National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Qualité de votre microforme

Qualité de votre microforme

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Separate absorption, grading, charge and multiplication avalanche photodiodes

by

Lawrence E. Tarof, B. Sc., M. Eng.

**A thesis submitted to the Faculty of Graduate
Studies and Research in partial fulfillment of the
requirements for the degree of**

Doctor of Philosophy

**Ottawa-Carleton Institute for Electrical Engineering
Department of Electronics
Faculty of Engineering
Carleton University**

September, 1993

© copyright

1993, Lawrence E. Tarof

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-89730-9

Canada

Name Lawrence E Jarof

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Electronics Engineering

SUBJECT TERM

0544

SUBJECT CODE

U-M-I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
General 0289
Ancient 0290
Linguistics 0291
Modern
Literature 0401
General 0294
Classical 0295
Comparative 0297
Medieval 0298
Modern 0316
African 0591
American 0305
Asian 0352
Canadian (English) 0355
Canadian (French) 0593
English 0311
Germanic 0312
Latin American 0315
Middle Eastern 0313
Romance 0314
Slavic and East European

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
General 0318
Biblical Studies 0321
Clergy 0319
History of 0320
Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
Archaeology 0324
Cultural 0326
Physical 0327
Business Administration
General 0310
Accounting 0272
Banking 0770
Management 0454
Marketing 0338
Canadian Studies 0385
Economics
General 0501
Agricultural 0503
Commerce-Business 0505
Finance 0508
History 0509
Labor 0510
Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
General 0615
International Law and Relations 0616
Public Administration 0617
Recreation 0814
Social Work 0452
Sociology
General 0626
Criminology and Penology 0627
Demography 0938
Ethnic and Racial Studies 0631
Individual and Family Studies 0628
Industrial and Labor Relations 0629
Public and Social Welfare 0630
Social Structure and Development 0700
Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
General 0473
Agronomy 0285
Animal Culture and Nutrition 0475
Animal Pathology 0476
Food Science and Technology 0359
Forestry and Wildlife 0478
Plant Culture 0479
Plant Pathology 0480
Plant Physiology 0817
Range Management 0777
Wood Technology 0746
Biology
General 0306
Anatomy 0287
Biostatistics 0308
Botany 0309
Cell 0379
Ecology 0329
Entomology 0353
Genetics 0369
Limnology 0793
Microbiology 0410
Molecular 0307
Neuroscience 0317
Oceanography 0416
Physiology 0433
Radiation 0821
Veterinary Science 0778
Zoology 0472
Biophysics
General 0786
Medical 0760
EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
General 0566
Audiology 0300
Chemotherapy 0992
Dentistry 0567
Education 0350
Hospital Management 0769
Human Development 0758
Immunology 0982
Medicine and Surgery 0564
Mental Health 0347
Nursing 0569
Nutrition 0570
Obstetrics and Gynecology 0380
Occupational Health and Therapy 0354
Ophthalmology 0381
Pathology 0571
Pharmacology 0419
Pharmacy 0572
Physical Therapy 0382
Public Health 0573
Radiology 0574
Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
Chemistry
General 0485
Agricultural 0749
Analytical 0486
Biochemistry 0487
Inorganic 0488
Nuclear 0738
Organic 0490
Pharmaceutical 0491
Physical 0494
Polymer 0495
Radiation 0754
Mathematics 0405
Physics
General 0605
Acoustics 0986
Astronomy and Astrophysics 0606
Atmospheric Science 0608
Atomic 0748
Electronics and Electricity 0607
Elementary Particles and High Energy 0798
Fluid and Plasma 0759
Molecular 0609
Nuclear 0610
Optics 0752
Radiation 0756
Solid State 0611
Statistics 0463
Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
General 0537
Aerospace 0538
Agricultural 0539
Automotive 0540
Biomedical 0541
Chemical 0542
Civil 0543
Electronics and Electrical 0544
Heat and Thermodynamics 0348
Hydraulic 0545
Industrial 0546
Marine 0547
Materials Science 0794
Mechanical 0548
Metallurgy 0743
Mining 0551
Nuclear 0552
Packaging 0549
Petroleum 0765
Sanitary and Municipal 0554
System Science 0790
Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

The undersigned hereby recommend to the Faculty
of Graduate Studies and Research acceptance of the
thesis,

**Separate absorption, grading,
charge and multiplication
avalanche photodiodes**

submitted by

Lawrence E. Tarof, B. Sc., M. Eng.

in partial fulfillment of the requirements for the
degree of Doctor of Philosophy

Thesis Supervisor

External Examiner

Chair, Department of Electronics

Carleton University
September, 1993

Abstract

This work concerns the study of separate absorption, grading, charge and multiplication (SAGCM) avalanche photodiodes (APDs), which have certain advantages over separate absorption, grading and multiplication (SAGM) APDs.

The use of partial charge in device periphery for an SAGCM APD is proposed, modeled and demonstrated for the first time. Detailed experimental investigations were undertaken for these devices, with many new results. For example, the gain is easily measured independent of responsivity for SAGCM APDs, a measurement which is difficult for SAGM structures. It is predicted and demonstrated for the first time that the gain in an APD depends on the input wavelength. A novel technique was developed to measure electrically the integrated doping using actual SAGCM APDs. It is found for the first time that the bandwidth-gain characteristic has an open "window" of usable bandwidth which moves towards higher or lower gain depending on the breakdown voltage, which in turn depends on the integrated doping. For a given required performance, it is demonstrated for the first time that greater fabrication tolerances can be achieved with an SAGCM APD than a SAGM APD. Very high performance can be achieved in the best case, with a gain-bandwidth product of 122 GHz obtained, the highest published result to date for an InP/InGaAs APD.

In an effort to understand these results, new analytical expressions were developed to predict various characteristics of δ -doped APDs, which are a reasonable approximation to SAGCM APDs. Initially bulk-InP only ionization was considered, and subsequently InP/InGaAs ionization. A variety of experimental results were obtained for SAGCM APDs with respect

to fabrication parameters for the first time, which indicate that the InP/InGaAs ionization model describes well many of these results. Areas of agreement include breakdown voltage and minimum gain for usable bandwidth.

Artificially enhanced ionization (e.g. superlattice multiplication region) APDs should be able to outperform bulk-InP multiplication APDs. Some novel methods by which this may be achieved are presented, and novel design curves are presented for III-V materials combinations not generally thought of for use in these types of devices.

Acknowledgements

I would like to first thank my thesis supervisor, Dr. Garry Tarr, for sticking by me for the nine years it has taken to commence and complete both my Masters and Ph. D. work through part time studies while working at Bell-Northern Research. Both my Masters and Ph. D. projects were outside his area of academic focus. I thank him for the encouragement needed to reach the end of this journey.

My wife Cathy and my children Daniel and Jenny, both born during the course of my graduate work, have been unwavering in their support of my efforts to obtain first my Masters and subsequently my Ph. D., even though this has meant a sacrifice of precious family time. My thanks go to my loving family.

I would like to thank the Bell-Northern Research organization, and in particular my string of managers over the past nine years, Dr. Frank Shepherd, Dr. George Chik, Dr. Jan Glinski and Dr. Gen Ribakovs, for making it possible for me to pursue and complete my graduate work part time. The organization supports graduate work in general, both financially and through moral encouragement. My managers have made it possible for my thesis work to be carried out by giving me the necessary freedom and challenge in the scope of my projects at BNR. In particular, for my Ph. D., I am very fortunate that my job function has permitted the focused study required to obtain my Ph. D. while being in the Ph. D. program for less than four years. I hope this has been a two-way street – i.e. that BNR has derived some benefit from the knowledge contained in my Masters and Ph. D. work.

I owe a special debt to my co-workers at BNR, who have on many occasions gone out of their way, above and beyond their job requirements, to

assist me in the in-depth study necessary to obtain my Ph. D. My greatest debt in this regard is to Jeffrey Yu, who has on many occasions stayed after hours and otherwise gone out of his way to make many of the experimental characterization measurements necessary and to help me to tie up various loose ends. Many of the measurements presented here were performed by Jeffrey or under his supervision. Other co-workers in particular who contributed significantly to this work include Robert Bruce (processing), Tim Baird (assembly), Dr. Gordon Knight and Dr. Norbert Puetz (MOCVD crystal growth), Dr. Carla Miner and Bruno Emmerstorfer (analysis), Forrest Ma (dark current noise measurements), Karen Fox (analysis and processing), Agnes Margittai and Cheryl Maritan (processing), and Dr. Mikelis Svilans and Kumar Visvanatha (characterization).

Table of contents

1. Introduction.....	1
1.1 Photodetectors in receivers.....	1
1.1.1 Optical fiber limitations.....	1
1.1.2 Stringent device requirements.....	2
1.1.3 Receiver sensitivity.....	3
1.1.4 Comparison of photodetectors.....	5
1.2 APD background.....	6
1.2.1 APDs: appropriate materials.....	6
1.2.2 III-V APDs.....	6
1.2.3 Operation under reverse bias.....	7
1.3 Organization of the thesis.....	8
2. Review of APD fundamentals.....	11
2.1 Sources of leakage current in APDs.....	11
2.2 Breakdown and multiplication mechanisms.....	14
2.2.1 Avalanche vs. tunneling breakdown: description.....	14
2.2.2 Impact ionization.....	14
2.2.3 Ionization theories.....	16
2.2.4 Breakdown dependence on E-field, Eg, T and dopant concentration.....	17
2.3 Avalanche rate equations.....	19
2.3.1 Fundamental equations.....	19
2.3.2 Standard expressions for multiplication, noise and avalanche buildup time.....	22
2.3.3 Single and dual carrier ionization.....	30
2.4 Breakdown in planar junctions.....	32
2.5 Fundamental limitations on photodetector performance.....	37
2.5.1 Quantum efficiency.....	37
2.5.2 Average absorption length for photodiodes.....	38
2.5.3 Bandwidth considerations.....	39
2.5.3.1 Transit time.....	39
2.5.3.2 RC.....	42
2.5.3.3 Avalanche buildup.....	42
2.5.3.4 Hole trapping.....	43
2.5.3.5 Diffusion tails.....	44
2.5.3.6 Overall photodetector bandwidth.....	44
2.5.4 Quantum efficiency – bandwidth product.....	49
2.5.5 Optimum depletion layer thickness and maximum bandwidth for devices of finite diameter.....	50
2.5.6 Practical bandwidth limitations for photodiodes.....	57
3. Evolution of APD structures.....	59
3.1 Homogeneous APDs.....	59

3.2 Separate absorption and multiplication (SAM) for III-V APDs.....	60
3.2.1 Concept.....	60
3.2.2 Charge layer requirement.....	60
3.3 Separate absorption, grading and multiplication (SAGM) APDs.....	63
3.4 Constraints in fabrication.....	65
E-field.....	65
Tolerances in fabrication.....	65
4. Separate absorption, grading, charge and multiplication (SAGCM) APDs.....	68
4.1 Candidate device structures.....	68
4.1.1 Planar and mesa geometries.....	68
4.1.2 Standard solution with guard rings.....	70
4.2 Initial SAGCM concept, assumptions.....	71
4.3 Charge sheet in periphery.....	74
4.3.1 Comparison of DC E-fields.....	74
4.3.2 Derivation of numerical equation for cylindrical E-field.....	76
4.3.3 Modeling results.....	79
4.4 Modeling of ideal SAGCM performance: bulk-InP multiplication.....	87
4.4.1 Multiplication equations for bulk-InP ionization.....	87
4.4.2 δ -doped APD: assumptions.....	90
4.4.3 Analytical derivations for d-doped APD.....	92
4.4.4 E-field, voltage and k: modeling results.....	97
4.4.5 Sensitivity analysis for breakdown field and voltage.....	101
4.4.6 Avalanche buildup time.....	102
4.4.7 Displacement correction factor.....	105
4.5 Possible fabrication sequences.....	110
5. Device fabrication and characterization.....	112
5.1 Fabrication.....	113
5.1.1 Mask set.....	113
5.1.2 MOCVD first growth.....	113
5.1.3 Mesa formation.....	116
5.1.4 MOCVD overgrowth.....	116
5.1.5 Zn diffusion.....	117
5.1.6 AR coating.....	118
5.1.7 p-contact.....	118
5.1.8 n-contact.....	119
5.1.9 Dicing.....	119
5.1.10 Mounting.....	119
5.2 Evaluation prior to full processing.....	120
5.2.1 Photoluminescence intensity - nondestructive.....	120
5.2.2 Epitaxial layer thickness measurement - nondestructive.....	121

5.2.3 Hall measurements for doping level.....	125
5.2.4 I-V measurement following Zn diffusion.....	127
5.3 Evaluation of fully processed wafers.....	128
5.3.1 I-V, M-V, C-V.....	128
5.3.2 SIMS.....	129
5.3.3 Frequency response.....	131
5.3.4 Noise.....	132
6. Initial experiments.....	133
6.1 Preliminary structures.....	133
6.1.1 Preliminary SAGM APD.....	133
6.1.2 Preliminary SAGCM APDs.....	135
6.1.2.1 MOCVD charge sheet.....	135
6.1.2.2 Ion implantation.....	137
6.1.3 Choice for further work.....	137
6.2 SAGCM initial results and discussion.....	139
6.2.1 Comparison of RIE and WCE with no etch stop, SL and no SL for mesa formation.....	139
6.2.2 RIE vs. WCE with no etch stop: process control.....	139
6.2.3 DC characteristics: leakage current and breakdown voltage.....	141
6.2.4 SL vs. no SL: speed.....	141
6.2.5 Choices for future processing.....	143
6.2.6 Introduction of etch stop.....	143
7. SAGCM APD performance.....	147
7.1 I-V, M-V and C-V dependencies.....	147
7.2 Frequency response.....	156
7.3 Noise.....	165
8. SAGCM APD measurements vs. fabrication parameters.....	168
8.1 Experiment to vary x_d , measure V_{ann} , V_{mesa} and BV	168
8.2 Whole wafer measurements.....	175
8.2.1 Characteristic voltages vs. wafer position.....	175
8.2.2 f-M vs. wafer position.....	179
8.3 Inadequacy of InP-only multiplication model.....	184
9. Ionization outside the "multiplication" region.....	185
9.1 Multiplication.....	185
9.2 Modeling results: E-field.....	191
9.3 Voltage.....	200
9.3.1 Modeling results.....	200
9.3.2 Comparison with experiment.....	204
9.4 $M(x)$	206
9.4.1 Modeling results.....	206
9.4.2 Comparison with experiment.....	212
9.5 Minimum gain for full bandwidth.....	215
9.5.1 Modeling results.....	215
9.5.2 Comparison with experiment.....	218
9.6 Summary.....	223

10. Artificially enhanced ionization (AEI) APDs	224
10.1 Motivation	224
10.2 InGaAs/InAlAs experimental results	227
10.3 Ideal AEI, Type $\frac{1}{2}$ APD	227
10.4 Band offsets, Type $\frac{1}{2}$ APD in real materials systems	230
10.5 Design of AEI APDs	231
11. Conclusions	235
References	238

List of tables

2.5.1	Literature values of electron and hole velocities.....	41
2.5.2	Optimum bandwidths and thicknesses for p-i-n diodes.....	55
4.3.1.	Modeling parameters assumed for the E-field profiles in Section 4.3.....	80
4.4.1.	Ionization coefficients for different InGaAsP compositions from the literature.	94
7.2.1	Device parameters for high GBW devices.	163
7.3.1	Device parameters for the noise measurements.	165
8.1.1	Parameters used in wafer for experiment to vary x_d	169
8.2.1	Parameters used in three whole wafers from the P reactor.....	179

List of figures

1.1	Schematic diagram of optoelectronic repeater.....	3
2.3.1	Illustration of multiplication and current density vs. position for an APD with $\alpha > \beta$	20
2.3.2	Excess noise factor $F(M)$ vs. M for constant k , from McIntyre's equation.....	27
2.4.1	Schematic diagrams of a (a) plane and (b) planar p-n junction.....	33
2.4.2	Illustration of approximation of a planar junction as a cylindrical depletion region of inner boundary r_0 and outer boundary R	35
2.5.1	Illustration of expected bandwidth vs. gain for an APD.....	48
2.5.2	Illustration of a depletion region of diameter D and thickness t	51
2.5.3	Electrical frequency response vs. $f \cdot t_e$ for transit time limited response of a p-i-n photodiode.....	56
3.1	Layer sequence for a SAM APD.....	61
3.2	(a) Layer sequence and (b) E-field vs. x for a SAGM APD.....	64
4.1.1	Illustration of (a) mesa and (b) planar geometries.....	69
4.1.2	Illustration of an APD with a guard ring.....	71
4.2.1	Schematic of Webb's SAGCM APD.....	72
4.2.2	(a) Layer sequence and (b) E-field vs. x for a SAGCM APD.....	73
4.3.1	Schematic diagram of a SAGCM APD with partial charge sheet in the periphery.....	75
4.3.2	Illustration of approximation of a planar junction as a cylindrical depletion region of inner boundary r_0 and outer boundary R for a SAGCM APD.....	77
4.3.3	Approximate cylindrical representation of the peripheral region of a SAGCM APD.....	77
4.3.4	Representative slab (Cartesian) E-field profiles for an SAGCM APD for the case of no, some and full peripheral charge sheet.....	81
4.3.5	Representative E-field profiles for the cylindrical peripheral region of an SAGCM APD for the case of no, some and full peripheral charge sheet.....	81
4.3.6	Maximum cylindrical E-field in InP and InGaAs vs. fractional charge sheet for planar APD structure, various x_d	83
4.3.7	Maximum cylindrical E-field in InP and InGaAs vs. fractional charge sheet for planar APD structure, various x_j	83
4.3.8	Maximum cylindrical E-field in InP and InGaAs vs. fractional charge sheet for planar APD structure, various t_{InGaAs}	84
4.3.9	Voltage vs. t_{InGaAs} in the center of the device (Cartesian).....	86
4.3.10	Voltage vs. x_d in the center of the device (Cartesian).....	86
4.4.0	Illustration of multiplication and current density vs. position for an APD with $\beta > \alpha$	88

4.4.1	E-field for breakdown and $M=10$ vs. x_d for an ideal δ -doped APD with bulk InP ionization only.....	98
4.4.2	Voltages for breakdown and $M=10$ vs. x_d for an ideal δ -doped APD with bulk InP ionization only.....	99
4.4.3	Calculated k for breakdown and $M=10$ for an ideal δ -doped APD with bulk InP ionization only.....	100
4.4.4	Gain-bandwidth product vs. w for an ideal δ -doped APD with bulk InP ionization only.....	105
4.4.5	Displacement correction factor vs. k for different carrier velocity ratios.....	107
4.4.6	Displacement correction factor vs. carrier velocity ratios for different k	109
4.4.7	Displacement correction factor vs. x_d for carrier velocity ratio of 0.5.....	109
4.5.1	Schematic diagram of ion implanted SAGCM APD.....	111
4.5.2	Schematic diagram of MOCVD charge sheet SAGCM APD.....	111
5.1.1	Schematic cross-section of a fully processed SAGCM APD.....	112
5.1.2	Schematic diagram of a full 2" InP wafer showing the conventions used in this work for describing quarter wafers and tiles on the whole wafer. The direction of MOCVD gas flow is also shown.....	114
5.2.1	Typical fringe map of InP top layer thickness. The difference in thickness between adjacent fringes (black or white) is $0.19\text{ }\mu\text{m}$	124
5.2.2	Doping profile from Hall measurements for the P reactor.....	126
5.2.3	Doping profile from Hall measurements for the R1 reactor.....	126
5.3.1	Typical SIMS profiles showing x_d measurement.....	130
5.3.2	x_j measurement across a 2" wafer.....	131
6.1.1	I-V characteristics of preliminary SAGM structures from O618A with SL grading layer.....	134
6.1.2	I-V curve for preliminary planar blanket MOCVD charge sheet structure.....	136
6.1.3	I-V curves for planar blanket ion implantation structure.....	138
6.2.1	Histograms of measured etch depth on quarter wafers for WCE with no etch stop (a-c) and RIE (d).....	140
6.2.2	I-V characteristics for SAGCM APDs processed as quarter wafers as indicated from the combination of WCE with no etch stop or RIE, SL grading layer or no SL grading layer.....	142
6.2.3	Histogram of measured etch depths on whole wafers using RIE. The results from three wafers are included here.....	144
7.1.1	I-V and M-V curves for eight individual devices on R1-406L/R1-355T.....	148
7.1.2	Schematic diagram of the depletion region for SAGCM APDs under the annular and mesa punchthrough conditions.....	149
7.1.3	C-V curves for R1-406L/R1-355T for devices of various sizes. Note the sharp drop in capacitance at V_{ann} and V_{mesa}	152

7.1.4	M-V measurements over four decades of input photocurrent for O1029/R1-147L.....	154
7.1.5	I-V measurement for P555-3d2.	155
7.1.6	M-V measurement for P555-3d2.....	155
7.2.1	High frequency response for R1-406L/R1-355T, various gains.....	158
7.2.2	Typical f-M curves for P405/396 (1Q grading layer with n InGaAs) and P555 (9Q grading layer with n InGaAs).	159
7.2.3	Schematic diagram of an optimized 9Q APD.	160
7.2.4	GBW measurement for O1029/R1-147L. The GBW is 107 GHz.	161
7.2.5	GBW measurement for P623. The GBW is 122 GHz.....	163
7.2.6	f-M curve for P555-3d2.	164
7.3.1	F vs. M curve for various SAGCM APDs.....	166
7.3.2	Shot noise power vs. I_d curves for a typical SAGCM APD.....	166
8.1.1	Schematic diagram of SAGCM APD showing relevant quantities for x_d experiment.....	169
8.1.2a	SIMS profiles of Zn for Zn diffusions of various times into an APD structure.	170
8.1.2b	Junction depth vs. time for the SIMS profiles of Fig. 8.1.2a.	170
8.1.3	Characteristic voltages vs. x_d for samples from a single wafer.....	173
8.1.4	Tunneling threshold, breakdown and edge breakdown voltages and gain vs. x_d for samples from a single wafer.....	174
8.2.1	I-V and M-V curves for P389/382 for (a) center breakdown and (b) edge breakdown devices.	176
8.2.2	Characteristic voltages vs. position for P-242/R1-366T.....	177
8.2.3	Characteristic voltages vs. position for P555/543 and P623/614.....	177
8.2.4	GBW vs. position for P623/614.....	180
8.2.5	f-M curves for many devices from P555/543.	180
8.2.6	GBW vs. BV: composite results from P555/543, P623/614 and P624/616.	182
8.2.7	Mmin for 2 GHz bandwidth vs. BV: composite results from P555/543, P623/614 and P624/616.....	182
8.2.8	GBW vs. Mmin: composite results from P555/543, P623/614 and P624/616.	183
9.1.1	Illustration of multiplication vs. position for a multilayer APD.....	188
9.2.1	E-field in the InP at breakdown and at operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization.....	192
9.2.2	E-field in the InP at breakdown and at operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$ for InP/InGaAs ionization.	192
9.2.3	E-field in the InP at breakdown and at operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$: comparison between InP-only and InP/InGaAs ionization.....	194
9.2.4	E-field in the InP at breakdown and at operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$: comparison between InP-only and InP/InGaAs ionization.....	194

9.2.5	E-field at the InP/InGaAs heterointerface at breakdown and at operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$: comparison between InP-only and InP/InGaAs ionization.....	196
9.2.6	E-field at the InP/InGaAs heterointerface at breakdown and at operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$: comparison between InP-only and InP/InGaAs ionization.....	196
9.2.7	E-field vs. x	198
9.2.8	Ionization coefficients vs. x	198
9.3.1	Breakdown and operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization.	201
9.3.2	Breakdown and operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$ for InP/InGaAs ionization.....	201
9.3.3	Breakdown and operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$: comparison between InP-only and InP/InGaAs ionization.	203
9.3.4	Breakdown and operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$: comparison between InP-only and InP/InGaAs ionization.	203
9.3.5	Breakdown and operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$: comparison between InP-only and InP/InGaAs ionization.	205
9.4.1	Multiplication for various injection conditions vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization.....	207
9.4.2	Ionization coefficients in InP and InGaAs vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$	207
9.4.3	Multiplication for various injection conditions vs. charge for constant x_d of $0.3 \mu\text{m}$ for InP/InGaAs ionization.....	209
9.4.4	Ionization coefficients in InP and InGaAs vs. charge for constant x_d of $0.3 \mu\text{m}$	209
9.4.5	Multiplication for injection at the heterointerface for both InP/InGaAs and InP-only multiplication vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for the same conditions which give a gain of 10 for injection with $1.3 \mu\text{m}$ wavelength light for InP/InGaAs ionization.	211
9.4.6	Multiplication for injection at the heterointerface for both InP/InGaAs and InP-only multiplication vs. charge for constant x_d of $0.3 \mu\text{m}$ for the same conditions which give a gain of 10 for injection with $1.3 \mu\text{m}$ wavelength light for InP/InGaAs ionization.	211
9.4.7	Experimental measurement of ratio of gain at $1.55 \mu\text{m}$ wavelength to gain at $1.3 \mu\text{m}$ wavelength at a gain of 10 at $1.3 \mu\text{m}$ wavelength.....	214
9.5.0	Illustration of the effect of constant x_d and varying charge for constant E_{het}	216
9.5.1	Voltages for breakdown, $M=10$, $E_{\text{het}}=9 \text{ V}/\mu\text{m}$, and $E_{\text{het}}=7 \text{ V}/\mu\text{m}$ vs. x_d for constant charge of $3 \times 10^{12} \text{ cm}^{-2}$	217

9.5.2	Voltages for breakdown, $M=10$, $E_{het}=9 \text{ V}/\mu\text{m}$, and $E_{het}=7 \text{ V}/\mu\text{m}$ vs. charge for constant x_d of $0.3 \mu\text{m}$	217
9.5.3	Minimum gain at full bandwidth vs. x_d for constant charge of $3 \times 10^{12} \text{ cm}^{-2}$. assuming $E_{het}=9 \text{ V}/\mu\text{m}$, $E_{het}=7 \text{ V}/\mu\text{m}$	219
9.5.4	Minimum gain at full bandwidth vs. charge for constant x_d of $0.3 \mu\text{m}$ assuming $E_{het}=9 \text{ V}/\mu\text{m}$, $E_{het}=7 \text{ V}/\mu\text{m}$	219
9.5.5	Minimum gain at full bandwidth vs. breakdown voltage for constant charge of $3 \times 10^{12} \text{ cm}^{-2}$ assuming $E_{het}=9 \text{ V}/\mu\text{m}$, $E_{het}=7 \text{ V}/\mu\text{m}$	220
9.5.6	Minimum gain at full bandwidth vs. breakdown voltage for constant x_d of $0.3 \mu\text{m}$ assuming $E_{het}=9 \text{ V}/\mu\text{m}$, $E_{het}=7 \text{ V}/\mu\text{m}$	220
9.5.7	Minimum gain at full bandwidth vs. breakdown voltage for three experimental wafers. The theoretical curve is simulated for $x_d = 0.3 \mu\text{m}$ and $E_{het} = 4.5 \text{ V}/\mu\text{m}$	222
10.1	Schematic diagram of artificial enhancement ionization APD.....	226
10.2	Type 1 and 2 heterojunctions.....	229
10.3	Type $1\frac{1}{2}$ heterojunction.....	229
10.4	ΔE_c vs. Elow.....	232
10.5	ΔE_v vs. Elow.....	232
10.6	$\frac{\Delta E_c}{\Delta E_{low}}$ vs. Elow.....	233

List of acronyms

AEI	artificially enhanced ionization
APD	avalanche photodiode
HJ	heterojunction
PL	photoluminescence
RIE	reactive ion etching
RS	reflectance spectroscopy
SAM	separate absorption and multiplication
SAGM	separate absorption, grading and multiplication
SAGCM	separate absorption, grading, charge and multiplication
SIMS	secondary ion mass spectroscopy
WCE	wet chemical etching

List of symbols

BV	breakdown voltage
C	capacitance
D_p	hole diffusivity
e	electronic charge
E_B	breakdown field
E	E-field (usually) or energy, depending on context
E_g	bandgap energy
f	fractional charge sheet used in E-field simulations in Section 4.3.
f_{-3dB}	-3 dB electrical bandwidth
f_{RC}	RC limited electrical bandwidth
f_e	electron transit time limited electrical bandwidth
f_h	hole transit time limited electrical bandwidth
h	Planck's constant
I_d	dark current
I_{diff}	diffusion dark current
I_{g-r}	generation-recombination dark current
I_{tun}	tunneling dark current
J_{diff}	diffusion current density
J_e	electron current density
J_{gr}	generation-recombination current density
J_p	hole current density
J_{tun}	tunneling current density
k	ratio of ionization coefficients
k_B	Boltzmann's constant
$K_{displacement}$	displacement correction factor
L	inductance
L_p	hole diffusion length
l_a	absorption length
m_e	electron mass
m_h	hole mass
M_e	multiplication due to electron injection
M_h	multiplication due to hole injection
n	refractive index
n_i	intrinsic carrier concentration
N_D	doping concentration (per volume)
q	electronic charge
R	responsivity or resistance, depending on context
t	thickness or time depending on subscripted context: e.g. t_{InGaAs} is the InGaAs thickness; t_e is the electron transit time
T	absolute temperature

v_n	electron velocity
v_p	hole velocity
V_{ann}	annular punchthrough voltage
V_b	breakdown voltage
V_{mesa}	mesa punchthrough voltage
x	distance coordinate in the direction of growth (perpendicular to the layers) or mole fraction of group III element, depending on context
x_d	undoped multiplication region width, usually within one layer
x_j	p-diffused junction depth
y	distance coordinate not in the direction of growth (parallel to the layers) or mole fraction of group V element, depending on context
w	multiplication region width
α	electron ionization coefficient
β	hole ionization coefficient
γ	absorption coefficient
ϵ	dielectric constant, which may be appropriately subscripted
λ	wavelength of light
λ_g	bandgap wavelength
σ	charge sheet density (areal), which may be appropriately subscripted
τ	time constant, appropriately subscripted
τ_{eff}	effective carrier lifetime

1. Introduction

1.1 Photodetectors in receivers

The increasing requirements demanded of long haul, high bit rate optical fiber telecommunications systems have led to the necessity of developing high performance optoelectronic components suitable for operation in the 1.3-1.55 μm wavelength regime. Research concerning one such optoelectronic component forms the basis of this work. This initial section discusses the basic requirements of optical fiber telecommunications systems, and in particular addresses the receivers. The optoelectronic component of interest within the receiver is the photodetector, which converts incoming light from the optical fiber into electrical current. Various possible photodetectors are compared, demonstrating the need for an avalanche photodiode (APD) in a state-of-the-art high performance receiver. An appropriate set of materials must be used to fabricate these devices. The suitability of the III-V materials systems, and in particular those compounds lattice matched to InP, is discussed. Some important concepts concerning APDs are presented.

1.1.1 Optical fiber limitations

As the bit rate and transmission distance along the fiber become large, certain effects from the fiber can impair the operation of an optical fiber communications system [1-10]. For ideal operation, the fiber would have an attenuation of zero and a constant refractive index with respect to wavelength. However, real optical fibers are not ideal. Attenuation in the fiber, which is caused by the presence of residual amounts of O-H chemical groups, places an upper limit on how far a signal can be transmitted and still

be receivable. The non-zero material dispersion – i.e. the change in refractive index with respect to wavelength – causes light of different wavelengths to travel at different velocities down the fiber. For conventional fiber, the attenuation minimum is at $1.55\ \mu\text{m}$ and the dispersion passes through zero at $1.3\ \mu\text{m}$, which is why these are the two most commonly used wavelengths in fiber optical communications systems. Thus there is a constraint which links the maximum bit rate, maximum distance and maximum linewidth for which the initial transmitted light pulse is still receivable.

1.1.2 Stringent device requirements

A typical arrangement for propagation of the signal in a fiber optical transmission system as illustrated in Fig. 1 involves a transmitter, the fiber itself, a receiver and electronics [1-12]. The transmitter converts the original electrical information into an optical signal. This signal is input through one end of the fiber, which propagates the signal. At the output end of the fiber, the light emerges from the fiber into a receiver. The receiver converts the optical signal to an electrical signal, which is processed for the application of interest using various electronics. For low bit rate, small distance applications such as local area networks, the device performance requirements are only moderately stringent. For high bit rate, "long reach" applications such as bulk data transmission or high definition television, the device performance requirements become more stringent. In many cases it may be possible to use many devices of moderate performance in place of few devices of premium performance. However, the manufacture, installation and maintenance of devices is expensive, and so there is considerable economic incentive to develop higher performance devices, which if produced in volume would be of comparable cost to moderate performance devices.

1. Introduction

Figure 1.1 Schematic diagram of optoelectronic repeater.

1.1.3 Receiver sensitivity

High speed electronics has made it possible to encode and decode information at ever faster rates. The state of the art in narrow linewidth laser development and low dispersion optical fibers has made it possible to transmit light at high bit rates for long distances without the initial pulses, which were separated when launched into the fiber, spreading out to the point where they overlap, which would prevent the different bits from being distinguished. However, as the distance along the fiber increases for long reach applications, the initial power which was launched into the fiber is attenuated from its initial value [1-10]. As laser linewidths become smaller in order to avoid the dispersion penalty, stimulated Brillouin scattering in the optical fiber limits the maximum power which can be launched into the fiber [8]. At the same time, it is doubtful whether the fiber attenuation performance can be easily improved much beyond its present state of 0.05-0.2 dB/km at 1.55

1. Introduction

μm wavelength. With these developments the state of the art in fiber optical systems is limited by the sensitivity of the receiver at high bit rates. Receiver sensitivity is defined as the minimum amount of light power incident upon the receiver to guarantee a specified bit error rate (BER). Examples of typical BERs are 10^{-9} or 10^{-12} .

A receiver generally consists of a detector, which receives the light and converts the energy to a photocurrent, and one or more transistors to amplify the current. Several receiver configurations are possible. One of the most often seen configurations uses a photodetector to detect the light, and any of several standard state-of-the-art amplifier configurations to amplify the photocurrent signal. In this case the sensitivity is limited by the photodetector properties, including the photocurrent obtainable, leakage current present and response time.

No receiver can be more sensitive than the shot noise limit for detection, which is unavoidable given the presence of a photocurrent, even if the leakage current is zero [1,5]. The shot noise from the photocurrent is proportional to the photocurrent and the bandwidth. As the bandwidth increases, the minimum receivable power also increases proportionally. As the bandwidth decreases, the minimum receivable power also decreases proportionally until the point where the photocurrent becomes comparable to the leakage current. For the case of no leakage current, if the signal is digital and Gaussian statistics can be used to describe the noise, the shot noise limit is proportional to the photon energy and bandwidth, inversely proportional to the quantum efficiency, and for a given bit error rate is proportional to $\text{erfc}^{-1}(\text{BER})$ [1, 5, 6, 10]. For non-Gaussian statistics, the situation is more complicated [13, 14].

1. Introduction

1.1.4 Comparison of photodetectors

The most common candidates for photodetectors are the p-i-n photodiode and the APD. They both have the same basic operating principle, whereby light is absorbed and converted to photocurrent. The p-i-n photodiode is a serious photodetector candidate for certain optoelectronics applications, and best suited to applications of almost any bandwidth conceivable today where reasonably high signal levels are present. Reports of very high bandwidths of up to 40 GHz exist in the literature [15-18]. The dark current in a well designed p-i-n photodiode is very low [19], and less than that of an APD. The dark current is low enough to influence the receiver noise significantly only for low bandwidth applications where the minimum received power is also low. The low dark current of the p-i-n photodiode is of limited benefit at high bandwidths, where the amplifier noise is likely to dominate the performance.

By using a detector which has built-in gain, it is possible to extend the sensitivity of the receiver to lower levels of received light. An APD is often used for this purpose [1-7, 9-12, 20, 21]. Initially proposed by Read in 1958 [20], an APD operates in a manner similar to a p-i-n photodiode, with the exception that the primary photocurrent caused by the photoelectric effect is amplified within the same device. An APD based receiver is more sensitive than a p-i-n photodiode based receiver, provided the APD has sufficient bandwidth for the application [5, 6]. However, an APD has significantly less bandwidth than an otherwise equivalent p-i-n photodiode. Unlike the p-i-n photodiode, there is excess noise associated with the APD multiplication. This limits the maximum useful gain of the APD, to the point where the APD noise is comparable to the first stage amplifier noise [6].

1.2 APD background

1.2.1 APDs: appropriate materials

For wavelengths of light less than $1.1\ \mu\text{m}$, it is possible to use a Si APD [1, 22, 23]. Such APDs have been under development for a long time and are well understood. However in this spectral regime both the dispersion and attenuation in the optical fiber are high, which makes the use of such wavelengths impractical for long haul telecommunications. The dispersion in the fiber is zero near $1.3\ \mu\text{m}$, and the attenuation is at a minimum near $1.55\ \mu\text{m}$, which necessitates the development of optical components which operate efficiently at these wavelengths [1]. Both Ge and III-V APDs operate in this wavelength regime. Since the Ge APDs are known to have high leakage current and comparable hole and electron ionization coefficients (to be described in more detail in Section 2), it is generally agreed that the state of the art performance is likely to come from a III-V APD [1-12].

1.2.2 III-V APDs

Most III-V APDs operating in this wavelength regime are based on the InP materials system – i.e. InP or materials which can be grown on an InP substrate. Owing to the poor reliability and dislocations which can result from the stress of epitaxial layers of different lattice constant from the substrate, a high performance APD of interest should contain primarily epitaxial layers which are lattice matched to the substrate. Such a materials system is the set of quaternary alloys $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$, where $y=2.197x$, and x and y are mole fractions of the group III and V elements, respectively [24]. This quaternary materials system allows for alloys which are lattice matched to InP over the bandgap wavelength (λ_g) and energy (E_g) range of $0.92\text{-}1.65\ \mu\text{m}$ and $1.35\text{-}0.75\ \text{eV}$, respectively, for $y=0\text{-}1$. For any given composition, light of

wavelength less than λ_g is absorbed by the particular alloy, whereas light of wavelength greater than λ_g is transmitted.

The absorbing material in both p-i-n photodiodes and APDs is usually $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ (InGaAs), in which the absorption length at $1.3\ \mu\text{m}$ is $0.87\ \mu\text{m}$ [25]. For this material both 1.3 and $1.55\ \mu\text{m}$ light are readily absorbed. As the product of E_g and λ_g is a constant, given by

$$E_g \lambda_g = 1.24\ \text{eV}\cdot\mu\text{m} \quad (1.1),$$

InGaAs is the composition in the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ materials system for which E_g is at its lowest energy and λ_g is at the highest wavelength. It is possible to use other quaternary alloys lattice matched to InP to absorb light at 1.3 and/or $1.55\ \mu\text{m}$. However InGaAs is almost always used because the absorption length for a given wavelength increases as y increases, which would require thicker layers to absorb a given fraction of the light [24,25].

1.2.3 Operation under reverse bias

In order to operate a p-i-n photodiode or an APD at high speed it is necessary to create an electric field (E-field) to facilitate rapid transport of carriers from the absorbed area across the depletion region because drift is orders of magnitude faster than diffusion. This usually involves the depletion of a region which includes the absorption region by applying a reverse bias to the diode. For maximum depletion at a given voltage in the area of interest, a one-sided p⁺-n junction is generally used [1, 6, 10, 12, 26]. The one-sided abrupt junction model will be used in all the modeling presented in this work.

1. Introduction

1.3 Organization of the thesis

Chapter 2 describes the theory of the APD. Most of the treatment is based on the literature, except for Section 2.5.5 where the maximum bandwidth of p-i-n photodiodes and APDs for devices of non-zero diameter is analytically treated in a novel way. In Chapter 3, the evolution of the APD is reviewed from the literature, with emphasis on the rationale for separation of the absorption, grading and multiplication regions into different areas of the device. The often seen separate absorption, grading and multiplication (SAGM) APD is discussed.

In Chapter 4, the separate absorption, grading, charge and multiplication (SAGCM) APD is introduced, upon which most of the remainder of this work is based. After a discussion on preventing edge breakdown in Section 4.1, Webb's initial SAGCM concept [27] is presented in Section 4.2. An original SAGCM APD design with partial charge in the device periphery [28-30, 162-164] is introduced in Section 4.3, together with numerical modeling of the E-field and voltage of the device based on Webb's assumptions. In Section 4.4, novel modeling results for E-field, voltage, avalanche buildup time and k (a parameter related to the ratio of ionization coefficients) are presented, based on an exact analytical treatment of an ideal δ -doped APD. These results are compared to recent numerical simulations by Kuchibholta and Campbell [31]. In Section 4.5 possible fabrication sequences for such a device are discussed.

In Chapter 5, the device fabrication and characterization techniques are discussed. These techniques are largely standard, except for those presented in Section 5.2. In particular, the optical techniques for determining epitaxial layer thickness are novel [32-34], based on the work presented in Ref. [35-37].

Some initial experiments are described in Chapter 6. In Section 6.1, experimental results are presented for preliminary SAGCM APD structures with no attempt to avoid edge breakdown. In Section 6.2, some initial experimental results are presented for complete SAGCM APDs with different methods of fabrication.

In Chapters 7 and 8, the experimentally measured performance of SAGCM APDs is presented. In Chapter 7, the current-voltage, multiplication-voltage, capacitance-voltage dependencies, the frequency response and noise are presented for individual devices. It is found that unlike SAGM APDs, a unity gain regime exists in SAGCM APDs which allows for independent measurement of gain and responsivity for the first time. In addition, certain characteristic voltages of SAGCM APDs can be defined which relate to the operational device physics. Also, a gain-bandwidth product in excess of 100 GHz is achieved, the highest obtained to date in a bulk-InP multiplication APD.

In Chapter 8, the SAGCM APD performance is evaluated in terms of specific fabrication parameters. Section 8.1 describes measurements of the characteristic voltages as a function of InP multiplication region width. Based on the device physics, it is shown that the charge sheet in the actual device may be inferred from the device characteristics themselves in a novel way, rather than relying on doping calibrations from other wafers. Section 8.2 describes measurements of the characteristic voltages, bandwidth vs. multiplication, minimum gain with usable bandwidth and gain-bandwidth over whole wafers. The trends in device performance with respect to known variations in fabrication parameters are found to be inconsistent with the δ -doped APD theory presented in Chapter 4. In particular, the experimental

evidence suggests that ionization within the InGaAs absorption region may not be neglected, as is usually done in theoretical treatments of III-V APDs.

An original analytical model for the performance of SAGCM APDs is developed in Chapter 9 based on standard formulations, taking into account ionization in the InGaAs. The case of the ideal δ -doped APD is considered, similar to Section 4.4. Novel modeling results for the E-field, voltage, gain and minimum gain with usable bandwidth are presented. Comparisons with experiment are given where possible, and semi-quantitative agreement is obtained. Also, it is predicted and measured for the first time that the gain in the APD depends on the wavelength of light absorbed, within the wavelength range for which the device is intended.

Artificially enhanced ionization (AEI) APDs, the APDs of the future, are discussed in Chapter 10. Sections 10.1 and 10.2 discuss the motivation and experimental results to date on these devices. In Section 10.3, the Type 1-1/2 heterojunction APD is introduced, a novel concept in AEI APD design. In Section 10.4, an original method of designing AEI APDs is presented, which takes into account conduction and valence band discontinuities and the well bandgap energy. Novel design curves are presented for some examples of materials which might in practice be used in an AEI APD.

Conclusions of this work are presented in Chapter 11.

2. Review of APD fundamentals

In this chapter the fundamental phenomena which govern APD operation are reviewed.

2.1 Sources of leakage current in APDs

There are several fundamental sources of leakage current. The intent of this section is to give a brief description of these mechanisms, which are often considered when analyzing the sources of leakage current in APDs [38-42].

Minority carriers diffusing towards or away from the junction in the neutral regions cause diffusion current [23, 26, 43]. The fundamental source of diffusion current is gradients in carrier concentration, which give rise to diffusion. In the limit of a large p⁺-n diode where both the diode radius and n region thickness are much larger than minority hole diffusion length, the diffusion current density J_{diff} is given by

$$J_{diff} = \frac{e n_i^2 D_p}{N_D L_p} [\exp(e V / k_B T) - 1] \quad (2.1.1)$$

where e is the electronic charge, L_p is the hole diffusion length, D_p is the hole diffusivity, V is the voltage, k_B is Boltzmann's constant, T is the absolute temperature and n_i is the intrinsic carrier concentration given by

$$n_i = 2 \frac{(4\pi m_e m_h)^{3/4}}{h^3} (k_B T)^{3/2} \exp\left(\frac{-E_g}{2k_B T}\right) \quad (2.1.3)$$

where m_e and m_h are the electron and hole mass, h is Planck's constant and E_g is the energy bandgap.

2. Theory

Generation and recombination of electrons and holes takes place in the depletion region of the diode, and is associated with the filling and emptying of energy levels within the bandgap (traps). Generation current arises from minority carriers generated from thermal excitation of electron hole pairs, and the subsequent transport of the minority carriers to the other side of the junction (where they become majority carriers) under the influence of the E-field. Recombination current arises from majority carrier flow in the opposite direction to that of generation current. These carriers recombine on the side of the junction where they are minority carriers. Only those carriers which have sufficient thermal energy to surmount the potential barrier contribute to the recombination current, so the recombination current is sensitive to the applied voltage. The generation-recombination current density J_{gr} is given approximately by

$$J_{gr} = \frac{en_i W}{\tau_{eff}} [\exp (eV/2k_B T) - 1] \quad (2.1.4)$$

where τ_{eff} is the effective carrier lifetime, assuming the traps are located at midgap [23, 26, 43].

The ratio $\frac{J_{gr}}{J_{diff}}$ is of interest [43]. Under reverse bias such that $V/2k_B T \gg 1$, the temperature dependence of this ratio is proportional to $1/n_i$ which is in turn proportional to $\exp(E_g/2k_B T)$. It is therefore expected that J_{gr} should dominate at low temperatures and J_{diff} should dominate at high temperatures.

For sufficiently high E-field in sufficiently low bandgap material, tunneling current flows. The tunneling current density J_{tun} is given by

2. Theory

$$J_{\text{tun}} = \sqrt{\frac{2m_e}{E_g}} \frac{e^3 E_{\text{max}}^2 w}{h^2} \exp\left(-\frac{2\pi\beta_{\text{tun}}m_e^{1/2}E_g^{3/2}}{e h E_{\text{max}}}\right) \quad (2.1.5)$$

where E_{max} is the maximum E-field, w is the depletion thickness and β_{tun} is a dimensionless number which depends on the detailed shape of the tunneling barrier. For a triangular barrier $\beta_{\text{tun}}=1.11$ and for a parabolic barrier $\beta_{\text{tun}}=1.88$ [23,43].

The origin and physical path of travel of the leakage current is of importance. For a planar device, there are some current paths which pass through the multiplication region and other current paths which avoid the multiplication region. Leakage current which passes through the multiplication region is said to be multiplied leakage current. Leakage current which does not pass through the multiplication region is said to be unmultiplied leakage current. For unmultiplied leakage current, the noise is that associated with the leakage current only. For multiplied leakage current, there is performance degradation in a receiver beyond that of the leakage current alone [44]. The equivalent noise is that of the primary leakage current multiplied by the excess noise factor and the square of the gain, which will be discussed in more detail in Section 2.3.

Generation-recombination current which arises within the central diode region passes through the multiplication region. Diffusion current and tunneling current are multiplied or unmultiplied depending on the path of leakage current.

In addition to these mechanisms, there are other sources of leakage current which are not fundamental. For example, excess generation-recombination current which may be field dependent can result from traps or

2. Theory

defects in the material. As well, surface leakage is possible, which would almost always be unmultiplied.

2.2 Breakdown and multiplication mechanisms

2.2.1 Avalanche vs. tunneling breakdown: description

As the reverse bias increases the E-field throughout the depletion region increases. In the extreme case breakdown can occur by either of two mechanisms: avalanche or tunneling [1, 4, 11, 12, 23, 26, 45]. Avalanche breakdown occurs as a result of impact ionization: a carrier acquires sufficient energy from the E-field that upon impact with the lattice an ion, or electron-hole (e-h) pair, is produced. These e-h pairs in turn pass through the E-field and generate more e-h pairs, and so on. In this way gain occurs as the initial carrier passes through the high E-field region. At high enough E-field the gain becomes in theory infinite, as will be discussed in Section 2.3. Tunneling occurs when sufficient band-bending or sloping has occurred that the carriers pass from the valence band to the conduction band, as described in Eqn. (2.1.5). The resultant current flowing through the device from the tunneling process is not dependent on the photocurrent, unlike impact ionization. Tunneling is a source of leakage current, and is therefore undesirable in an APD.

2.2.2 Impact ionization

The avalanche photodiode is based on the phenomenon of avalanche multiplication. Electrons and/or holes (carriers) can acquire large amounts of energy from a high E-field or other means. When a collision between a carrier and the lattice occurs, the energy from the carrier can be transferred to the

lattice. Sufficient energy can be absorbed by the lattice that an electron is promoted from the valence to the conduction band, creating an electron-hole pair. This process is called impact ionization. These new carriers are swept out by the E-field and can themselves acquire high energy, causing further electron-hole pairs to be created. The entire process in which many carriers are created from one initial carrier is called avalanche multiplication [11,12].

The impact ionization process in a semiconductor is a "discrete" process – either it does or does not occur. If insufficient energy is transferred to the lattice to cause ionization, even by a small margin, the energy is dissipated, usually by heating, and is lost to the impact ionization process. Carriers in the process of gaining energy can lose their energy through non-ionizing scattering events before acquiring sufficient energy to ionize. The scattering rates can be different for electrons and holes. This process differs among metals, semiconductors and insulators. In a metal the energy states are continuous and so any amount of energy, even a small amount, can be transferred from a carrier to the lattice. This scattering leads to Joule heating with the result that avalanche breakdown in metals is rarely observed. In a semiconductor or insulator, the bandgap energy is such that a certain minimum energy must be transferred to the lattice, and given the right conditions (e.g. sufficiently high E-field), it is in certain cases possible for carriers to acquire this much energy, although this process is in competition with non-ionizing scattering events. The main difference between a semiconductor and an insulator is the amount of energy which must be gained by an electron to promote from the "valence" to "conduction" band, which is larger for an insulator than for a semiconductor.

2. Theory

2.2.3 Ionization theories

Semiconductor ionization has been studied since the 1950's. Early work by Wolff [46] in 1954 introduced the concept of a threshold energy, greater than or equal to the bandgap energy, below which ionization would not occur and above which ionization would always occur. Others demonstrated that once carriers have a certain energy they are more likely to cause multiplication [47-52]. Shockley and Wolff [46, 47] also pointed out that for carriers in fields so high that ionization is significant, optical phonons are expected to dominate the scattering.

The probability of ionization within a particular path length in the semiconductor is inversely related to the mean length over which ionization occurs. Ionization coefficients, of dimension inverse length, are defined as α for ionization caused by electron collisions with the lattice and β for ionization caused by hole collisions with the lattice. Both α and β depend on the electric field (E-field) in a manner to be discussed later. Understanding the ionization coefficients is central to an understanding of any APD.

The simplest theory pertaining to ionization probability is the lucky drift model, first proposed by Schockley [47]. In this model, similar to Wolff's observation, a carrier causes an ionization event if and only if the carrier manages to acquire a threshold energy (greater than or equal to the bandgap energy), and α and β depend in detail on this process. Later embellishments to this model include concepts such as "quasi-ionization rates" which depend on both α and β [53, 54] and a "soft threshold" energy [55], as well as further refinement to the scattering calculations [48-52]. The effect of hydrostatic strain in Si and Ge has been studied [56]. Monte Carlo simulations are often used in these calculations [51].

2. Theory

Experimentally the ionization coefficients have been determined for a number of materials systems. Capasso [12] has previously reviewed ionization coefficients in Si, Ge, GaAs, InAs, $\text{Ga}_x\text{As}_{1-x}\text{Sb}$, GaP and $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ compounds lattice matched to InP, and studied the ionization coefficients in $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ (lattice matched to InP) [57]. As well, Cook [58], Windhorn [59], Osaka [60,61], Umebu [62], Armiento [63] and Pearsall [64] have studied ionization coefficients for various $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ compounds lattice matched to InP.

2.2.4 Breakdown dependence on E-field, E_g , T and dopant concentration

It is worthwhile to consider the effect of the electric field (E-field), bandgap energy (E_g), temperature (T) and dopant concentration (N_D) on breakdown, for both the avalanche and tunneling breakdown mechanisms [26, 38-65]. For a given E_g , the greater the E-field, the more sloped the band diagram is and the closer the conduction band is to the valence band. Therefore the probability of either avalanching or tunneling increases. There are not surprisingly exponential relationships between the E-field and both the probability of avalanching and tunneling. By the same argument, for a given E-field, there is also an exponential relationship between E_g/kT and the probability of tunneling. At first glance it appears that there should also be an exponential relationship between E_g and the probability of avalanching since less impact energy must be transferred from the initiating carrier to the lattice atom for ionization if the bandgap is smaller, but other events must also occur for impact ionization. The carrier which is acquiring energy from the E-field must not collide prematurely with the lattice, which would cause the carrier to lose its energy from scattering without initiating an ionization

2. Theory

event. Once a collision with the lattice occurs, the probability of impact ionization increases if E_g is smaller, but this is a relatively weak effect compared with the effect of E_g on the tunneling probability. This means that if E_g is increased, there is a marginal reduction in the probability of impact ionization but a large reduction in the probability of tunneling, allowing for some selectivity between the impact ionization process (desirable) and the tunneling process (undesirable).

The probability of premature scattering is related to the lattice properties and the doping level in the material. To first order and for a given semiconductor material, if the lattice is of high quality the regular spacing of the lattice atoms leads to high mobility and therefore a relatively low probability of premature collision; the carrier has a relatively high chance of acquiring sufficient energy to cause ionization. However, a dopant is a perturbation of the lattice which can lead to carrier deflection and therefore reduced mobility; increasing amounts of dopant increase the probability of premature scattering and therefore reduce the ionization coefficients for a given E-field.

Increasing T increases the phonon vibration of the lattice, thereby (similarly to doping) increasing the probability of premature scattering and reducing the ionization coefficients for a given E-field. The opposite is true for tunneling: as T increases the tunneling probability increases exponentially. Other sources of leakage current, such as generation-recombination and diffusion also tend to increase with increasing T or smaller E_g . Ideally, the lower the temperature the greater the selectivity toward the avalanche and away from the tunneling process, but environmental considerations "in the field" in telecommunications do not always allow for operation at low temperatures. Experimentally it is observed

2. Theory

that the breakdown voltage increases for increasing temperature for both Si [65] and InP/InGaAs [21] APDs.

2.3 Avalanche rate equations

2.3.1 Fundamental equations

The mathematical description of the avalanche gain mechanism within the multiplication region has been treated in the literature and is well known. Most literature definitions [e.g. 11] are from the point of view of Si APDs for which electron injection is desirable, since $\alpha > \beta$. In this geometry, the p contact of the diode is traditionally located at $x \leq 0$ and the n contact of the diode is located at $x \geq w$. This is illustrated schematically along with the hole and electron current densities as functions of x in Fig. (2.3.1). The direction of the E-field is such that electrons travel from 0 to w , which form the boundaries of the depletion region. The electron and hole current densities $J_n(x)$ and $J_p(x)$ are related to each other and the ionization coefficients by [10]

$$\frac{dJ_n(x)}{dx} = \alpha(x)J_n(x) + \beta(x)J_p(x) + eG(x) \quad (2.3.1)$$

$$-\frac{dJ_p(x)}{dx} = \alpha(x)J_n(x) + \beta(x)J_p(x) + eG(x) \quad (2.3.2)$$

where $G(x)$ the generation rate of electron-hole pairs. $G(x)$ is usually associated with the absorption of a photon at position x within the depletion region. Note that α and β are functions of position insofar as the E-field is a function of position. Defining the total current $J = J_n(x) + J_p(x) =$

Figure 2.3.1 Illustration of multiplication and current density vs. position for an APD with $\alpha > \beta$.

2. Theory

constant, Eqns. (2.3.1) and (2.3.2) can be rewritten

$$\frac{dJ_n(x)}{dx} = [\alpha(x) - \beta(x)] J_n(x) + \beta(x) J + eG(x) \quad (2.3.3)$$

$$-\frac{dJ_p(x)}{dx} = [\alpha(x) - \beta(x)] J_p(x) - \alpha(x) J - eG(x) \quad (2.3.4)$$

These first order differential equations are solved by using the integrating factor $\exp[-\int_0^x (\alpha - \beta) dx']$ and integrating from 0 to w to obtain two equivalent expressions.

$$J = \frac{J_p(w) + J_n(0) \exp\left[\int_0^w (\alpha - \beta) dx'\right] + e \exp\left[\int_0^w (\alpha - \beta) dx'\right] \cdot \int_0^w G \exp\left[-\int_0^x (\alpha - \beta) dx'\right] dx}{1 - \int_0^w \beta \exp\left[\int_x^w (\alpha - \beta) dx'\right] dx} \quad (2.3.5)$$

$$J = \frac{J_p(w) \exp\left[-\int_0^w (\alpha - \beta) dx'\right] + J_n(0) + e \int_0^w G \exp\left[-\int_0^x (\alpha - \beta) dx'\right] dx}{1 - \int_0^w \alpha \exp\left[-\int_0^x (\alpha - \beta) dx'\right] dx} \quad (2.3.6)$$

2. Theory

Eqn. (2.3.5) or Eqn. (2.3.6) is the starting point for all multiplication calculations. The condition for avalanche breakdown is that the denominator of either equation go to zero.

2.3.2 Standard expressions for multiplication, noise and avalanche buildup time

The multiplication due to generation of carriers at position x within the multiplication region may be extracted from the above equations. The simplest way to treat this is to assume that $G(x)=G\delta(x)$, where $\delta(x)$ is the Dirac delta function, and assume both $J_n(0)$ and $J_p(w)$ are zero. In this case, the multiplication $M(x)$ due to generation of carriers at position x is given by

$$M(x) = \frac{J}{eG} = \frac{\exp \left[- \int_0^x (\alpha - \beta) dx' \right]}{1 - \int_0^w \alpha \exp \left[- \int_0^{x'} (\alpha - \beta) dx' \right] dx} \quad (2.3.7)$$

from Eqn. (2.3.6), with a similar formulation possible from Eqn. (2.3.5) [11].

The multiplication due to hole and electron injection are given from Eqn. (2.3.7) by the special cases $M_n=M(0)$ and $M_p=M(w)$. For electron injection,

$$M_n = \frac{1}{1 - \int_0^w \alpha \exp \left[- \int_0^{x'} (\alpha - \beta) dx' \right] dx} \quad (2.3.8)$$

and for hole injection

$$M_p = \frac{\exp \left[- \int_0^w (\alpha - \beta) dx' \right]}{1 - \int_0^w \alpha \exp \left[- \int_0^{x'} (\alpha - \beta) dx' \right] dx} \quad (2.3.9)$$

Note that Eqns. (2.3.8) and (2.3.9) are the same as would be obtained from assuming $G(x)=0$ everywhere, and using Eqn. (2.3.6) and assuming

$$M_n = \frac{J}{J_n(0)}, M_p = \frac{J}{J_p(w)} \quad (2.3.10)$$

It is evident that using this convention for the coordinate system, the expression for multiplication for electron injection is less complicated than for hole injection. Note that a change in axis orientation would not change the presence of an integral in the numerator of Eqn. (2.3.10) and the absence of an integral in the numerator of Eqn. (2.3.9). This convention is useful for analysis of Si-based APDs, in which it is usually desirable to design the device in such a way as to cause electron injection.

Other quantities such as the noise and avalanche buildup time can be calculated once the gain $M(x)$ is known.

There has been a lot of study on APD noise to date [66-75]. The statistical nature of the multiplication process leads to shot noise in excess of the shot noise which would be associated with the current from a p-i-n

2. Theory

photodiode. The noise is often expressed in terms of the excess noise factor F from McIntyre's formulation [66,67]:

$$F = M \left[1 - (1-k) \left(\frac{M-1}{M} \right)^2 \right] \quad (2.3.11)$$

$$\langle i_n^2 \rangle = 2e((I_{po} + I_{do})M^2F + (I_{pu} + I_{du}))B \quad (2.3.11a)$$

$$\approx 2eI_{po}M^2FB \text{ for } I_{po} \gg I_{pu}, I_{do}, I_{du} \quad (2.3.11b)$$

where M is the gain, $\langle i_n^2 \rangle$ is the root mean square noise current, I_{po} and I_{do} are primary multiplied photocurrent and dark current respectively (e.g. the actual multiplied photocurrent is $M \cdot I_{po}$), I_{pu} and I_{du} are unmultiplied photocurrent and dark current respectively, B is the bandwidth and k is the k -factor related to the ratio of ionization coefficients in the following way.

$$k_{eff} = \frac{k'_{eff}}{k_1^2}, \text{ where} \quad (2.3.12)$$

where

$$k'_{eff} = \frac{k_2 - k_1^2}{1 - k_2} \quad (2.3.13)$$

$$k_1 = \frac{\int_0^w \beta M(x) dx}{\int_0^w \alpha M(x) dx} \quad (2.3.14)$$

2. Theory

$$k_2 = \frac{\int_0^w \beta M^2(x) dx}{\int_0^w \alpha M^2(x) dx} \quad (2.3.15)$$

The two values k_{eff} and k'_{eff} are the effective k-factors to be used in Eqn. (2.3.11) for electron and hole injection, respectively. For a constant E-field in the multiplication region, α and β are constant. In this case $k_{eff} = \beta/\alpha$ and $k'_{eff} = \alpha/\beta$.

If the ionization coefficients α for electrons and β for holes are both non-zero then the initial carrier type can generate e-h pairs. The electrons and holes drift in opposite directions, and both may cause further impact ionizations. This can continue as a back-and-forth process within the avalanche region [1,11,12]. Thus via positive feedback one initial carrier can eventually cause many e-h pairs to be generated from several passes through the avalanche region. At first glance this may appear to be an advantage since the gain is increased over the case where only one carrier type initiates the multiplication. However (as will be discussed below) this increased gain is at the expense of both more avalanche multiplication noise and more avalanche buildup time of the device, and is therefore undesirable [1,11,12].

As is evident from inspection of Eqn. (2.3.11), a lower k leads to a lower F. Accordingly, the ratio of ionization coefficients k is defined in this work by

$$k = \begin{cases} \alpha/\beta & \text{if holes injected into the avalanche region} \\ \beta/\alpha & \text{if electrons injected into the avalanche region} \end{cases} \quad (2.3.16)$$

2. Theory

In this way, $k \leq 1$ if the carrier of higher ionization coefficient is injected into the avalanche region. This is consistent with much of the published literature on Si or InP-based APDs, for each of which k is traditionally defined as being less than one when the correct carrier is injected into the avalanche region.

It is interesting to consider the effect of k on the excess noise in the APD. The simplest case to consider is that of constant k , for which the noise equations become greatly simplified. A plot of $F(M)$ vs. M from McIntyre's theory is shown in Fig. 2.3.2 for k ranging from 0.001 to 1000. The plots are seen to be approximately symmetric on a log-log scale about the axis $F(M)=M$ for k and $1/k$. The physical significance is as follows. At any given gain, lower excess noise is obtained if the carrier with the higher ionization coefficient is injected into the multiplication region. For low k , if the correct carrier is injected, the excess noise can be quite low, with a limiting value of 2 for $k=0$ at high gain. However, if the wrong carrier is injected, the excess noise becomes very high, with the penalty becoming increasingly more severe as the disparity between ionization coefficients increases. In summary, it is important to inject the carrier with the higher ionization coefficient into the multiplication region. The lower the k , the lower the excess noise.

The effect of gain variation in the lateral direction of the device on the noise can be important [71, 72]. Inspection of Eqns. (2.3.11)-(2.3.15) indicates that the noise power increases superlinearly (between M^2 and M^3 dependence) with respect to the signal power (M^2 dependence). Suppose there is some distribution of gains across a real device. Then there is an average gain across the device, with some distribution of gain higher and lower than the average. The noise power, superlinear with respect to signal power, becomes weighted towards the higher gain regions. The resultant noise for the given average gain (which is what would be measured

2. Theory

experimentally for a spot size comparable to the device size) is greater than the noise for a device with the same uniform gain. In other words, inhomogeneity in the gain profile of the device results in higher noise than were the gain profile uniform.

Most noise calculations and experimental results which appear in the literature follow McIntyre's equations [66, 67]. The k and F work described in this thesis also follows McIntyre's equations. However, formulations other than McIntyre's are possible. For example, Marsland [68-69] and Hayat [70] have recently considered the "dead space" between ionization events. They conclude that McIntyre's calculations overestimate the excess noise factor for a given k .

Figure 2.3.2 Excess noise factor $F(M)$ vs. M for constant k , from McIntyre's equation.

2. Theory

The avalanche buildup time is an important quantity in the study of APDs [1-7, 9-12, 76-86]. The primary avalanche build-up time $\tau_{aval n}$ for electron initiated multiplication is

$$\tau_{aval n} = \frac{K_{displacement}}{v_n + v_p} \int_0^w \exp \left[- \int_0^x (\alpha - \beta) dx' \right] dx \quad (2.3.17)$$

where $K_{displacement}$ is the correction factor and v_n, v_p are the electron, hole velocities. The physical origin of $K_{displacement}$ is displacement current which arises from the space-charge induced E-field resulting from the carrier motion (i.e. electron and hole currents) [79]. The primary avalanche build-up time τ_p for hole initiated multiplication is

$$\tau_{aval p} = \frac{K_{displacement}}{v_n + v_p} \exp \left[\int_0^w (\alpha - \beta) dx' \right] \int_0^w \exp \left[- \int_x^w (\alpha - \beta) dx' \right] dx \quad (2.3.18)$$

These quantities are in the ratio

$$\frac{\tau_{aval p}}{\tau_{aval n}} = \exp \left[\int_0^w (\alpha - \beta) dx' \right] \quad (2.3.19)$$

The gain-bandwidth product (GBW) is given by

$$GBW = \frac{1}{2\pi\tau_{aval n}} \text{ for electron injection} \quad (2.3.20)$$

2. Theory

$$= \frac{1}{2\pi\tau_{\text{aval}p}} \text{ for hole injection}$$

It is evident upon inspection of Eqns. (2.3.19) and (2.3.20) that electron injection gives a higher GBW for the case $\alpha > \beta$ and hole injection gives a higher GBW for the case $\beta > \alpha$. The existence of a constant GBW at high gain arises from the positive feedback of multiplying carriers within the multiplication region. Accordingly, the GBW limitation on APD bandwidth only applies to gains in excess of $1/k$, for it is only for these gains that positive feedback is significant. The most common approximation to the above equation is given by Emmons [76], who has shown that for high gain and frequency, for electron injection and equal hole and electron velocities for the case $\alpha > \beta$,

$$\text{GBW} \approx \frac{1}{N(w/v_n)k} \quad (2.3.21)$$

where N is a slowly varying number that is $\frac{1}{3}$ for $k=1$ and 2 for $k=0.001$ which takes into account the displacement correction. This is the same trend in performance as was the case for the noise. Thus there is not a tradeoff between low noise and high GBW. Injecting the correct carrier – that with the higher ionization coefficient – into the multiplication region is beneficial to both the GBW and the excess noise. Furthermore, the lower the k , the higher GBW for constant w . Qualitatively, the less positive multiplication feedback the more beneficial for device performance.

2. Theory

2.3.3 Single and dual carrier ionization

The equations presented above are in general very complicated because of the variation of both ionization coefficients with position. To give feeling for the effects of the ionization coefficients on the avalanche gain mechanism, three special cases may be considered for the case of electron injection: (1) $\beta=0$, (2) $\alpha=\beta$ and (3) $\alpha=0$. In all these cases a uniform electric field will be assumed which makes the ionization coefficient(s) constant throughout the multiplication region of width w . These considerations apply to hole injection if α and β are reversed.

The first case ($k=0$) is the ideal case for the highest performance APD. As seen above, the lowest noise and highest GBW are possible for this condition. M_n is given from Eqn. (2.3.5) as

$$M_n = \exp(\alpha w) \quad (2.3.22)$$

In this case, the multiplication increases as α increases. Also, there is no true avalanche breakdown since the denominator of Eqn. (2.3.5) is always 1 and never goes to zero.

The second case of equal ionization coefficients ($k=1$) is given from Eqn. (2.3.8).

$$M_n = M_p = \frac{1}{1 - \alpha w} \quad (2.3.23)$$

In this case a true avalanche breakdown is obtained for the case $\alpha w=1$. Webb has shown that the avalanche breakdown is sharper with respect to E-field as k increases from 0 (Eqn. (2.3.22)) to 1 (Eqn. (2.3.23)).

2. Theory

The third case is similar to the first case in that there is no true avalanche breakdown. However, inspection of Eqn. (2.3.8) shows that there is no gain either. This is the same case as the first, but injecting the wrong carrier.

Thus if $k \gg 1$, then the avalanche process does not take place because the wrong carrier is injected. If $k \sim 1$, then some positive feedback occurs. If $k \ll 1$ the positive feedback is not present. It is always more efficient to inject carriers from the direction where $k < 1$, even for $k \sim 1$ [11, 12, 66, 67].

2.4 Breakdown in planar junctions

The doping, E-field (E) and electrostatic potential drop (V) within the depletion region of a reverse biased diode are determined from Poisson's equation and Gauss' Law:

$$\nabla \cdot \mathbf{E} = \frac{N_D - N_A}{\epsilon} \quad (2.4.1)$$

$$\nabla V = -\mathbf{E} \quad (2.4.2)$$

where ϵ is the dielectric constant. Very often the characteristics of an APD may be solved using one dimensional (1-D) versions of these equations. Except where noted, it will be assumed throughout this work that the potential drop outside the depletion region is negligible. In addition, the built-in potential from the p-n junction will be neglected unless otherwise stated, because the voltages of interest are usually tens of volts, much greater than the built-in potential of typically less than 1 V. For these assumptions, the voltage and electrostatic potential difference across the device are equal.

Two types of p-n junctions are typically found in a wide variety of semiconductor materials: plane and planar junctions [26]. These are illustrated schematically in Fig. 2.4.1. For a plane junction, in which the p-n junction is a straight line, the 1-D solution of Poisson's equation in Cartesian coordinates is appropriate. In that case, Eqns. (2.4.1) and (2.4.2) simplify to

$$\frac{dE}{dx} = \frac{N_D - N_A}{\epsilon} \quad (2.4.3)$$

$$\frac{dV}{dx} = -E \quad (2.4.4)$$

Figure 2.4.1 Schematic diagrams of a (a) plane and (b) planar p-n junction.

The solution to these equations is very straightforward.

$$E(x) = \int_0^x \frac{N_D(x') - N_A(x')}{\epsilon(x)} dx' + E(0) \quad (2.4.5)$$

$$V(x) = \int_0^x E(x') dx' + V(0) \quad (2.4.6)$$

Usually $V(0)$ is taken to be either zero or the bias applied to the device, because in a real experimental situation what matters is the voltage drop across the device rather than a reference to some other external voltage. The constant $E(0)$ is less arbitrary and is in reality whatever value is necessary to satisfy the known electrostatic potential across the device, as given in Eqn. (2.4.6). This point may be clarified further as follows. At the very edge of the

2. Theory

depletion region, $E(0^-)=E(w^+)=0$ since it is assumed that there is no potential drop outside the depletion region. However, the E-field increases abruptly from zero to its value $E(0^+)$ or $E(w^-)$ within the depletion region, with a ramp distance related to the doping of the n^+ and p^+ regions adjacent to the depletion region. For the purposes of modeling in this work, this ramp distance will be neglected. At the p^+-n junction ($x=0$), the E-field is at its highest point within the device. This E-field decreases from 0^+ throughout the depletion region to its value at w^- according to Eqn. (2.4.3). As noted above, the integral of this E-field profile is the electrostatic potential across the device, as given in Eqn. (2.4.6).

A planar junction differs from a plane junction in that the edges are curved. This is the junction typically formed from a diffusion through a selectively open mask. For the central area of a planar junction, it is often a good approximation to treat the junction as a plane junction and apply the 1-D Cartesian coordinate solution to Poisson's equation [26, 87-91]. The curved region of a planar junction may be approximated as a cylindrical boundary in many cases. This is illustrated in Fig. 2.4.2 for a cylindrical depletion region of inner boundary r_0 and outer boundary R . In this case a 1-D solution may still be used, but this time in cylindrical rather than Cartesian coordinates [58-62]. The appropriate differential equations (2.4.1) and (2.4.2) become:

$$\frac{1}{r} \frac{d}{dr} (rE) = \frac{N_D - N_A}{\epsilon} \quad (2.4.7)$$

$$\frac{dV}{dr} = -E \quad (2.4.8)$$

2. Theory

Figure 2.4.2 Illustration of approximation of a planar junction as a cylindrical depletion region of inner boundary r_0 and outer boundary R . The E-field may be considered at any coordinate r within the depletion region.

Eqn. (2.4.7) may be rewritten as

$$\frac{dE}{dr} + \frac{1}{r}E = \frac{N_D - N_A}{\epsilon} \quad (2.4.9)$$

which is a first order differential equation which can be solved using the integrating factor $\exp[-\int \frac{1}{r} dr]$ with solution

$$E(r) = \frac{1}{r} \int_{r_0}^r \frac{N_D - N_A}{\epsilon} r' dr' + E(r_0) \quad (2.4.10)$$

$$V(r) = \int_{r_0}^R E(r) dr \quad (2.4.11)$$

The case of a one sided abrupt junction with constant doping in a semi-infinite material may be readily considered from these equations. The E-field and voltage are determined from Eqns. (2.4.10) and (2.4.11):

$$E(r) = \frac{N_D}{2\epsilon} \left(r - \frac{r_0^2}{r} \right) \quad (2.4.12)$$

$$V(r) = \frac{N_D}{2\epsilon} \left(\frac{r^2 - r_0^2}{2} - r_0^2 \ln\left(\frac{r}{r_0}\right) \right) \quad (2.4.13)$$

The variable r may be eliminated between these two equations to give [87]

$$V = \frac{N_D r_0^2}{4\epsilon} \left(\left\{ 1 + \frac{2\epsilon E}{N_D r_0} \right\} \cdot \ln \left\{ 1 + \frac{2\epsilon E}{N_D r_0} \right\} - \frac{2\epsilon E}{N_D r_0} \right) \quad (2.4.14)$$

where E is the maximum E-field present in the device. This relationship holds between the breakdown field and breakdown voltage. This may be compared with the expression which may be obtained from similar integration of Eqns. (2.4.5) and (2.4.6) in Cartesian coordinates, again for semi-infinite material and an abrupt one-sided junction:

$$V = \frac{\epsilon E}{2N_D} \quad (2.4.15)$$

The plane junction is the limiting case of a planar junction with infinite radius of curvature. Repeated applications of L'Hospital's rule on Eqn. (2.4.14) can be used to show that in the limit of $r_0 \rightarrow \infty$, V approaches the value in Eqn. (2.4.15).

2. Theory

2.5 Fundamental limitations on photodetector performance

In this section some of the fundamental limitations on photodetector performance are discussed. Most of this information comes from or can be derived from the literature: quantum efficiency, average absorption length and bandwidth limitations. In Section 2.5.5 an analytical derivation is presented for optimum photodetector bandwidth as a function of device diameter and absorption layer thickness not previously reported in the literature.

2.5.1 Quantum efficiency

The fraction of light absorbed by the photodiode depends on the wavelength of the light and the thickness of the absorption material. Such absorption follows Beer's Law. The internal quantum efficiency QE_{int} of the device is given by

$$QE_{int} = 1 - \exp(-\gamma x_{abs}) \quad (2.5.1)$$

where γ is the wavelength-dependent absorption coefficient and x_{abs} is the thickness of absorbing material. Define

$$l_a = 1/\gamma \quad (2.5.2)$$

as the "absorption length" – the amount of material required so that $1/e$ of the light would be transmitted in the absence of reflections. The product γx_{abs} or quotient x_{abs}/l_a is the number of absorption lengths in the device. Note that both γ and l_a are functions of the wavelength of light.

2. Theory

A further consideration is the fraction of the light incident upon the photodetector which reaches the absorption layer, which may be limited by reflections encountered prior to the absorption layer. The external quantum efficiency QE_{ext} includes this effect and is given by

$$QE_{\text{ext}} = (1 - \text{Refl}) QE_{\text{int}} \quad (2.5.3)$$

where Refl is the wavelength-dependent reflectance of the photodetector. The reflectance of the air-InP interface is roughly 30 % in the wavelength regime of interest. Often an antireflection layer can be used to reduce this effect. The simplest way to proceed is to use a quarter wavelength thick layer of dielectric. If the APD structure is a heterostructure, the effect of reflections within the heterostructure may be included as well [92]. Even more sophisticated, a stack of dielectric materials may be used as a wavelength filter [93] or to broaden out the wavelengths for which nearly zero reflection takes place. With this in mind, the fundamental limitation on quantum efficiency is QE_{int} , as $QE_{\text{ext}}/QE_{\text{int}}$ is mainly influenced by factors which can be designed for.

2.5.2 Average absorption length for photodiodes

The average coordinate at which light is absorbed is also the average distance traveled by the holes $\langle x_h \rangle$ for a semi-infinitely thick InGaAs layer. This quantity is given from statistical considerations [94] by

$$\langle x_h \rangle = \gamma \int_0^{\infty} x e^{-\gamma x} dx = 1/\gamma \quad (2.5.4)$$

where $1/\gamma$ is the absorption length ($0.87 \mu\text{m}$) of the InGaAs. For finite thickness of InGaAs and front or rear entry, Eqn. (2.5.4) may be generalized:

$$\langle x_{h,\text{front}} \rangle = \frac{\gamma}{QE} \int_0^{x_{\text{abs}}} x e^{-\gamma x} dx = \frac{1}{\gamma} \left(1 - \gamma x_{\text{abs}} \frac{1-QE}{QE} \right), \quad (2.5.5)$$

$$\langle x_{h,\text{rear}} \rangle = x_{\text{abs}} - \langle x_{h,\text{front}} \rangle \quad (2.5.6)$$

where QE is the quantum efficiency of the device given by Eqn. (2.5.1). In this case it is assumed that for the front entry device, light is absorbed from initial x coordinate 0 to the thickness x_{abs} . The average absorption lengths may be used for bandwidth estimation and other calculations.

2.5.3 Bandwidth considerations

There has been considerable work in the literature on the study of photodetector bandwidth [1-7, 9-12, 80-86, 95-113]. The speed of the device is affected by a number of considerations, which is the subject of this section. A subset of the APD bandwidth limitations is contained within the p-i-n detector limitations [15-18, 95]. There are further considerations which are specific to APDs, beyond that of a p-i-n detector. In this section several individual limitations on photodetector performance are discussed. Then the combination of these effects is discussed first for a p-i-n photodiode and then for an APD.

2.5.3.1 Transit time

There are two transit times for an APD. The first is the primary carrier transit time for the initial carriers created from the absorption of the light to reach the multiplication layer. The second is the secondary carrier transit

time for the multiplied carriers, of opposite polarity to the primary carriers, to retrace the steps of the primary carriers, continuing on until the appropriate contact is reached. Note that the time for avalanche buildup is not part of the transit time, and is considered separately below. Define τ_h as the transit time for holes and τ_e as the time for electrons. In a III-V APD, the holes initiate the ionization, so the holes are the primary carriers which travel from the InGaAs to the avalanche region and the electrons are the secondary carriers which travel from the avalanche region through the InGaAs to the n boundary of the depletion region. It is assumed that the device is under sufficient reverse bias so that this full distance is under depletion.

The transit times are given by the ratio of the distance $d_{e,h}$ traveled by the carrier (may be unequal) to the "saturated" velocity of the appropriate carrier, i.e.

$$\tau_e = \frac{d_e}{v_e} \quad (2.5.7)$$

for electrons, with τ_h similarly defined for holes. As a first approximation, the average travel distance may be used for the primary carriers. This is determined from the average location of light absorption, which is given in Eqns. (2.5.5) and (2.5.6). Various saturated values of electron and hole velocities are used by different authors. Some of these results are given in Table 2.5.1. The experimental values range from 20-60 $\mu\text{m}/\text{ns}$ for holes and 50-70 $\mu\text{m}/\text{ns}$ for electrons, with calculated values somewhat higher. In this work, E-field independent values of 70 and 50 $\mu\text{m}/\text{ns}$ (7×10^6 and 5×10^6 cm/s) are used for the electron and hole velocities, respectively unless stated otherwise [81]. Then the associated bandwidth depends on the type of modulation used and pulse response assumed. For the standard assumptions

2. Theory

of sinusoidal modulation and a rectangular pulse, the bandwidth is given by [1]

$$f_e = \frac{0.443}{\tau_e} \quad (2.5.8)$$

for electrons, with f_h similarly defined for holes.

Table 2.5.1 Literature values of electron and hole velocities.

Author	Ref.	carrier (e or h)	details	velocity ($\mu\text{m}/\text{ns}$)
Campbell	[109]	h	assumed	60
Brennan and Hess	[51]	h	calculated, InP	100
Brennan	[115]	h	calculated, InGaAs	70
Shiba	[96]	h	inferred from data	20
Hill	[114]	h		
Hollenhorst	[81]	h	assumed	50
Windhorn	[59]	e	measured, InP	80, declining
Windhorn	[116]	e	measured, InGaAs	70, declining
Hollenhorst	[81]	e	assumed	70

In reality, mobility measurements of bulk material show that the carrier velocity is a function of both T and the E -field [59]. The reduction of velocity with increasing T is due to increased lattice scattering. The origin of the E -field dependence is the existence of two conduction band minima within the Brillouin zone of differing effective masses. Effective mass is inversely proportional to the second derivative of energy with respect to wavevector within the Brillouin zone. As the E -field increases from zero, the velocity increases linearly at first within the first minimum, until some electrons have sufficient energy to transfer into the second minimum, of

2. Theory

higher effective mass than the first. At this point the velocity reaches a peak and subsequently decreases owing to the higher effective mass. This is the Gunn effect. In principle, at sufficiently high E-field, all the electrons are in the second minimum and the velocity saturates. Since the E-fields under consideration in the multiplication region of an APD are tens of times greater than the peak velocity E-field, it would seem a reasonable assumption that a constant velocity could be used for APD transit time estimations. However, experimental work shows a continuing decrease of electron velocity with increasing E-field, which may be due to increased scattering. For detailed calculations involving either layers of non-zero doping (and therefore non-constant E-field) or different gain values (for which the E-fields vary), the velocity variation with respect to E-field must be included.

2.5.3.2 RC

There is a fundamental limit to bandwidth owing to the capacitive transient charging effects which arise from the capacitance C associated with the depletion region of the device and the combined resistance R of the load and the device. Define τ_{RC} as the RC time constant. Then τ_{RC} and the associated -3 dB electrical bandwidth f_{RC} are given by [1, 11, 12]

$$\tau_{RC} = RC \quad (2.5.9)$$

$$f_{RC} = \frac{1}{2\pi\tau_{RC}} \quad (2.5.10)$$

2.5.3.3 Avalanche buildup

In an APD there is a buildup time τ_{aval} associated with the avalanche gain process which is given by $M\tau_n$ for electron injection into the multiplication region and $M\tau_p$ for hole injection into the multiplication

region. For single carrier ionization, there is only the transit time through the multiplication layer to consider as the multiplication is in one direction only. However for dual carrier ionization there is a back and forth multiplication process. This feedback effect takes more time than simple travel through the multiplication region. Not surprisingly, the greater the feedback, the greater the associated τ_{aval} , as discussed in Section 2.1. For sinusoidal modulation and a rectangular pulse response, the associated bandwidth is given by [1]

$$f_{aval} = \frac{0.443}{\tau_{aval}} \quad (2.5.11)$$

This effect was first studied for Si APDs by Emmons [76], and will be discussed in more detail later.

2.5.3.4 Hole trapping

In an heterostructure APD there is the possibility of delay caused from an impediment to either carrier crossing a heterojunction, which is described by thermionic emission. Such an effect would be related to the abruptness of the heterojunction, the barrier height, the temperature and the effective mass of the carrier. Since the valence band discontinuity in the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ materials system is 60% valence band and 40% conduction band and the effective mass of holes is an order of magnitude greater than that of electrons, it is more likely that this would occur for holes than for electrons, a phenomenon known as hole trapping. Define $\tau_{pile-up}$ as the time constant associated with this effect.

2. Theory

This effect was first observed by Matsushima [105, 106] and Forrest [107, who recommended the use of a graded layer between the InP and InGaAs. This has the effect of smoothing the barrier between the different materials.

2.5.3.5 Diffusion tails

Until this point it has been assumed that the device is fully depleted. Thus the associated transit times are that of drift. For any undepleted portion of the device, carrier transport must be by diffusion rather than drift, with a resultant slowing of device response. An oscilloscope trace of a device with both drift and diffusion present would show both a "fast" and "slow" component. The fast component is due to drift and the slow component is due to the diffusion from the undepleted portion of the InGaAs, often referred to as a "diffusion tail".

2.5.3.6 Overall photodetector bandwidth

The basic limitations to the speed of a p-i-n photodetector are RC and the transit time of the primary carriers from the point at which they are generated from light absorption to the boundary of the depletion region at which they are collected, assuming that the absorption layer is fully depleted.

The BW of a p-i-n photodetector is limited both by f_{RC} (where the load resistance is typically $50\ \Omega$ plus whatever series resistance is associated with the device) and by $f_{e,h}$. Both these bandwidths have been discussed above. Since the electrons and holes are drifting simultaneously towards the appropriate contact under the influence of the E-field, the transit time bandwidth will be dominated by the lower of f_e or f_h . To precisely quantify the effect of the capacitance and the transit time together, the pulse response must be considered. While many authors gloss over the precise method of

combination of these two effects, this subject has been treated elsewhere by multiplying the two transfer functions [109] and has also been the object of some recent attention [97, 98]. Somewhat less accurately but much more intuitively, the capacitive time constant and the transit time constant are independent, so the overall pin time constant (inverse of BW) is well approximated by the sum of squares of RC and transit time constants [15]. That is,

$$\begin{aligned}\tau_{\text{pin}}^2 &= \text{Max}(\tau_h, \tau_e)^2 + \tau_{\text{RC}}^2 \\ f_{\text{pin}}^{-2} &= \text{Min}(f_h, f_e)^{-2} + f_{\text{RC}}^{-2}\end{aligned}\tag{2.5.12}.$$

where $\text{Max}(a', b')$ is the maximum of a' or b' and $\text{Min}(a', b')$ is the minimum of a' or b' .

For the APD there are two different carrier transit times arising from primary carriers (holes) traveling to and secondary carriers returning from (electrons) the multiplication region. In addition, there is an avalanche buildup time associated with the multiplication process, proportional to the gain, which is characterized by the gain-bandwidth product (GBW); the higher the better. There is potentially further delay associated with hole trapping at the InGaAs/InP interface, which is avoided by the use of a grading layer. The actual APD frequency response is a complicated function of all these processes, too detailed to discuss here but given in other treatments of this specific problem. For example, initially, Campbell treated this problem by multiplying all the transfer functions together, which is approximated by the sum of squares [109]. Later, Campbell summed the "effective transit time factors" for the primary and secondary carriers, multiplying the resultant transfer function by the RC transfer function [97]. The transit time factors are

2. Theory

complicated expressions. Recently Hollenhorst [98], Kahraman [99] and Roy [100] have developed complicated matrix expressions for arbitrary structures, a detailed presentation of which is beyond the scope of this work. For estimation purposes in this work, the RC time constant will be treated as uncorrelated from the rest of the time constants, the primary hole transit, hole trapping, avalanche buildup and secondary electron transit events will all be treated in series, one following the other. This corresponds to a physical picture of an initial photon being absorbed, following the hole through to the multiplication region and following the electrons back to the n-contact, while treating the RC time constant as independent process. In this case,

$$\tau_{APD}^2 = (\tau_{aval} + \tau_h + \tau_e + \tau_{pile-up})^2 + \tau_{RC}^2 \quad (2.5.13).$$

$$f_{APD}^{-2} = (f_{aval}^{-1} + f_h^{-1} + f_e^{-1} + f_{pile-up}^{-1})^2 + f_{RC}^{-2}$$

If the absorption layer is only partially depleted the fall time of a photocurrent pulse exhibits both a "fast" and a "slow" component. As discussed above, the fast component is limited by the combination of effects described so far, associated with the drift in the E-field from the depleted portion of the InGaAs (desired); the slow component is related to carrier diffusion. This situation could arise in an APD if too much charge sheet density σ were present or insufficient reverse bias were applied.

As only the avalanche buildup bandwidth limitation is a direct function of the multiplication M , it is expected that avalanche buildup should dominate the APD bandwidth at very high gain. As τ_{aval} is proportional to the gain of the device [76], the quantity f_{aval} represents the gain-bandwidth product (GBW) of the APD if the remaining components can be neglected. The remaining components determine the intrinsic bandwidth of the APD —

2. Theory

i.e. without the avalanche buildup time. Unfortunately, it is difficult to decouple the RC response time from the transit times or hole pile-up time. However, a conservative (i.e. lower bound) estimate of the GBW product may be derived from assuming a sum-of-squares relationship among all the response times:

$$\begin{aligned}\tau_{\text{APD}}^2 &= \tau_{\text{aval}}^2 + \tau_{\text{intrinsic}}^2 \\ f_{\text{APD}}^{-2} &= \left(\frac{\text{GBW}}{M}\right)^{-2} + f_{\text{intrinsic}}^{-2}\end{aligned}\tag{2.5.14}$$

where $\tau_{\text{intrinsic}}^2$ is the sum of squares of all the appropriate characteristic times [29, 30].

All the different limitations on the APD bandwidth can be summarized into an expected bandwidth as a function of gain, as illustrated in Fig. 2.5.1 [21]. At low gain, there are limitations due to hole trapping or incomplete depletion, which arise from the relatively low E-field at the heterointerface E_{het} . At modest gains beyond which the InGaAs is fully depleted and hole trapping is not an issue, the bandwidth is dominated by the transit times and RC, and the maximum bandwidth for the device is achieved. At high gains, the avalanche buildup time becomes increasingly dominant, and this portion of the f-M curve is limited by the GBW.

2. Theory

Figure 2.5.1 Illustration of expected bandwidth vs. gain for an APD

2.5.4 Quantum efficiency – bandwidth product

The quantum efficiency depends on the thickness of absorbing material, usually (and in this case) InGaAs. For a device with no reflections, the quantum efficiency (QE) is given by $1 - \exp(-t_{\text{InGaAs}}/l_a(\lambda))$. A thinner absorption region trades off faster transit times for lower QE. In the limit of no capacitance, for a thin absorption layer, the QE-BW product for a p-i-n photodiode assuming sinusoidal modulation is given approximately by

$$\text{QE} \cdot \text{BW} = (1 - \exp(-t_{\text{InGaAs}}/l_a(\lambda))) \cdot 0.443 \frac{\text{Min}(v_e, v_h)}{(t_{\text{InGaAs}}/2)} \quad (2.5.15)$$

In this case it is assumed that the p-n junction is located at the upper InP/InGaAs heterointerface, so that the depletion region is of thickness t_{InGaAs} . The factor of 2 in the denominator, generally not seen in other treatments of this subject [160], is included here because on average a given carrier must travel only half the thickness of the layer. More will be said on this point later. For a sufficiently thin absorption layer, the absorption may be approximated as uniform within the layer and this approximation should be valid. It is also assumed that the holes travel more slowly than the electrons, and therefore are the carrier which limit the frequency response. Using the approximation $\exp(-x') \approx 1 - x'$, assuming $t_{\text{InGaAs}}/l_a \ll 1$,

$$\begin{aligned} \text{QE} \cdot \text{BW} &\approx \frac{t_{\text{InGaAs}}}{l_a(\lambda)} \cdot 0.443 \frac{v_h}{(t_{\text{InGaAs}}/2)} \\ &= \frac{0.886 \cdot v_h}{l_a(\lambda)} \end{aligned} \quad (2.5.16)$$

The $\tau \approx \text{QE} \cdot \text{BW}$ product is independent of t_{InGaAs} . The electron and hole velocities were discussed above. Using Humphrey's data [25], $1/l_a$ is given by 1.15 and $0.68 \mu\text{m}^{-1}$ for 1.3 and $1.55 \mu\text{m}$ wavelength respectively, assuming $n\text{-InGaAs}$. This corresponds to $l_a(1.3 \mu\text{m}) = 0.87 \mu\text{m}$ and $l_a(1.55 \mu\text{m}) = 1.47 \mu\text{m}$. Using $50 \mu\text{m}/\text{ns}$ for the saturated hole velocity [81], the $\text{QE} \cdot \text{BW}$ limit is 51 and 30 GHz for 1.3 and $1.55 \mu\text{m}$ wavelength, respectively.

Note that these estimates ignore any reflections which could provide transmitted light with a second pass through the absorption region. This would have the effect of increasing the $\text{QE} \cdot \text{BW}$ limit.

A recent development is that of resonant absorption, in which a Fabry-Perot cavity is formed with a thin layer of absorption material [117, 118]. This allows enhancement of the QE in the vicinity of a particular wavelength (e.g. 1.3 or $1.55 \mu\text{m}$ but not both) far above that of merely of a double pass. This development holds promise for reduced transit times for a given QE than was previously possible, and thus an increased effective $\text{QE} \cdot \text{BW}$.

2.5.5 Optimum depletion layer thickness and maximum bandwidth for devices of finite diameter

While the $\text{QE} \cdot \text{BW}$ limit is often quoted in the literature as a fundamental limitation on device performance, it is impossible to make a photodiode of zero capacitance, which is necessary for the $\text{QE} \cdot \text{BW}$ limit to be valid. In reality, the photodiode must be of finite dimensions, both diameter D and thickness t . In this section analytic expressions for the bandwidth of a p-i-n photodiode are presented, taking into account the finite size. The optimum thickness for a given device diameter and the maximum bandwidth are derived. Extrapolations are made to the fundamental

2. Theory

limitations of APD performance. These analytical results are not available in the literature.

Consider a reverse biased photodiode of diameter D and depletion layer thickness t , as illustrated in Fig. 2.5.2. The RC time constant is given from Eqn. (2.5.10) by

$$\tau_{RC} = \frac{\epsilon R \pi D^2}{4t} \quad (2.5.17)$$

Therefore the frequency response of the p-i-n photodiode is given from Eqns. (2.5.7), (2.5.8) and (2.5.12) by

$$C = \frac{\epsilon \pi D^2}{4t}$$

Figure 2.5.2 Illustration of a depletion region of diameter D and thickness t .

2. Theory

$$\begin{aligned}
 f_{\text{pin}}^{-2} &= f_{\text{RC}}^{-2} + f_h^{-2} \\
 &= \left(\frac{\epsilon R \pi^2 D^2}{2} \right)^2 \frac{1}{t^2} + \left(\frac{1}{f_h \tau_h v_h} \right)^2 t^2
 \end{aligned} \tag{2.5.18}$$

where the quantity $f_h \tau_h$ is 0.443 for strong absorption, 0.886 in Eqn. (2.5.16).

For fixed D , it is worth considering the variation of frequency response as a function of t . This may be done by considering Eqn. (2.5.18), which is of the form

$$g(t) = A \frac{1}{t^2} + B t^2 \tag{2.5.19}$$

where g is an arbitrary function. The derivatives with respect to t are

$$\frac{dg}{dt} = -2A \frac{1}{t^3} + 2B t \tag{2.5.20}$$

$$\frac{d^2g}{dt^2} = 6A \frac{1}{t^4} + 2B > 0 \tag{2.5.21}$$

where A and B are constants. Since the second derivative is greater than zero, g is a minimum and therefore $1/g$ is a maximum at the optimum t . The optimum t , t_{opt} , may be found by setting the derivative in Eqn. (2.5.20) to zero and solving for t :

$$t_{\text{opt}}^2 = \sqrt{\frac{A}{B}} \tag{2.5.22}$$

Note that this is the same result which is obtained if Eqn. (2.5.19) is differentiated with respect to t^2 instead of t . Substitution of Eqn. (2.5.22) into Eqn. (2.5.19) gives

2. Theory

$$g_{\text{opt}} = 2\sqrt{AB} \quad (2.5.23)$$

In this case, the function g is actually f_{pin}^{-2} , and so the limiting depletion layer thickness t_{opt} where the bandwidth is at a maximum occurs when the RC and transit time BWs are equal, and is given as $f_{\text{opt}} = \frac{f_{\text{RC}}}{\sqrt{2}} = \frac{f_h}{\sqrt{2}}$

Completing the derivation,

$$t_{\text{opt}} = \sqrt{\frac{f_h \tau_h \pi^2 v_h \epsilon R}{2}} D \quad (2.5.24)$$

$$\begin{aligned} f_{\text{opt}} &= \frac{1}{\sqrt{2 \sqrt{\left(\frac{\epsilon R \pi^2 D^2}{2}\right)^2 \left(\frac{1}{f_h \tau_h v_h}\right)^2}}} \\ &= \sqrt{\frac{f_h \tau_h v_h}{\epsilon R \pi^2}} \frac{1}{D} \end{aligned} \quad (2.5.25)$$

Eqns. (2.5.24) and (2.5.25) express the fundamental limitations on bandwidth achievable for a practical p-i-n photodiode of finite dimensions.

The bandwidth is obviously a function of t (Eqn. 2.5.18). One of the important implications of these results is that an optimal bandwidth can be predicted which depends only on fundamental device parameters and the diameter D , with no explicit dependence on t .

It is worth discussing the pulse response as a function of the details of the light absorption. As mentioned above, the condition for transit time limited bandwidth, assuming equal carrier velocities, is usually stated as $f_h \tau_h = 0.443$. This condition assumes a rectangular pulse response, which would be the case for strongly absorbed light which is absorbed on the p^+ side

2. Theory

of the depletion region, causing electrons to flow the entire distance to the n-side contact. However, the QE-BW product is of interest only for the case of thin absorption layers, for which the absorption is assumed to be uniform. For this case, a triangular pulse response would result for equal carrier velocities, and a two-tiered triangular pulse response would result for different carrier velocities, with the two different slopes corresponding to the two different velocities. Conradi has shown [161] that the Fourier transform of this pulse response is given by

$$I(\omega) = \frac{(1+b^2) - b^2 \cos(\omega\tau_e/b) - \cos(\omega\tau_e)}{(\omega\tau_e)^2} - i \frac{(1+b) - b^2 \sin(\omega\tau_e/b) - \sin(\omega\tau_e)}{(\omega\tau_e)^2} \quad (2.5.26)$$

where $b=v_h/v_e$. This may be compared with the Fourier transform of the pulse response for strongly absorbed light

$$I(\omega) = \frac{\sin(\omega\tau_e/2)}{(\omega\tau_e/2)} \quad (2.5.27)$$

The electrical frequency response, given by $|I(\omega)|^2$, is plotted against $f\tau_e$ in Fig. 2.5.3. Five cases of uniformly absorbed light are illustrated with solid lines, namely $v_h/v_e = 0.2, 0.4, 0.6, 0.8$ and 1 . The dotted line corresponds to strongly absorbed light, which is the underlying assumption behind the condition $f_h\tau_h=0.443$. Although there are differences in the shapes, it is seen that the strongly absorbed light corresponds fairly closely to the case $v_h/v_e = 0.7$. For each trace $f_e\tau_e$ corresponds to the $f\tau_e$ value where $|I(\omega)|^2/|I(0)|^2 = 1/2$ ($=-3.01$ dB). The dependence of $f_e\tau_e$ on v_h/v_e is shown in Fig. 2.5.4. It is seen that as v_h/v_e increases, so too does $f_e\tau_e$. The condition $f_e\tau_e = 0.443$,

2. Theory

corresponding to strongly absorbed light, is met for $v_h/v_e = 0.7$. For any particular v_h/v_e , the value of $f_e\tau_e$ may be used in Eqns. (2.4.25) and (2.4.25) to obtain appropriate t_{opt} and f_{opt} values, or substituted for 0.886 in Eqn. (2.5.16) to obtain a $QE \cdot BW$ value.

These analytical results may be compared with the theoretical results of Bowers [16] for the -3 dB electrical bandwidth of a p-i-n photodiode as a function of t . Bowers used a complicated expression for the transit time transfer function based on pulse response [95], combined with an unspecified RC transfer function in an unspecified manner. The results are given in Table 2.5.2. For all the calculations, $R=50\Omega$ is assumed.

Table 2.5.2 Optimum bandwidths and thicknesses for p-i-n photodiodes.

D (μm)	t_{opt} (μm) (Bowers)	t_{opt} (μm) (Note 1)	t_{opt} (μm) (Note 2)	f_{opt} (GHz) (Bowers)	f_{opt} (GHz) (Note 1)	f_{opt} (GHz) (Note 2)
10	0.3	0.28	0.32	77	69	77
20	0.6	0.55	0.65	38	35	38
40	1.2	1.1	1.3	19	17	19

1. this work, $v_h=50 \mu m/ns$, $v_e=70 \mu m/ns$, $\epsilon_r=12.9$ (previous assumptions)
2. this work, $v_h=65 \mu m/ns$, $v_e=70 \mu m/ns$, $\epsilon_r=13.6$ (Bowers' assumptions)

The values from Bowers [16] are the best values which could be determined from estimating the maximum bandwidth and optimal thickness from the figures by hand, and not actual numbers supplied by Bowers. It is evident that there is very good agreement between this work and that of Bowers, especially when Bowers' assumptions v_h , v_e and ϵ_r are used. Thus the dependencies indicated in Eqns. (2.5.24) and (2.5.25) are confirmed using an exact pulse response calculation. The $(t_{opt}/D, f_{opt} \cdot D)$ values from Eqns. (2.5.24) and (2.5.25) are (0.0265, 666 GHz- μm) for the assumptions in Note 1 and (0.0323, 768 GHz- μm) for the assumptions in Note 2.

2. Theory

Figure 2.5.3 Electrical frequency response vs. $f\tau_e$ for transit time limited response of a p-i-n photodiode. The solid lines correspond to uniform absorption with the v_h/v_e values indicated, which range from 0.2 to 1. The dashed line corresponds to strong absorption.

Figure 2.5.4 Values of $f_{-3dB}\tau_e$ vs. v_h/v_e for uniform absorption. This function corresponds to the -3 dB point of the traces in Fig. 2.5.3.

2. Theory

Eqns. (2.5.24) and (2.5.25) are valid for a p-i-n photodiode. For an APD there are two transit times, so the maximum BW is smaller owing to increased transit time. On average the primary carriers must travel only half the layer thickness, but the secondary carriers must travel the entire layer thickness, which would correspond to a bandwidth reduction of roughly two-thirds, assuming equal carrier velocities. However, the secondary carriers (electrons) are faster than the primary carriers (holes) by a factor of 1.05-3, depending on which literature estimates are used.

Above and beyond the electron velocity discussed here, additional bandwidth can in principle be had by tailoring the E-field in the InGaAs to the field corresponding to the peak velocity in InGaAs ($\sim 3 \text{ V}/\mu\text{m}$). This field must be controlled to within a tolerance ΔE of $\sim 1 \text{ V}/\mu\text{m}$ (the approximate width of the velocity peak in InGaAs on the v-E curve) for maximum benefit. The upper limit on background doping level in the InGaAs corresponding to a given ΔE change within the InGaAs thickness may be numerically represented by ($\epsilon_r = 12.9$ for InGaAs) as

$$N_D = \frac{10^{15} \text{cm}^{-3}}{1.40 \text{ V}/\mu\text{m}^2} \frac{\Delta E}{t_{\text{InGaAs}}} \quad (2.5.30)$$

For $t_{\text{InGaAs}} = 1.3 \mu\text{m}$, Eqn. (2.5.30) implies a maximum doping level of $5 \times 10^{14} \text{cm}^{-3}$. This represents stringent, but achievable (state of the art) requirements from the crystal growth capability [19, 119-121].

2.5.6 Practical bandwidth limitations for photodiodes

In the absence of lenses, a practical photodetector must be at least as large as the fiber core in order to receive the light. Assuming a fiber core

2. Theory

diameter of $10\text{ }\mu\text{m}$, the maximum BW of a p-i-n photodiode is estimated as 89 GHz and the maximum BW of an APD is estimated as 46 GHz assuming an electron velocity of $70\text{ }\mu\text{m/ns}$ and a hole velocity of $50\text{ }\mu\text{m/ns}$, using Eqn. (2.5.27). These estimates ignore both the unavoidable parasitic capacitances and the avalanche buildup time, which would reduce f_{opt} . The APD BW is likely to be limited by these effects. In particular, prior to this work no GBW has been reported higher than 90 GHz, which means that at a gain of 10 the APD bandwidth is limited to 9 GHz by GBW alone. However, by improving GBW it is possible to improve the BW of the APD. There is no fundamental limitation on the APD BW which should prevent operation at a bit rate of 10 Gb/s or even beyond.

2. Theory

3. Evolution of APD structures

In this section, the evolution of APD structures from homogeneous Si APDs through to separate absorption, grading and multiplication InGaAsP APDs is reviewed.

As discussed in Section 2.2, edge breakdown occurs prior to central area breakdown for a planar APD. This phenomenon is typically avoided in planar APDs by the use of one or more guard ring(s). How they work is described in more detail in Section 4.1.2. The remainder of this section is devoted to the evolution of the central area of the device.

3.1 Homogeneous APDs

In the early stages of APD research, the main material of interest was Si. This is not surprising, as research concerning Si semiconductors dates back to the 1940's. In this case, the light was absorbed and multiplication took place in the same material. Si APDs work very well, largely because there is a large disparity between the ionization coefficients – i.e. k can be made very low [22]. This has positive implications for noise performance of these devices. As interest in wavelengths beyond the practical range of Si absorption wavelengths increased, it became desirable to investigate III-V materials, and in particular InGaAsP lattice matched to InP. Unfortunately, k cannot be made as low as for Si in these materials [6, 11, 12, 58-64, 73, 74], which places limitations on noise performance not found in Si. When investigating InP-based materials, it seemed natural to start with InGaAs, the composition with lowest bandgap energy for this materials system. Unfortunately, the tunneling leakage current is very high for a homogeneous InGaAs APD [38-43] as would be expected from Section 2, and other solutions had to be found.

3.2 Separate absorption and multiplication (SAM) for III-V APDs

3.2.1 Concept

It is apparent that for an APD, there is competition between a low E_g to allow absorption to occur and a high E_g to inhibit tunneling while allowing impact ionization. These considerations are not present in a p-i-n diode. For a p-i-n diode, an unintentionally doped (typical background doping levels are of the order of 10^{15} cm^{-3} [19, 119-121]) 2-3 μm thick InGaAs layer can usually be depleted with less than 10V reverse bias, resulting in a peak E-field less than 10 V/ μm . For significant multiplication to occur, the maximum E-field must be much larger than this. However, because E_g is so small for InGaAs, significant tunneling occurs at lower E-fields; estimates of the E-field at the onset of tunneling range from 10-25 V/ μm [38-42]. Therefore InGaAs is not a suitable material for avalanche multiplication.

3.2.2 Charge layer requirement

The low bandgap energy of InGaAs, however, provides a clue for designing III-V APDs. Other materials in the InP materials system are of higher bandgap energy, although they do not absorb the light of wavelengths 1.3 and 1.55 μm . As illustrated in Fig. 3.1, by arranging the epitaxial layers in such a way that light can be absorbed in the InGaAs where the E-field is kept relatively small, generating carriers which travel through the depletion region to another region of higher E-field and E_g where multiplication can occur, the tunneling leakage current problem can be solved. Such a structure, containing separate absorption and multiplication (SAM), is considered necessary in order to realize high performance III-V APDs [1-6, 11, 12, 38-41]. The operation of the device may be explained using Gauss' Law.

3. Evolution of APD structures

Fig. 3.1 Layer sequence for a SAM APD.

The integrated space-charge density (or charge sheet density) σ is defined by

$$\sigma = \int_{x_1}^{x_2} N_D(x) dx \quad (3.2.1)$$

where x_1 and x_2 are the limits of integration and N_D is the doping level at x . These quantities are related to the E-field drop ΔE between x_1 and x_2 by Gauss' Law

$$\nabla \cdot \mathbf{E} = N_D / \epsilon \quad (3.2.2)$$

3. Evolution of APD structures

$$\sigma = \epsilon \Delta E \quad (3.2.3)$$

where ϵ is the dielectric constant. Eqn. (3.2.2), which is more general, is true for any coordinate system and is one of Maxwell's equations. Eqn. (3.2.3) is a solution of Eqn. (3.2.2) for the one-dimensional (1-D) case which assumes Cartesian coordinates.

Since literature values for the E-field to support avalanche breakdown in InP range from 45 V/ μm up [39-41] and literature values for the E-field at the onset of tunneling in InGaAs range from 10 V/ μm up [38-41], the E-field must drop by approximately 35 V/ μm from the avalanche region to the InP/InGaAs heterointerface to ensure there is no tunneling in the InGaAs. The key requirement for the SAM APD is that a specified σ of n-type doping, usually given as $2.2\text{-}2.7 \times 10^{12} \text{ cm}^{-2}$ in the literature [39-41], be present between the avalanche region and the InP/InGaAs heterointerface to bring about the required E-field drop. If there is insufficient σ then tunneling in the InGaAs could occur. If there is too much σ then at the point of InP breakdown, the InGaAs would only partially depleted, and in the extreme case undepleted. If there is no depletion in the InGaAs, there is no photoresponse, because the carriers generated from the absorption of light in the InGaAs will be unable to overcome the barrier of the heterointerface band discontinuity. If the InGaAs is partially depleted there is no loss in the DC photoresponse, but the high frequency response suffers as discussed below. There are many combinations of $N_D(x)$ which will bring about a given σ . However, the E-field drop is independent of the details of the doping profile except insofar as the breakdown field depends on the doping level, about which more is said later.

In the higher bandgap half of the $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ materials, system k is least for InP [61]. Thus the fortuitous situation exists where InP is the material

3. Evolution of APD structures

of both highest E_g and lowest k , and thus suitable as a multiplication layer without unduly compromising other characteristics of the device performance. In this case $\beta > \alpha$, so it is necessary to inject holes into the avalanche region. This differs from the case discussed so far of Si APDs. In order to achieve hole injection into the InP it is necessary to arrange the layers as shown in Fig. 3.1, with the InGaAs adjacent to the n-contact and the InP adjacent to the p-contact.

3.3 Separate absorption, grading and multiplication (SAGM) APDs

InP has the highest E_g in the InP materials system which is beneficial for multiplication and InGaAs has the lowest E_g in the InP materials system which is beneficial for absorption. However, if the InP multiplication layer is adjacent to the absorbing InGaAs layer, the largest possible valence and conduction band discontinuity within the InP materials system exists at the heterointerface. The result is a hole pile-up at the heterointerface related to thermionic emission which degrades the speed of the device [1, 12, 105-113]. In order to avoid this problem, a grading region (Fig. 3.2) consisting of one or more epilayers can be used. This region grades the composition from InGaAs to InP, lowering the band discontinuities and therefore reducing or eliminating the hole pile-up effect. This refinement, consisting of separate absorption, grading and multiplication (SAGM), is generally used in III-V APDs.

There are different possible ways to form a grading layer. The first tried, and the simplest, is to use a composition of InGaAsP intermediate between InP and InGaAs [105-108]. More than one grading layer is often used

to provide a more gradual ramp between the InP and InGaAs, with three layers being common in the mid to late 1980's [109-110] and even more layers or continuous grading seen in the 1990's [111, 112]. Another possibility, first suggested by Capasso, is to use InP and InGaAs alternating in graded thicknesses to approximate a ramp from InP to InGaAs. This type of grading layer is referred to as "pseudoquaternary" [113].

Fig. 3.2 (a) Layer sequence and (b) E-field vs. x for a SAGM APD. The p-n junction must be accurately positioned or else tunneling in the InGaAs (too deep) or incomplete depletion or hole trapping (too shallow) can occur.

3. Evolution of APD structures

3.4 Constraints in fabrication

E-field

An E-field profile of an SAGM APD juxtaposed with the appropriate layers is shown in Fig. 3.2. The p-n junction must be accurately positioned. If the p-n junction position is too deep, there is less charge between the p-n junction and the InGaAs, with the result that the E-field in the InGaAs is higher. In this case, tunneling in the InGaAs could occur. If the p-n junction position is too shallow, there is less charge between the p-n junction and the InGaAs, with the result that the E-field in the InGaAs is higher. In this case hole trapping or incomplete depletion can occur. This has direct implications on the tolerances in fabrication of these devices. All the arguments in this section apply as well to the SAM APD.

Tolerances in fabrication

In the active region of the device the correct charge sheet σ is provided by accurate control of both the doping level in the InP (N_D) and the distance between the p-diffused junction and the upper InGaAs heterointerface (x_d). There are many different combinations of N_D and x_d for any given σ . In this case, the correct choice depends on how well N_D and x_d can be controlled. Neglecting the breakdown E-field dependence on N_D , a first order differential expansion can be performed to estimate this uncertainty:

$$d\sigma = N_D dx_d + x_d dN_D \quad (3.4.1)$$

Dividing through,

$$\frac{d\sigma}{\sigma} = \frac{dx_d}{x_d} + \frac{dN_D}{N_D} \quad (3.4.2).$$

I. e., the relative error in σ is the sum of the relative errors in N_D and x_d .

If the thickness of the upper InP layer can be measured prior to the p-diffusion, the uncertainty in x_d is merely the uncertainty in the p-diffused junction depth x_j , which is typically 0.1 μm , the limit of the Zn diffusion process reproducibility. For the growth reactors used in this work the uncertainty in N_D is 10^{16} cm^{-3} , which is mainly a limitation on the linearity of mass flow controllers and the concentration of dopant gas in the source cylinder. Using these parameters, and assuming the desired σ is $2.5 \times 10^{12} \text{ cm}^{-2}$, the appropriate choice to minimize the error in σ is $N_D = 5 \times 10^{16} \text{ cm}^{-3}$ and $x_d = 0.5 \mu\text{m}$, which results in an uncertainty in σ of $1.0 \times 10^{12} \text{ cm}^{-2}$ (actually $7.1 \times 10^{11} \text{ cm}^{-2}$ if sum of squares is used). This N_D and x_d are representative of what is seen in the literature, but note that the uncertainty in the charge sheet density is greater than tolerable to fall within the above mentioned range $2.2\text{--}2.7 \times 10^{12} \text{ cm}^{-2}$. Hollenhorst finds that for $N_D = 6 \times 10^{16} \text{ cm}^{-3}$ it necessary to control the location of the p-n junction to within $\pm 200 \text{ \AA}$ [81, 82], which is difficult to achieve in practice. Even if there were no uncertainty in N_D , the uncertainty in x_d alone is enough to cause an uncertainty in σ of $0.5 \times 10^{12} \text{ cm}^{-2}$, or 20% of the total charge. In this case, if N_D were reduced and x_d were increased to reduce the uncertainty in σ the avalanche length would become longer and the GBW of the device degraded. It is partially for these reasons that it is so difficult to make a high performance SAGM APD such as that in Fig. 3.2. Furthermore, as the doping increases beyond 10^{16} cm^{-3} , impurity scattering becomes an increasingly important contribution to the mean distance a carrier travels before collision, with the result that the

3. Evolution of APD structures

ionization coefficients themselves decrease with increasing doping. Thus E_B as well increases because of both the decreased multiplication region length and the decreased ionization coefficients. Thus unintentional fluctuations in doping can alter the quantity of σ required in the SAM APD. Researchers are becoming increasingly more aware of the need to design an APD which is not so difficult to fabricate [27, 122-124].

3. Evolution of APD structures

4. Separate absorption, grading, charge and multiplication (SAGCM) APDs

4.1 Candidate device structures

Although it is possible to show that certain performance is theoretically possible from a device, obtaining high performance experimental devices has proven difficult for APDs, and is the object of considerable worldwide research. In this section, some candidate APD structures are presented. A new device structure is presented which can in principle alleviate some of the difficulties other workers have encountered.

4.1.1 Planar and mesa geometries

Two principal standard geometries exist for an APD: mesa or planar, as illustrated in Figs. 4.1.1 [11, 12]. The mesa configuration is simpler to design and fabricate. In particular, there is no curvature to the p-n junction. In the planar geometry, a p-diffusion is usually performed through a window in a mask, resulting in a planar p-n junction which is curved at the edge of the p-diffused area. The E-field under reverse bias grows more rapidly under a curved than a flat p-n junction, which leads to premature breakdown in the edge region of the device [1-6, 23, 26, 87-91]. Therefore special and sometimes elaborate precautions, most commonly involving guard rings [1-6, 81, 82, 124, 125], must be taken in the planar geometry to reduce the E-field at the device periphery. However, even for a mesa structure, there is no guarantee that the E-field at the edge of the mesa will not be higher than in the central portion of the device unless the mesa is perpendicular to the epitaxial layers. For other angles the details need careful examination because of the fringing edge fields in the mesa [126].

Figure 4.1.1 Illustration of (a) mesa and (b) planar geometries.

There is a considerable amount of initial performance data available on both mesa [12, 38-40, 58-64, 72-74, 105-107, 109, 110] and planar [4, 21, 80, 81, 108, 96, 122-126] geometries. In view of the fabrication difficulties involved with the planar geometry, early work typically employed the mesa geometry. In fact, it is only recently (1988) that planar devices have matched the initial performance of mesa devices [81].

In the mesa geometry, a p-n junction is exposed in planes other than the [100] crystal planes, which are difficult to passivate. Therefore mesa devices are considered to be of poorer reliability than planar devices [129]. Also planar devices are more easily integrated with other devices. To be of any practical use at all the device must be reliable. This leaves the options of learning how to effectively passivate a mesa or designing a planar device around the p-n junction curvature considerations. In this work the planar approach has been taken, and many other workers have come to a similar

4. Separate absorption, grading, charge and multiplication APDs

conclusion. All the commercially available APDs to date that this worker is aware of (e.g. Fujitsu, AT&T, Hitachi, Epitaxx) use a planar geometry.

4.1.2 Standard solution with guard rings

Having decided on a planar APD, the problems to be solved are how to distribute the doping in the charge layer and how to prevent premature edge breakdown. The most common solution in the literature is to use a guard ring in a structure such as is indicated in Fig. 4.1.2 [1, 4, 81, 82, 125, 126]. The basic idea for the guard ring functionality is that the maximum E-field for a given applied voltage may be reduced by any or all of the following: non-abrupt (graded) junction, two sided junction or increased radius of curvature. In a non-abrupt junction the doping changes gradually from p to n-type. In a two sided junction there is depletion on both the p and n sides of the junction, distributing the voltage, but causing a lower peak E-field than if the junction were one-sided. The guard ring may be provided by ion implantation (most common) or a second Zn diffusion. In either case processing difficulties arise. With ion implantation the issues are how to reproducibly activate the implant and how to minimize the damage from the ion bombardment, which can have potential impact both on initial performance and reliability of the device. With a second p-diffusion the issues are how to achieve a difference in the profiles and/or abruptness of the p-n junctions between the two diffusions.

Added to profiles and/or abruptness is the further complication of accurately controlling the guard ring location relative to the main p-n junction in order to prevent premature edge breakdown. If the guard ring is too deep then tunneling in the InGaAs could occur, whereas if the guard ring is too shallow then premature edge breakdown could still occur. Also some

4. Separate absorption, grading , charge and multiplication APDs

Figure 4.1.2 Illustration of an APD with a guard ring.

workers find that one guard ring [4] is not enough, and two or more guard rings become necessary, each of whose characteristics must be very accurately controlled [81, 82, 125, 126].

4.2 Initial SAGCM concept, assumptions

Webb was the first to show that guard rings could be eliminated from a planar APD if the charge sheet was selectively confined to the center of the device [27]. Webb's APD structure is shown in Fig. 4.2.1. The charge sheet is selectively confined to the center of the device, avoiding the need for guard rings, thus allowing separate absorption, charge sheet, gradings and multiplication (SAGCM). At the same time, the multiplication region is decoupled from the charge sheet itself, easing the fabrication tolerances over that of SAM or SAGM APDs.

4. Separate absorption, grading, charge and multiplication APDs

Fig. 4.2.1 Schematic of Webb's SAGCM APD. The charge sheet is selectively confined to the center of the device, avoiding the need for guard rings. At the same time, the multiplication region is decoupled from the charge sheet itself, easing the fabrication tolerances over that of SAM or SAGM APDs.

This may be understood from comparing the E-field profile in the central region of an SAGCM APD, as illustrated in Fig. 4.2.2, to that of an SAGM, as in Fig. 3.3. In addition to all the features presented in Fig. 3.3, Fig. 4.2.2 shows the heterointerface E-field E_{het} is also shown, about which more will be said later. The very tight fabrication tolerance of $\pm 200 \text{ \AA}$ in SAGM APDs arises from the need to keep the E-field in the multiplication region very tightly controlled [81, 82]. The p-n junction position affects the E-field from Gauss' Law, and is directly related to the doping in the multiplication region. For a factor of 50 less doping in the multiplication region for a

4. Separate absorption, grading, charge and multiplication APDs

SAGCM APD than for a SAGM APD, the fabrication tolerance of $\pm 200 \text{ \AA}$ increases to $\pm 1 \text{ \mu m}$ for the same change in E-field within the multiplication region. This is strong incentive to pursue these devices further. Note that as the multiplication region thickness changes for a given E-field, the gain of the device will also change for both the SAGM and the SAGCM APD.

Fig. 4.2.2 (a) Layer sequence and (b) E-field vs. x for a SAGCM APD. For a given peak E-field, even if the p-n junction is not accurately positioned, tunneling in the InGaAs (too deep) or incomplete depletion or hole trapping (too shallow) will not occur.

4.3 Charge sheet in periphery

Eliminating the charge sheet in the device periphery is beneficial because there is less peripheral E-field in the InP, thus preventing edge breakdown in a planar structure. The main benefit of this approach is to decouple the multiplication region from the charge sheet, easing fabrication tolerances.

In this section, an original concept is presented as a further move in that direction. Two lateral charge sheets may be incorporated as illustrated in Fig. 4.3.1: the complete charge sheet density (σ_{active}) below the plane region of the p-diffusion and a partial charge sheet density ($\sigma_{\text{periphery}}$) in the remainder of the APD, for SAGCM in two dimensions. This extra degree of freedom permits independent control of the E-field profiles in the active region and the device periphery. As in other APDs, the E-field profile of the active region of the device may be tailored to permit breakdown in the InP while preventing tunneling in the InGaAs. By using a partial charge sheet in the device periphery, the E-field in the InGaAs in the remainder of the device may also be independently tailored to minimize the tunneling in the device periphery.

4.3.1 Comparison of DC E-fields

To better understand the advantage of a partial charge sheet in the device periphery, it is necessary to consider the E-field profile of the APD structure. In the plane (along lines A and B of Fig. 4.3.1) and curved (along lines C and D of Fig. 4.3.1) regions of a planar APD structure the E-field profile may be approximated from a one-dimensional (1-D) solution of Poisson's equation in Cartesian and cylindrical coordinates, respectively [28]. This

Figure 4.3.1 Schematic diagram of a SAGCM APD with partial charge sheet in the periphery. The E-field profiles along the lines indicated by A, B, C and D may be used to predict device performance.

concept was discussed in Section 2.4. These determine upper and lower bounds of certain device parameters. The values of E-field at other points are intermediate between the Cartesian and cylindrical solutions, and do not represent extreme cases.

4. Separate absorption, grading, charge and multiplication APDs

2

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

The Cartesian solutions are straightforward. However, some thought is required to correctly analyze the cylindrical portion of the device. A planar junction with a charge sheet is shown in Fig. 4.3.2, with the inner portion of the depletion region of coordinate r_0 , the upper edge of the charge sheet of coordinate r_1 , the lower edge of the charge sheet of coordinate r_2 and the lower edge of the n^- region of coordinate r_3 . The depletion region lower boundary is denoted by the coordinate R . Analogous to the discussion surrounded by Fig. 2.4.2, the assumptions involved for calculating the E-field along the line indicated by C in Fig. 4.3.2 (similar to the line indicated by C in Fig. 4.3.1) are shown schematically in Fig. 4.3.3 where the coordinates just discussed become radii of curvature [28]. As will be demonstrated below, the maximum E-field is highest for the entire structure along this line. While the charge sheet is not actually curved in the real device, this is a reasonable approximation because: (1) the depletion region is curved and (2) a curved charge sheet will if anything lead to a higher (i.e. worst case) estimate of the maximum E-field along this line than would be the case for a non-curved charge sheet.

4.3.2 Derivation of numerical equation for cylindrical E-field

The E-field in a SAGCM structure may be determined by numerical simulation of Poisson's equation. In this section, a numerical derivation for simulation of the E-field in the cylindrical portion of the device is presented. This may or may not correspond to known techniques of numerical analysis, a study of which does not form a part of this work.

The starting point for the derivation is Eqn. (2.4.7) which may be written (including a sign change for the E-field in InP for convenience) as

Figure 4.3.2 Illustration of approximation of a planar junction as a cylindrical depletion region of inner boundary r_0 and outer boundary R for a SAGCM APD. The E-field may be considered at any coordinate r within the depletion region.

Figure 4.3.3 Approximate cylindrical representation of the peripheral region of a SAGCM APD.

4. Separate absorption, grading, charge and multiplication APDs

$$d(rE) = -\frac{N_D}{\epsilon} r dr \quad (4.3.1)$$

Consider the position $r+\Delta r$, at which the E-field is of value $E(r+\Delta r)$. Expansion of both differentials gives

$$E(r + \Delta r) = \frac{r}{(r + \Delta r)} E(r) - \frac{r}{(r + \Delta r)} \frac{N_D}{\epsilon} \Delta r \quad (4.3.2)$$

The quantity $\frac{r}{(r + \Delta r)}$ is given by Taylor series expansion.

$$\begin{aligned} \frac{r}{(r + \Delta r)} &= \frac{1}{(1 + \frac{\Delta r}{r})} \\ &\approx 1 - \frac{\Delta r}{r} + \left(\frac{\Delta r}{r}\right)^2 - \left(\frac{\Delta r}{r}\right)^3 + \dots \end{aligned} \quad (4.3.3)$$

Substituting Eqn. (4.3.3) into (4.3.2), keeping terms of up to first order only gives

$$E(r+\Delta r) = E(r) \left(1 - \frac{\Delta r}{r}\right) - \frac{N_D}{\epsilon} \Delta r \quad (4.3.4)$$

This expression for $E(r+\Delta r)$ differs from the expression for the Cartesian case by the quantity $E(r) \Delta r/r$. Note that the same result is obtained by considering the first order differential expansion of Eqn. (4.3.1). However, Eqn. (4.3.3) contains the error in the first order expansion.

For the APD under consideration, the device may be partitioned into a number of regions i , each of which have a particular volume doping and

4. Separate absorption, grading, charge and multiplication APDs

dielectric constant. If the edge of the depletion region is specified by R , then the doping within the region is defined by

$$N_D(r) = N_i, \quad r_{i-1} < r < r_i, \quad x_j < r < R \quad (4.3.5)$$

Within each region, Eqn. (4.3.4) may be used over many points for a small incremental thickness Δr to determine the E-field throughout the region. The boundary condition $V = -\int E(r') dr'$ is used as well. As a test, similar to any numerical procedure, the E-field profile should be independent of Δr . If the change in E-field profile is less than some tolerance for different Δr values, then Δr is small enough. In practice, the grid spacing Δr should be particularly small in the charge sheet itself, as is evident from inspection of Eqn. (4.3.4), where N_D is high. Thus Eqns. (4.3.4) and (4.3.5) may be used for numerical simulations.

4.3.3 Modeling results

In this section, Eqn. (4.3.4) has been used to simulate SAGCM APDs of varying charge sheets in the device periphery for the first time with the following assumptions and conditions. An E-field in the central region of the device of $50 \text{ V}/\mu\text{m}$ was assumed, as well as a charge in the active mesa region σ_{active} of $2.7 \times 10^{12} \text{ cm}^{-2}$, in line with literature reports of appropriate values [27, 39-41]. The cylindrical E-field profile was varied until the electrostatic potential difference across the device was the same as in the central region, which assumes a conducting p^+ diffusion. Fractional charge sheets from 0 to 1 in increments of 0.2 were used. Undoped InGaAs thicknesses of 2.5, 2.75 and $3 \mu\text{m}$ were used above a $1.3 \mu\text{m}$ thick undoped buffer-InP layer. It was assumed that the device planarized following the second growth as shown

4. Separate absorption, grading , charge and multiplication APDs

schematically in Fig. 4.3.1, with a mesa height of $0.2\text{ }\mu\text{m}$. As well, x_d was simulated for 0.2 , 0.4 and $0.6\text{ }\mu\text{m}$. Finally, x_j was simulated for 1.5 , 2.0 , 2.5 and $3.0\text{ }\mu\text{m}$. Note that all the x_j values are less than Webb's $3.5\text{ }\mu\text{m}$ [27]. This is because deeper Zn diffusions are more difficult to control (e.g. $\pm 0.1\text{ }\mu\text{m}$ represents 2.9% of $3.5\text{ }\mu\text{m}$ and 5% of $2\text{ }\mu\text{m}$, which allows for more tolerance), and it would be desirable to reduce x_j from $3.5\text{ }\mu\text{m}$.

Representative E-field profiles in the active region and the slab region of the device annular periphery (where the labeling corresponds to lines A and B in Fig. 4.3.1) are illustrated in Fig. 4.3.4. An abrupt p^+-n junction and the modeling parameters given in Table 4.3.1 were assumed. The E-field terminates abruptly at the n^+ substrate. The E-field throughout the InGaAs is greater when annular peripheral charge σ_{ann} is zero, increasing the probability of tunneling which can cause unacceptably high dark currents. A small displacement of the p^+-n junction does not substantially change σ_{active} (which controls the multiplication properties), but only the breakdown voltage (BV) of the device.

Table 4.3.1. Modeling parameters assumed for the E-field profiles in Section 4.3.

Breakdown field in InP	$50\text{ V}/\mu\text{m}$
Background concentration N_D	10^{15} cm^{-3} (n-type)
σ_{active}	$2.6 \times 10^{12}\text{ cm}^{-2}$
$\sigma_{\text{periphery}}$	$0, 0.4\sigma_{\text{active}}, \sigma_{\text{active}}$
x_j	$2.0\text{ }\mu\text{m}$
x_d	$0.4\text{ }\mu\text{m}$
t_{undoped}	$3.8\text{ }\mu\text{m}$

Figure 4.3.4 Representative slab (Cartesian) E-field profiles for an SAGCM APD for the case of no, some and full peripheral charge sheet.

Figure 4.3.5 Representative E-field profiles for the cylindrical peripheral region of an SAGCM APD for the case of no, some and full peripheral charge sheet.

4. Separate absorption, grading, charge and multiplication APDs

Representative E-field profiles for the portion of the periphery where the p-diffusion front is curved (line C in Fig. 4.3.1) are illustrated in Fig. 4.3.5 for the same reverse bias and other assumptions as in Fig. 4.3.4 [28]. When the complete charge density is present (curve C1), the maximum E-field in the InP is greater than the breakdown field, as expected from above. When no partial charge sheet density is present ($\sigma_{ann}=0$, curve C3) the E-field in the InP is at its lowest, but as in the slab region, the E-field in the InGaAs is increased. By using a partial charge sheet density ($\sigma_{ann}= 0.4 \sigma_{active}$, curve C2) the E-field in the InP can be kept below the breakdown E-field for InP while the E-field throughout the InGaAs is reduced. The BV along line D should not dominate the device characteristics (i.e. no premature breakdown along the side) for sufficiently low doped background InP. This may be more precisely quantified by comparing the voltage in the slab region to the voltage in Eqn. (2.4.14) for the same E-field. Provided the voltage in the cylindrical region is higher, there will be no premature breakdown from this effect.

It is evident that a maximum cylindrical E-field for the planar portion of the device in both the InP and InGaAs may be defined above which the edge region will dominate the device characteristics for each simulation. That is, if the maximum cylindrical E-field in the InP is greater than that in the central region, there will be premature breakdown at the edge; if the maximum cylindrical E-field in the InGaAs is beyond the tunneling threshold, then the InGaAs tunneling leakage current will be prohibitively large. As well, the voltage is uniquely determined for each simulation.

The dependence of the cylindrical E-field in the planar portion of the device is shown in Figs. 4.3.6 - 4.3.8. In each figure E-field is plotted vs. f , the fractional charge sheet. In each figure two of (x_d , x_j and t_{InGaAs}) are kept at constant values of (0.4 μm , 2.0 μm and 2.75 μm) and the third varied.

4. Separate absorption, grading , charge and multiplication APDs

Figure 4.3.6 Maximum cylindrical E-field in InP and InGaAs vs. fractional charge sheet for planar APD structure, various x_d .

Figure 4.3.7 Maximum cylindrical E-field in InP and InGaAs vs. fractional charge sheet for planar APD structure, various x_j .

4. Separate absorption, grading, charge and multiplication APDs

Figure 4.3.8 Maximum cylindrical E-field in InP and InGaAs vs. fractional charge sheet for planar APD structure, various t_{InGaAs}

Fig. 4.3.6 shows the maximum cylindrical E-field vs. f dependency for various x_d . For a given x_d , the E-field in the InP increases and the E-field in the InGaAs decreases with increasing f . Increasing x_d has the effect of steepening the decrease of the maximum E-field in the InGaAs and decreasing the permitted maximum f . The permitted f regime is considered to be that where the cylindrical E-field in the InP is less than the E-field in the main region of the device, while at the same time avoiding substantial tunneling leakage current. As the tunneling threshold in the InGaAs varies significantly within the literature, it is difficult to state the permitted lower f boundary.

4. Separate absorption, grading, charge and multiplication APDs

Fig. 4.3.7 shows the maximum cylindrical E-field vs. f dependency for various x_j . For a given x_j , as for a given x_d , the E-field in the InP increases and the E-field in the InGaAs decreases with increasing f . Increasing x_j has little effect on the maximum E-field in the InGaAs and increases the permitted maximum f for which the maximum E-field in the InP is still less than that in the central breakdown regime. Also, the curves are more closely spaced on the E-field vs. f graph as x_j increases incrementally. The physical significance of this is that for infinite x_j (the plane case) the E-field in the InP at $f=1$ should be exactly $50 \text{ V}/\mu\text{m}$, which is approached slowly as x_j increases. This also means that in attempting to decrease x_j there is a point where the permitted maximum f starts to decrease rapidly; this starts to occur as x_j decreases from 2.0 to $1.5 \mu\text{m}$.

Fig. 4.3.8 shows the maximum cylindrical E-field vs. f dependency for various t_{InGaAs} . It is seen that there is little effect of a change in t_{InGaAs} on the maximum E-field in either the InP or the InGaAs, which is dominated by f alone for constant x_d and x_j .

The voltage applied to a device for a given maximum E-field in the central region is determined by the plane portion of the device, and therefore independent of x_j . The dependencies of voltage vs. x_d and t_{InGaAs} are shown in Fig. 4.3.9 and 4.3.10. As seen in Fig. 4.3.9, the effect of increasing x_d is to increase the voltage. This is expected from consideration of the E-field vs. x dependency shown in Fig. 4.2.3. The effect of increasing x_d is to move the p-n junction to $x < 0$ (i.e. to the left on the x -axis), increasing the area under the curve and therefore increasing the voltage. The value of t_{InGaAs} is seen to have little effect on the voltage in Fig. 4.3.9, also expected from the E-field vs. x dependency shown in Fig. 4.2.2. The effect of increasing t_{InGaAs} is to move the n^+ substrate to the right, increasing the area under the curve and

therefore increasing the voltage, but at a much slower rate than would be the case for the same increase in x_d because the E-field in the InGaAs is much less than the E-field in the InP. These trends are also clearly shown in Fig. 4.3.10, where the voltage is nearly flat with respect to t_{InGaAs} , but depends in large measure on x_d . Note that the gain will depend on x_d .

Figure 4.3.9 Voltage vs. t_{InGaAs} in the center of the device (Cartesian).

Figure 4.3.10 Voltage vs. x_d in the center of the device (Cartesian).

4. Separate absorption, grading, charge and multiplication APDs

4.4 Modeling of ideal SAGCM performance: bulk-InP multiplication

In this section the performance parameters of an ideal δ -doped APD are simulated using analytical techniques. Recently, Kuchibhotla and Campbell [31] have presented numerical simulations for an ideal δ -doped SAGCM APD. In this section, for the first time, analytical results for the same structure are obtained and compared with the numerical results of Kuchibhotla and Campbell. Results are obtained for the E-field, voltage and GBW as functions of gain and x_d , as well as displacement correction factor as a function of k , x_d and carrier velocity ratio. The analytical formulas presented in this section lend themselves to an analytical sensitivity analysis for both the E-field and voltage, which is also presented for the first time.

4.4.1 Multiplication equations for bulk-InP ionization

For bulk-InP APDs, unlike Si, $\beta > \alpha$ in the InP multiplication region, so hole injection is preferable. One convenient way to approach the mathematical formalism is to redefine the position of electron and hole injection opposite to the Si convention used in Section 2.3. Opposite to the previous geometry, the n contact of the diode is located at $x \leq 0$ and the p contact of the diode is located at $x \geq w$. This is illustrated schematically along with the hole and electron current densities as functions of x in Fig. (4.4.0). Mathematically, the idea is to mimic the Si-APD treatments, with the exception that α and β become reversed. This is not usually done in the literature. With this in mind, the multiplication $M_{\text{InP}}(x)$ due to generation of carriers at position x is given by

Figure 4.4.0 Illustration of multiplication and current density vs. position for an APD with $\beta > \alpha$.

4. Separate absorption, grading, charge and multiplication APDs

$$M_{\text{InP}}(x) = \frac{\exp \left[- \int_0^x (\beta - \alpha) dx' \right]}{1 - \int_0^w \beta \exp \left[- \int_0^{x'} (\beta - \alpha) dx' \right] dx} \quad (4.4.1)$$

where again w is the width of the multiplication region and α and β are in general position dependent.

The multiplication due to hole and electron injection in an InP-based APD are given by the special cases $M_{\text{p InP}} = M(0)$ and $M_{\text{n InP}} = M(w)$. For hole injection,

$$M_{\text{p InP}} = \frac{1}{1 - \int_0^w \beta \exp \left[- \int_0^{x'} (\beta - \alpha) dx' \right] dx} \quad (4.4.2)$$

which may be written in the form

$$1 - \frac{1}{M_{\text{p InP}}} = \int_0^w \beta \exp \left[- \int_0^x (\beta - \alpha) dx' \right] dx \quad (4.4.3)$$

Here the advantage of reversing the conventions of hole and electron injection positions relative to Si becomes clear in that the numerator of Eqn. (4.4.2) is 1 rather than an integral, which makes for easier analysis below.

4. Separate absorption, grading, charge and multiplication APDs

This would not be the case if the Si conventions from Section 2.3 were used, merely switching the location of 0 and w (i.e. reversing the direction of the x -axis). Also, using these conventions (reversed conventions relative to Si) also allows for direct association of equations with their analogs from Section 2.3.

It is of importance to know the E-field profile in the multiplication region for the SAGCM APD. This involves solving the avalanche multiplication equation (e.g. Eqn. (4.4.1) or (4.4.2)) for the actual p-doping profile on the p-side and the non-zero thickness of the charge sheet on the n-side of the multiplication region.

4.4.2 δ -doped APD: assumptions

As a first approximation, the case of a constant E-field multiplication region may be considered. For this case, it is assumed that the $p^+ - n^-$ junction is perfectly abrupt and the charge sheet drops the E-field over a distance of zero – i.e. ideal δ -doping. Kuchibhotla and Campbell have recently provided a numerical treatment of a p-i-n APD using equations (2.3.7)-(2.3.17) [27]. While a numerical treatment provides the flexibility to accommodate a non-ideal δ -doped structure, one should be able to obtain the same results by analytical methods, which can provide some physical insight. In this section an analytical treatment of the same problem is provided, with comparison to the numerical results previously published.

In order to understand the analytical derivations and simulations below, it is necessary to be clear on the assumptions with respect to the device structure. Only the Cartesian portion of the device is simulated; it is assumed that the device periphery is arranged so as to avoid edge breakdown as discussed in Section 4.3 or that the geometry is mesa type, which might well

avoid the problem altogether. The E-field in the multiplication region is assumed to be constant. Implicit in the assumption is zero background doping. As a rough guide, a background doping of 10^{15} cm^{-3} , typical of what is achievable today in an MOCVD reactor, corresponds to an E-field drop within the multiplication region of $1.5 \text{ V}/\mu\text{m}^2$ – i.e. a drop of $1.5 \text{ V}/\mu\text{m}$ over a distance of $1 \mu\text{m}$. This E-field drop is neglected in this work, but might be included in a highly accurate treatment. More fundamentally, as technology improves, the doping is limited by the intrinsic carrier concentration (n_i) in InP, roughly eight orders of magnitude down from this value. For this case, the donor and acceptor concentrations are equal and the corresponding E-field drop is zero. Thus in principle a constant E-field is well justified. Another assumption includes an abrupt one-sided p-n junction. This implies that the Zn diffusion profile falls off infinitely fast from its p^+ level, inconsistent with the known profile [165,166], which is the solution of the diffusion equation with concentration dependent coefficients (which would be the erfc function for only one diffusion mechanism and concentration independent coefficients [167]). Nevertheless, this approximation is not bad and is often used in device modeling. Unlike Section 4.3 where the thickness of the charge sheet was taken into account in the modeling, in this section an abrupt δ -doping of zero thickness is assumed for an ideal charge sheet. It is further assumed that ionization may be neglected in areas other than the high-field multiplication region. This set of approximations is consistent with those of Kuchibhotla and Campbell [31].

In the remainder of this section, mathematical analysis of the δ -doped APD is done using analytical, not numerical, techniques.

4. Separate absorption, grading , charge and multiplication APDs

4.4.3 Analytical derivations for δ -doped APD

If α and β are constants, Eqn. (4.4.3) may be written in the form

$$1 - \exp [-(\beta - \alpha)w] = (1 - \frac{1}{M_p \text{InP}}) (1 - k) \quad (4.4.4)$$

where $k = (\alpha / \beta)$. Because k depends on β , Eqn. (4.4.4) is a transcendental equation which cannot analytically be solved for β . Even worse, α , β and k themselves are functions of the E-field, which makes analytical evaluation with respect to the E-field especially difficult. However, this equation can be analytically solved for w or $M_p \text{InP}$:

$$w = \frac{-\ln[1 - (1 - \frac{1}{M_p \text{InP}}) (1 - k)]}{\beta - \alpha} \quad (4.4.5)$$

$$M_p \text{InP} = \frac{1 - k}{\exp [-(\beta - \alpha)w] - k} \quad (4.4.6)$$

Eqns (4.4.5) and (4.4.6) are not found elsewhere in the literature, although an expression similar to Eqn. (4.4.6) appears in Ref. [1]. They may be used as follows. For any value of E-field, $\alpha(E)$ and $\beta(E)$ may be substituted into Eqn. (4.4.5) or (4.4.6). The resulting w may be calculated for a given $M_p \text{InP}$ (Eqn. (4.4.5)) and the resulting $M_p \text{InP}$ may be calculated for a given w (Eqn. (4.4.6)). Eqn. (4.4.6) is used extensively in Section 4.4.4.

At this point, it is well worth noting some observations connected with these expressions before proceeding to the modeling results.

The condition for breakdown (infinite gain) is that the denominator of Eqn. (4.4.6) go to zero, which gives the result

4. Separate absorption, grading , charge and multiplication APDs

$$\exp [-(\beta-\alpha)w] = k \quad (4.4.7)$$

as the condition for breakdown. This expression has been noted by Conradi [65] as approximately correct for high gain. Eqn. (4.4.7) is transcendental, so numerical solutions must be used. However, this is a simpler closed form expression than Eqns. (4.4.4) - (4.4.6), and easily amenable to analysis either by Taylor series or by differentials of both sides of the equation. Note that for a constant Γ -field, these results can also be obtained from Eqn. (2.3.7) in the same manner as they were derived from Eqn. (4.4.2).

A number of observations are possible from considering the analytical expressions in Eqns. (4.4.5) and (4.4.6). Both the numerator and denominator of Eqn. (4.4.5) retain the same sign as each other, but change signs from both positive when $\beta > \alpha$ to both negative when $\beta < \alpha$, always yielding a positive value for w . The sign in Eqn. (4.4.6) for $M_p \ln P$ changes from positive to negative, going through infinity, as $\exp [-(\beta-\alpha)w]$ changes from greater than k to less than k . For the case of an infinitely thick multiplication region or zero ionization coefficients, $M_p \ln P$ is unity. For a practical APD, the denominator ranges from 0^+ (breakdown) to k (gain of 1).

The multiplication region width for Eqn. (4.4.5) or the gain for Eqn. (4.4.6) for hole injection may be calculated from the literature experimental dependencies of α and β on E . Usually these are expressed in the following form [58-64]:

$$\begin{aligned} \alpha &= \alpha_0 \exp\left(\frac{-E_\alpha}{E}\right) \\ \beta &= \beta_0 \exp\left(\frac{-E_\beta}{E}\right) \end{aligned} \quad (4.4.8)$$

4. Separate absorption, grading, charge and multiplication APDs

where α_0 , β_0 , E_α and E_β are constants. There are a number of literature values obtained for these constants, as summarized in Table 4.4.1. Note that one problem within the literature is the presence of figures supplied without a best line fit provided. For these cases, the dependencies must be determined by a best line fit through the graphical data.

Table 4.4.1. Ionization coefficients for different InGaAsP compositions from the literature. The [As] mole fraction is 0 for InP, between 0 and 1 for quaternary InGaAsP compositions lattice-matched to InP and 1 for InGaAs.

[As] mole fraction	Ref.	Author	α_0 (μm^{-1})	β_0 (μm^{-1})	E_α (V/ μm)	E_β (V/ μm)
0	[58]	Cook	293	162	264	211
0	[62]	Umeba	736	204	345	242
0	[63]	Armiento	555	321	310	256
0	[73]	Osaka	555	198	310	229
0.7	[73]	Osaka	337	294	229	240
0.7	[61]	Osaka	350	231	231	241
1	[73]	Osaka	227	395	113	145
1	[64]*	Pearsall's data	5130	7300	195	220
0	[61]**	Osaka's data	340.5	171.1	290.5	223.6
0.28	[61]**	Osaka's data	722.9	416.0	309.1	274.4
0.39	[61]**	Osaka's data	777.0	506.3	315.4	289.6
0.7	[61]**	Osaka's data	363.1	301.4	230.3	239.9
1	[61]**	Osaka's data	32.4	60.2	120.2	152.4

* interpreted by Osaka

** best graphical line, interpreted in this work

A further observation is possible, from Eqn. (4.4.6). The denominator contains an exponential of α and β which themselves are exponential functions of the E-field. In a practical APD, w remains constant and the gain

4. Separate absorption, grading, charge and multiplication APDs

is varied by changing the bias, which in turn changes the E-field within the device, affecting the ionization coefficients. One should expect a strong increase in the multiplication for E-fields close to the breakdown condition. This observation is experimentally well known, but it is interesting that one can be led to this conclusion intuitively upon inspection of Eqn. (4.4.6).

For these experimental values, the k dependence is given by

$$k = \frac{\alpha_0}{\beta_0} \exp\left(\frac{-(E_\alpha - E_\beta)}{E}\right) \quad (4.4.9)$$

From Eqn. (4.4.9) it is evident that k changes substantially as a function of E .

It is possible to consider the ratio of gains from electron and hole injection. From Eqn. (4.4.1), using $M_p = M(0)$ and $M_n = M(w)$,

$$\frac{M_n}{M_p} = \exp \left[- \int_0^w (\beta - \alpha) dx' \right] \quad (4.4.10)$$

Eqn. (4.4.10) is a general result, noted previously by Stillman and Wolfe ([11], Eqn. (70)) and Lee ([48], Eqn. (8)). If α and β are constants, then Eqn. (4.4.10) simplifies to

$$\frac{M_n}{M_p} = \exp [-(\beta - \alpha)w] \quad (4.4.11)$$

The expression for w from Eqn. (4.4.5) can be substituted into Eqn. (4.4.11) to give

4. Separate absorption, grading , charge and multiplication APDs

$$\frac{M_n}{M_p} = k \left(1 + \frac{1-k}{k} \frac{1}{M_p} \right) \quad (4.4.12)$$

As it is assumed the correct carrier is injected, $k < 1$ and therefore $M_n/M_p \geq k$. Thus the ratio of gains from electron and hole injection is given by $k(1 + \zeta_{\text{corr}})$, where ζ_{corr} is a correction factor which is less than 1 for hole injection. This correction is most significant for small gains. For a practical APD, both M_n and M_p can range from 1 (unity gain) to ∞ (breakdown). These two limits are represented by

$$\begin{aligned} \left(\frac{M_n}{M_p} \right)_{M_p \rightarrow 1} &= 1 \\ \left(\frac{M_n}{M_p} \right)_{M_p \rightarrow \infty} &= k \end{aligned} \quad (4.4.13)$$

The first expression above, namely that the ratio of gains from hole and electron injection is 1 for unity gain, is intuitively obvious, and lends partial validation to the mathematical treatment. The second statement that the ratio of gains at high gains is k is perhaps less obvious but very useful. This gives a fundamental method of measuring k for the device. The most common way to measure k is to measure the excess noise, usually as a function of gain, and extract k , assumed to be constant with respect to gain. This method assumes that k does not vary with gain, yet k is a function of the E-field (since α and β are) and the way to vary gain is to vary the E-field. Worse, there are calibration [75] and other errors in the excess noise measurement which further compound the problem. Eqn. (4.4.13) says that if a method of pure electron injection and pure hole injection can be found [12, 21] then the ratio of gains approaches k as multiplication approaches ∞ . This k is of course the k at breakdown, which is the highest possible k for the

4. Separate absorption, grading, charge and multiplication APDs

device. At lower gains, k will be less with a value which is easily obtained from Eqn. (4.4.12):

$$k = \frac{M_n - 1}{M_p - 1} \quad (4.4.14)$$

This relationship has been presented previously by McIntyre ([67], Eqn. (A5)). Note that Eqn. (4.4.14) approaches Eqn. (4.4.13) in the limit of $M_n, M_p \gg 1$.

Note that the extraction of k is purely academic exercise. A lower k value should result in superior device performance. For example, all other things being equal, the GBW should be higher and the excess noise should be lower for lower k . These performance features themselves are the ones which really matter for device performance. Also, in many cases k does not vary by very much over the gain region of interest, and in this case excess noise measurements in themselves can give a reasonable estimate of k .

4.4.4 E-field, voltage and k : modeling results

Simulations have been carried out on Eqn. (4.4.6) using Microsoft Excel calculations with no special programming skills required. The E-field was adjusted to the nearest 0.01 V/ μm in order to produce either breakdown or a gain of 10, which is a typical gain at which a practical APD might operate in an optoelectronic receiver. For breakdown, it was found easiest to calculate $1/M_p \text{ InP (E)}$ rather than $M_p \text{ InP (E)}$ itself, as the reciprocal gain passes smoothly through zero at the breakdown field. The ionization coefficients used were those in [61] as given in Table 4.4.1 in order to permit comparison between these analytically obtained values and those of [31], obtained by numerical integration.

Figure 4.4.1 E-field for breakdown and $M=10$ vs. x_d for an ideal δ -doped APD with bulk InP ionization only.

The E-fields for these two conditions (breakdown and $M=10$) as a function of x_d (using $w=x_d$ in the preceding equations) are plotted in Fig. 4.4.1. It is evident that as x_d becomes smaller, ever increasing E-field is required for breakdown or for a gain of 10. Qualitatively, the entire curve agrees with that of Ref. [31] to best visual inspection. As quantitative references, the E-field at $M=10$ for $x_d=\{0.1, 0.2\} \mu\text{m}$ of $\{74.67, 62.43\} \text{V}/\mu\text{m}$ agrees exactly to three significant digits with Kuchibhotla and Campbell's published values of $\{74.65, 62.4\} \text{V}/\mu\text{m}$ [31]. The conclusion is that the analytical expressions of Eqns. (4.4.5)-(4.4.7) can accurately reproduce a result obtained from numerical integration for the same set of assumptions.

4. Separate absorption, grading , charge and multiplication APDs

Figure 4.4.2 Voltages for breakdown and $M=10$ vs. x_d for an ideal δ -doped APD with bulk InP ionization only. A $3.2\ \mu\text{m}$ thick InGaAs layer of negligible ionization below the InP with a doping of $10^{15}\ \text{cm}^{-3}$ was assumed.

The voltages for breakdown and $M=10$ as a function of w are plotted in Fig. 4.4.2. Obviously the voltage depends on the E-field profile for the material below the InP. A $3.2\ \mu\text{m}$ thick InGaAs layer of negligible ionization below the InP with a doping of $10^{15}\ \text{cm}^{-3}$ was assumed. The prediction is that the breakdown voltage in a practical APD should be a strong function of x_d for $w < 0.3\ \mu\text{m}$, decreasing from 115.3 to 60.9 V as w increases from 0.1 to 0.3 μm . As x_d increases beyond 0.3 μm , the breakdown voltage decreases gradually to its minimum of 48.4 V at $x_d = 0.7\ \mu\text{m}$, rising gradually beyond this point to 64.8 V at $w = 1.5\ \mu\text{m}$, which is the upper bound of the simulation.

4. Separate absorption, grading, charge and multiplication APDs

Figure 4.4.3 Calculated k for breakdown and $M=10$ for an ideal δ -doped APD with bulk InP ionization only.

Also k for breakdown and $M=10$ as a function of x_d are plotted in Fig. 4.4.3. Note that k is a function of the gain. Although most literature reports imply a constant value of k , independent of the gain, this is not reasonable because the E-field varies with gain, and k is a function of the E-field. As in the simulations of Kuchibhotla and Campbell [31], k increases strongly with decreasing x_d . Most commercial manufacturers of InP/InGaAs APDs will guarantee $k \leq 0.54$. The theoretical prediction is that $k \leq 0.54$ occurs only for $x_d \geq 0.7 \mu\text{m}$. Again, this is consistent with the simulations of Kuchibhotla and Campbell, and suggests that noise performance problems may be encountered with SAGCM APDs for smaller x_d for the assumption of InP-only ionization.

4. Separate absorption, grading, charge and multiplication APDs

Note that Fig. 4.4.3 predicts $k \geq 1$ for $x_d < 0.1 \mu\text{m}$. This is based on an extrapolation of Osaka's values beyond the E-field range for which they were measured, which is not legitimate. In all likelihood, the α and β dependencies vs. E never really cross, with both ionization coefficients being subject to increased scattering as the E-field increases.

4.4.5 Sensitivity analysis for breakdown field and voltage

As mentioned above, the analytical expressions obtained for the breakdown condition are amenable to a sensitivity analysis, which is presented for the first time in this section. In particular, if both sides of Eqn. (4.4.7) are differentiated with respect to E and arranged to solve for $\frac{dw}{dE}$,

$$\frac{dw}{dE} = \frac{\frac{d\beta}{dE} \left(\frac{1}{\beta} - w\right) - \frac{d\alpha}{dE} \left(\frac{1}{\alpha} - w\right)}{\beta - \alpha} \quad (4.4.15)$$

where the expressions are evaluated at the breakdown field E_B . This expression may be written in the form

$$dE_B = \frac{\beta - \alpha}{\frac{d\beta}{dE} \left(\frac{1}{\beta} - w\right) - \frac{d\alpha}{dE} \left(\frac{1}{\alpha} - w\right)} dw \quad (4.4.16)$$

which evaluates the change in breakdown field E_B as a function of changes in multiplication region width w . This is an important result for practical applications because in the manufacture of SAGCM or δ -doped APDs one of the critical steps is the control of w by means of a p-diffusion. For a given doping profile the E-field profile as a function of position (i.e. $E(x)$) is determined to within a constant which depends only on the bias. In the vicinity of a particular E_B the change in E_B and hence the change in the entire

E-field profile of the device can be determined analytically once known dependencies for α and β are substituted into Eqn. (4.4.16). The ionization coefficients are often expressed in the form given by Eqn. (4.4.8). In this case the derivatives of α and β can be easily eliminated from Eqn. (4.4.16), and the change in the breakdown field dE_B for small changes in the multiplication region width dw may be expressed as

$$dE_B = \frac{\beta - \alpha}{E_\beta (1 - \beta w) - E_\alpha (1 - \alpha w)} E_B^2 dw \quad (4.4.17)$$

In addition, the expected change in breakdown voltage dV_B for a change in multiplication region width can be written as

$$\begin{aligned} dV_B &= dE_B t_{\text{total}} + E_B dw \\ &= \left(\frac{\beta - \alpha}{E_\beta (1 - \beta w) - E_\alpha (1 - \alpha w)} t_{\text{total}} E_B^2 + E_B \right) dw \quad (4.4.18) \end{aligned}$$

where t_{total} is the total thickness of the APD, assumed to already be fully depleted. Note that Eqns. (4.4.17) and (4.4.18) result from a linear expansion of the derivative – i.e. terms up to first order in the Taylor series – and the result would be valid only for small dw and dE_B .

4.4.6 Avalanche buildup time

Analogous to Eqn. (2.3.16), in bulk-InP the primary avalanche build-up time τ_p for hole initiated multiplication is

$$\tau_p = \frac{K_{\text{displacement}}}{v_n + v_p} \int_0^w \exp \left[- \int_0^x (\beta - \alpha) dx' \right] dx \quad (4.4.19)$$

For the case of an idealized p-i-n multiplication region for a bulk-InP APD, α and β are constant, and Eqn. (4.4.19) simplifies to

$$\tau_p = \frac{K_{\text{displacement}}}{v_n + v_p} \frac{1 - \exp [-(\beta - \alpha)w]}{\beta - \alpha} \quad (4.4.20)$$

At the limit of breakdown, Eqn. (4.4.7) may be substituted into Eqn. (4.4.20).

$$\tau_p = \frac{K_{\text{displacement}}}{v_n + v_p} \frac{1}{\beta} \quad (4.4.21)$$

This equation does not appear elsewhere in the literature.

Eqn. (4.4.21) is an easily evaluated analytical expression. The primary avalanche buildup time is related to the gain-bandwidth product (GBW) through the simple relation

$$\text{GBW} = \frac{1}{2\pi\tau_p} \quad (4.4.22)$$

There is no explicit dependence in Eqn. (4.4.21) on the multiplication region width, which is different from other treatments of this subject. The usual claim is that GBW varies with carrier velocity and inversely with multiplication region width and k with a proportionality constant that varies

4. Separate absorption, grading, charge and multiplication APDs

slowly as k varies from 1 to 0. Comparing this to Eqns. (4.4.21) and (4.4.22), the carrier velocities are present, but k and w are absent. Actually k is present in the expressions derived here through $K_{\text{displacement}}$, to be discussed below in more detail. However at first glance it appears that w is not present in Eqn. (4.4.22). Actually, β depends on the E-field at breakdown, which in turn depends on w . As w decreases, β increases, in agreement with other treatments. Eqn. (4.4.22) would be consistent with standard literature treatments provided an inverse proportional relationship exists between w and β . This may be seen for two limiting cases. For the case of $\alpha=\beta$, the expression for the gain in Eqn. (2.3.23) depends on the product βw , and thus for a given gain, an inverse proportional relationship exists between w and β . At the other extreme, for the case of $\alpha=0$, the gain is seen from Eqn. (2.3.22) (substituting hole for electron injection since α , not β is zero) to be $\exp(\beta w)$, which again shows an inverse proportional dependence between w and β . For the real case of $\beta>\alpha$, where α is not zero, it is much more difficult to establish whether or not the same relationship exists between w and β .

The GBW has been plotted for the ideal δ -doped bulk InP multiplication APD in Fig. (4.4.4). The electron and hole velocities were taken to be 60 and 30 $\mu\text{m}/\text{ns}$ respectively to permit comparison with [31]. For $w=0.2 \mu\text{m}$, $\text{GBW}=108 \text{ GHz}$, three-quarters of the 143 GHz reported previously by Kuchibhotla and Campbell. At other values of w , the three-quarters relation seems to hold as well. It is not apparent why there is a difference. As mentioned above, there is agreement between these simulations and those of Kuchibhotla and Campbell [31] with respect to the E-field.

4. Separate absorption, grading , charge and multiplication APDs

Figure 4.4.4 Gain-bandwidth product vs. w for an ideal δ -doped APD with bulk InP ionization only.

4.4.7 Displacement correction factor

In this section the displacement correction factor $K_{\text{displacement}}$ is discussed, which arises from the displacement current resulting from the space charge associated with the free carriers which pass the current through the depletion region. $K_{\text{displacement}}$ is given by rearrangement of the equations of Kuvas and Lee [79] for the case of the p-i-n structure with constant ionization coefficients to be

$$K_{\text{displacement}} = \frac{(1 + \frac{v_p}{v_n})^2}{\frac{v_p}{v_n}} \frac{k}{(1 - k)^2} \left[\frac{1 + k}{1 - k} \ln\left(\frac{1}{k}\right) - 2 \right] \quad (4.4.23)$$

4. Separate absorption, grading, charge and multiplication APDs

It is worth noting that the value of $K_{\text{displacement}}$ is independent of the direction from which the carriers are injected or whether α is greater or less than β - i.e. $K_{\text{displacement}}(v_p, v_n) = K_{\text{displacement}}(v_n, v_p)$ and $K_{\text{displacement}}(1/k) = K_{\text{displacement}}(k)$. Note that $K_{\text{displacement}}$ is indeterminate at $k=0$ and $k=1$, so it is worth considering these limits, which has not been done in the literature.

$$\begin{aligned}
 \lim_{k \rightarrow 0} (K_{\text{displacement}}) &\sim \lim_{k \rightarrow 0} (-k \ln(k)) = \lim_{k \rightarrow 0} \frac{-\ln(k)}{\frac{1}{k}} \\
 &= \lim_{k \rightarrow 0} \frac{-\frac{1}{k}}{-\frac{1}{k^2}} \\
 &= k
 \end{aligned}$$

where L'Hospital's rule for indeterminate limits has been used. So

$$\lim_{k \rightarrow 0} (K_{\text{displacement}}) = 0 \quad (4.4.24)$$

Define $\lim_1$ such that

$$\begin{aligned}
 \lim_1 &= \lim_{k \rightarrow 1} \left(\frac{k}{(1-k)^2} \left[\frac{1+k}{1-k} \ln\left(\frac{1}{k}\right) - 2 \right] \right) \\
 &= \lim_{k \rightarrow 1} \left(- \frac{(1+k) \ln(k) + 2(1-k)}{(1-k)^3} \right) \\
 &= \lim_{k \rightarrow 1} \left(- \frac{\ln(k) + k \ln(k) + 2 - 2k}{(1-k)^3} \right) \\
 (\text{L'Hospital}) &= \lim_{k \rightarrow 1} \left(\frac{\frac{1}{k} + \frac{k}{k} + \ln(k) - 2}{3(1-k)^2} \right) = \lim_{k \rightarrow 1} \left(\frac{\frac{1}{k} + \ln(k) - 1}{3(1-k)^2} \right)
 \end{aligned}$$

4. Separate absorption, grading, charge and multiplication APDs

Figure 4.4.5 Displacement correction factor vs. k for different carrier velocity ratios.

$$\begin{aligned}
 (\text{L'Hospital}) &= \lim_{k \rightarrow 1} \left(\frac{-\frac{1}{k^2} + \frac{1}{k}}{-6(1-k)} \right) = \lim_{k \rightarrow 1} \left(\frac{-\frac{1}{k^2}(1-k)}{-6(1-k)} \right) \\
 &= \frac{1}{6}
 \end{aligned} \tag{4.4.25}$$

Then

$$\lim_{k \rightarrow 1} (K_{\text{displacement}}) = \frac{1}{6} \frac{\left(1 + \frac{v_p}{v_n}\right)^2}{\frac{v_p}{v_n}} \tag{4.4.26}$$

$K_{\text{displacement}}$ is plotted vs. k for different carrier velocity ratios in Fig. (4.4.5), with treatment similar to that of Kuvas and Lee [79]. The various constant carrier velocity ratio curves are spaced by nearly constant logarithm

4. Separate absorption, grading, charge and multiplication APDs

(1, 0.5, 0.2, 0.1, 0.05, 0.02 and 0.01). It is evident the curves are evenly spaced except for the curves for ratios of 0.5 and 1, which are more closely spaced than the rest. To see a substantial decrease in $K_{\text{displacement}}$ it is necessary to decrease k substantially below 0.5.

As k tends to zero (the single carrier ionization limit), $K_{\text{displacement}}$ tends to zero in a constant slope logarithmic relationship which makes τ_p (Eqn. (4.4.21)) correspondingly small and GBW (Eqn. (4.4.22)) correspondingly large. This is one of the strong reasons for ongoing research into low k materials for APDs, the other reason being the noise. Note that there is not a tradeoff between GBW and noise: both improve as k decreases.

For constant v_n and v_p , $K_{\text{displacement}}$ increases from zero with increasing k until the limit expressed in Eqn. (4.4.26). With respect to v_n and v_p , $K_{\text{displacement}}$ (and hence GBW) is at a minimum for equal carrier velocities. Note, however, that upon substituting the expression for $K_{\text{displacement}}$ (Eqn. (4.4.23)) into the expression for GBW (Eqn. (4.4.22)), it becomes evident that increasing either the v_n or v_p will increase GBW despite the increase in $K_{\text{displacement}}$. The dependence of $K_{\text{displacement}}$ on the carrier velocity ratio is plotted in Fig. (4.4.6) for various k . The curves are spaced further apart than for $K_{\text{displacement}}$ vs. k . For the velocity ratio < 0.5 there is a constant logarithmic slope dependence, with $K_{\text{displacement}}$ decreasing as the carrier velocity ratio decreases. Very little difference is seen from a ratio of 0.5 to 1, indicating that if the carriers are of comparable velocity there will be little effect on $K_{\text{displacement}}$ for any uncertainty in that ratio.

$K_{\text{displacement}}$ is plotted against x_d for an ideal δ -doped APD with ionization assumed to come only from the InP in Fig. (4.4.7). It is seen that a maximum of 0.75 is present at $x_d = 0.1 \mu\text{m}$, for which $k=1$ using Osaka's ionization coefficients. As mentioned above, $K_{\text{displacement}}$ is symmetric with

4. Separate absorption, grading, charge and multiplication APDs

Figure 4.4.6 Displacement correction factor vs. carrier velocity ratios for different k .

Figure 4.4.7 Displacement correction factor vs. x_d for carrier velocity ratio of 0.5.

4. Separate absorption, grading, charge and multiplication APDs

respect to k or $1/k$, so as k deviates from 1 in either direction, $K_{\text{displacement}}$ decreases. Note that even as k decreases to 0.41 for $M=10$ and $x_d=1.5\text{ }\mu\text{m}$, $K_{\text{displacement}}$ decreases to 0.69, only a 7.6 % decrease. Within the operating range discussed in this work for InP/InGaAs APDs, $K_{\text{displacement}}$ could well be treated as a constant.

4.5 Possible fabrication sequences

There are a number of alternatives for fabricating an SAGCM APD. The two most likely candidates are ion implantation [27] and MOCVD charge sheet incorporation [28-30].

For ion implantation, a blanket implant may be done for the peripheral charge sheet following the growth of the InGaAs and grading layers and some thickness of InP, as illustrated in Fig. 4.5.1. The energy of the implant may be adjusted to the thickness of InP within reasonable bounds. It is difficult to implant Si or any other n-type dopant into InP for a depth greater than $1\text{ }\mu\text{m}$, so the InP should not be thicker than this. As pointed out by Webb, having some of the implant distribution overlapping with the InGaAs can be beneficial. Following this step, a selective implant for the difference in charge between the total and peripheral charge sheets is done in order to place the total charge within the center of the device. In the case of no peripheral charge, this is the only implantation step. A second growth is performed at this time.

For growth incorporation, a blanket charge equivalent to the total charge is deposited, as illustrated in Fig. 4.5.2. A mesa is formed of such depth that the total charge remains in the central region of the device and the peripheral charge remains in the device periphery. A second growth is performed at this time.

4. Separate absorption, grading , charge and multiplication APDs

Figure 4.5.1 Schematic diagram of ion implanted SAGCM APD.

Figure 4.5.2 Schematic diagram of MOCVD charge sheet SAGCM APD.

4. Separate absorption, grading, charge and multiplication APDs

5. Device fabrication and characterization

In order to experimentally evaluate the validity of the ideas developed in the previous sections, APDs have been fabricated. In this section, a description of the fabrication details is presented, along with some of the analysis used to assist with and/or evaluate the fabrication. The APD process is in part based on a previously developed p-i-n photodiode process [19]. A schematic diagram of the complete structure is shown in Fig. 5.1.1.

Figure 5.1.1 Schematic cross-section of a fully processed SAGCM APD.

5.1 Fabrication

5.1.1 Mask set

Two mask sets were used, known as T79 and Z01. T79, the first mask set used included devices of various sizes (all given here as nominal dimensions). Mesas were of diameters 30, 40, 50, 60, 70, 80, 100, 120 and 150 μm , which define the active region (i.e. the region with gain). The diffusion masks were of annulus 5, 10, 15 and 20 μm larger than the mesa size, thus providing a 9x4 matrix of devices. To ease the tolerance of the photolithographic alignment the mesa sizes were "biased" inward by 2 μm on each side with the result that the actual mesa sizes were 4 μm smaller in diameter than the nominal sizes. The dropouts to ease photolithographic alignment were as small as possible, minimizing the area used on the mask. Z01, the second mask set used, consisted mainly of devices of actual mesa diameter 30 μm and annulus 5 μm . The dropouts were 5 x 5 mm, partitioned into areas which permitted post-fabrication analysis. This area used was at the expense of potential device area. Either whole or quarter wafers were processed. The wafer quarters were designated A, B, C or D in counterclockwise fashion with respect to the wafer flat as indicated in Fig. 5.1.2.

5.1.2 MOCVD first growth

There are many possible growth techniques, including liquid phase epitaxy, MOCVD, vapor phase epitaxy, chemical beam epitaxy and molecular beam epitaxy. Each technique has its own merits. In this work, MOCVD was used.

Figure 5.1.2 Schematic diagram of a full 2" InP wafer showing the conventions used in this work for describing quarter wafers and tiles on the whole wafer. The direction of MOCVD gas flow is also shown.

Epitaxial growth was performed in a modified CVT horizontal reactor. Lattice matching was achieved to within $\Delta a/a = 0.07\%$, where a is the lattice constant, as verified by double crystal x-ray diffractometry. A run number (e.g. O770) was assigned to each growth run for easy reference. The background carrier concentration of undoped material was found to have an

5. Device fabrication and characterization

upper limit of 10^{15} cm^{-3} . This is an improvement of a factor of three over Webb's devices, allowing for an E-field slope of $\leq 1.5 \text{ V}/\mu\text{m}^2$ in the multiplication and absorption regions. The direction of gas flow was such that the wafer flat formed the trailing edge, as indicated in Fig. 5.1.2; the A and D quarters were downstream of the B and C quarters. Morphology of the layers was monitored by visual inspection under the microscope. The layer sequence was as follows: an InP buffer layer of nominally $1 \mu\text{m}$, an InGaAs absorption layer of nominally $2.5\text{--}3.3 \mu\text{m}$ and a grading region (consisting of one or more layers). These layers were all of unintentional (background) doping for the initial tests and first full mesa devices. The grading layer was either a superlattice (SL) stack graded in thicknesses from the InGaAs to the InP, or one or more quaternary layers. On top of the grading layer, the charge sheet of InP was formed, sometimes by two- or three-step doping, for reasons which are discussed below. For the case of the SL grading layer, the growth temperature was 605°C ; otherwise the growth temperature was 620°C .

Two different reactors were used in this work, the O/P and R1 reactors, and wafers are appropriately titled. The O/P reactor can grow one wafer at a time, whereas the R1 reactor can grow two wafers, denoted as the leading (L) and trailing (T) wafers respectively. The O reactor underwent major renovations and was afterward known as the P reactor. This reactor in either incarnation possesses the ability to switch immediately between any two InGaAsP compositions, but is unable to switch from InGaAs to InGaAsP to InP without some stagnant (non-growth) time in the second transition. This is undesirable from the point of view of growing a single quaternary (1Q) grading layer APD. The R1 reactor possesses the ability to switch from InGaAsP₁ to InGaAsP₂ arbitrarily fast (where InGaAsP₁ is any InGaAsP composition lattice matched to InP), and the ability to switch from InGaAsP₁

5. Device fabrication and characterization

to InGaAsP₂ to InGaAsP₃ within one minute, which is the stabilization time required for the mass flow controllers which govern InGaAsP₁ to switch to the new composition InGaAsP₃. The growth rate is approximately 300 Å/min, which means that InGaAsP₂ must be at least 300 Å thick.

5.1.3 Mesa formation

The mesas were patterned by standard photolithographic techniques. The 405 or 290 nm line of a Hg arc lamp in a Karl Suss MJB3 aligner was used to expose Shipley Microposit positive photoresist through a light field (LF) mask with mask set T79 or a dark field (DF) mask with mask set Z01. A standard NaOH based developer was used, which dissolves exposed photoresist. Following the patterned definition, the mesas were transferred from the photoresist mask to the semiconductor material by means of either a wet chemical etch (WCE) or reactive ion etching (RIE). In the case of RIE either a Cl-based etch or a CH₄-H₂ etch was used. Following the mesa formation the remaining photoresist mask was removed using PRS-2000 photoresist stripper and rinsed in deionized (DI) water. In some cases an O₂ descum was used prior to the etch.

5.1.4 MOCVD overgrowth

The wafers were cleaned and placed in the MOCVD reactor. Prior to the overgrowth a thin Si layer was unintentionally incorporated into the charge sheet. The existence of this effect is well known although the exact mechanism by which this occurs is not understood [168, 169]. This can be an uncontrolled process, and extremely minute amounts of Si can cause a substantial effect. Also the cleaning procedure removed a thin uncontrollable portion of the charge sheet. These two offsetting effects contributed an uncertainty of 5-10% to the total charge sheet dose that was initially

5. Device fabrication and characterization

unforseen. One way around this problem was to spend time optimizing the cleaning procedure for minimal unintentional Si incorporation, and including a 20-25-nm thick undoped InP layer above the charge sheet. This thickness is slightly greater than the maximum measured amount of material removed with the cleaning procedure. Providing the cleaning procedure consistently removed less than 20 nm, the material unintentionally removed had almost no effect on the charge sheet. Following the growth of nominally 2.5 μm of unintentionally doped InP, the wafer was removed from the reactor and examined under the microscope. In many instances concentric growth rings oriented off axis from the direction of gas flow were observable under a Nomarski microscope.

Prior to further processing, a fringe map of the InP thickness [32-37] was generated and the photoluminescence efficiency of the InGaAs layer was assessed using a scanning photoluminescence apparatus [34]. As a fringe map gives a relative measure of thickness across the wafer, the reflection spectroscopy technique [32-37] was used for at least one point on the wafer to provide an absolute measure of the InP thickness across the wafer.

The nomenclature used in this work is to specify a wafer by two run numbers corresponding to the second and first growths respectively: e.g. P1/R1-1 would have a first growth of R1-1 and the second growth of P1. If only one reactor is specified, then both growths took place in the same reactor: e.g. P2/1 means a first growth of P1 and a second growth of P2.

5.1.5 Zn diffusion

A dielectric mask consisting either of P-doped SiO_2 deposited by an Applied Materials SILOX reactor, SiO_2 deposited by a PlasmaTherm plasma enhanced chemical vapor deposition (PECVD) reactor or SiN_x deposited

either by an electron cyclotron resonance (ECR) reactor or a PlasmaTherm PECVD reactor was applied to the overgrown wafer. Using the same photolithographic procedure as above, vias in the photoresist were patterned through a dark field (DF) mask and transferred to the Zn diffusion mask either by WCE using 1:10 $\text{NH}_4\text{F}:\text{HF}$ in the case of SiO_2 or RIE using CF_2H_2 in the case of SiN_x . Following a wafer cleaning step, the Zn diffusion was performed. The Zn diffusion chamber was either purged with N_2 or evacuated and held under vacuum. The wafer was probed for I-V characteristics following the Zn diffusion to determine suitability for further processing. The Zn diffusion conditions were determined by reference to calibration samples, whose junction depths were measured either by secondary ion mass spectroscopy (SIMS) profiles or electrochemical C-V profiling using a Polaron machine. In either case, the crater depth created by the destructive technique was measured by stylus profilometry using a Dektak machine.

5.1.6 AR coating

An antireflection (AR) coating of SiO_2 or SiN_x was applied to the wafer, again either by PECVD or ECR. The thickness was chosen as nominally a quarter wavelength in the material. Small deviations from this value have a small effect on the responsivity of the photodetector.

5.1.7 p-contact

Using the same photolithographic procedure as above contact windows were etched into the dielectric coating by RIE. A standard liftoff technique using a bilayer of polymethyl-methacrylate (PMMA) and photoresist was used to pattern the p-metal, which consisted of 40nm of Cr followed by 250 nm of Au, into contact with the wafer surface. The metals were evaporated by line of

5. Device fabrication and characterization

sight using an e-beam evaporator. The background pressure of the evaporator was nominally 10^{-7} Torr. In some runs a p-metallization consisting of Pd/Zn/Pd/Au deposited by e-beam evaporation and/or Ti/Pt/Au deposited by sputtering was used.

5.1.8 n-contact

The wafer was lapped and polished to a thickness of 150-200 μm to facilitate cleaving of the finished devices. Prior to this step the wafer was quartered to facilitate an even polishing thickness and to avoid wafer breakage. An n-metal of Cr/Au or Au-Ge/Cr/Au was evaporated onto the opposite side of the wafer to that which had been previously processed. The wafer was alloyed for 2 minutes at a temperature of 425°C . The whole wafer could then be probed for I-V, M-V and C-V characteristics.

5.1.9 Dicing

The finished wafers, once probed, were diced into $500 \times 500 \mu\text{m}$ dies. This was accomplished by cleaving the quarters into bars and the bars into chips. Sometimes this was done on an adhesive ("blue") tape using a Loomis dicing machine.

5.1.10 Mounting

The dies were extracted from either a chip tray or blue tape by means of a vacuum "pencil" and mounted onto a ceramic carrier, typically $2 \times 2 \times 4 \text{ mm}$ in size. Au/Sn solder was thermally compressed between the die and the carrier to form a bond between the n-contact and a gold pad on the carrier. 1.0 or 1.4 mil (25 or 35 μm diameter) wire was used to connect the p-contact with another contact on the carrier. The wire length was typically $\sim 2 \text{ mm}$. The chip on carrier could be characterized for frequency response and noise.

5.2 Evaluation prior to full processing

Once the wafers are grown and processed, they must be evaluated. In order to improve the feedback time and omit unnecessary processing steps which consume valuable processing resources, there has been an emphasis in industry and in the literature on developing techniques for rapid evaluation of wafers, before processing is completed. In this way more information comes to be available and more statistics can be accumulated than otherwise. Of course many key measurements still require fully processed devices mounted on a ceramic carrier.

5.2.1 Photoluminescence intensity - nondestructive

It has been recently shown that the photoluminescence (PL) intensity may be correlated with the leakage current of a p-i-n detector, and is a measure of the quality of the InGaAs layer [34, 129-131]. The leakage in the bulk InGaAs of a p-i-n structure is directly related to the photoluminescence (PL) of InGaAs detected near the bandgap wavelength when the top surface of the wafer is illuminated with light of a wavelength less than the bandgap wavelength. If 0.633 μm HeNe radiation is used, the light is absorbed in the InP above the InGaAs, creating carriers which diffuse into the InGaAs, where they are recombined due to the lower bandgap. The carriers recombine radiatively or non-radiatively. The key contributors to this intensity are the surface/heterojunction recombination velocity and the lifetime of the carriers in the bulk InGaAs. If the surface/heterojunction contribution is negligible, then the higher the PL intensity the lower the I_d in p-i-n detectors. This technique allows screening of wafers prior to any processing, and at various stages in the process to evaluate both the material and the influence of processing steps.

5. Device fabrication and characterization

The applicability of this technique to APDs, in which the multiplied dark current comes from the InP, could in principle be a separate issue. Note that since InGaAs has a much lower bandgap the initial expectation is that the major portion of the dark current contribution from generation-recombination and diffusion mechanisms should come from the InGaAs [43].

5.2.2 Epitaxial layer thickness measurement - nondestructive

The control of Zn diffused junction depth is important for the p-i-n, more critical for the SAGCM structure, and absolutely crucial for the SAGM structures. The epilayer thicknesses, and for the APDs, the top InP thickness in particular, must be accurately known in order to determine how deep the Zn diffused junction should be. Traditionally this has been evaluated destructively by scanning electron microscopy (SEM). Notwithstanding the question of interpretation of some SEM photographs, a whole wafer cannot be processed following SEM analysis, and moreover the SEM provides information in one dimension only. To overcome these difficulties the technique of reflection spectroscopy has been recently introduced by L. Tarof et al. [32-37]. This technique is non-destructive and makes use of the optical thin film properties of semiconductors at wavelengths above their bandgap.

For the APD this technique can in principle be used in several situations: to measure

- 1) InP charge sheet thickness measurement on and off the mesa before regrowth (neglecting grading layer (s))
- 2) total InP thickness (above the InGaAs) on and off the mesa after regrowth
- 3) InGaAs thickness (neglecting grading layers)

5. Device fabrication and characterization

4) InGaAs and quaternary grading layer(s) thicknesses.

For measurement of the initial charge sheet thickness and total InP thickness above the InGaAs following the regrowth, the reflectance is measured in the spectral range 0.9-1.5 μm , where the InP is transparent but the InGaAs is opaque [32-37]. This is done at a few selected points on the wafer. To map the entire wafer, the reflectance at 1.15 μm wavelength (measured using a He-Ne laser) is recorded by scanning the entire wafer using a recently described scanning photoluminescence system [34].

The InGaAs thickness may be measured in addition to the above mentioned InP thickness following the regrowth by using wavelengths in the infrared. An algorithm for measuring both thicknesses of an "A-B-A" double heterostructure has been recently developed by the author, suitable for this purpose [32,33]. The technique has been demonstrated successfully for several compositions of InGaAs(P) as the B layer on unpatterned wafers, including "standard" APD structures, using a Fourier Transform Infrared (FTIR) spectrometer with a 6 mm spot size. In order to apply the technique to measure a patterned wafer (such as after overgrowth), the spot size must be sufficiently small to resolve the pattern, so this measurement has not been done on actual processed mesa APD wafers. The inability to resolve the pattern is not a limitation of the technique but only of the apparatus available to the author: FTIR permits a 25 μm spot size with a proper focusing system.

For a single quaternary grading layer or sequence of quaternary grading layers monotonically varying in As mole fraction from InGaAs to InP (assumed lattice-matched), it is in principle possible to use a spectral range with a lower limit of 0.9 μm and an upper limit greater than the bandgap wavelength of InGaAs (e.g. 2 μm). As the wavelength increases, first the InP and then the grading layer(s) (in sequence) become transparent, which in

5. Device fabrication and characterization

principle allows for the determination of each layer thickness. If the thickness is to be determined neglecting the complex index of refraction, the individual layers must be thick enough to be absorbing below the bandgap wavelength, which would not be the case for the APDs. Otherwise the complex refractive index must be used.

The routine use of this technique for measuring the InP thickness alone relaxes somewhat the tolerance for reproducibility of the MOCVD epilayer thicknesses themselves since x_d is the parameter of interest and x_j is independently controllable. The requirement for uniform epilayers still remains intact. The implementation of the reflection spectroscopy (or other) technique to determine the other layer thicknesses would be a further improvement.

In this work, the reflectance topography of the InP top layer thickness was usually obtained by using a scanning reflectance setup which mapped the reflectance on a tight grid of $500 \times 500 \mu\text{m}$. The "fringe map" was used to evaluate the InP thickness uniformity. A typical fringe map is shown in Fig. 5.2.1. It was generally found that the thickness pattern was nearly radial, with a typical variation of $0.2\text{-}0.3 \mu\text{m}$ from the center to 5 mm from the edge and a variation of $\pm 0.1 \mu\text{m}$ from the center to 8 mm from the edge all in the direction of gas flow, with somewhat more variation parallel to the direction of gas flow. It was found that the sharpness of the interference fringes was reduced as the quality of the grading layer improved. This is not surprising, as for an ideal ramp between the InGaAs and InP, there would be a distributed rather than a sharp interface for the location of the reflection on which this technique relies.

5. Device fabrication and characterization

Figure 5.2.1 Typical fringe map of InP top layer thickness. The difference in thickness between adjacent fringes (black or white) is $0.19\text{ }\mu\text{m}$.

5.2.3 Hall measurements for doping level

The Hall measurement technique is well known for measuring doping levels in semiconductors. The particular setup used incorporated a van der Pauw measurement of cleaved semiconductor approximately square of length 4 mm. The contacts were 250 μm diameter Sn solder, and a geometrical correction (the van der Pauw "f-factor") was applied for asymmetric currents owing to the sample not being a perfect square. The magnetic field strength was 0.3 T.

In this case, two items are of interest: the background doping and the intentional doping.

Background doping measurements indicate a carrier concentration of 10^{15} cm^{-3} , to within a factor of 2. This reflects the variation from wafer to wafer as well as the difficulty of measuring carrier concentrations so low, which are at the limit of the measurement technique.

The reactor doping profiles are shown for both the P and R1 reactors in Figs. 5.2.2 and 5.2.3, respectively, for the same nominal range of carrier concentrations. The layer thickness for these samples was 2-3 μm . It is evident that there is a trend from the leading to trailing edge on any wafer for greater dopant incorporation. The absolute doping level is a function of the mass flow controller settings, and is monotonic but not quite linear. For the reactor doping profiles shown here, the MOCVD gas flow was not the same for every run. For the P reactor, the doping dependence may be described as $\pm 10\%$ in the direction of gas flow from the center to 5 mm from either edge. For the R1 reactor, the doping dependence may be described as $\pm 10\%$ in the direction of gas flow from the center to 5 mm from either edge for the leading wafer, and $\pm 5\%$ for the trailing wafer.

Figure 5.2.2 Doping profile from Hall measurements for the P reactor.

Figure 5.2.3 Doping profile from Hall measurements for the R1 reactor.

5. Device fabrication and characterization

The reactor doping profile is correlated with the susceptor temperature profile. The temperature is constant to within ± 1 °C over the entire range studied, which represents the limit of temperature uniformity possible in these reactors. These observed variations across the reactor reflect the difficulties in maintaining sufficiently constant temperature. Experimentally it is found that if the growth temperature is reduced, the doping profiles become even more non-uniform.

5.2.4 I-V measurement following Zn diffusion

Experimentally it has been shown that I-V characteristics of p-i-n photodiode wafers may be probed at room temperature following the Zn diffusion step. This permits evaluation to determine whether or not it is worthwhile to carry on with further processing, or better to expend processing resources on another wafer. For APDs the same measurement could be done, but not as reliably as with p-i-n wafers. Similarly the responsivity and gain may be evaluated crudely as a function of voltage. For high power levels a pulsed measurement could sometimes be done, but this was not a routine process.

5.3 Evaluation of fully processed wafers

Fully processed devices with metalized contacts are required to measure the capacitance and final I-V. Speed and gain are best measured on a fully processed device mounted on a ceramic carrier.

5.3.1 I-V, M-V, C-V

The I-V and M-V characteristics were initially measured on the wafer using a Hewlett-Packard (HP) 4142 meter, which permits measurement down to the pA range. The wafer was kept in a box shielded from unwanted room light during the measurement. The I-V curves were measured without light, and the M-V curves were measured using 1.3 μm wavelength light incident from a fiber which was manually aligned to the device using a microscope attachment. As the light was of non-normal incidence owing to space restrictions, this measurement indicates less gain for the smaller devices than would be evident for normal incidence, due to the spot size. The I-V and M-V measurements for a chip on carrier were done at normal incidence using a non-automated setup.

The capacitance was measured using a number of different setups. The C-V measurements shown in this work were taken using a HP 4275 parameter analyzer at a frequency of 1 MHz. It was found that there were very small differences, if any, between measurements taken at different frequencies up to 10 MHz. As well, there were only minor differences associated with the small signal test voltage magnitude, from 0.025 to 0.1 eV.

5.3.2 SIMS

SIMS is a useful technique for accurately measuring x_d after the Zn diffusion. The depth axis of the SIMS profile is calibrated by a subsequent Dektak measurement of the sputtered crater. Profiles of Zn and Ga or Si were taken. A typical set of profiles is shown in Fig. 5.3.1. The decrease in Zn concentration marks x_j and the increase in Ga shows the grading layer location, from which the known thickness of the InP on the mesa can be subtracted to give x_d . The increase in Si concentration provides an unambiguous location of the top of the charge sheet. Obviously the Si profile gives a more direct indication of x_d than does the Ga profile, so one might wonder why the Ga profile is pertinent. There are two reasons. First, the Ga marker is useful as a reference for the Si to indicate any unintentional process deviations which might otherwise go undetected. The second is that Zn profiles are collected using positive ion bombardment from the Cs gun of the SIMS machine, whereas Si profiles are collected using negative ion bombardment. This requires two separate runs, and the Ga marker is used for both runs to ensure no unintentional thickness differences are registered. In this case x_d is measured as accurately as possible. The Z01 mask set was designed to permit both measurements, but the T79 mask set was not; therefore x_d measurements are more accurate from the Z01 mask set than the T79 mask set. With reference to the SIMS profiles of Fig. 5.3.1, it is evident that the falloff in the Zn concentration, which in theory obeys a "modified erfc" [165, 166] (erfc function for a single diffusion mechanism and doping independent diffusion coefficients [167]) profile down to the limit of measurement, does not fall off very rapidly. Therefore an arbitrary choice of x_j must be made. Although the standard definition of x_j would be the point

Figure 5.3.1 Typical SIMS profiles showing x_d measurement.

where the diffused impurity concentration is the same as the background, there are a number of reasons to view this definition with caution in interpreting the SIMS profiles. First, there is a broadening of the profile associated with the SIMS measurement itself. Second, there is some uncertainty as to the absolute concentrations levels of Zn and Si, so the point of exact compensation cannot be precisely determined. Finally, the definition of an ideal SAGCM structure assumes a p^+-n^- junction of junction depth x_j , whereas the actual junction is p^-n^- : a point where the E-field is falling off abruptly – i.e. the electrically active Zn concentration is sufficiently high – might be more appropriate for comparison with the model than the p^-n^- junction. Thus the accuracy of x_d is $\pm 0.1 \mu\text{m}$.

5. Device fabrication and characterization

Figure 5.3.2 x_j measurement across a 2" wafer.

A typical plot of x_j with respect to wafer position compiled from many SIMS profiles is shown in Fig. 5.3.2. It is evident from that the Zn diffused junction depth is uniform across a 2" wafer to within $\pm 0.1 \mu\text{m}$ for a depth of $2 \mu\text{m}$. This is approximately the resolution of the measurement itself, so the conclusion is that the junction depth is uniform to within the limit of the measurement across the wafer. This also implies that the thickness profile as shown in Fig. 5.2.1 is in practice also the x_d profile of the wafer.

5.3.3 Frequency response

The frequency response was usually measured using a HP Lightwave Component Analyzer. Essentially this is a network analyzer with the capacity

for electrical/electrical, electrical/optical, optical/electrical or optical/electrical input/output of the signal, compared with the electrical/electrical configuration possible for a network analyzer alone. The APD frequency response was measured in the optical/electrical configuration, and compared against a calibrated reference p-i-n photodiode within the unit supplied by HP. Two models were used in this work, the HP 7302 unit which has a frequency response of 3 kHz to 6 GHz, and the HP 7303 unit which has a frequency response of 130 MHz to 20 GHz. In all cases the small signal analysis was done using sinusoidal modulation of the light.

5.3.4 Noise

The noise was measured with a HP noise figure meter, using "Special Function #9.3". Light at 1.3 μm wavelength was used to measure the noise in the frequency vicinity of 20 MHz and with a bandwidth of 4 MHz. In this way the noise measurement would be nearly independent of the frequency response of the device. If higher frequencies are used to measure the noise, it is possible that less noise would be measured not because of the low noise in the device, but rather because the noise measurement is itself limited by the frequency response of the device. The output of the noise figure meter (called "noise figure" by HP, but not really the noise figure in other uses of the term) was measured vs. the gain.

6. Initial experiments

In this section some initial experiments on different possible fabrication methods and different structures are described.

6.1 Preliminary structures

In this section the results on some preliminary structures are described. No attempt to avoid edge breakdown was made, as this made it easier to affect certain comparisons with the aim of making processing/design choices for complete structures.

6.1.1 Preliminary SAGM APD

Both O613 and O618 are preliminary standard APD structures, without guard rings. The processing was similar to that of a p-i-n detector, with the exception that the Zn diffusion was placed as accurately as possible into the InP, 0.5 μm above the InGaAs/InP interface. The principal difference between O613 and O618 was the presence of a SL grading layer in O618 at the InGaAs/InP interface. For O618 the growth temperature was reduced by 20° C to obtain a sufficiently high photoluminescence. One quarter of each wafer (the "A" piece) was processed.

The I-V characteristics of the wafers are shown in Fig. 6.1.1. The leakage current was typically in the 1-10 nA range at 0.9 V_b . A maximum gain of 3-5 could be measured in the central device region prior to breakdown for a selected area of the wafer, but no higher as corner breakdown dominated. There was a spatially rapid change of gain/photoresponse at a given reverse bias which was random on O613A, monotonic across O618A. This variation was correlated with variation in x_d , as determined by SIMS. For the "a" trace in this figure, no photoresponse was observed, and $x_d = 1 \mu\text{m}$. For the region corresponding to the "b" trace in this figure, photoresponse was observable

Figure 6.1.1 I-V characteristics of preliminary SAGM structures from O618A with SL grading layer. There was no attempt to suppress edge breakdown. The traces are taken from three different areas of the wafer.

for reverse bias greater than 5-10 V, which varied rapidly in this region. A maximum gain of 3-5 could be measured in the central device region prior to breakdown for this area of the wafer, but no higher as corner breakdown dominated. For this region of the wafer, $x_d = 0.5 \mu\text{m}$. For the "c" trace in this figure, photoresponse was observed at nearly zero bias, but no gain was observed. For this region of the wafer, $x_d = 0$. The observations connected with these three regions are consistent with the discussion surrounding Fig. 3.2. For O613A, the I-V curves were similar to the "b" trace in this figure. Note that on O618A, x_d varies by $1 \mu\text{m}$, an order of magnitude greater than the known epilayer variation. Therefore the variation appears to be due to Zn diffusion non-uniformity in this case. Previous and subsequent to this wafer

6. Initial experiments

Zn diffusion/SIMS studies were done which show far more uniform Zn diffusion than in O618A. The cause of the poor uniformity in this case is still unknown, but appears to be an anomaly rather than a trend.

Some high speed measurements were done on mounted devices from both O613A and O618A. The fall time for the SAGM SL structure (O618A) was 0.3 ns, compared with 2 ns for the SAGM structure without the SL (O613A). This is consistent with Forrest and Kim's observation that a grading layer can improve the device response time. It should be pointed out, however, that since the variation in x_d was random in O613A it is not clear that the E-field at the upper InP/InGaAs heterointerface was similar in both devices. Therefore the comparison, while indicative, is not absolutely conclusive in itself.

6.1.2 Preliminary SAGCM APDs

A preliminary SAGCM structure was investigated with O619B (mesa, MOCVD doped charge sheet) and O620 (implant). In both cases the doped region was blanket rather than selective, distributed approximately 80% in the InP, 20% in the InGaAs. The implanted wafer required two growths, but only one growth was used for O619. Similar to O613A,618A, these wafers were processed as p-i-n detectors with the Zn diffusion placed in the InP, not the InGaAs.

6.1.2.1 MOCVD charge sheet

The I-V characteristics for O619B (mesa, MOCVD doped charge sheet) are shown in Fig. 6.1.2. The I-V characteristics are uniform within a given "tile" (5 mm square), and a slow variation across the wafer is observed. The leakage current is excellent, rising from 10 pA at low voltage, where the depletion region is confined to the InP, to 0.1-1 nA at higher voltages (still prior to breakdown), where the depletion region extends into the InGaAs.

6. Initial experiments

Figure 6.1.2 I-V curve for preliminary planar blanket MOCVD charge sheet structure. There was no attempt to suppress edge breakdown.

These are the very low leakage currents, and the lowest of any APD material processed within this preliminary evaluation, and suggest that MOCVD doping may be an effective way to control σ while not having an adverse effect on the leakage current.

As this was only a preliminary structure, the gain on O619B in the central device region was no more than 1.1 prior to breakdown. It is not surprising that O619 should show considerably less gain than O613 or O618. For similar Zn diffused radius of curvature, corner breakdown is more rapid in lower doped material [23,26,87-91], and so insufficient E-field builds up in the active region of the device for substantial gain to occur. For a device with edge breakdown suppressed, the gain in the center of the device would be

6. Initial experiments

expected to dominate the multiplication characteristics, so the gain prior to edge breakdown should not be an issue.

6.1.2.2 Ion implantation

The leakage for O620C, shown in Fig. 6.1.3 (a), is poor, indicating that MOCVD overgrowth at 630°C is insufficient to anneal the known implant damage. Accordingly, different MOCVD anneal conditions were investigated by N. Puetz. Two wafers, O646 and 652, were grown similarly to O620 and implanted with a dose of $4 \times 10^{12} \text{ cm}^{-3}$, a greater dose than would be used in an actual APD, as a "worst case" benchmark. These wafer were then overgrown at 680°C for 1 hr (O659) and 2 hrs (O664) in such a way as to keep the InP thickness constant. The I-V characteristics of these wafers is shown in Fig. 6.1.3 (b) and (c). Note the improvement in leakage as the amount of annealing is increased. While not as good as O619 they are an encouraging intermediate result. Annealing conditions as high as 700°C could be investigated to attempt further improvement, and also the dose used in an actual APD would be less. Moreover, this was a two-growth structure (so the ion implantation could be carried out in between the growths), which may not be a fair comparison to the single growth MOCVD charge sheet structure. However, it is important to realize that if the leakage current for these devices is dominated by generation-recombination associated with defects from the ion implantation, there is every reason to expect that the reliability of such devices might be worse than for those of the MOCVD charge sheet.

6.1.3 Choice for further work

Pursuing the SAGM structure, although most common in the literature, would mean developing a guard ring process (risky) and coping with the extremely stringent process controls required as discussed

6. Initial experiments

previously. This leaves the SAGCM structure, with a choice to be made between ion implantation and MOCVD grown charge sheet. As the leakage current for O619B is the lowest, the choice was made to pursue the mesa SAGCM APD structure with the MOCVD grown charge sheet.

Figure 6.1.3 I-V curves for planar blanket ion implantation structure.

- (a) InP overgrowth at 630°C. It is evident that the leakage current is high, especially at low voltages.
- (b) InP overgrowth for one hour at 680°C. Avalanche breakdown is not evident.
- (c) InP overgrowth for two hours at 680°C. The leakage current is the lowest of the group, especially at low voltages, and avalanche breakdown is evident.

6. Initial experiments

6.2 SAGCM initial results and discussion

In this section the initial performance characterization of full SAGCM structures will be described.

6.2.1 Comparison of RIE and WCE with no etch stop, SL and no SL for mesa formation

Having decided on the mesa SAGCM APD structure, certain questions remain concerning the fabrication details. In this section, an experiment to distinguish among RIE and WCE for mesa formation, as well as the necessity for the inclusion of a grading (SL) layer, is described.

6.2.2 RIE vs. WCE with no etch stop: process control

First growths of two wafers were performed: O768 (no SL grading layer) and O770 (containing a SL grading layer). In these wafers the charge sheet was divided 80-20 % between the InP and InGaAs respectively. The wafers were quartered. An initial assessment of which etching technique to use was done by etching mesas on quarter-wafers (O768 and O770) to depths of nominally 500, 700 and 900 Å by WCE and 700 Å by RIE. The processing results are presented in Fig. 6.2.1. It is evident that the spread of values obtained was far greater for all the WCE cases than for the RIE case, indicating the superior RIE uniformity. Furthermore it is not evident from the data that there is a clear difference in the mean etched depth for the three different WCE cases, indicating the poor depth control using WCE. For the RIE case the mesas were consistently somewhat (10%) thinner than targeted. This offset could be subsequently adjusted in future runs.

Figure 6.2.1 Histograms of measured etch depth on quarter wafers for WCE with no etch stop (a-c) and RIE (d). The target depths for each case were (a) WCE - 500 Å, (b) WCE - 700 Å, (c) WCE - 900 Å and (d) RIE - 700 Å. It is evident that there is little difference in the achieved etch depths for WCE despite the different target depths, and as well the spread of achieved depths is much less for RIE than for WCE with no etch stop.

Following the etching, the RIE quarter and one of the WCE quarters were overgrown and processed into APDs. It was initially anticipated that an additional alignment layer would be necessary, as the MOCVD overgrowth would be expected to planarize the mesa. However, the mesas could still be observed following the overgrowth, so this additional photolithography and etch was not necessary. It was observed that the photolithographic alignment of the RIE wafers was considerably easier than the WCE wafers despite the

6. Initial experiments

somewhat greater etch depth of the WCE wafers. This indicates a greater contrast for the RIE, either due to better definition of the mesa or better visual sensitivity to the angle of the mesa etch for RIE.

6.2.3 DC characteristics: leakage current and breakdown voltage

The I-V characteristics of the four quarter wafers are shown in Fig. 6.2.2. To first order, the magnitude of the leakage current is typically in the 1-10 nA range for all four quarter wafers, indicating that the details of the mesa etching and grading layer does not cause in itself a major impact on the leakage current. It is evident that the breakdown characteristics of the two RIE quarter wafers are more uniform than those of the WCE quarter wafers. As mentioned above, the RIE wafers were etched to a smaller depth than originally targeted with the result that there was more partial charge in the periphery than intended. It is not surprising that the devices broke down prematurely at the edge of the device rather than in the central region. This was verified by translating the fiber across the device and measuring the AC photocurrent signal for 1.3 μm laser emission pulses using an oscilloscope; an increased signal was observed at the edges. The WCE wafers showed non-uniform characteristics, with the result that some devices broke down in the central region (desired). If the RIE quarter wafers were etched deeper, the devices would break down in the central region.

6.2.4 SL vs. no SL: speed

The effect of the SL could only be investigated in center breakdown devices. This limited the investigation to the WCE samples, as the RIE samples broke down prematurely. A pulsed measurement for the WCE samples indicated a fall time of 0.3 ns for the SL wafer and 2 ns for the no SL wafer, circumstantially confirming the role the SL was expected to play in

reducing hole trapping at the upper InGaAs interface over the case where no grading layer was present. As the WCE was so non-uniform, it was difficult to assign a specific E-field, and therefore the two speed results are not necessarily with the same E-field at that interface. Therefore this speed measurement is not absolutely conclusive in itself. However, one would presume a hefty E-field offset would be necessary to compensate for the factor of 7 difference in speed between the two cases, and it is reasonable to assume that the presence of the SL assists the speed.

Figure 6.2.2 I-V characteristics for SAGCM APDs processed as quarter wafers as indicated from the combination of WCE with no etch stop or RIE, SL grading layer or no SL grading layer.

6. Initial experiments

6.2.5 Choices for future processing

The leakage current seems to be unaffected to first order by the choice of RIE vs. WCE, SL vs. no SL. The relatively small leakage currents indicate that it is possible to place some of the charge sheet into small thicknesses of the top of the InGaAs layer without inducing a large tunneling current. It appears that the SL offers more speed than the non-SL case, although the measurement is not absolutely certain. The uniformity of RIE appears to be much better than for WCE, and the mesas were more visible for photolithographic purposes. Therefore the choice of RIE/SL is used in the next section.

6.2.6 Introduction of etch stop

Following the experiments on quarter wafers to determine the better way to etch the samples, whole wafers were processed using RIE. Unfortunately the mesa depths were inconsistent. This situation was found to arise primarily from the loading effect. A factor of four more material is removed when etching a whole wafer relative to a quarter wafer. Normally this might not be a problem, but the mesa mask level for mask set T79 is light field, which means that the majority of material is removed when the mesa is etched. This combination of a factor of four more material on a light field mask proved to be detrimental to the device fabrication: the RIE etch depth varied by 10 % center to edge, with less material etched in the center than at the edge due to the depletion of reactive species at the center relative to the edge. A histogram of RIE etch depths for whole wafers is shown in Fig. 6.2.3. At the same time, the MOCVD thickness varies by 10 % center to edge, with the thicker material at the center. This is the worst possible situation for trying to maintain uniform peripheral charge sheet, as the magnitude of the peripheral charge sheet is determined by the difference between the deposited

6. Initial experiments

charge sheet and the removed charge sheet. The grown material is thicker at the center than at the edge, yet less material is removed. There are two possible solutions to the problem using the given mask set.

Figure 6.2.3 Histogram of measured etch depths on whole wafers using RIE. The results from three wafers are included here.

6. Initial experiments

First, there is no requirement on the uniformity of the doping within the charge sheet in the x direction. A practical idea would be to have a region of high doping near the top of the charge sheet and low doping near the bottom. In this way, if the difference between the grown thickness and etched depth – i.e. mesa bottom – comes to rest anywhere within the low doped region, the peripheral charge sheet is relatively unaffected by the actual location within the low doped region. In the extreme case of a zero doped region, there would be no effect at all. This doping scheme was used in some of the wafers described in the next section.

A second solution would be to incorporate an etch stop layer within the charge sheet. The idea here would be to return to WCE, but with an etch stop to define absolutely the location of the mesa bottom. The differences between WCE with and without an etch stop are shown schematically in Fig. 6.2.4. This is even better for process control than a high-low doping using RIE. A potential drawback would be the etch stop layer itself, which would be a thin InGaAs(P) layer lattice-matched to InP. There would be the possibility of hole trapping within the quantum well arising from the layer, and moreover the ionization within a high-field region of a material other than InP might be unfavorable for device performance. The best choice for the etch stop material for processing would be InGaAs, as the farthest removed from InP. However, the greater the difference from InP, the greater the potential drawback for the reasons just described. As a compromise, a 50Å etch stop layer of InGaAsP with bandgap wavelength 1.2 μm (As mole fraction 0.5) was used in some of the wafers described in the next section. It was experimentally found that the etch depth correlated well with the known MOCVD epilayer variation, indicating that the peripheral charge sheet was

6. Initial experiments

controlled by the MOCVD doping rather than unintentional processing variations associated with defining the mesa.

Figure 6.2.4 Schematic diagram of APD with and without an etch stop layer. Without an etch stop layer the idea is to locate the mesa bottom anywhere within the low doped n-InP.

7. SAGCM APD performance

7.1 I-V, M-V and C-V dependencies

Several full SAGCM wafers were processed using both mask sets T79 and Z01. In this section the I-V, M-V and C-V dependencies are discussed.

A typical set of I-V and M-V characteristics is shown in Fig. 7.1.1. This wafer, R1-406/R1-355T, was processed using mask set T79 and the devices shown are of 46 μm diameter with 7 μm annulus. Strictly speaking, at this point there has been no independent measure of the gain, so the two curves are more properly designated as I_{dark} -V and I_{photo} -V curves than I-V and M-V curves. These particular I_{photo} -V curves do not subtract out the dark current; however, the I_{dark} -V curve is more than two orders of magnitude lower than the I_{photo} -V curve, so there is less than a 1 % error from this omission. For lower primary photocurrents presented later in this section, the dark current is subtracted from the measured current to give the true photocurrent.

Several features are apparent from the I_{dark} -V and I_{photo} -V curves. In order to appreciate the voltage dependence of the curves it is necessary to consider the depletion region, shown schematically in Fig. 7.1.2.

At low bias, no photocurrent is observed initially. This is because the depletion region does not extend into the InGaAs, and the valence band offset between the InGaAs and InP (0.36 eV, which is 14 $k_{\text{B}}T$ at room temperature) prevents any significant photoresponse.

At the "annular punchthrough" voltage V_{ann} (8 V in Fig. 7.1.1), the depletion region reaches the InGaAs in the device periphery and an abrupt increase in photocurrent is observed (Fig. 7.1.2).

Figure 7.1.1 I-V and M-V curves for eight individual devices on R1-406L/R1-355T. The I-V curves are the lower traces and the M-V curves are the upper traces. Two distinct punchthrough voltages, V_{ann} and V_{mesa} , are evident, as well as the breakdown voltage. The portion of the M-V curve which does not change with respect to voltage corresponds to a gain of 1.

7. SAGCM APD performance

Figure 7.1.2 Schematic diagram of the depletion region for SAGCM APDs under the annular and mesa punchthrough conditions.

As the bias further increases, annular depletion continues rapidly owing to the low doping of the InGaAs, and there is a flat region of the $I_{\text{photo}}-V$ curve (8-25 V in Fig. 7.1.1) in which the responsivity is found to be the same as for similarly processed p-i-n photodiodes whose p-n junctions are located in the InGaAs rather than the InP [19, 28-30]. Thus the gain in this bias regime is 1. This occurs because long lived electron-hole pairs are generated in the InGaAs absorption layer under the active mesa region of the device; these carriers diffuse laterally to the depletion region in the annulus [28-30, 129] and are separated out by the E-field present. The holes proceed toward the undoped InP layer above the annulus. Even though the E-field is high enough above the active mesa region for avalanche multiplication to occur,

7. SAGCM APD performance

no carriers are transported through this region, so no increase in the photocurrent is observed as the bias is increased until the "mesa area punchthrough" voltage V_{mesa} (28 V in Fig. 7.1.1) is reached. At this point, the depletion region in the active mesa region reaches the InGaAs and holes which were generated under the active region of the device are transported across the heterojunction into the undoped InP multiplication layer above the active region (Fig. 7.1.2). An abrupt increase in photocurrent is observed because of the high E-field in the multiplication region, and above this bias the gain continues to increase until the avalanche breakdown voltage V_B is reached. The entire process, including the existence of the region of flat photoresponse, allows for self-calibration of the gain of the device, unlike SAGM APDs. At the highest bias, it was verified that the avalanche breakdown was due to the active area, not the edge, for all the samples. This was done by translation of the optical fiber across the device. Also, it was found that all the characteristic voltages were independent of device size in the range of 30-150 μm diameter photosensitive area over several wafers.

It is important to point out the significance of the regime between V_{ann} and V_{mesa} . It was found for devices mounted on ceramic carriers over a range of wafers that the responsivity in the "M=1" regime is the same as that of p-i-n photodetectors to the measurement accuracy of $\pm 2\%$. Further, the photocurrent is constant to within $\leq 2\%$ through the M=1 regime. Also, on the wafer it is apparent from the photocurrent traces that on a log scale there is no observable difference in the responsivity among the various sizes of devices, which might be expected from a diffusion length reported to be 140 μm in high quality InGaAs [129]. Therefore the $I_{\text{photo}}-V$ curve is in fact an M-V curve, and will be so referred to through the remainder of this work.

7. SAGCM APD performance

This situation differs substantially from that of typical SAGM APDs. In these devices, there is no way to independently measure the gain and responsivity of the devices. For example, a device with a quantum efficiency of 50% and a gain of 20 would have the same photocurrent as a device with a quantum efficiency of 100% and a gain of 10. Some authors have tried to use a two wavelength technique corresponding to hole and electron injection into the multiplication region to independently measure the gain. From Chapter 2, the ratio of gains from electron and hole injection is a function of the gain. For example, the Fujitsu technical manual states that at a gain of 2 for hole injection the gain for electron injection should be 1.58. However, this presupposes a certain k , which can only be measured if the gain is known. This is a circular reasoning process with the end result that there is no way to independently measure both the gain and responsivity of the devices. With a SAGCM APD, this problem is alleviated, and for the first time, the gain is independently known [28-30].

The C-V curves for devices of various sizes is shown in Fig. 7.1.3. The annulus on these devices is 7 μm and the mesa diameters are 26, 36, 46, 56, 66, 76 and 96 μm from the smallest to the largest values respectively. It is evident that as the device size increases so does the capacitance. It is also evident that there is a sharp drop in capacitance at V_{ann} and V_{mesa} , independent of the device size, confirming the observations made above with respect to the device physics of the unity gain regime.

M-V curves over four decades of input photocurrent are shown in Fig. 7.1.4 for wafer O1028/R1-147L. Attenuator readings were used to verify that the responsivity was constant over the input photocurrent range used here. The $M=1$ regime is seen to be independent of input photocurrent, which is

7. SAGCM APD performance

Figure 7.1.3 C-V curves for R1-406L/R1-355T for devices of various sizes. Note the sharp drop in capacitance at V_{ann} and V_{mesa} .

consistent with the device physics outlined. The gain, however, is a strong function of input photocurrent. Previous authors have pointed out empirically and semi-theoretically that the maximum gain saturates in proportion to the square root of the input photocurrent [11, 22]. The measurements here are qualitatively consistent with this observation.

Figs. 7.1.5 and 7.1.6 show the I-V and M-V characteristics for a particular device from wafer P555 labeled 3d2. In later sections within this chapter, further measurements from this particular device will be shown so as to have a complete set of measurements on one particular device. It must be emphasized that all the measurements shown in this chapter are typical results, seen on many wafers. The particular measurements for inclusion within this thesis have usually been chosen because they are particularly detailed for that particular characteristic. For example, the amount of data reported in Figs. 7.1.3 and 7.1.4 are not acquired in this much detail on every wafer, but these characteristics are always observed no matter from which wafer the measurements are taken. The more recently processed wafers from mask set Z01 do not include multi-sized devices, so the type of measurement in Fig. 7.1.3 (capacitance of many sizes of device) can only be done for one size of device.

7. SAGCM APD performance

Figure 7.1.4 M-V measurements over four decades of input photocurrent for O1029/R1-147L. The leakage current is shown as well.

Figure 7.1.5 I-V measurement for P555-3d2.

Figure 7.1.6 M-V measurement for P555-3d2.

7. SAGCM APD performance

7.2 Frequency response

The electrical response vs. frequency for R1-406L/355T is shown in Fig. 7.2.1 for various gains. As expected from the discussion in Section 2.5 and from Fig. 2.5.1, the electrical response is expected to depend on the gain. A marker shows the -3 dB bandwidth, defined as the frequency for which the electrical response drops to 3 dB lower than the near-DC response. It is seen that for this particular device, the maximum bandwidth of 4.3 GHz occurs near $M=7$, declining as the gain varies in either direction from this value. Also, an inductive peak is evident which tends to increase the -3 dB bandwidth for devices whose "unaided" bandwidth is less than the inductive peak frequency of $\frac{1}{2\pi} \sqrt{\frac{1}{LC}}$. For this device, the peak is located near a frequency of 3.4 GHz, independent of the device gain. It is also evident that the device response is decreasing near 2.4 GHz before the inductive peak boosts the electrical response. This particular device has a 60 μm diameter diffused area. Other devices were fabricated with a 40 μm diameter diffused area. For this case an electrical bandwidth as high as 5.5 GHz has been observed.

The bandwidth vs. gain is shown in Fig. 7.2.2 for two devices from a wafer with a 1Q grading layer fabricated in the P reactor (P405/396) and a device with a 9Q grading layer fabricated in the R1 reactor (P242/R1-366T). It is evident that the R1 device has a higher maximum bandwidth than either of the two devices with the 1Q grading layer fabricated in the P reactor. As well, there are significant differences in the maximum bandwidth obtained in the P reactor for a 1Q grading layer. As will be shown later, between the two 1Q grading layer devices, the device with the higher breakdown voltage also

has a higher E-field at the heterointerface. The difference in maximum bandwidth between these two devices is related to hole trapping, with the higher heterointerface E-field being necessary to avoid hole trapping. There is also a difference, albeit less marked, between the R1 device and the better P device, where the R1 bandwidth increases more steeply at the low gain end of the curve. Again, this is probably attributable to some hole trapping in the P device. This suggests that a 1 Ω grading layer with a stagnation time in between the growth of the InGaAsP and InP in the P reactor leads to inconsistent and poor frequency response.

Figure 7.2.1 High frequency response: response vs. frequency for R1-406L/R1-355T, various gains. An inductive peak near 4.5 GHz is apparent.

Figure 7.2.2 Typical f-M curves for P405/396 (1Q grading layer with n InGaAs) and P555 (9Q grading layer with n InGaAs).

Various experiments were tried to correct the situation. In the end, a nine quaternary layer formulation graded from InGaAs to InP (9Q) in the P reactor was used for optimum performance. The idea is that although rapid switching between arbitrary compositions is difficult with this reactor due to the damping time associated with "large signal" switching of the mass flow controllers, incremental switching among closely spaced quaternary compositions is analogous to "small signal" switching of the mass flow controllers, where the gas flow oscillations subside more quickly.

7. SAGCM APD performance

In addition, it is beneficial to include some doping in the upper portion of the InGaAs. While tunneling is a possibility, if the thickness is restricted to 200 Å, there is relatively small volume for tunneling current contribution. The benefit comes from increasing the E-field at the heterointerface to prevent hole trapping.

The combination of the 9Q grading layer with a very thin doped portion at the top of the InGaAs layer is used for the optimized APD for which frequency response results over whole wafers are reported on in the next chapter. A schematic diagram is presented in Fig. 7.2.3.

Figure 7.2.3 Schematic diagram of an optimized 9Q APD.

7. SAGCM APD performance

Figure 7.2.4 GBW measurement for O1029/R1-147L. The GBW is 107 GHz.

It is possible to obtain a very high GBW using a SAGCM APD. Fig. 7.2.4 shows the bandwidth vs. gain dependency for such an APD. In order to ensure operation in the GBW regime and to eliminate the optimistic effect of the inductive peaking, the bandwidth was restricted to values less than 1.5 GHz, which required gains above 65. The input power for the data in Fig. 7.2.3 was varied from 10-530 nW, as the highest gains are achievable only with very low input powers (Fig. 7.1.4). Eqn. (2.5.14) was used to extract the GBW from the measured and intrinsic bandwidths. The avalanche bandwidths (and hence the GBWs) are conservative estimates since $f_{\text{intrinsic}}$

was actually less than the 3.6 GHz used in Eqn. (2.5.14). The equation of best fit to the data in Fig. 7.2.3 was

$$\text{Gain (dB)} = 20.3 - 10.03 \log (f_{\text{avalanche}} \text{ (GHz)}) \quad (7.2.1).$$

The correlation coefficient was 0.986. There is only a 0.3% deviation from the ideal slope of 10, which indicates both the presence of the GBW regime and the high quality of the data. The GBW may be estimated from Eqn. (7.2.1) at a frequency of 1 GHz (which falls within the measurement range) as 107 GHz. Also the individual GBW for each data point in Fig. 2 was calculated from Eqn. (2.5.14). The average ($\pm$ standard deviation) for all the individual points was 107 (± 10) GHz, in excellent agreement with the value obtained from Eqn. (7.2.1). Thus a GBW exceeding 100 GHz has been achieved for the first time to the author's knowledge [29, 30, 162].

A similar and more recent result is shown in Fig. 7.2.5. In this case the GBW was 122 GHz, presently the highest published GBW result the author is aware of for any InP-based APD. The device parameters for these two APDs is given in Table 7.2.1. Note that a GBW in excess of 100 GHz can be achieved with both a 3Q and 9Q grading layer. This is because the GBW should be independent of the grading layer.

Figure 7.2.5 GBW measurement for P623. The GBW is 122 GHz.

Table 7.2.1 Device parameters for high GBW devices.

Property	O1029/R1-147L	P623/614
x_j (μm)	2.2	2.2
x_d (μm)	0.1	0.2
σ ($\times 10^{12} \text{ cm}^{-2}$)	2.6*	3.3
t_{mesa} (μm)	0.1	0.2
grading layer type	1Q	9Q
t_{InGaAs} (μm)	2.8	3.3
$t_{\text{Buffer InP, undoped}}$ (μm)	1.0	0.5

* Best estimate at time of publication, and certainly the intended values. However, more recent evidence suggests that σ was much higher than this, but unquantifiable.

The bandwidth vs. M characteristics for P555 3d2 are shown in Fig. 7.2.6.

Figure 7.2.6 f - M curve for P555-3d2.

7.3 Noise

Noise measurements have been done on many wafers. In this section some typical data are presented.

Fig. 7.3.1 shows excess noise factor (F) measurements for five devices spanning a wide range of x_d and σ . The best estimates for the x_d and σ values of the individual devices, based on SIMS, infrared fringe maps and doping calibration maps as discussed in Chapter 5, are given in Table 7.3.1. The calibration was performed by reference to the shot noise of a p-i-n photodiode, and an LED was used as a light source [74]. In the same figure are plotted the McIntyre expressions [67] for $F(M)$ for the k values indicated. It is evident that the excess noise is nearly independent of the device parameters, and is well described by $k=0.48$. In these measurements, the spontaneous-spontaneous beat noise of the LED was corrected for [171, 172]. When this is not done, the excess noise of the p-i-n photodiode alone is overestimated relative to that of the APDs, and therefore the resultant excess noise factor of the APD underestimated. Note that there is no mention in Ref. [74] that this correction had been done. In addition, comparisons to commercial SAGM APDs have been made which show no significant difference in the noise between SAGCM APDs and commercial SAGM APDs, assuming the same responsivity for the commercial SAGM device as for the calibrated SAGCM device.

Table 7.3.1 Device parameters for the noise measurements.

Wafer, device #	x_d (μm)	σ ($\times 10^{12} \text{ cm}^{-3}$)	BV (V)
P757 #5	0.09	2.9	63.8
P757 #36	0.09	3.3	48.8
P555 #3	0.34	3.1	50.65
P555 #24	0.38	2.8	64.7
P526 #1	0.62	2.7	54.7

Figure 7.3.1 F vs. M curve for various SAGCM APDs. The solid lines are McIntyre F vs. M curves for the k values indicated.

Figure 7.3.2 Shot noise power vs. I_d curves for a typical SAGCM APD. The solid lines indicate power proportional to M (lower slope) and M^{2+x} (higher slope).

7. SAGCM APD performance

The conclusion based on these measurements is that the noise does not vary very much from device to device, wafer to wafer or even structure to structure, which makes it an "uninteresting" subject for further experimental study.

A dark current noise measurement is shown in Fig. 7.3.2 for a typical SAGCM APD [173]. For this case, the best x_d and σ estimates are $0.3 \mu\text{m}$ and $3.0 \times 10^{12} \text{ cm}^{-3}$, centered in the experimental range of these parameters. The primary dark current is estimated from the intersection of the two lines indicating power proportional to M (lower slope) and M^{2+x} (higher slope) as 1 nA . Using the $F=M^x$ approximation, which is often done with APDs [1], x is found to be 0.7 for this data. This is consistent with values for commercial APDs. The equation of best fit and correlation coefficient of 0.994 for $F=M^x$ is shown in the figure. The two data points to the right which indicate lower noise than $x=0.7$ may be due to saturation of the measurement apparatus.

This noise performance is in contrast with the predictions of Section 4.4 and Fig. 4.4.3 where the noise was expected to be a strong function of x_d . Furthermore, for hundreds of APDs for which $x_d \leq 0.5 \mu\text{m}$, a k value of ≤ 0.5 is obtained, in contrast to the expected value of k of ≥ 0.6 . This phenomenon is not restricted to SAGCM APDs. Most commercial APD manufacturers design their SAGM APDs in such a way as to have an InP multiplication region of $\leq 0.5 \mu\text{m}$. Kuchibhotla and Campbell calculate that for this case, $k \geq 0.6$ [31]. However, typical InP/InGaAs APD data sheets show $k < 0.6$, usually by a substantial margin. To reconcile these seeming contradictions, one possibility is to suppose that the assumptions behind these calculations are incorrect. The most likely prospect which presents itself is the InP-only ionization assumption, which will be discussed later.

7. SAGCM APD performance

8. SAGCM APD measurements vs. fabrication parameters

A significant portion of this work has been devoted to discussing the increased fabrication tolerances expected from SAGCM APDs. Accordingly, it would be worthwhile to undertake an experimental investigation of the practical fabrication tolerances associated with these devices. While it is difficult to control every variable but one due to inadvertent fluctuations in the process, some experiments can be usefully interpreted. Experimental results for SAGCM APD fabrication parameters are presented here for the first time.

8.1 Experiment to vary x_d , measure V_{ann} , V_{mesa} and BV

A wafer suitable for processing into SAGCM APDs was fabricated using two metalorganic chemical vapor deposition growth steps, as previously described. The diffused junction depth x_j and multiplication region length x_d assuming a one-sided abrupt p-n junction are inversely related; their sum is the second growth InP layer thickness t_{InP2} . As already discussed, the complete charge density (σ_{mesa}) is incorporated in the active mesa region below the p-diffusion and a partial charge density (σ_{ann}) is retained after the mesa etching in the annular region of the APD where the p-diffusion extends outside the mesa. The nominal device parameters are given in Table 8.1.1. For reference, a schematic diagram showing the relevant quantities for this section is given in Fig. 8.1.1.

The fabrication tolerance of the device was investigated by generating samples of various x_j from small pieces of a single wafer. This was accomplished by performing separate Zn diffusions for different times for the

Table 8.1.1 Parameters used in wafer for experiment to vary x_d .

Property	Value
x_d (μm)	0-0.7
σ_{mesa} ($\times 10^{12} \text{ cm}^{-2}$)	2.6
σ_{ann} ($\times 10^{12} \text{ cm}^{-2}$)	0.6
t_{mesa} (μm)	0.27
n^- upper InP thickness (μm)	2.1
t_{InGaAs} (μm)	2.8
$t_{\text{Buffer InP, undoped}}$ (μm)	1.0

Figure 8.1.1 Schematic diagram of SAGCM APD showing relevant quantities for x_d experiment.

8. SAGCM performance vs. fabrication parameters

Figure 8.1.2a SIMS profiles of Zn for Zn diffusions of various times into an APD structure. Unless otherwise indicated, the temperature for the diffusion was 450 °C.

Figure 8.1.2b Junction depth vs. time for the SIMS profiles of Fig. 8.1.2a.

8. SAGCM performance vs. fabrication parameters

various samples. By using the same initial wafer for the samples the main variable was x_j itself, rather than other wafer-to-wafer variations, which might tend to obscure the results. However, the small sample sizes generated precluded processing beyond this point. Therefore the samples were probed under reverse bias by direct contact with the n- and p- surfaces rather than by contact with the metallization. The uniformity of t_{InP2} as measured by reflectance spectroscopy was better than $0.2 \mu\text{m}$ for the samples used; this represents an upper limit on the experimental error in x_j or x_d . SIMS profiles of the active area of the devices are shown in Fig. 8.1.2a. The expected linear x_j^2 vs. time dependence for the diffusion was observed, as plotted in Fig. 8.1.2b.

As in the previous section, similar to Figs. 7.1.1 and 7.1.4, the characteristic voltages V_{ann} , V_{mesa} and BV were apparent from the M-V characteristics of these samples. In Fig. 8.1.3 the characteristic voltages are plotted vs. x_d . It is readily apparent that V_{ann} and V_{mesa} show a linear dependence with respect to x_d [162]. The breakdown voltage, however, is nearly independent of x_d . This is consistent with the following observations.

It is difficult to actually measure σ_{active} for an SAGM APD. Rather, this is usually inferred from calibration data of the volume doping concentration and the p-n junction depth. However, for an SAGCM APD, the E-fields in the high field regions correspond from Gauss' Law to the charge sheets σ_{active} and σ_{ann} at V_{mesa} and V_{ann} respectively. This may be understood from Fig. 4.2.2, which shows E-field profiles for a constant σ and varying x_d . The area under these profiles is the applied voltage. The punchthrough voltages depend on a certain heterointerface field E_{het} being present (zero for an ideal grading layer) at the upper edge of the InGaAs. It is evident that the slopes of

these functions with respect to x_d are E-fields, from which the σ values can be determined from Gauss' law (Eqn. (3.2.3)); these may differ slightly from the actual σ values present in the device since the E-field at the heterointerface at the punchthrough voltages may not be exactly zero, which may to a small extent influence the depletion depth. In Fig. 8.2.3, the dependencies of the characteristic voltages with respect to x_d are presented and the best straight line fits are shown. The correlation coefficients were >0.99 and 0.97 for V_{mesa} and V_{ann} respectively. The values obtained in this way are $\sigma_{\text{mesa}} = 2.65 \times 10^{12} \text{ cm}^{-2}$ and $\sigma_{\text{ann}} = 0.68 \times 10^{12} \text{ cm}^{-2}$, in close agreement with the nominal values of $2.6 \times 10^{12} \text{ cm}^{-2}$ and $0.6 \times 10^{12} \text{ cm}^{-2}$ respectively. Thus a novel experimental technique to measure the charge on actual APDs has been developed. The breakdown voltage does not seem to have any particular dependence with respect to x_d . This is in contrast to the breakdown voltage dependence to the simulations presented in Section 4.4, in which a strong dependence of breakdown voltage with respect to x_d is predicted. The conclusion is that this model does not adequately explain the breakdown characteristics of the device.

Additional characteristics are shown in Fig. 8.1.4. It is apparent that substantial gain is apparent for all samples of positive x_d . The maximum gain for samples of negative x_d was far less, but still greater than 1. This is due to ionization in the InGaAs. The edge breakdown voltage is greater than the central breakdown voltage for all the samples of positive x_d , indicating the peripheral charge sheet is fulfilling its function. Note that there was a 100 V measurement limit on the apparatus, and the edge breakdown in some cases may be greater than this value. The tunneling threshold voltage (determined from the edge breakdown samples) is greater than the breakdown voltage for all the samples of positive x_d except the very thinnest.

8. SAGCM performance vs. fabrication parameters

Representative E-field profiles for the portion of the periphery where the p-diffusion front is curved (line C in Fig. 4.3.1) are illustrated in Fig. 4.3.5 for the same reverse bias and other assumptions as in Fig. 4.3.4 [28]. When the complete charge density is present (curve C1), the maximum E-field in the InP is greater than the breakdown field, as expected from above. When no partial charge sheet density is present ($\sigma_{ann}=0$, curve C3) the E-field in the InP is at its lowest, but as in the slab region, the E-field in the InGaAs is increased. By using a partial charge sheet density ($\sigma_{ann}= 0.4 \sigma_{active}$, curve C2) the E-field in the InP can be kept below the breakdown E-field for InP while the E-field throughout the InGaAs is reduced. The BV along line D should not dominate the device characteristics (i.e. no premature breakdown along the side) for sufficiently low doped background InP. This may be more precisely quantified by comparing the voltage in the slab region to the voltage in Eqn. (2.4.14) for the same E-field. Provided the voltage in the cylindrical region is higher, there will be no premature breakdown from this effect.

It is evident that a maximum cylindrical E-field for the planar portion of the device in both the InP and InGaAs may be defined above which the edge region will dominate the device characteristics for each simulation. That is, if the maximum cylindrical E-field in the InP is greater than that in the central region, there will be premature breakdown at the edge; if the maximum cylindrical E-field in the InGaAs is beyond the tunneling threshold, then the InGaAs tunneling leakage current will be prohibitively large. As well, the voltage is uniquely determined for each simulation.

The dependence of the cylindrical E-field in the planar portion of the device is shown in Figs. 4.3.6 - 4.3.8. In each figure E-field is plotted vs. f , the fractional charge sheet. In each figure two of (x_d , x_j and t_{InGaAs}) are kept at constant values of (0.4 μm , 2.0 μm and 2.75 μm) and the third varied.

4. Separate absorption, grading , charge and multiplication APDs

3 of/de 3

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

Figure 8.1.3 Characteristic voltages vs. x_d for samples from a single wafer. V_{ann} and V_{mesa} correlate linearly with x_d . The slope of these two dependencies are E-fields, which are related to the peripheral and central charge sheets from Gauss' Law. The breakdown voltage does not obey the same dependency, indicating that the breakdown E-field is not constant with respect to x_d .

Note that the DC characteristics of SAGCM APDs are demonstrated over a range of $0.7 \mu\text{m}$ from this experiment alone and for the first time [162]. This is more than an order of magnitude greater tolerance than would be expected from SAGM APDs.

Figure 8.1.4 Tunneling threshold, breakdown and edge breakdown voltages and gain vs. x_d for samples from a single wafer. The breakdown voltage does not obey the same dependency, indicating that the breakdown E-field is not constant with respect to x_d .

8.2 Whole wafer measurements

8.2.1 Characteristic voltages vs. wafer position

Fig. 8.2.1 (a) shows I-V and M-V plots for four areas of wafer P389/382, arranged linearly over the direction of MOCVD gas flow. It is evident that the breakdown voltage decreases from the leading to the trailing edge of the wafer. Fig. 8.2.1 (b) shows the I-V and M-V plots for edge breakdown devices, which are devices without the central mesa region. These devices would be expected to break down at the edge, and ideally be of higher breakdown voltage than the comparable devices in Fig. 8.2.1 (a). Indeed, the edge breakdown voltage is higher in all cases, and does not change much with respect to position on the wafer.

The trend of characteristic voltages over a whole wafer is shown in Fig. 8.2.2 for wafer P242/R1-366T and Fig. 8.2.3 for wafers P555/543 and P623/614, in which the annular punchthrough, mesa punchthrough and breakdown voltages are determined as a function of wafer position for first growths in the R1 and P reactor respectively. It is known from the fringe maps of top InP thickness and excellent uniformity of the Zn diffusion that x_d is highest near the center of the wafer, with a decrease towards either edge. This is the same pattern observed for V_{ann} and V_{mesa} , consistent with the results of Section 8.1, thus verifying this trend for whole wafer analysis. The breakdown voltage is seen to be nearly monotonic, decreasing from the leading to the trailing edge. The monotonic trend in breakdown voltage from leading to trailing edge is not consistent with the radial pattern of x_d and not consistent with the pattern observed for V_{ann} and V_{mesa} .

Figure 8.2.1 I-V and M-V curves for P389/382 for (a) center breakdown and (b) edge breakdown devices. The breakdown voltage varies monotonically from the leading edge (tile 3i, high BV) to the trailing edge (tile 3b, low BV). The edge breakdown voltage is much less sensitive to location on the wafer. In all cases the edge breakdown voltage was greater than the center breakdown voltage.

8. SAGCM performance vs. fabrication parameters

Figure 8.2.2 Characteristic voltages vs. position for P-242/R1-366T.

Figure 8.2.3 Characteristic voltages vs. position for P555/543 and P623/614.

8. SAGCM performance vs. fabrication parameters

This confirms over a whole wafer the discrepancy noted in Section 8.1 concerning the variation of V_{ann} , V_{mesa} and BV , where V_{ann} and V_{mesa} varied linearly with x_d and the BV was nearly independent of x_d . At the trailing edge there is a sharp increase in the BV . This may be related to the sharp decrease in thickness at the edge evident from the InP fringe maps (similar to that of Fig. 5.2.1), for which case the Zn diffusion may enter the charge sheet itself, reducing the charge. This would be further confirmation of the dependence of the breakdown voltage on the charge.

There is a marked similarity in the patterns for BV and the Hall doping profiles from Figs. 5.2.2 and 5.2.3. It appears that the breakdown voltage depends mainly on the charge and not as much on x_d . This is inconsistent with the assumptions and results in Section 4.4 whereby the breakdown voltage is expected to be a strong function of x_d and independent of the charge, since for that model ionization in the InGaAs is assumed to be negligible.

A further confirmation of the dependence of BV on charge is the intentional differences between P555/543 and P623/614, as summarized in Table 8.2.1. The main intentional difference was a 5% increase in the charge from the mass flow controller settings in the MOCVD reactor. It is apparent that the BV vs. wafer position dependency is of the same shape for both wafers, but translated down by approximately 5 V as the charge sheet is increased by 5 %. The effects of the change on the other two characteristic voltages are not a significant, but it is apparent that on average, V_{mesa} is increased by 1-2 V from P555 to P623 for the same position on the wafer. This is the expected trend from Gauss' Law were x_d to remain constant and the charge increased, consistent with the results presented in Section 8.1.

8. SAGCM performance vs. fabrication parameters

Table 8.2.1 Parameters used in three whole wafers from the P reactor.

Property	P555/543	P623/614	P624/616
x_j (μm)	2.2	2.2	2.2
x_d (μm)	0.3 ± 0.1	0.3 ± 0.1	0.2 ± 0.1
σ ($\times 10^{12} \text{ cm}^{-2}$)	2.4-3.0	2.7-3.3	2.7-3.3
t_{mesa} (μm)	0.2	0.2	0.1
grading layer type	9Q	9Q	9Q
t_{InGaAs} (μm)	2.8	3.3	3.3
$t_{\text{Puffer InP, undoped}}$ (μm)	1.0	0.5	0.5

8.2.2 f-M vs. wafer position

The GBW vs. wafer position for P623/614 are shown in Fig. 8.2.4. In this case the GBW is defined as the product of the gain and bandwidth at a gain of 100, and for the highest GBW in this plot, probably includes some beneficial inductive peaking which enhances the apparent GBW. As with the breakdown voltage, the GBW correlates more strongly with charge than with x_d . Again this is in contradiction with the simulations on the bulk-InP ionization δ -doped APD. From Fig. 4.4.4 it is expected that GBW should increase with decreasing x_d . Certainly this is the case from the trailing edge to the center of the wafer, but the opposite trend is observed from the center to the leading edge of the wafer. Note again the similarity in the patterns for the GBW and the Hall doping profiles from Figs. 5.2.2 and 5.2.3. Moreover, the effect of the 5 % increase in doping from P555 to P623 was to translate the GBW profile on the wafer to the location where the doping is the same. This is strong evidence that the GBW, like the BV depends mainly on the charge and not as much on x_d .

8. SAGCM performance vs. fabrication parameters

Figure 8.2.4 GBW vs. position for P623/614.

Figure 8.2.5 f-M curves for many devices from P555/543. The tile numbers are indicated on the legend, which correspond to Fig. 5.1.2. Note that as the minimum gain for full bandwidth increases, so too does GBW.

8. SAGCM performance vs. fabrication parameters

Full f-M dependencies as a function of wafer position are shown for P555/543 in Fig. 8.2.5. There is a trend where the lower the minimum gain for which full bandwidth occurs, the lower the GBW, and as the minimum gain for full bandwidth increases, so does the GBW.

The GBW is correlated with BV as shown in Fig. 8.2.6. These are experimental results from three different wafers, whose parameters are summarized in Table 8.2.1. It is evident from the figure that as the breakdown voltage decreases, the GBW increases. Note that from the above discussion, BV is in turn is correlated primarily with charge and not x_d .

The minimum gain for 2 GHz bandwidth is plotted vs. BV in Fig. 8.2.7 for results over the same three wafers. Again a strong correlation is observed. As the BV increases, the minimum gain decreases.

The correlation between the GBW and $M_{\min}$ is presented in Fig. 8.2.8, again for the composite results over the same three wafers. It is evident that there is a strong correlation between increasing $M_{\min}$ and increasing GBW, with a correlation coefficient for a quantitative logarithmic dependence of 0.90, neglecting the two points in the upper left which do not belong to the distribution. This correlation is also qualitatively observable from the f-M plots of P555/543 (Fig. 8.2.6). There is a finite opening of high bandwidth in the f-M curve, which moves to the right (increasing M) as the breakdown voltage decreases and to the left as the breakdown voltage increases. Or put another way, the useful gain window moves to the right as the charge increases and to the left as the charge decreases.

Figure 8.2.6 GBW vs. BV: composite results from P555/543, P623/614 and P624/616.

Figure 8.2.7 Mmin for 2 GHz bandwidth vs. BV: composite results from P555/543, P623/614 and P624/616.

8.3 Inadequacy of InP-only multiplication model

It is clear from the discussion above that bulk-InP multiplication does not satisfactorily explain the breakdown voltage and f-M characteristics of SAGCM APDs. The trend of decreasing GBW with decreasing charge gives a clue to the explanation. The main effect of a change in charge for a given E-field in the InP is to increase the E-field in the InGaAs, also increasing the ionization in the InGaAs. Thus the assumption for bulk-InP multiplication δ -doped APDs that ionization in the InGaAs can be neglected is increasingly less justified as the charge sheet magnitude decreases. As ionization in the InGaAs becomes increasingly more important, the effective multiplication region of the device increases, and it would be expected that GBW would decrease for this situation. Hollenhorst has previously noted a correlation between a high heterointerface E-field and a low GBW for SAGM APDs [81], consistent with the possibility of ionization in the InGaAs. Kuwatsuka et al. have recently mentioned that the suppression of impact ionization in the InGaAs is necessary for a high GBW [111]. Osaka and Mikawa have previously observed that the ionization in the InGaAs may contribute significantly to the excess noise.

In the next chapter, an analytical model for ionization in the InGaAs to explain many of the experimental results of SAGCM APDs is presented.

9. Ionization outside the "multiplication" region

In Sections 4.4 theoretical predictions were made concerning the APD where ionization is restricted to the InP ("InP-only ionization") which were tested in Section 8. Among the difficulties encountered were the breakdown voltage variation with x_d and σ and the GBW variation with σ , which cannot be explained in terms of InP-only ionization. In this section multiplication in the lower field regions outside the InP is considered in order to attempt to explain some of these results.

9.1 Multiplication

The simplest case of an APD with multiplication in more than one region is an ideal δ -doped APD without a grading layer, comprised of a higher field InP region above the charge sheet and a lower field InGaAs region below the charge sheet. The assumptions used are similar to those of Section 4.4. As before, all the doping within the depletion region is confined to an infinitely thin charge sheet. The depletion region is bounded on one end (at $x=0$) by a p^+-n^- junction for which the E-field on the p-side of the junction falls to zero over an infinitely small distance and at the other end (at $x=w$) by an abrupt n^+ contact which causes the E-field to drop from whatever its value was at w^- to zero over an infinitely small distance. The InP and InGaAs layers within the depletion region are of zero doping, so the E-field is constant throughout each layer. The E-field in the InP is higher than that in the InGaAs, with the difference proportional to the charge sheet from Gauss' Law.

Let x_1 denote the InP multiplication region width and w the total APD thickness. In this section the coordinate system as described in Section 2 will be used (Fig. 2.1.1), where x_1 lies between 0 and w . Assuming both the multiplication and absorption regions to be undoped, the multiplication $M(x)$

may be written using the assumption of constant ionization coefficients from Eqn. (2.3.7) as

$$M(x) = \frac{\exp \left[- \int_0^x (\alpha - \beta) dx' \right]}{1 - \frac{\alpha_1}{\alpha_1 - \beta_1} (1 - \exp [-(\alpha_1 - \beta_1)x_1]) - \frac{\alpha_2}{\alpha_2 - \beta_2} \exp [-(\alpha_2 - \beta_2)x_1] (1 - \exp [-(\alpha_2 - \beta_2)(w - x_1)])}$$

(9.1.1)

where the subscripts 1 and 2 denote InP and InGaAs respectively in this case.

The condition for breakdown is that the denominator of Eqn. (9.1.1) go to zero; i.e.

$$1 - \frac{\alpha_1}{\alpha_1 - \beta_1} (1 - \exp [-(\alpha_1 - \beta_1)x_1]) - \frac{\alpha_2}{\alpha_2 - \beta_2} \exp [-(\alpha_2 - \beta_2)x_1] (1 - \exp [-(\alpha_2 - \beta_2)(w - x_1)]) = 0$$

(9.1.2)

In fact for any piecewise continuous E-field APD with a number of regions n_{regions} denoted by subscripts i ending at coordinates x_i and with constant ionization coefficients α_i and β_i the multiplication at position x is given by

$$M(x) = \frac{\exp \left[- \int_0^x (\alpha - \beta) dx' \right]}{1 - \sum_i^{n_{\text{regions}}} \left\{ \frac{\alpha_i}{\alpha_i - \beta_i} \exp [-(\alpha_i - \beta_i)x_{i-1}] (1 - \exp [-(\alpha_i - \beta_i)(x_i - x_{i-1})]) \right\}} \quad (9.1.3)$$

where x_1 is 0 and $x_{n_{\text{regions}}}$ is w . This is an analytical expression which contains an integral only in the numerator, and even that integral may be expressed as a sum of exponentials for any particular x . This analytical expression makes for comparatively easy physical understanding of a particular structure, as compared with the numerical methods traditionally used.

The condition for breakdown is that the denominator of Eqn. (9.1.3) go to zero; i.e.

$$\sum_i^{n_{\text{regions}}} \left\{ \frac{\alpha_i}{\alpha_i - \beta_i} \exp [-(\alpha_i - \beta_i)x_{i-1}] (1 - \exp [-(\alpha_i - \beta_i)(x_i - x_{i-1})]) \right\} = 1 \quad (9.1.4)$$

The multiplication of a multilayer δ -doped APD has not been treated in the literature, but can be explored mathematically with considerable ease using Eqns. (9.1.1)-(9.1.4). For the rest of this section the two layer InP/InGaAs structure is considered ("InP/InGaAs ionization"), and Eqns. (9.1.1) and (9.1.2) used.

9. Ionization outside the "multiplication" region

Figure 9.1.1 Illustration of multiplication vs. position for a multilayer APD.

It is of fundamental importance that in standard SAM APDs the difference in ionization coefficients switches signs at the InP/InGaAs heterointerface. For this situation, $M(x)$ does not vary monotonically from 0 to w , as is the case for the bulk Si or bulk InP only multiplication regions considered previously. Instead the multiplication is a maximum at the InP/InGaAs heterointerface, for the usual InP-based APD structure where the n-contact is adjacent to the InGaAs and the p-contact is adjacent to the InP. This is illustrated schematically in Fig. 9.1.1. Intuitively, this comes about because an electron-hole pair generated at the heterointerface is separated out whereby the holes are attracted through the InP to the p-contact, in which direction $\beta > \alpha$, and the electrons are attracted through the InGaAs to the n-contact, in which direction $\alpha > \beta$. At any point other than the heterointerface, one carrier or the other must travel at least part way in a direction where the opposite ionization coefficient is higher and the other

carrier travels less distance in the direction of favorable ionization coefficient than would be the case from the heterointerface. Interestingly, if the InP and InGaAs positions were reversed with respect to the p- and n-contacts, the opposite would occur, namely a minimum multiplication at the heterointerface from analogous arguments.

In order to illustrate the predictions made by considering ionization in the InGaAs, two sets of simulations have been performed: one taking the doping to be constant while varying x_d , and the other taking x_d to be constant while varying the doping. In all the cases below the InGaAs thickness is assumed constant at 3.2 μm .

In many cases below it is possible to compare the new result to an experimental result. The wafers P555/543, P623/614 and P624/616 are used for a number of the comparisons. The parameters for these wafers are given in Table 8.2.1. The main differences between a real SAGCM APD and an ideal δ -doped APD are as follows. First, the Zn diffusion profile is not abrupt (Fig. 5.3.1) and the charge sheet is of finite rather than zero thickness; this should have the effect of increasing the effective x_d over what was measured. The magnitude of this increase is related to the thickness of the charge sheet; thus P555/543 and P623/614 should have a greater increase in effective x_d than P624/616 from t_{mesa} in Table 8.2.1. Second, the background doping in the InGaAs is not really zero but rather $\sim 10^{15} \text{ cm}^{-3}$, which represents an integrated charge of $0.3 \times 10^{12} \text{ cm}^{-2}$ from Eqn. (2.5.28) and from Gauss' Law an E-field drop of 4.5 V/ μm throughout the InGaAs. This should have the effect of increasing the effective charge over what is inferred in the doping profiles. The effect of the triangular rather than flat E-field profile in the InGaAs is taken into account when calculating the breakdown voltage in these simulations, in order to permit comparisons between these results and those

9. Ionization outside the "multiplication" region

of Section 4.4. Also, the undoped buffer-InP has been neglected, since the undoped InGaAs is much thicker and of higher E-field. One of the effects of this is that the device is slightly thicker than simulated, which could increase the breakdown voltage just from the consideration of a larger area under the E-x curve.

9.2 Modeling results: E-field

The maximum E-field at breakdown and at the operating point of the device vs. x_d are plotted in Fig. 9.2.1 for $\sigma=3.0 \times 10^{12} \text{ cm}^{-2}$. The operating point is taken to be a gain of 10, described more completely below. Both E-fields vary smoothly in the x_d domain of 0.1-0.5 μm . For InP/InGaAs ionization, the breakdown field decreases from 59.75 to 52.56 V/ μm (a drop of 7.19 V/ μm) and the E-field at a gain of 10 decreases from 59.29 to 51.22 V/ μm (a drop of 8.07 V/ μm) as x_d varies. This is in contrast to the InP-only ionization result of the breakdown field decreasing from 78.58 to 52.62 V/ μm (a drop of 25.96 V/ μm) and the operating point field decreasing from 74.67 to 51.23 V/ μm (a drop of 23.44 V/ μm) over the same domain, as illustrated in Fig. 4.4.1. For InP/InGaAs ionization, the difference in the E-fields between breakdown and a gain of 10 in the x_d domain of 0.1-0.5 μm increases from 0.46 to 1.34 V/ μm . This is in contrast to the InP-only ionization result where the difference in the same E-fields decreases from 3.89 to 1.39 V/ μm . The physical explanation for the difference in E-field variation and magnitudes is as follows. The presence of ionization in the InGaAs reduces the E-field needed in the InP to produce any given gain. This reduction in E-field is most apparent where the InP is thinnest, where the highest ratio of multiplication from the InGaAs to multiplication from the InP is present. For this case, a given change in the undoped InP thickness represents a relatively large fraction of the InP thickness but a relatively small fraction of the total InP/InGaAs thickness, even when that thickness is weighted for different ionization coefficients. This point will be considered in more detail in Section 9.4. Therefore a change in InP thickness has less influence on the

Figure 9.2.1 E-field in the InP at breakdown and at operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization.

Figure 9.2.2 E-field in the InP at breakdown and at operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$ for InP/InGaAs ionization.

9. Ionization outside the "multiplication" region

breakdown or operating E-field for InP/InGaAs ionization than for InP-only ionization. Once x_d becomes as large as $0.5\ \mu\text{m}$, any differences between the InP-only and InP/InGaAs ionization are negligible for this particular charge sheet.

The maximum E-field at breakdown and at the operating point of the device vs. σ for $x_d=0.3\ \mu\text{m}$ are plotted in Fig. 9.2.2 for InP/InGaAs ionization. Both E-fields vary smoothly in the σ domain of $2.4\text{--}3.6 \times 10^{12}\ \text{cm}^{-2}$, rising sharply as σ increases from $2.4\text{--}3.0 \times 10^{12}\ \text{cm}^{-2}$ and leveling off as σ increases from $3.3\text{--}3.6 \times 10^{12}\ \text{cm}^{-2}$. This is in contrast to the InP-only ionization result where the E-fields should be independent of σ , with the E-field magnitude only negligibly different from the InP/InGaAs ionization case for $\sigma=3.6 \times 10^{12}\ \text{cm}^{-2}$. Again, the presence of ionization in the InGaAs reduces the E-field needed in the InP to produce any given gain. This reduction in E-field is most apparent where σ is the least, allowing for the least E-field difference between the InP and InGaAs from Gauss' Law. Once σ becomes as large as $3.6 \times 10^{12}\ \text{cm}^{-2}$, the difference between the InP-only and InP/InGaAs ionization is negligible.

Comparisons between the E-fields for InP-only and InP/InGaAs ionization are shown in Figs. 9.2.3 and 9.2.4. In Fig. 9.2.3, the E-field vs. x_d for $\sigma=3.0 \times 10^{12}\ \text{cm}^{-2}$ is shown and in Fig. 9.2.4, the E-field vs. σ for $x_d=0.3\ \mu\text{m}$ is shown. The discrepancies between the two theories are apparent, becoming wider as x_d or σ decreases. Recall the initial rationale for separating out the absorption and multiplication regions in Chapter 3. The main purpose is to provide for multiplication while at the same time keeping the E-field in the InGaAs within a certain range. If the E-field in the InGaAs is too high, then tunneling will occur. If there is insufficient E-field in the InGaAs, then lower

Figure 9.2.3 E-field in the InP at breakdown and at operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$: comparison between InP-only and InP/InGaAs ionization.

Figure 9.2.4 E-field in the InP at breakdown and at operating voltage vs. charge for constant x_d of 0.3 μm : comparison between InP-only and InP/InGaAs ionization.

9. Ionization outside the "multiplication" region

bandwidth, incomplete depletion or no photoresponse from the gain region will occur. If the E-field in the InP varies significantly with x_d or σ , then as x_d or σ varies on a wafer it will be hard to keep the E-field in the InGaAs within an acceptable range.

Comparisons between the E-fields at the InP/InGaAs heterointerface for InP-only and InP/InGaAs ionization are shown in Figs. 9.2.5 and 9.2.6. This corresponds to the E-field in the InGaAs. In Fig. 9.2.5, the E-field vs. x_d for $\sigma=3.0 \times 10^{12} \text{ cm}^{-2}$ is shown and in Fig. 9.2.6, the E-field vs. σ for $x_d=0.3 \text{ } \mu\text{m}$ is shown. Again, the discrepancies between the two models are apparent, becoming wider as x_d or σ decreases. It is evident from the figures that there is much more variation of E_{het} with respect to x_d or σ for the InP-only ionization theory than for the InP/InGaAs model. Thus the InP-only model predicts that near a given x_d and σ where E_{het} is acceptable, only a small variation of x_d or σ could be tolerated before E_{het} was no longer within the acceptable range. Also, for $x_d < 0.3 \text{ } \mu\text{m}$, substantial tunneling leakage current would be expected. In contrast, the InP/InGaAs model predicts that near a given x_d and σ where E_{het} is acceptable, a relatively large variation of x_d or σ could be tolerated while maintaining E_{het} within the acceptable range. Furthermore, throughout the entire range of x_d or σ values investigated, $E_{\text{het}} < 16 \text{ V}/\mu\text{m}$, implying that tunneling leakage current is far less likely to be a problem. The InP/InGaAs model is consistent with the experimental results presented in Chapter 8 wherein the device operation is relatively unaffected by substantial changes in the fabrication parameters.

Figure 9.2.5 E-field at the InP/InGaAs heterointerface at breakdown and at operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$: comparison between InP-only and InP/InGaAs ionization.

Figure 9.2.6 E-field at the InP/InGaAs heterointerface at breakdown and at operating voltage vs. charge for constant x_d of $0.3 \mu m$: comparison between InP-only and InP/InGaAs ionization.

9. Ionization outside the "multiplication" region

It may be worthwhile at this point to compare the results obtained from the analytical approximation of an abrupt p-i-n structure to a real experimental device with Zn and Si profiles which are not step functions, in order to evaluate the extent to which the analytical expression may be trusted. For example, the doping data in the SIMS plot from Fig. 5.3.1 could be used. Fig. 9.2.7 shows both the real and approximate E-field profile based on rectangular integration of this particular SIMS data using Microsoft Excel. The SIMS data was acquired only to a depth of $4.3\text{ }\mu\text{m}$ because it was evident that there was no change in the doping in the InGaAs layer. Accordingly, the same doping is assumed from this depth until $3.2\text{ }\mu\text{m}$ below the charge sheet in Fig. 9.2.7. There is one adjustable parameter for the integration, namely the constant of integration. This was used to determine the maximum E-field, as will be discussed below.

It is evident that the approximate E-field profile represents well the real E-field profile. In fact, it appears that the p^+ -n approximation is quite good, and better than the δ -doping approximation. The 10-90% rise and fall distances are 0.15 and $0.25\text{ }\mu\text{m}$ respectively. Note that the charge sheet could easily be made an order of magnitude thinner if desired.

Fig. 9.2.8 shows, for the high field InP region alone, both the real and approximate ionization coefficient profile for Osaka's ionization coefficients. The approximation for ionization coefficients is even better than for the E-field. This is because the ionization coefficients are exponential functions of the E-field. The ionization coefficients reach 10% of their value at 71% of the E-field for α and 64% of the E-field for β . The 10-90% fall distance is approximately $0.08\text{ }\mu\text{m}$, and the rise distance is less than this but indeterminate from the data due to the grid spacing. The expression for gain includes integrals involving exponentials of the ionization coefficients, so the

9. Ionization outside the "multiplication" region

Figure 9.2.7 E-field vs. x . The solid line is obtained from SIMS data. The dashed line is the approximation used in the analytical modeling.

Figure 9.2.8 Ionization coefficients vs. x . The solid line is obtained from SIMS data. The dashed line is the approximation used in the analytical modeling.

9. Ionization outside the "multiplication" region

level of approximation for gain would be expected to improve even further beyond that of the ionization coefficients alone.

Using the ionization coefficient profile, the denominator of Eqn. 2.3.7 was determined from numerical integration with a rectangular integration technique using Microsoft Excel. The maximum E-field (i.e. the constant of integration) was adjusted until the denominator was zero (i.e. breakdown). The E-field at which this occurred was 55.21 V/ μm , which is the average of the result from upper bound and lower bound rectangles. This agrees very well with the E-field of 55.57 V/ μm which is determined from the InP/InGaAs model for $x_d=0.375\text{ }\mu\text{m}$, $\sigma=3.13\times 10^{12}\text{ cm}^{-2}$ used for the approximation in Fig. 9.2.7.

The error with assigning an exact E-field for the approximation ($\pm 0.1\text{ V}/\mu\text{m}$ for the InP, $\pm 0.5\text{ V}/\mu\text{m}$ for the InGaAs) introduces an error of $\pm 1.3\%$ on σ . The error in x_d from the approximation is $\pm 0.05\text{ }\mu\text{m}$. These errors introduce an error of $0.2\text{ V}/\mu\text{m}$ on the calculated E-field – i.e. 55.57 (± 0.2) V/ μm , which makes the agreement with the numerically integrated value of 55.21 V/ μm even more remarkable.

The conclusion is that the analytical result can reproduce a result from a numerical integration of a SIMS profile very well. In the remainder of this chapter, the analytical model will be used without further qualification.

9.3 Voltage

9.3.1 Modeling results

The breakdown voltage has been calculated from Poisson's equation. For the case of undoped InGaAs, this simply involves combining the area of two rectangles of an E vs. x relationship. For this work, however, a doping of $1 \times 10^{15} \text{ cm}^{-3}$ was assumed for the InGaAs in order to make the breakdown voltage represent more realistically that of an actual experimental APD and allow for comparison with Section 4.4.

The breakdown and operating voltages of the device vs. x_d are plotted in Fig. 9.3.1 for $\sigma = 3.0 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization. Both voltages vary smoothly in the x_d domain of 0.1-0.5 μm : the breakdown voltage is nearly constant at just over 59 V from 0.1-0.2 μm , decreasing to 52.96 V at $x_d = 0.5 \mu\text{m}$ (a drop of 7 V), and the E-field at a gain of 10 decreases from 57.74 to 48.00 V over the domain of 0.1-0.5 μm (a drop of 12 V). This is in contrast to the InP-only ionization result where the breakdown field decreases from 115.3 to 50.7 V (a drop of 65 V) and the operating point field decreasing from 102.4 to 45.6 V/ μm (a drop of 57 V) over the same domain. The physical explanation for the difference in voltages is as follows. From Gauss' Law, the E-field in the InGaAs is uniquely determined from the maximum E-field in the InP and the magnitude of the doping. For a given doping, the large difference in maximum E-field for the bulk-InP ionization and the InP/InGaAs ionization necessarily translates into a large difference in breakdown voltage. Once x_d becomes as large as 0.5 μm , E-field differences between the InP-only ionization and the InP/InGaAs ionization are small, and so are the voltages. The difference in voltages between breakdown and a

Figure 9.3.1 Breakdown and operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization.

Figure 9.3.2 Breakdown and operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$ for InP/InGaAs ionization.

9. Ionization outside the "multiplication" region

gain of 10 in the x_d domain of 0.1-0.5 μm increases from 2.5 to 5.0 V/ μm . This is in contrast to the InP-only ionization result where the difference in the same voltages decreases from 12.9 to 5.1 V/ μm .

The breakdown and operating voltages of the device vs. σ for $x_d=0.3 \mu\text{m}$ are plotted in Fig. 9.3.2 for InP/InGaAs ionization. Both voltages vary smoothly as σ increases from 2.4 to $3.6 \times 10^{12} \text{ cm}^{-2}$. The breakdown voltage decreases from 66.89 to 34.67 V and the operating point of a gain of 10 decreases from 65.04 to 29.14 V over this domain. This is in contrast to the InP-only ionization result of 60.87 and 55.31 V breakdown and operating voltages respectively, independent of σ . Also, the difference between the breakdown and operating voltages increases from 1.85 to 5.53 V as σ increases from 2.4 to $3.6 \times 10^{12} \text{ cm}^{-2}$.

Comparisons between the applied bias for InP-only and InP/InGaAs ionization are shown in Figs. 9.3.3 and 9.3.4. In Fig. 9.3.3, the reverse bias vs. x_d for $\sigma=3.0 \times 10^{12} \text{ cm}^{-2}$ is shown. As noted above, for decreasing x_d the voltage necessary for breakdown or $M=10$ increases sharply for InP-only ionization. This is because the E-field in the InP for a particular gain increases, the E-field within the entire device also increases for a given charge. Since the InGaAs is much thicker than the InP and receives an increase in the E-field, the voltage increase is substantial.

In Fig. 9.3.4, the reverse bias vs. σ for $x_d=0.3 \mu\text{m}$ is shown. For InP-only ionization, the voltage necessary to achieve breakdown or $M=10$ increases linearly with decreasing σ as the E-field in the InGaAs increases linearly. This is expected upon inspection of Fig. 4.2.2.

9. Ionization outside the "multiplication" region

Figure 9.3.3 Breakdown and operating voltage vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$: comparison between InP-only and InP/InGaAs ionization.

Figure 9.3.4 Breakdown and operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$: comparison between InP-only and InP/InGaAs ionization.

9.3.2 Comparison with experiment

It is observed experimentally from InP thickness optical fringe maps and SIMS analysis that x_d varies by only $\pm 0.1 \mu\text{m}$ over most of a whole wafer, whereas the charge varies by $\pm 10\%$ over the same location. Therefore on a given whole wafer, the appropriate simulations are those of Fig. 9.3.2. The correlation of the BV to the known charge variation over a wafer as experimentally presented in Figs. 8.2.1, 8.2.2 and 8.2.3 is in agreement with the theoretical prediction from Fig. 9.3.2 in that if the doping is varied over a wafer, keeping x_d constant, the BV should decrease and high gain should be apparent at earlier voltages relative to breakdown as the doping increases. Also, it is apparent from the experimental M-V curves of Fig. 8.2.1 that as the BV increases (charge decreases) the voltage difference between a gain of 10 and breakdown decreases, again in agreement with the simulations of Fig. 9.3.2. Taken together, these observations confirm the validity of the InP/InGaAs ionization model.

At the same time, for the experiment presented in Section 8.1, x_d was varied by $0.7 \mu\text{m}$ with a best attempt made to keep σ constant. For this case, the appropriate simulations are those of Fig. 9.3.1. The correlation of the breakdown voltage to the known x_d variation among the samples as experimentally presented in Fig. 8.1.2 is in qualitative agreement with the theoretical prediction from Fig. 9.3.1 that the breakdown voltage should vary slowly with respect to x_d for InP/InGaAs ionization. This is in contrast with the large variations in breakdown voltage with respect to x_d predicted from InP-only ionization, as illustrated in Fig. 4.4.2. Note that although σ was made as constant as possible for this experiment, there is probably enough

variation in σ throughout the samples to have some influence over the results, so quantitative agreement is not expected.

This information is presented quantitatively in Fig. 9.3.5. The BV for both the case of InP-only and InP/InGaAs ionization was simulated for the thicknesses and charge given in Table 8.1.1. As these calculations involve only a high field InP region and a low field InGaAs region, the low field undoped buffer-InP layer was considered to be part of the InGaAs for the purposes of this simulation. It is evident upon inspection of Fig. 9.3.5 that the InP/InGaAs ionization describes the BV variation with respect to x_d much better than does the InP-only ionization.

Figure 9.3.5 Breakdown and operating voltage vs. charge for constant x_d of $0.3 \mu\text{m}$: comparison between InP-only and InP/InGaAs ionization.

9.4 M(x)

9.4.1 Modeling results

It is worthwhile exploring the multiplication as a function of where the electron-hole pair is created. In this section, five different locations are considered. First, injection from the p-contact is considered, otherwise known as M_e . Second, generation at the InP/InGaAs heterointerface is considered, defined as M_{het} . Third and fourth, generation at one absorption length below the heterointerface at 1.3 and 1.55 μm wavelength in n^- InGaAs is considered, defined as $M_{1.3}$ and $M_{1.55}$. Fifth, generation from the n-contact is considered, otherwise known as M_h .

The absorption profile for light at 1.3 and 1.55 μm wavelength is exponential, with the average absorption point one absorption length below the heterointerface for a front entry device with infinitely thick InGaAs (Eqn. (2.5.4)). For a proper treatment of the absorption profile, the integral of the generation profile convolved with $M(x)$ must be performed. However, the essential device physics may be more easily illustrated by assuming that the generation of electron-hole pairs occurs as a δ -function centered one absorption length below the heterointerface. For this simplified case, the analytical expression of Eqn. (9.1.1) may be used to determine M for each of the five injection conditions.

The multiplication for all five injection conditions vs. x_d at the operating point is plotted in Fig. 9.4.1 for $\sigma=3.0 \times 10^{12} \text{ cm}^{-2}$. The operating point is defined as a gain of 10 at one absorption length below the heterointerface for a wavelength of 1.3 μm – i.e. $M_{1.3}=10$. This definition of operating point has been used for all the InP/InGaAs ionization simulations

Figure 9.4.1 Multiplication for various injection conditions vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for InP/InGaAs ionization.

Figure 9.4.2 Ionization coefficients in InP and InGaAs vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$.

9. Ionization outside the "multiplication" region

above in this work but it is only at this point that the dependence of gain on injection wavelength has been considered. All the multiplications vary smoothly in the x_d domain of 0.1-0.5 μm : M_e varies from 10.4 to 6.2 and all the remaining four vary from 6.5 to 11.7 at $x_d=0.1 \mu\text{m}$ but converge to 10.0 at $x_d=0.5 \mu\text{m}$. This predicts that the multiplication should be a strong function of where the carriers come from within the InGaAs for x_d as small as 0.1 μm , with the dependence on location within the InGaAs decreasing as x_d increases to 0.5 μm . This may be understood in terms of the ionization coefficients, also plotted vs. x_d at the operating point in Fig. 9.4.2 for $\sigma=3.0 \times 10^{12} \text{ cm}^{-2}$. For $x_d=0.1 \mu\text{m}$, the InGaAs ionization coefficients are within 1-1.5 orders of magnitude of the InP ionization coefficients. Considering that $t_{\text{InGaAs}}/x_d=32$, the difference in ionization from injection at the n-contact to injection at the heterointerface is significant. As x_d increases to 0.5 μm the InGaAs ionization coefficients are down 3-4 orders of magnitude from the InP ionization coefficients. Combined with the observation that t_{InGaAs}/x_d is down to 6.4, the multiplication is at this point independent of origin within the InGaAs. It is worth noting that at $x_d=0.3 \mu\text{m}$, the InGaAs ionization coefficients are down 2-2.5 orders of magnitude from the InP ionization coefficients yet the location within the InGaAs of the generation is still not negligible. In order to neglect the InGaAs ionization for this doping, the ionization coefficients must be four orders of magnitude lower than the InP ionization coefficients. Also, the E-field in the InP is the greatest for $x_d=0.1 \mu\text{m}$ but the InP multiplication distance the smallest. In this case there is a substantial difference between M_e and M_{het} , although this depends on σ as well.

The multiplication for all five injection conditions vs. σ for $x_d=0.3 \mu\text{m}$ is plotted in Fig. 9.4.3. The results are consistent with the previous observations in Fig. 9.4.1. At the lowest doping, the multiplication is strongly

9. Ionization outside the "multiplication" region

Figure 9.4.3 Multiplication for various injection conditions vs. charge for constant x_d of $0.3 \mu\text{m}$ for InP/InGaAs ionization.

Figure 9.4.4 Ionization coefficients in InP and InGaAs vs. charge for constant x_d of $0.3 \mu\text{m}$.

9. Ionization outside the "multiplication" region

dependent on the generation point location, consistent with the arguments in the previous paragraph. As σ increases, M_e decreases from 8.6 to 7.1, with little change above $\sigma=3.3 \times 10^{12} \text{ cm}^{-2}$. As above, the ionization coefficients are plotted vs. σ at the operating point in Fig. 9.4.4 for $x_d=0.3 \mu\text{m}$. In order to neglect the InGaAs ionization for this x_d , the ionization coefficients must be five orders of magnitude lower than the InP ionization coefficients.

It is worthwhile to consider what fraction of the multiplication comes from the InP for these cases. Accordingly, for the E-field which gives a gain of 10 for $1.3 \mu\text{m}$ light, the gain for injection at the heterointerface for both InP-only and InP/InGaAs ionization is plotted in Figs. 9.4.5 and 9.4.6.

Note that $\alpha_{\text{InP}}/\beta_{\text{InP}}$ (i.e. k assuming InP-only ionization), as determined from substitution of the E-fields obtained from InP/InGaAs ionization into Eqn. (4.5.14), reaches a minimum value of 0.658 for $\sigma=3.6 \times 10^{12} \text{ cm}^{-2}$. This may be compared with $\alpha_{\text{InP}}/\beta_{\text{InP}}=0.668$ for InP/InGaAs ionization (a difference of 0.010) and $\alpha_{\text{InP}}/\beta_{\text{InP}}=0.702$ for the case of InP-only ionization for $x_d=0.3 \mu\text{m}$ (a difference of 0.044). The difference in the $\alpha_{\text{InP}}/\beta_{\text{InP}}$ values is because the maximum E-field in the InP depends on whether or not ionization in the InGaAs is considered.

Figure 9.4.5 Multiplication for injection at the heterointerface for both InP/InGaAs and InP-only multiplication vs. x_d for constant doping of $3 \times 10^{12} \text{ cm}^{-2}$ for the same conditions which give a gain of 10 for injection with $1.3 \mu\text{m}$ wavelength light for InP/InGaAs ionization.

Figure 9.4.6 Multiplication for injection at the heterointerface for both InP/InGaAs and InP-only multiplication vs. charge for constant x_d of $0.3 \mu\text{m}$ for the same conditions which give a gain of 10 for injection with $1.3 \mu\text{m}$ wavelength light for InP/InGaAs ionization.

9. Ionization outside the "multiplication" region

9.4.2 Comparison with experiment

It is worth considering some of the implications of the difference in multiplication depending on where the carriers are generated. From Figs. 9.4.1 and 9.4.3 it is evident that $M_{1.55} < M_{1.3}$ in all cases, becoming more apparent as either x_d or σ is reduced. This leads to the prediction that for a front entry device, for a given input photocurrent and for a given bias, the gain should depend on the wavelength of input light. (The same should be true for a rear entry device.) From the device physics presented here, this prediction is not restricted to SAGCM APDs but should apply to almost all standard InP-based APDs, including the SAM and SAGM configurations, although numerical simulations would be required to prove this. However, it is difficult to separate the quantum efficiency from the gain for typical SAM and SAGM APDs, making it difficult to verify this prediction. This may account for the lack of any previous experimental observations on this point. With the SAGCM structure in which there is a built-in unity gain regime, so it should be possible to verify the subtle differences in gain at 1.3 and 1.55 μm . Another prediction follows: the gain of an APD should usually differ depending on whether the light is incident from the front or the rear, because the average absorption point changes. For example, for $t_{\text{InGaAs}}=3 \mu\text{m}$, there should be a measurable difference in gain for 1.3 μm light since the average absorption points would be $\sim 0.8 \mu\text{m}$ below the heterointerface for front entry (more gain) and $\sim 2.2 \mu\text{m}$ below the heterointerface for rear entry (less gain). However, there should not be not very much difference in gain for 1.55 μm light, as the average absorption point in either direction is roughly halfway through the InGaAs layer.

It is possible to directly test the hypothesis relating to multiplication as a function of wavelength. A high and low breakdown voltage device can be measured at 1.3 and 1.55 μm input wavelength. From Fig. 9.3.2 (theory) and Section 8.2 (experiment) it has been shown that the APD with higher breakdown voltage has less charge than the APD with lower breakdown voltage. From the theoretical prediction of Fig. 9.4.3, for the same reverse bias and the same input photocurrent, there should be a difference in the measurable gain (less gain at 1.55 than at 1.3 μm wavelength) for the high breakdown voltage device but not for the low breakdown voltage device. This is in contrast with the bulk-InP multiplication theory which predicts that the gain is independent of the absorption wavelength.

This experiment has been done using four APDs from wafer P624, which has the thinnest charge sheet and lowest x_d of all the wafers investigated, from Table 8.2.1, and is therefore the closest approximation to a δ -doped APD. Fig. 9.4.7 shows a plot of $M_{1.55}/M_{1.3}$ vs. breakdown voltage, which is inversely related to σ , for the condition $M_{1.3}=10$. For a device with a breakdown voltage as high as 74, a gain as low as 9.4 was measured at 1.55 μm wavelength for the same reverse bias which gave a gain of 10 when 1.3 μm wavelength light was used. This is a further confirmation of the theory developed in this work. This is also the first published measurement of the dependence of gain on input wavelength. Note that in SAGM APDs this measurement would not be possible because of the difficulty in decoupling the gain from the responsivity.

Figure 9.4.7 Experimental measurement of ratio of gain at 1.55 μm wavelength to gain at 1.3 μm wavelength at a gain of 10 at 1.3 μm wavelength.

9.5 Minimum gain for full bandwidth

9.5.1 Modeling results

The minimum gain for which full bandwidth is achieved is limited by hole trapping or incomplete depletion below this gain. It therefore stands to reason that there is some minimum required value of E-field at the heterointerface (E_{het}), probably dependent on the characteristics of the grading layer, below which hole trapping or incomplete depletion is significant and above which it is not, permitting full bandwidth. For the case of constant charge, the effect of varying x_d is demonstrated in Fig. 4.2.2, on which E_{het} is indicated. Since the E-field in the InP "multiplication" region remains constant for a constant E_{het} , the multiplication increases as x_d increases. For the case of constant x_d , the effect of varying charge is demonstrated in Fig. 9.5.0. For a constant x_d and a constant E_{het} , the E-field in the InP "multiplication" region and hence the multiplication increases as x_d increases.

Simulations of the voltage for an ideal δ -doped APD with ionization in both the InP and InGaAs have been carried out for specific E_{het} values of 7 V/ μ m and 9 V/ μ m, which are arbitrary choices. For each E_{het} the voltages vs. x_d are presented in Fig. 9.5.1 and the voltages vs. charge are presented in Fig. 9.5.2. The simulated data points are presented as closed circles or squares. Also shown in the same figures are the breakdown and operating voltages, where the simulated data points are presented as open circles or squares. The useful operation of the device is restricted to voltages above some line of constant E_{het} . At the same time, the useful operation is also restricted to those voltages below the breakdown voltage, and ideally the device would be biased at the operating voltage, in this case (as throughout

Figure 9.5.0 Illustration of the effect of constant x_d and varying charge for constant E_{het} .

this section) assumed to be for $M=10$ for $1.3 \mu\text{m}$ light. The intersection of the operating voltage curve with the curve of constant E_{het} appropriate to a given device corresponds to the maximum charge or maximum x_d for which high speed operation can take place in the device. For the example simulated here of constant charge $= 3 \times 10^{12} \text{ cm}^{-2}$ (Fig. 9.5.1), the maximum x_d possible does not fall within the simulated range of $0.1\text{-}0.5 \mu\text{m}$; however, extrapolation by eye shows that the maximum x_d is $0.51 \mu\text{m}$ for $E_{het}=9 \text{ V}/\mu\text{m}$ and $0.55 \mu\text{m}$ for $E_{het}=7 \text{ V}/\mu\text{m}$. For the example simulated here of constant $x_d = 0.3 \mu\text{m}$ (Fig. 9.5.2), the maximum charge possible is $3.42 \times 10^{12} \text{ cm}^{-2}$ for $E_{het}=9 \text{ V}/\mu\text{m}$ and $3.57 \times 10^{12} \text{ cm}^{-2}$ for $E_{het}=7 \text{ V}/\mu\text{m}$. The lower the E_{het} required, presumably due to higher quality of the grading layer and lower doping in the absorption layer, the greater the operational range of the device.

9. Ionization outside the "multiplication" region

Figure 9.5.1 Voltages for breakdown, $M=10$, $E_{\text{het}}=9 \text{ V}/\mu\text{m}$, and $E_{\text{het}}=7 \text{ V}/\mu\text{m}$ vs. x_d for constant charge of $3 \times 10^{12} \text{ cm}^{-2}$.

Figure 9.5.2 Voltages for breakdown, $M=10$, $E_{\text{het}}=9 \text{ V}/\mu\text{m}$, and $E_{\text{het}}=7 \text{ V}/\mu\text{m}$ vs. charge for constant x_d of $0.3 \mu\text{m}$.

9. Ionization outside the "multiplication" region

Having established the minimum voltage for useful bandwidth, the minimum gain is calculated in Figs. 9.5.3 and 9.5.4 for the cases of constant charge = $3 \times 10^{12} \text{ cm}^{-2}$ with varying x_d and constant $x_d = 0.3 \mu\text{m}$ with varying charge respectively. For the case of constant charge = $3 \times 10^{12} \text{ cm}^{-2}$ it is predicted that the minimum gain is 1.3 for $x_d = 0.1 \mu\text{m}$ for either E_{het} simulated. As x_d increases, the minimum gain for full bandwidth depends in a sensitive way on E_{het} . For $x_d = 0.4 \mu\text{m}$, the minimum gain is 3 for $E_{\text{het}} = 7 \text{ V}/\mu\text{m}$ and 5 for $E_{\text{het}} = 9 \text{ V}/\mu\text{m}$. For higher x_d the disparity becomes more exaggerated. For the case of constant $x_d = 0.3 \mu\text{m}$ it is predicted that the minimum gain is 1.4 for charge = $2.4 \times 10^{12} \text{ cm}^{-2}$ for either E_{het} simulated. As charge increases, the minimum gain for full bandwidth depends in a sensitive way on E_{het} . For charge = $3.3 \times 10^{12} \text{ cm}^{-2}$, the minimum gain is 3 for $E_{\text{het}} = 7 \text{ V}/\mu\text{m}$ and 5 for $E_{\text{het}} = 9 \text{ V}/\mu\text{m}$. For higher charge the disparity becomes more exaggerated.

9.5.2 Comparison with experiment

The theory for minimum gain for useful bandwidth would be of more benefit if it could predict correlations based on measured parameters. Although it is convenient from a device physics point of view to simulate these properties as functions of x_d or charge, from an experimental point of view it is worth predicting the minimum gain for full bandwidth as a function of breakdown voltage, which itself is a function of charge from Section 9.3. The results are shown in Figs. 9.5.5 and 9.5.6 for the cases of constant charge = $3 \times 10^{12} \text{ cm}^{-2}$ and varying x_d and constant $x_d = 0.3 \mu\text{m}$ and varying charge respectively. Experimentally it is known that x_d varies by only $\pm 0.1 \mu\text{m}$ over most of a wafer, but the charge varies by $\pm 10\%$ over the same location. Therefore on a given wafer, the appropriate simulation is that of

Figure 9.5.3 Minimum gain at full bandwidth vs. x_d for constant charge of $3 \times 10^{12} \text{ cm}^{-2}$ assuming $E_{\text{het}} = 9 \text{ V}/\mu\text{m}$, $E_{\text{het}} = 7 \text{ V}/\mu\text{m}$.

Figure 9.5.4 Minimum gain at full bandwidth vs. charge for constant x_d of $0.3 \mu\text{m}$ assuming $E_{\text{het}} = 9 \text{ V}/\mu\text{m}$, $E_{\text{het}} = 7 \text{ V}/\mu\text{m}$.

9. Ionization outside the "multiplication" region

Figure 9.5.5 Minimum gain at full bandwidth vs. breakdown voltage for constant charge of $3 \times 10^{12} \text{ cm}^{-2}$ assuming $E_{het}=9 \text{ V}/\mu\text{m}$, $E_{het}=7 \text{ V}/\mu\text{m}$.

Figure 9.5.6 Minimum gain at full bandwidth vs. breakdown voltage for constant x_d of $0.3 \mu\text{m}$ assuming $E_{het}=9 \text{ V}/\mu\text{m}$, $E_{het}=7 \text{ V}/\mu\text{m}$.

9. Ionization outside the "multiplication" region

Fig. 9.5.6. The simulations predict that as the breakdown voltage decreases, the minimum gain for full bandwidth increases substantially. This curve looks quite similar to the experimental results from Fig. 8.2.8. The strong similarity of the two curves may be considered as further evidence for the correctness of this theory of ionization outside the InP multiplication region.

The apparent similarity between Figs. 9.5.6 theoretically and 8.2.8 experimentally may be explored quantitatively for the thicknesses and doping regime specified in Table 8.2.1. These results are presented in Fig. 9.5.7. The theoretical curve corresponds to $E_{\text{het}} = 4.5 \text{ V}/\mu\text{m}$. It is evident that although there is some scatter in the data, the theoretical curve represents a reasonable fit to the data. Some scatter in the data is to be expected, since the results are from multiple wafers and individual devices are of different x_d values and depletion layer thicknesses, owing to the variations across a wafer. With this in mind, Fig. 9.5.7 must be construed as strong evidence that InP/InGaAs multiplication can correctly predict important device parameters not just for DC measurements such as breakdown voltage and multiplication as a function of wavelength of injected light, but also for high frequency parameters such as the minimum gain for high bandwidth. While complex numerical simulations can be used to further refine the simulations presented here, it is evident that simple algebraic expressions such as those in Eqns. (9.1.1) and (9.1.2) describe well the important characteristics concerning the behavior of SAGCM APDs.

9. Ionization outside the "multiplication" region

Figure 9.5.7 Minimum gain at full bandwidth vs. breakdown voltage for three experimental wafers. The theoretical curve is simulated for $x_d = 0.3 \mu\text{m}$ and $E_{het} = 4.5 \text{ V}/\mu\text{m}$.

9.6 Summary

In this chapter a new analytical model of a δ -doped APD with ionization in both the InP and the InGaAs was developed. Comparison to a numerically integrated result based on an experimental SIMS profile indicates the analytical expressions are very accurate. Many theoretical predictions of this model were made. The experimentally verifiable predictions involve relationships between two measurable parameters. Quantitative agreement between theory and experiment was obtained for the breakdown voltage as a function of x_d and minimum gain for usable bandwidth as a function of breakdown voltage. Qualitative agreement was obtained for the breakdown voltage as a function of charge and gain vs. wavelength as a function of breakdown voltage.

10. Artificially enhanced ionization (AEI) APDs

10.1 Motivation

One of the main limitations of the bulk-InP multiplication layer APD is its k -factor, which is a material parameter which can in principle range from 1 (worst case) to 0 (best case). It is readily appreciated that in III-V materials lattice matched to InP, it is difficult to find bulk materials with a substantially low k . For example, in the InGaAsP materials system, $k \geq 0.3$ for any reasonable thickness of materials and is typically 0.4-0.5 in commercial APDs or the APDs presented in this work. The highest GBW in excess of 100 GHz [30] is achievable only under stringent fabrication tolerances.

Suppose two different materials are brought together. In general, the differing bandgaps and electron affinities of the two materials will result in a particular alignment of conduction and valence bands, as shown in Fig. 10.1. For clarity, electrons are considered here although the arguments are equally valid for holes. Consider the presence of an E-field. As electrons pass from material 1 to material 2, they gain energy equivalent to the conduction band discontinuity ΔE_c over a distance scale much smaller than the phonon scattering mean free path. This gives artificially enhanced ionization (AEI) in addition to whatever natural ionization may occur from the E-field. This possibility gives rise to artificial discrimination between hole and electron ionization not possible from considering bulk ionization alone, and offers intriguing possibilities for reducing k . However, in general a given electron can cause only one ionization event before losing its energy, thus limiting this beneficial effect to the regime where $M \leq 2$ for a single heterointerface. Suppose a stack of alternating materials is constructed, and suppose further that the E-field was sufficiently large to prevent carrier trapping from the low

to the high bandgap material. Then multiple ionization events can occur from a given initial electron, as artificial ionization is possible every time the electron falls into the potential well. In this case, the maximum gain is 2^N from N pairs of layers from materials 1 and 2. For this arrangement it is possible to envisage a practical AEI APD operating at high gains and low k .

This idea, first suggested by Capasso [12], is of active research interest in an attempt to improve the performance of InP-based APDs both from a theoretical [22, 134-136] and experimental [137-141] approach. It has been recently demonstrated that artificially enhanced ionization (AEI) of carriers can lead to an improvement in the k -factor of an InP-based APD [137-141]. This has the effect of increasing the optimum gain (at a given bit rate) and the GBW (for a given multiplication region length). The signal to noise ratio is optimized at the optimum gain, at which the noise increases exactly as fast as the signal; if the gain is further increased the noise (superlinear) increases faster than the signal (linear). If the optimum gain is increased, the receiver sensitivity improves, provided the GBW is sufficient to provide enough bandwidth at this higher gain. It is necessary to increase the GBW in order both to take maximum advantage of the higher optimum gain and to guarantee a high enough GBW ($\gg 100$ GHz) for a range of multiplication region lengths which could be achieved in practice for manufacturing.

While there are promising experimental results for AEI APDs [137-142], there is no published design methodology to date. This chapter review existing experimental work, presents original explicit design criteria for AEI APDs, discusses existing experimental work in the context of these criteria and presents predictions of new materials combinations for improved performance [154].

Figure 10.1 Schematic diagram of artificial enhancement ionization APD.

A further benefit of maximizing $\Delta E_C/E_{\text{low}}$ is that as much of the ionization as possible is due to the heterojunctions rather than a high E-field. This in turn reduces the magnitude of charge required for the charge sheet while maintaining a similar absolute tolerance – i.e. a higher relative tolerance. It is expected that the charge sheet doping would no longer be as critical a step as it is in the present APD structure.

In these particular AEI structures electrons rather than holes must be injected, opposite to the case of bulk InP multiplication. This has implications for device geometry. The simplest way to realize these AEI devices is to grow on p^+ InP substrates and form mesa rather than planar devices. As the p-n junction could be in a high bandgap material, the effect on leakage current should not be too severe.

10.2 InGaAs/InAlAs experimental results

There are many materials options possible in order to realize AEI devices. The goals of such work are the lowest possible k with sufficiently low dark current (I_d) (100 nA at the operating gain is probably good enough), sufficiently large size to couple light from a single mode fiber (of core diameter 8 μm) and sufficiently high yield. The parameters of importance are the conduction and valence band discontinuities (ΔE_c and ΔE_v) and the ratio of ΔE_c to the well bandgap E_g . A low k can be achieved only by selective ionization. The application of a high E-field tends to ionize both carriers, as does a discontinuity in both the valence and conduction bands of an AEI structure.

The InAlAs/InGaAs system lattice matched to InP has been investigated for a few years as a candidate for AEI [137-140]. In this case ΔE_c is 0.5 eV, compared with ΔE_v of 0.2 eV. The large ΔE_c is usually given as the reason why this is an attractive structure. For this structure $k=0.1$ has been achieved together with bandwidths in excess of 10 GHz, but with high I_d . The origin of I_d is tunneling in the low bandgap InGaAs. This is impossible to avoid in the presence of a high field. These structures will not fulfill their potential for improved sensitivity until a way is found to combine a low k -factor with a low dark current. Other solutions are clearly needed.

10.3 Ideal AEI, Type $\frac{1}{2}$ APD

An AEI region typically consists of an alternating stack of two materials of different conduction or valence bands, as shown in Fig. 10.1. As electrons and holes pass from high bandgap to the low bandgap material (of bandgap energies E_{high} and E_{low}), they gain energy equivalent to the conduction and valence band discontinuities ΔE_c and ΔE_v . This energy gain results in AEI; by

Artificially enhanced ionization APDs

controlling ΔE_c and ΔE_v it should be possible to selectively enhance the ionization of one or the other carrier. This effect is not possible from bulk ionization alone, for which an increasing E-field increases the ionization coefficients of both carriers.

For the case of AEI of electrons, there are three main parameters of interest for AEI APDs: ΔE_c , ΔE_v and $\Delta E_c/E_{low}$. The greater the ΔE_c , the greater the energy artificially gained by the electron. The greater the $|\Delta E_v|$, the greater the unwanted hole ionization. One possible drawback for high ΔE_c or ΔE_v is electron or hole trapping, the latter observed in some InP/InGaAs APDs [22]. For $\Delta E_c/E_{low} \geq 1$, the electron acquires enough energy to cause ionization merely by "falling" into the low bandgap material [47], so ionization should occur for an E-field large enough only to provide transport for the carriers. As $\Delta E_c/E_{low}$ decreases, the difference in ionization energy to provide multiplication must be made up by increasing the E-field, which is not selective between electron and hole ionization and therefore unfavorable for k.

AEI of both carriers is present in both the "Type 1" (nested band alignment) and "Type 2" (staggered band alignment) HJs, as illustrated in Fig. 10.2. In the Type 1 HJ, all the enhancement takes place in the lower bandgap material, whereas in the Type 2 HJ the enhancement of the different carriers takes place in different materials. The most desirable situation for AEI would be the "Type 1 $\frac{1}{2}$ " heterojunction [154], where either the valence or conduction band remains continuous across the heterointerface, as illustrated in Fig. 10.3 for the case of continuous valence band. This alignment has also been referred to as "Type III" [170]. For this case the electron ionization only would be artificially enhanced, with the hope of providing the lowest possible k-factor. There is a second condition for ideal AEI: namely $\Delta E_c/E_{low} \geq 1$. For

these conditions (Type $1\frac{1}{2}$ HJ and $\Delta E_c/E_{low} \geq 1$), the APD should operate at nearly zero E-field (i.e. the E-field needed is that only large enough to provide transport for the carriers), with ionization coming exclusively from one carrier, which should result in a very low k .

Figure 10.2 Type 1 and 2 heterojunctions.

Figure 10.3 Type $1\frac{1}{2}$ heterojunction.

10.4 Band offsets, Type $\frac{1}{2}$ APD in real materials systems

There is considerable controversy over how to measure conduction and valence band offsets at a heterojunction, with different measurement methods yielding different results on the same heterojunction [143-153]. Nevertheless, it appears that there is general agreement that $\text{In}_{.52}\text{Al}_{.48}\text{As}/\text{In}_{.53}\text{Ga}_{.47}\text{As}$ is a "Type 1" heterojunction [151-153] whereas $\text{In}_{.52}\text{Al}_{.48}\text{As}/\text{InP}$ is a "Type 2" heterojunction [149, 150, 152].

It is well known that the discontinuities within the InGaAsP system lattice matched to InP the bandgap discontinuity is 40% in the conduction band and 60% in the valence band [143, 151]. Since the conduction and valence band edges vary continuously and monotonically relative to the vacuum level from InP to $\text{In}_{.53}\text{Ga}_{.47}\text{As}$, it follows that for some quaternary composition of $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ lattice matched to InP the valence band discontinuity relative to $\text{In}_{.52}\text{Al}_{.48}\text{As}$ should be zero. Also, since the quaternary is of a higher bandgap energy than $\text{In}_{.53}\text{Ga}_{.47}\text{As}$, the tunneling probability should be less for a given E field.

The InGaAs/InAlAs/InP measurements are, however, but a subset of the work done recently on determining the conduction and valence band discontinuities in some different III-V materials systems. For example, Lee et al. [153] have found that for $\text{InAlAs}_{\text{strained}}/\text{InGaAs}_{\text{unstrained}}$ the increase in the bandgap from the strained InAlAs is 70% in the conduction band and 30% in the valence band. Also, Schuermeyer et al. [148] have done interpolation of some of the known parameters for binary systems to obtain estimates of discontinuities for various ternary systems. They demonstrated reasonable

agreement with certain measured cases, for example InAlAs/InGaAs as mentioned above.

10.5 Design of AEI APDs

To illustrate the application of the three main parameters of interest discussed above, design curves for AEI structures are presented for the first time in Figs. 10.4-10.6 for some specific materials systems [154]. Three high bandgap materials are chosen: InAlAs (unstrained and 0.3 % strain) and unstrained AlAsSb. The low bandgap materials are chosen to be InGaAsP lattice matched to InP. Literature values for ΔE_c and ΔE_v [148, 149, 153] are used in these calculations. ΔE_c (Fig. 10.4), ΔE_v (Fig. 10.5) and $\Delta E_c/E_{low}$ (Fig. 10.6) are plotted as a functions of E_{low} , which ranges from 0.75 eV for $In_{.53}Ga_{.47}As$ to 1.35 eV for InP.

Both ΔE_c (Fig. 10.4) and $\Delta E_c/E_{low}$ (Fig. 10.6) decrease as the bandgap of the InGaAsP increases for each high bandgap material. For each InGaAsP composition, both ΔE_c and $\Delta E_c/E_{low}$ increase from unstrained InAlAs to strained InAlAs to AlAsSb. Above the horizontal line at $\Delta E_c/E_{low}=1$ in Fig. 10.6, ionization can occur without a high E-field. ΔE_v (Fig. 10.5) passes through zero and changes sign for a particular composition, different for each high bandgap material. These "zero-crossing" compositions correspond to a Type $1\frac{1}{2}$ HJ. Note that quantum effects associated with thin layers would alter ΔE_c and ΔE_v from what would be expected for bulk HJs; such effects are not taken into account in this work. Computer simulations can be done to obtain the more accurate values for finite well and barrier size. In fact, algorithms for numerical work have recently been published which are applicable to this specific problem [155, 156]. Also, the detailed device physics

Figure 10.4 ΔE_c vs. E_{low} .

Figure 10.5 ΔE_v vs. E_{low} .

Figure 10.6 $\frac{\Delta E_c}{\Delta E_{low}}$ vs. E_{low} .

of strained layers, superlattices and the like is an emerging field [157-159]. Nevertheless the basic trends are well represented by Figs. 10.4-10.6.

It is evident that for these materials systems there is a tradeoff between minimizing $|\Delta E_v|$ and maximizing $\Delta E_c/E_{low}$ with respect to E_{low} . The most commonly pursued SL APD structure is $\text{In}_{.52}\text{Al}_{.48}\text{As}/\text{In}_{.53}\text{Ga}_{.47}\text{As}$ [137-140], lattice matched to InP, for which $\Delta E_c=0.5$ eV, $\Delta E_v=0.2$ eV and $\Delta E_c/E_{low}=0.67$ on Figs. 1-3. The high bandwidths approaching 10 GHz for these APDs indicates that neither electron nor hole trapping is a problem. A typical experimental k is 0.1, but there is substantial tunneling leakage current due both to the low bandgap $\text{In}_{.53}\text{Ga}_{.47}\text{As}$ and E-field required for ionization. Very recently k of 0.2 and a GBW of 110 GHz was proposed [154] and independently demonstrated [141] for a Type $1\frac{1}{2}$ HJ unstrained $\text{In}_{.52}\text{Al}_{.48}\text{As}/\text{InGaAsP}$

Artificially enhanced ionization APDs

($E_{\text{low}}=1.03$ eV) APD, for which $\Delta E_v=0.03$ eV and $\Delta E_c/E_{\text{low}}=0.38$ from Figs. 2 and 3, while avoiding tunneling. For some InGaAsP composition between these two values, a lower k than 0.2 should be possible while still avoiding the tunneling problem.

Based on Figs. 10.4-10.6, improved performance might be expected using strained InAlAs/InGaAsP, where E_{high} is increased. This is because the design goals of high $\Delta E_c/E_{\text{low}}$, low ΔE_v and high enough E_{low} to prevent tunneling can be more easily obtained from strained than from lattice-matched InAlAs/InGaAs(P) structures. The amount of strain which can be incorporated depends on the critical thickness (e.g. Matthews-Blakeslee) permitted before dislocations degrade the structure. The use of offsetting strain in the InGaAsP low bandgap material may also be considered, not only to increase the critical thickness but also because E_{low} is reduced, increasing $\Delta E_c/E_{\text{low}}$.

An even higher value of $\Delta E_c/E_{\text{low}}$ is possible by using AlAsSb as a barrier layer. For InGaAsP where $E_{\text{low}}=0.85$ eV, the desired situation $\Delta E_c/E_{\text{low}}=1$ arises. Increasing E_{low} to 1.23 eV yields a type $1\frac{1}{2}$ HJ where $\Delta E_c/E_{\text{low}}=0.57$, greater than the unstrained InAlAs/InGaAsP case where $\Delta E_c/E_{\text{low}}=0.34$. Any AlAsSb AEI APD within this InGaAsP composition range should have a very low k .

In conclusion, design criteria for AEI APDs have been presented which consider both band discontinuities and the bandgap energy of the lower bandgap material. Recent independent experimental results [141] confirm the ideas presented here. These new criteria can serve as an aid in the design of new AEI APDs. Based on these criteria, strained InAlAs or unstrained AlAsSb barrier layers in combination with InGaAsP well layers should produce improved performance over unstrained InAlAs layers.

Artificially enhanced ionization APDs

11. Conclusions

1. It was previously pointed out by Webb [27] that the use of an SAGCM APD has advantages over SAGM APDs. In this work, the use of partial charge in device periphery for an SAGCM APD was proposed, modeled and demonstrated [28]. The partial charge is beneficial to performance and fabrication issues for APDs. To demonstrate this, simulations were carried out for the first time to investigate various diffused p-n positions, junction depths, InGaAs thicknesses and fractional charge sheets in the periphery. For a given required performance, it was experimentally verified for the first time that greater fabrication tolerances can be achieved with an SAGCM APD than a SAGM APD [162-164].

2. Detailed experimental investigations were undertaken for SAGCM devices with partial charge in the periphery, with many novel results. One beneficial property of these devices is that for the first time the gain can be measured independent of responsivity, which is not possible for SAGM structures [28-30, 162-164]. Also, a novel experimental technique was developed to measure the integrated charge using actual SAGCM APDs, rather than relying on calibration wafers. It was found experimentally for the first time that the bandwidth-gain characteristic has an open "window" of usable bandwidth which moves towards higher or lower gain depending on the breakdown voltage [164]. Very high performance can be achieved in the best case, with a gain-bandwidth product of 122 GHz obtained [164], the highest published value to date.

3. Analytical expressions were used in order to predict various characteristics of bulk-InP only ionization δ -doped APDs, which are a

reasonable approximation to SAGCM APDs. Agreement with the recently published numerical results of Kuchibhotla and Campbell [31] of these analytical expressions, exact for the same assumptions as for the numerical work, was demonstrated for E-field and noise. An unexplained disagreement between these analytical results and the published numerical results by a constant factor of 1.3 was found for the gain-bandwidth product. The breakdown voltage dependence on multiplication region width was predicted for the first time. A variety of experimental results were obtained for SAGCM APDs with respect to fabrication parameters, which were compared to the theoretical results for the first time. After considering the experimental results, it appears that bulk-InP ionization is not solely responsible for the multiplication in SAGCM APDs; ionization sources exist elsewhere in the device. Accordingly, ionization in the InGaAs was considered.

4. Novel analytical expressions were developed to predict various characteristics of δ -doped APDs with ionization in both the InP and InGaAs, which are a reasonable approximation to SAGCM APDs. These expressions are exact for an ideal δ -doped APD. In most instances the observed behavior of SAGCM APDs could be accounted for. Quantitative agreement between theory and experiment was obtained for the breakdown voltage as a function of x_d and minimum gain for usable bandwidth as a function of breakdown voltage for the first time. Qualitative agreement was obtained for the breakdown voltage as a function of charge and gain vs. wavelength as a function of breakdown voltage for the first time. Also, for the first time it has been shown theoretically that the gain in an APD depends on the input wavelength. This suggestion has not been previously made. The

11. Conclusions

demonstration is possible for SAGCM APDs but would not be possible for SAGM APDs, for which the gain and responsivity are not easily decoupled.

5. Artificially enhanced ionization (e.g. superlattice multiplication region) APDs should be able to outperform bulk-InP multiplication APDs. Some novel methods by which this may be achieved are presented. Novel design curves are presented for III-V materials combinations not generally thought of for use in these types of devices. The Type 1-1/2 heterojunction superlattice APD was proposed [154]. This device has recently been experimentally demonstrated elsewhere [141].

11. Conclusions

References

- [1] J. Gower, Optical Communication Systems, Prentice/Hall, London, 1984.
- [2] J. E. Midwinter, "Current status in optical communications technology", IEEE J. Lightwave Technology LT-3, 927 (1985).
- [3] P. S. Henry, "Lightwave primer", IEEE Quantum Electronics QE-21, 1862 (1985).
- [4] H. Imai, T. Kaneda, "High speed DFB lasers and InGaAs APDs", IEEE J. Lightwave Technology LT-6, 1634 (1988).
- [5] S. R. Forrest, Ch. 4 in Lightwave Communications Technology, part D, ed. W.T. Tsang, in Semiconductors and Semimetals Vol. 22D, ed. R. K. Willardson, A.C. Beer (1985).
- [6] J. C. Campbell, Ch. 14 in Optoelectronic Technology and Lightwave Communications Systems, ed. Chinlon Lin (1989).
- [7] James L. Gimlett, "Ultrawide bandwidth optical receivers", IEEE J. Lightwave Technology LT-7, 1432 (1989).
- [8] A. R. Chraplyvy, Ch. 5 in Optoelectronic Technology and Lightwave Communications Systems, ed. Chinlon Lin (1989).
- [9] Claude Rolland, Larry E. Tarof and Azmina Somani, "Multigigabit networks: the challenge", IEEE Magazine Lightwave Telecommunications Systems LTS-2 (12), 16 (1992).
- [10] G. E. Stillman and C. M. Wolfe, Ch. 5 in Infrared Detectors II, in Semiconductors and Semimetals Vol. 12, ed. R. K. Willardson, A.C. Beer (1977).
- [11] G. E. Stillman, V. M. Robbins and N. Tabatabaie, "III-V compound semiconductor devices: optical detectors", IEEE Trans. Electron Devices

- ED-31, 1643 (1984).
- [12] Federico Capasso, Ch. 1 in Lightwave Communications Technology, part D, ed. W.T. Tsang, in Semiconductors and Semimetals Vol. 22D, ed. R. K. Willardson, A.C. Beer (1985).
 - [13] Jan Conradi, "A simplified non-Gaussian approach to digital optical receiver design with APDs: theory", IEEE J. Lightwave Technology LT-8, 1019 (1991).
 - [14] Jan Conradi, "A simplified non-Gaussian approach to digital optical receiver design with APDs: experimental", IEEE J. Lightwave Technology LT-8, 1027 (1991).
 - [15] J. Schlafer, C. B. Su, W. Powazinik and R. B. Lauer, "20 GHz bandwidth InGaAs photodetector for long-wavelength microwave optical links", Electronics Lett. 21, 469 (1985).
 - [16] J. E. Bowers, C. A. Burrus and R. J. McCoy, "InGaAs pin photodetectors with modulation response to millimetre wavelengths", Electronics Lett. 21, 812 (1985).
 - [17] D. Wake, L. C. Blank, R. H. Walling and I. D. Henning, "Top-illuminated InGaAs/InP p-i-n photodiodes with a 3-dB bandwidth in excess of 26 GHz", IEEE Electron Device Lett. EDL-9, 226 (1988).
 - [18] A. J. Moseley, A. C. Carter, M. Q. Kearley and C. A. Park, "High speed InGaAs/InP photodiodes for applications to 40 GHz", SPIE Vol. 995, High Frequency Analog Communicatrions, 61 (1988).
 - [19] M. Gallant, N. Puetz, A. Zemel, F. R. Shepherd, "MOCVD planar InGaAs p-i-n photodiodes with extremely low dark current", Appl. Phys. Lett. 52, 733 (1988).
 - [20] W. T. Read, Jr., "A proposed high-frequency, negative-resistance diode", Bell System Technical Journal 401, March, 1958.

References

- [21] "Technical report of InGaAs APDs", Revision A, Fujitsu technical report # FPD-87-032, 1987.
- [22] P. P. Webb, R. J. McIntyre and J. Conradi, "Properties of avalanche photodiodes", RCA Review **35**, 234 (1974).
- [23] S. M. Sze, Physics of Semiconductor Devices, 2nd ed., Wiley, 1981.
- [24] Y. Suematsu, K. Iga and K. Kishino, GaInAsP Alloy Semiconductors, Ch. 14, ed. T. Pearsall, Wiley, 1982.
- [25] D. A. Humphreys, R. J. King, D. Jenkins and A. J. Moseley, "Measurements of absorption coefficients of InGaAs over the wavelength range 1.0-1.7 μm ", Electronics Lett. **21**, 1187 (1985).
- [26] A. S. Grove, Physics and Technology of Semiconductor Devices,
- [27] P. Webb, R. McIntyre, J. Scheibling, M. Holunga, "A planar InGaAs APD fabricated using Si implantation and regrowth techniques", OFC, poster paper, New Orleans, 1988.
- [28] L. E. Tarof, D. G. Knight, K. E. Fox, C. J. Miner, N. Puetz and H. B. Kim, "Planar InP/InGaAs avalanche photodetectors with partial charge in the device periphery", Appl. Phys. Lett. **57** 670-672, 1990.
- [29] L. E. Tarof, "Planar InP/InGaAs avalanche photodetectors with n-multiplication layer exhibiting a very high gain-bandwidth product", IEEE Photonics Technology Letters **PTL-2**, 643-646, 1990
- [30] L. E. Tarof, "Planar InP/InGaAs avalanche photodetector with a gain-bandwidth product in excess of 100 GHz", Electronics Letters **27**, 34-35, 1991.
- [31] R. Kuchibhotla and J. C. Campbell, "Delta-doped avalanche photodiodes for high bit-rate lightwave receivers", IEEE J. Lightwave Technol. **LT-5**, 900-905, 1991.
- [32] L. E. Tarof, C. J. Miner and C. Blaauw, "Epitaxial layer thickness

- measurement of double heterostructures using reflectance spectroscopy", J. Appl. Phys. **68**, 2927, (1990).
- [33] L. E. Tarof, C. J. Miner and C. Blaauw, "InGaAsP compositional determination from reflectance spectroscopy of InGaAsP/InP heterostructures", J. Appl. Phys. **68**, 2939, (1990).
- [34] C. J. Miner, D. G. Knight, L. E. Tarof, N. Puetz, M. Gallant and A. J. SpringThorpe, "Scanning photoluminescence and reflectance characterization of InGaAs epitaxial layers on InP", presented at Defect Recognition and Image Processing III, p. 56, Tokyo, Japan, 1989.
- [35] L. E. Tarof, "Nondestructive III-V epitaxial layer thickness measurements in the visible and near infrared by reflection spectroscopy", Applied Optics **27**, 4798 (1988).
- [36] L. E. Tarof, C. J. Miner and A. J. SpringThorpe, "Epitaxial layer thickness measurement by reflection spectroscopy", Journal of Electronic Materials **18**, 361 (1989).
- [37] L. E. Tarof, Epitaxial Layer Thickness Measurement Using Reflection Spectroscopy, M. Eng. thesis, 1989.
- [38] H. Ando, H. Kambe, M. Ito, T. Kaneda, "Tunneling current in InGaAs and optimum design for InGaAs/InP APDs", Japanese J. Appl. Phys. **14**, L277 (1980).
- [39] S. R. Forrest, M. Didomenico, R. G. Smith, H. Stocker, "Evidence for tunneling in reverse-biased III-V photodetector diodes", Appl. Phys. Lett. **36**, 580 (1980).
- [40] S. R. Forrest, O. K. Kim and R. G. Smith, "Analysis of the dark current and photoresponse of InGaAs/InP APDs", Solid State Electronics **26**, 951 (1983).
- [41] S. R. Forrest, R. G. Smith and O. K. Kim, "Performance of

In_{0.53}Ga_{0.47}As, 'InP APDs', IEEE J. Quantum Electronics **QE-18**, 2040 (1982).

2.2-2.8 E12 cm⁻².

- [42] Aicha Elshabini-Riad, Jianqing He, Gary Lee Stirk and Aamira Al-Mazrooh, "Analytical model for leakage current in InGaAsP APDs", Solid State Electronics **31**, 1413 (1988).
- [43] A. Zemel and M. Gallant, "Current-voltage characteristics of metalorganic chemical vapor deposition InP/InGaAs p-i-n photodiodes: the influence of finite dimensions and heterointerfaces", J. Appl. Phys. **64**, 6552 (1988).
- [44] R. S. Fyath and J. J. O'Reilly, "Performance degradation of APD-optical receivers due to dark current generated within the multiplication region", IEEE J. Lightwave Technology **LT-7**, 62 (1989).
- [45] K. Nishida, K. Taguchi, Y. Matsumoto, "InGaAsP heterostructure APDs with high avalanche gain", Appl. Phys. Lett. **35**, 251 (1979).
- [46] P. A. Wolff, "Theory of electron multiplication in Si and Ge", Phys. Rev. **95**, 1415 (1954).
- [47] W. Schockley, "Problems related to p-n junctions in Si", Solid State Electronics **2**, 35 (1961).
- [48] C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleimack and W. Wiegmann, "Ionization rates of holes and electrons in Si", Phys. Rev. **134**, A761 (1964).
- [49] H. David Law and Charles A. Lee, "Interband scattering effects on secondary ionization coefficients in GaAs", Solid State Electronics **21**, 331 (1978).
- [50] H. Shichijo and K. Hess, "Band-structure-dependent transport and impact ionization in GaAs", Phys. Rev. B **23**, 4197 (1981).
- [51] Kevin Brennan and Karl Hess, "High field transport in GaAs, InP and

- InAs", Solid State Electronics **27**, 347 (1984).
- [52] Y.-Z. Chen and T.-W. Tang, "Impact ionization coefficient and energy distribution function at high fields in semiconductors", J. Appl. Phys. **65**, 4279 (1989).
 - [53] Masanori Ito, Takashi Mikawa and Osamu Wada, "An approach to determining impact ionization rates in semiconductor junctions", Solid State Electronics **30**, 969 (1987).
 - [54] Masanori Ito, Takashi Mikawa and Osamu Wada, "Theoretical analysis of the $-3/4$ power law in semiconductor avalanche breakdown", Semicond. Sci. Technol. **3**, 1112 (1988).
 - [55] J. S. Marsland, "Power series approximation used in soft threshold lucky drift model of impact ionization", Semicond. Sci. Technol. **5**, 177 (1990).
 - [56] I. K. Czajkowski, A. Allam, M. Silver, A. R. Adams and M. A. Gell, "Impact ionisation thresholds in Si and Ge under hydrostatic pressure and strain", IEE Proceedings **137-L**, 79 (1990).
 - [57] F. Capasso, K. Mohammed, K. Alavi, A. Y. Cho and P. W. Foy, "Impact ionization rates for electrons and holes in $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ ", Appl. Phys. Lett. **45**, 970 (1984).
 - [58] L. W. Cook, G. E. Bulman and G. E. Stillman, "Electron and hole impact ionization coefficients in InP determined by photomultiplication measurement", Appl. Phys. Lett. **40**, 589 (1982).
 - [59] T. Windhorn, L. W. Cook, M. A. Haase and G. E. Stillman, "Electron transport in InP at high electric field", Appl. Phys. Lett. **42**, 725 (1983).
 - [60] Fukunobu Osaka, Takashi Mikawa and Takao Kaneda, "Electron and hole ionization coefficients in (100) oriented $\text{Ga}_{0.33}\text{In}_{0.67}\text{As}_{0.70}\text{P}_{0.30}$ ", Appl. Phys. Lett. **45**, 292 (1984).

- [61] F. Kikunobu Osaka, Takashi Mikawa and Takao Kaneda, "Impact ionization coefficients of electrons and holes in (100)-oriented $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$ ", IEEE J. Quantum Electronics **OE-21**, 1326 (1985).
- [62] I. Umebu, A. N. M. M. Choudhury and P. N. Robson, "Ionization coefficients measured in abrupt InP junctions", Appl. Phys. Lett. **36**, 302 (1980).
- [63] C. A. Armiento and S. H. Grove, "Impact ionization in (100)-, (110)-, and (111)-oriented InP APDs", Appl. Phys. Lett. **43**, 198 (1983).
- [64] T. P. Pearsall, "Impact ionization rates for electrons and holes in InGaAs", Appl. Phys. Lett. **36**, 218 (1980).
- [65] J. Conradi, "Temperature effects in Si APDs", Solid State Electronics **17**, 99 (1974).
- [66] R. J. McIntyre, "Multiplication noise in uniform avalanche photodiode", IEEE Trans. Electron Devices **ED-13**, 164 (1966).
- [67] R. J. McIntyre, "The distribution of gains in uniformly multiplying APDs: theory", IEEE Trans. Electron Devices **ED-19**, 703 (1972).
- [68] J. S. Marsland, "On the effect of ionization dead spaces on avalanche multiplication and noise for uniform electric fields", J. Appl. Phys. **67**, 1929 (1990).
- [69] J. S. Marsland, R. C. Woods and C. A. Brownhill, "Lucky drift estimation of excess noise factor for conventional APDs including the dead space effect", IEEE Trans. Electron Devices **ED-39**, 1129 (1992).
- [70] Majeed M. Hayat, Bahaa E. A. Saleh and Malvin C. Teich, "Effect of dead space on gain and noise of double carrier multiplication APDs", IEEE Trans. Electron Devices **ED-39**, 546 (1992).
- [71] Tsugunori Takanohashi, Tatsunori Shirai, Susumu Yamazaki, Saroshi Komiya and Itsuo Umebu, "Inhomogeneous distribution of avalanche

- multiplication in InP APDs", Jpn. J. Appl. Phys. **23**, 270 (1984).
- [72] Y. K. Jhee, J. C. Campbell, J. F. Ferguson, A. G. Dentai and J. K. Plourde, "The effect of nonuniform gain on the multiplication noise of InP/InGaAsP/InGaAs APDs", IEEE J. Quantum Electronics **QE-21**, 1858 (1985).
- [73] Fukunobu Osaka and Takashi Mikawa, "Excess noise design of InP/InGaAsP/InGaAs APDs", IEEE Quantum Electronics **QE-22**, 471 (1986).
- [74] J. C. Campbell, S. Chandrasekhar, W. T. Tsang, G. J. Qua and B. C. Johnson, "Multiplication noise of InP/InGaAsP/InGaAs APDs", IEEE J. Lightwave Technology **LT-7**, 473 (1989).
- [75] L. D. Westbrook, M. D. A. MacBean and P. M. Rodgers, "Improved figure of merit for noise in avalanche photodiodes", Electronics Lett. **24**, 853 (1988).
- [76] R. B. Emmons, "APD frequency response", J. Appl. Phys. **38**, 3705 (1967).
- [77] R. B. Emmons and G. Lucovsky, "The frequency response of APDs", IEEE Trans. Electron Devices **ED-13**, 297 (1966).
- [78] C. A. Lee, R. L. Batdorf, W. Wiegmann and G. Kaminsky, "Time dependence of avalanche processes in Si", J. Appl. Phys. **38**, 2787 (1967).
- [79] R. Kuvas and C. A. Lee, "Quasistatic approximation for semiconductor avalanches", J. Appl. Phys. **41**, 1743 (1970).
- [80] J. C. Campbell, W. T. Tsang, G. J. Qua and J. E. Bowers, "InP/InGaAsP/InGaAs APDs with 70 GHz gain-bandwidth product", Appl. Phys. Lett. **51**, 1454 (1987).
- [81] J. N. Hollenhorst, D. T. Ekholm, J. M. Geary, V. D. Mattered, Jr. and R. Pawalek, "High frequency performance of planar InGaAs/InP APDs", in SPIE Vol. 995, High Frequency Analog Communications, (1988).

- [82] J. N. Hollenhorst, "Fabrication and performance of high speed InGaAs APDs", Digest of Conference on Optical Fiber Communication, 1990 Technical Digest Series, Vol. 1, 148 (Optical Society of America, Washington, DC, 1990).
- [83] Y. K. Jhee, J. C. Campbell, J. F. Ferguson, A. G. Dentai and W. S. Holden, "Avalanche buildup time of InP/InGaAsP/InGaAs APDs with separate absorption, grading, and multiplication regions", IEEE J. Quantum Electronics QE-24, 496 (1988).
- [84] H. C. Hsieh and Winslow Sargeant, "Avalanche buildup time of InP/InGaAsP/InGaAs APD at high gain", IEEE J. Quantum Electronics QE-25, 2027 (1989).
- [85] Masonori Ito, Takashi Mikawa and Osamu Wada, "Gain-bandwidth product of AlGaSb APDs analyzed by using equivalent multiplication region method", IEEE Electron Device Lett. EDL-10, 230 (1989).
- [86] M. Ito, T. Mikawa and O. Wada, "Optimum design of δ -doped InGaAs avalanche photodiode by using quasi-ionization rates", IEEE J. Lightwave Technol. LT-4, 1046-1050, 1991.
- [87] H. L. Armstrong, "A theory of voltage breakdown of cylindrical p-n junctions, with applications", IRE Trans. Electron Devices ED-4, 15 (1957).
- [88] D. P. Kennedy and R. R. O'Brien, "Avalanche breakdown characteristics of a diffused P-N junction", IRE Trans. Electron Devices ED-9, 478 (1962).
- [89] O. Leistiko, Jr. and A. S. Grove, "Breakdown voltage of planar silicon junctions", Solid State Electronics 9, 847 (1966).
- [90] V. Anantharam and K. N. Bhat, "Analytical solutions for the breakdown voltages of punched-through diodes having curved

- junction boundaries at the edges", IEEE Trans. Electron Devices ED-27, 939 (1980).
- [91] C. Basavanagoud and K. N. Bhat, "Effect of lateral curvature on the breakdown voltage of planar diodes", IEEE Electron Device Lett. EDL-6, 276 (1985).
- [92] David M. Braun, "Design of single layer antireflection coatings for InP/In_{0.53}Ga_{0.47}As/InP photodetectors in the 1200-1600 nm wavelength range", Applied Optics 27, 2006 (1988).
- [93] M. D. A. MacBean, N. P. Hewett, M. D. Learmouth and I. Reid, "Planar InP/InGaAs APD with integrated dielectric wavelength filter", Electronics Lett. 26, 804 (1990).
- [94] A. Papoulis, Probability, Random Variables, and Stochastic Processes, 2nd ed., McGraw-Hill, 1965.
- [95] G. Lucovsky, R. F. Schwarz and R. B. Emmons, "Transit-time considerations in p-i-n diodes", J. Appl. Phys. 35, 622, (1964).
- [96] T. Shiba, E. Ishimura, K. Takahashi, H. Namizaki and W. Susaki, "New approach to the frequency response analysis of an InGaAs APD", IEEE J. Lightwave Technology LT-6, 1502 (1988).
- [97] J. C. Campbell, B. C. Johnson, G. J. Qua and W. T. Tsang, "Frequency response of InP/InGaAsP/InGaAs APDs", IEEE J. Lightwave Technology LT-7, 778 (1989).
- [98] James. N. Hollenhorst, "Frequency response theory for multilayer photodiodes", IEEE J. Lightwave Technology LT-8, 531 (1990).
- [99] D. L. Crawford, R. L. Nagarajan and J. E. Bowers, "Comparison of bulk and quantum wire photodetectors", Appl. Phys. Lett. 58, 1629 (1991).
- [100] B. C. Roy and Nirmal B. Chakrabarti, "Pulse response of APDs", IEEE J. Lightwave Technology LT-10, 169 (1992).

- [101] James. N. Hollenhorst, "Frequency response theory for multilayer photodiodes", IEEE J. Lightwave Technology **LT-8**, 531 (1990).
- [102] B. C. Johnson, J. C. Campbell, A. G. Dentai, C. H. Joyner and G. J. Qua, "Interaction of hole trapping and transit effects in the temporal response of InP/InGaAs p-type insulator n-type photodiodes", J. Appl. Phys. **68**, 5343 (1990).
- [103] T. Miyoshi, H. Tsuchiya and M. Ogawa, "Quantum hole transport at the heterointerface of long wavelength APDs", IEEE J. Quantum Electronics **QE-28**, 25 (1992).
- [104] Gohalp Kahraman, Bahaa E. A. Saleh, Winslow L. Sargeant and Malvin C. Teich, "Time and frequency response of APDs with arbitrary structure", IEEE Trans. Electron Devices **ED-39**, 553 (1992).
- [105?] Y. Matsushima, K. Sakai and Y. Noda, "New type InGaAs/InP heterostructure APD with buffer layer", IEEE Electron Device Lett. **EDL-2**, 179 (1981).
- [105?] Y. Matsushima, K. Sakai and K. Utaka, "New type InGaAs/InP heterostructure APD with buffer layer", IEEE-Electron Device Lett. **EDL-2**, 179 (1981).
- [106] Y. Matsushima, S. Akiba, K. Sakai, Y. Kushiro, Y. Noda, K. Utaka, "High-speed-response InGaAs/InP heterostructure APD with InGaAsP buffer layers", Electronics Lett. **18**, 945 (1982).
- [107] S. R. Forrest, O. K. Kim and R. G. Smith, "Optical response time of In₅₃Ga₄₇As/InP APDs", Appl. Phys. Lett. **41**, 95 (1982).
- [108] Masahiro Kobayashi, Susumu Yamazaki and Takao Kaneda, "Planar InP/InGaAsP/InGaAs buried structure APD", Appl. Phys. Lett. **45**, 759 (1984).
- [109] J. C. Campbell, W. S. Holden, G. J. Qua and A. G. Dentai, "Frequency

- response of InP/InGaAsP/InGaAs APDs with separate absorption 'grading' and multiplication regions", IEEE J. Quantum Electronics **QE-21**, 1743 (1985).
- [110] J. C. Campbell, W. T. Tsang, G. J. Qua and B. C. Johnson, "High speed InP/InGaAsP/InGaAs APDs grown by chemical beam epitaxy", IEEE J. Quantum Electronics **QE-24**, 496 (1988).
- [111] H. Kuwatsuka, Y. Kito, T. Uchida and T. Mikawa, "High speed InGaAs/InP APDs with a compositionally graded quaternary layer", European Conference on Optical Communications Proceedings, 249, (1991).
- [112] W. Kuebart, H. Eisele, J. Scherb, J. Kimmerle, P. Wiedemann and W. Korber, "Continuous graded 10 Gb/s planar InGaAs/InP SAGM-APD grown by LP-MOVPE", European Conference on Optical Communications Proceedings, 253, (1991).
- [113] F. Capasso, H. M. Cox, A. L. Hutchinson, N. A. Olsson and S. G. Hummel, "Pseudo-quaternary GaInAsP semiconductors: a new $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ graded gap superlattice and its applications to APDs", Appl. Phys. Lett. **45**, 1193 (1984).
- [114] P. Hill, J. Schlafer, W. Powazinik, M. Urban, E. Eichen and R. Olshansky, "Measurement of hole velocity in n-type InGaAs", Appl. Phys. Lett. **50**, 1260 (1987).
- [115] Kevin Brennan, "Theory of the steady-state hole drift velocity in InGaAs", Appl. Phys. Lett. **51**, 995 (1987).
- [116] T. H. Windhorn, L. W. Cook and G. E. Stillman, "The electron velocity-field characteristic for n-InGaAs at 300K", IEEE Electron Device Lett. **EDL-3**, 18 (1982).
- [117] Albert Chin and T. Y. Chang, "Multilayer reflectors by molecular-beam

- epitaxy for resonance enhanced absorption in thin high-speed detectors", J. Vac. Sci. Technol. B **8**, 339 (1990).
- [118] Albert Chin and T. Y. Chang, "Enhancement of quantum efficiency in thin photodiodes through absorptive resonance", IEEE J. Lightwave Technology **LT-9**, 321 (1991).
 - [119] A. Okamoto, H. Sumuhawa, H. Terao and H. Watanabe, J. Cryst. Growth **70**, 140 (1984).
 - [120] D. G. Knight, C. J. Miner and A. Majeed, Mater. Res. Symp. Proc. **90**, 249 (1987).
 - [121] N. Puetz, G. Hillier and A. J. SpringThorpe, "The inverted horizontal reactor: growth of uniform InP and GaInAs by LP MOCVD," Journal of Electronic Materials, **17**, 381 (1988).
 - [122] Y. Liu, S. R. Forrest, V. S. Ban, K. M. Woodruff, J. Colosi, G. C. Erickson, M. J. Lange and G. H. Olsen, "Simple, very low dark current, planar long wavelength APD", Appl. Phys. Lett. **53**, 1311 (1988).
 - [123] Y. Liu, S. R. Forrest, J. Hladky, M. J. Lange, G. H. Olsen and D. E. Ackley, "Simple, very low dark current, planar long wavelength APD", Appl. Phys. Lett. **53**, 1311 (1988).
 - [124] D. E. Ackley, J. Hladky, M. J. Lange, S. Mason, G. C. Erickson, G. H. Olsen, V. S. Ban, Y. Liu and S. R. Forrest "In_{0.53}Ga_{0.47}As/InP floating guard ring APDs by double diffusion", IEEE Photon. Technol. Lett. **PTL-2**, 571 (1990).
 - [125] K. Taguchi, T. Torikai, Y. Sugimoto, K. Makita, H. Ishihara, S. Fujita and K. Minemura, "Planar InP/InGaAs APD with preferential lateral extended guard ring", IEEE Electron Device Lett. **EDL-7**, 257 (1986).
 - [126] K. Taguchi, T. Torikai, Y. Sugimoto, K. Makita, and H. Ishihara "Planar-structure InP/InGaAsP/InGaAs APDs with preferential lateral

References

- extended guard ring for 1.0-1.6 μm wavelength optical communication use", IEEE J. Lightwave Technology **LT-6**, 1643 (1988).
- [127] V. P. Sundersingh and A. A. Ghatol, "Concentration profiling for high voltage p^+-n-n^+ diodes", Int. J. Electronics **54**, 127 (1983).
- [128] C. P. Skrimshire, J. R. Farr, D. F. Sloan, M. J. Robertson, P. A. Putland, J. C. D. Stokoe and R. R. Sutherland, "Reliability of mesa and planar InGaAs PIN photodiodes", IEE Proceedings **137-J**, 74 (1990).
- [129] M. Gallant and A. Zemei, "Long minority hole diffusion length and evidence for bulk radiative recombination limited lifetime in InP/InGaAs/InP double heterostructures", Appl. Phys. Lett. **52**, 1686 (1988).
- [130] S. Komiya, Akio Yamaguchi and Itsuo Umebu, "Characterization of radiative efficiency in double heterostructures of InGaAsP/InP by photoluminescence intensity analysis", Solid State Electronics **29**, 235 (1986).
- [131] R. Muralidharan and K. Ploog, "Determination of diffusion length of minority carriers in epitaxial layers by photocurrent spectroscopy" Appl. Phys. A **49**, 527 (1989).
- [132] U. Koren, B. I. Miller, T. L. Koch, G. Eisenstein, R. S. Tucker, I. Bar-Joseph and D. S. Chemia, "Low-loss InGaAs/InP MQW optical electroabsorption waveguide modulator", Appl. Phys. Lett. **51**, 1132 (1987).
- [133] C. Blaauw, F. R. Shepherd, C. J. Miner and A. J. SpringThorpe, "Silicon incorporation in InP during LP-MOCVD using disilane", Journal of Electronic Materials **19**, 1 (1990).
- [134] K. Brennan, T. Wang and K. Hess, "Theory of electron impact ionization including a potential step: application to GaAs-AlGaAs",

- IEEE Electron Device Lett. EDL-2, 179 (1981).
- [135] Kevin F. Brennan, "Optimization and modeling of APD structures: application to a new class of superlattice photodetectors, the p-i-n, p-n homojunction, and p-n heterojunction APDs", IEEE Trans. Electron Devices ED-34, 1658 (1987).
 - [136] Raad Sami Fyath and John J. O'Reilly, "Effect on the performance of staircase APDs of electron impact ionization within the graded-gap region", IEEE Trans. Electron Devices ED-35, 1357 (1988).
 - [137] Y. Miyamoto, K. Hagimoto and T. Kagawa, "Gigabit high sensitivity using InGaAs/InAlAs superlattice APD", Electronics Lett. 26, 491 (1990).
 - [138] Yutaka Miyamoto, Kazuo Hagimoto and Toshiaki Kagawa, "A 10 Gb/s high sensitivity optical receiver using an InGaAs-InAlAs superlattice APD at 1.3 μm /1.5 μm ", IEEE Photon. Technol. Lett. PTL-3, 372 (1991).
 - [139] Toshiaki Kagawa, Yuichi Kawamura, Hiromitsu Asai and Mitsuru Naganuma, "InGaAs/InAlAs superlattice APD with a separated photoabsorption layer", Appl. Phys. Lett. 57, 1895 (1990).
 - [140] H. Nakamura, S. Hanatani, S. Tanaka, T. Notsu, H. Nakano and K. Ishida, "An InGaAs/InAlAs superlattice APD with thin well width for 10 Gb/s optical transmission systems", European Conference on Optical Communications Proceedings, 261, (1991).
 - [141] Toshiaki Kagawa, Yuichi Kawamura and Hidetoshi Iwamura, "InGaAsP-InAlAs superlattice APD", IEEE Quantum Electronics QE-28, 1419 (1992).
 - [142] G. Duggan, "Measurements of band discontinuities using optical techniques", in F. Capasso and G. Margaritondo, editors, Heterojunction band discontinuities: physics and device applications, North-Holland, 1987.

- [143] S. R. Forrest, Measurement of energy band offsets using capacitance and current measurement techniques", in F. Capasso and G. Margaritondo, editors, Heterojunction band discontinuities: physics and device applications, North-Holland, 1987.
- [144] R. W. Grant, E. A. Krout, J. R. Waldrop and S. P. Kowalczyk, "Interface contributions to heterojunction band discontinuities: x-ray photoemission spectroscopy investigations", in F. Capasso and G. Margaritondo, editors, Heterojunction band discontinuities: physics and device applications, North-Holland, 1987.
- [145] G. Margaritondo, "Controlling heterojunction band discontinuities: a surface physicist's approach", Solid State Electronics **29**, pp. 123-132, 1986.
- [146] G. Margritondo and P. Perfetti, "The problem of heterojunction band discontinuities", in F. Capasso and G. Margaritondo, editors, Heterojunction band discontinuities: physics and device applications, North-Holland, 1987.
- [147] A. G. Milnes, "Heterojunctions: some knowns and unknowns", Solid State Electronics **30**, pp. 1099-1105, 1987.
- [148] Fritz L. Schuermeyer, Paul Cook, Edgar Martinez and John Tantillo, "Band-edge alignment in heterostructures", Appl. Phys. Lett. **55**, 1877 (1989).
- [149] Leye Aina, Mike Mattingly and Lisa Stecker, "Photoluminescence from AlInAs/InP quantum wells grown by organometallic phase epitaxy", Appl. Phys. Lett. **53**, 1620 (1988).
- [150] J. R. Waldrop, E. A. Kraut, C. W. Farley and R. W. Grant, "Measurement of AlAs/InP and InP/In_{0.52}Al_{0.48}As heterojunction band offsets by x-ray photoemission spectroscopy", J. Vac. Sci. Technol. B **8**,

768 (1990). $\Delta E_v = 0.16$ eV

- [151] J. R. Waldrop, E. A. Kraut, C. W. Farley and R. W. Grant, "Measurement of InP/In_{0.53}Ga_{0.47}As and In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As heterojunction band offsets by x-ray photoemission spectroscopy", Appl. Phys. Lett. **69**, 372 (1991).
- [152] Mark S. Hybertsen, "Band offset transitivity at the InGaAs/InAlAs/In_{0.53}Ga_{0.47}As heterointerfaces", Appl. Phys. Lett. **58**, 1759 (1991).
- [153] P. Z. Lee, C. L. Lin, J. C. Ho, L. G. Meiners and H. H. Wieder, "Conduction-band discontinuities of In_xAl_{1-x}As/In_{0.53}Ga_{0.47}As n-isotype heterojunctions", J. Appl. Phys. **67**, pp. 4377-4379, 1990.
- [154] L. Tarof, "Proposal for new APD", BNR internal memo, January, 1991.
- [155] D. C. Hutchings, "Transfer matrix approach to the analysis of an arbitrary quantum well structure in an electric field", Appl. Phys. Lett. **55**, 1082 (1989).
- [156] D. P. Dave, "Numerical technique to calculate eigenenergies and eigenstates of quantum wells with arbitrary potential profile", Electronics Lett. **27**, 1735 (1991).
- [157] C. Priester and M. Lannoo, "Analog of the $k \cdot p$ theory for a localized-orbital description of the band structure of zinc-blende-structure semiconductors", Phys. Rev. B **44**, 10559 (1991). (includes strain)
- [158] G. T. Einevoll and P. C. Hemmer, "Superlattice minibands—explicit formulae for band gaps and effective masses", Semicond. Sci. Technol. **6**, 590 (1991).
- [159] G. Hendorfer and J. Scheider, "g-factor and effective mass anisotropies in pseudomorphic strained layers", Semicond. Sci. Technol. **6**, 595 (1991).
- [160] J. E. Bowers and C. A. Burrus, "High-speed zero-bias waveguide

- photodetectors", Electronics Lett. 22 905 (1986).
- [161] J. Conradi, Course notes on optical fiber communications and technology, Chapter 4 (1979).
- [162] L. E. Tarof, "Planar InP/InGaAs APDs with a gain-bandwidth product exceeding 100 GHz", Optical Fiber Communications Conference, 1991 Technical Digest Series, Vol. 4, 187 (Optical Society of America, Washington, DC, 1991).
- [163] L. E. Tarof, D. G. Knight, F. R. Shepherd and N. Puetz, "Avalanche photodiodes and methods for their manufacture", Canadian Patent #1,298,640, assigned to Northern Telecom Ltd.
- [164] L. E. Tarof, J. Yu, R. Bruce, D. G. Knight, T. Baird and B. Oosterbrink, "High frequency performance of separate absorption, grading, charge and multiplication InP/InGaAs avalanche photodiodes", IEEE Photon. Technol. Lett. PTL-5, 672-674 (1993).
- [165] H. Ando, N. Susa and H. Kanbe, "Carrier density profiles in Zn- and Cd- diffused InP", Jpn. J. Appl. Phys 20, L197 (1981).
- [166] M. Yamada, P. K. Tien, R. J. Martin, R. E. Nahory and A. A. Ballman, "Double Zn diffusion fronts in InP – theory and experiment" Appl. Phys. Lett. 43, 594 (1983).
- [167] B. Tuck, Introduction to diffusion in semiconductors, Peregrinus, Stevenage, England (1974).
- [168] Y. Iimura, H. Takasugi and M. Kawabe, "Reduction of a highly-resistive layer at an interrupted interface of GaAs grown by MBE", Jpn. J. Appl. Phys 25, 95 (1986).
- [169] M. L. Gray and C. W. Ebert, "The influence of substrate-epitaxial layer chemical impurities on heterostructure electrical characteristics", J. Appl. Phys 71, 3294 (1992).

- [170] J. C. Bean, Chapter 1 in High-Speed Semiconductor devices, ed. S. M. Sze, Wiley (1990).
- [171] N. A. Olsson, "Lightwave systems with optical amplifiers", IEEE J. Lightwave Technol. LT-5, 1071-1082, 1989.
- [172] R. S. Fyath and J. J. O'Reilly, "Performance of optically preamplified direct-detection receivers in the presence of laser chirp, Part 1", IEE Proceedings 136, Part J, 249-255, 1989.
- [173] C. L. F. Ma, M. J. Deen and L. E. Tarof, "D.C. and noise characteristics of InP-based APDs for optical communication applications", 1993 Canadian conference on electrical and computer engineering, Vancouver, Canada.

END

1 | 8-07-94

FIN