THERMAL LIMITATIONS OF InP HBT’S IN 80 AND 160 Gbits⁻¹ IC’s

I.Harrison†, M. Dahlstrom, S. Krishnan, Z. Griffith, Y.M Kim, M.J.W. Rodwell
†School of Electrical and Electronic Engineering, University of Nottingham, Nottingham, NG7 2RD, UK
Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA

A 3D thermal model based on finite elements has been developed for the analysis of the thermal resistance of
InP heterojunction bipolar transistors. The model was verified by comparing simulated and experimental
results. The simulations also show that the maximum temperature in the device can be significantly higher than
the experimentally determined base-emitter junction temperature. By applying scaling laws, a road map for
80Gbit/s and 160Gbit/s devices is presented. Simulations show that devices suitable for 160Gbit/s circuits will
be thermally possible if the InGaAs etch stop or contacting layer is removed from the sub collector.

I. Introduction

40Gbit/s integrated circuits (ICs) are becoming commercially available. Two major competing technologies
are SiGe heterojunction bipolar transistors (HBTs) and InP HBTs. While very high mm wave power gains have been
obtained with single heterojunction devices (SHBT), they suffer from low breakdown voltages and high thermal
resistance because of the narrow band gap InGaAs collector.

To overcome these inherent problems, a double heterostructure bipolar transistor (DHB Ts), which has an InP
collector, can be used. Devices of this kind have been reported with a max in excess of 450GHz [2,3]. These results indicate
that DHB Ts have potential applications in 80Gbit/s and 160Gbit/s ICs. This is reinforced by recent observations of
frequency dividers operating above 80GHz [4,5]. As device design is modified for 80-160Gbit/s applications, the current
density within the device will greatly increase resulting in higher device temperatures and so accurate thermal design is
critical for accurate circuit simulation and reliability studies.

Thermal modeling of HBTs has, to date, centered on power
amplifiers [6]. These studies have concentrated on modeling the heat flow in the substrate using a semi-analytical approach
developed for Si BJTs [7]. The method does not account for
(1) the temperature gradients within the device which are
significant in sub-micron HBT’s or (2) heat flow through the
emitter metallisation which can be high as 20% of the total heat
dissipated.

Before circuits can be marketed, the device reliability must
be assessed and this has driven an interest in the thermal resistance of InP HBTs. [9,10] For 80 and 160 and probably
40Gbit/s applications, DHB Ts will be used because of their lower thermal resistance. This paper examines the thermal
performance of devices for the 80 and 160 Gbit/s applications
by modeling the heat flow in 3D.

II. Device Scaling

To achieve high-speed performance, an increase in the
current density (J) will allow a device with a smaller collector-base junction area to be used in the circuit, so reducing CbC.
However, J cannot be increased indefinitely because of the
Kirk Effect.

If the collector doping N0 is chosen so that the collector is
fully depleted at zero bias current and the applied VCE the maximum current density through the emitter junction is
approximately given by

\[ J_{\text{max}} = 4eV_{\text{sat}}V_{\text{CE}}T_{\text{C}}^2 \]

where ε is the dielectric constant of the collector, Tc is the
collector thickness, and Vsat the saturated emitter velocity in
the collector [1]. The minimum time required (τ) by the
transistor to switch the output voltage by ΔV can be approximated by [1]

\[ \tau = \frac{C_{bb} \Delta V}{I_0} = \frac{A_C \Delta V}{A_e V_{\text{sat}} V_{\text{CE}}} \]

where I0 is the switched current and is the product of Jmax ,
given by (1) and the area of the emitter (Ae). To obtain the last
term of (2), the collectors-base junction area is assumed to be
Ae. Although the performance of digital circuits requires
additional terms [1] and depends significantly on the layout of the
circuit, (2) is the dominant term in InP HBT digital
circuits. By reducing the collector thickness the switching time can
de be decreased.

III. Thermal Model

The temperature distribution within a device has been
determined by using a commercial finite element solver (ANSYS) to solve 3D the heat diffusion equation. In 3D
problems, there is a requirement to use symmetry to reduce the
number of nodes so that storage limitations are not exceeded
and the computational times are acceptable. There is a natural
plane of symmetry along the length of the emitter and if the
base contact pad is ignored there is another orthogonal plane
of symmetry. The collector of a DHB T normally consists of an
InGaAs set back layer, and a grading layer, followed by InP
collector. As the electrons leave the base they experience a
large electric field that accelerates them. In this region close to
the base, the electrons move quasi-ballistically with little
scattering and so there will be little dissipation of heat. In this
model, the ballistic transport region is assumed to extend over
the whole of the setback and grading layer.

Thermal conductivity (k) of semiconductors decreases
when the temperature is increased and can be modeled using
the equation:

\[ k_T = k_{100}(300/T)^n \]

The values of the room temperature k and the exponential
term (n), used in this work are shown in table 1.

In GaAs, k reduces with heavy doping and a similar effect
will also occur in InP and InGaAs. Unfortunately there is no
experimental data for InP to incorporate into the model so k of
Fig. 1. The simulated device structure used for the verification of the thermal model. (Symmetry allows quarter of the device to be used). All measurements are in \( \mu m \). In order to show the detail, the diagram is not drawn to scale. The thickness of the collector, base and emitter metallisation are 0.3\( \mu m \), 0.1\( \mu m \) and 0.4\( \mu m \). The layer structure is given in the Table II.

Fig. 2. The physical device structure used in the assessment of the thermal performance of a current switch. Two lines of symmetry are assumed: one midway between the devices and the second through the middle of the devices so that only half of the emitter is simulated. V marks a thermal via connecting the emitter interconnects with the InP substrate. All dimensions are in \( \mu m \).

<table>
<thead>
<tr>
<th>Material</th>
<th>Values used ( (Wm^{-1}K^{-1}) )</th>
<th>Experimental Range ( (Wm^{-1}K^{-1}) )</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>InP</td>
<td>0.68</td>
<td>1.42</td>
<td>[11]</td>
</tr>
<tr>
<td>InGaAs</td>
<td>0.048</td>
<td>1.375</td>
<td>[12]</td>
</tr>
<tr>
<td>Au</td>
<td>3.17</td>
<td>3.17</td>
<td>[13]</td>
</tr>
</tbody>
</table>

\( 1 \) of gold was assumed to be temperature independent.

Table II

<table>
<thead>
<tr>
<th>Layer Dimensions</th>
<th>Label</th>
<th>Length (( \mu m ))</th>
<th>Sub-layer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub</td>
<td>500</td>
<td></td>
<td>Substrate</td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>0.20</td>
<td></td>
<td>InP sub collector</td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>0.05</td>
<td></td>
<td>Etch stop layer</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Temperature profile inside the simulated device when the power dissipation was 10mW. The lower surface of the substrate was fixed at 300K. The contours are shown as alternating light and dark gray. The collector is the hottest part of the device.

Fig. 4. Temperature profile through the center and edge of the device. The different parts of the device are shown at the bottom of the figure and the labels are given in Table II.

<table>
<thead>
<tr>
<th>Collector Thickness (( \mu m ))</th>
<th>Length of Emitter (( \mu m ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>0.15</td>
<td>0.3</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>0.9</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Fig. 5. The variation of junction temperature (circles), maximum device temperature (squares) and emitter length (diamonds) with the collector thickness. (This is the measured reduction in GaAs).

The heat flow in the emitter and emitter metalisation will be vertical before flowing in a horizontal direction in the emitter interconnect. Rather than model the layers making up the emitter and emitter metalisation separately a composite \( k \) was used. This can be founded by summing the thermal resistances of each layer. The collector and base metalisations

161
were also multilayer, however the heat flow in these layers is not significant and so these were modeled using $k$ of gold.

Two different device structures have been simulated. The first, shown in fig 1, represents the test transistors used in our laboratory and allows the verification of the simulation results with experiments. The layer structure and the device dimensions are shown in table II. The device is passivated using a polyimide, which has a very low $k$, and so heat flow through the polyimide is assumed to be zero.

High-speed digital circuits have either emitter coupled or current mode topology. In both of these a bias current is switched between two transistors of a differential pair. These transistors need to be in close proximity to each other for high speed. The second structure models one of the devices in such a differential pair. Under normal operating conditions the heat dissipation in each device is the same and so there is a line of symmetry, modeled using an adiabatic plane midway between the devices. The simulated structure is shown in fig 2.

### IV. Verification

The thermal model was verified by comparing the measured and simulated thermal resistance of devices. The temperature profile through the device is shown in fig 3. To obtain this plot the input power was set to 10mW, which was the same power used in the experimental determination of the thermal resistance and so a direct comparison between the two can be made. As expected, there is significant temperature variation through the device. In the center of the device the emitter-base junction temperature rise is 21.0K, but at the edge of the emitter the increase in junction temperature rise is higher (29.5K). This temperature difference is caused by the limited size of the emitter interconnect. The average temperature rise of the junction is 26.2K, which corresponds to a thermal resistance of 2620K/W. This is in good agreement with experimental value (2600K/W). The collector is the hottest part of the device with a temperature 60% higher than the average temperature of the emitter-base junction. The temperature rise will increase the transit time of the electrons traveling through the collector because of the reduction in the saturated velocity ($v_{sat}$). (For a 50K temperature rise, $v_{sat}$ will reduce by 10%. [14].) From (1), this change in $v_{sat}$ will also reduce the Kirk current density limit by the same amount.

In fig 4, the temperature through the center of the device and at the edge is shown. There is a large temperature gradient on both sides of the collector caused by the poor $k$ of the InGaAs used in the base and the sub collector etch stop layers. Reduction of the thickness of the InGaAs etch stop layers will significantly reduce the temperature of the collector. This will be discussed in section V.

As the device speed increases the thickness of the base decreases and the control of the base metalisation diffusion becomes important both during manufacture (yield) and during operation (reliability). Diffusion rate increases with temperature. Therefore if the temperature of the base metallisation is higher than expected, this may cause the lifetime of the devices to be below the expected value.

In some parts of the device, especially near the ends, the temperature of the base metalisation is higher than the average emitter—base junction temperature. For reasons given above this will have an adverse effect on reliability. Improving the heat sinking of the emitter at the edge of the device by increasing the width of the emitter interconnects will reduce the maximum temperature of the base metallisation and so improve the reliability.

In fig 5, the junction and maximum temperatures of the simulated device are presented. The structure of the device was the similar to that shown in fig 1 with the exception of the width of the emitter interconnection. To generate a $\Delta T$ of 300mV across a 50Ω load impedance requires a 6mA bias current. In generating fig 6, the current flowing through the device is assumed to be 6mA with a 50% switching duty cycle. Furthermore, when the device was conducting, $V_{CE}$ was assumed to be 1V. This equates to an average power dissipation of 3mW. The length of the emitter was chosen so that the device was operating at the Kirk limit given by (1), and the emitter junction width (fig 1) remains constant at 0.5 $\mu$m. The length is also given in fig 5. The difference between the junction temperature and the maximum temperature also increases. At a collector thickness of 1000A the maximum temperature rise is nearly 60K. For a 100% duty cycle the temperature rise would be approximately 120K. Clearly this temperature rise is unacceptable. The increase in the width of emitter interconnect also reduces the temperature variation across the device.

### V. Devices for 80 and 160Gbit/s applications

Results of the previous section show that the operating temperature of the device rapidly increases as the collector thickness is reduced, given that the device is operated at Kirk-effect-limited current density. It is therefore important to consider the thermal performance of devices suitably scaled for use in 80 and 160Gbit/s applications. Before these simulation can be performed the key physical parameters for 80 and 160 Gbit/s IC's need to be examined. Master-slave latch speed is an important figure of merit because these circuits are heavily used in fiber transceivers. In conservative system-level designs the maximum clock rate is 1.5 times the operating bit rate. To determine the physical parameters of the devices, stated in table III, the methodology of [1] was used. The base contact resistance and base sheet resistivity are all consistent with values obtained in our laboratory. [2]. The 160 Gbit/s devices have a very small base-emitter separation as well as a small base mesa. Devices with dimensions similar to the proposed devices have already been manufactured in the GaAs system. (e.g. [15]). Note that the current densities are 1, 2.3 and 9.8 mA/$\mu$m² at the three bit rates and this drives the
thermal design.

Table IV gives the temperatures at the base-emitter junction and the maximum temperature with and without the collector etch stop layer. In calculating the device temperatures, the same assumptions as in the previous section were used and so the total power dissipated in the device was set to 3mW. The temperature rise for the 40 and 80 Gbit/s devices is clearly within acceptable bounds even when the etch stop layer was used. The 80Gbit/s device would benefit thermally by a reduction in the thickness of the etch stop layer. With the etch stop layer the temperature rise in the 160Gbit/s device is 49K and 23% of the heat flows through the emitter. Under idle conditions the dissipated power will be doubled and to the first approximation the temperature rise will be 98K. This value is too high for long-term operation of the device. With no etch stop layer the maximum temperature rise within the device is significantly lower (20.5K). Under these conditions the flow of heat out of the emitter is reduced to only 0.3mW (i.e. 10% of the total). Under idle conditions, the temperature rise becomes 41K, which is acceptable. Note, the etch stop layer is used to control the base mesa process and so to get a device operating at these low temperatures will require the development of an alternative base mesa process or etch stop layer. The InGaAs etch stop layer also provides an additional benefit. The resistance of contacts on n+ InGaAs is smaller than on n+ InP. Therefore, the removal of the etch stop layer will increase the collector resistance and the effect of this on circuit performance needs to be considered.

VI. Conclusions

An experimentally validated thermal model of InP HBT has been presented and has been used to demonstrate that operation of devices for 80 and 160Gbit/s applications are thermally possible. To achieve this, the etch stop layer in the sub-collector needs to be removed.

The model also shows that the maximum temperature within the device is higher than the measured value at the emitter-base junction. This observation needs to be taken into account when the reliability of devices is considered.

Acknowledgements

The authors would like to acknowledge Walsin Corp. and the University of California CORE program for part funding this work. I.Harrison would also like to thank the Royal Academy of Engineering for an Engineering Foresight Award.

References


163