Self-Heating and Isothermal Characterization of Heterojunction Bipolar Transistors

by

Todd Christopher Kleckner

B.Sc. Engineering Physics, University of Alberta, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Applied Science in THE FACULTY OF GRADUATE STUDIES (Department of Electrical and Computer Engineering)

we accept this thesis as conforming to the required standard

The University of British Columbia

August 1998

© Todd Christopher Kleckner, 1998
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical & Computer Engineering

The University of British Columbia
Vancouver, Canada

Date 25/8/98
Abstract

Self-heating, the process by which power dissipation causes a rise in the operating temperature of the device, causes significant changes in the operation of bipolar transistors and can compromise the accuracy of device models if the operating environment of the device is not identical to the one in which it was characterized. As conventional methods utilized to construct device models are based in part upon measurements of device performance in which the static power dissipation results in an elevated device temperature, the extracted model contains information pertaining to the thermal environment in which it was characterized. This hidden temperature dependence is often ignored, and results in erroneous parameter extraction. Extending a technique developed by workers at Hewlett-Packard, a pulsed bias measurement system has been developed which addresses the need for temperature-independent device characterization. This system can be used to measure the isothermal collector characteristics at constant base-emitter voltage, and the isothermal Gummel plot of collector current. The resulting data can be used to extend the region of model validity by de-embedding the thermal effects from the inherent isothermal electrical response of the device. The system can also be used to determine the values of the thermal resistance and capacitance, macroscopic
thermal parameters that can be used with device model thermal subcircuits.

Using this measurement technique, the first experimental studies of thermal parameter scaling in deep-trench Si/SiGe and non-trench-isolated AlGaAs/GaAs heterojunction bipolar transistors has been performed. Scaling coefficients were determined for the dependence of thermal resistance and capacitance on emitter area. Comparison of measured thermal resistances to model predictions show substantial deviations for small devices and for large devices. Results for small devices suggest the interconnect metallization plays a significant role, and those for large devices show that electrothermal feedback is important and must be accounted for. For each of the two technologies studied, measured thermal capacitances vary linearly with the square root of the emitter area and depend only weakly on the emitter geometry. The result is a simple thermal capacitance scaling law which allows estimates for unknown structures in a given technology from a single measurement.
Contents

Abstract ............................ ii
Contents ................................ iv
List of Tables ........................ vii
List of Figures ........................ viii
Acknowledgements ...................... xi
Dedication ................................ xiii

1 Introduction ........................ 1

1.1 Introduction to Thesis ................. 1
  1.1.1 Overview .......................... 1
  1.1.2 Summary of Results ................. 2
  1.1.3 Outline of Chapter .................. 3
1.2 Self-Heating and its Effect on Device Operation ............... 4
1.3 Motivation ........................... 10
1.4 Lumped Thermal Model ................. 12
  1.4.1 Thermal Resistance Fundamentals .... 15
1.4.2 Thermal Capacitance Fundamentals .................................. 20
1.4.3 Temporal Response of Lumped Thermal System ............. 23
1.4.4 Application of Thermal Resistance to Device Self-Heating .... 24
1.5 Existing Experimental and Theoretical Techniques and Results ... 28
  1.5.1 Experimental Optical Methods .......................... 28
  1.5.2 Experimental Electrical Methods .......................... 30
  1.5.3 Theoretical Methods .................................. 34
1.6 Outline of Thesis .......................................... 36

2 Experimental Measurement of Thermal Subcircuit Model Parameters .......................... 38
  2.1 Experimental Method ........................................ 39
      2.1.1 Overview of Approach ................................ 39
      2.1.2 Sample Description and Preparation .................. 41
      2.1.3 Measurement Circuitry ................................ 44
  2.2 Calculation of Thermal Resistance .......................... 47
  2.3 Results and Discussion ................................... 51
      2.3.1 Transient Results and Isothermal Characterization .... 51
      2.3.2 Thermal Resistance .................................. 65
      2.3.3 Experimental Limitations ............................ 72
  2.4 Conclusions .............................................. 79

3 Thermal Parameter Scaling in Si/SiGe Heterojunction Bipolar Transistors .................. 82
  3.1 Description of Devices Tested ............................. 82
  3.2 Theoretical Considerations ................................. 83
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1</td>
<td>Thermal Resistance Model</td>
<td>83</td>
</tr>
<tr>
<td>3.3</td>
<td>Experimental Results and Discussion</td>
<td>85</td>
</tr>
<tr>
<td>3.4</td>
<td>Conclusions</td>
<td>91</td>
</tr>
<tr>
<td>4</td>
<td>Role of Layout in Thermal Performance of AlGaAs/GaAs Heterojunction Bipolar Transistors</td>
<td>95</td>
</tr>
<tr>
<td>4.1</td>
<td>Description of Devices Tested</td>
<td>96</td>
</tr>
<tr>
<td>4.2</td>
<td>Thermal Resistance Calculations</td>
<td>96</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Analytical Predictions of Rectangular Device Thermal Resistance</td>
<td>98</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Thermal Resistance Comparison of Rectangular and Square Emitter Devices</td>
<td>101</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Model Limitations for Large Area Devices</td>
<td>103</td>
</tr>
<tr>
<td>4.3</td>
<td>Experimental Results</td>
<td>109</td>
</tr>
<tr>
<td>4.4</td>
<td>Comparison Between Theory and Measurement</td>
<td>114</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>121</td>
</tr>
<tr>
<td>5</td>
<td>Conclusions</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Appendix A Pulse Circuitry</td>
<td>131</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Measured thermal time constant and thermal resistance, and theoretical thermal resistance values for deep trench Si/SiGe heterojunction bipolar transistors ................................................. 86

4.1 Effective area normalized to total emitter area for rectangular and square emitter devices of varying sizes .................................................. 105

4.2 Measured thermal parameters for one-finger AlGaAs/GaAs heterojunction bipolar transistors ................................................................. 109

4.3 Measured thermal parameters for two-finger AlGaAs/GaAs heterojunction bipolar transistors ................................................................. 110

4.4 Measured thermal parameters for square AlGaAs/GaAs heterojunction bipolar transistors ................................................................. 110

4.5 Measured thermal parameters for round AlGaAs/GaAs heterojunction bipolar transistors ................................................................. 112

4.6 Regression coefficients for AlGaAs/GaAs structures .................. 118
List of Figures

1.1 Physical origin of self-heating ........................................... 5
1.2 Temperature dependence of collector current factors ............. 9
1.3 Thermal subcircuit representation ..................................... 14
1.4 One dimensional insulated bar .......................................... 16
1.5 Parallelepiped buried in a semi-infinite medium .................... 18
1.6 Temporal response of lumped thermal system ....................... 25

2.1 Schematic of pulsed measurement ....................................... 39
2.2 Typical sample fixturing ................................................ 43
2.3 Generation of pulsed base bias ......................................... 45
2.4 Collector bias network .................................................. 46
2.5 Analytical temperature contours of a quadrant of an AlGaAs/GaAs device .................................................. 49
2.6 Experimental pulsed base-emitter voltage and corresponding collector current transient ........................................ 52
2.7 Collector current transient in response to exponentially relaxing device operating temperature ................................. 54
2.8 Collector current transient and exponentially relaxing regression .... 56
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.9</td>
<td>Determination of isothermal collector current</td>
<td>59</td>
</tr>
<tr>
<td>2.10</td>
<td>Steady-state and isothermal Gummel plot</td>
<td>62</td>
</tr>
<tr>
<td>2.11</td>
<td>Steady-state and isothermal collector characteristics</td>
<td>64</td>
</tr>
<tr>
<td>2.12</td>
<td>Isothermal collector characteristic at elevated temperature</td>
<td>66</td>
</tr>
<tr>
<td>2.13</td>
<td>Temperature rise as a function of power dissipation</td>
<td>69</td>
</tr>
<tr>
<td>2.14</td>
<td>Experimental pulsed base voltage data</td>
<td>73</td>
</tr>
<tr>
<td>2.15</td>
<td>Base-emitter voltage and power dissipation transients</td>
<td>76</td>
</tr>
<tr>
<td>2.16</td>
<td>Temperature rise as a function of base bias rise time</td>
<td>77</td>
</tr>
<tr>
<td>2.17</td>
<td>Base-emitter voltage transients for differing collector-emitter drives</td>
<td>80</td>
</tr>
<tr>
<td>3.1</td>
<td>Si/SiGe device representation</td>
<td>84</td>
</tr>
<tr>
<td>3.2</td>
<td>Scaling dependence of experimental and theoretical thermal resistance parameters</td>
<td>88</td>
</tr>
<tr>
<td>3.3</td>
<td>Thermal capacitance as a function of the square root of emitter cross-sectional area for Si/SiGe heterojunction bipolar transistors</td>
<td>92</td>
</tr>
<tr>
<td>4.1</td>
<td>Device geometries</td>
<td>97</td>
</tr>
<tr>
<td>4.2</td>
<td>Analytic predictions of rectangular device thermal resistance</td>
<td>99</td>
</tr>
<tr>
<td>4.3</td>
<td>Three-dimensional analytical predictions of rectangular and square emitter thermal resistances</td>
<td>102</td>
</tr>
<tr>
<td>4.4</td>
<td>Effective area of square and rectangular devices</td>
<td>107</td>
</tr>
<tr>
<td>4.5</td>
<td>Experimentally-determined thermal resistances for various emitter geometries</td>
<td>111</td>
</tr>
<tr>
<td>4.6</td>
<td>Experimentally-determined thermal capacitances for various emitter geometries</td>
<td>115</td>
</tr>
<tr>
<td>4.7</td>
<td>Comparison of experimental and theoretical thermal resistances</td>
<td>116</td>
</tr>
</tbody>
</table>
A.1 Pulsed measurement circuit schematic ................. 132
Acknowledgements

I would like to thank M. K. Jackson for the encouragement and focused direction he has provided as my supervisor. That which I have learned from him extends far beyond the confines of this thesis. I would also like to thank D. L. Pulfrey for the opportunity to work with him. I recognize and greatly appreciate the effort they have both taken to ensuring my future success.

I thank my co-workers, R. Coenen, M. Movassaghi, R. Rosales, R. Ruo, and J. Zhang for their friendship and assistance. I appreciate A. St. Denis and M. Vaidyanathan for the many lengthy and enlightening discussions, and for entertaining my queries, however hypothetical they may have seemed.

I gratefully acknowledge funding from the Natural Science and Engineering Research Council of Canada (NSERC), the Advanced Systems Institute of British Columbia, and Hughes Aircraft of Canada. The project was funded by an NSERC Strategic Grant, and by Micronet. The work was supported by the many collaborators indicated in the text of this work, in addition to others; namely, J. R. Long at the University of Toronto, D. Marchesan and S. Voinigescu of Nortel Technology, J. Sitch and B. Surridge of the Advanced Technology Group at Nortel Technology, D. J. Walkey at Carleton University, and P. J. Zampardi of Rockwell Semiconductor.

I am especially grateful to my friends for prescribing lucid moments: thanks
to D. Jez for the peach and the orange, to C. Nally for keeping alive the spirit, and to R. St-Aubin for Thursdays, and the support of a true friend. C. McTait has earned an endless merry-go-round ride for her incessantly thoughtful, caring, and understanding nature. And finally to my parents, my feelings are warm enough to rival an Edmonton winter; they not only gave me vision but taught me to question all that I saw with it. To them I owe everything.

TODD CHRISTOPHER KLECKNER

The University of British Columbia
August 1998
For my best friend Dean, who taught me to look within for what is true.
Chapter 1

Introduction

1.1 Introduction to Thesis

1.1.1 Overview

This thesis describes the experimental methodology with which a complete and consistent thermal model can be constructed for bipolar semiconductor devices, and presents measured thermal parameters for a variety of material systems and device geometries. The pulsed-bias measurement scheme allows for determination of the parameters for a lumped thermal model. In contrast with other methods, this method requires no a priori assumptions regarding the temperature dependence of various device parameters. This technique was used to perform the first experimental study of thermal resistance scaling of deep-trench Si/SiGe heterojunction bipolar transistors, and an investigation of geometry-dependent thermal-parameter dependence in non-trench isolated AlGaAs/GaAs heterojunction bipolar transistors.
1.1.2 Summary of Results

The major focus of the work described in this thesis consists of the development of a pulsed-bias measurement system to determine thermal subcircuit parameters, namely the thermal resistance and capacitance, for inclusion in device models to account for variations in device operating temperature with changes in operating point. The magnitude and time response of this power-dependent operating temperature are described by two device parameters known as the device thermal resistance and capacitance. These parameters are determined with the application of a fast rise-time voltage signal to the base of the device to be characterized while simultaneously capturing the collector current dynamics that result with a signal-averaging, digital oscilloscope.

The pulsed-bias measurement technique is applied to a study of scaling of thermal parameters in deep-trench isolated, single rectangular finger Si/SiGe heterojunction bipolar transistors (HBTs). It is found that, while the model predicts accurate values of thermal resistance in the limit of long length devices, it overestimates the thermal resistance for devices with smaller aspect ratios. This effect is believed to be the result of a failure to account for the additional heat flow path introduced by the interconnect metallization with which electrical bias is applied to the device. In agreement with general results on the capacitance of three-dimensional structures, the thermal capacitance of the devices is found to vary linearly with the square root of the emitter area.

A study of the layout dependence of the thermal resistance and capacitance of AlGaAs/GaAs HBTs is performed to determine the device geometry with the optimum thermal performance. An investigation of fifteen different devices, of four differing geometries and a number of differing emitter areas has shown that for a
fixed emitter area, a two-finger rectangular emitter device has the best thermal performance of those studied. The modeled thermal resistance has been shown to be erroneous in the limits of small and large area devices; these errors have been attributed to neglect of interconnect metallization and temperature non-uniformity, respectively. The thermal capacitance of the devices has been shown to be geometry independent within the resolution of the experimental method, and linearly dependent on the square root of the emitter area.

1.1.3 Outline of Chapter

The remainder of Chapter 1 will deal with background material relevant to self-heating in bipolar devices and experimental methods used to investigate these effects. Section 1.2 will describe the physical origins of the self-heating process, the temperature dependence of device operation, as well as introducing the electrothermal feedback effect. Section 1.3 will detail the motivation for accurate characterization of the self-heating process. Section 1.4 will outline the approximations inherent to using macroscopic thermal parameters such as the thermal resistance and capacitance to model the temperature behavior of a system to steady-state or transient thermal phenomena, in addition to detailing the considerations and limitations of applying such a lumped thermal model to describe device self-heating. Section 1.5 will outline existing experimental techniques developed for thermal characterization of electronic devices, as well as presenting the thermal resistance calculation of non-trench-isolated rectangular source structures. Finally, Section 1.6 will outline the remaining chapters of the thesis.
1.2 Self-Heating and its Effect on Device Operation

Self-heating is the process by which the power dissipation of a device alters its own temperature. Because heat is generated only in a localized portion of a bipolar device, and the conduction of heat away from this region into the surroundings is limited, device temperature rises. In general, this temperature rise is nonuniform. The process is shown in Figure 1.1, which depicts a simplified picture of a typical integrated bipolar transistor. In this figure, the power dissipation is assumed to be confined to a volume $s$; in the case of a bipolar transistor, this will usually be the reverse-biased base-collector junction. Associated with device operation is a time-averaged power dissipation $P_D$, which results in a heat flow $Q$ from the thermal source $s$ to the backside of the substrate. The backside of the substrate is considered to be constrained by the external environment to a substrate temperature $T_{\text{sub}}$. Heat which originates at the thermal source and flows out the top of the device via the interconnect metallization is usually assumed to be negligible. The heat flow to the backside of the wafer, $Q$, coupled with the finite thermal conductivity, $\kappa$, of the semiconductor wafer on which the device is fabricated, results in a temperature rise above that of the backside of the wafer. In many cases it is necessary to account for the temperature dependence of the thermal conductivity, and also for the possibility that the power dissipation is temperature dependent; this last effect causes electrothermal feedback, which will be discussed in the next paragraph. Solution of the equations describing the non-linear system illustrated in Figure 1.1 gives the temperature distribution throughout the device under study; it is this temperature rise that is the direct result of self-heating.

In order to fully characterize the effects of self-heating, it is helpful to understand the way in which increases to the operating temperature at the device as
Figure 1.1: Device operation results in a thermal source $s$ of power dissipation $P_D$. The resultant heat flow $Q$ coupled with the substrate thermal conductivity $\kappa(T)$ results in a temperature rise of the source $\Delta T_s$ above the externally-constrained substrate temperature $T_{\text{sub}}$. 
described previously will subsequently alter device performance. In the following
discussion a simple, analytic model is used to describe the temperature dependence
of an AlGaAs/GaAs heterojunction bipolar transistor operation at constant base-
emitter voltage. This model will be used for a qualitative description of electrother­
mal feedback, the thermal runaway process, and will provide a means by which
experimental limitations can be determined and subsequent analysis of experi­
mental results performed. While the following equations and discussion are specific to an
AlGaAs/GaAs heterojunction bipolar transistor, the same overall trend is observed
in InGaP/GaAs and Si/SiGe heterojunction bipolar transistors.

To quantify the effects of increasing temperature on a double heterojunction
bipolar transistor operated at constant base-emitter voltage, consider the form of
the collector electron current density $J_n$ as expressed by St. Denis et al.

$$J_n = -\frac{q}{N_B} \left[ \exp\left(\frac{V_{BE}/kT}{T}\right) - \exp\left(\frac{V_{BC}/kT}{T}\right) \right]$$

where $q$ is the electronic charge, $k$ is the Boltzmann constant, $T$ is the absolute
temperature of the device, $N_B$ is the base doping density, and $V_{BE}$ and $V_{BC}$ are the
biases applied to the base-emitter and base-collector junctions, respectively[1]. For
uniform base doping, the three terms in the denominator of Equation 1.1 are given
by

$$\Theta_I = \frac{W_B}{D_n n_i^2}$$
$$\Theta_{EB} = \frac{1}{n_i^2 v_E \exp(-\Delta E_{nE}/kT)}$$
$$\Theta_{CB} = \frac{1}{n_i^2 v_C \exp(-\Delta E_{nC}/kT)}$$

where $W_B$ is the quasi-neutral base width, $D_n$ is the electron diffusion coefficient
in the base, $n_i$ is the intrinsic carrier concentration in the base, $v_E$ and $v_C$ are
the electron thermal velocities in the emitter and collector regions, respectively, and \( \Delta E_{nE} \) and \( \Delta E_{nC} \) are the electron energy barriers as seen by electrons exiting the base through the base-emitter and base-collector junctions, respectively, if the band-bending in the base is very small. These terms describe the various barriers to electron transport in the device, namely, diffusion across the quasi-neutral base, and thermionic transport over the base-emitter and base-collector junctions of the device.

Equation 1.1 can be simplified by considering the temperature dependence of charge flow in an AlGaAs/GaAs single heterojunction bipolar transistor, where the base-emitter junction aluminum content has been sufficiently graded to remove the discontinuity in the conduction band, as is usually the case for improved device current gains\[2\] and reproducibility of base-emitter junction turn-on potential\[3\]. For a single heterojunction bipolar transistor, \( \Delta E_{nC} = 0 \) as the base and collector regions of the device are fabricated from the same material and, as concluded by St. Denis et al., adequate grading of the base-emitter junction effectively reduces \( \Delta E_{nE} \) to zero. Therefore, for such a device, \( \Theta_I \) is the main impediment to charge flow in the device. With this, investigation of the temperature dependence of the resulting simplified equation can be performed numerically; the temperature-dependent energy bandgap \( E_g(T) \) is taken to be

\[
E_g(T) = E_{g0} + \frac{\alpha T^2}{T + \beta}
\]

(1.2)

where \( E_{g0} \) is the energy bandgap at zero Kelvin, taken to be 1.52 eV, \( \alpha \) is \(-5.4 \cdot 10^{-4} \) eV/K\(^{-2}\) and \( \beta \) is 204 K\[4\]. The temperature dependence of the dielectric constant \( \varepsilon(T) \) of GaAs is

\[
\varepsilon(T) = \varepsilon_0[1 + B\varepsilon(T - 300)]
\]

(1.3)
where $B_c$ is taken to be $0.9 \cdot 10^{-4} \text{ K}^{-1}$ and $\epsilon_{300}$ is the dielectric constant at 300 K, which is taken to be 13.1 times the permittivity of free space\cite{4}. The temperature dependence of the electron mobility in n-type material is

$$\mu_n(T) = \mu_{n300} \left(\frac{300}{T}\right)^{2.3}$$

(1.4)

where $\mu_{n300}$ is the electron mobility at 300 K\cite{5} and is estimated to be 0.315 m$^2$/Vs\cite{6}. For the purposes of this study, the mobility temperature dependence in p-type material is assumed to be the same as that observed in n-type material.

The calculation is performed for a structure with emitter, base, and collector doping concentrations of $5 \cdot 10^{17}$, $5 \cdot 10^{19}$ and $5 \cdot 10^{16}$ cm$^{-3}$, respectively. The effective masses of electrons and holes are taken to be 0.067 $m_e$, and 0.48 $m_e$, respectively, where $m_e$ represents the electron rest mass. The metallurgical base width is 600 Å. The structure is biased in the active region, with $V_{BE} = 1.3 \text{ V}$ and $V_{CE} = 2.0 \text{ V}$, and is based on the device as described by Lester et al.\cite{2}, for which later measurements will be presented in Chapter 2.

Figure 1.2 depicts the variations for the three main constituents of Equation 1.1, namely the temperature variation of the square of the intrinsic carrier concentration, the quasi-neutral base width, and the exponential term, each normalized to their room temperature values. While the temperature dependence of the quasi-neutral base width does not play a significant role in the overall temperature dependence of the current, the competing effects of an increased intrinsic carrier concentration and the exponential dependence of collector current on the thermal voltage are substantial. Over a temperature increase of 200 K, the square of the intrinsic carrier concentration increases by more than 11 orders of magnitude, while the term describing the effect of the change in thermal voltage decreases by more than 9 orders of magnitude. The total temperature dependence of the collector cur-
Figure 1.2: Temperature dependence of components of collector current over a temperature range from 300 K to 500 K. Shown is the square of the intrinsic base carrier concentration (solid line), the quasi-neutral base width (dashed-dotted line), and the exponential term (dotted line). The temperature variation of $n_i^2$ is greater than that of the exponential term, resulting in an increase of collector current for increasing temperature as shown by the product of the terms (dashed line).
rent, given by the product of the three temperature-dependent terms, increases with increasing temperature owing to the strong temperature dependence of the intrinsic carrier concentration. While these calculations are specific to an AlGaAs/GaAs heterojunction bipolar transistor, the observed trend is not material dependent; InGaP/GaAs, AlGaAs/GaAs, and Si/SiGe heterojunction bipolar transistors exhibit increased collector currents for increasing temperatures when operated at constant base-emitter voltage.

This tendency of collector current to increase with increasing temperatures at constant base-emitter voltages is the origin of electrothermal feedback, which arises when increasing temperatures result in a larger collector current which increases the device power dissipation. This increase in power dissipation results in more self-heating, which again causes a larger temperature rise and a corresponding increase in collector current. This cyclic effect, termed electrothermal feedback, can compromise device performance by subjecting the device to significantly elevated temperatures and in extreme cases can lead to device instability and failure through a process referred to as thermal runaway.

1.3 Motivation

From the discussion in the previous section, it is apparent that self-heating can cause substantial changes in device operation if temperature rises are significant; it will be shown in later chapters that this is the case. Despite this, parameters for use in circuit simulator device models are usually based on steady-state measurements of device performance at a single substrate operating temperature. Because of the temperature rises that occur due to self-heating, the resulting steady-state measurements constitute a temperature-dependent characterization. Attempting to fit such
temperature-dependent data with models based on isothermal device operation will lead to errors. For example, a number of factors contribute to the complicated dependence of current-gain cutoff frequency, $f_t$, on collector current. If self-heating plays a role in the decrease of $f_t$ at high collector currents, but is not included in the models, then the magnitudes of other effects such as base pushout will have to be over- or under-estimated to obtain agreement with the experimental data. Another example is in the extraction of emitter resistance, where thermal effects play a large role. The resulting models are no longer physically based, and their predictive capabilities can be severely restricted if the final operating point and thermal environment of the device differ from the one in which it was characterized. The unphysical nature of the parameter extraction can also make it difficult to develop truly scalable models, in which the emitter dimensions of the device are given as model parameters. The absence of such a scalable model has been a significant restriction in the development of bipolar circuits, and models are typically available only for a fixed set of specific device sizes.

The magnitudes of the errors described in the previous paragraph will be lowest for small-signal circuits, where the device currents and voltages remain close to a given operating point. In this case, it is sufficient to determine model parameters at operating points close to the final circuit operating conditions; self-heating will occur, and the device temperature will rise, but the rise will be the same in characterization and in operation. Assuming the model still presents a reasonable description of the device physics at elevated temperatures, results can be expected to be reasonably accurate with this approach. The results with large-signal circuits, where currents vary dramatically during operation, are expected to be less successful. In a digital circuit, for example, device currents often vary from a relatively
low level in the off state to a high level in the on state; typically in high-speed bipolar circuits, this on-state power dissipation is high enough to cause substantial self-heating. During circuit operation this switching between states happens at the clock frequency of the circuit, which can be as high as 10 Gb/s for the devices considered in the present work. On this time scale, the temperature of the device cannot change, but will come to a nearly constant value based on the time-averaged power dissipation. In this situation, it is necessary to be able to predict device operation for particular operating points at device temperatures that are never attained in steady-state operation at the same currents and voltages. As such, an accurate description of the self-heating process for bipolar devices is of critical importance to circuit designers who rely on accurate device models.

Not only is self-heating of importance to integrated circuit designers, but is of paramount concern for reliability engineers because device operation at elevated temperatures generally results in a shorter device lifetime. Knowledge of the degree of self-heating is needed to extrapolate expected product lifetimes from elevated-temperature accelerated-failure testing. This allows calculation of the maximum power dissipation allowable while maintaining device temperature within safe operating limits based on statistical estimates of device lifetime.

1.4 Lumped Thermal Model

For a small class of thermal problems, it is possible to simplify greatly the analysis of the dynamic response of the system by invoking the assumption that the system can be adequately described by a lumped thermal model. This approach is based on the direct analogy that can be drawn between electrical and thermal problems. In general, the resulting thermal subcircuit must be quite complex to describe the
full problem accurately. However, it is often simplified into a few elements, from which the lumped thermal model is derived. The resulting simplification is so great that it is almost a prerequisite for any practical circuit simulation with models including self-heating. The most common equivalent circuit representation is shown in Figure 1.3. As shown in Figure 1.3(a), the system response is described by three main elements, the first of which describes the instantaneous power dissipation of the device being considered and is modeled via a current source with dimensions of Watts. The second component, and the focus of Section 1.4.1, is the thermal resistance. It is a parameter which describes the steady-state temperature rise above ambient for a given power dissipation, and is expressed in units of Kelvin per Watt. Finally, the thermal capacitance is a parameter which, together with the thermal resistance, models the transient response of the device temperature when the power dissipation changes either from self-heating, a change in operating point, or a combination of both. A discussion of the thermal capacitance will be provided in Section 1.4.2. The dimensionality of the thermal capacitance is Joules per Kelvin. The voltage developed across the parallel $R_{TH} C_{TH}$ circuit is equal to the temperature rise of the device. Incorporation of a voltage supply equal in magnitude to the ambient temperature of the surrounding environment and located at the negative node of the $R_{TH} C_{TH}$ network allows for the generation of a voltage signal equal to the absolute device temperature.

The relevant theory along with the underlying assumptions fundamental to the thermal resistance concept will be introduced in Section 1.4.1. The corresponding theory and limitations underlying the thermal capacitance will be detailed in Section 1.4.2. The dynamical response of such a process will be detailed in Section 1.4.3. While the present discussion has been limited to a general description of
Figure 1.3: Thermal subcircuits to model static and dynamic response of lumped thermal systems; (a) describes the temperature rise of a thermal source with power dissipation $P_D$ and (b) models the absolute temperature of a system in an ambient environment of $T_{REF}$. 
the thermal source, there are certain limitations when the above model is applied to electronic devices; Section 1.4.4 will discuss the limitations involved with applying the thermal resistance concept to bipolar electronic devices.

1.4.1 Thermal Resistance Fundamentals

Justification for the basic idea of using a linear proportionality constant to relate the temperature rise to the power dissipation can be found by investigating the steady-state solutions to the heat flow problem. The Fourier field equation[7],

\[ C\rho \frac{\partial T(r, t)}{\partial t} = \nabla(\kappa \nabla T(r, t)) + Q \]  

(1.5)

describes the relation between the spatially- and temporally-varying temperature \( T(r, t) \) and \( Q \), the heat generated per unit time per unit volume. In this equation, \( C \) is the heat capacity per unit mass, \( \rho \) is the density, and \( \kappa \) is the thermal conductivity. As a demonstration of the thermal resistance concept, consider the heat flow through an insulated bar of cross-sectional area \( A \) and thickness \( W \), as shown in Figure 1.4[7]. The boundaries at \( x = 0 \) and \( x = W \) are assumed to be maintained at constant temperatures \( T_H \) and \( T_C \), respectively. For such a geometry, steady-state heat flow is one-dimensional, and if the thermal conductivity is independent of temperature, the temperature distribution is

\[ T(x) = T_H - (T_H - T_C) \frac{x}{W} \]  

(1.6)

The total heat flow through the bar, \( Q \), is related to the temperature difference as expressed in the one dimensional Fourier rate equation

\[ Q = -\kappa A \frac{dT}{dx} \]  

(1.7)
Figure 1.4: One-dimensional insulated bar. Heat flow $Q$ originates at the hot interface at $x = 0$ and terminates at the cold interface at $x = W$; the cross-sectional area of the bar is $A$.

It is easy to show that the total heat flux into the hot boundary $Q$ is related to the temperature difference

$$\Delta T = T_H - T_C \quad (1.8)$$

by

$$\Delta T = QR_{TH} \quad (1.9)$$

where

$$R_{TH} = \frac{W}{\kappa A} \quad (1.10)$$

Comparison of Equation 1.9 to Ohm's Law, $\Delta V = IR$, allows a useful analogy to be drawn between the thermal and electrical problems, by considering the heat flow to be analogous to the electric current, and temperature difference to be analogous to potential difference. In a similar fashion to electric circuit theory, thermal resistances can be combined for various heat transfer paths depending on whether the heat flow
occurs in series or parallel. Equation 1.10 shows that the thermal resistance for an insulated bar is strictly a function of the geometry of the problem and the material properties.

While the example illustrated is a simple, one dimensional problem, it is possible to determine thermal resistances for more complicated geometries in which multidimensional heat flow occurs. As the thermal resistance is a function of both system geometry and material properties, it is common to consider the thermal resistance to be comprised of two factors such that

$$R_{TH} = \frac{1}{\kappa S}$$  \hspace{1cm} (1.11)

where the shape factor $S$ defines the contribution of geometry to the thermal resistance, and the thermal conductivity $\kappa$ takes into account the material property governing heat flow. Shape factors have been calculated for many differing geometries\[8, 9\] which are of much greater complexity than the insulated bar considered previously. The first of two geometries that approximates the thermal behavior of heterojunction bipolar transistors is the isothermal rectangular parallelepiped, buried in a semi-infinite medium, shown in Figure 1.5. The corresponding shape factor $S_p$ is\[10]  

$$S_p = 1.685L \left[ \log_{10} \left( 1 + \frac{b}{a} \right) \right]^{-0.59} \left( \frac{b}{c} \right)^{-0.078}$$  \hspace{1cm} (1.12)

where $a$, $L$ and $c$ are the width, length, and height of the thermal source and $b$ is the depth of the thermal source from the isothermal plane. This expression is accurate only for devices where $L$ is much larger than $a$.

The second approximation to an electrical device consists of an identical geometry as that depicted in Figure 1.5, save for a negligible thermal source thickness.
Figure 1.5: Thermal system geometry for calculation of the shape factor for a parallelepiped buried in a semi-infinite medium. The thermal source of width $a$, length $L$ and thickness $c$ is a distance $b$ from the isothermal interface.
The shape factor $S_r$ is given as [10]

$$S_r = 1.45L \left[ \log_{10} \left( 1 + \frac{b}{a} \right) \right]^{-0.59} \quad (1.13)$$

where $L$ is the length of the thermal source, assumed to be much larger than the width $a$; $b$ represents the distance of the thermal source from the isothermal boundary, and for our calculation is given by the substrate thickness. Equation 1.13 is a one dimensional representation of the true problem, where longitudinal heat flow is neglected by assuming that the length of the thermal source is much larger than the width; heat flow originating from the sides of the thermal source is also neglected through the assumption of a negligible source thickness. While considering that the above two equations are empirically based, it is worth noting that the two relations do not converge in the limit that the source thickness $c$ approaches zero.

The only assumptions underlying the calculation of a thermal resistance for a given structure are a uniform source temperature, temperature-independent geometry, and temperature-independent material properties. Violation of any of these underlying assumptions will result in a non-linear temperature rise as a function of device power dissipation, or a failure of the basic idea that the device temperature can be described by a single number.

Joy et al. have performed theoretical studies on the variation of thermal resistance of a rectangular parallelepiped thermal source buried in a semi-infinite medium as a function of the source dimensions [11]. Their study is based on integration of the instantaneous point source over the volume of the thermally active region to yield an estimate of device thermal resistance, and provides a comprehensive study of scaling of thermal resistance. The calculation assumes negligible effect from the boundaries of the problem, save for the presence of the adiabatic interface at the top surface of the semiconductor wafer, which is accounted for with an im-
age source reflected about this plane. The model neglects the effect of conduction through interconnect metallization; the magnitude of this effect, estimated to be in excess of 15 K/mW for their devices, does not affect the thermal resistance of the large area devices studied in their work with corresponding thermal resistances on the order of 0.4 K/mW. Finally, the model assumes the power dissipation of the thermal source is independent of position inside the source. Joy et al. show that using the above assumptions and model description, the thermal resistance of most devices can be shown to vary as

\[ R_{TH} = \frac{1}{4\kappa(LW)^{1/2}} \]  

(1.14)

where \( L \) and \( W \) are the thermal source length and width, respectively. This approximate solution was arrived at by numerical calculations of the temperature at a particular location at the center of the top of the heat generating region for various magnitudes of the source dimensions. Joy et al. show that, for most reasonably shaped sources, the product of the effects due to the finite lateral dimensions, the thickness, and the buried depth of the thermally active region suggests that the thermal resistance scales linearly with the inverse of the square root of the thermal source cross-sectional area. Not only does Equation 1.14 suggest a simple scaling law, but also provides an absolute estimate of the thermal resistance for a rectangular, thermally active region. This scaling law will be tested in Chapters 3 and 4 on experimentally determined data for both Si/SiGe trench and AlGaAs/GaAs non-trench isolated devices.

1.4.2 Thermal Capacitance Fundamentals

While it has been shown that a number of prerequisites must be satisfied in order to apply the thermal resistance concept to a given thermal problem, full application
of the lumped thermal model also requires that certain assumptions inherent to the concept of a lumped thermal capacitance be satisfied. The fundamental assumption underlying the lumped thermal capacitance model is that the temperature of the thermal source is spatially uniform throughout the heat transfer process. This assumption presupposes that the temperature gradients within the source are negligible, and that the vast majority of the temperature variation occurs outside of the source. From Fourier's Law, heat conduction in the absence of temperature variation implies an infinite source thermal conductivity, which, while impossible to achieve, can be well approximated for a limited class of systems where the external surroundings have a much larger thermal resistance than the source itself.

The theory originally devised for estimation of whether or not a lumped thermal capacitance adequately describes the relevant dynamics was based on the cooling of a molten or hot solid within a fluid bath, usually water[7]. From this, applicability of the lumped thermal capacitance was estimated by quantifying the internal source thermal resistance to that resistance external to the source. Here, one considers the Biot number $Bi$, defined as

$$Bi = \frac{hL_c}{\kappa}$$

(1.15)

where $h$ represents the convection heat transfer coefficient between the source of thermal conductivity $\kappa$ and the external surroundings; $L_c$ is the characteristic length,

$$L_c = \frac{V}{A_s}$$

(1.16)

and is quantified by the ratio of the thermal source volume $V$ to its surface area $A_s$. The Biot number provides a measure of the temperature drop in the solid relative to the temperature difference between the solid surface and the fluid bath, assumed to remain at a constant temperature[7]. The lumped thermal capacitance model can
be used without incurring appreciable error in predicting dynamic responses when the Biot number is much less than unity, so that for \( Bi = 0.1 \), an error of less than 5% is incurred[12]. The thermal capacitance is given by[7]

\[
C_{TH} = \rho V C
\]  

(1.17)

where \( \rho \), \( V \), and \( C \) are the density, volume, and specific heat of the hot material, respectively. As the volume of material heated during the self-heating process in a semiconductor die is not known exactly, estimation of a thermal capacitance based on Equation 1.17 is difficult.

As a result of the lack of information pertaining to the dynamics of localized heating of a solid, the applicability of the thermal capacitance concept will be investigated experimentally. As will be shown in Section 1.4.3, one measure of the applicability of a thermal capacitance is the exponentially-decaying dynamics in response to a step change in the power dissipation. This test will provide evidence of the applicability of the thermal capacitance for the devices studied within this work.

Theoretical studies of the electrical capacitance of three-dimensional structures, the results of which will be applied in subsequent chapters, were reported by Y. L. Chow and M. M. Yovanovich[13]. By considering a sphere as the basis for their comparative study, they show that small perturbations made to the shape of the structure only change the capacitance of the structure slightly, given that the surface area of the object remains constant. They also show that the capacitance for a given geometry scales linearly with the square root of the surface area of the structure. From this analysis, it is possible to predict the capacitance of another similar structure \( C_s \), with surface area \( A_s \), via knowledge of the capacitance of another structure, \( C_a \) with corresponding area \( A_a \). The first step is to account for the
differing shape of the conductor, which is facilitated by tabulated factors[13]. These factors demonstrate, for instance, that the capacitance of a thin, circular disc and a thin, square disc only differ by 0.4% for identical surface areas. They also show that large perturbations to the square disc, such as those required to elongate the structure into a 4:1 aspect ratio rectangular disc, only result in an 11% variation in the structure capacitance. The second step is to account for the change to the surface area of the structure, which is easily accomplished by normalizing by the square root of the surface area of the known structure, \( A_s \) and multiplying by the square root of the surface area of the desired structure \( A_a \). Overall, their studies indicate a maximum error of 3%, and they show that at most a 5% error in incurred in using the analytic forms they derive, over a wide range of aspect ratios.

While the model of Reference [13] does not allow for an absolute calculation of the thermal capacitance, it does provide valuable information pertaining to the scaling of thermal capacitance for similarly-shaped structures. Tests of this scaling law will be applied to the experimental measurements of Chapters 3 and 4.

### 1.4.3 Temporal Response of Lumped Thermal System

The utility of the lumped thermal model comes in the simplicity in predicting the temporal evolution in response to thermal transients. The transient response of the source temperature \( T(t) \) to a step change in power dissipation is given by

\[
T(t) = T_{SS} + (T_o - T_{SS}) \exp(-t/\tau_{TH})
\]  

(1.18)

where \( T_o \) and \( T_{SS} \) are the initial and steady-state temperatures of the thermal source, respectively, and \( \tau_{TH} \) is the thermal time constant, given by

\[
\tau_{TH} = R_{TH} C_{TH}
\]  

(1.19)

23
where \( R_{TH} \) and \( C_{TH} \) are the thermal resistance and capacitance of the system, respectively. The temporal response of an ideal lumped thermal system to a step increase in the power dissipation at time \( t = 0 \) is shown in Figure 1.6. Both the normalized power dissipation and the fractional rise in temperature to steady state have been plotted as a function of time normalized to the thermal time constant. As with single-pole RC circuits, the system response relaxes to 63% of its steady state value after one thermal time constant. With Equation 1.18, it becomes simple to predict the temporal evolution of the system to changes in the levels of excitation, as the steady-state temperature can be predicted via the level of power dissipation and the thermal resistance, and the initial temperature of the body is assumed known. When the power dissipation of the thermal source is a function of time, the differential equation governing the voltage rise of a parallel RC network must be solved.

The use of the lumped model and the associated exponential step response is valid for those cases where application of the thermal resistance and capacitance concepts is justified. As stated previously, a lumped formulation requires the primary barrier to heat flow to exist outside the thermal source, and that the temperature of the thermal source to be spatially uniform. However, as will be detailed in Section 1.4.4, the lumped formulation may not always apply to practical devices.

1.4.4 Application of Thermal Resistance to Device Self-Heating

The focus of this section is to highlight the main difficulties in applying a lumped thermal resistance to electronic devices. The first suspect requirement is that of a uniform device temperature. Oettinger et al.[14] have shown that a spatially uniform junction temperature is not achieved for large-area power transistors. This effect
Figure 1.6: Temporal response of lumped thermal system to excitation by step increase in power dissipation at time zero. The solid line shows the fractional temperature rise where zero corresponds to the initial temperature, and unity the steady-state temperature. The dashed-dotted line show the normalized power dissipation as a function of time from excitation, normalized to the thermal time constant.
is especially pronounced for situations in which large temperature rises occur, such as high-power devices that undergo substantial electrothermal feedback. A non-uniform junction temperature can cause systematic deviations between theory and experiment for a measure of the thermal resistance determined at a single operating point.

A related difficulty that occurs with non-uniform device temperature is that for reliability reasons, device designers are interested in the peak junction temperature. Lumped element models for large-area devices fail to ensure safe device operation as the peak junction temperature can be significantly higher than that predicted by the product of the power dissipation and the experimentally determined, mean thermal resistance.

Another issue related to the appearance of a non-uniform junction temperature is the dependence of the thermal resistance on the terminal bias conditions for a fixed power dissipation. This can cause discrepancies between multiple measurements of thermal resistance for the thermal operating points that are apparently identical. Oettinger et al.[14] show that experimentally determined values of thermal resistance depend on the collector current and collector-emitter voltage that govern the device power dissipation. Inherent to the lumped thermal model is the assumption that the temperature rise is a function only of the power dissipation and not a function of the device terminal biases themselves. However, in Reference [14] it was shown that $R_{TH}$ not only depends on device power dissipation, but is observed to vary even for the case where differing values of the collector current and collector-emitter voltage give the same power dissipation. From this, it is apparent that the collector current distribution plays a role in determining the overall temperature rise of the junction. It was also shown that for low levels of collector
current at high collector-emitter voltages, the thermal resistance is a strong function of the device operating conditions, while for the converse case the thermal resistance varies to a much lesser degree. This fact is a result of the pronounced electrothermal feedback that takes place at large collector-emitter voltages. Under low voltage and higher current operation, the current is contained near the emitter periphery and the maximum junction temperature rise is less than the converse scenario at high voltages and low currents[15]. Operation with strong electrothermal feedback results in current constriction and eventual thermal runaway, also referred to as second breakdown; this greatly alters the effective thermal source volume by focusing the collector current through a fraction of the available cross-sectional area of the emitter-base junction.

Finally, non-linear temperature rises occur when temperature-dependent material parameters such as the energy bandgap, thermal conductivity[5], saturation velocities and carrier mobilities[16] are considered. Usually instances where the thermal resistance is found vary with power dissipation are attributed to the decrease of thermal conductivity with increasing temperature[17, 18, 19].

From this discussion, it is apparent that most of the limitations imposed on measuring thermal resistances for bipolar devices involve either large-area devices or the comparison of measurements performed at greatly differing operating points. Large-area devices are particularly prone to violating the assumptions of a thermal resistance because over the base-emitter junctions of such devices, appreciable temperature gradients can occur. Deviations between theory and experiment for such large-area devices will be shown in Chapters 3 and 4. Finally, in order to draw meaningful comparisons between experimental measurements, all of the measurements to appear in the following chapters will be performed at moderate collector-emitter
voltages with small amounts of collector current. These small levels of power dissipation result in small temperature rises of the device, over which the previously described non-linear tendencies of the bandgap, thermal conductivity, saturation velocities and mobilities can be ignored.

1.5 Existing Experimental and Theoretical Techniques and Results

Methods used to measure the temperature rise as a function of power dissipation of electronic devices fall into two general categories, optically- and electrically-based techniques, each of which have associated strengths and drawbacks. This section will outline the basis of these measurements, and will summarize their merits and weaknesses. Also included is a brief overview of theoretical methods used to calculate device thermal resistance. To date, no extensive experimental or theoretical study has been provided for thermal capacitances of electronic devices.

1.5.1 Experimental Optical Methods

Optical techniques have been investigated to ascertain device operating temperature owing to the direct observation nature of the measurement process. Two main optical techniques that have been developed for measurement of the thermal resistance are the liquid crystal method and infrared imaging[20].

The liquid crystal measurement technique is based on observation of the change in polarization state of a liquid crystal when it undergoes a phase change at a specific, characteristic temperature[4]. From an experimental standpoint, the procedure involves coating the device top surface with a liquid crystal, and varying
the power dissipation of the transistor while monitoring the polarization of light reflected from the surface. As the power dissipation is increased, the temperature on the surface of the wafer increases; by noting the power dissipation at which the polarization state changes, the thermal resistance can be determined. As each device must be prepared specially for this measurement, this technique is usually used to corroborate thermal resistance measurements based on less destructive electrical schemes[21] that will be the focus of the discussion of Section 1.5.2. The spatial resolution of such a technique is, in principle, limited to that of optical microscopes, and as such, resolutions on the order of 500 nm are achievable. This resolution limit allows the method to be applied to small-area devices. However, such a measurement probes the top surface of the device, which is not necessarily the same as, for example, the base-emitter junction. It is also possible that the presence of the liquid crystal on the top surface is sufficient to alter the thermal resistance of the device under test[4]. Finally, as a single device can only be coated with one liquid crystal, only one temperature transition can be detected per device. In order to measure more temperature contours, many identical devices would need to be prepared, each with a differing liquid crystal.

The second optical method consists of infrared imaging of the transistor structure to determine the junction temperature. By coating the top surface of the wafer with a thermographic phosphor to equalize the emissivity of the various device regions, post-measurement correction factors that account for differing emissivities can be avoided. While such a technique allows for investigation of the spatial non-uniformity of device operating temperature[14], the low spatial resolution of the long-wavelength infrared radiation is a problem. Bailbe et al.[22] approach the spatial resolution limit of 5-10 µm by using this technique on a 10 by 200 µm² device.
However, it was concluded by Negus et al.[20] that the resolution limit of infrared imaging leads to non-repeatable estimates of the thermal resistance for small-area devices. This conclusion was arrived at by studying a multi-finger structure, with 1 \( \mu \text{m} \) wide, 25 \( \mu \text{m} \) long emitter fingers separated by 4 \( \mu \text{m} \). Therefore, this method would not be useful to study the majority of devices investigated in the present work. Furthermore, as with the liquid crystal case, deposition of the thermographic phosphor is a preparation-intensive and invasive process. The possibility exists that the presence of a foreign body in contact with the device being studied may affect the measured thermal resistance.

Both optical methods are limited by the fact that only the temperature distribution of visible surfaces can be measured, and as such, these optical methods would exclude devices packaged with flip-chip bonding and make measurements of temperatures internal to the die impossible. Ideally, one would like to have the ability to measure the emitter junction temperature because the temperature distribution is not uniform within the device, and the most significant impact of temperature rise on device performance occurs at this location. However, as it is impossible to investigate the temperature variation with depth, the temperature of the base-emitter junction cannot be determined. The physical separation of the base-emitter junction and the top surface of the semiconductor wafer will place an upper limit on the error incurred by approximating the junction temperature as that of the top surface.

1.5.2 Experimental Electrical Methods

Electrical characterization techniques are based on measurements during device operation, and have the advantage of being non-invasive. Some steady-state[20, 23]
techniques rely on determining device temperature via calibration of a tempe­

sensitive parameter such as the forward current gain, $\beta_F$, or the base-emitter voltage

required to maintain a small constant collector current. These quantities are char­

acterized at low power dissipation as a function of temperature. Measurements

are then made as a function of power dissipation for at least two heat sink tem­

peratures. First, the difference in power dissipation required to result in the same

junction temperature, as probed by $\beta_F$, for example, is determined. Because the

differences in heat sink temperature are known, and this difference is due to the

increased power dissipation, the thermal resistance can be determined. These tech­

niques ignore changes in the probed quantities, such as $\beta_F$, with base-collector bias

due to effects such as base-width modulation. This assumption has been shown to

be incorrect in the literature for bipolar junction transistor characterization[14] but

has not been perceived to be an issue for heterojunction bipolar transistors because

of the the high base doping in these devices.

A similar calibration-based approach has been used in pulsed techniques

[11, 14, 17], which are based on switching the device from a high- to a low-power-

dissipation state, and sampling of a temperature-sensitive parameter just after

switching. If the sampling occurs quickly enough, this can be used to deduce the

device temperature prior to switching. These techniques require fast electrical mea­

surements at low signal levels, and in practical measurements often require correction

of the temperature-sensitive parameter for non-zero device power dissipation of the

low power dissipation state to achieve reasonable experimental sensitivity.

Other steady-state techniques have been demonstrated. One approach in­

volves measurements of the base current as a function of base-collector bias[24]. As­

suming that the base-collector space-charge-region thermal generation component
of the base-current in negligible, an observation of the change of base current with changing base-collector bias can be attributed to device self-heating. Another of the main assumptions, as indicated by Tran et al.[25], is a linear approximation of the natural logarithm of the ratio of the temperature-dependent to room-temperature base current. This assumption has been shown to be problematic due to the tendency of series resistances to decrease the slope of this ratio as a function of power dissipation.

Another approach is based on simultaneous extraction of the parasitic emitter resistance and thermal resistance[25, 26]. These methods rely on an iterative method consisting of extraction of the emitter resistance from the high-current region of the Gummel plot, and calculation of the thermal resistance to de-embed thermal effects from the Gummel plot. This allows a successive calculation of the emitter resistance from the latest estimate of the isothermal Gummel plot. This procedure continues until convergence is reached, which usually occurs after five to ten iterations. This method assumes linearity of the forward current gain parameter with temperature, which has been shown in the literature[23] to be in error be approximately 10%, and assumes the major temperature variation of the base saturation current is due to the intrinsic concentration.

More recently, Schaefer et al. described a pulsed measurement system based on the measurement of device operation immediately after switching from a low-power state [27]. By making the measurement quickly enough after switching, the operation under bias, but with the device at the ambient heat-sink temperature, can be characterized. As this technique for isothermal characterization forms the basis for the techniques developed in this thesis, a more detailed description of its operation is deferred to Chapter 2.
The techniques described above have been used to measure the thermal resistances of a number of devices, most of which are power devices. Dawson et al. applied a steady-state scheme to an eight-finger 1.5 W AlGaAs/GaAs power transistor, with each finger measuring $2 \times 20 \ \mu m^2$[23]. Negus et al. measured thermal resistance of a 0.5 W silicon bipolar device, where the principal issue was the thermal resistance of the packaging materials[20]. Adlerstein et al. show thermal resistance measurements for $3 \times 10$ and $3 \times 20 \ \mu m^2$ AlGaAs/GaAs heterojunction bipolar transistors[17]. Reisch et al. show thermal resistances for two Si bipolar transistors with $16 \times 16$ and $2 \times 128 \ \mu m$ emitters[24]. Oettinger et al. measured thermal resistances for multifinger Si bipolar transistors at power dissipations in excess of 50 W[14]. Only a few results pertinent to relatively small-area bipolar transistors have been presented. Joy et al. provided pulsed electrical measurements of thermal resistance for a $2.5 \times 15 \ \mu m^2$ Ge and a $2.5 \times 12.5 \ \mu m^2$ Si transistor[11]. Hanington et al. show thermal resistance measurements for the same $1.4 \times 3 \ \mu m^2$ AlGaAs/GaAs heterojunction bipolar transistor studied in the present work[26]. Tran et al. show thermal resistance measurements for a $0.5 \times 10 \ \mu m^2$ trench-isolated Si/SiGe heterojunction bipolar transistor, and a $0.5 \times 20 \ \mu m^2$ non-trench-isolated Si bipolar device[25]. Marchesan has measured thermal resistance for the same $0.5 \times 20 \ \mu m^2$ Si/SiGe HBT studied in the present work[28]. As of yet, no experimental study has been performed on the scaling or geometry-dependence of thermal resistance of bipolar transistors, and, as mentioned earlier, no measurements have been reported for thermal capacitances.
1.5.3 Theoretical Methods

Many differing theoretical and numerical models have been investigated for the estimation of device thermal resistance. One-dimensional models have been arrived at by coupling the heat-transfer equation with Poisson's equation and the continuity equations[29]. A self-consistent, numerical two-dimensional model has been developed that allows for the investigation of the electrothermal effect in AlGaAs/GaAs heterojunction bipolar transistors[5]. Higgins developed a numerical thermal model, which has shown that from a thermal standpoint, the ideal emitter finger for a single finger device is long and narrow[30]. Liou et al. developed a numerical, two-dimensional model to study the temperature dependence of AlGaAs/GaAs heterojunction bipolar transistors[16]. Finally, an analytic form of the temperature rise of a rectangular-shaped source has been found by solving the boundary value problem via representation of the thermal source by a Fourier series distribution[16, 17]. As this calculation will be used in Chapters 2 and 4, it will be described in detail below.

The solution for a planar, rectangular thermal source of uniform power dissipation on the surface of a finite thickness medium has been detailed previously in the literature[16, 17]. Calculation of the spatial temperature distribution relies on the placement of image sources reflected about the adiabatic surfaces assumed to exist far from the active area at the sidewalls of the semiconductor die; this ensures no heat flow across these interfaces. The infinite set of thermal sources, with periodicity given by the dimensions of the top surface of the semiconductor die, can be expressed in terms of a Fourier series. The temperature rise above ambient \( \Delta T \) at a point \( P(x, y, z) \) relative to the center of the thermal source of width \( D_x \) and length
$D_y$ can be expressed as

$$
\Delta T = \frac{P_D L_z}{\kappa L_x L_y} + \sum_{n,m=0}^{\infty} B_{nm} \cos \frac{2\pi n x}{L_x} \cos \frac{2\pi m y}{L_y} \left( \cosh(C_{nm} z) - \frac{\sinh(C_{nm} z)}{\tanh(C_{nm} L_z)} \right) \tag{1.20}
$$

where the $m = n = 0$ term is excluded. $L_x$, $L_y$ and $L_z$ represent the width, length, and thickness of the semiconductor die, $\kappa$ is the thermal conductivity, and $P_D$ is the total power dissipation in the thermal source. The coefficients $C_{nm}$ are given by

$$
C_{nm} = 2\pi \sqrt{(n/L_x)^2 + (m/L_y)^2} \tag{1.21}
$$

and for $n,m > 0$, the $B_{nm}$ coefficients are

$$
B_{nm} = \frac{4P_D \tanh(C_{nm} L_z)}{D_x D_y C_{nm} \kappa n m \pi^2} \sin \frac{n \pi D_x}{L_x} \sin \frac{m \pi D_y}{L_y} \tag{1.22}
$$

and for $m=0$ or $n=0$,

$$
B_{n0} = \frac{P_D L_x \tanh \left( \frac{2\pi n L_x}{L_z} \right)}{D_x L_y \kappa (n \pi)^2} \sin \frac{n \pi D_x}{L_x} \quad B_{0m} = \frac{P_D L_y \tanh \left( \frac{2\pi m L_y}{L_z} \right)}{D_y L_z \kappa (m \pi)^2} \sin \frac{m \pi D_y}{L_y}
$$

By using Equation 1.20, it is possible to calculate the temperature rise over the active area of the device. It should be noted that the predicted temperature distribution is a function of the $x$ and $y$ coordinates, and therefore Equation 1.20 predicts a non-uniform source temperature, in contrast to the assumptions used to justify the use of a thermal resistance. In order that the result of Equation 1.20 can be compared to the experimental values, which represent an average of the source temperature rise, it is necessary to average the resulting distribution. From the
spatial temperature distribution, one can calculate a mean thermal resistance $R_{TH}$ for the device modeled from

$$R_{TH} = \frac{\Delta T}{P_D}$$  \hspace{1cm} (1.23)

by defining an average device temperature rise $\Delta T$, or an average absolute device temperature $\bar{T}$. Liou et al.[16] suggest determination of the thermal resistance by performing an area average of the device temperature distribution

$$\Delta \bar{T} = \frac{1}{D_x D_y} \int_A \Delta T(x, y) dA$$  \hspace{1cm} (1.24)

where $A$ represents the active area of the device. Equation 1.24 provides a means by which the non-uniform temperature distribution predicted by Equation 1.20 can be condensed into a single number, representing the average temperature rise of the active region of the device for a given power dissipation $P_D$. This averaging scheme will be applied to a calculation of the thermal resistance for an AlGaAs/GaAs heterojunction bipolar transistor in Chapter 2, and the limitations of this approach will be outlined there.

### 1.6 Outline of Thesis

The remainder of this thesis will outline the pulsed measurement procedure developed for characterization of device thermal parameters, and will discuss the results obtained with this method applied to both III-V and Si/SiGe heterojunction bipolar transistors. Chapter 2 will detail the functionality of the pulsed measurement system and the corresponding analysis used to determine the thermal subcircuit parameters for a $3\times3 \, \mu m^2$ AlGaAs/GaAs heterojunction bipolar transistor, and will compare the experimental results with those predicted by the analytic model described in Section 1.5.3. Chapter 3 will describe an investigation of scaling issues associated
with thermal parameters for Si/SiGe heterojunction bipolar transistors in which one lateral dimension varies between measurements; devices in this technology are isolated by deep trenches, which strongly influence the thermal behavior. Chapter 4 examines the layout dependence of the thermal behavior of AlGaAs/GaAs heterojunction bipolar transistors to investigate the optimum emitter geometry for a given constant cross-sectional area to minimize device thermal resistance. Finally, Chapter 5 will briefly summarize the conclusions.
Chapter 2

Experimental Measurement of Thermal Subcircuit Model Parameters

This chapter outlines the experimental equipment and methodology used to determine the thermal response of an AlGaAs/GaAs heterojunction bipolar transistor. The first portion of the chapter details the experimental mechanics, functionality and limitations of the measurement process, and contrasts them with existing techniques detailed in Chapter 1. The second part of the chapter describes results specific to the 3×3 μm² AlGaAs/GaAs heterojunction bipolar transistor for which the technique is demonstrated, detailing the extraction procedures, and comparing the experimental results to analytic and other experimental results performed for the same transistor structure.
Figure 2.1: Representative schematic of pulsed base bias measurement setup for a device maintained in a temperature controlled environment. A fast rise time voltage signal $V_{BE}$ is applied to the base $B$ of the device, while the collector-emitter bias $V_{CE}$ is held constant. The collector current that results, $I(t)$, is monitored via a signal-averaging, digital oscilloscope.

2.1 Experimental Method

2.1.1 Overview of Approach

The premise underlying the pulsed measurement scheme is based on the approach described in [27], where the isothermal measurement technique is demonstrated. In the present work this technique is simplified experimentally, and then extended to allow for measurements of thermal resistance. As shown in Figure 2.1, the device under test is contained in a thermally-controlled environment, and attached to various signal sources and monitoring equipment. $V_{BE}$ represents a pulsed base-
bias source, $I(t)$ current monitoring by a digital oscilloscope, and $V_{CE}$ a constant voltage source applied to the collector-emitter junction of the transistor inside the oven. The measurement procedure consists of switching the device from a state of negligible power dissipation to one of high power dissipation with application of a low-duty-cycle, fast-rise-time, pulsed base bias. The collector current dynamics that result are measured with a 100 MHz signal averaging digital oscilloscope. The collector current transient that results contains information that pertains to device self-heating, owing to the large change in power dissipation that results from the pulsing of the base-emitter bias. For an ideal, very fast base-emitter pulse, there is insufficient time for the temperature to change over the duration of the rise time. Since the electrical device response is much faster, the result is switching of the device from the initial, low-power operating point to the final, high-power-dissipation operating point, without a change in device temperature. Subsequent observation of the changes in the magnitude of collector current allows for a direct observation of the effects of a self-heating as the device temperature approaches its steady state value. Measurement of the isothermal collector current, that current that would result if the device temperature could be maintained constant, allows for the removal of the self-heating effect from the measured device response. The resulting response represents the behavior of the device measured at the temperature of the ambient environment.

Using the collector current dynamical response to a step increase in power dissipation, the thermal time constant can be determined. Experimental transient results will be shown in Section 2.3.1, where extraction of the thermal time constant and the isothermal collector current will be discussed. Finally, the ability of the technique to determine the junction temperature rise above ambient for a given
power dissipation has been developed, and will be explained and demonstrated in Section 2.3.2.

To be able to contrast the resulting isothermal data to that taken in steady-state operation, it is necessary to determine both sets of data with negligible variation of the base-emitter terminal bias between measurements. In order to minimize differences between the steady-state and isothermal measurements, the steady-state curve is determined by increasing the duty cycle of the pulsed base-emitter voltage transient to 100%. This ensures that the parasitic access resistances are identical for both measurements, and that for a given set of applied terminal biases, the voltages that appear at the junctions of the device are identical in the pulsed and steady-state measurements.

2.1.2 Sample Description and Preparation

The device tested is a self-aligned AlGaAs/GaAs heterojunction bipolar with a $3\times3\ \mu m^2$ emitter, and was fabricated at Nortel Technology. The aluminum content in the emitter has been compositionally graded to increase the short circuit current gain; this structure has a typical cutoff frequency of approximately 80 GHz, and current gains in excess of 100. Other details regarding the performance and structure of the transistor can be found in [2].

Of particular importance in the present work is the metallization used to provide an electrical bias to the emitter contact. The device is fabricated with a cylindrical emitter post that is in contact with the top of the emitter; the dimensions of the post are on the order of two microns in diameter and one micron in length. This post is connected in series with a three-micron-wide metal strip measuring nine microns in length and one micron thick, which itself connects to the
center of a coplanar strip transmission lines that allows for electrical contact to be provided to the semiconductor die. These striplines measure fifty microns wide, and four millimeters long. The smaller relative sizes of the base and collector contacts allows for the assumption that the predominant thermal parasitic effect is the heat flow that occurs through the emitter post and towards the metallization of the coplanar strip transmission line in which the transistor is fixtured. The effect of the emitter post and metallization will be considered during comparison of the experimentally-extracted value of thermal resistance to that predicted by simplified heat-flow analysis.

The device under test is contained in a controlled thermal environment, which allows temperature-dependent measurements to be made. The temperature-controlled environment consists of a variac-controlled oven, in which the temperature is constant to within a tenth of a degree Celcius over the duration of a measurement. The temperature of the oven is monitored with a K-type thermocouple, usable over the temperature range of 73 to 1100 K with a 0.1% accuracy and precise to 0.1 K.

The sample is prepared such that the backside of the semiconductor wafer is in direct thermal contact with a heat sink; the heat-sink fins are in thermal equilibrium with the oven environment through free convection. Typical sample fixturing is shown in Figure 2.2. The device to be studied is bonded to the heat sink with a thermally conducting compound with a thermal resistance less than 0.06 K/W, which results in a worst-case error of 0.01% for the device studied in the present work. The electrical connections to the device are provided with wire bonds that extend from a printed circuit board (PCB) that interfaces the pulsed test circuitry with the sample.

As a result of the tendency of the tested devices to oscillate in the GHz
Figure 2.2: Typical sample preparation for device measurements. The semiconductor die that contains the device to be characterized is bonded to the heat sink with a low thermal resistance compound. Electrical connections to the device are provided through wire bonds to a PCB that interfaces the device and the pulsed bias circuitry.
range, the collector-emitter and base-emitter junction are shunted with high-pass microwave terminations. The terminations consist of a series combination of 47 pF capacitors, and 50 Ω resistors located on the surface of the die, away from the active devices. To high-frequency oscillations, these terminations look resistive and serve to terminate the device transmission lines. However, to lower-frequency effects like the self-heating effects of interest to this work, the effect of the terminations is negligible. This termination scheme has been shown to be effective at stabilizing high frequency oscillations in Si/SiGe, AlGaAs/GaAs, and InGaP/GaAs heterojunction bipolar transistors.

2.1.3 Measurement Circuitry

The functional representation of the base circuitry is shown in Figure 2.3. A function-generator-triggered emitter-coupled logic gate is used to generate a pulse with a fast rise time that ranges in magnitude from -230 mV to 0 V. Addition of a precision DC offset voltage $V_{BE}$ supplied by a Hewlett-Packard 4155A semiconductor parameter analyzer (SPA) allows for the reproducible generation of a fast pulsed base bias with a well quantified magnitude that ranges from $V_{BE} - 230$ mV to $V_{BE}$. Once this voltage pulse has been generated, it is applied to the base of the device under test (DUT) with a bipolar push-pull feedback driver. This driver, capable of driving capacitive loads, ensures circuit stability when driving the base of the transistor, and was used because of problems using direct drive by the high slew-rate operational amplifier.

Figure 2.4 depicts the functionality of the circuitry used to apply the collector-emitter bias to the device under test. The precision collector bias, also supplied by the SPA, is applied to the device via a high-slew-rate operational amplifier; this bi-
Figure 2.3: Functional diagram of the pulsed base bias generator. A function-generator-triggered emitter-coupled logic gate is level shifted with a precision offset, and this signal is applied to the base of the device under test (DUT) with a feedback driver. The detailed circuit diagram is shown in Figure A.1.
Figure 2.4: Functional representation of the network used to simultaneously apply a precision collector-emitter voltage and measured the transient collector current during switching. A precision collector-emitter voltage $V_{CE}$ is applied to the device under test with feedback, which allows for simultaneous monitoring of the collector current.
asing scheme allows one to determine the collector current transient in response to application of the pulsed base bias by monitoring the current driven by the feedback driver via the voltage signal developed across the 50 Ω source impedance. By measuring this differential voltage with a signal-averaging 100 MHz digital oscilloscope, one can capture the resulting self-heating dynamics when the device is switched from a negligible power dissipation state to a high power dissipation state.

2.2 Calculation of Thermal Resistance

This section provides a determination of the thermal resistance for the 3x3 \( \mu \text{m}^2 \) AlGaAs/GaAs heterojunction bipolar transistor. The calculation is based upon the method described earlier in Section 1.5.3, and in the literature[16, 17]. The section begins with a calculation of the spatial temperature distribution over the active region of the device, and follows with analysis regarding the estimated average temperature rise based on arguments presented earlier in Section 1.5.3. This calculation will allow for a comparison to the measurement result that is the focus of Section 2.3.2.

To estimate the thermal resistance of the 3x3 \( \mu \text{m}^2 \) AlGaAs/GaAs structure, the analytic model of Section 1.5.3 is applied to determine the spatial temperature distribution across the top surface of the wafer. The total power dissipation is taken to be 5.09 mW; this level of power dissipation is chosen to facilitate comparison with the experimental data to appear in Section 2.3.2. The thermal conductivity of the GaAs substrate is taken to be 46 Wm\(^{-1}\)K\(^{-1}\), and the width and length of the substrate are taken to be 0.9 mm. This last choice is somewhat arbitrary, and does not affect the results significantly. Taking advantage of the symmetry of the device, the temperature rise at 25 points in an octant of the device is calculated by
summing over the first five thousand terms of each index in Equation 1.20. A plot showing temperature contours for one quarter of the device is shown in Figure 2.5. Based on this scheme, the peak temperature rise found at the center of the device is determined to be 20.6 K. In order to assess convergence of the sum in Equation 1.20, the temperature rise was recalculated by summing over the first ten thousand terms of each index. The peak temperature rise only changed by 0.02% between these two calculations, indicating convergence was attained. The temperature contour map shown in Figure 2.5 has contour lines that do not intersect the symmetry planes of the device in a perpendicular fashion. This is due to the finite number of points used to interpolate the temperature map, and could be rectified by doing a more extensive calculation. However, the computational effort required to calculate the data of Figure 2.5 is already quite substantial, and it is not believed that more terms will change the results appreciably.

To facilitate comparison of the theoretical value of the thermal resistance to that measured experimentally, it is necessary to quote an average temperature rise of the device based on the spatially varying temperature distribution of Figure 2.5. In the following, two methods to estimate the average temperature rise of the device are considered. As will be shown for the device considered here, a simple area average of the device temperature produces results in good agreement with more complicated calculations that aim to account for temperature non-uniformity. The first is the method of Liou et al.[16], as shown earlier in Section 1.5.3, which presupposes that the different temperatures weight equally in the overall average temperature of the device. However, one could argue that the mean temperature of the device is that which would result in the same collector current as in the non-uniform-temperature case. The second method is based upon this approach, and
Figure 2.5: A contour map of the temperature rise of the device for a power dissipation of 5.09 mW, the level of power dissipation at which the experimental data of Section 2.3.2 is measured. The peak temperature rise is 20.6 K at the center of the device.
uses the simple temperature dependent transistor model of Section 1.2.

The calculation for the second averaging scheme proceeds as follows. First, using the data shown in Figure 2.5, the temperature is interpolated at one hundred equally spaced points across the top surface of the device. To find the single, weighted-average temperature that would result in the same current flow as the non-uniform distribution of Figure 2.5, one considers the device collector current $I_C$ to be a function of the mean temperature $\overline{T}$. The current flow is calculated at each of the hundred elements based on Equation 1.1 given the area of each of the individual elements, and the temperature at that particular location; mathematically, the current $I_i$ at each point $i$, of cross-sectional area $A_i$, is a function of the localized temperature $T_i$. Since the total device current must be the sum of the current supported by the individual elements,

$$I_C(\overline{T}) = \sum_i I(T_i)$$  \hspace{1cm} (2.1)

and by expressing the current in terms of the product of a current density and the area of the element $A_i$ or device emitter $A_E$,

$$J_C(\overline{T}) = \sum_i \frac{A_i}{A_E} J(T_i)$$  \hspace{1cm} (2.2)

By using the calculated current density as a function of temperature, shown previously in Figure 1.2 to represent the function $J(T_i)$, together with Equation 2.2, the total current density of the device $J_C(\overline{T})$ can be determined. From this, interpolation of the current density data in Figure 1.2 allows one to determine the mean device temperature $\overline{T}$. This is the uniform device temperature that would result in the same current flow as the non-uniform temperature shown in Figure 2.5.

Calculation using the method described above results in an estimated $\overline{T}$ of 316.8 K. The unweighted average of Equation 1.24 results in an estimated mean
temperature of 316.7 K and it is seen to differ only slightly from the weighted average predicted by Equation 2.2. While the difference between the estimated mean temperatures for the two techniques increases for larger power dissipations, it is nevertheless a small effect for the $3 \times 3 \mu m^2$ structure studied here. For example, for a three-fold increase in device power dissipation, to approximately 15 mW, the uniform power dissipation model with a weighted average calculation of temperature rise only results in a 1.2% increase in thermal resistance. Thus one concludes, in accordance with Reference [16], that an area average is sufficient to estimate a device thermal resistance; this thermal resistance, obtained by normalizing the temperature rise of the device by the 5.09 mW power dissipation, results in an estimated thermal resistance of 3.3 K/mW.

2.3 Results and Discussion

2.3.1 Transient Results and Isothermal Characterization

Figure 2.6(a) depicts the experimentally measured pulsed base-emitter voltage and (b) shows the resulting transient collector current for a collector-emitter voltage of 3.63 V at room temperature. Figure 2.6(b) shows clearly the effect of device self-heating. Without the self-heating effect, one would expect that once the base-emitter voltage has reached its nominal level, the collector current would follow on a time scale fast compared to the 250 ns rise time of the pulse. However, examination of the collector current for times greater than 1.2 $\mu$s clearly indicates a collector current that rises as a function of time, despite the fact that all the terminal biases are held constant. After the device is switched to its high power dissipation state at approximately 0.7 $\mu$s, the power dissipated in the reversed biased base-collector
Figure 2.6: Experimentally measured base-emitter voltage and collector current, measured at a collector-emitter voltage of 3.63 V and a heat-sink temperature of 316.1 K. Effects of self-heating can be seen for times after 1.2 \( \mu \text{s} \), where the base-emitter drive is constant but the collector current continues to rise.
space-charge-region results in a heat flow to the backside of the semiconductor wafer, where the temperature is maintained constant by the room temperature surroundings. As detailed previously, the effect of this heat flow is an increased junction temperature, which for device operation at constant base-emitter voltage, results in an increased collector current through electrothermal feedback.

The thermal time constant can be extracted from the transient current data, but two conditions must be satisfied for accurate results. The first condition relates to the assumption that the device collector current responds with the same time constant as that which predicts the evolution of the device temperature. By application of the lumped thermal model, the device response to a step power increase is such that the temperature will relax with a time constant $\tau_{TH}$. As the levels of power dissipations at which this extraction performed are small, the corresponding temperature rise of the device is also usually small and is typically less than 10 K. For small temperature rises such as this, one can approximate the collector current as a linear function of temperature, and therefore the relaxation of the collector current will take place with the same thermal time constant as that describing the temperature dynamics themselves. Figure 2.7 shows the accuracy of this approximation and is based on model calculations for the response to a step increase in power at time $t = 0$. The solid line shows the calculated transient response of the device collector current, via Equation 1.1, based on assuming the base-emitter junction temperature relaxes with a 1.3 $\mu$s time constant, as was found with preliminary measurements. The symbols show the current calculated by knowing the initial and steady-state device temperatures, with the response for times between these two conditions calculated by assuming that the collector current relaxes with the time constant of 1.3 $\mu$s. From the good agreement shown in Figure 2.7, it is concluded
Figure 2.7: Calculated collector current transient in response to a device operating temperature that is relaxing to its equilibrium value with a thermal time constant \( \tau_{TH} \) (solid line). The fit to the curve (symbols) is that predicted by assuming the current itself relaxes with time constant \( \tau_{TH} \).
that extraction of the thermal time constant via regression of the collector current transient is accurate for the case of small temperature changes.

The second condition for valid extraction of the time constant relates to the degree to which the experimental power dissipation resembles a step function. For the data shown previously in Figure 2.6, the magnitude of the self-heating effect is large, and as such the power dissipation is changing significantly over the duration of the transient. In this case, extraction of the thermal time constant by exponential fitting of the collector current will yield an artificially high estimate for $\tau_{TH}$ as a result of the slower response of the electrothermal feedback effect. Therefore, accurate extraction of $\tau_{TH}$ must be done when the degree of self-heating is small. Such a case is shown in Figure 2.8, where only a small amount of self-heating is exhibited. Here the power dissipation can be approximated as constant over the duration of the pulse, and accurate extraction of the time constant can be expected.

Having established the conditions required for accurate $\tau_{TH}$ extraction, the data of Figure 2.8 will now be analyzed. An exponential fit to the collector current transient is performed over the range of data for which the base-emitter voltage is essentially constant; by 1.2 $\mu$s, the base-emitter voltage has approached to within 0.5 mV of its steady-state limit, and as such the fit is performed from 1.2 to 6.5 $\mu$s. The fit is of the form

$$I(t) = I_{SS} + (I_O - I_{SS}) \exp(-t/\tau_{TH})$$

(2.3)

where $I(t)$ represents the instantaneous value of the collector current, and $I_{SS}$ and $I_O$ are the steady state and initial collector current for the transient. The least-squares fit to the experimental data, also shown in Figure 2.8, is excellent and yields an estimate of the thermal time constant. The error in the extracted thermal time constant is estimated from results of fits to similar data taken with small changes
Figure 2.8: Experimental collector current transient (solid line) and corresponding exponential regression (symbols), performed at a base-emitter bias of 1.391 V, a collector-emitter bias of 0.87 V at 316.1 K. The thermal time constant, determined from such a fit, is $0.7 \pm 0.1 \mu s$. 
in the collector-emitter bias. For the 3×3 AlGaAs/GaAs heterojunction bipolar transistor studied here, the thermal time constant was determined to be $0.7 \pm 0.1 \, \mu s$.

As mentioned above, if a similar fit and extraction were performed at higher power dissipations and for more pronounced self-heating, the quality of fit will decrease, and the extracted thermal time constant is expected to increase due to electrothermal feedback. This was verified experimentally by extraction at the same base-emitter voltage, and a collector-emitter bias of 4.76 V; the resulting time constant is $1.90 \, \mu s$. However, the discrepancy in comparison to the previous result does not imply that the device is poorly described by a lumped thermal system; instead, one must realize that the device power dissipation is a function of time, and clearly this will affect the mathematical form of the resulting transient. For the scenario in which the coefficient that relates the increase of device temperature to increases in power dissipation itself increases with temperature, it is possible that the simple thermal subcircuit can predict the onset of thermal runaway, an unbounded, transient response without a discernible time constant.

Having discussed the accurate extraction of the thermal time constant, in the following the determination of the isothermal device response is considered. Ideally, this is accomplished by switching to the high-power-dissipation state without appreciable rise of the junction temperature over the switching process. In this ideal case, the rise time of the base-emitter voltage would be sufficiently short that the isothermal collector current could be sampled directly from the digital oscilloscope. However, in the present experimental setup, this is not the case: typical rise times of the base-emitter voltage are approximately $0.25 \, \mu s$, while typical measured time constants range from 0.4 to 0.7 $\mu s$. Despite the fact that these two time scales are
comparable, estimates of isothermal device response can be made quite accurately, as described below.

Figure 2.9 illustrates the issues involved in the extraction of the isothermal collector current. Figure 2.9(a) shows the exponential fit to the experimental collector current and (b) shows the base-emitter voltage transient. These are the same data shown earlier in Figure 2.6; here the leading edge of both curves are shown by the solid lines, on expanded time scales. It is obvious from Figure 2.9(b) that while the 10 to 90% rise time is on the order of 0.25 $\mu$s, even after the 90% level is reached, the variation of base-emitter bias is still very significant. In fact, in the absence of any self-heating, the base-emitter bias has to stabilize to within approximately 1mV before the collector current will be constant to within a few percent. Therefore, the portion of the collector current transient where the base-emitter bias is effectively constant begins some time after the transient appears to be completed.

In order to perform a first-order correction for the self-heating that occurs over the rise time of the pulse, the time at which the device is switched from its low to high power dissipation state is estimated to be the 50% power point. By using the base-emitter voltage data and not the collector current data for extraction of the 50% power point, an extraction methodology is proposed that is independent of the degree of self-heating in the device, as self-heating only changes the response of the collector current data. One can see that in the limit of very short rise times, the 50% power point accurately captures the time at which the pulse is effectively applied to the device under test. In the limit of longer pulse rise times, the 50% power point allows for a first-order correction to be performed to the experiential data; this is understood by realizing that the amount of self-heating that has taken place prior to the 50% power point of the isothermal device response balances the
Figure 2.9: (a) Collector current transient (solid line) and regression (dashed dotted line) and (b) corresponding base-emitter voltage transient, measured at a collector-emitter bias of 3.63 V, and a temperature of 316.1 K. The nominal value of the base-emitter bias is 1.379 V. The three dotted vertical lines appearing left to right correspond to the minimum, nominal, and maximum estimates of the isothermal collector current, respectively.
amount of self-heating that has yet to take place by the time the pulse reaches its nominal magnitude at the device under test. The location of the 50% power point from the base-emitter voltage data relies on assuming a device ideality factor of unity, such that the time at which the base-emitter voltage pulse is 18 mV lower than its asymptotic value. In the data of Figure 2.9, this occurs at approximately 0.69 µs; this position is indicated by the position of the central dotted lines in both panels of Figure 2.9. Extraction of the isothermal device response proceeds by determining the location of times for which the transient collector current is entirely determined by self-heating under constant bias conditions; an exponential fit to the collector current in this region is performed, and is shown by the dash-dotted line of Figure 2.9(a). Based on this fit, estimates can be made of the current that would have been observed at earlier times had the base-emitter shown a perfect step behavior. Because of the finite rise time of the base bias signal, there is some uncertainty about the appropriate time to which this extrapolation should be performed. Therefore in the present work the uncertainty in this time has been assessed as approximately ± 25% of the base-emitter bias rise time. With this, three estimates of the isothermal collector current are made and used to establish the error bars of the measured results; the three dotted vertical lines shown in Figure 2.9 show the times corresponding to these minimum, nominal, and maximum estimates. For the data shown, the isothermal collector current determined in this way was determined to be 2.4 ± 0.1 mA.

It is apparent, therefore, that the error bounds are fairly small in the present situation. However, the larger the self-heating effect, the more pronounced the slope of the leading edge of the collector current transient, and therefore the larger the error bound on the isothermal collector current.
Using the approach outlined above, isothermal device characterization can be performed; the Gummel plot of collector current is considered first. This plot, which describes the variation of collector current as a function of base-emitter voltage, is one of the prime sources for determining device model parameters. Because some parameter extraction takes place at large collector currents where self-heating can be significant, it is important to remove the thermal effects from the inherent electrical effects that the model is attempting to account for. Repeated application of the technique described above is used to measure the variation of isothermal collector current as a function of the applied base-emitter voltage; the symbols in Figure 2.10 show the result of such a measurement, performed at a base-collector voltage of zero volts and at room temperature. For comparison, the steady-state data are shown by the solid line. As expected, the discrepancy between the isothermal data and the data acquired under steady-state conditions increases for larger power dissipations, as the amount of self-heating increases. For collector current densities on the order of $6 \times 10^4$ A/cm$^2$, deviations between the isothermal data and that obtained under steady state measurement conditions exceed 20%. For extraction of parameters from measurements at high power dissipation, such a discrepancy presents a substantial error that could compromise the predictive capabilities of device models. Clearly the explicit assumption that models extracted from such measurements are performed at a single device operating temperature[24, 31] is violated. Such a scenario arises when extracting the Gummel-Poon forward knee current $I_K$, which models the decrease of the forward gain $\beta_F$ due to the onset of high-level injection. Extraction of this model parameter involves finding the intersection of the low and high current asymptotes of the collector current density, plotted as a function of the base-emitter drive[31]. If this extraction would be performed on the steady state data as shown in Figure 2.10,
Figure 2.10: Steady state (solid line) and isothermal (symbols) Gummel plot of collector current density, for $V_{BC} = 0$ V at room temperature. For high current densities nearing $6 \cdot 10^4$ A/cm$^2$, greater than 20% error is incurred by assuming that the steady-state measurement is performed at a single device operating temperature.
it can be seen that the resultant value would be overestimated due to the increased collector current that results from an elevated device operating temperature.

In principle, it is possible to measure the isothermal base current in the same manner in which the isothermal collector current was determined. This would allow one to construct a complete isothermal Gummel plot, from which most of the parameters that would compose an isothermal device model could be determined. Owing to the large gains of the devices studied, this would necessitate precision electronics for time-resolved measurements of currents on the order of tens of microamps, and as such, has not been pursued in the present work.

Another form of isothermal device characterization is determination of the collector output characteristic. The solid lines in Figure 2.11 show the steady state variations of device collector current measured at room temperature, for base-emitter voltages of 1.370 and 1.390 V. The steady-state measurements show considerable output conductance; this is conventionally attributed solely to the self-heating effect as the highly-doped base in heterojunction bipolar transistors is expected to suppress base-width modulation[23]. For collector-emitter biases in excess of three volts and for the higher base-emitter drive levels, the steady-state collector characteristics show substantial curvature. Steady-state measurements of the base current at these drive levels have shown a decrease of the base current as the device comes out of saturation, followed by a subsequent increase of base current for increasing collector-emitter voltages. This has shown that this effect is solely due to self-heating, and not due to the onset of avalanche breakdown in the reverse biased base-collector junction. A decrease and subsequent reversal of base current for increases in collector-emitter voltage would have been indicative of impact ionized holes being swept back into the p-type base of the device[32]. For comparison, the
Figure 2.11: Steady state (solid line) and isothermal (symbols) collector characteristics for base-emitter voltages of 1.370 and 1.390 V, measured at a temperature of 304 K.
isothermal variations of collector current as a function of the collector-emitter voltage are also shown in Figure 2.11. The curvature of the steady-state measurements is no longer observable in the isothermal data, which also suggests the curvature of the steady-state data is the result of the self-heating process. This conclusion is supported by the fact that the curvature of the steady-state data increases with the level of base-emitter drive. With more base-emitter drive, the level of device power dissipation increases, and so does the amount of self-heating. It should be noted that the sizes of the error bars increase as the collector-emitter voltage increases, as noted previously.

The non-zero output conductance of the isothermal data would tend to imply that, apart from self-heating, there are other physical effects apart taking place that contribute to the output conductance of the device. However, it should be noted that the shape of the pulse applied to the base-emitter junction for small collector-emitter voltages is not ideal. As such, the base-emitter voltage at which the isothermal measurement is performed may be varying slightly. In addition, there is some device heating during the rise time of the base-emitter bias, which will result in a small but systematic overestimate of the isothermal currents. Considering all of these effects it would seem that the experimental apparatus would have to be improved before precise conclusions about the output conductance could be made.

2.3.2 Thermal Resistance

One of the main results in the present work is a fully experimental method to determine the base-emitter junction temperature at a particular power dissipation. This is achieved by extending the technique described in the previous section. The idea is illustrated in Figure 2.12, which shows two curves. The solid line shows the
Figure 2.12: Steady state (solid line) and isothermal (symbols) collector characteristic. The steady-state data was determined at a heat sink temperature of 304.9 K, while the isothermal data was measured at a temperature of 316.1 K. For both curves, the base-emitter bias was held at 1.379 V. The intersection of the two curves occurs at a collector-emitter voltage of 2.26 V and a collector current of 2.26 mA.
steady-state collector current data, measured at an ambient temperature of 304.9 K. The second curve, shown by the symbols and the error bars, shows the isothermal results taken at an elevated temperature of 316.1 K. While the device temperature along the steady-state curve is increasing with power dissipation, the temperature for the isothermal curve is, by definition, fixed. Because both curves were taken at the same base-emitter bias, the device temperature must be the same at the intersection point of the steady-state and isothermal data, where the steady-state power dissipation is known. This allows for the determination of the temperature in steady state at the intersection point seen in Figure 2.12; the temperature at this point on the curve is 316.1 K, and therefore the temperature rise is 316.1 - 304.9 K, or 11.2 K.

The preceding qualitative discussion can be justified on more rigorous grounds, by considering that the collector current $I_C$ in the forward active region is a function $f$ of the terminal bias conditions and the temperature of the base-emitter junction, assumed to be at a single uniform temperature. Then

$$I_C = f(V_{BE}, V_{CE}, T)$$

At the intersection point of the steady state and the isothermal measurements, the collector current, the collector-emitter voltage, and the base-emitter voltage are identical. From Equation 2.4, the only way this is possible is if the junction temperature of the steady-state measurement at the intersection point is the same as the temperature at which the isothermal measurement was performed. Thus, in Figure 2.12, the temperature rise of the base-emitter junction by 11.2 K above ambient is known to be for a collector-emitter voltage of 2.26 V, and a collector current of 2.26 mA. Furthermore, the power dissipation at this point is known, and is given by the product of the collector-emitter voltage and the collector current to be 5.09
mW. Therefore, we can construct the plot shown in Figure 2.13, which shows the temperature rise as a function of device power dissipation. From the ratio of this temperature rise to the power dissipation, the thermal resistance is determined to be $2.2 \pm 0.1 \text{ K/mW}$. The error bounds on the quoted thermal resistance value are based on the uncertainty in the intersection point due to the finite error bars on the isothermal data. Figure 2.13 summarizes the results of the thermal resistance measurement, showing the result measured above with the open circle. Also shown is a straight line that represents an independent measure of the device thermal resistance performed by co-workers at Nortel on similar devices[33]. It is seen that the two measurements do not agree within the experimental error of the pulsed measurement scheme. It is possible that the discrepancies between these measurement results could be due to uncertainties in the Nortel measurement, which are not quoted, or to the potentially different operating points at which the measurements were performed.

Having determined the thermal resistance, the thermal capacitance can now be determined, as it is given by the ratio of the thermal time constant to the thermal resistance,

$$C_{TH} = \frac{\tau_{TH}}{R_{TH}}$$  \hspace{1cm} (2.5)

Normalizing the thermal time constant, stated previously as $0.7 \pm 0.1 \mu\text{s}$, by the thermal resistance of $2.2 \pm 0.1 \text{ K/mW}$, the thermal capacitance is found to be $0.32 \pm 0.06 \text{ nJ/K}$.

As was shown earlier in Section 2.2, a theoretical consideration for the structure considered in the present work was performed. A power dissipation of $5.09 \text{ mW}$ resulted in an estimated thermal resistance of $3.3 \text{ K/mW}$. This theoretical value does not agree within error with the experimental value of $2.2 \pm 0.1 \text{ K/mW}$. 

68
Figure 2.13: Plot of the junction temperature rise above ambient as a function of the device power dissipation as determined with the pulsed measurement (circle), and an independent measure of the thermal resistance (solid line).
As was shown, the spatially non-uniform temperature distribution does not change the estimate of the thermal resistance appreciably; a weighted average of the source temperature resulted in only a 0.2% change of the estimated thermal resistance value.

In order to account for the discrepancy in the measured and modeled results, one must consider the effect of the emitter post metallization as described earlier in Section 2.1.2. Assuming negligible heat flow through the oxide which surrounds the emitter post allows one to perform a simple one dimensional heat flow analysis of the cylindrical emitter post and series emitter contact metallization pattern. Defining the parasitic thermal resistance $R_P$ as the series contributions of the emitter post and the contact metallizations, one can write

$$R_P = R_{EP} + R_{EC}$$  \hspace{1cm} (2.6)

where $R_{EP}$ and $R_{EC}$ are the contributions of the post and contact metallization, respectively. As each individual component can be quantified according to the one-dimensional thermal resistance result of Equation 1.4.1, a first order estimation of the effective parallel heat flow path is a simple function of the post and contact metallization geometry, and the thermal conductivity of the material with which they are composed. Taking the thermal conductivity of gold to be $310 \, \text{Wm}^{-1}\text{K}^{-1}$ \cite{15}, and the post dimensions given previously in Section 2.1.2, the effective parasitic thermal resistance is approximately $10.7 \, \text{K/mW}$. This calculation assumes that once past the emitter contact metallization, the coplanar stripline is maintained at 300 K. Due to the massive size, higher thermal conductivity compared to GaAs, and the fact that heat flow is no longer one dimensional, the thermal resistance of the coplanar stripline should be small compared to that of the parasitic resistance already calculated. Once the heat flow has arrived at the top surface of the wafer where the
coplanar stripline connect to the device, the heat can be removed either through surface convection from the large area coplanar lines, or through conduction back into the substrate. The effective thermal resistance of both of these subsequent processes is believed to be small due to the large area conductor involved.

Combining the calculated parasitic resistance in parallel with the analytical result which quantifies heat flow to the backside of the semiconductor wafer yields an effective thermal resistance of the transistor structure of 2.5 K/mW, which is in much better agreement with the experimental value of 2.2 K/mW. In order to obtain even better agreement between experiment and theory, one would need to account for the additional parasitic effects introduced by the base and collector metallization as well, which would necessitate a detailed knowledge of the layout of the structure studied.

Using the thermal resistance and capacitance, it is possible to generate a thermal subcircuit of the form shown earlier in Figure 1.3. Such a thermal subcircuit can be incorporated into integrated circuit simulators, allowing for the inclusion of self-heating in device models. The other constituent model parameters, which must be acquired at a single device temperature, can be determined with one of two methods. For parameters that can be extracted with low power dissipation measurements, conventional steady-state measurements can be performed. However, for parameters that require non-negligible power dissipations, the method of choice is that illustrated for generating isothermal Gummel plots and collector characteristics. Overall, the results of the pulsed measurement together with steady-state measurements can be used for the full construction of a device model that includes heating effects, and does not require assumptions regarding device behavior.
2.3.3 Experimental Limitations

Experimental pulsed base bias data for a $3 \times 3 \mu m^2$ AlGaAs/GaAs heterojunction bipolar transistor with $V_{CE}$=1.97 V is shown in Figure 2.14 where (a) shows the whole base voltage pulse applied to the device under study, while (b) and (c) depict the finite rise time and the variations of the base-emitter voltage pulse, respectively. The low collector-emitter bias emphasizes the experimental limitations, and was chosen to illustrate these limitations most clearly; however, this low-bias regime is not used in measurements presented in subsequent chapters. The non-idealities in the experimental pulse applied to the device under test are the principle sources of experimental limitations. It is seen from Figure 2.14(b) that the rise time of the signal applied to the heterojunction bipolar transistor is on the order of 250 ns. Because the rise time of this signal is on the order of the thermal time constant, it is necessary to understand the effects of extracting the isothermal collector current at differing locations along this curve. As suggested previously, extraction at the 50\% power point allows for a first-order correction to be performed to the extraction procedure. However, one could envision extracting from the point at which the base-emitter bias reaches its nominal level at the device under test. The following discussion quantifies the error incurred by incorporating such an extrapolation procedure over that already described.

To quantify the effects of the finite rise time, consider the source of the base-emitter voltage to be the slew-rate limited output from an operational amplifier, switching from the low power dissipation state to the high power dissipation state. The transient power dissipation $P_D(t)$ of the device is approximated as

$$P_D(t) = I_C(t)V_{CE}$$  \hspace{1cm} (2.7)
Figure 2.14: Experimental pulsed base voltage data for a 3×3 μm² AlGaAs/GaAs heterojunction bipolar transistor with $V_{CE}=1.97$ V; (a) shows the entire transient data, (b) shows the leading edge of the voltage pulse, and (c) shows the variations in amplitude over the duration of the pulse.
where the collector current transient \( I_C(t) \) itself is a function of the base-emitter voltage transient, and \( V_{CE} \) represents the applied collector-emitter voltage. Owing to the large gain of the devices studied, the power dissipation arising from the base current has been neglected in Equation 2.7.

By considering the thermal subcircuit as depicted previously in Figure 1.3(a), one can write a Kirchhoff current law relation at the upper node of the parallel \( R_{TH}C_{TH} \) network. Realizing that the current flow into this node is the device power dissipation \( P_D(t) \), and the nodal voltage at the top of the network is the temperature rise \( \Delta T \), the transient power dissipation can be expressed as

\[
P_D(t) = \frac{\Delta T}{R_{TH}} + C_{TH} \frac{d(\Delta T)}{dt} \tag{2.8}
\]

from which one can write the differential equation

\[
R_{TH}C_{TH} \frac{d(\Delta T)}{dt} + \Delta T - P_D(t)R_{TH} = 0 \tag{2.9}
\]

which, by re-writing the first term, can be cast into a function of the dimensionless time parameter \( t/\tau_{TH} \)

\[
\frac{d(\Delta T)}{d(t/\tau_{TH})} + \Delta T - P_D(t/\tau_{TH})R_{TH} = 0 \tag{2.10}
\]

where the thermal time constant has replaced the \( R_{TH}C_{TH} \) product. The mathematical form of the base-emitter transient is considered to be of the form

\[
V_{BE}(t) = \frac{V_H + V_L}{2} + \frac{V_H - V_L}{2} \text{erf} \left( \frac{1.8124t}{\tau_V} \right) \tag{2.11}
\]

where \( V_H \) and \( V_L \) are the high and low base-emitter drives, and \( \tau_V \) is the 10-90% rise time of the base-emitter voltage transient. The normalized power dissipation \( P_{DN} \), the instantaneous power dissipation normalized to the maximum power dissipation, as determined from Equation 1.1 can be expressed as

\[
P_{DN} = \frac{P_D}{P_{D\text{max}}} = \exp \left( \frac{V_{BE}(t) - V_H}{V_T} \right) \tag{2.12}
\]
where $V_T$ is the thermal voltage. By solving Equation 2.12 for the times corresponding to normalized power dissipations of 0.1 and 0.9, one can describe the 10-90% rise time of the power dissipation $\tau_P$ as

$$
\tau_P = \frac{\tau_V}{1.8124} \left[ \text{erf}^{-1} \left( \frac{2V_T}{V_H - V_L} \ln(0.9) + 1 \right) - \text{erf}^{-1} \left( \frac{2V_T}{V_H - V_L} \ln(0.1) + 1 \right) \right] \tag{2.13}
$$

where it has been assumed that the power dissipation of the low state is zero. For a room-temperature value of the thermal voltage and for the 230 mV pulse used in the experiment, this approximation results in errors that are a fraction of a percent, and are therefore negligible. From Equation 2.13, the rise time of the power pulse increases linearly with the base-emitter pulse, and is a slowly varying function of the pulse magnitude. Figure 2.15 depicts the modeled base-emitter voltage transient and the corresponding rise in power dissipation of the device studied. Owing to the exponential dependence of collector-current on base-emitter voltage, the effective speed of the power dissipation pulse increases, such that a 250 ns rise time base-emitter voltage transient results in a 157 ns rise time power pulse, as predicted by Equation 2.13.

Given the form of $V_{BE}$ expressed in Equation 2.11, Equation 2.10 is solved for the percent rise in temperature that occurs during the rise time of the pulse. Here, the pulse is considered to have reached its steady-state value by the time it settles to within 0.5 mV of its nominal, high value. Figure 2.16 shows the results of the numerical integration for ratios of $\tau_V/\tau_{TH}$ ranging from 0.01 to 100. Relatively small increases in temperature of the device occur for base-emitter rise times on the order of half the thermal time constant. As has already been shown, this can be explained by considering that the effective speed of the device power dissipation, which causes the temperature rise of the device, is faster than the speed of the base-emitter pulse itself. For a $3 \times 3 \mu m^2$ AlGaAs/GaAs heterojunction bipolar transistor,
Figure 2.15: Part (a) shows the modeled base-emitter voltage transient applied to the device, and (b) shows the resulting power dissipation, normalized to the steady state value.
Figure 2.16: Percent rise of the device temperature to its steady state value as a function of the ratio of the base-emitter pulsed base bias rise time to the device thermal time constant.
a $\tau_V/\tau_{TH}$ ratio on the order of 0.31, results in a predicted temperature rise to 15.6% of the steady-state temperature by the time the base-emitter voltage has settled to within 0.5 mV of its nominal high value. However, from Figure 2.16, it is seen that to reduce this effect to five percent would only necessitate a three-fold increase in the speed of the base-emitter pulse. Because of the magnitude of this error for the present rise time of the base-emitter voltage pulse, the extrapolation is performed at the 50% power point for the experimental data shown in this work.

The other principal experimental limitation, shown in Figure 2.14(c), is that the base-emitter voltage applied to the device varies over the duration of the voltage pulse. Ideally, after some small rise time associated with the switching of the base-emitter voltage from its low to high state, the base-emitter voltage would reach its nominal high state and remain at that state for the duration of the pulse. However, Figure 2.14(c) shows a substantially slower settling time on the order of a microsecond. This settling time, which itself exceeds the thermal time constant of the device studied in this chapter, will have the overall effect of delaying the time at which the device temperature begins to rise due to self-heating. The trend of this limitation opposes that of the finite rise time; while the finite rise time causes a temperature rise prior to application of the nominal pulse, the slow asymptotic rise seen in Figure 2.14(c) tends to delay the onset of substantial device self-heating.

The asymptotic approach of the signal to the nominal level for the first 1.5 $\mu$s is believed to be the result of the decreased driver speed when operation changes from slew-rate limited to linear amplification. The effect of the systematic variations seen on the leading edge of the base-emitter voltage transient are hard to quantify, owing to the complicated nature of the device electrothermal feedback. However, for the devices studied, it is possible to avoid the slow rise by testing at high collector-
emitter voltages. The shape of the base-emitter voltage pulse as a function of the collector-emitter voltage is shown in Figure 2.17 as applied to a $0.5 \times 10 \mu m^2$ Si/SiGe heterojunction bipolar transistor. From the data it is apparent that as the collector-emitter voltage increases, the systematic deviation of the pulse from an ideal pulse lessens. For applied collector-emitter voltages which result in moderate levels of impact ionization, the transient reaches its asymptotic value much more quickly than for lower applied collector-emitter voltages. This is presumably due to the asymmetric ability of the feedback driver to switch when sourcing or sinking current and as such, the long settling time of the base-emitter pulse for low collector-emitter voltages is strictly an experimental limitation. While the results of Figure 2.17 are specific to a $0.5 \times 10 \mu m^2$ Si/SiGe heterojunction bipolar transistor, the trend of the data is generally observed in all the devices studied. The error incurred during the isothermal measurement is reduced by extracting the relevant information at high collector-emitter voltages; the measurements of thermal resistance to be presented in subsequent chapters have been measured under these conditions.

2.4 Conclusions

An experimental technique, based on pulsed base bias measurements, has been presented. This experimental technique allows one to ascertain the required model parameters for inclusion of self-heating into integrated circuit simulator environments. The thermal resistance and time constant of a $3 \times 3 \mu m^2$ AlGaAs/GaAs heterojunction bipolar transistor have been determined as $2.2 \pm 0.1$ K/mW and $0.7 \pm 0.1 \mu s$, respectively. While the uniform power dissipation model of Section 2.2 is unable to predict the experimentally observed value, inclusion of parasitic heat flow paths such as the emitter contact metallization and emitter post result in much
Figure 2.17: Base-emitter voltage transients for a 0.5×10 μm² Si/SiGe heterojunction bipolar transistor, for collector-emitter voltages of 1.5 V(solid line), 3.5 V(dashed dotted line) and 4.0 V(dotted line). The slow, asymptotic approach over 1.25 μs seen in the 1.5 V drive case no longer present at 4.0 V case. To emphasize the trends of the three differing cases, only every 25th data point has been plotted.
closer agreement to be achieved between theory and measurement. The uniform power dissipation model suggests substantial temperature gradients exist across the lateral dimensions of the transistor emitter structure; however, for the small transistor structure studied, the estimated degree of temperature non-uniformity does not affect the overall thermal resistance of the structure.
Chapter 3

Thermal Parameter Scaling in Si/SiGe Heterojunction Bipolar Transistors

This chapter presents both a theoretical and experimental investigation of thermal parameter scaling for deep trench Si/SiGe heterojunction bipolar transistors which are described in Section 3.1. The theoretical predictions of thermal resistance, performed by D.J. Walkey et al. of the Department of Electronics at Carleton University, will be briefly detailed in Section 3.2. The results of the experimental study, based on the scheme presented earlier in Chapter 2, will be shown in Section 3.3.

3.1 Description of Devices Tested

The devices studied in the present work are Si/SiGe heterojunction bipolar transistors (HBTs). Unlike in the AlGaAs/GaAs HBTs that will be described in the
following chapter, isolation between adjacent devices is accomplished by fabrication of insulator-filled deep trenches that surround the active area of the device. The device geometry is shown in Figure 3.1(a), where (b) shows a vertical cross-section of the device, where the cross section is taken at plane A-A' shown in part (a) of the figure. A trench surrounds the emitter area, and extends vertically for a depth of several microns. The trench is approximately 1 μm from the end of the emitters in the long direction, and the emitter is asymmetrically located in the other direction. A significant feature of the trenches, in addition to their electrical isolation, is the low thermal conductivity of the trench; this is due to their composition, which is usually silicon dioxide with a polysilicon interior. As will be discussed later, the presence of the trench structures substantially alters the thermal behavior of the device in comparison to non-trench devices. Emitters of constant 0.5 μm width, and lengths of 1.3, 2.5, 5.0 and 20.0 μm were studied; the selection of a range of emitter lengths that vary over more than an order of magnitude was made to allow testing of the scaling of the thermal resistance and thermal capacitances. Device layouts were performed at Nortel Technology, and devices fabricated at IBM. Additional details regarding device structure, performance, and layout can be found in the literature[34, 35].

3.2 Theoretical Considerations

3.2.1 Thermal Resistance Model

While there are a number of models of thermal resistance that have been developed, only one attempt has been made to incorporate the effects of trench isolation. As this work has not yet been published, and will be used as one basis for compar-
Figure 3.1: Schematic representation of Si/SiGe heterojunction bipolar transistors; (a) shows the entire device, where the active emitter region is enclosed within deep trench isolation, and (b) shows a cross-section of the device at cutting plane A-A', depicted in part (a).
ison with the experimental results presented later, this section outlines the main features of the model. This model was developed by D.J. Walkey et al. of the Electronics Department at Carleton University[36]. Calculations are based on a Green's function approach to estimating the thermal resistance. As a result of the oxide trenches, conventional Green's functions cannot be used because of the non-uniform vertical structure of the modeled region. This difficulty is dealt with by assuming the presence of an isothermal plane at the bottom of the trench, depicted as plane iso in Figure 3.1(b). This assumption was justified by numerical simulations, and allows one to estimate the total thermal resistance of the deep trench structure as the sum of the thermal resistances for two individual regions.

Calculation for the first region, that portion of the device encased laterally within the deep oxide trenches, is modeled with adiabatic walls at the locations of the trenches; the vertical extent of the region is that of the trench itself. The top surface of the region is also assumed to be adiabatic. Integration of the point-source Green's function over the heat-generating region, in combination with the use of the method of images to satisfy the combination of adiabatic and isothermal boundary conditions at different locations, leads to a calculation of the thermal resistance. The second region is composed of an isothermal plane at the bottom of the trenches; this buried plane is modeled as a thermal source of rectangular dimensions equal to that of the lateral extent of the trench isolated-area. The thermal resistance of this second region is calculated by a similar Green's function integration.

3.3 Experimental Results and Discussion

In an identical fashion to that shown in Chapter 2, experimental studies of thermal parameters were performed for the devices described in Section 3.1. Results
Table 3.1: Measured thermal time constant and thermal resistance, and theoretical thermal resistance values for deep trench, 0.5 \( \mu m \) wide Si/SiGe heterojunction bipolar transistors.

<table>
<thead>
<tr>
<th>emitter length ( L_E [\mu m] )</th>
<th>time constant ( \tau_{TH} [\mu s] )</th>
<th>experimental ( R_{TH} [K/mW] )</th>
<th>modeled ( R_{TH} [K/mW] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3</td>
<td>0.47 ± 0.07</td>
<td>3.0 ± 0.2</td>
<td>6.372</td>
</tr>
<tr>
<td>2.5</td>
<td>0.54 ± 0.07</td>
<td>2.4 ± 0.1</td>
<td>4.630</td>
</tr>
<tr>
<td>5.0</td>
<td>0.6 ± 0.1</td>
<td>2.0 ± 0.1</td>
<td>3.018</td>
</tr>
<tr>
<td>20.0</td>
<td>0.64 ± 0.07</td>
<td>1.01 ± 0.07</td>
<td>1.003</td>
</tr>
</tbody>
</table>

are summarized in Table 3.1, which shows the experimentally-determined thermal time constants and thermal resistances. From these two sets of data, the thermal capacitance can be determined. For comparison, the thermal resistances predicted using the approach described in Section 3.2 are shown[37].

The experimental thermal resistance results show substantial decrease with increasing emitter length, varying by a factor of approximately 3 over the range of emitter lengths studied. In contrast with this, the time constants observed are only weakly dependent upon the emitter length. This unanticipated observation implies a relationship between the scaling of thermal resistance and thermal capacitance. The origin of this link from a mathematical point of view is given in Equation 1.19. A nominally-constant \( \tau_{TH} \) implies that any increase in \( R_{TH} \) is offset by a corresponding decrease in \( C_{TH} \).

As expected, the larger area devices have reduced thermal resistance. This trend can be justified qualitatively by considering the amount of material the thermal source is able to transfer heat into; the longer-emitter devices have a larger perimeter and area with which to transfer the heat away from the device. To examine more quantitatively the scaling behavior, the data have been plotted in
Figure 3.2. Varying predictions of power-law dependence of thermal resistance have been made, with results that decrease with the square root of emitter area [11, 38], or decrease linearly with the area [9, 17]. Accordingly, the data are presented in log-log form. The open circles show the measured thermal resistances tabulated in Table 3.1. The solid line shows a best fit through the experimental points, and corresponds to a scaling law with an exponent of -0.40. This fit to the experimental data is quite good, passing through all points within experimental error. Also included in Figure 3.2 are open and solid square symbols, which respectively show results of a similar measurement by Marchesan[28], who studied a device on the same wafer as that studied in the present work, and Tran et al.[25], who reported results for an earlier version of the IBM Si/SiGe HBT technology. Considering the differences in experimental technique, in particular, the use of wire bonding in the present work, and RF probing in References [25, 28], the results are in excellent agreement.

We now turn to a comparison between the experimentally-observed and theoretically-predicted scaling of thermal resistance. The observed scaling exponent of -0.40 is remarkably close to that predicted by Joy et al.[11] and Zweidinger et al.[38], both of whom predict an exponent of approximately -0.5. All other predictions give scaling exponents further from the experimental observations. It is tempting, therefore, to conclude that the experimental data support the theory of References [11] and [38]. This motivated an examination of the scaling laws claimed; in particular, Equation 9 of Reference [11] was numerically integrated for device dimensions similar to those studied here, as well as those studied in Reference [11]. It was not possible to reproduce the scaling laws claimed for rectangular devices of aspect ratios greater than approximately 5:1. In fact, for such high aspect ratio devices, the exponent predicted by Equation 9 of Reference [11] is -1. Therefore it
Figure 3.2: Experimental and theoretical thermal resistance parameters as a function of device area, plotted in a log-log fashion. The measured and calculated quantities appear as open and closed circles respectively, while linear regressions to the data are shown with a solid line. The theoretical prediction is seen to asymptotically approach the experimentally measured quantities for large area devices; the parasitic theory, which includes a first-order estimate to top-side heat removal, is shown with the dotted line. Also plotted are the results of Tran et al., and Marchesan as described in the text.
appears that the results of References [11, 38] should be viewed with caution; they will not, therefore, be considered further.

The theoretical approach described in Section 3.2 was used to perform calculations for the device geometries studied in the present work, including a detailed description of the trench isolation structure and its dimensions [37]. The results are shown in Table 3.1, and shown by the solid circles in Figure 3.2; the dashed line serves as a guide to the eye, connecting the results of the five calculations. Comparing these results with experimental values, it is apparent that the theoretical prediction is excellent for the largest device studied, agreeing within the experimental error. In judging the success of this model for the large devices, it is important to point out that such agreement would not be obtained without inclusion of the effects of trench isolation. In fact, calculation of the thermal resistance for the present geometry, ignoring the trenches, would underestimate the thermal resistance by approximately a factor of two.

Notwithstanding the model's ability to predict the thermal resistance of the large device, the divergence from the experimental results seen for the smaller devices is substantial. It is apparent that the theoretical results cannot be described by a single scaling law, showing a greater dependence on area for large devices than for small ones. This is not unexpected, however, because the smallest devices shown in Figure 3.2 have aspect ratios approaching that of a square. Numerical calculations for non-trench-isolated devices, which will be presented in Chapter 4, show a -0.5 exponent for scaling in square devices. The dependence of the theoretical curve on area shown in Figure 3.2 can therefore be interpreted as showing a transition between square and rectangular device scaling, where an exponent of -1 is expected [10, 17].

There are a number of physical effects which could account for the observed
discrepancy. The first one, relevant for the smaller devices which have higher predicted thermal resistance, would be a contribution to the measured values from a parasitic heat conduction path not included in the device model. The most obvious origin of such a path would appear to be the omission of heat flow out the top of the emitter; in fact, the top surface is assumed to be adiabatic, allowing no heat flow at all, in the model. It is difficult to assess the contribution of top-surface conduction without a detailed examination of the metallization structure. As was suggested in Chapter 2, the role of such metallization can be estimated using the one-dimensional isothermal bar model presented in Chapter 1. Such an estimate of the thermal resistance was made, assuming a 1 μm thick aluminum interconnect metallization, 50 μm wide and 150 μm in length, connected to two large emitter pads assumed to be at 300K. The resulting thermal resistance estimate is 6 K/mW, and the contribution of such a metallization-related path would be expected to be similar for devices of varying area. Hypothesizing that a parasitic path such as that described plays a role in the measurements, the theoretical results were modified to incorporate a fixed parallel conduction path. The value of approximately 6 K/mW gave the best agreement with the smallest device where such an effect would play the greatest role. The resulting theoretical curve is shown in Figure 3.2 by the dotted line. Agreement with experiment is greatly improved, as the impact on the larger devices, which have lower thermal resistances, is much less than that seen for small devices. While the constant-parallel-path hypothesis made is reasonably successful in reconciling theory and experiment, it is clear that a more detailed model would be required for accurate predictions. In fact, the conclusion of the present work that parasitic paths may play a role strongly motivates an experimental study of the role of metallization upon thermal resistance in small-area devices.
Having examined the scaling and theoretical prediction of thermal resistance, the discussion now turns to a similar study of the thermal capacitance data. Figure 3.3 shows the experimental variations of thermal capacitance with the square root of the emitter area. The dotted line shows a linear regression to the data, as suggested by the theory outlined in Section 1.4.2. The fit of the theory to the data is excellent, especially considering that the thermal capacitance values are plotted versus the square root of the emitter area and not the square root of the surface area of the heat-generating region itself, as suggested by Chow et al.. Estimation of the thickness of the thermal source as the width of the base-collector space charge region results in a very small error when the square root of the emitter area is approximated as the product of the square root of the surface area and a constant prefactor. While the slope of the linear regression changes when this approximation is made, the linear variation is maintained and the approximation results in only a 3% error. As this error is much smaller than the error in the experimental data, the thermal capacitance is still observed to vary linearly with the square root of the emitter area. Experimental evidence supporting the trend of thermal capacitance to increase linearly with the square root of the emitter area lends itself to a very simple yet accurate scaling law. While this model does not allow for the calculation of an absolute measure of the thermal capacitance, it does provide valuable information pertaining to the scaling of thermal capacitances for similarly shaped structures.

3.4 Conclusions

This chapter presented experimental and theoretical data for the scaling of thermal parameters for deep trench, Si/SiGe heterojunction bipolar transistors. The devices studied consisted of constant 0.5 μm width structures with emitter lengths varying
Figure 3.3: Thermal capacitance as a function of the square root of emitter cross-sectional area for Si/SiGe heterojunction bipolar transistors. The dotted line, which shows a linear regression to the data, is an excellent fit to the data.
from 1.3 to 20.0 μm long.

Based on the experimental evidence provided for the thermal resistance, a simple scaling law was devised that predicts the thermal resistance of the devices studied varies as $A_{E}^{-0.40}$, where $A_{E}$ represents the area of the emitter. The experimental thermal resistance data depicts a clear trend, in qualitative agreement with that predicted via a two-stage, composite Green’s function approach developed by D.J. Walkey et al. at Carleton University. The experimental thermal resistance values of small area devices, observed to be much smaller than those predicted theoretically, are believed to be influenced by alternate heat flow paths not accounted for in the device model. It was shown that a simple estimate based on one-dimensional heat flow calculations for the emitter metallization could account for the discrepancy observed between experiment and theory. While such an analysis provides evidence that interconnect metallization may be playing a role in the measurements, a more extensive calculation would be required to place a quantitative bound on the effects of top-side metallization on device thermal performance.

The results of Chow et al. concerning scaling of the capacitance of arbitrary three-dimensional structures was applied to the thermal capacitance measurements for the bipolar devices studied. The experimental data was found to vary linearly with the square root of the emitter area. While strictly the theory suggests a linear variation with the square root of the surface area, approximation of the square root of the surface area as that of the square root of the emitter area and a constant prefactor results in a maximum error of 3%. The excellent agreement between experimental data and the least-squares, linear regression provides strong evidence for a simple and accurate scaling law for the thermal capacitance. Based on the slowly-varying nature of the thermal capacitance with device shape for a given area,
it is expected that this scaling applies to devices of differing geometry as well; such an investigation will be performed in Chapter 4.
Chapter 4

Role of Layout in Thermal Performance of AlGaAs/GaAs Heterojunction Bipolar Transistors

This chapter is devoted to an experimental and theoretical study of the effect of device layout on thermal performance of AlGaAs/GaAs HBTs. The goal is to provide data to guide designers in the choice between layouts of equal area with differing geometries. Clearly, there are compromises that take place when optimizing the thermal performance of the device, and these should also be considered when choosing an overall optimum structure. Quantitative comparisons of one-, two- and three-dimensional thermal resistance models for non-trench isolated devices are used to evaluate the accuracies of predictions of the simpler models. Experimental measurements of thermal resistance and capacitance are made for varying device
area and for square, circular and one- and two-finger rectangular geometries. The experimental results are compared with three-dimensional analytical predictions.

4.1 Description of Devices Tested

The techniques described in Chapter 2 are used to make measurements on a series of AlGaAs/GaAs heterojunction bipolar transistors of varying emitter areas and geometries. The devices were fabricated at the Rockwell International Science Center; details of the device process are given in [26]. The four geometries investigated are shown in Figure 4.1, and have square, circular, or one- and two-finger rectangular emitters. Square devices studied have dimensions in microns of 1.2x1.2, 1.4x1.4, 2x2, 4x4 and 8x8 μm²; circular devices have diameters of 2.4 and 3.9 μm. All rectangular devices have finger widths of 1.4 μm, and finger lengths of 3, 6, 9 and 12 μm. The interfinger separation for the two finger devices is 3.4 μm. All devices were fabricated on the same wafer, and have the same doping and compositional profiles; the thickness of the wafer is 635 μm.

4.2 Thermal Resistance Calculations

In the present chapter a variety of device geometries will be characterized. The theoretical models described in Chapter 1 are based on one-, two- or three-dimensional descriptions of the heat-flow problem. While one- and two-dimensional analytical predictions of thermal resistance are available for circular geometries, the three-dimensional analytic approach described in Chapter 2 can only be applied to one-finger rectangular or square geometries. It will be shown that the one- and two-dimensional models are substantially in error for rectangular geometries. Because
Figure 4.1: Four possible device layouts, each with identical emitter area but differing thermal performance; (a) shows the square emitter, (b) the one-finger or rectangular emitter, (c) the two-finger emitter and (d) the circular emitter with all relevant dimensions labeled. The interfinger separation on the two finger device is 3.4 μm.
development of three-dimensional calculations needed for accurate modeling of these more complex geometries is beyond the scope of the present work, the analysis is restricted to one-finger devices with rectangular and square geometries.

Quantitative comparison of one-, two- and three-dimensional models is presented in Section 4.2.1 to evaluate their respective accuracies. Section 4.2.2 presents quantitative predictions of thermal resistance as a function of the total emitter area for the devices described above. Finally, Section 4.2.3 describes the role of temperature non-uniformity, and quantifies this effect through the introduction of an effective emitter area.

4.2.1 Analytical Predictions of Rectangular Device Thermal Resistance

Based on tabulated shape factors, as shown earlier in Chapter 1 for the isothermal rectangular parallelepiped and the thin rectangular thermal source in Equations 1.12 and 1.13, respectively, the thermal resistance for rectangular geometries, 1.4 μm wide in a 635 μm thick, GaAs substrate are calculated. The shape factor relations are taken from Reference [10] which, for the short rectangular isothermal plate, agree with those found in [9]. The fully three dimensional calculation of thermal resistance can be performed via the methods discussed earlier in Section 2.2; unlike the previous two shape factor calculations, this one is capable of quantifying the thermal resistance of rectangular structures for device lengths that approach and even equal the width of the device. For this reason, the three-dimensional predictions will be used as the benchmark with which the accuracy of the previous two methods will be evaluated.

Figure 4.2 shows the calculated results for rectangular emitters, 1.4 μm wide,
Figure 4.2: Analytic predictions of rectangular device thermal resistance. The one-dimensional model corresponds to a long, thin heat source and ignores corresponding lateral and longitudinal heat flow that occurs in the true structure. The two-dimensional calculation assumes the length of the device to be much greater than the width, and therefore agrees well with the three-dimensional summation calculation in the limit of long devices.
ranging in length from 3 to 24 $\mu$m. The calculation was performed for a GaAs substrate, where the thermal conductivity is taken to be $46 \text{ Wm}^{-1}\text{K}^{-1}$. The thermal source thickness, chosen to be representative of the base-collector space charge region width, is assumed to be 0.2 $\mu$m. Five different emitter lengths were investigated, and the trend of the results is shown with the dotted line. It is seen that the one-dimensional calculation does not agree with the three-dimensional calculation, even in the limit of devices that are over seventeen times as long as they are wide. Clearly, the heat flow that originates from the side of the devices is appreciable, as seen by the difference between the one- and two-dimensional shape factor calculations. Contrary to the one-dimensional case, the two-dimensional case shows good agreement with the three-dimensional calculation for relatively long devices, and it is seen that the two methods converge at the same value for the longest device studied. However, for shorter devices the two methods diverge, as the two-dimensional calculation ignores the longitudinal heat flow of the device. For shorter devices, this component can become comparable to the lateral component, and therefore the two-dimensional calculation greatly overestimates the thermal resistance of the shorter structures.

As the three-dimensional calculation has no assumptions regarding the relative dimensions of the thermal source, use of this method as the benchmark to evaluate the accuracy of the lower order methods is justified. However, comparison of the calculation predictions to experimental results must be approached with some caution as there are two main assumptions not based on the length scales of the thermal source that are inherent to the above calculations. Firstly, and shown earlier in Chapter 2, these results only represent the thermal resistance of the structure if no other heat flow path exists. In addition to conduction of heat from the top surface to the backside of the wafer, there will also be some parallel heat transfer processes.
that serves to cool the device; while the role of radiation at these temperatures is small, conduction through the device metallization to the top surface is one process that is believed to compromise the theoretical predictions presented here, as was shown previously in Chapter 2. In order to account for these effects, a detailed knowledge of the device geometry would be required. Secondly, the absence of electrothermal feedback in the calculation is especially important for large devices, or those operated at appreciable power densities. Inclusion of electrothermal feedback would result in the current being constrained to a localized portion of the active device, such that the effective emitter area is smaller than the nominal emitter area.

4.2.2 Thermal Resistance Comparison of Rectangular and Square Emitter Devices

Due to the inability of the one- and two-dimensional analytic models to accurately evaluate the thermal resistance for devices with comparable lateral emitter dimensions, the three-dimensional, uniform power dissipation model is used to perform comparisons between rectangular and square emitter devices. The rectangular and square devices studied theoretically are of the same dimension as those measured experimentally. Details regarding device size were discussed previously in Section 4.1. Figure 4.3 shows the predicted variations of thermal resistance for emitter areas ranging from 1.4 to 64 \( \mu m^2 \). Both the rectangular and square values approach the same thermal resistance values for small area emitters, as expected. For small dimension devices, the rectangular device begins to take on the shape of the square device, so both trends coincide for small area emitters. However, for larger area devices, the predicted thermal performance for the rectangular device is better than its equal area, square counterpart. This is due to the ability of the rectangular
Figure 4.3: Three-dimensional analytical predictions of rectangular and square emitter device thermal resistance. In the asymptotic limit of small devices, the rectangular and square devices are identical. However, for larger area devices, the thermal resistance of the rectangular device is smaller than the square device.
device to shed heat more efficiently to the surrounding area via its larger periphery to area ratio. For large area power devices, a rectangular topology is much more efficient at reducing the mean temperature rise for a given power dissipation, and thus the rectangular device could be operated at higher power densities and still remain within its safe operating limit. While the effect is small for small area devices, it is an appreciable effect for large area devices; at the maximum size of rectangular devices tested, there is on average a 65% increase in the temperature rise of the square device over that of the rectangular device for the same power dissipation.

4.2.3 Model Limitations for Large Area Devices

The difference predicted by the three-dimensional model for rectangular and square devices of identical emitter areas represents only a first-order estimate of the thermal benefits of the rectangular structure over the square structure. Because the heat flow originating for the large, square emitter device must travel through appreciable distances of active emitter region before it enters the non-active substrate, large temperature gradients exist across the active region. These gradients, indicative of temperature non-uniformity, result in more electrothermal feedback taking place for the square device than for the rectangular device for a fixed power dissipation. The effect of this temperature non-uniformity from a thermal perspective is that the effective thermal source appears to be smaller than the nominal emitter dimensions of the device. Associated with this effective smaller area is a rise in the thermal resistance of the device. The thermal resistance calculations for the devices presented in Section 4.2.2 represent a lower bound on the thermal resistance; the estimate is more accurate for the rectangular devices owing to the less severe temperature non-uniformity.
In order to quantify the extent of temperature non-uniformity taking place in a device, a new figure of merit termed the effective emitter area can be introduced. The effective emitter area is that which dissipates the majority of the power, defined here as 90% of the total power. By using the spatial temperature distribution arrived at with the three-dimensional analytical calculation, the current flow across the surface of the device can be calculated via Equation 1.1 when specialized to the devices of this study. The area of the device that supports 90% of the power dissipation can be determined by an integration of the current across the active region of the device. In order to provide a simple explanation for the results, the ability of a given structure to shed heat into the substrate is quantified by the ratio of the emitter periphery to emitter area. For small aspect ratio devices such as the square, with corresponding small periphery to area ratios, heat must flow for significant distances across the active region of the device before reaching the surrounding thermal medium; this results in significant temperature non-uniformity, and an associated small effective emitter area. Conversely, large aspect ratio devices with large periphery to area ratios such as long rectangular devices are able to shed heat efficiently into the nearby substrate without imposing a large degree of temperature non-uniformity to the source itself. From this, one expects that large emitter periphery to area ratios are indicative of small amounts of temperature non-uniformity, and that small periphery to area ratios are indicative of larger amounts of temperature non-uniformity. This premise will be tested in the following discussion, and will be extended to the experimental data of Section 4.3 as indication of the level of temperature non-uniformity of the experimental devices.

The effective area calculation was performed by quantifying the temperature rise at 25 locations in a quarter of the device. Using device parameters and doping
Table 4.1: Effective area normalized to total emitter area $E_a/E_A$ for power densities of 0.1 and 1.0 mW/μm². For large area square devices with associated small periphery to area ratios $E_P/E_A$, one sees considerable current constriction in the form of small effective emitter areas, defined as the area of the emitter that dissipates 90% of the power.

<table>
<thead>
<tr>
<th>device [μm²]</th>
<th>$E_P/E_A$ [μm⁻¹]</th>
<th>$R_{TH}$ [K/mW]</th>
<th>$E_a/E_A$ (0.1mW/μm²)</th>
<th>$E_a/E_A$ (1mW/μm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4x3</td>
<td>2.04</td>
<td>4.76</td>
<td>0.98</td>
<td>0.87</td>
</tr>
<tr>
<td>1.4x6</td>
<td>1.76</td>
<td>3.04</td>
<td>0.97</td>
<td>0.83</td>
</tr>
<tr>
<td>1.4x9</td>
<td>1.65</td>
<td>2.30</td>
<td>0.97</td>
<td>0.82</td>
</tr>
<tr>
<td>1.4x12</td>
<td>1.60</td>
<td>1.88</td>
<td>0.97</td>
<td>0.82</td>
</tr>
<tr>
<td>1.2x1.2</td>
<td>3.33</td>
<td>8.29</td>
<td>0.98</td>
<td>0.86</td>
</tr>
<tr>
<td>2x2</td>
<td>2.00</td>
<td>5.05</td>
<td>0.98</td>
<td>0.87</td>
</tr>
<tr>
<td>4x4</td>
<td>1.00</td>
<td>2.53</td>
<td>0.96</td>
<td>0.77</td>
</tr>
<tr>
<td>8x8</td>
<td>0.50</td>
<td>1.26</td>
<td>0.93</td>
<td>0.68</td>
</tr>
</tbody>
</table>

concentrations found in Reference [26] in a simple analytic model shown earlier in Equation 1.1 together with the spatial temperature distribution, the current density as a function of position in the active region of the device was calculated. Using the 25 points across a quarter of the device, the current density was interpolated at ten thousand points by assuming a linear variation between adjacent points. The contributions of successive, discretized elements were summed in order of dominant contribution to the collector current until their contribution was equal to 90% of the total current flow, which for this simple model is equal to 90% of the total power dissipation.

Table 4.1 shows the results for rectangular and square emitter devices; the fractional effective emitter area, $E_a/E_A$, given by the ratio of the effective area normalized to the total emitter area, is shown at two differing power densities, as a function of the device size and the periphery to area ratios $E_P/E_A$. At small power densities, very little temperature non-uniformity occurs, as virtually the entire de-
vice supports the power dissipation. However, at high power densities of operation and for the larger devices, only a portion of the active area supports any real current. For example, in a 8×8 \( \mu m^2 \) transistor only 68% of the active region is required to drive 90% of the current. Effectively then, one sees that modeling such a device with an 8×8 \( \mu m^2 \) thermally generating region with uniform power dissipation is erroneous; a more accurate result could be achieved by modeling a device with 68% of this area, with only 90% of the power dissipation. While this neglects the temperature rise associated with the 10% of power located around the periphery, concentration of the remaining 90% of power dissipation near the center of the device may result in a larger overall thermal resistance. Such a calculation was performed, where the device structure studied consisted of a 6.6×6.6 \( \mu m^2 \) thermal source. The effective thermal resistance calculated for this structure, accounting for the 90% power dissipation, was 1.38 K/mW. By comparing this to the 1.26 K/mW thermal resistance associated with the 8×8 \( \mu m^2 \) device, one sees that the extensive spatial non-uniformity of the temperature profile and the corresponding power dissipation results in an underestimate of the thermal resistance by the uniform power dissipation model. Again, one realizes that this corrected value of the thermal resistance of 1.38 K/mW places a refined estimate on the lower bound of the thermal resistance; the temperature dependent thermal conductivity and electrothermal feedback will result in an increased thermal resistance from the value quoted here.

The effective area results presented in Table 4.1 are shown in graphical form in Figure 4.4, which shows the effective area normalized to the total emitter area for various sizes of rectangular and square devices. The rapid decrease of effective area with increasing emitter dimensions for square devices operated at high power densities is evident by the trend depicted. For the rectangular devices, the pro-
Figure 4.4: Effective area normalized to total emitter area for rectangular and square emitter devices at two different power densities of operation. At lower power densities, the effective emitter area which contains 90% of the power dissipation is similar for both the square and rectangular devices. However, at higher power densities, one sees appreciable current constriction in the square device, evidenced by the rapid decrease of the effective emitter area with increasing emitter size.
nounced lateral heat flow compared to the longitudinal flow results in the device
shedding most of the generated heat across the non-active portion of the substrate,
owing to the small lateral dimensions of the emitter. As the heat originating near
the center of the active region can reach the non-active material by traveling the
short, lateral dimension of the rectangular emitter, much of the temperature gradi­
ent occurs external to the active region. Conversely for the square device, the lateral
and longitudinal heat flow are equal, and this results in a pronounced level of heat
conduction across the active region of the device, resulting in appreciable tempera­
ture gradients and non-uniformity for the emitter region of the device. Clearly, the
notion that one can scale current densities linearly with increasing emitter area is
dependent on the emitter geometry, and could lead to accelerated failures of square
emitter devices.

One must also be aware that the calculation shown is not self-consistent in
terms of accounting for the resulting non-uniform power dissipation. In reality, the
effective area will be even smaller than that shown, owing to the electrothermal
feedback that causes current constriction in the device. The spatially non-uniform
temperature distribution suggested by the effective area figures of merit only leads
to more pronounced non-uniformity, as the hot areas of the device will support
more current and increasing levels of power dissipation. Full investigation of this
effect would rely on a recursive method that could account for a non-uniform power
dissipation in the active region of the device. This amounts to an understanding
that the calculation presented provides a first-order correction to the uniform power
dissipation model. The effective area calculation, while of limited accuracy, shows
clearly that temperature non-uniformity can be playing a significant role in small
periphery to area ratio devices.
emitter length \( L_E [\mu m] \) | emitter area \( A_E [\mu m^2] \) | time constant \( \tau_{TH} [\mu s] \) | thermal resistance \( R_{TH} [K/mW] \) \\
--- | --- | --- | --- \\
3.0 | 4.2 | 0.52 ± 0.07 | 3.1 ± 0.3 \\
6.0 | 8.4 | 0.6 ± 0.1 | 2.6 ± 0.2 \\
9.0 | 12.6 | 0.7 ± 0.1 | 2.2 ± 0.1 \\
12.0 | 16.8 | 0.55 ± 0.05 | 2.0 ± 0.1 \\

Table 4.2: Measured thermal parameters for one-finger AlGaAs/GaAs heterojunction bipolar transistors.

### 4.3 Experimental Results

An experimental investigation using the techniques presented in Chapter 2 was performed for a variety of device geometries with differing areas, as outlined in Section 4.1. Overall, fifteen devices were studied. This section presents experimental results obtained for the thermal resistance and capacitances of these devices, and compares the results to those expected based on theoretical considerations.

Tables 4.2 through 4.5 summarize the experimental results obtained for the devices studied. Each table contains information pertaining to a specific emitter geometry and overviews the experimental values measured for the thermal resistance and time constant as a function of device area. Table 4.2 summarizes the results for the one-finger emitter devices, with device lengths ranging from 3.0 to 12.0 \( \mu m \). Table 4.3 shows the results for the two-finger emitter devices, where the same length emitter fingers as those of the one-finger rectangular devices facilitates direct comparison between these two sets of devices. The one-finger 1.4×6 and 1.4×12 \( \mu m^2 \) devices have identical areas to the two-finger 2×1.4×3 and 2×1.4×6 \( \mu m^2 \) devices, respectively. From the data that appears in the above tables, it is shown that for a given emitter area, the two-finger devices generally exhibit a smaller thermal...
<table>
<thead>
<tr>
<th>emitter length $L_E$ [$\mu m$]</th>
<th>emitter area $A_E$ [$\mu m^2$]</th>
<th>time constant $\tau_{TH}$ [$\mu s$]</th>
<th>thermal resistance $R_{TH}$ [K/mW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>8.4</td>
<td>$0.60 \pm 0.09$</td>
<td>$2.2 \pm 0.1$</td>
</tr>
<tr>
<td>6.0</td>
<td>16.8</td>
<td>$0.67 \pm 0.06$</td>
<td>$2.0 \pm 0.1$</td>
</tr>
<tr>
<td>9.0</td>
<td>25.2</td>
<td>$0.7 \pm 0.1$</td>
<td>$1.71 \pm 0.08$</td>
</tr>
<tr>
<td>12.0</td>
<td>33.6</td>
<td>$0.7 \pm 0.1$</td>
<td>$1.54 \pm 0.07$</td>
</tr>
</tbody>
</table>

Table 4.3: Measured thermal parameters for two finger AlGaAs/GaAs heterojunction bipolar transistors.

<table>
<thead>
<tr>
<th>emitter length $L_E$ [$\mu m$]</th>
<th>emitter area $A_E$ [$\mu m^2$]</th>
<th>time constant $\tau_{TH}$ [$\mu s$]</th>
<th>thermal resistance $R_{TH}$ [K/mW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>1.4</td>
<td>$0.39 \pm 0.04$</td>
<td>$3.1 \pm 0.3$</td>
</tr>
<tr>
<td>1.4</td>
<td>2.0</td>
<td>$0.46 \pm 0.03$</td>
<td>$2.8 \pm 0.2$</td>
</tr>
<tr>
<td>2.0</td>
<td>4.0</td>
<td>$0.46 \pm 0.05$</td>
<td>$2.8 \pm 0.3$</td>
</tr>
<tr>
<td>4.0</td>
<td>16</td>
<td>$0.57 \pm 0.08$</td>
<td>$2.3 \pm 0.1$</td>
</tr>
<tr>
<td>8.0</td>
<td>64</td>
<td>$0.7 \pm 0.1$</td>
<td>$1.7 \pm 0.1$</td>
</tr>
</tbody>
</table>

Table 4.4: Measured thermal parameters for square AlGaAs/GaAs heterojunction bipolar transistors.

Equivalently, one could interpret the results in the form of an increased periphery to area ratio for the two-finger devices over that of the identical area, one-finger device. Tables 4.4 and 4.5 show similar results for square and round emitter devices. Unlike the one- and two-finger emitter results, no direct comparisons can be made because of the different emitter areas of the devices involved.

Figure 4.5 shows all of the experimental thermal resistance measurements as a function of the device area. The dotted lines depict only the trend of the data. The tendency of the experimental data to be constrained to a maximum thermal resistance of 3.1 K/mW for small emitter area devices is believed to be the effect of
Figure 4.5: Experimentally determined thermal resistances for circular, square, one- and two-finger emitter devices. For small devices, the thermal resistances are indistinguishable from each other within the resolution of the experiment. For larger devices, the thermal resistance of the two-finger devices is appreciably smaller than the other emitter geometries. The dotted lines group results according to device geometry, and are a guide to the eye. Also plotted is the result of Hanington et al. for the small one-finger rectangular device studied.
emitter diameter $E$ [μm]  
emitter area $A_E$ [μm²]  
time constant $\tau_{TH}$ [μs]  
thermal resistance $R_{TH}$ [K/mW]  

<table>
<thead>
<tr>
<th>$E$</th>
<th>$A_E$</th>
<th>$\tau_{TH}$</th>
<th>$R_{TH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>4.5</td>
<td>0.51 ± 0.04</td>
<td>2.6 ± 0.3</td>
</tr>
<tr>
<td>3.9</td>
<td>12</td>
<td>0.64 ± 0.07</td>
<td>2.5 ± 0.2</td>
</tr>
</tbody>
</table>

Table 4.5: Measured thermal parameters for round AlGaAs/GaAs heterojunction bipolar transistors.

an additional heat flow path that limits the temperature rise of the active region. This effect, described previously in Chapter 2 for a $3 \times 3 \mu m^2$ AlGaAs/GaAs device, was shown to limit the theoretically-predicted value of thermal resistance. The emitter metallization of the structure, which could provide a means by which the hot areas of the active device could release heat to the top surface of the wafer, is a possible explanation for this observation. For small emitter area devices, the thermal performance is very similar for differing emitter geometries.

Also plotted in Figure 4.5, shown as an open triangle, is the result of Hanington et al.[26], which shows the thermal resistance of a $1.4 \times 3 \mu m^2$ device determined via a simultaneous extraction of the parasitic emitter resistance and thermal resistance. From the plot, it is evident that the two measurements for the same device do not agree within experimental error. At present, possible explanations for the discrepancy observed between the experimental and theoretical data would be the level of power dissipation at which the measurement was performed, or if current constriction was playing a role in the measurement process. As reported by Oettinger et al.[14], the thermal resistance if a function of the power dissipation at which the thermal resistance measurement is performed, and significant discrepancy can be observed. Clearly, additional investigation of this difference is required in order to explain this apparently large deviation between experimental methods.
From the data corresponding to the square and round emitter devices, the two sets of devices are indistinguishable within the resolution of the measurement. Based on estimation of the emitter periphery to area ratio, this result is expected as the shape of a square or round device does not change appreciably for a fixed emitter area. Indeed, the emitter periphery to area ratio for a round device varies as $4/\Phi_E$, where $\Phi_E$ represents the diameter of the emitter, while for the square device the ratio is given as $4/L_E$. For equal area devices, these quantities only differ by 13%.

Looking to the single and double finger data, it is seen that for smaller area devices, the two-finger devices have a substantially reduced thermal resistance compared to the one-finger devices. However, for larger devices, the difference between the two emitter geometries begins to lessen. Again the emitter to periphery ratios are compared for a rectangular device of dimension $l \times w$ and a two finger device of dimension $2\times l/2 \times w$, such that both devices have identical area. The emitter periphery to area ratio for the one-finger device, $E_{Ps}/E_{As}$, is given as

$$\frac{E_{Ps}}{E_{As}} = \frac{2l + 2w}{lw} \quad (4.1)$$

while the ratio for the two-finger emitter device, denoted $E_{Pt}/E_{At}$ is

$$\frac{E_{Pt}}{E_{At}} = \frac{2l + 4w}{lw} \quad (4.2)$$

Clearly, for small devices for which the emitter width is comparable to the emitter length, the two-finger device exhibits enhanced heat transfer compared to the one-finger device owing to the additional heat conducted through the two additional ends of the device. However, in the limit that the length of the device becomes large compared to the width, the ratios approach the same asymptotic limit of $2/w$. For the data point that corresponds to measurements of both a one- and two-finger
device of identical area, there is only a 9% difference in the emitter periphery to area ratios of the two devices. While the lateral flow is impeded slightly in the two-finger devices because of the adjacent emitter finger, the 3.4 μm finger separation is probably sufficient to make this effect small.

Overall, it is seen that the best structure of those studied from a thermal standpoint, defined as one in which the temperature rise is minimized for a given power dissipation, consists of a two-finger device. However, for large-area devices, the benefits of a two-finger structure over a one-finger device are limited. Choice of an overall optimized structure in this case would include consideration of the geometry dependent electrical performance as well.

An examination of the variations in thermal capacitance with device area and geometry is shown in Figure 4.6, which depicts the measured variations of thermal capacitance as a function of the square root of the emitter area. The thermal capacitance shown was determined by taking the ratio of the measured thermal time constant to the measured thermal resistance, and the error estimates shown in Figure 4.6 are a composite of the limitations of each of these individual measurements. Discussion of the agreement between theory and experiment will be deferred until the end of Section 4.4 as discussions regarding the comparison between thermal resistance theory and experiment will be similar in content.

4.4 Comparison Between Theory and Measurement

In order to facilitate comparison between theoretical predictions and experimental measurements of thermal resistance, the relevant data has been plotted in Figure 4.7. From a qualitative perspective, the theoretical and experimental data support the same trend for devices of moderate size. For devices of sufficiently large size, the ex-
Figure 4.6: Experimentally determined thermal capacitances for circular, square, one- and two-finger emitter devices. Within the limitations of the measurement, the thermal capacitance appears to be only a function of the device area and not the shape. The dashed line shows a linear fit to the experimental data, where the largest square device has been excluded.
Figure 4.7: Comparison of experimental and theoretical thermal resistances for differing emitter geometries. Discrepancies seen for small area devices are believed to be the result of parasitic heat flow paths not accounted for in the model calculation. The trend for the experimental data to lie above the theoretical predictions for large area devices is believed to be due to the model assumption of uniform power dissipation, and its inability to account for electrothermal feedback.
Experimental data supports the theoretical trend that the thermal resistance of square devices is larger than that of rectangular devices; for the smaller devices studied, the ability to make this conclusion is impeded by the accuracy of the experimental thermal resistance data. Within the error of the experimental data, the square and rectangular devices approach the same value of thermal resistance for small-area devices, as is suggested by the three-dimensional calculation results of Section 4.2.2.

However, from a quantitative standpoint, the theoretical prediction of the thermal resistance is much greater than the measured results for small-area devices. This effect, as suggested earlier, is believed to be due to additional heat transfer through the emitter metallization that is not taken into account in the model calculation. This supposition is supported by the tendency of the experimental data to plateau around 3.1 K/mW. It was also shown in Chapter 2 that an emitter post and series metallization of reasonable dimensions predicts a magnitude of additional heat flow that is sufficient to approximately predict the observed plateau in the present experimental data. This effect is more pronounced for devices with large thermal resistances and as such, is believed to account only for the discrepancy observed for small-area devices.

For large-area devices, the experimental data lies above the theoretical results calculated with the three-dimensional analytical model. In this extreme, and as suggested by the effective area calculations of Section 4.2.3, it is believed that temperature non-uniformity is the origin of the differences observed between theory and experiment. As shown with calculation for a 8×8 μm² square emitter device, the 90% of the power dissipated is constrained within 68% of the device and successive calculation of the thermal resistance corresponding to this structure resulted in a 10% increase in the predicted value of the thermal resistance. While the magnitude
Table 4.6: Regression coefficients for AlGaAs/GaAs non-trench isolated structures. The coefficient represents the exponential of device area power law that describes the scaling present in the experimental and theoretical data of Figure 4.7.

of the predicted increase is not sufficient to bring into agreement the experimental and theoretical results, electrothermal feedback, which is not accounted for in the model, is going to amplify this effect. If incorporated into the model calculation, it would reduce the observed discrepancy further. Temperature non-uniformity is also more severe with the square emitter devices compared to the rectangular devices, as shown previously in Table 4.1. This could well account for the larger discrepancies observed between the theoretical predictions and experimental results for the square device, as compared to those of the rectangular device.

In a similar fashion to the experimental thermal resistance results of Chapter 3, Figure 4.7 has been plotted in a fashion to facilitate the extraction of power law scaling constants for the experimental and theoretical data. Comparisons of the experimental scaling coefficients to those predicted theoretically will show support for the conclusion that temperature non-uniformity and interconnect metallization are playing a significant role in the thermal behavior of the devices studied. Table 4.6 shows the results of the linear regressions performed on the experimental and theoretical curves for the one-finger rectangular and square devices. The tabulated coefficients represent the exponent of the emitter area power law that best describe the scaling of the thermal resistance of the AlGaAs/GaAs non-trench-isolated structures; for example, the theoretical coefficient for the square devices of -0.497 repres...
sents a scaling that is nearly linear with the inverse of the square root of the emitter area. As suggested by the theoretical calculations performed here, and the results presented in the literature[10, 17], the scaling coefficient for an ideal thermal source appears to take on values ranging from -0.5 to -1; the small aspect ratio devices, such as the square devices of the present calculation, show a scaling coefficient of -0.497, and very large aspect ratio devices correspond to scaling coefficients that approach -1. The theoretical scaling coefficient of the rectangular devices considered in this study lies between these two possible extremes, as expected.

The fits to the experimental data, while passing through all of the data points, predict scaling coefficients that fall outside the apparently possible range. In order to account for this discrepancy, the two main limitations of the model are considered here, each in turn, to ascertain their effect on the nominally correct predictions for the ideal thermal source. Firstly, the primary limitation of the thermal resistance model for small-area devices is considered. This limitation, already outlined in Chapters 2 and 3, is the model omission of the heat flow that occurs down the interconnect metallization. The effect of this additional heat flow path will be to limit the temperature rise of the device, thereby lowering the thermal resistance for the small devices. This effect, if incorporated into the device model, would result in the prediction of a decreased scaling coefficient. The second effect, and the primary one affecting the large-area devices, is temperature non-uniformity. This effect, already outlined, will result in an increased thermal resistance of the large-area devices. The overall result of temperature non-uniformity is a decreased scaling constant, such that interconnect metallization and temperature non-uniformity act in a similar manner; these effects, if incorporated into the model calculation, would result in a decreased discrepancy between experimentally and theoretically predicted
scaling coefficients. It is seen then, that investigation of the scaling coefficients allows for an alternate interpretation of the data, which is consistent with quantifying the effects of interconnect metallization and temperature non-uniformity, as was shown previously.

Having performed a comparison between theory and experiment for the thermal resistance data, the discussion now turns to an investigation of the geometry and area dependence of thermal capacitance. The experimental and theoretical results for the device thermal capacitance were shown previously in Figure 4.6. As suggested by the theory of Section 1.4.2, the data has been plotted as a function of the square root of the emitter area. The dashed line shows a linear fit to the data and is forced to pass through the origin; the largest square device has been excluded from the fit to the data. The failure of this largest, square device to follow the scaling law suggested in the literature[13] is believed to be the result of appreciable temperature non-uniformity, as cited earlier in the discussion of thermal resistance. Significant levels of temperature non-uniformity would result in an effectively smaller device; from Figure 4.6 it is seen that about a 60% reduction in the device size would account for the apparent discrepancy between the experimental measurement and the simple scaling law. As suggested by the effective emitter area calculation of Section 4.2.3, the area of the 8x8 μm² square device is substantially less than the nominal 64 μm². Thus, for large-area devices, there appears to be a systematic error incurred by plotting the thermal capacitance as a function of the nominal emitter area. One can visualize plotting the 64 μm² device at an effective area of 36 μm², which would allow one to extend the trend of linear increases of thermal capacitance with the square root of the emitter area. This value of effective emitter area is of the same order as that suggested by the calculations shown in Fig-
ure 4.4, and could easily reconcile the difference between experiment and theory for the largest device. Finally, as suggested by Chow et al.

Chow, et al.,[13], the tendency is for the capacitance to be virtually independent of the device geometry. This is supported by the fact that the linear fit to the data essentially passes through all of the data, independent of device geometry. While Chow et al. show a five to ten percent difference among flat conductors of differing shape, this effect is too small to be seen within the error bounds of the experimental data of Figure 4.6; a more accurate investigation of this small effect would necessitate more stringent error bounds on the experimental data.

4.5 Conclusions

This chapter presented experimental and theoretical results for various emitter geometry and area AlGaAs/GaAs heterojunction bipolar transistors. In particular, the variations of thermal resistance and capacitance as the area of the emitter region increases for one-finger, two-finger, square and round emitter devices is examined in order to ascertain the optimum structure from a thermal standpoint. This represents the first geometrical study of experimentally determined thermal parameters.

It was found that the effective emitter area, that area of the active region that supports 90% of the total power dissipation, is a measure of temperature non-uniformity. It was shown that square devices had a smaller effective area than rectangular devices of identical emitter area. It was also found, that in situations where detailed knowledge about the current distribution within a device is unknown, as is the case with experimental data, the use of the emitter periphery to area ratio provides a similar measure of temperature non-uniformity.

Experimental studies of differing emitter geometry devices showed that for
a fixed emitter area, the two-finger rectangular emitter device exhibits the lowest thermal resistance. The circular and square emitter devices, with emitter periphery to area ratios that differ on the order of 10% are shown to behave identically within the resolution of the experimental method. It was found that, for smaller area devices, the structure of choice for optimizing the thermal performance was the two-finger device. For larger area devices, negligible differences are observed between one- and two-finger devices, and as such, choice of an overall optimum structure should include consideration of the electrical performance as well.

The three-dimensional uniform power dissipation theory was shown to predict accurate values of thermal resistance for both rectangular and square geometries for moderately sized devices. Through the use of the effective emitter area and the emitter periphery to area ratio, it was shown that the deviation between theory and experiment for large area, square devices is the result of substantial temperature non-uniformity.

Comparison of theory and experiment of thermal capacitance data for the AlGaAs/GaAs heterojunction bipolar transistors has provided valuable scaling information. In accordance with the theory developed by Chow et al., it was shown that the thermal capacitance was linearly dependent on the square root of the emitter area, and was insensitive to the exact geometry of the thermal source. While data for a 64 $\mu$m$^2$ square transistor was found to not support the theory within the error of the measurement, it was shown that temperature non-uniformity is likely playing a role in the thermal capacitance, which is consistent with similar conclusions derived from the thermal resistance of the structure.
Chapter 5

Conclusions

The work described in this thesis consisted of the development of a pulsed base bias system for study of the thermal response of bipolar transistors, and the application of this approach to a variety of modern heterojunction bipolar transistors (HBTs). From this work a number of conclusions were made.

It was shown that the thermal resistance and capacitance can be extracted from results of transient and steady-state characterization. Using isothermal extraction techniques at a variety of temperatures, the thermal resistance can be determined in a manner that is independent of many of the assumptions used for such measurements in the past.

A number of broad conclusions concerning thermal resistance can be made from studies of a wide variety of AlGaAs/GaAs and Si/SiGe HBTs. For small emitter areas, two-finger devices have lower thermal resistance than one-finger devices; however, this difference becomes smaller and eventually negligible with increasing emitter area. Considering devices of equal total area, there is no advantage of a circular over a square layout, with both having thermal resistances larger than
the two-finger device, which gives the best thermal performance for the geometries studied.

Theoretical comparison between the simple analytical models often used for prediction of thermal resistance and a more rigorous three-dimensional model showed that the former are substantially in error. However, this three-dimensional calculation is only accurate for very low power dissipations, because at reasonable powers, temperature non-uniformity plays an important role. A novel approach to quantifying this effect in a simple way was developed, based on introduction of the concept of an effective emitter area. At moderate power levels in large devices the effective area can be as little as 50% of the physical emitter area. The overall conclusion of the theoretical work is that accurate inclusion of the effects of temperature non-uniformity will be an essential feature of a successful device model.

Measured thermal resistances consistently show a weaker dependence on area than predicted. This was found to be true over quite a wide range of device sizes, geometries, and in technologies which differed substantially. In small devices it was concluded that omission of parasitic heat flow paths due to contact metallization causes the models to overestimate the measured values. For large devices it was observed that the measured thermal resistances exceed the predicted ones; it was concluded that the origin is the temperature non-uniformity mentioned above. The overall conclusion of the comparison between calculated and measured thermal resistance is that three effects play a role: conduction to the backside of the wafer, parasitic conduction through metallization, and temperature non-uniformity. It appears that for the devices studied, at least two of the three effects must be included for any given device.

Measured thermal capacitances were found to be very well described by a
simple scaling law. The square-root dependence upon emitter area, and the relative independence of device geometry were concluded to be general features of thermal capacitance, consistent with previous studies of the electrical capacitance of variously-shaped conductors. Deviation from this scaling law was observed for at least one large-area device. This was attributed to a decrease in effective emitter area due to temperature non-uniformity, which is consistent with similar conclusions outlined above that were based on theoretical predictions, or thermal resistance measurements.

Taken together, the conclusions presented in this thesis motivate a variety of further studies. First, improvement of the experimental apparatus in terms of pulse rise time would reduce the errors in measuring isothermal current and thermal resistance. It would also be very convenient to make such measurements with on-wafer probes rather than the diced and bonded samples used in the present work. Experimental verification of the role of parasitic metallization could be investigated by studying identical devices in differing metallization patterns. It seems that the information to be gained from finite-element simulations of realistic experimental devices will be invaluable in guiding future work, both in the development of simpler models and in the interpretation of experimental results.
Bibliography


129


[37] D.J. Walkey (Carleton University), private communication, Aug. 1998.

Appendix A

Pulse Circuitry

Figure A.1 shows the pulsed circuitry used to perform the experimental characterization of the self-heating effect in bipolar devices. The circuit can function either as the collector circuitry, or the base circuitry, depending on the input to the clock. For a clock signal of duration $t_{CK}$, the circuit will switch between voltage levels $V_{be,max}$ and $V_{offset}$, such that the voltage signal at $V_{out}$ will be $V_{be,max}$ for the time period defined by $t_{CK}$. The current flowing through the device is obtained by measuring the voltage signal at $I_{out}$, where a measured signal of 50 mV represents a current flow of 1 mA to the base of the device under test (DUT). When used as a pulsed source, the base of the DUT should be connected to $V_{out}$.

For use as the collector voltage source, one uses a clock input of zero. In this case, when the collector of the DUT is connected to node $V_{out}$, the voltage signal at $I_{out}$ represents the current flow through the collector contact of the device, where again a 50 mV signal is representative of a current flow of 1 mA.

All power supplies are decoupled with a 3.3 µF capacitor in parallel with a 100 nF capacitor, at their entry point to the circuit.
Figure A.1: Pulsed measurement circuit schematic. \( V_{\text{out}} \) represents the node to which the device is connected. \( I_{\text{out}} \) is a voltage signal in mV equal to 50x the current flow through the device, in mA. \( V_{\text{be,max}} \) is the high state of the pulse applied to the device, and \( V_{\text{offset}} \) is the low state.