7-8-2015

InAs/GaSb Tunnel Diodes

Matthew J. Filmer

InAs/GaSb Tunnel Diodes

Matthew J. Filmer

InAs/GaSb Tunnel Diodes

Matthew J. Filmer
July 8, 2015

A Thesis Submitted
in Partial Fulfillment
of the Requirements for the Degree of
Master of Science
in
Microelectronic Engineering

R·I·T | Kate Gleason
College of ENGINEERING

Department of Electrical and Microelectronic Engineering

InAs/GaSb Tunnel Diodes

Matthew J. Filmer

Committee Approval:

Dr. Sean Rommel *Advisor* Date
Associate Professor

Dr. Karl Hirschman Date
Professor

Dr. Santosh Kurinec Date
Professor

Dr. Robert Pearson Date
Program Director

Dr. Sohail Dianat Date
Department Head

Acknowledgments

There are many people who have helped along the way with this project. Most importantly is my advisor, Professor Sean Rommel. His guidance and support were invaluable in the completion of this project. I would also like to thank my committee members, Professor Karl Hirschman and Professor Santosh Kurinec, for the guidance they have likewise provided.

Members of Professor Rommel's research group have also been of great aid in this project. Brian Romanczyk for an initial introduction to III-V device fabrication, as well as Paul Thomas and Abhinav Gaur for continued support throughout the project.

Assistance in the design of the RITD structures was provided by Kausik Majumdar of SEMATECH. Growth of these structures, as well as the HTD structures, was done at Texas State University by Ravi Droopad. Without them this project would not have been possible.

There are also many people working for the SMFL who have greatly assisted with this project. Sean O'Brien and Patricia Meller for tool training and certification. As well as John Nash, Richard Battaglia, Bruce Tolleson and David Yackoff for maintaining the tools that were crucial for the success of this project.

This work was funded in part by the national science foundation grant number ECCS-1202054.

Abstract

The Tunneling Field-Effect Transistor (TFET) has shown promise as a possible replacement for the MOSFET, especially in low power applications where power supplies continue to shrink. One of the biggest developmental challenges with the TFET is the relatively low drive current density compared to that of a MOSFET. Increasing tunneling probability will increase drive current density. Determining the peak current density in a tunnel diode can give insight into the tunneling probability for the material system. This can then be used to predict what material systems might support high current density TFETs.

This study focuses on the InAs/GaSb broken gap system. Heterojunction Tunnel Diodes (HTDs) and Resonant Interband Tunnel Diodes (RITDs) were fabricated from Molecular Beam Epitaxy (MBE) grown substrates using varying doping concentrations, including undoped. The HTD devices were designed to investigate the effect doping has on peak current density. To determine a minimum current density, an undoped HTD was fabricated. The RITDs were designed to investigate an alternative structure and determine if it could be used to further improve peak current density.

This study has shown that changes in dopant concentration only have a significant impact on tunneling current when the dopants are near the junction. No differences in current density was observed between the four HTD samples. The undoped sample exhibited an average peak current density of 16 kA/cm^2. The RITDs samples did not show any significant differences in peak current density with varying well width. Additionally, the control HTD device performed significantly better, in terms of peak current density, than any of the RITD samples.

Contents

List of Figures

List of Tables

List of Symbols

Term	Description	Units/Value
ϕ_n	Fermi potential from the conduction band edge in an n-type material	V
ϕ_p	Fermi potential from the valence band edge in a p-type material	V
χ	Electron affinity	V
a_x, a_y	Alignment offsets in the x, y direction	mm
E_C	Bottom edge of conduction band	eV
ΔE_C	Conduction band offset of a heterojunction	eV
E_F	Fermi level	eV
E_g	Band gap energy	eV
E_{Fn}, E_{Fp}	Quasi-Fermi level for electrons, holes	eV
E_V	Top edge of valence band	eV
ΔE_V	Valence band offset of a heterojunction	eV
E_{Vac}	Vacuum energy level	eV
$f_F(E)$	Fermi-Dirac distribution	
g_c, g_v	Density of states in the conduction, valence band	$\text{cm}^{-3}\text{eV}^{-1}$
I	Current	A
I_P, I_V	Peak, Valley current	A
J	Current density	A/cm^2
J_P, J_V	Peak, Valley current density	A/cm^2
J_d, J_t, J_x	Diffusion, Tunneling, Excess current density	A/cm^2
L	Length	mm
n	Electron carrier concentration	cm^{-3}
n_x, n_y	Number of die in a row, column	count
N_A, N_D	Acceptor, Donor concentration	cm^{-3}
N_C, N_V	Effective density of states in the conduction, valence band	cm^{-3}
p	Hole carrier concentration	cm^{-3}
S	Subthreshold swing	V/decade of current
s_x, s_y	Die step distance in the x, y direction	mm
V	Voltage	V

Term	Description	Units/Value
V_A	Applied voltage	V
V_{bi}	Built-in voltage	V
V_P, V_V	Peak, Valley voltage	V
W	Width	mm

Chapter 1

Introduction

It is desirable to scale down MOSFETs in order to increase the number of transistors on a chip, as well as reduce power consumption of that chip. Given the load capacitance C, the operating frequency f, and the supply voltage, V, power consumption in a MOSFET is proportional to CfV^2 [1]. It is unlikely that clock frequencies are going to be reduced, and the ability to change load capacitance is minimal. Therefore, it is of critical importance to reduce the supply voltage in order to reduce power use in CMOS circuits.

Despite the desire to reduce supply voltages, it is becoming difficult to do so. In order for a MOSFET to be effective, it must support high drive currents. This has historically not been much of an issue in digital cirtuits, but with the trend toward low supply voltages it has become one. Now, drive currents are becoming limited by the subthreshold swing (S), and are not high enough to meet the needs when supply voltages are dropped below ~ 0.5 V, and a new design is needed [2].

The fundamental limit of subthreshold swing in a MOSFET comes from the method a MOSFET uses to control current. Current can flow through a MOSFET when energetic carriers in the source can, through thermionic emission, travel over the barrier that is the gate. It is the process of thermionic emission, and the shape of the band tails, that limit S [3].

The subthreshold swing limit in a MOSFET is generally considered to be as given

in (1.1), or about 60 mV/dec at 300 K [1].

$$S(T) = \frac{kT}{q} \ln 10 \tag{1.1}$$

The Tunneling Field-Effect Transistor (TFET) does not control current by controlling thermionic emission through the adjustment of a barrier height, instead a tunneling current is controlled. This current is made up of carriers located near the band edges rather than the band tails. The result is a considerably lower subthreshold swing limit. Using the Zener tunneling current (1.2) [4], a theoretical representation of the subthreshold swing of a TFET can be calculated: (1.3) [5].

$$I = aV_{\text{eff}}\xi \exp(-\frac{b}{\xi}) \tag{1.2}$$

$$S = \ln 10 \left[\frac{1}{V_{\text{eff}}} \frac{dV_{\text{eff}}}{dV_{GS}} + \frac{\xi + b}{\xi^2} \frac{d\xi}{dV_{GS}} \right]^{-1} \tag{1.3}$$

In these equations, V_{eff} is the bias applied at the tunnel junction, and ξ is the electric field. The terms a and b are coefficients determined from the properties of the materials (1.4a) and (1.4b). This equation allows for a TFET subthreshold swing lower than 30 mV/dec [5].

$$a = Aq^3 \sqrt{2m^*/E_g}/4\pi^2\hbar^2 \tag{1.4a}$$

$$b = 4\sqrt{m^*}E_g^{3/2}/3q\hbar \tag{1.4b}$$

The current TFETs cannot, however, be used as a drop-in replacement for the MOSFET in digital circuits. While they may have a low subthreshold swing, today's TFETs suffer from low drive current [6]. Therefore, work must be done to improve this current before the TFET becomes a viable MOSFET replacement. This work investigates the performance of broken gap Heterojunction Tunnel Diodes (HTDs) in

order to find possible TFET design improvements.

1.1 Document Structure

Chapter 2 discusses background information relating to this work. This includes a discussion of heterojunctions, as well as tunnel diode structure, the I-V characteristic, as well as HTDs and Resonant Interband Tunnel Diodes (RITDs). Information regarding the design of the diodes fabricated here, as well as the fabrication process is found in Chapter 3. This includes the mask layout, how the devices were tested, and detailed process steps that should allow another researcher to fabricate similar devices if desired.

Chapter 4 covers the work that was done to improve the optical lithography process on pieces. This includes a detailed explanation of the process for exposing a piece considering the shape of the piece, as well as alignment for the first, as well as subsequent, levels. Also included are mask design considerations as they relate to exposing a piece.

Finally, Chapters 5 and 6 discuss the RITD and HTD experiments, respectively. Each of these chapters includes the layer structures used, the expected results, as well as the actual results. All of this is then summarized in Chapter 7.

Chapter 2

Theory

2.1 Materials

Electrons in the conduction band and holes in the valence band do not have energies equal to the energy of their respective band edge. Instead, there is a distribution of energies. In this distribution, the lower limit of energy is the conduction band edge for electrons in the conduction band, and the valence band edge for holes in the valence band.

As the semiconductors in question are 3D semiconductors, a 3D density of states must be used. The 3D density of states is the maximum number of electrons in the conduction band, or holes in the valence band, that can exist at a given energy per volume. The density of states can be calculated from (2.1), where E is the energy from either the conduction band edge, in the case of electrons (2.1a), or valence band edge, in the case of holes (2.1b) [7].

$$g_c(E) = \frac{4\pi(2m_n^*)^{3/2}}{h^3}\sqrt{E - E_C} \tag{2.1a}$$

$$g_v(E) = \frac{4\pi(2m_p^*)^{3/2}}{h^3}\sqrt{E_V - E} \tag{2.1b}$$

Electrons are fermions, and therefore follow the Fermi-Dirac distribution when

looking at the probability of a given energy level being occupied. This distribution, as it applies in doped semiconductors, is provided in (2.2).

$$f_F(E) = \left[1 + \exp\left(\frac{E - E_F}{kT}\right)\right]^{-1} \tag{2.2}$$

Therefore, the electron energy distribution, $n(E)$ is simply the product of the occupancy probability, given as the Fermi-Dirac distribution, $f_F(E)$, and the conduction band density of states, $g_c(E)$. Similarly, the hole energy distribution in the valence band is the product of the vacancy probability and the valence band density of states, $g_v(E)$. These relationships are given in (2.3) [7].

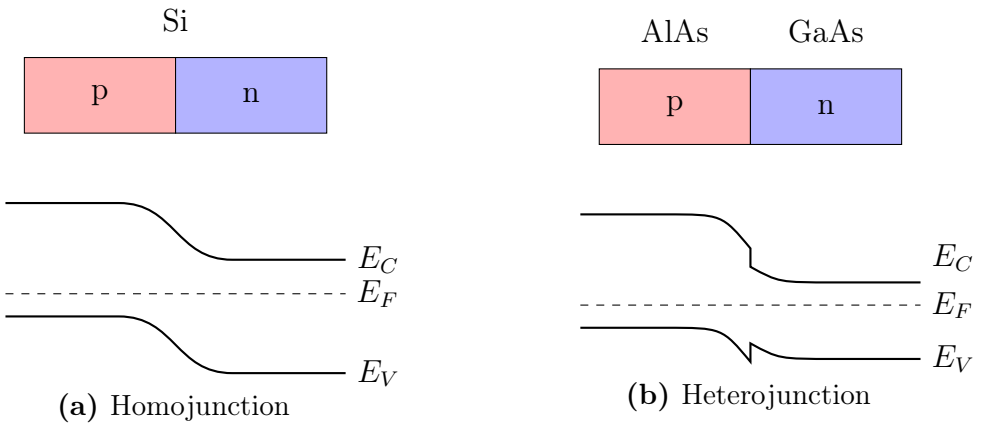
$$n(E) = g_c(E) f_F(E) \tag{2.3a}$$

$$p(E) = g_v(E)[1 - f_F(E)] \tag{2.3b}$$

2.1.1 Heterojunctions

One of the most basic building blocks for any semiconductor device is the p-n junction as shown in Figure 2.1a. A homojunction is characterized by a constant band gap, and has different doping concentrations on either side of the junction. A heterojunction however, consists of two different materials, each with a different band gap energy. An example heterojunction consisting of AlAs and GaAs is shown in Figure 2.1b. The difference in band gap energy causes a step in one or both of the band edges at the junction.

Heterojunctions fall into one of three categories, depending on the alignment of the band gap regions. These categories are straddling gap, staggered gap and broken gap, or Types I, II and III respectively. Example band diagrams for each type of band alignment is shown in Figure 2.2. The relative alignment of the energy gap is related



(a) Homojunction **(b)** Heterojunction

Figure 2.1: Schematic energy band diagrams for: **(a)** Si p-n homojunction and **(b)** AlAs/GaSb heterojunction.

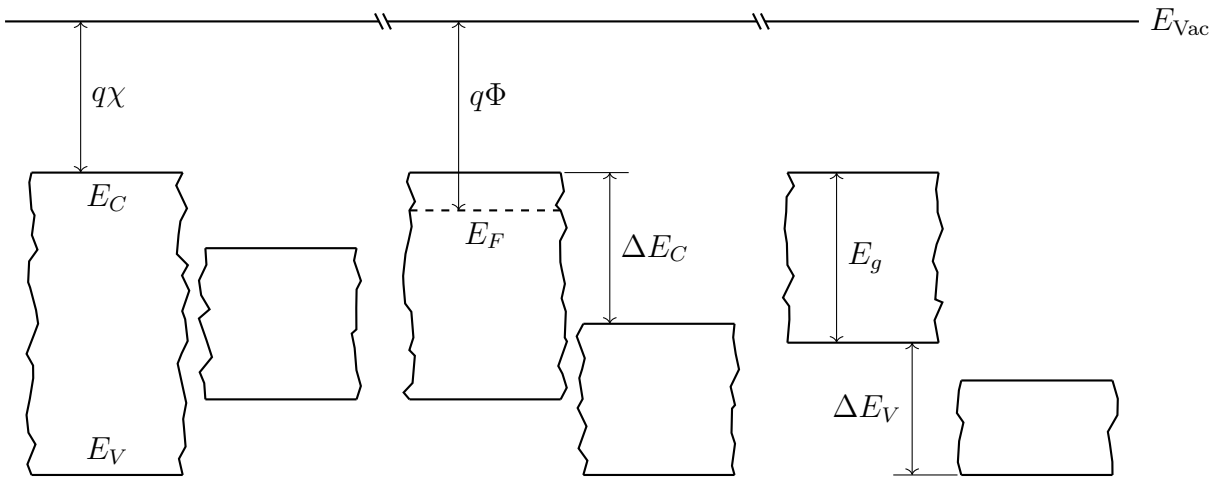


Figure 2.2: The three types of heterojunction band alignment. In straddling gap (left) one band gap is entirely contained within the other. In a staggered gap system (center) there is partial overlap between the band gaps of each material. In the broken gap system there is no overlap between the two forbidden regions.

to the electron affinity (χ) of each material, which is the energy required to move an electron from the bottom of the conduction band to the vacuum level.

For a straddling gap system, the forbidden region of one material is entirely contained within the forbidden region of the other. Examples of this type are GaAs/AlAs and GaSb/AlSb heterojunctions.

In a broken gap system there is no overlap between the forbidden region of the two semiconductors. A consequence of this, is some overlap between either the conduction bands of the two materials, or the valence bands of the materials. The InAs/GaSb heterojunction is an example of this type of band alignment.

A staggered gap system, is one where there is partial overlap between the forbidden regions of each material. If they were completely overlapped it would instead be a straddling gap system, and if there was no overlap it would instead be a broken gap system. An example of this type of junction is InAs/AlSb.

The band gap energy, as well as the band alignment, of various III-V materials are shown in Figure 2.3. Note that any heterojunctions formed from the materials presented here in either the 5.65 Å or the 5.87 Å systems would be straddling gap heterojunctions. However, any of the three heterojunction types could be formed using the materials is the 6.1 Åsystem.

When dealing with heterojunctions there are a two parameters of importance, the conduction band and valence band offsets or ΔE_C and ΔE_V. As defined in (2.4), the conduction band offset is the magnitude of the difference in energies of the two conduction band edges. Likewise, the valence band offset is the magnitude of the difference in energies of the two valence band edges. These parameters are useful for designing quantum wells, as they will directly control the depth of the well.

$$\Delta E_C \equiv |q\chi_1 - q\chi_2| \tag{2.4a}$$

$$\Delta E_V \equiv |(q\chi + E_g)_1 - (q\chi + E_g)_2| \tag{2.4b}$$

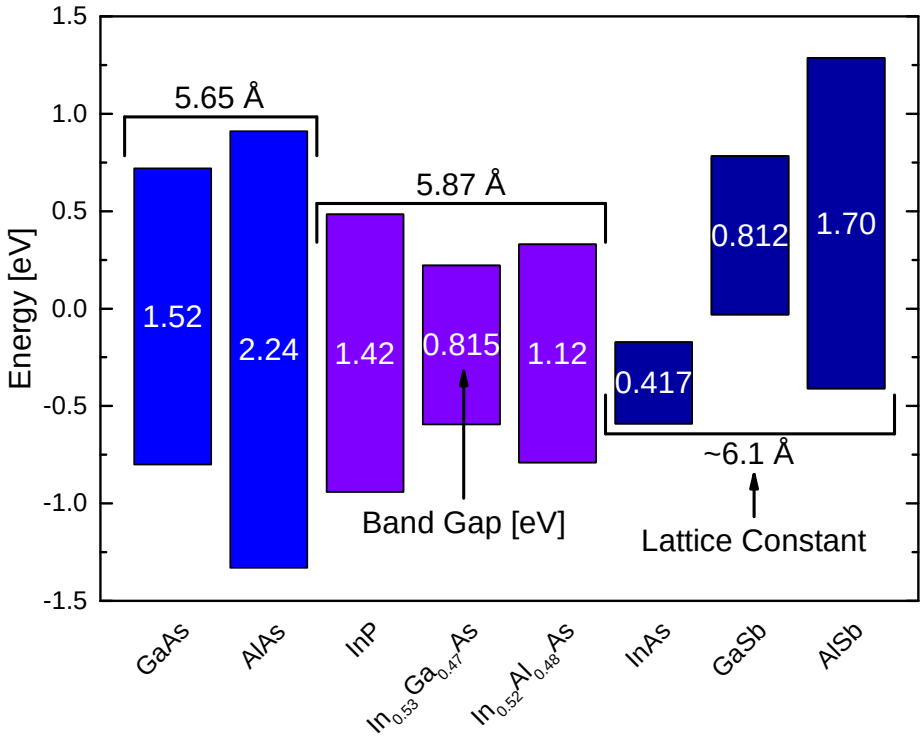


Figure 2.3: Band alignment of various III-V binary and ternary semiconductors, grouped by approximate lattice constant. Materials of similar lattice constant are usually used together. This study deals exclusively with the ~ 6.1 Å system.

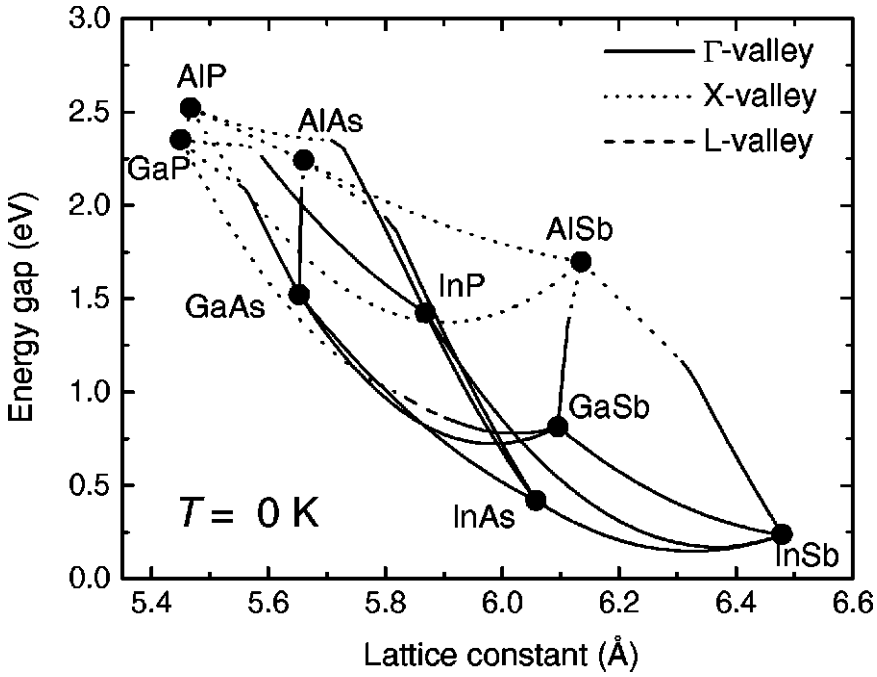


Figure 2.4: A plot of various III-V binary semiconductors band gap energy and lattice constant. The curves between each material correspond to a ternary semiconductor with varying material ratios.[1] [8]

Heterostructures are often designed using a lattice matched material system. Such a system is a set of materials with the same lattice constant. If the lattice constants differ by too much, the lattice will be under stress, or will relax, creating defects at the junction which will likely degrade device performance. A chart, such as the one shown in Figure 2.4, can be very useful in determining what materials are available for a given lattice constant.

The materials used in this study are exclusively from the ~ 6.1 Å, approximately lattice matched, InAs/GaSb/AlSb system. This system was chosen for this study because it contains the materials necessary to create a broken gap junction (InAs and GaSb), as well as a wide band gap material to create a barrier (AlSb).

[1]Reprinted with permission from I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, Journal of Applied Physics, Vol. 89, Page 5815, (2001). Copyright 2001, AIP Publishing LLC.

2.2 Tunnel Diodes

A tunnel diode, or an Esaki diode, is a two terminal device that exhibits Negative Differential Resistance (NDR) under forward bias, and is conductive in reverse bias. Tunnel diodes were originally discovered by Leo Esaki in 1957 and were the first devices found to exhibit NDR [9]. A device is said to exhibit NDR when an increase in voltage causes a decrease in current, or dI/dV is negative.

2.2.1 Structure

The structure of a tunnel diode is nearly identical to that of a typical p-n junction diode. The critical difference between these two devices is degenerate doping of both sides of the junction in the tunnel diode, and non-degenerate doping of the same in the standard diode. By definition, this means the Fermi level on the n side is in the conduction band, and the Fermi level on the p side is in the valence band. This is in contrast to the typical p-n junction diode, where the Fermi level is somewhere in the forbidden region, between the conduction and valence bands. An important consequence of these doping levels is an overlap of the conduction and valence bands in the tunnel diode at equilibrium.

2.2.2 Current-Voltage Characteristic

The current-voltage characteristic of a tunnel diode is unlike that of conventional diodes. Instead, they are highly conductive in reverse bias and exhibit NDR in forward bias. An example I-V characteristic for a typical tunnel diode is provided in Figure 2.5. Note the peak and the valley regions shown on this plot. The peak current, I_P, is the maximum current at the peak, and likewise, the valley current, I_V, is the minimum current in the valley. The peak and valley voltages, V_P and V_V are the voltages at which the peak and valley current occur. NDR occurs in the entire

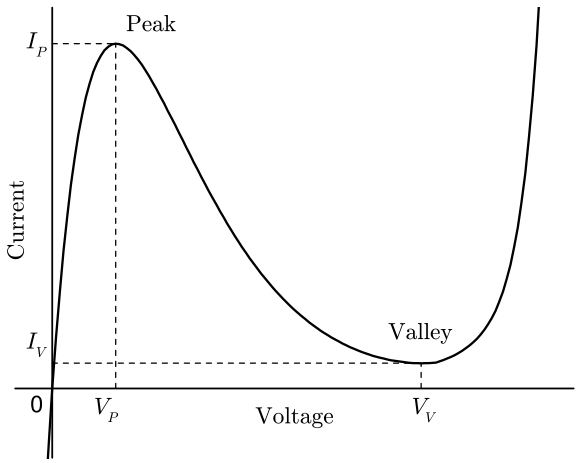


Figure 2.5: Example I-V characteristic of a tunnel diode. The plot indicates the peak and valley regions, as well as the corresponding peak/valley current (I_P/I_V) and peak/valley voltage (V_P/V_V).

region between the peak and the valley, and does not occur outside this region.

An important parameter of a tunnel diode, is the Peak to Valley Current Ratio (PVCR). As it sounds, this is simply the ratio of the peak and valley currents, as shown in (2.5). Since both the peak and valley currents should be proportional to area, PVCR can also be written as the ratio of the peak and valley current densities, J_P and J_V.

$$\text{PVCR} = \frac{I_P}{I_V} = \frac{J_P}{J_V} \tag{2.5}$$

2.2.3 Band Structure

The energy band diagram of a tunnel diode looks very similar to that of a typical p-n junction diode. An example band diagram for a typical tunnel diode in thermal equilibrium is given in Figure 2.6. In a tunnel diode, both the n and p sides of the junction are degenerately doped. A consequence of this is negative Fermi potentials

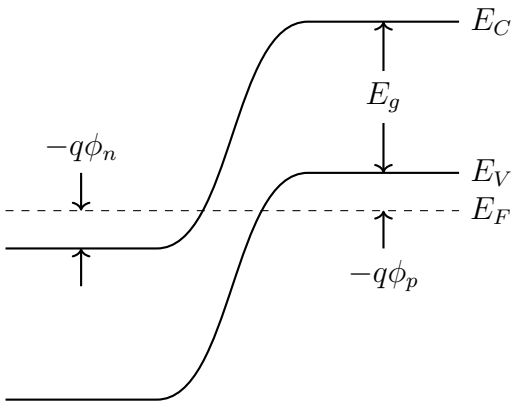


Figure 2.6: Energy band diagram of a typical tunnel diode at thermal equilibrium.

ϕ_n and ϕ_p, where the Fermi potentials are defined as in (2.6).

$$q\phi_n = E_C - E_F \tag{2.6a}$$

$$q\phi_p = E_F - E_V \tag{2.6b}$$

To understand the current-voltage characteristic of a tunnel diode, we first look at band diagrams for different applied biases, as given in Figure 2.7. For each of the four cases shown here, a corresponding point is shown on the I-V curve (Figure 2.7e).

The first case, Figure 2.7a, shows a tunnel diode under reverse bias. As a negative voltage is applied to the p side of the junction relative to the n side there is an increase in the overlap between filled states in the valence band on the p side, and empty states in the conduction band on the n side. When this happens, Zener tunneling, or Band to Band Tunneling (BTBT), can occur. The overlap continues to increase with increasing reverse bias, accounting for the tunnel diode's conductivity under reverse bias.

The second case is for a forward bias less than $-(\phi_n + \phi_p)$, and is shown in Figure 2.7b. Under these conditions, there is overlap between filled states in the conduction band of the n region, and empty states in the valence band of the p region.

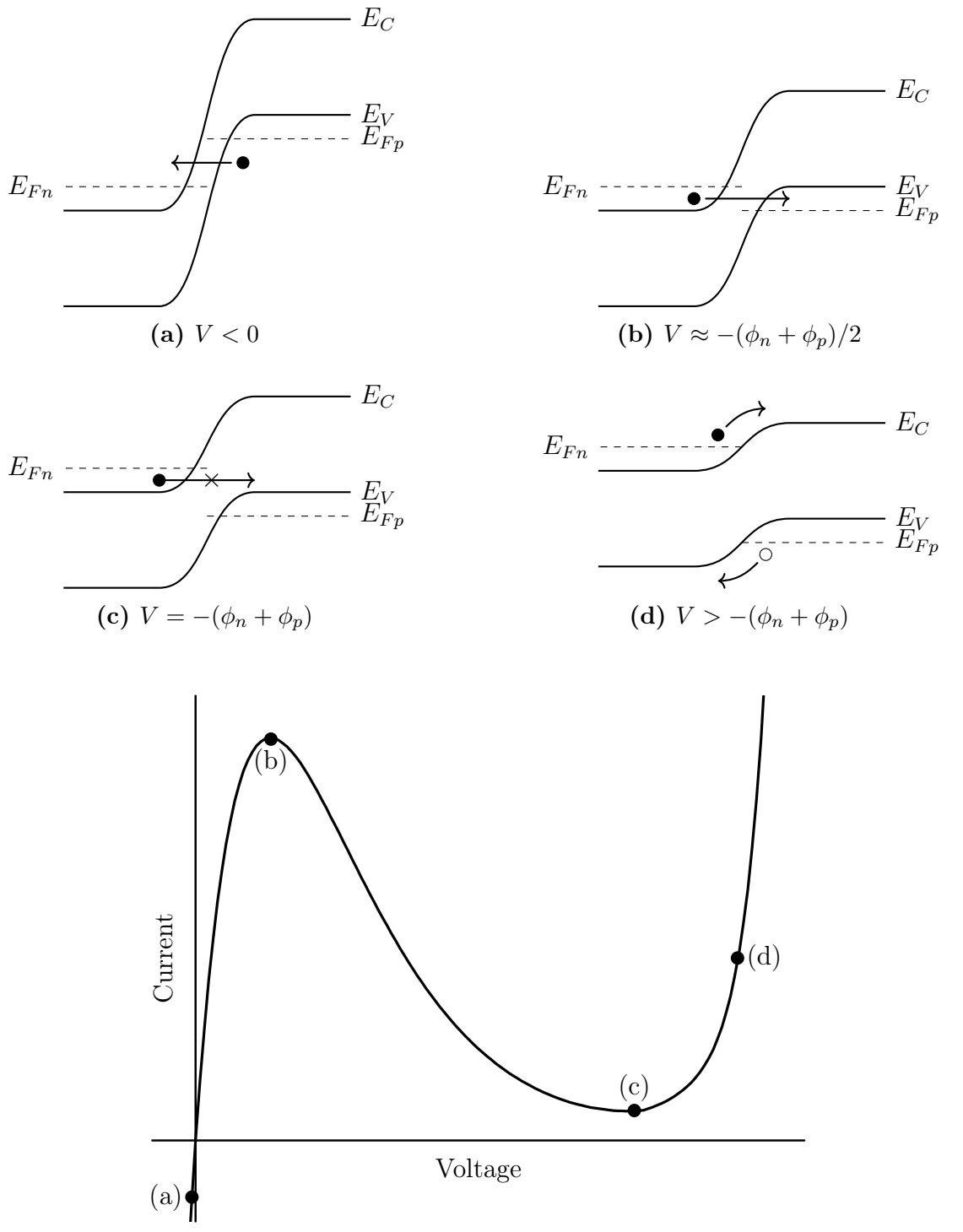


Figure 2.7: Tunnel diode energy band diagrams under **(a)** reverse bias, enabling reverse tunneling current, **(b)** small forward bias, where forward tunneling current begins, **(c)** forward bias, where the bands are no longer aligned and therefore there can be no tunneling current, **(d)** large forward, bias where diffusion current becomes significant. **(e)** indicates where these band diagrams fall on the I-V curve.

When this is the case, tunneling can occur. The example band diagram shown here is for $V = -(\phi_n + \phi_p)/2$, which is approximately where the peak current occurs.

The third case is for a bias greater than $-(\phi_n + \phi_p)$, but not high enough for significant forward diode current. Shown in Figure 2.7c is the band diagram for $V = -(\phi_n + \phi_p)$. Here, there is no longer an overlap between the conduction and valence bands, and therefore direct BTBT can not occur. The only current here is excess current, and a very small amount of diffusion current from thermionic emission. At this bias there is no longer any tunneling current, and since excess current and diffusion current increase with an increasing V, there can be no voltage higher than $_(\phi_n + \phi_p)$ that results in a lower current. Therefore, the valley current occurs at a slightly lower bias than this, where:

$$-\frac{dI_t}{dV} = \frac{dI_x}{dV} + \frac{dI_d}{dV} \tag{2.7}$$

The final case is for an applied bias high enough to turn the diode on. In this case, the tunnel diode is operating exactly the same as a typical p-n junction diode. As shown in Figure 2.7d, energetic electrons in the conduction band on the n side, and energetic holes in the valence band on the p side, contribute to current.

2.2.3.1 Components of Current in a Tunnel Diode

There are three main components of current in a tunnel diode: tunneling current (J_t), excess current (J_x) and diffusion current (J_d). The contributions of these components are shown in Figure 2.8. The initial increase, followed by a decrease in the tunneling current is responsible for the NDR seen in a tunnel diode. Diffusion current, is responsible for the increase in current at large forward biases. The final component of current in a tunnel diode is excess current. This is the reason the valley current is not equal to the diffusion current, as would be expected when direct BTBT cannot occur.

The current density in a tunnel diode can be modeled as the sum of each of

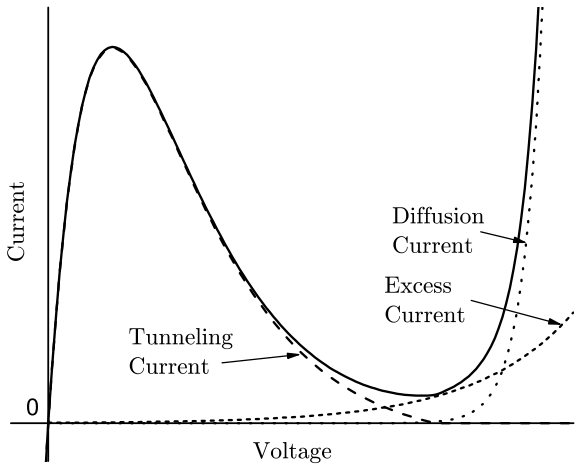


Figure 2.8: Example I-V characteristic of a tunnel diode. This plot shows the components of total current: tunneling current (J_t), excess current (J_x), and diffusion current (J_d). Tunneling current is dominant for 0 V $< V_A < V_V$, excess current is dominant for $V_A \approx V_V$, and diffusion current is dominant for $V_A > V_V$.

the three components of current (2.8d). The tunneling current density (J_t) can be modeled as (2.8a) and the excess current density as (2.8b), where the term C_4 is an experimentally determined fitting factor. Finally, the diffusion current can be modeled using the traditional diode current equation (2.8c) [4].

$$J_t = \frac{J_P V}{V_P} \exp\left(1 - \frac{V}{V_P}\right) \tag{2.8a}$$

$$J_x = J_V \exp(C_4(V - V_V)) \tag{2.8b}$$

$$J_d = J_0 \exp\left(\frac{qV}{kT}\right) \tag{2.8c}$$

$$J = J_t + J_x + J_d \tag{2.8d}$$

Excess current is generally undesired in a tunnel diode as it reduces the PVCR by increasing the valley current. The majority of excess current is a tunneling current [10] and usually comes from BTBT of carriers, through trap or defect states, from the conduction band on the n side, to the valence band on the p side.

2.2.3.2 Origin of NDR in Tunnel Diodes

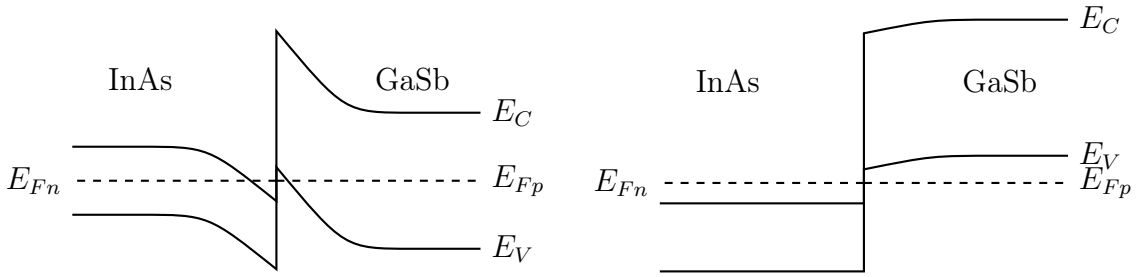
NDR is defined as a negative dI/dV, which is a decrease in current through the device for an increase in voltage across the terminals. In a tunnel diode, this occurs in the region between the peak and the valley.

As the forward bias increases from zero, the current will initially increase. This is due to the increasing overlap of the filled states in the conduction band on the n side, and the empty states in the valence band on the p side. Eventually however, as the bias continues to increase, the overlap will begin to decrease, resulting in a decreasing current. The current will eventually reach a minimum, when the bias equals $-(\phi_n + \phi_p)$. For forward bias greater than this value, there is no longer any direct BTBT current. All current at this point is composed of excess current, and diffusion current.

2.2.4 Heterojunction Tunnel Diode

A similar device to the tunnel diode is the HTD. Like the tunnel diode, the HTD relies on an overlap between the conduction band on one side of the junction, and the valence band on the other side. Example band diagrams for an InAs/GaSb HTD are shown in Figure 2.9. The difference between a normal tunnel diode and an HTD is simply the tunnel diode is composed of a single material while the HTD is made up of two materials.

These structures operate the same way a traditional tunnel diode does. But, because there are two different materials that comprise the junction, there will be some band misalignment. Because of the misalignment between the bands, degenerate doping is not always required to create a tunnel diode and observe NDR. In the extreme case of a broken gap system, such as InAs/GaSb, NDR can be observed without any doping, as the bands already overlap. When these structures are biased such that there is a carrier available to tunnel, and an available energy state to tunnel to, there



(a) Undoped **(b)** Degenerately doped

Figure 2.9: Schematic band edge diagram of **(a)** undoped and **(b)**, degenerately doped InAs/GaSb HTDs. The band misalignment in the undoped structure comes from the band offsets, not doping. Therefore, degenerate doping is not necessary to observe NDR in these devices.

is current. When the bands are no longer appropriately aligned, there is only excess current, and at high enough biases, diffusion current.

2.2.5 Resonant Interband Tunnel Diode

The RITD is a similar structure to the HTD, but instead of consisting of two layers, as does the HTD, it consists of four layers. In the RITD, a thin region of a wide bandgap material is added in between two layers of one of the two materials that make up the junction. This barrier layer is added in order to create a quantum well. For the devices discussed here, the well is in GaSb, formed between AlSb and InAs. The full material stack is InAs/GaSb/AlSb/GaSb. A schematic energy band diagram, is shown in Figure 2.10.

These structures were originally discussed by Ting *et al.* as a way to increase the peak current density of tunnel diodes [11]. At resonance, the reflection coefficient of the junction drops to nearly zero. In comparison, there is significant reflection at all carrier energies in the HTD structure [11]. This reduction in reflection leads to an increased tunneling probability, and therefore increased current.

The peak current density benefits of the RITD structure were demonstrated in papers by Ting *et al.* [11] and Collins *et al.* [12]. A plot of the data presented there

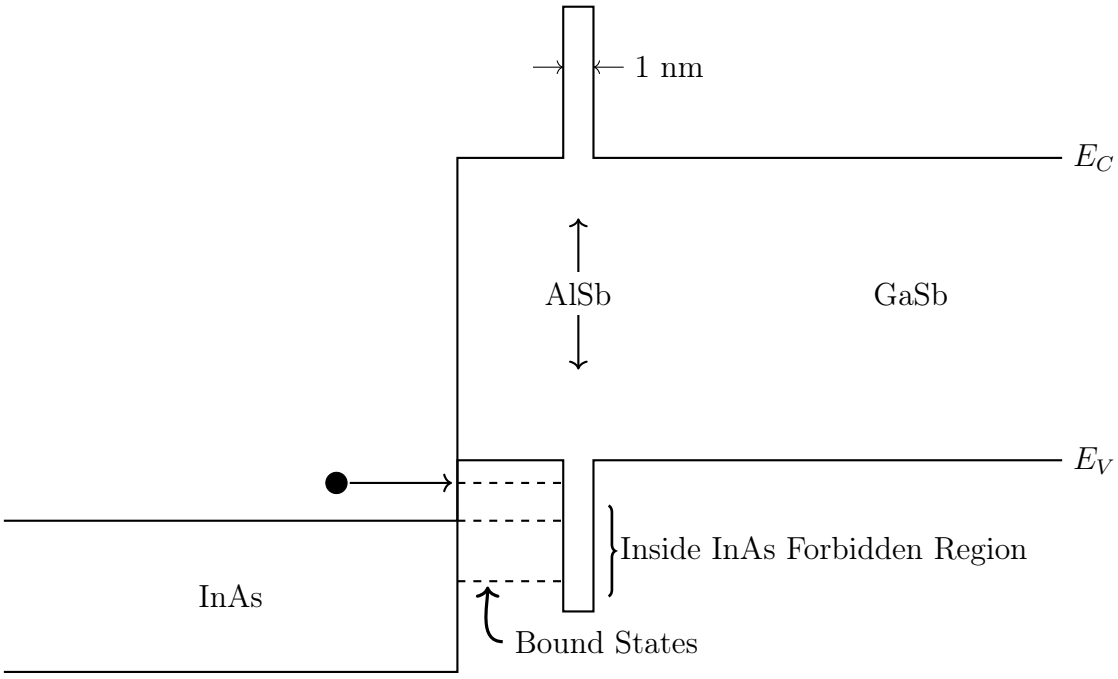


Figure 2.10: Schematic energy band diagram of an RITD. The AlSb barrier creates a quantum well in the GaSb which quantizes energy states. Electrons can tunnel through these states, resulting in reduced reflection at those energy levels. The reflection is increased however, at energy levels not corresponding to a bound state. The number of bound states is a function of the well width and is not necessarily the 3 that is shown. The offset between the AlSb and GaSb valence bands is 0.38 nm.

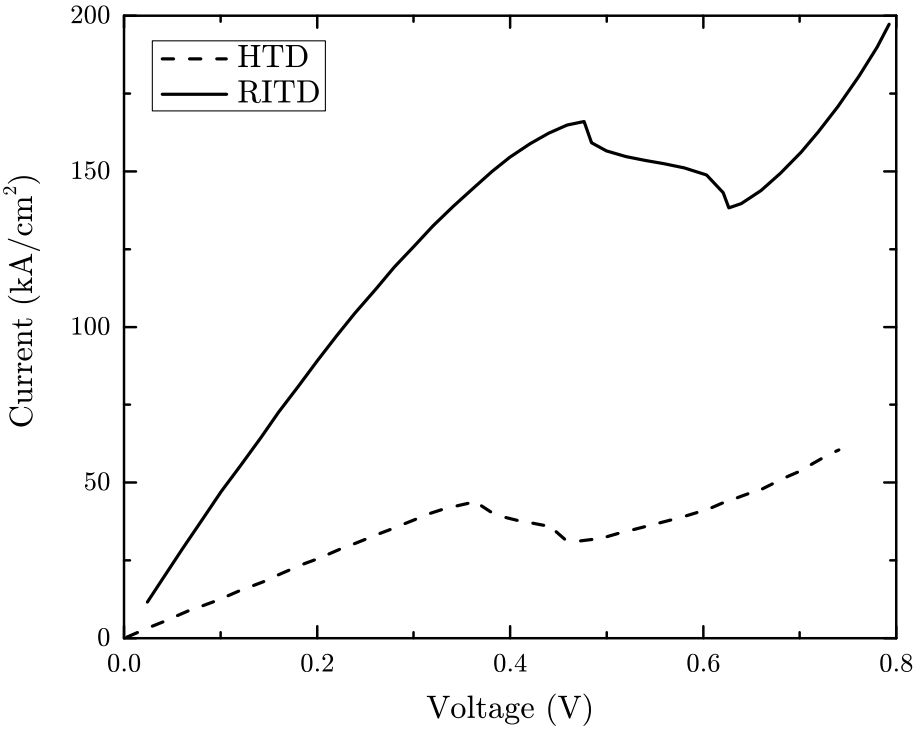


Figure 2.11: An *I-V* plot based on data presented in a paper by Collins *et al.* of an InAs/GaSb HTD and an InAs/GaSb/AlSb/GaSb RITD with a well width of 51 Å and a barrier width of 12 Å [12]. These curves compare the *I-V* characteristic of an InAs/GaSb HTD with that of an InAs/GaSb/AlSb/GaSb RITD. It can be seen that the RITD device exhibits a peak current around 3–4 times that of the HTD device.

is given in Figure 2.11. It can be seen that both the peak current density and the valley current density increased by a factor of ~ 4. Ting and Collins also reported on another HTD with a higher current density, twice that of the HTD shown here. But the RITD they reported still had double the peak current density of that HTD.

Chapter 3

Design and Fabrication

3.1 Design

The diodes fabricated for this study are vertical devices. In vertical devices, current flow is perpendicular to the substrate. The diode layout, shown in Figure 3.1, consists of three main structures: mesa, grid and probe pad.

The mesa is a rectangular area, created by etching away the surrounding top contact metal and n++ layer. The mesa makes up the diode itself, and defines the area through which current can flow.

On the same mask layer as the mesa, is the grid. The grid is, in actuality, a very large area diode covering almost the entire die and surrounding each of the mesas. This structure is used as a contact to the p++ side of the junction.

On top of the mesa, and supported by an Inter-Layer Dielectric (ILD), is the probe pad. This is an additional metal layer which is used to make probing the smaller area diodes possible. This layer, along with the supporting ILD, is optional and serves only to aid in the probing of many of the small area devices.

A cross section of a single diode is shown in Figure 3.2. This cross section is taken at the region indicated in Figure 3.1.

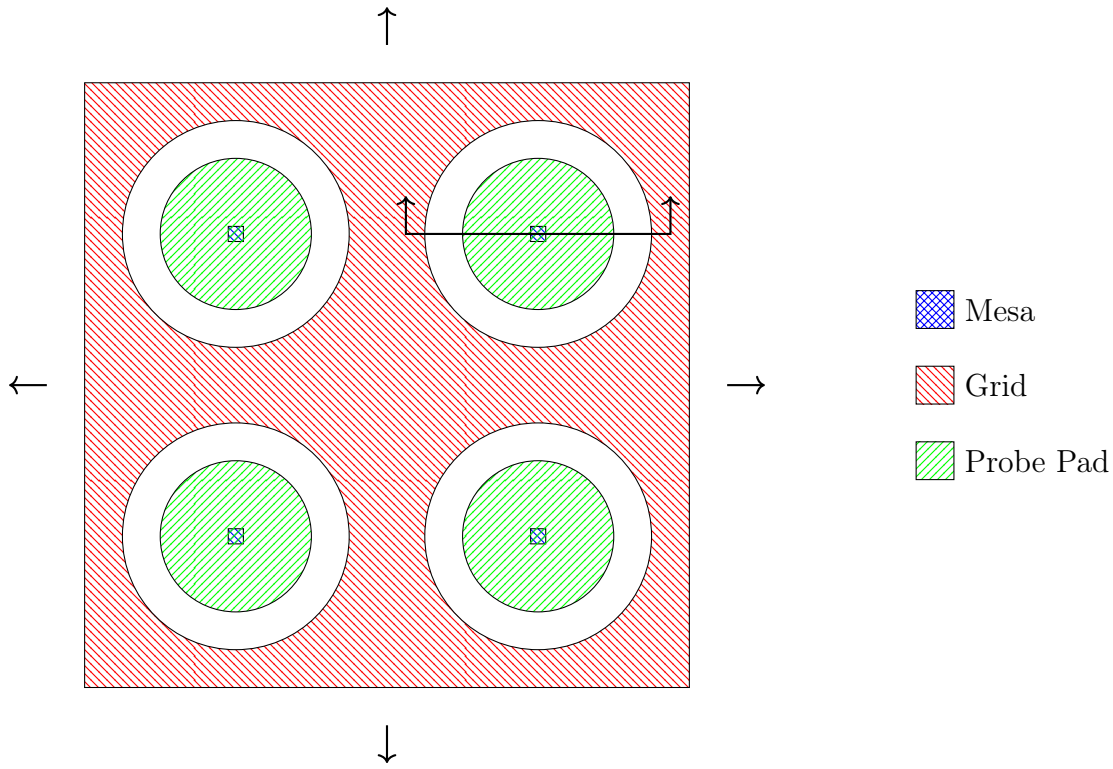


Figure 3.1: Section of the diode mask layout. The full mask is a 5×5 array of the shown piece, with varying diode size, and is provided in Appendix B.

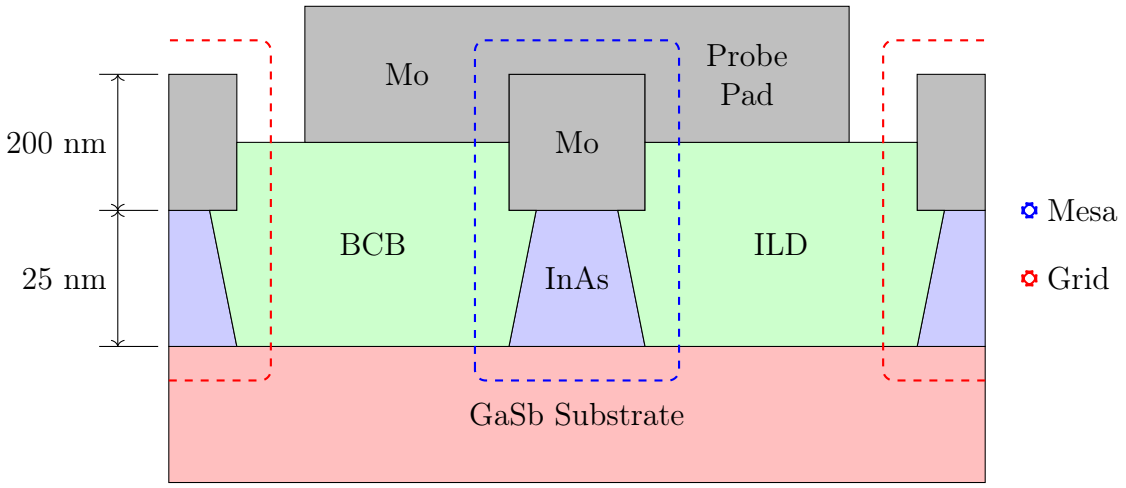


Figure 3.2: Cross section (not to scale) of a single diode taken from the region indicated by the line in Figure 3.1. The mesa (○) and grid (○) structures are indicated.

3.1.1 Electrical Testing

In order to test a diode, such as those discussed here, contact must be made to both sides of the junction. It would be convenient if the substrate could be used as one of these contacts. However, it is often the case that the diode layer structures are grown on non-conductive substrates, or use an insulating buffer layer. Both of these prevent the substrate from being a viable contact. Even though the substrate can sometimes be used as a contact, a standard process was developed that does not rely on the substrate as a contact.

The contact to the n++ region is simply the probe pad. A contact to the p++ region is more complicated. Usually, in a vertical structure such as this, a via would be needed through the ILD in order to contact the underlying layer. In this case however, the mesa structure already goes through the ILD. Could this somehow be used to create a contact to the p++ layer? The answer is yes, and this is where the grid structure comes from. Tunnel diodes are very conductive in reverse bias. When the mesa is forward biased, which is the region of interest, the grid is reverse biased. Additionally, the grid is a few orders of magnitude bigger than the mesa. This further increases the conductivity of the grid relative to the mesa. Together, all of this results in a contact that has minimal impact on the measured data.

3.1.2 Area Bias

When these devices are fabricated, the area of the mesa is smaller than was defined on the mask. The deviation in size as fabricated from the mask size is due to three main factors: e-beam direct write bias, Mo etch undercutting and InAs etch undercutting. If a GaSb etch is performed there well be undercutting of this etch as well. The cause and impact of each of these are discussed in the following subsections.

3.1.2.1 Direct Write Bias

The electron beam used in e-beam direct write is not perfectly uniform. Instead, it is strongest in the middle, and falls off at the edges. When using such a beam to expose patterns in a negative photoresist, some resist outside the features will be exposed to a low dose. Under normal conditions, this resist will develop away. Likewise, some of the resist inside the features will be slightly under exposed. The result of this, after developing, is a thinner resist near the edges than the center of every feature. This is usually not a problem, but if a sample is underexposed, the resist can become too thin to completely mask an etch. It is not uncommon for the electron beam to experience current fluctuation during a write. In order to avoid underexposed features, it is common practice to slightly overexpose everything. This has the effect of increasing the size of the features. A photoresist with a higher contrast will exhibit less change in feature size than a lower contrast resist given the same change in dose.

The fluctuations in electron beam current, and actions taken to address them, account for the entirety of the direct write bias. Were there not fluctuations in beam current, the dose to size could be easily determined and used for every run, resulting in zero bias.

3.1.2.2 Mo Etch Undercut

The molybdenum etch is a dry etch, but that does not mean there is no undercutting. The Mo etch does create a relatively straight sidewall profile, but does not prevent undercutting of the resist. This undercutting is the same on all sides and reduces the area of the mesa relative to the size and shape of the mesa.

3.1.2.3 InAs and GaSb Etch Undercut

Finally, the last part of the process that can affect the area of the junction is the final wet etch to form the mesa. The InAs is etched in 20:1 Citric Acid:H_2O_2. This etch is

a crystallographic etch, meaning it preferentially etches along certain crystal planes. Because of this, the undercut along the top and bottom edges is different than the undercut along the left and right edges.

3.1.2.4 Area Correction

In order to as accurately as possible calculate current density from the measured current, an accurate area must be used. As was just discussed, there are multiple sources of area bias, meaning the mask defined areas are insufficient, and will usually result in lower current densities than more accurate areas.

To obtain accurate junction areas a representative sample of diodes are imaged in an SEM to obtain top down micrographs of the etched top contact metal. This imaging is done after the mesa etch, after the photoresist is removed, and before any further processing. The area of the top metal is only part of the information needed to calculate an accurate area. Side profile images are also used to measure undercut. As the undercut varies much less than the size of the metal, only a few undercut profiles are measured.

In order to calculate the area from the top down micrographs, an automated system created by Brian Romanczyk is used [13]. This system is a MATLAB program that loads each micrograph and, using image processing techniques, identifies the metal region and counts the number of pixels. Using predetermined pixel areas for each magnification, the total area of the metal can be calculated. Following the metal area calculation the undercut area is subtracted and the final junction area is determined.

3.2 Fabrication Process

This section will outline the processing steps taken to fabricate the tunnel diodes discussed in this document. Unless noted otherwise, all starting material stacks

discussed in this document are grown by Molecular Beam Epitaxy (MBE) on GaSb substrates. For the figures shown here, it will be assumed that the diodes are InAs on GaSb heterostructures, but other materials can be easily substituted, and etch times and chemistry adjusted accordingly. The starting structure, after MBE, is shown in Figure 3.3. The structures used in this study were grown at Texas State University in a DCA 450 solid source MBE chamber. This system is equipped with As and Sb valved crackers and Si and Te effusion cells used for n type doping of InAs, as well as Be for p type doping of GaSb. The InAs layer, at 25 nm thick, is below the critical thickness. This is done in order to prevent the crystal lattice from relaxing and introducing defects.

Due to the high cost of the III-V substrates and MBE growths needed, these diodes are fabricated on small pieces, approximately 1 cm by 1 cm, of the starting wafer. This also allows multiple process runs to be done on the same layer structure without concern for variations between MBE runs.

3.2.1 Top Contact

A 200 nm thick layer of molybdenum is used as the top contact to the diodes. Molybdenum was chosen because it makes a good contact to the III-V materials in use. Additionally, Mo is readily available in the SMFL.

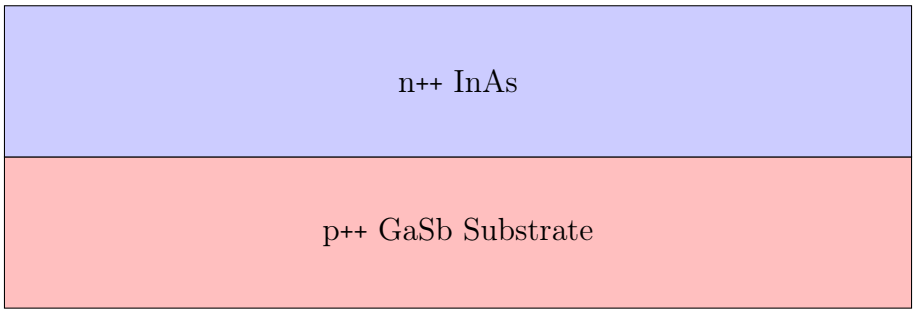


Figure 3.3: Cross section of the starting substrate after MBE growth

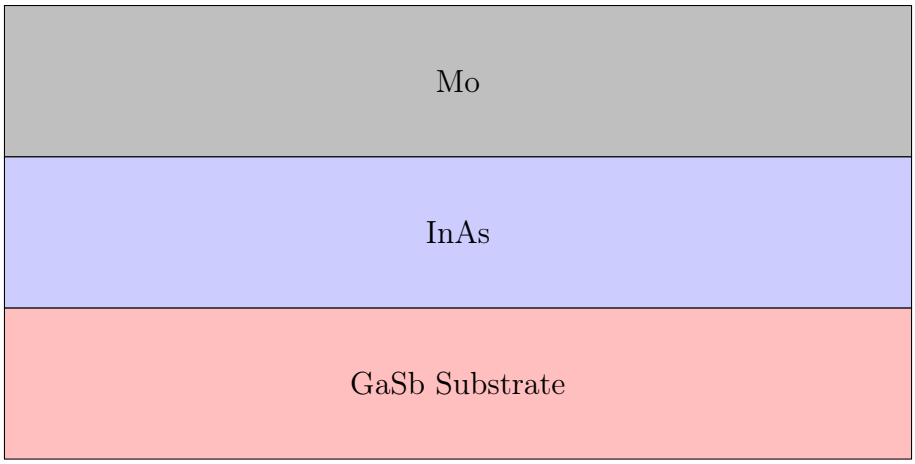


Figure 3.4: Cross section of an in-process diode after molybdenum deposition

3.2.1.1 Molybdenum Deposition

The deposition is done in a DC sputter system using an 8 inch Mo target. If there is any oxygen present during the sputter, exposed molybdenum will react to form an undesired Mo_xO_y oxide. For this reason, it is crucial that the sputter system be pumped down for at least 15 hours, in order to reach a base pressure in the 10^{-7} Torr range, before starting the deposition.

Every time the sputter system used for this process is loaded or unloaded, the target is exposed to atmosphere. When this happens, the surface of the target can react with the atmosphere, forming undesired compounds. To reduce potential contamination, it is important to perform a pre-sputter[1], before the actual deposition. The pre-sputter is done with the shutter in place using the process in Table 3.1a. Following this, the deposition is done without the shutter in place, using the process in Table 3.1b. As the pieces being coated here are much smaller than the sputter target the pieces are placed over the center of the target and held stationary during the deposition. A cross section of the process up to this point is shown in Figure 3.4.

[1]A pre-sputter is a sputter with a shutter in place between the target and the substrate. This sputter removes any reacted material from the surface of the sputter target, in order to prevent incorporation of undesired compounds in the deposited film.

Table 3.1: 200 nm Mo sputter conditions

(a) Pre-sputter

Ar Flow Rate	17 sccm
Pressure	2.2 mTorr
Power	1000 W
Time	120 s

(b) Deposition

Ar Flow Rate	17 sccm
Pressure	2.2 mTorr
Power	200 W
Time	420 s

3.2.2 Mesa

The diodes are formed by first patterning and etching the top metal, removing the photoresist, and then etching the mesa with the top metal as an etch mask. This has the advantage of reducing the exposure of the etched sidewalls to plasmas or chemistry, during the resist strip, that could potentially cause undesired reactions and effect device performance.

3.2.2.1 Lithography

For the first lithography level, an e-beam direct write system called NPGS, developed by JC Nabity Lithography Systems, is used [14]. This system allows for a standard SEM to be modified for direct write. It is therefore a much cheaper alternative to standalone direct write systems. The NPGS system used for this study is used in combination with a LEO SEM that uses a LaB_6 filament.

The first level was designed such that less than half the die area will be covered by photoresist after the develop. Therefore, in order to minimize write time, a negative tone resist is used for this level. Unlike conventional optical lithography systems, where an entire die is exposed at once, in an e-beam direct write system the electron beam must be rastered across the entire area of any feature that is to be exposed. This is substantially slower than in optical systems where the entire die is exposed at once. Therefore, since the desired features make up less area than the field, using a negative resist and exposing the areas where resist should remain, rather than using

Table 3.2: Process details for nLof 2020 (diluted 1:1 with PGMEA) **(a)** coat and **(b)** develop.

(a) Coat

Spin	Speed	3500 RPM
	Time	45 s
Bake	Temp.	110°C
	Time	60 s

(b) Develop

Bake	Temp.	110°C
	Time	60 s
Develop	Developer	CD-26
	Time	30 s

a positive resist and exposing the areas that should be removed, will save time.

The resist used for this level is AZ nLof 2020, diluted 1:1 by weight with PGMEA. This resist can be used as an e-beam resist without dilution, but it will coat to a little less than 2 µm thick at 3500 RPM [15]. When diluted however, it coats to around 300 nm. The thinner resist film is desired in order to reduce write times, as well as improve the minimum feature sizes that can be resolved. The coat process is given in Table 3.2a.

When fabricating tunnel diodes using this process, multiple die are exposed in an array, as is common in semiconductor fabrication. The standard array used for this process is 6×4, along with a 3×3 array of pre-alignment marks used to roughly correct rotation before starting a second level exposure.

When doing e-beam writes, the devices are exposed with a dose of 35 nC/cm^2. An acceleration bias of 25 kV, and a beam current between 100 pA and 500 pA are used, with lower currents requiring a longer write time, but improving resolution. After exposure, the resist is baked and developed in CD-26 developer (2.4% TMAH in H_2O [16]) with process details given in Table 3.2b.

3.2.2.2 Molybdenum Etch

An RF plasma etch tool is used to etch the molybdenum. Prior to running the etch, the chamber is cleaned (Table 3.3a), and seasoned (Table 3.3b). The etch process is given in Table 3.3c. All of these steps are done with the process chamber heated to 60°C. At this point, the etched devices are inspected under an optical microscope to

Table 3.3: **(a)** Chamber clean, **(b)** chamber season, and **(c)** 200 nm molybdenum etch process details.

(a) Clean		**(b)** Season		**(c)** Etch	
O_2 Flow	100 sccm	SF_6 Flow	125 sccm	SF_6 Flow	125 sccm
Pressure	300 mTorr	Pressure	125 mTorr	Pressure	125 mTorr
Power	280 W	Power	150 W	Power	150 W
Time	180 s	Time	180 s	Time	130 s

verify the etch has completely cleared. In the rare case where it has not, the devices are etched again for an amount of time, determined on a case-by-case basis, and is based on how much Mo appears to remain. Usually this is about 15 seconds. After this etch has cleared, the resist is stripped in an oxygen plasma.

3.2.2.3 InAs Etch

Using the top metal as an etch mask, the InAs is etched in a 20:1 citric acid:H_2O_2 solution. The etch rate of InAs in this chemistry has been experimentally determined to be 6.25 Å/s. The layer structures used in this experiment usually have a top InAs layer that is 250 Å thick. Using this etch rate, an etch time of 40 s would be used. Because GaSb does not etch in this chemistry, it is not of critical importance to avoid over etching. The result would only be a smaller area diode. When other materials are to be etched, the same etch solution is used, and etch rates are taken from a paper by DeSalvo *et al.* [17].

After the InAs etch, the diodes (Figure 3.5) are complete and can be tested. Since many of the diodes fabricated in this manner are quite small, some as small as 50 µm squares, probing them at this point can be difficult. Therefore, if a few devices are demonstrated to work, more fabrication work is done to add bigger probe pads on top of the mesa.

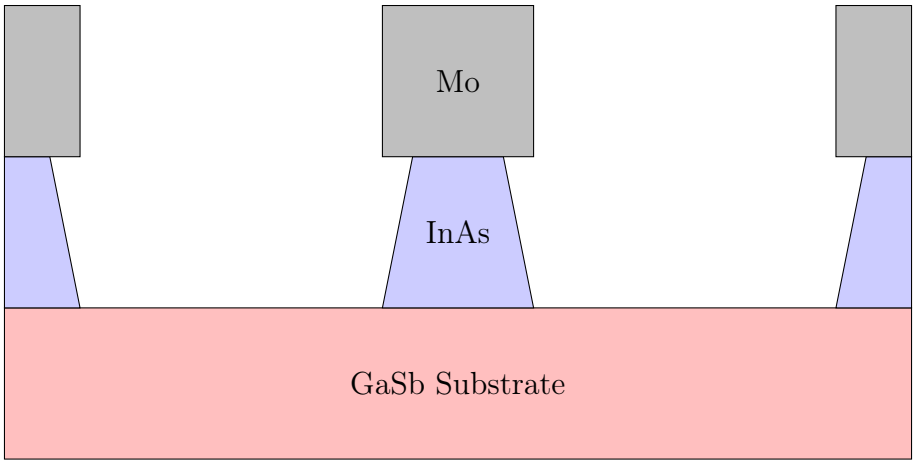


Figure 3.5: Diode after meta lithography and etch

3.2.3 Inter-Layer Dielectric

An ILD is needed to support the addition of a probe pad. The ILD used here is divinylsiloxane-bis-benzocyclobutene (BCB), specifically, Dow Cyclotene 3022-35.

3.2.3.1 BCB Coat

The BCB is spin coated, in the same manner as a photoresist. The sample is spun at 3000 RPM for 45 s. This results in a film approximately 1.3 µm thick [18]. After coated, the BCB must be cured. First, the samples are baked on a hot plate at 140°C for 5 min. Then, they are transfered into an oven at 140°C in an N_2 ambient. The oven is ramped to 250°C over 30 min. Then, the samples are baked at 250°C for 1 hr. If the BCB is cured in an O_2 ambient, rather than an N_2 ambient, the BCB will oxidize, which is generally undesired.

3.2.3.2 BCB Etch Back

The BCB film, as coated, is much thicker than the mesa height. Therefore, in order to expose the top contact, it must be etched back. This etch is performed in an RF plasma etch system using the process given in Table 3.4c. After running the initial etch back process, the samples are inspected under an optical microscope. If the top

Table 3.4: Process details for BCB **(a)** coat, **(b)** cure and **(c)** etch back.

(a) Coat

Prime	Primer	AP 3000
	Speed	3000 RPM
	Time	45 s
Spin	Speed	3000 RPM
	Time	45 s
Bake	Temp.	140°C
	Time	300 s

(b) Cure

Load	Temp.	140°C
	Ambient	N_2
Ramp	Temp.	250°C
	Time	50 min.
Bake	Temp.	250°C
	Time	1 hour
Unload		

(c) Etch back

Pressure	300 mTorr
Power	25 W
Gap	1.65 cm
O_2 Flow	80 sccm
He Flow	100 sccm
SF_6 Flow	20 sccm
Time	300 s

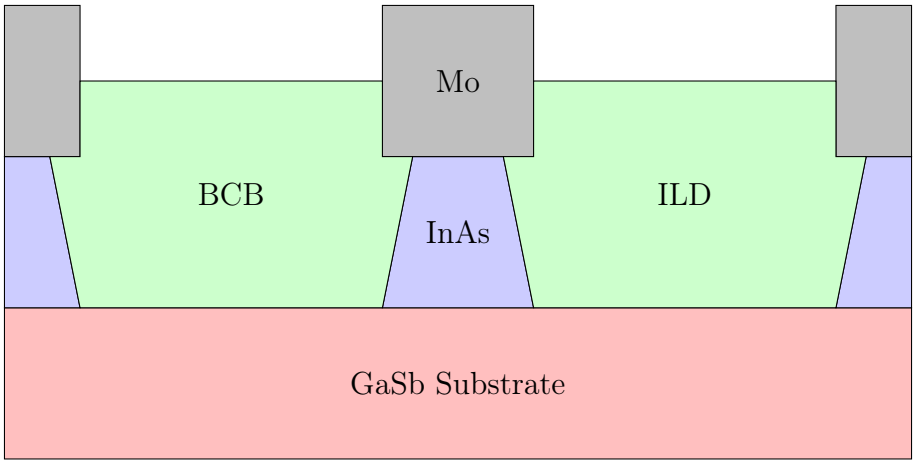


Figure 3.6: Diode after BCB coat and etch back

metal is not yet cleared, the samples are etched again for an amount of time dependent on the thickness of BCB remaining (usually between 30 s and 60 s). This is repeated until the BCB is clear. A cross section at this point is given in Figure 3.6.

It is critical that the BCB not be over etched such that the sidewalls of the n++ region are exposed to the etch plasma. Even though an acceptable ILD could be fabricated with an additional BCB coat and etch back, the exposed sidewall would oxidize in the oxygen plasma, which could create leakage paths that degrade diode performance.

Table 3.5: Resist coat conditions for **(a)** LOR-5A and **(b)** dilute AZ-8250. **(c)** Develop conditions for the LOR-5A/AZ-8250 resist stack used for Mo liftoff.

(a) LOR-5A Coat

Spin	Speed	3000 RPM
	Time	45 s
Bake	Temp.	140°C
	Time	300 s

(b) Dilute AZ-8250 Coat

Spin	Speed	3000 RPM
	Time	45 s
Bake	Temp.	110°C
	Time	60 s

(c) Develop

Bake	Temp.	110°C
	Time	60 s
Develop	Developer	CD-26
	Time	45 s

3.2.4 Probe Pad

3.2.4.1 Lithography

This lithography step is a bit different than the one in Section 3.2.2.1. First, the samples are coated in a bi-layer resist stack of diluted AZ-8250 on LOR-5A. The coat recipes for these resists are given in Table 3.5a,b. Unlike the resist used for the first level lithography, nLof 2020, AZ-8250 is a positive resist. Since this is a liftoff process, and a positive resist is being used, the areas where metal features are wanted are the areas that get exposed.

The dose used for this resist is 7 nC/cm^2. The remaining electron beam details, such as beam current and energy, are the same as the previous exposure.

3.2.4.2 Molybdenum Deposition

The Mo deposition for liftoff uses the same sputter process as outlined in Section 3.2.1.1. A sputter deposition, unlike evaporation, is not a line of sight deposition process. Generally, depositions for liftoff should be line of sight in order to maximize the chance of having a disconnect between the metal on the resist, and the metal on the substrate. In this process however, the undercutting of the resist prevents the Mo from forming a continuous layer, which allows liftoff to successfully occur. This is still not ideal however. Because the process is not line of sight, thin vertical *fences* are formed at the edge of the pad. These can fall over and cause shorts.

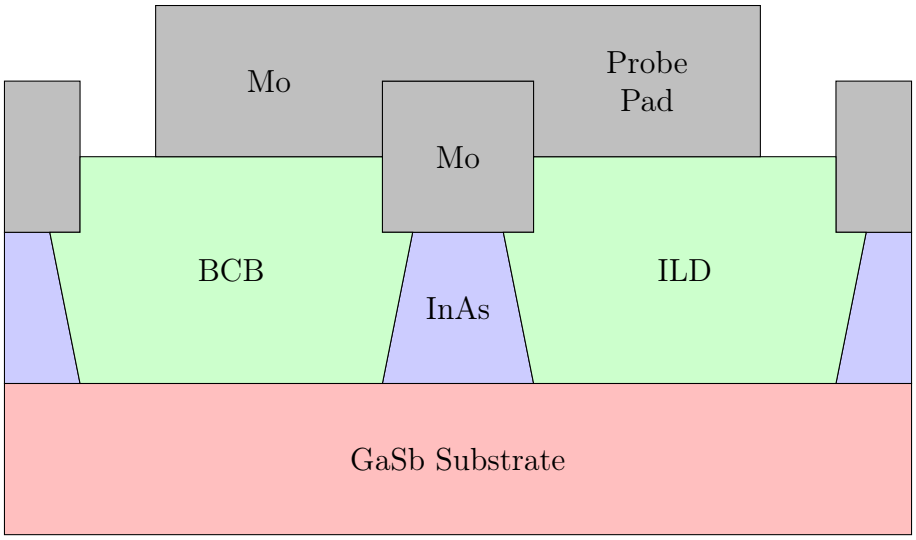


Figure 3.7: Final device structure after probe pad lift-off

3.2.4.3 Liftoff

Liftoff is done in Remover PG, heated to 75°C. In order to minimize redeposition, and speed up the liftoff, the samples are agitated periodically, throughout the liftoff process. The samples remain in the Remover PG until they appear to be complete, which usually takes 15 minutes, and is a function of the size of the field that must be undercut. To remove residual Remover PG the samples are rinsed in isopropyl alcohol, and then water. In order to verify complete liftoff, the samples are inspected under an optical microscope. If liftoff is not complete, the samples are returned to the Remover PG for a few more minutes, then rinsed and inspected again. This is repeated until liftoff is complete.

The devices are now complete. The final device cross section is shown in Figure 3.7. At this point the devices can be easily tested using manual probes and a semiconductor parameter analyzer.

3.3 Optical Lithography Process Development

Previously, all lithography levels had been done by e-beam direct write. This is the process that was described in detail in Section 3.2.2.1. The reason for the use of e-beam direct write was flexibility. Direct write allows researchers to change the design whenever a change is desired without any care for the time or costs associated with producing a new mask.

Optical lithography, however, has its own advantages. In optical lithography, an entire die is exposed at once. This differs from an e-beam direct write, where the beam is slowly rastered across the sample to form the desired image. Because of this, the time to expose a single sample is substantially higher in an e-beam direct write system than in an optical lithography system.

Another disadvantage of an NPGS based lithography process compared to an optical lithography process has to do with alignment. In a standalone lithography system, the stage is, unsurprisingly, designed for use in a lithography system. These stages often employ a laser interferometer in order to precisely position the stage after stepping to a new die location. The stages used in an SEM however, were not designed for precise positioning. Therefore, each die must be aligned individually. This has a considerable impact on the time to expose a sample at level 2 and beyond.

In addition to the need to align each die individually when using NPGS, alignment is substantially more difficult in an e-beam direct write system than in an optical lithography system. In an e-beam system, any area that is imaged, is also exposed. Therefore, it is not possible to image the entire die in order to find the alignment marks. In order to align, small alignment scan windows are specified. The windows used for the 1st lithography level are shown in Figure 3.8. These windows are not much bigger than the alignment marks themselves. It is not uncommon to begin alignment and have no alignment marks within their windows. At this point, it becomes very

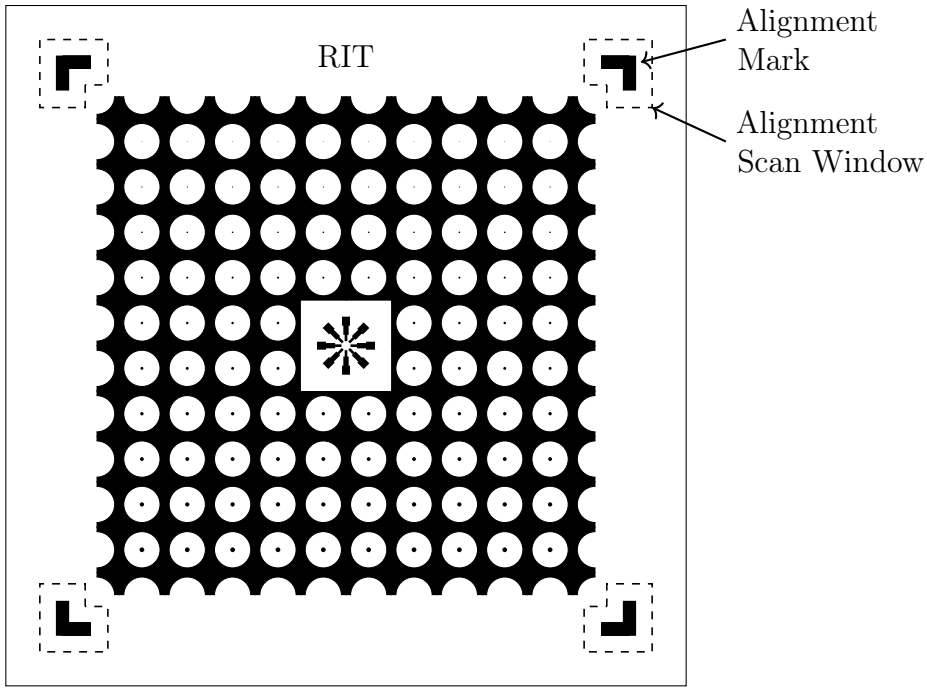


Figure 3.8: Diagram of the first level pattern for a single die as written by e-beam direct write. One of the four alignment marks and associated scan window is indicated.

difficult to locate the alignment marks without accidentally exposing a part of the die that should not be exposed. The need to limit the observable area during alignment using these windows further complicates the alignment process.

All of these problems associated with e-beam direct write lithography can be solved by moving to an optical only lithography process. However, this will lose the flexibility to change diode sizes between runs. Therefore, a hybrid e-beam and optical lithography process was developed where the mesa level is written by e-beam, and the remaining features are done by optical lithography. This hybrid process is described in Section 3.3.1.

3.3.1 Hybrid Lithography Process

In order to take advantage of the benefits of optical lithography over e-beam direct write, without losing the flexibility of direct write, a hybrid optical/e-beam process was designed with the following objectives:

- Minimize e-beam direct write steps
- Allow for mesa size customization for each run
- Minimize distance from mesa to the p++contact

Direct write capabilities are still desired for the imaging of the very small mesa structures. Therefore, for the mesa level, e-beam direct write is used. The remaining levels do not need to change between process runs. In keeping with the design objective to minimize e-beam direct write steps, these remaining levels are patterned by optical lithography.

As the process was undergoing some major changes anyway, this time was taken to improve upon the p++ contact. Under the old, e-beam only, process, the contact to the p++ side of the diode was through a very large reverse biased tunnel diode referred to as the *grid*. The biggest disadvantage of this was the distance from the grid to the mesa. This distance adds series resistance that degrades the measured I-V characteristics of the diodes. The new design moves the bottom contact much closer to the mesa, and provides direct contact to the p++ layer, without going through a tunnel diode.

A portion of the hybrid layout is shown in Figure 3.9, with the full layout given in Appendix C.

3.3.1.1 Shadowed Metal Process Option

A shadowed deposition process is a self aligned process. When using a line of sight deposition process, such as evaporation, a material is deposited only on the surface that is directly visible from above the substrate. If there is any undercutting the material will not be deposited under these features.

This process as described here can easily substitute a shadowed metal deposition process in place of the grid liftoff. Doing so will reduce resistance through the p++ layer by moving the metal contact closer to the junction. It also removes the need

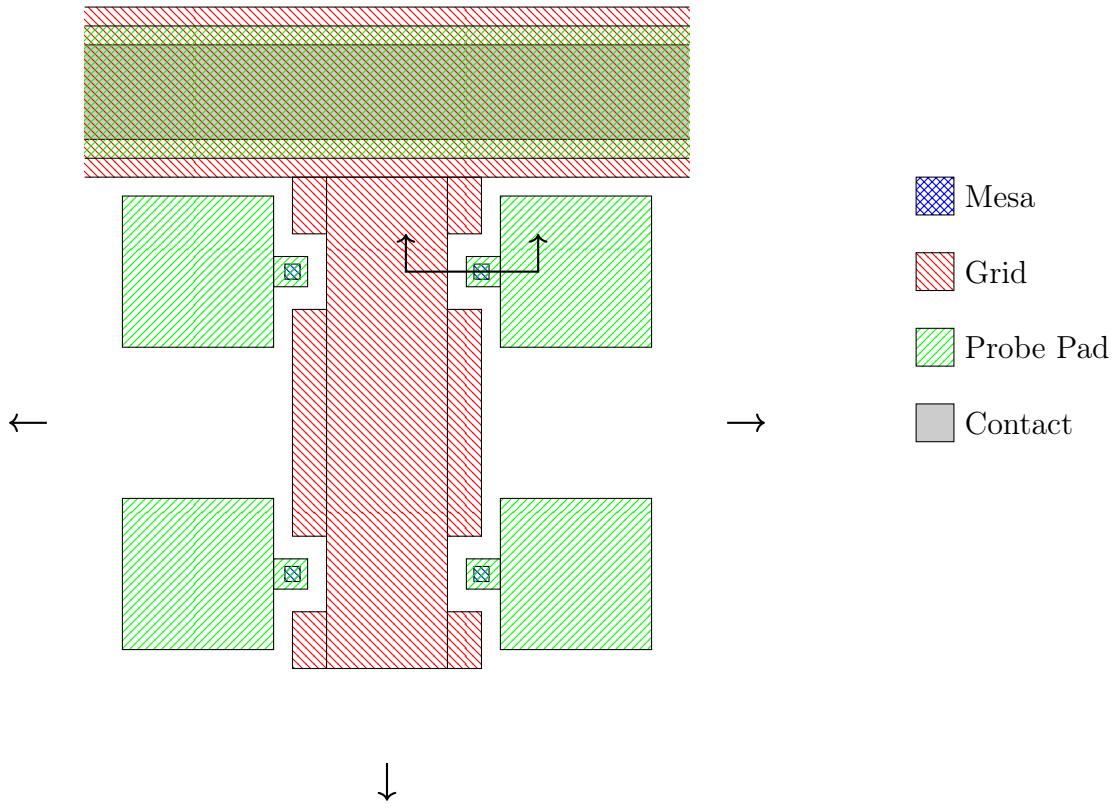


Figure 3.9: Section of the new hybrid optical/e-beam diode mask layout. The full layout is an expanded version of what is shown here, and has varying diode sizes as defined by the e-beam step. The full layout is provided in Appendix C.

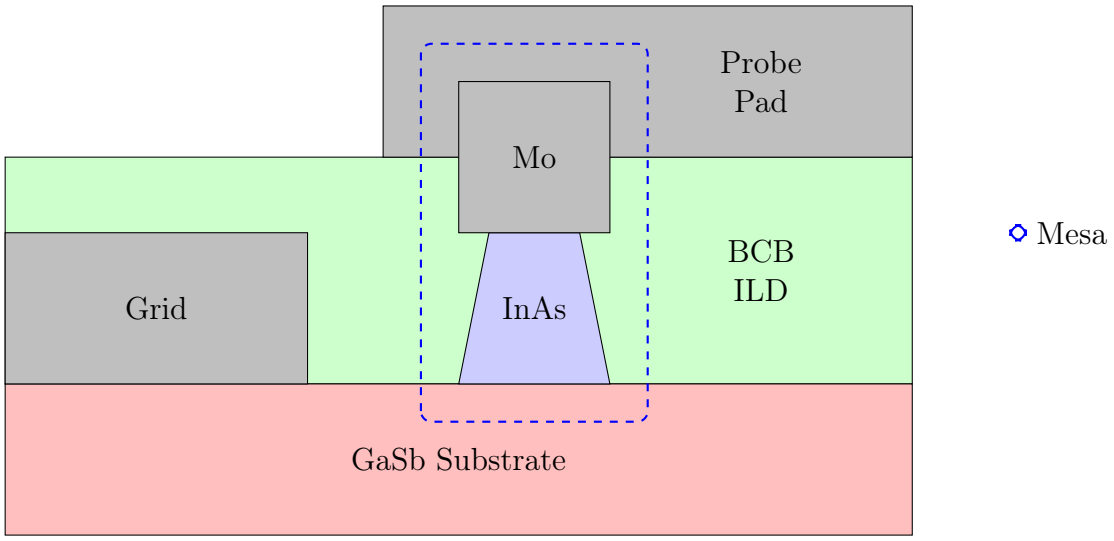


Figure 3.10: Cross section (not to scale) of a single diode taken from the region indicated by the line in Figure 3.9. The mesa (⬡) structure is indicated.

for a lithography or liftoff step to pattern the metal. The process that has been successfully executed is a 50 nm molybdenum deposition done by e-beam evaporation with a deposition rate of approximately 3 Å/s and is discussed in more detail by Gaur [19].

It is important to remember that when using this shadowed process instead of the liftoff process, the metal will end up everywhere, including under the probe pads. Because of this, the ILD thickness will be decreased resulting in an increase in capacitance. If the shadowed process is desired, but the increase in capacitance is prohibitive, an extra lithography and etch step could be done to remove metal from under the probe pad.

Another important thing to consider regarding the shadowed process is the film thicknesses involved. The shadowed process developed by Gaur was designed for thicker films. The thicker films resulted in more undercutting, which allowed the shadowed deposition to work appropriately. When working with thinner films, such as is common in tunnel diodes, there may not be enough undercutting of the top metal to shadow appropriately. If this is the case, and metal is deposited on the junction, the diodes will be shorts. In order to account for this, the etch time could be increased to cause more undercutting, as long as the process can handle the over etch.

Chapter 4

Lithography on Pieces

Optical lithography is substantially faster than e-beam direct write. However, it is still desirable to work on small pieces of a wafer, rather than an entire wafer. This may be for cost reasons, or the desire to preserve pieces of a wafer for future process runs or materials characterization.

A process was designed for exposing partial wafers, or *pieces*, on the GCA 6300B DSW wafer stepper (GCA), a g-line stepper for 4 or 6 inch wafers, in the SMFL at RIT. This process uses a special pieces *paddle* to hold the pieces, rather than the standard 4 inch or 6 inch paddles. A diagram of this paddle is shown in Figure 4.1. There are three sample stops which can help to roughly align the rotation of the sample to the system, as well as a single small vacuum hole for holding the piece firmly in place during the exposure.

4.1 First Level

When exposing the first lithography level, the location of an array of die must be chosen. If exposing an entire wafer, the stepper automatically determines where the array will be. However, when exposing the first lithography level on a piece, the die array must be positioned manually. This is done using a procedure similar to the alignment procedure used to align to existing features.

When aligning to a substrate, rather than to a pattern, the image that is to be

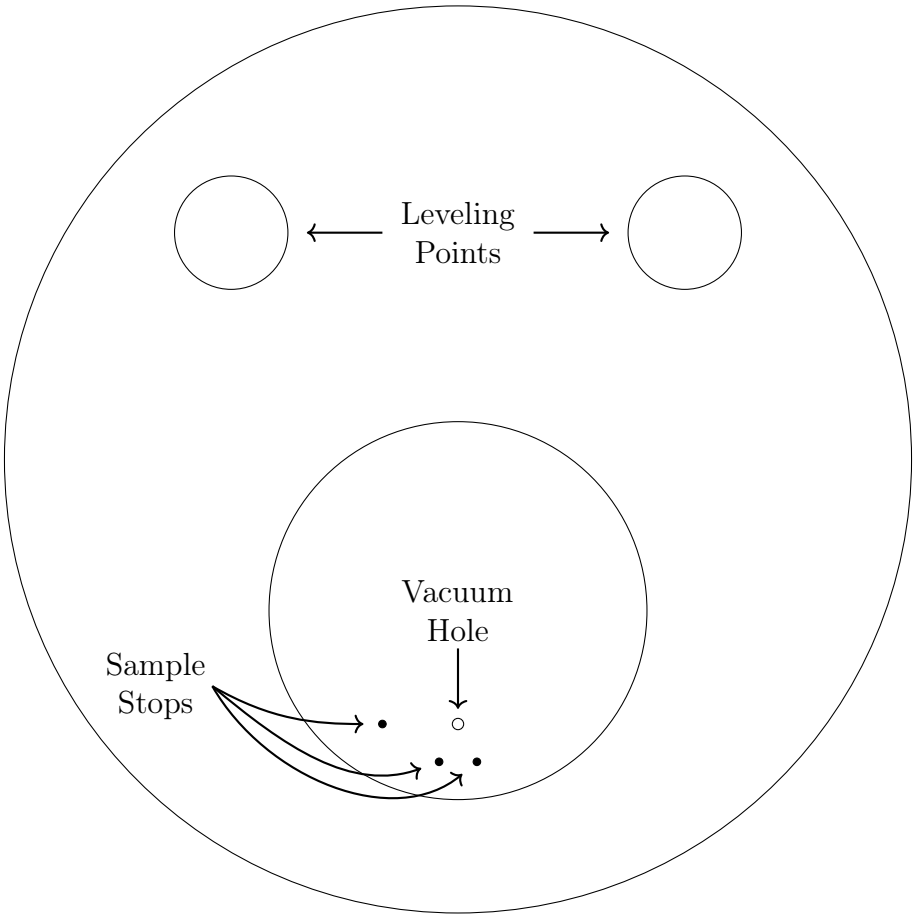
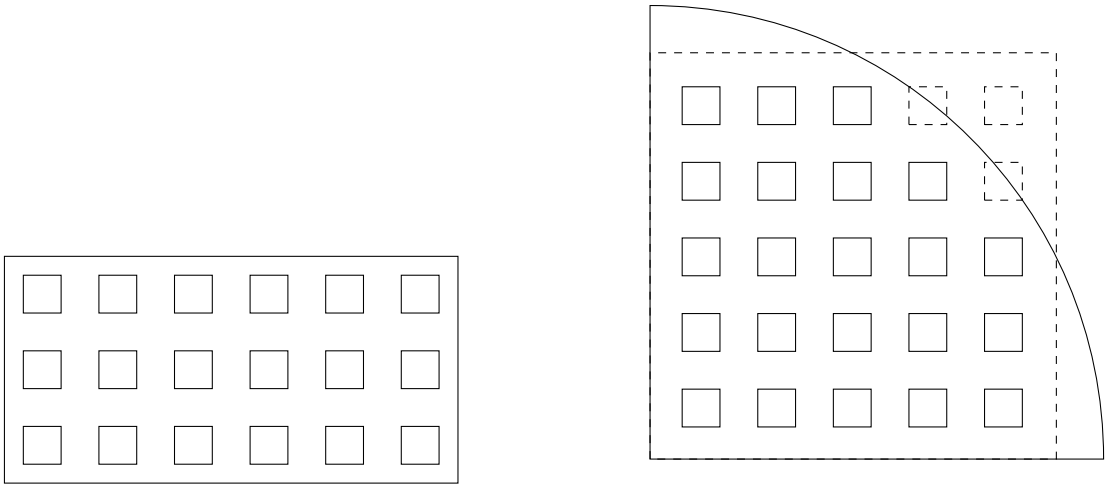


Figure 4.1: Drawing of the pieces paddle used when doing lithography on pieces on the GCA. Special locations that may be able to be used for leveling are indicated, as well as the stops used to help position the sample, and the vacuum hole over which the sample must be placed.



(a) Rectangular piece that fits a rectangular die array

(b) Non-rectangular piece that does not fit a rectangular die array

Figure 4.2: Examples of varying piece shapes that require different die placement consideration. **(a)** is a rectangular piece allowing for a perfect rectangular array and **(b)** is a quadrant of a wafer which cannot perfectly fit a rectangular array.

exposed must be aligned to the die as the die will be after it is exposed. This is easiest when the alignment offset is (0,0) for the 1st level, as this will put the center of the right alignment die directly below the right objective. If there is a nonzero alignment offset for the 1st level, then the pattern will be shifted off center by the programmed amount.

4.1.1 Rectangular Piece

The simple case is for a rectangular piece, such as the piece shown in Figure 4.2a. In this case, a rectangular array can fit perfectly within the piece. When preparing to align to such a piece, make sure to set up the number of rows and columns in the job file, and follow all other piece setup requirements as outlined in Section 4.5.

To align to the substrate, align the right alignment objective to the lower right corner of the piece. Note the coordinates displayed on the stepper screen. These are the stage coordinates relative to what the stepper has decided is the center of the substrate. It is very unlikely that $(0, 0)$ actually corresponds to the center of anything,

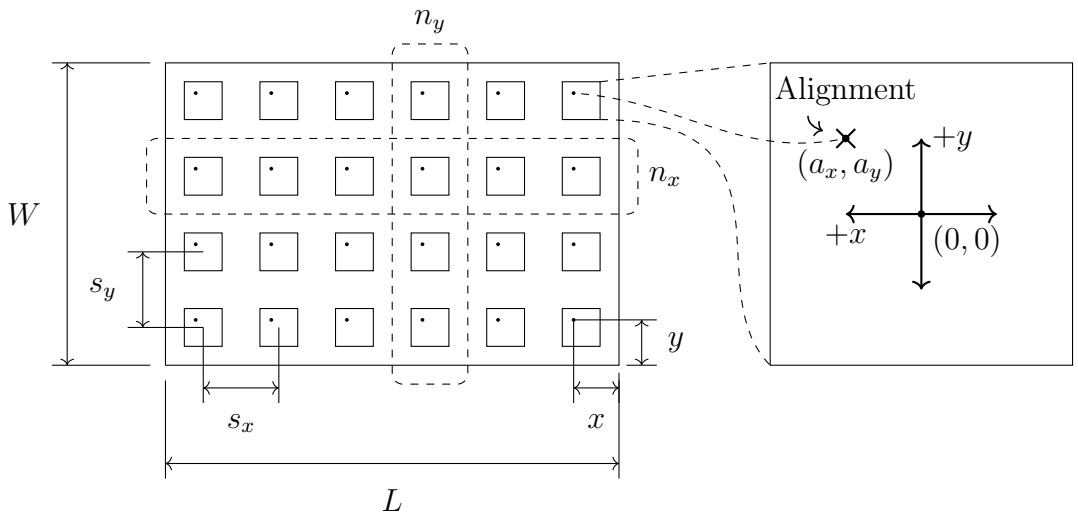


Figure 4.3: Various parameters used in the calculation of the alignment offset from the bottom right corner. L/W are the length and the width of the piece, n_x/n_y are the number of die in the x/y direction, s_x/s_y are the spacing between die in the x/y direction, a_x/a_y are the x/y alignment offsets and x/y are the desired alignment offsets.

unless the job was set up specifically for the size piece that is being used. Using (4.1), with parameters as shown in Figure 4.3, an offset can be calculated from this corner.

$$x = \frac{L - (n_x - 1)s_x}{2} + a_x \tag{4.1a}$$

$$y = \frac{W - (n_y - 1)s_y}{2} + a_y \tag{4.1b}$$

Starting with the right alignment objective directly above the lower right corner, and by referencing the coordinates displayed on the screen, move the stage by the calculated amount. It is not required that the lower right corner be used for initial alignment, or that the lower right die be used as the right alignment die. However, the provided equations only work when using the lower right corner and the lower right die. A researcher who desires to use a different corner or die should apply basic concepts of geometry in order to calculate the correct offset.

4.1.2 Non-Rectangular Piece

Not all pieces are rectangular, such as the one shown in Figure 4.2b. These pieces can be aligned to, but are a bit more complicated. When determining the die placement, do so as if working with a rectangular piece, as shown by the dashed rectangle in the figure. This rectangular piece need not fit entirely within the actual piece. It may be necessary to use die in a row other than the bottom row as the alignment die. When doing so, add multiples of the die step to the calculated alignment offsets x and y from (4.1) to account for this. Additionally, the bottom right corner of the rectangle may not fall exactly on the bottom right corner of the piece. If so, also add this offset to the calculated x and y offsets.

It is possible to use a corner other than the bottom right for initial alignment, or any other feature with a known location. It is only important that the right objective is positioned over the center of the right alignment die, offset by the 1^{st} level alignment offset programmed into the job file, before starting the exposure. It does not matter how you actually got there. Again, equations for all imaginable cases are not included here, but can be easily derived with a basic understanding of geometry.

When the stepper attempts to expose a die that does not fall on the wafer, processing will stop and a focus error will be shown on the screen. The stepper will then prompt you to press Q to stop processing. If instead Enter is pressed, processing will continue, skipping that die. This makes it easy to set up a die array to fill the piece, without needing to worry about the die that may fall off of the piece. If, however, many runs will be done with the same size piece, it could be beneficial to instruct the stepper to skip these die. During the job setup, dropout die can be specified which will not be exposed. This way, the job should run through to completion without ever stopping for a focus error. There is more information on dropout die in Section 4.5.3

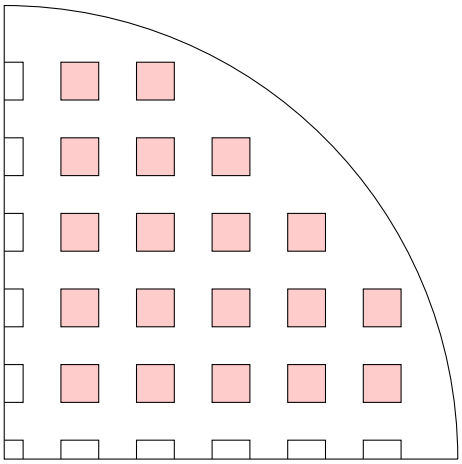


Figure 4.4: An example of a piece of a wafer. This wafer was previously patterned and was then cleaved into multiple pieces. The die previously patterned are shown. The subset of these die that will be patterned for the next level are shaded (■)

4.2 Second Level

Features on a piece can be aligned to, regardless of whether or not they were originally aligned to a piece or a full wafer. The procedure is the same if the first level lithography was performed on a piece, or on a full wafer which was subsequently broken. The latter can be useful for performing experiments late in the process. In the case of the wafer that was patterned and then cleaved however, there may be partial die. In some cases the remaining parts of these die can be patterned with the upcoming level, but they usually cannot.

Exposing a piece that already has patterns uses almost the exact same procedure as exposing a full wafer. Assuming the job is correctly set up for a piece, which is explained further in Section 4.5, the only thing required is to align to the correct die. This often requires moving the stage quite far until the right objective is over the right alignment die. If the die to be exposed are not in a rectangular array, the job should specify a rectangular array that would contain all the die. In the case of the piece in Figure 4.4 a 5×5 array would be sufficient. When the stepper attempts to expose a die that is not on the piece, an error will be displayed. As was done for the

1st level, press Enter to continue processing. This will be required for each die in the array that does not fall on the piece.

4.3 Quadrant Masks

One way to reduce the cost of purchasing a mask set, is to divide up a single mask plate into four quadrants and put a different mask level in each quadrant.[1] An example of such a mask is provided in Figure 4.5. When using this mask, the stepper must be provided alignment offsets for each quadrant, as well as masking aperture settings to mask off the other quadrants.

When exposing an entire wafer, quadrant masks usually work as expected. There are, however, limitations to using quadrant masks that can become a problem when trying to expose a piece. Unlike die placement on a full wafer, when working with a piece, die are frequently very close to the edge. There are multiple reasons why this is the case. First, if the piece being exposed is close to the size of a single die, there is no option but to place die close to the edge. Second, if the piece being exposed was cleaved part way through the process, after there had already been at least one lithography level, die are very likely to end up close to the edge unless exceptionally wide streets are used.

When exposing a die that is close to the edge, the stepper is not always able to focus on the sample properly, and the die may be exposed with an incorrect focus, or not exposed at all. Die exposed out of focus are very rarely usable, and must generally be reworked.

The reason this focus issue occurs is related to how the GCA stepper focuses on substrates. The automatic focus system reflects a beam of light off the substrate, and detects the reflection [20]. The height of the microreduction printer tube, containing

[1] The mask can be divided up into many more than four parts, but this document will only discuss the most common case of four square sections.

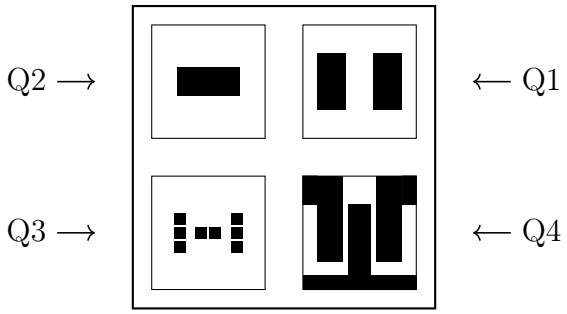


Figure 4.5: Example quadrant mask with four different lithography levels all contained within the same physical mask. When using such a mask, the stepper must be configured to mask off the unwanted sections in order to prevent exposure of the entire mask on the sample. Quadrant numbers are indicated.

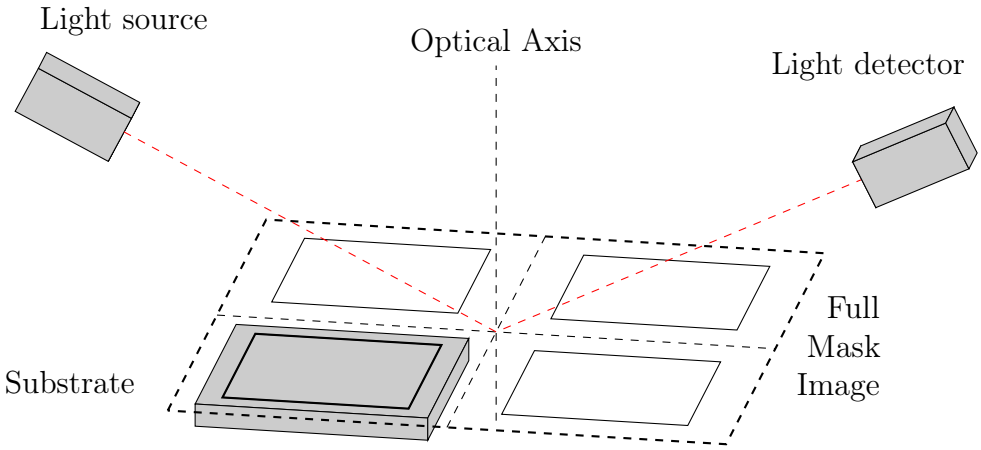
the reduction lens, is adjusted until optimal focus is achieved. The beam of light that is reflected off the substrate is centered on the optical axis of the reduction lens. A schematic diagram of this system is shown in Figure 4.6. As shown in the figure, for substrates the size of, or smaller than, a single mask quadrant, the substrate will not be on the optical axis. Therefore, the stepper will not be able to correctly focus on the top of the substrate, resulting in an exposure that is severely out of focus.

4.3.1 Solutions to Quadrant Mask Focus Difficulties

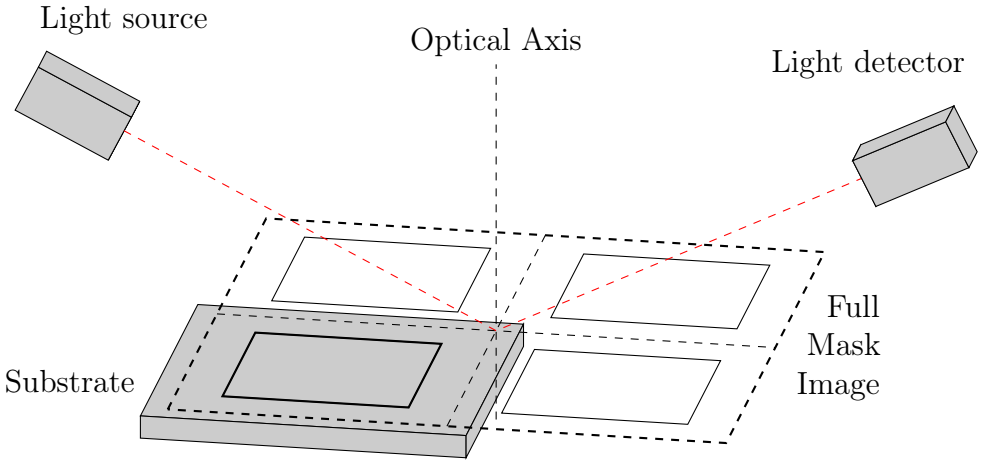
Quadrant masks are still desirable, as they can reduce costs. Therefore, presented here are some ways to reduce the occurrence of out of focus die. The focusing difficulties when using quadrant masks are caused by the fact that the die on the mask is not on the optical axis, as it normally would be in a standard single level mask. These solutions attempt to reposition the piece onto the optical axis in order to allow focusing to occur.

4.3.1.1 Shift Die Off Center

Sometimes it is possible to use a quadrant mask with a piece smaller than would normally work just by shifting the location of the die on the piece. However, there are some restrictions on which mask quadrants can be used in future lithography steps.



(a) Stepper is unable to focus because the sample is not on the optical axis.



(b) Stepper is able to focus because the sample is on the optical axis

Figure 4.6: Schematic diagram of the GCA stepper automatic focusing system. A representative image of a quadrant mask is shown as it would be exposed without the masking aperture blades in place. In the case of **(a)**, the substrate is too small, and therefore does not overlap the optical axis. This prevents the stepper from focusing on the substrate. This can be fixed by using a bigger substrate, such as shown in **(b)**.

The simplest case is when there is only one mask level to be exposed (or every mask level is in the same quadrant, on different quadrant masks). In this case, the die can be shifted in both the x and y directions without causing problems for future levels, as there are no future levels (or the future levels are in the same place on other masks). If, for example, the level to expose is in quadrant one (as indicated in Figure 4.5), the die, or array can be shifted on the sample up and right to the top right corner of the piece. This will move the sample closer to the optical axis, and hopefully allow the system to properly focus.

There are more restrictions on how the die, or array, can be shifted on the piece when additional levels will be aligned to the first. Another simple, although unhelpful, case is the case where quadrants on opposite corners will both need to be exposed on the same piece. These would be quadrants 1 and 3, or 2 and 4. In this case, no shift can be performed as a shift in either the x or the y direction to move the piece closer to the optical axis, would actually end up moving the piece away from the optical axis in a future level. For this case, other solutions should be investigated.

The last case is the case where multiple levels will need to be exposed, but all levels are on the same side of the mask. Specifically, for this to work, all masking levels to be used must be in quadrants (1 and 2), (2 and 3), (3 and 4) or (4 and 1) only. In this case, the die can be shifted in either the x or the y direction, depending on which quadrants are used, without causing problems for future levels. The reason for this is because both quadrants used are off axis in the same direction in one of the two axes. For quadrants (1 and 2) and (3 and 4) the die can be shifted in the y direction for the first level, which will move the die toward the optical axis for both levels. For quadrants (2 and 3) and (4 and 1) the die can be shifted in the x direction for the first levels, which will likewise move the die toward the optical axis for both levels.

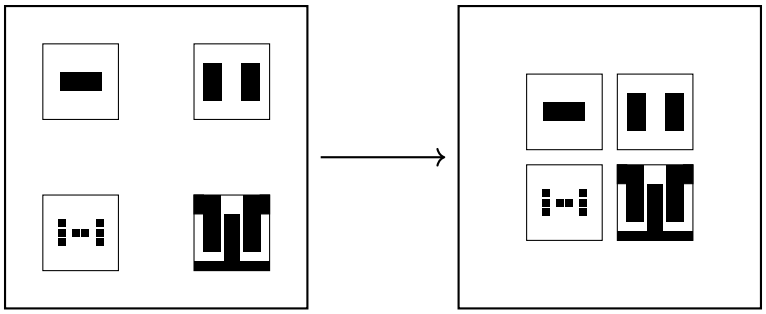


Figure 4.7: Example quadrant masks (left) each pattern centered in its respective quadrant, (right) patterns moved toward the center to reduce the distance from the optical axis. This will reduce difficulty focusing on small samples.

4.3.1.2 Reposition Mask Quadrants

The quadrant masks created at RIT are laid out such that each level is centered in its respective quadrant. This means if the maximum mask area is w wide and h high, the center of each quadrant is located at ($\pm w/4$,$\pm h/4$) from the center of the mask. This convention makes writing stepper jobs easier, as specific offsets do not need to be remembered for each mask that is created, they are all the same. These offsets were chosen as they allow for the maximum die size, and can therefore be used for all quadrant masks, regardless of the size of their die, assuming the die actually fit in the space allowed.

This convention does does however, also move the pattern relatively far away from the optical axis. This is something that is an unavoidable consequence of large die sizes, but is something that can be adjusted when using smaller die. If each quadrant was moved toward the center of the mask, as shown in Figure 4.7, this would decrease the distance of the die from the optical axis. In doing so, assuming the previously discussed die shifting method is not an option, the minimum piece size that can be exposed is increased.

4.4 GCA Alignment Commands

There are a few keyboard commands that can be of use during alignment of a piece that are not normally used during alignment of a full wafer. These commands are presented here, along with any precautions that must be taken when using them. In each of these cases, the command will be a single key that should be pressed during alignment. These commands are used in the same manner as the [O] command that is used to change the stage speed.

4.4.1 Switch Alignment Objective

The GCA stepper has two alignment objectives. The objective on the right is used to align in the x and the y axes, and the combination of the two objectives are used to correct for rotation. The general procedure for creating a job on the GCA is to choose a step size such that die are placed directly below each of these objectives. This allows both alignment marks to be visible at the same time. Alternatively, each level can contain two alignment marks, such that a more space efficient die step can be chosen, while still resulting in an alignment mark directly below each objective.

When aligning a piece, the alignment marks cannot always be placed under each objective. In fact, it is not uncommon to try to align to patterns on a piece smaller than the distance between the two objectives. In this case, it is useful to use the alignment command: [A]. Pressing this key will move the stage to the right an amount equal to the distance between the alignment objectives less the die step size in the x direction times the number of steps between the alignment marks. The practical effect of this, assuming a rotation-corrected sample, is to position the left alignment mark under the left objective. It is, however, uncommon for the alignment mark to actually be positioned correctly on the first try. When the stage is positioned for alignment under the left objective, rotation correction should be attempted. A second press of

A will move the stage by the same amount, in the opposite direction. When the stage is in this position, the x and y alignment should be corrected. It will likely take multiple iterations of switching between the two objectives before correct alignment is achieved.

4.4.2 Die Indexing

It is sometimes of use to be able to step between die. The GCA has a command to do just this: D. Upon pressing this key, the alignment joysticks will no longer jog the stage, but instead step between die. To return to jog mode, the command E is used.

It is important to note that stepping between die does not adjust the alignment offsets that will be used when exposing the current piece. This means that die index mode cannot be used to move to the intended alignment die after finding the alignment mark on another die. The effect of doing so would be as if alignment had been performed on the die that was under the objective prior to stepping between die.

While die indexing does not allow you to move to the correct alignment die after finding another, it does have a use for alignment. As previously stated, if die indexing mode is used, the alignment offsets will be calculated as if alignment was performed on the die before die indexing. If the alignment mark is damaged or otherwise unusable on the intended alignment die, die indexing can be used to move to another die and align with that die, saving the operator from the need to re-write the job using different alignment die.

An interesting aspect to die indexing is how it interacts with the alignment objective switch command: A. A die offset is maintained for each alignment objective. If die indexing mode is used to change the die being aligned to for the right objective, when pressing A, the normally used alignment die will be positioned under the left objective.

4.4.3 Moving Between Objectives

Sometimes it is of use to change which alignment objective is being used, without changing which alignment mark is being observed. This is often the case when samples have poor contrast as the left objective seems to perform better than the right one.

Pressing [T] will allow for easy moving of the alignment mark under one objective to the other. It works in a similar manner to the die step mode. Once pressed, the right joystick will now move the stage in increments of the distance between the alignment objectives. To move the alignment mark under the right objective to the left one, simply press [T] and then tap the joystick left once. To return to normal die jog mode, press [E], as would be done when leaving die index mode.

When using this movement mode care should be taken to not accidentally move in the wrong direction. Due to the relatively high distance between the alignment objectives, it is possible that a single step in the wrong direction will result in hitting the stage limit, which may require restarting the entire alignment process.

4.5 GCA Job Creation

There are two major things to take into consideration when creating a job intended for pieces that are different from normal job creation. The first is wafer leveling, and the second is setting up alignment die.

4.5.1 Leveling

Normally, the GCA will measure a z offset at three points around the wafer prior to alignment. This allows the focus to be adjusted for each die in case the wafer is not perfectly level. Because a piece of a wafer is not a full wafer, it is likely that at least one of these three points will no longer be over the piece. In this case, leveling will fail and the stepper will not allow the piece to be exposed. To get around this, leveling

must be disabled. This is done by setting the *Leveler Batch Size* in the job to -1.

It may be possible to level using three points on the piece paddle. The points used would be the two circles indicated in Figure 4.1 as *Leveling Points* as well as a part of the main sample stage. This has yet to be investigated, so further work on that will be required.

4.5.2 Alignment Die

When setting up a job it is common to use the standard alignment die. These would be whichever die fall under the two alignment objectives. In the case of a piece, these die can rarely be used. Therefore, two die must be chosen. It is possible to use any two die, or even use the same die for both. It is best however, to use two die on the same row that are as far apart as possible, in order to make rotation correction easier.

4.5.3 Dropout Die

Dropout die are die that will not be exposed. Their location in the array will instead be skipped. This can be useful for exposing pieces that do not fit a rectangular array. However, setting up dropout die requires defining one or more rectangles of die to exclude. It is uncommon to have many process runs with the same size and shape piece. Unless this is the exception, it is usually easier to manually instruct the stepper to continue processing every time a focus error is displayed when the stepper attempts to expose a die that does not fall on the piece.

Chapter 5

Well Width in Resonant Interband Tunnel Diodes

The RITD was initially described in papers by Collins *et al.* [12] and Ting *et al.* [11] as a way to improve the peak current density in HTDs. This study was performed to try to optimize the well width in order to achieve the highest peak current density.

Three InAs/GaSb/AlSb/GaSb RITD structures were grown (RITD-1–RITD-3) as well as a control InAs/GaSb HTD (CONTROL). The RITD structures were modeled after those discussed by Collins and Ting [11,12]. These papers described two different resonant interband structures. The first structure has a quantum well in the conduction band of the InAs, which is formed between an AlSb barrier, and the GaSb side of the junction. The other design consists of a well in the valence band of the GaSb layer. This well exists between an AlSb barrier and the InAs side of the junction. The latter structure, with the quantum well in the valence band of the GaSb layer, is the design used in this study. Details about the operation of these devices can be found in Section 2.2.5.

While the materials used are the same, there are two major differences between the devices discussed here and those discussed by Collins and Ting. First, the n++ dopant concentrations used in this study were two orders of magnitude higher than the referenced devices. This should increase current density by increasing the carriers available for tunneling. Second, the references used a well width of 5.1 nm. For this study, devices with well widths of 3 nm, 5 nm and 7 nm were fabricated and tested.

The well width effects the location of the discrete energy levels within the well. Energy levels are spaced farther apart in narrow wells than they are in wide wells. Adjusting the well width should therefore adjust the energy level where reflection is minimized, and, depending on the carrier distribution, could therefore have an impact on the tunneling current.

5.1 Layer Structures

The exact layer structures for each of the three RITD samples are given in Figures 5.1a–c, and the control HTD structure in (Figure 5.1d). These structures were grown at Texas State university by MBE. Energy band diagrams for each of these devices are provided in Figure 5.2. These band diagrams were simulated in BandProf [21].

For all structures, the top layer of InAs is n doped with a Te concentration of 10^{19} cm^{-3}, and the bottom GaSb is p doped with a Be concentration of 5×10^{18} cm^{-3}. As was previously stated, the Ting and Collins devices had a much lower n+ dopant concentration. The InAs layer for their devices was n doped with Si at a concentration of 2×10^{17} cm^{-3}, and the GaSb layer was p doped, also with Si,[1] at a concentration of 5×10^{18} cm^{-3}, the same concentration used here.

The effective density of states in the conduction band (N_C) of InAs is 1.18×10^{17} cm^{-3} [22]. Therefore, at a dopant concentration of 10^{19} cm^{-3}, the InAs is degenerately doped. Likewise, the effective density of states in the valence band (N_V) of GaSb is 6.63×10^{18} cm^{-3} [22]. Therefore, being doped at 5×10^{18} cm^{-3}, the GaSb is not degenerately doped. However, as previously discussed, this is not a problem because, being a broken gap system, there is already an overlap between the bands even without any doping. Additional material properties can be found in Appendix D.

[1] Group IV elements can act as either n type or p type when used as dopants in III-V semiconductors. Which they act as is dependent upon what element they replace in the lattice.

InAs	25 nm	1×10^{19} cm^{-3}
GaSb	3 nm	undoped
AlSb	1 nm	undoped
GaSb	300 nm	5×10^{18} cm^{-3}
GaSb	Substrate	

(a) RITD-1

InAs	25 nm	1×10^{19} cm^{-3}
GaSb	5 nm	undoped
AlSb	1 nm	undoped
GaSb	300 nm	5×10^{18} cm^{-3}
GaSb	Substrate	

(b) RITD-2

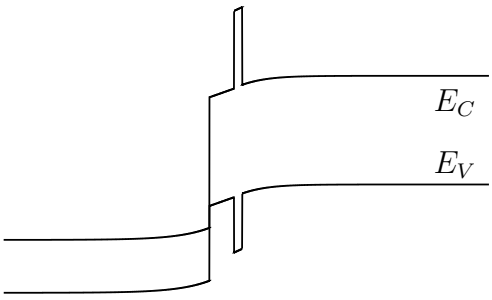
InAs	25 nm	1×10^{19} cm^{-3}
GaSb	7 nm	undoped
AlSb	1 nm	undoped
GaSb	300 nm	5×10^{18} cm^{-3}
GaSb	Substrate	

(c) RITD-3

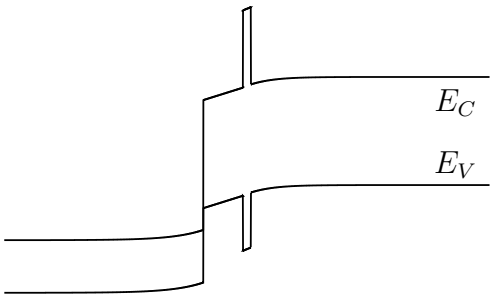
InAs	25 nm	1×10^{19} cm^{-3}
GaSb	3 nm	undoped
GaSb	300 nm	5×10^{18} cm^{-3}
GaSb	Substrate	

(d) CONTROL

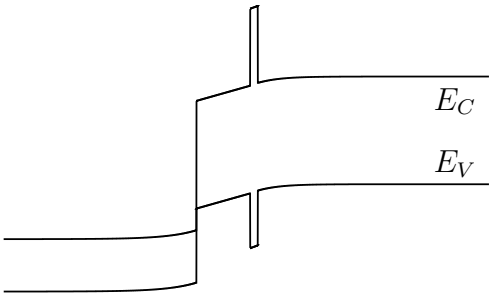
Figure 5.1: **(a)**, **(b)** and **(c)** Three RITD structures to be used in this experiment with varying well width. **(d)** Control tunnel diode structure. The undoped well is highlighted in blue (■) and the barrier is highlighted in red (■, –). For all structures, the n type dopant is Te and the p type dopant is Be



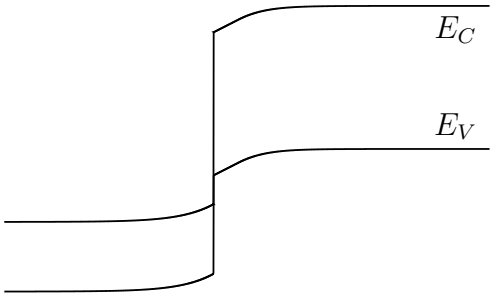
(a) RITD-1



(b) RITD-2



(c) RITD-3



(d) CONTROL

Figure 5.2: Simulated band edge diagrams for each of the three **(a–c)** RITD structures, and the control HTD **(d)**.

The GaSb well layers, as well as the AlSb barrier layers are undoped, as is the case with the Ting and Collins devices. The barrier thickness was a constant 1 nm for all three RITDs, where Ting and Collins used 1.2 nm. Finally, as previously mentioned, the GaSb well width was varied from the 5.1 nm in the reference. The well width in RITD-1 is 3 nm, RITD-2 is 5 nm and RITD-3 is 7 nm.

5.2 Fabrication Process

The process used to fabricate these devices is very similar to the process used to fabricate HTDs, outlined in Section 3.2. This process has been modified, with the addition of a few etch steps, to account for the AlSb layer that is present in RITDs, that is not present in HTDs. The process steps are summarized here, with the additional, or modified, steps marked (∗):

1. Top contact metalization
2. ∗Mesa lithography and etch
   (a) Lithography
   (b) Contact etch
   (c) ∗Mesa etch
3. ILD deposition and etch back
4. Probe pad lithography, metalization and liftoff

After top metalization, lithography and etch comes the mesa etch. The InAs is etched in 20:1 citric acid:H_2O_2 and the 25 nm layer is etched for 40 s. Following the mesa etch, the GaSb well region is etched, as well as the AlSb barrier. This etch is done in 1:4 NH_4OH:H_2O, and is stopped after clearing the AlSb barrier layer.

5.3 Current-Voltage

The basic operation of an RITD is the same as an HTD. The only difference being a change in the reflection coefficient of electrons approaching the junction. Therefore,

the current-voltage characteristic of an RITD is very similar to that of an HTD. Given in Figure 5.3 are I-V plots of each of the three RITD samples, as well as the CONTROL sample, for a variety of mask defined device sizes ranging from (500 nm)^2 to (2 µm)^2.

NDR is clearly visible on these curves, indicating that tunneling is occuring. Additionally, the current scales with area, as is expected for any diode. As is the case for a standard tunnel diode, these devices are highly conductive under reverse bias.

5.4 Peak Current Density

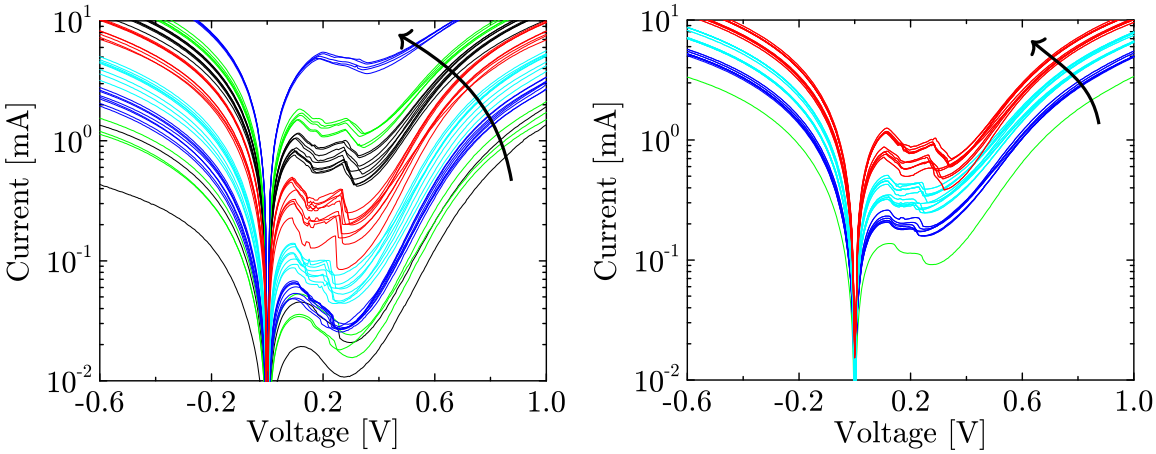
For all devices that exhibited NDR, the peak current density was extracted. The measured device area was used for this extraction, rather than the mask defined area, in order to improve the accuracy of the current density. Devices that did not exhibit NDR were assumed to be nonfunctional due to defects introduced during processing, or defects originating during the MBE growth and were therefore ignored. A histogram of these current densities for each sample is given in Figure 5.4a, and a box plot of the same in Figure 5.4b.

5.4.1 Stastical Analysis

From the box plot, it is obvious that CONTROL sample had a higher peak current density than any of the other RITD samples. The other samples however, were not so clear. In order to determine what, if any, other differences exist, a statistical analysis was performed on this data set.

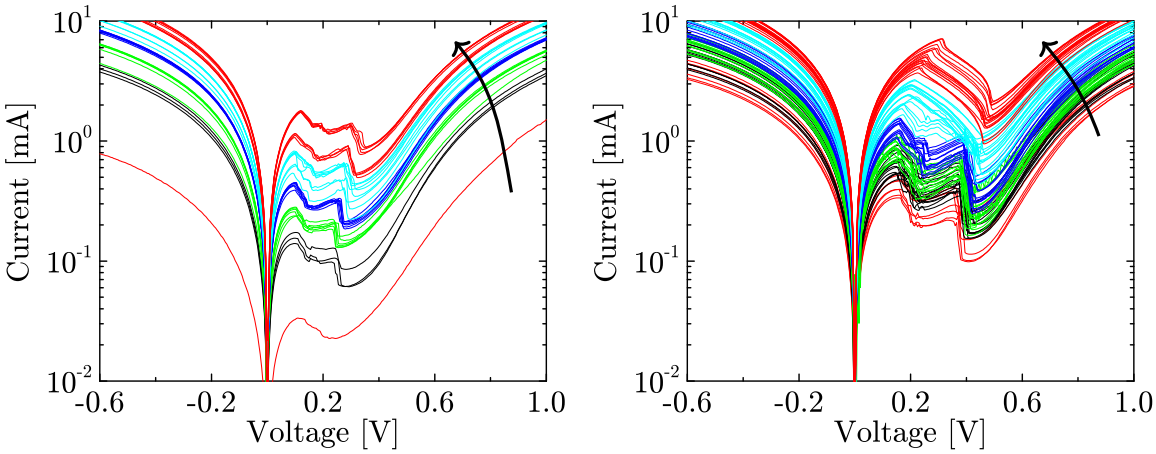
The test that was used to determine if there exists a significant difference between the means of this data set is the one-way ANOVA. Unsurprisingly, at $p < 10^{-10}$ there are significant differences between the means in this set.

The ANOVA results only show that there is a difference between some of the means, but this was already clear. CONTROL clearly has a higher average current than the other three. In fact, there is very little overlap between CONTROL and any of the



(a) RITD-1
Area: $(600\ \text{nm})^2$ – $(2\ \mu\text{m})^2$

(b) RITD-2
Area: $(700\ \text{nm})^2$ – $(1\ \mu\text{m})^2$



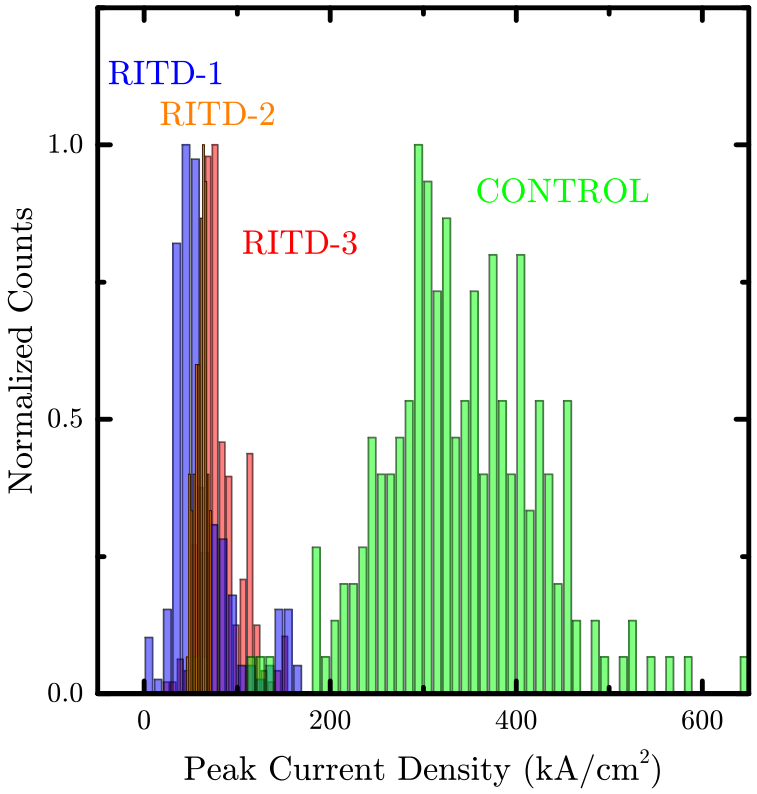
(c) RITD-3
Area: $(500\ \text{nm})^2$ – $(1\ \mu\text{m})^2$

(d) CONTROL
Area: $(500\ \text{nm})^2$ – $(1\ \mu\text{m})^2$

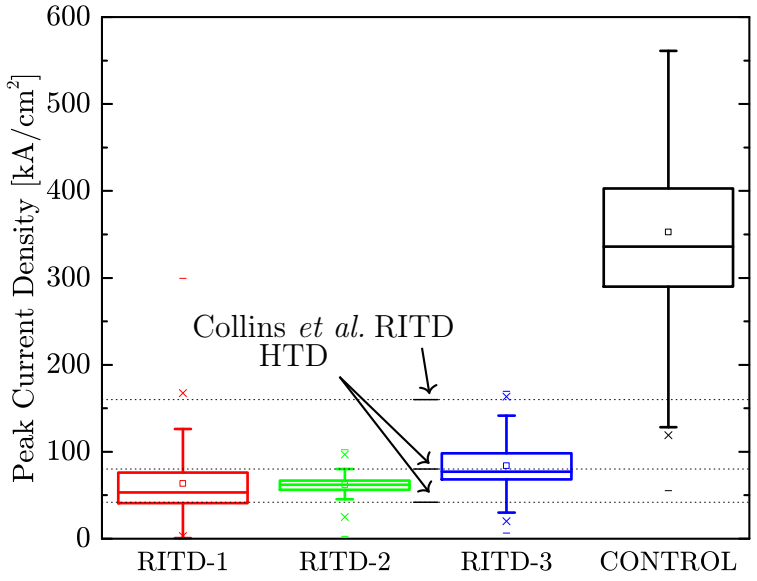
Figure 5.3: Current-voltage plot of for each of the RITD samples. Increasing area is indicated in each plot by an arrow. The areas shown here range from $(500\ \text{nm})^2$ to $(3\ \mu\text{m})^2$ and are mask defined areas. The actual areas are slightly smaller.

Table 5.1: One way ANOVA table for RITD mean peak current density indicating a significant difference between means ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$).

	DF	Sum of Squares	Mean Square	F Value	p Value
Model	3	1.246×10^{13}	4.154×10^{12}	668.2	$< 10^{-10}$
Error	756	4.700×10^{12}	6.217×10^{9}		
Total	759	1.716×10^{13}			



(a) Histogram of RITD peak current densities



(b) Box plot of RITD peak current densities

Figure 5.4: A histogram **(a)** and a box plot **(b)** comparing the peak current densities in each of the RITD samples. It is clear that the control HTD has the highest current density, and the other three RITD devices differ by very little, if at all. Additionally, the box plot has data from one RITD device and two HTD devices published by Collins *et al.* [12].

Table 5.2: Results from Tukey's range test on the peak current density of the RITD samples. These results show that the CONTROL sample is different from the rest. Additionally, there might be a significant difference between RITD-3 and RITD-1, depending on the acceptable level of type I error.

Samples	Difference	SEM	q Value	p Value	Significant?
RITD-2−RITD-1	-1472	9752	0.2135	0.9988	
RITD-3−RITD-1	20230	7676	3.727	0.04253	Maybe
RITD-3−RITD-2	21700	9258	3.315	0.08917	
CONTROL−RITD-1	289100	7860	52.00	$< 10^{-10}$	✓
CONTROL−RITD-2	290500	9412	43.66	$< 10^{-10}$	✓
CONTROL−RITD-3	268800	7239	52.52	$< 10^{-10}$	✓

other RITDs. What is of interest now, is identifying if any of the other three samples are significantly different from each other. This is much less clear, as they show significant overlap.

In order to determine which pairs of samples have significantly different means, Tukey's Range Test was used. At $p < 10^{-10}$, the average peak current density of CONTROL is significantly different than each of the other three samples. In the other extreme is the difference between RITD-2 and RITD-1, with a p value of 0.9988, this difference is not significant. The other two cases, however, are not so clear. For each the p value is close to 0.05, and depending on how much type I error is acceptable, either of these could be considered significant or not significant.

5.5 Comparison to Literature

Overlaid on top of Figure 5.4b are three data points. One corresponds to the peak current density of an RITD, and the other two correspond to the peak current density of HTDs. All three of these are from data published by Collins *et al.* [12]. In all three device structures, the InAs was doped at 2×10^{17} cm^{-3} and the GaSb at 5×10^{18} cm^{-3}. The higher current density HTD had 10 nm undoped InAs and GaSb layers at the junction.

First, the two Collins HTD samples had peak current densities right around those of the three RITD samples fabricated in this study. But then there is the Collins RITD sample, with twice as much current as the highest current density HTD reported. While the current density is certainly greater than the mean current density in the RITD-1–RITD-3 samples, there are a few outliers that actually had a higher current density.

All of the samples reported by Collins have significantly lower current densities than the CONTROL sample. At first this may be unexpected, as all of the HTDs reported by Collins had lower current densities than the RITD [12]. However, these results may make sense if the impact doping has on the electron distribution in the InAs conduction band is considered.

5.6 Explination of Results

Since this experiment did not perform as was originally desired, it is important to try to figure out why it did not. The only major difference between the samples fabricated for this study, and those reported by Collins, is an increase in the InAs dopant concentration of about two orders of magnitude. Normally, an increase in dopant concentration should increase the density of electrons available for tunneling. Overall, it is expected that an increase in dopant concentration would cause an increase in current.

A comparison of CONTROL with the HTDs reported by Collins suggests that the increase in dopant concentration did in fact result in an increase in current density, at least for the HTD structure. A comparison between the Collins RITD and the RITDs fabricated here did not show the same change.

In an HTD, electrons of any energy can generally contribute to the tunneling current, as long as they are in the conduction band/valence band overlap region, and there is a state to tunnel into. In an RITD however, this is not the case. There is

an increasing reflection coefficient for electrons the further away their energy is from one of the discrete levels within the well. Because of this, many of the electrons that could normally contribute to current in an HTD, cannot contribute to current in the RITD.

Increasing the dopant concentration does not have a uniform increase in electron concentration for all energies. The electron concentration is the product of the electron density of states, and the Fermi-Dirac distribution. Increasing the dopant concentration shifts the Fermi distribution, but the electron density of states still limits the total electron concentration at a given energy. Due to the nature of the Fermi distribution, the density of electrons very close to the conduction band edge will approach the density of states limit at lower dopant concentrations than would be required for higher energy electrons.

If the bound energy level in the quantum well is close to the InAs conduction band edge, it is unlikely that an increase in dopant concentration would have a significant increase in the electron concentration at that level. This is demonstrated in Figure 5.5 where three electron energy distributions are shown that are calculated from three different dopant concentrations. Since it is likely that the energy level in the quantum well is quite close to the InAs conduction band edge, it is likely that at increase in dopant concentration has very little effect on the number of electrons available for tunneling. This is especially true when the lower of the two dopant concentrations is already quite high. It is because the increase in dopant concentration does not cause an increase in carrier concentration at the required energy level that results in no current density improvement when using higher dopant concentrations.

The difference in current density in RITD-3 compared to RITD-1 or RITD-2 was borderline significant. If we assume it was a significant difference, this increase in current could make sense. Note however, that more data should be collected to be sure before making any design decisions based on this information.

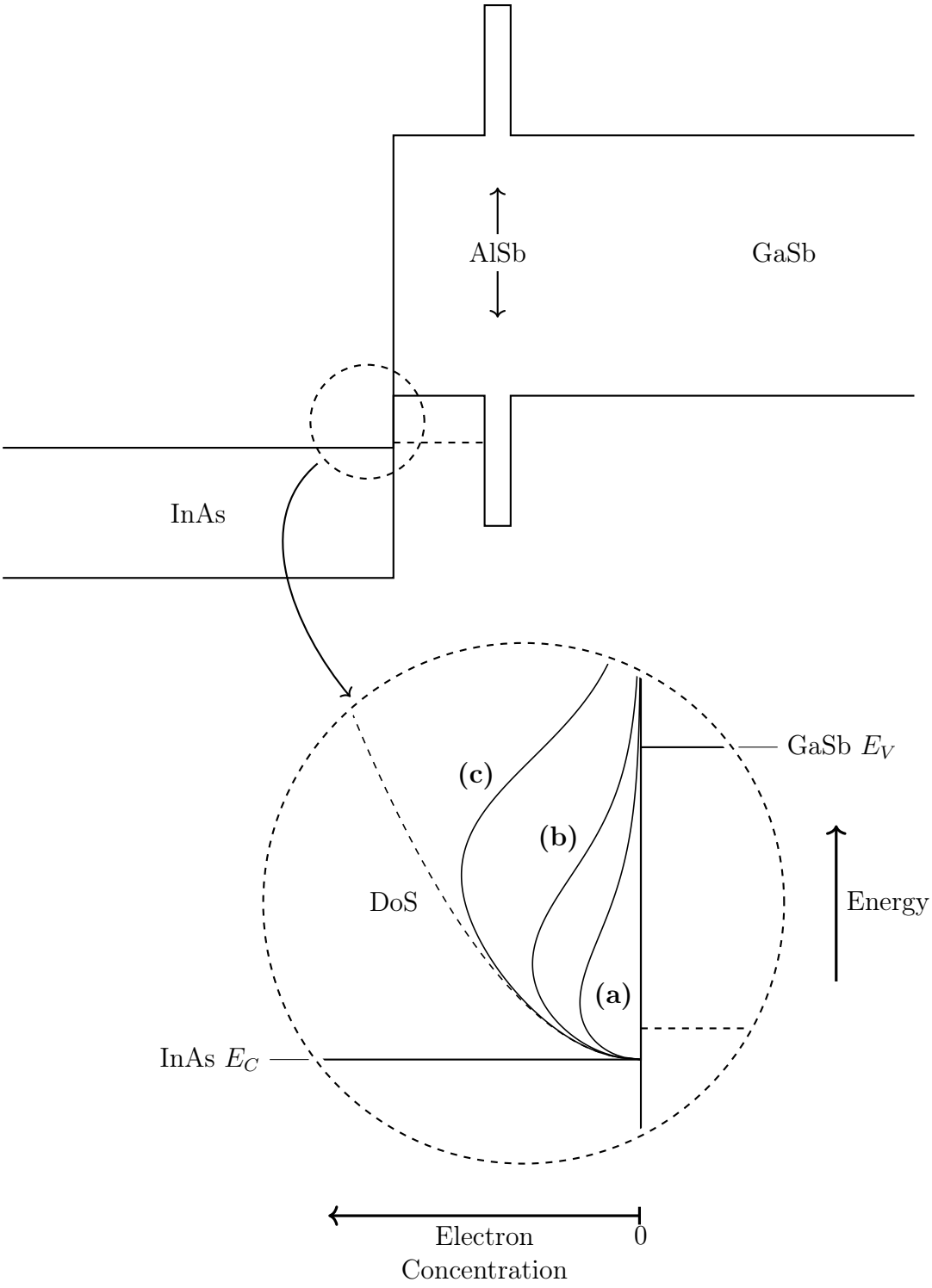


Figure 5.5: RITD band diagram showing calculated electron distributions in InAs for varying dopant concentrations. It is assumed that the valence band edge of GaSb is 0.15 eV higher than the conduction band edge of InAs. For **(a)** $N_D = 2 \times 10^{17}$ cm^{-3}, for **(b)** $N_D = 4.5 \times 10^{17}$ cm^{-3} and for **(c)** $N_D = 10^{18}$ cm^{-3}. The density of states (DoS) is also shown which serves as the maximum of the electron distribution. For the energy domain shown here, the DoS is indistinguishable from $N_D = 2 \times 20^{18}$ cm^{-3}.

The well width for RITD-3 was the widest of the three RITD samples. It is well known that as a quantum well increases in width, the energy levels shift to lower energies. In the case of a quantum well in the valence band, as is the case for these RITD devices, this is a shift toward the valence band edge. Since this energy level must be above the conduction band edge of the InAs for these devices to function properly, the result of increasing the well width is a shift away from the InAs conduction band edge. As the energy level moves in this direction, it lines up with an increasing electron concentration in the InAs conduction band, due to the increasing electron density of states. The result is more electrons that can now tunnel through this state, and therefore an increase in tunneling current. It should be considered though that as more electrons become available to tunnel, fewer states in the valence band of the GaSb are made available to tunnel into. Therefore, there is an optimal well width that depends on the dopant concentrations on either side of the well.

Chapter 6

Undoped InAs/GaSb Heterojunction Tunnel Diode

An experiment was designed to test the impact n and p contact doping on InAs/GaSb HTDs with undoped junctions. In each case there was a 10 nm undoped InAs layer, and a 100 nm undoped GaSb layer to form the junction. The structures designed here are similar to the p-i-n structure used to make a TFET. Learning more about these structures could help design better TFETs in the future.

6.1 Layer Structures

The device design was modeled after a similar device published by Collins *et al.* [23]. In that paper, InAs/GaSb HTDs were fabricated with 10 nm undoped InAs and 10 nm undoped GaSb spacers, as well as without those spacers. In both cases, Si was used as both an n type and a p type dopant. The InAs was doped at 2×10^{17} cm^{-3}, and the GaSb at 5×10^{18} cm^{-3}.

For this experiment a 100 nm undoped GaSb layer was used in addition to the same 10 nm undoped InAs layer used in the Collins devices. The GaSb doping was reduced to 10^{18} cm^{-3}, which, like in the case of the RITDs is not degenerate. For the InAs doping, one device is doped at 10^{7} cm^{-3}, which is just below N_C in InAs, making this device not degenerately doped on either side of the junction. The other two devices however are doped above N_C and are therefore degenerately doped.

The layer structures for the four samples are provided in Figure 6.1. These samples

have the following two parameters varied:

- 200 nm of GaSb doped with Be at 10^{18} cm^{-3} or undoped. ( )
- Top 15 nm of InAs doped with Te at 10^{17}, 10^{18}, 10^{19} cm^{-3} or undoped. ( )

These four structures were simulated in BandProf to produce the band diagrams shown in Figure 6.2 [21]. Diodes were fabricated from these structures using the process outlined in Section 3.2.

6.2 Electrical Characteristics

It is of first importance to make sure these devices act as tunnel diodes. Representative I-V curves for each of the four layer structures are provided in Figure 6.3. These plots show diodes with mask defined areas of (1.35 µm)^2 through (3 µm)^2. In all cases NDR is clearly visible and therefore these devices are functioning as tunnel diodes.

6.2.1 PVCR

One characteristic of tunnel diodes that is commonly compared is the PVCR. While the measured current is a function of area, the PVCR is not, and is therefore useful when comparing devices of varying size.

In order to assess the effect the changes in doping had on the devices, the diodes were tested and PVCR was extracted from the measured I-V curves. A box plot of the extracted PVCR for each of these four structures is shown in Figure 6.4. There was difficulty taking I-V measurements of HTD-0 due to residual photoresist on the mesas during probing. This likely accounts for the high variation in extracted PVCR in HTD-0.

In order to determine if there are significant differences between the average PVCR of each sample, a one way ANOVA test was run [24]. The results of this test are provided in Table 6.1. As shown in this table, the p value is 6.438×10^{-6}, and the null hypothesis that all samples share the same mean is therefore rejected.

InAs	15 nm	1×10^{19} cm^{-3}
InAs	10 nm	undoped
GaSb	100 nm	undoped
GaSb	200 nm	1×10^{18} cm^{-3}
GaSb	Substrate	

(a) HTD-0

InAs	15 nm	1×10^{18} cm^{-3}
InAs	10 nm	undoped
GaSb	100 nm	undoped
GaSb	200 nm	1×10^{18} cm^{-3}
GaSb	Substrate	

(b) HTD-1

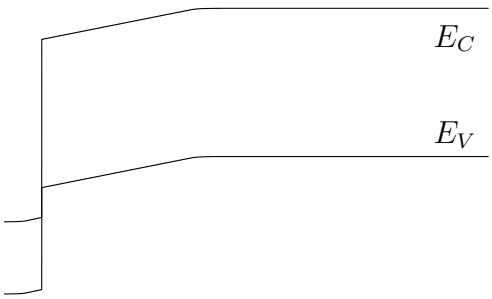
InAs	15 nm	1×10^{17} cm^{-3}
InAs	10 nm	undoped
GaSb	100 nm	undoped
GaSb	200 nm	undoped
GaSb	Substrate	

(c) HTD-2

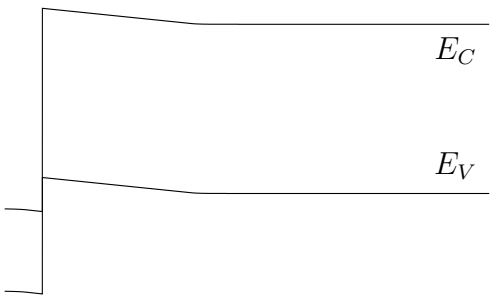
InAs	15 nm	undoped
InAs	10 nm	undoped
GaSb	100 nm	undoped
GaSb	200 nm	undoped
GaSb	Substrate	

(d) HTD-3

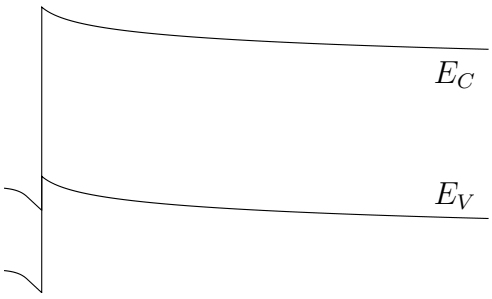
Figure 6.1: **(a)** and **(b)** are HTD structures with a 100 nm undoped GaSb region at the junction on a 200 nm p+ (10^{18} cm^{-3}) GaSb layer (■). **(c)** and **(d)** do not have the doped p+ GaSb layer, instead the 200 nm GaSb layer is undoped. Structures **(a)**–**(c)** have a 15 nm n+ InAs layer with varying dopant concentrations (■) on top of a 10 nm undoped InAs region at the junction. **(d)** is completely undoped. In all cases, Te is used as the n type dopant and Be as the p type dopant.



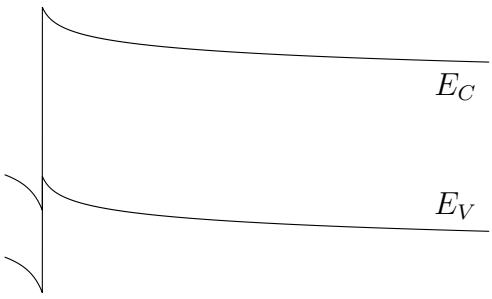
(a) HTD-0



(b) HTD-1

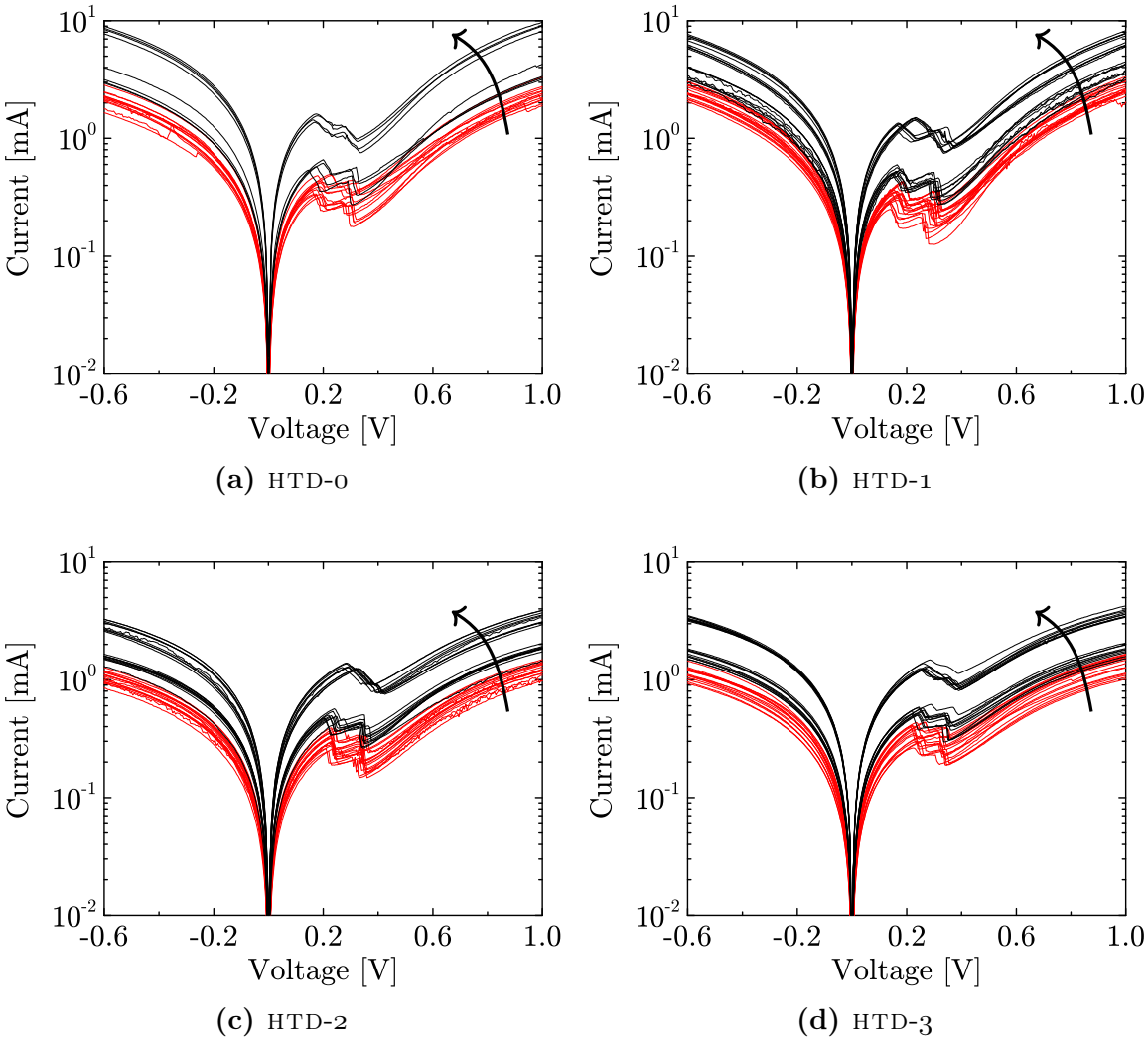


(c) HTD-2



(d) HTD-3

Figure 6.2: Simulated band edge diagrams for each of the four HTD structures.



(a) HTD-0 **(b)** HTD-1 **(c)** HTD-2 **(d)** HTD-3

Figure 6.3: Current-voltage plot of for each of the HTD samples. Increasing area is indicated in each plot by an arrow. The areas shown here range from (1.35 µm)^2 to (3 µm)^2 and are mask defined areas. The actual areas are slightly smaller.

Table 6.1: One way ANOVA table for HTD mean PVCR indicating a significant difference between means. ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$)

	DF	Sum of Squares	Mean Square	F Value	p Value
Model	3	1.984	0.6613	14.38	6.438×10^{-9}
Error	403	18.53	0.04599		
Total	406	20.52			

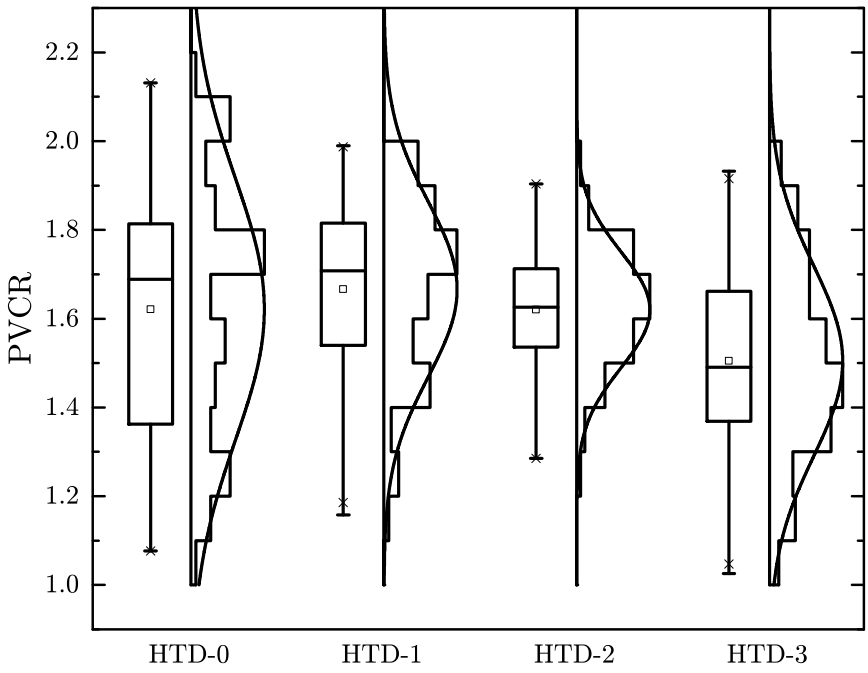


Figure 6.4: A box plot of the PVCR for each HTD sample. Whiskers indicate maximum and minimum value within 1.5 IQR of the box.

Next, to determine which means differ, Tukey's range test was used. The results of this test are provided in Table 6.2. It is shown here that the PVCR of HTD-3 is different from all other HTDs, and none of the other differences are significant.

The results of this experiment indicate that the doping in the top 15 nm InAs layer increases PVCR, over layer structures that have an undoped InAs layer instead. The amount of doping however, has not been shown to have any effect. Due to the

Table 6.2: Results from Tukey's range test on the PVCR of the HTD samples. These results show that the PVCR of HTD-3 is different from all other HTD samples. No other PVCR differences are significant.

Samples	Difference	SEM	q Value	p Value	Significant?
HTD-1−HTD-0	0.0449	0.0328	1.94	0.519	
HTD-2−HTD-0	−0.0013	0.0384	0.0486	1.00	
HTD-2−HTD-1	−0.0462	0.0337	1.94	0.517	
HTD-3−HTD-0	−0.117	0.0316	5.21	0.00149	✓
HTD-3−HTD-1	−0.1614	0.0257	8.88	9.222×10^{-10}	✓
HTD-3−HTD-2	−0.115	0.0326	5.01	0.0025	✓

relatively heavy doping concentration used for all device structures that had doping, it is expected that changes in PVCR would be noticeable for doping concentrations much lower than those used here.

The doping of the bottom 200 nm GaSb layer has not been shown to have any impact on diode performance. This is likely due to the much greater distance from the junction to the doped GaSb layer relative to the distance from the junction to the doped InAs layer.

The range of the PVCR for HTD-0 was wider than the others. Likewise, the PVCR histogram for HTD-0 was much more flat than the histograms for the other samples, which fit a normal distribution much more closely. After etching the mesa, a film remained on top the metal which became apparent during electrical testing. This film prevented accurate measurements from being taken and this was reflected in the data shown here. For this reason, data collected from HTD-0 will be excluded from any future analysis.

6.2.2 Peak and Valley Current Densities

The PVCR does not show everything. It is possible for there to be no change in PVCR but still a large change in the peak or valley current densities. Doping conditions, for example, can have a large impact on the peak and valley current densities, but have little to no effect on the PVCR [10]. When changing doping conditions, the tunneling probability is changed. When the tunneling probability is increased the peak current density is increased, but so is the valley current density. As excess current is primarily a tunneling current [10], increasing the tunneling probability to achieve a higher peak current density will also increase the excess current density, which will have a corresponding effect on the valley current density.

The peak and valley currents for each sample are shown in the boxplot in Figure 6.5. Again, a one-way ANOVA is used to determine if there are significant differences. A

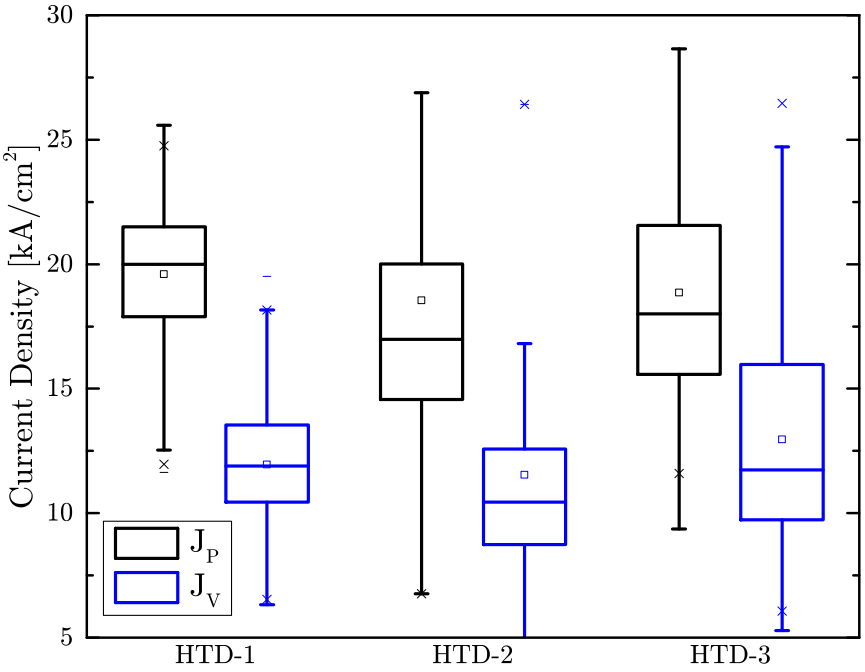


Figure 6.5: Peak and valley current densities of HTD-1, HTD-2 and HTD-3

separate test was run for the peak current density and the valley current density. The results from these tests are shown in Tables 6.3a and 6.3b. For the peak current density, a p value of 0.342 was calculated. Therefore, it is determined that there is no significant variation in the peak current densities of these three samples. However, for the valley current density the result is different. Here, a p value of 0.0276 was calculated. This is a borderline value that indicates possibly significant results.

Now, considering the results of the valley current density ANOVA, we look to Tukey's test again to determine which specific samples were different. The results of this test are given in Table 6.4. It is interesting that while the base ANOVA gave a significant result for $\alpha = 0.05$, Tukey's test did not show any individual differences, although the HTD-3–HTD-2 comparison is quite close.

At first it may seem strange that the peak and valley current densities did not show any significant differences from each other, but the PVCR, a value calculated from the peak and valley currents, did. This is likely due to increased error in the current densities as compared to the PVCR.

Table 6.3: ANOVA tables for HTD **(a)** peak and **(b)** valley current densities

(a) Peak current density

	DF	Sum of Squares	Mean Square	F Value	p Value
Model	2	5.76×10^7	2.88×10^7	1.08	0.342
Error	337	9.02×10^9	2.68×10^7		
Total	339	9.08×10^9			

(b) Valley current density

	DF	Sum of Squares	Mean Square	F Value	p Value
Model	2	1.17×10^8	5.87×10^7	3.63	0.0276
Error	337	5.45×10^9	1.62×10^7		
Total	339	5.57×10^9			

Table 6.4: Tukey's range test table for HTD valley current density

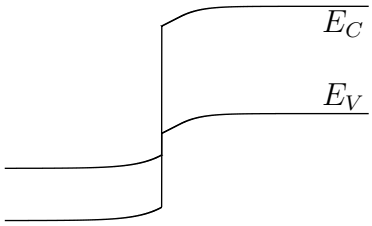
	Difference	SEM	q Value	p Value	Significant?
HTD-2–HTD-1	-407	632	0.911	0.796	
HTD-3–HTD-1	1010	483	2.97	0.0919	
HTD-3–HTD-2	1420	611	3.29	0.0540	

The PVCR is calculated from the peak and valley currents, not the peak and valley current densities. Because of this, there is no measurement error arising from the area measurements in the PVCR data, while there is in the current density data. Even if the PVCR was calculated from the current densities, the measured area would divide out and wouldn't add to the PVCR error anyway.

6.3 Undoped vs. Doped HTD

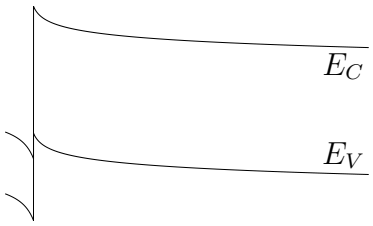
Since the CONTROL sample is really just an HTD, it is of interest to compare it to the HTD samples. Since a statistical analysis failed to show any significant differences among the peak or valley current densities of the HTD samples, any one of them can represent the group in a comparison against the CONTROL sample. In this case, the undoped HTD (HTD-3) will be used. For convenience, the layer structures, as well as

InAs	25 nm	1×10^{19} cm^{-3}
GaSb	3 nm	undoped
GaSb	300 nm	5×10^{18} cm^{-3}
GaSb	Substrate	



(a) Doped HTD (CONTROL)

InAs	15 nm	undoped
InAs	10 nm	undoped
GaSb	300 nm	undoped
GaSb	Substrate	



(b) Undoped HTD (HTD-3)

Figure 6.6: Layer structure and band diagrams for **(a)** doped and **(b)** undoped HTDs.

the simulated band diagrams, are given again in Figure 6.6.

A comparison is made between the peak and valley current densities of the HTD-3 sample and the CONTROL sample. This is shown as a boxplot in Figure 6.7. The three data points from Collins *et al.* are also included for reference. Here it is clear that the highly doped HTD had a significantly higher peak current density than the Collins HTDs. However, the Collins HTDs had higher peak current densities than the undoped HTD.

These results can be interpreted to mean that tunneling current density increases as dopant concentration increases, as long as the dopant is near the junction. The Collins devices were doped at 2×10^{17} cm^{-3} and the CONTROL device was doped at 10^{19} cm^{-3}. It is likely that the peak current density could increase even more with a higher dopant concentration, but this is likely close to the highest peak current density possible in these devices. The undoped HTD serves as a minimum peak current density. This information could be of use in the design of TFETs in order to make educated guesses about TFET current densities. However, a TFET is a surface current device and a tunnel diode is a bulk current device. Therefore, it is still difficult to directly compare the two current densities. But estimates can still be made by considering the thickness of the channel region.

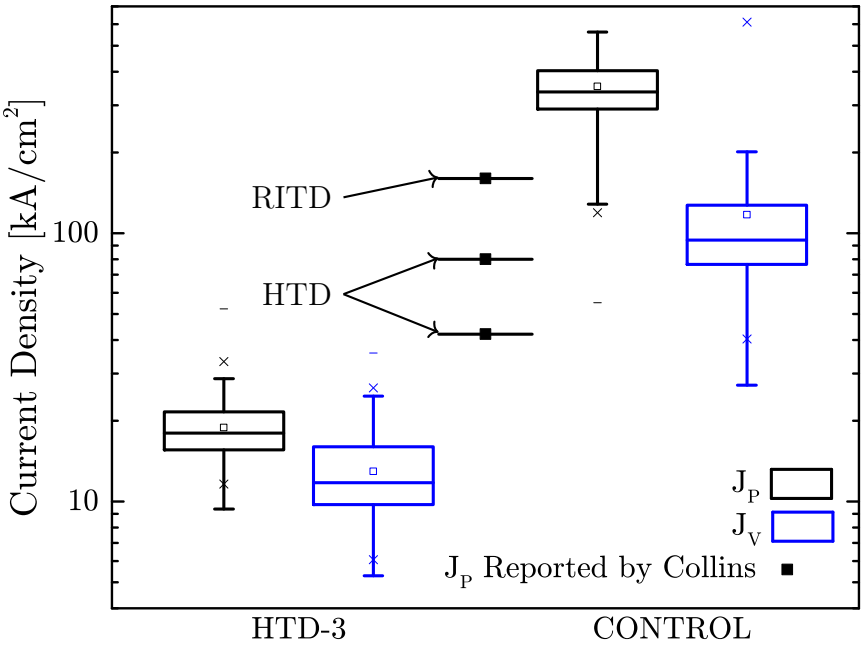


Figure 6.7: A box plot comparing the peak and valley current densities of HTD-3 with CONTROL. Note, the currents shown for HTD-3 are representative of HTD-1–HTD-3, not just the undoped one. Also indicated is the current density from the RITD, and the two HTDs, reported by Collins *et al.* [12]

Chapter 7

Final Remarks

7.1 Summary of Work

This work has refined the process for fabricating tunnel diodes. Specifically, a greater understanding of the alignment system in the GCA stepper lead to a more robust process that allows future researchers to rely on optical lithography, even when exposing only a single die. Any such processes were previously done by e-beam direct write, out of necessity, and not necessarily because of any of the specific benefits of using e-beam. This can reduce the time required to process a piece, and thereby increase throughput. It also will give researchers more flexibility in their process design, which could lead to better devices in the end.

An investigation into an alternate interband tunnel device, the Resonant Interband Tunnel Diode (RITD), was completed. This experiment attempted to investigate how changes in the width of the quantum well effect the peak current density in the hope of improving current density over the more standard HTD design. This work failed to show any improvements in peak current density. It did however, show borderline significant changes in current density as a function of well width.

Another investigation into HTDs with large undoped regions near the junction was performed. This experiment attempted to investigate possible TFET designs, and how changes in the dopant concentration away from the junction could effect peak current

density. This experiment however failed to show any differences in the performance of the four layer structures, except for the structure without any doping whatsoever. This structure showed a slightly lower PVCR than the other three structures. Besides this, there were no other differences.

Finally, a comparison was made between an HTD with dopant near the junction (CONTROL), and one without dopant near the junction (HTD-3). This comparison clearly showed an improvement in current density as dopant was added near to the junction. A comparison was also made between these samples and values published in literature. The literature devices were doped, but with a lower concentration used in this study. The devices from literature presented a lower current than the doped sample used here (CONTROL), but higher current than the undoped sample (HTD-3).

The results of the undoped vs. doped experiment do provide a minimum expected current density for InAs/GaSb tunnel diodes. This could be used to estimate a minimum expected current density from TFETs, if they were to be fabricated using these materials.

7.2 Future Work

As the RITD experiment did not yield concrete results in terms of well width and its impact on peak current density. It appears unlikely that these devices would be able to yield a higher current density than the standard HTD design. However, to be sure, additional experiments could be performed.

The individual layers, when grown by MBE, may not have been the exact thickness that was desired. If, for example, the barrier layer was 2 nm instead of 1 nm thick, it could result in a considerable decrease in current. In order to verify layer thicknesses, SIMS should be used. This can, and should, be done on the existing pieces, and not on newly grown samples. Newly grown samples could have different layer thicknesses than the ones used here, which would diminish the importance of the results.

As previously mentioned, the 7 nm well width sample had a borderline significant increase in current density. A targeted experiment could be designed around this size. For this experiment, well widths of 6, 7 and 8 nm could be used. The variation in well width should not be too great, as extreme values are unlikely to reflect how the device would behave with only slight variations in well width.

The layer structures used for the HTD experiment had a large undoped GaSb layer near the junction, as well as a smaller undoped InAs layer. An experiment could be ran that reduces the undoped InAs layer while maintaining the large undoped GaSb region. This might better reflect how a TFET would actually be fabricated. Additionally, moving the dopant closer to the junction is likely to increase current density.

Bibliography

[1] K. P. Cheung, "On the 60 mV/dec @300 K limit for MOSFET subthreshold swing," in *International Symposium on VLSI Technology Systems and Applications*. IEEE, Apr. 2010, pp. 72–73.

[2] A. M. Ionescu and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches," *Nature*, vol. 479, Nov. 2011.

[3] A. C. Seabaugh and Q. Zhang, "Low-voltage tunnel transistors for beyond cmos logic," *Proc. IEEE*, vol. 98, no. 12, pp. 2095–2110, Dec. 2010.

[4] S. M. Sze, *The Physics of Semiconductor Devices*, 3rd ed. A John Wiley & Sons, Inc., 2007.

[5] Q. Zhang, W. Zhau, and A. Seabaugh, "Low-subthreshold-swing tunnel transistors," *IEEE Electron Device Lett.*, vol. 27, no. 4, pp. 297–300, Apr. 2006.

[6] H. Zhao *et al.*, "Improving the on-current of $In_{0.7}Ga_{0.3}As$ tunneling field-effect-transistors by p++/n+ tunneling junction," *J. Appl. Phys. Lett.*, vol. 98, p. 093501, 2011.

[7] D. Neamen, *An Introduction to Semiconductor Devices*. New York, NY: McGraw-Hill, 2006.

[8] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III-V compound semiconductors and their alloys," *Journal of Applied Physics*, vol. 89, no. 11, pp. 5815–5875, Jun. 2001.

[9] L. Esaki, "Discovery of the tunnel diode," *IEEE Trans. Electron Devices*, vol. ED-23, no. 7, pp. 644–647, Jul. 1976.

[10] A. G. Chynoweth, W. L. Feldmann, and R. A. Logan, "Excess tunnel current in silicon and Esaki junctions," *Physical Review*, vol. 121, no. 3, Feb. 1961.

[11] D. Z.-Y. Ting *et al.*, "Large peak current densities in novel resonant interband tunneling heterostructures," *Appl. Phys. Lett.*, vol. 57, no. 12, pp. 1257–1259, Sep. 1990.

[12] D. A. Collins *et al.*, "Interband tunneling in InAs/GaSb/AlSb heterostructures," *Journal of Crystal Growth*, vol. 111, no. 1, pp. 664–668, 1991.

[13] B. R. Romanczyk, "Fabrication and characterization of III-V tunnel field-effect transistors for low voltage logic applications," MS thesis, Rochester Institute of Technology, Aug. 2013.

[14] J. C. Nabity, *Nanometer Pattern Generation System*, Sep. 2005.

[15] "AZ nLof 2000 series i-line photoresists," http://www.microchemicals.com/micro/AZ_nLOF2000_bulletin.pdf, AZ Electronic Materials, Dec. 2002.

[16] "Microposit MF CD-26 developer MSDS," Rohm & Haas Electronic Materials LLC, Jan. 2004.

[17] G. C. DeSalvo, W. F. Tseng, and J. Comas, "Etch rates and selectivities of citric acid/hydrogen peroxide on GaAs, $Al_{0.3}Ga_{0.7}As$, $In_{0.2}Ga_{0.8}As$, $In_{0.53}Ga_{0.47}As$, $In_{0.52}Al_{0.48}As$, and InP," *J. Electrochem. Soc.*, vol. 139, no. 3, pp. 831–835, Mar. 1992.

[18] "Cyclotene 3000 series advanced electronic resins," http://www.dow.com/cyclotene/docs/cyclotene_3000_dry_etch.pdf, Dow Chemical Company, Apr. 2008.

[19] A. Gaur, "Surface treatments to reduce leakage currant in homojunction $In_{0.53}Ga_{0.47}As$ PIN diodes for TFET," Master's thesis, Rochester Institute of Technology, 2015.

[20] "6000 series DSW wafer stepper instrument instruction manual," GCA Corporation, Apr. 1985.

[21] W. R. Frensley, "Bandprof," https://courses.ece.ubc.ca/480/priv/BandProf_11.jar, UT Dallas, Jun. 2003.

[22] D. J. Pawlik, "Comprehensive mapping and benchmarking of Esaki diode performance," PhD dissertation, Rochester Institute of Technology, 2013.

[23] D. A. Collins *et al.*, "Experimental observation of negative differential resistance from an InAs/GaSb interface," *Appl. Phys. Lett*, vol. 57, no. 7, pp. 683–685, Aug. 1990.

[24] T. B. Barker, *Quality by Experimental Design*, 3rd ed. CRC Press, 2005.

Acronyms

ANOVA Analysis of Variance.

BCB divinylsiloxane-bis-benzocyclobutene.

BTBT Band to Band Tunneling.

CMOS Complementary Metal-Oxide-Semiconductor.

GCA GCA 6300B DSW wafer stepper.

HTD Heterojunction Tunnel Diode.

ILD Inter-Layer Dielectric.

MBE Molecular Beam Epitaxy.

MOSFET Metal-Oxide-Semiconductor Field-Effect Transistor.

NDR Negative Differential Resistance.

NPGS Nanometer Pattern Generation System.

PGMEA Propylene Glycol Methyl Ether Acetate.

PVCR Peak to Valley Current Ratio.

RIT Rochester Institute of Technology.

RITD Resonant Interband Tunnel Diode.

SEM Scanning Electron Microscope.

SIMS Secondary Ion Mass Spectrometry.

SMFL Semiconductor and Microsystems Fabrication Lab.

TFET Tunneling Field-Effect Transistor.

TMAH Tetramethylammonium Hydroxide.

Appendix A: Internal names for devices discussed in this document

The names given to the samples discussed in this document were chosen for ease of understanding by the reader. Listed in Table A.1 are the names used internally, the grower, as well as the growth number.

Table A.1: Internal names for each sample discussed in this document.

Name	Internal Name	Grower	Growth Number
RITD-1	TD11-C	TSU	7-961
RITD-2	TD11-E		7-962
RITD-3	TD11-B		7-963
CONTROL	TD11-D		7-960
HTD-0	PIN7-B	TSU	7-1003
HTD-1	PIN7-C		7-1004
HTD-2	PIN7-E		7-1005
HTD-3	PIN7-D		7-1010

Appendix B: Tunnel diode layout for e-beam direct write.

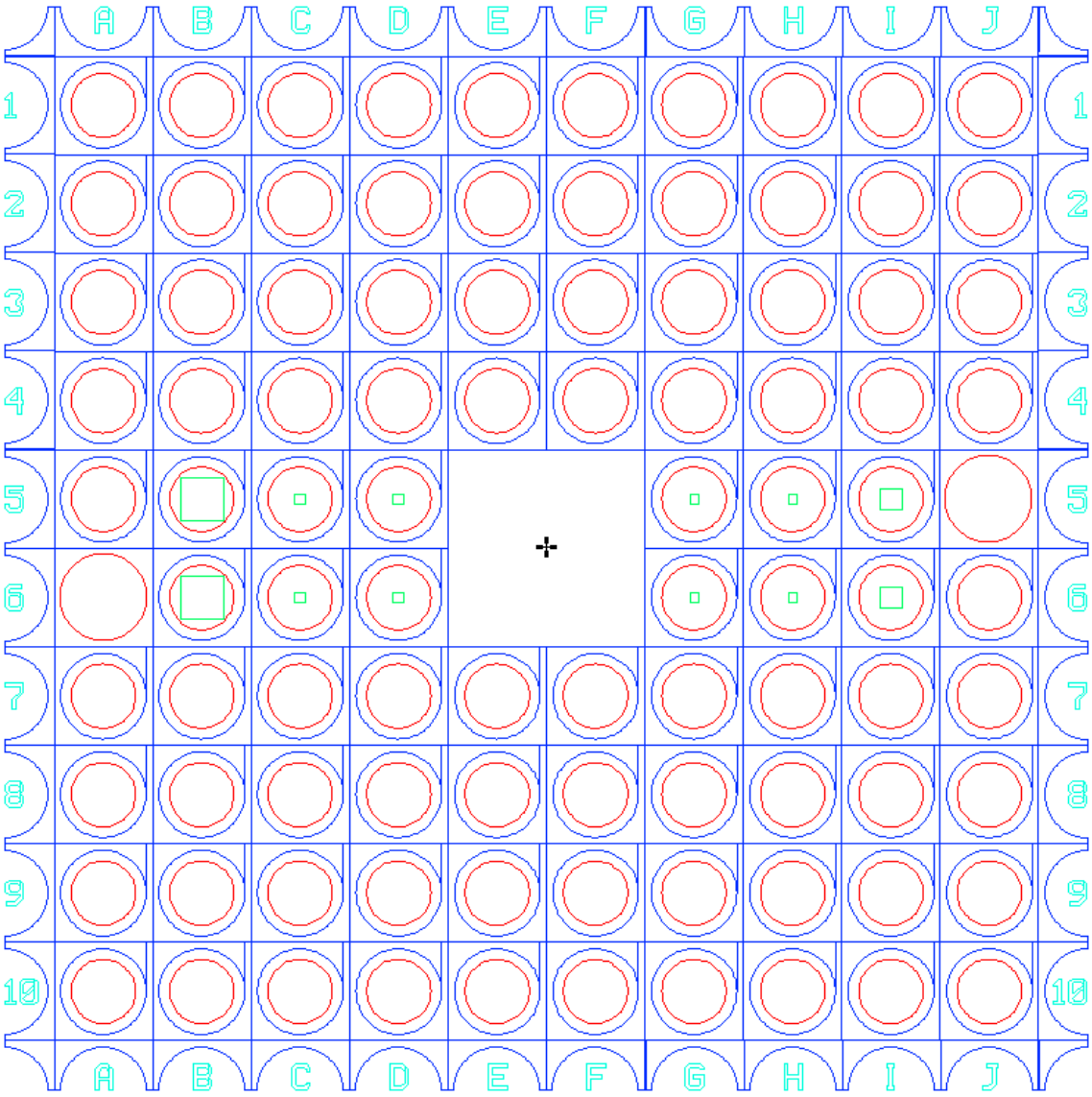


Figure B.1: Original tunnel diode e-beam layout, without alignment marks.

Appendix C: Tunnel diode hybrid lithography layout.

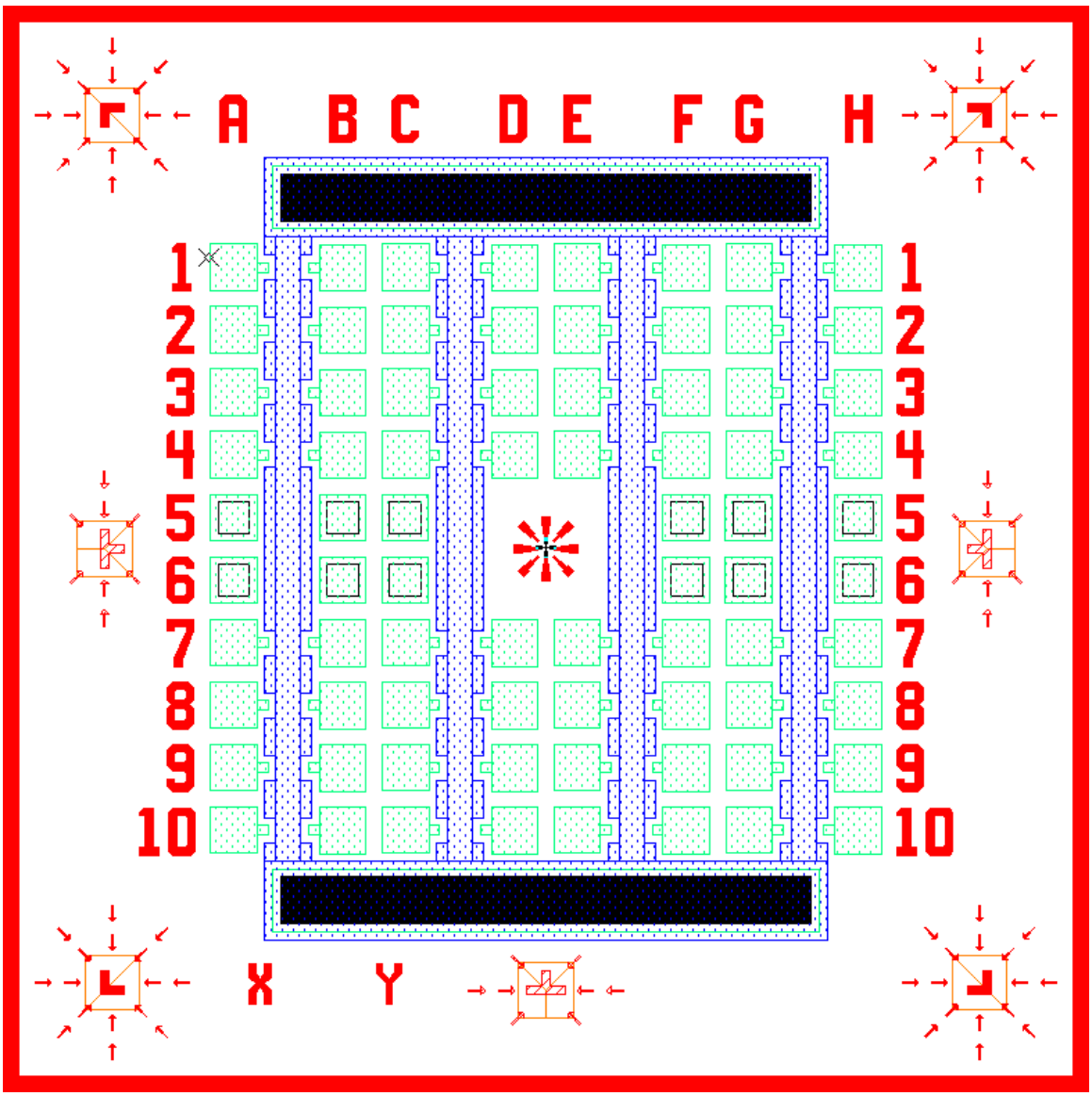


Figure C.1: New tunnel diode hybrid optical/e-beam layout.

Appendix D: Properties of materials

Table D.1: Electronic properties of some common semiconductor materials [22]

	E_g	ϵ_r	m_e^*		m_h^*			m^*_{BTBT}	N_C	N_V
	[eV]		DOS	Cond.	DOS	Cond.	Light	Heavy		$[10^{18}\ \mathrm{cm}^{-3}]$	$[10^{18}\ \mathrm{cm}^{-3}]$
Si	1.12	11.68	1.08	0.26	0.81	0.39	0.16	0.49	0.16	28.2	18.3
Ge	0.66	16	0.56	0.12	0.29	0.21	0.044	0.28	0.076	10.5	3.92
GaAs	1.42	12.91	0.063	0.063	0.53	0.39	0.082	0.51	0.054	0.397	9.73
GaSb	0.73	15.7	0.041	0.041	0.41	0.31	0.050	0.40	0.036	0.208	6.63
InAs	0.35	15.15	0.028	0.028	0.41	0.33	0.026	0.41	0.026	0.118	6.69
InSb	0.18	16.8	0.02	0.02	0.43	0.36	0.015	0.43	0.019	0.0710	7.12