Modelling and Characterisation
of 2-Terminal
Heterojunction Phototransistors

Hwee Ngoh ANG

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

UniS

School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey, GU2 7XH, UK

July 2005

© JULIE ANG 2005
To my better half and soul-mate

\[ X + X \rightarrow X^2 \]

The wisdom of the masses exceeds that of the wisest individual
ABSTRACT

A Heterojunction Phototransistor (HPT) is an attractive high-speed photodetector for optical communications, as it can provide high gain and high power. It can also be used to generate millimetre-wave signals using optical heterodyne for radio astronomy and wireless broad-band communications. Edge illuminated 2-terminal HPTs (2T-HPTs) integrated with an optical waveguide are designed and characterised. A small signal model is developed and it simulates the 2T-HPT frequency responses with good accuracy and supplements the measurement to determine its gain and cut-off frequency. This overcomes the requirement to measure the input photocurrent to the 2T-HPT using an exact replica of its PIN structure.

The high frequency responses for the longer devices, which give higher saturation current, are limited by the loss of the transmission line. This high loss is also observed by the 2-port characterisation of the back-to-back configured 2T-HPT and gives a trade-off between the cut-off frequency and saturation current in the design. Additional analysis shows that the cut-off frequency is also limited by the large device capacitances. Using scaling of the model, the cut-off frequency is shown to increase to 59.4GHz with a gain of 39dB for a 1μm×10μm 2T-HPT.

The above result shows that the 2T-HPT could be best utilised in a periodic structure (P-TWHPT), with small device area sections connected by short transmission lines. This increases the saturation current level without reducing the cut-off frequency and gain. A P-TWHPT model is developed to simulate and analyse its frequency response. It is shown that the effects of phase matching and reverse input termination are not critical for 2T-HPT structure, as its 3dB bandwidth is limited by the base discharge time constant. A four 5μm×10μm 2T-HPT section P-TWHPT is simulated with the phase mismatch and without reverse termination. It has a gain of 49dB and a cut-off frequency of 67.5GHz.
ACKNOWLEDGEMENTS

I would like to thank many people for their help in completing this thesis, but my foremost appreciation goes to my supervisors, Professor Bernard L. Weiss and Professor Ian D. Robertson, for their continuous guidance, encouragement and support throughout my graduate studies at the University of Surrey.

I gratefully acknowledge again all those who provided help and encouragement, and in particular, the IEE for awarding the Vodafone Scholarship; the Sheffield University for fabricating the phototransistors; Dr. Luis Gomez-Rojas for helping to initiate the early stages of the characterisation; the Electronics workshop for fabricating the text fixtures for my measurements; the technical staffs and all other members of the School of Electronics and Physical Sciences.

Special thanks are to all friends here, home and abroad, for their encouragement and support. Ultimately, it is the understanding, continued love and patience of my family, which provided me the courage to embark on this project. Especially my husband, his encouragement and support of the project never faltered, and without it this endeavour could not have been fulfilled.

This work was performed as part of a joint project between the Universities of Surrey, Leeds and Sheffield. The project was funded by the UK Engineering and Physical Sciences Research Council (GR/R34035). We also thank Bookham Technology for their support of this project.
CONTENTS

Abstract iii

Acknowledgements iv

Glossary of Symbols viii

Chapter 1 Introduction
1.1 Project Rationale 1-1
1.2 Research Aim and Objectives 1-3
1.3 Thesis Structure and Contents 1-3

Chapter 2 Literature Review
2.1 Various Photodetector Structures 2-2
  2.1.1 Vertically Illuminated Photodetector 2-2
  2.1.2 Waveguide Photodetector 2-7
  2.1.3 Travelling Wave Photodetector 2-9
  2.1.4 Periodic Travelling Wave Photodetector 2-18
2.2 Saturation Current 2-21
2.3 Summary 2-25

Chapter 3 Heterojunction Phototransistor
3.1 Principle of Operation 3-1
  3.1.1 Phototransistor Gain 3-6
  3.1.2 Phototransistor Cut-off Frequency 3-7
3.2 2 Terminal and 3 Terminal HPTs 3-8
  3.2.1 Effect of Input Optical Power on Gain 3-9
  3.2.2 Effect of Base Terminal on Frequency Response 3-11
Chapter 6  Periodic Model & Improved Devices

6.1  Smaller Devices  6-2
   6.1.1  Simulation and Measurement Results  6-2

6.2  Periodic Travelling Wave Devices  6-9
   6.2.1  Phase Matching and Reverse Input Termination  6-11
   6.2.2  Time Responses  6-15
   6.2.3  Frequency Responses  6-17
   6.2.4  Periodic Model Incorporated with SSM  6-22

6.3  Summary  6-26

Chapter 7  Conclusions & Further Developments

7.1  Conclusions  7-1

7.2  Further Developments  7-3

References
GLOSSARY OF SYMBOLS

\( a_1 \)  
incident wave on port 1

\( a_2 \)  
incident wave on port 2

\( A \)  
device active area

\( A_E \)  
emitter active area

\( A_C \)  
base-collector junction area

\( b_1 \)  
reflected wave on port 1

\( b_2 \)  
reflected wave on port 2

\( c \)  
velocity of light in free space

\( C \)  
junction capacitance of the photodetector

\( C' \)  
capacitance per unit length of the electrical transmission line

\( C_{BE} \)  
base-emitter junction capacitance

\( C_{BC} \)  
base-collector junction capacitance

\( d \)  
thickness of the photoabsorption layer

\( d n_B(x)/dx \)  
gradient of the doping concentration in the \( x \) direction

\( D \)  
minority-carrier diffusion coefficient

\( D_{nB} \)  
base minority-carrier diffusion coefficient

\( E(x) \)  
electric field in the \( x \) direction

\( E_g \)  
absorption layer bandgap energy

\( E_{g(InGaAs)} \)  
energy gap in the InGaAs material

\( E_{g(InP)} \)  
energy gap in the InP material

\( \Delta E_g \)  
differences in the emitter and base energy gaps

\( \Delta E_{v(InP/InGaAs)} \)  
valance gap discontinuity between the InP/InGaAs heterojunction

\( f_{3dB} \)  
3dB electrical bandwidth

\( f_{ATL} \)  
artificial transmission line cut-off frequency

\( f_{CR} \)  
capacitance resistance time constant limited bandwidth

\( f_{LPF} \)  
cut-off frequency of the low pass filter

\( f_t \)  
carrier transit time limited bandwidth

\( f_r \)  
unity gain cut-off frequency

\( f_{VM} \)  
velocity mismatch bandwidth

\( f_a \)  
frequency at which the current transfer ratio dropped
to 0.707 times its low frequency value

\( G \)  
optical gain

\( h \)  
Plank's constant

\( I_B \)  
base bias current

\( I_C \)  
collector bias current

\( I_E \)  
emitter bias current

\( I_{PH} \)  
AC photocurrent produced by the RF power of the input optical light

\( I_{PH} \)  
photocurrent produced by the optical power illumination

\( I_{PH} \)  
assuming only bandgap transitions

\( I_{PH} \)  
DC photocurrent produced by the input optical signal

\( J_{diffusion}(x) \)  
diffusion current density

\( J_{drift}(x) \)  
Drift current density

\( kT \)  
thermal voltage (in eV) at 300K

\( kT/q \)  
thermal voltage at 300K
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l )</td>
<td>photon absorption length</td>
</tr>
<tr>
<td>( l_e )</td>
<td>length of the electrical transmission line</td>
</tr>
<tr>
<td>( l_o )</td>
<td>length of the optical waveguide</td>
</tr>
<tr>
<td>( L )</td>
<td>inductance of per section of the transmission line</td>
</tr>
<tr>
<td>( L' )</td>
<td>inductance per unit length of the electrical transmission line</td>
</tr>
<tr>
<td>( L_E )</td>
<td>device length</td>
</tr>
<tr>
<td>( L_I )</td>
<td>insertion loss</td>
</tr>
<tr>
<td>( L_{nB} )</td>
<td>base minority-carrier diffusion length</td>
</tr>
<tr>
<td>( n(x) )</td>
<td>doping concentration in the x direction</td>
</tr>
<tr>
<td>( n_1 )</td>
<td>refractive index of medium 1</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>refractive index of the medium 2</td>
</tr>
<tr>
<td>( n_1 )</td>
<td>intrinsic carrier concentration</td>
</tr>
<tr>
<td>( n_{GaAs} )</td>
<td>refractive index of GaAs</td>
</tr>
<tr>
<td>( n_{InAlGaAs} )</td>
<td>refractive index of In_{0.53}Al_{0.14}Ga_{0.33}As</td>
</tr>
<tr>
<td>( n_{InGaAs} )</td>
<td>refractive index of In_{0.55}Ga_{0.47}As</td>
</tr>
<tr>
<td>( n_{Polyimide} )</td>
<td>refractive index of Polyimide</td>
</tr>
<tr>
<td>( N )</td>
<td>doping concentration</td>
</tr>
<tr>
<td>( N_B )</td>
<td>base doping concentration</td>
</tr>
<tr>
<td>( N_C )</td>
<td>collector doping concentration</td>
</tr>
<tr>
<td>( N_E )</td>
<td>emitter doping concentration</td>
</tr>
<tr>
<td>( N_{Emcap} )</td>
<td>emitter cap layer doping concentration</td>
</tr>
<tr>
<td>( N_{SC} )</td>
<td>sub-collector doping concentration</td>
</tr>
<tr>
<td>( P_{in} )</td>
<td>input optical power</td>
</tr>
<tr>
<td>( q )</td>
<td>electronic charge</td>
</tr>
<tr>
<td>( r_s )</td>
<td>small signal junction resistance</td>
</tr>
<tr>
<td>( R_{bi} )</td>
<td>base intrinsic resistance</td>
</tr>
<tr>
<td>( R_C )</td>
<td>collector resistance</td>
</tr>
<tr>
<td>( R_{C(Epitaxial)} )</td>
<td>collector epitaxial resistance</td>
</tr>
<tr>
<td>( R_F )</td>
<td>emitter resistance</td>
</tr>
<tr>
<td>( R_{Fresnel} )</td>
<td>Fresnel reflection coefficient</td>
</tr>
<tr>
<td>( R_{L} )</td>
<td>load resistance</td>
</tr>
<tr>
<td>( R_S )</td>
<td>parasitic resistance</td>
</tr>
<tr>
<td>( R_{SHSC} )</td>
<td>sub-collector layer sheet resistance</td>
</tr>
<tr>
<td>( R_L )</td>
<td>return loss</td>
</tr>
<tr>
<td>( S_{11} )</td>
<td>input reflection coefficient when the output is matched</td>
</tr>
<tr>
<td>( S_{12} )</td>
<td>reversed transfer coefficient when the input is matched</td>
</tr>
<tr>
<td>( S_{21} )</td>
<td>forward transfer coefficient when the output is matched</td>
</tr>
<tr>
<td>( S_{22} )</td>
<td>output reflection coefficient when the input is matched</td>
</tr>
<tr>
<td>( S_{SC} )</td>
<td>CPW gap (emitter to sub-collector)</td>
</tr>
<tr>
<td>( \tan \delta )</td>
<td>loss tangent</td>
</tr>
<tr>
<td>( v_e )</td>
<td>velocity of the electrical signal</td>
</tr>
<tr>
<td>( v_o )</td>
<td>velocity of the optical signal</td>
</tr>
<tr>
<td>( V_{BC} )</td>
<td>base-collector voltage</td>
</tr>
<tr>
<td>( V_{BE} )</td>
<td>base-emitter voltage</td>
</tr>
<tr>
<td>( V_{CE} )</td>
<td>collector-emitter voltage</td>
</tr>
<tr>
<td>( w )</td>
<td>active device width</td>
</tr>
<tr>
<td>( W_E )</td>
<td>emitter width</td>
</tr>
<tr>
<td>( W_{SC} )</td>
<td>collector contact width</td>
</tr>
</tbody>
</table>
\( X_b(\text{dep}) \) base depletion thickness
\( X_{bc}(\text{dep}) \) base-collector depletion thickness
\( X_B \) base thickness
\( X_c(\text{dep}) \) collector depletion thickness
\( X_C \) collector thickness
\( X_{\text{dep}C} \) base-collector depletion thickness
\( X_{\text{dep}E} \) base-emitter heterojunction depletion thickness
\( X_E \) emitter thickness
\( X_{\text{E cap}(\text{InGaAs})} \) InGaAs emitter cap thickness
\( X_{\text{E cap}(\text{InP})} \) InP emitter cap thickness
\( X_{\text{NeutralBase}} \) neutral base thickness
\( X_{\text{spacer}} \) spacer thickness
\( X_{sc} \) sub-collector thickness
\( Z_{\text{ATL}} \) artificial transmission line characteristic impedance
\( Z_L \) load impedance
\( Z_0 \) photodetector impedance
\( Z_{\text{OUT}} \) output impedance
\( \alpha_B \) common-base current gain
\( \alpha_T\text{ac} \) current transfer ratio
\( \alpha_T0 \) base transport factor
\( \alpha_a \) absorption coefficient
\( \beta \) current gain
\( \gamma(\omega) \) input reflection coefficient
\( \gamma_{\text{BJT}} \) BJT emitter injection efficiency
\( \gamma_{\text{HPT}} \) HPT emitter injection efficiency
\( \Gamma \) waveguide confinement factor
\( \Gamma_L \) load reflection coefficient
\( \Gamma_S \) source reflection coefficient
\( \varepsilon \) permittivity of the intrinsic layer
\( \varepsilon_0 \) free space permittivity
\( \varepsilon_r \) relative dielectric constant of the intrinsic layer
\( \varepsilon_{rC} \) collector relative dielectric constant
\( \varepsilon_{rE} \) emitter relative dielectric constant
\( \varepsilon_{r(\text{avg})} \) weighted average relative dielectric constant
\( \varepsilon_{r(eff)} \) effective relative dielectric constant of the transmission line
\( \varepsilon_{r(wg)} \) relative dielectric constant of the optical waveguide
\( \varepsilon_{r(\text{GaAs})} \) relative dielectric constant for GaAs
\( \varepsilon_{r(\text{In}0.53\text{Al}0.14\text{Ga}0.33\text{As})} \) relative dielectric constant for In\(_{0.53}\)Al\(_{0.14}\)Ga\(_{0.33}\)As
\( \varepsilon_{r(\text{InGaAs})} \) relative dielectric constant for InGaAs
\( \varepsilon_{r(\text{InP})} \) relative dielectric constant for InP
\( \varepsilon_{r(\text{SiN})} \) relative dielectric constant for SiN
\( \varepsilon_{r(\text{SiF}_2)} \) relative dielectric constant for SiF\(_2\)
\( \eta \) quantum efficiency
\( \eta_{obs} \) absorption efficiency
\( \lambda \) optical wavelength
\( \mu \) carrier mobility
\( \mu_n \) electron mobility
\( \mu_{nB} \) base electron mobility
\( \mu_{pB} \) base hole mobility
\( \mu_nC \) collector electron mobility
\( \mu_nE \) emitter electron mobility
\( \mu_{nSC} \) sub-collector electron mobility
\( \mu_n(\text{Ec}_{\text{cap}} \text{GaAs}) \) InGaAs emitter cap electron mobility
\( \mu_n(\text{Ec}_{\text{cap}} \text{InP}) \) InP emitter cap electron mobility
\( v_e \) electron velocity
\( v_h \) hole velocity
\( v_{SAT} \) carrier saturation velocity
\( \rho \) resistivity
\( \rho_{\text{contact}} \) contact resistance
\( \rho_B \) base resistivity
\( \rho_C \) collector resistivity
\( \rho_E \) emitter resistivity
\( \rho_{\text{Ec}_{\text{cap}} \text{GaAs}} \) InGaAs emitter cap resistivity
\( \rho_{\text{Ec}_{\text{cap}} \text{InP}} \) InP emitter cap resistivity
\( \rho_{SC} \) sub-collector resistivity
\( \rho_{\text{AE}} \) specific emitter contact resistance
\( \rho_{\text{AC}} \) specific collector contact resistance
\( \rho_{\text{ASC}} \) specific sub-collector contact resistance
\( \tau_B \) base transit time
\( \tau_C \) collector charging time
\( \tau_E \) emitter charging time
\( \tau_{\text{eff}} \) effective transit time
\( \tau_{SC} \) space-charge transit time
\( \tau_{\text{transit}} \) transit time
\( \tau_{\text{Carrier Lifetime}} \) minority-carrier lifetime
\( \tau_{OE} \) output electrical signal time delay
\( \tau_O \) input modulating optical signal time delay
\( \Phi_{BC} \) base-collector junction built-in voltage
\( \omega \) radian frequency
CHAPTER 1

INTRODUCTION

Optical fibre communications developed tremendously after the first low-loss fibres were produced in 1970. It is now a mature technology and is used widely in long-haul telecommunications systems, metropolitan areas and high data rate local area networks. The success of the optical fibre communications is mainly due to its massive potential bandwidth, low transmission loss, potential low cost, small size and light weight. In addition, it has several other attractive features, such as electrical isolation, signal security, ruggedness, flexibility, system reliability, immunity to interference and crosstalk.

Today, there is a demand for a larger bandwidth to meet the steadily increasing data rates and traffic, particularly due to the heavy usage of broadband Internet. In addition, new higher-end applications, such as radio over fibre systems, medical imaging and MPEG video, which require enormous bandwidth, are continually being introduced. This leads to a need to realise optical fibre communication systems with potential large bandwidth, which is limited primarily by the photodetector.

1.1 PROJECT RATIONALE

In optical fibre communication systems, the photodetector is an important component, which can dominate its overall performance. As a result, the development of the photodetector must keep ahead of the communication industry requirement. The function of the photodetector is to convert the received optical signal into an electrical signal. The photodetector output current is proportional to its incident optical power, which is a replica of the transmitted electrical signal. The basic requirements of today's photodetector are large bandwidth, high efficiency, high sensitivity, low bias...
voltage, and high output current saturation, in addition to being low cost, reliable, small size and light weight.

Semiconductor photodetectors are used in optical fibre communications, mainly due to its low cost, small size, high reliability and potential large bandwidth. Various types of semiconductor photodetectors have been devised in parallel with the development of the optical fibre communications. Semiconductor photodetectors can broadly be divided into two categories, without gain and with internal gain. The most popular photodetectors, PIN and MSM, belong to the category without gain, while the avalanche photodiode and phototransistor belong to the category with internal gain.

Most of the recent research activities are focused on the PIN and MSM photodetector because of their massive bandwidth of hundreds of GHz coupled with low operating voltage, simple design and fabrication. However, their conversion efficiency are usually low as they do not have gain, and require an additional amplifier to boost the output current to drive other circuits. This offsets the advantage of simple design.

On the other hand, the avalanche photodiode is able to provide internal gain through avalanche multiplication. However, it requires a high operating voltage and suffers from a high noise level; these disadvantages are inherent to the avalanche process. Conversely, the phototransistor, which utilises transistor action to provide the high internal gain, operates with low voltage and low noise level. The fundamental operating principle of phototransistors is similar to normal bipolar transistors; hence the development of the heterojunction technology for the bipolar transistor (HBT) to improve its performance can be applied to the phototransistor as well. A phototransistor of this type is then known as heterojunction phototransistor (HPT).

Due to their similarity, recent breakthroughs in HBTs development into the millimetre-wave frequencies region have benefited the HPTs directly. The advances in epitaxy and device fabrication technology in the HBTs help to overcome otherwise complicated fabrication and design processes of the HPT. Thus, the HPT is now a very attractive option to meet the growing requirements of the optical fibre communications.
1.2 RESEARCH AIM AND OBJECTIVES

The purpose of the project is to investigate the design and characterisation of a high cut-off frequency, high gain, and high saturation power 2-terminal HPT (2T-HPT). The device fabrication processes were developed by Sheffield University, who also fabricated the devices.

The main research area of the project is to develop equivalent circuit models for the 2T-HPT, which are used to analyse the effects of various physical parameters. The model will be further developed to investigate travelling wave structures and the effects of phase matching and reverse input termination conditions. These models help to optimise and design the 2T-HPT circuits. Thus it plays an important part in the success of designing a high performance 2T-HPT. The second part of this research is to perform various RF-on-wafer characterisation of the 2T-HPT, such as responsivity, IV curve, frequency response and electrical loss to validate the equivalent circuit models.

1.3 THESIS STRUCTURE AND CONTENTS

The thesis is divided into 7 chapters. Following the introduction, Chapter 2 reviews the development of various types of photodiodes. Best reported figures are given along with various approaches to extend beyond the typical bandwidth-efficiency product. The other figure of merit, saturation current, is also discussed. An overview of the various photodetector performances is presented.

In Chapter 3, the basic operation principle of the HPT, its gain and cut-off frequency, are described. A review of some of the reported HPTs with significant developments is given. The design of an edge illuminated 2T-HPT, integrated with an optical waveguide, is presented. The transmission line loss is also analysed.
In Chapter 4, the characterisation of the 2T-HPT transmission line loss is discussed with its measurement set-up. The various reported methods used to characterise the photodetector frequency responses are presented. From the characteristic of the HPT and the available resources, the most appropriate method is described.

In Chapter 5, a small signal model (SSM) for the 2T-HPT is developed and the detailed calculations for all the parameters used in the model are given. The electrical performances of the 2T-HPTs are characterised and analysed. The SSM is used to determine the gain and cut-off frequency.

In Chapter 6, the SSM is used to analyse the effects of various parameters. This gives an overview of the 2T-HPT performances and limitations. A periodic travelling wave 2T-HPT (P-TWHPT) model is subsequently developed using the SSM. The various phase matching and reverse input termination conditions are analysed through the P-TWHPT model. Using scaling of the model, smaller discrete and periodic devices are designed.

The final Chapter 7 consists of a discussion section, in which the experimental and simulation results of previous chapters are drawn together in a coherent whole with a number of suggestions for future study arising from the current work.
CHAPTER 2

LITERATURE REVIEW

Photodetectors are important devices for optical communication systems operating in the near infrared region (0.8 to 1.6μm). They are semiconductor devices that convert incident optical signal (photons) into electrical signal (electrons and holes) in the photoabsorption layer. The photodetector must satisfy stringent requirements, such as high sensitivity at operating wavelengths and high response speed. In addition, it should be compact in size, using low bias voltage and current, as well as being reliable.

The optical communications primary operating wavelengths are 0.8 to 0.9μm (short wavelength region) and 1.3 to 1.6μm (long wavelength region). For semiconductor photodetectors to respond to optical signals at these wavelengths, the absorption layer bandgap energy \( E_g \) must be smaller than the photon energy \( 1 \):

\[ E_g < \frac{hc}{\lambda} \]

(2.1)

where \( h \) is the Planck's constant, \( c \) is the velocity of light in free space and \( \lambda \) is the optical wavelength. The semiconductor alloy In_{0.53}Ga_{0.47}As, whose bandgap energy is 0.75eV (1eV=1.6x10^{-19}J) [2] has a cut-off wavelength of 1.66μm and is suitable to be used as the absorption layer of the photodetector operating at the long wavelength region. On the other hand, the semiconductor material GaAs, with a bandgap energy of 1.424eV [2] has a cut-off wavelength of 0.871μm and is used widely in the short wavelength region.

The chapter reviews the various photodetector structures, their bandwidth and efficiency, with a focus on In_{0.53}Ga_{0.47}As and GaAs photodetectors operating at 1.55μm and 0.85μm wavelengths, respectively. Best reported figures are given together with various approaches to exceed beyond the typical bandwidth-efficiency product. The improvements of the saturation current level, with the variation of the
photodetector structures, are also outlined. This is followed by a more thorough discussion on the saturation current. An overview of the various photodetector performances is presented at the end of the chapter.

2.1 VARIOUS PHOTODETECTOR STRUCTURES

The main figure of merit for a photodetector is the bandwidth-efficiency product that imposes a bound on the speed and responsivity of the detector. In addition, the saturation current, which determines the spurious-free dynamic range and affects the signal-to-noise ratio of the photodetector, is another important parameter. Several photodetector structures and their performances are discussed as follow.

2.1.1 VERTICALLY ILLUMINATED PHOTODETECTOR

For conventional vertically illuminated photodetector (VPD), light is incident vertically on the photoabsorption layer where the thickness of the photoabsorption layer \(d\) is also the maximum photon transit path length \(l\), as shown in Fig 2.1.

![Figure 2.1: Vertically Illuminated Photodetector (VPD)](image)

The photoabsorption layer is the intrinsic layer, which is depleted of mobile carriers by the reverse bias voltage. When a photon is absorbed in the photoabsorption layer, an electron-hole pair is generated. The electrons and holes then drift to the p and n layers, respectively, assisted by the electric field. These carriers contribute to the output electrical signal.
The 3dB electrical bandwidth of the VPD is given as [3]:

\[ f_{3dB} = \frac{f_t}{\sqrt{1 + \left(\frac{f_t}{f_{CR}}\right)^2}} = \frac{f_{CR}}{\sqrt{1 + \left(\frac{f_{CR}}{f_t}\right)^2}} \]  

(2.2a)

where the carrier transit time limited bandwidth is given by:

\[ f_t \approx \frac{3.5u_{Sat}}{2\pi d} \]  

(2.2b)

where \( u_{Sat} \) is the carrier saturation velocity and \( d \) is the thickness of the photoabsorption layer;

and the capacitance resistance (CR) time constant limited bandwidth is given by:

\[ f_{CR} = \frac{1}{2\pi C(R_s + R_L)} \]  

(2.2c)

where \( R_L \) is the load resistance;

and \( R_s \) is the parasitic resistance, dominated by the contact resistance and is given as:

\[ R_s = \frac{\rho_{contact}}{A} \]  

(2.2d)

where \( \rho_{contact} \) is the contact resistivity and \( A \) is the surface area;

and \( C \) is the junction capacitance of the photodetector. Due to the high doping concentration of the p and n layers, the depletion region is assumed to reside entirely in the intrinsic layer and hence \( C \) is given as:

\[ C = \frac{\varepsilon A}{d} \]  

(2.2e)

where \( \varepsilon \) is the permittivity of the intrinsic layer.

The bandwidth of the VPD is limited by the CR time constant and the carrier transit time. The CR time constant arises due to the finite time required to charge the intrinsic junction capacitance through the load resistance. Practically, parasitic components are also involved and increase the time constant. The carrier transit time arises due to the finite time required for the photogenerated carriers in the absorption region to travel to their respective electrodes.
The load resistance \((R_L)\) for all standard systems is 50\(\Omega\), hence the junction or photoabsorption capacitance \((C)\) must be decreased to improve the bandwidth. This can be achieved by reducing the surface area of the photoabsorption layer \((A)\), however, this will increase the parasitic resistances \((R_S)\), which limits the improvement. The photodetector capacitance \((C)\) can also be decreased by increasing the thickness of the photoabsorption layer \((d)\), which in turn improves the CR time constant limited bandwidth \((f_{CR})\). However, this will increase the carrier transit distance \((l=d)\), which in turn degrades the carrier transit time limited bandwidth \((f_t)\), hence again limiting the performance of the photodetector. Consequently, there is an optimum \(f_{CR}\) and \(f_t\) that yield a maximum bandwidth \((f_{3dB})\).

The output photocurrent is another important parameter of the photodetector. The photocurrent produced by the optical power illumination, assuming only bandgap transitions, is given as [1]:

\[
I_{ph} = \frac{\eta q \lambda P_{in}}{hc}
\]  

where \(\eta\) is the quantum efficiency, \(q\) is the electronic charge, \(\lambda\) is the optical wavelength, \(h\) is the Planck’s constant, \(c\) is the velocity of light in free space and \(P_{in}\) is the input optical power.

A high output photocurrent is desired so that it is able to drive other circuitries and gives a good sensitivity. In addition, it also ensures a high signal-to-noise ratio assuming the noise level remains the same. From Eqn 2.3, \(I_{ph}\) depends on the number of photons per second \((\lambda P_{in}/hc)\) incident on the absorption layer and the quantum efficiency. The quantum efficiency for the semiconductor photodetector, assuming that the entire incident optical signal falls within the cross sectional area of the absorption layer, is given as [4]:

\[
\eta = (1 - R_{Frame})\eta_{abs}
\]  

\(\eta_{abs}\) is the absorption efficiency and is given as:

\[
\eta_{abs} = \left(1 - e^{-\alpha d}\right)
\]
where $l$ is the photon absorption length and $\alpha$ is the absorption coefficient, which is a function of the absorption layer material properties and varies with the incident optical wavelength. $R_{\text{Fresnel}}$ is the Fresnel reflection coefficient and is given as:

$$R_{\text{Fresnel}} = \left\{ \frac{n_1 - n_2}{n_1 + n_2} \right\}^2$$

(2.4c)

where $n_1$ and $n_2$ are the refractive indices of the photodetector and input medium (air), respectively. The Fresnel reflection loss is the inherent input coupling loss due to the partial reflection and can be reduced by applying anti-reflection coating on the photodetector input facets. From Eqn 2.4, in order to achieve a maximum quantum efficiency, it is critical to ensure that the photon absorption length is sufficiently long for all the photons to be absorbed, i.e. $\alpha l >> 1$.

However, in the VPD structure, the photon absorption length is equal to the thickness of the photoabsorption layer ($l = d$). Hence the desire for a short photon absorption length (i.e. small $d$) to minimise the carrier transit time to increase the bandwidth, conflicts with the requirements for a large absorption volume (i.e. large $d$) for good quantum efficiency, so called the bandwidth and efficiency trade-off.

The bandwidth and efficiency of the VPD is studied with the following typical parameters [3,5-8]: for a 1.55$\mu$m wavelength In$_{0.53}$Ga$_{0.47}$As VPD with $n=3.544$, $\varepsilon_s=13.88$ and $\alpha_s=0.68\mu$m$^{-1}$, and a 0.85$\mu$m wavelength GaAs VPD with $n=3.655$, $\varepsilon_s=12.9$ and $\alpha_s=1\mu$m$^{-1}$, both have $v_{\text{SAT}}=5.3\times10^4$m/s, $\rho_{\text{contact}}=1\times10^{-10}$$\Omega$·m$^2$ and $R_L=50\Omega$.

With a large device area of say 15$\mu$m×15$\mu$m, $f_{3dB}$ is dominated by the large CR time constant. However, with the area reduced by 10 fold, $f_t$ becomes the main limitation. Though miniaturising the device surface area improves $f_{3dB}$, it lowers the device saturation current. Fig 2.2 shows the bandwidth and efficiency of a VPD, with a surface area of 2$\mu$m×2$\mu$m and including the Fresnel loss, as a function of the absorption layer thickness ($d$). Due to the bandwidth-efficiency trade-off, there is an optimum thickness that yields the maximum bandwidth-efficiency product, which is just less than 13GHz and 18GHz, for the 1.55$\mu$m and 0.85$\mu$m wavelengths VPD, respectively. The shorter wavelength VPD performs better due to the inherent higher absorption coefficient for GaAs, which gives a higher efficiency for the same device size.
The resonant-cavity-enhanced photodetector (RCE-PD) is used by Tan et al. [9] to improve the absorption efficiency of the VPD without compromising the bandwidth. Reflectors are placed at both ends of the absorption layer to reflect the remaining optical signal back. In this way, the optical signal is made to go through the absorption layer multiple times and the absorption efficiency is improved without increasing the absorption layer physical thickness. The RCE-PD set a record efficiency of 28% for the VPD. The same work reported the usage of mushroom-mesa and air-bridge to reduce the CR time constant. The mushroom-mesa, also known as the undercut mesa, has extended surface areas for its top and bottom semiconductor layers (p-doped and n-doped layers) to give a low contact resistance while maintaining a small intrinsic capacitance. The air-bridge helps to reduce the unwanted parasitic capacitance associated with the interconnections between the contacts and the probe pads. The photodetector is then dominated by $f_t$, which is limited by the thick absorption layer. Adopting all the above principles, the RCE-PD operating at 1.3μm wavelength is able to achieve a $f_{3dB}$ of 120GHz, giving a large bandwidth-efficiency product of 34GHz.
2.1.2 WAVEGUIDE PHOTODETECTOR

The edge illuminated photodetector was developed to overcome the bandwidth-efficiency trade-off, where its illumination is directed perpendicular to the carrier transit path, such that absorption and carrier drift are orthogonal, as shown in Fig 2.3.

![Figure 2.3: Edge Illuminated Photodetector](image)

This structure differs from the VPD's, such that photon transit path is the length of the absorption layer rather than its thickness (i.e. \( d \neq l \)), permitting the bandwidth and efficiency to be specified independently. Hence the thickness can be reduced for minimising the carrier transit time without compromising the absorption efficiency. However, the light disperses from the absorption layer as it propagates inside the device, leading to the reduction of the efficiency. To confine the light within the absorption layer, it is sandwiched between two transparent cladding layers to form an optical waveguide structure. This design is known as the waveguide photodetector (WGPD).

The absorption coefficient now becomes \( \Gamma \alpha_a \), where \( \Gamma \) is the waveguide confinement factor that indicates the amount of light confines inside the waveguide. The quantum efficiency for the WGPD is then expressed as [10]:

\[
\eta = (1 - R_{\text{Fresnel}})(1 - e^{-\Gamma \alpha_a l})
\]  

(2.5)

The bandwidth and efficiency of the WGPD are analysed using the same parameters given earlier for the VPD, except for the additional \( \Gamma \), reasonably taken as 0.4 [10]. Like the case of the VPD, \( f_{3dB} \) is dominated by \( f_{CR} \) when the device area is large, and \( f_i \)
when the device area reduces. As the bandwidth and efficiency can be optimised separately, the ideal bandwidth-efficiency product exceeds 90GHz, as shown in Fig 2.4, for a 2μm wide WGPD with a 0.2μm thickness. Like the case for the VPD, the shorter wavelength WGPD performs better due to the inherent higher absorption coefficient for GaAs, which gives a higher efficiency for the same device size.

Although the WGPD can achieve near 100% absorption efficiency by ensuring the absorption length is sufficiently long, it suffers from poor input coupling efficiency. This is due to the small input aperture limited by the thickness of the absorption layers (~0.5μm or less), which is usually lesser than the spot size of a lensed fibre. Increasing the absorption thickness, improves the coupling efficiency but at the expense of the bandwidth, having a bandwidth-efficiency trade-off like the VPD.

A plausible solution to overcome the poor coupling efficiency is to design a separate leaky optical waveguide along the absorption layer and the structure becomes a waveguide-fed-photodetector (WG-fed-PD). The optical input is coupled to the passive waveguide with evanescent coupling of light to the absorption layer. The
The optical signal is then continuously feed (evanescently coupled) along the absorption layer. This allows a large waveguide aperture to improve the coupling efficiency while maintaining a thin absorption layer for large bandwidth.

The best reported bandwidth-efficiency product is 57GHz by Giboney et al. [11] for a 2µm×9µm WG-fed-PD operating at 0.83µm wavelength, having a thin 0.17µm absorption layer and hence a transit time limited bandwidth of 118GHz. It used anti-reflection coating at the facet to reduce Fresnel reflection and increased the efficiency to 49%. Another significant achievement with a bandwidth-efficiency product of 55GHz at 1.55µm wavelength is by Kato et al. [12] who employed the mushroom-mesa to reduce the CR time constant of its 1.5µm×12µm WG-fed-PD to obtain a transit time limited bandwidth of 110GHz and again using anti-reflection coating at the facet to achieve an efficiency of 50%.

2.1.3 TRAVELLING WAVE PHOTODETECTOR

The CR bandwidth limitation of the photodetector can be resolved using the travelling wave structure, as shown in Fig 2.5. The principle of the travelling wave photodetector (TWPD) is to distribute the bandwidth limiting junction capacitance along the length of the electrical transmission line formed by the electrodes. The capacitance is balanced by the transmission line inductance such that the ratio is equal to the impedance of the load. It is crucial to get these impedances matched otherwise the frequency response of the TWPD will still be affected by the junction capacitances, like the WGPD. Basic transmission line structures are two parallel metal plates sandwiching the semiconductor layers, coplanar strips (CPS) and coplanar waveguide (CPW).
The CR time constant is no longer present in the 3dB electrical bandwidth, which is now given as [3]:

\[
f_{3dB} = \frac{f_t}{\sqrt{1 + \left(\frac{f_t}{f_{TM}}\right)^2}}
\]  

(2.6)
where $f_{VM}$ is the velocity mismatch bandwidth due to the electrical and optical signal travelling at different velocity. Velocity mismatch prevents the photogenerated current along the length of the device to add in phase and reduces the bandwidth. Another factor that governs $f_{VM}$ is the interference on the forward travelling wave (towards the load) caused by the reflected backward travelling wave (towards the input end) when the reverse input is not terminated with a matched load (i.e. open circuit).

The normalised velocity mismatch frequency response, obtained from the Fourier transform of an impulse response is given as [13]:

$$I(\omega) = \frac{1}{2} \left[ \frac{w_f}{w_f - j\omega + \gamma(\omega)} \frac{w_r}{w_r + j\omega} \right] e^{-j\omega l/v_e} \tag{2.7}$$

where $w_f = \frac{\Gamma \alpha_s v_e}{1 - (v_e/v_0)}$ and $w_r = \frac{\Gamma \alpha_s v_e}{1 + (v_e/v_0)}$.

$\gamma(\omega)$ is the input reflection coefficient, $\omega$ is the radian frequency, $l$ is the length of the device, and $v_e$ and $v_o$ are the electrical and optical velocities, respectively.

The electrical velocity is dependent on the propagation mode of the electrical field, which changes with the doping concentration and the thickness of the semiconductor layers directly beneath the electrodes, and the planar transmission line structure, as explained in the following.

![MSM TWPD Structure with CPW Electrical Field Pattern](image)
Metal-semiconductor-metal (MSM) TWPD has its electrodes on top of the absorption layer, which usually has a very low doping concentration and gives a high resistivity. This allows the metal contacts to realise a CPW field pattern, as shown in Fig 2.6, which assumes about half of the electrical fields travels in the air while the other half travels in the lower doped semiconductor layer. The CPW transmission line has an electrical velocity that is higher than the optical velocity and can be approximated as [14]:

\[
\nu_e \approx \frac{c}{\sqrt{\frac{\varepsilon_r(\text{avg}) + 1}{2}}}
\]  

(2.8a)

where \(c\) is the velocity of light in free space and \(\varepsilon_r(\text{avg})\) is the weighted average relative dielectric constant of the absorption and the substrate layers. A useful feature of CPW transmission line is that the electrical velocity and the impedance can be altered by varying the width (\(S\)) and gap (\(W\)) of the CPW to achieve velocity and impedance matching. The above analysis is also true for MSM TWPD designed in CPS transmission line structure.

On the other hand, the PIN TWPD has highly doped p-type and n-type semiconductor layers beneath its electrodes. The free carriers from these layers screened the electrical fields and confined them inside the intrinsic absorption layers. As a result, the structure cannot form a CPW field pattern, as shown in Fig 2.7.

The low resistive semiconductor layer (n layer) beneath the absorption layer basically shorted the two ground contacts together and formed a microstrip-liked structure. However, as most of the electrical fields are confined inside the thin intrinsic layer, a
"parallel-plate" structure is formed instead. The electrical velocity is then approximated as [15]:

\[ v_e \approx \frac{d}{\varepsilon_r \varepsilon_0 w Z_0} \]  

(2.8b)

where \( d \) is the thickness of the photoabsorption layer, \( \varepsilon_r \) is the relative dielectric constant of the intrinsic layer and \( w \) is the width of the active device. \( Z_0 \) is the photodetector impedance and must be matched to the load impedance, which is generally 50\( \Omega \). It is stated by Giboney et al. [13] that the PIN TW PD width has to be reduced to 1\( \mu \)m in order to achieve 50\( \Omega \). The electrical velocity is smaller than a parallel-plate waveguide filled with homogeneous dielectric and cannot be increased for a monolithic PIN TW PD. Hence its electrical velocity is always mismatch with the optical velocity, which is typically higher [3]. The optical field velocity can be approximated using the following equation since most of the optical field are travelling within the optical waveguide [8]:

\[ v_o \approx \frac{c}{\sqrt{\varepsilon_r(wg)}} \]  

(2.9)

where \( \sqrt{\varepsilon_r(wg)} \approx n_{wg} \) is the refractive index of the optical waveguide.

The frequency response of the various velocity matching and reverse input termination conditions are plotted in Fig. 2.8 using Eqn 2.7 for an In_{0.55}Ga_{0.47}As PIN TWPD operating at 1.55\( \mu \)m wavelength, with a 1\( \mu \)m width, 0.15\( \mu \)m absorption layer and other parameters given earlier for the VPD and the WGPD.

The plot shows that at low frequency, the output amplitude of the reverse input terminated photodetector is half of the reverse input open circuit photodetector. This is because the reverse input termination absorbs the backward travelling wave, which consists half of the photogenerated power. On the other hand, with the reverse input open circuit, it allows the backward travelling wave to be reflected towards the load and thus gives a ~3dB higher power at low frequency. However, the frequency response changes more gradually with the reverse input terminated, therefore, it gives a larger \( f_{cm} \) even under velocity mismatch condition.
Figure 2.8: Frequency Responses for Reverse Input Terminated (dashes) and Open Circuit (solid lines) with various $v_r/v_o$ ratios (0.3 – triangles; 1.3 – squares & 1 – circles)

Figure 2.9: Velocity Mismatch Bandwidth as a function of Velocity Ratio ($v_r/v_o$) with Reverse Input Terminated and Open Circuit
The velocity mismatch bandwidth as a function of the ratio of the electrical and optical velocities, with the reverse input terminated and open circuit, are plotted in Fig 2.9. Again, as seen from the plot, with the reserve input terminated instead of open circuit, the bandwidth is higher throughout the whole range of different velocity ratios, and is infinite (ignoring microwave loss and dispersion) when the velocity is matched. Conversely with the reverse input open circuit, the bandwidth increases with $v_e$ and saturates beyond the velocity match condition, at a velocity ratio around 1.4. Hence a higher $v_e$ is preferred.

The lowest velocity mismatch bandwidth from Fig 2.9, is 400GHz, which is twice that of $f_t$ calculated using Eqn 2.2b. Hence the overall $f_{3dB}$, as stated in Eqn 2.6, is mainly limited by $f_s$, which is a function of carrier saturation velocity and absorption thickness. Following the above computation, reverse input termination and velocity matching are not necessary if the targeted bandwidth is less than 176GHz. This allows the quantum efficiency to be optimised near 100%, whereas only a maximum of 50% is obtainable with the reverse input terminated.

It has been concluded so far that the bandwidth of the TWPD and the small area WGPD are comparable, both being dominated by $f_s$. However, the quantum efficiency of the TWPD can be improved by increasing the device length without compromising the bandwidth, which is completely independent of the CR time constant. Unfortunately the additional device length will reduces the $f_{CR}$ of the WGPD and hence its bandwidth. As shown in Fig 2.10, the absorption efficiency can approached near 100% when $\Gamma a l >> 1$ for both the TWPD and the WGPD, assuming no input coupling loss, also ignoring microwave loss and dispersion. A 2$\mu$m wide WGPD and a 1$\mu$m wide TWPD are considered with all other parameters, as stated earlier.
There is an optimum length, when exceeded the bandwidth-efficiency product of the WGPD begins to drop, that is when the bandwidth reduction is greater than the efficiency improvement. On the other hand, the TWPD can afford to have longer length and hence the absorption coefficient can be distributed to give a lower current density; consequently a higher saturation current level can be achieved here than the WGPD and the VPD. However, the TWPD has its electrical transmission line laid above lossy semiconductor layers and this microwave loss may be significant when the device is excessively long, outweighing the saturation current improvement.

The best reported TWPD by Giboney et al. [11] has a bandwidth of 172GHz and an efficiency of 45%, giving a bandwidth-efficiency product of 76GHz. The bandwidth is limited by the transit time across the absorption layer thickness of 0.17μm. The PIN TWPD is not velocity match and has no reverse input termination. It operates at 0.86μm wavelength and has a 1μm width to give a 50Ω impedance matched to the load. The travelling wave structure is independent of the CR time constant, having a relatively constant bandwidth as the device length increases.
In another breakthrough, record bandwidths of 560GHz and 570GHz operating in the short wavelength region were achieved with a PIN and a MSM TWPD, by Chiu et al. [16] and Shi et al. [17], respectively. However, both the designs have a poor efficiency of only 8% (without anti-reflection coating). Focus was on the bandwidth improvement, employing low-temperature grown (LTG) GaAs as the absorption layer, which has a very short carrier lifetime of 300fs [16]. This introduces an additional carrier extraction mechanism, carrier trapping, which helps to reduce the transit time bandwidth limitation. The new transit time bandwidth and effective transit time are now given as [14]:

\[ f_t = \frac{1}{2\pi \cdot \tau_{eff}} \] (2.10a)

where

\[ \frac{1}{\tau_{eff}} = \frac{1}{\tau_{transit}} + \frac{1}{\tau_{CarrierLifetime}} \] (2.10b)

\( \tau_{CarrierLifetime} \) is the carrier lifetime and \( \tau_{transit} \) is the transit time given as [3]:

\[ \tau_{transit} = \frac{d}{3.5 \cdot v_{sat}} \] (2.10c)

with a saturation velocity (\( v_{sat} \)) of \( 5.3 \times 10^4 \text{m/s} \) [3] and a absorption layer thickness (\( d \)) of 0.17\( \mu \text{m} \), the transit time is calculated to be 916fs. Hence the effective transit time is dominated by the smaller carrier lifetime of 300fs. Effectively, this concept can be seen as the reduction of the photoabsorption thickness to provide a large \( f_t \), where those carriers with longer transit time than the carrier lifetime will recombine before reaching the electrodes. Nevertheless, the physically thickness is being maintained to give a large \( f_{CR} \). However, as the quantum efficiency is a function of the carrier lifetime and the transit time, the devices have very poor efficiency, as seen in the reported results.

The impedance of Chiu et al. [16] PIN TWPD is 50Ω while that of Shi et al. [17] MSM TWPD is 30Ω. Therefore, to be exact, the MSM TWPD should be classified as a WGPD as it is not matched to the output impedance. However, since the device area is small and the MSM structure inherently has a low capacitance, the CR time constant limitation has very little effect on the overall bandwidth. Hence like the PIN TWPD, the MSM TWPD bandwidth is dominated by the effective transit time. To
date there are only a few reports of TWPD operating in the long wavelength and the best performance is by Alles et al. [18], having a 60GHz bandwidth with a 24% efficiency and an output power of -24dBm at 60GHz. The detailed designs of this MSM TWPD are not given and thus further analysis is not possible.

### 2.1.4 PERIODIC TRAVELLING WAVE PHOTODETECTOR

The TWPD can be structured into a periodic travelling wave photodetector (P-TWPD) consisting of periodically loaded discrete photodiodes serially connected by a long optical waveguide and an electrical transmission line, as shown in Fig 2.11. The new structure allows the photodiodes, the optical waveguide and electrical transmission line to be optimised separately but the fabrication is more complicated as more mesas and electrical contacts are required.

![Figure 2.11: Periodic Travelling Wave Photodetector (P-TWPD)](image)

The capacitance of each photodiode \(C\) is balanced with the inductance of each section of transmission line \(L\), to give the required impedance to match with the next section and subsequently to the load. In this way, it formed an artificial transmission line (ATL) and its equivalent electrical circuit is, as shown in Fig 2.12.
The ATL formed by the capacitances of the photodiodes and inductances of the transmission line sections is essentially a low-pass filter (LPF), where the series inductors and the shunt capacitances stop the high frequency component while passing the low frequency component. When both ends of the ATL are terminated with the matched impedance, it has a cut-off frequency given by [19]:

$$f_{ATL} = \frac{1}{\pi \sqrt{LC}}$$  \hspace{1cm} (2.11)

where $L$ is the inductance of per section of the transmission line and $C$ is the capacitance of the individual photodiodes.

With the capacitance distributed to form an ATL, the overall bandwidth is now dependent on the transit time of the individual photodiodes, where the maximum is limited by $f_{ATL}$. On the other hand, the total absorption volume is the sum of all the discrete photodiodes. Hence it can give a higher quantum efficiency when the outputs of the array of lumped photodiodes add coherently. This requires the optical and the electrical signals to take the same time to travel in each section. For the monolithic P-TWPD, most of the electrodes are laid on the substrate or lowly doped semiconductor layers in order to achieve a lower loss. As a result, the propagating electrical fields are not confined by the semiconductor layers and are allowed to travel freely in the substrate and the air. This gives an electrical velocity, similar to that stated in Eqn 2.8a, which is higher than the optical velocity. Therefore, velocity matching can be achieved by having the electrical transmission line longer than the optical waveguide. In spite of this, there will still be a certain degree of velocity mismatch due to the multimode of the optical waveguide, where different mode propagates at different
velocities. The optical waveguide is not usually designed as a single mode waveguide because a larger input aperture is preferred for more efficient input coupling.

Both the TWPD and P-TWPD can improve their saturation current by reducing their current densities with a longer device length, which can be achieved through careful distribution of the optical waveguide absorption characteristics. Unlike the TWPD, since P-TWPD has lower microwave loss, as explained above, it can have more sections to increase the overall active device area and hence the saturation current. Another simpler way to increase the saturation current is to have a larger discrete photodiode in each section. The larger capacitance incurred can be offset by increasing the inductance of the transmission line section to maintain the impedance matching, however, this reduces the $f_{tgt}$ limitation.

Another factor that affects the bandwidth is the reverse input termination. When the reverse input is open circuit, the backward travelling wave is reflected and reduces the bandwidth. On the other hand, the reverse input termination will half the efficiency. A plausible solution to overcome this trade-off is impedance tapering, which cancels the reflected backward travelling wave from the input through a multi-sections transmission line [20,21]. The line impedances of the different sections are tapered down toward the output end, which causes a portion of the forward propagating current to be reflected back, where the amplitude depends on the impedance ratio of the consecutive section. The generated backward wave in each section is then cancelled out by this reflected wave when both the wave amplitudes are equal. Thus no reverse input termination is required and a higher bandwidth can be achieved without sacrificing the efficiency.

To date the best reported bandwidth-efficiency product for the P-TWPD is 7.8GHz by Droge et al. [22], where the 78GHz bandwidth and 10% efficiency is achieved without reverse input termination and velocity matching. Another work by Hirota et al. [23] reported a bandwidth of 115GHz but with an efficiency of only 6% due to the reverse input termination. Both P-TWPDs operate in the 1.55μm wavelength and neither the output power nor the current are given. On the other hand, a velocity matched but without reverse input termination P-TWPD by Lin et al. [24], operating
in the 0.86μm wavelength, gives a bandwidth of 49GHz with an efficiency of 12.3%. It has a high output current of 56mA. Recently, a P-TWPD by Murthy et al. [21] with impedance tapering and operating at 1.55μm, gives a bandwidth of 38GHz with an efficiency of 19.22% and a linear output power of −1dBm at 40GHz is reported.

2.2 SATURATION CURRENT

The saturation current is defined as the DC photocurrent where the AC output power is compressed from the linear response by 1dB [25]. When saturation occurs, the AC output power is no longer proportional to the input optical power. In the time domain, the peak of the photocurrent is reduced and the pulse width is increased. In the frequency domain, the power of the fundamental harmonic is reduced while the power of other undesirable harmonics are increased. This non-linearity limits the maximum available output current.

Under intense illumination, the high concentration of photogenerated carriers increases the photocurrent density considerably. This causes a space-charge effect or an electrical-field screening effect, which results in a non-uniform redistribution of electric field across the depletion region, as shown in Fig 2.13 [3]. The electric field near the centre depletion region collapses as the current density increases. There is a minimum critical electric field, assumed to be 50kV/cm for carriers to travel at saturation velocity, while a lower electric field will cause the carrier to travel below the carrier saturation velocity. The lower carrier velocity will increase the transit time and non-linearity occurs, limiting the maximum output current.

To increase the saturation current, the space-charge effect must be reduced. One way is to increase the carrier saturation velocity using the uni-travelling-carrier photodetector (UTC-PD). Its principle is to use only the electrons as the carriers because they travel at a higher velocity than the holes, as seen in Fig 2.14.
Figure 2.13: Electric Fields of PIN Photodetector [3]

Figure 2.14: InGaAs and GaAs Carrier Velocities [10]
In the conventional photodetector, the carrier saturation velocity is dependent on both the electron velocity \( (u_e) \) and the hole velocity \( (u_h) \) is given as [3]:

\[
\frac{1}{u_{sat}^4} = \frac{1}{2} \left( \frac{1}{u_e^4} + \frac{1}{u_h^4} \right)
\]  

In order to suppress the slower travelling holes from the carriers, a p-doped layer instead of an intrinsic photoabsorption layer is used together with a transparent wide-bandgap depletion layer, as shown in Fig 2.15. In the p-typed photoabsorption layer, the photogenerated majority holes are annealed at the rate of the dielectric relaxation time, which is the time constant for the space charge to neutralise, after being disturbed by a random fluctuation of carrier density. While the electrons diffuse towards the depletion layer. The electrons can then drift across the thin depletion layer at their overshoot velocity, which is significantly higher than \( u_{sat} \) and thus reduces the current density. In this way only the fast travelling electrons contribute to the photocurrent and the saturation current is increased. However, the slow diffusion process in the photoabsorption layer reduces the transit time limited bandwidth. To increase the speed of the diffusion, the doping profile of the photoabsorption is graded to introduce a gradient on the conduction band and a bandwidth comparable to that of the conventional photodetector can be achieved. The best reported 1.55\( \mu \)m wavelength UTC-PD by Ito et al. [26] has a 30mA peak output current with a 3dB bandwidth of 185GHz, however, its efficiency was not given.
Another way to increase the saturation current is to decrease the average photocurrent density by increasing the effective photoabsorption area. However, this increases the capacitance, thereby reducing the CR time constant limited bandwidth. Therefore, there is a trade-off between the bandwidth and saturation current. The solution is through the use of the distributed device scheme. For the TWPD and P-TWPD, the optical signal is fed via a leaky waveguide, where the light is coupled evanescently to the photoabsorption layer. The distribution of the optical signal is controlled by the confinement factor of the optical waveguide. This helps to reduce the photocurrent density and thus increases the saturation current. As the TWPD distributes the junction capacitance and the current density concurrently, the saturation current can be increased without compromising the bandwidth by having a longer device length. For the P-TWPD, the output photocurrent is the sum of all the discrete photodiodes. Similarly, the saturation current can be increased by having more photodiodes, without degrading the bandwidth. Unfortunately this coupling from the optical waveguide to the photoabsorption layer is difficult to control, especially for the monolithic fully distributed TWPD. Also, as the photodetector is not uniformly illuminated, the saturation current will appear at the point where the photocurrent density is highest.

In addition, unwanted parasitic resistances can also cause non-linearity. At high current level, significant voltage is developed across these resistances, which reduces the bias voltage across the intrinsic region and consequently reduces the carrier velocity. Both the space-charge effects and the non-linearity of parasitic resistances can be reduced by increasing the applied bias voltage. Another factor that causes non-linearity is the photogeneration of undesirable carriers in other semiconductor layers, with a bandgap energy smaller than the photon energy, when illuminated. These carriers are outside the intrinsic region. They are not assisted by the electrical field and cannot travel at the saturation velocity, causing non-linearity in the output current.

Besides saturation current, the maximum output current of the photodetector is also limited by the electric field. As seen from Fig 2.13, the electric field at both the junctions increase as the photocurrent density increases and breakdown occurs when the maximum critical field level is exceeded. During breakdown, the junction is
forward biased instead of reverse biased and causes a very large current flowing through, which may damage the photodetector. Another limiting factor is the thermal effect caused by the electrical power consumed inside the photodetector. It increases with the output current and ultimately leads to thermal destruction.

2.3 SUMMARY

The performances of the reported photodetectors are summarised in Fig 2.16. The bandwidth-efficiency trade-off of the VPD limits its bandwidth-efficiency product to around 30GHz despite various techniques like RCE, mushroom-mesa and air-bridge have been employed to extend its performance. The WGPD and WG-fed-PD have a better bandwidth-efficiency product of greater than 50GHz by using edge illumination to enable the bandwidth and efficiency to be optimised separately. However, it is the ability of the TWPD to distribute its capacitance that boosts the performance by another 50% to a staggering 76GHz. Though the P-TWPD has a poor bandwidth-efficiency product of less than 10GHz, it has a high output power of -1dBm.
CHAPTER 3

HETEROJUNCTION PHOTOTRANSISTOR

The photodetectors discussed in Chapter 2 have a very low output signal level and require an amplifier in order to drive the subsequent circuits. The heterojunction phototransistor (HPT) is an attractive high-speed photodetector for optical communications, as it can provide high gain, low noise, and low operating power requirements [27]. Moreover, its structure and fabrication processes are similar to the established heterojunction bipolar transistor (HBT), which is capable of ultra-high speed operations [28].

The chapter begins by describing the principle of operation for the bipolar junction transistor (BJT), HBT and HPT. The main figure of merit of the phototransistor, its gain and cut-off frequency, are also discussed. The effects of having a high base bias and the significance of the base terminal of the HPT are presented next. This is followed by a review of novel 2T-HPT structures and the integration with optical waveguides. These works led to a new design of an edge illuminated epitaxial waveguide fed 2T-HPT, which is able to distribute the photoabsorption profile along the length of the device. Finally, the layout design of the emitter and collector metal contacts for the phototransistor are presented together with the analysis of the transmission line loss.

3.1 PRINCIPLE OF OPERATION

Fig 3.1 shows an npn bipolar junction transistor (BJT). The emitter-base junction is forward biased and the electrons diffuse through the junction, where the electron diffusion current density is a function of the electronic charge \( q \), the base minority-carrier diffusion coefficient \( D_{nB} \) and the gradient of the doping concentration in the \( x \) direction \( n(x) \) [29]:

\[
J = \frac{q n(x) D_{nB}}{2 \mu_n V_{BE}} \frac{d n(x)}{dx}
\]
\[ J_{\text{diffusion}}(x) = qD_{nB} \frac{dn_n(x)}{dx} \]  

(3.1a)

where \[ D_{nB} = \frac{kT}{q} \mu_{nB} \]  

(3.1b)

and \( kT/q \) is the thermal voltage at 300K and \( \mu_{nB} \) is the base electron mobility.

Figure 3.1: npn Bipolar Junction Transistor (BJT)

Therefore, the doping concentration of the emitter layer has to be significantly higher than its base layer. In this way, the amount of majority holes from the lightly doped base diffusing into the emitter layer, which formed the minority-carrier in the emitter is minimised. These unwanted holes are supplied by the base terminal, therefore increases the input base current, which leads to a decrease in the gain. Conversely, the amount of desired majority electrons, from the highly doped emitter diffusing into the base layer, is maximised. The ratio of these injections is described by the emitter injection efficiency. For a homojunction npn BJT, it is the ratio of the desired forward electron current to the overall emitter current and is given as [30]:

\[ \gamma_{BJT} = \left[ 1 + \frac{D_{pE}L_{nB}N_{B}}{D_{nB}L_{pE}N_{E}} \tanh\left( \frac{X_{\text{NeutralBase}}}{L_{nB}} \right) \right]^{-1} \]  

(3.2)

where \( D, L, N \) and \( X \) are the minority-carrier diffusion coefficient, minority-carrier diffusion length, doping concentration and width of the neutral base, respectively. The subscript \( B \) and \( E \) denote the base and emitter layers of the transistor, whereas \( n \) and \( p \) denote the electrons and holes, respectively.
On the other hand, the electron drift current density through the reversed biased base-collector junction, where the drift is a function of the electronic charge \( q \), the electron mobility \( \mu_n \), the doping concentration \( n(x) \) and the electric field \( E(x) \) in the \( x \) direction [29]:

\[
J_{\text{drift}}(x) = q\mu_n n(x)E(x) \tag{3.3}
\]

At sufficient high electric field the term \( \mu_n E(x) \) should be replaced by the saturation velocity \( u_{sat} \). Together with a significantly higher doping concentration of the base than the collector, the collector layer is fully depleted and the electrons can drift across the entire collector at saturation velocity. Moreover, any undepleted region of the collector will cause an epitaxy resistance and may leads to the quasi-saturation effect at high operating current. In addition, a significantly higher base doping level with respect to the collector helps to reduce the base width modulation or Early effect, when the collector-emitter voltage is increased. Both effects will be discussed in details later in Chapter 5.3.1.

The desired electrons injected from the emitter must travel across the base layer so that they are collected in the collector layer. This is achieved by making the base layer thickness significantly smaller than the diffusion length \( X_b/L_{\text{diff}} \), which is a function of the minority-carrier lifetime \( \tau_{\text{Carrier Lifetime}} \). The ratio of the amount of the electrons reaching the collector and the amount of electrons injected from the emitter is determined by the base transport factor given as [31]:

\[
\alpha_{\tau_0} = \frac{1}{\cosh \left( \frac{X_{\text{Natural Base}}}{L_{\text{diff}}} \right)} \approx \left( 1 - \frac{X_{\text{Natural Base}}^2}{2L_{\text{diff}}^2} \right) \tag{3.4a}
\]

where \( L_{\text{diff}} = \sqrt{D_{\text{diff}} \cdot \tau_{\text{Carrier Lifetime}}} \) \( \tag{3.4b} \)

The above analysis shows that the base layer doping concentration must be significantly higher than the collector's but significantly lower than the emitter’s \( (N_B >> N_B >> N_C) \). The restricted base doping concentration gives an undesirable high base resistance while the high emitter doping concentration gives an undesirable high
base-emitter junction capacitance. These make the transistor not suitable to operate in the high frequency region.

A solution is to use a heterostructure base-emitter junction, as shown in Fig 3.2. The capital N (in Np) rather than a lowercase letter n is used to reflect the characteristic that the emitter is made of a larger-energy-gap material. The emitter and base layers are of dissimilar materials having different energy gaps ($E_g$). Due to the large valence band discontinuity ($\Delta E_V$) of the heterojunction, the holes are trapped in the base and are prevented from being injected into the emitter region. On the other hand, the conduction band discontinuity ($\Delta E_C$) is small enough, so that the electrons in the emitter are able to inject into the base region. This allows the base doping concentration to be higher than the emitter’s to reduce the base resistance, while the emitter doping concentration to be lowered to reduce the base-emitter junction capacitance.

![Figure 3.2: Band Diagram of Np Heterojunction](image)

The emitter injection efficiency of the BJT in Eqn 3.2 is modified by the heterojunction of the HBT as [2]:

$$\gamma_{HBT} = \left[1 + \left(\frac{D_{pB} X_{nB} N_B}{D_{nB} X_{pB} N_E} \exp\left(-\frac{\Delta E_g}{kT}\right)\right)^{-1}\right]$$

(3.5)

The emitter injection efficiency is now dominated by the difference in the emitter and base energy gaps ($\Delta E_g = \Delta E_V - \Delta E_C$), which has an exponential relationship. As a result, the HBT can have a heavily doped base and a lightly doped emitter. This allows its frequency response and gain to surpass that of the homojunction BJT.
HBT can be used as a phototransistor (HPT) by incorporating the photoabsorption function in the reverse biased base-collector junction, which is similar to the junction of a PIN photodiode. Hence instead of applying an electrical signal directly into the base, as in the HBT, the base-collector depleted junction generates the necessary base current from the input optical power, which is then amplified by the transistor action.

In order for the reverse biased base-collector junction to absorb the optical signal, the energy gap \( E_g \) of the base and collector semiconductor layers must be smaller than the product of the optical frequency \( f \) and Planck's constant \( h \), as stated in Eqn 2.1. The absorption process that converts the optical signal to electrical signal by the generation of electron-hole pairs, occurs in the depleted base-collector junction. However, the base layer is usually very thin and highly doped; hence the depletion layer is assumed to reside entirely in the collector layer. At sufficiently high electric field, the collector layer is fully depleted and the electrons can drift across the entire collector at saturation velocity. For an Npn HPT, the electric field inside the depletion layer will sweep the electrons toward the collector layer and the holes toward the base layer, as shown in Fig 3.3.

Subsequently, the electrons will be collected by the collector's electrode whereas the holes upon entering the base become majority-carriers and begin to diffuse toward the base-emitter junction. However, it will soon be trapped by the large energy band-gap formed by the base-emitter heterojunction. Very few holes actually managed to overcome this energy barrier to reach the emitter's electrode while some of it will recombine with the electrons.
The optical gain or amplification, through transistor action, is achieved if the lifetime of the electrons in the base, injected from the emitter, is longer than the transit time across the base. The majority of the holes accumulated at the base-emitter junction, as seen in Fig 3.3, increase the base-emitter voltage, which further forward biased the base-emitter junction. This helps to increase the number of electrons flowing from the emitter to the collector, hence increases the gain.

### 3.1.1 PHOTOTRANSISTOR GAIN

The optical gain of the Npn HPT, which simply relates the number of electrons (or holes for Pnp HPT) in the collector current to the number of incident photons, can be approximately expressed as [32]:

\[
G = \eta \beta \quad (3.6a)
\]

\[
\beta = \frac{I_C}{I_B} \quad (3.6b)
\]

where \( \eta \) is quantum efficiency and is usually low due to poor input coupling efficiency, and \( I_C \) and \( I_B \) are the collector and base bias current, respectively. While \( \beta \) is the transistor base-to-collector amplification factor, also known as the common-emitter current gain. With proper transistor design, a high current gain can be achieved, which helps to offset the poor quantum efficiency. The common-emitter current gain can also be expressed as [32]:

\[
\beta = \frac{\alpha_B}{1 - \alpha_B} \quad (3.7a)
\]

where \( \alpha_B \) is the common-base current gain given as:

\[
\alpha_B = \gamma_{HPT} \cdot \alpha_{T0} \quad (3.7b)
\]

and \( \gamma_{HPT} \) is the emitter injection efficiency and \( \alpha_{T0} \) is the base transport factor. To maximise the desired high common-emitter gain, both the emitter injection efficiency and the base transport factor must be near unity. The base width is typically less than one-tenth of the diffusion length, which gives a near unity \( \alpha_{T0} [30] \). When the emitter-base junction is forward biased, the potential barrier created by the heterojunction
effectively eliminates the hole injections from the base into the emitter. This gives a near unity emitter injection efficiency. Thus the HPT is able to provide a very high current gain.

### 3.1.2 PHOTOTRANSISTOR CUT-OFF FREQUENCY

The most important and useful figure of merit of the HPT is the gain-bandwidth product, also known as the cut-off frequency ($f_T$). It is the frequency where the gain has dropped to unity, under short circuit output condition. $f_T$ is determined from the h-parameters, which are extracted from the measured 2-port S-parameters. HPTs are electrical 1-port devices and their $f_T$ cannot be determined directly by measurement. Though the cut-off frequency measured is 50Ω loaded ($f_C$), the frequency response limitations for a HPT can still be studied through $f_T$.

The equation for determining the BJT and HBT cut-off frequency is based on Cooke’s method and was modified by Allsopp *et al.* [33] for the HPT analysis:

$$f_T = \frac{1}{2\pi (\tau_b + \tau_{sc} + \tau_c + \tau_e)}$$

The base transit time ($\tau_b$) is the time required for the minority-carriers to diffuse through the base from the emitter to the collector [2]:

$$\tau_b = \frac{X_B^2}{2.43 D_{nB}}$$

where $X_B$ is the base thickness, $D_{nB}$ is the base diffusion coefficient and the factor 2.43 is due to the uniform base doping profile. The space-charge transit time ($\tau_{sc}$) is the time required for the carriers to drift across the base-collector depletion thickness ($X_{depC}$) [2]:

$$\tau_{sc} = \frac{X_{depC}}{2\nu_{Sat}}$$

where the carriers is assumed travelling at its saturation velocity ($\nu_{Sat}$). The factor 2 is due to the averaging of the sinusoidal current over the period.
The collector charging time ($\tau_c$) is the time required to charge the collector junction capacitance ($C_{BC}$) through the emitter and collector resistances ($R_E$ and $R_C$) [2]:

$$\tau_c = C_{BC}(R_E + R_C) \quad (3.11)$$

The emitter charging time ($\tau_e$), which is derived from the forward response time of the device, is the time required to charge both the emitter and collector junction capacitances ($C_{BE}$ and $C_{BC}$) through the small signal junction resistance ($r_e$) [2]:

$$\tau_e = r_e(C_{BE} + C_{BC}) \quad (3.12a)$$

and

$$r_e = \frac{kT}{qI_C} \quad (3.12b)$$

where $kT/q$ is the thermal voltage at 300K and $I_C$ is the output collector current.

Allsopp et al. [33] also take into consideration of the diffusion capacitance. It is excluded here, as it is not applicable at high frequency, i.e. those approaching $f_T$ [2]. For lightwave applications, the $f_T$ of HPT is mainly dominated by the emitter charging time ($\tau_e$) and collector charging time ($\tau_c$) due to the high capacitances associated with the large device area required for effective input coupling and photoabsorption. However, the heterostructure allows a heavily doped base to reduce the base resistance and a lightly doped emitter to reduce the base-emitter junction capacitance; therefore, a better frequency response.

### 3.2 2-TERMINAL AND 3-TERMINAL HPTs

The first HPT was designed with only two terminals (2T). Its base terminal is floating, as the base current is generated optically. The 2T-HPT is optically biased by the induced base current generated directly by the time average input optical power. This gives the advantage of low power consumption. In addition, it has the advantage of having no base contact parasitic capacitance. It also has a lower shot noise level, as it is not biased electrically. In addition, the 2T-HPT also gives a very simple design, which allows straightforward and low-cost fabrication process.
3.2.1 EFFECT OF INPUT OPTICAL POWER ON GAIN

However, with this configuration, the gain of the device is reduced when the input optical power is low. This is because of the undesirable recombination centres, which are defects that formed inherently in the heterojunction [27]. These recombination centres causes a defect current, which is not injected into the base. As a result, the emitter injection efficiency and hence the gain are reduced. When the input optical power is low, the defect current is significant with respect to the total emitter current. The reduced output collector current increases the charging time for the junction capacitance, as shown in Eqn 3.12, which consequently leads to a decrease in the cut-off frequency.

This problem can be overcome by biasing the 2T-HPT by a separate optical source, as shown in Fig 3.4 [34]. The light emitting diode (LED) integrated onto the same chip as the HPT generates an additional base current optically, which helps to increase the output collector current. As a result, it gives a threefold reduction in the rise time and twofold increment in the output pulse amplitude, which reflects an increase in both the cut-off frequency and gain of the HPT.

![Figure 3.4: HPT Externally Optical Biased by an LED [34]](image-url)
Alternatively this additional base bias current can be supplied by an external current source through the base terminal of a 3T-HPT [35]. The result in Fig 3.5a shows, with the input optical power maintained constant at 1μW, the base current is used to increase the bias and hence the output collector current. Using the same 3T-HPT but with the base left floating, the output collector current as a function of the input optical power for a 2T-HPT was also measured, as shown in Fig 3.5b. The gain for both HPT set-ups were then computed using the expression [35]:

$$G = \frac{hc}{\lambda q} \times \frac{I_{C_{\text{opt}}}}{P_{IN}}$$  \hspace{1cm} (3.13)

where $hc/\lambda$ is the energy of the incident photons, $I_{C_{\text{opt}}}$ is the component of the collector current generated due to the incident power $P_{IN}$ and $q$ is the electronic charge. The result shows that with an input optical power of 1μW, the external base current enhanced the gain of the HPT by fivefold.

Figure 3.5: (a) Gain of 3T-HPT as a function of the Output Collector Current, which is varied by its External Base Current (with $P_{IN}=1\mu W$)  
(b) Gain of the 2T-HPT as a function of the Input Optical Power  
(c) Gain of the 3T-HPT as a function of the Input Optical Power (with $I_C=5.5mA$) [35]
The constant gain dynamic range for the 3T-HPT is shown in Fig 3.5c (inset). The device was biased near the peak of the gain curve, at an output collector current of 5.5mA with a base current of 11μA. The gain is fairly constant around 190 over the range of input optical power. Hence it has been demonstrated that the gain dependence on the optical power can be overcome by having an external base current. This is due to saturation of the recombination centre by the external base current, which is responsible for the lowering of the gain at low input optical power. In addition, the higher gain brought about a higher collector current, and increases the cut-off frequency, by reducing the emitter and collector charging time.

3.2.2 EFFECT OF BASE TERMINAL ON FREQUENCY RESPONSE

As defined earlier, the frequency where the gain drops to unity is the cut-off frequency or the gain-bandwidth product of the transistor. On the other hand, the 3dB bandwidth ($f_{3dB}$) is the frequency range where the output response drops by 3dB, and it is an inverse function of its gain [36]. For frequency greater than $f_{3dB}$, the gain rolls off at the rate of 20dB/decade [37]. The inherent gain of the typical 2T-HPT is high due to the increase of the forward base bias by the accumulation of the charges in the base. This gives a small $f_{3dB}$ for the 2T-HPT. An additional base terminal is added to form a 3T-HPT, which can give a larger $f_{3dB}$ through reducing the gain. The $f_{3dB}$ can be further improved by optimising either the base terminal circuit or the base bias voltage. Nevertheless this $f_{3dB}$ increment does not affect the maximum gain-bandwidth product, which mainly depends on the operating current and the junction capacitances.

The base terminal circuit can be optimised using additional passive components, as shown in Fig 3.6 [38]. The serially connected inductor and resistor formed a low pass filter to allow further degree of optimisation. The capacitor in the circuit is mainly for biasing purpose only. The simulated response of the base terminal connected with different value of resistances ($R$) and inductances ($L$) is shown in Fig 3.7. In addition, the simulated response of the discrete HPT, where its base is terminated with a 50Ω
load, is also shown on the same plot for comparison. The result shows that the HPT with the base terminal circuit has a significantly higher $f_{3dB}$ than the discrete HPT but has a lower gain at low frequency. In the base terminal circuit, the resistance determines the low frequency gain while the inductance determines the high frequency gain.

![Figure 3.6: HPT Base Terminal with Additional Passive Components [38]](image)

![Graph showing relative output power vs. frequency for different load resistances](image)
The $f_{3dB}$ can also be increased through optimising the base bias voltage. The high gain of the HPT at low frequency is determined by the charges stored in the base, which set-up the base-emitter voltage. At the same time, the bandwidth is also determined by the rate of the elimination of these charges, which are either extracted via the base terminal or carrier recombination. The time taken for the latter process is determined by the carrier lifetime ($\tau_{\text{CarrierLifetime}}$), which is a function of the doping concentration and the material properties, independent of the device geometry and DC bias [34]. The effect of this long decaying time for the charges in the base can be observed in the impulse response of a 2T-HPT, which has an exponential decaying response [39]. Hence these excess charges in the base are extracted via the base terminal for a 3T-HPT at a faster rate, where this rate can be optimised by the base-emitter bias voltage [40]. This is shown in the time domain measurement in Fig 3.8. The result shows that at a base-emitter bias voltage of 0.7V, the impulse response settle to the initial level.
fastest, which indicates the complete removal of the excess hole. This yields a $f_{3dB}$ of 73GHz.

![Figure 3.8: Measured Emitter Photocurrent Transients at an Incident Optical Energy of 1.1pJ/pulse. The curves are displaced vertically from one another for clarity [40](image)](image)

3.3 EDGE ILLUMINATION

Edge illumination or edge coupling is a very excellent design to overcome bandwidth-efficiency trade-off, especially for the long wavelength photodetector, which requires a longer absorption length. Unlike the vertically illuminated photodetector, the photon travels along the length of the device instead of the thickness of the photoabsorption layer. In this way, the photon and carrier transit path are orthogonal. This enables the electrical bandwidth, which is also inversely proportional to carrier transit time, and the absorption efficiency, which is proportional to the absorption length, to be optimised separately as a function of the parameters of the layer. This feature is critical when the HPT is designed for high power operations, as shown later.
The 2T-HPT simple design of only two electrical contacts, which is similar to the PIN photodetector, makes it suitable for edge illumination. The edge coupled 2T-HPT was first demonstrated by Wake et al. [41] with the active device area designed in a planar waveguide geometry. The InGaAs base and collector layers formed the core of the waveguide, while the InP emitter and substrate layers below formed the claddings of the waveguide. The unity gain cut-off frequency, as shown in Fig. 3.9, is determined by comparing the frequency response with a reference edge coupled photodiode.

![Graph showing frequency response of Edge Coupled Phototransistor and Photodiode](image)

Figure 3.9: Frequency Response of Edge Coupled Phototransistor and Photodiode [41]

It is crucial to ensure the coupling efficiency of the edge coupled HPT is equal or less than the edge coupled photodiode, validating the cut-off frequency determined by using the photodiode output as a reference for unity gain. Assuming the above condition is adhered, the edge coupled HPT is observed to have a unity gain cut-off frequency of approximately 40GHz. It also shows that it has a very high gain at low frequency, which is an undesirable effect as it can cause the device to be thermally damaged, especially for small devices. The dimensions of the device are small, about 5μm by 10μm, this reduces the device capacitance to achieve a high unity gain cut-off frequency.
3.3.1 DISTRIBUTED PHOTOABSORPTION PROFILE

The main advantage of the edge illumination is its ability to distribute the photoabsorption profile so that the photocurrent density can be reduced to give a high saturation current level. For the conventional WGPD, the absorption layer of the photodetector also functions as the core layer of the waveguide. In order to achieve a high bandwidth, the absorption layer has to be thin so that the transit time is small. However, a thick core layer is required to achieve a good input coupling efficiency. This causes a trade-off between the bandwidth and the coupling efficiency, which can be overcome by physically separating the core layer and the absorption layer. This is achieved through the integration of a leaky optical waveguide along the length of the photodetector. The waveguide will feed the optical signal to the absorption layer along the length of the photodetector through evanescent coupling. The core layer thickness of the waveguide can be increased to give a larger input aperture so that an efficient input coupling is achieved. While the absorption layer thickness can be minimised independently to reduce the transit time. This design is similar to the WG-fed-PD, as discussed in Chapter 2.1.2.

In a later work, the power handling capability of the distributed photoabsorption coefficient HPT was demonstrated by Scott et al. [42]. The design integrated a leaky Polyimide optical waveguide on top and along the length of the HPT. As the refractive index of the Polyimide waveguide \( n_{\text{Polyimide}} = 1.520 \) is less than that of the InGaAs semiconductor layer beneath \( n_{\text{InGaAs}} = 3.368 \), the optical signal continuously leaks as it propagates. The rate of leakage from the waveguide, i.e. the HPT absorption coefficient, is controlled by the dimensions of the waveguide. In this way, the current density is distributed along the length of HPT, instead of concentrating at the front end, so that the saturation current density is increased greatly. As shown in Fig 3.10, the RF output power measured with a 60GHz modulated input optical power increases linearly up to 50mA of the DC photocurrent without any sign of saturation. The frequency response was not given, however, the unity gain cut-off frequency is expected to be poor due to the large junction capacitances associated with the large surface area device of 20μm by 200μm.
Figure 3.10: Measured RF output power vs DC photocurrent for (a) Lumped Element HPT (b) Distributed 200μm HPT [42]
3.3.2 EPITAXIAL OPTICAL WAVEGUIDE

Another simpler way of integrating an optical waveguide with the HPT is to utilise the different refractive index of its epitaxial layers. For ternary or quaternary semiconductors, their refractive indices can be varied by changing their compositions. In this design, the photocurrent density in the absorption layer is reduced by increasing the confinement of the light inside the waveguide, so that the evanescent field decreases. This is achieved either by increasing the refractive index differences between the core layer and the top cladding layer of the waveguide, or increasing the cladding layer thickness between the absorption layer and the core layer.

The optical characteristic of an epitaxial waveguide fed 2T-HPT is simulated using the BeamPROP™ software from RSoft. The structure model and epitaxial layers are given in the next section (Fig 3.14 & Table 3.1). The integrated waveguide consists of a lattice-matched undoped In$_{0.53}$Al$_{0.14}$Ga$_{0.33}$As as the core layer, sandwiched between the InP sub-collector and the InP SI substrate layers, which formed the cladding layers, as shown in Fig 3.11.

![Figure 3.11: Proposed WG-fed-PD Partial Structure Model with its Refractive Index Profile](image-url)

---

**Bottom Cladding Layer**: InP SI SUBSTRATE

**Core Layer**: InAlGaAs WAVEGUIDE

**Top Cladding Layer**: InP SUB-COLLECTOR

**Absorption Layers**: InGaAs COLLECTOR

**InGaAs BASE**

Refractive Index

3.544

3.441

3.350

3.167

---
As the refractive index of the In$_{0.53}$Al$_{0.14}$Ga$_{0.33}$As ($n_1=3.441$) [43,44] is higher than that of the InP ($n_2=3.167$) [45], most of the optical signal propagates within the core layer in the longitudinal direction. The other part of the optical signal propagating in the InP layers is known as the evanescent field. The power of the evanescent field increases as the difference between the core and cladding layer refractive index decreases. The top cladding layer, which is the thin sub-collector layer (0.2µm), allows the evanescent field to extend into the InGaAs absorption layers above. In this way, the photodetector is feed continuously along its length by the evanescent field as the optical signal propagates along the waveguide. On the other hand, the bottom cladding layer, the thick InP substrate layer (350µm), contained the evanescent field within it.

![Simulated Waveguide Modal Fields](image)

**Figure 3.12: Simulated Waveguide Modal Fields**
The excessively wide waveguide leads to multiple modes propagating in the waveguide, as seen in the Fig 3.12. The simulated input optical source has a spot size of 1.8µm, which is similar to the lensed fibre used in the measurement. It is directed at the centre of the waveguide for maximum input coupling. The fundamental and first order modal fields of the TE and TM are similar, hence only that of the TE are shown in the figure. This indicates that the waveguide has very low polarisation dependence. This is also validated by the very close computed effective index of the TE and TM modes, of 3.41611 and 3.41497, respectively. It is also noted that the higher order modal field spreads laterally beyond the region where the evanescent couplings to the photoabsorption take place and hence, reduces the overall efficiency.

The length of the photodetector has to be considered carefully. Sufficient length is required to fully absorb the optical power while excessive length induces unnecessary electrical loss. The desired length can be determined from the simulated result of the optical power in the waveguide as a function of its length, given in Fig 3.13.

Figure 3.13: Simulated Total Optical Power and Waveguide Power vs the Length of Waveguide
The result shows that, inside the waveguide the power falls off following an exponential law \(e^{-\alpha z}\). A device length of \(\sim 200\mu \text{m}\) and \(\sim 400\mu \text{m}\) is required to absorb 80% and 95% of the optical signal, respectively. A small portion of the power is lost in the substrate during the initial few wavelengths before the propagating mode is established. In addition, a portion of the optical signal is lost as the evanescent field trapped in the substrate. These powers in the substrate are not available for current generation. It is also important to align and focus the optical beam at the centre of waveguide as any misalignment will reduce the coupling efficiency. The most critical is the upward offset on the vertical axis, which can cause large amount of optical power to be absorbed directly by the photoabsorption layer, instead of evanescent coupled from the waveguide. In this case, a high photocurrent will be generated at the photodetector input end and saturates the photodetector. This reduces the saturation power level of the photodetector and diminishes the merit of the photoabsorption distribution.

### 3.4 METAL CONTACTS

The design of the electrodes or metal contacts of the WG-fed-PD and TWPD are important. These are used to extract the carriers from the semiconductors and to transfer them to the output terminals; hence they are required to have low electrical loss. High loss will reduce the quantum efficiency and gives a poor frequency response. The loss is more significant toward the TWPD, as it utilises a long device length to achieve the travelling characteristics of large bandwidth and high saturation power.
3.4.1 TRANSMISSION LINE

In order to achieve a good ohmic contact, the metal contacts for the emitter and the collector of the transistor are deposited onto highly doped semiconductor layer, usually doped to its maximum level (the solubility of the dopant) [2]. This ensures that the depletion width at the contact interface is sufficiently narrow so that the carriers can tunnel through the barrier by field emission and reduces the parasitic resistance. However, the collector has to be lowly doped to ensure that it is fully depleted at high electric field, so that the carriers can travel at their saturation velocity. As a result, an additional sub-collector layer, which has a very high doping concentration, is introduced between the collector layer and the collector metal contacts. The sub-collector layer guarantees a low collector metal contact series resistance. On the other hand, the emitter metal contact is laid on its cap layer, which presence is to increase the heterostructure of the base and emitter layers critical thickness. The cap layer is usually highly doped and hence no additional intermediate layer is required. Similarly, the metal contacts for the PIN photodetector are deposited onto its highly doped p and n layers.

As the device length is extended in the WG-fed-PD and TWPD designs, the metal contacts also function as a transmission line for the carriers to propagate along the device length towards the load. For low loss transmission lines, the substrate or semiconductor underneath need to have a high resistivity. Good substrate has a very low loss tangent and is given by [46]:

\[
\tan \delta = \frac{1}{\omega \varepsilon \rho}
\]  

(3.14a)

where \( \omega \) is the radian frequency, \( \varepsilon \) is the permittivity and \( \rho \) is the resistivity given by:

\[
\rho = \frac{1}{q \mu N}
\]

(3.14b)

where \( q \) is the electronic charge, \( \mu \) is the carrier mobility and \( N \) is the doping concentration. As noted from Eqn 3.14, the doping concentration, which is inversely proportional to the resistivity, has to be low in order to give a low loss transmission line. Unfortunately, this contradicts with the requirement of a good ohmic contact, therefore these parameters need to be compromised.
3.4.2 ELECTRICAL FIELD PATTERN

The uniplanar design of the WG-fed-PD and TWPD metal contacts physically formed a coplanar waveguide (CPW) or coplanar strips (CPS) transmission line structure. CPW is preferred as it is directly compatible with the standard symmetrical structure of the transistor. In addition, it has a lower radiation loss compared with CPS. During normal operation of the HPT, the base-emitter (BE) junction is forward biased while the base-collector (BC) junction is reverse biased. Both the sub-collector layer and the forward biased BE junction have a low resistance, while the depletion layer formed by the reverse biased BC junction has a very high resistance. As a result, most of the electric field (E-field) is screened inside the BC junction, which essentially is the depleted collector layer. This formed a microstrip-liked electric field pattern, as shown in Fig 3.14. This effect is similar to the PIN photodetector, as described previously in Chapter 2.1.3, where the E-field is screened by the highly doped p and n layers. If the centre metal contact width is significantly greater than the depletion or intrinsic layer thickness, then together with the high relative dielectric constant of this semiconductor layer, the external fringing of the E-field is small. This gives a quasi parallel-plate electrical field pattern instead. Hence the propagating characteristic of the structure can either be microstrip or parallel-plate.

![Figure 3.14: Electrical Field Pattern of Typical HPT](image-url)
3.4.3 TRANSMISSION LINE LOSS

The transmission line microwave loss, is mainly dependent on the conductivity of the sub-collector layer immediately beneath the ground metal contacts, which determines the propagation mode. Using the electromagnetic simulator, SONNET version 6a, a 2-port CPW transmission line laid on an HPT structure, as shown in Fig 3.14, is simulated at 10GHz to analyse the effect of the conductivity of the sub-collector layer. The width and gap of the CPW transmission line are $W=15\mu m$ and $S=20\mu m$, with a device length of 1mm. The collector layer is assumed fully depleted with zero conductivity and the rest of the parameters are as shown in Table 3.1. Due to the excessive large memory and long processing time required, the thin spacer and cap layers of the photodetector are not included in the simulation.

<table>
<thead>
<tr>
<th>Material</th>
<th>Function</th>
<th>Thickness (µm)</th>
<th>Relative Dielectric Constant</th>
<th>Doping Concentration (cm$^{-3}$)</th>
<th>Conductivity (S/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGaAs</td>
<td>Cap</td>
<td>0.03</td>
<td>13.88</td>
<td>$N=1*10^{19}$</td>
<td>-</td>
</tr>
<tr>
<td>InP</td>
<td>Cap</td>
<td>0.01</td>
<td>12.56</td>
<td>$N=1*10^{19}$</td>
<td>-</td>
</tr>
<tr>
<td>InP</td>
<td>Emitter</td>
<td>0.08</td>
<td>12.56</td>
<td>$N=8*10^{17}$</td>
<td>$2.69*10^4$</td>
</tr>
<tr>
<td>InGaAs</td>
<td>Spacer</td>
<td>0.005</td>
<td>13.88</td>
<td>Undoped</td>
<td>-</td>
</tr>
<tr>
<td>InGaAs</td>
<td>Base</td>
<td>0.08</td>
<td>13.88</td>
<td>$P=1*10^{19}$</td>
<td>$1.44*10^4$</td>
</tr>
<tr>
<td>InGaAs</td>
<td>Collector</td>
<td>0.3</td>
<td>13.88</td>
<td>$N=5*10^{16}$</td>
<td>0</td>
</tr>
<tr>
<td>InP</td>
<td>Sub-Collector</td>
<td>0.2</td>
<td>12.56</td>
<td>$N=1*10^{19}$</td>
<td>$1.92*10^5$</td>
</tr>
<tr>
<td>InAlGaAs</td>
<td>Waveguide</td>
<td>1.5</td>
<td>13.31</td>
<td>Undoped</td>
<td>0</td>
</tr>
<tr>
<td>InP</td>
<td>Substrate</td>
<td>350</td>
<td>12.56</td>
<td>SI</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1: HPT Structure Epitaxial Layers

As seen in Fig 3.15, the variation of conductivity of the sub-collector layer does not give a constant relationship with the device propagation loss. When the conductivity is less than $2*10^4$S/m, the sub-collector layer behaves as a lossy substrate due to its high loss tangent. It supports a CPW propagation mode where higher conductivity increases the loss. Increasing the sub-collector conductivity further, makes it behaves as a lossy conductor, which leads to a quasi parallel-plate propagation mode. Now, the
higher conductivity helps to reduce the loss. In the latter propagation mode, the current travels across the sub-collector layer to the collector contacts, and then through these contacts towards the load. Meanwhile, the waveguide layer beneath the sub-collector layer functions as the substrate. Since the waveguide layer does not have any doping, it has a low conductivity and work as a good substrate for the quasi parallel-plate transmission line. In this way, the requirements of a low conductivity substrate for low loss transmission line and a high doping concentration layer for good ohmic contact are met. The typical doping concentration of an n-type InP sub-collector is $1 \times 10^{19} \text{cm}^{-3}$, this gives a mobility of $\sim 1200 \text{cm}^2/\text{Vs}$ and a conductivity of $\sim 1.92 \times 10^5 \text{S/m}$. From Fig 3.15, it has a quasi parallel-plate propagation mode with a loss of about $-2.5 \text{dB/mm}$.

![Figure 3.15: Loss as a function of Sub-Collector Conductivity @10GHz](image)

With the above computed conductivity level, the loss as a function of the width and gap of the CPW is also simulated at 10GHz. As shown in Fig 3.16, the loss increases with the CPW dimensions as the resistances associated with the transverse current paths across the sub-collector layer to the collector contacts increase. As a result, a smaller CPW dimension is preferred to minimise the loss. However, this increases the
parasitic resistance associated with the active region and eventually outweighs the initial loss reduction achieved.

**Figure 3.16: Loss as a function of the Widths and Gaps of the CPW**

3.4.4 GROUND METAL CONTACTS DESIGN

The previous analysis shows that it is feasible to lay the ground metal contact on the sub-collector layer, which can give a low resistance ohmic contact and support a low loss quasi parallel-plate transmission line. However, due to fabrication limitation, the entire ground metal contacts cannot be laid on the sub-collector layer. For common transistor design, the sub-collector layer is typically more than 1 μm thick. However, it has to be very thin for this design, 0.2 μm, in order for the optical signal to couple evanescently to the collector absorption layer above it. As a result, the etching process of removing the collector layer at both the sides requires a high precision depth control. Without a precision control, a slight over etching will remove the thin sub-
collector layer completely. Hence the etching process is stopped earlier, before reaching the sub-collector layer, and a thin collector layer is maintained as a safe margin. Consequently, the ground metal contacts are laid on the collector layer instead, as shown in Fig 3.17. This leads to an inevitable higher resistance (in the region of $10^4 \Omega \text{cm}^2$) for the ohmic contact.

![Diagram of the WG-fed-PD Structure Model with majority of its Ground Metal Contacts on the Collector Layout](image)

(NB: NOT DRAWN TO SCALE)

Figure 3.17: WG-fed-PD Structure Model with majority of its Ground Metal Contacts on the Collector Layout

The ground metal contacts can be redesigned, as shown in Fig 3.18, to reduce the ohmic contact resistance (to the region of $10^7 \Omega \text{cm}^2$) though requiring additional fabrication steps. The majority of the ground collector contacts are now laid on the un-doped optical waveguide layer or the SI substrate, with minimum contact on the highly doped sub-collector layer. However, the contact cannot be etched and stopped directly at the collector ridge because of notch formation at the ridge. Hence a small amount is overlaid onto the collector layer. However, the majority of the current only flows through the sub-collector contact due to its smaller resistance. Also, with majority of the metal laid on the un-doped waveguide layer or SI substrate, it gives a lower loss tangent than on the sub-collector layer for the current to propagate...
longitudinally along the length of the device towards the load. Between the waveguide layer and the SI substrate, the latter has a better loss tangent. However, due to the steep etch plane edge profile in the 1.5µm thick waveguide layer, the ground metal contacts cannot transit smoothly from the sub-collector layer onto the SI substrate and is prone to discontinuity. Therefore, the majority of the metal contact is laid on top of the waveguide layer instead. The proposed design maintains a good ohmic contact while minimising the transmission line loss.

Figure 3.18: WG-fed-PD Structure Model with majority of its Ground Metal Contacts on the Waveguide Layout

3.4.5 RESULTS AND ANALYSES

Back-to-back photodetectors with structure models, as shown in Figs 3.17 and 3.18, and epitaxy layer design, as shown in Table 3.1 are fabricated for characterising the transmission line loss. Two sets of CPW transmission line dimensions are used, $S=20\mu m$, $W=15\mu m$ and $S=10\mu m$, $W=7\mu m$, and a device length of 1 mm. The losses of the transmission line are measured using a network analyser from 1GHz to 20GHz, as described later in Chapter 4.1.3.
The measured result in Fig 3.19 confirms that the latter design, with majority of the ground metal contacts on the waveguide layer, has a lower loss. It also shows that the smaller CPW dimension with $S=10\mu m$ and $W=7\mu m$ has a lower loss.

The losses of the transmission line are simulated using the SONNET simulator. Due to the very large memory and long processing time required, the thin spacer and cap layers of the photodetector are not included in the simulation. Similarly, simulation is only carried out on the smaller CPW dimension ($S=10\mu m$ and $W=7\mu m$). The measured and simulated results are plotted in Figs 3.20 & 3.21. The measured losses are significantly higher than the simulated results with zero collector conductivity. Even though the collector layer is reversed biased, it still has a finite conductivity. When this is taken into account in the simulations, the loss increases and roughly matches the measured result. The mismatch remained is probably due to the exclusion of the thin spacer and cap layers in the simulation. In addition, the metal conductor loss is also not included.
Figure 3.20: Measured and Simulated Results for Design with Majority of its Ground Metal Contacts on the Collector Layer with CPW dimensions of $S=10\mu m$ & $W=7\mu m$

Figure 3.21: Measured and Simulated Results for Design with Majority of its Ground Metal Contacts on the Waveguide Layer with CPW dimensions of $S=10\mu m$ & $W=7\mu m$
The loss of the smaller CPW dimension ($S=10\mu m$ and $W=7\mu m$) for the two design are next simulated at 10GHz as a function of the collector layer conductivity. The simulated result in Fig 3.22 demonstrates that a slight increment of the collector layer conductivity incurred a great loss. The simulations also show that the design with the majority of the ground metal contacts on the waveguide layer has a lower loss, concurring with the measurements in Fig 3.19.

![Figure 3.22: Simulated Result for Design with Majority of its Ground Metal Contacts on the Collector Layer (solid line) and the Waveguide Layer (dashes) with CPW dimensions of $S=10\mu m$ & $W=7\mu m$](image)

### 3.5 SUMMARY

The basic operation principle of the HPT is derived from it correspond HBT and BJT. The gain and cut-off frequency, which are the main figure of merit of the HPT are analysed. The intrinsic gain of the HPT is limited by the quantum efficiency in addition to the emitter injection efficiency and base transport factor, which are similar
to the HBT. The cut-off frequency of the HPT is limited by the same factors as HBT, which are the base transit time, space-charge transit time, collector charging time and emitter charging time. However, the 2T-HPT has no base terminal and causes it to have a long discharging time constant limited by the carrier lifetime, which gives a small 3dB bandwidth. For the 3T-HPT, the base terminal gives an additional feature to trade-off between the gain and the 3dB bandwidth. This is achieved either by applying an optimised bias voltage or designing additional reactance circuit to the base terminal.

Another important figure of merit of the HPT is the saturation power, which requires distributing the photoabsorption profile of the absorption layer. This is achieved through edge illumination design and the integration of an optical waveguide with the HPT. The distributed design causes an increase in the device length, and the HPT electrical contacts become a transmission line. As the transmission line is laid on high conductivity semiconductor layers, it suffers heavy losses. As a result, there is a trade-off between the contact resistance and transmission line loss. This is shown in the design of an edge illuminated 2T-HPT integrated with an optical waveguide.
Characterisation of the photodetectors, which are the devices under test (DUTs), is indispensable for the design process. It measures the output performance and the parameters of the DUT, which are important for analysis and future design. However, the measured result depends not only on the DUT but also the measurement method used, equipment and the testing conditions. As a result, it is important to ensure that the measurement methods and the test conditions match the operating condition of the DUT, and the measurement set-up is calibrated.

The first half of this chapter discussed on the transmission line loss with details of the measurement set-up. The second half of the chapter concentrated on the frequency response, where various techniques are discussed.

### 4.1 S-PARAMETERS

The objectives for this part of the characterisation are to determine the transmission line loss of the photodetector. This is an important parameter for analysing the performance of the photodetector and is a function of frequency. It is obtained from the S-parameters measured by a network analyser. As shown in Fig 4.1, the four set of definitions for a 2-port device are expressed as follows [47]:

![Figure 4.1: S-parameter Definitions](image-url)
Input reflection coefficient when the output is matched, \( S_{11} = \frac{b_1}{a_1} \mid_{a_2=0} \) \hspace{1cm} (4.1a)

Output reflection coefficient when the input is matched, \( S_{22} = \frac{b_2}{a_2} \mid_{b_1=0} \) \hspace{1cm} (4.1b)

Forward transfer coefficient when the output is matched, \( S_{21} = \frac{b_2}{a_1} \mid_{a_2=0} \) \hspace{1cm} (4.1c)

Reversed transfer coefficient when the input is matched, \( S_{12} = \frac{b_1}{a_2} \mid_{b_1=0} \) \hspace{1cm} (4.1d)

where \( a_1 \) and \( a_2 \) are the incident wave on ports 1 and 2; \( b_1 \) and \( b_2 \) are the reflected wave on ports 1 and 2, respectively.

### 4.1.1 TRANSMISSION LINE LOSS

The metal electrodes of the photodetector also act as its transmission line and have an important effect on the performance of the photodetector. The transmission line loss is proportional to its length; hence the absorption efficiency improvement can be outweighed when the loss is excessive. The transmission line loss is characterised by the insertion loss (\( L_I \)) and is expressed in dB as [47]:

\[
L_I = 20 \log |S_{21}| \hspace{1cm} (4.2)
\]

As a photodetector is electrically a 1-port device, a back-to-back configuration, as shown in Fig 4.2, is fabricated to measure its \( S_{21} \). The output terminal of the photodetector is designed with a co-planar probe pad to facilitate the measurement using on-wafer probing. The dimension of the probe pad is designed for a microwave ground-signal-ground (GSG) probe with a pitch of 200\( \mu \)m. The device is subsequently cleaved to the required length to form the edge illuminated photodetector.
The impedance of the measured transmission line has to be matched to the standard 50Ω of the network analyser, as shown in Fig 4.3 [47]; otherwise the measured S-parameters include loss due to the impedance mismatch in addition to the loss due to the transmission line structure:

\[
S_{21|\text{mismatch}} = \frac{(1 - \left|\Gamma_L\right|^2)S_{21}(1 - \left|\Gamma_S\right|^2)}{\left|(1 - S_{22}\Gamma_L)(1 - S_{11}\Gamma_S) - S_{12}S_{21}\Gamma_L\Gamma_S\right|^2} \quad (4.3a)
\]

\[
S_{11|\text{mismatch}} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \quad (4.3b)
\]

where \(\Gamma_L\) and \(\Gamma_S\) are the load and source reflection coefficients, respectively. As the transmission line is a symmetrical 2-port device, its \(S_{12|\text{mismatch}}\) and \(S_{11|\text{mismatch}}\) are equal to its \(S_{21|\text{mismatch}}\) and \(S_{22|\text{mismatch}}\), respectively. The required transmission line loss \(S_{21}\) can then be determined by applying the measured 2-port S-parameters data into the
commercial Agilent Advanced Design System (ADS) software. The measurement is carried out with the DUT biased at its operating voltage, as the transmission line loss is a function of its bias voltage.

### 4.1.2 MEASUREMENT SET-UP

Like common microwave transmission line structures, the $S_{21}$ for the photodetector’s transmission line is measured using the network analyser. As mentioned earlier, a back-to-back configuration is used for the 2-port measurement utilising a microwave-probe station, as shown in Fig 4.4a. However, a DC voltage is required to be applied to the DUT, which can be done through the network analyser internal bias-T, as shown in Fig 4.4b. The bias-T prevents the DC voltage from interfering with the network analyser port, while the RF signal flows freely between the network analyser port and the DUT, but blocks the DC power supply.

The microwave-probe station has 2 positioners and a chuck in between. The GSG probes are mounted on the positioners, each having 3 axes of control: a vertical Y-axis, a lateral X-axis and a lateral Z-axis. The DUT is placed on the chuck, which has a yaw control to align the x-z planes of both the DUT and the GSG probes. It is also important to ensure that the probe tip plane aligns to the DUT probe pad plane, so that all the 3 tips of the probe touch down simultaneously and sit flat on the DUT probe pads. Otherwise it will cause an impedance mismatch and will affect the accuracy of the measurement. If the misalignment is severe, the probe tips may be damaged permanently. The probe planarisation is verified by observing the probe tips landing on the contact substrate, ensuring that the 3 tip markings left on the metal of the contact substrate have the same size and depth. The planarisation alignment is simply adjusted via the probe arm planarisation knob. In addition, the microwave-probe station has a microscope, to help with the alignment process.
It is vital to do a calibration to correct the instrument and set-up errors to ensure the accuracy of the measurement within the required bandwidth. The calibration method used is Short-Open-Load-Through (SOLT), which is based on the discrete short, open, load and through standards in the Impedance Standard Substrate (ISS) provided with the GSG probe. The calibration is done using the Cascade Microtech WinCal™ software, which automatically computes the calibration coefficients and verifies the set-up accuracy after measuring all the standards. Once the calibration is completed, the set-up is ready to measure the DUT 2-port S parameters.
4.2 FREQUENCY RESPONSES

The other part of the characterisation is focused on determining the frequency response of the photodetector. For photodetector without gain, such as PIN and MSM photodiodes, the main figure of merit is its 3dB bandwidth, which is the frequency range from DC to where the power drops by 3dB. On the other hand, for photodetector with gain, such as heterojunction phototransistor and avalanche photodiode, the main figure of merit is its cut-off frequency ($f_c$), which is the frequency where the gain drops to unity with the output terminated with a 50Ω load.

4.2.1 MEASUREMENT TECHNIQUES

The measurement techniques commonly used for characterising the frequency response of the photodetector can be broadly classified into two groups, depending on their input optical excitation: pulse or continuous wave (CW).

For the pulse excitation, optical pulses with a width significantly smaller than the inverse of the bandwidth of the DUT, is used to stimulate the DUT. The output pulse width is broadened due to the limited bandwidth of the DUT. The sampling oscilloscope captures a single output pulse and the frequency response is obtained from its Fourier transform. The measured time domain output response allows those undesirable spurious reflections cause by the set-up to be removed using time gating. The set-up is shown in Fig 4.5.

![Figure 4.5: Basic Block Diagram of Pulse Excitation using Sampling Oscilloscope or Spectrum Analyser](image)
On the other hand, a spectrum analyser takes in the periodic output pulses and determines its frequency response directly. However, the measured output is a discrete frequency spectrum, where the frequency interval is determined by the input pulse period. For pulse excitation, it is essential to ensure that the peak power of the input optical pulses do not cause space charge effects leading to a non-linear response of the DUT, which will degrade the accuracy of the measurement. Presently the maximum measurable bandwidth for the above two set-ups are limited by the measurement equipment bandwidth, approximately 50GHz.

For higher frequency measurements, electro-optic sampling is commonly used. It utilises the electro-optic effect, where the optical refractive index in an electro-optic crystal changes when an electric field is applied to it. If an optical beam passes through the crystal, it will experience a phase shift between the two orthogonal polarisation components, which is proportional to the electrical field strength. The set-up is shown in Fig 4.6, where the optical pulses with width significantly shorter than the impulse response of the DUT are used to excite it. At the same time the optical pulses are used to sample the DUT output response after a delay. The average power of the sampling beam is then measured using a slow photoreceiver. By adjusting the delay time, the equivalent impulse response of the DUT is mapped out by the sampling beam. The frequency response of the DUT is obtained by computing its Fourier transform. The measurable bandwidth using this method can attain up to THz, and is limited by the input optical pulse width instead of the measurement equipment bandwidth. However, the set-up is very complicated, requiring several additional systems and precision alignments of the sampling optical beam.

Figure 4.6: Basic Block Diagram of Pulse Excitation using Electro-Optic Sampling
On the other hand, for the CW analysis, a CW optical source, where its amplitude is modulated by a microwave signal directly or externally, is used to excite the photodetector, as shown in Fig 4.7.

Using a frequency synthesiser to generate the modulating microwave signal, the measurement bandwidth is swept across and a spectrum analyser is used to measure the continuous output frequency response of the photodetector. Similarly, a network analyser can be used to replace the frequency synthesiser and spectrum analyser. This set-up has the additional advantage of having the phase response of the photodetector. However, the output of the network analyser is usually not strong enough to drive the optical modulator directly and requires an additional low noise amplifier. The Lightwave component analyser is commercially available in the market that integrates the whole set-up. Nevertheless, the measurable bandwidth of these set-ups is mainly limited by the modulation bandwidth of the CW source or optical modulator, presently at about 40GHz. The accuracy of the measurement is mainly dependent on the stability of the optical source power within the frequency range of interest.

An alternative technique for measuring broader bandwidth is the optical heterodyne detection. CW optical beams from two linearly polarised lasers, with the same polarisation and amplitude, are mixed through a polarisation maintaining coupler and are used to stimulate the DUT, as shown in Fig 4.8. The output from the DUT is an
intermediate frequency signal that is the frequency difference between the two lasers. By sweeping the optical frequency of one of the lasers while maintaining both the lasers output power constant, a continuous frequency response of a very wide bandwidth DUT can be measured. However, this set-up method is as complicated as the electro-optic sampling, requiring several additional measurement systems.

Figure 4.8: Basic Block Diagram of CW Analysis using Optical Heterodyne Detection

### 4.2.2 MEASUREMENT SET-UP

Both the pulse spectrum and CW analyses are suitable for determining the DUT relative frequency response. However, the pulse excitation is not appropriate for output power measurement because the high peak input power of the pulses may saturate the DUT, even though the average power is low. Therefore, the CW excitation is preferred for the characterisation. Considering the available equipment and the expectation of a bandwidth not exceeding 20GHz for the large-area prototype DUTs, the measurement is carried out using the spectrum analyser with an external modulated CW laser. The basic set-up is shown in Fig 4.9.

Figure 4.9: Frequency Response Measurement Set-Up
A 1550nm CW laser modulated by an external optical modulator, with a nominal 3dB bandwidth of 40GHz, is used to stimulate the DUT. The modulating signal is provided by a frequency synthesiser and the modulated optical signal is coupled into the DUT waveguide using a lensed fibre with a nominal spot size of 1.8μm. It is unfavourable for the 2T-HPT to operate in low input optical power condition, as it suffers from poor emitter injection efficiency and has a low gain. In addition, the optical bias of the 2T-HPT must be large so that it gives a higher output current, which reduce the junction capacitances charging time constant and maximise the cut-off frequency. As discussed in Chapter 3.2.1, this can be achieved by injecting an additional optical power to the 2T-HPT to increase the base current, which leads to higher output current. In this set-up, the additional bias optical power is achieved by simply increasing the CW laser power, as the bias of the HPT is dependent on the time average input optical power. The DUT output signal is then measured by a spectrum analyser. The DUT is biased with an external bias-T, connected between the DUT and the spectrum analyser.

Unlike the case of the 2-port S-parameters measurement, the frequency response measurement requires an optical input. Hence an optical-microwave-probe station is used instead, as shown in Fig 4.10. It consists of an input positioner to align the optical lensed fibre and an output positioner to align the GSG probe. Like the microwave-probe station, both positioners have 3 axes of control. However, the optical input positioner has higher resolution, where the vertical Y-axis, the lateral X-axis and the lateral Z-axis actuators have resolutions of 0.02μm, 0.5μm, and 1μm, respectively. The precision actuators are required to facilitate the focusing of the
optical beam at the centre of the optical waveguide, which is 1.5\(\mu\text{m}\) thick and has widths ranging from 1-20\(\mu\text{m}\).

The DUT, which is the cleaved photodetector, is mounted on a chuck placed between the two positioners. Unlike the microwave-probe station, this chuck does not have any yaw control to align the input facet of the DUT perpendicular to the tip of the lensed fibre. This is especially crucial if the input facet of the DUT is not perpendicularly cleaved. As a result, the alignment has to be repeated many times to obtain the best input optical signal coupling, which is verified when the DC output current measured by the ammeter is maximum. However, the input coupled power for DUTs of the same size should be constant, with the assumption that the losses for all the DUTs are the same. Hence, after the alignment, the input optical power level is adjusted until the desired DC output current is obtained.

As the output of the DUT is connected to the spectrum analyser via an RF cable, bias-T and several adaptors, it is important to calibrate out the total loss incurred here. In addition, the output power of the frequency synthesiser must maintain constant within the measurement bandwidth. This power is measured with the spectrum analyser, with and without the output RF section, as shown by the dotted connections in Fig 4.9. The difference between the two results gives the RF loss of the set-up, which is used to calibrate the measured output frequency response of the DUT. The frequency response is obtained by measuring the output power of the DUT using the spectrum analyser, while the frequency synthesiser is swept across the measurement bandwidth.

4.3 SUMMARY

A back-to-back configuration of the photodetector is fabricated and used to measure its 2-port S-parameters to determine its transmission line loss. The device is subsequently cleaved to form the edge illuminated photodetector. Different techniques for the frequency response measurement are analysed. A detailed measurement set-up to determine the bandwidth of the photodetector is also given.
CHAPTER 5

SMALL SIGNAL MODEL & CHARACTERISATION

Heterojunction bipolar phototransistors (HPTs) combine the functions of photodetection and amplification in one device. They have traditionally been pre-eminent in high speed and high power performances. In this chapter, the proposed design of an edge coupled 2-terminal (2T) InGaAs/InP HPT integrated with an optical waveguide is first given. The small signal model for the frequency response of the 2T-HPT is presented next. The small signal model is developed from the standard heterojunction bipolar transistor (HBT) model, which has a similar principle of operation, as described in Chapter 3.1. The main elements of the model are derived from physical device parameters and an exemplar calculation is given.

Next, the 2T-HPT electrical performance is characterised. The IV characteristic, the responsitivity and the frequency responses of the 2T-HPT with various lengths are presented. Together with the measured result, the non-ideal responses are analysed to determine the possible causes. The measured frequency responses are used to optimise the small signal model so that the simulated and measured results matched. The cut-off frequency and the gain of the 2T-HPT is then determined from the simulation optimised result.

5.1 PROPOSED DEVICE STRUCTURE

An edge illuminated 2T-HPT designed with a monolithically integrated leaky optical waveguide is proposed. The transistor gain of the HPT gives a high output response with low noise level while the length of the HPT together with the distribution of the optical signal by the waveguide gives a high saturation current level. However, the increasing junction capacitance limits the cut-off frequency. The 2-terminal device is a direct replacement for the PIN diode commonly employed and uses a very simple
and straightforward fabrication process, which is important for low-cost manufacturing.

The HPT's emitter is wide bandgap InP while the base and collector are narrow bandgap In_{0.53}Ga_{0.47}As. The former pair formed the heterojunction. The base-collector depletion junction is the main photoabsorption volume for the 1.55μm wavelength input optical signal. The optical waveguide core is In_{0.53}Al_{0.2}Ga_{0.27}As, where the percentage of the Al determined the refractive index.

The input optical signal is edge coupled to the optical waveguide core beneath the HPT. The optical signal is then continuously leaks laterally into the base and collector layers above, along the entire length of the device. This helps to reduce the high photocurrent density near the input edge and distribute the current density along the length of the HPT. As a result, a higher saturation current level is achieved.

The electrodes of the emitter and collector are then conveniently designed into a coplanar waveguide (CPW) configuration for direct RF on-wafer probing. In addition, the width and spacing of the electrodes for the emitter and collector can be varied to achieve various output impedance values.

The miniaturisation of the device dimensions reduces the junction capacitance and allows large bandwidth operations. This is achieved either by reducing the active device length (L) or width (S), which defined the transistor junction area. However, in order to facilitate external coupling, a larger device was first fabricated to investigate the performance of the monolithic waveguide HPT. The detailed epitaxial layer design and dimensions of the HPT are shown in Table 5.1 and Fig 5.1, respectively.
### Table 5.1: Epitaxial Layer Design

<table>
<thead>
<tr>
<th>Material</th>
<th>Function</th>
<th>Doping Concentration (cm⁻³)</th>
<th>Thickness</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGaAs</td>
<td>Cap</td>
<td>N=1*10¹⁷</td>
<td>300Å</td>
<td>Xₑ&lt;sub&gt;caps(InGaAs)&lt;/sub&gt;</td>
</tr>
<tr>
<td>InP</td>
<td>Cap</td>
<td>N=1*10¹⁷</td>
<td>100Å</td>
<td>Xₑ&lt;sub&gt;caps(InP)&lt;/sub&gt;</td>
</tr>
<tr>
<td>InP</td>
<td>Emitter</td>
<td>N=8*10¹⁷</td>
<td>800Å</td>
<td>Xₑ</td>
</tr>
<tr>
<td>InGaAs</td>
<td>Spacer</td>
<td>Undoped</td>
<td>50Å</td>
<td>X&lt;sub&gt;spacer&lt;/sub&gt;</td>
</tr>
<tr>
<td>InGaAs</td>
<td>Base</td>
<td>P=1*10¹⁹</td>
<td>450Å</td>
<td>X&lt;sub&gt;B&lt;/sub&gt;</td>
</tr>
<tr>
<td>InGaAs</td>
<td>Collector</td>
<td>N=5*10¹⁶</td>
<td>3000Å</td>
<td>X&lt;sub&gt;C&lt;/sub&gt;</td>
</tr>
<tr>
<td>InP</td>
<td>Sub-Collector</td>
<td>N=1*10¹⁷</td>
<td>2000Å</td>
<td>X&lt;sub&gt;SC&lt;/sub&gt;</td>
</tr>
<tr>
<td>InAlGaAs</td>
<td>Waveguide</td>
<td>Undoped</td>
<td>15000Å</td>
<td>X&lt;sub&gt;wg&lt;/sub&gt;</td>
</tr>
<tr>
<td>nP</td>
<td>Substrate</td>
<td>SI</td>
<td>350µm</td>
<td>X&lt;sub&gt;sub&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

5.2 SMALL SIGNAL MODEL

A basic small signal model for the 2T-HPT is presented, as shown in Fig 5.2. It was developed from the standard T-model for the Heterojunction Bipolar Transistor (HBT) [2], due to their similar operating principle, as discussed in Chapter 3.1. However, the input signal (which is the photocurrent for this case) is modelled by an
AC current source connected in parallel to the base and collector. The main elements of the model are calculated using the physical device parameters. Calculations for a device length \((L_E)\) of 200\(\mu m\) are given. The material properties given are the same for other device dimensions.

### 5.2.1 EMITTER AND COLLECTOR JUNCTION CAPACITANCES

The base-emitter and the base-collector depletion regions can be assumed to reside entirely in the emitter and collector, respectively, since the doping concentration in the base is significantly higher than those of the emitter and collector layers.

The emitter junction capacitance is given by:

\[
C_{BE} = A_E \frac{\varepsilon_E}{X_{dpe} + X_{space}}
\]  

(5.1a)

where the base-emitter heterojunction depletion thickness is given as:

\[
X_{dpe} = \sqrt{\frac{2\varepsilon_E}{qN_E} \left(\Phi_{BE}' - V_{BE}\right)}
\]  

(5.1b)

and the base-emitter junction built-in voltage is approximated as:

\[
\Phi_{BE}' = \left(E_{g(InP)} - \Delta E_{e(InP/InGaAs)}\right)/q
\]  

(5.1c)
The typical values of the energy gap in the InP material ($E_{g(InP)}$) and the valance gap discontinuity between the InP/In$_{0.53}$Ga$_{0.47}$As heterojunction ($\Delta E_{v(InP/InGaAs)}$) are 1.35eV and 0.37eV, respectively [2].

The base-emitter junction is forward biased by the DC component of the optical signal, and the base-emitter voltage ($V_{BE}$) is determined to be 0.6V by the Sheffield University. They used a 3T-HPT, which structure design is similar to the 2T-HPT except it has an additional base contact. It is very important to properly forward bias the base-emitter junction in order to achieve the maximum bandwidth. This is because a large collector current will minimise the emitter charging time ($r_e$). With an emitter doping concentration ($N_E$) of 8x10$^{17}$cm$^{-3}$, a relative dielectric constant ($\varepsilon_{rE}$) of 12.56, an area ($A_E$) of 20µm×200µm, a spacer thickness ($X_{space}$) of 0.005µm, and where $q$ is the electronic charge, the calculated value of $C_{BE}$ is 14.49pF.

Similarly, the collector junction capacitance is given by:

$$C_{BC} = A_C \frac{E_C}{X_{depC}}$$  \hspace{1cm} (5.2a)

where the base-collector homojunction depletion thickness is given as:

$$X_{depC} = \frac{2 \varepsilon_C}{q N_C} \sqrt{\Phi_{BC}^t - V_{BC}}$$  \hspace{1cm} (5.2b)

and the base-collector junction built-in voltage is given as:

$$\Phi_{BC}^t = \frac{E_{g(InGaAs)}}{2q} + \frac{kT}{q} \ln \frac{N_C}{n_i}$$  \hspace{1cm} (5.2c)

The collector to emitter bias voltage ($V_{CE}$) is set at 1.5V (will be discussed in Chapter 6.1). Since this is a 2-terminal device, the base-collector voltage ($V_{BC}$) is the difference between $V_{BE}$ and $V_{CE}$, which is −0.9V. The base-collector junction area ($A_C$) is the same as the emitter area and the energy gap in the In$_{0.53}$Ga$_{0.47}$As material ($E_{g(InGaAs)}$) is 0.75eV [2]. With a collector doping concentration ($N_C$) of 5x10$^{16}$cm$^{-3}$, an intrinsic carrier concentration ($n_i$) of 6.31x10$^{11}$cm$^{-3}$ [2], a relative dielectric constant ($\varepsilon_C$) of 13.88, and where $kT/q$ is the thermal voltage at 300K, the calculated value of $C_{BC}$ is 2.24pF.
5.2.2 EMITTER, COLLECTOR AND BASE RESISTANCES

The emitter and collector resistances consist of the resistance of the epitaxial layers, the metal semiconductor interfaces and the metal contact layers. However, the metal line resistances are disregarded here because it is usually significantly smaller than the other resistances.

\[ R_E = \rho_{E \text{cap}} \frac{X_{E \text{cap}}}{W_E L_E} + \rho_E \frac{X_E - X_{\text{dep}E}}{W_E L_E} + \rho_{\text{eff}} \frac{1}{L_E W_E} \]  \tag{5.3}

The first term in Eqn 5.3 is the emitter cap epitaxial resistances, followed by the emitter epitaxial resistance and the emitter contact resistance. The resistivity can be estimated using the following standard equation [31]:

\[ \rho \approx \frac{1}{q \mu_n N} \]  \tag{5.4}

With a doping concentration \((N_{\text{Ecap}})\) of \(1 \times 10^{19} \text{ cm}^{-3}\), the \(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}\) emitter cap layer has an electron mobility \((\mu_{\text{Ecap}})\) of 2500 cm\(^2\)/V\(\cdot\)s [48] while the \(\text{InP}\) emitter cap layer’s is 1200 cm\(^2\)/V\(\cdot\)s [49]. Hence the estimated emitter cap resistivities \((\rho_{E \text{cap}})\) are 2.50\(\times\)10\(^{-4}\) \(\Omega\).cm and 5.21\(\times\)10\(^{-4}\) \(\Omega\).cm, respectively. Similarly, with a doping concentration \((N_E)\) of 8\(\times\)10\(^{17}\) cm\(^{-3}\), the \(\text{InP}\) emitter layer has an electron mobility \((\mu_{\text{E}})\) of 2100 cm\(^2\)/V\(\cdot\)s [49], giving an estimated emitter layer resistivity \((\rho_E)\) of 3.72\(\times\)10\(^{-3}\) \(\Omega\).cm and a specific emitter contact resistance \((\rho_{\text{E eff}})\) of 7.50\(\times\)10\(^{-7}\) \(\Omega\).cm\(^2\) [50]. The thickness of the \(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}\) emitter cap \((X_{E \text{cap}(\text{InGaAs})})\), \(\text{InP}\) emitter cap \((X_{E \text{cap}(\text{InP})})\) and \(\text{InP}\) emitter layer \((X_E)\) are 0.03\(\mu\)m, 0.01\(\mu\)m and 0.08\(\mu\)m, respectively. With a device width \((W_E)\) of 20\(\mu\)m and length \((L_E)\) of 200\(\mu\)m, \(R_E\) is computed to be 0.02\(\Omega\).

\[ R_C = \rho_C \frac{X_C - X_{\text{dep}C}}{L_E W_E} + \rho_{\text{SC}} \frac{S_{\text{SC}}}{2L_E X_{SC}} + \sqrt{\frac{R_{\text{SHSC}} \rho_{\text{SC}}}{2L_E}} \coth \left( \frac{W_{\text{SC}} \sqrt{R_{\text{SHSC}}}}{\rho_{\text{SC}}} \right) \]  \tag{5.5a}

where \( R_{\text{SHSC}} = \frac{\rho_{\text{SC}}}{X_{\text{SC}}} \) \tag{5.5b}

The first term in Eqn 5.5a is the collector epitaxial resistance, followed by the sub-collector epitaxial resistance and the collector contact resistance. The ground contacts have a width \((W_{\text{SC}})\) of 260\(\mu\)m with its majority laid on the optical waveguide layer.
However, the active contact is on the sub-collector layer and this gives a gap (S_{EC}) of 35\mu m from the emitter contact, as shown in Fig 5.1. The usual spreading resistance is not included as the device structure has uniform base and collector areas.

With a doping concentration \((N_C)\) of 5\times10^{16}\text{cm}^{-3}, the In_{0.53}\text{Ga}_{0.47}\text{As} collector layer has an electron mobility \((\mu_n)\) of 6300cm²/Vs [49], giving an estimated collector layer resistivity \((\rho_C)\) of 1.98\times10^{-2}\Omega.cm. The InP sub-collector resistivity \((\rho_{SC})\) and specific contact resistance \((\rho_{sec})\) is the same as \(\rho_{Eccp(InP)}\), as both have the same material and doping concentration, which is 5.21\times10^{-4}\Omega.cm and 7.50\times10^{-7}\Omega.cm², respectively. The thickness of the collector \((X_C)\) and sub-collector \((X_{SC})\) are 0.3\mu m and 0.2\mu m, respectively. The sub-collector layer sheet resistance \((R_{shsc})\) is calculated to be 26\Omega and for the same active device area, \(R_C\) is computed to be 1.20\Omega.

The base, without any contacts, has only the epitaxial resistance that is given by:

\[
R_B = \frac{\rho_B X_B}{L_B W_B} \tag{5.6}
\]

With a doping concentration \((N_B)\) of 1\times10^{19}\text{cm}^{-3}, the In_{0.53}\text{Ga}_{0.47}\text{As} base layer has an hole mobility \((\mu_{pB})\) of 70cm²/Vs [49], giving an estimated base layer resistivity \((\rho_B)\) of 8.93\times10^{-3}\Omega.cm. For a base thickness \((X_B)\) of 0.045\mu m and the same active device area, the base intrinsic resistance \((R_{bi})\) is computed to be 0.001\Omega.

### 5.2.3 SMALL SIGNAL JUNCTION RESISTANCE

The small signal junction resistance is the inverse of the transconductance \(g_e\) and is given by:

\[
r_e = \frac{kT}{q I_E} \tag{5.7}
\]

The emitter current \((I_E)\) is measured directly with both the optical signal and the collector-emitter basing voltage \((V_{CE})\) applied. For the 2T-HPT, \(I_E\) is the same as the collector current \((I_C)\) and can be expressed as the product of current gain \((\beta)\) and base current \((I_B)\):
\[ I_C = I_B \beta \]  \hspace{1cm} (5.8)

where \( I_B \) is the DC photocurrent \( (I_{ph}) \) generated by the time average input optical power. The computation of \( I_E \) is complicated and inaccurate as the optical bias has too many unknown variables including the coupling efficiency, absorption coefficient, laser output intensity at the lensed fibre output, etc.

For a 20µm x 200µm length HPT, with an optical intensity of ~0.8mW measured at the lensed fibre output and \( V_{CE} \) set to 1.5V, the measured \( I_E \) was 18mA. This gives a \( r_e \) of 1.44Ω.

### 5.2.4 CURRENT TRANSFER RATIO

The current transfer ratio is the product of the current gain and the phase delay caused by the finite space-charge transit time \( (\tau_{SC}) \) taken to travel through the base-collector depletion region and is given by:

\[ \alpha_{\tau_{ac}} = \frac{\gamma_{HPT} \alpha_{ph} e^{-j \omega \tau_{ac}}}{1 + j \frac{f}{f_a}} \]  \hspace{1cm} (5.9)

where \( \tau_{ac} = \frac{X_{depC}}{2 \nu_{SAT}} \)  \hspace{1cm} (5.10)

The carrier is assumed travelling at its saturation velocity \( (\nu_{SAT}) \) and the factor 2 is due to the averaging of the sinusoidal current over the period.

The frequency at which \( \alpha_{\tau_{ac}} \) dropped to 0.707 times its low frequency value is given by:

\[ f_a = \frac{2.43 D_{bulk}}{2 \pi X_{S^2}} \]  \hspace{1cm} (5.11a)

where the factor 2.43 is due to the uniform base doping profile and the base minority-carrier diffusion coefficient is given by:

\[ D_{nB} = \frac{kT}{q \mu_{nB}} \]  \hspace{1cm} (5.11b)
and $kT/q$ is the thermal voltage at 300K. For an In$_{0.53}$Ga$_{0.47}$As base with doping concentration ($N_B$) of $1\times10^{19}$cm$^{-3}$, the base electron mobility ($\mu_{nB}$) is 2500cm$^2$/Vs [48], and $D_{nB}$ is calculated to be 65cm$^2$/s.

The base transport factor is given by:

$$\alpha_{T0} \approx 1 - \frac{X_B}{2L_{nB}^2}$$

(5.12a)

where the factor 2 is due to the uniform base charging profile (DC analysis) and the base minority-carrier diffusion length is given as:

$$L_{nB} = \sqrt{D_{nB} \tau_{Carrier Lifetime}}$$

(5.12b)

and the base minority-carrier lifetime ($\tau_{Carrier Lifetime}$), for the In$_{0.53}$Ga$_{0.47}$As material with a doping concentration ($N_B$) of $1\times10^{19}$cm$^{-3}$, is measured by Cui et al. [48] to be 60ps. $L_{nB}$ is calculated to be 0.62µm.

The emitter injection efficiency ($\gamma_{HPT}$) is not included in the original HBT model. The parameter is added here to account for the loss due to the defect in the vicinity of the heterojunction, as discussed in Chapter 3.1. Characterisation of the physical materials will be required to obtain the various parameters to determine the practical value for the emitter injection efficiency. Hence, the computed ideal value of approximately 1 is used for the initial simulation, where the actual value will be obtained from the simulation optimised result. With a base doping concentration ($N_B$) of $1\times10^{19}$cm$^{-3}$, thickness ($X_B$) of 0.045µm, base-collector depletion width ($X_{depC}$) of 0.219µm [Eqn 5.2], and $\nu_{MT}$ of 5.3x10$^4$cm/s [3], the computed $\tau_{SC}$, $f_a$ and $\alpha_{T0}$ are 2.07ps, 1.24THz and 0.9974, respectively.

5.2.5 PHOTOCURRENT

The AC photocurrent produced by the RF power of the input optical signal, assuming only bandgap transitions is given by [1]:

$$i_{ph} = \frac{\eta q \lambda P_{RF}}{hc}$$

(5.13)
where $q$ is the electronic charge, $h$ is the Planck's constant and $c$ is the velocity of light in free space. The optical wavelength ($\lambda$) is 1.55\,\mu m and the RF power of the input optical signal ($P_{RF}$) is 0.1 mW. The quantum efficiency ($\eta$), from Eqn 2.5 [10]:

$$\eta = (1 - R_{Fresnel})(1 - e^{-\alpha_d l})$$

(5.14a)

where the Fresnel reflection coefficient is given as:

$$R_{Fresnel} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2$$

(5.14b)

and the refractive index of the In$_{0.53}$Al$_{0.14}$Ga$_{0.33}$As waveguide ($n_1$) is 3.441 [44], whereas the refractive index of the input medium (air -- $n_2$) is 1. Assuming the optical signal is input normally with respect to the waveguide, the magnitude of reflection is 0.3021, and gives a Fresnel loss of 0.6979. With a waveguide confinement factor ($\Gamma$) of 0.4 [10], an absorption coefficient ($\alpha_d$) of 0.68\,\mu m$^{-1}$ [7], a photon absorption length ($l$) of only 20\,\mu m, the term $(1 - e^{-\alpha_d l})$ is computed to be 0.9957 and can be approximated as 1. Therefore, $\eta$ is approximately 0.6979, which is an overestimated value as coupling loss is not included. Hence, using Eqn 5.13, the maximum $i_{ph}$ is estimated to 87.14\,\mu A. The value used in the simulation is lower mainly due to poor coupling efficiency.

### 5.3 CHARACTERISATIONS AND ANALYSES

The electrical performance of the fully integrated optical waveguide 2T-HPTs are measured using the CW analysis, as discussed in Chapter 4.2.2. Four 20\,\mu m wide 2T-HPTs with different lengths, 87.5\,\mu m, 200\,\mu m, 400\,\mu m and 600\,\mu m, are measured. The detailed epitaxial layer design and dimensions of the 2T-HPT are as given in Table 5.1 and Fig 5.1, respectively.

The 1.55\,\mu m long wavelength optical source has a maximum output power of 4.5 mW. However, the final maximum output power is reduced to approximately 0.8 mW, when it is connected to the modulator. This is due to the insertion loss of the connection using the patch core and the modulator itself.
5.3.1 IV CHARACTERISTIC

The IV characteristic is used to show the phototransistor DC response under various bias conditions. The measurement set-up is shown in Fig 5.3. The laser used is at 1.55μm wavelength and the input optical power to the DUT is measured from the lensed fibre using a wide area power sensor. The lensed fibre is then aligned to the DUT to achieve maximum coupling efficiency and the output photocurrent is taken from the ammeter.

![Figure 5.3: IV Characteristic Measurement Set-Up](image)

The DC power supply gives the necessary collector-emitter bias voltage \(V_{CE}\), which is limited by the low breakdown voltage of the In_{0.55}Ga_{0.47}As collector. The low breakdown voltage is due to the thermally generated reverse saturation or leakage current at the collector-base junction \(I_{CBO}\) and the high gain of the transistor. It is also believe to be caused by the low bandgap and high intrinsic carrier concentration of the In_{0.55}Ga_{0.47}As material [51]. This breakdown voltage under open-base condition \((I_B=0A)\) is less than 2.5V at room temperature (25°C), and reduces with increasing temperature [51]. With an input signal \((I_B\neq0A)\), the breakdown voltage is generally smaller at higher current densities [52]. Therefore, a maximum bias voltage of 1.5V is used here.

Fig 5.4 shows the measured output emitter current \(I_E\) as a function of collector-emitter bias voltage \(V_{CE}\) at input optical power up to 1.6mW for the 200μm HPT. The IV characteristic curve deviates from the ideal, where the current rises sharply to its maximum value and then remains constant with increasing voltage. The effects of the soft transition in the saturation region and the non-constant output current in the active region, increases as the input optical power is increased. Similar IV characteristic have been observed in other 2T-HPTs as well as 3T-HPTs [41,53].
QUASI-SATURATION

The poor saturation characteristic is due to the quasi-saturation, which is caused by the finite collector resistance ($R_C$). The collector resistance consists of the collector contact resistance, the collector and the sub-collector epitaxial resistances. Although $R_C$ should be small, processing mishaps can cause it to be significant. Under this condition, when the operating current is high, the voltage drop across the collector resistance ($I_C R_C$) reduces the reverse bias voltage ($V_{BC}$) of the base-collector junction, as shown in Fig 5.5.
The reverse bias voltage is then given as:

\[ V_{BC} = V_{BC} + I_C R_C \]  \hspace{1cm} (5.15)

As a result, a larger bias voltage is required to operate the phototransistor into the active region, and the saturation region is extended to a higher value of \( V_{CE} \). The effect is greater at higher collector current because the voltage drop across the collector resistance is larger. If the operating current is continued to increase further, the transistor goes into the saturation mode, where both the base-emitter and the base-collector junction are forward biased. This quasi-saturation effect can be reduced by increasing \( V_{CE} \) or reducing the operating current.

**NON-CONSTANT OUTPUT CURRENT IN ACTIVE REGION**

This characteristic, where \( I_E \) increases as \( V_{CE} \) increases is very similar to the Early effect, which is also known as base narrowing or base-width modulation. Due to the very high doping concentration of the base \( (N_B) \) with respect to the collector \( (N_C) \), it is assumed that the base-collector depletion region resides completely in the collector layer with negligible depletion width in the base layer. It is also assumed that when the base-collector voltage \( (V_{BC}) \) increases, the increase of the depletion layer width only extends in the collector layer and not in the base layer. However, when \( V_{BC} \) increases above a certain value, the depletion region in the base region becomes significant. As a result, the neutral base width (that is the un-depleted base width, \( X_{Neutral(Base)} \)) is reduced and the current gain is increased.

The base-collector depletion layer width is given by \([29]\):

\[
X_{bc(deep)} = \left[ \frac{2 \varepsilon_{InGaAs} (\Phi_{BC} - V_{BC})}{q} \left( \frac{N_B + N_C}{N_B \cdot N_C} \right) \right]^{1/2} \]  \hspace{1cm} (5.16a)

\[
\Phi_{BC} = \frac{kT}{q} \ln \frac{N_B N_C}{n_i^2} \]  \hspace{1cm} (5.16b)

where \( q \) is the electronic charge and \( kT/q \) is the thermal voltage at 300K. The relative dielectric constant of InGaAs \( (\varepsilon_{InGaAs}) \) is 13.88 \([5]\) and the intrinsic carrier concentration \( (n_i) \) is \( 6.31 \times 10^{11} \text{cm}^{-3} \) \([2]\). With \( N_B \) of \( 1 \times 10^{19} \text{cm}^{-3} \), \( N_C \) of \( 5 \times 10^{16} \text{cm}^{-3} \) and
$V_{BC}$ of $-0.9\text{V}$, the base-collector junction built-in voltage ($\Phi_{BC}$) is $0.722\text{V}$ and the depletion layer width is $210\text{nm}$. The penetration of the depletion layer into the base and collector layers are given by [29]:

$$X_{b\text{-(dep)}} = \left(\frac{N_C}{N_B + N_C}\right)X_{bc\text{-(dep)}}$$  \hspace{1cm} (5.17a)$$

$$X_{c\text{-(dep)}} = \left(\frac{N_B}{N_B + N_C}\right)X_{bc\text{-(dep)}}$$  \hspace{1cm} (5.17b)$$

Figure 5.6: Minority-Carrier Distribution in the Base (Changes as the Neutral Base Thickness Reduces)

When $V_{BC}$ increases, the minority-carrier concentration distribution in the base is forced to become steeper as $X_{\text{NeutralBase}}$ becomes thinner, as shown in Fig 5.6. Consequently, the electron diffusion current density, which is proportional to the gradient of the doping concentration in the $x$ direction ($n_B(x)$) increases. From Eqn 3.1a [29]:

$$J_{\text{diffusion}}(x) = qD_{nB} \frac{dn_B(x)}{dx}$$  \hspace{1cm} (5.18)$$

where $q$ is the electronic charge and the base minority-carrier diffusion coefficient ($D_{nB}$) is $65\text{cm}^2/\text{s}$ [Eqn 5.11b]. As a result, $I_C$, and therefore $I_E$, increase rather than remain constant as $V_{BC}$ increases.
The Early effect appeared often in transistors where the base thickness is small and the doping concentration is low. However, HPTs do not usually suffer from the Early effect because the base doping is significantly higher than the collector doping. They are also not restricted to be lower than the emitter, as in homojunction BJT, in order to ensure a higher emitter-injection efficiency.

Table 5.2 shows the variations of $X_{bc(dop)}$ for these devices as $V_{BC}$ is varied. With a base layer thickness ($X_B$) of 0.045$\mu$m, the changes of $X_{NeutralBase}$ with respect to the applied voltage $V_{BC}$ are very small.

<table>
<thead>
<tr>
<th>$V_{BC}$ (V)</th>
<th>$X_{bc(dop)}$ (nm)</th>
<th>$X_{id(dop)}$ (nm)</th>
<th>$X_{5(dop)}$ (nm)</th>
<th>$X_{NeutralBase}$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.7</td>
<td>209.51</td>
<td>208.47</td>
<td>1.04</td>
<td>43.96</td>
</tr>
<tr>
<td>-0.9</td>
<td>223.76</td>
<td>222.65</td>
<td>1.11</td>
<td>43.89</td>
</tr>
<tr>
<td>-1.1</td>
<td>237.16</td>
<td>235.98</td>
<td>1.18</td>
<td>43.82</td>
</tr>
</tbody>
</table>

Table 5.2: Depletion Layer Thicknesses for values of $V_{BC}$

With the above calculated $X_{NeutralBase}$ variations and using Eqn 5.18, the percentage increase of current density caused by the increase of $V_{BC}$ from -0.7V to -1.1V, is less than 0.4%. This implies that $I_C$ should remain approximately constant as $V_{BC}$ increases within the values. In addition, when the IV curves are extrapolated back along the voltage axis they do not converge to a point, which is used to determine the Early voltage. Consequently it can be concluded that the IV curves are not the classical curves of the Early effect. Therefore, this increase of $I_C$ is not caused by the Early effect.

The non-constant output collector current when $V_{CE}$ increases in the active region, is probably caused by the large reverse saturation or leakage current ($I_{CBO}$) in the In$_{0.53}$Ga$_{0.47}$As collector region. The low bandgap energy (0.75eV) and high intrinsic carrier concentration of In$_{0.53}$Ga$_{0.47}$As causes a high $I_{CBO}$, which increases with the bias voltage. In addition, its poor thermal conductivity (0.05W/cm.K at 300K) leads to high self-heating effect and causes a higher $I_{CBO}$ [52]. This additional base-collector component, $I_{CBO}$, travels to the neutral base and gets amplified by the transistor high gain. As a result, the output current increases with $V_{CE}$ instead of maintaining at a
constant level when the optical power is constant. Hence, it gives a high output conductance in the active region, which increases with the input optical power.

### 5.3.2 Responsivity

The responsivity of the phototransistor is the ratio of the output photocurrent to the input optical power. It is determined using the same set-up as the IV characteristic measurement in the previous section. The emitter current ($I_E$) as a function of input optical power for a 200μm HPT, with a collector-emitter bias voltage ($V_{CE}$) of 1.5V, is shown in Fig 5.7. With a Fresnel loss of 0.6979 (given in Eqn 5.14), the responsivity of the phototransistor is approximately 31.52A/W. This approximated value is an underestimated figure as the input coupling loss is not included. It also shows that the output DC current ($I_E$) increases linearly with the input optical power, with a maximum DC current of more than 30mA. This demonstrates that the distributed HPT is able to operate at very high input optical powers.

![Figure 5.7: Measured Responsivity](image-url)
5.3.3 FREQUENCY RESPONSE

One of the important characteristics of the HPT is its cut-off frequency ($f_c$), which is the frequency where the gain drops to unity with a 50Ω load. However, the input of the HPT is an optical signal and cannot be compared directly with the electrical output to determine the cut-off frequency. The general standard is to use the input photocurrent of the HPT as a reference for the unity level [41,42]. A reference PIN photodiode is required to measure the input photocurrent to the 2T-HPT.

The reference PIN photodiode must have the same physical characteristics as the HPT, such as input optical aperture, optical absorption volume and efficiency. The design of this reference PIN photodiode is identical to the HPT except that it has no emitter layer so that there is no transistor amplification. The reference PIN photodiode is hereinafter referred to as the HPT-PIN. Ideally, for the same active device length, the output photocurrent of the HPT-PIN within its 3dB bandwidth is equal to the HPT photocurrent without amplification. Therefore, the HPT cut-off frequency is the frequency where the output response of the HPT is equal to the HPT-PIN response within its 3dB bandwidth.

Apart from having the same optical-to-electrical conversion physical characteristics between the HPT and its HPT-PIN, it is also crucial to maintain the same input coupling efficiency to the devices. This is to ensure that they have the same input optical power. Therefore, precision alignment of the input lensed fibre is required. The frequency responses for the 2T-HPTs and their respective HPT-PIN photodiodes of the same length are measured and their characteristics are compared. The measured frequency responses are also compared with those simulated using the small signal model described in Section 5.2.

HPT-PIN

The frequency responses of the HPT-PINS for device lengths of 200μm, 400μm and 600μm are first characterised and analysed. They are biased at -0.9V so that the
electric field across the base-collector junction ($V_{BC}$) of the HPT-PIN and the HPT are similar. The frequency responses are measured from 100MHz to 20GHz with an input optical power of ~0.8mW. The minimum frequency is limited by the optical modulator while the maximum frequency is limited by the RF signal generator. In addition, a top illuminated commercial PIN photodiode (PDSC12T by OPTO SPEED), which has a relatively flat response up to 20GHz, is measured to determine the accuracy of the measurement set-up, as shown in Fig 5.8.

![Figure 5.8: Frequency Responses of the Top illuminated Commercial PIN photodiode (PDSC12T by OPTO SPEED)](image)

The measured result of the commercial PIN photodiode, as compared with its given response, shows that there is a maximum measurement error of ±1.2dB. This fluctuation error is not caused by the equipment used as it is not present during the calibration for the measurement set-up, as given in Chapter 4.2.2. This error was frequency independent, as shown in Fig 5.8, and was thought to be due to the instability of the input optical coupling during the measurement process. As the measurement is carried out on a normal work bench, it is very sensitive to vibrations, which affect the input lensed fibre alignment during the measurement. This fluctuation error was minimised by repeating the measurement several times and the taking the average.
The measured frequency responses of the HPT-PINs are plotted in Fig 5.9 and the response of the commercial PIN is included for comparison. The HPT-PINs have a relatively flat response over its 3dB bandwidth ($f_{3db}$). The measured $f_{3db}$ are 800MHz, 600MHz, and 400MHz for the 200μm, 400μm, and 600μm devices, respectively. With a fixed active device width of 20μm, the respective active device areas are 4nm$^2$, 8nm$^2$ and 12nm$^2$. The $f_{3db}$ is inversely proportional to the active device area due to the associated junction capacitance. On the other hand, the commercial PIN photodiode has an $f_{3db}$ of more than 20GHz. It has a diameter of 12μm and a significantly smaller active device area of 0.11nm$^2$. Hence, it has the smallest associated junction capacitance and the largest $f_{3db}$.

At low frequency, the transmission line loss due to the length of the device can be ignored since it is very short in comparison to the wavelength. Hence, from the principle that the photocurrent generated is proportional to the absorption layer length; the longer devices should have a higher output electrical power than the shorter devices at 100MHz. However, if all the photons have been absorbed within the shortest device length (i.e. 200μm for this case), the photocurrent generated
should then be constant at 100MHz. Therefore, the inconsistency in the response at 100MHz in Fig 5.9 indicates that there is a variation in the input coupling efficiency. This is caused by misalignment and poor cleaving at the input facet. As a result, different input optical powers are coupled into the devices. Hence, the device characteristics for various lengths cannot be compared directly here. However, every effort was made to achieve the maximum input coupling for all the devices by optimising the alignment of the optical fibre and the device such that maximum output electrical power is obtained at the initial 100MHz frequency.

Assuming all the devices have the same input coupling efficiency from the lensed fibre with a spot size of only 1.8µm, the low output power of the vertically illuminated commercial PIN photodiode indicates that it does not have sufficient length to absorb all the input optical power. Hence, it has a wide bandwidth but a poor efficiency, which is the classical bandwidth-efficiency trade-off for the vertical illuminated design.

\[ HPT \]

![Figure 5.10: Measured and Simulated Results of the 87.5µm HPT (a) with ideal parameters (b) with optimised parameters](image)
The frequency response of an 87.5µm long HPT is measured and compared with the simulated result, as shown in Fig 5.10. The HPT is biased at 1.5V with an input optical power of ~0.8mW and measured from 100MHz to 20GHz. The amplitude of the input photocurrent \( (I_{ph}) \) changes the output response linearly throughout the whole frequency range. The loss of the metal contact, which formed the transmission line, is assumed to be negligible at low frequency. Therefore, the measured result should have the same output response with the simulated result at 100MHz.

The input photocurrent of 87.139µA calculated in Section 5.2.5, is an overestimated value as coupling loss is not included. The input optical coupling loss due to the optical misalignment, especially the angular displacement in the measurement is unaccounted. In addition, the input facet of the device is poorly cleaved and is not polished, which further increases the coupling loss. Furthermore, the edge illuminated optical waveguide of the HPT is only 1.5µm thick, which is less than the spot size of the lensed fibre, incurring additional coupling loss. All these factors lead to a lower input photocurrent than the computed value. The input photocurrent in the simulation is determined by timing the simulated result to coincide with the measured result at 100MHz. With an input photocurrent \( (I_{ph}) \) of 0.63µA, the amplitude of both results is matched at 100MHz, as shown in Fig 5.10(a).

At high frequency, the measured result, which includes parasitic effects and microwave loss, should generally be lower or close to the simulations with ideal conditions. However, the measured result is higher than the simulated result (a) in Fig 5.10. Therefore, the input photocurrent must be higher than the value approximated in (a), in order to have a higher simulated output response. This increases the output response at 100MHz as well, and the common-emitter current gain \( (\beta) \) is reduced to match the simulated and the measured responses. The gain of the HPT from Eqn 3.7:

\[
\beta = \frac{\alpha_b}{1 - \alpha_b} \quad (5.19a)
\]

\[
\alpha_b = \gamma_{HPT} \cdot \alpha_{T0} \quad (5.19b)
\]

where the common-base current gain \( (\alpha_b) \) is a function of the emitter injection efficiency \( (\gamma_{HPT}) \) and base transport factor \( (\alpha_{T0}) \). The common-base current gain of 0.9974 calculated in Section 5.2.4 assumed that the emitter injection efficiency to be
unity. However, for practical heterojunction, it is always less than unity due to defect and fabrication problems. Hence the actual common-base current gain is optimised simultaneously with the input photocurrent until the simulated output curve matches the measured curve, as shown in Fig 5.10(b), with a common-base current gain of 0.953 and a photocurrent of 0.954μA.

Similarly, from the above analysis, the simulated frequency responses of the 200μm, 400μm, and 600μm long HPTs are optimised and compared with the measured result. The simulated result shows that, as the device length increases, it becomes inaccurate, especially at high frequency. As shown in Fig 5.11, the simulated result of the 200μm long HPT matches the measured result, while the simulated result of both the 400μm and 600μm long HPTs deviates at higher frequency beyond 2GHz, as shown in Figs 5.12 and 5.13, respectively.

![Figure 5.11: Measured and Simulated Results of the 200μm HPT](image)

*Figure 5.11: Measured and Simulated Results of the 200μm HPT (I_{ph}=1.714μA, α_B =0.937)*
CHAPTER 5 SMALL SIGNAL MODEL & MEASUREMENTS

Figure 5.12: Measured and Simulated Results of the 400\mu m HPT
\( i_{\mu}=3.550\mu A, \quad \alpha_B=0.807 \)

Figure 5.13: Measured and Simulated Results of the 600\mu m HPT
\( i_{\mu}=7.170\mu A, \quad \alpha_B=0.643 \)
For the HPTs with the same physical design but longer length, theoretically the gain should remain constant, while the input photocurrent increases due to the longer absorption length. On the contrary, the simulation optimised result of the longer HPT has a larger photocurrent, i.e. it has a lower gain. The common-base current gain ($\alpha_b$) obtained from the simulation optimised result for the 87.5$\mu$m, 200$\mu$m, 400$\mu$m, and 600$\mu$m are 0.953, 0.937, 0.807 and 0.643, respectively. The result shows that as the length of the HPT increases, the common-base current gain reduces. This is due to insufficient input optical power to bias the longer 2T-HPTs and causes them to operate in the low gain region. When biasing is low, the recombination current in the heterojunction, quasi-neutral base bulk and at the surface may be large compared with the useful diffusion current of the minority-carrier across the base. As a result, the emitter injection efficiency and base transport factor are lower, hence a smaller common-base current gain.

The input photocurrent ($i_{ph}$) obtained from the optimised simulated result for the various device length of 87.5$\mu$m, 200$\mu$m, 400$\mu$m, and 600$\mu$m are 0.954$\mu$A, 1.714$\mu$A, 3.550$\mu$A, and 7.170$\mu$A, respectively. The result shows that the input photocurrent increases with the device length. This gives a difference with the BeamPROP analysis on the waveguide in Chapter 3.3.2. The simulated result from Fig 3.13 shows that 95% of the optical power is coupled from the waveguide within the first 400$\mu$m of the device; hence, the photocurrent should not increase significantly beyond 400$\mu$m. The likely cause of the continuing increment of the photocurrent is due to the increment of the dark current, which increases with the length of the device. In addition, the inconsistent input coupling efficiency, which gives different input optical power to different devices as mentioned earlier, can also contribute to the error.

The simulation optimised result matches the measured result closely at frequency below 2GHz but have a higher response at higher frequency. The difference increases for the longer 400$\mu$m and 600$\mu$m devices. This is due to the loss of the metal contact, which formed the transmission line, increases as the length and frequency increase. Also inserted in Figs 5.11 – 5.13 are the frequency responses of their respective HPT-PINs, which are used to determine the cut-off frequency of the HPT. The cut-off...
frequency of the HPT is the frequency where the response is equal to the HPT-PIN response (within its 3dB bandwidth), as shown in Figs 5.11 – 5.13. The measured output response at 100MHz of the various HPT and respective HPT-PIN are tabulated in Table 5.3.

<table>
<thead>
<tr>
<th>Device Length</th>
<th>200µm</th>
<th>400µm</th>
<th>600µm</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPT Output Power (dBm)</td>
<td>-2.27</td>
<td>-3.20</td>
<td>-1.65</td>
</tr>
<tr>
<td>&quot;HPT&quot; PIN Input Power (dBm)</td>
<td>-24.50</td>
<td>-23.60</td>
<td>-28.40</td>
</tr>
<tr>
<td>Gain (dB)</td>
<td>22.23</td>
<td>20.40</td>
<td>26.75</td>
</tr>
<tr>
<td>Cut-off Frequency (GHz)</td>
<td>1.7</td>
<td>1.7</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Table 5.3: Measured Output Power, Gain @100MHz and Cut-off Frequency

The result shows that the longest 2T-HPT of 600µm has the highest cut-off frequency of 3.3GHz. This contradicts the basic principle that the cut-off frequency reduces as the device capacitance increases with its size. In addition, the gains of the 2T-HPT vary inconsistently when the length increases. These errors arise due to the inconsistency in the input optical power coupled to the various photodetector during the measurement, as analysed earlier for the HPT-PIN frequency response measurements. If the amount of input optical power coupled to the HPT-PIN is lower than that of the HPT, the cut-off frequency of the HPT will be higher. On the other hand, if the amount of input optical power coupled to the HPT-PIN is higher, the cut-off frequency of the HPT will be lower. As a result, although the measurement has an accuracy of ±1.2dB, the cut-off frequency and the gain of the 2T-HPT cannot be determined accurately.

On the other hand, the input photocurrent to the 2T-HPT can be obtained readily in the simulation, without the need of its respective HPT-PIN. The cut-off frequency and the gain of the 2T-HPT are obtained from the simulation optimised result. The simulated result tabulated in Table 5.4, shows that the cut-off frequency reduces as the device length increases. In addition, the gain at 100MHz also reduces as the device length increases. This is because the bandwidth limiting junction capacitance increases with the device length. The junction capacitance reduces the output response amplitude as the frequency increases.
### 5.4 SUMMARY

The measured result of the edge coupled 2T-HPT integrated with an optical waveguide is presented. The measured DC characterisation for the 2T-HPT with an active device area of $20 \mu m \times 200 \mu m$ shows that the output emitter current increases linearly up to more than 30mA, which is limited by the laser output power. The non-ideal shape of the IV characteristics is analysed. It was found that the continuous increase of the emitter current as $V_{CE}$ is increased, is probably due to the high leakage current. The high intrinsic carrier concentration and the low bandgap of the In$_{0.53}$Ga$_{0.47}$As collector region cause leakage current, which get amplify by the transistor gain.

A 2T-HPT small signal model (SSM) is developed from its equivalent HBT model. The input photocurrent is modelled as a constant current source, which connects the base and collector in parallel. With the measured frequency responses for the 87.5$\mu m$, 200$\mu m$, 400$\mu m$ and 600$\mu m$ long 2T-HPTs, the SSM is used to analyse and determine the causes of the non-ideal characteristic. This is achieved by optimising the model parameters, the input photocurrent and the common-base current gain, so that the simulated response matches the measured response. The simulated input photocurrent is reduced to account for the lower input optical coupling efficiency. The common-base current gain is reduced to account for the lower emitter injection efficiency and base transport factor due to fabrication mishaps, which increases the rate of carrier recombination.

The close match of the measured and optimised simulation result show that the SSM can accurately give the frequency response of the 2T-HPT. The input photocurrent

**Table 5.4: Simulated Gains @100MHz and Cut-off Frequency**

<table>
<thead>
<tr>
<th>Device Length</th>
<th>87.5$\mu m$</th>
<th>200$\mu m$</th>
<th>400$\mu m$</th>
<th>600$\mu m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain at 100MHz (dB)</td>
<td>24.17</td>
<td>19.30</td>
<td>11.80</td>
<td>7.40</td>
</tr>
<tr>
<td>Cut-off Frequency (GHz)</td>
<td>2.5</td>
<td>1.2</td>
<td>0.58</td>
<td>0.39</td>
</tr>
</tbody>
</table>
can also be obtained from the model directly to determine the device gain and cut-off frequency. This could not be obtained through measurements due to the inconsistency input optical coupling to the devices. The simulation optimised result shows that the output power increases with the device length and has a maximum gain of 24dB. The largest cut-off frequency of the 2T-HPT is 2.5GHz, which is limited by the large capacitance associated with the large active device area of 20μm×87.5μm.
CHAPTER 6

PERIODIC MODEL & IMPROVED DEVICES

High bandwidth and high saturation power are two main requirements for modern photodetectors. The high bandwidth increases the speed of the photodetector, which leads to a higher bit rate for the optical communication system whilst the high saturation power increases the photodetector input optical power dynamic range and gives a large linear output current range. More importantly, the higher output current of the photodetector can drive subsequent RF detection circuitry directly without going through an RF amplifier.

In this chapter, a 2T-HPT with shorter device length was measured, and analysed using the small signal model. The reduced length is mainly to lower the high junction capacitance and transmission line loss, which are associated with the active device length. This helps to increase the cut-off frequency of the 2T-HPT. The frequency response of the 2T-HPT is also analysed as a function of the device length, device width, bias current and collector depletion thickness. Using scaling of the model, smaller discrete devices are designed.

A periodic travelling wave photodetector (P-TWPD) model is developed to analyse the effect of phase matching and reverse input termination conditions. The P-TWPD consists of multiple photodetectors connected serially through an optical waveguide and a transmission line so that their outputs add coherently. This gives the advantages of high cut-off frequency and high saturation power simultaneously. The 2T-HPT small signal model is then incorporated into the periodic travelling wave structure for further analysis and design.
6.1 SMALLER DEVICES

In Chapter 5, the analysis of the large device area 2T-HPT using the small signal model shows its cut-off frequency is limited by the large junction capacitance, which increases with the active device area, as tabulated in Table 6.1. In addition, the loss of the transmission line, which increases with the length and operating frequency, also restricts the cut-off frequency. Hence, a shorter 2T-HPT, which has a smaller junction capacitance and lower transmission line loss, is preferred for high frequency operations.

![Table 6.1: Simulated Gains @100MHz and Cut-off Frequency for a 20μm Wide HPT with various Device Lengths](image)

The basic design of this smaller 20μm x 20μm HPT is similar to the longer HPTs presented in Chapter 5, except that the base doping concentration is 2 x 10^{19} cm^{-3} instead of 1 x 10^{19} cm^{-3}. This reduces the carrier lifetime from 60ps to 16.8ps [48] and lowers the base transport factor from 0.9974 to 0.9894, calculated using Eqn 5.12. The detailed epitaxial layer design and dimensions of the 2T-HPT are given in Table 5.1 and Fig 5.1, respectively.

6.1.1 SIMULATION AND MEASUREMENT RESULTS

The measured and simulated results for the 20μm long HPT are shown in Fig 6.1. The simulated result is optimised similarly as in Chapter 5.3.3 to match it to the measured result. The shorter device length gives a shorter absorption length (L), therefore a smaller photocurrent (i_ph) of 0.134μA as compared with 0.954μA obtained from the 87.5μm long HPT. The small variation of the common-base gain (α₂) in the simulation, from the ideal 0.9894 to 0.9854, indicates that the input optical power is sufficient to bias the short device in the ideal region.
The simulation optimised result matches the measured result closely up to 20GHz. The better accuracy at the high frequency range, as compared to the longer devices simulated in Chapter 5.3.3, is because the transmission line loss, which is not included in the simulation, is smaller for the shorter 20μm device. This good agreement demonstrates the accuracy of the small signal model for short devices. The gain at 100MHz and cut-off frequency determined from the simulation are 35.21dB and 8.3GHz, respectively.

![Graph showing measured and simulated results](image)

**Figure 6.1: Measured and Simulated Results of the 20μm×20μm 2T-HPT**

\[ i_{ph} = 0.134 \mu A, \ \alpha_B = 0.9854 \]

Using the small signal model, the 2T-HPT is simulated and analysed with various dimensions and bias conditions. The simulation assumes that the transmission line has negligible loss and the input optical coupling efficiency is constant for all devices. The bias voltage for all the simulation is set at 1.5Vdc.

**Length and Width**

The 20μm×20μm 2T-HPT is simulated with various active device lengths (L) and widths (S), which defined the transistor junction area, as shown in Figs 6.2 and 6.3,
respectively. The unity gain cut-off frequency for both cases increases as the device length and width reduce.

![Figure 6.2: Simulated Unity Gain Cut-off Frequency as a function of the Active Device Width (S) for a 20µm long 2T-HPT](image)

![Figure 6.3: Simulated Unity Gain Cut-off Frequency as a function of the Active Device Length (L) for a 20µm wide 2T-HPT](image)
Bias Current

The effect of the emitter bias current, $I_E$, on the unity gain cut-off frequency is also simulated for the 20μm×20μm 2T-HPT, as shown in Fig 6.4. It shows that the cut-off frequency increases by 2GHz (from 6.7GHz to 8.7GHz), when $I_E$ increases from 5mA to 20mA. This increment is due to the reduction of the emitter charging time, which is an inverse function of the $I_E$. From Eqn 3.12:

$$\tau_e = \frac{kT}{qI_E} \left( C_{BE} + C_{BC} \right)$$

(6.1)

However, the improvement becomes insignificant as $I_E$ is increased further. It increases only by 0.4GHz (from 8.7GHz to 9.1GHz) when $I_E$ is increased again from 20mA to 35mA.

Figure 6.4: Simulated Unity Gain Cut-off Frequency as a function of the Emitter Bias Current for a 20μm×20μm 2T-HPT
Collector Depletion Thickness

The cut-off frequency is limited by the base transit time ($\tau_b$), the space-charge transit time ($\tau_{sc}$), the collector charging time ($\tau_c$), in addition to the emitter charging time ($\tau_e$), as explained in Chapter 3.1.2. For lightwave applications, the cut-off frequency is mainly dominated by $\tau_e$ and $\tau_c$ due to the high capacitances associated with the large device area required for effective input coupling and photoabsorption. Therefore, the saturation of the improvement is because the cut-off frequency is now limited by $\tau_c$ instead. From Eqn 3.11:

$$\tau_c = C_{sc}(R_e + R_C)$$  \hspace{1cm} (6.2a)

where

$$C_{sc} = \frac{A_e E_C}{X_{depC}}$$  \hspace{1cm} (6.2b)

However, the collector space-charge region transit time must be taken into consideration as well. This is because it is proportional to the depletion thickness, which contradicts the requirement of the collector charging time, as given in Eqn 3.10:

$$\tau_{sc} = \frac{X_{depC}}{2\nu_{sat}}$$  \hspace{1cm} (6.3)

As a result, there is an optimum depletion thickness where the sum of the collector charging time and the collector space-charge region transit time is minimised, as given below:

$$\tau_e + \tau_{sc} = \frac{A_e E_C}{X_{depC}} (R_e + R_C) + \frac{X_{depC}}{2\nu_{sat}}$$  \hspace{1cm} (6.4)

The above overall time constant is plotted as a function of the depletion thickness, as shown in Fig 6.5. For the given sum of $R_e$ and $R_C$ (12.179Ω), the result shows that the optimum depletion thickness is 25nm for the minimum delay time of ($\tau_e + \tau_{sc}$). This small depletion thickness restricted the collector thickness to be thin as any undepleted region will reduce the carrier velocity. As a result, the breakdown voltage, which is a function of collector thickness [2], will be reduced as well, and the transistor ceases to be useful. Therefore, a sufficiently thick depletion layer is required for a reasonable breakdown voltage, which is a trade-off with the cut-off frequency.
In the collector layer, any undepleted region \((X_c - X_{depC})\) will incur additional collector epitaxial resistance, as given by [2]:

\[
R_{C(epitaxial)} = \rho_C \frac{X_c - X_{depC}}{L_{E}W_{E}}
\]  

(6.5)

For a well designed transistor, the collector thickness \((X_c)\) is designed to be equal to the depletion thickness \((X_{depC})\). This can be achieved by either reducing the collector doping concentration so that the collector is fully depleted at a reasonable reverse base-collector bias voltage or changing the collector thickness. However, for the design of the 2T-HPT here, the collector resistance is dominated by the sub-collector epitaxial and contact resistances instead.

Concurrently, the base layer is designed to be very thin in order to achieve a high current gain. This leads to a very small base transit time, given by Eqn 3.9:

\[
\tau_b = \frac{X_B^2}{2.43D_{nB}}
\]  

(6.6)
Consequently, the base transit time is significantly smaller than the space-charge transit time, the collector charging time and the emitter charging time. As a result, the cut-off frequency is normally not limited by the base transit time.

**Devices with Narrower Width**

The unity gain cut-off frequency can be optimised, as discussed above. However, to increase the cut-off frequency further, it is necessary to reduce the active device area of the 2T-HPT to reduce the junction capacitances. The frequency responses of a smaller 2T-HPT with a constant length of 10μm but different active device width of 1μm, 2μm and 5μm are simulated, as shown in Fig 6.6. Their respective cut-off frequencies are 59.4GHz, 49.2GHz and 32.8GHz, with a gain of 39dB.

![Simulated Frequency Responses of small active device area 2T-HPTs](image)

**Figure 6.6: Simulated Frequency Responses of small active device area 2T-HPTs**

The simulated 3dB bandwidths of the 2T-HPT with 1μm, 2μm and 5μm active device widths are 710MHz, 618MHz and 463MHz, respectively. It decreases as the width increases due to the higher capacitance. In order to achieve a wider 3dB bandwidth,
the input and output capacitances of the 2T-HPT must be matched over a wide bandwidth. However, the 2T-HPT has no electrical input base contact and there is no electrical path for the excess charges in the base to discharge. Neither can the charges be discharged via the emitter because of the high level of the heterojunction energy band-gap nor the collector because of the polarity of the electric field in the depletion region. This electric field is sustained by the DC photocurrent and the external bias voltage. Since charge cannot be destroyed, the only way for the holes to be extracted is through various carrier recombination processes. The time taken for these processes is determined by the carrier lifetime ($\tau_{\text{carrier lifetime}}$), and it is a function of the doping concentration, dopant type, fabrication process, and the material properties, independent of the device geometry and DC bias [35]. The effect of this long discharging process in the base can be observed in the impulse response of the device, which has a long exponential decaying response [5]. Therefore, the 3dB bandwidth of the 2T-HPT is limited by this base discharge time constant.

6.2 PERIODIC TRAVELLING WAVE DEVICES

The above analysis shows that the cut-off frequency can be increased by reducing the active device area. However, this reduces the saturation current level, as the current density increases when the active device area reduces. To overcome this trade-off, the concept of the microwave distributed or travelling wave amplifier is employed on the photodetector design. Two types of travelling wave designs are the fully distributed and the periodic distributed devices, as discussed in Chapter 2.1.3.

For the fully distributed travelling wave photodetector (TWPD), the capacitance of the photodetector is distributed along the length of the electrical transmission line formed by the electrodes. The long transmission line, built on the highly conductive semiconductor layers, suffered very high attenuation especially at high frequency. This is demonstrated in the measurements of the CPW transmission line formed by the back-to-back 2T-HPT in Chapter 3.4.5, and the frequency responses of those long 2T-HPTs with 400µm and 600µm device length.
On the other hand, the periodic distributed design enables the transmission line to be built on the SI substrate, which has lower loss. Hence, the periodic distributed design is preferred here and the travelling wave structure is known as the periodic travelling wave photodetector (P-TWPD). The principle of this travelling wave concept is to connect a number of photodetectors serially such that their output currents are added constructively. The distributing optical waveguide feeds the input optical signal to the individual photodetector while the electrical transmission line summed the output currents of the photodetectors coherently. This is achieved by phase matching, where the length of the electrical transmission line and the optical waveguide are designed such that the optical and electrical signals take approximately the same time to travel through each section. In addition, the output capacitance of the photodetector has to be integrated into an artificial transmission line (ATL) structure, where its characteristic impedance is given by:

\[
Z_{\text{ATL}} = \sqrt{\frac{L}{C}} \quad (6.7)
\]

where \(L\) and \(C\) are the inductance and capacitance of per section of transmission line, respectively. The characteristic impedance of the ATL has to be matched to the load impedance, which is usually 50\(\Omega\).

The output power of the P-TWPD is the sum of all the photodetectors while the 3dB bandwidth is determined by the individual photodetector. However, half of the power is loss due to the backward travelling waves, which are absorbed by the reverse input termination. If the reverse input is not terminated, i.e. it is open circuit, the backward travelling wave will be reflected back to the load and interferes with the desired forward travelling wave. This causes a reduction in the 3dB bandwidth. For an ideal P-TWPD, the maximum 3dB bandwidth is limited by the artificial transmission line cut-off frequency. From Eqn 2.11:

\[
f_{\text{ATL}} = \frac{1}{\pi \sqrt{LC}} \quad (6.8)
\]

As seen from Eqn 6.7, for a given characteristic impedance of the ATL and a given photodetector capacitance, the inductance is fixed. Consequently, the cut-off frequency of the ATL is fixed and is ideally independent on the number of sections, as
seen from Eqn 6.8. Therefore, to increase the saturation current level, the number of photodetector can be increased without sacrificing the 3dB bandwidth. On the other hand, the 3dB bandwidth can be increased by using a smaller area photodetector that gives smaller capacitance, while the output saturation power can be maintained by increasing the number of sections; that is increasing the number of photodetectors. However, practically the losses of the electrical transmission line and the optical waveguide limit the number of sections that can be employed usefully. This technique has been successfully applied on various types of photodiode to increase the saturation current level [19,22,24].

6.2.1 PHASE MATCHING AND REVERSE INPUT TERMINATION

A basic P-TWPD model is developed to analyse the various effects of phase matching between the input optical signal and the output electrical signal, and the reverse input termination of the electrical transmission line.

![Periodic-TWPD Circuit Model](image)

Figure 6.7: Periodic-TWPD Circuit Model
The commercial Advanced Design System (ADS) software from Agilent, which is used widely to simulate RF circuits, is used here to model the P-TWPD structure. However, the complete model of a photodetector consists of both RF and optical circuits, which is not available in ADS. To overcome this, the input optical signal is modelled here as an RF signal with a frequency of 194THz, which corresponds to its free space wavelength of interest at 1.55µm. The basic circuit with three photodetectors connected serially are shown in Fig 6.7.

**Optical Waveguide Model**

As the optical signal is modelled as an RF signal, the optical waveguide is simulated using the standard electrical transmission line model in the ADS library. For ideal condition, all the optical signal travels inside the optical waveguide, hence the effective relative dielectric constant is approximately the waveguide relative dielectric constant ($\varepsilon_{\text{wg}}$) of 13.48.

The transmission line has to be terminated, so that there is no distortion to the input signal due to reflection at the end of the transmission line. In practical situations, the optical waveguide is not terminated as any uncoupled input optical signal will either leak or be transmitted into the free space, and the power of the input signal reflected back is negligible. In addition, the loss of the transmission line is set to zero.

**Photodetector Model**

The basic operation of the photodetector is to detect the baseband signal, which is intensity modulated on the optical carrier. The principle of intensity modulation in the optical communication is similar to the amplitude modulation (AM) in RF communication. Therefore, an AM detector is used to detect the baseband signal from the amplitude modulated RF carrier. The AM detector is principally analogous to the photodetector, and is used here to model the photodetector in the simulation.
CHAPTER 6
PERIODIC MODEL & IMPROVED DEVICES

The phase matching and reverse input termination are two important factors that affect the bandwidth of the travelling wave design. In order to analyse the effect of these two factors, the bandwidth and gain of the photodetector model are set to infinity and unity, respectively, so that it does not affect the response of the travelling wave structure. For the monolithic design, a small photodetector device area is preferred. This gives a small junction capacitance, and the inductance required to balance it can be realised using the printed transmission line. The simulated photodetector model is set with an active device area of $5\mu m \times 10\mu m$, which gives an output capacitance ($C$) of $0.028pF$, computed using Eqn 5.2.

Artificial Transmission Line

The output capacitance of the photodetector is integrated into the electrical transmission line to form an artificial transmission line (ATL) structure. The characteristic impedance of the ATL ($Z_{ATL}$) is set to $50\Omega$ to match to the load impedance. It is modelled using discrete inductors and capacitors, where the inductance ($L$) and the capacitance ($C$) are given as follows [54]:

$$L \approx \frac{l_e \sqrt{\varepsilon_{r(\text{eff})}}}{c} Z_{ATL} \quad (6.9a)$$

$$C \approx \frac{l_e \sqrt{\varepsilon_{r(\text{eff})}}}{c} \frac{1}{Z_{ATL}} \quad (6.9b)$$

where $c$ is the free space velocity, $l_e$ and $\varepsilon_{r(\text{eff})}$ are the length and the effective relative dielectric constant of the transmission line, respectively.

As the design of the physical transmission line is an open structure, it is assumed that about half of the electric field propagates in the free space above while the rest will propagates within the absorption and substrate layers. Therefore, the effective relative dielectric constant ($\varepsilon_{r(\text{eff})}$) of the transmission line is approximated by [54]:

$$\varepsilon_{r(\text{eff})} \approx \frac{\varepsilon_{r(\text{avg})} + 1}{2} \quad (6.10)$$
where the weighted average relative dielectric constant of the absorption and substrate layers \(\varepsilon_{r(\text{avg})}\) is approximated to be 13, and \(\varepsilon_{r(\text{eff})}\) is computed to be 7. For an output capacitance \(C\) of 0.028\(\mu\)F for each photodetector, the required inductance \(L\) to give a 50\(\Omega\) characteristic impedance is 0.070\(n\)H. This inductance corresponds to a 159\(\mu\)m long transmission line section \(l_e\). Like the optical waveguide model, the ATL is assumed lossless.

In order for all the photodetector outputs to add coherently, the output from the two consecutive photodetectors must have approximately the same time delay; that is the time delay for the input modulating optical signal \((\tau_0)\) and the output electrical signal \((\tau_E)\) are equal, assuming the time delay \((\tau_{PD})\) for all the photodetectors are the same. Fig 6.8 shows the basic circuit model and the time delay of the periodic structure with 3 serially connected photodetectors.

Figure 6.8: Optical and Electrical Time Constants

The time delay for the input modulating optical signal \((\tau_0)\) can be computed using the following equations:

\[
\tau_0 = \frac{l_e}{v_o} \quad \text{(6.11a)}
\]

\[
v_o = \frac{c}{\sqrt{\varepsilon_{r(\text{wg})}}} \quad \text{(6.11b)}
\]
where \( c \) is the free space velocity, \( l_0 \) and \( \varepsilon_r(\text{wg}) \) are the length and relative dielectric constant of the waveguide, respectively, and \( v_0 \) is the velocity of the modulating optical signal.

The time delay for the output electrical signal \((\tau_E)\) is computed similarly, using the following equations:

\[
\tau_E = \frac{l_e}{v_e} \quad (6.12a)
\]

\[
v_e = \frac{c}{\sqrt{\varepsilon_r(\text{eff})}} \quad \text{or} \quad v_e = \frac{1}{\sqrt{L'C'}} \quad (6.12b)
\]

where \( l_e \) and \( \varepsilon_r(\text{eff}) \) are the length and effective relative dielectric constant of the electrical transmission line, respectively, and \( v_e \) is the velocity of the electrical signal. \( L' \) and \( C' \) are the inductance and capacitance per unit length of the artificial transmission line, respectively. A 159\( \mu \)m long electrical transmission line section \((l_e)\) results in a time delay \((\tau_E)\) of 1.4ps. For phase matching, the length of the optical waveguide \((l_0)\) with \( \varepsilon_r(\text{wg}) \) of 13.48, has to be 114\( \mu \)m to give the same time delay.

### 6.2.2 TIME RESPONSES

The time responses of the P-TWPD were first simulated with an input modulating pulse width set to 2ps. The response of the reverse input terminated P-TWPD is shown in Fig 6.9, which shows no reflected pulse from the reverse input. In addition, the phase mismatch conditions, with a phase ratio \((\tau_O/\tau_E)\) of 0.5 and 2 are simulated by varying the length of the optical waveguide \((l_0)\) accordingly.

The result shows that for phase match condition \((\tau_O/\tau_E=1)\), the output pulse width is similar to the input pulse width of 2ps. On the other hand, with the phase mismatch, the output pulse widths increase and the amplitudes reduce. The phase mismatch leads to pulse spreading and the bandwidth of the P-TWPD is reduced. It is noted that in the phase mismatch situation, a smaller optical delay \((\tau_O/\tau_E=0.5)\) corresponding to a shorter optical waveguide is preferred over a longer optical delay \((\tau_O/\tau_E=2)\), as it gives a smaller pulse spreading and lesser amplitude reduction.
Figure 6.9: Output Time Response for the Reverse Input Terminated P-TWPD with various Phase Matching conditions

Figure 6.10: Output Time Response for the Phase Match P-TWPD with Reverse Input Terminated and Open Circuit
The effect of the reverse input termination on the phase match P-TWPD is simulated. In Fig 6.10, the output response with the reverse input open circuit has multiple pulses due to reflections, which results in a wider pulse width but slightly higher amplitude. The reflected pulse is the backward travelling wave reflected by the open circuit reverse input, and travels toward the output load. The undesirable reflected pulse is added incoherently to the main output pulse along the transmission line. On the other hand, the output response with the reverse input terminated has no reflected pulses, as the reverse input termination absorbs all the backward travelling waves. Although with the reverse input terminated, it does not suffer from pulse spreading, it has lower amplitude.

6.2.3 FREQUENCY RESPONSES

In order to determine the bandwidth of the P-TWPD structure with various phase matching and reverse input termination conditions, a 0.1ps pulse is used as the modulating signal, which simulates an impulse. The outputs of the time domain signal are then Fourier transformed to give the frequency response.

Phase Matching

The simulated 3dB bandwidth of the P-TWPD structure with the reverse input terminated and phase match, as shown in Fig 6.11, is approximately 227GHz. It is the same as the cut-off frequency of the artificial transmission line calculated using Eqn 6.9. At this frequency, the impedance of the artificial transmission line changes from real to imaginary and a ripple can be observed. If the 3dB bandwidth of the artificial transmission line is larger than the photodetector, the bandwidth of the P-TWPD is then limited by the photodetector.
Figure 6.11: Frequency Response of the Phase Match ($\tau_\phi/\tau_E=1$) P-TWPD with Reverse Input Terminated

Figure 6.12: Frequency Response of the Phase Mismatch ($\tau_\phi/\tau_E=0.5$) P-TWPD with Reverse Input Terminated
The responses of the phase mismatch designs with the reverse input terminated are shown in Figs 6.12 and 6.13. The 3dB bandwidth for the phase mismatch ratio ($\tau_o/\tau_E$) of 0.5 and 2, are approximately 192GHz and 140GHz, respectively. It is noted again that in the phase mismatch situation, a smaller optical delay ($\tau_o/\tau_E=0.5$) gives a higher 3dB bandwidth than a longer optical delay ($\tau_o/\tau_E=2$).

**Reverse Input Termination**

Figs 6.14 – 6.16 show the responses of the reverse input open circuit P-TWPD with various phase matching ratio ($\tau_o/\tau_E$) of 1, 0.5 and 2, respectively. Their 3dB bandwidths are 45GHz, 44GHz and 39GHz, respectively, which are lower than the reverse input terminated P-TWPD. However, its output amplitude is approximately 6dB higher than the P-TWPD with the reverse input terminated. It is noted once again that in the phase mismatch situation, a smaller optical delay ($\tau_o/\tau_E=0.5$) gives a higher 3dB bandwidth.
Figure 6.14: Frequency Response of the Phase Match ($\tau_O/\tau_E=1$) P-TWPD with Reverse Input Open Circuit

Figure 6.15: Frequency Response of the Phase Mismatch ($\tau_O/\tau_E=0.5$) P-TWPD with Reverse Input Open Circuit
It is observed from the above simulations that the P-TWPD with reverse input terminated and phase match ($\tau_o/\tau_E=1$) has the highest 3dB bandwidth, which is limited by the ATL cut-off frequency. However, due to fabrication and design constraints, the reverse input is not always terminated. The 3dB bandwidth with reverse input open circuit and various phase matching conditions is summarised in Table 6.2. Also inserted in Table 6.2 is the simulated result of another P-TWPD with a smaller photodetector output capacitance of 0.014pF.

<table>
<thead>
<tr>
<th>P-TWPD</th>
<th>Output Capacitance</th>
<th>$3\text{dB Bandwidth (GHz)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\tau_o/\tau_E=0.5$</td>
</tr>
<tr>
<td>1</td>
<td>0.028pF</td>
<td>44</td>
</tr>
<tr>
<td>2</td>
<td>0.014pF</td>
<td>88</td>
</tr>
</tbody>
</table>

Table 6.2: Summary of the Reverse Input Open Circuit P-TWPDs with various Phase Matching Conditions
The results show that the 3dB bandwidth with a smaller optical delay \((\tau_o/\tau_e=0.5)\) approaches the 3dB bandwidth with the phase match \((\tau_o/\tau_e=1)\). Therefore, when the reverse input is open circuit, it is not critical to achieve phase matching if the optical delay is smaller than the electrical delay \((\tau_o/\tau_e<1)\).

### 6.2.4 PERIODIC MODEL INCORPORATED WITH SSM

This P-TWHPT small signal model was formed by incorporating the HPT small signal model into the basic P-TWPD circuit model, as shown in Fig 6.17. Instead of using the AM detector to model the photodetector in the periodic circuit mode, the 2T-HPT small signal model presented earlier in Chapter 5 is used. This enables the 2T-HPT designed in a periodic travelling wave structure to be simulated.

![Figure 6.17: P-TWHPT Small Signal Model with 3 HPT Sections](image)

MODELLING AND CHARACTERISATION OF 2T-HPTs

6-22
As the 2T-HPT small signal model cannot simulate the optical to electrical conversion, the photocurrent was simulated instead of the optical signal. The input photocurrent was injected at the input end of the electrical transmission line, which is used to model the optical waveguide. As the input photocurrent travelled along the transmission line, it is coupled to the input of the HPT. This is similar to the process of the optical signal coupling from the optical waveguide to the HPT. In addition, the losses of the electrical transmission line and optical waveguide of this model are also not simulated.

![Figure 6.18: P-TWHPT Design with 3 2T-HPT Sections](image)

The P-TWHPT shown in Fig 6.18 is designed with 3 sections of 2T-HPT, each with an active device area of 10μm×5μm and has an output capacitance; i.e. base-collector capacitance \( C_{bc} \) of 0.028pF. The design of the artificial transmission line and the length of the optical waveguide are the same as in the basic P-TWPD design, as given in Section 6.2.1. Fig 6.19 shows the simulated result of the phase match P-TWHPT with 3 and 4 periodic sections. The gain \( G \), unity gain cut-off frequency \( f_c \) and 3dB bandwidth \( f_{3dB} \) for the various number of HPT sections are summarised in Table 6.3.
CHAPTER 6    PERIODIC MODEL & IMPROVED DEVICES

Figure 6.19: Frequency Responses of the Phase Match P-TWHPT with various number of HPT sections

<table>
<thead>
<tr>
<th>Reverse Input</th>
<th>No. of HPT</th>
<th>Gain (G) (dB)</th>
<th>Unity Gain Cut-off Frequency ($f_c$) (GHz)</th>
<th>3dB Bandwidth ($f_{3dB}$) (MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminated</td>
<td>3</td>
<td>43.0</td>
<td>45.1</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>45.5</td>
<td>60.6</td>
<td>263</td>
</tr>
<tr>
<td>Open Circuit</td>
<td>3</td>
<td>49.0</td>
<td>67.5</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>51.5</td>
<td>74.0</td>
<td>186</td>
</tr>
</tbody>
</table>

Table 6.3: Summary Performances of the Phase Match P-TWHPT with various number of HPT sections

The result shows that the 3dB bandwidths are small and they decrease as the number of periodic section increases. Though the output capacitance of the photodetector can be distributed through an artificial transmission line, the 3dB bandwidth of the P-TWHPT is still determined by the individual HPT, which is limited by the long base discharge time constant, as discussed in Section 6.1.2. This large input capacitance increases as the number of cascading HPT section increases.

MODELLING AND CHARACTERISATION OF 2T-HPTs
The gain of the reverse input open circuit P-TWHT is the sum of all the HPT's gain and has a 6dB higher response than the reverse input terminated P-TWHT at low frequency. The reverse input termination absorbed half of the input power, which is the backward travelling wave, hence reduces the gain by half. The gain increases with the number of HPT sections for both reverse input termination conditions and causes the output response to increase throughout the whole frequency range. As a result, the cut-off frequency also increases with the number of HPT sections. However, the 3dB bandwidth reduces as the gain increases. The above arise from the assumption that the optical waveguide distributes the input optical signal to all the HPT, so that there is sufficient input optical power till the last HPT. In addition, the transmission line is assumed lossless, hence the signal is not attenuated as the length is increased with the additional HPT sections.

The P-TWHT is simulated with the phase mismatch. It is noted that in Section 6.2 that a smaller optical delay ($\tau_d/\tau_E<1$), corresponding to a shorter optical waveguide, gives a higher 3dB bandwidth than a longer optical delay ($\tau_d/\tau_E>1$). Hence the length of the optical waveguide is reduced to a minimum of 50μm, which is limited by the resolution of the fabrication process. This gives a phase mismatch ratio ($\tau_d/\tau_E$) of 0.88. The simulated result is summarised in Table 6.4. The gain and the 3dB bandwidth are the same for both phase matching conditions. The phase mismatch reduces the 3dB bandwidth of the artificial transmission line cut-off frequency, as demonstrated in Section 6.2.3. Therefore, the unity gain cut-off frequency of the P-TWHT reduces when the phase is mismatched.

<table>
<thead>
<tr>
<th>Reverse Input</th>
<th>No. of HPT</th>
<th>Unity Gain Cut-Off frequency ($f_c$) (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Phase Match</td>
</tr>
<tr>
<td>Terminated</td>
<td>3</td>
<td>45.1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>60.6</td>
</tr>
<tr>
<td>Open Circuit</td>
<td>3</td>
<td>67.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>74.0</td>
</tr>
</tbody>
</table>

Table 6.4: Performance Comparison for the P-TWHT with various number of HPT sections
Although the above simulation demonstrated that the response of the P-TWHPT improves with the number of sections, the actual improvement is limited by the loss of the artificial transmission line and the optical waveguide. In a practical situation, the loss is generally high for the transmission line fabricated on semiconductor substrate, and it is proportional to the length. Hence, for long transmission line, the loss may overcome the additional gain achieved. For the optical waveguide, it must be able to distribute the input optical signal to all the HPTs. At the same time, the input optical signal also loses its power as it propagates down the waveguide. As a result, this imposes a restriction on the length of the P-TWHPT, and limits the number of HPT sections that can be added.

6.3 SUMMARY

The close match between the measured and simulated result of the 20µm by 20µm 2T-HPT demonstrates the accuracy of the small signal model. The small signal model is used to simulate the effects of various design parameters on the frequency response of the 2T-HPT. The result agrees with the basic principle that a smaller active device area gives a better frequency response due to smaller junction capacitances. In addition, higher bias current reduces the emitter charging time and gives a slight increase in the cut-off frequency. However, there is a limit where further increase of the bias current gives insignificant improvement. Furthermore, smaller collector depletion thickness also improved the frequency response. However, a sufficient thickness is required for a reasonable breakdown voltage for the HPT.

The small signal model is used next to simulate the frequency responses of a smaller HPT with a constant length of 10µm but different active device widths of 1µm, 2µm and 5µm. The result shows that the cut-off frequency is now mainly limited by the base discharge time constant, which is determined by the carrier lifetime. Nevertheless, the simulations show that cut-off frequency of 59.4GHz, with gain of 39dB, can be obtained with an active device size of 1µm×10µm.
To increase the gain without compromising the cut-off frequency, the travelling wave concept is used. In addition, this helps to increase the saturation current level. The periodically distributed design is preferred over the fully distributed design because of its lower electrical transmission line loss. A periodic travelling wave photodetector (P-TWPD) model is developed to analyse the various phase matching and reverse input termination conditions. The P-TWPD with reverse input terminated and phase match has the largest 3dB bandwidth. On the other hand, the reverse input open circuit P-TWPD has a higher response of 6dB at low frequency but lower bandwidth than the reverse input terminated P-TWPD for all phase matching conditions. For the phase mismatch condition, a smaller optical delay ($\tau_o/\tau_E < 1$) corresponding to a shorter optical waveguide is preferred, as it gives a higher bandwidth for both reverse input termination conditions.

The small signal model is subsequently incorporated into the P-TWPD model for further simulation. The simulated P-TWHPT structure consists of multiple 2T-HPTs, each with an active device area of $5\mu m \times 10\mu m$. The gain of the HPT can readily be cascaded, which leads to a higher cut-off frequency. The P-TWHPT also has a higher saturation current level as its design helps to reduce the current density in each HPT. As the 2-terminal P-TWHPT has no electrical input base terminal, its 3dB bandwidth is limited by the long base discharge time constant. Hence phase matching and reverse input termination are not critical for the 2-terminal P-TWHPT. A phase match and reverse input terminated P-TWHPT with 3 periodical section can achieve a gain of 43dB and a cut-off frequency of 45.1GHz. On the other hand, for the same P-TWHPT design but with the phase mismatch and the reverse input open circuit, a gain of 49dB and a cut-off frequency of 67.5GHz are achieved.
CHAPTER 7

CONCLUSIONS & FURTHER DEVELOPMENTS

7.1 CONCLUSIONS

The 2-terminal heterojunction phototransistor (2T-HPT) has a simple design where the bias current is determined by the input optical power. The absence of the base contact helps to remove the parasitic base capacitance and thus gives a higher cut-off frequency ($f_c$). The 2T-HPT is integrated with a leaky optical waveguide longitudinally, so that the optical signal is fed evanescently to the active layers of the HPT. In this way, the waveguide fed 2T-HPT, which is edge illuminated, does not suffer from bandwidth-efficiency trade-off as the carrier and the photon transit paths are independent.

Various frequency response measurement techniques were studied, which determined that the CW excitation techniques are preferred as they can give both the AC and DC output power of the photodetector, which is not available in the pulse excitation technique. The set-up of the direct CW excitation here can measure up to 40GHz, limited by the optical modulator. The optical heterodyne set-up can be used for higher frequency measurements. The IV characteristics, responsitivities and frequency responses of the 2T-HPTs are measured.

The measured IV characteristic shows that the 2T-HPT has high conductance due to high leakage current. This is inherent to the low energy bandgap, poor thermal conductivity and high intrinsic carrier concentration of the $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ collector layer. These also resulted in a low breakdown voltage for the device. On the other hand, the measured responsivity shows a linear response up to 30mA output current.

A small signal model (SSM) for the 2T-HPT is developed. The main elements of the model are computed using the physical device parameters and subsequently optimised
from the measured result. The input photocurrent of the 2T-HPT can be obtained from the model directly to determine the device gain and cut-off frequency. This could not be obtained through measurements due to the inconsistency input optical coupling to the devices.

The frequency responses for the 20µm wide 2T-HPTs with various lengths (200µm, 400µm, and 600µm) showed that the loss of the continuous transmission line, which is formed by the emitter and collector contacts, limit the response at high frequency. The high transmission line loss is also observed in the 2-port characterisations of the transmission line formed by the 2T-HPT back-to-back configuration. Hence, for high frequency operation, it is not practical to design a 2T-HPT with long length to increase the saturation current level or the absorption efficiency. These indicate the 2T-HPT is not suitable to be designed into a fully-distributed travelling wave structure. Additional analysis shows that the cut-off frequency is also limited by the large device capacitances, which are associated with the device area.

Further measurement and simulation demonstrated that the 20µm wide 2T-HPTs with shorter lengths of 20µm and 87.5µm, have a cut-off frequency of 8.3GHz and 2.5GHz, respectively. Using scaling of the model, it was then shown that the cut-off frequency can be increased to 59.4GHz for a 1µm×10µm 2T-HPT and a 3dB bandwidth of 700MHz.

The above result shows that the 2T-HPT could be best utilised in a periodic structure (P-TWHPT), with small device area sections connected by short transmission lines. In this way, the saturation current level can be increased without reducing the cut-off frequency and gain. The simulated optical power distribution of the optical waveguide shows that it decays exponentially along the device length. Sufficient length is required to absorb the optical power whereas excessive length induces unnecessary electrical losses. The P-TWHPT is designed with four 5µm×10µm device area 2T-HPT sections.

A P-TWHPT model is developed to simulate and analyse its frequency response. The model is used to analyse the effects of phase matching and reverse input termination.
conditions, which shows that they are not critical for the 2T-HPT structure. The simulations have demonstrated that the P-TWHPT without phase matching and the reverse input open circuit, a gain of 49dB and a cut-off frequency of 67.5GHz are achieved. However, its 3dB bandwidth is only 219MHz due to the long base discharge time constant.

7.2 FURTHER DEVELOPMENTS

Although the cut-off frequency of the 2T-HPT can be increased to several tenths of GHz, the 3dB bandwidth is limited to less than a GHz. The emitter-base heterojunction accumulates holes in the base region, which further forward biased the junction and increases the current gain. However, without a base terminal, there is no electrical path for the holes to discharge. These base charges can only be discharged via carrier recombination process, which is determined primarily by the carrier lifetimes. This long base discharge time constant, limits the 3dB bandwidth rendering the 2T-HPT unsuitable for modern day communication systems. Therefore, it is necessary to add an electrical base input terminal to provide an electrical path to discharge the excess carriers. This formed a 3 terminal HPT (3T-HPT). However, the additional electrical base contact causes parasitic capacitance, and reduces the gain as part of the photocurrent is conducted away from the base. Nevertheless, the simulated result in Fig 7.1 shows that a 5μm×10μm 3T-HPT with a 280Ω resistor and a 3.5nH inductor connected serially to the base terminal, reduces the gain to 14dB but increases the 3dB bandwidth of 10GHz.
Another possible solution to increase the 3dB bandwidth is to reduce the common-emitter current gain by reducing the base carrier lifetime or increasing the base transit time. The base carrier lifetime can be reduced by using carbon dopant or by increasing the doping concentration, as demonstrated by Cui et al. [48]. On the other hand, the base transit time can be increased by increasing the base width. Although the gain of the HPT is reduced for this method, it does not require a base terminal and therefore, does not suffer parasitic capacitance effect. In addition, it maintains a simple design of 2 terminal and can be used as a drop-in replacement for photodiode.

The frequency responses of a 5μm×10μm 2T-HPT with a common-base gain of 0.9874 and 0.8 are simulated, and shown in Fig. 7.2. The former has a 3dB bandwidth of 400MHz with a gain of 39dB while the latter has a 3dB bandwidth of 7.7GHz but a gain of 14dB. The periodic travelling wave design (P-TWHPT) is subsequently used to increase the gain.
It has been demonstrated that the periodic distributed design can help to increase the gain without reducing the 3dB bandwidth. However, when the 3dB bandwidth is wide the reverse input termination is necessary to reduce the reflected backward travelling wave, which causes distortion and reduces the 3dB bandwidth. Unfortunately, the reverse input termination causes a 6dB reduction in the output response. The impedance tapering design, which has been demonstrated in the microwave amplifier, can be used to cancel the reflected backward travelling wave without having a reverse input termination. In this design, the impedance of the artificial transmission line is not constant. It tapered from a high impedance at the input section to the required 50Ω at the last output section. The mismatch impedance in each section then induces a reflected wave, which cancels the backward travelling wave of each section. In this way, the P-TWHPT can achieve a higher output power without reducing its 3dB bandwidth.

For applications which require high frequency response but low power handling, small device area HPTs are desirable. The edge-illumination configuration for such
devices resulted in poor input coupling efficiency. Hence the vertical illumination configuration, which has a better input coupling efficiency, is preferred. In this configuration, the emitter electrode is made using the optical transparent metal, Indium Tin Oxide (ITO), allowing the optical signal to be illuminated from the device top. As the emitter is a wide bandgap InP material, the optical signal will be only be absorbed by the base and collector layers.
REFERENCES


REFERENCES


REFERENCES


REFERENCES


52. JS Yuan, “SiGe, GaAs, and InP Heterojunction Bipolar Transistors”, USA: John Wiley & Sons, 1999.
