pushes out to reveal the retrograded Ge profile which is concealed by the field in the CB depletion region under standard operation. Shown in Fig. 4.12, the Ge retrograde then becomes visible in the energy band diagram of the device creating a pseudo-potential well within the neutral base region.

At first glance the cause of the measured behavior of the trap formation at the STI oxide interface as a function of voltage was unclear. Simulations at high current reveal what happens to the electric field within the device as the current is increased under base push-out conditions. Fig. 4.13 shows the electric field within the collector for increasing current densities. Strong base push-out is evident, as expected, and the electric field is pushed deeper into the collector region. At sufficiently high current density, the field will be pushed all the way to the sub-collector region. Once the field is pushed to the sub-collector, it will begin to compress and the applied $V_{CB}$ potential...
Figure 4.12: TCAD simulated conduction band energy within the neutral base region for a range of current densities near the onset of base push-out. The shape of the germanium profile is overlaid to help illustrate the formation of the pseudo-potential well.

Figure 4.13: TCAD simulated electric field in the collector region for $V_{CB} = 4 \text{ V}$ for a range of emitter current densities demonstrating the compression of the electric field at the sub-collector under heavy base push-out conditions. In addition, the field at $J_E = 20 \text{ mA/\mu m}^2$ and $V_{CB} = 1 \text{ V}$ is shown for comparison. The sub-collector location is shown by magenta dashed line.
drop will occur over a narrow region, leading to enhanced hot carrier generation and impact ionization at the STI oxide, consistent with the measured data. The enhanced hot carrier generation is evident in the inverse mode data in Fig. 4.9, which shows a sharp increase in inverse-mode damage, relative to the forward-mode damage, as a function of $V_{CB}$ for $J_E > 10 \text{ mA/µm}^2$.

### 4.5 Summary

The various modes of hot carrier degradation in SiGe HBTs have been investigated in this study, and the nature of the new Auger hot carrier generation mechanism prevalent for high current density stresses has been examined in detail. Measured stress data was used to make comparisons between the field-dependent mixed-mode mechanism and the current-dependent Auger mechanism to clarify the differences in aging behavior for different areas of the output plane for a SiGe HBT. TCAD simulations were used to further elucidate unique features in the high current degradation behavior in the featured technology. High-current stress damage increases with current density and exhibits a threshold behavior attributed to the formation of a potential well in the neutral base during base push out, which leads to a sharp increase in Auger recombination within the neutral base. Trap formation at the STI oxide interface was also observed to have a stronger than expected voltage dependence as a result of a compression of the electric field at the collector/sub-collector interface under extreme base push-out conditions. Additionally, Auger hot carrier degradation displays a positive temperature dependence due to the physical nature of the Auger recombination mechanism, where mixed-mode stress damage decreases with increasing temperature due to an increase phonon scattering behavior that reduces impact ionization events.
CHAPTER 5

COMPACT AGING MODELING FOR SIVE HBTS

This chapter presents the formulation and calibration methods for a predictive aging model designed for incorporation in a circuit simulation environment. The analysis presented here resulted in a publication [5], which established a framework for creating a compact model compatible aging model to enable reliability simulations that can quickly and easily be carried out at the circuit or system level.

5.1 Motivation

When each component of a complex system must be qualified for reliability over its intended lifetime, the time required for testing can become quite long, particularly when design re-spins are considered. How to accurately and efficiently predict end-of-life electrical performance, however, remains an open question in the world of circuit simulation, and the ability to quickly look at the end-of-life response for a circuit remains as a sort of “holy grail” for circuit designers. Predictive reliability simulation tools are therefore essential in order to reduce the burden of reliability testing. The underlying reliability problem is highly sophisticated, though, and while circuit designers often treat the transistor as a perfectly modeled black box, it is not that simple. Device reliability and aging is a complex response to the dynamic operating conditions of a device over its lifetime.

A multitude of efforts have been made using technology computer aided design (TCAD) to look at the internal carrier densities and electric fields of the device and capture the aging behavior of specific degradation mechanisms, such as BTI [62] and hot carrier injection (HCI) [80,81] in MOS devices, as well as mixed-mode and Auger hot carrier degradation in bipolar devices [78,83,90]. This is truly the only way
to make predictions that account for all the relevant physical processes, but accurate reliability simulation for a single device can take a long period of time, making TCAD simulation of a full circuit prohibitively long for use by circuit designers.

The most practical step towards achieving fast circuit reliability information in a realistic circuit simulation environment is the development of aging models that integrate with a device compact model. While it will fail to capture the full physics of the stress environment that TCAD can offer, the aging model should be physics-based and can be informed by calibrated TCAD stress simulations. Similar approaches have been made in the past with some success, such as for hot carrier breakdown of CMOS devices [91,92]; however, most models fail to present the full picture for circuit aging, either through the use of simple physical models or incomplete representation of aging effects in the device [82,93,94]. Effective empirical aging models such as in [95–98] have been developed, but they represent a non-ideal solution, as they fail to provide an intuitive look at the causes of the aging and may not have the generality to be usable across multiple technologies.

This present work embodies the first steps in the development of a physics-based damage model as a wrapper for SiGe HBT compact models that will capture the characteristic behavior of hot carrier degradation. The framework of the coupled lucky-electron model (LEM) and R-D model is used to formulate analytic degradation equations for mixed-mode stress. As part of this framework, this aging model is incorporated into the Cadence Spectre circuit simulator and simulated the effects of aging on simple circuit structures to demonstrate the efficacy of the approach.

5.2 Model Formulation

Over the years, the most commonly used and successful modeling methodology for hot carrier degradation effects in SiGe HBTs has been the lucky-electron model (LEM) [80] using a reaction-diffusion (R-D) framework for surface state generation [62], as
discussed in the previous chapter. These models can be linked to a SPICE simulator, or any other circuit simulation environment, via pre/post-processing to enable aging predictions for arbitrary circuits and operating conditions [91].

In recent years, R-D models have been questioned for their ability to accurately model surface state creation in highly scaled MOS systems, particularly during recovery and at extremely short stress time periods [84, 99]. The leading explanation for the observed discrepancies with R-D theory has been fast de-trapping using a hole-trapping model [100, 101]. Dispersive transport models have taken over as the preferred modeling framework for BTI in MOS systems, though most differences with R-D theory can be primarily attributed to how initial and boundary conditions are chosen at the Si/SiO$_2$ interface [84]. Dispersive models take an energy-focused approach, with the dispersion of bond-breakage energies contributing the power-law time dependence of the degradation. Such dispersive models have been well implemented in TCAD environments for modeling NBTI, using spherical harmonic expansion (SHE) solutions to the Boltzmann transport equation (BTE) [102] to find accurate carrier energy distributions and modeling the Si-H bond as a truncated harmonic oscillator with coupled rate equations for each oscillator state [103]. This approach has also been taken to model SiGe HBT degradation for operation near SOA boundaries in TCAD [104]. Given the thick oxides in SiGe HBTs that have small fields across them relative to a scaled CMOS gate oxide, such trapping/de-trapping and dispersive transport consideration are generally unnecessary for modeling SiGe HBTs, and the R-D actually does an excellent job of capturing both the damaging and recovery behaviors of SiGe HBTs while maintaining a computationally simple form [98].

The general methodology used for aging implementation in circuit simulation environment was originally presented by Hu in 1992 for hot carrier injection in CMOS devices [91]. The general methodology used involves integrating the instantaneous degradation that a device experiences for a given operating condition or input signal.
The aging simulator will then recreate the circuit netlist with modified device parameters to represent the degradation. Re-simulation using the aged netlist will then allow comparison of a fresh and aged circuit. Most models, however, fail to present the full picture for circuit aging, either through the use of simple physical models or incomplete representation of aging effects in the device [82,92–94].

The primary result of mixed-mode degradation in a SiGe HBT is an increase in the non-ideal base current due to an increase in interface traps at the EB spacer oxide and the shallow-trench isolation (STI) oxide regions. This degradation is easily monitored by measuring the Gummel characteristics of a device periodically over a stress period, as shown in Fig 5.1. Using a version of the LEM adapted for simulating SiGe HBTs in a TCAD environment [83,90], the components of the device operation contributing to the LEM probability can be analyzed, and combining these simulated results with measured stress data, the individual probabilities can be translated to simpler forms based on the device operating point. To find a starting point, a modified version of (4.1) can be used to break down the components of mixed-mode degradation.

Figure 5.1: Gummel characteristic evolution of a SiGe HBT biased under typical mixed-mode stress conditions for 10,000 s.
The first probability to consider, adapted from (4.3), is the probability that a charge carrier will become hot

\[ P_{\text{hot}} = \int_{\Phi_{\text{hot}}}^{\infty} \frac{1}{\lambda E} e^{-\frac{\epsilon}{\lambda E}} d\epsilon = e^{-\frac{\Phi_{\text{hot}}}{\lambda E}} \]

where \( \lambda \) is the carrier mean-free-path length, \( E \) is the electric field, and \( \Phi_{\text{hot}} \) is minimum energy required to break a dangling bond at the oxide/semiconductor interface. By inspection, (5.2) has the same form as the probability for a carrier to gain sufficient energy to initiate impact-ionization events, which leads to the conclusion that the two probabilities can be directly related, using the peak electric field \( E_{\text{max}} \) to account for acceleration effects. As the primary source of mixed-mode degradation is secondary carriers generated by the impact ionization process, the avalanche generated current \( I_{\text{avl}} \) can be used as a measure of the carriers that may become hot, and the total number of hot carriers generated is given by

\[ HC = \frac{|J_{n,p}|}{q} \cdot P_{\text{hot}} \approx \frac{I_{\text{avl}}}{q} \cdot e^{-\frac{\Phi_{\text{hot}} - \Phi_{II}}{\lambda E_{\text{max}}}} \]

where \( \Phi_{II} \) is the threshold energy for initiating impact ionization. By approximating the CB junction as a one-sided \( p^+-n \) junction, an approximation for the peak electric field \( E_{\text{max}} \) can be made to simplify (5.3) and condense it to an external-bias dependent form

\[ HC = \frac{I_{\text{avl}}}{q} \cdot e^{-b \sqrt{\frac{1+J_{n,p}/J_0}{V_{bi}+V_{CB}}}} \]

where \( b \) and \( J_0 \) become model fitting parameters for the aging model and \( V_{bi} \) is the
built-in voltage of the CB junction. The $J_E$ dependence is included as an approximation for the lowering of the peak electric field at high current densities when the Kirk and/or heterojunction barrier effects engage.

The remaining probabilities for the LEM are the probability that a hot carrier is redirected toward an oxide interface, and the probability that a carrier reaches the oxide interface without undergoing an energy-robbing collision. For typical forward-active mode operation, the non-ideal base current is primarily affected by the interface traps along the EB spacer oxide, so the effects of trap creation along the STI oxide, which primarily affects reverse-active mode operation, may be neglected. In an $n pn$ SiGe HBT, hot holes are naturally directed toward the EB spacer from the CB depletion region, and measurements done to determine the redirection probability have found that the fraction of electrons redirected toward the EB spacer can be simplified to a small constant [76]. Under these assumptions, the total redirection probability can be taken as a constant, when considering the device as a whole. The collision probability decreases exponentially according to the distance the carrier travels and $\lambda$, the mean-free-path length [81]. Again considering the averaged condition within the device, the distance from the peak electric field to the EB spacer $d$ will remain constant at low current densities; however, for high current densities, base push-out will push the peak field location deeper in the device [39]. The collision probability can then be approximated as

$$P_{coll} = e^{-\frac{d}{\lambda}(1+\frac{J_E}{J_1})} \quad (5.5)$$

where $J_1$ is an additional model fitting parameter to account for the shifting of peak field location. Combining (5.4) and (5.5), the total hot carrier impingement rate at the EB spacer becomes
Using a TCAD-generated hot carrier model, the hot carrier impingement rates for a 1st generation SiGe HBT technology were calculated across a wide range of stress bias conditions using the field-dependent hot carrier probabilities [83], shown in Fig. 5.2. As expected, the hot carrier rate is small for lower voltages but increases greatly at high voltages, and the hot carrier rate rolls off with increasing current when the Kirk effect engages around 500 $\mu$A/$\mu$m$^2$.

Assuming mixed-mode stress is a forward-reaction-dominated (damaging) process, $HC_{EB}$ can be plugged into (4.6) and used as the forward reaction rate within the R-D model, giving an approximate solution for the time evolution of interface traps $N_{it}$ as
\[ N_{it}(t) \approx \sqrt{\frac{HC_{EB}N_0}{2K_R}} (D_H t)^n \]  

(5.7)

where \( N_0 \) is the total dangling bond density at the interface, \( K_R \) is the reverse reaction rate and assumed to be constant, \( D_H \) is the diffusion coefficient of hydrogen in the oxide, and \( n \) is a constant time dependence [62, 83].

For long stress periods, the approximate solution to the R-D model given by (5.7) is not valid as the interface trap density saturates to \( N_0 \). This phenomena has been explored in [94, 96] and modeled using a time varying aging rate. For the particular technology used in the present work, deviation of the time exponent \( n \) was not seen until base current changed by several orders of magnitude. Because such a level of degradation is unlikely for circuits operated inside or near the SOA limits, saturation modeling has been left as a topic for future work.

Based on SRH recombination theory [105, 106], the non-ideal base recombination current \( I_{B,\text{non-ideal,sat}} \) is directly proportional to the interface trap density. Combining (5.6) and (5.7), an equation for the degradation of the non-ideal base saturation current in the device is then directly obtainable as

\[ \Delta I_{B,\text{non-ideal,sat}}(t) = A \cdot \sqrt{I_{\text{avl}}} \cdot e^{-\frac{b}{\sqrt{\frac{1 + J_P/J_0}{\Delta V_{GB}}} + V_{CB}}} \cdot e^{-\frac{J_P}{I_1}} \cdot t^n \]  

(5.8)

where \( A \) is constant proportionality parameter for fitting. \( I_{\text{avl}} \) can be approximated by the product of the emitter current and the \( M - 1 \), which is easily measurable [13]; however, modern bipolar compact models such as HICUM and MEXTRAM accurately model avalanche current in the device, so \( I_{\text{avl}} \) can be extracted directly from the simulated device operating point. The changes in the base current seen in Fig. 5.1 is readily related to the degradation by

\[ I_B(t) = I_{B,\text{sat}} e^{\frac{qV_{BE}}{kT}} + I_{B,\text{non-ideal,sat}}(t)e^{\frac{qV_{BE}}{mT}} \]  

(5.9)
where $I_{B,sat}$ is the ideal base saturation current, $kT/q$ is the thermal voltage, and $m$ is the ideality factor for the recombination current with a typical value of 2.

5.3 Stress Measurement and Model Calibration

By measuring the aging of a SiGe HBT across a range of stress bias conditions, the aging model formulated in the previous section can be well-calibrated using only measured data. The stress measurements shown here were performed on-wafer for a 2nd generation SiGe HBT technology with $BV_{CEO} = 2.2$ V and $f_T = 100$ GHz using an automated dc probing system. Transistors were stressed using a Keysight 4156C Semiconductor Parameter Analyzer by pulling a set current density through the emitter and applying a set reverse-bias voltage across the CB junction with the base grounded for 10,000 s at room temperature, as described in Fig. 4.3. Device degradation was monitored by periodically interrupting the stress and measuring the device Gummel characteristics, as in Fig. 5.1. Each stress condition was measured on a fresh unstressed device. To accommodate the measurement of many stress conditions in a reasonable time frame, only stress voltages beyond $BV_{CEO}$ were used, as degradation is slow at low voltages and instrument resolution limits the detection of small current changes.

To properly calibrate the aging model parameters, the rate at which the non-ideal base saturation current ($I_{B,\text{non-ideal, sat}}$) increases must be extracted for a range of stress conditions. The first step in this process is extracting $\Delta I_{B,\text{non-ideal, sat}}$ at each time step of the stress. As any change in the base current is due to trap creation that contributes to the non-ideal recombination current, this process can be performed with a simple linear regression of $\log \Delta I_B$ vs $V_{BE}$, but consistently extracting $\Delta I_{B,\text{non-ideal, sat}}$ over a full stress measurement can be difficult. When measured in an semi-open environment, slight changes in the ambient temperature over a stress period can vary the curve’s slope over time, and a change in instrument resolution at different current
ranges introduces additional uncertainty when fitting data that covers a wide range of currents. To help remedy this, a linear least squares regression utilizing dummy variables, as shown in (5.10), was used to simultaneously extract the degradation at each time step and force a fixed temperature for all time steps.

\[
\begin{align*}
\begin{bmatrix}
\ln \Delta I_B(t_1) \\
\ln \Delta I_B(t_2) \\
\vdots \\
\ln \Delta I_B(t_n)
\end{bmatrix}
&= 
\begin{bmatrix}
1 & 0 & \cdots & 0 & V_{BE}(t_1) \\
0 & 1 & \cdots & 0 & V_{BE}(t_2) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & V_{BE}(t_n)
\end{bmatrix}
\begin{bmatrix}
\ln \Delta I_{B, \text{non-ideal, sat}}(t_1) \\
\ln \Delta I_{B, \text{non-ideal, sat}}(t_2) \\
\vdots \\
\ln \Delta I_{B, \text{non-ideal, sat}}(t_n) \\
\frac{q}{mkT}
\end{bmatrix}
\end{align*}
\] (5.10)

The result of this multiple regression process for a single stress measurement is shown in Fig. 5.3. Assuming the dynamics of the base current degradation to take the form, \( \Delta I_{B, \text{non-ideal, sat}} = f(I, V) \cdot t^n \), the log-plot of the degradation, in the form
Figure 5.4: Evolution of \( \Delta I_{n, \text{sat}} \) over time for a mixed-mode stress condition. Also shown is a fitted aging line used to extract the aging rate.

\[
\log \Delta I_{B, \text{non-ideal,sat}} = \log f(I, V) + n \log t_{\text{stress}},
\]

(5.11)

allows for simple extraction of the bias dependent aging rate and time dependence can be found via linear regression. Fig. 5.4 shows that this extraction method provides excellent results that match the expected form. A stress time exponent of \( n = 0.5 \) was found across the measured bias range, which falls within the expected range [95]. Individually extracting each \( \Delta I_{B, \text{non-ideal,sat}} \) instead of performing a simultaneous extraction leads to a larger variance in values and less confident age rate extraction.

A wide range of stress conditions have been measured, and using this extraction process, a map of the associated age rates for each stress condition has been created, shown in Fig. 5.5. Using the extracted age rates and \( I_{\text{av}} \) data taken from simulation at stress bias as inputs, an iterative non-linear least squares solver was used to find the best-fit parameters \( A, b, J_1, \) and \( J_0 \) from (5.8). The model-predicted age rates using those best-fit parameters are compared to the data-extracted values in Fig. 5.5 and show good agreement.
Figure 5.5: Measurement extracted and model simulated age rates for a range of mixed-mode stress bias conditions. The simulated (line) and measured (line) values show good agreement across the full data range.
5.4 Aging Simulations

The aging model has been implemented in the Cadence Spectre circuit simulation environment via the Unified Reliability Interface (URI) framework [107]. This tool allows externally defined reliability models to be implemented as wrappers around the simulation process. The Spectre simulator will pass the needed operating point parameters, such as \( J_E \), \( V_{CB} \), and \( I_{avl} \), to the URI models, which will iteratively age each device within the circuit. The process flow, shown in Fig. 5.6, consists of:

1. Simulation of a circuit in pre-stress condition.
2. Transient simulation over several circuit operation periods to iteratively accumulate the net aging of a device over a circuit period.
3. Extrapolation of aging from a short simulation time to a full lifetime.
5. Re-simulation of the circuit in post-stress condition with the aged parameter set.

![Simulation flow diagram for the aging process within Spectre.](image)

Figure 5.6: Simulation flow diagram for the aging process within Spectre.
Determining the aging rate during the transient stress simulation is performed by summing the stress contribution at each discrete time step according to the formulated model. In performing this stress accumulation, the previously derived degradation equation (5.8) takes the form

\[
Age = \sum_{0}^{t_{sim}} \left( A \cdot \sqrt{I_{avl} \cdot e^{-b \cdot \frac{V_{bi} + V_{CB}}{V_{bi} + V_{CBO} + J_{E}/J_{1}}} e^{-\frac{J_{E}}{J_{1}}}} \right)^{\frac{1}{n}} \cdot \Delta t \tag{5.12}
\]

where \(Age\) is the amount of degradation accumulated over the simulation cycle \(t_{sim}\).

After the transient simulation completes, the \(Age\) parameter is linearly extrapolated to a full circuit lifetime \(t_{life}\) and used to scale to the necessary compact model parameter. Using the HICUM model peripheral B-E recombination saturation current parameter \(I_{reps}\), which is equivalent to \(I_{B,non-ideal,sat}\) in (5.9), the current is scaled with the equation

\[
I_{reps} = I_{reps,0} + \left( Age \cdot \frac{t_{life}}{t_{sim}} \right)^{n} \tag{5.13}
\]

where \(I_{reps,0}\) is the default, un-aged value of the saturation current. As this model is implemented external to the circuit simulation and is not specific to any particular compact model, this aging model can easily be added to any bipolar compact model that has accurately modeled avalanche current.

Across the range of stress conditions measured, good model-to-data fit is achieved using this method. A simulation example is shown in Fig. 5.7, which shows the simulated and measured Gummel characteristic before and after 10,000 s of mixed-mode stress. Previously, only stress voltages above \(BV_{CEO}\) were measured due to the time required to observe significant degradation at low voltages, but this aging simulation allows for observation of the aging effects at much lower voltages. The contour shown in Fig. 5.8 shows very small degradation at low voltage that rapidly increases at higher stress voltages. The degradation roll-off at high currents associated with the
Figure 5.7: Simulated Gummel characteristics showing degradation of a SiGe HBT under mixed-mode stress conditions.

Figure 5.8: Simulated degradation map for a SiGe HBT under mixed-mode stress conditions showing the normalized change in $I_{reps}$, the HICUM peripheral B-E recombination saturation current.
Kirk effect is also clearly visible. Even though they simulate different technologies, the general shape of this degradation map and the hot carrier rate map in Fig. 5.2 match, which shows that this model should be applicable and easily adaptable to other SiGe HBT technologies in general.

5.5 Circuit Aging

With good calibration of model to data for single device stresses, this aging model should be usable to monitor the impact of mixed-mode stress on realistic circuits. To validate the model and explore circuit aging effects, a simple two device current mirror, designed in the same technology used for model calibration, was chosen for measurement and simulation of aging effects. The current mirror was chosen due to its simplicity and ubiquitous use in analog and RF circuit design. Measurements and simulations were set up as in Fig. 5.9, with a stress current being applied to the input device and a fixed voltage being held at the output device. Monitoring of the circuit performance was performed by sweeping the input current with fixed $V_{out} = 1\, V$ and tracking the output current and mirror ratio ($I_{out}/I_{in}$), shown in Fig. 5.10.

As can be seen in Fig. 5.10, both measurement and simulation data show a similar trend; namely, that aging of a current mirror results in a decreasing mirror ratio, or rather a drop in the output current for the same input current level. As expected, degradation of the mirror ratio is greater at lower current levels due to the influence

![Figure 5.9: Schematic of current mirror stress setup in both simulation and measurement. A constant current density $J_{in} = 1\, mA/\mu m^2$ is supplied at the input device, and a constant voltage $V_{out} = 4\, V$ is applied on the output device.](image-url)
of the increased non-ideal base current. If the output device sees a large voltage from the applied load, it will undergo mixed-mode degradation while the input device will degrade minimally due to a small CB electric field. This will create an imbalance in the two device’s current gain $\beta$ and degrade the mirror ratio. A clear mismatch exists between the measured and simulated degradation in Fig. 5.10, though, which is evidence of a secondary degradation mechanism in addition to the mixed-mode mechanism, likely being hot carrier degradation of the polysilicon in the emitter near the EB junction, which could cause degradation in the form of the parallel shift seen in Fig. 5.10 [78].

Calculating the effects of aging on a simple circuit such as a current mirror can be straightforward and intuitive, but when moving to more complex circuits and systems, the effects of components aging at different rates becomes convoluted, making the identification of critical components more difficult. An encouraging takeaway for circuit designers though is that the degree of performance degradation of the current
mirror in Fig. 5.10 is significantly less than the degradation seen by a single device under similar stress conditions, as in Fig. 5.7. This means that the boundaries of transistor reliability may be pushed further than expected for reliable circuit operation. Understanding of device-to-circuit interactions in regards to aging is still incomplete, but this tool provides a framework to build upon and continue exploring the effects of device aging on circuit performance.

5.6 Summary

A physics-based aging model for mixed-mode stress in SiGe HBTs was formulated that is based on the coupled lucky-electron model and reaction-diffusion equation. This model is suitable for implementation in compact models and integration with circuit simulation environments. This model has been calibrated using measured stress data, and an effective method for extracting aging rates from measured data was detailed. This degradation model has been implemented in the Cadence Spectre circuit simulator via the Unified Reliability Interface framework. The URI enables for calculation of the device damage accumulated over a stress period, creation of aged model parameter sets, and extrapolation to end-of-life aged parameter sets. Simulated device stress data was shown to be in good agreement with measured data and matches the trends predicted by reliability physics theory. Additionally, stress measurements and simulations of a simple SiGe HBT current mirror were performed, showing that degradation of the output device over time leads to a decreased current mirror ratio, but additional modeling work is necessary to capture the full stress response.

The success of this model in predicting device-level aging and replicating the trends seen in simple circuit aging is encouraging for its future use for full circuit aging simulations. This model provides a valuable framework to build upon and to explore the convoluted device-to-circuit interactions present in aging. While this model only
simulates base-current degradation due to mixed-mode stress, the methodology described here will allow for the integration of additional degradation mechanisms, such as emitter-interface degradation, high-current degradation, and annealing, and for accurate end-of-life performance predictions for full analog and RF circuits, something long desired by circuit designers.
CHAPTER 6

AGING MODELING OF SIGE HBTS FOR RELIABILITY-AWARE CIRCUIT DESIGN

This chapter builds upon the modeling framework created in chapter 5 to develop a more fully featured aging model and explore the concept of reliability-aware circuit design. This analysis has resulted in one publication [6] and looks to re-evaluate the traditional understanding the SOA.

6.1 Motivation

Returning to the topic of how device scaling and decreasing breakdown voltages are seen from the perspective of reliability, the tightening area within usable output plane of a device necessitates a closer look at the safe operating area (SOA) and standard reliability guidelines. From the perspective of a circuit designer, designing with reliability in mind normally means restricting device operation to a foundry-defined SOA coded into the process design kit (PDK). An SOA will define limits for voltage bias and current flow, under which the foundry will guarantee minimal device degradation. These SOA rules tend to be fairly simple, generally providing a maximum voltage and current for a device and electromigration guidelines for the metal lines, and these rules are calculated based on \( dc \) characterization of single devices. These SOA-imposed limits are then applied as “one-size-fits-all” rules to all devices, regardless of application and configuration. Devices in PAs, LNAs, switches, current mirrors, etc. are all treated by the same rules, though devices in each circuit may see wildly different operating conditions and have disparate reliability concerns.
Figure 6.1: Simulated load line of a device in a driver amplifier showing schematic representation of the device aging regions, SOA, and failure boundaries.

Such an approach is, of course, overly simplistic, and ignores the nuances of device-to-circuit interactions, and potentially sacrifices circuit and system performance [?].

Taking the example of an output stage device from a high-swing cascode amplifier, for which a simulated device load line is shown in Fig. 6.1, the device will swing throughout the entire output plane, passing through areas within the standard SOA, mixed-mode stress, high-current stress, and annealing regions. The reliability regions painted on Fig. 6.1 are purely schematic in nature but illustrate the need to consider device reliability as more than a problem that can be treated with simplistic rules. The hard boundaries for reliability are the points where catastrophic device failure due to electrothermal runaway or extreme electromigration occur, and the foundry-defined SOA boundaries will be some distances away from these hard boundaries, leaving a tempting gray area for circuit designers. The objective in advancing reliability-aware design is to build an understanding of where in a circuit and for how long these SOA rules can be safely violated, which circuit designers need more and more as breakdown
voltages shrink. Additionally, individual device degradation may not necessarily lead to changes in circuit performance, so building predictive aging simulation capabilities to identify reliability critical and non-critical devices and ultimately maximize performance will help usher in a paradigm shift in reliability-aware design in the SiGe world.

Mixed-mode stress causing the degradation of the EB spacer oxide in a SiGe HBT, as modeled in chapter 4, is the primary long-term aging concern, but in moving to a model that can provide true predictive capability for arbitrary circuits, more comprehensive modeling capabilities must be developed. Additional modes of hot carrier degradation in SiGe HBTs due to high-current (HC) stress and polysilicon degradation must be considered as well. These effects have all been modeled to various extents in TCAD environments [78, 104, 108], but TCAD solutions remain a non-ideal solution for the simulation of circuits and systems.

6.2 Testing Setup

The devices under test were high-performance (HP) SiGe HBTs from GLOBAL-FOUNDRIES 90 nm 9HP SiGe BiCMOS technology with unity current and power gain frequencies \( f_T/f_{\text{MAX}} \) of 300/360 GHz, \( J_C \) at peak \( f_T \) of about 20 mA/\( \mu \)m\(^2\), and open-base breakdown voltage \( BV_{CEO} = 1.7 \) V [19]. Initial model calibrations were made using devices with a single drawn emitter size of 0.1 \( \times \) 4.0 \( \mu \)m\(^2\), but additional measurements were made using devices with drawn emitter lengths \( L_E = 2, 10, 20, \) and 30 \( \mu \)m for additional geometry calibrations. All tests were performed on-wafer at room temperature \( T = 27 \) °C, and stress bias was applied as a constant \( V_{CB} \) and constant emitter current density \( (J_E) \) for up to a stress time \( t_{\text{stress}} = 100,000 \) s. Stress was periodically interrupted to measure forward Gummel characteristics of the devices and thereby monitor degradation. In order to isolate the contributions to degradation of each stress mechanism, measurements are focused at
\( V_{CE} > BV_{CEO} \) and \( J_C < J_C(\text{peak } f_T) \) for mixed-mode stress and at \( V_{CE} < BV_{CEO} \) and \( J_C > J_C(\text{peak } f_T) \) for high current stress.

### 6.3 Model Formulation

From measured data, the methodology described in chapter 4 is used to extract the bias-dependent aging rate. The base current degradation is characterized as a power law time dependence with a bias dependent aging rate.

\[
\Delta I_B(t) = f(V_{CB,Stress}, J_{E,Stress}) \cdot t^n. \tag{6.1}
\]

Using the time evolution of the non-ideal base saturation current, the aging rate \( f(V_{CB}, J_E) \) and exponent \( n \) are easily isolated. In measuring over-stress conditions above \( BV_{CEO} \) and at modest current densities, peak mixed-mode stress points can be identified. The previously created physics-based model for mixed-mode stress represents the hot carrier generation rate as

\[
K_{MM} = A_{MM} \cdot J_{avl} \cdot e^{-b \sqrt{\frac{1 + J_E/J_1}{V_{CB} + V_{bi}}} \cdot e^{-\frac{J_E}{J_2}}}, \tag{6.2}
\]

where \( A_{MM}, b, J_1, \) and \( J_2 \) are constants, \( J_{avl} \) is the avalanche current density, and \( V_{bi} \) is built-in CB junction potential. In Fig. 6.2, a surface visualizing the trends of the aging rate changing with stress bias for this technology is given along with cut lines along \( V_{CB} \) and \( J_E \) showing model-to-data fit. Using this methodology has again shown a good model to data fit. This continued quality of fit is encouraging for helping to validate the general methodology being used, as the 90 nm device with \( f_T > 300 \) GHz being evaluated here and the 130 nm device with \( f_T \approx 100 \) GHz are very different devices when the details of their construction is examined closely. The ability to achieve model calibration with a small number of parameters can be attributed to the physics-based nature of the model.
Figure 6.2: Calculated aging rate due to mixed-mode stress at room temperature with (a) a surface showing a full stress output plane and (b) cut-lines showing model-to-data fit at $J_E = 1 \, mA/\mu m^2$ and at $V_{CB} = 3 \, V$. 
In characterizing the effects of high-current stress, measurements are performed at low-$V_{CB}$ bias where the mixed-mode mechanism is negligible and a high stress current density beyond peak $f_T$ where the three-carrier-interaction Auger generation process becomes more significant. Relating to the free electron and hole densities, the Auger recombination rate is given as

$$U_{Auger} = \Gamma_n n^3 + \Gamma_p p^3 \quad (6.3)$$

where the $\Gamma_n$ and $\Gamma_p$ prefactors represent the electron-electron-hole and electron-hole-hole processes, respectively. Due to the high carrier densities present, peak Auger-hot-carrier generation locations in a SiGe HBT will be near the EB junction particularly at the poly/crystalline-Si emitter interface and in the neutral base, meaning the emitter current density can serve as a good substitute for free carrier densities, and when moving to current densities above peak $f_T$, where free carrier densities exceed the background doping, (6.3) simplifies as $n \approx p >> n_i$, giving

$$U_{Auger} \propto n^3 = p^3 \propto J_E^3. \quad (6.4)$$

As Auger-hot-carrier driven high-current stress does not become prevalent until current densities beyond peak $f_T$, the simplified expression in (6.4) can be used to model the high-current stress reaction rate as

$$K_{HC} = A_{HC} \cdot J_E^\alpha, \quad (6.5)$$

where $A_{HC}$ and $\alpha$ are constants. Using the same methodology employed to fit and evaluate the mixed-mode stress equation, the results for high-current stress in Fig. 6.3 clearly show the expected power law dependence on the stress current density, closely matching the expected $J_E^3$ dependence with Auger generation.

As the mixed-mode and high-current hot carrier driving mechanisms are operative
Figure 6.3: Hot carrier aging rate measured as a function of stress emitter current density at $V_{CB} = 0$ V. A power law current density is observed as expected from the Auger-hot-carrier mechanism.

Figure 6.4: Combined hot carrier aging rate for the mixed-mode and high-current stress mechanisms at room temperature. Mixed-mode dominates at high voltages and moderate current, while high-current becomes apparent beyond peak-$f_T$ current density.
in separate regions of the device, namely in the vicinity of the CB and EB junctions, respectively, and to first order are independent, the net hot carrier reaction rate can be found by a simple sum of (6.1) and (6.4). The combined effects of these mechanisms at room temperature can be seen in Fig. 6.4, where compared to Fig. 6.2(a), a small increase in aging rate can be seen at high current. It may be noted, however, that the high-current aging mechanism is much weaker than the mixed-mode mechanism and will likely contribute less to aging under most “normal” circuit operating conditions. Such a disparity in the strength of each effect may be reduced or even reversed, however, at elevated temperatures given the opposite behavior of each mechanism with scaling temperature [78].

To further enable a practical universal model for use in circuit design, rules must be established for how these hot carrier aging rates scale with device geometry. Intuitively, a larger device will produce a larger volume of hot carriers for the same $J_E/V_{CB}$ stress, and there exists a larger oxide perimeter to be potentially damaged.
To directly compare aging rates, several representative stress conditions were characterized in Fig. 6.5 for several devices with fixed $W_E = 0.1 \, \mu m$ and $L_E = 1, 2, 4, 10,$ and $20 \, \mu m$, presenting an aging rate that scales linearly with the device perimeter. When normalized to a minimum geometry device, the simple relation

$$\frac{\Delta I_B(P - P_0)}{\Delta I_B(P_0)} = c_{geo}(P - P_0),$$

is arrived at, where the factor $c_{geo} \approx 0.5$ is a constant. A value $c_{geo} < 1$ shows that shorter devices undergo proportionally larger degradation. For example, four $0.1 \times 1.0 \, \mu m^2$ devices in parallel would be expected to undergo more degradation than a single $0.1 \times 4.0 \, \mu m^2$ device. This behavior is expected, because as P/A drops, the contribution of perimeter effects to the total current decreases. Additionally, testing the geometry scaling behavior across several different stress conditions showed the same trends, indicating that geometry scaling behavior does not meaningfully change across the output plane.

### 6.4 RF Device Stressing

To this point, all measurements and modeling have been based on dc measurements of single devices. For broader use in circuit design for mixed-signal and RF applications, however, these modeling efforts must be validated for time-varying voltages and currents. From a practical standpoint, the best approach is to assume a quasi-static approximation, modeling time-varying stresses as a sequence of successive dc stresses, and find the frequency at which that approximation falls apart to apply adaptations for high-frequency effects. From a simulation standpoint, an effective aging rate for a RF stress condition can be found by integrating over a transient waveform of the device operating condition. The degradation in this approach is calculated to take the form
as described in [109, 110]. Using this methodology, this model was integrated into the Cadence Spectre circuit simulator by coding the model equations for use with the Spectre unified reliability interface (URI) post-processing utility [107], allowing for simulation of arbitrary stress conditions and comparison of pre- and post-stress performance. The specific implementation used here is the same as the modeling efforts described in chapter 4. The limitations of this model come about for cases where non-equilibrium conditions become prominent and there exists a dissociation between carrier and energy distributions within the device. For current SiGe HBT technologies, quasi-static approximations are generally successful in capturing charge transport within the device [111], but continued scaling and operation of devices in high-field, velocity-saturated conditions may limit their validity. The goal here is to lay the groundwork for testing the usable range of conditions for these models. RF measurements were performed by applying a constant continuous wave (CW) RF stress to a device and measuring the base current degradation over time. The setup used is pictured in Fig. 6.6. Load-pull and measurements were performed to stress
at a point tuned for maximum power. A dc bias of $V_{BE} = 0.85 \, V$ and $V_{CE} = 2 \, V$ was applied to bias near peak-$f_T$ and provide a good accelerated test condition. RF power was calibrated to deliver $P_{in} = -15 \, \text{dBm}$ at $f_0 = 5 \, \text{GHz}$. Based on simulations performed separately in Keysight Advanced Design System (ADS), this power and dc bias level keep the device in class-A operation. Care to keep operation as class-A was taken to avoid any effects that non-conductive stress may introduce, as the effects of non-conductive stress remain unknown but will be a critical component in providing predictive reliability capabilities when considering large-signal stress.

To validate the model for predicting this RF stress, the measurement setup of Fig. 6.6 was created in Cadence and predictive aging simulations were run. When considering the transient stress condition, the instantaneous stress condition at each time step is evaluated, as shown in Fig. 6.7. Analyzing these waveforms, the aging rate appears to, at least for this specific stress condition, primarily follow $V_C$. A trough in the aging rate does exist though for the minimum $V_B$ bias, likely due to a reduction in the current within the device.
Figure 6.8: Normalized hot carrier aging rate showing a linear dependence of device aging on emitter perimeter.

A comparison of the measured and simulated degradation in the base current is shown in Fig. 6.8. The two show an identical time dependence, developing as \( \sqrt{t} \), and the aging rates are similar and within an expected region of variance consistent with the variation seen device-to-device for identical stresses. Additionally, the de-embedding efforts in the simulated were fairly basic, representing room for discrepancy between the two results.

### 6.5 Full Aging Simulations

With calibrated aging rules covering a full range of forward-mode bias conditions, the present aging model should be viable for use in evaluating the hot-carrier reliability of full circuits. EB oxide degradation has been described here, but additional degradations may occur within polysilicon, modifying the emitter resistance \( R_E \) and ideal base saturation current \( I_{be_{is}} \) [78]. As these are all hot-carrier degradation effects, we can simply adapt (6.7) with individual scaling parameters, \( A \), per mechanism when aging multiple parameters within HICUM. Given the current incomplete understanding of hot-carrier polysilicon degradation effects, the aging of such parameters has
been bounded and implemented only to illustrate the effects their degradation can have on circuit performance. Such a simulation can be useful to allow adaptation of a circuit design to be more robust to such changes should they occur.

An example simulation is provided here of a wireline driver circuit, representing a common application for a highly scaled SiGe HBT technology, where long-term reliability due to hot carrier effects is a concern. A simplified schematic of the circuit is provided in Fig. 6.9. For simulation of a 20 GHz CW input signal of $P_{in} = 10 \, dBm$ and over-stress bias, shown in Fig. 6.10, the circuit gain can be seen to begin dropping over time. Observing the individual degradation of each transistor, $Q_4$ sees the largest shift in bias and could be identified as the primary reliability concern in the circuit. While the base of $Q_4$ degrades the most, at closer look, it’s current drop can be primarily attributed to bias shifts due to the degradation of $Q_0$ and $Q_2$. Due to it seeing the largest $V_{CB}$ bias, device $Q_0$ actually undergoes the largest degree of base current degradation. This point represents the largest disconnect between a basic device-level approach to reliability and a reliability-aware design approach accounting for device-to-circuit interactions. The degradations in performance seen at the circuit

![Figure 6.9: Simplified schematic of driver circuit used for aging simulation.](image-url)
Figure 6.10: Circuit response degradation over time due to stress of a 20 GHz CW input signal with $P_{in} = 10\ dBm$. The individual bias shifts over time for devices $Q_0, Q_2,$ and $Q_4$ are shown in the inset.

level are primarily due to cascading effects of device degradation leading to a shifting operation point of key devices within the circuit. With analysis tools such as this, maximizing the trade-off between performance and reliability becomes a much easier task, and the operating conditions under which the circuit will drift out of spec can be identified. By identifying which devices impacting circuit reliability, techniques such as building compensation loops to dynamically adjust for bias shifts may be used [112], but at the same time the decision may be made to bias devices more aggressively beyond typical SOA boundaries (in the “forbidden” zone between the SOA and device failure) to improve performance if such degradation was deemed acceptable.

### 6.6 Summary

An aging compact model based on a physical description of the hot carrier generation mechanisms in a SiGe HBT during mixed-mode and high-current stress has been developed here for a highly scaled SiGe BiCMOS technology and can predict the
degradation of devices over a wide bias range. Previously developed aging models have been extended to model high current stress effects and the effects of device geometry on aging. Additionally, measurements and simulation data have been presented to validate quasi-static approximations in aging modeling. Given the ever-shrinking voltage headroom for highly scaled SiGe HBTs, the classical understanding of a “one-size-fits-all” SOA must be re-evaluated as SiGe BiCMOS technologies push into mm-wave applications, and the use of aging models present an opportunity to easily evaluate reliability on a case-by-case basis, not only for circuits, but also the individual transistors driving them. Such a models provide for tightening the design loop for reliability and advancing the development of techniques for reliability-aware circuit design.
CHAPTER 7

CONCLUSION

7.1 Summary of Contributions

This work explores electrical reliability phenomena within silicon germanium (SiGe) bipolar technologies. While SiGe HBTs have enjoyed quickly growing use in a number of high-frequency RF and mm-wave applications, the relatively low breakdown voltage of highly scaled SiGe technologies presents a limitation for their usefulness in comparison to many III-V technologies. Therefore, the ability to design more robust devices or provide methodologies to boost available performance is highly desirable. The specific contributions of this include

1. An exploration and design parameterization of a vertical superjunction collector SiGe HBT for enhanced breakdown. These devices attempt to break the classical “Johnson limit” but present a challenging design space. Several superjunction devices based on a 1st generation SiGe HBT for use in power amplifiers were fabricated and when compared with a control device showed instances of both significant performance gains and degradations. An optimized device showing significant gains in breakdown voltage and minimal degradation in RF performance relative to a control device presents a strong case for the potential of this device concept.

2. An exploration of the various modes of degradation in a SiGe HBT, introducing for the first time, and characterizing the Auger hot carrier generation mechanism as a source of device degradation under high-current stress. Both measurements and simulations were used to illustrate the differences between
the Auger hot carrier mechanism and the classically understood mixed-mode mechanisms prevalent under high electric field conditions.

3. The development of the first physics-based predictive aging model for SiGe HBTs that is fully integrable in a compact modeling and circuit simulation environment. The model was developed based on a coupled lucky-electron/reaction-diffusion approach and focused on the classically understood mixed-mode degradation mechanism. Additionally, a methodology for setting up and calibrating a model for a given SiGe HBT technology was developed.

4. An extension of previously developed predictive aging models with an emphasis on providing a comprehensive model usable for practical circuit design. The model was extended to capture the effects of high-current stress and device geometry scaling. Additionally, measurements and simulations of RF stress of a device provide validation for the use of this $dc$-calibrated model for predicting the effects of time-varying $ac$ stresses.

### 7.2 Future Work

The research presented here has represented initial capabilities in several areas and has exposed several interesting topics for future research.

1. The superjunction devices designed in this work were built in a $1^{st}$ generation technology intended for lower frequencies, but the need to fully investigate the superjunction collector in a highly scaled device is readily apparent. Given the voltage limitations of highly scaled technologies, the potential gains in voltage headroom present a significant potential for improving circuit performance.

2. The superjunction devices designed here were evaluated for their $dc$ and small-signal $ac$ performance. The effects of the superjunction collector on large-signal linearity remains a question for general use in a commercial platform. While
conductivity modulation effects will likely degrade device linearity, a well-designed superjunction should expect some gains to large signal performance with the additional breakdown voltage.

3. The modeling of temperature and recovery effects as they concern device aging is an area with lots of room for exploration. In the efforts to create a truly universal aging model, the temperature dependencies of each mechanism must be modeled independently to capture the shifting dominance of each mechanism on the overall stress with environmental conditions.

4. To extend the usefulness of aging models for all circuit applications, effects specific to RF stresses must be investigated. For example, the effects of non-conductive stress are an open question, and efforts to evaluate the large-signal reliability of devices [113] have exposed discrepancies between \( dc \)-defined reliability metrics and actual circuit reliability under RF operation.

5. With predictive aging models in place, space opens for new reliability-aware circuit design methodologies. With the ability to identify the reliability-critical components and devices within a circuit, designs can be altered to improve the sensitivity to reliability effects or to intelligently push beyond standard reliability rules when individual device degradation shows no effect on the overall circuit performance.
REFERENCES


[70] Fujihira, T. and Miyasaka, Y., “Simulated superior performances of semiconductor superjunction devices,” in *Power Semiconductor Devices and ICs,*


[77] Rieh, J. S., Jagannathan, B., Chen, H., Schonenberg, K. T., Angell, D., Chinthakindi, A., Florkey, J., Golan, F., Greenberg, D., Jeng,


[96] Fischer, G. and Sasso, G., “Ageing and thermal recovery of advanced SiGe heterojunction bipolar transistors under long-term mixed-mode and reverse


VITA

Brian R. Wier was born in Atlanta, Georgia in 1991. He received the B.S. degree with Highest Honors in electrical engineering from the Georgia Institute of Technology in 2012. In 2013, Brian joined the silicon-germanium devices and circuits group of Dr. John D. Cressler, under whose advisement he earned the M.S. degree in electrical and computer engineering in 2015 and passed his dissertation defense in completion of the Ph.D. degree in electrical and computer engineering in November 2018.

While at Georgia Tech, Brian has authored and co-authored more than 30 refereed journal publications and conference proceedings. His research interests while at Georgia Tech were primarily focused on the characterization and modeling of long-term aging and reliability effects for both devices and circuits in advanced SiGe BiCMOS technologies. Additional topics of interest have included the design of novel SiGe HBT devices, cryogenic operation of SiGe HBTs, and the modeling of unique device-to-circuit interactions. Following the completion of his doctorate, Brian will be joining the Electro-Optical Systems Laboratory (EOSL) at the Georgia Tech Research Institute (GTRI) in Atlanta, Georgia.