UNIVERSITY OF ADELAIDE LIBRARY
17-10-96

GaAs MESFET Photodetectors for Imaging Arrays

by

Derek Abbott, B.Sc. (Hons.)

Thesis submitted for the degree of

Doctor of Philosophy

The University of Adelaide
Adelaide, South Australia

Department of Electrical and Electronic Engineering
Faculty of Engineering

December 1995

Contents

Abstract . . . viii

Declaration . . . x

Preface . . . xi

Acknowledgement . . . xiii

Resume . . . xviii

Publications . . . xix

List of Figures . . . xxi

List of Tables . . . xxvii

Glossary . . . xxix

1 Introduction **1**

1.1 Aims and Significance . . . 2

1.1.1 Long Term Aim . . . 2

1.1.2 Immediate Aim of this Work . . . 2

1.1.3 State of the Art and Present Knowledge in the Area of GaAs Arrays 3

1.1.4 Significance of GaAs . . . 5

1.2 Overview of XY Array Strategy . . . 6

1.2.1 Photocollection Modes . . . 6

1.2.2 Pixel Array . . . 7

1.3 Research Philosophy . . . 8

1.4 Field of this Research . . . 10

1.5 Achievements from this Research 11
1.6 GaAs Background and Future Vision 11
1.7 Summary and Chapter Overview 15

2 GaAs versus Silicon **23**
2.1 Introduction 24
2.2 History 24
2.2.1 Gallium 24
2.2.2 Arsenic 25
2.2.3 Gallium Arsenide 25
2.3 Competing Technologies 26
2.3.1 Diamond 26
2.3.2 Silicon Carbide (SiC) 29
2.3.3 Gallium Nitride (GaN) 30
2.3.4 Polymer Electronics 31
2.3.5 Porous Silicon 31
2.3.6 Silicon-on-Insulator (SOI) 31
2.3.7 Silicon-Germanium (SiGe) Heterostructures 32
2.4 The Semi-insulating (SI) GaAs Substrate 35
2.5 GaAs Characteristics 40
2.5.1 Bandgap 40
2.5.2 Speed 41
2.5.3 Radiation Hardness 42
2.5.4 Dark Current 42
2.5.5 Absorption Coefficients 43
2.5.6 Atmospheric Bands 48
2.5.7 Quantum Efficiency Formulae 48
2.5.8 Quantum Efficiency & Responsivity 50

2.5.9 Spatial Degradation by Diffusion 58

2.6 Summary 60

3 MESFET Photodetection 67

3.1 GaAs MESFET Optical Model 68

3.1.1 Background 68

3.1.2 Photoresponse Model 69

3.2 GaAs MESFET Photoresponse Measurements 73

3.2.1 Monochromator Measurements 73

3.2.2 Laser Flood Measurements 78

3.2.3 Focussed Laser Measurements 81

3.2.3.1 Two Dimensional Laser Scan Set Up 86

3.2.3.2 Two Dimensional Scan Results 86

3.2.3.3 Discussion 90

3.2.3.4 Experimental Refinement 94

3.3 Summary 94

4 Noise 103

4.1 Introduction 104

4.2 A Brief Early History of Noise 104

4.3 A New Derivation of the Thermal Noise Formula 109

4.3.1 Overview of Methods in the Literature 109

4.3.2 The Lumped Model 110

4.3.3 The Windowed Fourier Transform (WFT) 110

4.3.4 The Thermal Noise Formula 114

4.4 The Classical Energy Catastrophe 118

4.5 The Quantum Energy Catastrophe 118

4.6 The Steak Grilling Debate 119

4.7 Quantum Cut-Off Experimental Status 120
4.8 MacDonald's Objection 122
4.9 The Case of Limiting R & C 122
4.10 kTC Noise 125
4.11 Power in a Matched Load 126
4.12 Distributed RC 127
4.13 The Capacitor Paradox 129
4.14 Forward's Photoamplifier 129
4.15 Measurement of Noise 130
4.16 Thermal Noise in MOSFETs 134
4.17 Flicker Noise in MOSFETs 134
4.18 Thermal Noise in MESFETs 135
4.19 Flicker Noise in MESFETs 136
4.20 Shot Noise in MESFETs 136
4.21 Noise Analysis of Imager Output Circuit 137
4.21.1 Silicon output circuit 139
4.21.2 GaAs output circuit 140
4.22 Summary 142

5 Imagers **153**

5.1 Introduction 154
5.2 Pixel Design 154
5.2.1 Illumination Modes 154
5.2.2 Photocollection Modes 154
5.2.3 Channel to Substrate Mode 156
5.2.4 Gate to Channel Mode 158
5.2.5 Pixel Active Area 159
5.2.6 Pixel Design 159

5.3 Array Organisation 162
5.4 Address Circuitry 166
5.4.1 Case 1: The Capacitively Bootstrapped Shift Register 169
5.4.2 Case 2: Conventional D Flip-Flop Shift Register 171
5.4.3 Case 3: Edge Triggered D Flip-Flop Shift Register 174
5.4.4 Register Comparison 174
5.5 Column-to-video line multiplexer 176
5.6 Summary 177

6 Design Methodology **181**
6.1 Introduction 182
6.1.1 Technology 183
6.1.2 Parasitic Oscillations 184
6.1.3 Layout Approaches 186
6.1.3.1 Symbolic mask level layout 189
6.1.3.2 Practical constraints 190
6.1.3.3 CAD tools 191
6.1.3.4 Modelling of gallium arsenide circuits 192
6.1.3.5 Interface to fabrication 193
6.1.4 Performance Study 194
6.1.4.1 Adders 194
6.1.4.2 Multipliers 198
6.2 Summary 201

7 Visions & Applications **207**
7.1 Introduction 208
7.2 Case 1: Aerospace paradigm 208
7.2.1 Image Processing 208

7.2.1.1 Image registration 209
7.2.1.2 Image enhancement 209
7.2.2 Polynomial Evaluator 221
7.2.2.1 Behavioural Model 222
7.2.2.2 Algorithmic Solution 223
7.2.2.3 Functional Specification 224
7.2.2.4 Structural Description (Floor Plan) 225
7.2.2.5 Symbolic Physical Description 225
7.2.2.6 Physical Layout 228
7.3 Case 2: The Mobile Multimedia Paradigm 235
7.3.1 Complementary Gallium Arsenide 236
7.3.2 Objectives 237
7.3.3 Current State-of-the-Art 237
7.3.4 Sub-systems 237
7.3.4.1 High-density high-speed RAM 239
7.3.4.2 Image compression 239
7.3.4.3 Modulo Multiplier for Data Encryption/Decryption 240
7.3.4.4 Solid-state imager 240
7.3.4.5 RF/digital interface 241
7.4 Case 3: The Motion Detector Paradigm 242
7.4.1 Introduction 243
7.4.2 Comparison with other approaches 247
7.4.3 Motion Detector Chip 248
7.4.4 Range Estimation 253
7.4.4.1 Gallium Arsenide Implementation 255
7.5 Case 4: The Spatial Light Modulator Paradigm 257
7.5.1 Significance 259

7.5.1.1 FLC/Si Technology 259

7.5.1.2 GaAs Technology 260

7.5.1.3 Devices 261

7.5.1.4 Optical Neural Network 263

7.6 Techniques and Methods 264

7.6.1 FLC/Si Devices 264

7.6.2 LED/GaAs and FLC/GaAs Devices 266

7.6.3 Smart Pixel Circuitry 266

7.6.4 SLM and ONN Testing 266

7.7 Summary 267

8 Conclusions and Recommendations **277**

A Finite Difference Method **283**

A.1 FDM Voltage Potential Solution 284

A.2 FDM Electric Field Solution 285

B Quantum Efficiency **289**

B.1 Quantum Efficiency in a Semi-Infinite Slab 290

B.2 Quantum Efficiency in a Finite Slab 291

C Heat Equations **295**

D Complex Integration **299**

D.1 General Complex Integral for Capacitor Problem 300

D.2 Noise Analysis of 2-Stage RC Ladder 301

E Polynomial Equations **303**

ABSTRACT

Is it possible for an XY addressable imaging array to be realised in standard or near-standard gallium arsenide digital IC technology? This was the original question at the heart of this work. The motivation was the ability to embed such an imager with digital GaAs processing on one IC chip. Integration of an imager with a processor offers the promise of smart sensors that exploit the excellent power-delay product of GaAs, useful for mobile applications – where, for instance, image compression may need to take place in real-time on the same chip. Compact mobile systems of the future can also exploit the mixed digital/RF capability of GaAs.

Measurements of the photoresponse of GaAs MESFETs are used to understand the photodetection mechanisms to aid future study and design. A significant new internal gain effect, in planar MESFETs has been discovered which we call the "photovoltaic self-biasing edge-effect." The edge-effect can be exploited to attain up to a factor of ten improvement in photosensitivity. In order to project further the advantages of GaAs and fully understand the challenges, the key differences in device physics between silicon and GaAs are critically examined.

Theoretical quantum efficiency calculations are carried out and compared against the photoresponse measurements. A new result is obtained showing that the quantum efficiency in a finite slab (which models an epitaxial layer) is equal to the quantum efficiency in a semi-infinite slab (which models a bulk substrate) plus a simple *additive* second order term.

A new optical technique for extracting semi-insulating (SI) substrate parameters, including the carrier diffusion length, is demonstrated showing diffusion lengths of an order of magnitude *lower* than in silicon. This is significant for improved spatial resolution in high definition television (HDTV) imagers. From the perspective of silicon device physics this is a surprising result, but is explained in terms of the deep levels in the GaAs SI substrate.

Various new pixel schemes and layouts are presented for the realisation of a GaAs XY addressable imaging array. A number of shift register designs for readout of the imager are

analysed, simulated and compared for the first time. Detailed calculations for optimising the noise performance of imager readout circuitry are developed and this is based on a detailed review of a number of conumdrums and surprises in noise theory.

Future visions and novel applications of the imager are carefully considered. A review of various image enhancement, image compression and image registration schemes are presented for possible future integration with the imager. Potential applications in mobile multimedia communications, artificial insect vision and spatial light modulators are reviewed.

The edge-effect discovery is shown to be particularly suited to producing increased photosensitivity in a smart sensor array based on insect vision.

In conclusion, a number of important new principles have been uncovered by this research leading to the potential realisation of a simple GaAs XY addressable imager that can be integrated with digital GaAs processing circuitry. This enables the 'smart sensor' paradigm to shift from silicon and reap the advantages of GaAs.

DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of the author's knowledge and belief, contains no material previously published or written by another person, except where due acknowledgement has been made in the text.

The author gives consent to this copy, when deposited in the University Library, being available for loan or photocopying.

Derek Abbott.

PREFACE

This work has an interesting history behind it giving rise to a thesis with a number of significant perspectives that are highlighted as follows.

Firstly, like many engineers today, the author originally came from a firmly entrenched silicon background, via working many years in industry, and thus originally approached the field of gallium arsenide with extreme skepticism. A change in belief took place, as concrete experimental results from this research gradually verified that GaAs is an excellent medium for imaging systems.

The initial research questions and challenges concerning a GaAs implementation are thus tackled and presented from the point of view of a 'silicon mind.' Interesting discoveries and surprises concerning GaAs device physics are viewed from a silicon perspective.

This work faced the interesting challenge that technology was not static over the duration of this research. The status quo was constantly changing throughout this research period. Due to the many unknowns in GaAs device physics, unknowns in substrate behaviour and immaturity of the technology, only a few years ago, the idea of a GaAs XY array imager was quite a courageous one. The known 'facts' at the time dictated that such an imager would not outperform silicon. However, the adopted attitude has been "what do we need to change to *make* it work?" The next question was "*why* should we expend the effort to do this?" The initial motivation was a vision that a simple GaAs XY array could be integrated with high speed image processing circuitry on the one chip, making making an ideal solution for mobile multimedia applications. This in itself was also a rather bold vision as *mobile* multimedia, in the telecommunications community then, was to our knowledge not spoken of and easy to contest based on the near-term foreseeable technology then.

Today, the understanding of the GaAs substrate has greatly improved, the maturity of the technology can support large VLSI designs, the research described in this work shows 'surprisingly' (from the silicon mindset) good prospects for GaAs XY arrays based both on theoretical analysis and actual experimental results. Furthermore, telecommunications

technology has matured to the point where concepts such as BISDN *microcell* networks are now within foreseeable reach and therefore the eventual realisation of a communications infrastructure required to support mobile multimedia is no longer a dream but now inevitable.

The philosophy employed in writing this thesis has been to communicate in a clear and non-obfuscating manner. As part of creating a clear presentation some special attention has been made to the history, vision and motivation that the work is built on. For clear understanding the "why" is as important as the "how."

In any pioneering research program, the meandering path is seldom straight. More often than not other discoveries are made along tangent paths, that can turn out to be more valuable than the original goals. This is the nature of true research. Therefore a number of these 'side alleys' have been carefully woven into the main fabric of this thesis – with hopefully minimal disturbance to the logical flow. The famous engineer Thévenin was surprised when he was told, just before his death, that his theorem had been accepted worldwide, some 43 years after he posed it in 1883. For this reason nothing has been prejudged.

Alexander Graham Bell once said **"Leave the beaten track occasionally and dive into the woods. You will be certain to find something you have never seen before."** This has been one of the principles underlying this work.

ACKNOWLEDGEMENTS

I warmly thank my parents for their vision and sacrifice in providing an education for me. Also for the strong work ethic, they taught me by their example. They did not have the money to buy toys, so as a baby I was given old electrical plugs to play with. This early contact with electrical devices must have worked on my subconscious to bring me to this point today. Electricity was actually my worst physics subject at high school, so the plugs must be thanked for at least having a delayed effect on shaping my future.

I thank Dr. Percy Baxter and Dr. Denis Scotter for employing me at the GEC Hirst Research Centre, UK, which set me on the path of interest in electronic devices. They also awarded me a Bursary to help pay my way for obtaining my first degree. I would not be here today without their foresight.

At Hirst, David Burt was my mentor and generously guided me through many semiconductor principles. Ray Bell taught me the basics of electronics, how to use instruments and even how to solder. I thank them both for their patience and generosity. An important milestone, was at about the age of 19 when I asked the 'silly' question: "why do electric field lines stop existing outside a depletion region?" Stephen Lovatt provided the answer: "because field lines begin and end on *fixed* charges." The realisation of this simple fact, helped my understanding of device physics to suddenly blossom. I thank Stephen for providing the vital catalyst.

My mentors at Loughborough University, during undergraduate days, were Dr. Mike Brown and Nick Phillips. Mike Brown taught me how to look at solid state physics in simple ways and Nick Phillips sparked my curiosity in unsolved noise theory. These seeds, that were then implanted in my mind, have now grown and found expression within this thesis. Without their inspiration, this work would not be the same.

Moving to the present, I heartily thank my supervisors Prof. Kamran Eshraghian and Associate Prof. Bruce Davis for their guidance, patience and belief in me. Prof. Eshraghian has taught me so much through his tireless energy, loyalty and vision. He has been the rock upon which this work has been built, in terms of raising funding, springboarding

ideas and never prejudging a 'wild' idea. His belief in me and in the vision of the project, carried us through the early days when I could not see that far ahead. His boundless enthusiasm, deep belief in people & teamwork, willingness to encourage risks & taking them, devotion to long term thinking, ability to engineer & redefine reality, commitment to high expectations along with readiness, virtue and vision have been marks of his true leadership. Prof. Davis also has generously given of himself and has guided me through many mathematical concepts and has been an inspirational springboard for discussion of noise concepts.

There are many people that I must thank of their input into this work. Due to the obvious breadth of this thesis in contrast to my limited time & capabilities something has had to 'give.' Consequently, I kept myself totally 'software dumb' throughout this work. Michael Liebelt, Andrew Beaumont-Smith and Ali Moini have been extremely patient and generous by rescuing me from many software problems. The scan software for the Laser Prober turned out to be particularly tricky and although I originally wrote most of the code, Andrew must get the credit for making it work properly.

Dr. Don Griffin and Michael Needle must be thanked for the early work they did in getting the Laser Prober made up. A special thanks goes to Dr. Tony Pugatschew of the University of South Australia for arranging the laser unit to work through the microscope optics of the prober. Stephen Guest, Ian Linke and all the guys from the workshop did a marvellous job of constructing the antivibration table, for the prober, out of a granite gravestone, steel tubing and industrial vibration isolators. Thanks also to Geoff Pook for making up numerous PCB test heads for all my samples.

The Physics Department also generously lent me their monochrometer equipment for a year.

A special thanks is due to Dr. Eithne McCabe, Oxford University (now at Trinity, Dublin), for her support and encouragement with some of the early laser based measurements. She was the first of a long line of international visitors to witness the 'photovoltaic self-biasing edge-effect,' uncovered by this work.

I warmly thank Dr. Jerry Hallmark and Dr. Jon Abrokwah, of Motorola, for recently hosting me in Arizona and answering all my questions about their CGaAs HIGFETs. I would also like to thank Dr. Tim Coe, of Vitesse Semiconductor Corporation, for hosting me in Camarillo, California and also answering my questions.

Prof. Antonio Nuñez, Prof. Roberto Sarmiento and Dr. Jose Lopez have hosted me many times at the University of Las Palmas, Spain – their hospitality has been selfless and lavish. They have been a delight to work and collaborate with. Much of the credit for

Chapter 6 must be given to their team. I cannot praise them too highly.

Dr. Bruce Odekirk of Triquint Semiconductor generously spent a lot of his time unlocking some of the mysteries of the semi-insulating substrate for me. Much credit is owed to him for this very tricky area. Discussions with Prof. Doug S. McGregor, University of Michigan, Ann Arbor, and Prof. Robert B. Darling, University of Washington, Seattle, on this vexing subject have also been tremendously influential and helpful.

Dr. Bobby Yu of the University of Reading, UK, receives my deepest thanks for flying out to meet me both in Adelaide and Paris and has provided much of the inspiration for the Spatial Light Modulator (SLM) paradigm in Chapter 7.

Our Head of Department, Dr. Tony Parker must be thanked for his marvellous support, finding lab space and releasing me from teaching duties on many occasions for me to engage in international conferences and collaboration.

Dr. Neil Burgess has been a great springboard for discussions on the *mobile multimedia* concept and he has generously pointed me to some useful references on encryption and compression for which his inspiration must be given credit.

The assistance of Sam Mosel with the polynomial evaluator and Mrs. L. Gu with the background work on image processing & use of KHOROS is gratefully acknowledged. Much credit is due to their tireless help.

The following figures were put together by the following people and appeared in joint publications with myself: (1) Figs. 1.2 and 1.3, Andrew Beaumont-Smith, (2) Fig. 3.3, Song Cui (3) Figs. 7.27, 7.30 and 7.31, Dr. André Yakovleff and (4) Figs. 7.34, 7.35 and 7.36, Ali Moini.

The photographs in Figs. 4.1 and 4.2 are by courtesy of AT&T Archives. Dr. Bob Anholt, of Gateway Modeling, Minnesota, kindly ran G-PISCES-2B for me, to produce Fig. 3.16. Clive Fuller kindly took the photograph in Fig. 3.12.

The section on insect vision would not be possible without the ingenuity and energy of the whole 'Bugeye' team: Ali Moini, Dr. André Yakovleff, Andrew Blanksby, Richard Beare, Ook Kim, Gyudong Kim, Thong Nguyen, Dr. Abdesselam Bouzerdoum, Prof. Adrian Horridge, Prof. Bob Bogner and Prof. Wonchan Kim.

Thanks are due to Dr. Murray W. Hamilton, University of Adelaide; Prof. B. L. Hu, University of Maryland; Prof. Laszlo B. Kiss, JATE, Hungary; Dr. Roger H. Koch, IBM Watson Research Center; Prof. Matt McIrvin, Harvard University; Prof. Tony Siegman, Stanford University for helpful discussions on ZPF.

A special thanks is due to Prof. Paul C. W. Davies, University of Adelaide, and Prof. Walter T. Grandy, Jr., University of Wyoming, who as leading experts in the field, kindly checked through my section on ZPF.

Also thanks is due to Dr. Andrew Mehnert, of Queensland University of Technology, for checking through this thesis and filtering out my errors.

Assistance with historical references from Hugh H. Wigg, Adelaide; Claudia Bruning, Fachhochschule Hamburg Fachbereich Bibliothek und Information; Dr. Jan W. Nienhuys, Eindhoven University of Technology; Prof. Gary A. Zimmerman, Antioch University, Seattle; Prof. Arthur van Rheenen, University of Minnesota, and Prof. Chapin C. Cutler, Stanford University is gratefully acknowledged.

Thanks is due to Ninette Ellis, Walter Foid and Chris Holshuysen for translation of papers from Dutch and German.

A special thanks to my wife, Stevie, for translating a number of papers from French. I warmly thank my wife, children and family for their support and coping with my long hours day & night away in the lab or grafted to a computer terminal. The true credit for this work belongs to them – and this work is dedicated to them.

A final thanks to my friends over the years for their support and encouragement. I have been particularly influenced by Peter & Helen McKenna, Bill & Brioney Smart, Bill & Jane Francis and especially Lyndon & Jenny Dolling who have taught me to always challenge the prevailing paradigm and to think without limits.

The initial funding for this work came from the Sir Ross & Sir Keith Smith [1] Foundation and credit is due to Prof. Sam Luxton who encouraged us to apply for this funding. Further funding then came from the Australian Research Council (ARC) and the Australian Department of Education Employment & Training (DEET) Targeted Institutional Links (TIL) program (in conjunction with Seoul National University, Korea) and the Australian Asean Educational Cooperative Program (AAECP).

[1] Sir Ross & Sir Keith Smith were famous Australian aviators of the 1930s and the aerospace application overview in Chapter 7 is in their honour. An anecdote about Sir Ross & Sir Keith Smith can be found in Dale Carnegie's classic book on interpersonal relations.

Typesetting: This thesis has been generated using LaTeXsoftware and graphical output is displayed using MATLAB. Processed images are via KHOROS. Numerous different drawing packages were used as departmental standards kept changing – regularising the postscript format generated from these different sources was a nightmare and was achieved with difficulty.

Spelling: The English spelling convention has been adopted, except for the words "programme" and "analogue" have been replaced by their American equivalents "program" and "analog." Where more than one spelling variant is allowed in English such as "biassing" or "biasing" and "infra-red" or "infrared," the option with the *fewest* characters has been adopted.

Quotations: At the beginning of each chapter, various quotes are displayed. The quotes are a mixture of pro and con. Contrary quotes are provided to reflect on the paradigm-shift in thinking. Some are tongue-in-cheek.

Transistors: Unless otherwise stated all, GaAs MESFET analysis, experimental results and discussion apply only to n-channel and *non-self-aligned* transistors on an SI GaAs:EL2 substrate with no p-buffer layer.

Mathematics: All mathematical results in this thesis have been carefully double checked using the MAPLE symbolic math editor. In some cases where this was not readily possible, MAPLE was used to numerically solve an equation and this was cross-checked for a number of specific examples with the by-hand analytical solution.

RESUME

Derek Abbott obtained his B.Sc. (HONS) in physics at Loughborough University of Technology, UK, and completed his PhD thesis in Electrical & Electronic Engineering at the University of Adelaide. He worked for a number of years at the GEC Hirst Research Centre, London, on infrared and visible image sensors – requiring discipline in VLSI design, optoelectronics, device physics, semiconductor noise, fabrication and testing. He has worked with both novel and standard technologies including nMOS, CMOS, CCD, SOS, GaAs and vacuum microelectronics (ie. field emission devices). Since his arrival in Australia, he worked for Austek Microsystems, Technology Park, and has been associated with the University of Adelaide since 1987 and is a Lecturer within the Department of Electrical Electronic Engineering. He has been an invited speaker at a number of international institutions and has appeared on national radio and television. He is a founder member of the Centre for Gallium Arsenide VLSI Technology, Adelaide, instituted in 1987. He has been consultant to various industry and defence organisations. His current research interests are in the areas of VLSI, GaAs, photodetectors, biologically inspired sensors, device physics and noise. His most recent contribution is the discovery of a new photovoltaic self-biasing internal gain effect within planar GaAs MESFETs. He is on international committees of both the SPIE, USA and UPoN, Hungary. He was an invited keynote speaker at the SPIE conference in Philadephia, 1995, and is an invited speaker for Unsolved Problems in Noise (UPoN '96) in Hungary. He has been invited to speak at a number of international institutions including: EPFL, Lausanne; University of Geneva; MIT, Cambridge; Los Alamos National Research Labs (LANL), New Mexico; Motorola, Tempe, Arizona; Seoul National University and the University of Las Palmas, Spain.

PUBLICATIONS

- "Gallium Arsenide MESFET Imager," *World Intellectual Property Organisation, International Bureau*, **Patent WO 93/07643**, 15 April 1993. (with Eshraghian.)

- "Photovoltaic Gate Biasing Edge Effect in GaAs MESFETs," *Electronics Letters*, IEE, **vol. 27**, no. 21, Oct. 1991, pp. 1900-1991, (with Eshraghian, Cui & McCabe).

- "Derivation of the thermal noise formula using window limited Fourier transforms," *IEEE Trans Ed.,* **vol. 39,** no. 1, Feb. 1996, pp. 1-13.

- "An Introduction to Gallium Arsenide Devices," *Proceedings of the Australian International Electronics Technology Conference*, Adelaide, Oct. 1988, pp. 19-37.

- "Towards Ultra High Speed Systems," *Electronics '90*, Oct. 1990, pp. 193-202. (With A. Nunez.)

- "Conception of a Robot Eye Using a New Technology," *ANZAAS*, Microfilm, University of Sydney, Oct. 1991.

- "Ultra high speed gallium arsenide systems: design methodology, CAD tools and architectures," *Microelectronics Conf.*, Melbourne, June 1991, pp. 85-90.

- "An analog implementation of early visual processing in insects," *International Symposium on VLSI Technology, Systems, and Applications,* Taipei, pp. 283-287, May 12-14, 1993. (With Moini *et al.*)

- "An implementation of smart visual micro-sensor based upon insect vision," *Proc. 12th Aust. Micro. Conf.,* 5-8 Oct 1993, Gold Coast, Queensland, pp. 129-134. (With X.T. Nguyen *et al.*)

- "The stair-step tracking algorithm for velocity estimation," *Proc. 1st Australia & NZ Conf. on Intelligent Info. Sys.,* 1-3 Dec 1993, Perth, pp. 412-416. (With Moini *et al.*)

- "A new VLSI smart sensor for collision avoidance inspired by insect vision," *Proc. SPIE,* **vol. 2344**, 2-4 Nov 1994, Boston, pp. 105-115.

- "Biologically inspired obstacle avoidance – a technology independent paradigm," *Proc. SPIE,* **vol. 2591**, Oct 1995, Philadelphia, pp. 2-10, **Invited Keynote Paper.**

- "VLSI implementation of an insect vision based motion detector chip,"*IEEE J. Sol. St. Circ.,* 1996. (With Moini *et al*, accepted.)

- *Natural and Artificial Compound Eye Sensors,* Ed. J. Sanders, SPIE Milestone Series, 1996. (With A. Yakovleff *et al.*)

- "Obstacle Avoidance and Motion-Induced Navigation," *CAMP 95 Computer Architectures for Machine Perception Workshop*, Como, Italy, pp. 384-393, 1995. (With A. Yakovleff.)

- "GaAs versus SiGe – is there a challenge?" *GAAS 96 – European Gallium Arsenide and Related III-V Compounds Application Symposium,* CNAM, Paris, 1996, pp. 4A1 ff.

- "Current gain mechanism in planar GaAs MESFETs due to new photovoltaic self-biasing effect," *GAAS 96 – European Gallium Arsenide and Related III-V Compounds Application Symposium,* CNAM, Paris, 1996, pp. 3B3 ff.

- "Quantum vacuum fluctuations, zero point energy and the question of observable noise," *UPoN 96,* Szeged, Hungary, 1996, **(Invited).**

- "Is the thermal noise coefficient really 4?," *UPoN 96,* Szeged, Hungary, 1996, (accepted).

- "A new circuit theory paradox in the noise analysis of a 2-stage RC ladder," *UPoN 96,* Szeged, Hungary, 1996, (accepted).

- "Photovoltaic edge-effect in planar GaAs MESFETs," *ICO-XVII,* Taejon, Korea, 1996, (accepted).

List of Figures

1.1 Photonics versus electronics. The scope of this research is indicated in the shaded region. 10

1.2 Optical interconnects. 13

1.3 Partitioning of high speed systems. 14

2.1 Victor Moritz Goldschmidt 27

2.2 Number of published GaAs papers on the INSPEC data base. Only the keywords 'GaAs' and 'gallium arsenide' were searched for. 28

2.3 Understanding the meaning of 'semi-insulator.' 36

2.4 Carrier concentration plots, electric field lines and band bending diagrams for a silicon pn junction an a GaAs hi-lo n/SI junction. 38

2.5 Channel/substrate hi-lo junction field solved by FDM using the 4-level deep donor compensation model due to George *et al.* 39

2.6 Channel/substrate hi-lo junction field solved by FDM using the 2-level deep donor compensation model due to McGregor *et al.* 39

2.7 Energy Band Structures of GaAs v. Si 41

2.8 Lambert's Law of Absorption for GaAs. Solid line: $\lambda = 0.4$ μm. Dashed line: $\lambda = 0.6$ μm. Chained line: $\lambda = 0.8$ μm. 44

2.9 GaAs Absorption Length Dependence on Wavelength 45

2.10 GaAs v. Si Absorption Length Comparison 46

2.11 MTF due to diffusion. A comparison of the traditional formula due to Seib (dashed line) and the formula due to Barbe (solid line). 51

2.12 Planck's law 52

2.13 Measured absorption length against wavelength for silicon. 53

2.14 Measured absorption length against wavelength for GaAs. 54

2.15 Internal quantum efficiency v. wavelength 55

2.16 Weighted responsivity versus wavelength for silicon. Solid line, $W = 0.5\ \mu$m. Dashed line, $W = 2\ \mu$m. Chained line, $W = 5\ \mu$m 56

2.17 Weighted responsivity versus wavelength for GaAs 57

2.18 Theoretical MTF versus spatial frequency. Diffusion-only curves are: Si bulk, 550 nm (chained line); Si epi, 800 nm (dotted line); Si bulk, 800 nm (dashed line). Geometrical-only curve is for a 7/20 aperture to pitch ratio and the GaAs case matches this curve (solid line). 59

3.1 Equivalent circuit of MESFET showing photoresponsive elements. 70

3.2 Voltage across channel/substrate barrier against laser modulation frequency. The vertical dashed lines show the approximated time constant positions. 74

3.3 Drain current versus laser modulation frequency. Three curves represent three separate devices. 75

3.4 Wavelength response using xenon source monochromator. V_{gs}=-0.8 V, V_{ds}=0.6 V. Dual curve represents the looping effect. 76

3.5 I-V Characteristic variation using xenon source monochromator, V_{gs}=-0.8 V, '*'=700 nm, '+'=300 nm, 'o'=0 nm (dark). Dual curves represent the looping effect. There is no looping in the dark condition. 77

3.6 Transfer characteristic variation using xenon source monochromator. V_{ds}=0.6 V, '+'=700 nm, '*'=400 nm, 'o'=200 nm. 79

3.7 Transfer characteristic variation using laser flood. V_{ds}=0.6 V, '+'=815 nm @ 3 mW, '*'=632 nm @ 0.95 mw, 'o'=0 nm (dark). 80

3.8 Transistor length scan with focussed laser, 678 nm @ 0.125 mW, 2 μm diameter spot size. 1 MΩ series gate resistor. V_{gs}=-0.8 V, V_{ds}=0. 6V. . . . 82

3.9 Transistor width scan with focussed laser, 678 nm @ 0.125 mW, 2 μm diameter spot size. 1 MΩ series gate resistor. V_{gs}=-0.8 V, V_{ds}=0.6 V. Transistor edge corresponds to 20 μm on the x-axis. 83

3.10 Transistor width scan with focussed laser, 678 nm @ 0.125 mW, 2 μm diameter spot size. No series gate resistor. V_{gs}=-0.8 V, V_{ds}=0.6 V. Transistor edge corresponds to 20 μm on the x-axis. 84

3.11 Response against gate series resistor. 678 nm @ 0.125 mW, 2 μm diameter laser spot size. 1 MΩ series gate resistor. V_{gs}=-0.8 V, V_{ds}=0.5 V, 'o'=transistor edge, '+'=transistor middle. 85

3.12 Laser Prober with Computerised X-Y Platform 87

3.13 MESFET drain current response to 2-D laser scan, max. current 3.7 mA. No gate resistor. 88

3.14 MESFET drain current response to 2-D laser scan, max. current 29 mA. With 10 MΩ gate resistor. 89

3.15 Laser spot positions for high & low gain drain current response. 89

3.16 3-D Electric field plot using G-PISCES-2B with Spicer's UDM. Left half is field under the gate overhang. Right half is field under the Schottky gate. . 91

3.17 Convergence of Schottky gate, MIS gate and channel/substrate field lines at the transistor edge. Conceptual drawing with exaggerated features. . . . 92

3.18 Quantum efficiency versus gate resistor. Solid line: internal quantum efficiency (T=1) from first order theory. Dashed line: measured (low gain). Chained line: measured (high gain). 93

3.19 MESFET drain current response to 2-D laser scan, max. current 29 mA. With 10 MΩ gate resistor. 3-D plot. Performed on antivibration table. . . 95

3.20 MESFET drain current response to 2-D laser scan, max. current 29 mA. With 10 MΩ gate resistor. Contour plot. Performed on antivibration table. 96

4.1 Bert Johnson 106

4.2 Harry Nyquist 107

4.3 Lumped Circuit Model. 111

4.4 The window function. 113

4.5 Quantum equipartition energy 121

4.6 Noise in distributed RC line. 128

4.7 Two stage RC ladder – a source of an unsolved paradox. 129

4.8 Amplifier configuration. 130

4.9 Opamp noise sources 131

4.10 Imager output circuit configuration. 138

4.11 Imager output circuit noise (silicon). 140

4.12 Imager output circuit noise (GaAs). 142

5.1 Fingered Transistor Dual-Gate Pixel Conceptual Layout 160

5.2 Pixel Equivalent Circuit 161

5.3 Layout of 3 by 3 pixels. 163

5.4 Layout of 3 by 3 pixels, with antiblooming structure. 164

5.5 Array organisation. 167

5.6 Array organisation with antiblooming structure. 168

5.7 Shift register with bootstrapped stages 169

5.8 Conventional D flip-flop shift register stage 171

5.9 Conventional D flip-flop shift register layout segment 172

5.10 Conventional D flip-flop shift register simulation. Top trace demonstrates non-overlapping clocks, second trace is the serial input and subsequent parallel outputs follow. 173

5.11 Edge triggered D flip-flop shift register stage 174

5.12 Edge triggered D flip-flop shift register layout segment 175

5.13 Column-to-video line switching: equivalent circuit 177

6.1 DCFL logic gates 184

6.2 SDCFL logic gates 184

6.3 DCFL inverter circuit 185

6.4 SDCFL inverter circuit 185

6.5 Photograph qualitatively showing parasitic oscillations in I-V characteristic. Arbitrary axes. Notice region of negative resistance. 186

6.6 Layout styles: nMOS-type stick approach 187

6.7 Layout styles (a) Ring Notation using dual bus pair, (b) Ring Notation using single bus pair 188

6.8 Ring Notation showing topology of circuit 188

6.9 Ring Notation for 2-Input NOR gate, (a) initial representation, (b) after simplification 189

6.10 Translation from Ring Notation into Mask layout for a D-Latch 189

6.11 Gallium arsenide design process using CAD tools 192

6.12 1-bit Ripple Carry layout (a) nMOS type stick and (b) Ring Notation . . . 195

6.13 Floorplan of the GaAs implementation of carry look-ahead expressions . . 197

6.14 Area vs. delay of Ripple Carry and Carry Look-Ahead adders 197

6.15 Block diagram of Binary carry look-ahead adders 198

6.16 Comparison of different adders 199

6.17 Computational density of parallel multipliers 200

6.18 Shift register bit-slice (a) diagram, (b) simulation 201

7.1 Edge enhancement/thresholding edge detection system 210

7.2 A satellite image of Adelaide (band7). 213

7.3 The result of edge detection from Roberts operator. 214

7.4 The result of edge detection from Prewitt operator. 215

7.5 The result of edge detection from Sobel operator. 216

7.6 The result of a median filter. 219

7.7 The result of Sobel operator applied on the median filtered image. 220

7.8 SDCFL design example. 222

7.9 Block diagram of bit-serial algorithmic solution. 224

7.10 PE cell structural layout. 225

7.11 Generation of half-carry result. 226

7.12 Generation of sum result. 226

7.13 Generation of carry result. 226

7.14 Register logical representation. 227

7.15 Diplexor logical representation. 227

7.16 Adder floor plan block diagram. 228

7.17 Adder cell ring notation. 229

7.18 Register cell ring notation. 230

7.19 Diplexor cell ring notation. 230

7.20 PE cell structural block diagram. 230

7.21 Polynomial evaluator cell. 231

7.22 Adder and simulation. 232

7.23 Register and simulation. 233

7.24 Block diagram of bit-serial algorithmic solution with PE cell structural layout below. 234

7.25 System block diagram for basic mobile video device. Unshaded areas are standard tasks. Shaded areas need to be particularly addressed by CGaAs solutions. 238

7.26 Insect vision concept compared to conventional CCD camera. 244

7.27 Coding of templates 250

7.28 Template table. 250

7.29 Example template output pattern. X-axis represents bearing – each sample is 1°. Y-axis represents time – each sample is 20 ms. 251

7.30 Extraction of angular velocity from a template pattern. 252

7.31 Looming pattern. 252

7.32 Calculation for range 254

7.33 Theoretical MTF versus spatial frequency. Diffusion-only curves are for specific optical wavelengths. Geometrical-only curve is for the actual 7/20 aperture to pitch ratio used. 256

7.34 A Smart Pixel Architecture 262

7.35 A FLC/Si or FLC/GaAs Smart Pixel 265

7.36 A LED/GaAs Smart Pixel 265

B.1 Internal quantum efficiency versus absorption length. 292

B.2 Internal quantum efficiency versus absorption length. 293

C.1 Cylindrical coordinate system 296

List of Tables

2.1 Historical evolution of GaAs – a brief chronology. 29

2.2 Properties of Si, GaAs, SiC, GaN and Diamond. 30

2.3 Absorption coefficients 43

2.4 GaAs v. Si Responsivity. 58

4.1 History of fluctuation research – a brief early chronology. 108

4.2 Thermal noise over infinite bandwidth for different cases of limiting R and C. 125

5.1 Performance comparison of shift registers. Notice Case 1 has no power supplies as it is powered via the clocks. 175

6.1 Gallium Arsenide Symbolic Representation and Notations 190

6.2 Parameters of several ripple carry adders. 196

6.3 Parameters of the carry look-ahead adders. 196

6.4 Parameters of the binary carry look-ahead adders. 198

6.5 Parameters of several adders. 199

7.1 Some possible insect vision technology application areas 246

7.2 Evolution of 'Bugeye' Insect Vision Chip 253

GLOSSARY

ADC	Analog-to-digital converter
ATM	Asynchronous transfer mode
BISDN	Broadband ISDN
CCD	Charge coupled device
CDS	Correlated double sampling
CGaAs	Complementary GaAs
CHFET	Complementary high mobility GaAs FET (=CGaAs FET)
CID	Charge injection device
CMT	Cadmium mercury telluride
CW	Continuous wave (unmodulated laser)
DCFL	Direct coupled FET logic
DCT	Discrete cosine transform
DDS	Delayed double sampling
ESD	Electrostatic Damage/Discharge
EL2	"Electron level 2" – a deep donor
FDM	Finite difference method
FLC	Ferroelectric liquid crystal
FPN	Fixed pattern noise
HBT	Heterojuntion bipolar transistor
HDTV	High definition television
HEMT	High electron mobility transistor
HIGFET	High mobility insulating gate FET
IM^3PC	Interactive mobile multimedia personal communicator
IPN	Interactive personal navigation
ITB	Interactive telebanking
ITK	Interactive time keeping
ITO	Indium tin oxide
IQE	Internal quantum efficiency
LAN	Local area network
LWIR	Long wave infrared
MESFET	Metal semiconductor FET
MTF	Modulation transfer function
MWIR	Medium wave infrared
NED	Noise electron density
NN	Neural network
OEIC	Optoelectronic integrated circuit
ONN	Optical neural network
PAE	Power added efficiency
QE	Quantum efficiency
RAM	Random access memory
SI	Semi-insulating
S/D	Source or drain
S/N	Signal-to-noise ratio
SOI	Silicon on insulator
SRAM	Static random access memory
UDM	Unified defect model
WFT	Windowed Fourier transform
ZPE	Zero point energy
ZPF	Zero point fluctuations

Chapter 1

Introduction

"Research is the process of going up alleys to see if they are blind."
Marston Bates (1906-1974)
AMERICAN ZOOLOGIST

"Human knowledge is personal and responsible, an unending adventure at the edge of uncertainty."
Jacob Bronowski (1908-1974)
POLISH-BRITISH MATHEMATICIAN AND POET

1.1 Aims and Significance

UNIVERSITY OF ADELAIDE

1.1.1 Long Term Aim

The overall aim is to create a significant advance in the area of solid-state imaging via the research of an image sensor than can be ultimately integrated with high-speed gallium arsenide (GaAs) processing circuitry on a common substrate chip. This combination will facilitate a new generation of compact *smart sensors*, useful in mobile applications for detecting and transmitting full visual images. Spin-off application in the area of high definition television (HDTV) is also promising due to the larger absorption coefficients in GaAs leading to improved spatial resolution. The intention is for a GaAs image sensor, to be fabricated in a standard or near-standard GaAs digital IC process, enabling potential integration with a higher speed GaAs image preprocessor. This new paradigm offers the advantage of a preprocessor that can quickly access the image sensor in serial or parallel to efficiently perform low-level *early vision* tasks. The high-level vision tasks, if required, may be then carried out in real-time on a simpler host processor. The superior speed-power performance of GaAs over silicon and the excellent ongoing progress in GaAs fabrication maturity make GaAs a favourable choice for mobile applications where superior speed-power product is essential. Also the potential mixed digital/RF capability of GaAs, such as in the emerging CGaAs™ technology, is ideal for compact wireless mobile systems. The speed performance and other well known characteristics of GaAs are ideal for applications where harsh environment and increasing need for processing speed place high demands.

1.1.2 Immediate Aim of this Work

Due to very large throughput rates and vast amounts of data necessary for real time performance, there is an increasing demand for higher speed and lower power consumption in digital image and signal processors. This means that the choice of algorithms, and indeed that of the architecture to match the technology, becomes a crucial task in the realisation of such systems for real time applications.

Within the scope of the research to develop an imaging system for mobile applications it became obvious that many interrelated tasks had to be pursued in order to address numerous questions in relation to Ultra High Speed systems. The issues that required careful consideration included, the identification of suitable technology for Very High and Ultra High speed image processing, front end processing including the photodetector characteristics, architectural mapping and the design methodology.

Having identified gallium arsenide as a suitable contender, the complexities associated with this emerging technology became a focus for much of the initial research program. As the result of this investigation, an original approach to design methodology and hence architecture, referred to as the 'Ring Notation,' has been developed [Eshraghian 91].

The performance of GaAs optical sensors is already proven by our previous investigations [Abbott 91] and the work of others [Darling 93, Lange 94, Lin 93, Hibbs 91, Kozlowski 92], therefore one of the remaining major tasks in the development of a GaAs image sensor is to determine the various factors which affect the optical gain mechanisms in a GaAs MESFET. Thus, the need to accurately model, the behaviour of the MESFET transistor for implementation within the image sensor became the first major task pursued. It may be noted that considerable controversy in the literature exists over the nature of GaAs MESFET optical gain. As a result of our investigation, a consistent model was developed that overcame the difficulties raised by other researchers engaged in the controversy [Adibi 89].

This model is used to resolve apparent contradictions, between different results published in the literature, by observing that each of the reported effects come into play by varying degree depending on the wavelength of light used and hence penetration depth achieved. The results have been instrumental in providing a better understanding towards realisation of both on-chip and off-chip optical interconnects. Furthermore, in the course of our research, we uncovered an important effect that we have termed the "Edge Effect." This discovery may well have significant influence upon the pixel design of future imaging systems.

The research program for the gallium arsenide (GaAs) solid-state image sensor has been successful. The characteristics of the GaAs MESFET have been carefully investigated and we have demonstrated that the excellent characteristics enable its utilisation as the photosensitive element in an XY array imager, with particular advantages for mobile applications.

1.1.3 State of the Art and Present Knowledge in the Area of GaAs Arrays

An XY array image sensor is simply an array of transistors that is exposed to light and that can be addressed and read out in a similar manner to a memory architecture. The terms 'XY image sensor,' 'MOS imager,' and 'mosaic array' have all been used in the literature to describe essentially the same thing. The charge injection device (CID) also consists of an array of transistors, however this term is only used for those devices with

a specific patented read out technique championed by General Electric.

Presently the silicon CCD technology dominates the image sensor field. However, because CCDs require a complex esoteric process, they are not suitable for integration on a simple standard IC process. In the past, many workers, eg. [Dyck 68, Arnold 71, Kioke 80], developed simpler silicon XY array image sensors that are more suitable for integration on a standard IC process. These devices were popular in the early days of the hand-held video camera.

However, CCDs have now displaced XY arrays in the video camera market and consequently the XY array has been receiving little attention. Nevertheless, there are some recent studies in the literature with silicon XY arrays. A group at Edinburgh have integrated an XY array to create a smart sensor allowing simple low-level preprocessing functions such as adaptive thresholding and correlation/convolution – this scenario has been used with success for applications such as biometric verification [Mavor 92, Vellacott 94]. A group at AT&T have recognised the importance of the XY array for multimedia applications [Dickinson 95].

For mobile applications, such as in aerospace, where there is a greater demand for speed-power performance, our unique proposal is a paradigm-shift from silicon to GaAs. The inherently larger output node capacitance on silicon XY arrays, as compared with CCDs, gives rise to higher kTC noise which is a limiting feature, of the old paradigm. Consequently, the lower line and parasitic capacitances in GaAs mean that this technology is better suited to the XY array approach. Over the past decade there has been extensive research on the use of a simple GaAs MESFET as an optical detector. This activity is due to the MESFET's superior noise performance over, say, the *pin* diode and more so because the MESFET can be simply integrated on a standard IC process thus providing an ideal optical port for an OEIC, eg. [Hibbs 91]. Now that GaAs has considerably matured, over the last 5 years, there is renewed activity in GaAs for X-ray detectors with applications in medicine, astronomy etc. GaAs has proven advantageous in this area due to low thermal leakage current [McGregor 92, Beaumont 93, Bencivelli 95] and greater tolerance to permanent radiation damage. Work in the area of GaAs CCDs is ongoing, but is still immature due to the complex CCD process [Liu 80, Sahai 83, Kosel 85]. This CCD work has no bearing on our research other than highlighting the value of GaAs as an optical detector and also demonstrating that the Schottky gates can serve as convenient in-house antiblooming lines. This is not possible with MOS gates, in traditional optical imagers, due to the insulating oxide. The antiblooming facility, with Schottky gates, can be exploited in the proposed XY array.

In conclusion, on the international scene, there has been considerable interest in single

GaAs detectors and their optical performance is proven, with a number of advantages. Our novel proposal recognises the advantage of producing an array of such devices to form an imager in the optical spectrum.

GaAs is now at a level of maturity where chips with as many as 1.44 million transistors have been successfully built – this highlights the rapidly increasing advance in yield and at full maturity there is no fundamental reason why GaAs yield should not, in fact, exceed that of silicon (due to the fewer mask layers). Other supporting evidence towards a successful outcome is the fact that photodetector arrays, different in nature from our proposal, but nevertheless highlighting the ability of GaAs have been successfully fabricated. Examples are neural photodetector arrays [Darling 93, Lange 94, Lin 93], OEIC array connectors [Hibbs 91] and hybrid infrared arrays [Kozlowski 92].

1.1.4 Significance of GaAs

This work investigates fundamental GaAs principles that will ultimately enable an advance in the area of smart sensors, especially for mobile applications. Important spin-off areas include remote sensing, industrial robotic vision, forensic image restoration, aerospace vehicle guidance and control, mobile videophones, air surveillance and mapping and possibly HDTV. Significant future branch off areas of this proposed technology may include infrared (via hybrid techniques) and X-ray imaging which opens up vast application areas in medicine, astronomy etc. The economic benefit and social impact in all these areas is clearly not insignificant.

The **key reason** for developing the imager in a **standard** GaAs IC process, is the facility *to eventually integrate this sensor with a digital GaAs preprocessor for low-level vision tasks* such as thresholding, contrast enhancement, edge detection etc. and/or data compression/encryption for remote communication. The advantage of a GaAs preprocessor is to meet the demands for ever increasing speed-power performance. Improved speed-power together with the increased compactness from the circuit integration suits well for advanced mobile applications.

In addition, there are numerous interesting side benefits to an imager in GaAs. To summarise, firstly, we have already discussed the reported low thermal leakage current [McGregor 92, Bencivelli 95], lower line and parasitic capacitances better suited to the XY array approach, natural antiblooming provided by Schottky gates and greater tolerance to permanent radiation damage. In comparison to CMOS, it is well known that the benefits against radiation induced phenomena include: (a) no latchup (b) no gate oxide charging (c) no field inversion (d) shorter carrier lifetimes.

Secondly, further side benefits from using GaAs are: (a) simple manufacture (hence yield will be excellent when full maturity is achieved); (b) at visible wavelengths the absorption length in GaAs is less than one micron – this implies an up to factor ten improvement over silicon – also the visible absorption curve is favourably centred about the channel depth; (c) this implies better photocollection hence quantum efficiency – in CMOS useful photocollection is zero below the epitaxial layer (bulk substrates are not normally used due to pixel crosstalk); (d) in our imager the n+ on semi-insulator (SI) is a high-low junction and collects *majority carriers* – hence there is no Modulation Transfer Function (MTF) degradation due to diffusion – plotting from the well known equations shows that, if we assume the geometrical MTF curve for GaAs, this is typically better than the diffusion MTF of epitaxial silicon, especially at longer wavelengths – this is useful for high definition applications.

This work does not rely on the immediate realisation of any one of these side benefits – however, their significance lies in their potential impact on wider application areas. Future research will answer the question of exactly how large the anticipated side benefits are in reality.

1.2 Overview of XY Array Strategy

1.2.1 Photocollection Modes

All present-day silicon 2-D imagers are designed around a photovoltaic collection technique. Not only does the photovoltaic approach lend itself to the architecture of the imager, but the details of the physics reveals a lower generation noise. Also a further reduction in the effect of noise is achieved, over photoconductive devices, as photovoltaic devices intrinsically perform a capacitive temporal integration of the signal – thus improving the signal-to-noise ratio. S/N can be increased in photoconductive devices, by increasing their size – ie. effectively introducing a spatial integration. However any significant increase in area would render a 2-D imager array inordinately large. Therefore it is this fact that photoconductive devices are in the spatial domain, whereas photovoltaic devices are in the temporal domain, that leads to the fundamental reason why the photovoltaic approach is favoured for 2-D arrays, in the interest of a high S/N per area ratio.

Having chosen that the imager must be designed around the optimisation of photovoltaic mechanisms, we need to decide on which MESFET depletion region should be utilised. Making the correct choice is crucial as the photovoltaic mechanism relies on the electric

field, supported by a depletion region, to separate electron-hole pairs and thus collect charge. The following depletion regions are available in the MESFET: (1) surface depletion region, (2) S/D to substrate depletion region, (3) gate depletion region and (4) channel to substrate depletion region. The surface depletion region can be immediately disregarded, as it is not a 2-terminal structure and therefore charge readout would be impossible. S/D or channel to substrate depletion regions and gate depletion regions are both viable. Even though there is a long RC time constant associated with the S/D and channel cases, at TV rates this mode is still viable. However, a design based around the gate depletion region offers antiblooming possibilities.

Selection of this photocollection mode immediately suggests two useful constraints for the pixel design. Firstly, the pixel must include a dual-gate MESFET – one gate is to be shielded from the light and acts as a switch to enable/disable readout, and the other is used for photocollection. Secondly, as the gate material is opaque, the detector gate is to be designed with a 'fingered' structure to maximise the amount of depletion region that is exposed to generated photocharge.[1] The use of a fingered transistor will act as a simple 'proving ground' to demonstrate the working concept of the imager – with a standard unaltered digital GaAs IC process. Although the design relies solely on the gate depletion region, it is anticipated that the channel/substrate region may have an advantageous side effect – namely, stray electrons in this region will be swept into the channel. This effect will lead to improved Modulation Transfer Function (MTF) characteristics.

1.2.2 Pixel Array

In later sections it will be seen that this work experimentally verifies samples of fingered MESFETs as the basis for a detector pixel. However, The effect of transistor finger spacing on the collection efficiency of the GaAs device, needs to be fully quantified and further investigated experimentally. As the gates are opaque, the finger width will be kept to a minimum. A series of test structures with various finger spacings would be the best way to maximise the quantum efficiency – future research can also investigate transparent gates such as thinned gold or ITO [Sahai 83]. Hence, at this initial research phase, optimising fill factor will not be an immediate issue. The research challenge of surface leakage currents will be combated, if necessary, using annular structures.

The array organisation required to read out the pixels will follow the standard techniques used for silicon XY array imagers. Basically, one shift register addresses a series of paths

[1]Although the gate is opaque, photocollection can still take place as the depletion region laterally protrudes from under the gate. With an optimised structure this scheme can be surprisingly effective.

that control the pixel enable/disable gates. A second shift register is used to address an analogue multiplexer that converts the parallel data into a serial stream. A *charge sensing* approach will be adopted, for readout, where charge is transferred from pixel to data line and from data line to output bus by a process of capacitive charge sharing. This is a well understood and proven technique used in silicon XY imaging arrays.

Silicon imagers have been known to adopt either the technique of *voltage sensing* or *charge sensing* for image readout. Voltage sensing entails detecting a voltage at each pixel and charge sensing reads out the charge from each pixel. Basically, charge sensing is the superior technique because: (1) each pixel does not need to be explicitly reset, as the action of charge readout is the reset, (2) one capacitor is used for charge-to-voltage conversion, at the output, rather than one at every pixel, therefore fixed pattern noise is reduced and (3) charge-to-voltage conversion occurs at the final output and not at each pixel node, therefore pixel circuitry and hence space is reduced. Recent reports, for silicon arrays, have described an 'active pixel' technique [Dickinson 95] using voltage sensing at each pixel and an alternative approach that implements the sensing at every pixel column [Mavor 92, Vellacott 94]. Both approaches are fundamentally limited by the degree of matching between active circuits and the area consumed by the additional circuitry that would eliminate a realistic HDTV implementation.

The only real advantage with voltage sensing is the increased signal-to-noise ratio due to the small pixel capacitance used for charge to voltage conversion. In the charge sensing case, a lower signal-to-noise ratio results from the fact that the charge-to-voltage conversion capacitance is unavoidably larger (inherently due to the architecture of the array) and is required to ensure efficient charge transfer to the output bus. It is this very problem that work such as [Mavor 92, Vellacott 94] and [Dickinson 95] seeks to address. GaAs has the advantage that the lower metal track capacitances improve the S/N and therefore charge sensing lends itself to this technology. Hence the use of GaAs is an alternative solution to [Mavor 92, Vellacott 94] and [Dickinson 95], promising superior performance and space saving.

1.3 Research Philosophy

This work initially raised a number research questions which we highlight as follows.

If the GaAs imager is realised, where are the areas of innovation and originality in performing this research?

Our near-term goal is to fabricate the worlds first GaAs XY array imager in order to

characterise it and investigate the research challenges that include fixed-pattern-noise, surface leakage, etc. and definitively answer a number of fascinating research questions such as "is the MTF really good enough for HDTV?" or "how well does the antiblooming perform?" and how the results compare with theory. We know that this pilot chip will not outperform silicon in terms of fill-factor and defect density, but these are not the immediate issue, are not fundamental and will gradually improve with technology. Our immediate concern is whether the hypotheses connected with the short absorption lengths can be verified within an actual working imager and how well our circuit techniques can overcome the challenge of small signal swings.

Given the usual read-out times for imagers, is not the speed of GaAs made superfluous?

No. One future option is a low speed imager integrated with high speed image processing circuitry to produce a compact "smart" sensor and/or RF circuitry for wireless communications. Another option is for the imager itself to go to higher speeds for non-TV applications (eg. machine target tracking).

What are the advantages of using GaAs?

(1) A GaAs image sensor can be integrated with high speed GaAs image processing circuitry implying future increased compactness and reliability, (2) improved speed-power product, (3) simple manufacture, fewer mask layers than CMOS, hence *improved* defect density when full maturity is reached, (4) natural antiblooming provided by Schottky gates and hence ultimately more compact pixels, (5) good spatial resolution from shallow absorption depths; hence no need for epi-on-high doped layer approach *implying improved quantum efficiency*, (6) possible further sensitivity increase via our internal gain edge-effect discovery [Abbott 91], (7) higher temperature tolerance from larger bandgap, (8) lower dark current as evidenced by published results [McGregor 92], (9) greater radiation tolerance to *permanent* damage, (10) no latchup.

What are the research challenges?

(1) Fixed pattern noise – this will be addressed by CDS/DDS techniques, (2) surface leakage current – addressed by using annular structures, (3) threshold voltage drops making switch designs difficult – we have a number of published approaches within our design methodology to carry out (4) non-optimum fill factor – a demonstrator chip can be made to work by slightly increasing the lighting level to prove the concept — transparent gates are then perfectly possible for the future.

What are the future research questions to answer once the chip is fabricated?

(1) What is the smallest pixel pitch possible, measure the MTF curve and is this suitable for HDTV? (2) What is the fastest and lowest operation speed and the power consumption? (3) What is the best finger spacing? (4) Is the edge-effect usable? (5) Based on the measured kTC noise on the output, project what would be the largest array size possible? (6) Can the concept be proved with a demonstration with at least producing a crude image? (7) What light levels can the antiblooming withstand? (8) What is the measured responsivity?

1.4 Field of this Research

The merging of optical with electronic functions is a rapidly growing field and terms such as 'electrooptics,' 'optoelectronics,' 'quantum electronics,' 'quantum optics' and 'light wave technology' have been somewhat elastic as these paradigms have been in a formative stage. However, a consensus is beginning to emerge and it seems clear that 'photonics' and 'electronics' are now regarded as *parallel* terms, see Fig. 1.1, where electronics deals with the control of *charge* flow and photonics with the control of *photons*.

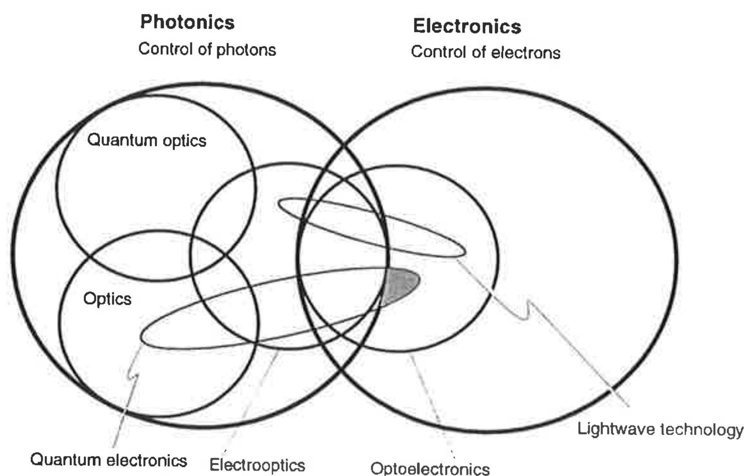


Figure 1.1: Photonics versus electronics. The scope of this research is indicated in the shaded region.

Electrooptics is a term for essentially optical devices where electrical effects have a role, for example as in the laser. On the other hand, optoelectronics refers to systems that are essentially electronic in nature but where light is involved, for example LCDs and imaging arrays.

The term quantum electronics is used where light interacts with matter and therefore, to

some extent, is a subset of the above terms. Quantum optics refers to the study of light itself, such as in its quantum and coherence properties. Lightwave technology is used to describe devices used in optical communications and optical signal processing.

Hence, it becomes clear that imaging arrays, the subject of this thesis, is in the field where quantum electronics, optoelectronics and electronics overlap – this is indicated by the shaded area in Fig. 1.1.[2]

1.5 Achievements from this Research

Considerable amount of work was conducted in the area of optical modelling of GaAs MESFETs to determine the advantages and feasibility of the proposed imager. This modelling has created a firm foundation to enable a first pass design of the imager pixel. The pixel is a transistor with a floating drain exposed to light. The concept and operation of the pixel follows identical well known principles proven in current silicon XY array technology. A notable difference is the presence of a special fingered gate over the photosensitive drain. The depletion regions around the fingers provide increased photocollection volume. Also during optical overload (blooming), the over charge will send the fingers into forward bias. By connecting all the fingers to a constant voltage source (with good current sink capability), a momentary path to sink the excess charge is created until reverse bias is reestablished.

Significant progress has also been made by our discovery of a photovoltaic gate biasing edge effect [Abbott 91, Abbott 93]. Here we reported, that under certain gate biasing conditions, dramatic peaks in drain photocurrent gain are observed where the gate overlaps the transistor edges. We have successfully secured a patent [Abbott 93] detailing a unique imager design where we utilised this gain effect to create an increased sensitivity.

1.6 GaAs Background and Future Vision

The rapidly emerging area of High Speed and Ultra High Speed processing that underpins the transformation of much of the ideas into working systems necessitates for evolutionary changes in both the technology and the strategy that would facilitate the ability for physical mapping of such systems. The systems that are mostly affected, and indeed in

[2]Note that Fig. 1.1 is not rigorously correct, but is a useful starting point for discussion. Full rigour is not appropriate as the terms are somewhat elastic and exact representation would probably be impossible on a 2-dimensional diagram. The beauty of Fig. 1.1 lies in a mature acceptance of ambiguity.

the next decade or so would require an ever increasing processing power include real time signal processors and image processors, computer vision, telecommunications, biomedical systems and personal super computers, just to name a few.

Since the invention of the transistor in 1947, and the development of the very first integrated circuit at the beginning of 1960, there has been four generation of ICs. Now, we are beginning to witness the emergence of the fifth generation of ICs that are characterised by complexities in excess of 1.8 million devices on a single chip. Over the past several years, CMOS technology has become the dominant fabrication process for relatively high performance and cost effective VLSI circuits and systems. For example, today's microprocessors are able to handle some 80-100 million operations per second, but the circuitry that manages communications with other processors and in particular with the memory, is just too slow to keep up with this data rate.

Paralleling the developments in CMOS technology some evolving technological changes are also beginning to take place in the area of BiCMOS where some of the advantages of the Bipolar transistor are being utilised as CMOS speed enhancer. Main advantages of Bipolar technology over MOS, are greater current drive per unit area, low delay sensitivity to load variation (make excellent line drivers, and decoders for memory elements and fast on-chip cache), better device matching, low sensitivity to process variation and linear performance (sense amplifiers and current mirrors).

Although progress in both silicon processing and technology has been somewhat remarkable during the last decade, we are still confronted with the speed/power limitations of the technology that are becoming apparent in fast digital systems and subsequently have brought about the need for development of other technologies such as gallium arsenide.

Gallium arsenide is rather a different material to silicon. For the same power consumption, GaAs is about half an order of magnitude faster than emitter-coupled logic (ECL), the fastest of the silicon family. The speed advantage is primarily the result of gallium arsenide's high electron mobility and the semi-insulating substrate. Other advantages of gallium arsenide over silicon include its higher temperature tolerance, radiation hardness and optoelectronic properties. The last point permits efficient integration of electronic and optical devices, on the GaAs IC, and is somewhat critical for Ultra High Speed systems of the future. Gallium arsenide has a direct bandgap. This means that energy level transitions take place with momentum conservation, allowing photon emission. Thus, GaAs is both a good emitter, as well as a good receiver of light. This creates new opportunities towards on-chip integrated optical communications. Optical interconnect can either take the form of optical fibres or thin film paths (eg. aluminium oxide) integrated on the chip carrier as illustrated in Figure 1.2. Another form is the free-space interconnect where

focussed or unfocused light is broadcast over an area. This can be utilised to supply a high speed global clock to a VLSI circuit without the usual routing complexity. Here an opaque masking layer would be used to shield parts of the chip from unwanted light.

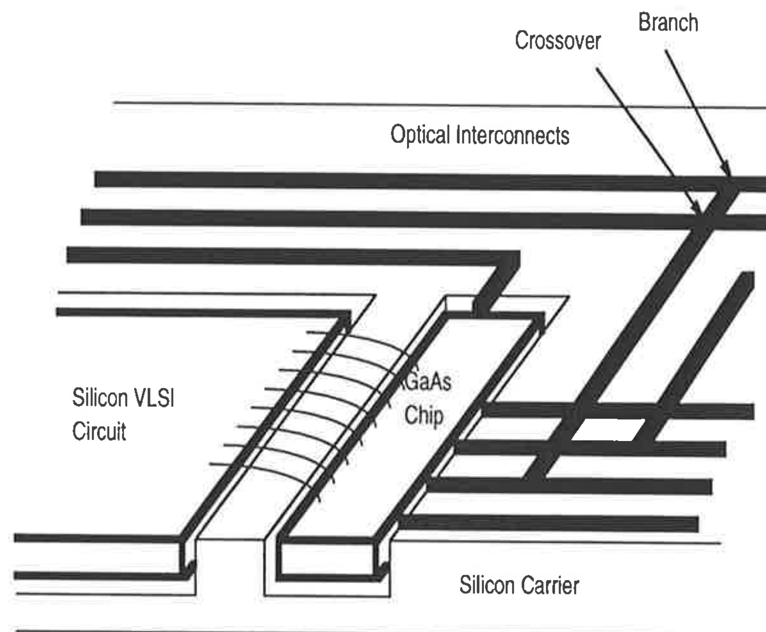


Figure 1.2: Optical interconnects.

One possibility is to utilise gallium arsenide technology for front-end processor sections of high-speed single stream processors for digital data (eg. 1-2 Gigasamples/second), usually generated by way of very wideband image sensor, detectors or the like. This fast data stream can subsequently be subdivided into lower rate parallel streams suitable for processing in silicon CMOS/BiCMOS subsystems at lower frequencies. Figure 1.3 illustrates such a concept and highlights the critical paths for communications.

The necessity for high clock rates in the 'front end' processors is compounded since it is very likely that 10-20 microcycles of the processor may be required to preprocess each incoming data sample. This means an input data rate of 10^8 bytes/second might demand a system clock rate of 2GHz. By mixing GaAs and Silicon CMOS/BiCMOS technologies, it becomes possible to exploit high system clock rates together with all the important reduction in power dissipation in a number of systems, including high bandwidth image and signal processors in particular in aerospace applications.

In order to take the advantages that can be gained by combining GaAs/BiCMOS/CMOS as the base for Ultra High Speed High performance imaging VLSI systems referred to as the 'unified technology', attention is also directed towards characterisation, optimisation

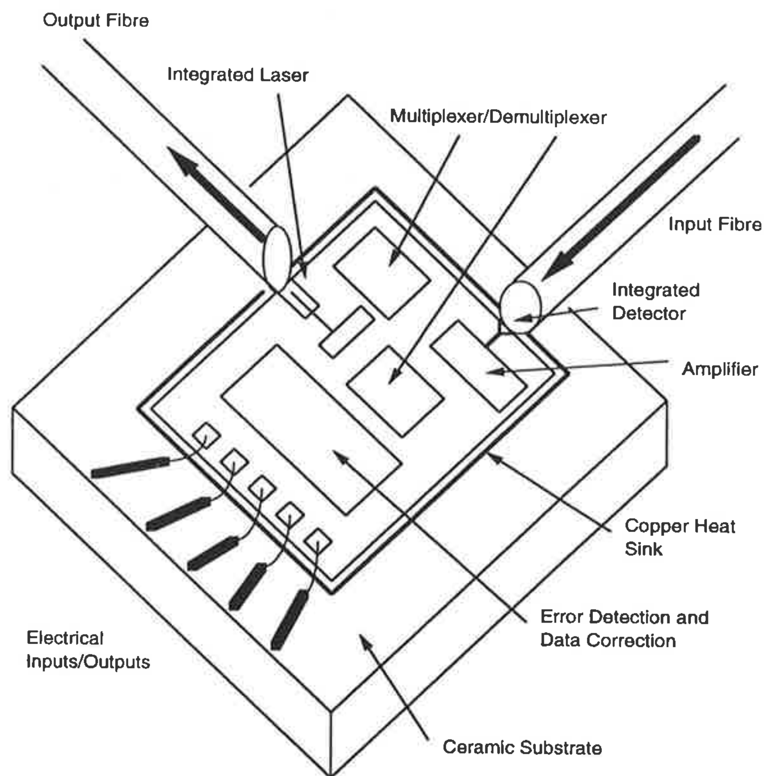


Figure 1.3: Partitioning of high speed systems.

as well as development of a design methodology for GaAs/BiCMOS/CMOS as the base technology for Ultra High Speed VLSI systems. The uniqueness of the approach provides the possibility to exploit high system clock rates in a number of systems including telecommunication, vision systems, and high bandwidth signal and image processing applications. The significance of this research is that it could provide the design community the all-important design methodology for the design of High Speed and Ultra High Speed, High Performance VLSI systems. In order to facilitate rapid progress and to create the foundation for future work towards algorithm transformation and mapping of Ultra High Speed, high performance imaging systems, initially two areas of research namely GaAs sensor development and some aspect of architectural issues associated with the image processing have been pursued within the scope of this program.

1.7 Summary and Chapter Overview

In this introductory chapter we have outlined the immediate and future aims of this work. We have highlighted the research challenges that will be addressed in later chapters, as well as those for future work. We have introduced the paradigm of a smart image sensor based on the gallium arsenide technology and have discussed the rationale and anticipated benefits. Future application areas have been briefly described – however the concept of of an interactive multimedia mobile communicator (IM^{3}C) or mobile videophone and imaging for aerospace are the main, though not exclusive, motivational forces for this work.

In Ch. 2 we first take a look at the GaAs technology. In order to understand why we selected GaAs, for this work, its history and comparison with other technologies is reviewed. It is shown that most competing technologies except for silicon are either inappropriate or too immature. Hence Ch. 2 is mainly devoted to a comparison with silicon in terms of mainly optical and substrate properties. Electrical circuit properties are well documented in the literature and are not repeated here except in brief summary.

Having accepted GaAs as the choice of technology for this work, Ch. 3 looks at the issues of photodetection in GaAs MESFETs. Detailed photoresponse measurements are described and a new photovoltaic 'edge-effect' is reported, giving an enhancement in photosensitivity. The results of Ch. 3 suggest two possible approaches for detectors:

- *Active pixel.* An active pixel exploits the internal gain of the GaAs MESFET to produce enhanced photosensitivity. Although promising, it is argued that this approach is not suitable for the imager array, as gain variations lead to fixed pattern

noise (fpn). However, Ch. 7 proposes that active pixels can be exploited in motion detectors, where only coarse thresholded optical data is required.

- *Passive pixel.* A passive pixel simply uses the drain junction of the MESFET as a photodetector. The MESFET itself is used as a switch to read out the photocollected charge at the drain node. Due to the simplicity and better uniformity of this approach, it is adopted for the imager design proposed in Ch. 5. A further possible embellishment is described where fingers of Schottky metal contacts are laid over the photosensitive drain area to act as antiblooming lines.

Having suggested these useful design constraints for the imager pixel, based on photodetector measurements in Ch. 3, Ch. 4 investigates noise issues in order to understand the design constraints of the imager's output circuit.

As the area of noise is somewhat obscure and poorly understood in the literature, Ch. 4 begins with an original survey of the history and areas of common misunderstanding in noise theory. This sets the framework for then discussing noise in the imager output circuit. Original theory and results in the area of noise are also presented in Ch. 4 – although this is not all directly relevant to the imager design, it is included as the results are startling and were part of the meandering path of this research. This is the nature of true research and the reader is asked to indulge in a tangent for several pages of Ch. 4.

Given pixel design constraints, from Ch. 3, and output circuit design constraints, from Ch. 4, an overall outline of a possible initial proof-of-concept imager design using a passive pixel approach is finally given in Ch. 5.

Having the proposed imager design scheme, Ch. 6 then proceeds to describe the early work carried out to define a suitable design methodology to realise VLSI circuits on GaAs. Using this methodology allows the efficient realisation of image processing circuits that can then be integrated with the imager array, thus making way for the smart imager paradigm in GaAs.

Ch. 7 then uses this methodology to present a case study of an actual GaAs design of a polynomial evaluator that can be used as part of a real-time image post-processor, with the imager at the front-end, for applications such as in aerospace. Ch. 7 then proceeds to survey other possible image enhancement functions for future GaAs realisation, as an illustration of the type of processing of the imager's output that can be performed in real-time for the next generation of aerospace or IM^{3}C smart sensors. Finally, an alternative type of vision system application is described: a motion detector based an insect vision for collision avoidance – this new concept uses coarse thresholded data and thus is an ideal application for exploiting the active pixel approach identified in Ch. 3.

In this way, Ch. 7 uses a two-pronged approach – it not only illustrates the vision of the smart imager paradigm, but also illustrates a possible exploitation for the promising active pixel mechanisms, uncovered by this work, within an alternative smart motion detector paradigm.

The branching of this work into these two equally significant approaches, was not foreseen at the outset of this work, but became apparent as the research evolved. Although the branching into the motion detector paradigm represents a departure from the original aims, it was a natural consequence from following through the interesting results from this work.

Bibliography

[Abbott 91] D. Abbott, S. Cui, K. Eshraghian and E. McCabe, "Photovoltaic gate biasing edge effect in GaAs MESFETs," *Electronics Letters,* **vol. 27**, no. 21, pp. 1900-1902, Oct. 1991.

[Abbott 92] D. Abbott and K. Eshraghian, "Imaging system for aerospace vehicle," *GaAs Centre Report,* vol. GAAS 92-01, 1992.

[Abbott 93] D. Abbott and K. Eshraghian, "Gallium arsenide MESFET imager," *World Intellectual Property Organisation, International Bureau*, Geneva, **Patent WO 93/07643**, 15 Apr. 1993.

[Adibi 89] A. Adibi and K. Eshraghian, "A generalised model for GaAs MESFET photodetectors," *IEE Proc. G, Circuits, Devices & Systems,* **vol. 136**, Part-G, no. 7, pp. 337-343, Dec. 1989.

[Arnold 71] E. Arnold *et al*, "Video signals and switching transients in capacitor-photodiode and capacitor-phototransistor image sensors," *IEEE Trans. Elec. Dev.,* **vol. 18,** no. 11, pp. 1003-1010, Nov. 1971.

[Beaumont 93] S.P. Beaumont *et al,* "GaAs solid state detectors for physics at the LHC," *IEEE Trans. Nuc. Sci.,* **vol. 40,** no. 4, 1993.

[Bencivelli 95] W. Becivelli *et al*, "Some new results on semi-insulating GaAs detectors for low energy X-rays," *Nuc. Inst. Methods Phys. Res. A,* Elsevier, **vol. 355,** pp. 425-427, 1995.

[Darling 93] R. B. Darling and W. T. Dietze, "Implementation of multiplicative lateral inhibition in a GaAs sensory neural-network photodetector array," *IEEE J. Quantum Elec.,* **vol. 29,** no. 2, pp. 645-654, Feb. 1993.

[Dickinson 95] A. Dickinson, B. Ackland, E. Eid and D. Inglis, "Standard CMOS active pixel image sensors for multimedia applications," *Advanced VLSI Conference,* Chapel Hill, March 28, 1995.

[Dyck 68] R.H. Dyck and G.P. Weckler, "Integrated arrays of silicon photodetectors for image sensing," *IEEE Trans. Elec. Dev.,* **vol. 15,** no. 4, pp. 196-201, April 1968.

[Eshraghian 91] K. Eshraghian, R. Sarmiento, P.P. Carballo, and A. Nuñez, "Speed-area optimization for DCFL and SDCFL class of logic using ring notation," *Microprocessing and Microprogramming,* **vol. 32,** pp. 75-82, 1991.

[Hibbs 91] M. Hibbs-Brenner *et al*, "Design, fabrication and performance of an integrated optoelectronic cellular array," *SPIE Oprical Enhancements to Computing Technology,* **vol. 1563,** pp. 10-20, 1991.

[Kioke 80] N. Kioke *et al*, "MOS area sensor: Part 1," *it IEEE Trans. Elec. Dev.,* **vol. 27,** no. 8, pp. 1676-1687, 1980.

[Kosel 85] P.B. Kosel *et al*, "Configurations for high speed gallium arsenide charge-coupled device imagers," *Optical Engineering,* **vol. 24,** no. 1, pp. 176-182, 1985.

[Kozlowski 92] L.J. Kozlowski and R.E. Kezer, "2 x 64 GaAs readout for IR FPA application," *SPIE Infrared Readout Electronics,* **vol. 1684,** pp. 131-138, 1992.

[Lange 94] E. Lange, Y. Nitta and K. Kyuma, "Optical Neural Chips," *IEEE Micro,* **vol. 14,** no. 6, pp. 29-41, 1994.

[Lin 93] S. Lin, A. Grot, J. Luo and D. Psaltis, "GaAs optoelectronic neuron arrays," *Applied Optics,* **vol. 32**, no. 8, pp. 1275-1289, 1993.

[Liu 80] Y.Z. Lui, I. Deyhimy, R.J. Anderson, R.A. Milano, M.J. Cohen, J.S. Harris and L.R. Tomasetta, "A backside-illuminate imaging AlGaAs/GaAs charge-coupled device," *Appl. Phys. Lett.,* **vol. 37,** no. 9 pp. 803-807, 1980.

[Mavor 92] J. Mavor, "Compact image processors employing analogue MOS techniques," *IEEE APCCAS 92,* Sydney, pp. 1-9, 8-11 Dec. 1992.

[McGregor 92] D. S. McGregor *et al*, "Development of bulk GaAs radiation detectors," *IEEE Nuc. Sci.,* **vol. 39**, no. 5, pp. 1226-1236, October 1992.

[Moini 93] A. Moini, K. Eshraghian, D. Abbott *et al*, "An analog implementation of early visual processing in insects," *International Symposium on VLSI Technology, Systems, and Applications,* Taipei, pp. 283-287, May 12-14, 1993.

[Sahai 83] R. Sahai *et al*, "GaAs CCDs with transparent (ITO) gates for imaging and optical signal processing," *IEEE Electron Device Lett., vol. 4,* no. 12, 463-464, 1983.

[Sensiper 93] M. Sensiper *et al*, "Modulation transfer function testing of detector arrays using narrow-band laser speckle," *Optical Engineering*, **vol. 32**, no. 2, pp. 395-400, 1993.

[Vellacott 94] O. Vellacott, "CMOS in camera," *IEE Review*, pp. 111-114, May 1994.

Chapter 2

GaAs versus Silicon

"All great truths begin as blasphemies."
George Bernard Shaw (1856-1950)
IRISH COMIC DRAMATIST AND LITERARY CRITIC

"This is the essence of science: ask an impertinent question, and you are on the way to a pertinent answer."
Jacob Bronowski (1908-1974)
POLISH-BRITISH MATHEMATICIAN AND POET

"The great tragedy of science – the slaying of a beautiful hypothesis by an ugly fact."
Thomas Henry Huxley (1825-1895)
BRITISH BIOLOGIST

2.1 Introduction

The purpose of this chapter is to examine the history of GaAs and its competing technologies, in order to illustrate the importance of the chosen technology. Given that silicon emerges as the main contender, we proceed to critically examine the unusual substrate and optical properties of GaAs, as compared to silicon, to understand and demonstrate the anticipated advantages of GaAs.

2.2 History

As part of the understanding of our choice of the GaAs technology and gauging its level of maturity for the proposed work, we shall briefly review its historical development.

2.2.1 Gallium

Gallium is a soft white silver metal found throughout the earth's crust and can be extracted from ores such as bauxite, coal and sphalerite. It has a low melting point (29.78°C) and can melt with the warmth of a hand. Consequently, its first application was in high temperature thermometers. Other applications have been in low melting solders, arc lamps, batteries and vanadium-gallium superconductors. Gallium is one of the few metals that expands as it freezes.

On the 27th of August 1875, between 3 and 4 am in the morning (!), a French chemist by the name of Paul Émile Lecoq de Boisbaudran (1838-1912) discovered gallium by isolating it from zinc blende ore found in the Pyrenees. With a spectroscopic analysis he confirmed that he indeed had a new element and Lecoq identified this with Mendeleyev's prediction of the existence of 'eka-aluminium.' [1] Lecoq made a significant discovery by showing that gallium = eka-aluminium, as this was the *first* of Mendeleyev's predictions of the periodic table to be verified.

There is some debate as to where the name 'gallium' comes from. According to some authorities it derives from the Latin *Gallia* meaning Gaul (France) and hence Lecoq named it in honour of his country. Other sources claim that it derives from the Latin *gallus* meaning a 'cock' and hence was named after Lecoq himself. The educated guess of this author is that Lecoq made a deliberate *double entendre*!

[1] *Eka* is Sanskrit for 'one.'

2.2.2 Arsenic

Arsenic compounds have a long history going back into antiquity. The main compounds are orpiment (As_2S_3) or 'yellow arsenic,' arsenolite (As_4O_6) or 'white arsenic,' arsenopyrite (FeAsS) and realgar (AsS). The pure metallic form of arsenic (As) is called 'grey arsenic' and comes in the form of a flaky grey brittle metalloid. The name 'arsenic' was thought to come from the Greek *αρσενικος* meaning 'male' – as the Greeks believed metals had different sex. However, the latest authorities now say it comes from the Arabic *az-zarnīk* meaning 'orpiment' – a compound of arsenic used as a yellow/gold pigment. In turn, *az-zarnīk* comes from the Persian word *zar*, meaning 'gold.'

The first documented description of orpiment and realgar is due to Aristotle (384-323 BC). A neoplatonist by the name of Olympiodorus of Alexandria (*fl. ante* 550 AD) is the first to document white arsenic. The German alchemist and Dominican philosopher Albertus Magnus (*c.* 1206-1280) was the first to actually document metallic arsenic and its discovery is usually attributed to him. An anonymous Spanish alchemist, using the pen name Geber (*fl.* 14th cent. AD), discovered that white arsenic can be obtained by burning realgar.

The Swedish chemist Georg Brandt (1694-1768) showed the connection between white and metallic arsenic in 1733 and was the first to examine metallic arsenic in detail. He was one of the first chemists to oppose alchemy.

Arsenic has had many uses: orpiment was used as a yellow pigment called "King's Yellow" and, mixed with quicklime, was used as a depilatory called "rusma" in the East. Arsenic was used as a hardening agent in the process of making lead shot, is used in rat poisons and in insecticides.

2.2.3 Gallium Arsenide

Gallium Arsenide was first grown and chemically characterised in 1926 by Victor Moritz Goldschmidt, Fig. 2.1. However it was not until the early 1960s that researchers first attempted to make devices with GaAs, see Table 2.1. In 1965 Carver Mead produced the first GaAs FET [Mead 66]. The GaAs FET began to find its first applications in the microwave area. The 1970s saw the first small scale integrated circuits but reliability was poor. The 1980s produced accessible and reliable GaAs foundries enabling realisation of large scale integration. By the late 1980s the scale of integration was sufficient for the contemplation of complex digital processing in GaAs. The early 1990s saw the first true VLSI GaAs designs. The present day quality and performance of GaAs has largely

displaced the preconceptions about GaAs reliability, though some stigma from the 1970s still remains. Today applications abound and we see GaAs in >90% of all mobile phones, in satellite communication systems, optical fibre transceivers, ATM switching in telecommunication systems and in high speed digital signal processing, to name a few. The main area where there has been a weakness is in the GaAs CPU – research programs in this area lead by SUN and CRAY, for example, have all collapsed. The key problem is the gate leakage current in the MESFET, making large CPU memory in GaAs difficult to realise. This may now change as the emerging CGaAs™ technology, using low leakage HIGFETs, begins to prove itself. The rate of growth in GaAs research has gathered considerable momentum and this is illustrated in Fig. 2.2, which shows a remarkable increase in the number of GaAs publications. Whilst the number of publications is not an ideal measure of the health of the GaAs industry, it does reflect the increasing effort going into research. The recent drop in publications (1993 & 1994), was probably caused by military *détente* and the latest increase in growth (1995) is most likely due to the rapid penetration of GaAs into communications systems.

2.3 Competing Technologies

Before we begin to discuss the special characteristics of GaAs over silicon, which make it particularly suited for the proposed imager, we briefly review some of the major competing technologies other than silicon, in order to highlight the importance of GaAs as the technology base for this work.

2.3.1 Diamond

There has been much interest in the idea of producing integrated circuits on diamond substrates. The characteristics of diamond appear to offer high power, high frequency and high temperature operation surpassing that of GaAs and silicon. However, diamond has a number of limiting problems, the worst being that only p-type crystals can be grown. Furthermore, [Collins 92] has identified a number of errors in the literature showing over-optimistic theoretical predictions for diamond. Another fundamental problem is that the p-dopant (boron) in diamond has a rather high activation energy (360 eV) at room temperature – this means that only 1-5% of the dopant is activated and thus the drain current is drastically low (in the order of nA) at room temperature. It should also be noted that the diamond technology is extremely immature, with only single transistors fabricated to date and with material that is still extremely difficult to grow with device

Figure 2.1: Victor Moritz Goldschmidt (b. Zurich 27th Jan. 1888, d. Oslo, Norway 20th Mar. 1947) was the only son of the distinguished physical chemist Heinrich Jacob Goldschmidt, who held professorships at Amsterdam, Heidelberg, and Oslo. His mother was Amelie Kohne. In 1926, he was the first to grow and report on the chemical properties of gallium arsenide, as part of a systematic study of many gallium compounds. Although he had no inkling of how gallium arsenide might be someday useful, the impact on engineering has been profound with applications in integrated circuits, solar cells, superlattices, quantum electronic devices, light emitting diodes and laser diodes. Of Jewish parentage he left for Norway in 1901 and obtained Norwegian citizenship in 1905. He matriculated in the same year at the University of Christiania. In 1911 he obtained his Ph.D degree, and in 1914 he became a professor of crystallography, mineralogy and geology. Even in the middle of his own absorbing work he was always glad to answer questions and give his opinion. His lectures were excellent and clear. In his judgement of persons and situations he would often exaggerate, and his reactions were sometimes violent and seemingly childish. But for his friends his great qualities overshadowed everything else. As a Jew, he suffered severely at the hands of the Nazis, was arrested twice, and in October 1942 was sent to a concentration camp, the whole of his property being confiscated. He carried a capsule of hydrocyanic acid for use as the final evasion of oppression. A university colleague in Oslo once asked Goldschmidt for a similar capsule. He replied, "This poison is for professors of chemistry only. You, as a professor of mechanics, will have to use the rope."

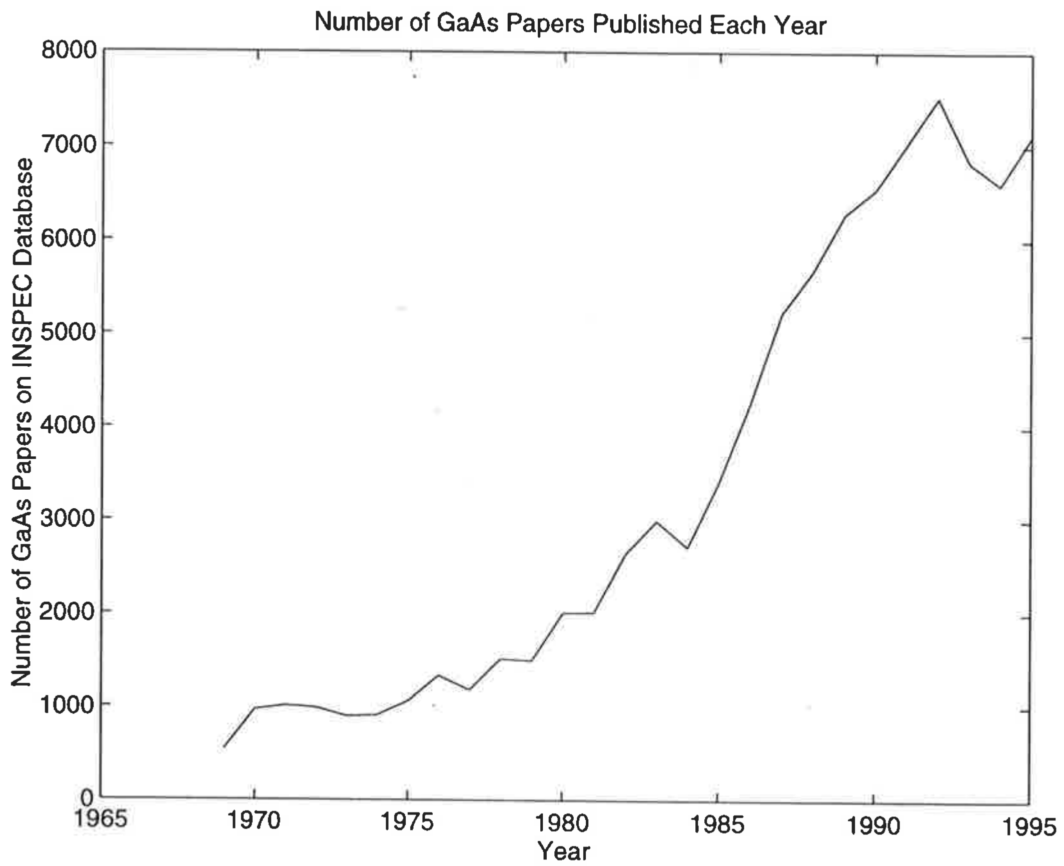


Figure 2.2: Number of published GaAs papers on the INSPEC data base. Only the keywords 'GaAs' and 'gallium arsenide' were searched for.

Name	Origin	Milestone	Date
Albertus Magnus	Germany	Isolated pure arsenic	c.1250
Jons Jacob Berzelius	Sweden	Isolated pure silicon	1824
Paul Émile Lecoq de Boisbaudran	France	Isolated pure gallium	1875
J. Czochralski	Poland	Invents CZ crystal growth method	1917
Victor Moritz Goldschmidt	Norway	First to grow & characterise GaAs	1926
Julius Lilienfeld	Germany	Files patent for copper sulphide FET	1930
Vereinigte Chemische Fabriken	Germany	First commercial gallium	1932
Oskar Heil	Germany	Files patent for FET	1935
Cetralab, Div. of Globe-Union	USA	First thick film circuit	1941
Sosnowski, Starkiewicz & Simpson	UK	MBE growth invented	1947
Bardeen, Brattain & Shockley	USA	Germanium point contact transistor	1948
G.K. Teal & J.B. Little	USA	Germanium bipolar junction transistor	1948
R.S. Ohl & W. Shockley	USA	Semiconductor ion implantation	1949
G.W.A. Dummer	UK	IC concept first described	1952
G.K.Teal & E. Buehler	USA	Silicon junction transistor	1952
N.H. Fletcher	USA	Interdigitated transistor	1954
Stanislas Teszner	Fr-Pol	First commercial germanium FET	1958
J.A. Hoerni	USA	Planar process	1959
J.S. Kilby	USA	First IC patent	1959
Various companies	USA	Birth of linear ICs	1960
Loor, Christensen, Kleimock & Theurer	USA	Epitaxial growth	1960
H. Nelson		LPE growth of GaAs	1961
J.W. Allen	UK	Concept of semi-insulator	1960
S.R. Holstein & F.P. Heiman	USA	First MOS IC	1962
N. Holouyak	USA	First practical GaAsP LED	1962
R.N. Hall et al	USA	Semiconductor laser	1962
M.I. Nathan & G. Lasler	USA	First GaAs pn junction	1962
Various	USA	Birth of TTL	1963
Carver A. Mead	USA	First GaAs FET	1965
Various	USA	Birth of CMOS	1968
Bell Labs	USA	E-Beam lithography	1974
A. Yaniv et al	USA	First GaAs OEIC	1978
Vittesse Semiconductor	USA	First semi-custom GaAs VLSI IC	1990
University of Adelaide	Australia	First full custom GaAs VLSI IC	1991

Table 2.1: Historical evolution of GaAs – a brief chronology.

quality.

2.3.2 Silicon Carbide (SiC)

SiC can occur in many polytype forms and over a 180 different polytypes have been identified. Polytype 6H is the most stable and has been the most studied. However, although less is known about polytype 4H, it has recently emerged as the favoured form of SiC – giving increased mobilities over 6H, by a factor of two, as shown in Tab. 2.2. Due to the good thermal conductivity, wide bandgap and large beakdown field, this wide bandgap semiconductor (WBS) material is targeted for power applications. For VLSI digital applications the mobility of Si and GaAs, still appear to be superior (see Tab. 2.2). SiC has only recently become commercially available to difficulty in finding a suitable growth technique. This together with its immaturity means that starting material

	Si	GaAs	SiC-6H	SiC-4H	GaN	Diamond
Crystal structure	diamond	zincblende	wurtzite	wurtzite	cubic	diamond
Lattice constant (Å)	5.43	5.65	4.36	4.36	4.51	3.57
Cost in 1995 ($/sq inch)	2.8	21	2200	>2200	200	2000
Energy gap (eV)	1.12 Indirect	1.42 Direct	2.86 Indirect	3.2 Indirect	3.4 Direct	5.6 Indirect
Melting point (°C)	1412	1238	2800	2800	1500	4000
Thermal conductivity at 300 K ($Wcm^{-1}K^{-1}$)	1.31	0.46	4.3	4.3	1.3	20
Breakdown field (kV/cm)	300	400	>1000	>1000	3500	5000
Relative permittivity	11.9	13.1	9.7	9.7	8.9	5.5
Effective electron mass	$0.92m_e$ (longitudinal) $0.19m_e$ (transverse)	$0.067m_e$	$1.5m_e$ (longitudinal) $0.25m_e$ (transverse)		$0.2m_e$	
Effective hole mass	$0.49m_e$ (heavy hole) $0.16m_e$ (light hole)	$0.50m_e$ (heavy hole) $0.076m_e$ (light hole)	$1.0m_e$		$0.8m_e$ (heavy hole)	
Refractive index	3.42	3.3	2.55		2.33	
Electron mobility at 300 K (cm^2/Vs)	1450	8500	500	1000	800	N/A
Hole mobility at 300 K (cm^2/Vs)	500	400	250	500	100(?)	500
Saturation velocity (cm/sec)	1×10^7	1×10^7	2×10^7	2×10^7	1.5×10^7	2×10^7

Table 2.2: Properties of Si, GaAs, SiC, GaN and Diamond.

prices are high and material quality is generally poor [Lande 96].

2.3.3 Gallium Nitride (GaN)

GaN is another example of an intensively studied WBS material. Parameters are shown in Tab. 2.2, though it should be noted that there is considerable uncertainty regarding GaN electronic transport parameters. A suitable substrate for expitaxial growth does not presently exist and an alternative substrate must be used. Sapphire starting substrates are generally used, however lattice and thermal matching with GaN are poor [Trew 95]. Latest work involves growing GaN on SiC substrates. This technology is still in the phase

of fundamental materials research and thus is too immature for consideration.

2.3.4 Polymer Electronics

Polymer electronics offers many exciting possibilities. LEDs, transistors, displays, detectors, integrated circuits can all be achieved with the promise of mechanically flexible circuits, by using doped polymers. Therefore, it is conceivable that a future television may be simply a thin flexible A4 sheet, which is portable and can be used in the bath! Polymers offer many advantages including better contrast displays, the ability to form new hetereostructures and the opportunity to design new materials with different bandgaps and electron affinities. In polymers, there are no dangling bonds at interfaces and thus high vacuum systems are not required for fabrication. Fabrication is further simplified in that spin coating replaces epitaxial growth. Apart from being an extremely immature technology, the disadvantages thus far include extremely low mobility and detector quantum efficiencies in the order of 1% [May 95].

2.3.5 Porous Silicon

Conventional bulk silicon has an indirect bandgap and therefore fundamentally cannot compete with the ability of GaAs to emit light and therefore promise full OEIC technology. However, porous silicon can emit light and, although this property is not fully understood, the possible threat is that this may enable silicon to compete with GaAs in the OEIC arena. However porous silicon is still very immature: stable operation for thousands of hours of operation, without degradation in quantum efficiency has not yet been achieved. Also modulation speeds in the hundreds of MHz range have not yet been attained. Monolithic ICs in porous silicon have not been achieved and progress is far off [Buda 94]. GaAs has already achieved monolithic circuits, integrated with both detectors and laser diodes.

2.3.6 Silicon-on-Insulator (SOI)

The main insulation techniques that are used in this technology are sapphire, SIMOX (Separation by IMplanted OXygen), DI (polysilicon handle Dielectric Isolation) or BESOI (Bond and Etchback SOI). Sapphire tends to have the a poorer quality semiconductor/insulator interface and DI suffers from large wafer warping $> 100\ \mu$m. Thus SIMOX has been the traditionally favoured choice. However as SIMOX suffers from defect densities as high as 10^4-10^5 defects/cm^2, the newly emerging BESOI technology appears to

be overtaking with only 10-100 defects/cm^2 but is immature and unproven in the arena of thin film SOI. For comparison traditional bulk silicon contains 1-5 defects/cm^2.

SOI shows promise in high-speed VLSI. There is high resistance to punchthrough, due to the insulating substrate. Low capacitive coupling between the inversion channel and the underlying film reduces the vertical electric field, with a consequent enhancement in mobility and an increase in drain saturation current, as compared to the equivalent bulk CMOS transistor. The major advantage of SOI, due to the insulating substrate, is radiation hardness.

However, a problem that emerges is the *floating body effect* which gives rise to a kink in the I-V characteristics of the transistor. This kink effect induces unwanted transients and can be reduced at the expense of lower breakdown voltage, increased lateral bipolar snapback effects and negative slope resistance effects.

If the silicon film thickness is reduced, so that it becomes fully depleted, the kink effect is eliminated. However, the drain depth is thereby reduced, giving rise to an increase in impact ionisation leading to reduced drain breakdown voltage. This places a premature limit on scaling. Holes from the impact ionisation also exacerbate the lateral bipolar snapback effect.

Also the thin film on insulator has thermal management problems in terms of reliability, ESD protection and local heating effects which cause a drop in mobility giving rise to negative slope resistance effects on the I-V characteristic.

Due to these various problems, SOI ICs have found only niche application areas where radiation hardness is of prime importance such as in SRAMs for satellites. Furthermore this technology would be unsuitable for the proposed imager as the trend is towards silicon thicknesses around 0.1 μm [Groeseneken 90], which would severely limit the depletion region volume for photocollection.

2.3.7 Silicon-Germanium (SiGe) Heterostructures

There has been considerable recent activity with SiGe technology and much speculation about whether it will soon displace GaAs.

Advances in conventional CMOS, over the years, have been based around device scaling. As submicron dimensions have been approached, further scaling of CMOS has become increasingly complex and fundamental limits will soon emerge. The significance of SiGe is that it is compatible with conventional silicon processing and offers improved material properties, so that continued advancement can be sustained.

Given that there are many alternative emergent technologies, other than SiGe, the question is 'why is SiGe presented in the literature as the main contender?' If we examine some of the main emergent technologies in Table 2.2 we see that in terms of cost and mobility, GaAs is clearly in the lead. SiGe becomes a closer competitor by virtue of the fact that costs are supposedly low, as it can be produced in existing silicon fabrication lines. Also SiGe attracts higher levels of R&D investment, as it rides on silicon funding sources. Unfortunately, it is difficult to enter a SiGe column in Table 2.2 – this is partly due to the fact that SiGe parameters will vary with mole fraction and particular heterostructure architecture, but is mainly because the open literature appears fairly silent for room temperature parameters. However, wherever sparse data is available, GaAs still remains superior. For instance, the effective electron mass of $Ge_{0.25}Si_{0.75}=0.264m_e$ [Sun 95] is 4 times heavier than in GaAs. Although, p-channel devices show higher mobility (around 1000 cm^2/Vs [Sadek 95]), peak electron mobility in n-channel devices (3000 cm^2/Vs [Sadek 95]) is still below that of GaAs.

There are two kinds of SiGe starting material: **Type I** is a strained layer of SiGe on a silicon substrate and **Type II** is a strained layer of silicon on SiGe.

As SiGe is a binary alloy, whereas III-V heterostructures employ ternary alloys, there is less control over tuning the lattice constant. [2] This, together with the 4% lattice mismatch between Si and Ge, means that SiGe is restricted to particular critical strained layer thicknesses depending on the mole fraction. It turns out for SiGe FET devices that the strained layer must be extremely thin (5-15 nm) and that it is susceptible to dislocations if processing temperatures go above 600°C. The trend is towards lower fabrication temperatures, but at the present state-of-the-art this means that yield is currently poor. Also limiting the thermal budget during processing often results in poorer oxide quality, incomplete dopant activation and poor passivation reflow resulting in reliability problems. Furthermore, field oxide growth and field implant activation requires higher temperatures and consequently SiGe is restricted to a mesa process [Konig 94] – this makes it difficult for SiGe to compete with planar processes for achieving VLSI densities.

Carrier transport takes place at the Si/SiGe heterointerface, however as gate voltage is increased parasitic conduction at the Si/SiO_2 interface becomes a problem in MOS devices.

Unfortunately, due to a fundamentally small conduction band offset in Type I devices, only p-channel MOS devices can be achieved. On the other hand, complementary MOS devices would be possible in Type II materials, however these suffer more from dislocations. There

[2] There have been recent attempts at attaining better control by adding carbon to obtain the SiGeC ternary combination. This may be successful, but work has only just started.

are no successful reports of fabrication of Type II MOS devices. Furthermore Type II is considered too complex for future mass production purposes [Voinigescu 93].

With Schottky gate devices, both Type I & II complementary devices are possible, however it turns out that the Schottky barrier height for the p-channel device is too low, resulting in high gate leakage current.

In summary, there are a number of unsolved problems preventing fully complementary SiGe technology. The technology is immature and it is suited to low temperature processing (600°C max.) that has not arrived yet as state-of-the-art. To put the maturity of SiGe into perspective: the first SiGe FET was reported in 1985 [Pearsal 85], whereas the first GaAs FET was documented in 1966 [Mead 66] – hence SiGe has a maturity lag of upto about 20 years! By contrast, complementary GaAs has now been reported with VLSI packing densities [Bernhardt 95] and densities of a million gates per chip are expected by the end of 1996.

Complementary MOS SiGe faces fundamental problems, however the complementary Schottky gate SiGe problem may be solved in the near future. In any case, the complementary GaAs (CGaAs™) HIGFET has emerged and has advantages over both MOS and Schottky SiGe. As the CGaAs™ HIGFET has a *semi-insulating* gate, it has a lower leakage current than Schottky SiGe devices and does not trap charge, as does the oxide in an MOS structure. Therefore the HIGFET has greater tolerance to radiation and hot electrons than the SiGe MOS device.

Furthermore, SiGe cannot compete with the GaAs semi-insulating substrate for reducing parasitic capacitances.[3] SiGe cannot presently compete with the OEIC capability of GaAs. In addition, the power added efficiency (PAE) and noise performance of microwave SiGe FETs cannot compete with GaAs. This means that GaAs is also the choice for mixed digital/RF mobile applications where compactness and hence a monolithic realisation becomes important, such as in the realisation of the single chip radio transceiver. Although it is often argued from an economics point of view that there is no market for the added expense of such a single-chip solution, the shift towards the mobile multimedia paradigm will make strict demands on compactness and thus the mixed digital/RF/optical capability of GaAs becomes even more attractive.

Another important factor with GaAs, that is often overlooked, is its inherent simplicity: no wells, no substrate ties, no latchup and over ten fewer masking layers than in BiCMOS. Conventional GaAs fabrication, such as through Vittesse, is now cheaper than BiCMOS. CGaAs™ also has over ten fewer masking layers than BiCMOS and a key feature is that

[3]SiGe/SOI is an option, but this would inherit the problems of both SiGe and SOI.

n and p-channel devices differ by only two implant steps.

SiGe HBT devices are the only SiGe devices showing signs of near-term success. However, GaAs MESFET power amplifiers show greater linearity and greater power-added efficiency. The SiGe HBT is essentially a performance enhancement of existing bipolar technology with all its attendant strengths and weaknesses. Also present SiGe technologies tend to be non-planar and use mesa isolation, to avoid the higher temperatures involved. Proposals to create a new BiCMOS comprising conventional CMOS with SiGe HBTs, will be faced with initially lower yields, greater complexity, greater number of masking layers and expense. As SiGe struggles to close the large gap in maturity against GaAs, facing all the economic and technical challenges described, GaAs in the meanwhile will have progressed further into the future.

2.4 The Semi-insulating (SI) GaAs Substrate

The word 'semi-insulator' is a somewhat confusing misnomer as it suggests an intermediate material between 'insulator' and 'semiconductor' on the horizontal axis of Fig. 2.3. The term is actually intended to describe a *semiconductor* with a reduced level of conductivity, but not as low as that of an insulator. A better term would be 'compensated semiconductor' as the reduced conductivity is a result of arranging a balance between shallow and deep levels called 'compensation.' In contrast to this a 'doped' semiconductor has increased conductivity due to a deliberate excess of donors or acceptors. So, in both cases, we have a *semiconductor* and the conductivity is a function of *impurity management*, as shown on the vertical axis of Fig. 2.3.

So a semi-insulator is in fact a semiconductor with low conductivity, due to a process called *compensation* and is not a function of bandgap as the name erroneously suggests. However, we shall continue to use the term 'semi-insulator' or 'SI' due to its widespread use and acceptance. The term should be understood in terms of *level* of conductivity rather than *type* of material in terms of bandgap.

There are two common forms of SI GaAs, namely GaAs:Cr and GaAs:EL2. GaAs:Cr is referred to as *intentionally doped* SI GaAs because chromium (Cr) impurities are deliberately introduced to compensate unavoidable silicon shallow donors in the substrate. By contrast, GaAs:EL2 is referred to as *unintentionally doped* SI GaAs as EL2 is a deep donor that already exists in the substrate, ie. is not externally introduced. EL2 is a native deep donor that compensates the unavoidable carbon shallow acceptors (10^{15} /cm^3) found in most wafers produced by present-day growth techniques.

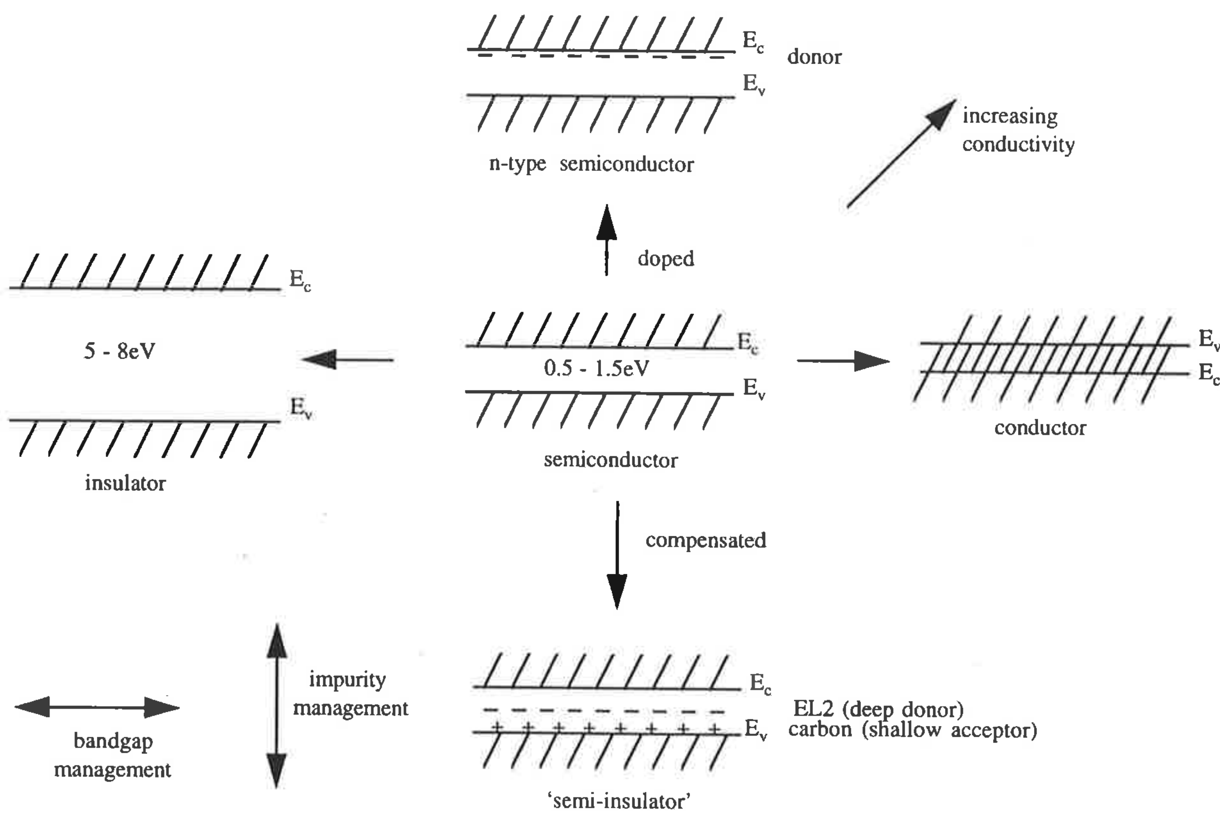


Figure 2.3: Understanding the meaning of 'semi-insulator.'

Modern GaAs wafer technologies now almost exclusively use GaAs:EL2 substrates and GaAs:Cr is largely outmoded. The reason for the fall in favour of the GaAs:Cr substrate was one of lack of controllability in that Cr was not thermally stable and hence out-diffused during processing.

EL2 has the property of increasing in concentration with increasing arsenic (As) concentration. Hence modern SI GaAs is grown slightly As rich to produce enough EL2 to compensate the carbon acceptors – this can result in SI wafers that are very slightly n-type.[4] Deep levels such as EL2 have a high activation energy and hence do not fully ionise at room temperature. The carrier concentration (10^8 /cm^3) in GaAs:EL2 is therefore extremely small. This results in typical SI substrate resistivities around 10^8 Ω/cm, in contrast to around 1 Ω/cm for typical CMOS substrates.

What is EL2? There has been a lot of controversy on the nature of EL2, however it is now accepted that EL2 is an As antisite defect. The controversy now centres around the smaller detail of whether it is an isolated defect or is associated with an interstitial or another vacancy.

During wafer production, EL2 is formed at 900°C, then quenched at 1200°C and then reformed at 800-900°C. This increases the wafer resistivity presumably because other defects are not reformed. It is strangely paradoxical that a defect such as EL2 is a 'mixed blessing' and is harnessed to provide the useful semi-insulating property. Similarly, oxygen impurities in GaAs are a 'mixed blessing' in that they are essential in the process of gettering to produce a denuded zone.

There is much consternation in the literature about the terminology used when discussing EL2. EL2 itself is not compensated – it is simply a deep donor and because of its depth into the energy band, it will not contribute electrons to the conduction band under normal conditions. EL2 is also *not* a trap; it is a deep donor. Some fraction of EL2 will be ionised by the excess shallow acceptor population – this is the compensation mechanism mentioned above. The ionised EL2 (called EL2+) does act as a trap, and is the primary trap of interest in most of the parasitic effects in GaAs, such as backgating, drain current lag, frequency dispersion of output conductance etc. Manufacturers are constantly improving the extrinsic defect populations while maintaining a slight excess of acceptors. The trend is now to reduce the EL2 concentration, although the EL2+ concentration goes down with $(N_a - N_d)$. Whether or not the reductions in N_a, N_d and consequently EL2+ now being achieved do provide the expected improvements in trap related problems should be

[4] It has been reported that sometimes SI wafers can turn out slightly p-type possibly due to carbon acceptors being fully ionised, whereas EL2 donors are not fully ionised, at room temperature. Unless stated otherwise, this work refers only to our test samples which have slightly n-type bulk substrates.

known in the near future.

Another source of confusion in the semi-insulating (SI) substrate is the nature of the depletion region. Referring to the concentration plot of electrons across an n/SI hi-lo junction, Fig. 2.4, we see that the concentration in the bulk is lower than in the depletion region! There is no consensus in terminology: some workers call the the whole of the bulk 'depleted' whereas some prefer to call the region of band bending the 'depletion region.' Strictly speaking the former is more correct, however the later is more useful as we tend to think of a depletion layer as that region which supports an electric field. Therefore, throughout this work, we shall use the term 'depletion region' as signifying the area of a junction that supports an electric field, corresponding to the band bending region. After all 'depletedness' is relative and we see in Fig. 2.4 that the n concentration curve is very similar to that in silicon, except that in silicon the presence of the p region distracts us from thinking that there is an apparent paradox.

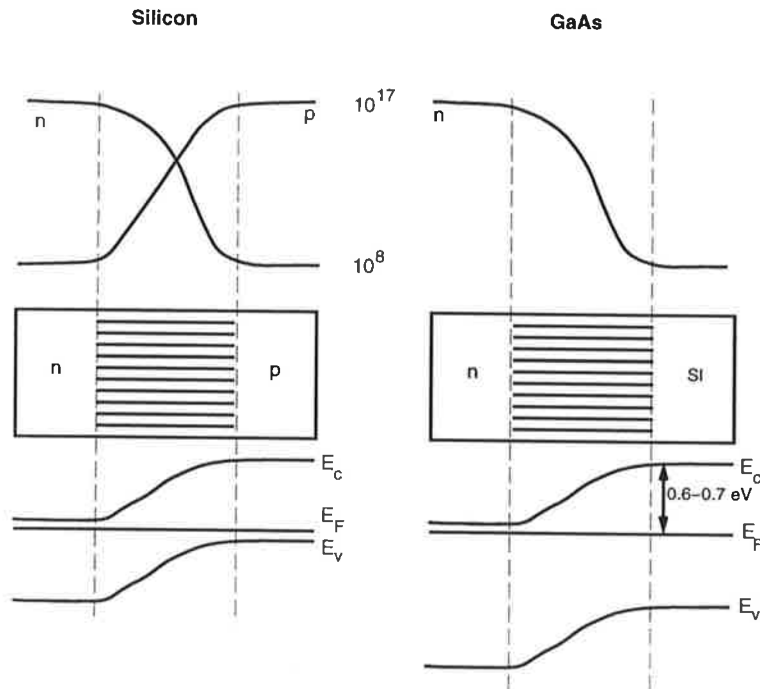


Figure 2.4: Carrier concentration plots, electric field lines and band bending diagrams for a silicon pn junction an a GaAs hi-lo n/SI junction.

Fig. 2.4 highlights another important feature of GaAs: the Fermi level is permanently pinned 0.6-0.7 eV below the conduction band, due to the effect of surface states. Therefore the surface cannot be inverted. This explains why all GaAs MESFETs are depletion devices and do not enhance an inversion layer as for MOSFET channels. A GaAs 'en-

hancement' device is therefore a misnomer - it gets its 'normally-off' characteristic by making the channel thin so that it is pinched-off by the built-in Schottky gate depletion region.

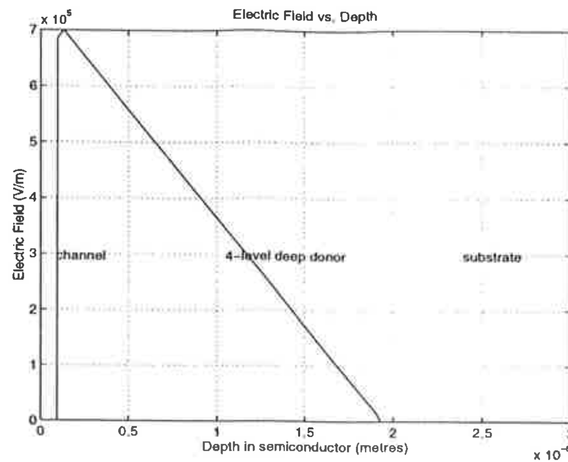


Figure 2.5: Channel/substrate hi-lo junction field solved by FDM using the 4-level deep donor compensation model due to George *et al.*

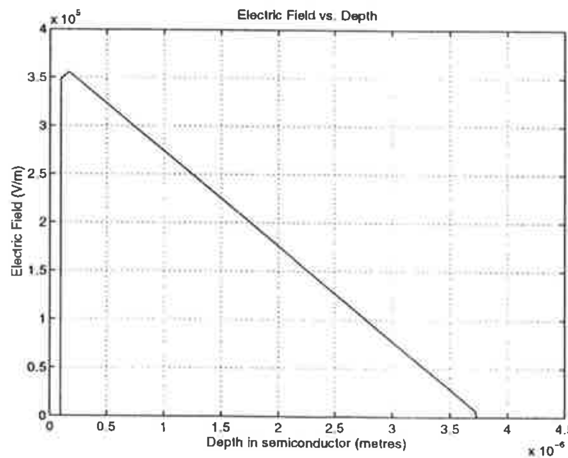


Figure 2.6: Channel/substrate hi-lo junction field solved by FDM using the 2-level deep donor compensation model due to McGregor *et al.*

The difference in depletion layer thickness between the GaAs and silicon junctions in Fig. 2.4 has not been indicated as this depends on specific factors such as doping and bias. However at zero bias we can typically expect a depletion thickness around 0.3-0.5 μm for silicon and 0.5-2.0 μm for GaAs. The larger spread of values in GaAs is due to the

number of different models available. For instance, if we use a 4-level model [George 89] using deep donor compensation we can obtain the field plot given in Fig. 2.5 and a 2-level model [McGregor 92] gives the plot in Fig. 2.6 showing a shorter penetration of field and hence indicating a thinner depletion region. These plots were obtained using a numerical finite difference method (FDM) detailed in Appendix A. Although analytical solutions were possible, this FDM solution sets the groundwork for researching more complex models in the future. A complete analysis of all the models is beyond the scope of this present work. The unified defect model (UDM) [Spicer 85] is regarded as the state-of-the-art model and we use this in later chapters as embodied within the commercial G-PISCES-2B software package.

2.5 GaAs Characteristics

Various key concepts such as bandgap, dark current, absorption coefficients etc. are considered and their impact on the feasibility and application areas of a GaAs imager are discussed.

2.5.1 Bandgap

Gallium Arsenide has a *direct bandgap*, whereas silicon has an *indirect* bandgap. This means that energy level transitions in silicon, as shown in Fig. 2.7, take place with momentum change – whereas energy level transitions, in GaAs, can take place with momentum conservation, allowing photon emission.

In a direct band gap semiconductor like GaAs, radiative recombination is efficient because energy and momentum are conserved in the transition. In silicon, however, electrons and holes do not have the same momentum. Radiative recombination can only take place if a third particle, such as a phonon, participates in the process. This means that light emission is a million times less likely in indirect bandgap semiconductors. Note that radiative recombination takes place on a nanosecond timescale in GaAs and on a millisecond timescale in silicon. Because, electrons and holes take so long to convert to photons in silicon, competing recombination mechanisms involving defects take over. Hence only a small fraction of the electrons and holes recombine to produce light. The quantum efficiency for conversion to light, in silicon, is typically 0.0001%.

Therefore GaAs is a good emitter of light, as well as a good receiver – whereas silicon cannot efficiently emit light. The ramification of this is that our study of the GaAs MESFET

as an imager element could well produce spin-offs in the area of optical communications between GaAs circuits. This is particularly important as there is the possibility of integrated optical wave guides on the GaAs substrate, forming compact ultra-high speed interconnects between circuit modules. This optical communications potential is relevant for future high-speed electronic systems.

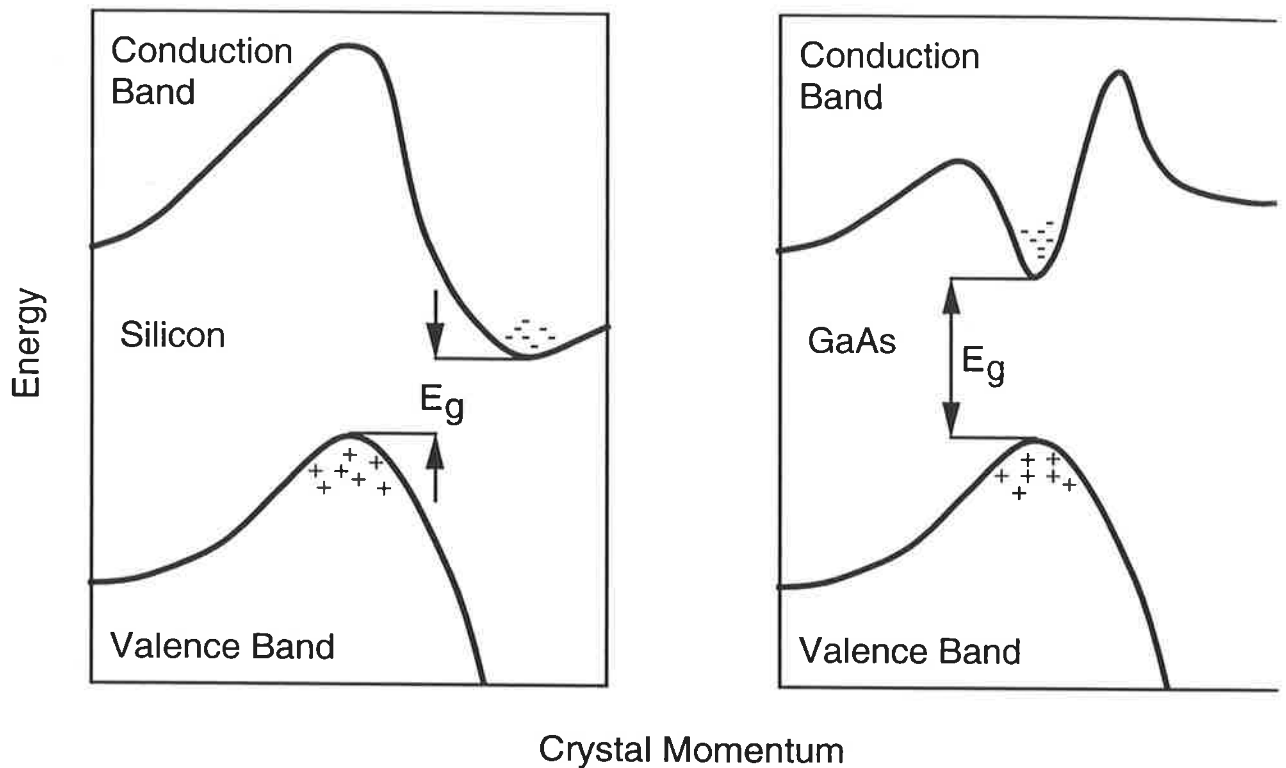


Figure 2.7: Energy Band Structures of GaAs v. Si

Notice the valence band shape in Fig. 2.7 for both GaAs and silicon is almost the same. This explains why the hole mobilities in both GaAs and silicon, as indicated in Table 2.2, are very similar. This has held back the development of a 'CMOS' style process for GaAs – however, this has at last emerged using pseudomorphic HIGFETs, where hole mobilities are successfully increased by lattice strain techniques.

2.5.2 Speed

At first sight the speed advantage of GaAs appears to be irrelevant, as imagers generally operate in the MHz range. However, the significant point is that the speed of gallium arsenide is required by the complex image processing algorithms that are becoming more

prevalent in vision systems. Furthermore, a gallium arsenide imager will offer the possibility of integration with the GaAs processing circuitry hence reducing cost, weight and increasing interconnect reliability. Alternatively, a high-speed imager, itself, could be realised for non-TV rate applications, such as in high-speed machine target tracking.

2.5.3 Radiation Hardness

The assertion that a good detector of light, such as a gallium arsenide imager, is also a good radiation hard circuit, appears at first sight paradoxical. However there is no paradox when we clarify the meaning of 'radiation hardness'. We mean a tolerance against *permanent* damage caused by radiation, brought about by the fact that a MESFET does not contain a sensitive gate oxide – but tolerance to transient radiation effects is only marginally better than silicon.[5]

This tolerance to permanent radiation damage is particularly important for the reliability of aerospace systems that do not have the protection against radiation offered by the Earth's atmosphere. The use of gallium arsenide circuits potentially offers the promise of lighter aerospace vehicles via the reduction or even elimination of metallic anti-radiation cladding necessary for present-day circuits.

2.5.4 Dark Current

Dark current refers to unwanted current present in the imager, caused by thermally excited carriers jumping a potential barrier – ie. the MESFET Schottky barrier, in our case. Dark current manifests itself by adding spatial noise and blemishes to the final image. For a Schottky barrier device, the dark current is given in A/cm^2, by the following equation:

$$I = A^*T^2e^{-\phi_B/kT}$$

– where, A^* is the *effective Richardson emission constant* (A/cm^2/K^2) neglecting the effects of optical phonon scattering and quantum mechanical reflections, T is the temperature (K), k is Boltzmann's constant and ϕ_B is the Schottky barrier height.

The barrier height is approximately 0.8 eV, for both the proposed GaAs MESFET imager and a typical n-Si Schottky barrier such as PtSi, so there is no apparent advantage.

[5]Note that new developments have recently shown a dramatic improvement in transient radiation hardness for GaAs. This has been achieved by using an LT GaAs buffer layer. The low temperature (LT) buffer layer exhibits extremely short carrier lifetimes in the 200-1000 fs range.

However, the advantage of GaAs is in the lower A^* at 8 $A/cm^2/K^2$, as opposed to 110 $A/cm^2/K^2$ in silicon, for electrons.

This means that the dark current in a GaAs imager could be about 10 times lower than in a similar silicon imager. However, in practice some of this gain may be lost due to higher defect densities in GaAs, giving rise to increased dark current. Nevertheless, superior dark current characteristics are presently obtainable and this is evidenced by the successful realisation of room temperature GaAs X-ray detectors, whereas silicon X-ray detectors require cooling [Beaumont 93, Bencivelli 95, McGregor 92, Sumner 92].

In conclusion, we have indicated that GaAs can exceed silicon in terms of dark current performance and that this may increase further as the technology progressively matures. This is relevant for aerospace applications at low light levels (eg. star tracking), where the long integration times required cause the image to saturate with dark current. This is overcome in silicon devices, by cooling (eg. to 77K) and thus GaAs offers the potential of more convenient operating temperatures. Obviously the removal of a cooler reduces the cost, weight and volume of the imaging system.

2.5.5 Absorption Coefficients

An analysis that compares the optical absorption coefficients of GaAs and silicon, shows that GaAs is optically the superior material for use in photodetection and imaging.

The governing equation for the penetration of light into the material is given by Lambert's law of absorption (also known as Bouguer's or Beer's law),

$$\Phi(x) = \Phi_o e^{-\alpha x}$$

– where α is called the absorption coefficient. Comparison of the absorption coefficients of GaAs and silicon against wavelength (λ), clearly highlights the superiority of GaAs.

Absorption Coefficients		
$\lambda(\mu m)$	GaAs $\alpha^{-1}(\mu m)$	Si $\alpha^{-1}(\mu m)$
0.4	0.01	0.1
0.6	0.20	2.0
0.8	0.82	12.0

Table 2.3: Absorption coefficients

The absorption length or penetration is given by α^{-1}. So for instance, the percentage of light absorbed within two absorption lengths, ie. $2\alpha^{-1}$, is $100(1-e^{-2}) = 86\%$. Lambert's law is graphically illustrated in Fig. 2.8 for GaAs.

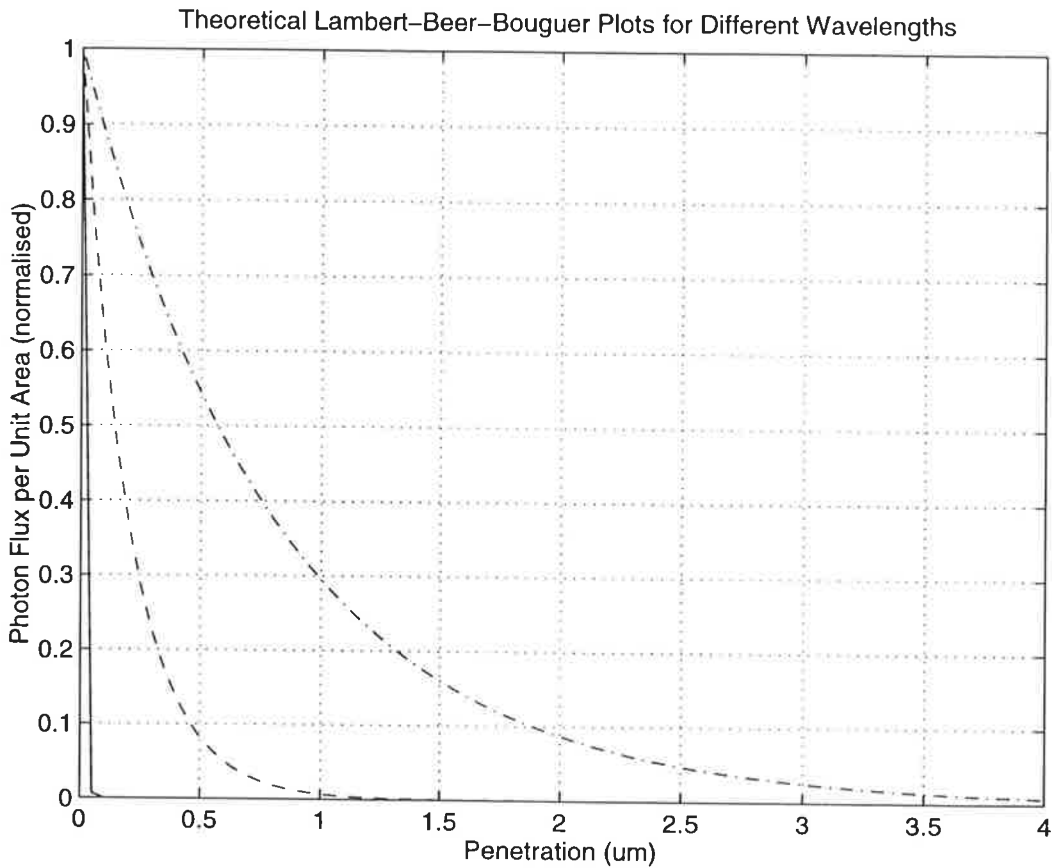


Figure 2.8: Lambert's Law of Absorption for GaAs. Solid line: $\lambda = 0.4\ \mu$m. Dashed line: $\lambda = 0.6\ \mu$m. Chained line: $\lambda = 0.8\ \mu$m.

To understand what this means, from Table 2.3 we deduce that, at $\lambda = 0.6\ \mu$m, 86% is absorbed within 0.4 μm for GaAs & within 4 μm for Si.

Yet again we see a key GaAs parameter with an order of magnitude improvement over silicon. The implication is that GaAs is a more efficient photocollector as most of the light is detected near the surface, where transistor action collects the photocharge. Improved spatial resolution is expected, as a result, and fewer wasted carriers imply improved quantum efficiency and hence greater responsivity.

The absorption length dependence on wavelength, in GaAs, is displayed in Fig. 2.9. The positions of the cut-off wavelengths in GaAs and Si, are indicated showing that GaAs is closer to the peak response of the human eye and is better placed with respect to the visible

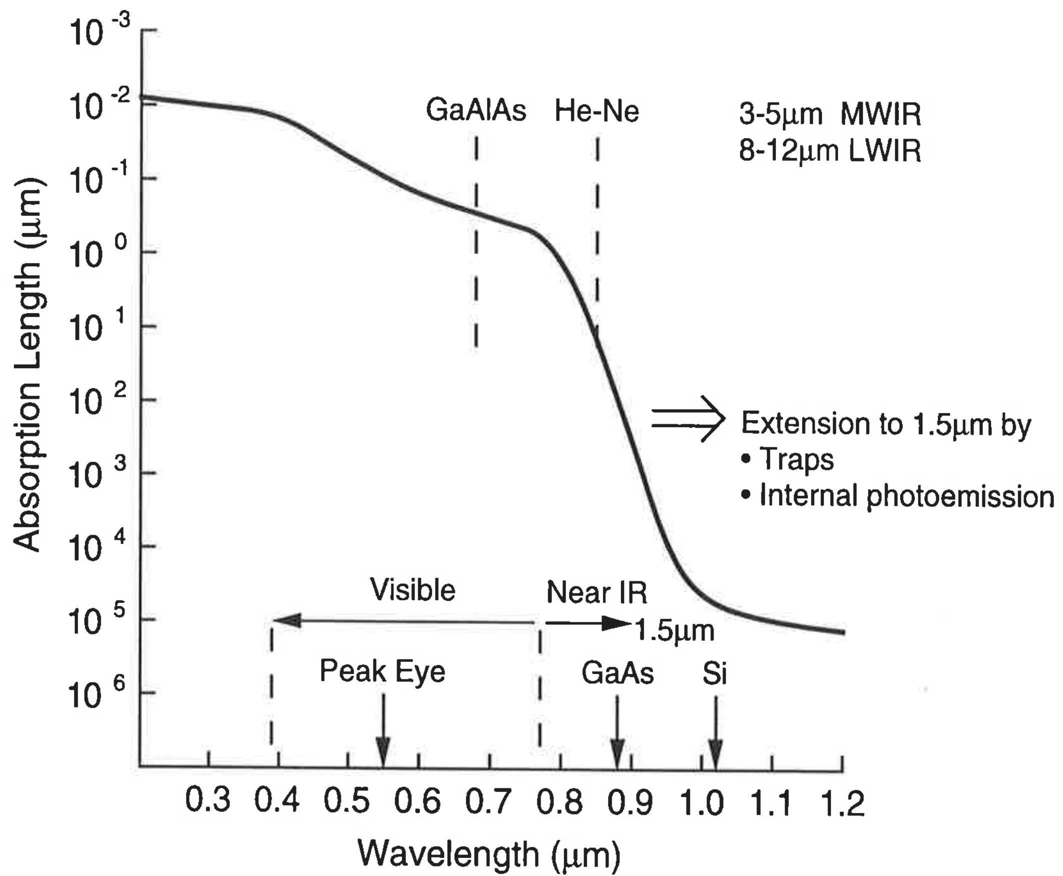


Figure 2.9: GaAs Absorption Length Dependence on Wavelength

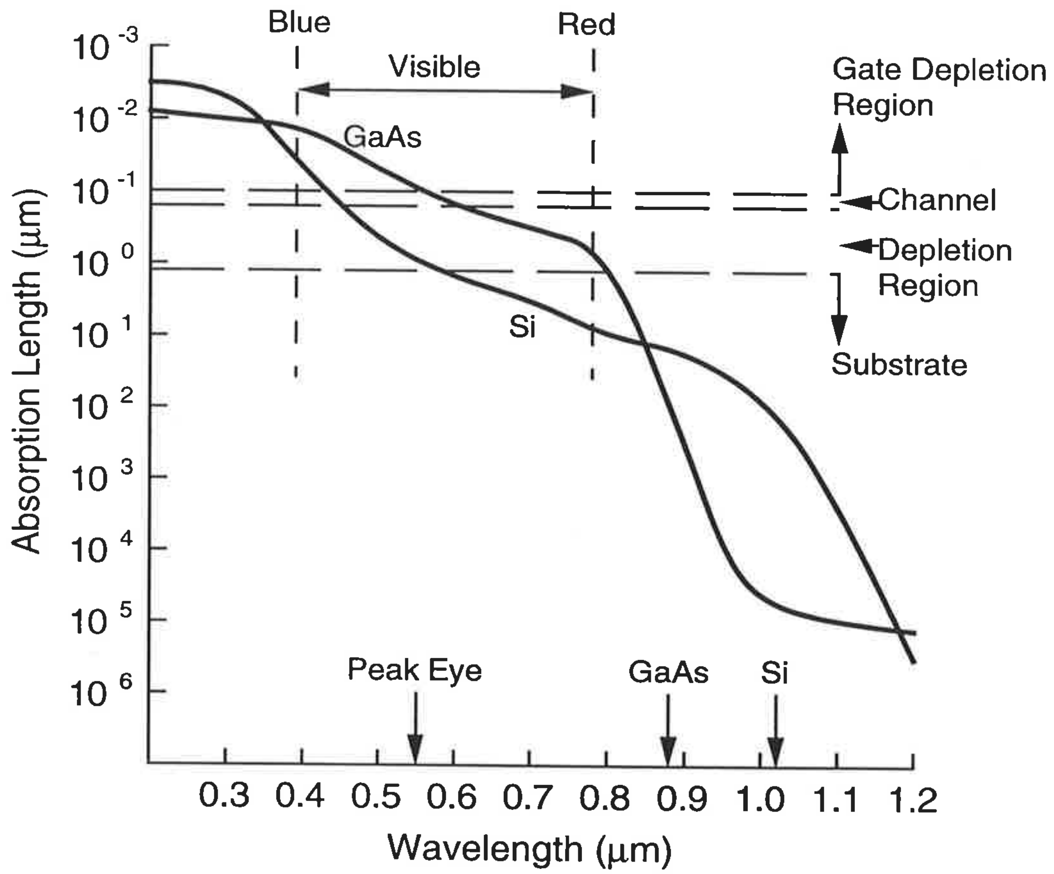


Figure 2.10: GaAs v. Si Absorption Length Comparison

spectrum. Indeed, at 5.5 μm, the peak eye response, the absorption length for GaAs is precisely in the channel region – whereas for silicon, absorption at this wavelength, occurs well into the substrate where many carriers recombine and thus are thus not collected.

The relative positions of GaAlAs and red He-Ne lasers are indicated on the curve showing that GaAlAs is suited to investigations at channel-substrate interface depths, whereas red He-Ne lends itself for observing substrate effects.

Fig. 2.9, clearly shows that neither GaAs nor silicon can intrinsically achieve sensitivity in the standard 3-5μm MWIR or 8-12μm LWIR infrared bands.

However, a GaAs imager can be extended to operate in the Near IR waveband, as indicated in Fig. 2.9. The introduction of states in the forbidden gap, introduced intentionally by impurities such as chromium, provide 'stepping stones' for electrons to traverse the bandgap via excitation from these lower energy wavelengths. This technique is well known in silicon infrared imagers, where the impurity used is typically indium. Such imagers are called *extrinsic* detectors.

As indicated, an alternative scheme to achieve Near IR sensitivity, is to utilise the *internal photoemission* effect. Internal photoemission occurs when the incident light frees electrons from the surface of the gate metal, which are then collected if they have enough energy to jump the gate Schottky barrier. Now, the Schottky barrier height for a MESFET is about 0.8 eV, whereas the GaAs and Si bandgaps are 1.4 eV and 1.1 eV, respectively – therefore less energy is required for an electron to jump the Schottky barrier than to jump a bandgap. Thus internal photoemission provides sensitivity to longer wavelengths of lower energy, namely in the Near IR band.

Commercial Si imagers, using internal photoemission, are available on the market – translating this scheme over to GaAs is beyond the scope of this work, but would make a fine basis for useful future proposals.

Fig. 2.10 highlights the superiority of GaAs over Si, for visible band detection, showing clearly that in GaAs the entire visible spectrum is absorbed close to where transistor action occurs, whereas for silicon the longer visible wavelengths are absorbed well into the substrate. Electrons that are generated by this absorption, in the silicon substrate, either recombine (and are therefore wasted) or diffuse to the channel region causing an unfortunate degradation in spatial resolution.

Note that size and position of the channel/substrate depletion region, in Fig. 2.10, is for a GaAs MESFET. For a typical silicon MOSFET, the depletion region will be a little smaller – this further highlights the superiority of the GaAs MESFET for photodetection.

2.5.6 Atmospheric Bands

Examination of the transmission of different wavelengths through the Earth's atmosphere reveals that some infrared wavelengths are blocked by the presence of water, CO_2, aerosols etc. The Near IR band has a large transmission widow centred at 1.4 μm and is therefore a potentially useful band. This is important in the area of astronomy, where clouds of cosmic dust tend to have strong emission in this band. This underscores the usefulness of future proposals for Near IR detectors.

Other useful and very popular infrared transmission windows are the standard 3-5 μm Medium Wave IR (MWIR) and 8-12 μm Long Wave IR (LWIR) bands. Present-day silicon imagers achieve MWIR capability via reducing the PtSi Schottky barrier height by using p-Si instead of n-Si substrates. There is scope for investigation of p-GaAs Schottky barrier imagers, in future proposals – changing the MESFET process would be an additional challenge and expense.

Silicon can achieve LWIR performance via the use of IrSi Schottky barriers – unfortunately, a major drawback is the requirement to cool to 50 K. Therefore there is scope for future studies to investigate more exotic GaAs Schottky contacts, that perhaps do not require such stringent cooling – this, of course, implies the expense of GaAs process changes.

Another technique employed, to exploit the MWIR and LWIR atmospheric windows, is the hybridisation of materials such as InSb or CMT on top of silicon circuitry. There is potential for applying this scheme to GaAs, with an advantage over GaAs Schottky barrier imagers, in terms of quantum efficiency – though this improvement would have to be weighed against the increased complexity, expense and uniformity degradation of the hybrid approach. At this early stage it is difficult to predict which direction future proposals should take, in this regard, but by judging from the recent surge in PtSi detectors, it would seem commercially wise to try to steer future GaAs proposals towards the cheaper monolithic Schottky barrier approach – rather than the complex hybrid approach.

2.5.7 Quantum Efficiency Formulae

Before we can compare the quantum efficiencies of GaAs and silicon imagers we must first carefully discuss a few esoteric subtleties.

Firstly, it should be noted that minority carriers in silicon substrates have diffusion lengths much larger than the pixel size. Hence, in order to reduce pixel crosstalk, silicon imagers are all exclusively fabricated on epitaxial substrates. The substrate under the epi-layer

is heavily doped so that stray carriers quickly recombine and do not find their way to neighbouring pixels. Unfortunately, the quantum efficiency is thereby reduced, as a large number of carriers are forced to recombine in this way.

Due to the unusual hi-lo junction[6] (n+/SI) situation present in the GaAs imager, it is the *majority* carriers that are being collected – hence a bulk substrate is appropriate for a GaAs imager. Therefore, for our GaAs case, it is appropriate to use the formula for quantum efficiency that approximates the bulk substrate to a semi-infinite slab giving (see Appendix B).

$$\eta_{bulk} = T(\lambda)(1 - \frac{e^{-\alpha W}}{1 + \alpha L_o}). \tag{2.1}$$

By contrast, in a typical silicon imager it is the *minority* carriers that are collected and there is a quantum efficiency degradation due to minority carriers recombining before they are collected. As this situation does not occur in our GaAs scenario, we may simulate this by setting the minority carrier diffusion length $L_o = 0$ in Eqn. 2.1. [7]

For the epitaxial silicon case, the semi-infinite slab assumption of Eqn. 2.1 is no longer valid as the epi-layer is thin and of the order of h = 15-25 μm. In Appendix B, we show that if we consider a finite slab, the appropriate quantum efficiency formula for epi is

$$\eta_{epi} = T(\lambda)(1 - \frac{e^{-\alpha W}}{1 + \alpha L_o} - \frac{\alpha L_o e^{-\alpha W}}{\alpha^2 L_o^2 - 1} \frac{e^{-h/L_o} - e^{-h\alpha}}{\sinh(h/L_o)}). \tag{2.2}$$

If we let

$$\eta_+ = T(\lambda)\frac{\alpha L_o e^{-\alpha W}}{\alpha^2 L_o^2 - 1} \frac{e^{-h/L_o} - e^{-h\alpha}}{\sinh(h/L_o)}$$

we see the remarkable result that $\eta_{epi} = \eta_{bulk} - \eta_+$, where η_+ is the quantum efficiency of the highly doped substrate beneath the epitaxial layer. Notice that this formula assumes that recombination in the highly-doped substrate is instantaneous – this approximation is accurate for most practical cases. The ability to separate η_{epi} into two clearly identifiable terms is significant, as the modulation transfer function (MTF) can then also be separated and the two terms can be analysed separately in a physically meaningful way.

[6]Note that our samples do not have a p-buffer layer and that the SI substrate is slightly n-type and hence we have a hi-lo junction situation.

[7]This equation should not be confused with the expressions for quantum efficiency that will be developed in a following chapter on the 'edge effect' which considers the collection of holes by the gate, producing a gate photocurrent. We are now considering the more general case of electron collection producing channel photocurrent.

Both Eqn. 2.2 and the realisation that it can be separated into two physically identifiable terms is an original contribution and has not been reported in the literature. The quantum efficiency formulae are generally poorly presented and ill-conceived in the literature. For instance, the latest attempt at a formula for η_{epi}, which can be found in [Theuwissen 95], is clearly dimensionally incorrect, does not have separable terms and under some conditions produces values greater than unity! Note that our expression (Eqn. 2.2) has all the expected features. Each term is dimensionless – ie. each argument is a ratio of two lengths or an absorption coefficient times a length. This is an important feature which is missing in erroneous attempts found in the literature. Also for $h \to \infty$ our expression reduces to the semi-infinite case. Our function is also well-behaved in that quantum efficiency goes down for increasing absorption length, as expected, and goes up for increasing diffusion length. Another important feature is that, by inspection, Eqn. 2.2 always stays below unity.

There is also some confusion in the literature over Eqn. 2.1. Our equation agrees with the derivation in the seminal work of [Seib 74] and with a fair body of literature. However, there is an erroneous formula due to [Barbe 75] that appears to have insidiously propagated, unchallenged, into some of the literature, such as [McCaughan 79]. Barbe's expression is in terms of MTF and therefore we plot the MTF expressions for both Barbe and Seib in Fig. 2.11, showing that Barbe's expression is clearly too optimistic.

2.5.8 Quantum Efficiency & Responsivity

Using Eqns. 2.1 and 2.2 we can now analyse the quantum efficiency and hence the responsivity in both the silicon and GaAs devices – these formulae were used to plot Fig. 2.15.

At the red end of the spectrum, ie. 8 μm, we find that the GaAs device has about twice the quantum efficiency of a comparable silicon device. Initially, this advantage will be lost due to the opaque gates in GaAs, but the potential is there for future investigation into transparent ITO gates. In our present device, the effect of transistor finger spacing on the quantum efficiency of the GaAs device, needs to be quantified and can be investigated experimentally. A series of test structures with various spacings would enable further maximisation of the quantum efficiency.

Responsivity is related to the quantum efficiency by,

$$R_\lambda = \frac{q\lambda}{hc}\eta \quad \text{A/W.}$$

For all wavelengths, the total is,

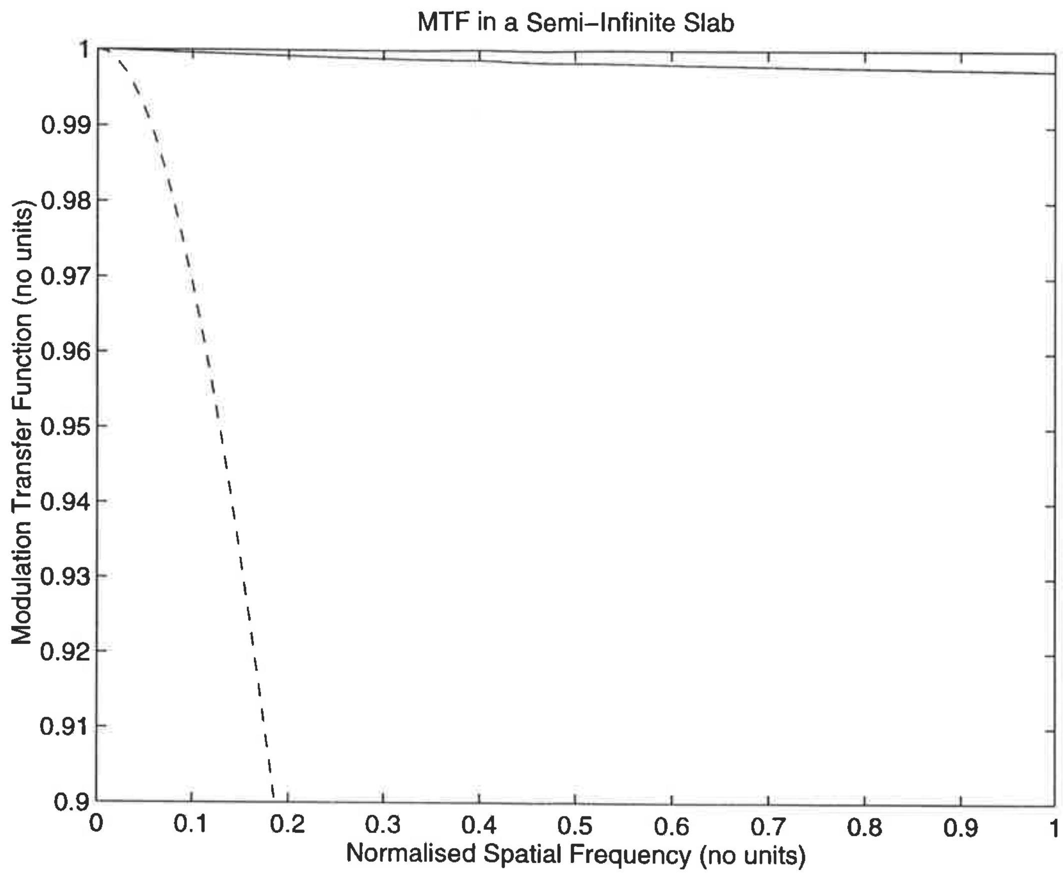


Figure 2.11: MTF due to diffusion. A comparison of the traditional formula due to Seib (dashed line) and the formula due to Barbe (solid line).

$$R = \frac{\int_0^\infty R_\lambda W_\lambda d\lambda}{\int_0^\infty W_\lambda d\lambda}$$

where, by Planck, radiated power,

$$W_\lambda = \frac{2\pi c^2 h}{\lambda^5 (e^{hc/\lambda kT} - 1)} \quad \text{W/m}^3$$

where this curve is illustrated in Fig. 2.12.

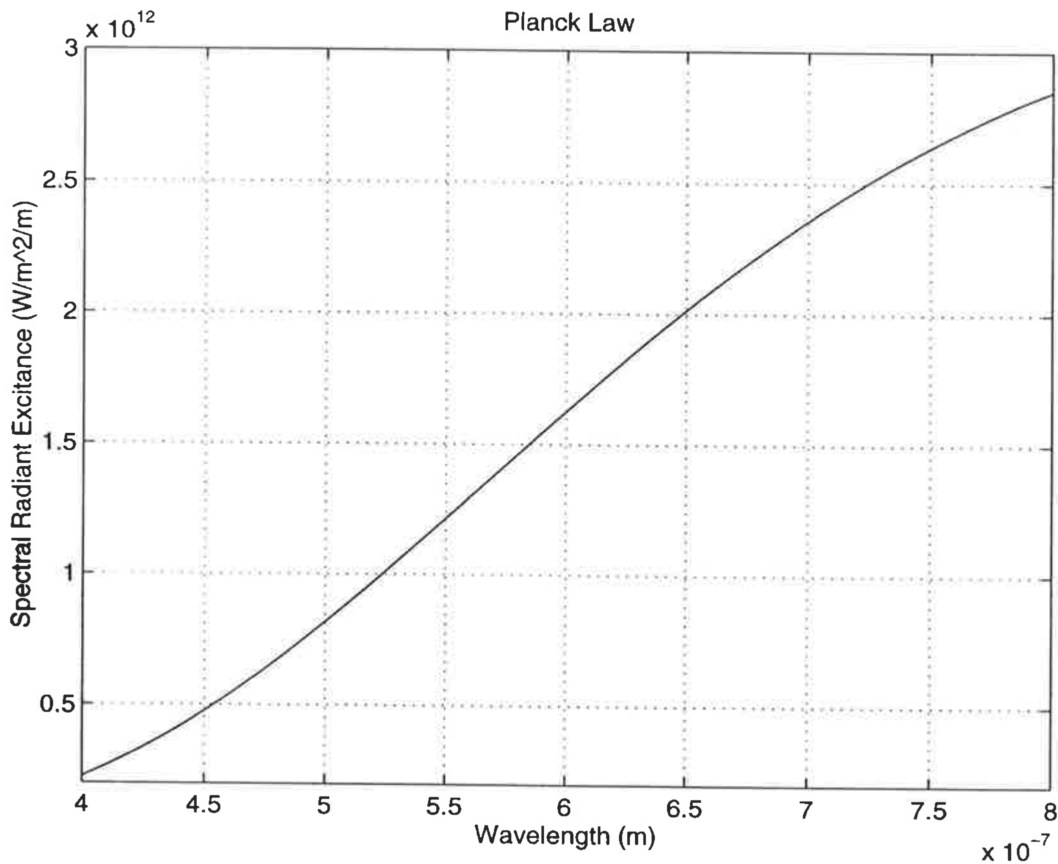


Figure 2.12: Planck's law curve, W_λ. Area under curve for visible wavelength range is 5.31×10^5 W/m^2. Total area under the curve for all wavelengths, given by Stefan's Law, is 46×10^5 W/m^2.

For reference, a silicon MOS CCD, has typically, $R \approx 50$ mA/W (3000 K tungsten source). We expect to exceed this with the GaAs imager and values are predicted as follows.

Figs. 2.13 & 2.14 show measured absorption lengths for silicon and gallium arsenide. From these curves the predicted quantum efficiencies are plotted in Fig. 2.15.

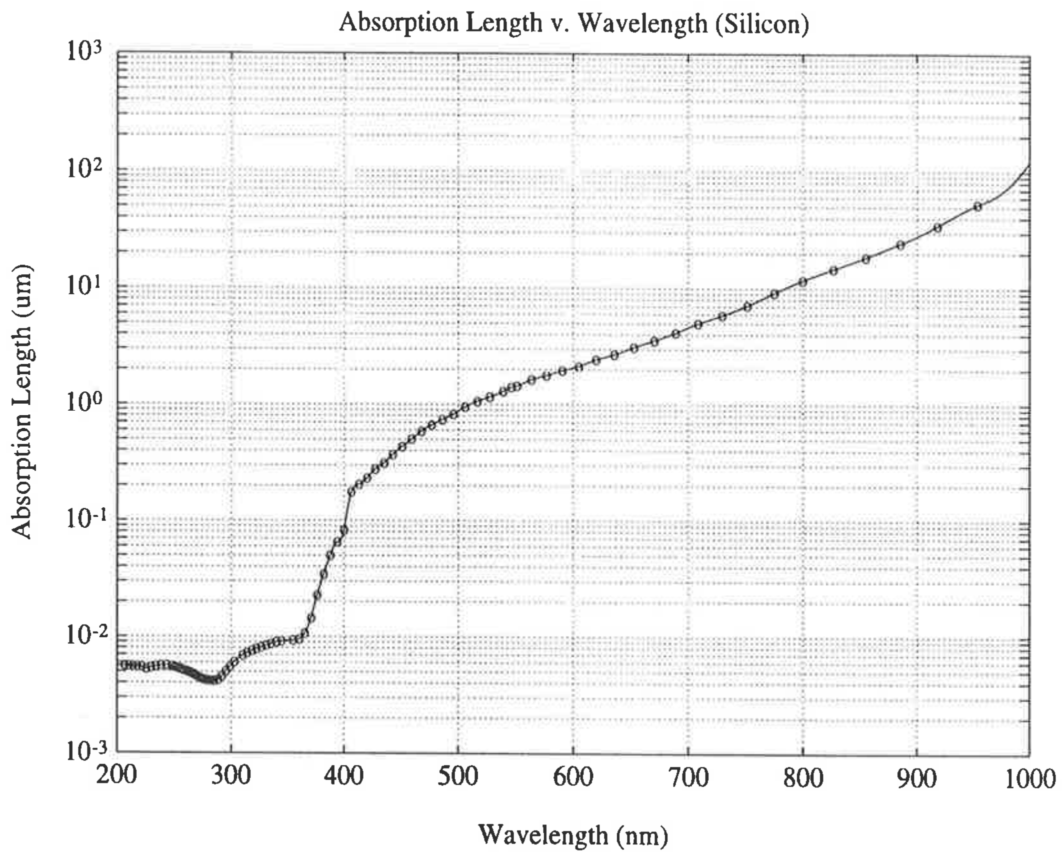


Figure 2.13: Measured absorption length against wavelength for silicon.

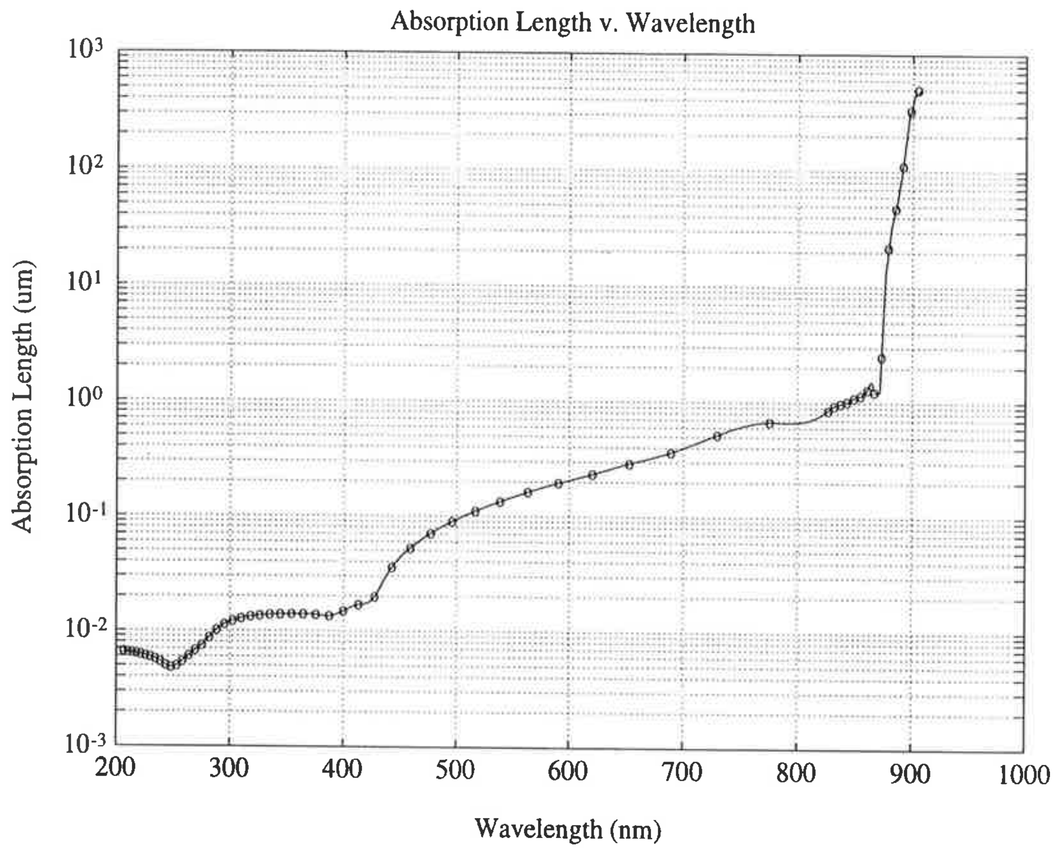


Figure 2.14: Measured absorption length against wavelength for GaAs.

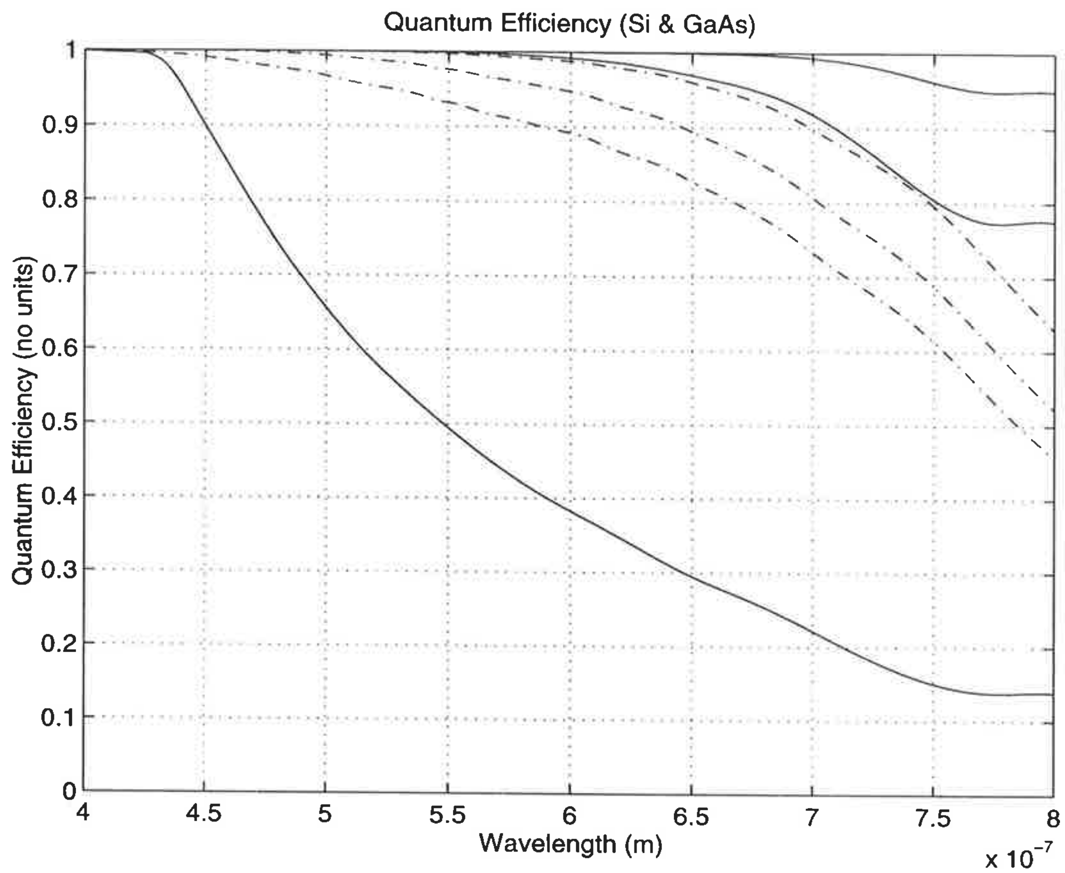


Figure 2.15: Internal quantum efficiency versus wavelength. Dashed lines, silicon W=5, 2 and 0.5 μm. Solid lines, GaAs W=2, 1 or 0.1 μm. Quantum efficiency increases with depletion width W. Transmission coefficient is set at unity.

The superior quantum efficiency of GaAs over silicon, as seen in Fig. 2.15, is due to the shallower absorption lengths and majority carrier collection process in the GaAs imager. In silicon imagers, minority carriers are collected and so there is a degradation in quantum efficiency due to recombination processes. For the purpose of comparison, we have assumed a transmission coefficient of unity – this simplification does not make a difference to our conclusions. In Fig. 2.15, a depletion region width of $W = 0.1$ μm corresponds to that under the gate. The improved quantum efficiency curves with W = 1-2 μm, simulate the expected effect of the substrate/channel depletion region working together with the gate depletion region. As the imager is a low-frequency device, we expect the substrate/channel depletion region to take part in the collection process, leading to an excellent quantum efficiency.

From the quantum efficiency curves we can now apply Planck's law to calculate the responsivities. Fig. 2.16 & 2.17 show the weighted responsivities for silicon and gallium arsenide.

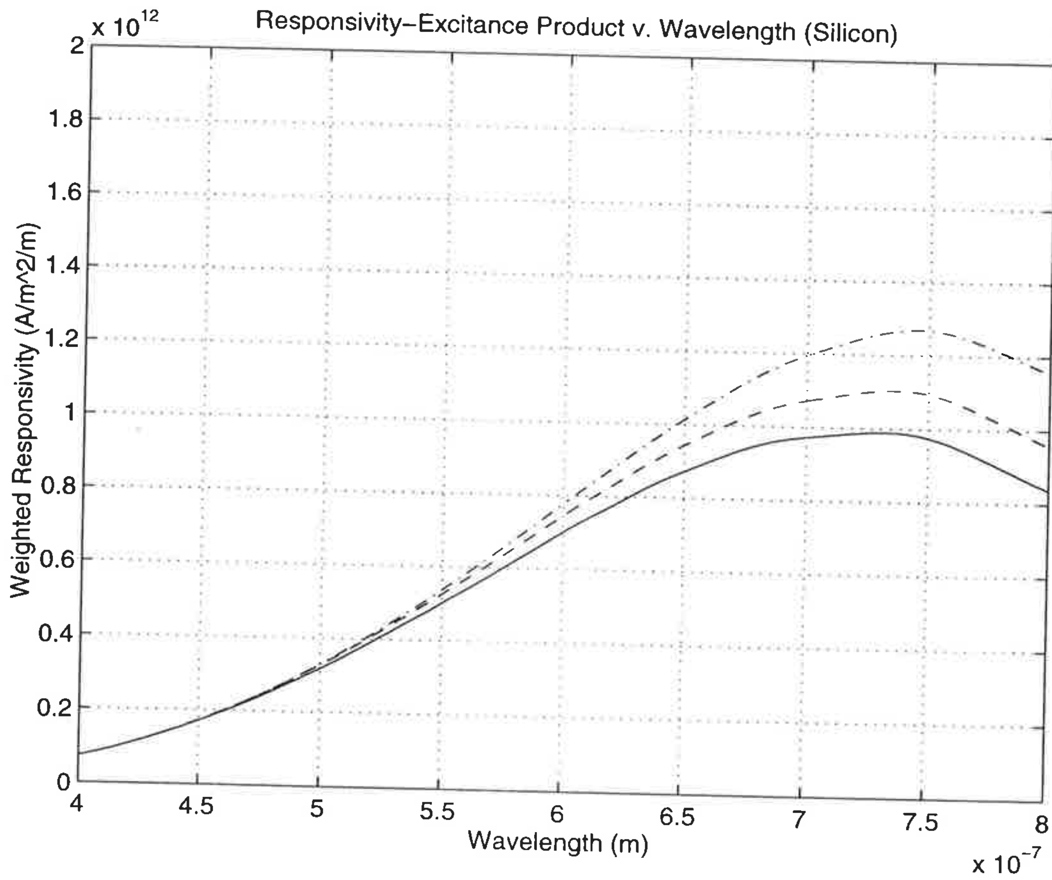


Figure 2.16: Weighted responsivity versus wavelength for silicon. Solid line, $W = 0.5$ μm. Dashed line, $W = 2$ μm. Chained line, $W = 5$ μm

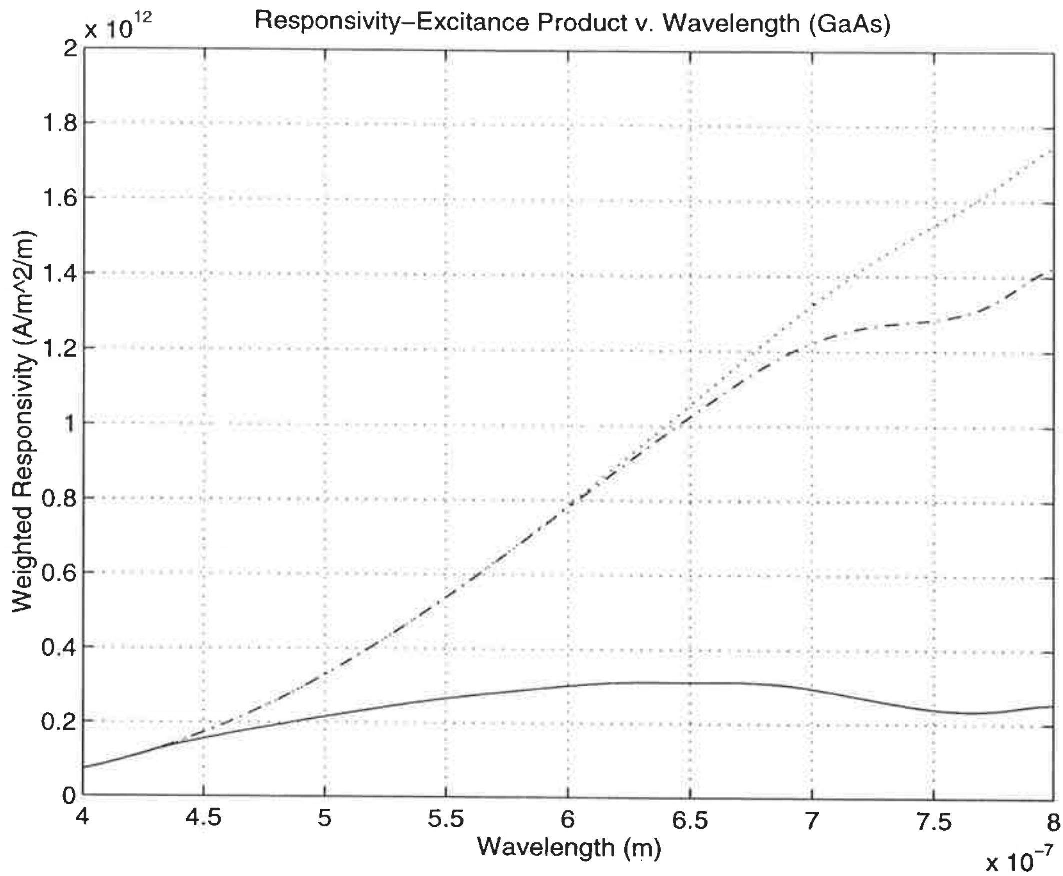


Figure 2.17: Weighted responsivity versus wavelength for GaAs. Solid line, W=0.1 μm. Chained line, W=1 μm. Dotted line, W=2 μm.

Total responsivity, over the visible spectrum, is determined by evaluating the areas under the curves in Figs. 2.16 & 2.17 and then dividing by the total area under the Planck's Law curve $= 46 \times 10^5$ W/m^2 (see Fig. 2.12). A typical tungsten light source can be simulated by inserting a temperature of 3000 K into the Planck equations and the resulting responsivities are displayed in Table 2.4.

The results for silicon, are a little larger compared to 50 mA/W for a typical silicon XY array imager with $W = 5$ μm. This is because we have omitted the effect of the transmission coefficient and furthermore we have assumed a fixed depletion width. In reality the depletion width diminishes as photocharge collects – however the comparison between GaAs and silicon is still valid if we make the simplification of a fixed depletion width. Table 2.4 shows that for GaAs operating with just the gate depletion region, $W = 0.1$ μm, the responsivity is inferior. However, if we include the channel/substrate region, with $W = 2$ μm, the responsivity in GaAs is about 20% larger than in silicon. Furthermore, if we take into account that the channel/substrate region has a fixed potential across it, in

Type	Depletion W (μm)	Area Under Curve (W/m^2)	Responsivity (mA/W)
Si	0.5	2.37×10^5	51
Si	2.0	2.95×10^5	56
Si	5.0	2.81×10^5	61
GaAs	0.1	1.00×10^5	22
GaAs	1.0	2.95×10^5	64
GaAs	2.0	3.25×10^5	71

Table 2.4: GaAs v. Si Responsivity.

contrast to the silicon XY array where the depletion region shrinks during photocollection, we can expect that GaAs would have upto a factor of 5 improvement in responsivity. This means that the proposed 'fingered gate' pixel design, may still equal or even out-perform silicon, despite the presence of opaque gates.

2.5.9 Spatial Degradation by Diffusion

According to the Lambertian exponential law for the absorption of light, we see that 63% is absorbed within one absorption length. We have seen that the absorption lengths, for visible light in GaAs, are all of comparable order to the vertical transistor dimensions. The implication of photocharge being collected at these shallow depths is that the GaAs MESFET is an efficient photocollector. Furthermore, if most of the carriers are efficiently collected then the number of stray carriers spreading, thereby causing spatial degradation of the image, is low. Hence we expect the spatial resolution of a GaAs imager to be excellent.

The channel/substrate depletion region is expected to assist in this regard, by deflecting minority carriers into the substrate and sweeping stray electrons into the collecting channel. The magnitude of this effect and its importance can be determined by monochromator based experiments. The effect of light indirectly altering the depletion width, via photo-assisted emission from substrate traps or photo-ionisation of traps, would need to be taken into consideration. In a later chapter, actual monochromator measurements are presented to show that the channel/substrate region does indeed act as a good photocollector.

The quantitative comparison of GaAs and silicon, at first sight, appears to be difficult due to the fact that our GaAs example is on a bulk substrate, whereas silicon imagers are on epitaxial substrates. Hence we are not comparing like with like. A thin epi-layer gives rise to a poor quantum efficiency but an excellent spatial resolution, whereas a bulk

substrate gives an improved quantum efficiency at the expense of spatial resolution. So how can we make a fair comparison that takes into account this trade-off? The answer is to compare GaAs with silicon, not in terms quantum efficiency or raw spatial resolution *per se*, but in terms of a special figure of merit known as the modulation transfer function (MTF). An ideal MTF has the value of unity and it has the behaviour that if either quantum efficiency or spatial resolution decreases, then MTF also decreases below unity. The definition of MTF is simply [Seib 74]

$$\mathrm{MTF} = \frac{\eta_k}{\eta}$$

where η_k is given by the same formula as the quantum efficiency, η, but with each L_o term substituted by $\sqrt{\frac{1}{L_o^2} + (2\pi k)^2}$ and k is the spatial frequency.

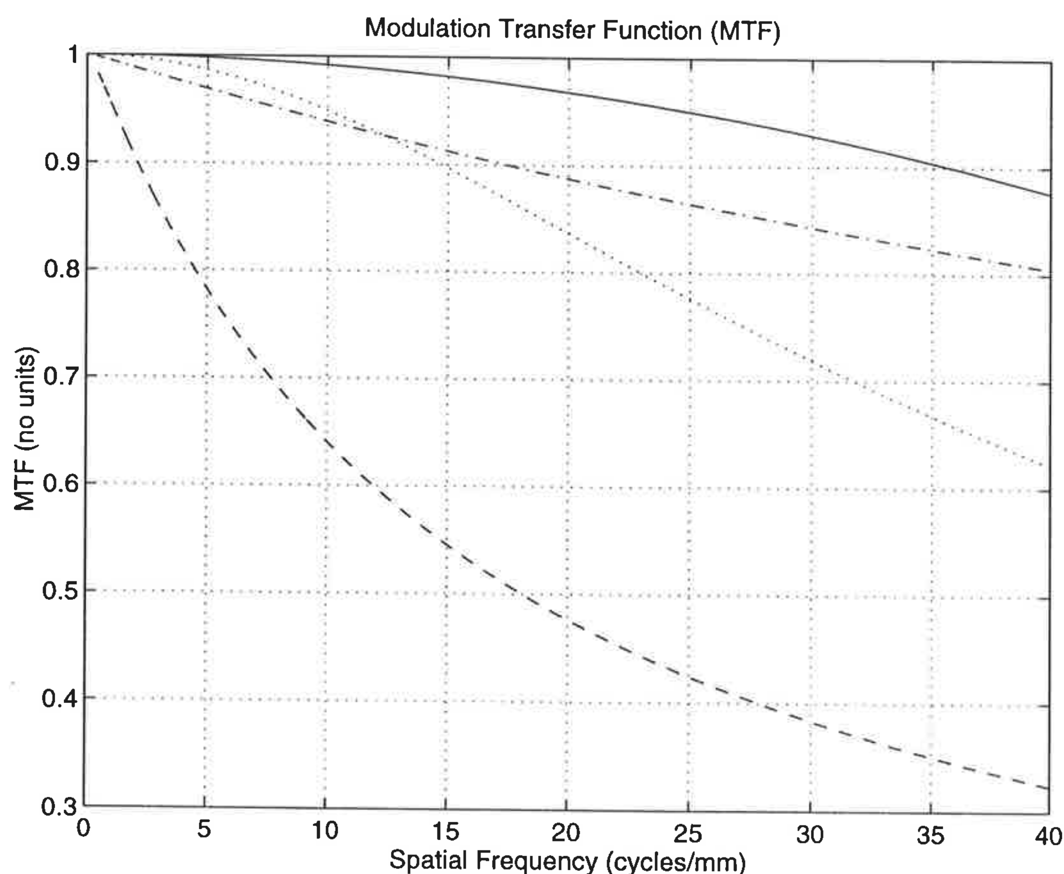


Figure 2.18: Theoretical MTF versus spatial frequency. Diffusion-only curves are: Si bulk, 550 nm (chained line); Si epi, 800 nm (dotted line); Si bulk, 800 nm (dashed line). Geometrical-only curve is for a 7/20 aperture to pitch ratio and the GaAs case matches this curve (solid line).

This is plotted in Fig. 2.18, where the importance of epi rather than bulk substrates for

silicon is demonstrated. Although the diffusion curve for a wavelength of 550 nm in bulk is acceptable, the MTF for 800 nm is severely degraded in bulk compared to epi. The geometrical-only curve is obtained from the usual sinc function expression and represents the ideal case in absence of diffusion effects. Due to the hi-lo n^+/SI junction majority carriers are collected, hence the diffusion term vanishes and our GaAs case corresponds to the ideal geometrical curve. This result is attractive for HDTV detector array applications, where high pixel densities are more susceptible to crosstalk by diffusion.

2.6 Summary

We have given a brief account of the history of gallium arsenide showing that it has built up considerable momentum, giving rise to applications in many areas of high-speed communications and mobile communications. Digital applications also abound but have been slower to progress due to the problems with realising large memories in GaAs. This situation promises to vastly improve as the rapidly emerging CGaAs™ technology matures. A brief comparison between GaAs and other possible competing technologies such as diamond, SOI and SiGe showed that GaAs is the best choice for the XY array approach due its greater maturity, lower complexity and presence of the semi-insulating substrate. For the interactive mobile multimedia communicator paradigm, GaAs is the best choice due to its capability for mixed digital/RF/optical designs.

In comparing the characteristics of GaAs and silicon, we have indicated that GaAs shows promise for X-ray detector arrays, near infrared arrays (by investigating internal photoemission), HDTV arrays, MWIR and LWIR arrays (by investigating new Schottky materials or via hybridisation with CMT) and arrays for aerospace & astronomy, as well as optical arrays for mobile communications. We have shown that GaAs outperforms silicon in a number of significant areas.

In the analysis of quantum efficiency, we have disputed the equation for quantum efficiency in a frontside illuminated finite slab given in recent literature. From first principles we derive the correct equation, which has not been previously reported. This original contribution is significant as the equation can be separated into two physically meaningful terms. This new perspective will be useful for simplifying MTF analysis in epi substrates. Furthermore, we have disputed the MTF equation in bulk due to Barbe and have confirmed, from first principles, the validity of Seib's bulk equation.

We have shown that due to the superior absorption coefficients in GaAs, improved quantum efficiency is expected. Furthermore for a hi-lo n/SI junction in GaAs, majority

carriers are collected and this leads to improved spatial resolution. This advantage is presented in the form of MTF plots, showing superior overall MTF of GaAs over silicon.

Having made the case for GaAs as the choice technology, for this work, in the next chapter we now proceed to perform actual optical measurements to study the photoresponse of the GaAs MESFET for optimisation of the pixels for the imaging array.

Bibliography

[Barbe 75] D.F. Barbe, "Imaging devices using the charge-coupled concept," *Proc. IEEE,* **vol. 63,** pp. 38-67, Jan. 1975.

[Beaumont 93] S.P. Beaumont *et al,* "GaAs solid state detectors for physics at the LHC," *IEEE Trans. Nuc. Sci.,* **vol. 40,** no. 4, 1993.

[Bencivelli 95] W. Becivelli *et al*, "Some new results on semi-insulating GaAs detectors for low energy X-rays," *Nuc. Inst. Methods Phys. Res. A,* Elsevier, **vol. 355,** pp. 425-427, 1995.

[Bernhardt 95] B. Bernhardt *et al,* "Complementary GaAs (CGaAs™): A high performance BiCMOS alternative," *IEEE GaAs IC Symposium,* San Diego, pp. 18-21, 29 Oct.-1 Nov. 1995.

[Buda 94] F. Buda and J. Kohanoff, "Porous silicon: a silicon structure with new optical properties," *Prog. Quant. Electr.,* **vol. 18,** pp. 201-226, 1994.

[Collins 92] A.T. Collins, "Diamond electronic devices – can they outperform silicon or GaAs?" *Materials Science and Eng.,* **vol. B11,** pp. 257-263, 1992.

[George 89] P. George *et al.*, "Modelling the substrate depletion region for GaAs FETs fabricated on semi-insulating substrates," *Solid-State Electronics,* **vol. 32,** no. 2, pp. 165-168, 1989.

[Groeseneken 90] G. Groeseneken *et al*, "Temperature dependence of threshold voltage in thin-film SOI MOSFETs," *IEEE Electron Dev. Lett.,* **vol. 11,** no. 8, pp. 329-331, 1990.

[Konig 94] U. Konig, "Challenges for a Si/Ge heterodevice technology," *Microelectronic Engineering,* **vol. 23,** pp. 3-13, 1994.

[Lande 96] S. Lande, "The status of silicon carbide and related materials," *III-Vs Review*, **vol. 9,** no. 1, pp. 58-62, 1996.

[May 95] P. May, "Polymer electronics – fact or fantasy?" *Physics World,* pp. 52-57, March 1995.

[Mead 66] C.A. Mead, "Schottky barrier gate field effect transistor," *Proc. IEEE Lett.,* pp. 307-308, Feb. 1966.

[McCaughan 79] D.V. McCaughan and B.R. Holeman, Chapter 5, pp. 241-295, found in *Charge-Coupled Devices and Systems,* Eds. M.J. Howes and D.V. Morgan, Wiley, 1979.

[McGregor 92] D. S. McGregor *et al*, "Development of bulk GaAs radiation detectors," *IEEE Nuc. Sci.,* **vol. 39**, No. 5, pp. 1226-1236, October 1992.

[Pearsal 85] T.P. Pearsal *et al,* "Ge_xSi_{1-x} Modulation-Doped p-Channel Feild Effect Transistors," *Proc. 1st Int. Symp. Silicon MBE,* Electrochemical Society, Pennington, N.J., pp. 400-405, 1985.

[Sadek 95] A. Sadek and K. Ismail, "Si/SiGe CMOS possibilities," *Solid-State Electronics,* **vol. 38,** no. 9, pp. 1731-1734, 1995.

[Seib 74] D.H. Seib, "Carrier diffusion degradation of modulation transfer function in charge coupled imagers," *IEEE Trans. E. D.*, **vol. ED-21**, no. 3, pp. 210-217, 1974.

[Spicer 85] W.E. Spice and S.J. Eglash, "Fundamental studies of interfaces: the unified defect model and its application to GaAs integrated circuits," *VLSI Electronics MIcrostructure Science,* **vol. 10**, pp. 79-117, 1985.

[Sumner 92] T.J. Sumner *et al*, "GaAs revisited as a room temperature X-ray detector," *Nuc. Inst. Methods Phys. Res. A,* Elsevier, **vol. 322,** pp. 514-517, 1992.

[Sun 95] G. Sun, L. Friedman and R.A. Soref, "Intersubband lasing lifetimes of SiGe/Si and GaAs/AlGaAs multiple quantum well structures," *Appl. Phys. Lett.,* **vol. 66**, no. 25, pp. 3425-3427, 1995.

[Theuwissen 95] A.J.P Theuwissen, *Solid-State Imaging with Charge-Coupled Devices,* Kluwer, 1995, p. 135.

[Trew 95] R. J. Trew and M. W. Shin, "High frequency, high temperature field-effect transistors fabricated from wide band gap semiconductors," *Int. Journ. of High Speed Elec.,* **vol. 6,** no. 1, pp. 211-236, 1995.

[Voinigescu 93] S. Voinigescu *et al,* "Emerging SiGe technology - a review," *Electron Technology,* **vol. 26,** no. 4, pp. 25-64, 1993.

Chapter 3

MESFET Photodetection

"Frequently, I have been asked if an experiment I have planned is pure or applied research; to me it is more important to know if the experiment will yield new and probably enduring knowledge about nature. If it is likely to yield such knowledge, it is, in my opinion, good fundamental research; and this is much more important than whether the motivation is purely esthetic satisfaction on the part of the experimenter on the one hand or the improvement of the stability of a high-power transistor on the other."
William Shockley (1910-1989)
AMERICAN CO-INVENTOR OF THE TRANSISTOR

"When one comes down to particular instances, everything becomes more complicated."
Albert Camus (1913-1960)
FRENCH NOVELIST AND PLAYWRIGHT

3.1 GaAs MESFET Optical Model

3.1.1 Background

From the outset we realised the importance of determining the various optical gain mechanisms in a GaAs MESFET in order to accurately predict transistor behaviour for implementation within an image sensor. It was noted that considerable controversy, in the literature, existed over the nature of GaAs MESFET optical gain. As a result of our investigation, a consistent model was developed [Adibi 89] that overcame the difficulties raised by other researchers engaged in the controversy.

The approach pursued in this research was to model the overall optical gain of a MESFET in terms of four unified effects:

(1) A photovoltaic effect in the gate-source and gate-drain depletion regions

(2) A photoconductive effect in the channel, with associated recirculation gain

(3) A photovoltaic effect in the channel-substrate depletion region

(4) A photovoltaic effect via photo-assisted carrier emission from substrate traps.

This model is used to resolve apparent contradictions, between different results published in the literature, by observing that each of these effects come into play, by varying degrees, depending on the wavelength of light used and hence penetration depth achieved.

Apparent discrepancies are reconciled by side effects such as the blocking of the electron recirculation effect caused by pulsing the light in coincidence with the depletion layer recovery time. Another effect not taken into account by other researchers is the sideways diffusion of carriers into unilluminated regions – this effect is further used to reconcile previously unexplained discrepancies.

Two proposed primary mechanisms are reported in the literature:

- Classical photoconductivity effect [Gammel 78, Osterwalder 79, Gammel 80]
- Classical photovoltaic effect

However the photovoltaic mechanism manifests via two reported secondary effects:

- Depletion modulation of channel (intrinsic) [Graffeuil 79, Edwards 80, Wieder 80, Sun 81, Chaturvedi 83, Umeda 85, Umeda 86]

- Depletion modulation of channel (extrinsic) [Sugeta 80]

Various combinations of the above effects are also reported [Pan 78, Flesner 82, Forrest 82, Noad 82]. The controversy centres around the lack of consensus amongst various researchers as to which combination of these mechanisms actually takes place in the GaAs MESFET. Developing a correct description of the optical mechanisms, in the GaAs MESFET, is considered important for design optimisation of an imager based on an 'active pixel' that exploits the gain mechanisms of the transistor.

The key experimental techniques to investigate the photoresponse characteristics involve MESFET illumination by:

- Unfocussed laser flood
- Scanned focussed laser
- Modulated laser (focussed & unfocussed)
- Monochromator.

Firstly, we shall describe our model [Adibi 89] and then the experiments and results will be detailed in a later section.

3.1.2 Photoresponse Model

The GaAs MESFET can be implemented as the photoresponsive element in the XY imager either passively or actively. The passive case requires the MESFET to behave as a simple switch and only the source terminal is illuminated so as to act as a photodiode. In the active case the whole MESFET is illuminated to take advantage of the resulting internal photogain mechanisms. The advantage of passive illumination is that it is easy to implement. Active illumination offers better sensitivity and is more of a challenge to implement, particularly due to the controversy surrounding the origin of these internal photogain mechanisms.

In order to design an imager utilising active illumination, the first step was to produce a generalised model that adequately explained the photoresponse mechanisms in the MESFET so that reproducible design calculations could be made. In summary, the model is encapsulated in the equivalent circuit shown in Fig. 3.1. The photoresponsive elements of the equivalent circuit are as follows:

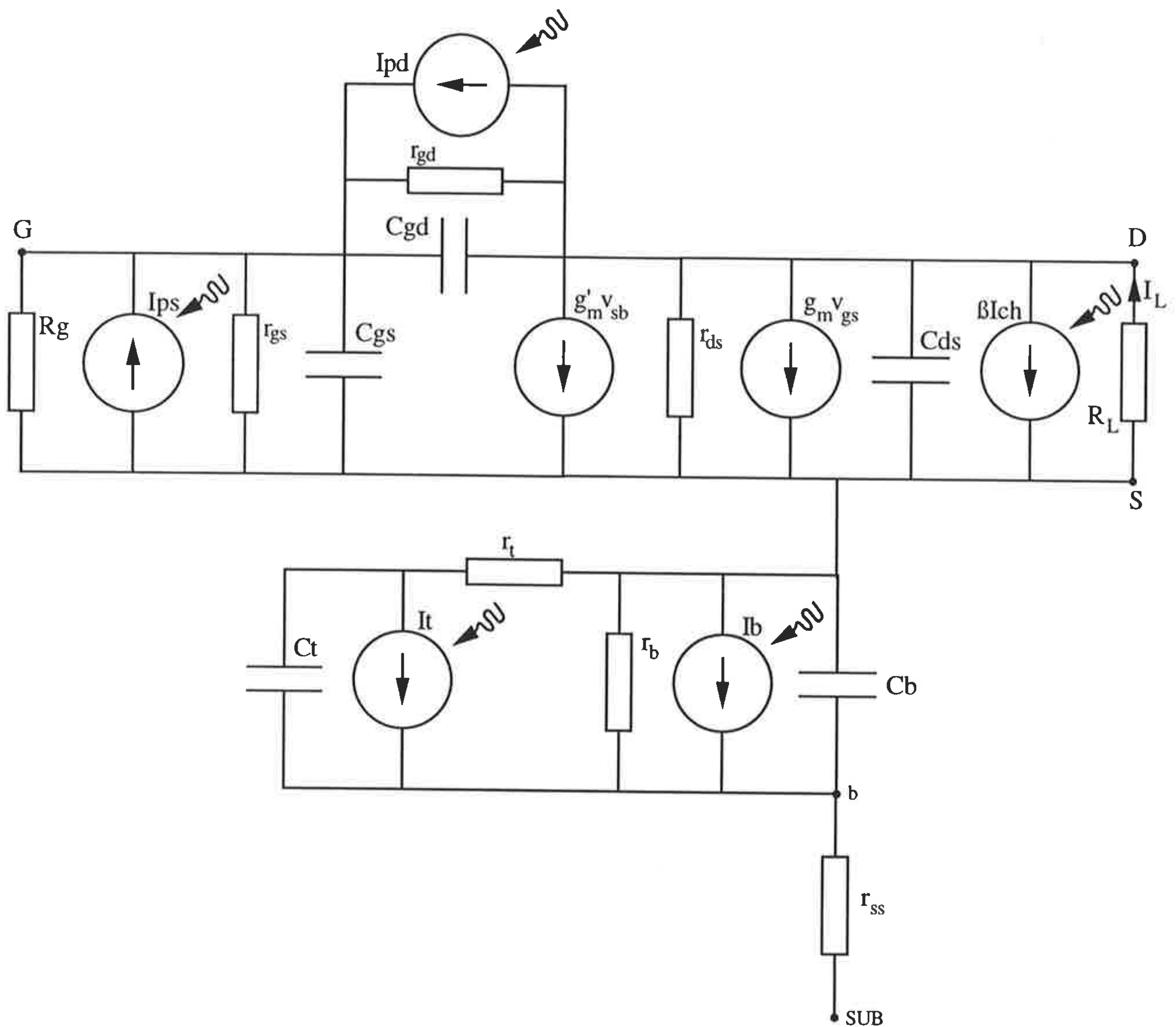


Figure 3.1: Equivalent circuit of MESFET showing photoresponsive elements.

- Photoconductive response in the channel is modelled by I_{ch}.
- Photovoltaic response takes place in the parts of the gate depletion region that protrude out from under the gate. This protrusion is larger on the side of the gate towards the drain than on the side towards the source. Response in these two regions is thus separately modelled by I_{pd} and I_{ps}.
- Photovoltaic response in the channel/substrate depletion region is modelled by I_b.
- Photo-assisted carrier generation from substrate traps is modelled by I_t.
- Modulation of the channel/substrate depletion region changes the drain current and this is modelled by $g'_m v_{sb}$. Modulation of the gate depletion region also changes the drain current and is modelled by $g_m v_{gs}$.

Due to electrons recirculating through the channel, until a photogenerated hole is recombined, there is an associated gain β that must be applied to I_{ch}. When the hole transit time, t_h, is comparable to the minority carrier (hole) lifetime, τ, the photoconductive current gain becomes,

$$\beta = \frac{\tau}{t_e}$$

for,

$$t_h \approx \tau$$

where t_e is the electron transit time. When t_h is small compared to τ the gain obviously becomes,

$$\beta = \frac{v_e}{v_h}$$

for,

$$t_h \ll \tau$$

where v_e and v_h are electron and hole velocities respectively. In our case $\tau \approx 20$ ns and $t_h = 0.4$ ns, for $V_{ds} = 0.6$ V with a 3.2μm drain/source separation. The first expression incorrectly gives $\beta \approx 1000$, for our case, but the second expression holds with $\beta \approx 20$.

Referring to Fig. 3.1, C_{gs} and C_{gd} can be modelled by the standard relation,

$$C_g = C_{go}[1 - \frac{v_g}{v_{bi}}]^{-M_g}$$

where,

$$M_g \approx \frac{1}{2}$$

and the Schottky dynamic resistances, r_{gs} & r_{gd}, can be found from suitable curve tracer measurements.

Under reverse bias the gate reverse saturation current $I_o \approx 10^{-13}$ A, and for gate photocurrent $I_{pg} \gg I_o$ we can say that $I_g = I_{pg}$. However, under forward bias,

$$I_g = I_o(e^{\frac{qV_g}{nkT}} - 1) - I_{pg}$$

and the gate open circuit voltage can be found by putting $I_g = 0$,

$$V_{oc} = \frac{nkT}{q} \ln(1 + \frac{I_{pg}}{I_o})$$

Terminating the gate with a resistor, R_g, to ground will cause a current

$$I_g = \frac{nkT}{qR_g} \ln(1 + \frac{I_{pg}}{I_o})$$

to flow. This can be tested by experiment and we expect the relation to be more accurate for shorter wavelengths. By varying R_g and I_{pg}, this could be used as a technique to extract n and I_o.

Substrate parameters can be evaluated by modulating the laser beam and sweeping through frequency. The frequency response will drop off at approximate positions corresponding to $\tau_b = 2\pi R_b C_b$ and $\tau_t = 2\pi(r_b + r_t)C_t$. The first time constant is due to the channel/substrate interface and the second is due to traps. Also $R_b = r_b \| r_t$, where r_b is the channel/substrate barrier interface resistance and r_t is the equivalent resistance due to traps. To determine the substrate transconductance, g'_m is given by,

$$g'_m = \frac{i_{ds}}{v_{sb}}$$

we stimulate the traps with a suitably long wavelength ($\lambda > 800$nm) and then

$$v_{sb} = \frac{r_b I_t}{j\omega C_t} \{ \frac{1}{r_b + r_t + 1/j\omega C_t} \}$$

By applying a voltage V_{ap} to the substrate, with no incident light, we can adjust V_{ap} to obtain the same i_{ds} achieved with the laser; then,

$$v_{sb} = \frac{V_{ap} R_b}{R_b + r_{ss}}$$

Using manufacturer's data C_b and r_{ss} can be calculated. V_{ap} and ω are known driven variables. i_{ds}, τ_t, τ_b and v_{sb} are measured. Thus there remains five equations and five unknowns r_b, r_t, C_t, I_t and g'_m. A spread of values against ω can be graphed to determine both experimental error and model limits.

For wavelengths that stimulate both substrate traps and the channel/substrate region, the full solution is given by,

$$v_{sb} = \frac{r_b I_t(\frac{1}{j\omega C_t R_t}) + r_b I_b(\frac{1}{j\omega C_t r_t} + 1)}{\frac{r_b}{r_t} + (j\omega C_b r_b + 1)(\frac{1}{j\omega C_t r_t} + 1)}$$

The time constants are given by the quadratic roots of the reduced denominator, which are somewhat complex but can be approximated by $\tau_b = 2\pi R_b C_b$ and $\tau_t = 2\pi(r_b + r_t)C_t$. Fig. 3.2 shows the frequency response for this case, indicative of a dual pole and single zero system – the dashed line at the lower frequency gives $1/\tau_t$ and at the higher frequency gives $1/\tau_b$. To illustrate the form of the curve, dummy values of $C_b = 20$ fF, $C_t = 1$ pF, $r_b = 10\text{k}\Omega$, $r_t = 100$ kΩ, $I_b = 50$ μA and $I_t = 20$ μA have been inserted. In practice, v_{sb} is a difficult quantity to measure; however changes in v_{sb} modulate the channel thickness and hence drain current. Therefore the time constants are evaluated from a plot of drain current versus laser modulation frequency. It should be noted that a roll-off in v_{sb} *increases* the channel thickness and hence drain current. Parameter extraction from this curve is more complex than the long wavelength case, as the two poles can influence each other. The advantage of using a longer wavelength is that the circuit reduces to a simpler one pole system. In the two pole case the top plateau is given by $v_t = r_b(I_t + I_b)$, the lower plateau by $v_b \approx R_b I_b$ and the zero position by $\tau_z = \frac{2\pi C_t r_t I_b}{I_t + I_b}$.

Figure 3.3 shows the measured drain current against laser modulation frequency. The three curves are for three separate devices, giving an indication of the spread in values. From this graph we deduce that the time constant due to substrate traps, $\tau_t = 0.1 - 1$ ms.

3.2 GaAs MESFET Photoresponse Measurements

3.2.1 Monochromator Measurements

The photoresponse of a MESFET at different depths, within the device, can be 'probed'

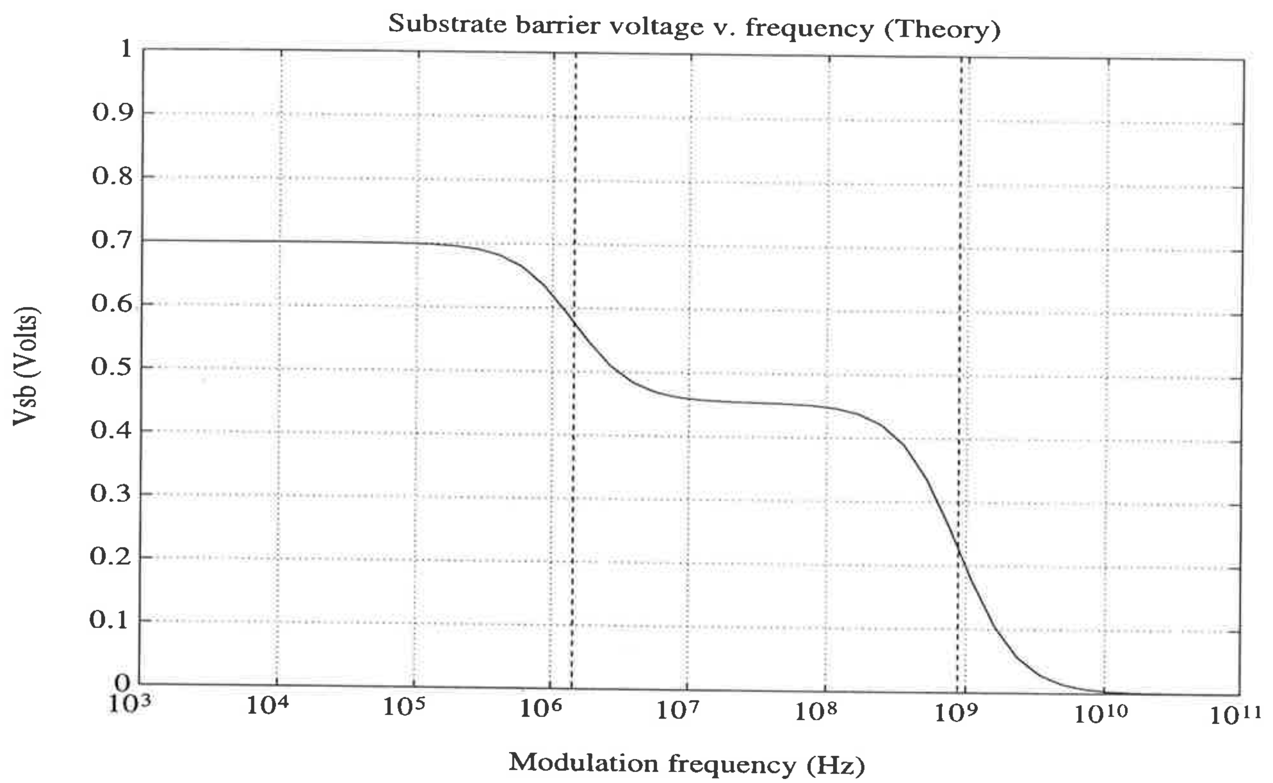


Figure 3.2: Voltage across channel/substrate barrier against laser modulation frequency. The vertical dashed lines show the approximated time constant positions.

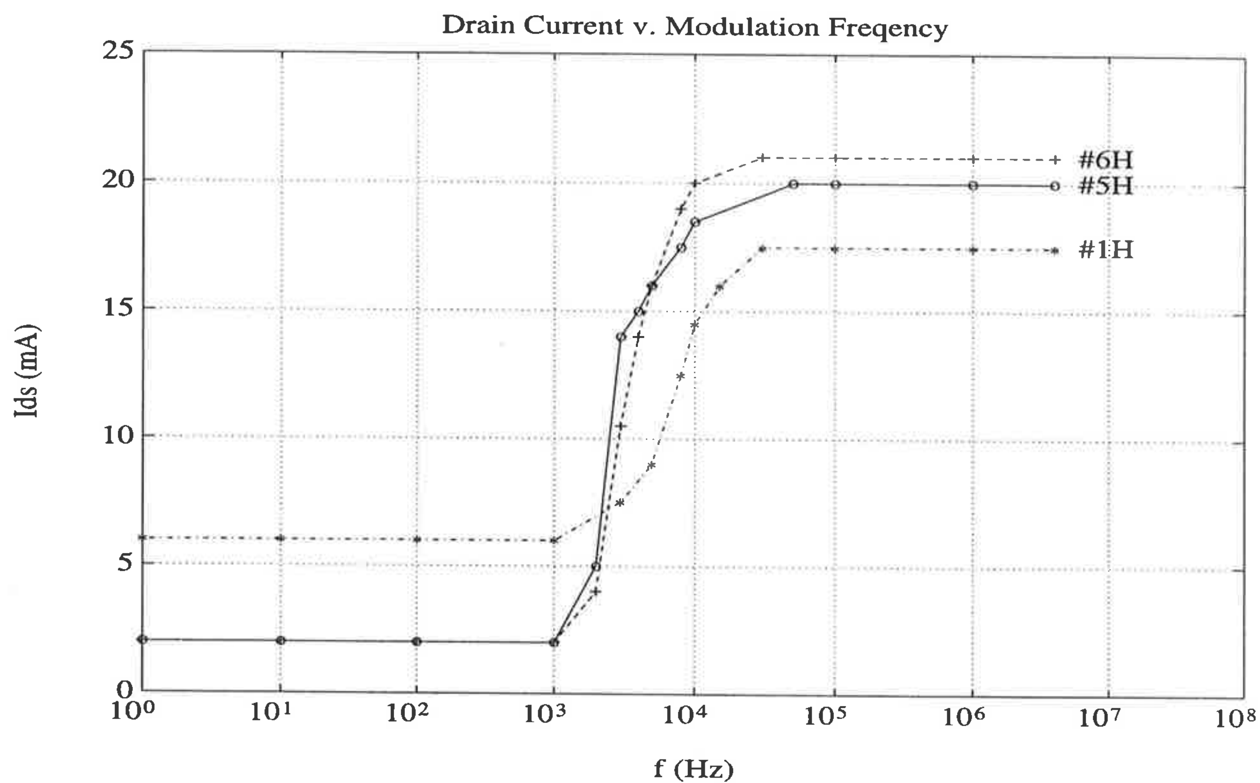


Figure 3.3: Drain current versus laser modulation frequency. Three curves represent three separate devices.

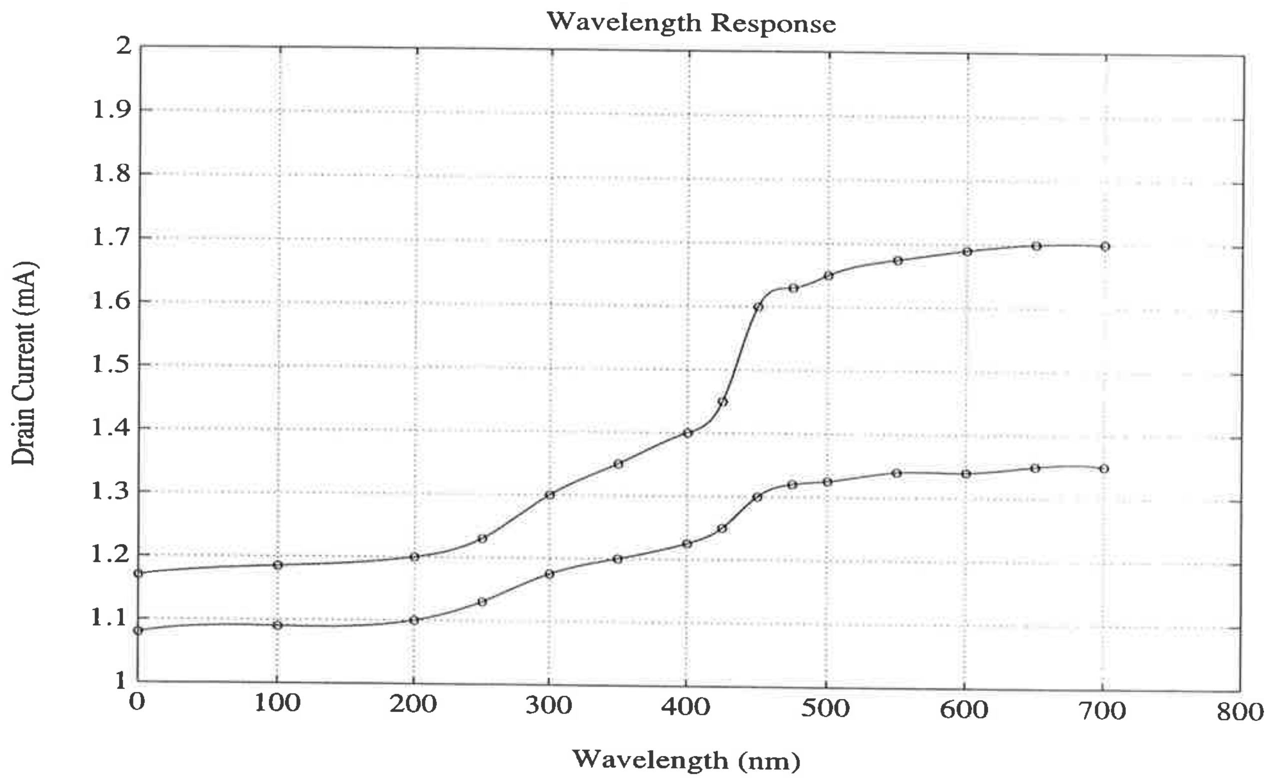


Figure 3.4: Wavelength response using xenon source monochromator. V_{gs}=-0.8 V, V_{ds}=0.6 V. Dual curve represents the looping effect.

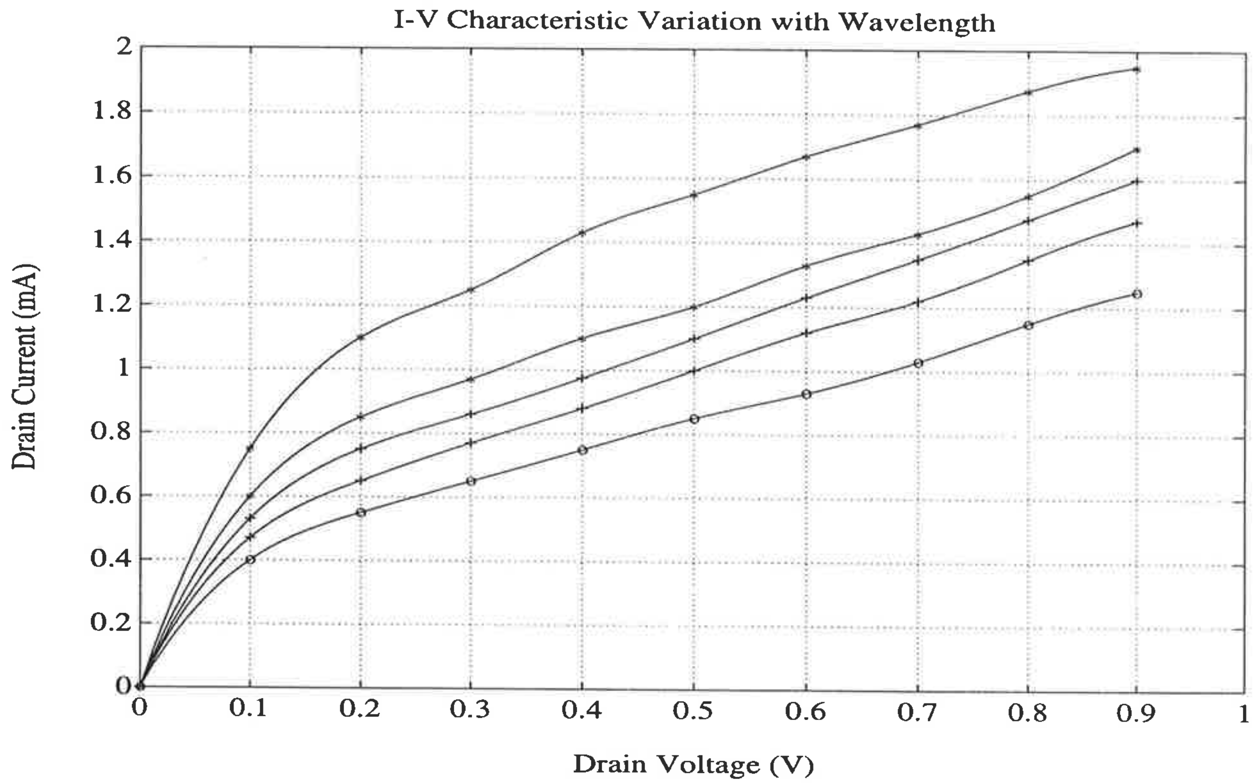


Figure 3.5: I-V Characteristic variation using xenon source monochromator, V_{gs}=-0.8 V, '∗'=700 nm, '+'=300 nm, 'o'=0 nm (dark). Dual curves represent the looping effect. There is no looping in the dark condition.

by using a variable wavelength source. The objective is to separate the gate response from the channel/substrate response, so as to establish their relative importance.

To this end we used a Bausch & Lomb monochromator with a xenon source and a 1200 groove/mm grating. Fig. 3.4 shows the drain current response to wavelength. The two curves represent the upper and lower portions of the hysteresis loop. An interesting observation is that the loop increases in width for longer wavelengths. Fig. 3.5 shows the looping in the I-V characteristic – notice there is no looping in the dark condition. The loop is widest at about 700 nm – however, on the contrary, with a 0.125 mW laser at 678 nm no looping is observed. This appears to suggest that both power and wavelength have an effect on the hysteresis. Further work would need to be carried out to conclusively tie in these observations with the accepted trap theory; however, as it does not appear immediately relevant to the imager design, further analysis is deferred for the future.

The conclusions drawn from Fig. 3.4 are:

- The sharp increase at 450 nm corresponds to light penetrating the gate depletion region. Below this wavelength, surface state recombination would be dominant.

- Between 600 nm and 700 nm the curve does not roll off. This is significant as these wavelengths correspond to the channel/substrate depletion region. Hence the channel/substrate depletion region is just as efficient a photocharge collector as the gate depletion region. This is only true for low frequency operation due to the large time constant associated with the channel/substrate interface.

The last observation is most important as we now expect the low frequency imager to efficiently collect photocharge generated from the whole visible spectrum – ie. the quantum efficiency should be excellent.

Fig. 3.6 shows only a small perturbation in the transfer characteristic with wavelength. This is advantageous as the transistors, in the imager, should not have greatly varying bias points with incoming light.

3.2.2 Laser Flood Measurements

In order to explore the effect of laser illumination, transfer characteristics are plotted in Fig. 3.7 with the whole MESFET region exposed to unfocussed laser light. The perturbation of the transfer characteristics is larger in Fig. 3.7 than Fig. 3.6, due to the higher powered laser sources. The lasers appear to create a shift in the transfer curve, equivalent

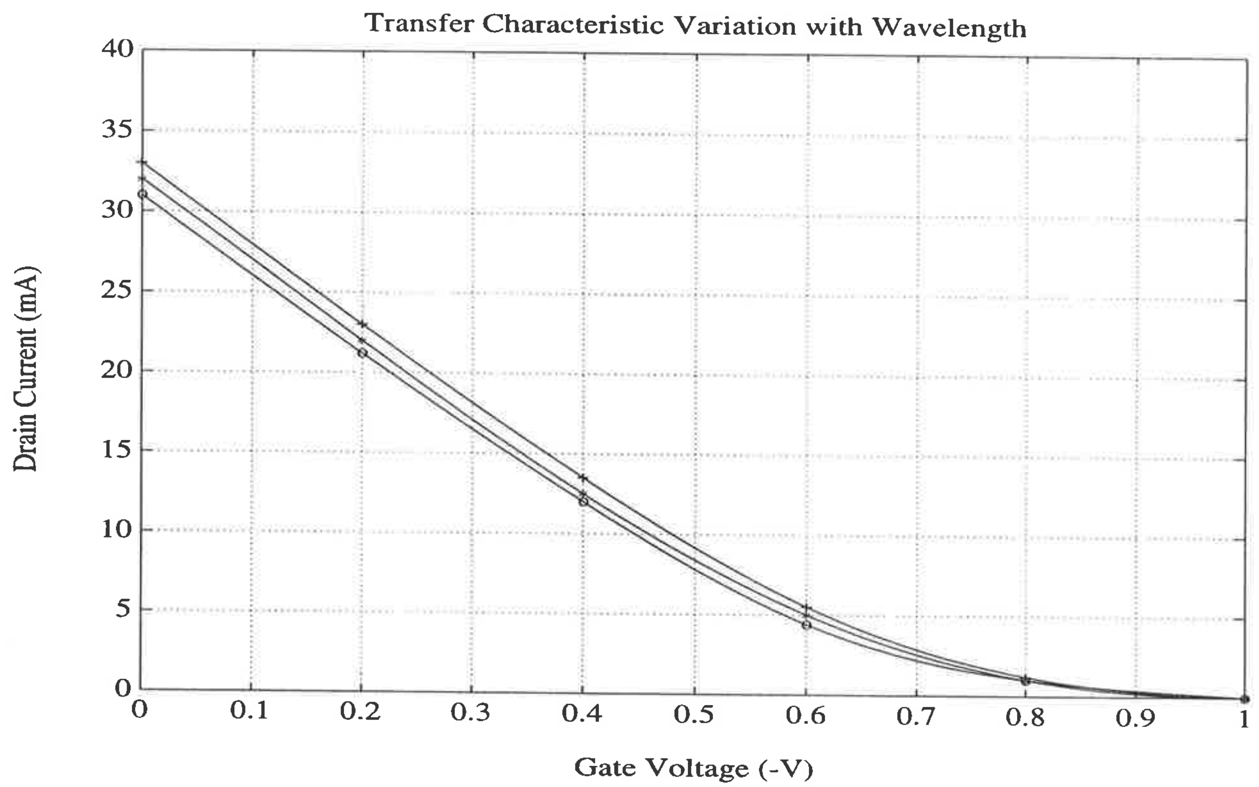


Figure 3.6: Transfer characteristic variation using xenon source monochromator. V_{ds}=0.6 V, '+'=700 nm, '*'=400 nm, 'o'=200 nm.

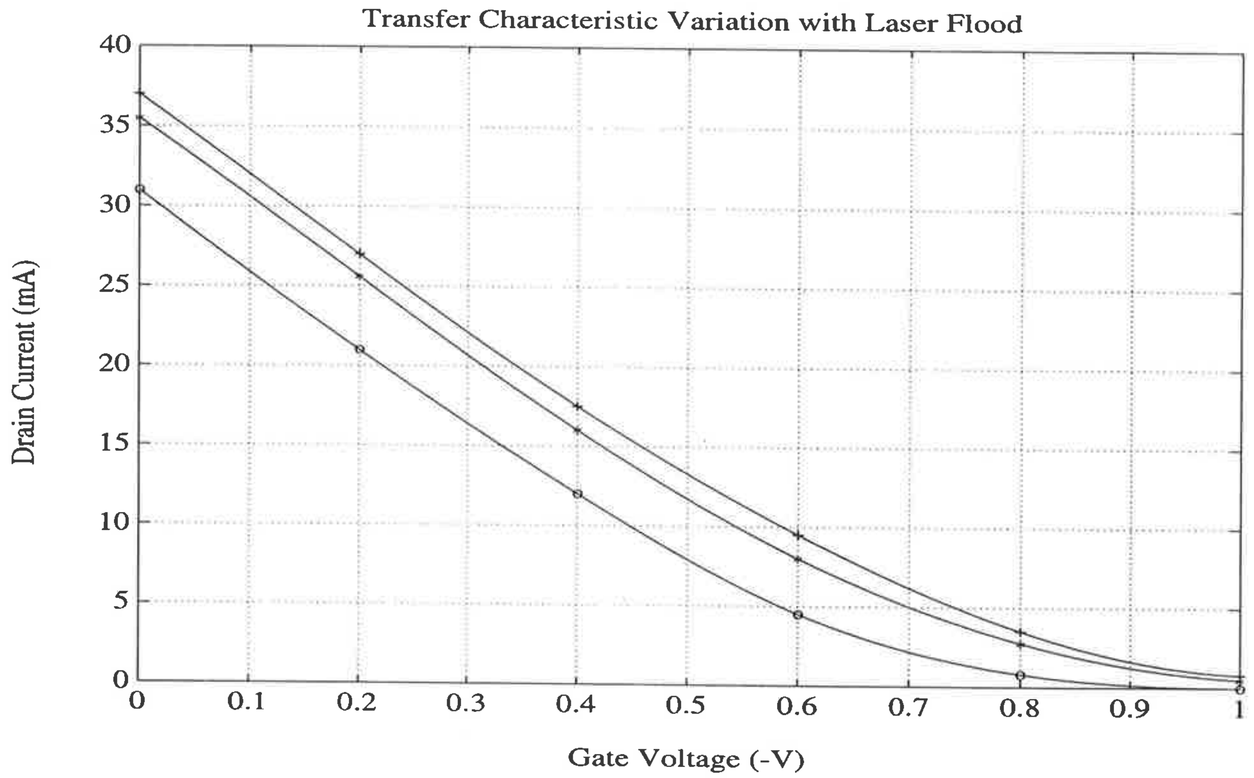


Figure 3.7: Transfer characteristic variation using laser flood. V_{ds}=0.6 V, '+'=815 nm @ 3 mW, '*'=632 nm @ 0.95 mw, 'o'=0 nm (dark).

to increasing the gate voltage by around 0.1 V. This could potentially be modelled by a contraction of the gate depletion region.

3.2.3 Focussed Laser Measurements

A new effect, in planar GaAs MESFETs, whereby a sharp increase in optical gain at the transistor edges occurs, is reported for the first time. This gain effect only manifests when a large resistor is inserted in series with the gate, to produce the conditions for photovoltaic gate biasing. The mechanism for increased gain, at the edges, is suggested to be due to carrier photogeneration in the substrate that is subsequently collected by the gate. Application in the area of X-Y addressable transistor array imagers, motion detectors, optical neural nets, GaAs X-ray detectors etc. is possible, for increasing photosensitivity.

Photoresponse of GaAs MESFETs has received much attention due to potential application in high speed optoelectronic communications, OEICs and optical tuning of microwave devices. Various optical gain mechanisms have been reported, including photovoltaic gate biasing. This effect occurs when the gate photocurrent follows through an external series gate resistor, R_g, thus increasing the gate voltage and hence drain current. To produce a significant increase in drain current, a large R_g introduces a large RC time constant which typically causes the response to roll off in the 10-100 MHz range. Consequently, as most researchers have concentrated on high speed applications, photovoltaic gate biasing has been regarded as being of limited merit [Darling 87] and has not received full attention. However, as the data rate of a typical imager is less than 10 MHz, photovoltaic gate biasing can potentially be used to increase the sensitivity of such devices. Furthermore, we discovered that the sensitivity is increased in the region where a gate overhangs the transistor edge. This new knowledge can help to maximise the benefits of this effect to imaging devices.

This new edge effect was discovered when a GaAs MESFET was scanned with a focussed laser spot. The planar MESFETs used have a fingered structure with 5 gates, L = 0.8 μm and overall W = 400 μm. The channel (depth, d = 0.1 μm and doping, N_d = 1.2×10^{17} cm^{-3}) is situated on a semi-insulating (SI) GaAs:EL2 substrate, with no buffer layer. The source-gate and drain-gate separations are both 1.3 μm.

Fig. 3.8 shows a length scan across a transistor with 5 fingers. As the gate length, 0.8 μm, is smaller than the laser spot diameter, 2 μm, the peaks of the curve correspond to when the spot exactly straddles the gate. A 1 MΩ resistor is included, in series with the gate, to produce an extrinsic gain effect by photovoltaic self-biasing. This gain was deliberately introduced to accentuate the peaks and troughs. It is notable that the gain at a trough is

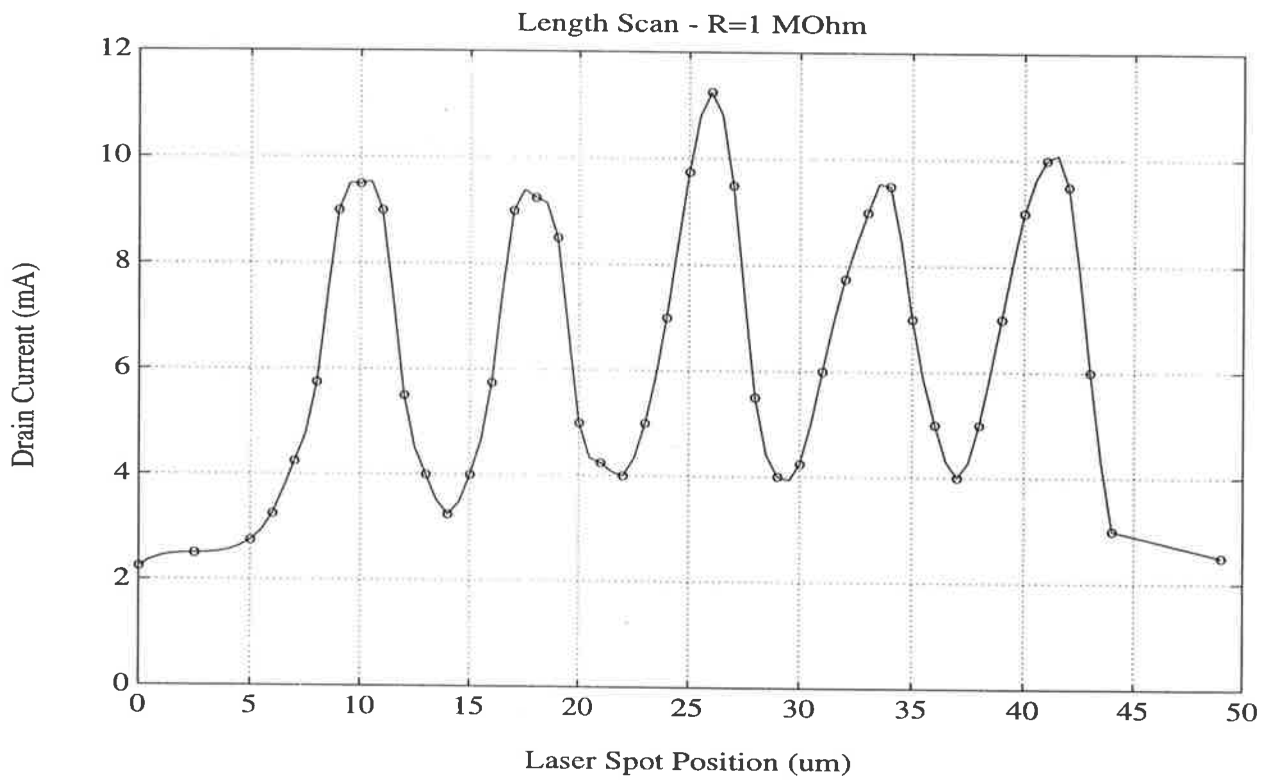


Figure 3.8: Transistor length scan with focussed laser, 678 nm @ 0.125 mW, 2 μm diameter spot size. 1 MΩ series gate resistor. V_{gs}=-0.8 V, V_{ds}=0. 6V.

at best 2, whereas the gain at a peak is at best 5, ie. the gain is more than doubled when the laser spot straddles the gate. This indicates that maximum sensitivity is obtained by illuminating the gate edges.

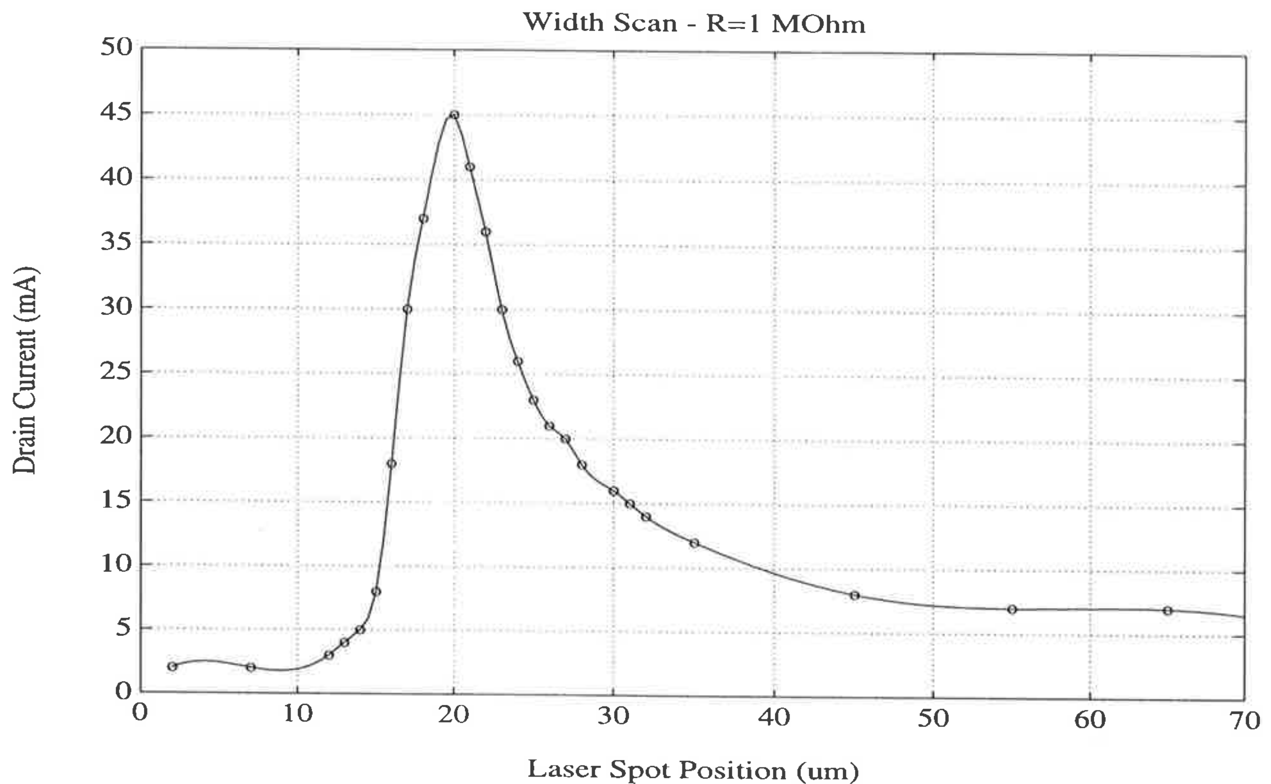


Figure 3.9: Transistor width scan with focussed laser, 678 nm @ 0.125 mW, 2 μm diameter spot size. 1 MΩ series gate resistor. V_{gs}=-0.8 V, V_{ds}=0.6 V. Transistor edge corresponds to 20 μm on the x-axis.

Fig. 3.9 was obtained by scanning the laser across the transistor width and highlights a significant gain effect. From a gain of about 3, the gain dramatically increases to over 20, at the transistor edge. This is a new effect that has not been reported in the literature, by other workers. We have patented a novel imager design that utilises this new effect [Abbott 91] to create an increased sensitivity.

An important result from Fig. 3.9, for the imager design, is that the left hand side of the curve decays rapidly over a distance 5 μm, as the spot moves away from the transistor edge and onto the substrate. This decay distance is much smaller than the anticipated imager pixel size (40 μm sq) and is an excellent feature required for an imager with low pixel crosstalk and good spatial resolution.

With reference to Fig. 3.10, we see that the gain drops off at the edge, in the absence of

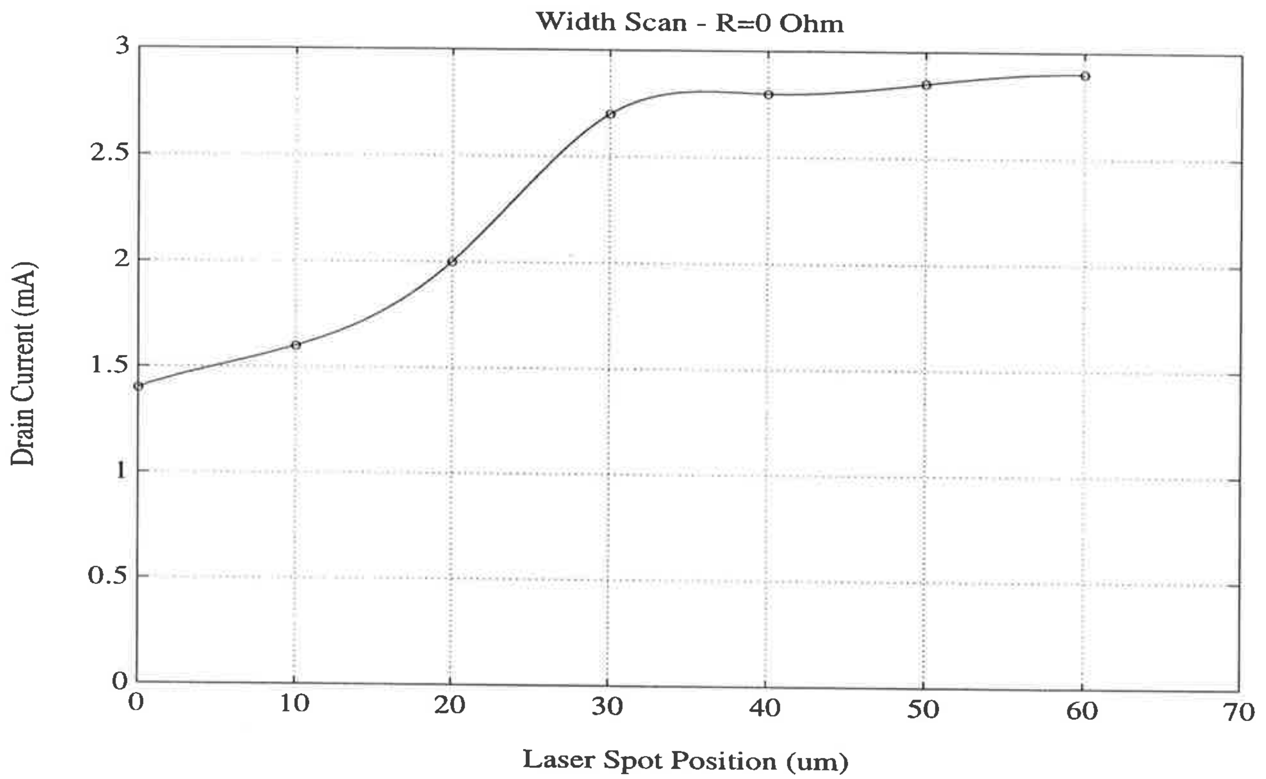


Figure 3.10: Transistor width scan with focussed laser, 678 nm @ 0.125 mW, 2 μm diameter spot size. No series gate resistor. V_{gs}=-0.8 V, V_{ds}=0.6 V. Transistor edge corresponds to 20 μm on the x-axis.

the series resistor. This demonstrates the photoductive component of the photoresponse, showing that it is small. The fact that roll-off commences a few microns before the edge is reached could possibly be explained by tapering of the channel.

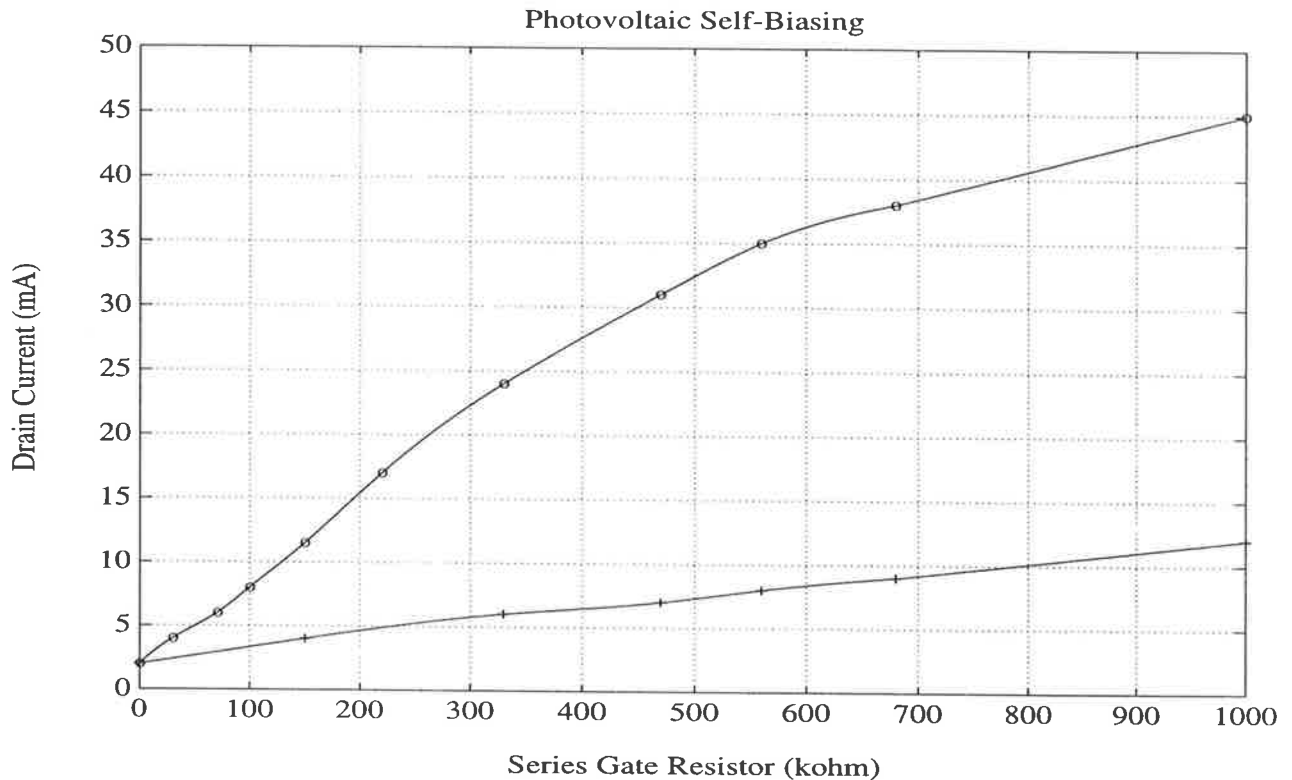


Figure 3.11: Response against gate series resistor. 678 nm @ 0.125 mW, 2 μm diameter laser spot size. 1 MΩ series gate resistor. V_{gs}=-0.8 V, V_{ds}=0.5 V, 'o'=transistor edge, '+'=transistor middle.

Fig. 3.11 shows the photovoltaic self-biasing (or 'gate biasing') effect, as a function of series resistance. The higher curve, corresponding to the transistor edge, begins to plateau at higher resistances, as the gate approaches the forward bias regime. The edge clearly produces a higher gain for all resistor values. Due to the large series resistor required to produce a significant gain, high frequency operation is not possible and so photovoltaic gate biasing has not received much attention in the literature.

However, photovoltaic gate biasing is of interest in a low frequency GaAs X-Y addressable transistor array for solid-state imaging. The pixels in such an imager can be configured to optimise the edge gain mechanism, thereby increasing the sensitivity of the device. With careful development, a GaAs imager utilising a conventional planar MESFET technology may potentially offer increased radiation hardness, reduced dark current and the

convenience of integration with high speed GaAs image processing cells.

We have discovered a new finding that the optical gain sharply increases where the gate crosses the transistor edge. This only occurs with a series gate resistor inserted, to produce the conditions for photovoltaic gate biasing. This finding suggests that carriers generated in the substrate, beyond the transistor boundary, are able to be collected by the gate depletion region. This edge effect is observed in a planar GaAs MESFET and is quite different to the edge gain effect observed in mesa GaAs MESFET structures [Flesner 82, Rouger 85].

We proposed a generalised model for the photoresponse of a GaAs MESFET [Adibi 89] which considered a photovoltaic (PV) effect in the gate depletion region, a photoconductive (PC) effect in the channel and two main substrate effects (trapping by defect centres and substrate/channel depletion region collection). In order to explain the observed effects, in our present case, we are able to invoke the gate depletion region and substrate/channel depletion region parts of the model.

To explore the edge effect further, we set up a computerised measurement system to produce a complete 2 dimensional scan of the whole transistor region.

3.2.3.1 Two Dimensional Laser Scan Set Up

The planar MESFETs used have a fingered structure with 5 gates, $L = 0.8$ μm and overall $W = 400$ μm. The channel (depth, $d = 0.1$ μm and doping, $N_d = 1.2\times10^{17}$ cm^{-3}) is situated on a semi-insulating (SI) GaAs:EL2 substrate, with no buffer layer. The source-gate and drain-gate separations are both 1.3 μm.

The device was mounted on a computerised X-Y platform, shown in Fig. 3.12, and illuminated by a 2 μm diameter CW laser spot with a wavelength of 678 nm. The laser power incident on the device was measured to be 1.4 μW. The X-Y platform was controlled to move the device through a 2-D raster sequence and the drain current was automatically logged so that a 3-D plot of the transistor response was generated.

3.2.3.2 Two Dimensional Scan Results

Example 3-D plots of transistor drain current response to laser illumination are shown in Figs. 3.13 & 3.14. For comparison, the case for no series gate resistor is shown in Fig. 3.13. Peak drain current is in the central region of the transistor, and the drain current rolls off towards the periphery of the transistor, as classically expected. However, with a 10 MΩ resistor in series with the gate, Fig. 3.13 shows dramatic drain current

Figure 3.12: Laser Prober with Computerised X-Y Platform

peaks at the positions where gate metal crosses the transistor edge. The five peaks, along an edge, correspond to the five gates. There are two sets of five peaks, as the gates overlap the transistor at both ends, of course.

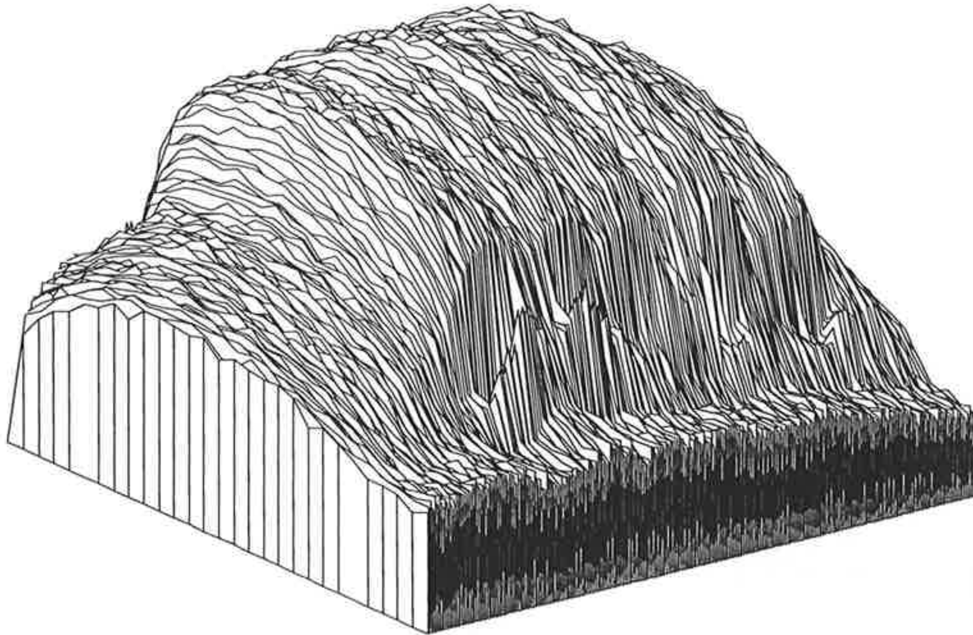


Figure 3.13: MESFET drain current response to 2-D laser scan, max. current 3.7 mA. No gate resistor.

These plots were repeated for a number of devices and it was observed that the peak heights did not vary substantially, within each set of five. However, there were some cases when a set peaks at one end of the transistor was as much as four times the height as the set at the opposite end. This effect could not be correlated with any visible transistor layer misalignment.

Note that although the gates are opaque, the peaks actually correspond to a position where the laser spot is exactly centred on the gate (Fig. 3.15). This is because the spot diameter is larger than the gate length and thus a central position maximises the amount of light to the gate edges, where the depletion region extends outwards, from under the gate, and is exposed. Also holes diffusing in the channel, that are generated close to the gate edges, can be collected by the gate depletion region. Even though the diffusion length

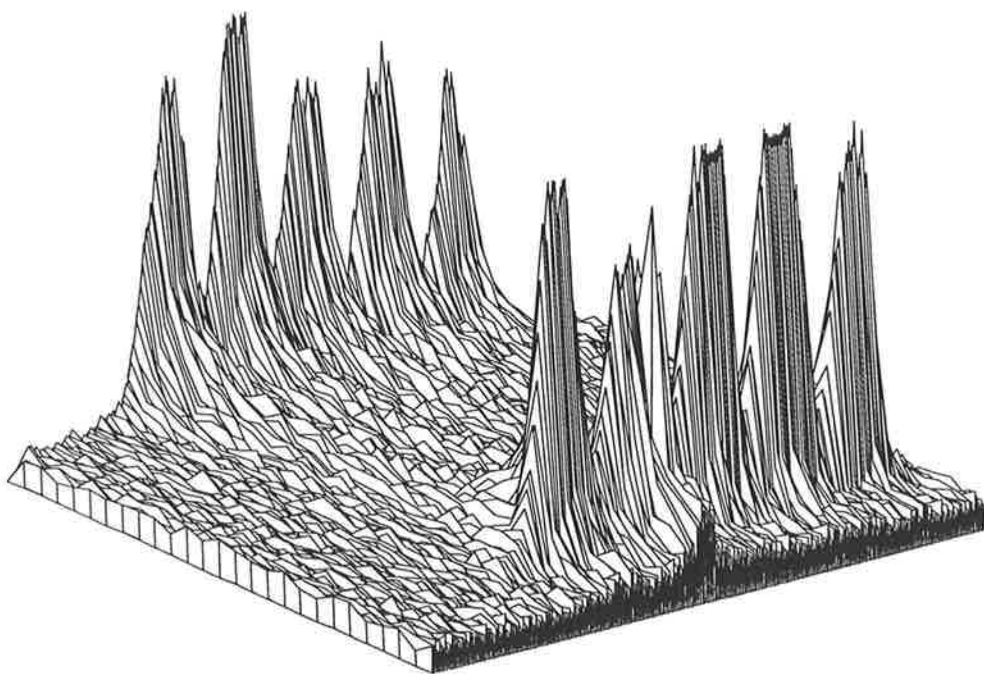


Figure 3.14: MESFET drain current response to 2-D laser scan, max. current 29 mA. With 10 MΩ gate resistor.

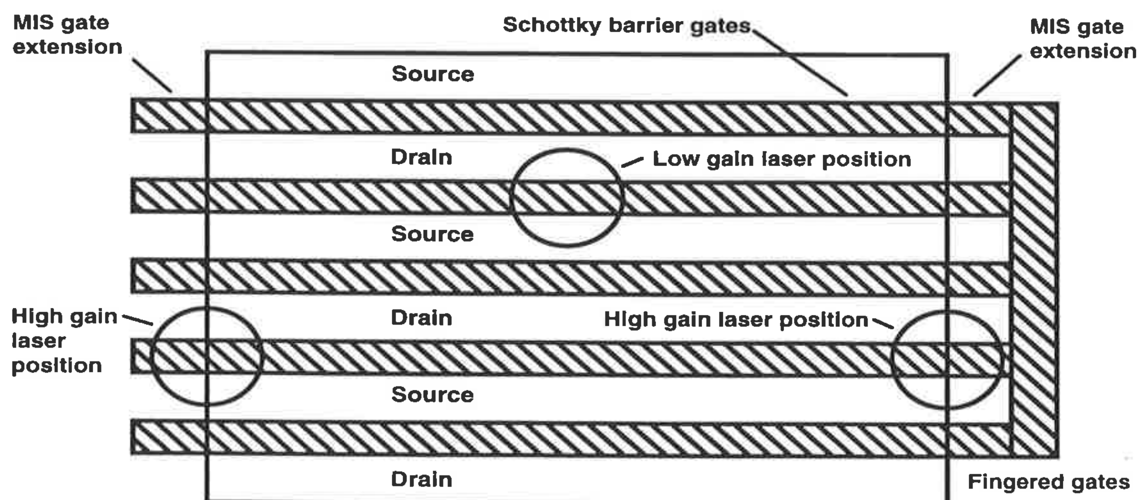


Figure 3.15: Laser spot positions for high & low gain drain current response.

of holes in the channel is 1.8 μm, holes that are far from the gate are more likely to be collected by the channel/substrate depletion region because the channel depth $d \ll 1.8$ μm – therefore there is a very narrow capture angle for gate collection.

3.2.3.3 Discussion

We established that the peak drain current could be accounted for by an increase in gate photocurrent, as measured by an electrometer. Fig. 3.18 shows gate photocurrent versus R_g, when the laser spot is centrally placed on the gate, for a position at the transistor edge (high gain) and towards the middle of the transistor (low gain). The theoretical curve is found by following the front illumination case of Seib [Seib 74], to evaluate the quantum efficiency, η. However, in a GaAs MESFET, the boundary conditions change and the equations are now solved with hole concentration $p = 0$ at the substrate/channel interface ($y = d$), instead of at $y = \infty$. This results in,

$$\eta = T(\lambda)A_o\{1 - \exp(-\alpha W_d) - A + B - C\} \tag{3.1}$$

where

$$A = \frac{\alpha^2 {L_o}^2 \exp(-\alpha W_d)}{1 - \alpha^2 {L_o}^2}$$

$$B = \frac{A}{\alpha L_o} \frac{1 - \exp\{-(d - W_d)(\alpha + \frac{1}{L_o})\}}{1 - \exp\{\frac{-2(d-W_d)}{L_o}\}}$$

$$C = \frac{A}{\alpha L_o} \frac{1 - \exp\{-(d - W_d)(\alpha - \frac{1}{L_o})\}}{1 - \exp\{\frac{2(d-W_d)}{L_o}\}}$$

and after some manipulation, this elegantly reduces to,

$$\eta = T(\lambda)A_o\{1 - \frac{e^{-\alpha W}}{1 + \alpha L_o} - \frac{\alpha L_o e^{-\alpha W}}{\alpha^2 L_o^2 - 1} \frac{e^{-h/L_o} - e^{-h\alpha}}{\sinh(h/L_o)}\} \tag{3.2}$$

which is identical to the finite slab equation given in Appendix B. L_o is the hole diffusion length in the channel and α is the absorption coefficient at the wavelength of interest. If we assume that the depletion region extension from under the gate is W_d and approximate that light is only captured in these regions, as a first order estimate, then the dimensionless geometry factor, $A_o = 2 \times 1.83 \times W_d/\pi$. The depletion region width, W_d, is given by the standard relation for a Schottky barrier,

$$W_d = [\frac{2\epsilon_o \epsilon_r}{qN_d}|V_{bi} - V_g - \frac{kT}{q}|]^{\frac{1}{2}} \tag{3.3}$$

where, the built-in voltage, $V_{bi} = 0.85$ V. If the external bias applied to R_g is -0.8 V, then the gate voltage, $V_g = -0.8 + I_g R_g$. By knowing the incident laser power, P_o, the

gate photocurrent is obtained from,

$$I_g = (\frac{q\lambda}{hc})\eta P_o \tag{3.4}$$

The theoretical curve for η, in Fig. 3.18, is obtained via an iterative solution of the above equations with the transmission coefficient set at $T = 1$, for simplicity. This is lower than the measured curve because our first order approximation has neglected sideways diffusion of holes into the gate. Another possibility is local heating effects, caused by the laser, have not been taken into account – however we have shown in Appendix C that this effect is negligible. Future work could consider a full three dimensional solution to the diffusion equation. We consider the following hypotheses in order to explain the increase in gate photocurrent that produces the sharp increase in drain current at the transistor edges:

HYPOTHESIS 1: *Gate metal thinning over the edge.* Possibly the gate metal thins at the edge, enough to become partially transparent, thereby letting in more light. This hypothesis is easy to reject, as such thinning would effectively increase the area factor A_o by 30% at most – this can only modulate the drain current by 2-3 mA at the most. We need an effect that is ten times that order.

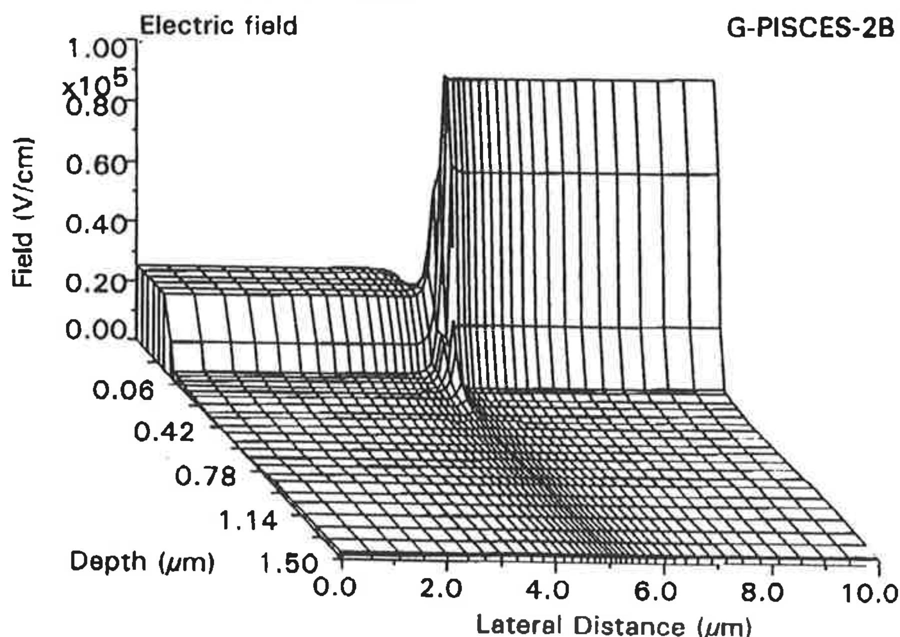


Figure 3.16: 3-D Electric field plot using G-PISCES-2B with Spicer's UDM. Left half is field under the gate overhang. Right half is field under the Schottky gate.

HYPOTHESIS 2: *Electric field at edge.* It is conceivable that the abrupt junction at the transistor edge is causing a region of high electric field that results in a higher current flowing through the gate. Figure 3.16 a 3-D electric field plot generated by G-PISCES-2B, that uses Spicer's unified defect model. The left half of the plot represents electric field

under the gate metal that overhangs the transistor and extends over Si_3N_4. The right half of the plot represents the field under gate Schottky contact. At the interface or edge between the two, the field is clearly well-behaved. Therefore we reject this hypothesis.

HYPOTHESIS 3: *Contiguous depletion regions.* From Figure 3.16, it can be seen that the depletion region that supports an electric field under the MIS gate overhang is contiguous with the gate depletion region. It is also quite obvious that the channel/substrate depletion region must also connect here at the transistor edge. Therefore the large volume of depletion region, just outside the transistor edge, at the meeting point of the gate, MIS and channel/substrate depletion regions greatly increases the photocollection capacity. Photocharge generated in this region can flow into the gate, via connecting field lines, giving rise to an increased gate photocurrent. A conceptual diagram of how the field lines possibly flow is shown in Fig. 3.17. In the last chapter we established that the channel/substrate depletion region is of the order of 2 μm and thus the resulting collection volume is large enough to explain the observed effect. Consequently this becomes our working hypothesis.

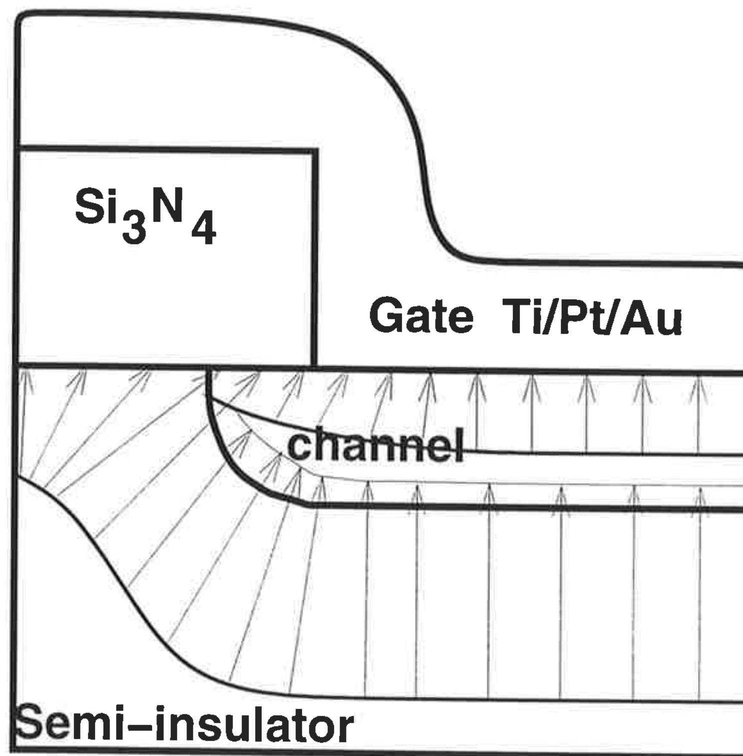


Figure 3.17: Convergence of Schottky gate, MIS gate and channel/substrate field lines at the transistor edge. Conceptual drawing with exaggerated features.

According to our working hypothesis, the high gain curve in Fig. 3.18 can be explained

if there are holes generated in the substrate, outside the transistor, being collected by the gate. Therefore we postulate the existence of a depletion region in the substrate that has connecting electric field lines with the edge of the gate depletion region. This new depletion region is assumed to be formed under the MIS structure, where the gate extends over Si_3N_4, outside the transistor. The channel/substrate depletion region must also connect at the edge and, given a depth of about 2 μm, we estimate that the depletion region would need to protrude from under the MIS electrode by about 0.2 μm to account for the high gain curve in Fig. 3.18.

In conclusion, we have reported a new optical edge gain effect in planar GaAs MESFETs, useful in low frequency applications, such as in an X-Y array imager. Photocollection under the MIS gate extension, outside the transistor, is suggested to explain the gain effect. Future models for the photoresponse in planar GaAs MESFETs, that attempt to generalise all operating conditions, must take into account this new effect.

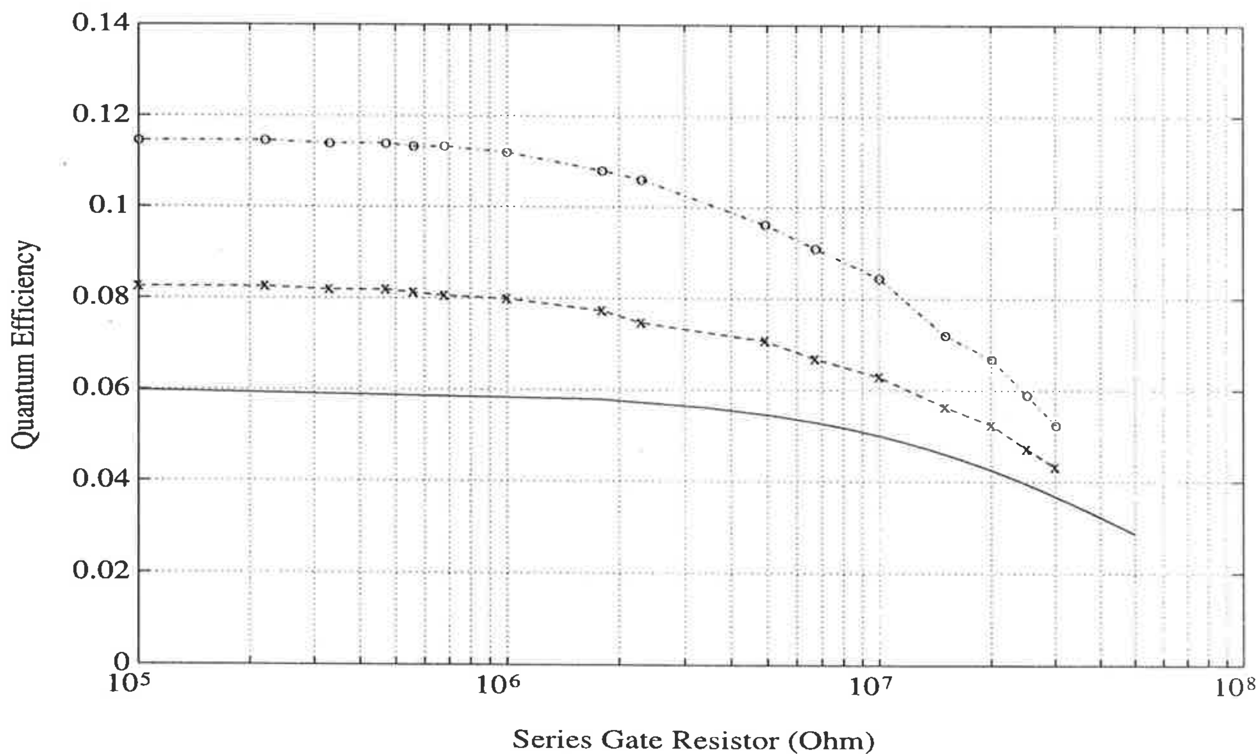


Figure 3.18: Quantum efficiency versus gate resistor. Solid line: internal quantum efficiency (T=1) from first order theory. Dashed line: measured (low gain). Chained line: measured (high gain).

3.2.3.4 Experimental Refinement

It can be seen that in Fig. 3.13, the peaks are not uniform and one peak appears to be 'smeared.' It was not known, at the time, whether the non-uniformity was fundamental to the edge effect or whether it was an artifact of the measurements. In an effort to further refine the experiment, to reduce the vibrational disturbance that was the likely cause of the smeared peak, we designed an antivibrational table. The table comprised a heavy marble top supported by a steel frame. The steel tubing, of the frame, was filled with glass beads to improve damping. The complete four-legged structure rested four separate industrial vibration isolator mounts. Each mount had an effective Hooke's Law spring constant of about 10^6 kg/m and the total supported mass was 387 kg. This results in a resonant frequency of the order of 100 Hz which is safely out of the 10-20 Hz region that is typical for building vibrations. For a rigorous analysis the torsional and shear resonant frequencies should also be evaluated – but this is a complex mechanical engineering problem beyond the scope of this work. However, it turned out the the table design was sufficient to produce the perfect result shown in Fig. 3.19. Notice that Fig. 3.13 and Fig. 3.19 are scans of the very same device. Not only has the smeared peak disappeared, but the peaks are now perfectly uniform. Hence, we can conclude that the peak non-uniformity in Fig. 3.13 was due to *mechanical* disturbance of the sample under test and thus is merely an artifact. Fig. 3.20 is a contour plot of the same data in Fig. 3.19. The contour plot clearly shows the integrity of the mechanical scanning, with no visible smearing of data.

3.3 Summary

In this chapter we have examined a number of issues surrounding photodetection with GaAs MESFETs. We have reviewed the controversy surrounding the modelling of the active MESFET photodetector and have presented Adibi's model in some detail. For the first time, we successfully measure the substrate time constant parameter in the Adibi model. We were unable to extract Adibi's channel/substrate time constant parameter, as future work must employ a laser with a higher modulation frequency than that presently available. We also discussed that by using a variable wavelength laser, in the future, a number of useful substrate parameters can be elegantly extracted using the Adibi model.

Another original contribution is that, with monochrometer and laser based measurements, we have established that the I-V hysteresis in MESFETs is effected by both the power and wavelength of incoming light.

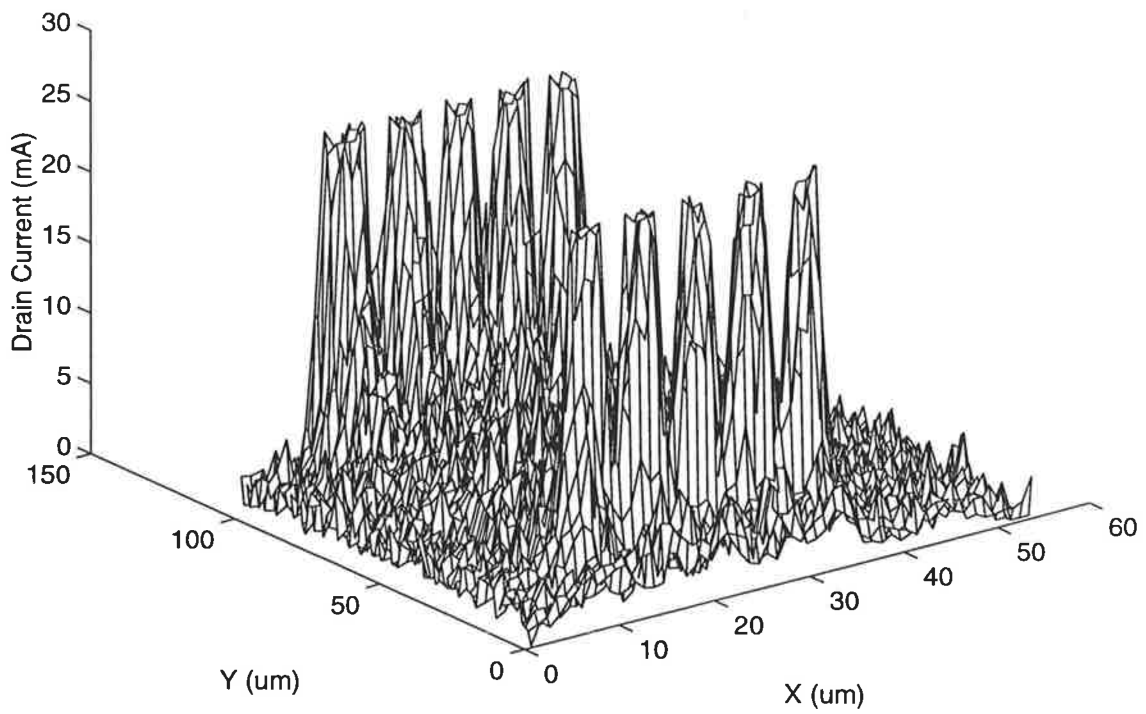


Figure 3.19: MESFET drain current response to 2-D laser scan, max. current 29 mA. With 10 MΩ gate resistor. 3-D plot. Performed on antivibration table.

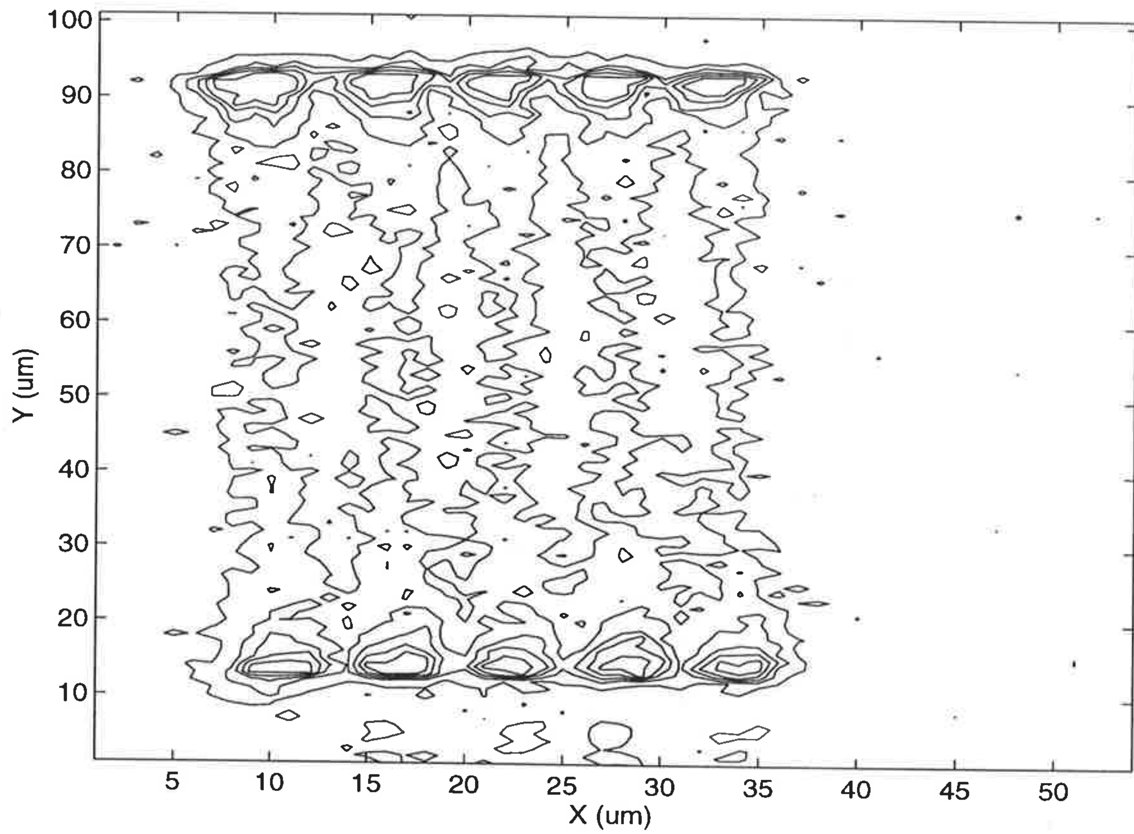


Figure 3.20: MESFET drain current response to 2-D laser scan, max. current 29 mA. With 10 MΩ gate resistor. Contour plot. Performed on antivibration table.

Some confusion exists over the nature of fields and depletion regions in semi-insulating substrates – however, we have clearly demonstrated, with a monochrometer, that photocollection in the channel/substrate depletion region does indeed take place. This is expected to enhance the performance of the proposed imager.

Another important original contribution is that we have discovered a new internal gain effect in GaAs MESFETs. As well as having implications for increasing the sensitivity of photodetectors we have exploited the gain effect to estimate that the carrier diffusion length in the semi-insulating substrate is less than 10 μm – the realisation of this fact is important for good spatial resolution in high definition television (HDTV) sensors.

We have discussed a number of hypotheses as to the origin of the internal gain edge-effect and the picture that emerges is that, where the gate crosses the edge of the transistor, there is a convergence of electric field lines from (a) the gate depletion region, (b) the MIS gate overhang depletion region and (c) the channel/substrate depletion region, giving rise to a larger photocollection volume. This results in a larger photocurrent that flows into the gate. To manifest this effect in a useful manner, a gate resistor is required so that the increased gate current modulates the drain current. This is the first time a detailed picture of the depletion region structure at the transistor edges has emerged. Future work may determine if this depletion region model, at the edge, explains any purely electrical parasitic effects in the GaAs MESFET.

The observation that there is some variation in the edge-effect from transistor to transistor and the discussion of the complicated active photoresponse of the MESFET immediately suggests an important design constraint for the imager: the photoresponsive element, in the first instance, should be a simple diode structure. Utilising a full active MESFET photodetector will present unwanted non-uniformities and thus fixed-pattern-noise (fpn) in the final imager. However, utilisation of the edge-effect would be appropriate in simple motion detectors where the signals are thresholded and thus pixel uniformity is not crucial – this possibility will be discussed in the penultimate chapter as part of our future vision.

Before we consider the design of an imager, in the light of this study on photodetectors, in the next chapter we turn to the important issue of understanding noise in the MESFET and how this will also impact on the imager design.

Bibliography

[Abbott 91] D. Abbott, "Gallium Arsenide MESFET Imager," *Prov. Patent*, PK8660, Aug. 1991.

[Adibi 89] A. Adibi and K. Eshraghian, "Generalised model for GaAs MESFET photodetectors," *IEE Proc.*, **vol. 136**, Pt. G, no. 6, pp. 337-343, 1989.

[Chaturvedi 83] G.J. Chaturvedi, R.K. Purohit and B.L. Sharma, "Optical Effect on GaAs MESFETs," *Infrared Phys.*, **vol. 23**, no. 2, pp 65-68, 1983.

[Darling 87] R.B. Darling and J.P. Uyemura, "Optical gain and large-signal characteristics of illuminated GaAs MESFETs," *IEEE J. Quantum Elec.*, **vol. QE-23**, no. 7, pp. 1160-1171, 1987. (refer p. 1167, para. 2)

[Edwards 80] W.D. Edwards, "Two and Three Terminal Gallium Arsenide FET Optical Detectors," *IEEE Elec. Dev. Lett.*, **vol. EDL-1**, no. 8, pp. 149-150, 1980.

[Edwards 82] W.D. Edwards and R.F. Haythornwaite, "Laser and Electron beam Scanning of GaAs FETs," *Microelectron. Reliab.*, **vol. 22**, no. 4, pp 735-746, 1982.

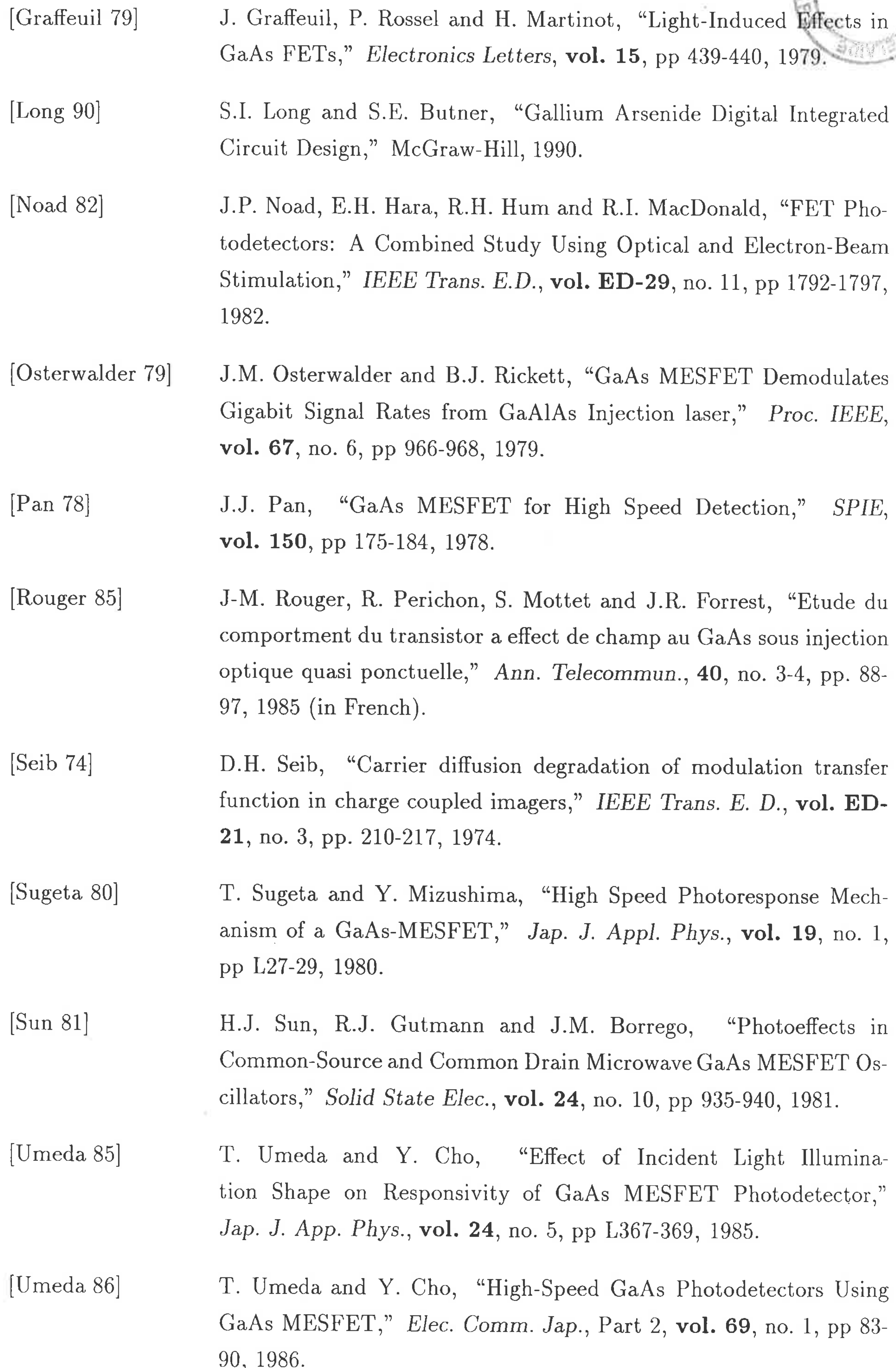
[Flesner 82] L.D. Flesner, N.M. Davis and H.H. Wieder, "High Speed Repsonse of a GaAs Metal-Semiconductor Field-Effect Transistor to Electron Beam Excitation," *J. Appl. Phys.*, **vol. 53**, no. 5, pp 3873-3877, 1982.

[Forrest 82] J.R. Forrest, F.P. Richards and A. Perichon, "The Microwave MESFET Optical Detector," *IEDM*, pp 529-532, 1982.

[Gammel 78] J.C. Gammel and J.M. Ballantyne, "The OPFET: a New High-Speed Optical Detector," *Proc. IEDM*, pp 120-123, 1978.

[Gammel 80] J.C. Gammel and J.M. Ballantyne, "Comments on 'High Speed Photoresponse Mechanism of a GaAs-MESFET,"' *Jap. J. Appl. Phys.*, **vol. 19**, no. 5, pp L273-275, 1980.

[George 89] P. George, P-K. Ko and C. Hu, "Modelling the substrate depletion region for GaAs FETs fabricated on semi-insulating substrates," *Solid-State Electronics*, **vol. 32**, no. 2, pp. 165-168, 1989.



[Graffeuil 79] J. Graffeuil, P. Rossel and H. Martinot, "Light-Induced Effects in GaAs FETs," *Electronics Letters*, **vol. 15**, pp 439-440, 1979.

[Long 90] S.I. Long and S.E. Butner, "Gallium Arsenide Digital Integrated Circuit Design," McGraw-Hill, 1990.

[Noad 82] J.P. Noad, E.H. Hara, R.H. Hum and R.I. MacDonald, "FET Photodetectors: A Combined Study Using Optical and Electron-Beam Stimulation," *IEEE Trans. E.D.*, **vol. ED-29**, no. 11, pp 1792-1797, 1982.

[Osterwalder 79] J.M. Osterwalder and B.J. Rickett, "GaAs MESFET Demodulates Gigabit Signal Rates from GaAlAs Injection laser," *Proc. IEEE*, **vol. 67**, no. 6, pp 966-968, 1979.

[Pan 78] J.J. Pan, "GaAs MESFET for High Speed Detection," *SPIE*, **vol. 150**, pp 175-184, 1978.

[Rouger 85] J-M. Rouger, R. Perichon, S. Mottet and J.R. Forrest, "Etude du comportment du transistor a effect de champ au GaAs sous injection optique quasi ponctuelle," *Ann. Telecommun.*, **40**, no. 3-4, pp. 88-97, 1985 (in French).

[Seib 74] D.H. Seib, "Carrier diffusion degradation of modulation transfer function in charge coupled imagers," *IEEE Trans. E. D.*, **vol. ED-21**, no. 3, pp. 210-217, 1974.

[Sugeta 80] T. Sugeta and Y. Mizushima, "High Speed Photoresponse Mechanism of a GaAs-MESFET," *Jap. J. Appl. Phys.*, **vol. 19**, no. 1, pp L27-29, 1980.

[Sun 81] H.J. Sun, R.J. Gutmann and J.M. Borrego, "Photoeffects in Common-Source and Common Drain Microwave GaAs MESFET Oscillators," *Solid State Elec.*, **vol. 24**, no. 10, pp 935-940, 1981.

[Umeda 85] T. Umeda and Y. Cho, "Effect of Incident Light Illumination Shape on Responsivity of GaAs MESFET Photodetector," *Jap. J. App. Phys.*, **vol. 24**, no. 5, pp L367-369, 1985.

[Umeda 86] T. Umeda and Y. Cho, "High-Speed GaAs Photodetectors Using GaAs MESFET," *Elec. Comm. Jap.*, Part 2, **vol. 69**, no. 1, pp 83-90, 1986.

[Wieder 80] H.H. Wieder, N.M. Davis and L.D. Flesner, "Electron Beam Modulation of GaAs Metal-Semiconductor Field-Effect Transistors," *Appl. Phys. Lett.*, **vol. 37**, no. 10, pp 943-945, 1980.

Chapter 4

Noise

"The paradoxical character of truth is in its objective uncertainty."
Søren Kierkegaard (1813-1855)
DANISH FOUNDER OF EXISTENTIALIST PHILOSOPHY

'Nothing in nature is random..... A thing appears random only through the incompleteness of our knowledge."
Benedict de Spinoza (1632-1677)
DUTCH PHILOSOPHER

"There is no absolute knowledge.... all information is imperfect. We have to treat it with humility."
Jacob Bronowski (1908-1974)
POLISH-BRITISH MATHEMATICIAN AND POET

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them."
William Henry Bragg (1862-1942)
BRITISH SOLID-STATE PHYSICIST

"La chose importante, c'est la théorie, qui est mind-blowing. Et si j'ai raison, ma théorie va produire un crises en world thinking et, avec luck, un Prix Nobel."
Miles Kington (1941-)
BRITISH EXPONENT OF FRANGLAIS

4.1 Introduction

In order to optimise the design of the imager output circuitry a careful noise analysis of the output is necessary. Given that noise is generally a poorly understood area and the literature tends to be obscure, we discuss a number of general issues, paradoxes and misconceptions with thermal noise. This then sets the framework for full noise analysis of the imager.

Firstly, this chapter takes a detour to present historical background and then discuss a new derivation of the thermal noise formula. This leads to discussion of a number of interesting paradoxes & misunderstandings surrounding this formula. We then describe an experimental set-up for measuring noise that will be required for future work to measure and characterise noise in the imager. Initial thermal noise tests on the equipment are shown to give results that challenge accepted theory. Analysis of this problem makes use of the new thermal noise derivation presented here.

This backbround work finally allows a discussion, with greater clarity, of noise problems in the design of the imager output circuit and makes comparisons with silicon.

4.2 A Brief Early History of Noise

Thermal noise caused by electrons jostled within a conductor's lattice is an electrical analogy of Brownian motion. The random motion of particles in a fluid is named after R. Brown in recognition of his work in 1827 [Brown 28]. Inspired by discoveries following a historic voyage to Australia [King 27], he was led to closely examine the structure of pollen under a microscope, where upon he became intrigued by their random motion in a fluid. Brown was not the first to see such motion, in fact, many had before such as W. F. Gleichen, J. T. Needham, G-L. Leclerc, A-T. Brogniart and L. Spallanzani [Brown 28, Brown 29], except that their ability for correct interpretation was clouded midst the ongoing debate on vitalism and spontaneous generation. Brown opened the door for research in microscopic fluctuations, by being the first to perform a major systematic experimental analysis convincingly demonstrating that the motion was not due to bubbles, release of matter, interaction between the particles themselves or organisms. It is interesting to note, however, that J. Ingen-Housz in 1784 [Ingenhousz 84] and then J. Bywater in 1819 [Bywater 19] independently came to the conclusion that the motion exists for inorganic particles, before Brown. For the next half a century a number of scientists including H. V. Regnault, L. C. Wiener [Wiener 63], Cantoni & Oehl and S. Enxer [Exner 67] debated the cause of the fluctuations, on whether it was heat, light,

or electrical forces. Finally, in 1877, R. J. Delsaulx for the first time suggested impact of liquid molecules on the particles [Delsaulx 77]. Evidence to support this hypothesis came gradually. The work of L-G. Gouy [Gouy 88] showed the motion slowed down in more viscous fluids and he made the further step of ascribing fluctuations to *thermal* motion of the fluid molecules. F. M. Exner, in 1900, established that the activity decreased with increasing particle size and decreasing temperature [Exner 00]. About this time M. R. Smoluchowski began theoretical work on the subject and published papers in the 1904-06 period. A. Einstein independently wrote a number of famous theoretical papers in 1905-07. The first theoretical discussion of electrons as Brownian particles [de Hass-Lorentz 13] came as early as 1912, by G. L. de Haas-Lorentz, [1] which inspired G. A. Ising, in 1925, to fully explain the problem of galvanometer fluctuations [Ising 26] observed by Moll & Burger [Moll 25]. With J. J. Thomson's discovery of the electron, in 1897, and P. K. L. Drude's classical model of electrical conduction in terms of an electron gas in an atomic lattice, both well established by this stage, the accumulated knowledge was ripe for the understanding of electrical noise.

J. B. Johnson (Fig. 4.1), drawing inspiration from W. Schottky's work [Schottky 18] of 1918, began during 1925 to characterise the thermal noise in various conductors via a vacuum tube amplifier, and published in 1927-28 his well-known formula [Johnson 28] for voltage noise, which is equivalent to Einstein's fluctuation formula for Brownian motion of charge. Johnson discussed his results with H. Nyquist (Fig. 4.2) who, about a month later, managed to produce a remarkably compact theoretical derivation based on the thermodynamics of a transmission line [Nyquist 28].

Because of the equivalence of Johnson's formula with the earlier theory (see also [Blum 73]), some authors prefer to use the neutral term *thermal noise*, whereas some prefer *Johnson noise* or *Johnson-Nyquist noise* to prevent the confusion between electrical thermal noise and *temperature fluctuations*. Similarly W. S. Jevons in 1878 attempted to coin the phrase *pedesis* (Gk. 'jump') [Jevons 78] as a neutral expression for Brownian motion; however, tradition prevailed. For a brief chronology see Table. 4.1.

[1] She was the eldest daughter of the physicist H. A. Lorentz, married his assistant W. J. de Haas and has the distinction of being the first woman in noise theory. In his 1912 series of lectures, H. A. Lorentz expounded her work within a statistical thermodynamics framework [Lorentz 16].

Figure 4.1: John [Erik] Bertrand Johnson (*né* Johan Erik Bertrand) was born in the Carl Johan parish of Goteborg, Sweden, on October 2nd, 1887 and christened October 7th, 1887. His birth certificate only records his mother Augusta Mathilda Johansdotter (b. 1866) and his surname derives from his assumed father Carl Bertrand Johansson. He emigrated to the USA in 1904 and attended Yale at the same time as Nyquist, producing a thesis entitled "*Total Ionisation of Slow Electrons*" in 1917. Johnson was a pioneer in the study of cathode ray tubes and studied the causes of noise in vacuum tubes in the 1925-30 period, working at Bell Telephone Laboratories until 1952. He then joined the Edison Research Laboratory until his retirement in 1969. He received a number of awards and medals and held over 30 patents. Johnson became a US citizen in 1938, was a Republican and Presbyterian. His interests included opera, plant life and woodwork. He married Clara Louisa Conger (d. 1961) in 1919 and Ruth Marie Severtson Bowden in 1961. By his first marriage he had two sons, Bertrand Conger and Alan William. Bert Johnson died at the age of 83 in Orange, NJ, on November 27th, 1970. *Photo permission: AT&T Archives.*

Figure 4.2: Harry Nyquist (*né* Harry Theodor Nyqvist) was born in the parish of Stora Kil in the county of Varmland, Sweden, February 7, 1889, the son of Lars Jonsson Nyqvist (b. 1847) and Katrina Eriksdotter (b. 1857). There were seven children altogether: Elin Teresia, Astrid, Selma, Harry Theodor, Amelie, Olga Maria and Axel; none of whom were christened. Harry emigrated to the USA in 1907 and, attending Yale, his 1917 thesis was on the Stark effect and therefore he would have been aware of the work of H. A. Lorentz; however, no historian has yet established if Nyquist knew of Lorentz's 1912 work [Lorentz 16] on the statistical thermodynamics of noise. Nyquist began working with the AT&T Company c.1917 and went on to produce 138 patents in the area of telephone and television transmission, as well as collecting many honours and awards. He arrived at his derivation of the thermal noise formula in about a month after discussions with Johnson. He is also credited with the Nyquist diagram, for defining stable conditions in negative feedback systems, and the Nyquist sampling theory in digital communications. Harry Nyquist was unique in that he was famous as a theoretician and yet was a prolific inventor. He retired in 1954, though continued as a consultant, and died at the age of 87 on April 4th, 1976, Harlingen, Texas. *Photo permission: AT&T Archives.*

Name	Background	b.-d.	Origin	Observation	Date
Sacharias Jansen	Optician, Coin forgery	1588-c.1631	Dutch	Invented compound microscope with father	$\leq$1609
Hans Lipperhey	Spectacle maker	c.1570-1619	Dutch	Independently invented compound microscope	$\leq$1609
Antoni van Leeuwenhoek	Anatomy, Microscopy	1632-1723	Dutch	Extensive microscope observations	-
William Derham	Bishop, Physician	1657-1735	English	Observed animacules in pepper water	1713
Georges Louis Leclerc (Comte de Buffon)	Naturalist	1707-1788	French	Observed the motion before Brown	-
John Turberville Needham	Naturalist, Clergyman	1713-1781	Eng-Bel	Observed the motion before Brown	-
Wilhelm F. von Gleichen-Russworm	Naturalist	1717-1783	German	Observed the motion before Brown	-
Lazzaro Spallanzani	Naturalist, Jesuit	1729-1799	Italian	Observed the motion before Brown	-
Jan Ingen-Housz	Physics, Medicine	1730-1799	Dut-Eng	Observed inorganic particle motion	1784
John Bywater	Optician, Philos.	c.1774-1836	English	Observed inorganic particle motion	1819
Robert Brown	Botany	1773-1858	Scottish	First systematic study of the motion	1827
Henri Victor Regnault	Physics, Chemistry	1810-1878	French	Thought the motion was due to light	1858
(Ludwig) Christian Wiener	Math., Physics, Philos.	1826-1896	German	Discarded evaporation as an explanation	1863
Cantoni & Oehl	Physics		Italian	Found motion persists after a year	1865
Sigmund Exner	Medicine, Physiol.	1846-1926	German	Found smaller particles move quicker	1867
(Rene) Joseph Delsaulx	Math., Physics, Priest	1828-1891	Belgian	First to suggest molecular impact	1877
William Stanley Jevons	Logic, Economics	1835-1882	English	Tried to coin the term *pedesis*	1878
Carl Wilhelm von Nageli	Botany, Microscopy	1817-1891	Swiss	Incorrectly discards molecular impact idea	1879
William Miller Ord	Anatomy	1843-1902	English	Argues against electrical cause	1879
Louis-Georges Gouy	General Physics, Optics	1854-1926	French	Motion more rapid if viscosity lowered	1888
(Richard) Meade Bache	Physics	c.1830-1907	USA	Motion persists after a week in darkness	1894
Joseph John Thompson	Physics	1856-1940	English	Discovered the electron	1897
Paul Karl Ludwig Drude	Physics	1863-1906	German	Electron gas model of conduction	1900
Felix Maria Exner	Meteorology	1876-1930	German	Motion increases with temperature	1900
Louis Jean Baptiste Alphonse Bachelier	Mathematics	1870-1946	French	Analyzed fluctuations in Paris stock exchange first to apply theory to Brownian motion	1900
Jean Baptiste Perrin	Physics	1870-1942	French	Began systematic experiments.	1900
Marian Ritter von Smolan Smoluchowski	Physics	1872-1917	Polish	First systematic theory began	1900
Albert Einstein	Physics	1879-1955	Ger-USA	Began publishing famous theoretical papers	1905
Geertruida Luberta de Haas-Lorentz	Physics	1885-1973	Dutch	First to discuss electrical noise and first woman in noise theory	1912
Hendrik Antoon Lorentz	Physics	1853-1928	Dutch	Statistical thermodynamics framework	1912
Walter Schottky	Physics	1886-1976	German	Classic paper on electrical noise	1918
Moll & Burger	Physics		Dutch	Amplified galvanometer fluctuations	1925
Gustav Adolf Ising	Physics	1883-1960	Swedish	Correctly explained galvanometer noise	1926
(John) Bert(rand) Johnson	Physical electronics	1887-1970	Swd-USA	Began work on circuit noise	1925
Harry (Theodor) Nyquist	Comms. Engineering	1889-1976	Swd-USA	Transmission line based derivation	1927
Norbert Wiener	Mathematics	1894-1964	USA	Began mathematical formalism	1928

Table 4.1: History of fluctuation research – a brief early chronology.

4.3 A New Derivation of the Thermal Noise Formula

A simple theoretical derivation for obtaining the Johnson thermal noise formula using window limited Fourier transforms is presented in detail for the first time, utilising the well known energy theorems. In the literature, a diverse range of alternative methods already exist and the pedagogical value of the Fourier transform approach is that it illustrates useful mathematical principles, taught at the undergraduate level, naturally highlighting a number of physical assumptions that are not always clearly dealt with.

As we shall see, this new transform technique for noise analysis clearly reveals the origin of the coefficient in the thermal noise formula. This will become useful when we later discuss our noise measurement apparatus for the imager – after careful calibration of the system we find surprising new evidence to challenge conventional theory. The simplicity of this transform technique also may lead to future application in analysing various noise problems in photodetector and imaging systems.

4.3.1 Overview of Methods in the Literature

The three most common methods found in the pedagogical literature for the derivation of Johnson's formula are (a) Nyquist's original proof [Nyquist 28] considering a transmission line in thermal equilibrium [Buckingham 83, Kittel 58, Lawson 50, Robinson 74, King 66, Beck 76, Bell 60], (b) a sharply tuned LCR circuit in thermal equilibrium [Wannier 66], and (c) the autocorrelation function technique [Beck 76, Bendat 58, Davenport 58].

Other techniques concentrate on starting from individual particle motion include the (a) Langevin equation approach considering particle mobilities and possible use of the Wiener-Khintchine theorem [Buckingham 83, Kittel 58], (b) kinetic theory derivation using the simple Drude model picture of conduction in metals in terms of a classical electron gas [Lawson 50, Freeman 58], (c) extension of this approach considering the modern Fermi-Dirac gas model of electron conduction [Dekker 91], and (d) a further generalised statistical proof independent of whether particles are classical or quantum [Robinson 74].

This remarkable diversity of proofs allows the pedagogue to draw upon whichever suits the particular course material at hand. However, none of the cited references present a proof in terms of Fourier transforms, making use of the well known *energy theorems*. The notion that the energy in a stationary random process is infinite is partially responsible for this omission. As pointed out by [Buckingham 83, King 66] the use of Fourier transforms is

nevertheless permissible as the *power* is finite, however they do not pursue the matter any further. Therefore, for the first time, we shall detail a proof using the Fourier transform energy theorems by considering them in terms of power.

Nyquist's original derivation has been criticised as it only considers *TEM* modes and part of the proof involves shorting out the resistors leaving an unanswered question of upset thermal equilibrium. The proof can be modified to overcome such objections [Blum 73, Beck 76], at the expense of brevity. A further pedagogical objection is that the Nyquist proof and the tuned LCR proof explicitly say very little about the statistical assumptions of the noise process; a list of further objections is given by [Moullin 38]. The present alternatives are either lengthy Lévy-Khintchine-Paley-Wiener type formalisms or kinetic theory. Thus our aim is for a simple 'engineering proof' based on Fourier transforms.

Section 4.3.2 introduces the lumped circuit model, section 4.3.3 discusses windowed Fourier transform concepts, section 4.3.4 derives Johnson's formula and the following sections review a number of the conundrums, debates and anomalies surrounding thermal noise that are generally not clearly discussed in the literature.

4.3.2 The Lumped Model

Consider a resistor in parallel with a capacitor. Any segment, dx, of this circuit loop consists of some continuous conducting medium, that has some finite resistance, eg. the resistor material, the metal wires or the capacitor dielectric. The electrons in these materials will experience random velocity fluctuations, due to thermal energy in the material. This Brownian-like motion of charge, leads to a voltage fluctuation across each segment of the circuit. In a given instant of time, the sum of this ensemble of fluctuations forms a net voltage e_n.

Assumption 1: This circuit is modeled in Fig. 4.3, by a random voltage generator $e_n(t)$, a pure capacitor C, a lumped resistor R and resistance free wires. Note that the capacitor is pure, so there is space between the plates, and therefore we expect the thermal noise formula to be independent of C.

4.3.3 The Windowed Fourier Transform (WFT)

By definition the Fourier transform $I_n(\omega)$ of the fluctuating current $i_n(t)$ through the circuit, as shown in Fig. 4.3, is given by

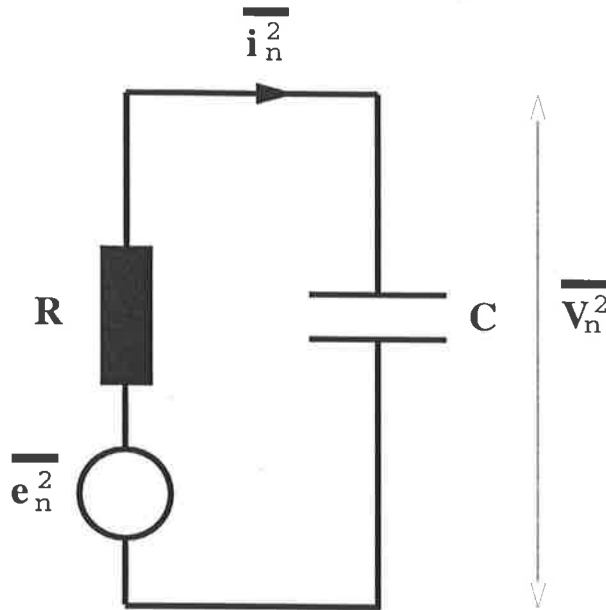


Figure 4.3: Lumped Circuit Model.

$$I_n(\omega) = \int_{-\infty}^{\infty} i_n(t) e^{-j\omega t} dt. \quad (4.1)$$

The Fourier transform is a useful tool for this noise problem, as undergraduate level students are widely taught convenient *energy theorems* – including how to express voltage power spectral density in terms of the transform. So the approach that we present here is to use the transforms to find an expression for the power spectrum in the noise generator in terms the power spectrum in the capacitor. This expression can then be reduced in terms of mean squared voltage fluctuations, followed by the standard equipartition theory arguments to finally yield the thermal noise formula.

However, $I_n(\omega)$ can only exist if $i_n(t)$ is absolutely integrable, ie.,

$$\int_{-\infty}^{\infty} |i_n(t)| dt < +\infty.$$

Unfortunately, this condition is not satisfied as $i_n(t)$ is a randomly varying function of time and does not decay to zero as $t \rightarrow \pm\infty$. The instantaneous values of $i_n(t)$ cannot be predicted and this type of function represents an example of a *stochastic process.*

A common error is to ignore the integrability problem from the outset and to proceed to express $I_n(\omega)$ in terms of $V_n(\omega)$, the transform of $v_n(t)$, by observing that

$$I_n(\omega) = \int_{-\infty}^{\infty} \frac{d(Cv_n)}{dt} e^{-j\omega t} dt$$

an then integrating by parts

$$I_n(\omega) = C[v_n e^{-jwt}]_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} (-j\omega C) v_n e^{-j\omega t} dt$$

which yields the 'correct' result

$$I_n(\omega) = j\omega C V_n(\omega)$$

by making a second mistake in assuming that $V_n(\omega)$ exists and the square bracket residual vanishes at the limits, which is clearly unfounded as the limiting values are unknown. At this point disapproval can be expressed, and the concept of windowing the function can be introduced. In practice, the random process can only be observed for a finite window of time τ, so a dimensionless time window [2] function $W(t, \tau)$ is defined in Fig. 4.4 and, provided a large τ is chosen to minimise statistical sampling error, the windowed version of equation 4.1 becomes

$$I_n(\omega, \tau) = \int_{-\infty}^{\infty} i_n(t, \tau) e^{-j\omega t} dt. \tag{4.2}$$

As the measured voltage v_n is only known over the observation time τ, it is tempting to define the windowed current as,

$$i_n(t, \tau) = \frac{d}{dt}[C v_n(t) W(t, \tau)].$$

Assumption 2: The stochastic process is independent of where the origin of $W(t, \tau)$ is placed on the time axis. This condition is referred to as *stationarity* and is reasonable in view of the observed nature of thermal noise.

This useful property means that 4.2 is invariant to the positioning of $W(t, \tau)$. Although the presence of W in 4.2 solves the integrability problem, it introduces the artifact of spectrum leakage. The leakage occurs essentially because the transform of W, itself, contains a continuous range of non-zero frequency components. This can be ignored as we will eventually be considering just the limiting case as $\tau \rightarrow \infty$. Another potential problem is that the differential term appears to create a discontinuity artifact at the edges of the window, however by substitution into 4.2 one finds that it eventually cancels, as follows

[2] Some texts, such as [Buckingham 83], replace a random variable $x(t)$ by a 'gated' random variable $x_T(t)$ which is zero outside the time window T. A Fourier transform is then performed on $x_T(t)$. Although, as we show, this does eventually lead to the correct result, the *implicit* window is introduced without discussion and the student is justified in questioning what becomes of any edge discontinuity effects. For this reason we introduce an *explicit* window function $W(t, \tau)$ as a 'book keeping' device to track and monitor the artifact.

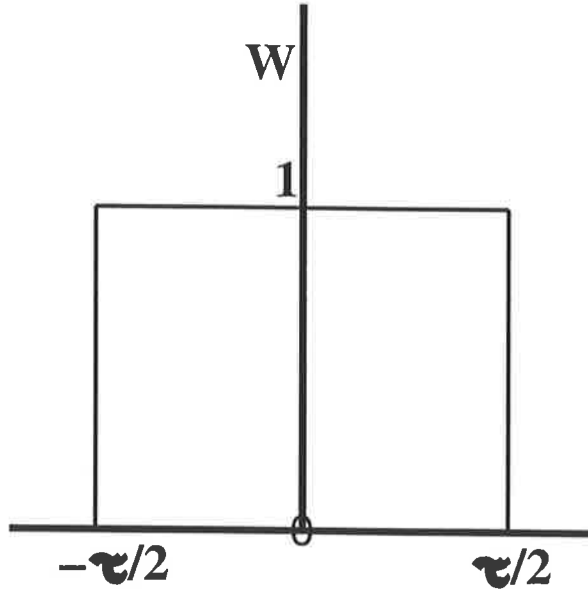


Figure 4.4: The window function.

$$I_n(\omega,\tau) = \int_{-\infty}^{\infty} \frac{d}{dt}[Cv_nW]e^{-j\omega t}dt \tag{4.3}$$

by expanding the differential we obtain,

$$I_n(\omega,\tau) = C\int_{-\infty}^{\infty} \frac{dv_n}{dt}We^{-j\omega t}dt + C\int_{-\infty}^{\infty} \frac{dW}{dt}v_ne^{-j\omega t}dt$$

integrating the left hand integral by parts,

$$\begin{aligned} I_n(\omega,\tau) &= C[v_nWe^{-j\omega t}]_{-\infty}^{+\infty} \\ &- C\int_{-\infty}^{\infty} v_n\left\{\frac{dW}{dt}e^{-j\omega t} - j\omega We^{-j\omega t}\right\}dt \\ &+ C\int_{-\infty}^{\infty} \frac{dW}{dt}v_ne^{-j\omega t}dt \end{aligned}$$

where the residual now legitimately vanishes, and the superfluous differential window terms cancel, again giving the 'correct result'

$$I_n(\omega,\tau) = j\omega CV_n(\omega,\tau).$$

Although this attempt has greater merit, as the integrals are now legitimate, it can be argued that the invoked definition of windowed current was somewhat *ad hoc*. Notice

as $\tau \to \infty$, $i_n(t,\tau)$ does not tend to $i_n(t)$ due to delta functions at the extremities. The artificial introduction of these spikes, from the outset, fortuitously cancels with the differential window term during the integration by parts, thereby producing the 'correct' result. Again disapproval must be expressed and the preferred method now follows.

Consider a windowed version of $e_n(t)$ as

$$e_n(t,\tau) = e_n(t)W(t,\tau).$$

We now define the signals $v_n(t,\tau)$ and $i_n(t,\tau)$ as the *responses* of the circuit to $e_n(t,\tau)$. Notice this time that as $\tau \to \infty$, $v_n(t,\tau) \to v_n(t)$ and $i_n(t,\tau) \to i_n(t)$. We have $i_n(t,\tau) = \frac{d}{dt}[Cv_n(t,\tau)]$, thus

$$I_n(\omega,\tau) = \int_{-\infty}^{\infty} \frac{d}{dt}[Cv_n(t,\tau)]e^{-j\omega t}dt$$

and integrating by parts,

$$I_n(\omega,\tau) = C[v_n(t,\tau)e^{-j\omega t}]_{-\infty}^{+\infty} + j\omega C \int_{-\infty}^{\infty} v_n(t,\tau)e^{-j\omega t}dt$$

which reduces to

$$I_n(\omega,\tau) = j\omega C V_n(\omega,\tau) \tag{4.4}$$

Although this result is trivial, it was important to show that there were no window artifact problems. General discussion of window problems can be found in [Papoulis 84] and [Dym72]. As $\frac{d^mW}{dt^m} = 0\Big|_{t=\pm\infty}$ it can be shown for the general case that

$$\frac{d^m v_n(t,\tau)}{dt^m} \leftrightarrow (j\omega)^m V_n(\omega,\tau)$$

is a windowed or time-limited Fourier pair.

4.3.4 The Thermal Noise Formula

We now proceed to use the eqn. 4.4 result to find the power spectrum in the capacitor and then produce the celebrated Johnson formula. Consider the voltages around the loop, in Fig. 4.3, by Kirchhoff

$$v_n(t) + i_n(t)R = e_n(t)$$

and viewed from the window $W(t,\tau)$ this becomes

$$v_n(t,\tau) + i_n(t,\tau)R = e_n(t,\tau).$$

Notice that by definition, $e_n(t,\tau)$ does not contain any delta function terms and therefore $i_n(t,\tau)$ and $v_n(t,\tau)$ must also be free of spikes. This can simply be demonstrated by *reductio ad absurdum*: if $i_n(t,\tau)$ contained a delta function pair, due to windowing, then $v_n(t,\tau)$ would need an identical pair of opposing sign to balance the above equation. But this would be impossible as $i_n(t,\tau)$ would then contain the second derivative and the reasoning continues inductively *ad infinitum*.

Now taking Fourier transforms we have

$$V_n(\omega,\tau) + I_n(\omega,\tau)R = E_n(\omega,\tau)$$

substituting in eqn. 4.4 gives

$$V_n = \frac{E_n}{1 + j\omega RC} \tag{4.5}$$

multiplying by complex conjugates

$$|V_n|^2 = \frac{|E_n|^2}{1 + (\omega RC)^2}. \tag{4.6}$$

By conservation of energy, the total energy in the time domain must equal that in the frequency domain, therefore

$$\int_{-\infty}^{\infty} e_n^2(t,\tau)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |E_n(\omega,\tau)|^2 d\omega.$$

This is known as *the energy theorem* or Plancherel's theorem (a special case of Parseval's theorem). Each side of the equation represents *total energy* and therefore $|E_n|^2$ represents the *energy density* with units of V^2s/Hz. Due to the Hermitian property of the Fourier transform, $|E_n|^2$ is always even therefore, we can write the one-sided form

$$\int_{-\infty}^{\infty} e_n^2(t,\tau)dt = \frac{1}{2\pi}\int_{0}^{\infty} 2|E_n(t,\tau)|^2 d\omega.$$

By definition of temporal average

$$\begin{aligned}\langle e_n^2 \rangle &= \lim_{\tau\to\infty} \langle e_n^2 \rangle_\tau \\ &= \lim_{\tau\to\infty} \frac{1}{\tau}\int_{-\infty}^{\infty} e_n^2(t,\tau)dt\end{aligned}$$

therefore

$$\langle e_n^2 \rangle = \lim_{\tau \to \infty} \frac{1}{2\pi} \int_0^\infty 2\frac{|E_n(\omega, \tau)|^2}{\tau} d\omega$$

where $\frac{|E_n|^2}{\tau}$ is called the *sample spectrum* or *periodogram.* It is permissible to move the $\lim_{\tau \to \infty}$ inside the integral provided the ensemble average is performed first [Buckingham 83, p. 32]. Thus,

$$\overline{e_n^2} = \frac{1}{2\pi} \int_0^\infty \left\{ \lim_{\tau \to \infty} \frac{2\overline{|E_n|^2}}{\tau} \right\} d\omega.$$

As we have a random process, the limit would be indeterminate had the ensemble average not been performed first.

Assumption 3: The process is ergodic, so temporal and ensemble averages are equivalent, ie. $\lim_{\tau \to \infty} \langle e_n^2 \rangle_\tau = \overline{e_n^2}$. Thus,

$$\langle e_n^2 \rangle = \frac{1}{2\pi} \int_0^\infty \left\{ \lim_{\tau \to \infty} \frac{2\overline{|E_n|^2}}{\tau} \right\} d\omega. \tag{4.7}$$

By definition, the one-sided power spectral density of e_n is, $S(\omega) = \lim_{\tau \to \infty} \frac{2\overline{|E_n|^2}}{\tau}$, therefore we can rewrite eqn. 4.7 as

$$\langle e_n^2 \rangle = \frac{1}{2\pi} \int_0^\infty S(\omega) d\omega. \tag{4.8}$$

Assumption 4: The noise spectrum is white, therefore $S(\omega) = S_o$, a constant.

For a practical measuring instrument bandwidth of $\Delta\omega$, eqn. 4.8 becomes

$$\langle e_n^2 \rangle = \frac{1}{2\pi} S_o \Delta\omega. \tag{4.9}$$

Using identical arguments for the capacitor voltage, v_n, we have

$$\langle v_n^2 \rangle = \frac{1}{2\pi} \int_0^\infty \left\{ \lim_{\tau \to \infty} \frac{2\overline{|V_n|^2}}{\tau} \right\} d\omega$$

and substituting in eqn. 4.6

$$\begin{aligned}
\langle v_n^2 \rangle &= \frac{1}{2\pi} \int_0^\infty \frac{1}{1 + (\omega RC)^2} \left\{ \lim_{\tau \to \infty} \frac{2\overline{|E_n|^2}}{\tau} \right\} d\omega \\
&= \frac{S_o}{2\pi} \int_0^\infty \frac{d\omega}{1 + (\omega RC)^2} \\
&= \frac{1}{2\pi} \frac{S_o}{RC} [\arctan(\omega RC)]_0^\infty = \frac{1}{2\pi} S_o \frac{\pi}{2RC}.
\end{aligned}$$

Putting this into eqn. 4.9, to eliminate S_o, gives

$$\langle e_n^2 \rangle = \frac{2}{\pi} C \langle v_n^2 \rangle R \Delta\omega. \tag{4.10}$$

Assumption 5: The system is in equilibrium with its surroundings.

According to equipartition theory, a general dynamical system in thermal equilibrium has on average a potential energy of $kT/2$ for each degree of freedom. One short hand method for counting up the degrees of freedom in a linear system is to count the number of independent quadratic variables in the energy expression [Reif 67]. By inspection of 4.10, we see that our system takes up energy as $\frac{1}{2}C\langle v_n^2 \rangle$ and therefore has one degree of freedom, hence

$$\frac{1}{2} C \langle v_n^2 \rangle = \frac{1}{2} kT. \tag{4.11}$$

Assumption 6: Let us assume the system is classical (ie. no quantum effects) so that the Maxwell-Boltzmann kT term holds.

Substituting 4.11 into 4.10 finally yields Johnson's formula for open circuit noise voltage

$$\begin{aligned} \langle e_n^2 \rangle &= \frac{2}{\pi} kTR\Delta\omega && (4.12) \\ &= 4kTR\Delta f. && (4.13) \end{aligned}$$

This unassuming equation is a source of a number of interesting conundrums and much consternation. One common question is "where does the coefficient of 4 *really* come from? That is, what is its physical significance?" This can now be quite easily traced from the above analysis, where the $\frac{2}{\pi}$ clearly comes the one-sided integral of the arctan function. Hence it is purely a 'geometrical' quantity. If the integral is modified by substituting the capacitor with a more complex network, the number of degrees of freedom of the system changes to always maintain the ubiquitous 4.

If the capacitor is replaced by an inductor L, the analysis can be repeated in the current domain and the generated short circuit noise current can be shown to be $\langle i_{sc}^2 \rangle = \frac{2}{\pi}\frac{L}{R}\langle i_n^2 \rangle \Delta\omega$, where $\langle i_n^2 \rangle$ is the observed noise current. The system now takes up energy as $\frac{1}{2}L\langle i_n^2 \rangle = \frac{1}{2}kT$, therefore

$$\langle i_{sc}^2 \rangle = \frac{4kT\Delta f}{R}. \tag{4.14}$$

which is the familiar current form for the Johnson noise formula.

Complete analysis and detailed discussion of the behaviour of 4.13 and 4.14 at limiting values of the main variables is lacking in the pedagogical literature, so we proceed for the first time to examine, in detail, the main problem areas in the following sections.

4.4 The Classical Energy Catastrophe

The most obvious problem with 4.13 is that it classically predicts infinite energy as $f \to \infty$. This is analogous to the black-body radiation problem where the Rayleigh-Jean's law suffers from the so-called ultraviolet catastrophe – the divergent curve having infinite area over all frequencies. Anticipating this, Nyquist [Nyquist 28] in 1928 suggested replacing kT with the one-dimensional form of Planck's law

$$\frac{hf}{e^{hf/kT}-1}$$

which reduces to kT as $f \to 0$ and rolls off for $hf > kT$. So far so good, however this quantum term predicts zero energy at $T = 0$ which is a violation of the Uncertainty Principle. As we shall see the solution to this creates a further conundrum.

4.5 The Quantum Energy Catastrophe

During 1911-12, Planck's 'second theory' produced the following modification to the quantum term [Kuhn 78]

$$\frac{hf}{e^{hf/kT}-1}+\frac{hf}{2}=hf\coth\left(\frac{hf}{2kT}\right).$$

The extra $hf/2$ term is called the *zero-point energy* (ZPE) and in this case, at $T = 0$, the Uncertainty Principle is not violated. This creates a further conundrum in that $hf/2$ is infinite when integrated over all frequencies, which is an apparent return to the type of 'catastrophe' problem we saw in the classical case. One can only assume that Nyquist accordingly did not suggest this form and probably would have been aware of Planck's own misgivings concerning the experimental objectivity of $hf/2$. The inclusion of $hf/2$ in standard noise texts only became popular after 1951 following the classic work of Callen & Welton [Callen 51] that produced the $hf/2$ ZPE term as a natural consequence of their generalised treatment of noise in irreversible systems using perturbation theory.

The solution to the catastrophe problem is that $hf/2$, in fact, turns out to be the *ground state* of a quantum mechanical oscillator. If n is the quantum number, which is a positive

integer, then the allowed energy states for a quantum oscillator are $(n + \frac{1}{2})hf$ and thus the ground state is given when $n = 0$. As there is no lower energy state than the ground state, there is no energy level transition available to release the ZPE. Therefore it can be argued that $hf/2$ should be dropped before integration of the quantum expression. This procedure is an example of *renormalisation*, which basically redefines the zero of energy. Renormalisation is a significant area of quantum field theory and is usually presented in a more formal manner. The problem of renormalisation is an open question in the theory of gravitation where there is the apparent catastrophe of *total* energy becoming infinite. For most laboratory measurements there is no catastrophe as we are only interested in energy *differences*.

The fact that the ground state energy, which we call ZPE, cannot be released means that texts that quote the Callen & Welton $hf/2$ term as an observable noise component are not strictly correct. However, by coincidence it turns out that the mean square of the zero point fluctuation (ZPF) also has the form $hf/2$ [Louisell 73]. The mean square does not vanish with renormalisation, of course, and this ensures the Uncertainty Principle survives renormalisation. The mean square fluctuation is a detectable quantity and represents the magnitude of the ZPF.

Each mode contributes $hf/2$ towards the mean square fluctuation and, for an infinite number of frequencies, the magnitude is infinite. It is considered that this infinity is not fundamental, since the measurement conditions have not been specified. It can be shown [Louisell 73] that for any finite observation bandwidth and volume of space the magnitude of the fluctuations of a quantum field is finite – if either the bandwidth is infinite or the measurement is evaluated at a *point* in space then the fluctuations become infinite.

4.6 The Steak Grilling Debate

In 1982, Grau & Kleen expressed the view that $hf/2$ is both unextractable and unobservable, adding their memorable rejoinder in the *Solid-State Electronics* journal that $hf/2$ is not "available for grilling steaks" [Grau 82]. Uncannily, about the same time Koch, Van Harlingen & Clarke (KVC) published noise measurements in superconductors reporting to have observed ZPF [Koch 81]. Over the next 3-4 years a number of independent superconductor papers followed, all nonchalantly quoting the KVC interpretation of ZPF as standard. In reply, Kleen (1987) essentially restated his case pointing out an unanswered question in the superconductor measurements [Kleen 87]. As far as we are aware there has been no published KVC reply.

The orthodox position, is that the effects of ZPF are observable such as in the Casimir effect [Itzykson 80]. ZPF also has an orthodox status in explaining the observations of Mullikan [Mullikan 24], Lamb [Welton 48] and the nature of liquid helium [Adkins 75]. On the other hand, consensus is not total as the school of Kleen has some support eg. [Beck 76, p. 173] and [Kiss 88], the commonly supposed link between spontaneous emission and ZPF has been criticised [Siegman 64] and the overall understanding of ZPF is also questioned as expressed, for example, in the following quote [Grandy 70]:

> "The obvious question, then, is whether the zero-point energy and the vacuum fluctuations are one and the same thing. If they are, why is it that the former can be eliminated from the theory? The answer is not yet clear, and a deeper significance has yet to be discovered. Therefore, we will adopt the view that the zero-point energies are to be formally removed from the theory...., and all physical effects of the type.... discussed are to be ascribed to quantum fluctuations of the vacuum.... It must be admitted that the vacuum is not completely understood, neither physically nor philosophically. Whether or not the vacuum fluctuations are intimately related to the (unobservable) zero-point energy remains an open question."

where the expression "vacuum fluctuations" is an alternative term for ZPF. The view that ZPF cannot give rise to a detectable noise power itself, but can indirectly modulate or induce a detectable noise power has been recently expounded [Senitzky 93].

As for grilling steaks, the debate still sizzles but has shifted away from electrical noise theory. Controversial attempts to harness ZPE are underway using the concept of system self-organisation [King 91] and presupposing the idea that the ground state is not the actual source of energy but is a 'pipeline' into some universal background source [Puthoff 89]. In an enterprising decade where there have been controversial attempts to consider superluminal velocity [Giakos 91] and quantum information theory (promising two bits of information from one physical bit [Bennett 92] and a form of teleportation [Bennett 93]), there is no doubt that ZPE research will thrive. It remains to be seen what concrete results are produced and, if any, what the implications are to noise theory. Until further evidence, the quantum zero-field should be regarded as a conservative field as far as the extraction of energy is concerned.

4.7 Quantum Cut-Off Experimental Status

Fig. 4.7 shows a theoretical plot of the quantum term for different temperatures. The $hf/2$ term is plotted to illustrate that at normal working frequencies and temperatures it

is vanishingly small, so for these conditions it can be neglected regardless of the status of debate. It can be seen from the Fig. 4.7 plot that experimental verification of the quantum cut-off point for electrical noise is rather difficult due to the Tera Hertz frequencies. If the temperature is reduced, to reduce the cut-off frequency, we see that the maximum energy of the curves falls, thus making noise detection more difficult. In 1981, van der Ziel [van der Ziel 81] proposed to make measurements in this region, at 100 GHz using Hanbury-Brown Twiss circuits; unfortunately, this research effort was never completed. The only curves we have today, for electrical noise, appear to be those of the type of KVC, which show no cut-off due to ZPF becoming significant. Therefore, as far as we are aware, there are no measurements that directly demonstrate the quantum cut-off for *electrical* thermal noise, to this day. Although the cut-off region, for electrical noise, has so far been obscured by ZPF it may become possible in the future to view at least part of this region, without violation of the Uncertainty Principle, if somehow the concept of *squeezed states* can be successfully employed for the electrical case (eg. [Baseia 93]).

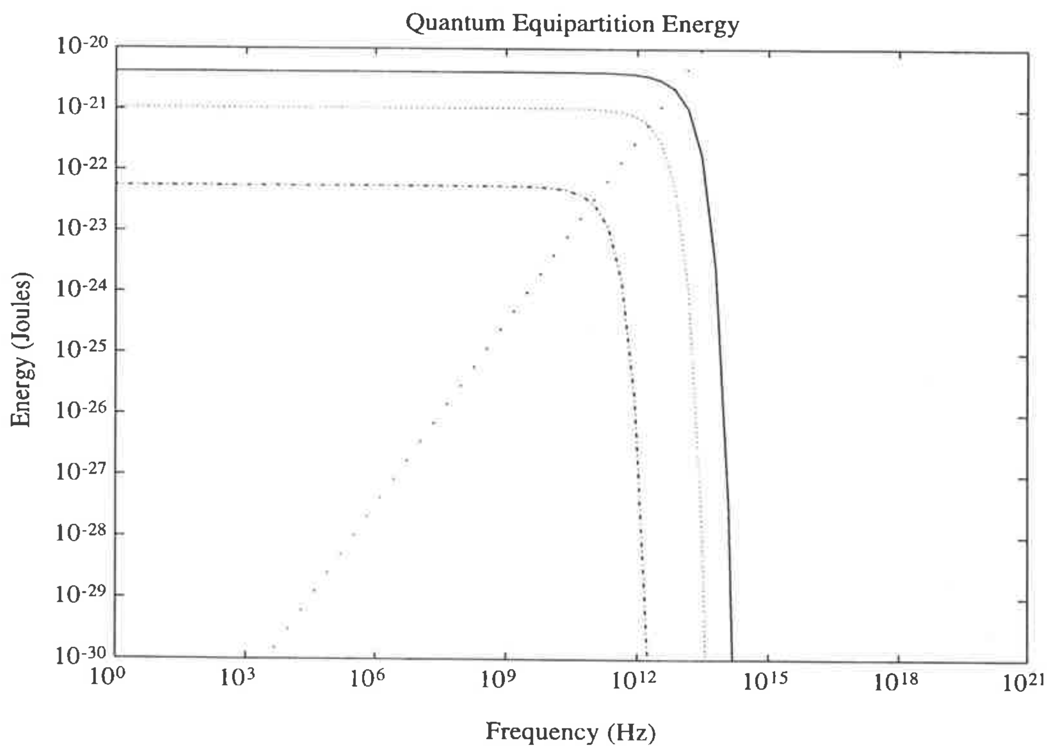


Figure 4.5: Quantum equipartition energy versus frequency for 300K, 77K and 4K. The line represents the hf/2 term plotted separately.

4.8 MacDonald's Objection

In 1962, MacDonald raised an interesting objection [MacDonald 62] concerning the quantum term. He correctly demonstrated that for $hf > kT$ the time-dependence characteristic of a *reversible* system is exhibited. For $hf \ll kT$ he showed that the system is *irreversible* as expected. Given that an electrical resistor is regarded as a dissipative irreversible system, a transition to a reversible regime for $hf > kT$ caused MacDonald to doubt the validity of the quantum term altogether.

The transition to the reversible regime can be simply thought of as taking place because at high frequencies, $hf > kT$, ie. for time intervals less than h/kT, the period is too short to achieve thermal equilibrium. As noise is a manifestation of a dissipative system maintaining thermal equilibrium, if the intervals are too short, then the dissipative process must roll off at these higher frequencies. This is precisely what is predicted by the quantum term.

Dissipation can be thought of as a process that eventually brings classical particles, in a closed system, to rest. This situation is not permissible for quantum particles as it would be a violation of the Uncertainty Principle. Dissipation does not play a role in microscopic description of quantum particles. It is a macroscopic concept whose relation to the quantum microscopic description is purely a statistical one in the classical limit. The 'sleight of hand' that turns non-dissipative equations of motion, into dissipative ones in the classical limit is hidden within the equation boundary conditions. For a more mathematical discussion see [Ullersma 66]. From the modern viewpoint, quantum Brownian motion (QBM) is now a major discipline area [Grabert 88] and MacDonald's objection is therefore clearly out-moded.

4.9 The Case of Limiting R & C

Remembering that the Johnson expression for $\langle e_n^2 \rangle$ is the case for an *open circuit* resistor, we now systematically illustrate, for the first time, how to examine the output voltage, current and charge fluctuations ($\langle v_n^2 \rangle$, $\langle i_n^2 \rangle$ and $\langle q_n^2 \rangle$) for the various limiting cases of R & C – the results are summarised in Table 4.2.

The case of finite R & C and $f \to \infty$

The voltage noise detected across the capacitor is simply

$$\langle v_n^2 \rangle = \int_0^\infty \frac{4kTRdf}{1 + (2\pi fRC)^2} = \frac{kT}{C} \tag{4.15}$$

which normally causes some surprise as the kT/C term is independent of R which is the source of the noise! This is easily explained by observing that as R increases, the corresponding increase in noise is exactly cancelled by a decrease in circuit bandwidth $\propto 1/R$. The circuit current noise is given by

$$\langle i_n^2 \rangle = \int_0^\infty \frac{(4kT/R)(2\pi fRC)^2 df}{1 + (2\pi fRC)^2} = \infty$$

and this divergent result is a consequence of classical theory breakdown.

The case of finite f, R & C

For a finite frequency band Δf, taken from zero, the voltage noise becomes

$$\begin{aligned} \langle v_n^2 \rangle &= \frac{2kT}{\pi C} \arctan(2\pi \Delta f RC) \\ &\approx \frac{kT}{C} - \frac{4kTR\Delta f}{(B_m/B_c)^2} \end{aligned}$$

by expanding arctan for $B_m \gg B_c$, where the measurement bandwidth $B_m = (\pi/2)\Delta f$ and the circuit bandwidth $B_c = 1/(4RC)$. This remarkable result is the difference between the familiar kT/C noise and the open circuit noise divided by the square of the ratio of the bandwidths. The analogous expression for the current noise can be shown to be

$$\langle i_n^2 \rangle \approx \frac{4kT\Delta f}{R} \left(1 + \left(\frac{B_c}{B_m}\right)^2\right) - \frac{kT}{CR^2}$$

which goes to infinity for $f \to \infty$, as expected for the classical theory. The solutions using the quantum term for finite R and C involve tedious integrals and are not that instructive; for such a treatment refer to ref. [Pyati 93].

Open circuit capacitor: finite C, $R \to \infty$, $f \to \infty$

Using the well-known delta function approximation of the form

$$\delta(\alpha x) = \frac{1}{\pi} \lim_{x \to \infty} \left\{ \frac{x}{1 + \alpha^2 x^2} \right\}$$

we observe that

$$\lim_{R \to \infty} \left\{ \frac{R}{1 + (2\pi fRC)^2} \right\} = \pi \delta(2\pi fC) = \frac{\delta(f)}{2C}$$

by applying the delta function identity $\delta(\alpha x) = \delta(x)/\alpha$. Therefore

$$\langle v_n^2 \rangle\Big|_{R=\infty} = \frac{1}{2}\int_{-\infty}^{\infty} 4kT\frac{\delta(f)}{2C}df = \frac{kT}{C} \tag{4.16}$$

where the integral limits are taken between $\pm\infty$ as $\delta(f)$ is centered about the $f = 0$ axis. The factor of 1/2 is introduced as we are dealing with frequencies in the positive domain – this is justified as the function is symmetrical about the $f = 0$ axis. This surprising result of kT/C has to be interpreted in terms of a dc voltage across the capacitor, because as $R \to \infty$ the circuit bandwidth$\to 0$. Therefore, the classical formulation predicts that an ensemble of capacitors will, on average, display a dc voltage of $\sqrt{kT/C}$ across their terminals. The source of the dc voltage can be thought of as the voltage that is sampled by the capacitor at the moment the finite resistor is removed. Note that substitution of the quantum term into the above integral also produces the same result.

This explanation in terms of a dc voltage is much more satisfactory than that of ref. [Pyati 93] which considers it as a time varying noise voltage and consequently proposes highly ingenious ways of making a pure capacitor dissipative! This is clearly unnecessary as any dissipative proposal for the capacitor can be modeled by an equivalent resistor.

Short circuit capacitor: finite C, $R \to 0$, $f \to \infty$

As $R \to 0$, all the integrals, classical and quantum, trivially go to zero – this is expected as R is the source of the noise.

Open circuit resistor: finite R, $C \to 0$, $f \to \infty$

For the classical case $\langle v_n^2 \rangle \to \infty$ as expected due to breakdown of the theory. The integrals for $\langle i_n^2 \rangle$ and $\langle q_n^2 \rangle$ trivially tend to zero. This is expected as there is no circuit loop for current to flow. The quantum case for the open circuit resistor, where Γ is the mathematical gamma function and ζ is the Weierstrass zeta function, now becomes [Rhyshik 63]

$$\begin{aligned}\langle v_n^2 \rangle = \int_0^{\infty} \frac{4Rhf\,df}{e^{hf/kT}-1} &= 4Rh\left(\frac{kT}{h}\right)^2 \Gamma(2)\zeta(2) \\ &= \frac{2R}{3h}(\pi kT)^2\end{aligned}$$

which implies that the open circuit thermal noise voltage across a 1 MΩ resistor, at room temperature, viewed with an instrument of infinite input impedance and infinite bandwidth (or at least $\gg kT/h$ =6250 GHz) with no parasitic capacitance present is 0.41 V_{rms}!

This is rather large, but fortunately these ideal conditions are unrealisable in practice. It also assumes that the result is directly observable and not swamped by $hf/2$ quantum noise.

Short circuit resistor: finite R, $C \to \infty$, $f \to \infty$

For the classical case, the voltage integral goes to zero whereas current and charge go to infinity. Zero voltage is expected as the terminals are shorted and infinite current in the loop is a breakdown of the classical theory. For the quantum case for current we have

$$\langle i_n^2 \rangle = \int_0^\infty \frac{(4hf/R)df}{e^{hf/kT} - 1} = \frac{2}{3hR}(\pi kT)^2$$

which is finite as expected. However, the quantum integral for charge turns out to be divergent, giving infinite charge. Notice we now have a random walk type non-stationary process. The infinite result may be seen, not as a breakdown in the classical or quantum theory, but due to the artificial construct of theoretical infinite capacitance. An infinite capacitor can be thought of as an infinite store of charge – this never occurs in practice which is another way of saying that there is no such thing as a perfect short circuit. Notice that in the limit as $C \to \infty$, the capacitor becomes simultaneously an infinite store of charge and a perfect short. This can be resolved by thinking of $C \to \infty$ as being modeled by an ideal voltage source.

		Classical			**Quantum**		
		$\langle v_n^2 \rangle$	$\langle i_n^2 \rangle$	$\langle q_n^2 \rangle$	$\langle v_n^2 \rangle$	$\langle i_n^2 \rangle$	$\langle q_n^2 \rangle$
$R \to 0$	Shorted Cap.	0	0	0	0	0	0
$R \to \infty$	Open Cap.	$\frac{kT}{C}$ (dc)	0	kTC (dc)	$\frac{kT}{C}$ (dc)	0	kTC (d
$C \to \infty$	Shorted Res.	0	∞	∞	0	$\frac{2}{3hR}(\pi kT)^2$	∞
$C \to 0$	Open Res.	∞	0	0	$\frac{2R}{3h}(\pi kT)^2$	0	0

Table 4.2: Thermal noise over infinite bandwidth for different cases of limiting R and C.

4.10 kTC Noise

A subtle source of confusion is that equations 4.15 and 4.16 appear identical – however they are fundamentally different. Equation 4.15 is the ac thermal noise of a resistor over a

measurement bandwidth greater than the circuit bandwidth, when loaded by a capacitor C. Whereas 4.16 is the average value of the dc voltage *sampled* by C, when the resistor supplying the thermal noise is removed. These are two fundamentally different quantities that coincidentally have the same numerical value.

Multiplication of eqn. 4.16 by C^2 gives the well known result $\langle q_n^2 \rangle = kTC$ used for analysing noise on switched capacitor circuits. For example, a 1 pF capacitor will sample 64 μV_{rms} across its terminals or in terms of electron number this is 400 rms electrons.

In practice the noise on a switched capacitor, when the transistor is 'on,' is given by 4.15 and the circuit bandwidth is large – but doesn't exceed the quantum cut-off for realistic RC values. When the transistor is 'off,' the average voltage sampled by the capacitor is given by 4.16 and the circuit bandwidth is very small. As the off-resistance of a transistor is finite the circuit bandwidth will be small but non-zero. Hence, the voltage on the capacitor will not be perfectly dc, but will be slowly varying. However, in most real applications the switching time of the capacitor is much shorter than the RC time constant, hence the sampled voltage appears as a steady dc level over the switching period.

4.11 Power in a Matched Load

If a resistor R develops an open circuit noise voltage of $4kTR\Delta f$, the power delivered to an equal load resistor is

$$P = \langle i_n^2 \rangle R = \frac{\langle v_n^2 \rangle}{(2R)^2} R = \frac{\langle v_n^2 \rangle}{4R} = kT\Delta f.$$

This causes some surprise as P appears to be independent of R, which is the source! To understand this, let us consider an arbitrary load R_L, so the power now becomes

$$P = 4kTR\Delta f \frac{R_L}{(R+R_L)^2}.$$

So we see that for small R, the noise term is small and therefore the delivered power P is small; whereas for large R the potential divider term becomes small, so the delivered power is still small. Maximum power transfer occurs when $\frac{dP}{dR} = 0$, which trivially yields $R = R_L$ – hence there is balance achieved between noise generation and the potential divider effect.

This analysis can, of course, be reproduced by considering a resistor R in parallel with a current noise source of $4kT\Delta f/R$. A common student error is to mechanically proceed the analysis with the power, $P = \langle i_n^2 \rangle R_L$, as before. This, of course, leads to the wrong result. For the case of a current source we must put $P = \langle v_n^2 \rangle / R_L$, giving

$$
\begin{aligned}
P &= \frac{\langle v_n^2 \rangle}{R_L} = \frac{\langle i_n^2 \rangle R^2 R_L}{(R+R_L)^2} \\
&= \frac{4kT\Delta f}{R} \frac{R^2 R_L}{(R+R_L)^2} \\
&= 4kTR\Delta f \frac{R_L}{(R+R_L)^2}
\end{aligned}
$$

which is the same result as before.

Another curious feature of the $P = kT\Delta f$ formula is that it appears to imply that for large observation times the power transfer tends to zero, whereas for small times the power transfer increases. This is simply explained by noting that the application of the thermal noise formula presumes that both resistors are in thermal equilibrium. Hence for long observation times we expect the net power transfer to be zero; otherwise a resistor would heat up and escape thermal equilibrium. However for a small 'snapshot' of time, as the fluctuations in each resistor are uncorrelated, there must be an instantaneous transfer of power. The momentary transfers of power back and forth between the resistors, on average, add to zero. This also explains why energy cannot be harnessed from the thermal noise in a resistor, cf. Brillouin's Rectifier Paradox [Brillouin 50], Penfield's Motor Paradox [Penfield 66], Feynman's Rachet Paradox [Feynman 63], Panse's Radiation Paradox [Browne 93] and Bogner's Microwave Isolator Paradox [Bogner 94]. Analogous arguments are used by some authors [Buckingham 83] to assert that energy cannot be extracted from ZPF, however this has been disputed [Puthoff 90]. The apparent infinite power as the snapshot of time approaches zero is, of course, due to breakdown of the classical kT term.

4.12 Distributed RC

Until this point, our analysis has only considered a lumped circuit model. In a given practical case, a resistor may have some distributed parasitic capacitance and thus it is instructive to analyze the noise in a distributed RC line. From standard transmission line theory, the impedance looking into an RC line with the other end shorted is

$$Z = \sqrt{\frac{R}{j\omega C}} \tanh \sqrt{j\omega RC}.$$

The voltage noise seen across the open circuit terminals is found by simply inserting $\Re(Z)$ into the Johnson noise formula [Davenport 58, Starr 53]. Therefore

$$\langle v_n^2 \rangle = 4kT\Re(Z)\Delta f$$

$$= \quad 4kT\Delta f \frac{R}{\sqrt{4\pi f RC}} \frac{\sinh \sqrt{4\pi f RC} + \sin \sqrt{4\pi f RC}}{\cosh \sqrt{4\pi f RC} + \cos \sqrt{4\pi f RC}}$$

which reduces to $4kTR\Delta f$ for small f, as expected. This is plotted in Fig. 4.6 with the simple RC case for comparison, showing that at low frequencies they are equal, but at high frequencies the simple RC curve rolls off faster. The quantum curve is included to show that there is a physical limit to the slow roll-off in the distributed case. Finally, the case for a transmission line with a *matched* load is also plotted to clearly show that this option does *not* model a discrete resistor with distributed parasitic capacitance. The curious phenomenon of the noise increasing, for low frequencies, in the matched load case, is a manifestation of noise in the load (which is physically unrealisable in this case) and not an anomaly in the line.

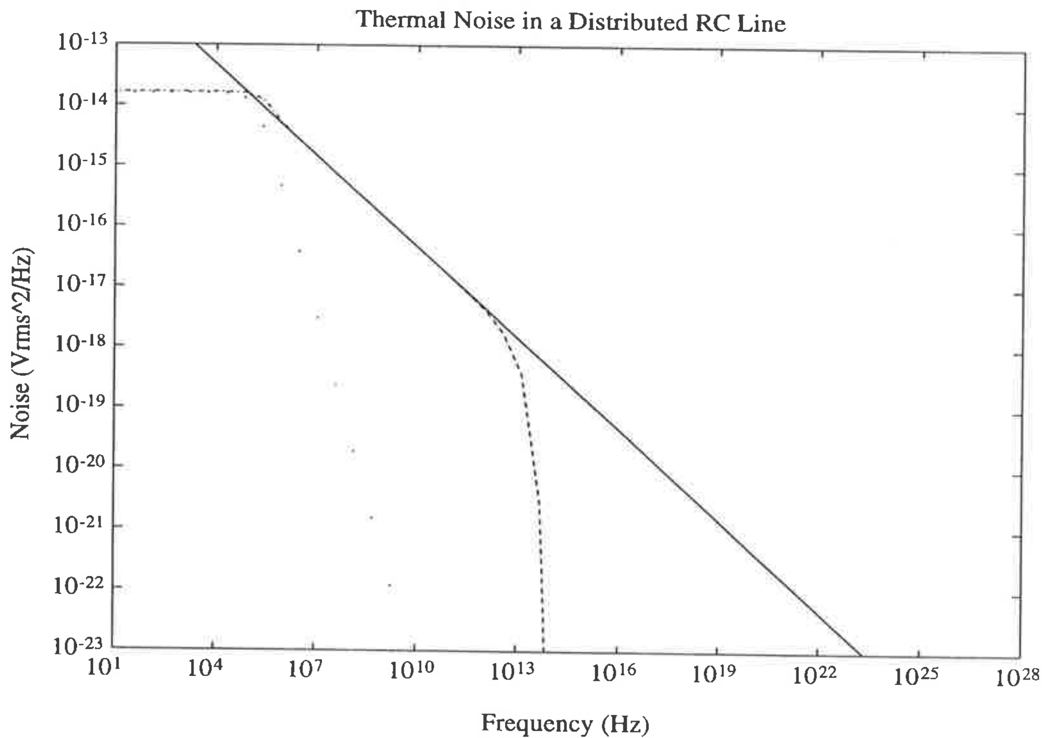


Figure 4.6: Noise in distributed RC line. Chained line: short circuit load. Solid line: matched load. Dashed line: short circuit load with Planck term. R=1 MΩ, C=1 pF. All curves plateau at 1.66×10^{-14} V^2_{rms}/Hz, except for the matched load case that increases due to noise in the load. For reference, the simple lumped RC case is shown as a dotted line.

4.13 The Capacitor Paradox

For the first time, we present an unsolved paradox in a simple two stage RC ladder (Fig. 4.7). In Appendix D, we solve the relevant complex integral for the circuit showing that correlation between the two capacitor noise voltages is zero, ie. $\overline{v_1 v_2} = 0$. However, we also show that if $R_2 \rightarrow 0$ or $R_1 \rightarrow \infty$ then $\overline{v_1 v_2}$ suddenly becomes non zero! This strange dilemma is unsolved and may highlight a limitation in the circuit noise theory formalism. If the capacitors are replaced by inductors, it may be that the problem has similar parallels to Penfield's motor paradox [Penfield 66].

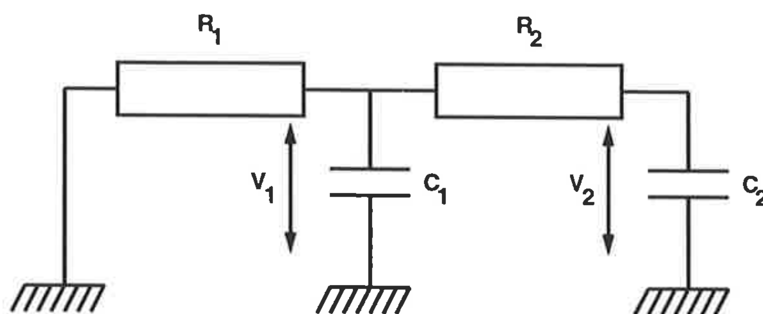


Figure 4.7: Two stage RC ladder – a source of an unsolved paradox.

4.14 Forward's Photoamplifier

A standard passive resistor in a photodiode circuit introduces thermal noise a can hence limit the signal to noise ratio. A room-temperature active resistor has been devised [Forward 82] that can maintain the same resistance, giving rise to the same photovoltage. However, the active resistor generates thermal noise as if it were at cryogenic temperature. This remarkable result may have applications in future areas of this work on GaAs photodetectors. The ability to maintain an equivalent resistance and yet reduce the thermal noise below the theoretical limit appears to be in violation of the fluctuation-dissipation theorem and the laws of thermodynamics. This apparent paradox can be explained when we consider the fact that the active resistor is powered and therefore there is an *external energy source* feeding the system – this preserves the thermodynamical laws.

4.15 Measurement of Noise

Measurement of noise is important both for the determination of noise parameters for the imager design and final performance testing of fabricated devices. A simple set up was devised to investigate some of the potential challenges. The amplifier configuration in Fig. 4.8 was connected to an HP3665A. In order to characterise the system, we set out to measure the thermal noise in R_p for a range of values. The output noise of the amplifier is given by,

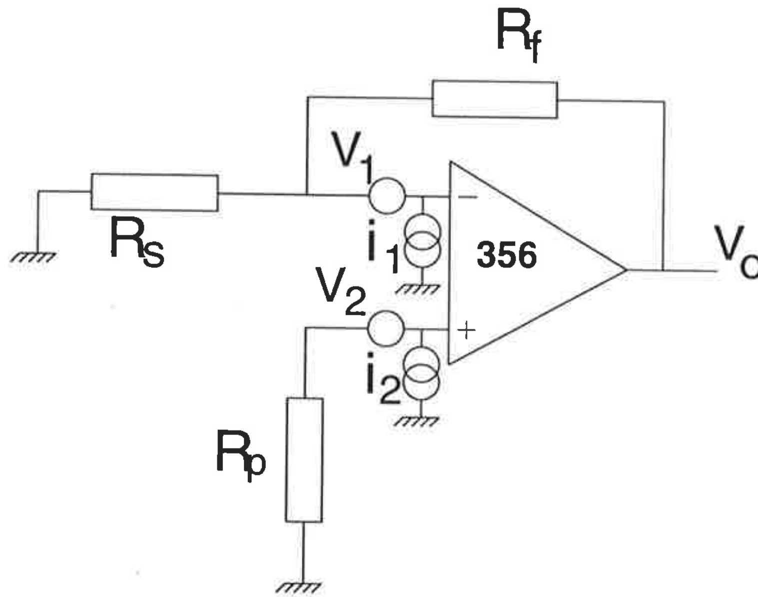


Figure 4.8: Amplifier configuration for noise measurement. R_p is the device under test (variable). R_s = 1K and R_f = 6.8M. Opamp voltage noise sources are uncorrelated but equal in magnitude v_1=v_2=10 nV/root-Hz. Opamp current noise sources are uncorrelated but equal in magnitude i_1=i_2=0.01 pA/root-Hz.

$$\overline{v_o^2} = \overline{v_f^2}+\overline{v_s^2}(\frac{R_f}{R_s})^2+\overline{v_2^2}(\frac{R_f+R_s}{R_s})^2+\overline{i_2^2}R_f^2+\overline{v_p^2}(\frac{R_f+R_s}{R_s})^2+\overline{v_1^2}(\frac{R_f+R_s}{R_s})^2+\overline{i_1^2}R_p^2(\frac{R_f+R_s}{R_s})^2$$

where $\overline{v_1^2}$ is equal to $\overline{v_2^2}$ in magnitude but uncorrelated and $\overline{i_1^2}$ is equal to $\overline{i_2^2}$ but uncorrelated. The terms in this expression are plotted in Fig. 4.9. Notice that only the data points in the central straight portion of the output noise curve are a valid measure of noise in R_p. This is because in this region the noise in R_p dominates all the other sources.

Back calculating from the valid data points we find that the coefficient '4' from the thermal noise formula comes to 3.7±0.3. This result is curious as it appears to predict noise levels

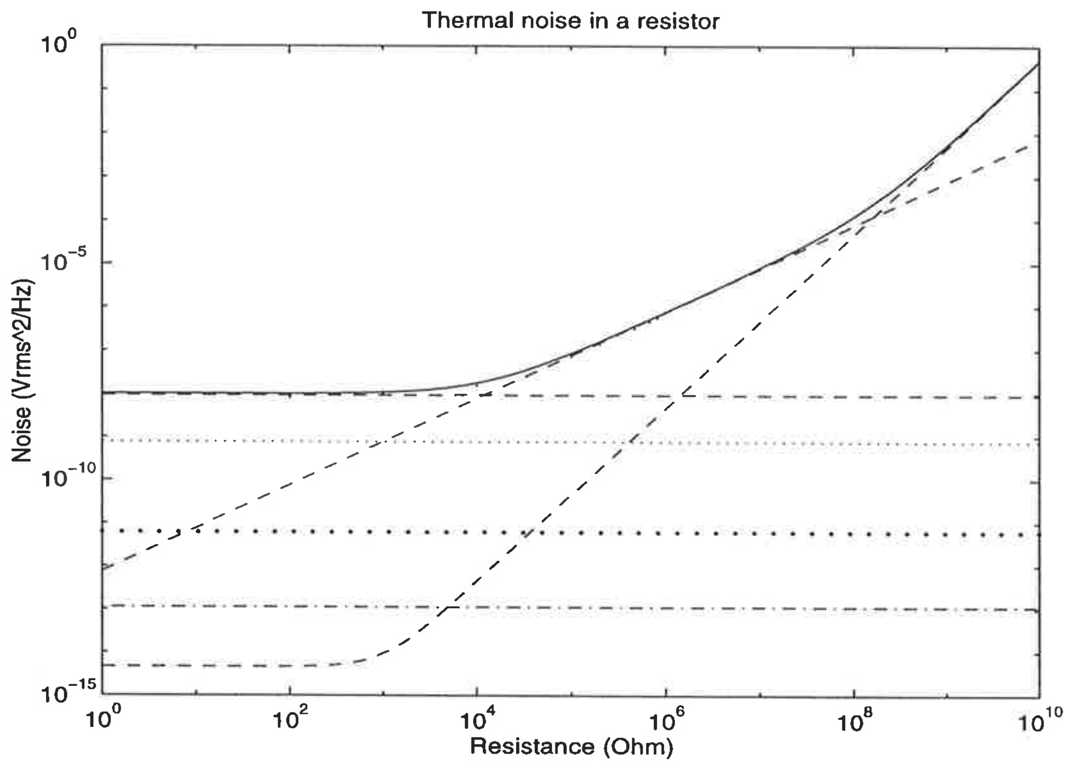


Figure 4.9: Operational amplifier noise sources. Solid line: total noise on output. Dashed lines: opamp voltage noise (upper trace), resistor under test noise (middle trace), opamp current noise (lower trace). Fine dotted line: input resistor noise. Coarse dotted line: spectrum analyser noise floor. Chained line: feedback resistor noise. Vertical dashes: actual measured data points – height of dashes represent error bars.

less than theory. It is tempting to dismiss this by pointing out that the result does include the value of 4 if the error is taken into consideration. However, only a small component of this error is experimental error, as such – it mostly represents the intrinsic variance of the noise taken over many rms averages. Hence the fact that the mean (3.7) is lower than 4, is something that must be critically examined in case we have discovered something fundamental. Let us now critically evaluate all the variable factors.

Repeatability: we found that the experiment could be repeated many times, on different days, with different circuit boards, different resistors and yet the mean thermal noise coefficient was always lower than 4 and about 3.7. It is interesting to note that in the seminal work on measuring thermal noise [Johnson 28] a mean of 3.7 was also obtained! However in 1928 this low value was ignored as it lay within experimental error.

Window Function: we operated the FFT spectrum analyser with a Hanning window which is theoretically the optimum for noise measurements and should introduce an error no more than 1%.

Component Values: the resistor values in the circuit and those under test were all measured rather than relying on colour coding. The maximum error for a Fluke meter reading is 1%. This was checked against several meters.

Frequency Response: the gain of the operational amplifier was not calculated from the resistor values, but instead was obtained from the frequency response of the operational amplifier which was measured with a swept sinewave. The response was flat over the measured noise band (100-300Hz). Maximum error introduced is 2%.

Temperature: the maximum error in temperature reading for insertion into the thermal noise formula is 1%.

Overdrive: if the amplified peak-to-peak noise is so large that it overdrives the operational amplifier into a non-linear region, the gain can drop thereby producing apparently lower noise readings. This was not the case as the noise was viewed simultaneously on an oscilloscope and a spectrum analyser – no sign of clipping was observed on the time trace and no unusual frequencies were observed in the frequency domain.

Batteries: the amplifier was powered by batteries to reduce extraneous noise. The battery voltage level was checked before each experiment.

Averages: the HP3665A was automatically logged by a PC running under LABVIEW. This enabled us to efficiently take 400 ensemble rms averages of the traces and 400 temporal rms averages of each trace (skipping multiples of 50 Hz). Thus each data point was obtained from 160,000 rms averages!

Automation: preliminary hand measurements were taken to check there were no artifacts introduced by the automated logging procedure.

Parasitic Resistance: the resistors under test were individually soldered into position and the resulting total resistance was rechecked. There was no significant extra resistance introduced from these connections.

Open Circuit Formula: the formula $4kTR\Delta f$ strictly gives the *open circuit* noise voltage across the resistor under test. The input capacitance of the amplifier is a few pF and therefore a possible objection is that we should use the closed circuit form, with load capacitance, $\frac{2kT}{\pi C}\arctan(2\pi\Delta f RC)$, and that this could be the source of error. However, for small arguments $\arctan(x) = x$ and the formula reduces to the open circuit form. In order to obtain an error of about 7.5% it would be necessary for $R > 400$ MΩ – this is clearly not the case for our data points. The R values we have used produce an error as little as 1×10^{-5}%.

Test Jig: the test jig was in a screened box with a star earth to the box. Screened BNC cable was used to feed the output to the spectrum analyser. Screening ensures that no pickup overdrives the amplifier, perhaps paradoxically producing lower noise readings.

Analyser: the calibration of the spectrum analyser was carefully checked by feeding in a known long pseudo-random bit sequence. The maximum error found was 2%.

Given the above obvious effects have been discounted we are now left with three possible hypotheses.

HYPOTHESIS 1: *Correlation of errors.* If all the possible error sources 'colluded' together to produce a lower noise value, their magnitude may just be enough to produce a low noise coefficient of 3.7. However, it is highly unlikely that so many uncorrelated events can synergise given that the experiment was done many times on different days, with different temperatures, different resistors, different amplifiers and cross checked with two different spectrum analysers! We reject this hypothesis.

HYPOTHESIS 2: *Boltzmann's constant is reduced.* At first sight this bold hypothesis doesn't seem too absurd. We must question why Boltzmann's constant (k) is not perhaps a little lower for an electron gas. After all, the particles are negatively charged and are repulsive, hence we might expect a different gas behaviour. However, we must reject this hypothesis – from the equipartition theorem we see that k is simply a constant of proportionality between the energy of a particle system and its temperature. Hence, the actual value of k arises from the way we conventionally define temperature and can thus have nothing to do with whether the gas particles are charged or not.

HYPOTHESIS 3: *Correlations between electrons.* As the electrons form a gas of negatively charged particles, and interact with negative repulsion, it is conceivable that this interaction results in a small amount of correlation which results in a slightly lower noise value than theory. As we have no evidence to dispute this idea and that it is not entirely absurd given that *space charge smoothing* is a well known effect that reduces *shot noise* below theory, we shall adopt this as our working hypothesis for future work.

4.16 Thermal Noise in MOSFETs

The thermal noise in the channel of a MOSFET can be referred back to the gate. This gate referred voltage noise is given by,

$$\overline{v_n^2} = \frac{8}{3}\frac{kTB}{g_m} \tag{4.17}$$

there the factor of $\frac{8}{3}$ only has a weak bias dependence and the formula is accurate for the saturation region. The thermal noise in the channel has a secondary effect in that it produces random fluctuations in voltage along the channel which, in turn, modulates the depletion layer encroaching on the width of the channel. Via capacitive coupling, this fluctuation is modelled by a gate noise current,

$$\overline{i_n^2} = \frac{8}{3}\frac{kTB\omega^2 C_g^2}{g_m} \tag{4.18}$$

where C_g is the gate capacitance. Although $\overline{v_n^2}$ and $\overline{i_n^2}$ fundamentally arise from the same thermal noise source, they are uncorrelated! The solution to this conundrum is due to the time constant associated with the capacitive coupling causing a phase shift between the two random waveforms. Detailed analysis shows that the square of the modulus of the correlation is only 0.16 and hence for most design applications we take the correlation as zero.

4.17 Flicker Noise in MOSFETs

$1/f$ noise results from a process (or set of processes) characterised by a continuous distribution of time constants over a very wide range. Two competing models that describe this process are well known:

(1) McWhorter's carrier density or number fluctuation model

(2) Hooge's mobility fluctuation model.

These models are invoked, either separately or in a unified form, to explain the variety of $1/f$ noise data in the literature. The number fluctuations model assumes that carriers in the conduction or valence bands communicate with the fast surface states, at the Si-SiO_2 interface, through a Shockley-Read-Hall (SRH) process. The carriers then tunnel into or out of traps, located further in the oxide, whose energies lie in the semiconductor bandgap. Noise is caused by carrier density fluctuations in the channel due to this random trapping of the free carriers. The second model assumes that mobility fluctuations are bought about by phonon scattering and interactions with interface charges. Various electron-interface charge interactions are reported:

(a) Coulomb scattering

(b) Scattering inhomogeneous surface potential due to non-uniform fixed interface charges

(c) Hopping between localised states in the interface.

It is clear that in both models surface states are instrumental in the $1/f$ noise mechanism. This means that the $1/f$ noise performance of a MOSFET is dependent on factors such as crystal orientation, hydrogen annealing, implantation and processing temperatures that affect the Si-SiO_2 interface.

For any given transistor the noise spectrum can be approximated by the following well-known expression,

$$\overline{v_f^2} = \frac{\alpha B}{WLC_o}(\frac{1}{f})^\gamma \tag{4.19}$$

where $\overline{v_f^2}$ is the gate referred flicker noise, W & L are the effective channel width and length, C_o is the gate oxide capacitance per unit area, α is the constant of proportionality and γ defines the power law. Since WLC_o, the active gate capacitance, is merely the factor that couples charge fluctuations in the channel with corresponding fluctuations in the gate voltage, this means that α and γ are the constants that essentially characterise $1/f$ noise in any device. Note that α increases with the number of fast surface states at the Si-SiO_2 interface and the exponent γ is a measure of the uniformity of the time constant distribution. Typically γ is in the region of 1.

4.18 Thermal Noise in MESFETs

The formulae for thermal noise in MESFETs are given by [Danneville 94]

$$\overline{v_n^2} = \frac{2kTB}{g_m} \tag{4.20}$$

and

$$\overline{i_n^2} = \frac{4kTB\omega^2 C_g^2}{g_m} \tag{4.21}$$

which appear similar to those for MOSFETs, except that the bias dependent coefficients are slightly different. Again the correlation between the two is small and therefore for worst-case analysis we shall treat them as totally uncorrelated.

4.19 Flicker Noise in MESFETs

Due to the absence of an oxide interface in MESFETs, flicker noise is considerably lower. Flicker noise in this case is attributed to carrier density fluctuation in the channel due generation and recombination or trapping. The formula is given by [Graffeuil 82]

$$\overline{v_f^2} = \frac{\alpha_c V_p^2 B}{2aN_d WL}(\frac{1}{f}) \tag{4.22}$$

where V_p is the pinchoff voltage, a is the channel depth, N_d is the channel doping and α_c is the flicker noise constant.

4.20 Shot Noise in MESFETs

Although MESFETs have the advantage of low flicker noise, the presence of gate leakage gives rise to shot noise not present in MOSFETs.

The origin of shot noise is generally attributed to the quantised nature of electric charge. The name 'shot' comes from the analogy of irregular impact of lead shot on a surface. It is not true that shot noise is present in all electrical currents. The flow of current through a resistor does not contain shot noise. Similarly freely falling lead shot has little opportunity to interact.

The analogy with lead shot is not totally adequate. In the movement of charged particles, large scale electrostatic interactions between individual current carriers smooth out the flow, given enough time for thermal equilibrium.

Shot noise is generated when current carriers cross a barrier, and in this sense is a non-equilibrium form of noise. Shot noise is encountered in solid-state devices, whenever a net current flows across a potential barrier such as the depletion layer of a junction. For the MESFET the formula is given by

$$\overline{i_s^2} = 2eI_gB \tag{4.23}$$

where e is the charge on the electron and I_g is the gate leakage current.

4.21 Noise Analysis of Imager Output Circuit

The preceding sections have created a framework for now discussing the noise in the output circuit of the XY array imager. Noise is usually poorly presented in the literature and thus our careful discussion of kTC noise and noise in transistors, in particular, will be used in the following analysis.

The output circuit of an XY array imager consists simply of the gate of a source follower connected to the output bus (called the video line), shown in Fig. 4.10. As signal charge is dumped on the video line, the resulting voltage variation is detected by the source follower. A second transistor is used to then reset the video line to a base dc voltage, thereby discharging the line in preparation for the next signal charge dump.

In order to achieve the best dynamic range, it is necessary to optimise the size of the source follower. As it turns out, this is a non-trivial problem due to a number of conflicting design trade-offs. The traditional rule of thumb, for silicon XY arrays, has been to set the width of the source follower such that the gate capacitance C_g equals the video line capacitance C_v. We shall show that this can give a workable result but is in fact not optimum.

Noise analysis can take place in the voltage, current or charge domain. Each domain is equally valid, and is selected on the basis of whichever results in the simplest analysis. In the case of an imager output circuit it is usually more convenient to operate in the charge domain. It is then conventional to divide the noise charge by the charge on the electron e, to produce a dimensionless result called the *rms electron number* or simply *rms electrons* or *noise electron density* (NED). In our case we shall plot all graphs in terms of *squared rms electrons* so that uncorrelated sources can be directly added. Another convenience we shall adopt is to use expressions for *gate referred noise* for the source follower, so that all sources of noise are compared 'as seen' from the video line.

The following conditions, in some cases contradictory, are required for optimum circuit

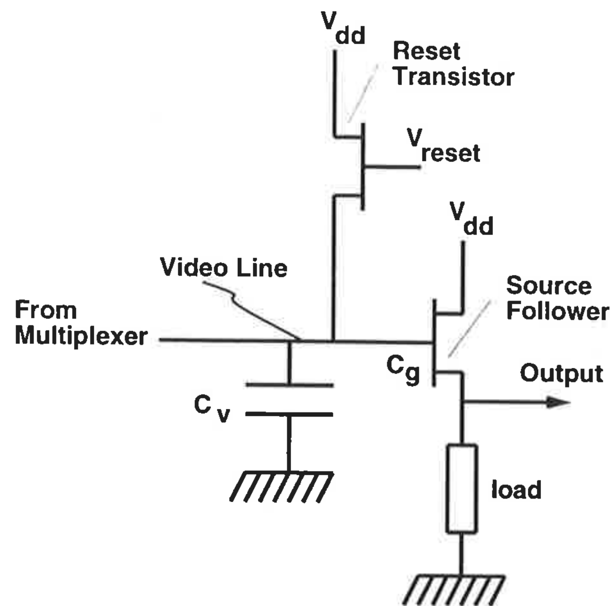


Figure 4.10: Imager output circuit configuration.

performance,

Reset transistor: (a) W/L to meet speed requirement (b) small WL to minimise channel charge and hence feedthrough spikes (sometimes called *partition noise*)

Source follower: (a) W/L to meet speed requirement (b) large WL for low gate referred $1/f$ noise (c) large W/L for low thermal noise voltage

Output node capacitance (C_v+C_g): (a) large node capacitance for low capacitor reset noise (kT/C) voltage (b) large node capacitance for low reset pulse feedthrough voltage (c) small node capacitance for large signal voltage swing.

These trade-offs turn a mere 3 component circuit into a surprisingly difficult optimisation problem that is too complex to solve analytically without some judicious simplifications.

Assumptions & simplifications:

(a) ignore the source follower gain reduction effect on input capacitance as this will affect signal and capacitor noise equally

(b) leave the reset transistor out of the optimisation – preset WL to the smallest area that still meets speed requirement

(c) use the maximum (worst case) value of source follower gate capacitance; ie. ignore any depletion layer capacitance in series with the gate

(d) neglect bias effects on flicker and thermal noise

(e) the thermal noise formulae, that incorporates g_m, assumes saturation

(f) $\sqrt{kT/C}$ is strictly the rms value of the noise sampled by the capacitor – this sample is a dc voltage – but as the instantaneous value is different for each time sample, we treat it as an ac noise source and add it in quadrature in the usual way

(g) we neglect any correlation between gate referred voltage thermal noise and gate current thermal noise, thus producing a worst case result

(h) we ignore any electrostatic effects between charge carriers producing any correlation (worst case)

(i) we assume the noise contribution of an off-chip preamplifier referred back to the source follower input is negligible.

4.21.1 Silicon output circuit

The charge-normalised noise formulae in terms of rms electron number are,

Thermal noise,

$$\overline{n_t^2} = \frac{8}{3}\frac{kTB}{gm}\frac{C^2}{e^2} + \frac{8}{3}\frac{kTB}{gm}\frac{(WLC_o)^2}{e^2};$$

Capacitor reset noise,

$$\overline{n_c^2} = \frac{kTC}{e^2}$$

Flicker noise,

$$\overline{n_f^2} = \frac{\alpha B}{WLC_o}\frac{1}{f}\frac{C^2}{e^2}$$

where $C = C_v + WLC_o$ and $g_m = 2\frac{W}{L}\mu C_o(V_{gs} - V_t)$. Inserting typical values: $(V_{gs} - V_t) = 3$ V, $L = 3\mu$m, oxide capacitance $C_o = 860\mu$F/m^2, bandwidth $B = 2$ MHz and mobility $\mu = 650 \times 10^{-4}$ m/V^{2}s we obtain the curves in Fig. 4.11. A video line capacitance $C_v = 4$ pF was estimated assuming a 128 by 128 size array of 40 μm pitch. A worst case $\alpha = 10^{-23}$ Joules was inserted.

The curves in Fig. 4.11 show the magnitude of the noise contributions as a function of source follower width. The dominant curve is the flicker noise at the bottom end of useful video bandwidth $f = 50$ Hz. The minimum flicker noise at the high end of the bandwidth is also shown and lies below the kTC noise. In practice, low frequency flicker noise and kTC noise are significantly reduced using correlated double sampling (CDS) [Barbe 75] or delayed double sampling (DDS) [Obho 88], therefore only the thermal noise curve

needs to be considered. This curve actually has a very shallow saddle point (minima) at $W = 1000\ \mu$m, however as the curve varies very slowly we can choose $W = 400\ \mu$m with an insignificant increase in noise. The traditional rule of thumb (extolled by [Burt 95], for example), suggests to match C_g with C_v, giving a large overestimate at $W = 1600\ \mu$m. In terms of noise this is a workable result but is wasteful in terms of power and real estate.

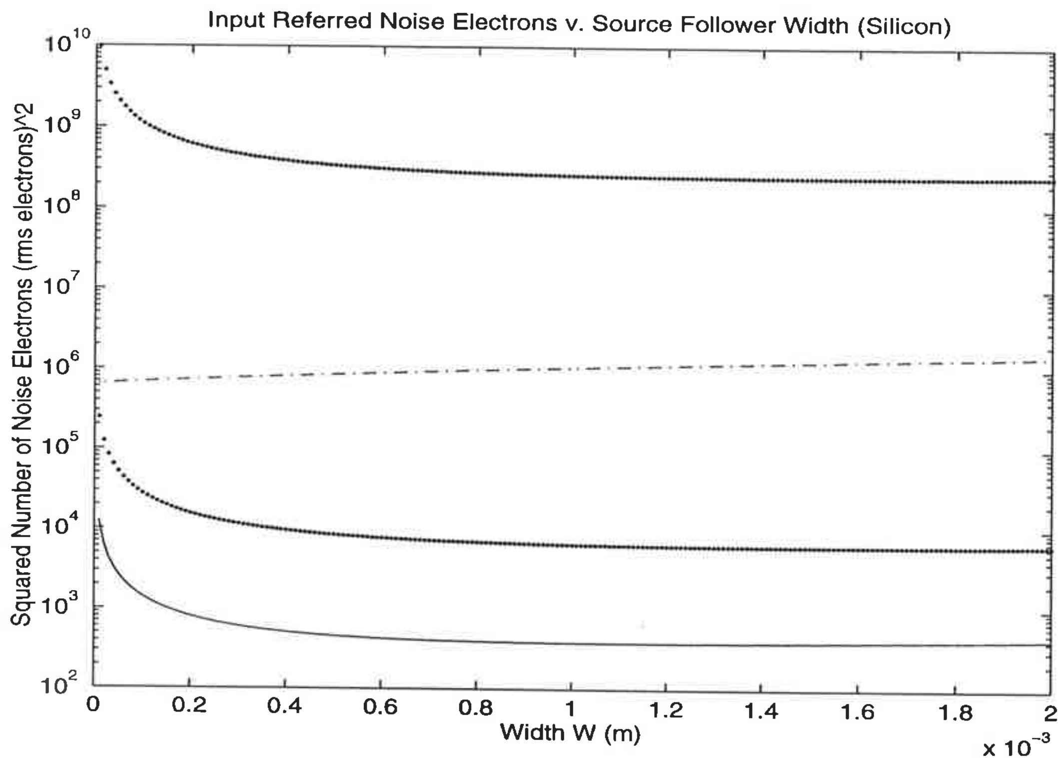


Figure 4.11: Imager output circuit noise sources – silicon. Solid line: thermal noise. Chained line: kTC noise. Dotted lines: maximum & minimum flicker noise (frequency dependent).

4.21.2 GaAs output circuit

The charge-normalised noise formulae in terms of rms electron number are,

Thermal noise,

$$\overline{n_t^2} = 4\frac{kTB}{g_m}\frac{C^2}{e^2} + 2\frac{kTB}{g_m}\frac{(WLC_o)^2}{e^2};$$

Capacitor reset noise,

$$\overline{n_c^2} = \frac{kTC}{e^2}$$

Flicker noise,

$$\overline{n_f^2} = \frac{\alpha_c V_p^2 B}{2aN_d WL}(\frac{1}{f})\frac{C^2}{e^2}$$

Shot noise,

$$\overline{n_s^2} = \frac{2I_g B}{e\omega^2}$$

where $C = C_v + WC_o$ and $g_m = 2W\beta(1 + \lambda V_{ds})\tanh(2V_{ds})(V_{gs} - V_t)$. Inserting typical values: $(V_{gs} - V_t) = 0.5$ V, gate capacitance $C_o = 1.3$ nF/m, bandwidth $B = 2$ MHz, channel doping $N_d = 1.2 \times 10^{23}$ m^{-3}, channel depth $a = 0.1$ μm, pinchoff voltage $V_p = -0.5$ V and beta $\beta = 200$ A/V^2/m we obtain the curves in Fig. 4.12. A video line capacitance $C_v = 0.5$ pF was estimated assuming a 128 by 128 size array of 40 μm pitch. A worst case dimensionless $\alpha_c = 5 \times 10^{-4}$ was inserted.

A notable feature of Fig. 4.12 is that the flicker noise curve ($f = 50$ Hz) is two orders of magnitude lower than in silicon – this is due to the absence of a gate insulator. The number of thermal and kTC noise electrons are also lower than in silicon – this is mainly due to the lower line capacitance in GaAs, leading to a potentially wider dynamic range. However, one set back is that the dominant curve is the MESFET gate shot noise. Unfortunately it is impossible to remove shot noise by correlated double sampling. This means that the circuit cannot be optimised for this MESFET technology. Hence to realise a proof-of-concept imager, in the short term, two options are available: (a) use off-chip output circuitry or (b) used ammonium sulphide, $(NH_4)_2S$, anealed MESFETs.[3] Option (a) may add 1-2 pF of parasitic package and board capacitance to the output node of the imager, however the total output node capacitance will still be less than in silicon. In option (b), anealing with ammonium sulphide is reported as reducing gate leakage currents by 3 orders of magnitude [Cooper 93] – this would reduce the shot noise of a minimum sized transistor down to the level of thermal noise in the silicon case. To achieve this comparable performance to silicon, a multiple stage output would be required; ie. a minimum sized transistor for low shot noise, followed by a larger transistor to drive the required load.

In the longer term, two emerging GaAs technologies are promising: the HIGFET and the anisotype FET. The HIGFET [Abrokwah 94] uses a semi-insulating AlGaAs layer in the gate to reduce leakage currents – the middle shot noise curve in Fig. 4.12 represents this case. It can be seen that the optimum trade-off between shot and thermal noise for this case occurs when $W = 5$ μm.

The anisotype FET [Abrokwah 93] uses a graded semiconductor InGaAs/GaAs layer in the gate, producing yet a further improvement in gate leakage – the shot noise for this

[3]However, there maybe a question over the stability of sulphide treatments.

case is the lowest dashed curve in Fig. 4.12. The optimum trade-off between thermal and shot noise is now at $W = 400\ \mu$m and the noise level is an order of magnitude lower than in the optimum silicon case.

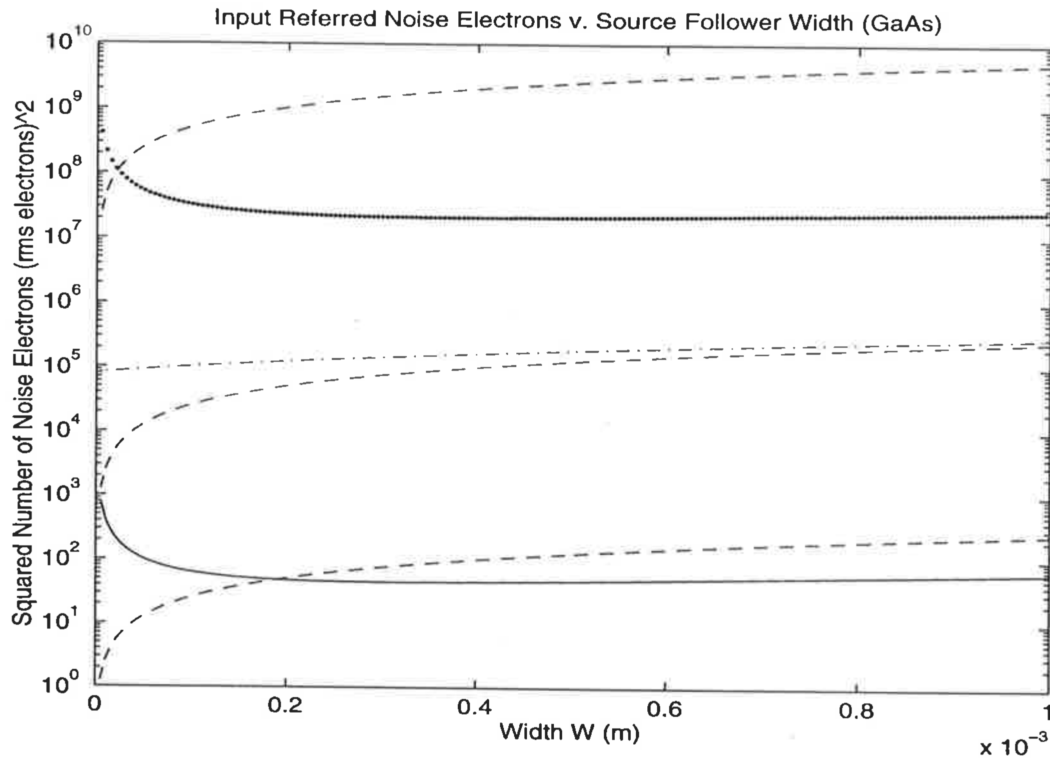


Figure 4.12: Imager output circuit noise sources – GaAs. Solid line: thermal noise. Chained line: kTC noise. Dotted line: maximum flicker noise. Dashed lines: shot noise (upper curve, MESFET, $I_g = 2 \times 10^{-8}$A/m; middle curve, HIGFET, $I_g = 10^{-12}$A/m; lower curve, anisotype FET, $I_g = 2 \times 10^{-15}$A/m.)

4.22 Summary

A brief history of the events leading up to the discovery of thermal noise has been covered, with some biographical information on Johnson & Nyquist, as these details have not been readily accessible in the pedagogical texts or encyclopaedias and in some cases are misleading or incorrect.

We have presented, for the first time, a simple 'engineering proof' of the thermal noise formula, based on Fourier transforms, that avoids lengthy kinetic theory or Wiener formalisms and illustrates the physical assumptions more clearly than the Nyquist proof.

This proof has enabled us to clearly verify and trace the physical origin of the thermal noise coefficient '4.' We have set up and described an automated noise measurement system, necessary for noise parameter extraction for future optimisation of imager designs and for performance testing. However, detailed calibration of the system has revealed a mean value of the thermal noise coefficient 7.5% *lower* than theory. This is the first time that below theory thermal noise has been claimed. The best working hypothesis is one of electrostatic interactions, causing some correlated behaviour between carriers, thereby reducing the noise.

We made a detour by discussing, in detail, the thermal noise formula at high frequencies. In the Tera Hertz region, acheivable by GaAs quantum well tunnel diodes, we show there is uncertainty and debate in the literature surrounding role of ZPF in observable noise. This unsolved problem will become more important as the technology at these frequencies matures. However, we have made a conceptual advance in that we have clearly identified that the idea that quantum mechanics satisfactorily resolves the catastrophe problem is a misconception. We state that the classical catastrophe is merely replaced by a different catastrophe. For the first time, we call this the *quantum catastrophe*. Such problems with quantum mechanics are known, but our original contribution is to clearly state and name this paradigm that is traditionally obfuscated in the literature.

We have also surveyed a number of other debates, misconceptions, conundrums and surprises regarding thermal noise that traditionally cause consternation. In particular we have disputed Pyati's assumption that kTC noise on an open circuited capacitor is time-varying and we have shown it to be a dc voltage. This has enabled us to correctly consider the noise analysis of the output circuit of the imager.

In regards to the output circuit in a silicon imager we have disproved the traditional conception that the optimum output transistor size is chosen when its gate capacitance equals the video line capacitance. In the case of GaAs, we have shown that although a two orders of magnitude reduction in flicker noise is possible, the gate leakage shot noise becomes dominant. One solution is to use an off-chip MOSFET amplifier, where noise performance is still predicted to be improved over a silicon imager – this solution defeats the objective of an imager that is totally monolithic, however serves as an initial proof-of-concept option. Another solution is to use a GaAs process that has been tailored for RAMs by ammonium sulphide annealing, for example, resulting in lower gate leakage. Comparable noise performance to silicon is expected by this approach.

Finally a longer term solution is to use a newly emerging GaAs technology that uses very low leakage anisotype FETs. With this approach we predict up to an order of magnitude improvement in total noise performance over silicon.

For an initial proof-of-concept, the solution of using an external amplifier seems the best solution. Consequently in the next chapter, we shall examine the imager design in terms of the pixels and address circuitry only and not the output circuit.

Bibliography

[Abrokwah 93] J.K. Abrokwah, "Anisotype-gate self-aligned p-channel heterostructure field effect transistors," *IEEE Trans. Elec. Dev.,* **vol. 40,** no. 2, pp. 278-284, 1993.

[Abrokwah 94] J.K. Abrokwah, "A manufacturable high-speed low-power complementary GaAs process," *Int. Conf. Solid State Devices and Materials,* pp. 592-594, Yokohama, Aug. 23-26, 1994.

[Adkins 75] C. J. Adkins, *Equilibrium Thermodynamics,* McGraw-Hill, 1975, p. 202.

[Barbe 75] D.F. Barbe, "Imaging devices using the charge-coupled concept," *Proc. IEEE,* **vol. 63,** pp. 38-67, Jan. 1975.

[Baseia 93] B. Baseia and A. L. De Brito, "Quantum noise reduction in an electrical circuit having a time dependent parameter," *Physica A,* **vol. 197,** pp. 364-370, 1993.

[Beck 76] A. H. W. Beck, *Statistical Mechanics, Fluctuations and Noise,* Edward Arnold, 1976.

[Bell 60] D. A. Bell, *Electrical Noise*, Van Nostrand, 1960.

[Bendat 58] J. S. Bendat, *Principles and Applications of Random Noise Theory,* Wiley, 1958.

[Bennett 92] C. Bennett and S. Wiesner, "Communication via one-particle and 2-particle operators on Einstein-Poldolsky-Rosen states," *Phys. Rev. Lett.*, **vol. 69,** no. 20, pp. 2881-2884, 1992.

[Bennett 93] C. Bennett *et. al.*, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.*, **vol. 70,** no. 13 pp. 1895-1899, 1993.

[Blum 73] A. Blum, "Ein Beitrag zur Begründung der Formel für das Widerstandsrauschen," *Arch. Elek. Uber.,* **vol. 27,** no. 5, pp. 208-212, 1973.

[Bogner 94] Private communication, Prof. R. E. Bogner, 1994. Consider a 2-port 'black box' with a resistor across the input port and another resistor across the output port. If a ferrite microwave isolator is the contents of the box then thermal noise from one resistor apparently heats up the other resistor.

[Brillouin 50] L. Brillouin, "Can the rectifier become a thermodynamical demon?" *Phys. Rev.,* **vol. 78,** p. 627-628, 1950.

[Brown 28] R. Brown, "A brief account of microscopical observations made in the months of June, July & August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic & inorganic bodies," *Philosophical Magazine,* **vol. 4,** no. 21, pp. 161-173, 1828.

[Brown 29] R. Brown, "Additional remarks on active molecules," *Philosophical Magazine,* **vol. 6,** no. 33, pp. 161-166, 1829.

[Browne 93] K. M. Browne, "Focused radiation, the second law of thermodynamics and temperature measurements," *Engineering Optics,* **vol. 6,** no. 2, pp. 127-130, 1993.

[Buckingham 83] M. J. Buckingham, *Noise in Electronic Devices & Systems*, Wiley, 1983.

[Burt 95] D. J. Burt, "Extending the performance limits of CCD image sensors," *GEC Journal of Research,* **vol. 12,** no. 3, pp. 130-140, 1995.

[Bywater 19] J. Bywater, *Physiological Fragments: to Which are Added Supplementary Observations, to Shew that Vital Energies are of the Same Nature, and Both Derived from Solar Light,* R. Hunter, London, 1824, pp. 127-128. (First published in 1819.)

[Callen 51] H. B. Callen and T. A. Welton, "Irreversibility and and generalized noise," *Phys. Rev.,* **vol. 83,** no. 1, pp. 34-40, 1951.

[Cooper 93] J.A. Cooper, "Recent advances in GaAs daynamic RAMs," *Advances in Electronics and Electron Physics,* **vol. 86,** pp. 1-79, 1993.

[Danneville 94] F. Danneville *et al*, "Microscopic noise modeling and macroscopic noise models," *IEEE Trans. Elec. Dev.,* **vol. 41,** no. 5, May 1994.

[Davenport 58] W. B. Davenport & W. L. Root, *An Introduction to the Theory of Random Signals & Noise*, McGraw-Hill, 1958, pp. 183-188.

[Dekker 91] A. J. Dekker, H. Hickman & T. M. Chen, "A tutorial approach to thermal noise in metals," *Am. J. Phys.,* **vol. 59,** no. 7, pp. 609-614, July 1991.

[Delsaulx 77] J. Delsaulx, "Thermo-dynamic origin of the Brownian motion," *The Monthly Microscopical Journal,* **vol. 18,** pp. 1-7, 1877.

[Dym72] H. Dym & H. P. McKean, *Fourier Series and Integrals*, Academic Press, 1972, pp. 121-132.

[Exner 00] F. M. Exner, "Notiz zu Brown's Molecularbewegung," *Ann. d. Phys.,* **vol. 2,** no. 8, p. 843-847, 1900.

[Exner 67] S. Exner, "Untersuchungen über Brown's Molecularbewegung," *Sitzungsber ichte der Kaiserlichen Academie der Wissenschaften,* **vol. 56,** pp. 116-123, 1867.

[Feynman 63] R. P. Feynman, R. B. Leighton & M. Sands, *The Feynman Lectures on Physics*, Addison-Wesley, 1963, **vol. I,** ch. 46 pp. 1-4.

[Forward 82] R.L. Forward, "Reducing noise in photoamplifier circuits with low noise artificial resistors," *J. Appl. Phys.,* **vol. 53,** no. 5, pp. 3365-3371, May 1982.

[Freeman 58] J. J. Freeman, *Principles of Noise,* Wiley, 1958, pp. 107-113.

[Giakos 91] G. C. Giakos and T. K. Ishii, "Rapid pulsed microwave propagation," *IEEE Microwave & Guided Wave Letters,* **vol. 1,** no. 12, pp. 374-375, 1991.

[Gouy 88] Gouy, "Note sur le mouvement brownien," *J. de Phys.*, **vol. 7,** no. 2, p. 561-564, 1888.

[Grabert 88] H. Grabert, P. Schramm and G-L. Ingold, "Quantum Brownian motion," *Phys. Rep.,* **vol. 168,** no. 3, pp. 115-207, 1988.

[Graffeuil 82] J. Graffeuil *et al*, "Low frequency noise physical analysis for the improvement of the spectral purity of GaAs FETs oscillators," *Solid-State Elec.,* **vol. 25,** no. 5, pp. 367-364, 1982.

[Grandy 70] W. T. Grandy, *Introduction to Electrodynamics,* Academic Press, 1970, p. 143 & p. 252.

[Grau 82] G. Grau and W. Kleen, "Comments on zero point energy, quantum noise and spontaneous emission noise," *Solid-State Electronics,* **vol. 25,** no. 8, pp. 749-751, 1982.

[Ingenhousz 84] J. Ingenhousz, *Vermischte schriften physisch medicinischen inhalts,* Wappler, Vienna, 1784, **vol. 2,** pp. 123-126.

[Ising 26] G. A. Ising, "A natural limit for the sensibility of galvanometers," *Phil. Mag.,* **vol. 51,** pp. 827-834, 1926.

[Itzykson 80] C. Itzykson and J-B. Zuber, *Quantum Field Theory,* McGraw-Hill, 1980, pp. 138-142.

[Jevons 78] W. S. Jevons, "Movement of microscopic particles suspended in liquids," *Quart. J. Sci.,* **vol. 15,** pp. 170-171, 1878.

[Johnson 28] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review,* **vol. 32,** pp. 97-109, 1928.

[King 27] P. P. King, *Narrative of a Survey of the Intertropical and Western Coasts of Australia Performed Between the Years 1818 and 1822,* John Murray, **vol. 2,** appendix B, pp. 534-565, 1827, (abbreviated title: "King's voyages to Australia.")

[King 66] R. King, *Electrical Noise*, Chapman & Hall, 1966.

[King 91] M. B. King, "Tapping the zero-point energy as an energy source," *26th Proc. IECE*, **vol. 4,** pp. 364-369, 1991.

[Kiss 88] L. B. Kiss, "To the Problem of Zero-Point Energy and Thermal Noise," *Solid State Communications,* **vol. 67,** no. 7, pp. 749-751, 1988.

[Kittel 58] C. Kittel, *Elementary Statistical Physics*, Wiley, 1958, pp. 141-146.

[Kleen 87] W. J. Kleen, "Thermal radiation, thermal noise and zero point energy," *Solid-State Electronics*, **vol. 30,** no. 12, pp. 1303-1304, 1987.

[Koch 81] R. H. Koch, D. J. Van Harlingen and J. Clarke, "Observation of zero-point fluctuations in a resistively shunted Josephson tunnel junction," *Phys. Rev. Lett.,* **vol. 47,** no. 17, pp. 1216-1219, 1981.

[Kuhn 78] T. S. Kuhn, *Black-Body Theory and the Quantum Discontinuity 1894-1912,* Oxford, 1978.

[Lawson 50] J. L. Lawson & G. E. Uhlenbeck, *Threshold Signals,* McGraw-Hill, 1950, pp. 64-71.

[Lorentz 16] H. A. Lorentz, *Les Théories Statistiques en Thermodynamique – Conférences Faites au Collège de France en Novembre 1912,* B. G. Teubner, 1916.

[Louisell 73] W. H. Louisell, *Quantum Statistical Properties of Radiation,* Wiley, 1973, p. 235 & p. 257.

[MacDonald 62] D. K. C. MacDonald, "On Brownian movement and irreversibility," *Physica*, **vol. 28,** pp. 409-416, 1962.

[Moll 25] W. J. H. Moll and H. C. Burger, "The sensitivity of a galvanometer and its amplification," *Phil. Mag.,* **vol. 51,** pp. 626-631, 1925.

[Moullin 38] E. B. Moullin, *Spontaneous Fluctuations of Voltage,* Oxford, 1938, p. 39.

[Mullikan 24] R. S. Mullikan, "The band spectrum of boron monoxide," *Nature*, **vol. 114,** pp. 349-350, 1924.

[Nyquist 28] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physical Review,* **vol. 32,** pp. 110-113, 1928.

[Obho 88] M. Obho and S. Yamamoto, "Integral delayed and differential noise suppresion method for CCD camera," *ICCE,* pp. 58-59, Chicago, June 8-10, 1988.

[Other 95] A. N. Other, "Test to see if anyone reads this thesis in detail," *Buried in the Noise,* **vol. 10,** pp. 1-100, 1995.

[Papoulis 84] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 1984, pp. 491-492.

[Penfield 66] P. Penfield, "Unresolved paradox in circuit theory," *Proc. IEEE,* **vol. 54,** no. 9, pp. 1200-1201, 1966.

[Puthoff 89] H. E. Puthoff, "Source of vacuum electromagnetic zero-point energy," *Phys. Rev. A,* **vol. 40,** no. 9, pp. 4857-4862, 1989.

[Puthoff 90] H. E. Puthoff, "The energetic vacuum: implications for energy research," *Spec. Sci. and Tech.,* **vol. 13,** no. 4, pp. 247-257, 1990.

[Pyati 93] V. P. Pyati, "An exact expression for the noise voltage across a resistor shunted by a capacitor," *IEEE Trans. Circ. Syst.*, **vol. FTA-39,** no. 12, pp. 1027-1029, Dec. 1992.

[Reif 67] F. Reif, *Statistical Physics,* McGraw-Hill, 1967, p. 248.

[Rhyshik 63] I. M. Ryshik & I. S. Gradstein, *Summen-, Produkt- und Integral-Tafeln*, Veb Deutscher Verlag Der Wissenschaften, 1963, p. 149.

[Robinson 74] F. N. H. Robinson, *Noise & fluctuations,* Clarendon, 1974.

[Schottky 18] W. Schottky, "Über spontane Stromschwankungen in verschiedenen Electrizitätsleitern," *Ann. Phys.,* **vol. 57,** pp. 541-567, 1918.

[Senitzky 93] I. R. Senitzky, "Detection of quantum noise," *Physical Review A,* **vol. 48,** no. 8, pp. 4629-4638, 1993.

[Siegman 64] A. E. Siegman, *Microwave Solid-State Masers,* McGraw-Hill, 1966, p. 210.

[Starr 53] A. T. Starr, *Radio and Radar Technique,* Pitman, 1953, pp. 552-553.

[Ullersma 66] P. Ullersma, "An exactly solvable model for Brownian motion," *Physica*, **vol. 32,** pp. 27-96, 1966.

[Wannier 66] G. H. Wannier, *Statistical Physics*, Wiley, 1966.

[Welton 48] T. A. Welton, "Some quantum observable effects of the quantum mechanical fluctuations of the electromagnetic field," *Phys. Rev.,* **vol. 74,** no. 9, p. 1157-1167, 1948.

[Wiener 63] C. Wiener, "Erklärung des atomistischen Wesens des tropfbar-flüssigen Körperzustandes, und Bestätigung desselben durch die sogenannten Molecularbewegungen," *Ann. Phys. Chem.,* **vol. 118,** p. 79-94, 1863.

[de Hass-Lorentz 13] G. L. de Haas-Lorentz, *Die brownsche bewegung und einige verwandte erscheinungen,* Braunschweig, F. Vieweg und Sohn, 1913. (In German. The Dutch original was published in 1912.)

[van der Ziel 81] A. van der Ziel, "Quantum noise at high frequencies and low temperatures," *Sixth Int. Conf. Noise Phys. Sys.,* pp. 364-368, April 6-10, 1981.

Chapter 5

Imagers

"An idea is putting the truth in check-mate."
José Ortega y Gasset (1883-1955)
SPANISH PHILOSOPHER AND AUTHOR

"It doesn't matter how new an idea **is***:*
what matters is how new it **becomes."**
Elias Canetti (1905-)
BULGARIAN-AUSTRIAN NOVELIST AND PLAYWRIGHT

"Mankind always sets itself only such problems as it can solve."
Karl Marx (1818-1883)
GERMAN SOCIOLOGIST, ECONOMIST AND POLITICAL THEORIST

5.1 Introduction

Ch. 4 demonstrated how future work can lead to a low noise on-chip output circuit, but in the present scope of this work shot noise was found to be too high. We therefore limit the study of the imager design, in this chapter, to only the pixels and readout registers. For initial proof-of-concept, a standard off-chip amplifier must be used. Ch. 3 concluded that the passive pixel approach is the most appropriate for the initial imager design and this is now elaborated in this chapter.

5.2 Pixel Design

5.2.1 Illumination Modes

There are three possible illumination modes for the MESFET,

- Backside
- Thin gold or ITO gates
- Fingered transistor

Backside illumination requires substrate thinning to ~100 μm and is a major complication and reliability hazard, for GaAs. Transparent gates (eg. thin gold or ITO) are attractive but require expensive process changes.

Our preferred mode is to select an approach that requires no process changes. The reduced sensitive area introduced by opaque gates is offset by the increased responsivity, as discussed in Ch. 2.

5.2.2 Photocollection Modes

All present-day silicon 2-D imagers are designed around a photovoltaic collection technique. Not only does the photovoltaic approach lend itself to the architecture of the imager, but the details of the physics reveals a lower generation noise. Also a further reduction in the effect of noise is achieved, over photoconductive devices, as photovoltaic devices intrinsically perform a capacitive temporal integration of the signal – thus improving the signal-to-noise ratio.

S/N can be increased in photoconductive devices, by increasing their size – ie. effectively introducing a spatial integration. However any significant increase in area would render a 2-D imager array inordinately large.

Therefore it is this fact that photoconductive devices are in the spatial domain, whereas photovoltaic devices are in the temporal domain, that leads to the fundamental reason why the photovoltaic approach is favoured for 2-D arrays, in the interest of a high S/N per area ratio.

Having chosen that the imager must be designed around the optimisation of photovoltaic mechanisms, we need to decide on which MESFET depletion region should be utilised. Making the correct choice is crucial as the photovoltaic mechanism relies on the electric field, supported by a depletion region, to separate electron-hole pairs and thus collect charge.

The following depletion regions are available in the MESFET:

- Surface depletion region
- S/D to substrate depletion region
- Gate depletion region
- Channel to substrate depletion region.

The surface depletion region can be immediately disregarded, as it is not a 2-terminal structure and therefore charge readout would be impossible. The source/drain to substrate or channel to substrate depletion regions may be used – however, in a following subsection, we show that the series resistance due to the substrate introduces a large RC time constant. Fortunately, it turns out that the time constant does not introduce a problem as standard TV rates are quite slow anyhow.

The gate depletion, as will be shown, has a much shorter RC time constant and therefore would become important in fast applications such as machine target tracking. However, use of the gate has the short-term disadvantage that it is opaque and, in any case, has a much smaller photocollection volume. In its favour, the fact that the gate forms a Schottky diode provides a means for antiblooming.

Selection of the gate photocollection mode, to enable antiblooming, immediately suggests two useful constraints for commencement of one of the pixel designs. Firstly, the pixel must include a dual-gate MESFET – one gate is to be shielded from the light and acts as a switch to enable/disable readout, and the other is used for photocollection. Secondly, as

the gate material is opaque, the detector gate is to be designed with a 'fingered' structure to maximise the amount of depletion region [1] that is exposed to generated photocharge. The area between gate fingers will contain source-to-substrate depletion region and will thus significantly add to the overall photocollection volume.

It is intended that the use of a fingered transistor will act as a simple 'proving ground' to demonstrate the working concept of the imager. Future proposals will address the concept of using transparent gates. Gates can be made transparent either by choice of material or by sufficient thinning.

It is anticipated that the channel/substrate region, under each gate finger, will have an advantageous side effect – namely, stray electrons in this region will be swept into the channel. This photocollection mechanism was established in Ch. 3 by monochrometer experiments.

5.2.3 Channel to Substrate Mode

In order to calculate the effect due to the RC time constant associated with the channel/substrate mode. We must firstly calculate the associated depletion region width. The depletion region width, W, is given by,

$$W = \sqrt{\frac{2\epsilon_o \epsilon V_{bi}}{qN}}$$

– where V_{bi} is the built-in voltage. Many researchers, in the past, have been guilty of calculating V_{bi} using the standard formula used in silicon. This is incorrect because the SI GaAs substrate is, in fact n-type, which forms a *hi-lo* junction with the n-type channel – whereas the standard silicon formulation assumes a standard *pn* junction. Furthermore, the traditional formula does not take into account the peculiar nature of the SI substrate with its various compensated deep and shallow levels. A good approximation is the four-level model of George *et al* [George 89] given by solving Poisson's equation, based on Fermi-Dirac statistics,

$$V_{bi} = \frac{E_g}{q} - \frac{kT}{q}\ln\left(\frac{N_c}{N_{ch}}\right) - \frac{kT}{q}\ln\left(\frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}\right)$$

where:

$$\alpha = (N_{sd} - N_{sa} - N_{da}) g_{dd} \exp\left(\frac{-E_{dd}}{kT}\right)$$

[1]Even though the gate is opaque, the depletion region laterally protrudes from under the gate and is thereby exposed.

$$\beta = N_{sd} + N_{dd} - N_{sa} - N_{da} + (N_{sd} - N_{sa})g_{da}g_{dd}\exp\left(\frac{E_{da} - E_{dd}}{kT}\right)$$

$$\gamma = (N_{sd} + N_{dd} - N_{sa})g_{da}\exp\left(\frac{-Eda}{kT}\right)$$

Approximate values in the above expression are,

E_g=1.42 eV

$N_c = 2.8 \times 10^{19}\text{cm}^{-3}$

$N_{ch} = 1.2 \times 10^{17}\text{cm}^{-3}$

$N_{da} = 10^{16}\text{cm}^{-3}$ @ 0.75 eV

$N_{sd} = 10^{15}\text{cm}^{-3}$ @ 1.35 eV

$N_{sa} = 10^{14}\text{cm}^{-3}$ @ 0.05 eV

$N_{dd} = 10^{14}\text{cm}^{-3}$ @ 0.65 eV

As α comes out negative, the rule is to then take the negative sign in the quadratic solution. Assuming no deep level degeneracy, g_{dd}=g_{da}=1, we obtain, the built-in voltage V_{bi}=0.59 V and thus the depletion region width W = 1.8 μm.

For a 20 μm^2 area, $C = \frac{\epsilon_o \epsilon A}{W} \approx 25$ fF and taking $R > 100$ MΩ, using,

$$f = \frac{1}{2\pi RC}$$

results in a maximum pixel frequency rate of, f <64 kHz.

Actual pixel rates must be *lower* than this limit. Therefore this mode of operation is appropriate for standard TV rates, which requires imager data rates of the order of MHz and hence pixel rates < 1 kHz. This conclusion equally holds for the source/drain to substrate mode – as the source/drain doping concentration is an order of magnitude greater than that of the channel.

We will now retrospectively compare the above formulation of V_{bi} with the following traditional formulae:

pn Junction

$$V_{bi} = \frac{kT}{q}\ln\left(\frac{N_A N_D}{{n_i}^2}\right)$$

Hi-lo Junction

$$V_{bi} = \frac{kT}{q} \ln(\frac{N^+}{N^-})$$

where N^+ and N^- are the two n-type doping concentrations across the hi-lo junctions – and, N^+ is the larger of the two.

For a typical silicon MOSFET,

$N_D = 10^{16}$ cm^{-3}

$N_A = 10^{19}$ cm^{-3}

$n_i = 1.45 \times 10^{10}$ cm^{-3}

therefore W=0.34 μm (pn) or W=0.15 μm (hi-lo)

For a typical GaAs MESFET,

$N_{sub} = 2.5 \times 10^{14}$ cm^{-3}

$N_{ch} = 1.2 \times 10^{17}$ cm^{-3}

($N_{drain} = 10^{18}$ cm^{-3})

$n_i = 1.79 \times 10^6$ cm^{-3}

upon substituting the above, once again we obtain, W=2.5 μm (pn), which is clearly an overestimate; similarly $\Rightarrow$ W=1.9 μm (hi-lo).

This clearly illustrates that the incorrect use of the pn formulation on SI GaAs MESFETs results in an *overestimate* of the depletion region width. Furthermore, for rough order-of-magnitude calculations, it would appear that the simple hi-lo formula, is adequate for GaAs MESFETs – and the more complex 4-level compensation model can be used for more precise modelling.

Comparing these calculations with a typical silicon MOSFET, confirms that the built-in Si junction depletion regions are typically smaller than in GaAs.

5.2.4 Gate to Channel Mode

The calculation of the gate depletion region width is much simpler as it only requires the standard formula for a Schottky diode. This yields, V_{bi} ≈0.85 V, thus W = 0.1 μm and C=100 fF for 10 μm sq.

Using $f = \frac{1}{2\pi RC}$ and R≈100 kΩ,

$\Rightarrow f \approx 16$ MHz.

Thus photocollection in the gate/channel depletion region mode can lead to GaAs imagers that operate significantly faster than at normal TV rates. This is useful for applications such as machine target tracking.

5.2.5 Pixel Active Area

The size that each detector pixel must be designed to, depends on the pixel storage capacity required to produce a detectable signal current.

The signal current is related to the pixel capacitance by,

$$I_s = \frac{1}{\tau}\int_o^{V_{max}} C dV$$

As C is voltage dependent, we substitute in the standard capacitance formula for a Schottky barrier diode, and rearranging the relation yields:

$$A = \frac{I_s \tau}{2\sqrt{V_{bi} - V_{max}} - \sqrt{V_{bi}}}\left(\frac{2}{q\epsilon_o \epsilon N_d}\right)^{\frac{1}{2}}.$$

Typical values are,

$V_{max} = 2.5$ V

$\tau = 0.4$ μs (for 128 by 128 imager at TV rates)

$I_s \approx 300$ nA (common in present CCDs)

resulting in an active pixel area of $\sim$10 μm by 10 μm, therefore a die size of $\sim$3-5mm sq is quite feasible for 128 by 128 pixels.

In conclusion, both the pixel size and overall die size are elegant and achievable.

5.2.6 Pixel Design

The pixel conceptual layout is shown in Fig. 5.1, for the fingered case. The salient features are as follows:

- The address line enables/disables pixel readout
- This switch is shielded from light by the data line metal

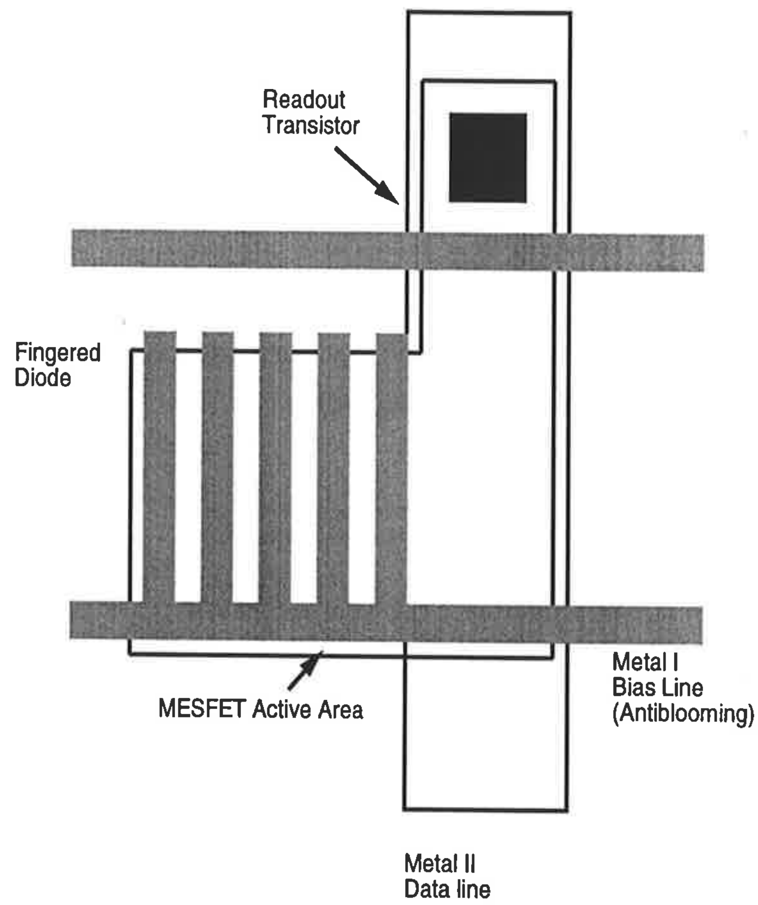


Figure 5.1: Fingered Transistor Dual-Gate Pixel Conceptual Layout

- The fingered layout optimises the amount of gate depletion volume presented to generated photocharge (optimum finger number and spacings are still to be determined)
- The layout can step-&-repeat to form an array (there is scope for making the layout more efficient in this regard)
- The fingered gates are connected to a bias line that can be be utilised to control blooming – this useful feature is made possible by the fact that the gate is a Schottky diode (– in MOS silicon arrays, the MIS structure does not allow this facility)
- The drawn contact between Metal II and Active Area, in reality would be done in two stages, using Metal I.

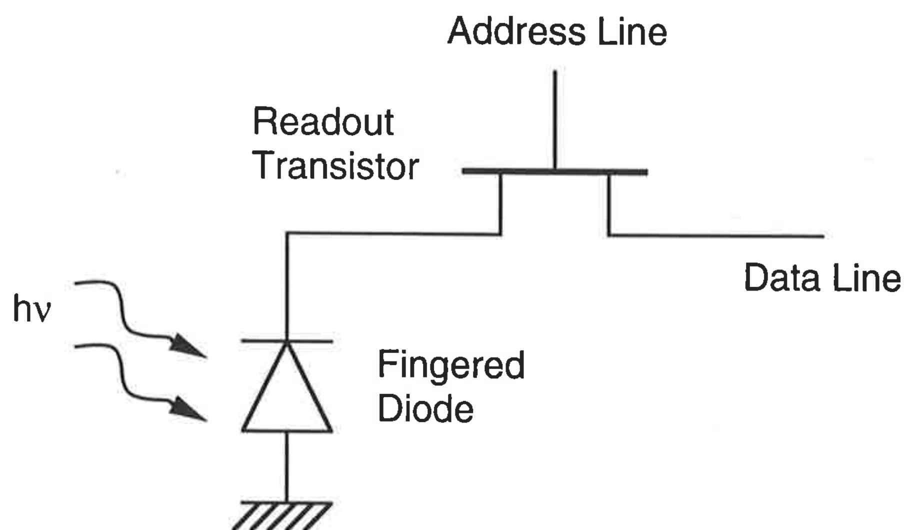


Figure 5.2: Pixel Equivalent Circuit

The proposed pixel layout can be modelled by the equivalent circuit shown in Fig. 5.2. In the fingered case, the diode represents the Schottky gate diode in parallel with the channel/substrate diode. The operation is as follows:

(1) The address line is pulsed to turn on the readout transistor

(2) Accumulated photocharge, in the diode, is dumped onto the data line by capacitive charge sharing

(3) When readout is complete the diode settles to the data bus voltage minus a threshold drop and the readout transistor is then turned off

(4) The diode is now in a reversed biased state

(5) Photons incident on the diode generate electron-hole pairs in the underlying GaAs

(6) Electron-hole pairs are separated by the field, supported by the depletion region. Electrons are swept up and accumulate on the diode's self-capacitance, causing the depletion region to contract

(7) The sequence repeats from (1).

Note that when photocharge collects this is called the *integration* period and it is then dumped in the *readout* period. The integration period is much longer than the readout period – so the effect of incident light during the readout period is negligible. Also as photocollection is much more efficient, than in silicon, the problem of photocharge finding its way to the depletion region of the readout gate is greatly reduced. This is yet another example of the inherent superiority of the GaAs array.

Fig. 5.3 shows the actual layout of 3 by 3 pixels. Each pixel consists of one transistor with an enlarged unconnected source terminal. This source terminal acts as the photosensitive source/substrate diode. The gate of the transistor is shielded from the light by a layer of second level metal.

An alternative layout is shown in Fig. 5.4, where the source has a number of Schottky metal fingers over its surface. As well as increasing the depletion volume for photocollection, these Schottky fingers provide antiblooming protection in that they provided a path for the forward bias current. This elegant mechanism is, of course, not possible in MOS imagers. The spacing of the fingers is unimportant for the first prototype, but can be optimised in the future by including a test structure with a number of spacing options.

Of the two alternative designs, we recommend the second option for fabrication. This option can reduce to the first case by appropriately biasing the Schottky fingers and accounting for the reduced fill factor. In this way, the fingered design will provide information on both approaches. It will also provide some preliminary information on utilising the edge gain effect, if a large resistor is placed in series with the common connection to the Schottky fingers.

5.3 Array Organisation

The array organisation required to read out 128 by 128 pixels will follow the standard techniques used for silicon XY array imagers. Basically, one shift register addresses a

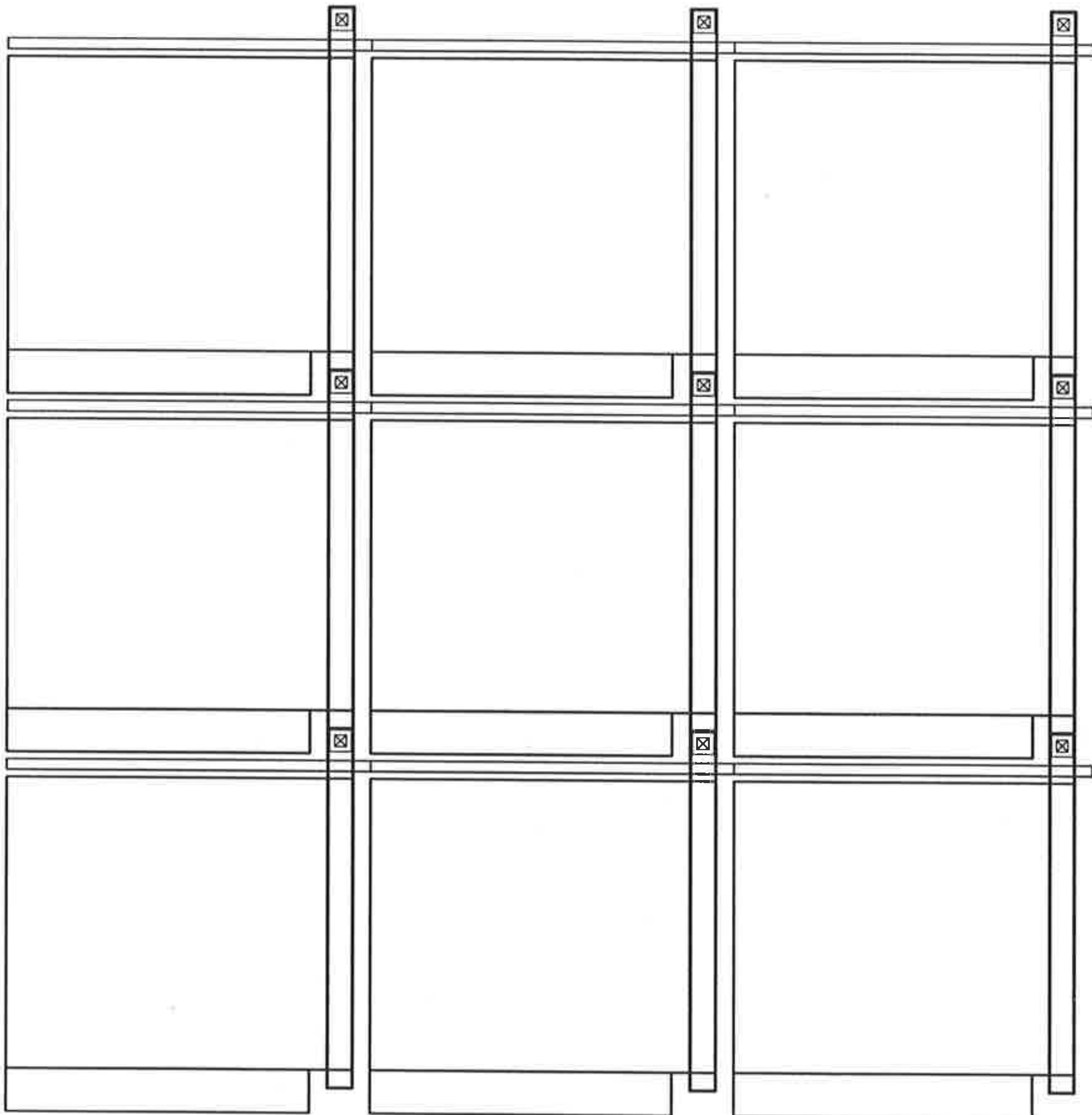


Figure 5.3: Layout of 3 by 3 pixels.

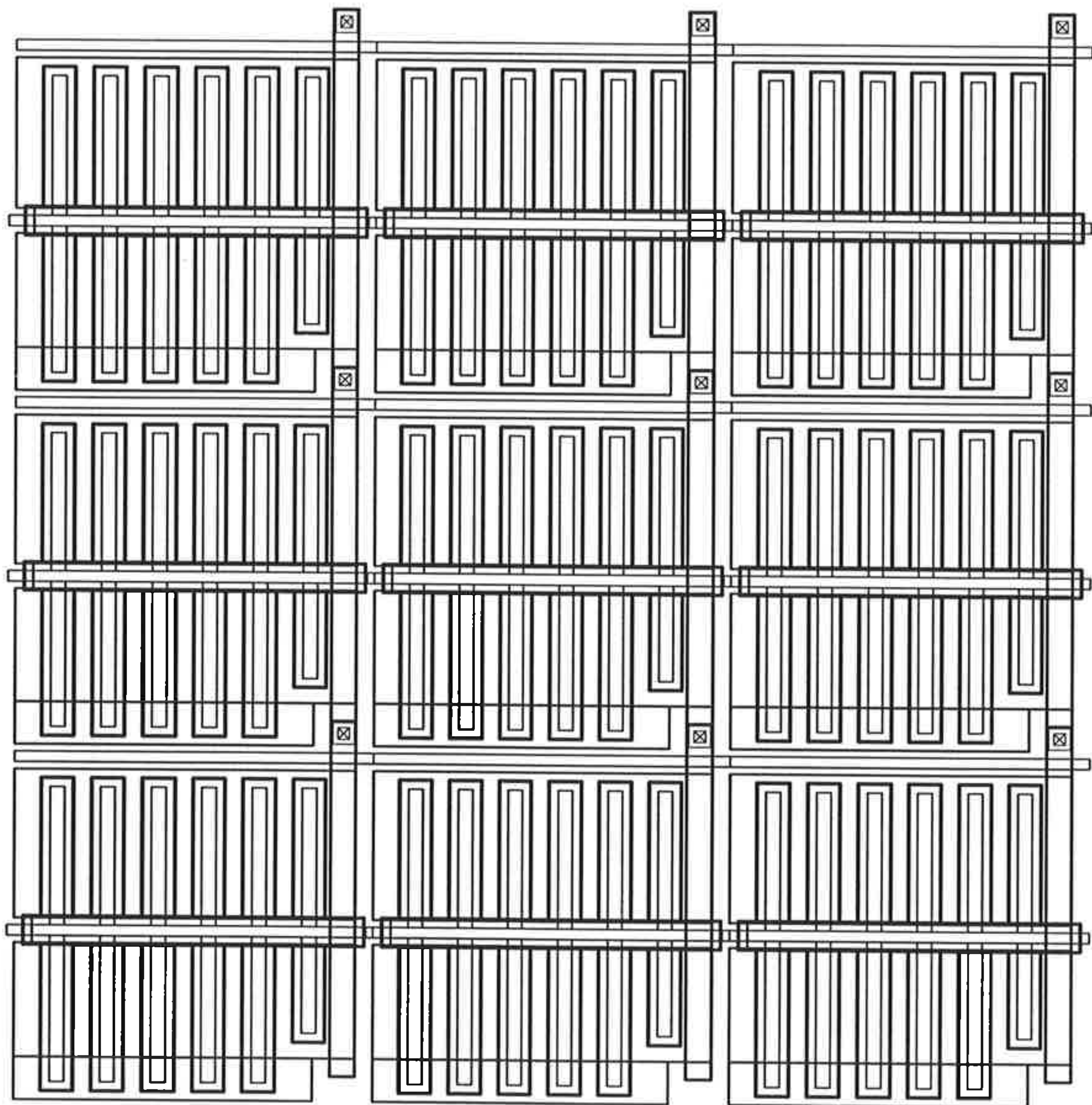


Figure 5.4: Layout of 3 by 3 pixels, with antiblooming structure.

series of paths that control the pixel enable/disable gates.

A second shift register is used to address a multiplexer that converts the parallel data into a serial stream.

If a *charge sensing* approach is adopted, for readout, charge is transferred from pixel to data line and from data line to output bus by a process of capacitive charge sharing. This is a well understood and proven technique used in silicon XY imaging arrays.

Silicon imagers have been known to adopt either the technique of *voltage sensing* or *charge sensing* for image readout. Voltage sensing entails detecting a voltage at each pixel and charge sensing reads out the charge from each pixel. Basically, charge sensing is the superior technique because:

- each pixel does not need to be explicitly reset (the action of charge readout is the reset)
- one capacitor is used for charge-to-voltage conversion, at the output, rather than one at every pixel, therefore fixed pattern noise is reduced
- charge-to-voltage conversion occurs at the final output and not at each pixel node, therefore pixel circuitry and hence space is reduced.

The only real advantage with voltage sensing is the increased signal-to-noise ratio due to the small pixel capacitance used for charge-to-voltage conversion. In the charge sensing case, a lower signal-to-noise ratio results from the fact that the charge-to-voltage conversion capacitance is unavoidably larger – inherently due to the architecture of the array and is required to ensure efficient charge transfer to the output bus. However, because of the space saving & uniformity advantages, charge sensing has predominated in silicon XY imagers and careful design overcomes any S/N problems.

GaAs has the advantage that the lower metal track capacitances improve the S/N and therefore charge sensing lends itself to this technology.

Although we expect an improved S/N, with GaAs, there is the question of increased fixed-pattern-noise (fpn) over silicon imagers. This may be due to transistor to transistor threshold variations. However, rapid advances are being made in GaAs processing and such variations are expected to continually decline as the technology matures.

Another potential problem area is that of threshold voltage drops due to MESFETs in series. The pixel and multiplexer switches are effectively in series, but this problem can be solved by

(a) capacitive bootstrapping techniques

(b) choosing a suitably high pixel reset voltage

(c) using depletion mode switches

(d) moving to CGaAs (ie. with GaAs HIGFETs).

Method (a) is fraught with problems due to gate leakage currents discharging the bootstrap capacitors – this technique would work in CGaAs but would then be redundant! In the case of (b), we have a viable option – due to the architecture of the array, this voltage never 'sees' a gate that would create a dangerous forward bias situation. Method (c) would work, but at the expense of increased power dissipation. Case (d) would give us proper complementary pass gates with no threshold drops and thus is the ideal solution for future work. In the short term, (b) therefore remains the best immediate option – in practice the optimum reset voltage would be determined by varying it until the best picture is obtained on a TV monitor.

A schematic of the standard structure of an XY array, utilising the charge sensing approach is shown in Fig. 5.5. Fig. 5.6 shows how this structure is modified when the Schottky fingers are employed to provide the antiblooming facility.

5.4 Address Circuitry

Having discussed the pixel array in detail, we now examine how these pixels are addressed. We saw in Figs. 5.5 & 5.6 that a 'row select register' is required enable each pixel and a 'column select register' is required to enable readout via enabling analog parallel-in/serial-out multiplexer. The two select registers can be of the same design.

The two options for select register archtecture are (1) a shift register or (2) a decoder. Shift registers are the usually favoured option in silicon XY arrays because pitch-matching is easier and the resulting pin-count is smaller. The ostensible advantage of decoders is 'random' access of the XY array for 'homing in' on a moving object, for target tracking purposes – however to the author's knowledge there are no successful reports of this in the literature. One obvious problem with non-sequential addressing is that non-sequential switching transients that feed through onto the output signal, by capacitive coupling, become more difficult to suppress – leading to increased fixed pattern noise (fpn). On the other hand, the synchronous switching transients caused by a shift register can be greatly reduced by a well-known charge integration technique [Ohba 80]. For this reason we will select the shift register approach. In the following subsections we look at 3 shift register case studies.

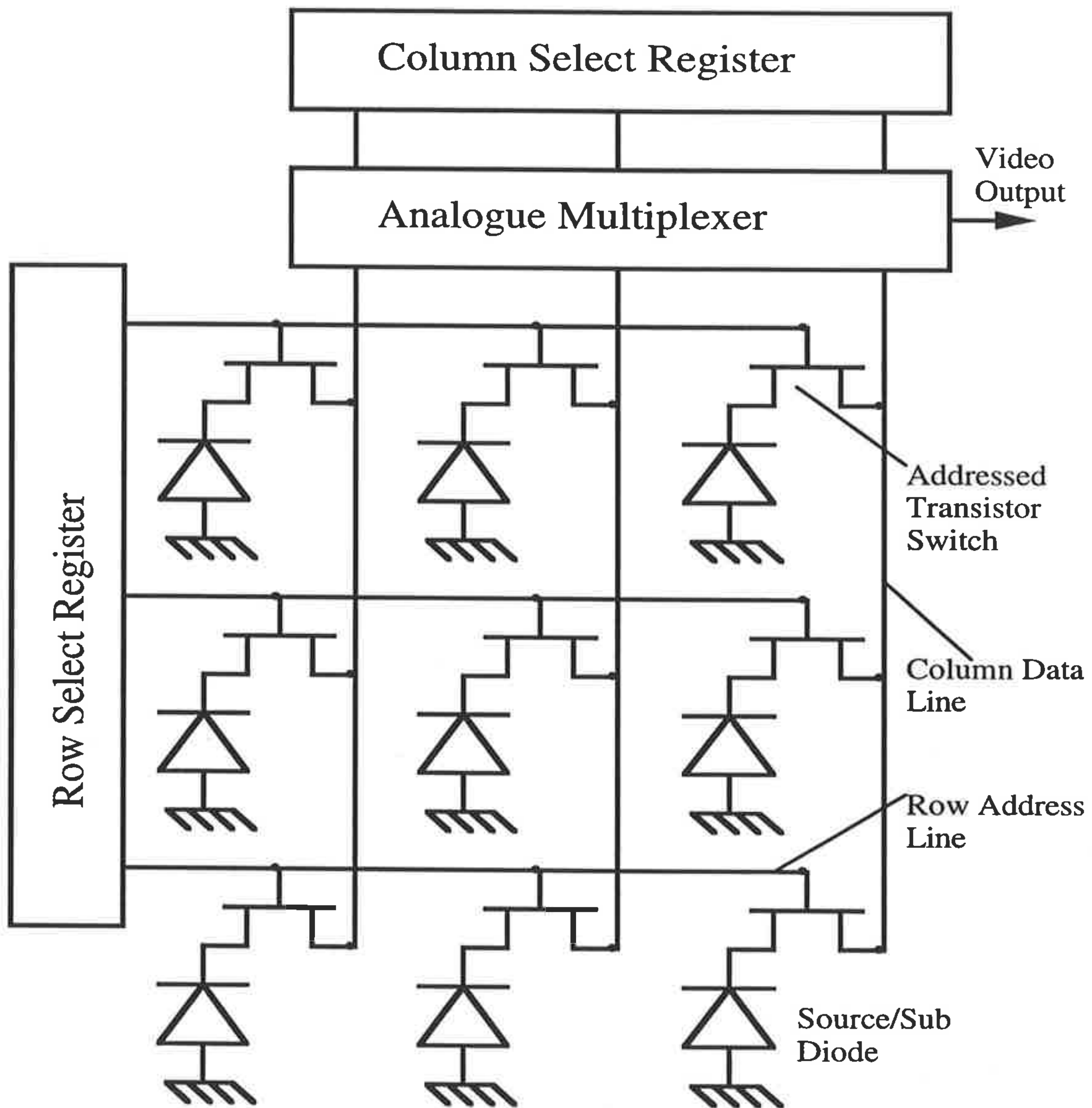


Figure 5.5: Array organisation.

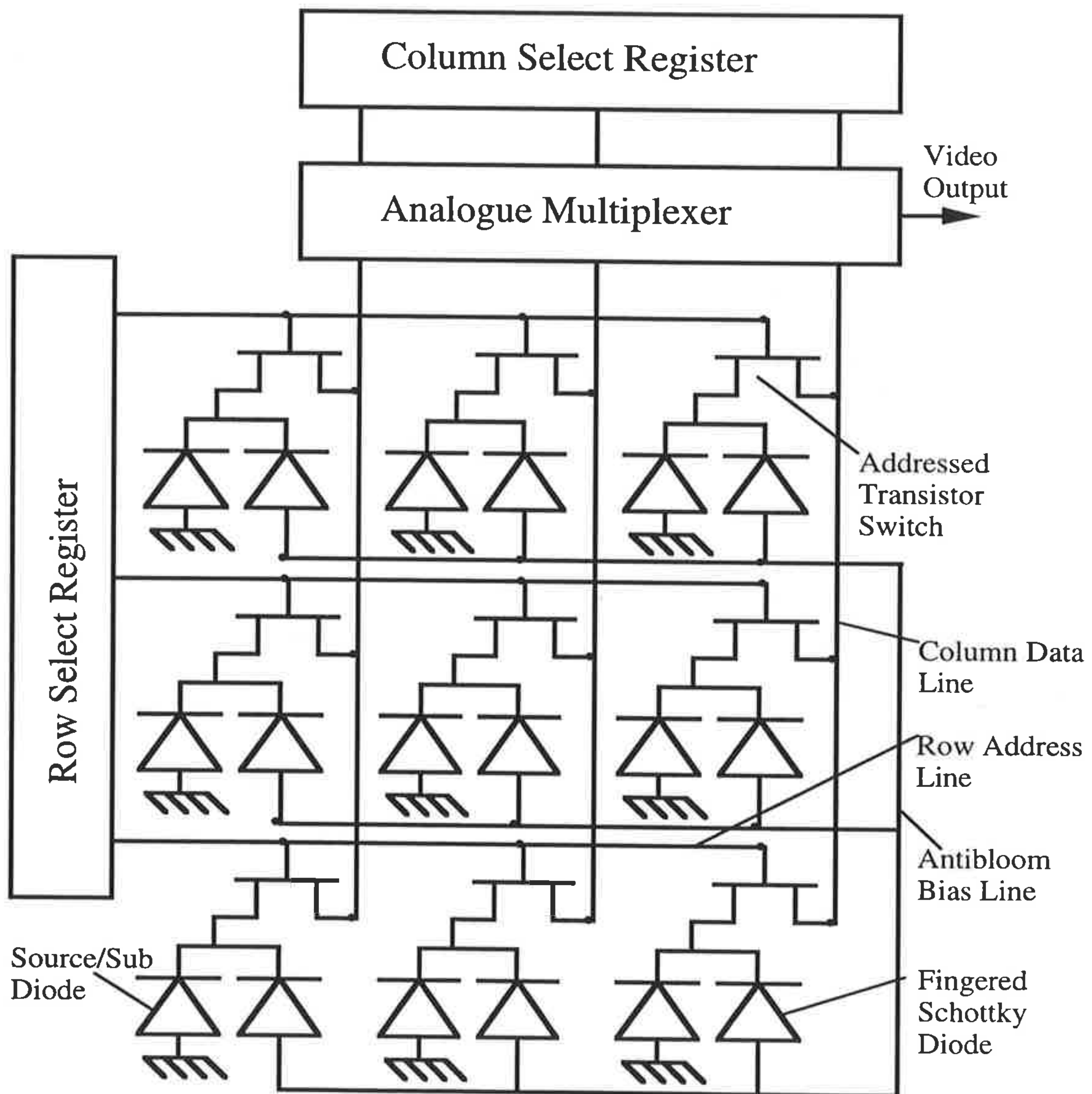


Figure 5.6: Array organisation with antiblooming structure.

5.4.1 Case 1: The Capacitively Bootstrapped Shift Register

Fig. 5.7 shows an original GaAs shift register with a 2-phase low power dynamic design. Each stage is non-inverting and consists of only 4 non-ratioed enhancement minimum sized MESFETs! An MIM capacitor provides bootstrapping action to overcome the threshold voltage drop problem for MESFET pass transistors. We shall refer to this new logic family as 'non-ratioed bootstrapped logic' (NRBL).

This register was devised and selected because of it's compact design and its ability to function with non-overlapping clocks. Non-overlapping clocks are a prerequisite of the synchronous clock transient cancellation scheme [Ohba 80], required to minimise fpn. Notice that the design in Fig. 5.7 also has no dc power busses! The register is entirely driven by the 2-phase clocks, leading to no static power dissipation and hence a low power design.

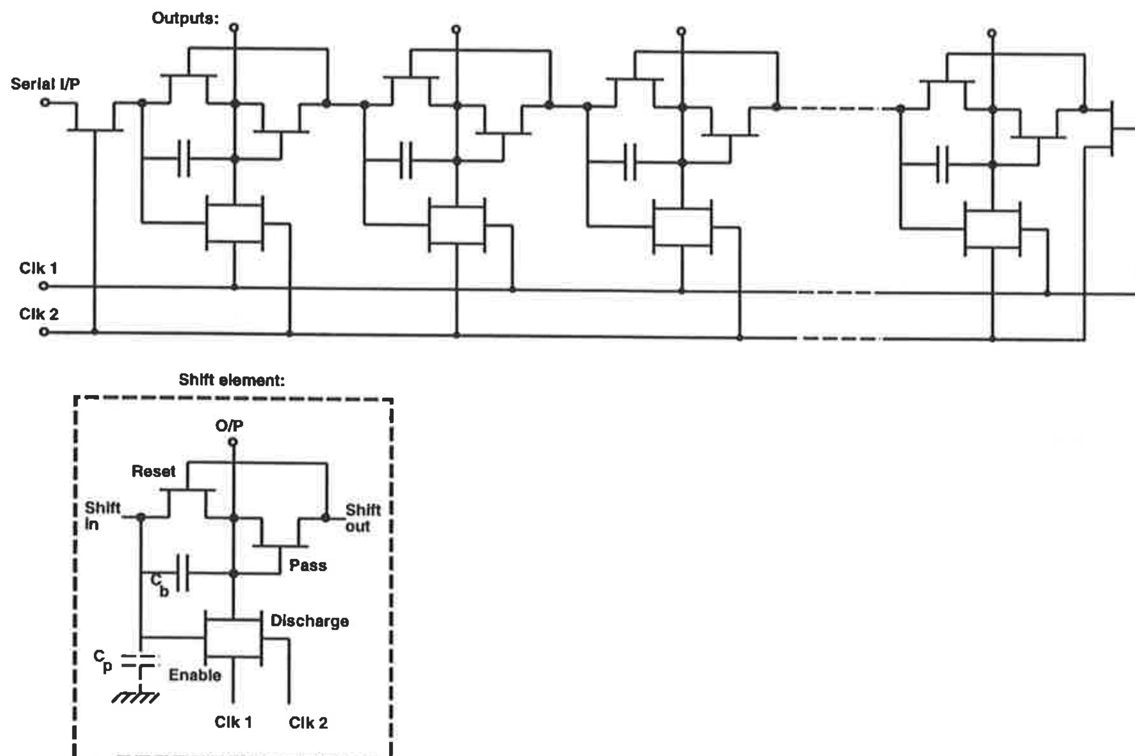


Figure 5.7: Shift register with bootstrapped stages

With reference to Fig. 5.7, the circuit operation is as follows. The drain of each enable transistor is pulsed every `Clk 1` period. The output of each shift element remains low until a high is enabled by the previous stage. When a high appears at the gate of an enable transistor, the output is charged up by the clock pulse. This pulse does not get

cut off at a threshold voltage (V_t) below the clock amplitude (V_{clk}) due to the action of the bootstrap capacitor C_b. C_b pulls up the gate of the enable transistor by V_{boot}, holding it at $V_{clk} + V_{boot}$, where

$$V_{boot} = \frac{C_b}{C_b + C_p} V_{clk}$$

and C_p is the stray capacitance to ground at the input of each stage. As the output is charged up to a full V_{clk}, $V_{clk} - V_t$ appears at the gates of the reset transistor and the following enable transistor via the pass transistor. When `Clk 1` starts going low the output drops. When the output is $2V_t$ below V_{clk}, the reset transistor turns on and resets the bootstrap capacitor. This causes the enable transistor to turn off when the output is at this level. When the next clock phase comes on (`Clk 2`) the discharge transistor turns on and completes the discharge of the output. (As the gate of the pass transistor is now at a lower potential than its source, the bootstap capacitor of the next stage is not discharged.) The same cycle repeats for the next stage and so on. The functions of each of the transistors are summarised as follows,

- *Enable transistor:* Connects clock to output when enabled by the previous stage
- *Reset transistor:* Resets bootstrap capacitor
- *Pass transistor:* (a) passes initial current to charge up bootstrap capacitor of next stage, (b) when output discharges, it turns off so that the bootstrap capcitor of the next stage does not discharge, (c) provides a threshold voltage drop so that the reset transistor is kept hard off until the output potentil drops.
- *Discharge transistor:* As the enable transistor turns off before the output settles, the discharge transistor provides the output with a conduction path to ground in the alternate clock cycle.

As can be seen in Fig. 5.7, extra transistors are reqired at the input and output ends of the shift register. These are required, respectively, to clock the serial input pulse in and to discharge the last bootstrap capacitor, which would otherwise remain permanently charged. A significant advantage of this type of register, is that the output scan pulses have exactly the same amplitude and width as the clock pulses, giving a considerable degree of external control over the operation of the XY array.

The power dissipated by the shift registers is due to the charging and discharging of the capacitance that each stage drives, C_L. As only one stage is active at any one time, the

total power dissipation is simply $C_L V_{clk}^2 f$ – where f is the clock frequency. For standard TV rates the dissipation is below 1 μW! This is favourably 5-6 orders of magnitude lower than that of ratioed-type shift registers that are sometimes known to give rise to thermal heating problems in large arrays.

HSPICE simulations disappointingly revealed that this shift register did not function at all. It appears that because each `shift in` node (see Fig. 5.7) has two transistor gates connected to it, the gate leakage current of MESFETs is so large that the bootstrap capacitor is prematurely discharged. This hypothesis was confirmed by reducing the gate leakage parameter in HSPICE by two orders of magnitude, at which point the register began to function normally.

However, because such a low leakage parameter is not readily available, this approach was abandoned. However, this study was instructive and will be of great merit for future work with GaAs HIGFETs. Due to the semi-insulating gate of HIGFETs the gate leakage current is reduced dramatically and this style of register will be ideal.

5.4.2 Case 2: Conventional D Flip-Flop Shift Register

Due to threshold voltage drops and small signal swings in GaAs circuits, it is well known that NAND gate structures are verbotten. Consequently, the design in Fig. 5.8 is based entirely on NOR gates. This design consists of two D flip-flops per shift register stage driven by two-phase clocks.

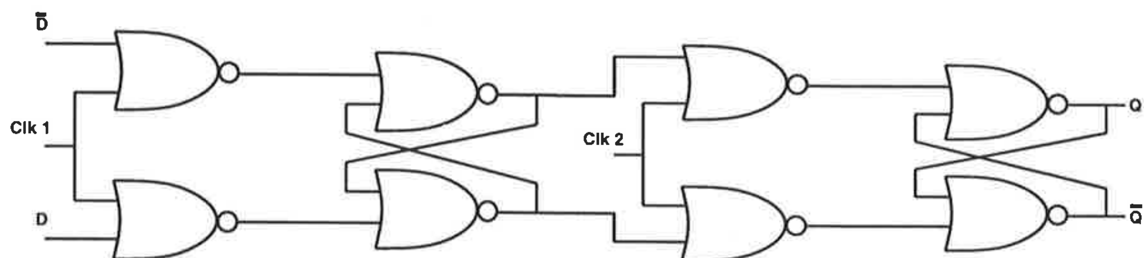


Figure 5.8: Conventional D flip-flop shift register stage

A layout segment of the register is displayed in Fig. 5.9 and a working HSPICE simulation is shown if Fig. 5.10.

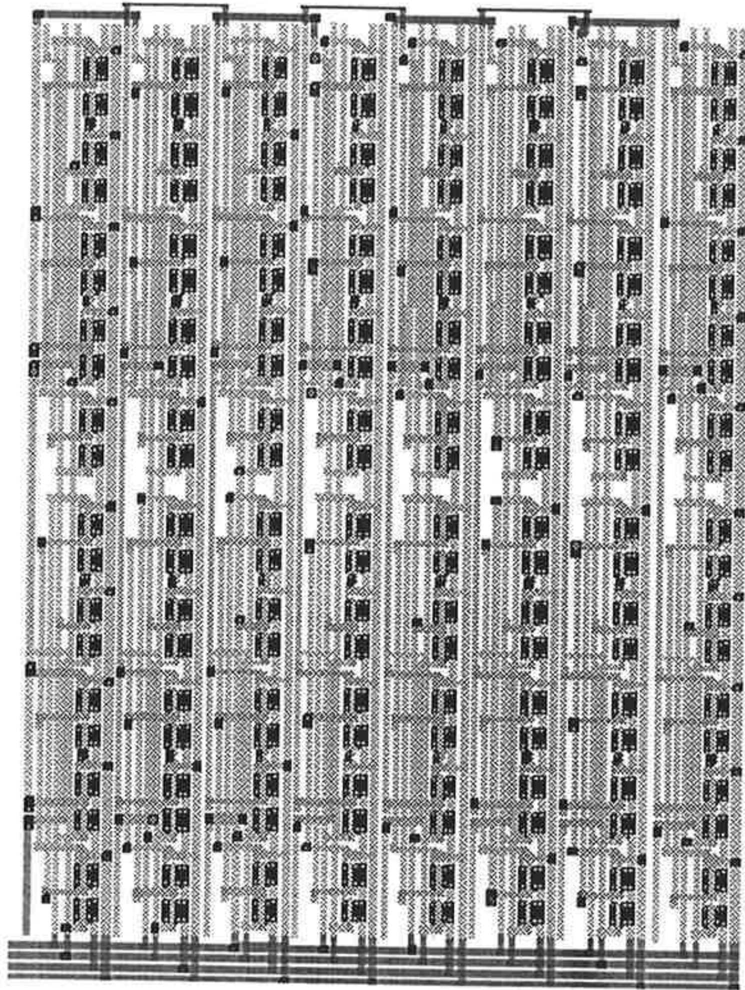


Figure 5.9: Conventional D flip-flop shift register layout segment

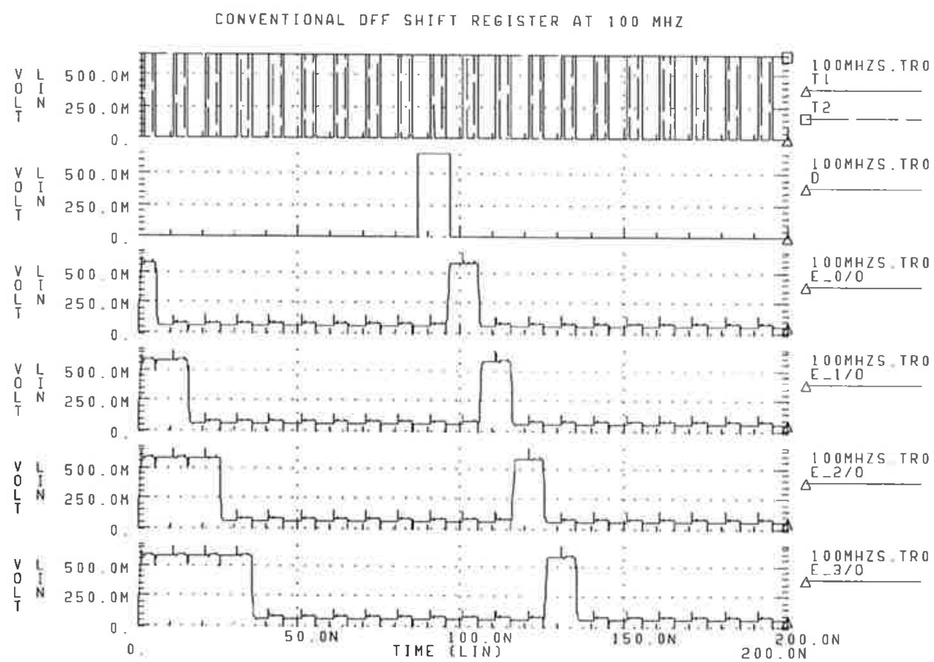


Figure 5.10: Conventional D flip-flop shift register simulation. Top trace demonstrates non-overlapping clocks, second trace is the serial input and subsequent parallel outputs follow.

5.4.3 Case 3: Edge Triggered D Flip-Flop Shift Register

Although Case 2 was found to work adequately, a NOR gate implementation using edge triggered flip-flops was investigated, in search of a more area efficient design. Fig. 5.11 shows that only a single flip-flop is required per register stage, whereas two flip-flops were required in Fig. 5.8, thus yielding an area saving. As will be shown later, another advantage is a dramatic power saving. Fig. 5.12 shows an example layout segment. Although Case 3 has a number of merits over Case 2, the main drawback of this approach was that the edge triggered design did not function with non-overlapping clocks. Consequently, Case 3 was abandoned, but is presented here as an example of a real 'dead-end' encountered, that may be useful in other digital applications.

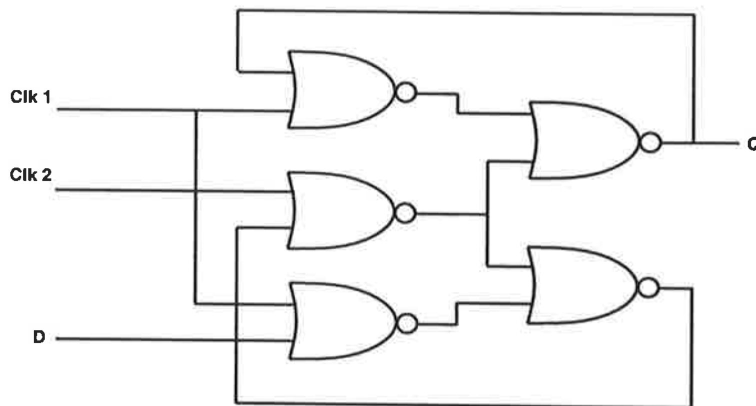


Figure 5.11: Edge triggered D flip-flop shift register stage

5.4.4 Register Comparison

Table 5.1 summarises the performance differences between our three case studies and compares them with three other designs available in the literature – Toshiba [Kikaura 88], OKI [Mori 88] and CMA [Sarmiento 92]. The first point to note is that the max. speed figures for our designs our irrelevant due to the low speed TV rates in our application. Our designs are therefore optimised for power and area, whereas the designs from the literature are clearly optimised for speed. Nevertheless the power per clock frequency (mW/MHz) is favourably low compared to the examples from the literature. Although our Case 1 and Case 3 have been abandoned for the present work, they may have valid application in other contexts. Case 3 has been rejected as it does not operate with non-overlapping clocks. The problem of Case 1 not working, due to gate leakage current, can be solved by

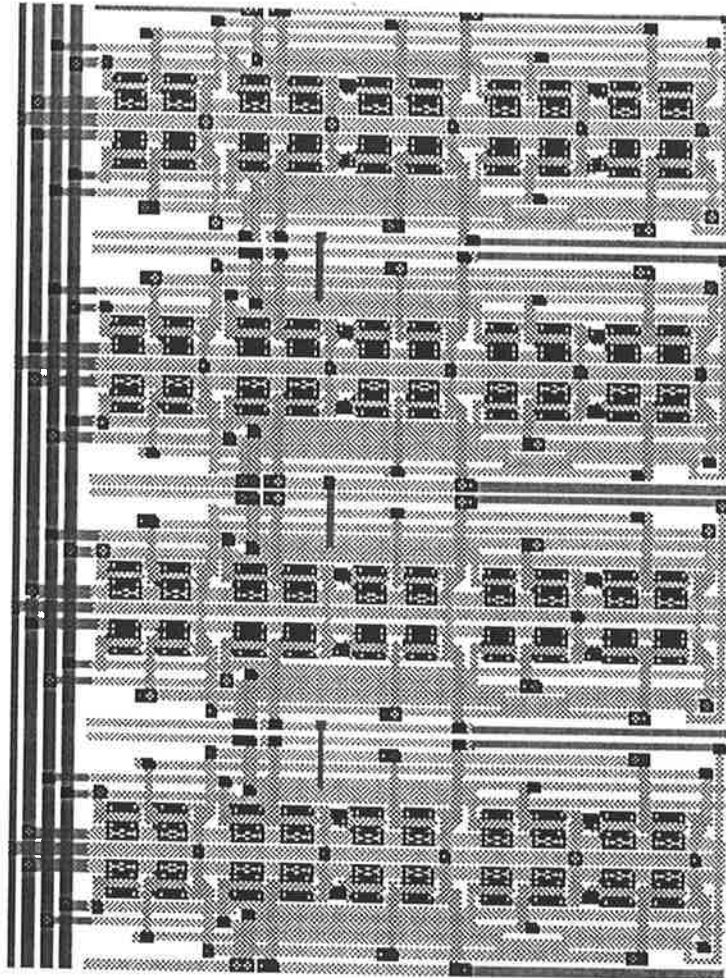


Figure 5.12: Edge triggered D flip-flop shift register layout segment

	Case 1	Case 2	Case 3	Toshiba	OKI	CMA
Max. speed (GHz)	-	0.5	0.8	2.4	1.2	1.2
Area/stage (μm^2)	2200	14450	8360	80294	20557	11000
No. of pwr supplies	0 (!)	1	1	2	1	1
Logic type	NRBL	DCFL	DCFL	SLCF	DCFL /SBFL	DCFL
mW/MHz	small	0.003	0.0009	0.01	0.004	0.003
Transistors/stage	4	24	15	?	12	20

Table 5.1: Performance comparison of shift registers. Notice Case 1 has no power supplies as it is powered via the clocks.

a move to the GaAs HIGFET technology. In the present context, Case 2 is recommended as being functional and of adequate performance for a proof-of-concept imager. Future work, however, should address HIGFET technology as it becomes increasingly available.

5.5 Column-to-video line multiplexer

The multipexer comprises a row of simple transistor switches, that connect the column lines, in turn, to the video line. The switching sequence is controlled by the vertical shift regsiter that is connected in parallel with the multiplexer. In design, the layout of the individual switches is optimised to (a) minimise scan pulse feedthrough onto the video line and (b) minimise the switch on-resistance. These requirements are considered in turn:

(a) Feedthrough

Feedthrough results from gate charge (Q_g) being drawn from and dumped on the video line as the multiplexer transistors turn on and off. Variations in this charge, ΔQ_g, give rise to fpn and this can be expressed as,

$$\Delta Q_g = \Delta C_g(V_{clk} - V_t - V_v) + C_g(\Delta V_{clk} - \Delta V_t).$$

Thus, fpn can be reduced by minimising the clock amplitude V_{clk} and maximising the video line bias V_v in operation – where V_t is the threshold voltage. Minimum sized transistors can be used to minimise the gate capacitance C_g.

(b) On-resistance

The transistor switches have to be minimum sized to satisfy requirement (a). Consequently, this results in an on-resistance of $R = 2$ kΩ. The limit that the resultant RC_v time constant imposes on the signal read out time (t_o) must be calculated. The differential equation that describes the equivalent circuit in Fig. 5.13 is,

$$\frac{d^2V}{dt^2} + \{\frac{1}{C_vR} + \frac{1}{C_cR_i} - \frac{1}{C_cR}\}\frac{dV}{dt} - \frac{V}{C_vC_cRR_i} = 0.$$

With appropriate boundary conditions, a solution can be approximated [Kioke 80] by putting $C_v \gg C_c$ giving,

$$t_o = RC_v \ln\{(1-\eta)(1-\frac{RC_c}{R_iC_v})\}$$

where η is the fraction of signal charge read out in t_o seconds. Taking C_c =5 pF and R_i =500 Ω, η = 99% and C_v=50 pF as severest estimates gives t_o=100 ns. This implies that all column lines can be favourably read out in less than a standard line time.

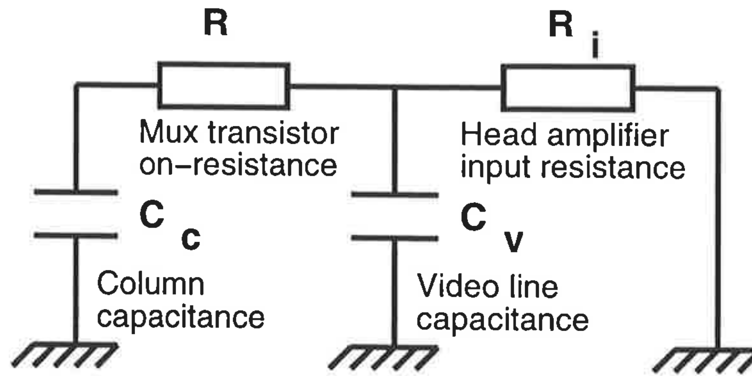


Figure 5.13: Column-to-video line switching: equivalent circuit

5.6 Summary

In this chapter we have defined a suitable imager array architecture. A *passive* pixel approach has been analysed showing that the internal photodiode RC time constants (even on semi-insulating GaAs) are fortunately fast enough for operation at standard TV rates. Two possible pixel layouts were proposed: (1) simply the diode formed by a transistors' source-to-substrate junction and (2) as before but with 'fingers' of gate contact across the surface of the diode – the fingers serve as antiblooming lines but at the expense of pixel fill-factor. The realisation that method (1) can operate on an SI substrate is an original contribution and so is the structure of (2) which conceptually reduces to a source-substrate diode in parallel with a gate diode. Either of these two pixel designs can be simply stepped into a 128 by 128 array for realisation of a working imager.

A number of approaches for array addressing and readout were discussed and rejected. Because of problems with gate leakage currents and the system requirement for non-overlapping clocks, the D flip-flop shift register design is recommended. Over an order of magnitude saving in power is demonstrated by an edge-triggered shift register design – although shown unsuitable for the imager, this result will be useful for future digital processingf blocks.

An new GaAs logic family which we call NRBL was proposed, for the first time, but this suffered from MESFET gate leakage currents. Future use in HIGFET circuits is highly recommended.

Finally multiplexer switches, for column-to-video line parallel/serial readout, were analysed. It was deemed that single minimum size transistors gave the lowest feedthrough

noise and were sufficiently fast for TV rates.

Having created a tailored design for the imager, in the next chapter we turn to examining a GaAs design methodology for general digital circuits – for future integration of digital processing elements with the imager.

Bibliography

[George 89] P. George *et al.*, "Modelling the substrate depletion region for GaAs FETs fabricated on semi-insulating substrates," *Solid-State Electronics,* **vol. 32,** no. 2, pp. 165-168, 1989.

[Kikaura 88] Y. Kikaura *et al,* "A 2 Gb/s GaAs 128-bit shift register using standard cells with 0.5 um WNx gate MESFETs," *IEEE GaAs IC Symposium,* Nashville, Nov. 6-9, 1988, pp. 127-130.

[Kioke 80] N. Kioke, "MOS area sensor: Part I," *IEEE Trans., ED,* **vol. ED-27**, no. 8, 1980.

[Mori 88] H. Mori *et al,* "A 1.2 GHz high-speed 256-bit shift register LSI," *IEEE GaAs IC Sytmposium,* Nashville, Nov. 6-9, 1988, pp. 135-138.

[Ohba 80] S. Ohba *et al*, "MOS area sensor: Part II," *IEEE Trans., ED,* **vol. ED-27,** no. 8, 1980.

[Sarmiento 92] R. Sarmiento *et al*, "High speed primitives of hardware accelerators for DSP in GaAs technology," *IEE Proc. G*, **vol. 139,** no. 2, 1992, pp. 205-216.

Chapter 6

Design Methodology

"There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things."
Niccolò Machiavelli (1469-1527)
ITALIAN WRITER AND STATESMAN

6.1 Introduction

Advances in the development of digital GaAs integrated circuits have progressed to the point that designers of signal and data processors can discern the system applications for which GaAs is best suited. Basic computation primitives in DSP and image processing systems are usually adders, multipliers and delay elements. In this chapter we present the results of a systematic study conducted to evaluate the influence of layout and design methodologies, both conventional and innovative ones, on the performance of digital computation primitives.

There is increasing demand for higher speed and lower power consumption in digital signal and image processors, due to very large throughput rates and vast amounts of data necessary for real time performance. It would appear that much could be gained by implementing both the image and signal processing algorithms with high speed digital gallium arsenide.

The choice of architecture, and indeed the choice of algorithm, is crucial to the performance of the signal processing systems [Gilbert 87]. The architecture is influenced by the well known design criteria for VLSI and in the case of gallium arsenide these criteria are more stringent. The designs for arithmetic and logic units in signal and data processors have to be reexamined in view of the strengths and constraints of GaAs devices.

CMOS-VLSI implementations of fixed precision and floating point arithmetic usually have employed as many logic gates as necessary to construct the function, because there is not major cost penalty for the use of large number of gates; however there are constraints imposed on the speed of this gates. Conversely, GaAs gates are much faster but more expensive at the current levels of complexity.

These tradeoffs favour single and well structured arithmetic implementations. This simply implies that the design must be highly regular with simple data and control paths, and local or minimum interconnections. It may also be possible to recast some arithmetic algorithms, which are recursive in nature, into iterative implementations, re-using gates again and again to execute a more simple arithmetic function.

Pipelining of arithmetic structures to increase throughput should be exploited whenever possible, since the addition of perhaps only 10-20% additional gates in the form of pipelining registers can enhance throughput for GaAs logic families, which can support the higher clock rates.

The design methodology for each class of logic together with the layout approach becomes critical as clock speed increases. The problem manifests itself in the form of signal skew

where there is a delay between arrival times of a given signal at two different destinations, and crosstalk which manifests as noise on buses.

Since basic computational primitives in DSP and image processing systems are usually adders, multipliers and delay elements, in this segment of research we have explored the influence of layout and design methodology on the performance of these modules for DCFL (Direct Coupled FET Logic) and SDCFL (Source-follower Direct Coupled FET Logic) classes of logic.

6.1.1 Technology

Gallium arsenide circuits built using DCFL logic implement the set of functions {NOR,NOT}, whereas SDCFL implements the augmented set {NOR,NOR-OR,NOT,NOT-OR}. The output from two SDCFL buffer stages (shown in Figs. 6.2b, 6.2d) be connected to produce a wired-OR function.

From a layout point of view, DCFL is the most compact GaAs logic and most suited to high performance VLSI systems. This class of logic uses depletion type and enhancement FETs in much the same manner as they are used in traditional nMOS Silicon technology where the ratio rule, which determine the device size, applies (Fig. 6.3).

One disadvantage, however, is the limited output voltage swing, which is due to limited barrier height of the Schottky diode gate and results in small noise margins (less than 100 mV). DCFL gates have also a weak load drive capability which influence the gate delay as fan-in, and fan-out, are increased. A further point that needs to be taken into consideration is the switching point of a DCFL gate which has a negative temperature coefficient which varies in the range -0.4 to -0.6 mV/C.

In order to overcome the complexities with small noise margins, fan-out limitations, and the ability to compensate for temperature variations, the SDCFL structure, which is simply DCFL gate followed by a source follower (Fig. 6.4), is considered as the basic primitive for realisation of computational subcircuits. SDCFL is fully compatible with DCFL.

Since the threshold voltage increases at the same time as channel conductance (these are two parameters that dictate the temperature behaviour) by "tuning" the process it is possible to compensate the net effect of temperature on drain current and realise SDCFL structures which have good tolerance to temperature variations.

A negative power supply ($-V_{ss}$) for the pulldown of the buffer stage ensures that when the buffer E-MESFET transistor is off the output low level can be made more negative

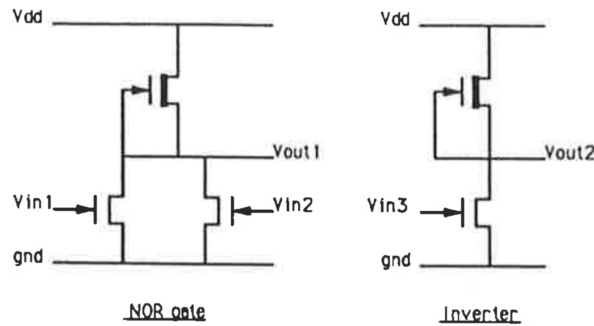


Figure 6.1: DCFL logic gates

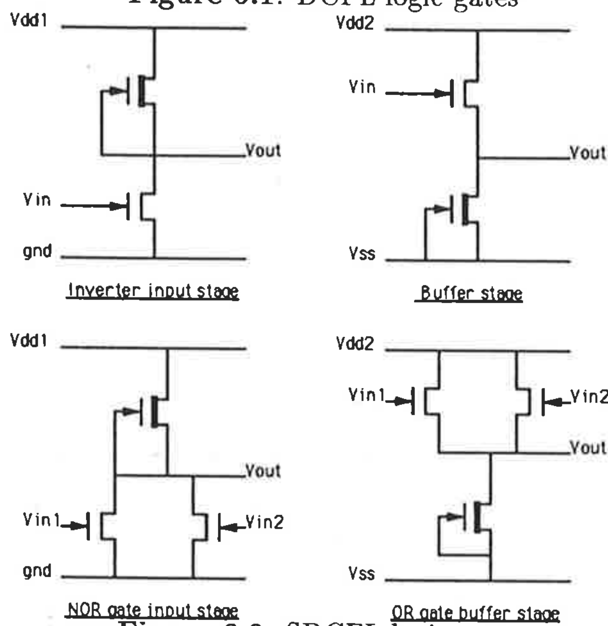


Figure 6.2: SDCFL logic gates

so that the following DCFL logic gate is fully off. Furthermore it provides for isolation of logic GND path from those of the buffer.

6.1.2 Parasitic Oscillations

Parasitic oscillations in GaAs MESFETs are not clearly discussed in the literature and thus we shall briefly describe this effect. It is the early study of this effect that inspired the development of the "Ring Notation" layout style that we shall introduce in the next section.

During early studies, it was found that if a curve tracer was used to plot the I-V char-

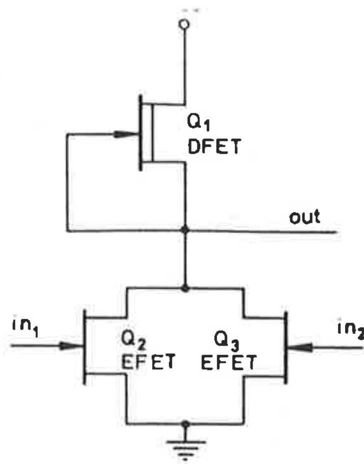


Figure 6.3: DCFL inverter circuit

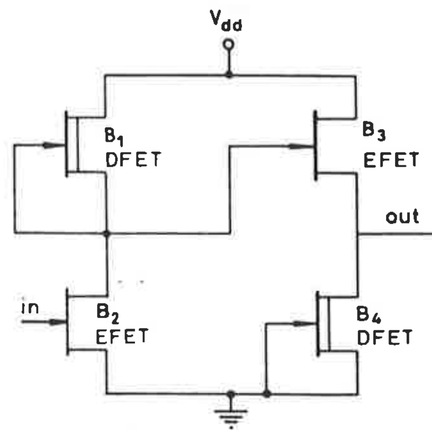


Figure 6.4: SDCFL inverter circuit

acteristics of a GaAs MESFET attached by conventional connectors, disastrous results occurred. An example photograph of the curve tracer screen is shown in Fig. 6.5, where it can be seen that there is a region with negative slope. This indicates negative resistance and hence indicates that oscillations are taking place. This is a good example of what is termed *parasitic oscillations* in GaAs MESFETs. The low nodal capacitances of the MESFET are of the order that a resonant circuit is formed with the parasitic capacitances and inductances of the conventional wire connectors. This in conjunction with the drain-gate capacitance providing feedback and high gain, inevitably produces an unwanted oscillator. This is not a problem particular to gallium arsenide, but is inherent for usually any very high speed device.

Hence, care must be taken to keep connections extremely short, using carefully terminated transmission lines on printed boards. For transistor measurements, this is not always possible if the test jig is, say, on a mechanical scanning device. Sometimes conventional wire connectors cannot be avoided. The solution is to package the transistor with the (grounded) source pin between the drain and gate pins. This serves to reduce the drain-

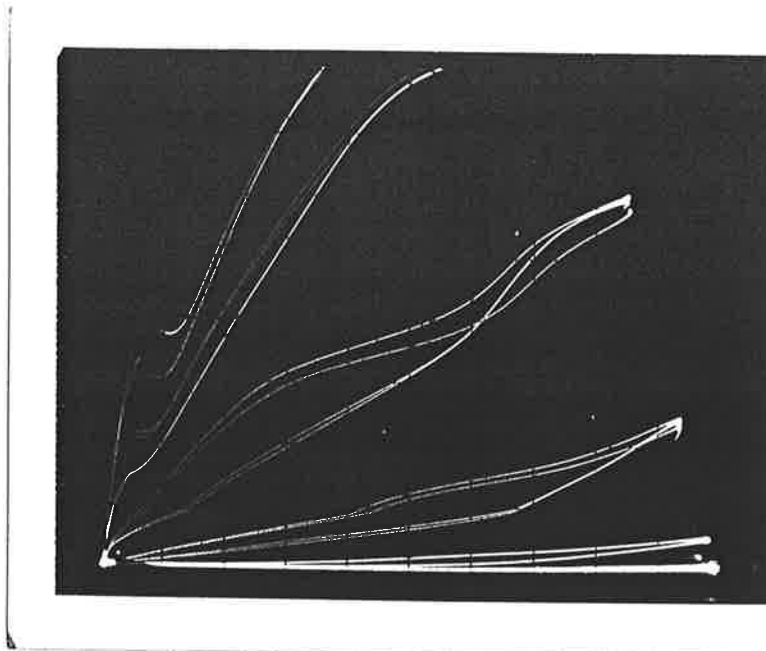


Figure 6.5: Photograph qualitatively showing parasitic oscillations in I-V characteristic. Arbitrary axes. Notice region of negative resistance.

gate feedback capacitance. A chip capacitor placed between drain and source can also be used to ensure any resonant oscillations are decoupled and not fed back to the gate. This is appropriate for dc measurements but for ac measurements the use of a chip capacitor is not possible. In this case the transistor must be continuously monitored by a spectrum analyser for oscillations and the resonance can be detuned by moving the orientation of the connecting wires by trial and error. This is not good practice, but is used when there is no alternative, and is reported eg. [Golio 90], with successful results.

It was the realisation that drain can be screened from the gate by the source connection, giving better results, that inspired the development of "Ring Notation." The key element of Ring Notation is that address busses (gates) are screened from V_{dd} power busses (drains) by ground busses (sources). This has the effect of reducing drain-gate capacitances throughout the whole VLSI layout and desirably increasing the drain-source capacitance (ie. decoupling). As we shall see in the following sections, another advantage is that more compact layouts can be achieved this way – which also results in lower propagation delay.

6.1.3 Layout Approaches

The large bandwidth requirement of high speed signals ($\sim$ 3 GHz) produces the following problems:

- crosstalk from high speed signal lines is increased
- inductive spikes due to current transients cause spurious noise voltages on busses
- reflection from unmatched impedances on long interconnects degrade rise times
- distortions occur due to transmission line effects

These problems can be reduced or minimised through layout considerations.

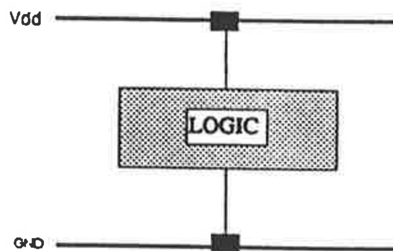


Figure 6.6: Layout styles: nMOS-type stick approach

As the basis of layout, two alternative styles have been used in this study. The first method is the conventional approach used by Silicon designers where $+V_{DD}$ bus is located at the top with the GND bus being located in the bottom of the layout. The circuit is then placed in-between these two rails (Fig. 6.6).

In order to reduce coupling between the fast signal lines and the power bus and furthermore to optimise the interconnections between circuits, a second approach, referred to as Ring Notation [Eshraghian 96], is adopted for comparisons. In "Ring" approach the GND bus is placed between the circuit and the +VDD bus as illustrated in Fig. 6.7(a). A second option for layout which uses "Ring" convention, and which provides for better packing density, is also illustrated (b).

In the "Ring Notation" as shown in Fig. 6.8(a), the dotted line (GREEN) represents the E-MESFET while the solid line (YELLOW) represents the D-MESFET. The approach is very much in-line with the colour coding used to differentiate between different layers in Silicon technology. Another rule which is implicit is that (GREEN) and (YELLOW) are connected only through either Metal 1 or Metal 2 (BLUE). Thus representation of this connection can be removed from layout at this level of abstraction (Fig. 6.8(b)).

In this class of logic, due to degradation of performance, we restrict ourselves to parallel branches in the input path, ie. NOR gates only, then representation of NOR gates can

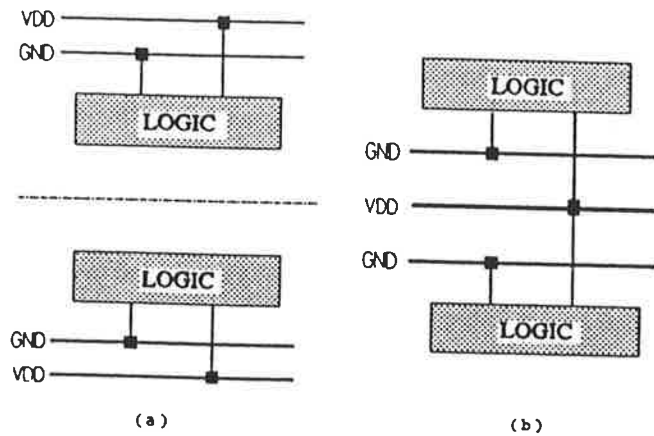


Figure 6.7: Layout styles (a) Ring Notation using dual bus pair, (b) Ring Notation using single bus pair

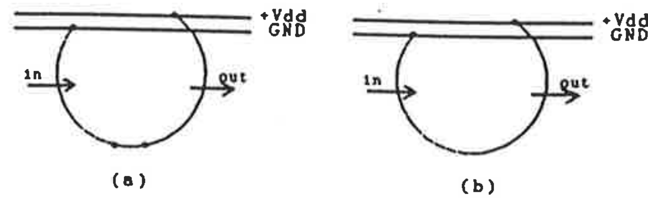


Figure 6.8: Ring Notation showing topology of circuit

further be simplified by eliminating the parallel branches of the input paths. Fig. 6.9 shows the process.

Translation from Ring to symbolic and hence to mask is rather straight forward. What becomes important in the Ring Notation, as an intermediate method of describing a layout, is that signal paths can be highlighted and the interconnect strategy formulated, before translation into a layout commences. Using this methodology complex structures can readily be mapped. One such case is the D-type latch where the structure is realised by reflection of half the logic gate about the x-axis (Fig. 6.10).

Once a design has been drawn in ring notation, interconnect paths may be optimised to create electrical symmetry and so minimise signal skew between critical paths. This is particularly important in the distribution of clock signals across a chip. An electrical equivalence in the form of a data flow diagram shows connectivity between cells or gates. A weight on each link describes its delay characteristic as a function of geometric length, width, adjacency (coupling), fan-in and fan-out. Geometric symmetry does not neces-

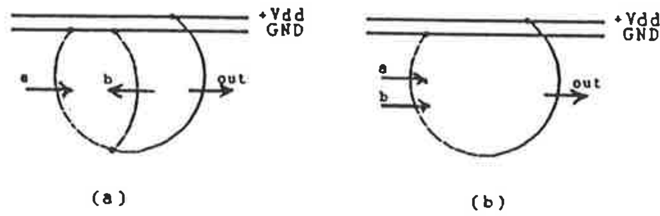
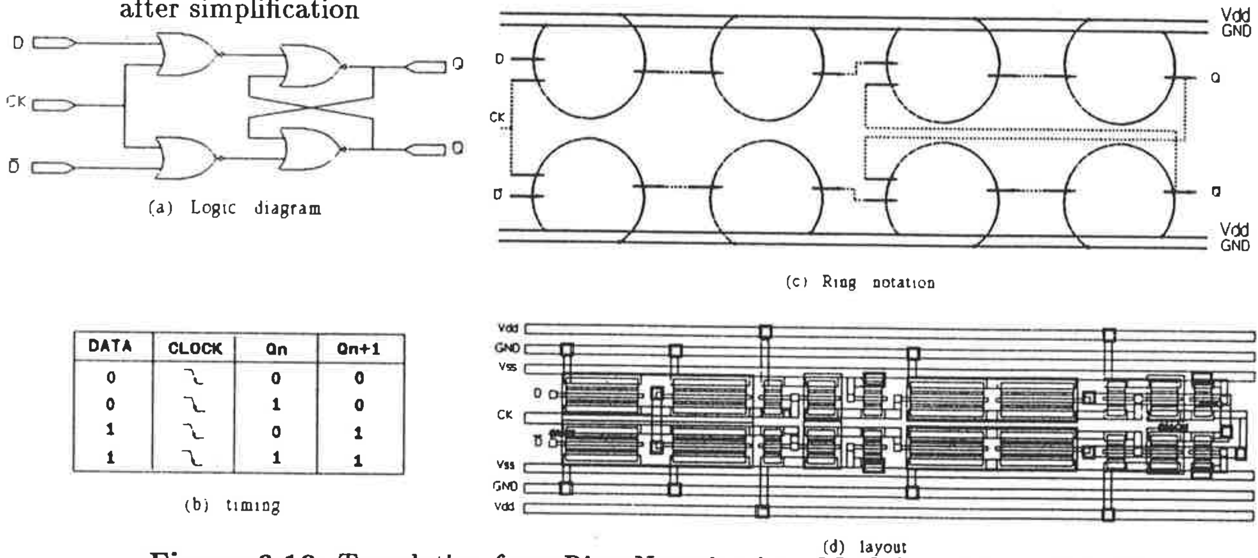


Figure 6.9: Ring Notation for 2-Input NOR gate, (a) initial representation, (b) after simplification



DATA	CLOCK	Qn	Qn+1
0	˥_	0	0
0	˥_	1	0
1	˥_	0	1
1	˥_	1	1

Figure 6.10: Translation from Ring Notation into Mask layout for a D-Latch

sarily imply electrical symmetry but by adjusting the aspect ratio of the layout, some optimisations may be possible.

6.1.3.1 Symbolic mask level layout

The gallium arsenide process layers are very complex, and inhibit the visualisation of all the mask levels used in the actual fabrication process. An approach has been developed in which both layer and topology information may be simply represented to the designer. Mask layers derived from the foundry specification are abstracted to a manageable number of layout levels to hide unnecessary layer complexity. Colour coding and symbolic representation greatly assist in interpreting gallium arsenide circuits, while allowing the VLSI designer to concentrate on more fundamental aspects of UHS VLSI design such as routing and optimisation of interconnection paths. The colour scheme, layer names and CIF codes adopted for gallium arsenide are shown in Table 6.1.

Layer	Colour	Symbolic	CIF Code	Comments
Diffusion *	Green	Implant	GD	active area, outside is substrate
Shallown-	Red	E-transistor	GP	creates the shallown- channel for the enhancement device
nplus	Yellow	n+ implant	GI	outside yellow, n+ implant takes place, for MESFET's only
n+mask	Orange	n+	GJ	inside, n+ implant takes place, for diodes only
Metal1	Blue	Metal1	GM	first level interconnect
Metal2	Purple	Metal2	GN	second level metal interconnect
Contact	Black	Contact	GC	diffusion to metal1 contact
Via1	Grey	Via1	GV	metal1 to metal2 connection
Symcap	Brown	MIM	GY	inside is MIM capacitor
Overglass	Cyan	-	GG	-

Table 6.1: Gallium Arsenide Symbolic Representation and Notations

** This layer is an implant region, but the term diffusion has been used to be consistent with nMOS and CMOS notation that many silicon designers are familiar with.*

Design rules are described on a Lambda-based grid while fine detail (micron-based rules) is embedded within the predefined objects and is hidden from the designer.

6.1.3.2 Practical constraints

The gallium arsenide lattice structure exhibits different conductance in different planes. To ensure a uniform transconductance between devices, it is required that all MESFET gates be laid out in the horizontal direction. Design rules dictate the minimum size and spacing of the geometric features in a layout to ensure a reasonable yield will be obtained. A trade off exists between performance and yield for a given chip size. The trend is towards the use of minimum design rules to maximise speed at the expense of yield. Even with this trend, the integration level of gallium arsenide is generally lower than silicon technologies, typically around 3000 devices per square millimetre as opposed to a 0.8 μm CMOS process with up to 8000 devices per square millimetre. The ratio of pull-up to pull-down device width is determined by maximising the noise margins, whereas the actual width is determined by gate delay.

Logic family implementation constraints which must be met to ensure successful high speed operation of the circuit include fan-in and fan-out. The fan-in of a gate should be

limited to 3, and is preferably 2 since the output high level of the input stage is degraded as fan-in is increased. The fan-out of a gate, which is characterised in terms of the number of driven gate capacitances and the total interconnect capacitance, is limited to 2 and the interconnect lengths which contribute parasitic capacity must be minimised.

Most gallium arsenide processes contain either two or three layer metal interconnect. The upper metal layers have lower parasitic and coupling capacitances due to the dielectric structure and the wire pitch. These layers are therefore preferred for signals where signal skew is to be minimised. Coupling between power supply busses and fast interconnects on any metal layer can be minimised by the use of additional ground wires.

Separate power busses for buffers and logic gates reduce the noise induced by switching transients.

6.1.3.3 CAD tools

The gallium arsenide VLSI design suite provides a complete design environment. It includes graphical layout of full custom chips using a library of predefined objects, design rule checking, network and parameter extraction and provides an interface to both functional and analogue circuit simulators. The design process is shown in Fig. 6.11 [ISD 88]. It consists of the following software tools:

PLAN a graphics editor for VLSI design for which technology files and a predefined objects library have been written for gallium arsenide implementing the design methodology presented in section 3.2.

CHECK a fast corner based design rule checker. The abstracted design rules are derived from the foundry rules and are used to verify circuit topology.

SEE a general purpose geometric display tool which shows a fully instantiated circuit on a variety of plotting devices.

GAASNET a gallium arsenide network extractor and circuit verification tool which extracts devices and circuit parameters from the CIF level description. Emphasis was placed on the accuracy of circuit extraction so that tightly toleranced designs could be realised. An interface is provided to various circuit simulators such as PSPICE [Microsim 90], SPICE [Quarles 89] and functional simulators such as THOR [Standford 86].

Technology independence allows multiple foundries and processes to be targeted.

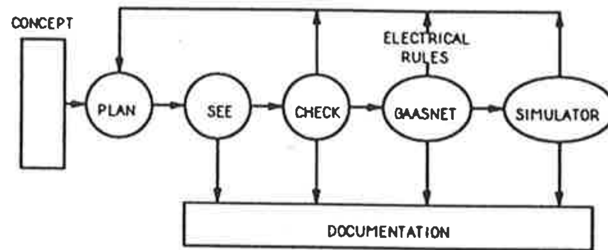


Figure 6.11: Gallium arsenide design process using CAD tools

6.1.3.4 Modelling of gallium arsenide circuits

Accurate modelling of gallium arsenide devices is critical in determining if a circuit will work. In very high speed and ultra high speed systems, logic circuits respond to repetitive signals and pulses that have periods similar to their propagation delay times. The modelling of these circuits may be done in two parts:

- device modelling of MESFET's, resistors and Schottky diodes
- interconnect modelling

Device models have been well characterised. The models employed are the Curtice [Curtice 80] and Statz models [Statz 87] combined with subthreshold models [Chang 87, Conger 88] to model MESFET operation in all regions [Chen 90]. Inaccurate noise margins may be obtained if leakage current in the subthreshold region is not considered as it degrades the output voltage swing. The parameters for the models are obtained by measurement at the frequencies of interest (typically 1 GHz) due to the non-linear frequency behaviour of MESFET's.

Interconnect modelling for gallium arsenide is not well characterised due to the impedance being a non-linear function of:

- distance to ground plane
- overcrossing capacitance of metal of different layers

- coplanar fringing capacitance
- geometry

The series resistance of an interconnect is around 45 mΩ/μm for 2 μm wide first layer metal. Therefore the impedance of an interconnect is mostly made up of capacitance to the ground plane and adjacent interconnects.

The rise time of a signal emerging from a gate is around 50 ps, if we choose an upper bound of 10 GHz for the bandwidth requirement and the velocity of propagation of a signal is 110 μm/ps, the wavelength is then 11 mm. For a transmission line approach to be appropriate, the length of the line needs to be at least a quarter wavelength long, 2.75 mm. The use of transmission line or stripline models is only required for interconnect lengths which approach this figure. Most interconnects are much less than 2.75 mm long. It should also be noted that stripline models are not likely to be appropriate since the substrate thickness to the ground plane ($\sim$ 600 μm) is two orders of magnitude larger than the line width. Short interconnects ($\ll$ 2.75 mm) are modelled simply as a lumped capacitance which represents the total distributed capacitance of an interconnect. The GHz frequencies dictate that some longer signal paths cannot be treated simply as lumped elements but as a transmission line terminated by a capacitance (mostly the gate-source capacitance of an enhancement MESFET.) We can model long interconnects as unterminated transmission lines which have a reflection coefficient of unity at the end of the line. The length of the line must be constrained by:

$$length \ll \frac{\tau_{rise}}{2 \times \tau_{pd}}$$

where τ_{rise} is the rise time of a signal and τ_{pd} is the propagation delay per unit length along the line. The velocity (V_p) of a signal along an interconnect is 110 μm/ps. If we include a factor of five ($\tau_{rise} = 50$ ps, $\tau_{pd} = 1/V_p = 1/110$ μm/ps), we have a rule that the length of an interconnect is less than 550 μm.

The output impedance of a gate is high compared to the characteristic impedance of the interconnect [Long 90] (modelled as a transmission line). Therefore a returning signal will not cause ringing as the reflection coefficient at the source for this signal is much less than 0.5. This justifies the use of a lumped capacitance model.

6.1.3.5 Interface to fabrication

Interface to the gallium arsenide foundry is through a mask level conversion program which produces foundry mask layers from the simplified designers layout.

The designer mask layers are a subset of the foundry mask layers which may be translated at the CIF level by the following operations:

- direct mapping
- bloat factor
- mask inversion
- logical operation of several mask layers

An example is the isolation damage layer between active regions which electrically isolate devices by damaging the surrounding semi-insulating substrate. This layer is derived from the inverted diffusion layer with a 'bloat factor' to account for lateral diffusion of the implant during processing.

6.1.4 Performance Study

Using the approaches shown in Fig. 6.6 and Fig. 6.7, a set of 8 bit and 16 bit adders and 8 bit multipliers were designed and their performance compared. Although several options were available to compare the performance of the different circuits and layouts, the approach that was adopted for the basis of our initial comparison defined a Figure of Merit or a performance indicator (P_{ind}) in terms of number of gates N, maximum clocking frequency of the circuit, and utilisation area, given by:

$$P_{ind} = \frac{N_{gates} * f_{max}}{Area}$$

where N_{gates} is the number of gates and f_{max} is the maximum clock frequency. This gives us a measure of computational density as gate-GHz/mm^2.

6.1.4.1 Adders

GaAs/Silicon differences (transistor count per chip as well as gate fanin and fanout) affect the choice of adders in a number of ways. Traditionally, the full carry look-ahead adder has been preferred for high speed applications, and ripple -carry adders have been discounted for all but the shortest word length.

Further evaluation of the adders show that the fastest adder in Silicon is not necessarily the fastest in GaAs technology (because of the fastest Si adder is based on high fanin and fanout capability, and GaAs technology is intolerant to high fanin and fanout).

Several adder architectures are available for GaAs technologies, ranging from high-speed, large area Binary-Carry-Lookahead adder to the low-speed, small area ripple carry adder. Other architectures having speed and resource requirements between these two extremes include Carry Look-Ahead and Carry-Select adders. Several versions of each architecture have been designed. In turn, each logic alternative has been implemented with different layout styles (Figs. 6.6 and 6.7). The aim is to study the different architectures as well as the layout techniques best suited for GaAs logic families.

Ripple Carry Adders

Table 6.2 summarises HSPICE simulations results of several ripple carry adders. The ripple carry adders have been designed with exclusive OR gates, taking advantage of the NOR-OR structure easily implementable in SDCFL. The adders have been optimised for low power and area consumption (VLSI applications). The results show that the Ring Notation methodology (Fig. 6.12(b)) produces circuits with higher speed that the nMOS type stick methodology (Fig. 6.12(a)), mainly because of the shortest interconnection length of the path. The Ring layout using single bus pair produces circuits with less area, but slightly slower than Ring layout using dual bus pair.

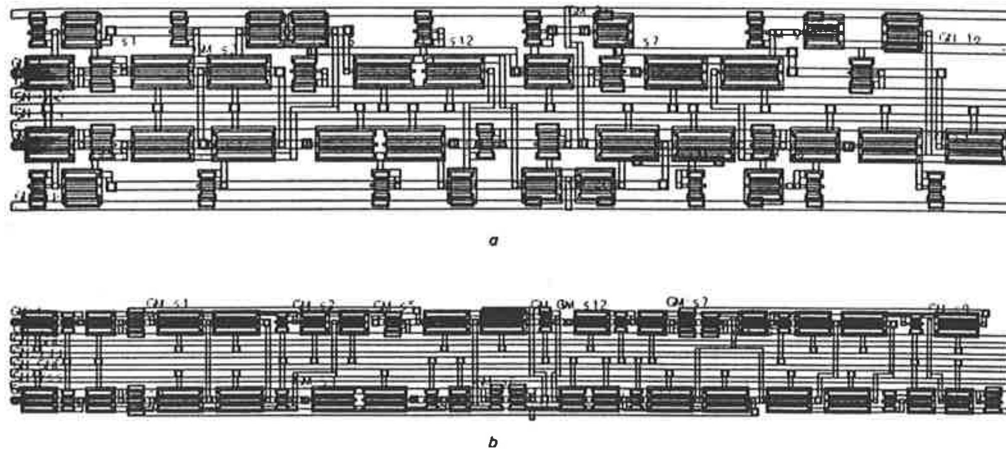


Figure 6.12: 1-bit Ripple Carry layout (a) nMOS type stick and (b) Ring Notation

From designers point of view, Ring Notation methodology is easier to implement than the nMOS type stick approach, allowing easy approach to structured design. Furthermore, it is well suited for symbolic layout.

Carry Look-ahead Adders

The limited fanin and fanout of GaAs gates affects both the area that is needed and the adder delay incurred. Single gates with high fanin and fanout must be implemented as a series of gates with low fanin and fanout, which increases transistor count. Because delay is highly dependent on load capacitance, high fanout devices have a relatively large delay.

Layout Style	No. of bits	Delay (ns)	Power (mW)	Area (mm^2)	P_{ind}
'nMOS'	8	3.8	83	0.133	269
'nMOS'	16	7.7	169	0.264	133
'Ring' dual bus	8	3.6	83	0.175	216
'Ring' dual bus	16	7.4	169	0.389	95
'Ring' single bus	8	3.7	83	0.150	245
'Ring' single bus	17	7.5	169	0.299	121

Table 6.2: Parameters of several ripple carry adders.

Layout Style	No. of bits	Delay (ns)	Power (mW)	Area (mm^2)	P_{ind}
'Ring' dual bus	9	2.9	80	0.211	337
'Ring' dual bus	16	4.9	140	0.368	204
'Ring' single bus	9	2.5	80	0.210	392
'Ring' single bus	16	4.7	120	0.363	302

Table 6.3: Parameters of the carry look-ahead adders.

In a tree built in a random fashion this dependence may cause the delay through N levels of high fanout gates to exceed the delay through $N+1$ levels of low fanout gates.

The carry expression in a Carry Look-Ahead adder is:

$$C_i = G_i + P_iG_{i-1} + P_iP_{i-1}G_{i-2} + ... + P_iP_{i-1}.....P_oC_o$$

Where $G_i = X_iY_{iy}P_i = X_i + Y_i$. The GaAs implementation of this expression bring about a critical path with no less gates than a Ripple Carry adder. A suitable implementation using GaAs is to incorporate $n+1$ cell serially. The carry expression will be divided in two terms: Ca_i and Cb_i. The first one will be fan-out loads between the inputs X_i and Y_i and their complements NX_i and NY_i each cell does the terms generation in an alternating way with either input type. Fig. 6.13 shows the floorplan for a 5 bits adder.

Table 6.3 shows the results of the GaAs implementation of the carry look-ahead expressions. In this case the single bus layout is faster than the double bus pair since the inter-cell communications are dominant in the critical path. A broad comparison between this adder implementation and the Ripple Carry one, as shown in Fig. 6.14, illustrates that the former is best suited for GaAs. The Ripple Carry adder implemented with only two-input gates remains too slow, as in the Silicon implementation, and with high are

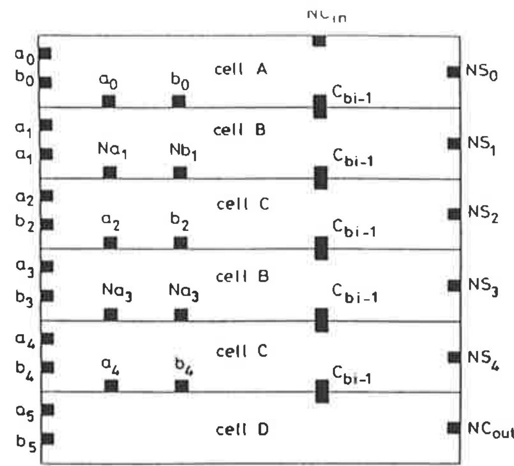


Figure 6.13: Floorplan of the GaAs implementation of carry look-ahead expressions

factor.

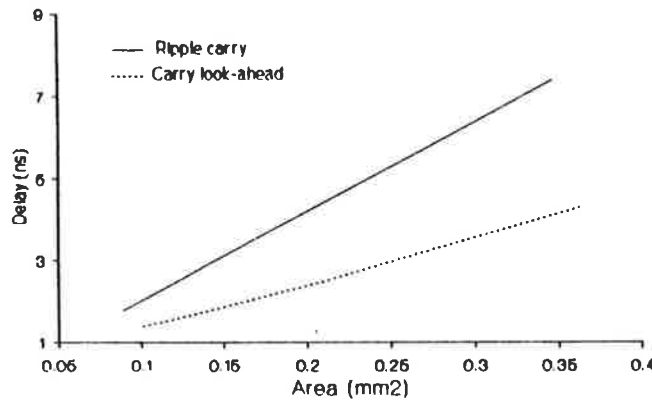


Figure 6.14: Area vs. delay of Ripple Carry and Carry Look-Ahead adders

Binary Carry Look-ahead Adders

The Binary Carry Look-ahead, sometimes called Brent & Kung adder, is based on the carry generation (G_i) and carry propagation (P_i) functions [Brent 83]. The main advantages of this adder implementation when used in GaAs technology is that it uses only two-point gates and the fan-out is extremely low. Fig. 6.15 represents the block diagram of an 8-bit Brent & Kung adder. The gp cells generate the G_i and P_i functions and the out cells provides the add operation. The carry generation block is a tree with two cell types: the processors (proca and procb)and the inverter-buffer cell (intera and interb).

Table 6.3 shows the results of the HSPICE simulations for the adder. As expected the double bus pair is slightly faster than single bus pair realisations, although with higher area utilisation factor.

The difference in transistor count limits the complexity of any adder that is going to be implemented as part of a processor. Fig. 6.16 shows the relationship between the number

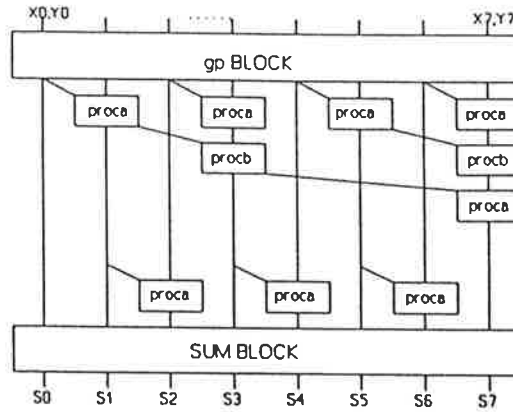


Figure 6.15: Block diagram of Binary carry look-ahead adders

Layout Style	No. of bits	Delay (ns)	Power (mW)	Area (mm^2)	P_{ind}
'Ring' dual bus	8	1.3	92	0.466	510
'Ring' dual bus	16	1.7	218	1.142	382
'Ring' single bus	8	1.4	92	0.434	507
'Ring' single bus	16	1.8	218	1.078	381

Table 6.4: Parameters of the binary carry look-ahead adders.

of bits and the B factor, where

$$B = \frac{no.bits}{(delay * area)}$$

Although the Brent & Kung adder is the fastest of the approaches pursued, the area may be too high for GaAs VLSI implementations. Therefore it is necessary to further optimise the Brent & Kung layout (ie. not using a cell concatenation).

6.1.4.2 Multipliers

Similar tradeoffs occur in the design of multipliers and other arithmetic and DSP subsystems. A variety of architectures are suitable for GaAs designs. The speed advantage of GaAs gives fast multipliers using straight forward parallel architectures. It is possible to implement inherently faster architectures.

On the other hand, bit-serial architectures present two keys advantages: Firstly, bit serial networks are easily routed on-chip without the problems of bit-serial busing. Secondly, all signals entering and leaving the chip do so via single pins.

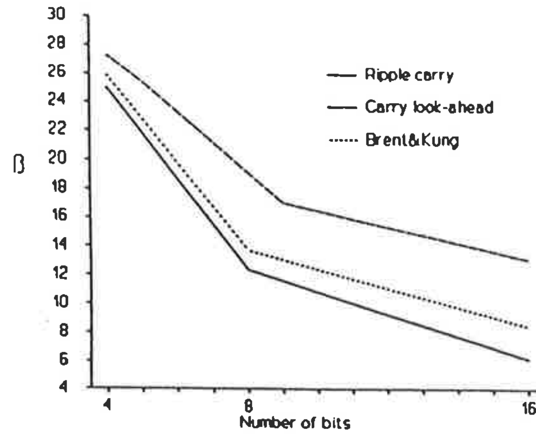


Figure 6.16: Comparison of different adders

Layout Style	Delay (ns)	Power (mW)	Area (mm^2)	P_{ind}
Booth/RC	7.4	255	1.542	95
Booth/CLA	6.4	252	1.550	116
Booth/B&K	5.4	2.67	1.820	126
Baugh/Wooley	5.3	279	1.970	111

Table 6.5: Parameters of several adders.

We have implemented two kind of parallel multiplier architectures: Booth and Baugh and Wooley; a serial multiplier and a serial/parallel multiplier.

Parallel multipliers

Table 6.5 shows the results of the 8*8 bit parallel multipliers designed. Booth multiplier have been designed with different adder types for the final sum stage. Using a Brent and Kung adder it can multiply two 8-bit operated in 5.4 ns. Then, use of fast adders to form the final sum eliminates many delays.

A comparison of this realisation with the GaAs Booth multiplier implementation in [Singh 90] shows the latter has a delay of 3.2 ns, with a power consumption of 510 mW and area occupation of 2.18 mm^2. The computed P_{ind} for it is 101 gate-GHz/mm^2. Our design demands less than a half power and area. The Baugh-Wooley algorithm is for multiplication of signed operands. It is inherently faster and power hungry than Booth algorithm. Fig. 6.17 represent the computational density of the parallel multipliers designed. The data obtained from reference [Singh 90] has been added for comparison purposes.

Serial and serial/parallel multiplier

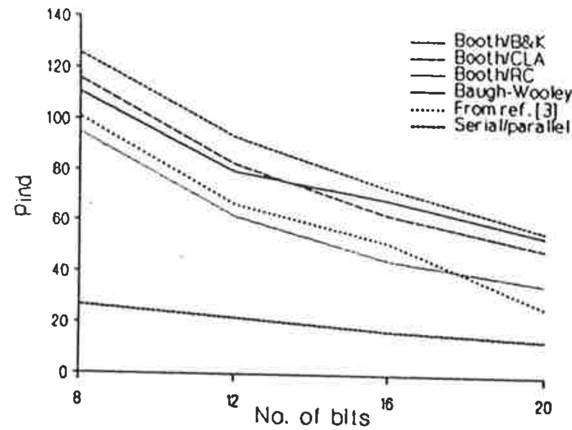


Figure 6.17: Computational density of parallel multipliers

On a broad comparison between bit-serial and bit-parallel operators assuming a common clock rate, we expect the serial part to process one word every N clock ticks compared with one clock tick in the parallel part, but to contain only 1/Nth of the hardware (one full adder cell opposed to NH in the case of an adder). Given a common clock rate and technology base for both implementations we would expect them to be broadly equivalent in terms of an area-time measure. The condition of a common clock rate may not hold however, since tight bit-level pipelining of the serial system should permit a shorter clock period than for the (non pipelined) parallel system, whose clock period must reflect an allowance for carry propagation in the multiply and add elements. This would be reflected in a reduction of the chip devoted to the arithmetic processes in the bit serial part.

However, the GaAs implementation of serial multipliers have a significant problem the shift register must be implemented with static logic, with an area consumption by bit similar to a full adder. Whereas a parallel multiplier needs about N*N full adders, a serial multiplier needs two N-bit shift registers (parallel-in serial-out), a full adder with latch and a 2N shift register (serial-in parallel-out). This means about 3N+2 equivalent full adders.

The serial multiplier designed runs at 1 GHz clock rate, producing a delay of 64 ns, with a power consumption lesser than 100 mW. In order to avoid glitches in the circuit the shift register with parallel inputs switching at 1 GHz has been designed with electrically equivalent path as shown in Fig. 6.18.

In order to decrease the processing time a serial/parallel multiplier could be used. The area consumption of such an 8x8-bit multiplier is 0.92 mm^2 and the delay incurred is 16 ns.

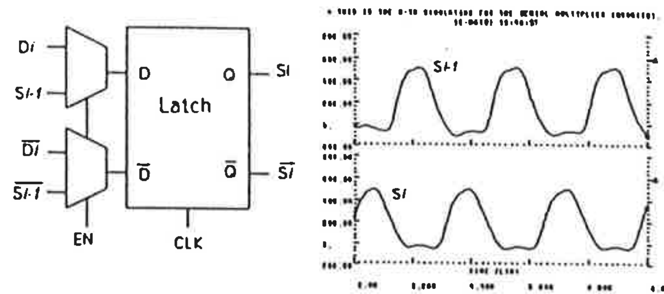


Figure 6.18: Shift register bit-slice (a) diagram, (b) simulation

6.2 Summary

A detailed discussion of parasitic oscillations is lacking in the literature. A careful discussion of the cause of this effect, leading to the realisation that optimal bus layout is achieved when ground busses are interposed between address and power busses was presented as an important original contribution. It is this key paradigm-shift from the old nMOS bus style, that lead to the systemised "Ring Notation" developed by Eshraghian.

The early work described in this chapter can now be greatly improved upon, as new GaAs processes have recently emerged such as CGaAs™ and HGaAs IV, promising greater speeds and much lower power dissipation. Also since this work was carried out, a 'modified Ring Notation' approach has been developed that uses vertically stacked transistors with shared source/drains [Cui 95]. This produces upto a factor of 2 saving in area. Nevertheless, the early work, presented here, was a stepping stone which established that the Ring Notation layout methodology was capable of saving area and reducing propagation delay.

In the next chapter we now proceed to apply this methodology in a case study design of a polynomial evaluator. This evaluator is an example of a useful circuit that can be integrated with the image sensor to achieve our aim for a GaAs smart sensor with high-speed processing.

Bibliography

[Brent 83] R.P. Brent and H.T. Kung, "A Regular Layout for Parallel Adders," *IEEE Trans. Computers,* **vol. 16**, no. 12, pp. 41-58, 1983.

[Chen 90] Y.D. Chen and A. Beaumont-Smith, "Evaluation of gallium arsenide subthreshold MESFET models," *Technical Report GAAS-90-4*, The University of Adelaide, November 1990.

[Conger 88] J. Conger, A. Peczalski, and M.S. Shur, "Subthreshold Current in GaAs MESFET's," *IEEE Electron Device Lett.*, **vol. 9**, no. 3, Mar. 1988.

[Cui 95] S. Cui *et al,* "A 32-bit GaAs IEEE floating point multiplier using trailing-1s rounding algoritm," *Proc. IEEE Elec. Tech. Directions 2000,* Adelaide, Australia, pp. 246-252, May 23-25, 1995.

[Curtice 80] W.R. Curtice, "A MESFET Model for use in the Design of GaAs Integrated Circuits," *IEEE Transactions on Microwave Theory and Techniques*, **vol. MTT-28**, pp 448-456, 1980.

[Eshraghian 96] K. Eshraghian, "Fundamentals of Ultra High Speed Systems: Gallium Arsenide VLSI Technology," Prentice-Hall, to be published 1996.

[Gilbert 87] B.K. Gilbert and G.W. Pan, "The application of Gallium Arsenide Integrated Circuit Technology to the Design and Fabrication of Future Generation Digital Signal Processors: Problems and Promises," *Proc. IEEE*, **vol. 76**, no. 7, pp. 816-834, 1987.

[Golio 90] J.M. Golio, "Frequency-dependent electrical charateristics of GaAs MESFETs," *IEEE Trans. ED,* **vol. 37,** no. 5, p. 1217ff, 1990.

[ISD 88] Integrated Silicon Design, "Phase 1 VLSI design suite user manual," version 4.2 edition, June 1988.

[Long 90] S.I. Long and S.E. Butner, "Gallium Arsenide Digital Integrated Circuit Design," McGraw-Hill, 1990.

[Microsim 90] MicroSim Corporation, "PSPICE reference manual," 1990.

[Quarles 89] T. Quarles, "SPICE3 user's manual," 1989.

[Rouger 85] J-M. Rouger, R. Perichon, S. Mottet and J.R. Forrest, "Etude du comportment du transistor a effect de champ au GaAs sous injection

optique quasi ponctuelle," *Ann. Telecommun.*, **40**, no. 3-4, pp. 88-97, 1985 (in French).

[Singh 90] H.P. Singh, R.A. Burrier and R.A. Sadler, "A 6.5-ns GaAs 20*20-b Parallel Multiplier with 67-ps Gate Delay," *IEEE J. Solid-State Circuits,* **vol. 25,** no. 5, pp. 1226-1231, 1990.

[Standford 86] VLSI-CAD group, "THOR tutorial," Stanford University, 1986.

[Statz 87] H. Statz, P. Newman, I.W. Smith, R.A. Pucel, and H.A. Haus, "GaAs FET Device and Circuit Simulation in SPICE," *IEEE Transactions on Electron Devices*, **ED-34**, pp. 160–169, 1987.

[Chang 87] C.T.M. Chang, T. Vrotsos, M.T. Frizzell, and R. Carrol, "A Subthreshold Current Model for GaAs MESFETs," *IEEE Electron Device Lett.*, **vol. EDL-8**, pp. 69-72, Feb. 1987.

Chapter 7

Visions & Applications

"Where there is no vision, the people perish."
Proverbs 29:18 (KJV)

"Imagination is more important than knowledge."
Albert Einstein (1879-1955)
GERMAN-AMERICAN PHYSICIST

"[Worse than blindness is]...sight without vision."
Helen Keller (1880-1968)
BLIND AND DEAF EDUCATOR/AUTHOR

"The statement of the cause is incomplete, unless in some shape or other we introduce all the conditions."
John Stuart Mill (1806-1873)
BRITISH UTILITARIANIST PHILOSOPHER

"When I point don't look at my finger! Look at where I point!"
Warren Sturgis McCulloch (1898-1971)
AMERICAN PILLAR OF CYBERNETICS

"Well informed people know it is impossible to transmit the voice over wires, and that were it possible to do so, the thing would be of no practical value."
Editorial, Boston Post (1865)

7.1 Introduction

In this chapter we now bring together the work of the previous chapters, by overviewing 4 key application areas for the GaAs imager concept developed. The four application paradigms discussed are:

- *Aerospace* – the radiation hard properties of the image sensor make it ideal for aerospace
- *Mobile Multimedia* – the ability to integrate low-power digital, optical and RF functions in GaAs makes the image sensor ideal for portable wireless video applications
- *Insect Vision* – the edge-effect discovery lends itself to a GaAs motion detector, based on insect vision principles, where coarsely thresholded data is merely required
- *Spatial Light Modulator* – a number of favourable optical and mechanical properties of GaAs make the image sensor highly suitable as the front end of an optically addressable spatial light modulator (OASLM).

Notice that these are all direct applications of the image sensor, except for the 'insect vision' example which is more about utilising the edge-effect discovery.

7.2 Case 1: Aerospace paradigm

7.2.1 Image Processing

Before we go on to utilise the design methodology, developed in the last chapter, to realise an example high speed digital image post-processing function for the imager, we must firstly identify image processing tasks that could be usefully integrated with the imager. Three important areas are,

- Image registration
- Image enhancement
- Image classification.

Image classification areas, such as texture classification, image identification etc. are highly useful in aerospace imaging systems, for example. However, these are high level functions that are beyond the scope of this study. We shall limit discussion to the low level functions.

7.2.1.1 Image registration

Image registration is the aligning of two images of the same scene from different perspectives. The change in perspective may be caused, for example, by movement of an aerospace vehicle or by virtue of two imaging systems mounted on different positions on the same craft. One image sensor may be infrared and the other may operate in the visible band, and so the two images may have different pixel resolutions – this is also regarded as an image registration problem.

Image registration is performed via a process of *image warping*, which simply means transforming an image from one coordinate system to another. A traditional example of warping is the cartographer's transformation of a globe map into a flat coordinate system. This class of processing can be performed by a polynomial evaluator, though, does not need to be performed on board the craft in real-time.

However, for smart aerospace vehicles of the future, where real-time registration would be required, for example, during the link-up between two vehicles, a high speed GaAs polynomial evaluator would be desirable. Furthermore, for irregular shaped objects, such as other aerospace craft, the same polynomial evaluator can be used to find the centre of mass of the object, as an origin or reference point for registration.

In section 7.2.2 we will proceed to design a polynomial evaluator architecture as a case study of a digital element that could be possibly integrated with the imager. It will also illustrate the design methodology principles developed in the last chapter.

7.2.1.2 Image enhancement

In practice, many enhancement operations such as contrast enhancement, fixed-pattern-noise correction etc. are carried out at the video pre-processing stage. An important image post-processing enhancement technique is edge detection, which we will survey in some detail.

Definition An edge in an image is a boundary or contour at which a significant change occurs in some physical aspect of an image, such as the surface reflectance, illumination, or the distances of the visible surfaces from the viewer. Changes in physical aspects manifest themselves in a variety of ways, including changes in intensity, colour, and texture. In most cases, intensity changes are considered.

Edge Detection Algorithms

Detecting edges is very useful in a number of contexts. For example, in a typical image

understanding task such as object identification, an essential step is to segment an image into different regions corresponding to different objects in the scene. Edge detection is often the first step in image segmentation. There are two basic approaches to image edge detection: the enhancement/thresholding method and the edge fitting method. In the former, discontinuities in an image attribute are enhanced or accentuated by some spatial operators. If the enhanced discontinuity is sufficiently large, or greater than some threshold level, an edge is deemed present. The edge fitting approach involves fitting of an ideal edge replica, a two-dimensional ramp or step function, to the image over some region. If the fit is close, an edge is judged present.

Enhancement/thresholding edge detectors

The enhancement/thresholding edge detection method is described in Fig. 7.1.

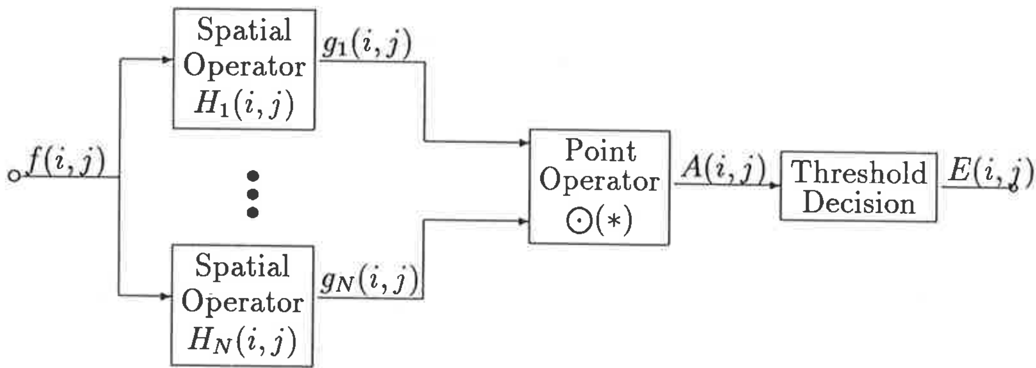


Figure 7.1: Edge enhancement/thresholding edge detection system

In this method, the discrete image array $f(i,j)$ is spatially processed by a set of N linear operators of masks $H_k(i,j)$ to produce a set of gradient functions.

$$g_k(i,j) = f(i,j) \bigotimes H_k(i,j)$$

where $\otimes$ denotes two-dimensional spatial convolution. Next, at each pixel, the gradient functions are combined by a linear or nonlinear point operator $\odot$ to create an edge enhanced array

$$A(i,j) = \bigodot[g_k(i,j)]$$

Typical forms of the point operator include the root mean square (rms), magnitude, and maximum. The enhanced array $A(i,j)$ provides a measure of the edge discontinuity at

the center of the gradient mask. An edge decision is formed on the basis of the amplitude of $A(i,j)$ with the respect to a threshold (t). If

$$A(i,j) \geq t$$

an edge is assumed present, and if

$$A(i,j) \leq t$$

no edge is indicated. The edge decision is usually recorded as a binary edge map $E(i,j)$ where a one value indicates an edge and a zero value, no edge. There are two types of spatial edge enhancement operators: the differential and the template matching operators.

A. Differential Operators

It is well known that the first derivative or the second derivative represents the change of a function. Therefore, possible edge points can be determined by looking for the local extrema of $f'(x,y)$ or by looking for a zero crossing of $f''(x,y)$.

For a two dimensional function, the first derivative is as following:

$$\nabla f(x,y) = \frac{\partial f(x,y)}{\partial x}\hat{i_x} + \frac{\partial f(x,y)}{\partial y}\hat{i_y}$$

For a 2-D sequence $f(i,j)$, the first partial derivatives can be replaced by some form of difference. For example, $\partial f(x,y)/\partial x$ may be replaced by

$$[f(i+1,j) - f(i-1,j)]$$

Some operators based on this idea are the Roberts, Prewitt and Sobel operators. Their mathematical expressions and mask forms are presented in the following.

Roberts

It is a 2 x 2 pixel mask in which

$$g_1(i,j) = f(i,j+1) - f(i+1,j)$$

$$g_2(i,j) = f(i+1,j+1) - f(i,j)$$

$$H_1 = \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix}$$

$$H_2 = \begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix}$$

Prewitt

It is a 3 X 3 pixel operator where

$$\begin{aligned} g_1(i,j) &= f(i+1,j-1) + f(i+1,j) + f(i+1,j+1) \\ &\quad -f(i-1,j-1) - f(i-1,j) - f(i-1,j+1) \end{aligned}$$

$$\begin{aligned} g_2(i,j) &= f(i-1,j+1) + f(i,j+1) + f(i+1,j+1) \\ &\quad -f(i-1,j-1) - f(i,j-1) - f(i+1,j-1) \end{aligned}$$

$$H_1 = \begin{vmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{vmatrix}$$

$$H_2 = \begin{vmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{vmatrix}$$

Sobel

It is also a 3 X 3 pixel operator where

$$\begin{aligned} g_1(i,j) &= f(i+1,j-1) + 2f(i+1,j) + f(i+1,j+1) \\ &\quad -f(i-1,j-1) - 2f(i-1,j) - f(i-1,j+1) \end{aligned}$$

$$\begin{aligned} g_2(i,j) &= f(i-1,j+1) + 2f(i,j+1) + f(i+1,j+1) \\ &\quad -f(i-1,j-1) - 2f(i,j-1) - f(i+1,j-1) \end{aligned}$$

$$H_1 = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix}$$

$$H_2 = \begin{vmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{vmatrix}$$

These operators usually utilise an rms nonlinearity to produce an edge enhanced array

$$A(i,j) = \sqrt{[g_1(i,j)]^2 + [g_2(i,j)]^2}$$

A magnitude point nonlinearity yielding

$$A(i,j) = |g_1(i,j)| + |g_2(i,j)|$$

is often used for computational simplicity.

Some results of these operators applied on a satellite image are give here. Fig. 7.2 is a satellite image (band7) of Adelaide. Figs. 7.3, 7.4 and 7.5 show the results of Roberts, Prewitt and Sobel operators using the rms point operator applied to the original image respectively. By comparison, it is clear that the Sobel operator is the best one among

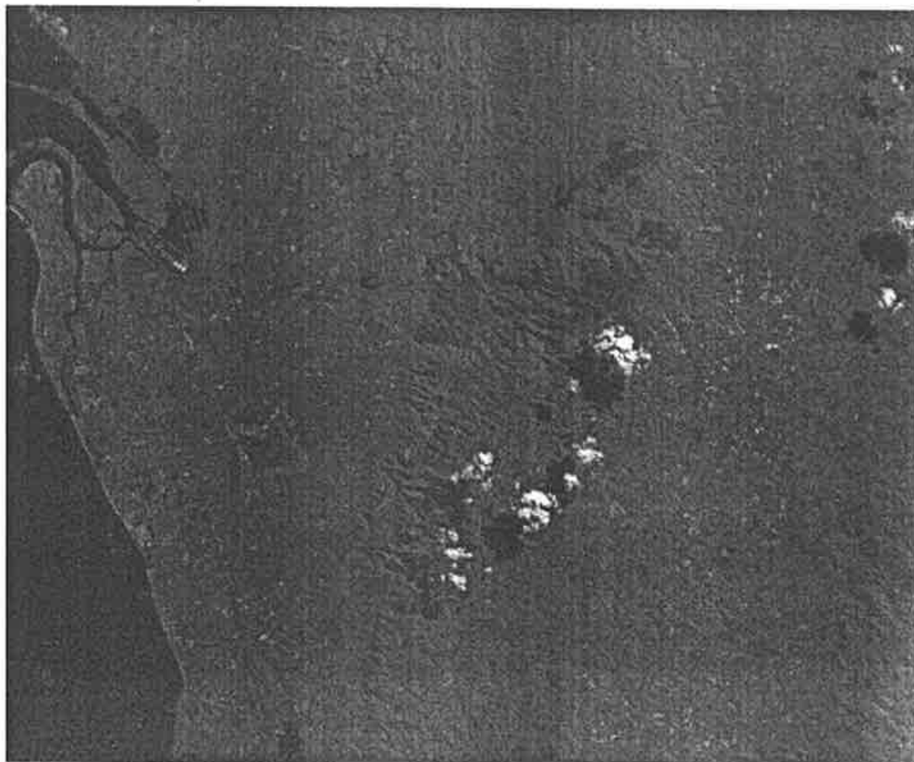


Figure 7.2: A satellite image of Adelaide (band7).

these operators.

There is another differential operator based on the second derivative $f''(x,y)$, ie. the Laplacian operator. It has three variations and their mathematical expressions and corresponding mask forms are given below.

$$\begin{aligned}\nabla^2 f(i,j) &= f(i+1,j) + f(i,j+1) + f(i-1,j) \\ &\quad + f(i,j-1) - 4f(i,j)\end{aligned}$$

Figure 7.3: The result of edge detection from Roberts operator.

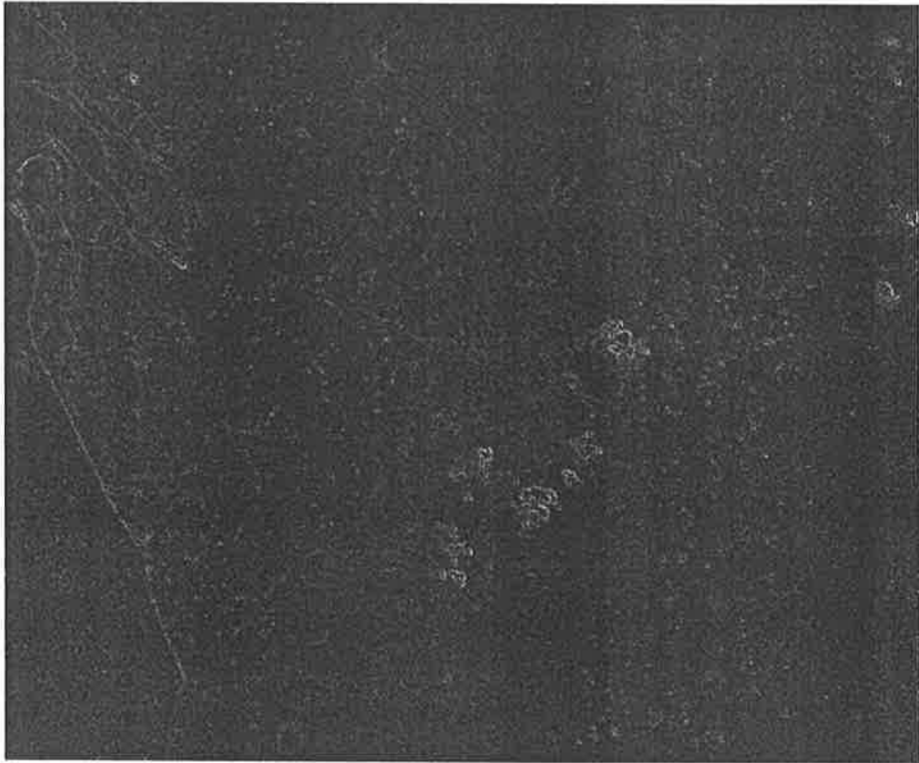


Figure 7.4: The result of edge detection from Prewitt operator.

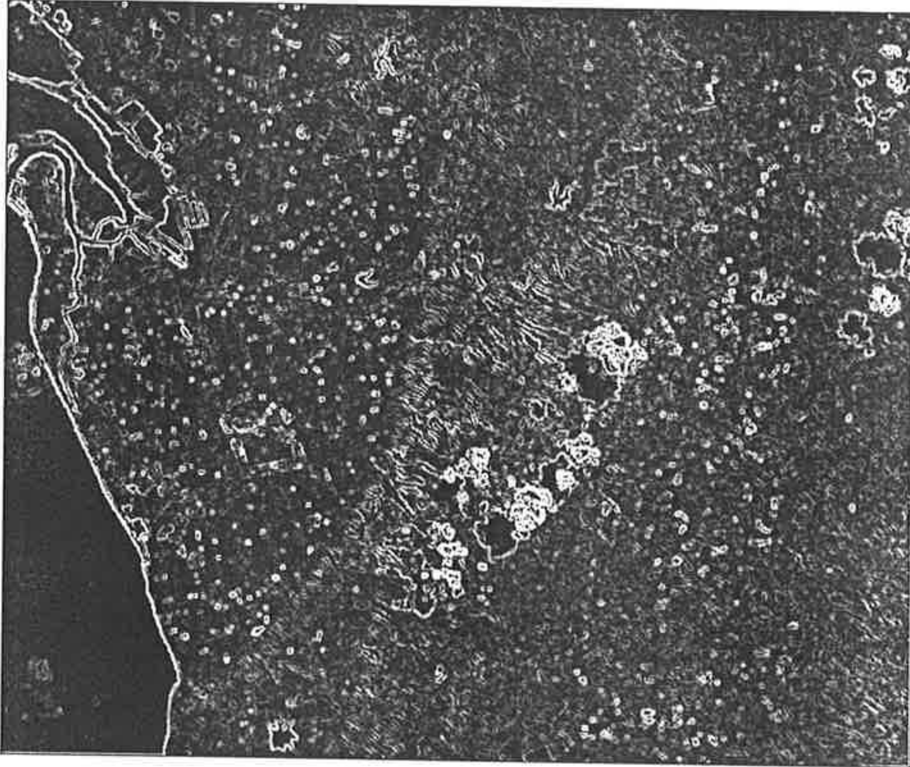


Figure 7.5: The result of edge detection from Sobel operator.

$$H = \begin{vmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

$$\begin{aligned} \nabla^2 f(i,j) &= f(i+1,j) + f(i,j+1) + f(i-1,j) \\ &\quad + f(i,j-1) - 8f(i,j) + f(i+1,j-1) \\ &\quad + f(i+1,j+1) + f(i-1,j+1) + f(i-1,j-1) \end{aligned}$$

$$H = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{vmatrix}$$

$$\begin{aligned} \nabla^2 f(i,j) &= 2f(i+1,j) + 2f(i,j+1) + 2f(i-1,j) \\ &\quad + 2f(i,j-1) - 4f(i,j) - f(i+1,j-1) \\ &\quad - f(i+1,j+1) - f(i-1,j+1) - f(i-1,j-1) \end{aligned}$$

$$H = \begin{vmatrix} -1 & 2 & -1 \\ -2 & -4 & 2 \\ -1 & 2 & -1 \end{vmatrix}$$

B. Template matching Operators

The template matching operators are a set of masks representing discrete approximations to ideal edges of various orientations. Two of eight possible compass orientations are given below for the compass gradient operator.

Mask H_1

$$\begin{vmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{vmatrix}$$

Mask H_2

$$\begin{vmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{vmatrix}$$

With these operators, the enhancement is formed as the maximum of the gradient arrays. Thus

$$A(i,j) = max|g_k(i,j)|$$

The edge orientation $\theta(i,j)$ corresponds to the compass direction of the largest gradient. No results are shown here because the algorithm is currently not available.

Edge fitting edge detectors

This method uses a local feature model to represent the grey scale distributions and uses parameters of the model as the edge measurements. One of the simplest and oldest method

was to assume that in each neighbourhood of the image the underlying gray tone intensity function f takes the parametric form of a polynomial in the row and column coordinates and that the sampling producing the digital image function is a regular equal interval grid sampling of the square plane which is the domain of f. Thus, in each neighbourhood f takes the form

$$\begin{aligned} f(r,c) &= a_1 + a_2 r + a_3 c + a_4 r^2 + a_5 rc + a_6 c^2 \\ &\quad + a_7 r^3 + a_8 r^2 c + a_9 rc^2 + a_{10} c^3 \end{aligned}$$

The problem here is to calculate the coefficients (or parameters) from the sampling data of the image. Using the least-square fitting method, they can be obtained from the following equations.

$$a_n = \frac{\sum_r \sum_c P_n(r,c) G(r,c)}{\sum_i \sum_j P_n^2(i,j)}$$

Where $P_n(r,c)$ is a set of orthogonal polynomial functions and $G(r,c)$ is the image data in this neighbourhood. After the coefficients are obtained, the gradient and the second derivative at any point can be easily calculated from the polynomial function. The gradient and the second derivative can be used as the edge measure.

Another commonly used model is to use a set of basis functions (eg. $\cos\theta$, $\sin\theta$) to approximate a step edge. If the fitted model is close to the local data distribution, a step edge is judged present and its strength can be measured from the coefficients of the basis functions. One of the advantages of the edge fitting method is that it can discriminate weak edges from the rapid fluctuating noise, if the model used is close enough to an ideal edge.

It is difficult to compare the performance of these operators and it depends on the application cases, whether they are differential, template or Laplacians. After these operators have been applied, more information about the local connectivity and threshold value should normally be added to reduce the random points (noise). The size of the operators also plays an important role in reducing noise and detecting fine details. The larger the operator size, more random noisy points are eliminated, but more details about edges are also lost. Therefore, some noise reducing algorithms should usually be used before applying edge detection operators. Lowpass filtering and median filtering are two examples for reducing noise.

Median Filter Median filtering is a nonlinear process useful in reducing impulsive, or salt-and-pepper noise. It is also useful in preserving edges while reducing random noise. In a median filter, a window slides along the image, and the median intensity value of the pixels within the window becomes the output intensity of the pixel being processed. For example, suppose the pixel values within a window are 5, 6, 55, 10, and 15, and the pixel

being processed has a value of 55. The output of the median filter at the current pixel location is 10, which is the median of the five values.

Fig. 7.6 shows the result of a 3x3 median filter applied on the original image.

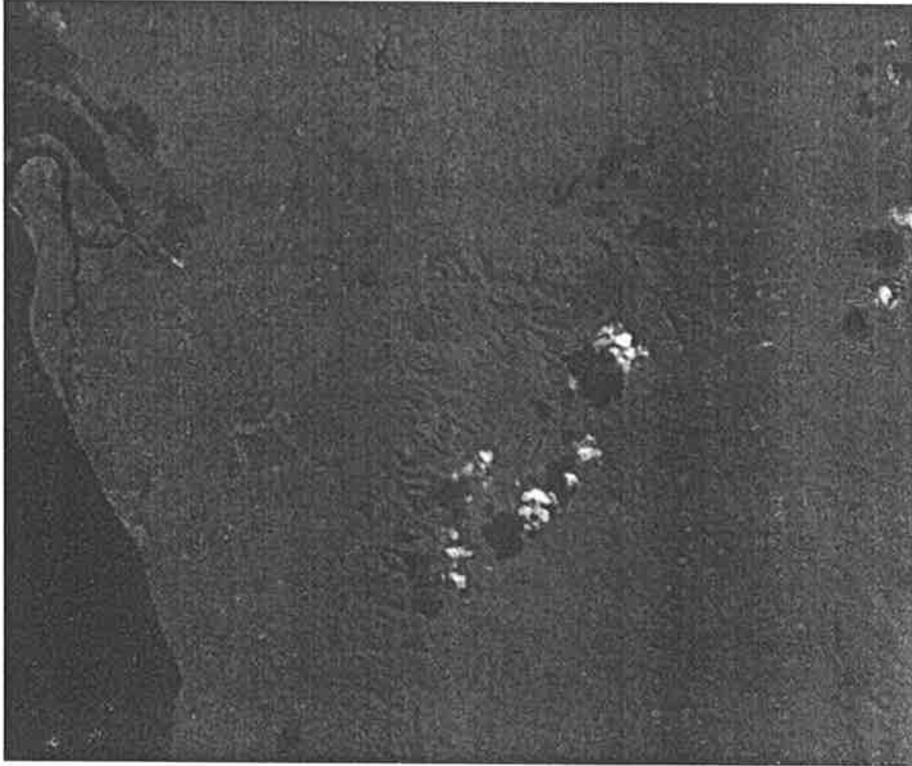


Figure 7.6: The result of a median filter.

Fig. 7.7 shows the result of Sobel operator applied on the median filtered image. From Fig. 7.7, it can be seen that most random points have disappeared, while some image details have also been lost. This is the disadvantage of most noise reducing operators.

Summary Several edge detecting operators have been reviewed and their results on a satellite image have been shown. Also some noise reducing algorithms have been mentioned. According to their mathematical expressions, it is not difficult to implement them through hardware, especially using GaAs because of its high processing speed. The processing speed of such ICs would be much faster than the conventional software implementation, therefore the era of processing huge amount of data in real time is indeed a reality.

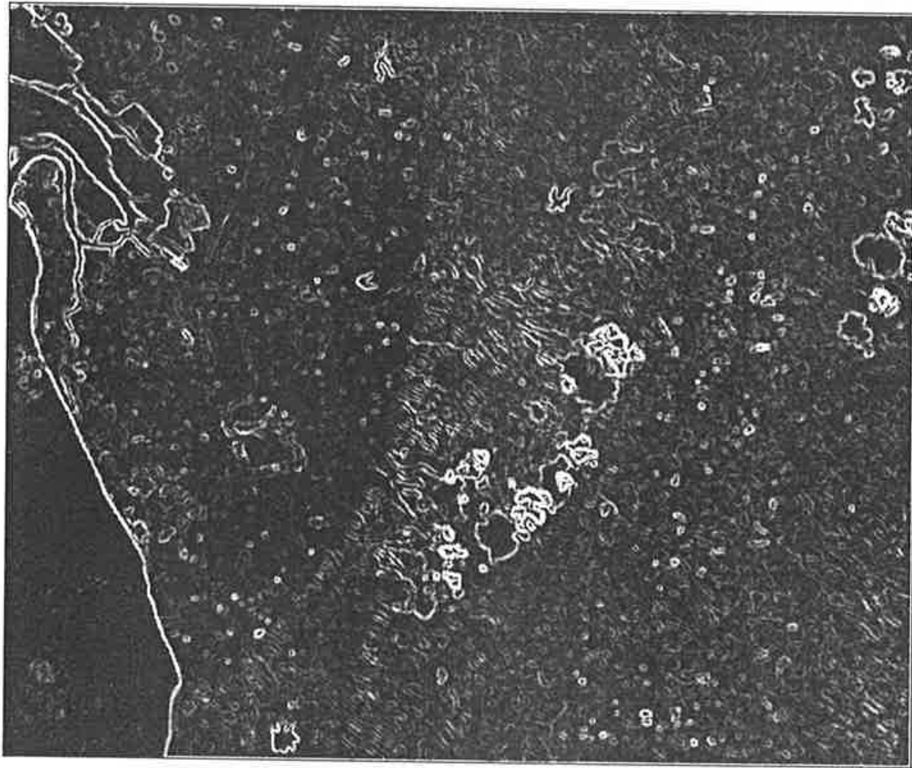


Figure 7.7: The result of Sobel operator applied on the median filtered image.

7.2.2 Polynomial Evaluator

In order to implement the 'registration' image processing function, discussed in Section 7.2.1.1, a polynomial evaluator was identified as a useful processing element. This structure would also have applications in edge detection for fitting corrected grey levels to the image. This section deals with the design of an architecture for a polynomial evaluator, as an example of a digital processing function that could be integrated with the image sensor within the aerospace application paradigm. Furthermore, it serves as a case study of an implementation of the design methodology developed in the last chapter.

The polynomial evaluator (PE) described hereafter is an example of a type of processing unit that performs a reasonably complex function whilst still being fundamentally a simple architectural element. Although the implementation decided upon for this particular application is by no means the only one applicable nor is it the fastest, it works and does so reliably, and with minimal chip area.

An important part of the design methodology in specific relation to this architectural entity is the need for a two-way design to take place simultaneously. Top-down design is necessary to allow the behavioural description to be realised correctly, and bottom-up design is needed to ensure that the algorithm chosen to implement the unit is such that it is easily realisable in Gallium Arsenide, and is consistent with the accepted design principles and limitations of this technology.

The technology that has been decided upon for the implementation of this functional element is Source-Follower Direct-Coupled FET Logic, or SDCFL. The reason that this particular technology has been chosen is the favourable physical characteristics that it displays, namely high noise-margins, good temperature stability, and also good fan-out capabilities. The power dissipation and switching delays are comparable to and also compatible with that of ordinary DCFL systems, so the advantages mentioned above are paid for only in chip area, since SDCFL systems generally require more transistors (for example, a DCFL inverter had only 2 devices, whilst an SDCFL inverter has 4). The drive voltages of SDCFL (and DCFL) are considerably lower than that of the usual silicon VLSI implementations, drain voltage connections (VDD) are generally about 1.4V, and source voltages (VSS) generally -0.2V.

In the design of the polynomial evaluator using GaAs SDCFL, as the logic family, the 'Ring Notation,' which considerably simplifies the 'pen-and-paper' stage of the device layout, was adopted. Fig. 7.8 an example of the notation being used in this exercise.

Since it is generally implicit to which power rails the different devices connect to, the notation is simplified further and a single 'power' rail is shown in later ring notation

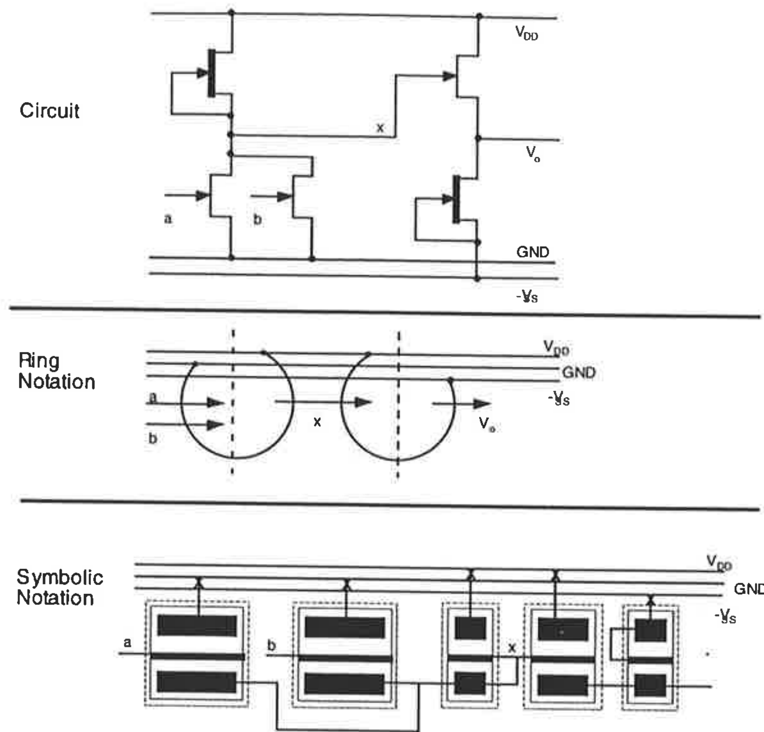


Figure 7.8: SDCFL design example.

diagrams. It is assumed that the correct connections are made, without explicit statements to that effect.

7.2.2.1 Behavioural Model

The behaviour of the PE is very simple to describe, it consists of supplying a set of coefficients $a_n, a_{n-1}, ..., a_0$ of an arbitrary order polynomial, and a step size, h, and proceeding with an execution cycle to obtain the next data point. The general representation of an arbitrary order polynomial is given by,

$$y(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0$$

Such an execution cycle is repeated until no more data is required from the polynomial, at which time new coefficients and step size is loaded and the process is simply repeated.

7.2.2.2 Algorithmic Solution

In keeping with the methodology discussed in the introduction, it is at this point where a certain component of bottom-up design is necessary.

It is required that the algorithm chosen can be mapped into a GaAs layout.

By examining the form of the polynomial, it is possible to transform it into a form in which the only operation necessary for the evaluation of any point on the curve from the initial starting point is found by the simple application of an recursive addition process on the individual bits of both the initial point and a set of coefficients. The decomposition of the algorithm, into the recursive structure, is detailed in Appendix E.

Commencing with the general form of a polynomial,

$$y(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$$

Then from Appendix E, the new, bit-serial form of the polynomial is described by:

$$\begin{aligned}
y(x_{i+1}) &= y(x_i) + d_1(x_i) \\
d_1(x_{i+1}) &= d_1(x_i) + d_2(x_i) \\
d_2(x_{i+1}) &= d_2(x_i) + d_3(x_i) \\
&\vdots \\
d_{n-1}(x_{i+1}) &= d_{n-1}(x_i) + d_n(x_i) \\
d_n(x_{i+1}) &= d_n(x_i)
\end{aligned}$$

where

$$x_{i+1} = x_i + h$$

and

$$d_j(x_{i+1}) = d_j(x_i) + h\dot{d}_j(x_i) + \frac{h^2}{2}\ddot{d}_j(x_i) + \cdot + \frac{h^{n-j}}{(n-j)!}d_j^{(n-j)}(x_i)$$

also

$$d_j(x_i) = h\dot{d}_{j-1}(x_i) + \frac{h^2}{2}\ddot{d}_{j-1}(x_i) + \cdot + \frac{h^{n-j+1}}{(n-j+1)!}d_{j-1}^{(n-j)}(x_i), d_0 = y$$

It should be noted that the manipulations required to convert an arbitrary polynomial into the form that is required by this algorithm are not trivial. This means the coefficients required by the hardware may take some time to calculate.

It may be necessary to have coefficients pre-computed for a number of common polynomials and starting values, and approximations may be made using interpolations for values that are required that do not coincide exactly with any stored set of coefficients. It is likely that for evaluations of polynomials requiring a large number of points from the same polynomial, significant speed increases in the evaluation would be obtained if the coefficients were calculated at run-time, since the time spent on this calculation would be offset by the large number of interpolations necessary if the 'approximate from stored values' method was used.

7.2.2.3 Functional Specification

The algorithm described in the previous section requires an iterative addition process, as shown in Fig. 7.9. The required functional blocks to realise the algorithm are adders and registers. The adders perform the calculations and the registers initially store coefficients, and then store intermediate values as the block is clocked in its execution cycle.

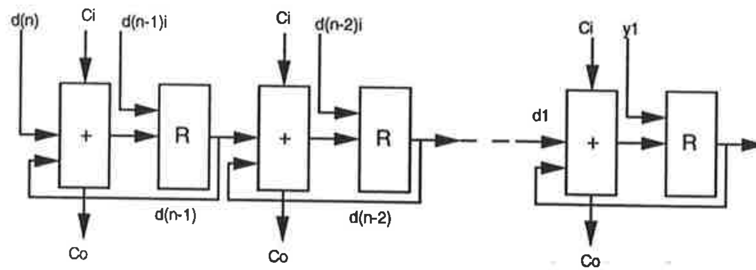


Figure 7.9: Block diagram of bit-serial algorithmic solution.

With reference to Fig. 7.9 the notation being implemented is as follows:

$d(n-j)$ = Coefficient $(n-j)$, all except $d(n)$ change as execution cycle is run.

$d(n-j)i$ = Initial coefficient $(n-j)$, loaded at start-up.

$y1$ = Initial evaluation point.

C_i = Carry in

C_o = Carry out.

The coefficients $d(n-j)i, d(n)$ must be loaded initially into the system. This can be achieved by the inclusion of a 2 way multiplexor (diplexor), so that when coefficients are being loaded into the system, the adders are by-passed and no execution cycles take place.

With the addition of this requirement, the system now has two states: Load, and Execute. An extra 1-bit signal may be used to realise this control.

7.2.2.4 Structural Description (Floor Plan)

A single PE cell consists of the three functional elements, an adder, a register and a diplexor. These elements may be connected as shown in Fig. 7.10 to form the PE cell for the polynomial evaluator.

The approach is illustrated by Fig. 7.10, which lends itself readily to replication. A single PE cell may be replicated both vertically to increase the bit size of the results, and horizontally to increase the order of polynomial.

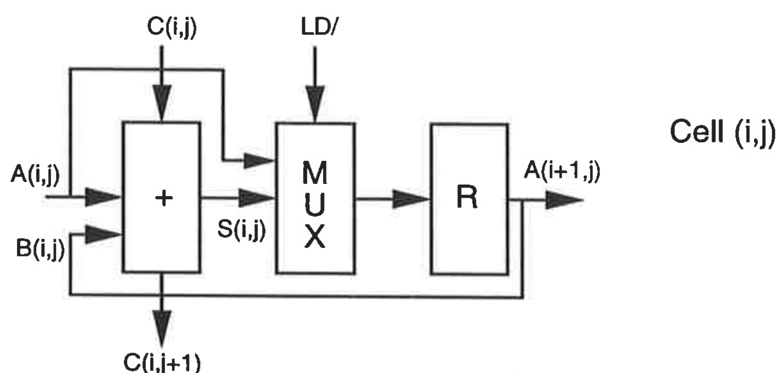


Figure 7.10: PE cell structural layout.

7.2.2.5 Symbolic Physical Description

At this point a decision had to be made about the implementation of the functional elements, ie. the adder, the register and the diplexor.

The Adder

It was decided that for the sake of simplicity, initially a simple ripple-through carry adder would be sufficient to realise the function required without overly increasing the complexity of the cell.

The adder can be described by three logic equations, ie. the half-carry (H_k), the sum (S_k), and the carry (C_k). Since the technology that has been decided upon for the implementation of the PE is SDCFL GaAs, it is necessary to manipulate the equation (using DeMorgan's Laws) into a form which contains only inverters and NOR gates. The

OR gate in SDCFL is of the 'wired-OR' form, and hence can only be used between NOR gates and inverters, ie. two OR gates in series cannot be used. Following is the logic equations for the three required results and their transformed equivalents.

Half-Carry: $H_k = \overline{A_k}B_k + A_k\overline{B_k} = \overline{A_k\overline{B_k}} + \overline{\overline{A_k}B_k}$

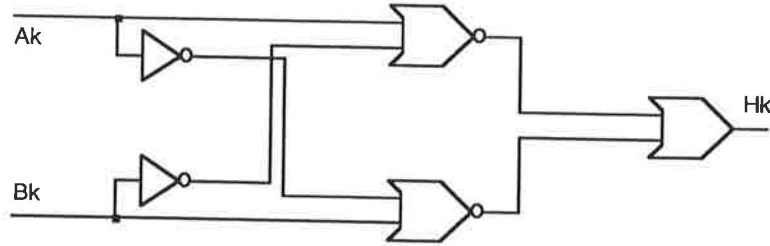


Figure 7.11: Generation of half-carry result.

Sum: $S_k = H_k\overline{C_k} + \overline{H_k}C_k = \overline{\overline{H_k}C_k} + \overline{H_k\overline{C_k}}$

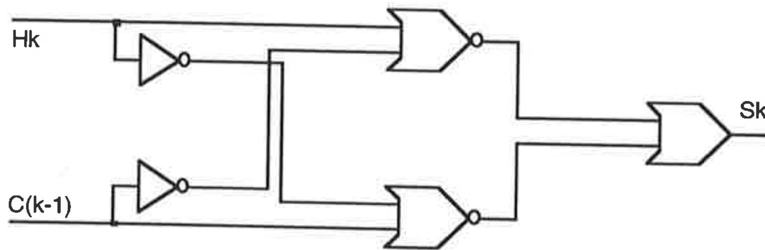


Figure 7.12: Generation of sum result.

Carry: $C_k = A_kB_k + H_kC_k = \overline{\overline{A_k}\,\overline{B_k}} + \overline{\overline{H_k}\,\overline{C_k}}$

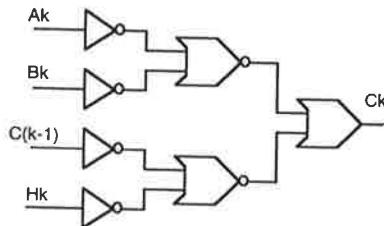


Figure 7.13: Generation of carry result.

Figs. 7.11, 7.12 and 7.13 illustrate the logical transformation of the three expressions.

The Register

The register, shown in Fig. 7.14, was by necessity an edge triggered device, since the data at the inputs would be changing while the data at the output was still needed for the next PE cell to the right, and transparent latches cannot realise this property without placing constraints on the clocking system. A negative edge-triggered 'D' type latch was used in the layout.

Figure 7.14: Register logical representation.

The Diplexor

The logic equation that describes the diplexor is given by:

$$\overline{Z} = I_0\overline{LD} + I_1 LD$$

The diplexor, shown in Fig. 7.15, was the simplest element in the cell and was realised using two SDCFL 'NOR' gates with their outputs in a 'Wired-OR' configuration.

Figure 7.15: Diplexor logical representation.

The PE Cell

The PE cell in its entirety consists of the adder, diplexor and latch connected in such a way so that the input to the latch is selectable between the A input to the adder, and

the S output of the adder. Hence the overall result of which is that depending on the state of the control input, LD, either a) the system passes constant values through the latches (latch input = A of adder), ie. loading data coefficients, or b) the system executes an evaluation cycle and produces a result (latch input = S of adder). Since the output of the diplexor is the complement of the selected input, it is necessary to connect the diplexor output to the $D/$ input of the latch in order to have the correct value latched. The D input of the latch is connected to the output of the NOR gate adjacent to the $D/$ input. This allows the latch to be operated without the normally required complementary input. The Q output of the latch is connected to the B input of the adder, to realise the feedback path and hence the recursive property of the cell.

7.2.2.6 Physical Layout

Each single PE cell contained the three functional elements, the adder, the D-latch and the diplexor. The adder cell was initially designed using the abbreviated ring notation. The initial design entailed 3 vertical rows of devices. This was not considered optimum, since a lot of area appeared to be wasted at the edges of the design. The floor plan finally arrived at for the final implementation is shown in Fig. 7.16. It encompasses all the necessary design criteria, including placement of the input and output signals so that the PE cell may be assembled with minimum length routing paths between sub-circuits.

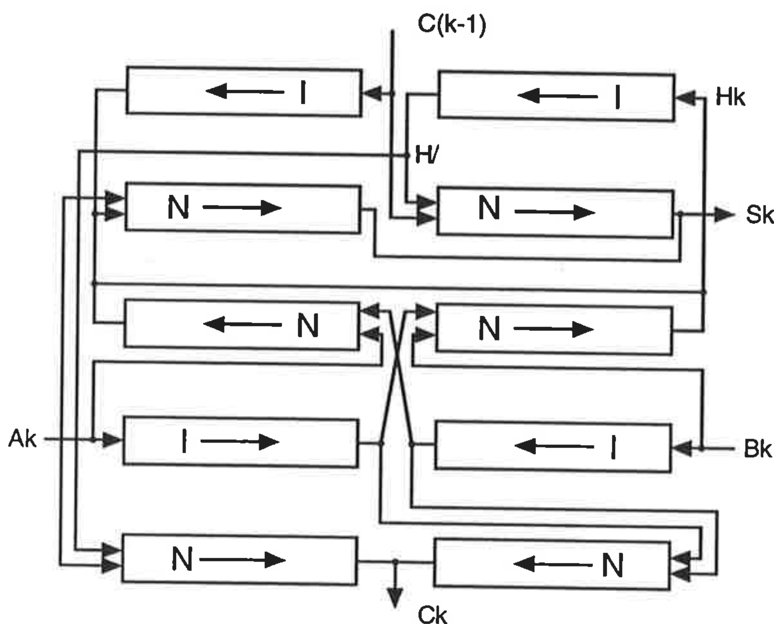


Figure 7.16: Adder floor plan block diagram.

Following the generation of the block diagram, ring representation was developed for all the sub-circuits. The Ring Notation diagrams for the three functional elements are shown in Figs. 7.17, 7.18 & 7.19.

In the above diagram, the thick horizontal lines represent the power and ground buses. All the buses have GND, VDD and VSS rails, Bus (1) also contains LD and $LD/$ signal, bus (2) has the Clk signal for the register, and bus (3) has the feed-forward by-pass path for the A_k signal through the adder, and the feedback path for the Q output to the B_k input through the register.

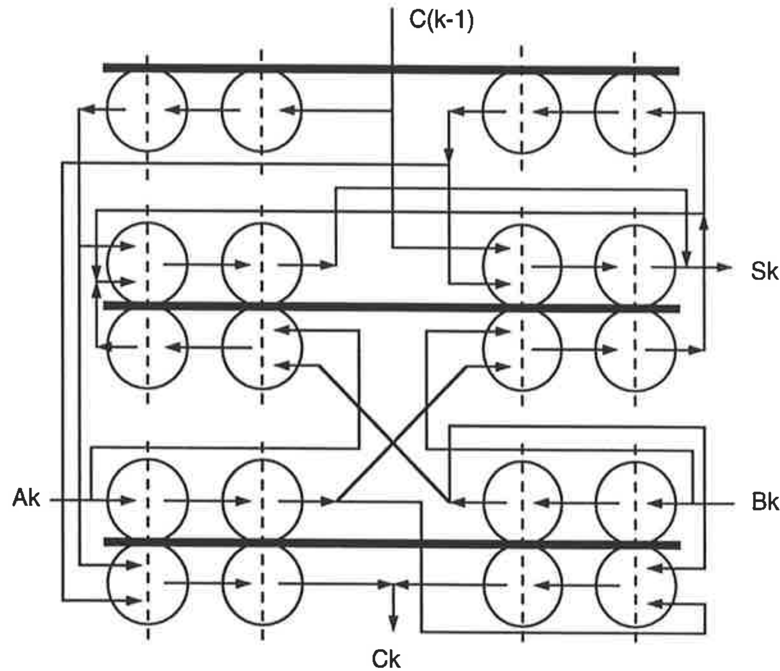


Figure 7.17: Adder cell ring notation.

In the layout, the adder occupied 5 rows of gates. This was fortunate due to the fact that the diplexor occupied one row, and the D-latch 4 rows. This allowed the Adder to be placed on the left of the cell, with the diplexor and the latch placed vertically (relative to each other) to the right of the adder. The full cell arrangement is illustrated by Fig. 7.20.

Since this placement of devices occupies 5 horizontal rows, it was not possible to stack these cells vertically, because two rows share a single set of power and ground buses. Therefore it is required that an even number of rows of devices be present in order to replicate any block vertically. Hence it was necessary to layout another PE cell below the first, and an anti-symmetric arrangement with respect to the power and ground buses. This arrangement occupied now a total of 10 rows.

In the 2 PE cell arrangement, the clock (clk) and control signals (LD and $LD/$), as well

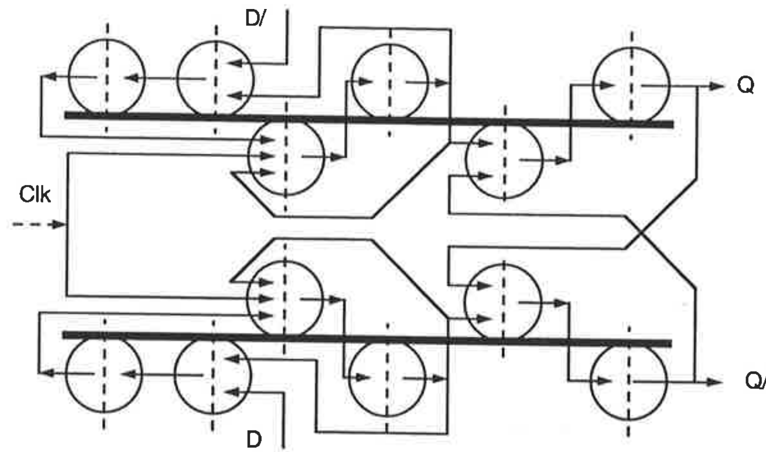


Figure 7.18: Register cell ring notation.

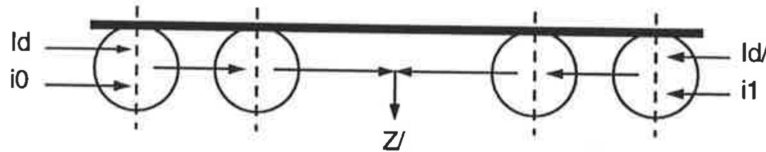


Figure 7.19: Diplexor cell ring notation.

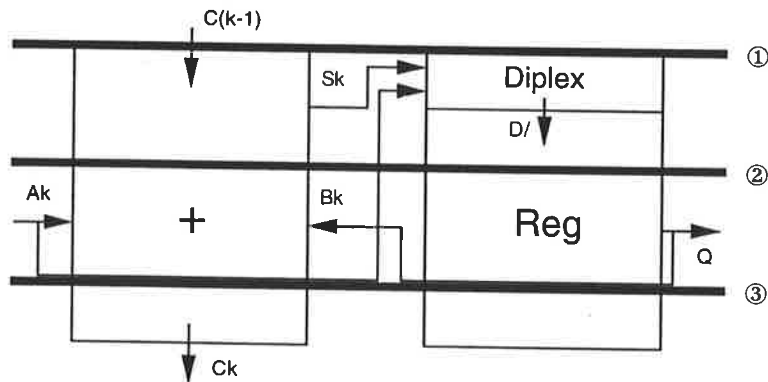


Figure 7.20: PE cell structural block diagram.

as the feed-forward A signal (to the input of the diplexor), and the feedback Q signal (to the B of the adder), are separated from the VSS and VDD rails. This is due to the high level of crosstalk and capacitive effects which introduce noise into the supply rails when there are fast switching transients on signal lines nearby.

Fig. 7.21 illustrates the layout of the completed 5th order polynomial evaluator having an 8-bit wide data path.

Figure 7.21: Polynomial evaluator cell.

The unit will only load coefficient data when the circuit connected to the input buffer has data available, and similarly the unit will only clock an execution cycle when the circuit

connected to the output has indicated that it is ready to receive data. The LogicWorks^{TM} circuit diagram of the adder, register unit and the PE cell, together with corresponding simulation results are shown in Figs. 7.22, 7.23 and 7.24.

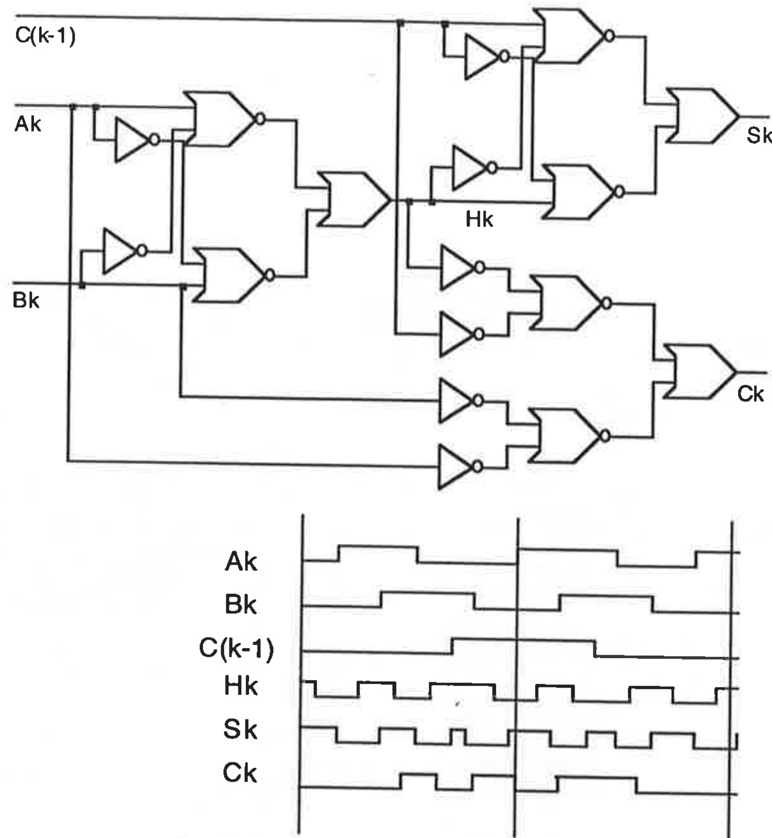


Figure 7.22: Adder and simulation.

The power dissipation of the device presents somewhat of a problem: The measurements of the current flow in the 241 cell indicate that power dissipation for the double cell is of the order of 16.5 mW, for a clock speed of 250 MHz (4 ns between clocks). Since the cell dimensions are approximately 900 λ by 700 λ, with a feature size (λ) of 0.4 mm, this equates to approximately 0.1 mm^2. With some simple extrapolation, it is clear that the dissipation of an arbitrary size PE cell is approximately 165 mW/mm^2. This is generally considered to be too high for use in GaAs, where the accepted limit for chips is approximately 100-120 mW/mm^2. The solution to the problem is to arrange that the layout have sufficient spacing between the individual cells in order to reduce this dissipation/area figure.

The propagation delay of each cell in a horizontal manner is approximately 2.8 ns. In the vertical direction, it is necessary to examine the propagation delay of the carry signal

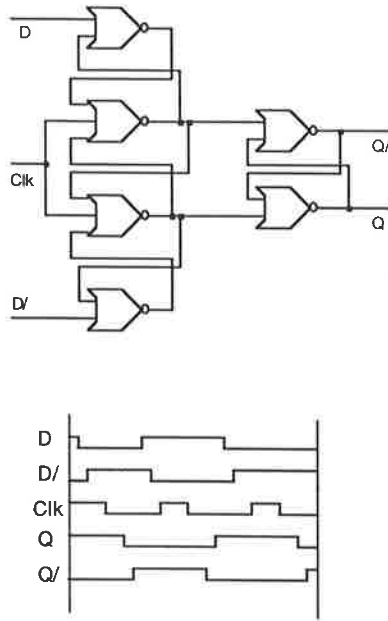


Figure 7.23: Register and simulation.

through each cell. Since during each evaluation cycle data is only moved from one cell to its immediate right neighbour, the horizontal delay is independent of the polynomial order. The vertical delay, however, is a different matter. The carry signal must propagate through the entire bit-depth of the array and allow the adder output data to settle in the most significant bit before any clocking of the MSB row can take place. It does not appear advantageous to clock the lesser significant bits before the entire arrays adder output is stable.

The estimation of the delay time can now be made. Using the data obtained from the adder simulation plot, the carry signal delay is in the worst case 1200 ps. Therefore we must wait $(n-1)$ times this long (for n bits deep array), plus the horizontal delay. As the bit depth increases, the horizontal delay contributes relatively less to the overall delay. Note that this is only the theoretical minimum clock delay for an evaluation cycle, and a data load cycle will have considerably less delay involved, since no carries are required, and furthermore the adder is by-passed resulting in a very fast load cycle (compared to the execution cycle). The delay (for the execution cycle) is therefore estimated to be 741200 ps (carry prop. delay) = 8.4 ns + 142800 ps (PE cell prop. delay) = 11.2 ns, giving a theoretical clock speed of 90 MHz.

Although this figure could be construed as being fairly slow for a technology such as GaAs, the most serious problem occurring here is the time complexity of the array being

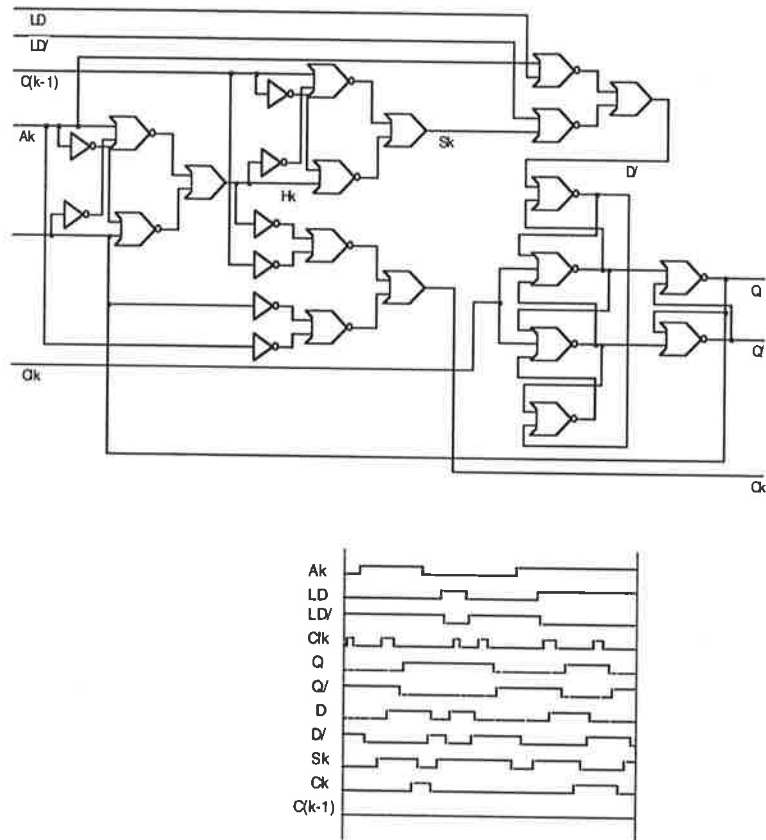


Figure 7.24: Block diagram of bit-serial algorithmic solution with PE cell structural layout below.

proportional to the bit-depth of the implementation. If this dependency could be removed, or at least reduced, significant savings could be made in the propagation delays of the array. One possible method of realising this would be to substitute the present ripple-through carry adder with a device such as a 4-bit carry lookahead adder. This would significantly reduce the time taken for the carry to propagate vertically through the array, and hence improve the overall performance of the circuit.

7.3 Case 2: The Mobile Multimedia Paradigm

Having explored some possibilities for the image sensor within the aerospace vision systems paradigm, we now review an application within a personal communication systems (PCS) paradigm.

Personal communication systems of the future will augment the mobile phone concept to include multimedia services such as: digitised speech, data, e-mail, paging, fax, GPS, still image compression, compressed/slow scan video and eventually real-time video. The many potential applications for this technology include use of mobile telecommunications in areas such as telemedicine, office teleworking, interactive telebanking, teleshopping, teleconferencing, personal communication/navigation, tele-education and interactive time keeping (ITK)[1] In summary the key visions, of this Author, that will drive this technology alongside video capture are:

- *Interactive Telebanking (ITB):* the IM^{3}PC will replace all types of bank card. On entering a PIN number into the phone, shopping transactions can automatically happen and the user's final bank balance can be checked on the phone's display
- *Interactive Time Keeping (ITK):* the time will be displayed in the corner of screen of the IM^{3}PC. This time display will automatically change for daylight saving and during interstate travel. For international travel, time will automatically speed up or even go backwards! When making an international call, the time in that region will be displayed
- *Interactive Personal Navigation (IPN):* when a 'locate' button is depressed on the phone, the nearest street name that you are located to will be displayed. This

[1] Interactive time will fastforward or reverse automatically during international travel. Time will automatically change for daylight saving, leap years, crossing interstate time zone boundaries etc. Local time for any overseas location can be retrieved by simply entering the area and country codes – all adjustments for local time zones etc. will be automatic. Even information as to whether there is a current public holiday in that time zone can be displayed.

essentially becomes a smart hand-held global positioning system (GPS).

This range of services, together with interactive video, will lead to the development of interactive mobile multimedia personal communicators (IM^3PCs) that require the integration of many sub-systems including real-time image and signal processors, computer vision, RF communications links, and high-speed networks. These systems demand very high speed processing, small physical size and low power consumption and in the next five years demand for such a combination of processor attributes will increase.

This will ultimately be a fully public system that is suited to operate in a microcell network environment with a fibre optic backbone. As an achievable intermediate step towards this vision, we hypothesise that IM^3PCs will first operate within a wireless local area network (LAN) office environment with a localised degree of mobility in mobile teleworking and teleconferencing applications. Given this first step, an enormous opportunity exists for companies that can exploit underlying microelectronics technology to provide IM^3PCs at a price acceptable to the business community. In particular, the following sub-systems must form part of any IM^3PC and their integration represents the major hurdle to realisation of IM^3PC technology:

- high-speed high-density RAM;
- very high-speed video compressor;
- very high-speed modulo multiplier encrypt/decrypt circuit;
- solid-state imager (this work);
- integrated RF/digital circuitry for wireless communication.

To facilitate hardware realisation and integration of these system elements, advances in both technology and implementation are required.

7.3.1 Complementary Gallium Arsenide

Motorola have developed a complementary gallium arsenide integrated circuit manufacturing process (CGaAsT™) that provides both n-type and p-type pseudo-morphic high electron mobility transistors with a semi-insulating gate (HIGFETs). Hole mobility in HIGFETs has been found to be very much higher than in MESFETs [Bernhardt 95], enabling CMOS-like gates to be built from dual networks of n-type and p-type transistors that have power dissipations two orders of magnitude lower than gates built from

MESFETs. Consequently, CGaAs™ overcomes the major difficulty associated with using gallium arsenide to fabricate integrated circuits – namely, its static power dissipation – permitting the integration of millions of very high-speed HIGFETs onto one chip, offering system performances equivalent to CMOS for lower power dissipation or superior system performances for the same power dissipation.

7.3.2 Objectives

Initially, the primary objective would be to develop the critical sub-systems for a future IM^{3}PC capable of operating in a wireless LAN environment using CGaAs™ VLSI technology in every case to demonstrate the flexibility and truly integrating nature of that technology. The critical sub-systems are listed above but it is worth noting that some standard parts are also required such as phase-locked loops, digital-to-analog converters, codecs, etc. A secondary objective would be to investigate the applicability of the Ring Notation layout style, developed at the University of Adelaide, to CGaAs™ integrated circuits and to ascertain a set of rules that determine which logic family (or families) should be used in which design in order to derive a rapid, easy-to-use, reliable design style for CGaAs™ .

7.3.3 Current State-of-the-Art

Internationally, many researchers have proposed and realised desk top multimedia applications; such as in [Ackland 95], [Beadle 93] and [Lin 95]. Mobile or wireless multimedia has been considered, such as in [Meng 95] and [Moura 96], proposing various algorithms for efficiently processing the data. However, there is little work that is specifically addressing the hardware issue of a compact and low power hand-held IM^{3}PC device – this is due to speed-power limitations in the current VLSI technology. Now with CGaAs™ as the emergent leading technology, that promises to cater for this niche, there is a window of opportunity for unique research in this area.

7.3.4 Sub-systems

We now finally proceed to briefly review some of the basic building blocks for the system.

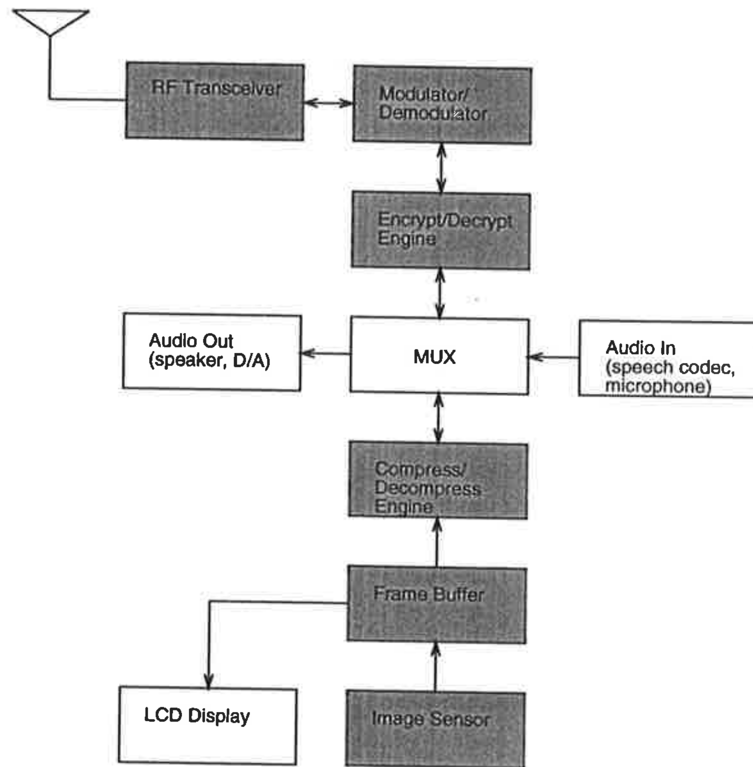


Figure 7.25: System block diagram for basic mobile video device. Unshaded areas are standard tasks. Shaded areas need to be particularly addressed by CGaAs solutions.

7.3.4.1 High-density high-speed RAM

High performance memory structures are required to implement data compression algorithms – eg. transpose memory. Also, cryptographic processing requires scratch-pad memory for partial results storage, and each frame captured by the imager will require a frame buffer. GaAs VLSI has long suffered from the high gate leakage current exhibited by MESFETs, so that GaAs RAM cells have been typically static in design. We have developed a 3-transistor dynamic RAM cell design, suitable for MESFET technology, that is smaller and less power-hungry than the static RAM cell [McGeever 95]. This cell design can be implemented in CGaAs™ in a small (1kb) RAM to optimise its performance and then larger RAMs can be built, selected for the above uses.

7.3.4.2 Image compression

Compressed video is critical to the development of multimedia applications, because the storage requirements and associated transmission times for uncompressed images (still or video) are prohibitive. Image compression methods reduce the amount of data that needs to be transmitted and stored and in recent years, there has been significant progress towards the development of image compression standards:

- CCITT H.261 for tele-conferencing
- ISO JPEG for high quality still image storage and retrieval
- ISO MPEG I for interactive video playback
- ISO MPEG II for entertainment quality video distribution

All of these standards employ transform coding based on the Discrete Cosine Transform (DCT) to remove spatial redundancy in the image data. The DCT is the most demanding part of the video encoding/decoding process in terms of the number of arithmetic operations required [Ackland 94]. Hence, the implementation in CGaAs™ of a DCT processor is required. Motion Compensation or Interpolation techniques need not be considered, because in the applications intended, there will be little variation between succeeding frames, so that only Frame Differencing need be employed to further reduce the amount of transmitted data. Our preliminary experiments indicate that using frame differencing only (and dispensing with interpolation and motion estimation) greatly reduces the complexity of video compression circuits for only a 3% increase in the compressed image data size [Wagnitz 95]. The design and fabrication of a 1-D DCT processor in CGaAs™

to establish performance constraints, and then the design a full 2-D DCT processor with associated transpose memory, is suggested for future work.

7.3.4.3 Modulo Multiplier for Data Encryption/Decryption

Global digital communications networks are growing at an exponential rate to facilitate the electronic transfer of data in areas as diverse as entertainment, business, law, government, personal communication, and banking, to name but a few. Data transmitted over these networks will often be of a sensitive nature and hence security of transmitted data is becoming increasingly important. Data encryption is employed to provide security and authentification of data: transmitted data must not be intercepted and deciphered by anyone unauthorised to do so; data to be transmitted must be "signed" so that fake received messages may not be invented by a receiver nor real sent messages disowned by a sender. Since its invention in 1978, RSA coding has been widely accepted as the best means of implementing public-key coding systems. For example, the Australian Standard for Electronic Data Transfer, AS28095.5.3 specifies RSA. It will be more widely used still in virtually all forms of secure data transmission now that the U.S. Government has accepted that RSA provides a means for electronically "signing" a document for electronic transmission.

Existing work has shown that radix-64 ("6 bits at a time") modulo exponentiation minimises the number of multiplications to perform an exponentiation of any length and use of Montgomery's modulo multiplication algorithm [Montgomery 85] enables long word length (512 bit) modulo multiplication rates in excess of 10 MHz to be achieved. (We are currently investigating a modification of Montgomery's algorithm to facilitate modulo squaring of long word length numbers, which is also needed in RSA processing.) Combination of these two techniques will lead to 512-bit processing on CGaAs™ in excess of 1 Mb/s, in comparison with a silicon-based processing rate of 64 kb/s [Kornerup 93].

7.3.4.4 Solid-state imager

Image capture is the vital ingredient in the development of IM^{3}PC devices. Considerable research and developmental work has already been directed towards realisation of a GaAs image sensor, herein, using a standard digital GaAs IC process, thus enabling integration with a high speed GaAs image preprocessor. The key reason for developing the sensor using a digital GaAs IC process is the facility to integrate this sensor with a digital GaAs preprocessor for low-level vision tasks such as thresholding, contrast enhancement, edge detection etc. and/or data compression for remote communication. The advantage of a

GaAs preprocessor is to meet the demands for ever increasing speed-power performance. Improved speed-power together with the increased compactness from the circuit integration suits advanced mobile applications well with their demand for minimised power consumption.

Our work on GaAs solid-state imagers to date includes extensive laser-based optical characterisation of GaAs detectors, extensive design and simulation studies of imager read-out circuitry, theoretical modelling of photo-detector characteristics and full noise analysis of imager output amplifier circuitry. This work uncovered the superiority of GaAs over Si especially with regard to spatial resolution, photo-collection efficiency, and anti-blooming capability. However, conventional GaAs MESFETs displayed disappointing levels of shot noise in the imager output circuit. This problem is totally solved by the move to CGaAs™ which uses a semi-insulating gate that dramatically reduces the gate leakage current and hence shot noise levels. We estimated, in Ch. 4 that CGaAs™ is ultimately expected to produce an factor of 10 improvement in overall noise performance over Si.

The results of our research on imagers in GaAs apply equally to CGaAs™ because all the anticipated benefits still hold for CGaAs™ . However CGaAs™ has the added advantage that proper pass gates are realisable, thus solving the problem of threshold voltage drops across analog switches. The semi-insulating gate also permits use of capacitive bootstrapped circuit techniques in the address circuitry, if desired, to provide variable pulse height control (this may be useful in the early stages of optimising the imager's performance). Finally it should be noted that Motorola's CGaAs™ process is truly planar, with no gate recess etching, thus providing the perfect medium for imager design work.

For future work, a small 32 x 32 pixel test imager will be designed and fabricated to provide experimental data to optimise the final design of the imager. A second larger imager (128 x 128 pixels) will then be designed and integrated with an output amplifier and analog-to-digital conversion module together with a frame buffer to demonstrate CGaAs™ capacity for integrating optical and digital circuitry on one substrate. The imager size of 128 x 128 pixels was chosen because that image size is deemed to be suitable for most IM^{3}PC applications.

7.3.4.5 RF/digital interface

CGaAs™ also offers the prospect of integrating RF and digital circuits onto one chip – a degree of integration not readily achieved in silicon technology. This represents a major step forward in IM^{3}PC technology because the number of chips needed in a communicator are reduced if some RF and digital functions are placed on the same IC. Most power

consumption in an IC is due to the i/o pads, which house relatively high power pad drivers. If there are fewer chips in a system, there will be fewer pads and consequently a lower power demand. Such a reduction in power requirements is vital to IM^{3}PC technology because it leads to longer battery life, as well as increased reliability and a reduction in size and weight.

Direct digital down conversion using under-sampled analog-to-digital converters removes all the IF circuitry found in a conventional superheterodyne receiver. None have been realised in silicon to date because sample-and-hold circuits with a sufficiently quick switching rate may not be realised in CMOS owing to the inadequate f_t of currently available MOSFETs (silicon MOSFETs would need a channel length of < 0.25 μm to be acceptable [Sheng 94]). The HIGFETs found in 0.75μm CGaAs™ circuits have an f_t of 20 GHz, comfortably exceeding that required for direct RF down conversion using an under sampled ADC.

The overall system characteristics, such as receiver/transmitter antenna requirements, RF coding and modulation schemes, analogue to digital performance requirements and the digital signal processing architecture will first be defined, as a future task. The design of a patch antenna for reception and transmission using a microstrip structure will be carried out. An analogue-to-digital converter specifically tailored for the transmitted signal coding and modulation schemes will then be effected. A high level simulation of the digital signal processing used to reconstitute the modulating signal waveform will need to be be carried out to confirm acceptable system performance before detailed design of the CGaAs™ circuits.

7.4 Case 3: The Motion Detector Paradigm

In this section we describe a novel motion detector, useful for robotic collision avoidance, inspired by the physiology of insect vision. A proof-of-concept in CMOS is reviewed first and then we discuss the advantages of a future GaAs implementation. As the VLSI implementation of insect vision only requires coarsely thresholded data, this creates an ideal application for the 'photovoltaic self-biasing edge-effect' that we uncovered in Chapter 3. Use of the 'edge-effect' would enable the realisation of a motion detector with up to a factor of 10 increase photosensitivity.

7.4.1 Introduction

An analog VLSI implementation of a smart microsensor that mimics the early visual processing stage in insects is described with an emphasis on the overall concept and the front-end detection. The system employs the 'smart sensor' paradigm in that the detectors and processing circuitry are integrated on the one chip. The integrated circuit is composed of sixty channels of photodetectors and parallel processing elements. The photodetection circuitry includes p-well junction diodes on a 2μm CMOS process and a logarithmic compression to increase the dynamic range of the system. The future possibility of gallium arsenide implementation is discussed. The processing elements behind each photodetector contain a low frequency differentiator where subthreshold design methods have been used. The completed IC is ideal for motion detection, particularly collision avoidance tasks, as it essentially detects time-to-impact, speed & bearing of an object. The Horridge Template Model [Horridge 91a] for insect vision has been directly mapped into VLSI and therefore the IC truly exploits the beauty of nature in that the insect eye is so compact with parallel processing, enabling compact motion detection without the computational overhead of intensive imaging, full image extraction and interpretation. This world-first has exciting applications in the areas of automobile anti-collision, IVHS, autonomous robot guidance, aids for the blind, continuous process monitoring/web inspection and automated welding, for example.

The Intelligent Vehicle Highway System (IVHS) is essentially a marriage between an electronic highway management system and cars with smart sensors, instrumentation & communication devices, to reduce traffic congestion and increase safety. In a number of countries, IVHS programs have built up considerable momentum and there are now a number of Australian initiatives [Longfoot 91]. The social benefits of this vision are well documented [Stafford 90] and include reduced traffic congestion, fuel consumption, road accidents and more efficient travel.

The developed insect vision sensor has a number of possible automotive applications, as shown in Table 7.1, and can either be used stand-alone or can be integrated as part of a future IVHS system. The sensor is essentially a smart motion detector, based on optical flow, and in an IVHS environment would be expected to feedback hazard information such a frequency of blind-spot hits, white line crossing hits, poor braking distance hits etc., to the central control hub, for identification of high-risk zones.

Conventional imaging devices and associated image processing have yielded only limited success in mobile real-time applications, due to the vast amounts of visual data and bulk of required processing hardware.

For collision avoidance tasks, where full imaging is unnecessary, simple motion detection of object boundaries offers an efficient solution. A world-first single-chip 4.5mm $\times$ 4.6mm device, based on insect vision principles, has been developed [Moini 93, Yakovleff 93, Bouzerdoum 93] in a 2 μm p-well CMOS process, that simply outputs the range, bearing and velocity of a detected object. The processing power of a commercial microcontroller is then sufficient for making decisions based only on such simple variables.

The chip accepts a real-time optical image and indicates the motion of edges in the visual field. From the outputs of the chip, we can infer the bearing, range, and speed of objects in the visual environment.

The advantage of a smart-sensor that can mimic insect vision is that the image processing is simplified and can be integrated on the detector chip, creating a compact device ideal for mobile applications. In addition, insects operate with no iris action or focusing adjustment required – this is also a feature of the developed sensor – leading to a truly solid-state vision system. Such simplicity is an important factor for high-volume automotive applications.

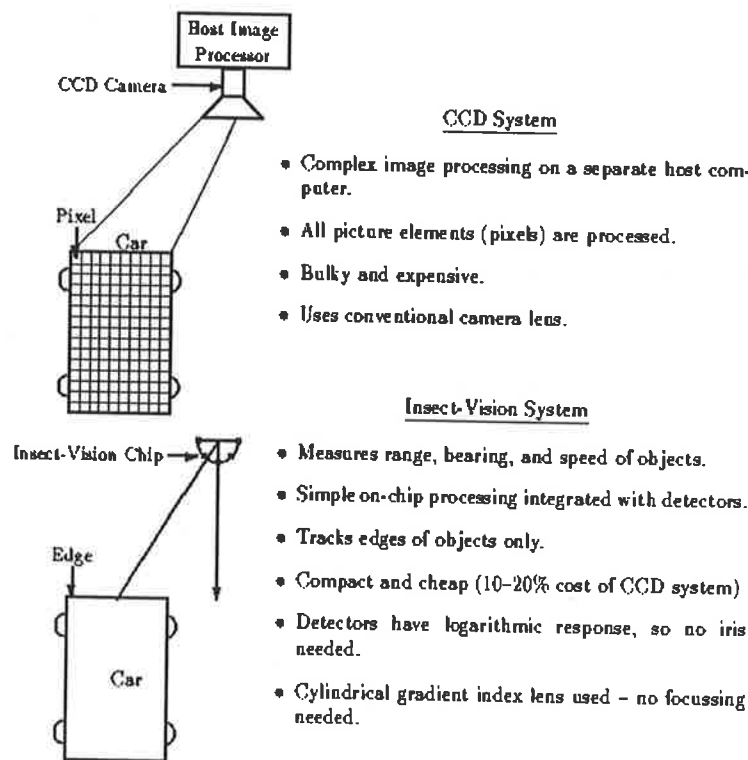


Figure 7.26: Insect vision concept compared to conventional CCD camera.

The concept and advantage of the system, in comparison to a conventional CCD camera is illustrated in Fig. 7.26, the compactness of insect vision being because it:

- Detects movement by optical flow
- Does not 'see' full image, just moving edges – hence small & compact
- No variable focusing needed
- No iris action needed.

The 'intelligence' part of the insect vision is independent of the technology used. So it can never be displaced or become obsolete as new technologies are discovered – in fact it can be used with any front-end technology:

- Optical (with silicon or gallium arsenide)
- Infrared (for night vision)
- Millimetre wave (excellent for poor weather conditions)
- Direct output from an electrical instrument (eg. ECG, echocardiogram, etc.).

Finally, a major advantage is that insect vision is multifunctional. In other words, on a car for instance, the very same chip design can do all the functions of braking distance, lane tracking, blind spot detection etc. Other competing non-insect detectors tend to be function specific and perform only one task.

The system is based on the use of *optical flow* to detect and locate obstacles. The optical flow information is derived via simple special purpose parallel processing combined with codified templates [Horridge 91a, Horridge 91b]. The system exploits this concept to produce a compact real-time motion & proximity detector.

Insects, compared to humans, possess a relatively simple visual system, yet are capable of performing complex visual tasks. The photoreceptors in the retina, sample the visual field and perform adaptation using a gain control mechanism for efficient operation in varying light conditions. The role of the large monopolar cells, the main output cells of the lamina, is coding of contrast to enable vision with large variation in background intensity.

Conventional robotics systems often utilise measurements such as range and velocity which require further interpretation in order to determine a course of action. In contrast, insect vision provides readily interpreted visual information, or 'percepts,' which are represented qualitatively (eg. obstacles are 'close,' 'moving fast,' or 'looming') instead of being

Area	Application	Description
Automotive	Blind spot detector	Sensor detects presence and *direction* of objects in blind-spot using *optical flow* techniques
	Braking distance warning	Time to impact is actually calculated using Hoyle's formula for looming objects
	White line tracking	Achieves lane tracking, for smart cruise control or swerving monitor under manual control
	Driver's head monitor	Monitors natural frequencies of of driver's head to detect alertness
	Automatic parking	In the future, a driver's key chain tag will remotely activate central locking, theft alarm system and auto parking sequence – sensors will guide against collision
	Auto road sign reading	If road signs are encoded with a coarse 'bar code,' this may become a viable low-cost technique
	IVHS sensors	Blind spot hits, critical braking distance hits etc. are sent to IVHS control hub – hit histograms will then identify hazardous zones
Physically handicapped	Blindness aid	Detector on walking stick gives supplementary feedback
	Wheelchair aid	Anticollision warning device
Domestic	Auto vacuum cleaner	Compact detectors allow vacuum cleaner to run autonomously
	Smart iron	As movement is detected, iron gives warning if it is motionless for too long, to prevent burning
	Security	Movement detectors for domestic security
	Electronic pet	Roams autonomously and can search & destroy unwanted pests – no mess, no chemicals and solar powered
Industrial	Autonomous robot	For hazardous areas or harsh environments
	Automated welding	Insect vision is ideal due to melting edges in continuous movement
	Process control	Continuous process monitoring/web inspection
Medical	Vital function monitor	Monitors waveforms through optical flow techniques
	Hospital robot	Autonomous slave

Table 7.1: Some possible insect vision technology application areas

expressed in terms of precise metric measurements. This implies that the approach to designing the control structure of an insect vision based system should differ from traditional computational schemes, where sensing and control are clearly distinct.

In fact, it appears that in some biological species, sensors and motor control are directly linked, at least at a low level, as exemplified by the insect's optomotor response [Reichardt 61]. For instance, the response observed in a number of insect species is generally in a different direction from that of the detected pattern motion [Lazzari 90, Borst 93], while guard bees seem to control a stable "hovering" position by responding to the small positional changes of a fixated pattern [Kelber 93]. Moreover, psychophysical evidence suggests that primates may extract heading direction and depth information simultaneously [Perrone 94].

Finally, many species, including humans, make use of the rate of expansion of an object relative to the visual receptor to estimate the time-to-impact. Mathematically, the time-to-impact is a function of the ratio of the angle subtended at the receptor by an object, to the rate of increase of that angle. Experimental studies (eg. [Reagan 78, Reagan 93]) suggest that channels of the visual pathways are sensitive to this 'looming' effect. Considering that neither the receptor's motion nor the distance between the observer and the object are known, the result is quite remarkable, as it can be utilised directly by collision avoidance mechanisms. However, it should be pointed out that it is unclear if the collision avoidance mechanism is itself triggered by the time-to-impact having decreased to a particular value. In fact, experimental studies on locusts show that the insect may alter its flight path when an obstacle subtends more than ten degrees of the field of view, apparently irrespective of the time-to-impact [Robertson 93].

7.4.2 Comparison with other approaches

The single-chip insect vision IC, for detecting motion, via realisation of the Horridge template model [Horridge 91a], is a world-first [Moini 93, Yakovleff 93, Bouzerdoum 93]. The use of insect vision principles enables a compact *smart-sensor* with the processing on the same chip as the detector array. Other workers have implemented other insect vision processing tasks, such as lateral inhibition [Darling 93], but these do not pertain to motion.

The work of John Tanner and Carver Mead at Caltech on the *correlating optical motion detector* [Tanner 84] in 1984, was inspirational and has led to several VLSI implementations of other motion detection algorithms. Unfortunately, most implementations were based on the gradient scheme [Tanner 86, Moore 91] and the Marr and Hildreth

zero-crossing detector [Bair 91], which are computationally intensive and inherently noise sensitive; both models are concerned with an accurate estimation of the velocity, which in turn requires an accurate estimate of the spatial and temporal derivatives of the intensity function. The result is that all the implementations were sensitive to light intensity; as the light intensity dropped the estimate of the velocity deteriorated quickly.

In regard to detectors on automobiles, other workers have gone from the solution of multiple CCD cameras [Mitsubishi] to discrete infrared detectors [Stuckman 89, Hyland]. The CCD camera approach is complex in that it extracts too much information for relatively simple tasks, such as lane sensing, which then has to be interpreted by extensive image recognition post-processing. On the other hand, the complexity problem is solved by the move towards simple discrete infrared detectors, for blind spot detection. However, as the principle is to coarsely detect hot car engines, this technology cannot be used for, say, white line sensing. Furthermore, high risk blind spot hazards of non-engine-hot objects such as pedal bikes must be identified from background clutter.

Our proposed insect vision motion detector is balanced between these two extreme ends. It is much more compact than the CCD approach and yet more visually discriminating than the discrete infrared approach, therefore represents the ideal automotive solution. [2] It also has the potential to perform multiple tasks such as lane sensing, blind spot detection and braking distance warning. For future research programs an infrared version of insect vision will be the next logical step, for poor driver visibility due to harsh weather conditions.

Microwave blind spot detectors are active systems requiring both transmitter and receiver, whereas insect vision is purely passive and therefore saves the need for a transmitter device. Millimetre wave passive detectors have been demonstrated [Lam 92], and the antennae are small enough to integrate on a GaAs insect vision chip. The drawback is the requirement for expensive 90 GHz technology and therefore is not an immediate commercial possibility. However, this option and realisation of this device is a logical step for a future research phase to improve harsh weather performance.

7.4.3 Motion Detector Chip

The developed chip contains a 1-D linear array of 60 photodetectors. Upon this foundation, the next stage will be to build a 2-D array chip. Several 2-D array chips can then

[2] For instance, as the insect vision chip is based on *optical flow* it can distinguish between the false alarm case of a parked car or a car passing in the opposite direction from the genuine case of a car approaching from the rear into the blind spot region.

be assembled to form a 180° or 360° near-omnidirectional sensor, for example. A possible scenario would be six chips in a hexagonal ring.

The design is a "smart" micro-sensor which accepts a time-varying pattern of stimuli and indicates the bearing, range, and speed of objects in the visual environment, ie. a sensor which identifies the 3-D arrangement of objects in space from motion rather than 3-D forms of the objects themselves from shading (which requires coding of relative intensities). Specifically the system is based on the use of *optical flow* to detect and locate obstacles. The optical flow information is derived via simple special purpose parallel processing combined with codified templates [Horridge 91a]. The system exploits this concept to produce a compact real-time motion & proximity detector.

Insects, compared to humans, possess a relatively simple visual system, yet are capable of performing complex visual tasks. The insect visual system has a highly parallel structure – the visual ganglia in the optic lobes are organised into columns and strata. The lamina is the first optic ganglion and and contains a large number of identical parallel channels

The photoreceptors in the retina sample the visual field and perform adaptation using a gain control mechanism for efficient operation in varying light conditions. The role of the large monopolar cells, the main output cells of the lamina, is coding of contrast to enable vision with large variation in background intensity.

The IC realisation is a linear array of 60 photodetectors with 60 channels of parallel analog differentiators. Processing at each node mimics the parallel structure of the insect visual system and provides improved edge discrimination in varying light conditions.

The light is focused onto the surface by a gradient index (GRIN) lens, with a visual angle of 72°. The lens has a 1.8 mm diameter and has been chosen due to it's flat surface that can be simply glued to the chip. The spacing between each of the 60 photodetectors has been designed so that each channel corresponds to a 1° angle, thus giving a total visual angle of 60°.

Photocurrent from the each detector is transformed into a voltage using a subthreshold circuit technique, thus providing a logarithmic response. This form of automatic gain control provides a wide dynamic range, as insects do not use a bulky iris mechanism.

The signal voltage is then buffered, time differentiated and thresholded to digitally 2-bit encode the detected light variations. These detected contrast changes are then multiplexed in pairs, from adjacent channels, together with the previous sample time states of the same pair, to form a Horridge template. The way Horridge templates are formed is illustrated in Fig. 7.27, where (↑) indicates an increasing contrast change, (↓) indicates a decreasing contrast change and (–) indicates no change at each photodetector – these

states are determined from one temporal sample to another. The temporal sampling time is adjustable, but is typically 10-20 ms.

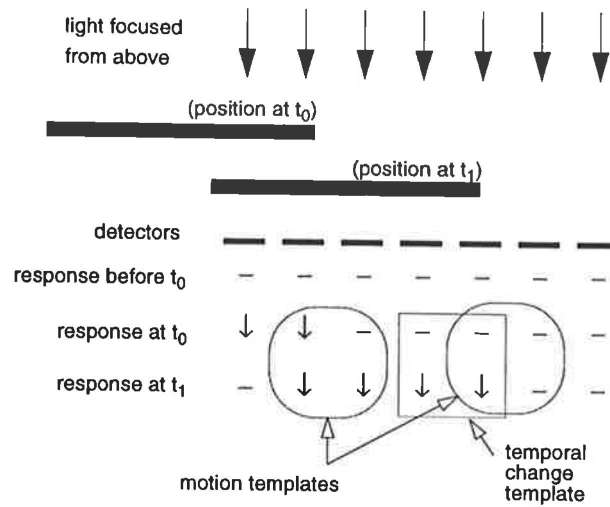


Figure 7.27: Coding of templates

To \ T1	– –	↓ –	↓ ↓	↓ ↑	– ↑	↑ ↑	↑ –	↑ ↓	– ↓
– –		←			→		←		→
↓ –	→		←						
↓ ↓		→		→				←	←
↓ ↑			←			→			
– ↑	←					→			
↑ ↑				←	←		→	→	
↑ –	→					←			
↑ ↓			→			←			
– ↓	←		→						

Figure 7.28: Template table.

Interpretation of the template is then found by a lookup table, see Table 7.28, stored in RAM and the outputs are then stored in an intermediate RAM. The template codes themselves provide information on object direction, whereas the orientation of a group of templates on a spatio-temporal plane provides information on angular velocity and

bearing. Each location in the template table is labelled with an arbitrary ascii character (`0123456789ABCDEFG....`) and a typical output from the chip is shown in Fig. 7.29 – where a '.' indicates no motion detected. Fig. 7.29 shows the motion due to two objects – where one object clearly changes direction twice.

```
...........................7.AAEFFFF60002..5......
.......................7AEFFFFF602......73........
..................7AAAAEFFF6F002........C4..4.....
...........7.AAAAA00000000020.........7AE4..4.....
.......7AAAEAF60002...................CFF4..4.....
....7AAEFF60F02......................7EF62..2.....
.7AEFFF602..2.......................CFF62.........
.C600002...........................7EFF4..........
.52...............................7EFF62..........
.................................7EFF62...........
................................7EF602............
.......7AAAAA..................7EF62..............
.......50DFFBAA3...............CF62...............
.........500DFFBAAA3..........7EF4................
............5000DFFBAAA3.....7EF62................
.................500DFFFBAAAAEFF4.................
......................50DFFFFFF62.................
........................500D0FFBAAAA3.............
..........................7E.00000002.7AAAAAAAA3..
.........................CF4............50DFFFFFF4
........................7E62E.............50.DFFF4
.......................7EF4..................50022
......................7EF4........................
.....................7EF62..2..............7.F4...
.....................CF62...............7AAEFF4...
...................7A002...........7AAAA0000002...
...................C62.........7AEFFF602..........
..................702......7AAAEF60002............
.................7E4...7AAAEA00002................
.................CFBAAAE60000.....................
```

Figure 7.29: Example template output pattern. X-axis represents bearing – each sample is 1°. Y-axis represents time – each sample is 20 ms.

Angular velocity is given by the inverse of the gradient of the output template pattern, Figure 7.30 and angular direction can be inferred from the fact that each horizontal location corresponds to 1° of visual angle.

Looming objects can be detected by exploiting the fact that moving edges are of identical polarity. For instance, if the object is dark with respect to the background, the motion templates detected to the left and right would indicate bright-to-dark changes in contrast. Fig. 7.31 shows an example of a looming picture – the angular velocity of the edge on the right hand side is, on average, slower than the edge on the right.

The detection of looming illustrates one example of interpreting velocity measurements by taking into account the polarity information conveyed by the templates. In general, if an object's contrast with respect to the background is reasonably constant, and if the object is wide enough or close enough to the sensor, relative motion should elicit the detection of at least two edges.

Table 7.2 indicates the chronological development and future vision for our insect vision chip dubbed the 'bugeye.' Bugeye I of 1992 vintage was the first design for proof-of-concept and it contained both analog and digital circuitry. The signal was differentiated for detecting changes in contrast. It successfully operated with bearing, velocity and time-to-impact being successfully extracted from the template output. However, the drawback was that it only worked under DC light sources. The Bugeye II redesign produced a chip

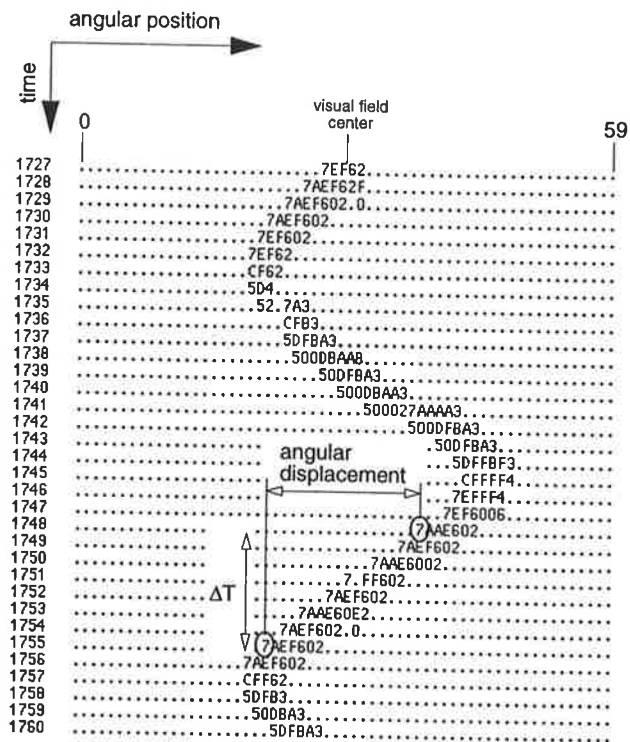


Figure 7.30: Extraction of angular velocity from a template pattern.

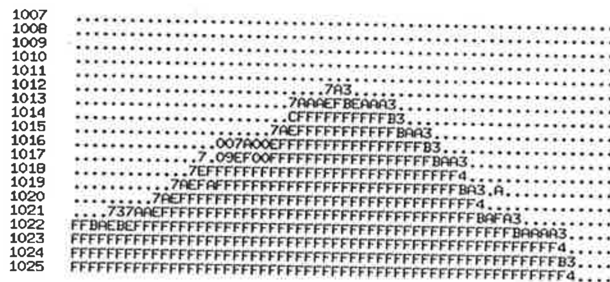


Figure 7.31: Looming pattern.

with improved dynamic range and contained a multiplicative noise cancellation circuit (MNC) [Moini 95]. The digital sections of the chip were discarded in favour of an external microcontroller. The MNC circuit allowed the chip to successfully operate under AC lighting conditions. The principle of MNC is to simply divide the signal in each channel by the spatial average over a number of channels. The circuit is designed so the averages over 3, 5 or 7 channels can be externally selected. As the detected signal luminance L is simply a function of the reflectance of an object ρ *times* the illuminance of the incident light E, division by the spatial average cancels the E terms (containing the unwanted AC noise component), resulting in a simple ratio of reflectances or *contrast ratio.* This has three benefits: (1) reduction in the effect of the 50 Hz or 60 Hz hum from AC light sources, (2) a data compression due a simple contrast ratio figure producing numbers close to unity and (3) and edge enhancement due to a reduction in spatial average near the edges.

The latest Bugeye III design contains a truly 2 dimensional array of detectors. AGC occurs at every node by virtue of logarithmic compression due to subthreshold detector circuits, however contrast change detection is carried out by only one row of differentiators. This is achieved by clocking signal out of the device as in conventional 2D arrays.

Bugeye IV will be an array of a larger size and perhaps exploit the qualities BiCMOS, being an analog device. Bugeye V will be a proof-of-concept in GaAs. This can be in a conventional E/D MESFET process or in the newly emerging complimentary CGaAs process.

Version	**Size**	**Technology**	**Year**
Bugeye I	64 by 12	2μm CMOS	1992
Bugeye II	64 by 2	1.2μm CMOS	1994
Bugeye III	64 by 32	0.8μm CMOS	1996
Bugeye IV	64 by 64	BiCMOS	1997
Bugeye V	64 by 64	GaAs or CGaAs	1998

Table 7.2: Evolution of 'Bugeye' Insect Vision Chip

7.4.4 Range Estimation

From Figure 7.4.4, it can be seen that $\theta \approx \phi$ and $R\frac{d\theta}{dx} = \sin\theta$, where R is the range. As the detectors are moving at a velocity V, the angular velocity is given by,

$$\dot{\theta} = \frac{d\theta}{dt} = \frac{d\theta}{dx}\frac{dx}{dt} = \frac{V\sin\theta}{R}$$

this formula was first considered in the context of aircraft navigation by Whiteside and Samuel [Whiteside 70].

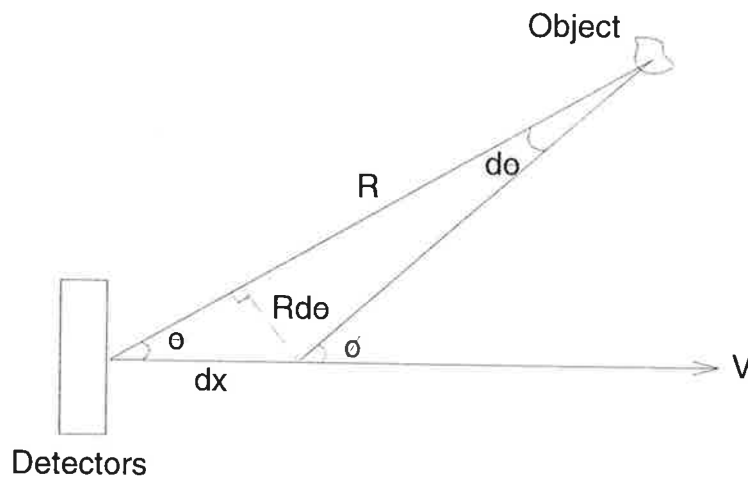


Figure 7.32: Calculation for range

In the context of in-lane traffic, however, the range of a looming object is the relevant quantity for braking distance warning. In this case the range is given by,

$$R = \frac{V}{\dot{\theta}} \sin\theta \cos\theta = \frac{V}{\dot{\theta}} \frac{\sin 2\theta}{2}$$

where V is the velocity of the detectors, θ is half the angle subtended by the car in front and $\dot{\theta}$ is the apparent angular velocity of it's rear wing tip. This formula assumes that the car ahead is stationary – if it is moving, it's velocity must also be known to calculate the range. This must be calculated by some other means or alternatively if *relative* velocity is considered, then 'time-to-impact' t can be calculated instead which is also a useful quantity,

$$t = \frac{\sin 2\theta}{2\dot{\theta}}$$

$\dot{\theta}$ is simply given by the inverse of the slope of the template pattern. As each template corresponds to a visual angle of 1°, then θ is simply half the number of template channels subtended by the object. Notice for small θ, the formula reduces to $t = \theta/\dot{\theta}$ which is the more usual form of Hoyle's formula used in the context of looming astronomical bodies [Hoyle 57].

7.4.4.1 Gallium Arsenide Implementation

The excellent progress and maturity of the Gallium Arsenide technology has lead to the development of high-speed systems, with as many as 1.44 million GaAs transistors per chip, suitable for working in harsh environments. Therefore, one of the next goals is to create a significant advance in the area of solid-state vision systems via the research of a high-speed Gallium Arsenide (GaAs) version of the chip. We propose a GaAs sensor, to be fabricated in a standard or near-standard GaAs digital IC process. The superior power-delay performance of GaAs over silicon and the excellent ongoing progress in GaAs fabrication maturity make GaAs a favourable choice [Moini 93, Darling 93]. The short absorption lengths and short carrier diffusion lengths in GaAs enable high sensitivity whilst maintaining low detector crosstalk and improved performance during optical overload (blooming).

Presently, the chosen 2μm CMOS process contains p-wells on an n-epi layer. The p-well/n-epi junction was chosen for the photodetection, rather than the n^+/p-well junction, because of greater quantum efficiency (QE) due to the greater depth of epi over well. The choice of a n-substrate process, over a p-substrate, was not critical for photodetection as the minority carrier diffusion lengths are much greater than the epi depth, in each case. The choice of epi over bulk is essential due to improved modulation transfer function (MTF), which implies a greater ability to detect higher spatial frequency detail. This is illustrated in Figure 7.4.4.1, where although the diffusion curve for 550 nm in bulk is acceptable, the MTF for 800 nm is severely degraded in bulk compared to epi. Although the diffusion curves are obtained from the well known function of quantum efficiency, the expression for QE in epi [Abbott 91] or Appendix B can be readily generalised for QE in a finite slab. The geometrical-only curve is obtained from the usual sinc function expression and represents the ideal case in absence of diffusion effects – we speculate that this could be achieved in GaAs where the hi-lo n^+/SI junction collects majority carriers, hence any diffusion terms vanish. Using the usual Planck's Law curve formulation we estimate the responsivity of our p-well/n-epi junction to be 53 mA/W, whereas for the GaAs case we obtain 71 mA/W. Therefore, in practice, for a scene illumination in the range 10^{-3} to 10^2 W/m^2, we expect a signal current from hundreds of femtoamps to tens of nanoamps, for our silicon element and slightly higher results for GaAs. For future work, GaAs may be promising for cases where demands are made on increasing the processing speed, whilst maintaining low power for mobile applications – as a by-product we would expect an improved MTF and responsivity.

The advantages of using gallium arsenide include:

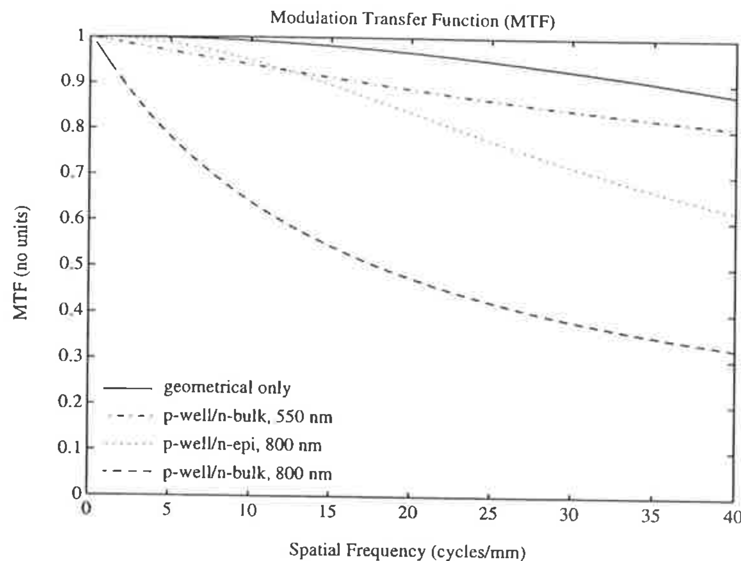


Figure 7.33: Theoretical MTF versus spatial frequency. Diffusion-only curves are for specific optical wavelengths. Geometrical-only curve is for the actual 7/20 aperture to pitch ratio used.

- GaAs has high speed for low power dissipation and is the choice technology for mobile communications, mobile global positioning systems (GPS) etc. – hence a GaAs insect vision chip will integrate well with this technology.

- GaAs high speed is also useful for digital circuitry that may be useful to integrate on the same chip, such as a high-speed customised Hough transform engine.

- GaAs chips actually have fewer manufacturing steps (about half) than silicon CMOS chips – hence at full maturity will be more reliable, have better yield and may even be cheaper.

- In GaAs, light is absorbed ten times closer to the surface, than in silicon. Therefore, light is more efficiently detected. Also if there is too much light overload (called 'blooming') this is better controlled in GaAs.

- The surface structure of GaAs is more suited for sandwiching together with an important infrared sensitive material called Cadmium Mercury Telluride (CMT). This will enable insect eyes to see in the dark.

- We discovered for the first time that GaAs transistors are around 10 times more sensitive to light at the edges [Abbott 91]. This is called the photovoltaic self-biasing edge-effect. This will enable insect eyes that work in very poor light conditions.

- GaAs is the choice material for the integration of millimeter wave microantenna structures [Parfitt 92]. This has two exciting possibilities: (1) mm wave antennas can replace the photodetectors to produce excellent performance in poor weather, (2) for an automobile, or any system, that uses multiple insect vision chips, wireless interchip communication can take place.

A world-first VLSI 4.5mm × 4.6mm CMOS integrated circuit has been developed that mimics the compact and yet powerful visual capabilities of an insect. The chip contains a linear array of 60 photodetectors. Future work will be to extend the present concept to a 2-D array and to also exploit the many advantages of a gallium arsenide implementation. The entomologist Thomas Eisner [Naumann94] said "Bugs are not going to inherit the earth – they own it now. So we might as well make our peace with the landlord." To this we can add that there is still much to be learned from the landlord.

7.5 Case 4: The Spatial Light Modulator Paradigm

A spatial light modulator (SLM) is essentially an image processor that outputs the data directly onto an integrated liquid crystal (LC) or a LED array display. The SLM can be either electrically addressed (EASLM) or optically addressed (OASLM). The GaAs imager we have developed provides an ideal optical input device for an OASLM and this aspect will be reviewed in this section.

There are many exciting applications of SLMs in combination with neural network (NN) and optical neural network (ONN) techniques. The driving vision includes:

- *Image recognition systems in vehicles/aircraft:* detection of motion, shape, direction etc. (road signs, moving objects, targets etc.); chip-level processing plus NN techniques; fast (sub millisecond), compact, robust, and not computing intensive. Its simplest form would be one smart pixel device: photodetectors do the detection, in-plane VLSI interconnects and circuits do the NN processing, and LC modulators output a low resolution feature-extracted image. A more complex form would be to have an ONN implemented by microlenses and holograms.
- *Civilian security systems:* face, fingerprint, signature recognition; security camera (motion detection, speed, shape etc.); picture, image scrambling; compact, portable, low-cost, NN intelligent. For face and motion recognition etc, the smart pixel circuit and interconnects will do the processing, and LC modulators output some meaningful pattern. Again, this can be abutted to an ONN. For picture scrambling, output

of any template can be represented by an LC modulator, for example, then digitally encoded for transmission.

- *Production line inspection:* number, size, fault and other error detection; from IC visual inspection to food packaging; low-cost, flexible. The output highlights the edge of a faulty object.

- *Personal Communications - multimedia:* picture, face recording and transmission; hand-held electronic drawing pad, head-mount display etc; powerful information throughputs (10 Gbits/sec) – compact & portable; image compression/decompression done in smart pixels (this is HOT). A small personal comms unit may consist of image detectors, NN interconnects plus circuits – this is where compression can be implemented. Again, an ONN can be added to do more complex operations such as object extraction – ie. recognise a circle and transmit a couple of bits instead of a whole chain of pixels. The idea is, to intelligently extract useful information for transmission. A very fast EASLM is needed.

- *Network Communications — optical/optoelectronic:* high bandwidth, fast optical switch; intelligent packet switching (ATM packet formatted into an optical page); optical switching is imminent, and devices are badly needed.

- *Parallel Computing:* extension processor unit to fit into computers (mainframe, PC etc); implement NN, to obtain parallel and intelligent processing; can be digital or analog; hardware image de/compression.

- *General Purpose Machine – NN machine – not a conventional computer:* has NN architecture throughout, highly interconnected; basic processors being smart pixels, which can have specialised functions; fast GaAs technology, micro-optics, laserdiode arrays etc. A completely new beast. Picture the brain. We can build specialist modules of smart pixels plus ONN, and they carry out specialist tasks. Smart pixels are the neural cells. Each module implements some NN model, and it need not be big or complex.

- *Virtual Reality:* an EASLM can be driven by computer generated holographic data to realise 3D motion pictures.

- *Electrically Reconfigurable Hologram:* an EASLM can be used to generate reconfigurable holographic interconnects and/or ONNs or to create an optical anaolg of an EPLD.

- *Optical Memory:* an OASLM can be used for reading from/writing to solid-state optical memory storage.

Due to the vastness of this area it would be advantageous to investigate and implement ferroelectric liquid crystal (FLC) on silicon (Si) and Gallium Arsenide (GaAs) VLSI technologies for smart pixels. FLC and GaAs are emerging technologies which show tremendous promise towards future two/three dimensional optoelectronic devices and systems. Smart pixels offer excellent prospects for integrating optics and microelectronics, whereby parallelism, connectivity, complexity and functionality can be exploited to their full potential.

In order to demonstrate the advantages and feasibility of the technologies, a spatial light modulator (SLM) for vision applications, can be built, containing smart pixels which perform local electronic processing and can be interconnected by optics. Vision systems are good applications to demonstrate complexity issues and processing abilities. An optical neural network (ONN) can be studied using the SLM. An ONN can implement high-level functionalities to complement the front-end processing in the smart pixel, highlighting the complexity achievable by the integration of optics with electronics. The overall architecture is one that does not rely on conventional software or interface, but is entirely contained in the VLSI-optics hardware.

7.5.1 Significance

7.5.1.1 FLC/Si Technology

Integration of FLC on Si VLSI has opened up possibilities of various two-dimensional electrooptic devices which are suitable for systems demanding high performances. Fast switching times (up to 10 μs) and high spatial bandwidths (up to 320×240 pixels in 1 cm^2) are obtainable. It is therefore possible to achieve information throughputs on the order of Gbits per second per cm^2. This is likely to break new grounds in realising systems which demand complex interconnections operating at very high speeds. Such systems will find applications in advanced computing, machine intelligence and communications. Assembly of FLC/Si devices has been a major challenge. High quality devices require careful control of positioning, FLC material alignment, component flatness, uniformity etc. in a clean room environment. However, once the assembly processes have been developed, it will be easily incorporated in standard Si CMOS or BiCMOS processes. The introduction of smart pixel structures will create a new set of challenges. Uniformity in optical and electrical properties becomes critical, but may also be relaxed if implemented in a suitable system such as neural network.

7.5.1.2 GaAs Technology

GaAs offers two significant possibilities: an LED/GaAs or FLC/GaAs. The advantages of GaAs are numerous. Firstly the larger absorption coefficient in GaAs leads to improved optical addressing characteristics [Moini 93]. Secondly, FLC technology is sensitive to surface planarity and wafer warp as little as 1 μm – GaAs offers excellent planarity due to the absence of field oxide steps and a warp of less than 0.5 μm has been achieved. In silicon, for typical devices, the warp is greater than 1 μm leading to critical non-uniformities in the FLC layer. GaAs also offers the potential to move away from LC technology altogether as LED arrays can be monolithically integrated [Grot 94]. With the emergence of low-power complementary GaAs (CGaAs), already available with VLSI maturity, the traditional problems with power dissipation are no longer of concern. Other advantages of GaAs include no latchup, fewer masking layers (simpler) and existing techniques for through-substrate-vias (this has implications for future 3-D mounted SLMs).

GaAs has a number of advantages with respect to the smart pixel concept. Firstly in order to enable optical input, a 2D image sensor must be vertically stacked with parallel connections to the display device using, for example, flip-chip techniques. Fortunately, the GaAs crystal has excellent crystal properties enabling etching of through-substrate-vias with relative ease and this is a known technique in GaAs MMICs that use vertical connections to a backside groundplane. Smart pixel functions can be embedded either in the sensor or display device or both.

Another feature of GaAs is that wafer warp across a chip is superior to silicon and <1 μm. Silicon has the disadvantage that a warp >1 μm degrades the performance of the FLC display, creating a deleterious dispersion in phase across the area of the device. The digital GaAs process is also intrinsically more planar than CMOS (eg. no field oxide) and therefore variations in FLC thickness are further reduced.

Due to the improved power-delay product of GaAs over silicon, either throughput rate can be increased or, for the same speed, power dissipation is reduced. This becomes important for stacked devices as greater care over thermal management is required.

For non-coherent applications, an alternative to an FLC output device is the use of VCSEL arrays. GaAs offers the unique advantage of smart VCSELs as emitting diodes and processing at each pixel can be integrated in GaAs. This has been successfully achieved by other workers. This approach has the advantage that it removes the need for an external light source and contrast is improved.

For the 2D sensor array, GaAs has the advantage that absorption lengths are 10 times shallower in GaAs over silicon, leading to great spatial resolution and greater photocol-

lection efficiency and better control over blooming. Extremely short substrate carrier diffusion lengths in GaAs (<10 μm) also lead to greatly improved spatial resolution – minority carrier diffusion lengths in standard CMOS can be as greater than 200 μm.

Due to the intrinsic simplicity of GaAs, there are no problems with latchup, no substrate ties, no well ties etc., thus allowing more real estate for smart functions at each pixel node and the through-substrate-vias. Also GaAs has over ten fewer masking layers than standard CMOS and with the current push towards 6" GaAs wafers GaAs is becoming extremely cost competitive with CMOS – for research batches where a number of design iterations are mandatory, GaAs has a better turn around time and cost savings can be made via fewer mask layers.

A key future research task would be to evaluate a number of logic family approaches with respect to smart pixels. Normally, Direct Coupled FET Logic (DCFL) would be used because of its better power-delay characteristics. However, with DCFL low noise margins (approx. 130 mV) are the main drawback. This is due to the barrier height of the MESFET Schottky diode, which limits the output voltage swing. Other problems are the large sensitivity of gate delay and noise margin with fan-in, fan-out and load capacitances as well as the poor temperature stability of the basic gates. These facts have spurred an interest in other circuital solutions that increase the logic swing or alleviate strict process requirements altogether. One of these solutions, named Pseudo Current Mode Logic (PCML), is based on current mirrors and allows a significant enhancement of noise margin (around 350 mV) and low sensitivity with fan-in. Super-Buffered FET Logic (SBFL) is used when fan-out, and load capacitance impose stringent conditions. For those parts of the system where extremely high speed is a requirement, Source-Coupled FET Logic (SCFL) is the solution, although its use should be limited because of its high power dissipation as compared with the others. This leads to the development of a merged logic where different logic families are mixed depending on the requirements of the whole system in order to get full benefit of each circuital solution.

7.5.1.3 Devices

There has been considerable demand in two-dimensional devices exhibiting electrooptic effects. These devices are important for any future systems to capitalise on the parallelism and connectivity in optics. The ability of electronics to drive and to process compensates for the difficulties in manipulating optical signals. It is therefore advantageous to combine optics for connectivity, with electronics for processing. Spatial light modulators show promise to meet this demand. Current electrically addressed SLMs, however, suffer the

electrical bottleneck of matrix-addressing. Liquid Crystal Light Valves, being optically addressed, are free from this bottleneck, but their functionality is limited. It is therefore desirable to have pixels which receive and modulate light in parallel, and provide electronic functionalities which are flexible to design. This calls for a new device design paradigm: Smart Pixels. A smart pixel generally consists of an optically detector, an optical modulator and a coupling circuit. In this way, the smart pixel allows input and output to be in parallel, and is capable of local processing which is also in parallel. This type of smart pixel architecture is illustrated in Fig. 7.34. The nature of the local processing is

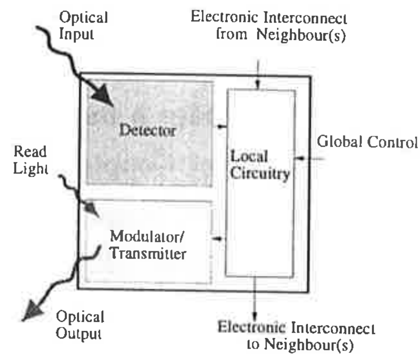


Figure 7.34: A Smart Pixel Architecture

significant. Each smart pixel is an individual processing unit, and hence both parallel and asynchronous in nature. If each smart pixel performs complex processing tasks, then the result is an array of asynchronous parallel processors. However, as indicated by neural networks, these individual units need not be complex, and yet high level operations can be performed when the units are interconnected to form a network. This is the approach to be adopted in this project. The smart pixel will perform low-level processing such as edge, texture and motion detection, noise reduction, light level adaptation etc. To complement this, an optical neural network implemented with these smart pixels will then perform high-level functionalities such as image recognition and transformation. Vision systems will be considered in this project, for they are good applications to demonstrate complexity and processing ability.

Smart pixels research has emerged rapidly in the past few years. Constrained as well as encouraged by the present VLSI technologies, devices are being demonstrated using Silicon related technologies (FLC/Si, deformable mirrors etc.) and GaAs based technologies (multiple quantum well modulators, Self Electrooptic Effect Devices and monolithic LEDs etc.). Most GaAs based demonstrators suffer from problems involving power dissipation, limited transistor circuits and low contrast ratios. However, the GaAs fabrication process

offers advantages such as high speeds, simpler steps, metal layers with higher flatness and good absorption of light in the visible spectrum. On the other hand, although FLC response times are relatively long, Silicon offers excellent circuit variety and range of components. Since this research group has expertise in both technology streams this project therefore provides a good platform to explore the possibilities available. It is conceivable, for instance, that the GaAs process is able to provide flatter reflective surfaces for FLC devices because of the absence of field oxide steps. It also allows higher modulator fill factor by having up to five metal layers. Integrating LEDs in the GaAs environment is another possibility. In the longer term, this investigation will enable this research group to evaluate three-dimensional optoelectronic devices. These may be in the form of a multi-chip module, which may allow chips to be bonded back-to-back so that the detector and the modulator are on opposite sides of the device. It is also possible to reduce the topographical undulation by building a stack of chips. These are not set as immediate options in this project, but have attractive possibilities for further advancement from the demonstrator SLM.

7.5.1.4 Optical Neural Network

Parallelism and Connectivity

Optics offers parallelism and high connectivity. These are the signatures of a sophisticated neural network. It has been established that optical signal processing has clear advantages over digital signal processing techniques in terms of speed and bandwidths. Optical correlators and neural networks have been demonstrated to a certain degree, but the electronic bottleneck and cumbersome off-line control remain apparent. It is clear that any system which includes electronic interface and software calculations detracts from the advantages in using optics. Therefore there is a need to find new architectures which preserve the parallelism and connectivity throughout the entire operation. This project proposes such a system architecture. The system to be investigated consists of an optical neural network (ONN) which is realised by Fourier optics in free-space. The dendrites and axons and their interconnections are defined by lenses and holograms. Neural network models such as Hopfield and various types of associative memories can then implemented. Armed with this ONN, the system therefore has the ability to perform difficult and abstract functions including recognition, association and other forms of high-level operations. Moreover, the ONN obtains the input information directly from the front-end smart pixels SLM. This is made possible by virtue of the smart pixel, which not only is able to detect images, but is also able to transmit/modulate light to reach the ONN. This whole operation again exhibits parallelism, and can be massively interconnected.

Partitioning of Functionality

Partitioning of functionality amongst layers or networks of neurons is important, and clearly evident in many biological nervous systems. Individual neurons can be specialised in certain operations (delay, differentiation, thresholding etc.), whilst interconnections can also be specialised or arranged in certain fashions (nearest neighbour interconnects, intra or inter-layer interconnects etc). Combinations of these enable various functionalities which can be simple, localised and low-level, or complex, extensive and high-level. In the proposed system, partitioning is achieved by assigning low-level functionalities to the smart pixels. These smart pixels will perform spatial or temporal differentiation or other localised operations. This is expected to significantly reduce the complexity and accuracy required of the ONN to subsequently process the information. The ONN then performs high-level functions such as recognition, association and transformation which require extensive connectivity. Hence the system architecture is efficient in engineering terms, and can be robust and flexible in application. There is an apparent mapping of this architecture to biological vision systems. The retina carries out optical detection, performs localised averaging, adapts to the light intensity, finds edges, responds to motion, sends off signals to the brain etc. These functions are reasonably local to the retina's cells, and can be implemented by VLSI circuitry. This is what the proposed SLM will be designed to do. Moreover, the brain analyses the transmitted information and attempts to recall memories, associate, recognise, manipulate, reiterate or discard the information. Some of these functions can be realised by the proposed ONN. It is hence clear that the overall concept is an autonomous vision architecture. This architecture does not rely on any external interface or software, but is entirely contained and self-sufficient in the VLSI-optics regime.

7.6 Techniques and Methods

7.6.1 FLC/Si Devices

Many properties of FLC/Si are advantageous for smart pixels devices. Switching speeds can be as fast as 10 μs. Contrast ratios are generally over 100:1. Most FLC materials show binary switching and are bistable [Wilkinson 94, Collings 89]. More advanced materials can be used to show analogue response in their Smectic A phase [Walba 93, Yu 95b]. These analogue materials are much sought after in applications like neural networks and fast display techniques, but are difficult to control and to maintain stability and temperature insensitivity. Both types of FLC materials will be considered in this project. The

operation principles of FLC are different from more commonly used nematic LCs, and indeed the FLC's response is 100 times faster. The FLC molecules orientate themselves according to the electric field present. The optic axis, due to the refractive index anisotropy, rotates with the molecules. The transmitted light's polarisation is therefore also rotated, and the effect is a rotatable halfwave plate. This gives rise to a change of phase in the light. This can be directly employed as phase modulation, or analysed by a polariser to give amplitude modulation. In a FLC/Si SLM (see Fig. 7.35) [Wilkinson 94, Collings 89], a thin film of FLC is placed between the Si backplane and a front glass electrode coated with ITO. A modulator metal pad is built on the backplane to act as a mirror. Voltages are then applied to this mirror to create an electric field across the FLC. This voltage can be determined by a digital driver if a binary operation is desired. Analogue circuitry can also be built to drive the mirror, which can act as a capacitive memory.

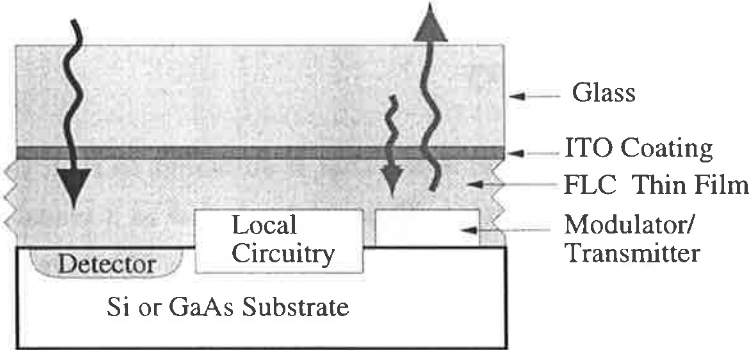


Figure 7.35: A FLC/Si or FLC/GaAs Smart Pixel

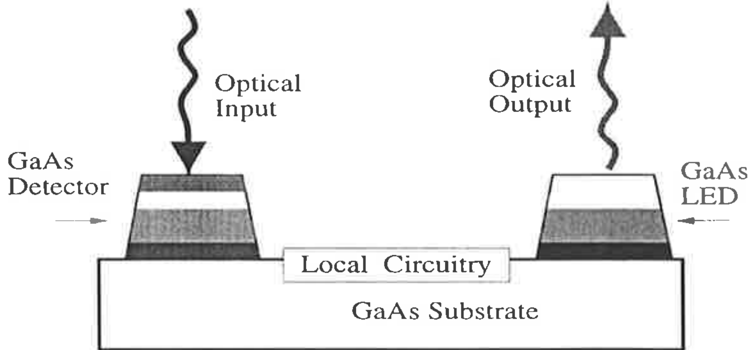


Figure 7.36: A LED/GaAs Smart Pixel

7.6.2 LED/GaAs and FLC/GaAs Devices

The maturity of GaAs is such that low-power complementary GaAs (CGaAs) is now a reality with VLSI integration levels. Transistor counts of a million transistors/chip are expected by the end of 1996. GaAs offers a more planar surface for the realisation of an FLC pixel, fast processing, no-latchup, greater absorption coefficient and simpler processing. State-of-the-art GaAs processing is now cheaper than BiCMOS. A further possibility is the monolithic integration of LED arrays, Fig. 7.36, which has been achieved [Kehrli 94, Grot 94] and offers a possibly more robust alternative to FLC and simpler optical systems. The power of each LED need not be high (100 nW is sufficient) – Ref. [Kehrli 94] showed 90 μW and a dissipation of 20 mW. The main research challenge with GaAs will be designing circuits tolerant to the small noise margins resulting from small signal swings.

7.6.3 Smart Pixel Circuitry

The minimum circuitry in a smart pixel is one that is sufficient to read the optical input and to drive the modulator or transmitter. This can be as few as 3 transistors in CMOS. Basic functionalities of a neuron have been attempted in two previous devices [Yu 95a, Yu 95b]. The smart pixel is ideal for performing low-level vision processing. An adaptive detector circuit can be built to enable a wide dynamic range [Moini 93]. Delay amongst neighbouring pixels can also be implemented. Spatial differentiation can be obtained by simple differentiators to enable edge detection or noise filtering. Simple buffers can also be built to provide short memory to connecting weights to realise cellular networks. To determine the circuit functionalities, detailed analysis and modelling will be carried out. This will include simulations of local and extensive neural networks, taking into account the role of partitioning between the SLM smart pixels and the ONN.

7.6.4 SLM and ONN Testing

In order to carry out electrical and optical testing of the SLM device, strategic test structures are needed to yield a comprehensive performance assessment. Preliminary optical systems are needed to characterise the optical performance and examine the feasibility for a full scale ONN. Realising an ONN is technically difficult, and addressing/interrogating smart pixels is not a trivial task. Certain ONNs (perceptron, Kohonen and Hopfield) have been investigated yielding useful results [Yu 95c, Yu 96, Collings 92, Collings 95]. The ONN we envisage will be simple. For a basic implementation, replication of the in-

put is obtained by use of holograms, and weight multiplication is achieved by amplitude modulation using an optical mask. Summation is performed either using the intrinsic integration of photons by photodiodes or phototransistors, or using lenses to control focusing. These different stages and components, to be used with the smart pixel SLM, need further investigation.

7.7 Summary

This chapter has ambitiously reviewed four main applications for this work. In the aerospace arena we reviewed a number of image processing functions to illustrate the kinds of processing that might be needed for future integration with the imager. We then focussed on the function of 'image registration' and discussed a design for an polynomial evaluator that could be used for this task. Although future work is required to develop an exact implementation, the polynomial evaluator provided a case study for the 'ring notation' design methodology outlined in the last chapter.

In the personal communication systems (PCS) arena we discussed some of the system level requirements such as compression, storage and encryption that would be required for a mobile phone augmented with interactive video using the GaAs imager. We defined a number of original paradigms, coining the terms: interactive telebanking (ITB), interactive personal navigation (IPN) and interactive time keeping (ITK). Often simply naming a paradigm is an important contribution that spurs further technological development.

In the motion detector arena, for collision avoidance tasks, we identified an architecture suitable for exploiting the 'edge-effect' gain mechanism uncovered in Chapter 3 to produce motion detectors with significantly increased photosensitivity. Even small variations in the edge-effect would not be tolerable for conventional imaging, because of fixed-pattern-noise (fpn). However, in the motion detector, based on insect vision, only coarsely thresholded data is required – thus making an ideal vehicle for demonstrating the advantages of the edge-effect.

In the spatial light modulator (SLM) and smart pixels arena we demonstrated that the optoelectronic and mechanical properties (warp etc.) of the GaAs imager were ideal, for the purposes of an optically addressed device. A rich selection of exciting SLM paradigms were defined and we showed that this opens up vast areas for future research.

The potential future research directions for the GaAs imager need not be limited to these 4 areas. Other areas that we have not even begun to discuss include HTDV sensors (due to the small diffusion length in GaAs), infrared arrays (either by hybrid techniques or

investigating new Schottky gate materials), X-ray detector arrays, particle detectors etc. The possibilities are staggering. These areas and the four, discussed in this chapter, will be recommended for future work in the next (concluding) chapter.

Bibliography

[Abbott 91] D. Abbott, S. Cui, K. Eshraghian & E. McCabe, "Photovoltaic gate biasing edge effect in GaAs MESFETs," *Electronics Letters,* **vol. 27**, no. 21, pp. 1900-1902, Oct. 1991.

[Abbott 93] D. Abbott & K. Eshraghian, "Gallium arsenide MESFET imager," *World Intellectual Property Organisation, International Bureau*, Geneva, **Patent WO 93/07643**, 15 Apr. 1993.

[Abbott 94] D. Abbott, A. Moini, A. Yakovleff, X.T. Nguyen, A. Blanksby, G. Kim, A. Bouzerdoum, R.E. Bogner & K. Eshraghian, "A new VLSI smart sensor for collision avoidance inspired by insect vision," *Proc. SPIE,* Boston, **vol. 2344**, pp. 105-115, Nov. 2-4, 1994.

[Abbott 95] D. Abbott, A. Yakovleff, A. Moini, X.T. Nguyen, A. Blanksby, R. Beare, A. Beaumont-Smith, G. Kim, A. Bouzerdoum, R.E. Bogner and K. Eshraghain, "Biologically inspired obstacle avoidance – a technology independent paradigm," *Proc. SPIE,* Philadelphia, **vol. 2591,** pp. 2-12.

[Ackland 94] B. Ackland, "The role of VLSI in multimedia," *IEEE Journal of Solid-State Circuits,* **vol. 29,** no. 4, pp. 381-388, 1994.

[Ackland 95] B. Ackland, "VLSI for multimedia applications," *Proc. 13th Australian Microelectronics Conference,* Adelaide, pp. 23-33, July 16-19, 1995.

[Bair 91] W. Bair and C. Koch, "Real-time motion detection using an analog VLSI zero-crossing chip," *Visual Information Processing: From Neurons to Chips*, Proc. SPIE, **vol. 1473,** pp. 59-65, 1991.

[Beadle 93] H.W.P. Beadle, "Multimedia communications and applications," *J. Electrical & Electronics Engineering, Australia – IE Aust. & IREE Aust.,* **vol. 13,** no. 3, pp. 181-190, 1993.

[Bernhardt 85] B. Bernhardt, et al., "Complementary GaAs (CGaAs): A high performance BiCMOS alternative," *IEEE GaAs IC Symposium,* San Diego, pp. 18-21, 1995.

[Borst 93] A. Borst, M. Egelhaaf and H.S. Seung "Two-dimensional Motion Perception in Flies," *Neural Computation*, **vol. 5**, pp. 856-868, 1993.

[Bouzerdoum 93] A. Bouzerdoum, R. E. Bogner, K. Eshraghian, A. Moini, A. Yakovleff & X. T. Nguyen, "A VLSI front-end visual processor based on insect vision," *Sensory Stratagems: From Living Eyes to Seeing Machines,* Canberra, 7–11 Feb. 1993.

[Bouzerdoum 94] A. Bouzerdoum, A. Moini, A. Yakovleff, X. T. Nguyen, R. E. Bogner, K. Eshraghian, "A smart visual micro-sensor," *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics,* 2–5 Oct. 1994.

[Collings 89] N. Collings, W.A. Crossland, P.J. Ayliffe, D.G. Vass and I. Underwood, "Evolutionary development of advanced liquid crystal light modulators," *Applied Optics,* **vol. 28,** no. 22, pp. 4740-4747, 1989.

[Collings 92] N. Collings, R. Sumi, K.J. Weible, B. Acklin and W. Xue, "The use of optical hardware to find good solutions of the Travelling Salesman Problem," *SPIE Optical Computing 92,* **vol. 1806,** pp. 637-641, 1992.

[Collings 95] N. Collings, A.R. Purzand and R. Völkel, "Construction of a programmable multiplier analogue neural network using space invariant interconnects," *SPIE,* **vol. 2565,** pp. 40-47, 1995.

[Darling 93] R. B. Darling and W. T. Dietze, "Implementation of multiplicative lateral inhibition in a GaAs sensory neural-network photodetector array," *IEEE J. Quantum Elec.,* **vol. 29,** no. 2, pp. 645-654, Feb. 1993.

[Grot 94] A.C. Grot and D. Psaltis, "Large Scale integration of LEDs and GaAs circuits fabricated through MOSIS," Tech. Digest, Optical Computing 94, MA2/3-4, 1994.

[Horridge 91a] G. A. Horridge, "Ratios of template responses as the basis of semivision," *Phil. Trans. Roy. Soc. London B,* **vol. 331,** pp. 189-197, 1991.

[Horridge 91b] G. A. Horridge, "The compromise between seeing spatial layout and making visual discriminations," *Current Science,* **vol. 60,** no. 12, pp. 686-693, June 1991.

[Hoyle 57] F. Hoyle, *The Black Cloud*, p. 20, Hinemann, 1957.

[Hyland] W. Hyland, "Sideminder: advanced blind spot monitoring," *Auto-Sense*, (company brochure), Denver, Colorado, USA.

[Kehrli 94] U. Kehrli, D. Leipold, K. Thelen, H.P. Schweizer, P. Seitz and B.D. Patterson, "Monolithically Integrated Optoelectronic Smart Pixels," Tech. Digest, *Optical Computing 94,* MC1/17-18, 1994.

[Kelber 93] A. Kelber and J. Zeil "A robust procedure for visual stabilisation of hovering flight position in guard bees of *Trigona (Tetragonisca) angustala* (Apidae, Meliponinae)," *Journal of Comparative Physiology A,* **vol. 167**, pp. 569-577, 1993.

[Kornerup 93] P. Kornerup, "High-radix modular multiplication for cryptosystems," *Proc. 11th IEEE Symp. Computer Arithmetic,* Windsor, Canada, June 1993.

[Lam 92] W. Lam *et al*, "Millimeter-wave imaging using preamplified diode detector," *IEEE Microwave Guided Wave Lett.,* **vol. 2,** No. 7, pp. 276-277, Jul. 1992.

[Lazzari 90] C. Lazzari and D. Varju "Visual lateral fixation and tracking in the haematophagous bug *Triatoma Infestans,*" *Journal of Comparative Physiology A*, **vol. 167**, pp. 527-531, 1990.

[Lin 95] H-D Lin, "Handling multimedia information with VLSI," *IEEE Circuits & Devices,* pp. 25-31, July 1995.

[Longfoot 91] J. E. Longfoot, "Australian initiatives and responses to IVHS," *Future Transportation Technology Conference, Portland, OR, USA,* SAE Technical Paper Series, pp. 61-66, Aug 5-7, 1991.

[McGeever 95] M.K. McGeever *et al,* "A 14kbit three transistor GaAs DRAM for an ATM switch," *Proc. 13th Australian Microelectronics Conference,* Adelaide, pp. 169-174, July 16-19, 1995.

[Meng 95] T.H. Meng *et al,* "Portable video-on-demand in wireless communication," *Proc. IEEE,* **vol. 83,** no. 4, pp. 659-680, 1995.

[Mitsubishi] "Mitsubishi HSR III," *Mitsubishi Motors Corporation,* (company brochure), Tokyo, Japan.

[Moini 93] A. Moini, K. Eshraghian, D. Abbott *et al*, "An analog implementation of early visual processing in insects," *International Symposium on VLSI Technology, Systems, and Applications,* Taipei, pp. 283-287, May 12-14, 1993.

[Moini 95] A. Moini, *et al*, "Multiplicative Noise Cancellation (MNC) in Analog VLSI Vision Sensors," *IEEE Proc. Elec. Tech. Directions to the Year 2000,* Adelaide, Australia, pp. 253-257, May 23-25, 1995.

[Montgomery 85] P.L. Montgomery, "Modular multiplication without trial division," *Mathematics of Computation,* **vol. 44,** pp. 519-521, 1985.

[Moore 91] A. Moore and C. Koch, "A multiplication based analog motion detection chip," *Visual Information Processing: From Neurons to Chips,* Proc. SPIE, **vol. 1473**, pp. 66-75, 1991.

[Moura 96] J.M.F. Moura *et al,* "Video over wireless," IEEE Personal Communications, pp. 44-54, Feb. 1996.

[Naumann94] I. D. Naumann, *Systematic and applied entomology,* Melbourne University Press, 1994.

[Nguyen 93a] X. T. Nguyen, A. Bouzerdoum, R.E. Bogner, K. Eshraghian, D. Abbott, A. Moini, "The stair-step tracking algorithm for velocity estimation," *ANZIIS-93,* Perth, pp. 412-416, 1–3 Dec., 1994.

[Nguyen 93b] X. T. Nguyen, K. Eshraghian, A. Moini, A. Bouzerdoum, A. Yakovleff, D. Abbott, R. E. Bogner, "An implementation of a smart visual micro-sensor based upon insect vision," *12th Australian Microelectronics Conference,* Queensland, pp. 129–134, 5–8 Oct., 1993.

[Nguyen 94a] X. T. Nguyen, A. Yakovleff, A. Moini, K. Eshraghian, A. Bouzerdoum, R. E. Bogner, "VLSI architecture of a low computational load processor for a visual system based on insect vision," *European design & test conf.,* pp. 85-89, 1994.

[Nguyen 94b] X. T. Nguyen, A. Bouzerdoum, A. Yakovleff, A. Moini, R. E. Bogner, K. Eshraghian, "VLSI robotic micro-sensor: range-finder from self motion," *IEEE Proc. mechatronics and machine vision in practice,* Toowoomba, pp. 78-83, 13–15 Sep., 1994.

[Parfitt 92] A. J. Parfitt, D. W. Griffin, P. H. Cole, "On the modelling of metal strip antennas contiguous with the edge of electrically thick finite size dielectric substrates," *IEEE Trans. Antennas and Propagation,* **vol. 40**, no. 2, Feb. 1992.

[Perrone 94] J.A. Perrone and L.S. Stone, "A Model of Self-Motion Estimation Within Primate Extrastriate Visual Cortex," *Vision Research*, **vol. 34**, no. 21, pp. 2917-2938, 1994.

[Reagan 78] D. Regan and K.I. Beverley, "Looming Detectors in the Human Visual Pathway," *Vision Research,* **vol. 18**, no. 4, pp. 415-421, 1978.

[Reagan 93] D. Regan and S.J. Hamstra "Dissociation of Discrimination Thresholds for Time to Contact and for Rate of Angular Expansion," *Vision Research,* **vol. 33,** no. 4, pp. 447-462, 1993.

[Reichardt 61] W. Reichardt "Autocorrelation, a Principle for the Evaluation of Sensory Information by the Central Nervous System," in *Sensory Communication,* Walter A. Rosenblith Ed., MIT Press and John Wiley & Sons (New York, London), pp. 303-317, 1961.

[Robertson 93] R.M. Robertson and A.G. Johnson "Collision Avoidance of Flying Locusts: Steering Torques and Behaviour," *Journal of Experimental Biology,* **vol. 183,** pp 35-60, 1993.

[Sheng 94] S. Sheng *et al,* "A monolithic CMOS radio system for wideband CDMA communications," *Proc. Wireless '94,* Calgary, Canada, June 1994.

[Stafford 90] F. P. Stafford, "Social benefits of IVHS systems," *SAE Transactions,* **vol. 99,** no. 6, pp. 1379-1384, 1990.

[Stuckman 89] B. E. Stuckman *et al*, "A solid state infrared device for detecting the presence of car in a driver's blind spot," *Proc. 32nd Midwest Symposium on Circuits and Systems,* Part 2, pp. 1185-1188, 1989.

[Tanner 84] J. Tanner and C. Mead, "A correlating optical motion detector," *Pro. Conference on Advanced Research in VLSI*, M.I.T., pp. 57-64, 1984.

[Tanner 86] J. Tanner and C. Mead, "An integrated analog optical motion sensor," In: *VLSI Signal Processing II*, pp. 59-76, IEEE Press, 1986.

[Wagnitz 95] D. Wagnitz, "Compressing images in real time on chip," *University of Adelaide: Internal Report,* 1995.

[Walba 93] D.M. Walba, D.J. Dyer, R. Shao, N.A. Clark, R.T. Vohra, K. More, W.N. Thurmes and M.D. Ward, "High performance electroclinic materials," *Ferroelectrics,* **vol. 148,** pp. 435-442, 1993.

[Whiteside 70] T. C. D. Whiteside and G. D. Samuel, "Blur zone," *Nature,* **vol. 225,** pp. 94-95, Jan. 1970.

[Wilkinson 94] T.D. Wilkinson, W.A. Crossland, S.T. Warr, T.C.B. Yu, A.B. Davey and R.J. Mears, "New applications for ferroelectric liquid crystals," *Liquid Crystals Today,* **vol. 4,** no. 3, pp. 1-6, 1994.

[Yakovleff 93] A. Yakovleff, A. Moini, A. Bouzerdoum, X. T. Nguyen, R. E. Bogner & K. Eshraghian, "A micro-sensor based on insect vision," submitted to the *Conference on Computer Architectures for Machine Perception (CAMP'93),* New Orleans, USA, 15–17 Dec. 1993.

[Yakovleff 94] A. Yakovleff, X. T. Nguyen, A. Bouzerdoum, A. Moini, R. E. Bogner, K. Eshraghian, "Dual-purpose interpretation of sensory information," *Proc. IEEE Int. Conf. Robotics and Automation,* San Diego, **vol. 2,** pp. 1635-1640, 8–13 May, 1994.

[Yakovleff 95] A. Yakovleff, D. Abbott, X.T. Nguyen & K. Eshraghian, "Obstacle avoidance and motion-induced navigation," *Proc. Comp. Arch. for Machine Perception (CAMP'95),* Como, Italy, pp. 384-393, 18-20 Sep. 1995.

[Yu 95a] T.C.B. Yu and R.J. Mears, A.B. Davey, W.A. Crossland, M.W.G. Snook, N. Collings and N. Birch, "Design of smart spatial light modulators for neural networks," *Journal of Optical Memory and Neural Networks,* **vol. 4,** no. 3, 1995.

[Yu 95b] T.C.B. Yu and R.J. Mears, A.B. Davey, W.A. Crossland, M.W.G. Snook, N. Collings and M. Birch, "Analogue optoelectronic implementation of a neural network with in-situ learning." *Proceedings of 1995 IEEE Int. Conf. on Neural Networks*, Perth, 1995.

[Yu 95c] T.C.B. Yu and N. Collings, "Optical neural systems using a smart pixel array device," *Internal Report of IMT, University of Neuchâtel (#395 DA 11/95),* 1995.

[Yu 96] T.C.B. Yu and N. Collings, "Optical Neural Systems with SASLM1: Initial Demonstrations," submitted to *Optical Computing 1996,* Sendai, 1996.

Chapter 8

Conclusions and Recommendations

"For this, indeed, is the true source of our ignorance – the fact that our knowledge can only be finite, while our ignorance must necessarily be infinite."
Karl Popper (1920-)
AUSTRIAN-BRITISH PHILOSOPHER OF SCIENCE

"Stupidity consists in wanting to come to a conclusion."
Elemire Zolla (1926-)
ITALIAN PHILOSOPHER AND ESSAYIST

"I never know what I say until I hear the response to it."
Norbert Wiener (1894-1964)
AMERICAN MATHEMATICIAN AND FATHER OF CYBERNETICS

"The future influences the present as much as the past."
Friedrich Nietzsche (1844-1900)
GERMAN PHILOSOPHER AND SCHOLAR

After dismissing a number of alternative technologies to silicon, we began by essentially comparing GaAs to silicon. GaAs was shown to be superior in a number of respects, thus underpinning our choice of investigating GaAs for an imager array. Initially, we saw the two most important characteristics of GaAs were that absorption coefficients were an order of magnitude shorter (leading to improved photocollection and blooming control) and that the power-delay product is increasingly becoming smaller, particularly with the emergence of CGaAs[1]. Small power-delay product is ideal for the vision of integrating high speed digital processing with the imager. Further to this we underscored the maturity of GaAs within its historical context and, for the first time, carefully clarified a number of difficult areas of consternation that are traditionally caused by obfuscation in the literature (eg. the GaAs semi-insulating substrate, hi-lo junctions on SI, physics of depletion regions on SI etc.).

Next we proceeded with an original examination of the Adibi model for photodetection in MESFETs. For the first time we extracted one of Adibi's substrate time constant parameters, using a modulated laser beam. The second substrate parameter requires future investigation with new equipment that can modulate the laser at higher frequencies (possibly several tens of MHz).

With monochrometer based measurements we established that surface recombination did not significantly affect the detection of wavelengths of interest. Also because of lack of knowledge of the physics of depletion regions in SI substrates, there was initial cause to question the photocollection capability of the channel/substrate depletion region. For the first time, we experimentally demonstrated, based on monochrometer measurements, that photocollection in this region did indeed take place. This is a desirable effect leading to a more efficient detector and, furthermore, it confirmed our theoretical discussion of depletion regions in SI substrates in this regard.

With a scanned focussed laser beam, we then discovered, for the first time, what we have named 'the photovoltaic self-biasing edge-effect.' This new effect is significant because, for applications not sensitive to fixed-pattern-noise (such as in coarse motion detectors), it promises up to a factor of ten increase in photosensitivity. The current working hypothesis for the explanation of this effect, of increased photogain at the transistor edges, is an increase in depletion region volume at the edges where (1) the gate depletion region (2) the channel/substrate depletion region and (3) the MIS gate overhang depletion region all converge. This is the first time that the convergence of these three depletion regions has been discussed or suggested and is consistent with the observation that the effect did not

[1]Power-delay product in CGaAs is presently around 0.1 μW/MHz/gate, whereas CMOS is about 4-5 times higher

occur in mesa MESFETs or in MESFETs with a p-buffer layer. Future work is required to demonstrate if this convergence produces any purely electrical parasitic effects. Also the effect of isolation implant-to-transistor spacing on the edge-effect, needs to be quantified. Future measurements with MESFETs from more vendors and with self-aligned MESFETs would also be valuable.

Another significant observation was that the decay of the photogain curve from the peak (at the transistor edge) to the substrate region was very short. From this decay length we were able to deduce to that the carrier diffusion length in SI GaAs is $< 10\ \mu m$. This short diffusion length is quite remarkable given that in silicon we typically get lengths 1-2 orders of magnitude higher. The combination of deep levels and the indirect GaAs bandgap leads to a much shorter carrier diffusion length than in silicon. The realisation of this fact is significant as it strongly points to GaAs as the ideal medium for HDTV image sensors, where good spatial resolution becomes of prime importance.

In the analysis of quantum efficiency in GaAs detectors, a significant side issue emerged. We found that the formulae for the quantum efficiency in a finite slab of material, given in the literature, are incorrect [2]. For the first time, we present the correct derivation and this can be found in Appendix B.

The next area of investigation was to examine issue of noise in the output circuit of the imager. As the whole topic of noise is poorly presented and obfuscated in the literature we firstly created a framework of basic theory and discussed a number of conundrums, paradoxes and difficulties that arise with noise theory. A number of original contributions were developed – as they are tangential to this work the interested reader is referred to the summary at the end of Chapter 4. With regards to the analysis of noise of the imager output circuit, we rejected the common misconception amongst imager designers that the gate capacitance of the output source follower should be tailored to equal the video line capacitance. Our improved analysis showed that the source follower could in fact be made much smaller with no degradation in noise performance. Specifically, with regard to GaAs design, our analysis unfortunately showed that MESFET gate leakage current lead to disastrous levels of shot noise. Our immediate recommendation is for an initial proof-of-concept GaAs MESFET imager to be fabricated without an output circuit and an off-chip output amplifier is to be used. Future work based on CGaAs HIGFETs will enable complete integration of the output circuit. Furthermore we showed that a future move to anisotype GaAs FETs is expected to produce a factor of 10 overall improvement in noise performance over silicon.

[2]Also we showed that some of the literature was incorrect for the QE formula in a semi-infinite slab

In considering the pixel design of the imager, we chose simple non-active pixels (ie. simple two terminal photodiode structures, rather than three terminal structures with internal gain) in order to minimise fixed-pattern-noise and other complex unknowns. An analysis of the various RC time constants for the various photocollecting depletion regions, showed that the semi-insulating substrate was fine for standard TV frequencies. Two types of pixel were eventually proposed (1) a pixel based on the source to substrate junction of a transistor (2) a similar pixel but with a gate diode in parallel. The second option has the advantage that a fingered gate diode can act as an antiblooming line, but it unfortunately reduces the pixel fill-factor. The relative performance of both approaches must be experimentally determined by future work, by realising the two different imagers.

For the column-to-video line analog parallel/serial multiplexer, single minimum sized transistors were found to be adequate in terms of speed performance and optimum in terms of minimised clock feedthrough spikes onto the video line. In terms of address circuitry for column and row select functions, a D flip-flop shift register was shown to have the required performance and in particular it operated with non-overlapping clocks. Non-overlapping clocks are a prerequisite of synchronous clock transient cancellation schemes. In this regard a design based on an edge-triggered flip-flop was rejected, as it could not work with non-overlapping clocks. However its mW/MHz rating was an order of magnitude lower and it took up less area and is thus a recommended approach for any future digital processing blocks that may require registers. Another approach based on an original non-ratioed bootstrapped logic (NRBL) was also rejected – unfortunately this register did not work due to MESFET gate leakage currents prematurely discharging the bootstrap capacitances. However as the power and area consumption is potentially very small for NRBL, this new approach is highly recommended for future work with HIGFETs (where gate leakage is dramatically reduced).

Having completed the analysis of the imager design, we then looked at the development of a design methodology for general digital processing blocks that could potentially be integrated with the imager. Beginning with an original discussion of parasitic oscillations in MESFETs we showed how this lead to a preferred bus layout style, which then lead to Eshraghian's systemised 'ring notation' which is basically a GaAs analog of the nMOS 'sticks' style. Various simulations then showed how this style lead to reduced area consumption and delay. Although this work is now largely obsolete as it has been superceded by 'modified ring notation' and 'unified ring notation' that offer even more efficient layouts, it nevertheless was part of the logical development path of this work and showed promise for ring notation in the early stages of our research.

All the work is finally crystalised in the penultimate chapter, where we examined four

potential application areas for the imager work. In the area of aerospace we reviewed a number of image processing algorithms and then developed a polynomial evaluator for the function of 'image registration.' Although future work is required to fully optimise the architecture further and develop the interface details to enable 'image registration,' the polynomial evaluator nevertheless served as an important case study clearly showing the application of the 'ring notation' design methodology.

The imager was also shown to be ideal for mobile video capture for personal communication systems (PCS) of the future. In this paradigm high speed compression, storage and encryption functions may be integrated with the imager. The HIGFET CGaAs technology was deemed necessary to obtain the best practical power-delay product.

A motion detector, based on insect vision principles, was reviewed and is ideal for demonstrating the 'photovoltaic self-biasing edge-effect.' The motion detector is ideal for collision avoidance tasks and uses merely coarsely thresholded data – thus is not sensitive to fixed-pattern noise. For this reason the 'edge-effect' is ideally demonstrated on this device, potentially giving rise to a motion detector with a factor of 10 increase in photosensitivity.

Finally the area of spatial light modulators (SLMs) was considered. For optically addressable SLMs (OASLMs), the optical and mechanical properties of the GaAs imager were shown to be ideal.

Future work need not be restricted to developing these four areas. The GaAs imager may prove advantageous in the areas of HDTV sensors, infrared detectors, X-ray detectors, particle detectors etc. – the possibilities are endless. This work has only scratched the surface in comparison with the opportunities it has laid open for future research with which to redefine reality.

Appendix A

Finite Difference Method

"I do not believe in mathematics."
Albert Einstein (1879-1955)
GERMAN-AMERICAN PHYSICIST

A.1 FDM Voltage Potential Solution

The one dimensional form of Poisson's equation is given by,

$$\frac{\partial^2 V}{\partial x^2} = -\frac{\rho}{\epsilon}. \tag{A.1}$$

Consider a solution divided into small increments h. A Taylor expansion of $V(x)$ gives,

$$V(x+h) \sim V(x) + h\frac{\partial V}{\partial x} + \frac{h^2}{2}\frac{\partial^2 V}{\partial x^2} \tag{A.2}$$

and,

$$V(x-h) \sim V(x) - h\frac{\partial V}{\partial x} + \frac{h^2}{2}\frac{\partial^2 V}{\partial x^2} \tag{A.3}$$

where eqn. A.2 is the forward difference equation and eqn. A.3 is the reverse difference equation. Adding these two together gives,

$$V(x+h) + V(x-h) = 2V(x) + h^2\frac{\partial^2 V}{\partial x^2} \tag{A.4}$$

rearranging,

$$\frac{\partial^2 V}{\partial x^2} = \frac{V(x+h) + V(x-h) - 2V(x)}{h^2} \tag{A.5}$$

and substituting this into eqn. A.1 gives,

$$V(x+h) - 2V(x) + V(x-h) = -\frac{\rho h^2}{\epsilon}. \tag{A.6}$$

The can be re-expressed in the following notation:

$$V_{n-1} - 2V_n + V_{n-1} = -\frac{\rho h^2}{\epsilon}. \tag{A.7}$$

A system of linear equations can be formed by substituting in various values of n. Provided that the initial, V_0, and final, V_n, boundary conditions are known, the equations can be solved using matrices. Matrix operations can be very conveniently performed using the commercial MATLAB software. The matrix equation can be written as,

$$
\begin{pmatrix}
-2 & 1 & 0 & 0 & . & . & 0 & 0 \\
1 & -2 & 1 & 0 & . & . & . & 0 \\
0 & 1 & -2 & . & . & . & . & . \\
0 & 0 & . & . & . & . & . & . \\
. & . & . & . & . & . & 0 & . \\
. & . & . & . & . & -2 & 1 & 0 \\
0 & . & . & . & . & 1 & -2 & 1 \\
0 & 0 & . & . & . & 0 & 1 & -2
\end{pmatrix}
\begin{pmatrix}
V_1 \\ V_2 \\ V_3 \\ . \\ . \\ . \\ V_{n-2} \\ V_{n-1}
\end{pmatrix}
=
\begin{pmatrix}
C - V_0 \\ C \\ C \\ . \\ . \\ . \\ C \\ C - V_n
\end{pmatrix}
\tag{A.8}
$$

where,

$$
C = \frac{\rho h^2}{\epsilon} \tag{A.9}
$$

and the above matrices of the form $[A][V] = [C]$ can be solved for $[V]$ in MATLAB with $[V] = [A]^{-1}[B]$.

A.2 FDM Electric Field Solution

Given that $\frac{\partial V}{\partial x} = -E$ we can substitute this into Poisson's equation giving,

$$
\frac{\partial E}{\partial x} = \frac{\rho}{\epsilon}. \tag{A.10}
$$

A Taylor expansion of $E(x)$ gives,

$$
E(x+h) \sim E(x) + h\frac{\partial E}{\partial x} \tag{A.11}
$$

and,

$$
E(x-h) \sim E(x) - h\frac{\partial E}{\partial x} \tag{A.12}
$$

where eqn. A.11 is the forward difference equation and eqn. A.12 is the reverse difference equation. Subtracting these two equations gives,

$$
E(x+h) - E(x-h) = 2h\frac{\partial V}{\partial x} \tag{A.13}
$$

rearranging,

$$
\frac{\partial E}{\partial x} = \frac{E(x+h) - E(x-h)}{2h} \tag{A.14}
$$

and substituting this into eqn. A.10 gives,

$$E(x+h) - E(x-h) = -\frac{2\rho h}{\epsilon}. \tag{A.15}$$

The can be re-expressed in the following notation:

$$E_{n+1} - E_{n-1} = -\frac{2\rho h}{\epsilon}. \tag{A.16}$$

A system of linear equations can be formed by substituting in various values of n. Provided that the initial, E_0, and final, E_n, boundary conditions are known, the equations can be solved using matrices. The matrix equation can be written as,

$$\begin{pmatrix} 0 & 1 & 0 & 0 & . & . & 0 & 0 \\ -1 & 0 & 1 & 0 & . & . & . & 0 \\ 0 & -1 & 0 & . & . & . & . & . \\ 0 & 0 & . & . & . & . & . & . \\ . & . & . & . & . & . & 0 & . \\ . & . & . & . & . & 0 & 1 & 0 \\ 0 & . & . & . & . & -1 & 0 & 1 \\ 0 & 0 & . & . & . & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ . \\ . \\ . \\ E_{n-2} \\ E_{n-1} \end{pmatrix} = \begin{pmatrix} C - E_0 \\ C \\ C \\ . \\ . \\ . \\ C \\ C - E_n \end{pmatrix} \tag{A.17}$$

where,

$$C = \frac{2\rho h}{\epsilon} \tag{A.18}$$

and the above matrices of the form $[A][E] = [C]$ can be solved for $[V]$ in MATLAB with $[E] = [A]^{-1}[B]$. However, leading diagonal of $[A]$ is all zeros which becomes problematic for finding the inverse of [A].

To get around this problem, the solution is to substitute A.10 into into the forward difference equation A.11, giving,

$$E(x+h) - E(x) = \frac{\rho h}{\epsilon} \tag{A.19}$$

or,

$$E_{n+1} - E_n = \frac{\rho h}{\epsilon} \tag{A.20}$$

which leads to a system of equations with the following matrix representation,

$$\begin{pmatrix} 1 & 0 & 0 & 0 & . & . & 0 & 0 \\ -1 & 1 & 0 & 0 & . & . & . & 0 \\ 0 & -1 & 1 & . & . & . & . & . \\ 0 & 0 & . & . & . & . & . & . \\ . & . & . & . & . & . & 0 & . \\ . & . & . & . & . & 1 & 0 & 0 \\ 0 & . & . & . & . & -1 & 1 & 0 \\ 0 & 0 & . & . & . & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ . \\ . \\ . \\ E_{n-2} \\ E_{n-1} \end{pmatrix} = \begin{pmatrix} C + E_0 \\ C \\ C \\ . \\ . \\ . \\ C \\ C \end{pmatrix} \quad \text{(A.21)}$$

where,

$$C = \frac{\rho h}{\epsilon} \quad \text{(A.22)}$$

and the solution is given by $[E] = [A]^{-1}[C]$. The original contribution is to clearly present the problem of the zero leading diagonal, as this is not clearly described in the literature.

Appendix B

Quantum Efficiency

"It is more important to have beauty in one's equations than to have them fit experiment."
Paul A.M. Dirac (1902-1984)
BRITISH QUANTUM PHYSICIST AND CO-INVENTOR OF FERMI-DIRAC STAISTICS

"God gave us the integers. All the rest is man's work."
Leopold Kronecker (1823-1891)
GERMAN MATHEMATICIAN

B.1 Quantum Efficiency in a Semi-Infinite Slab

Consider a semi-infinite slab of semiconductor material. Let the surface be at $x = 0$ and the backside be at $x = \infty$. The continuity equation balance for generation and recombination under steady-state is:

$$-D\nabla^2 n + \frac{n}{\tau} = G \tag{B.1}$$

where D is the diffusion constant, n is the minority carrier concentration in the undepleted region, τ is minority carrier lifetime and G the electron-hole pair generation rate. We shall use n to denote either n or p type carriers as this analysis is independent of whether electrons or holes are the minority carriers, so long as the correct coefficient values for holes or electrons are used in any particular instance. The generation rate G is given by,

$$G = \Phi \alpha e^{-\alpha x} \tag{B.2}$$

where Φ is the incident photon flux per unit area and α is the absorption coefficient. Combining the above equations with the expression for minority carrier diffusion length, $L_o = \sqrt{D\tau}$, we get,

$$\frac{d^2 n}{dx^2} - \frac{n}{L_o^2} = -\frac{1}{D}\Phi \alpha e^{-\alpha x}. \tag{B.3}$$

Using the solution of the form,

$$n = Ae^{-\alpha x} + Be^{-\frac{x}{L_o}} + Ce^{\frac{x}{L_o}} \tag{B.4}$$

we assert boundary conditions at the depletion edge $x = W$, $n = 0$ and at the slab backside $x = \infty$, $n = 0$, yielding

$$A = \frac{\alpha {L_o}^2 \Phi}{(1 - \alpha^2 L_o^2)D}, \quad B = -Ae^{(\frac{1}{L_o} - \alpha)W}, \quad C = 0.$$

The steady state carrier flux is given by,

$$J = \int_0^W G dx + D\frac{\partial n}{\partial x}\Big|_{x=W} \tag{B.5}$$

which becomes

$$J = \Phi(1 - e^{\alpha W}) - DA\alpha e^{-\alpha W} - \frac{DB}{L_o} e^{-\frac{W}{L_o}}$$

and by substituting in the expressions for A and B,

$$J = \Phi(1 - e^{\alpha W}) + \frac{\alpha L_o \Phi e^{\alpha W}}{1 + \alpha L_o}$$

hence, internal quantum efficiency which is given by, $\eta = J/\Phi$, becomes,

$$\eta = 1 - \frac{e^{-\alpha W}}{1 + \alpha L_o}. \tag{B.6}$$

B.2 Quantum Efficiency in a Finite Slab

Consider a slab of semiconductor material, with finite thickness, h. For mathematical convenience, let the surface be at $x = -W$, the depletion region edge be at $x = 0$ and the backside be at $x = h - W$. We can rewrite the continuity equation, in the last section, as

$$\frac{d^2 n}{dx^2} - \frac{n}{L_o^2} = -\frac{1}{D} \Phi \alpha e^{-\alpha(x+W)}. \tag{B.7}$$

Again, using the solution of the form,

$$n = Ae^{-\alpha x} + Be^{-\frac{x}{L_o}} + Ce^{\frac{x}{L_o}} \tag{B.8}$$

we assert boundary conditions at the depletion edge $x = 0$, $n = 0$ and at the slab backside $x = h - W$, $n = 0$, yielding

$$A = \frac{\alpha {L_o}^2 \Phi}{(1 - \alpha^2 L_o^2) D}, \quad B = -A \frac{1 - e^{-(h-W)(\alpha + \frac{1}{L_o})}}{1 - e^{-\frac{2(h-W)}{L_o}}}, \quad C = -A \frac{1 - e^{-(h-W)(\alpha - \frac{1}{L_o})}}{1 - e^{\frac{2(h-W)}{L_o}}},$$

The steady state carrier flux is given by,

$$J = \int_{-W}^{0} \Phi(1 - e^{-\alpha W}) + D \frac{\partial n}{\partial x} |_{x=0} \tag{B.9}$$

which becomes

$$J = \Phi(1 - e^{\alpha W}) - DA\alpha - \frac{DB}{L_o} + \frac{DC}{L_o}$$

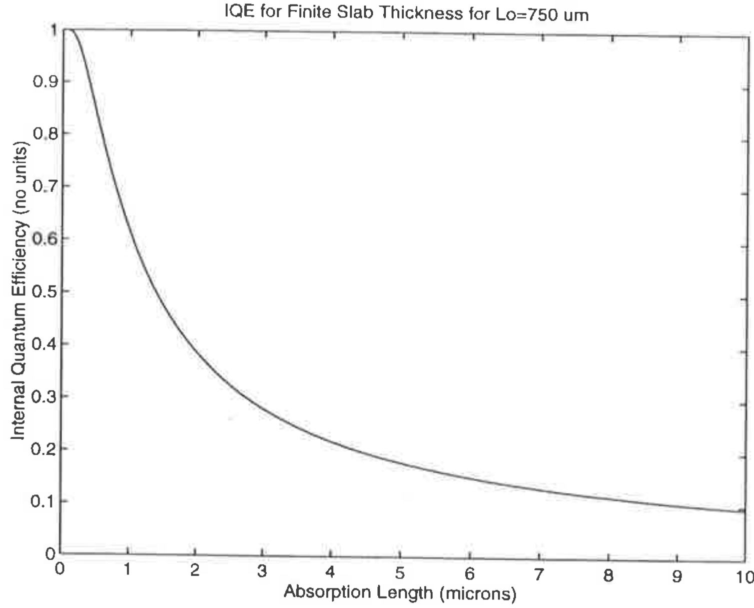


Figure B.1: Internal quantum efficiency versus absorption length.

and by substituting in the expressions for A, B and C, after some manipulation, using $\eta = J/\Phi$, we get,

$$\eta = 1 - \frac{e^{-\alpha W}}{1+\alpha L_o} - \frac{\alpha L_o e^{-\alpha W}}{\alpha^2 L_o^2 - 1} \frac{e^{-h/L_o} - e^{-h\alpha}}{\sinh(h/L_o)} \tag{B.10}$$

This equation shows that the quantum efficiency equals that of the semi-infinite case *plus* a (negative) second order term. It is quite remarkable and elegant that these two terms can be separately identified. Both this expression and the realisation that the semi-infinite case *plus* a second order term yields the finite slab case is an original contribution, and appears nowhere in the literature. Note that this expression has all the expected features. Each term is dimensionless – ie. each argument is a ratio of two lengths or an absorption coefficient times a length. This is an important feature that is missing in some erroneous attempts found in the literature. Also for $h \to \infty$ the expression reduces to the semi-infinite case.

The function is also well-behaved in that internal quantum efficiency (IQE) goes down, as expected, for increasing absorption length, Fig. B.1, and goes up for increasing diffusion length, Fig. B.2. Another important feature is that, by inspection, IQE always stays below unity – this is illustrated for specific examples in the graphs.

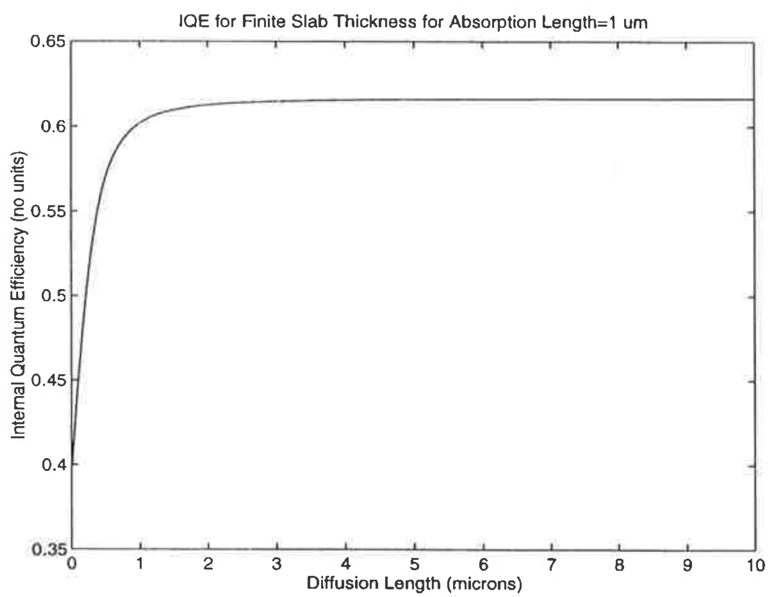


Figure B.2: Internal quantum efficiency versus absorption length.

Appendix C

Heat Equations

"As far as the laws of mathematics refer to reality they are not certain, and as far as they are certain they do not refer to reality."
Albert Einstein (1879-1955)
GERMAN-AMERICAN PHYSICIST

"The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we never can understand nor deserve."
Eugene Paul Wigner (1902-)
HUNGARIAN-AMERICAN PHYSICIST

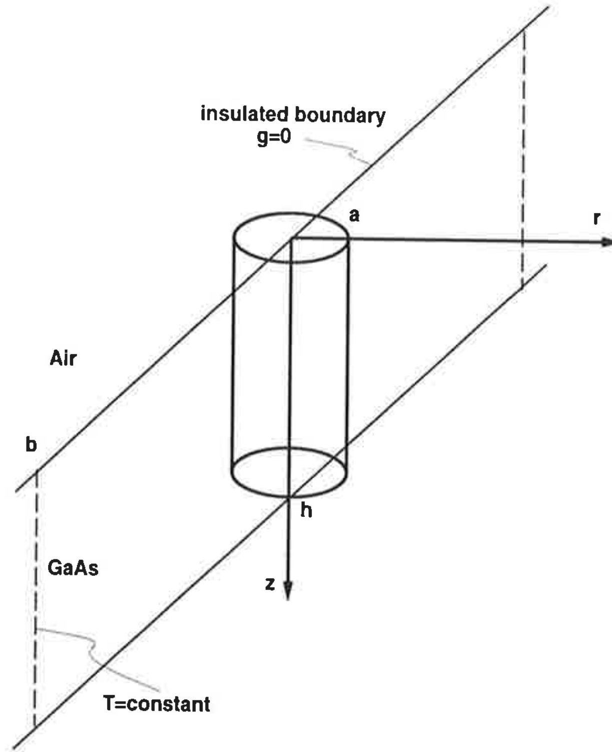


Figure C.1: Cylindrical coordinate system. a=radius of laser beam. b=wafer radius. h=wafer thickness.

The objective is to calculate the steady-state local temperature increase when a focussed laser beam of radius a impinges on a GaAs wafer of radius b and thickness h. We assume the mass surrounding the locally heated region acts as a good enough heat sink so that boundary conditions at $r = b$ and $z = h$ are at constant temperature. We assume that there is negligible heat loss by the top surface to the air and so $\partial T/\partial z = 0$ at $z = 0$.

The heat generation per unit volume is not constant but varies with z as, $g(z) \propto e^{-\alpha z}$ - where α is the absorption coefficient. If A is the constant of proportionality, then it is related to the incident power P by integrating over the volume of the cylinder,

$$P = A \int_0^h \int_0^{2\pi} \int_0^a e^{-\alpha z} r dr d\theta dz \tag{C.1}$$

which gives,

$$P = A \frac{a^2 \pi}{\alpha}(1 - e^{-\alpha h}) \tag{C.2}$$

but as $e^{-\alpha h} \ll 1$ the heat generation per unit volume can now be written as

$$g(z) = \frac{\alpha P}{a^2 \pi} e^{-\alpha z}. \tag{C.3}$$

Due to the temperature dependence of thermal conductivity on temperature given by the empirical expression $k(T) = 54.4 \times 10^3 T^{-1.2}$ W/mK we will invoke Kirchhoff's transformation. We must solve,

$$\nabla \cdot [k(T) \nabla T] + g(z) = 0 \tag{C.4}$$

and the dependent variable can be transformed by using the following substitution

$$U = \int_{T_o}^{T} \frac{k(T')}{k_o dT'} \tag{C.5}$$

where $k_o = k(T_o)$. To implement the substitution we can rewrite C.4 as

$$\nabla \cdot [k(T) \frac{\partial T}{\partial U} \nabla U] + g(z) = 0 \tag{C.6}$$

and from C.5

$$\frac{\partial U}{\partial T} = \frac{k(T)}{k_o} \tag{C.7}$$

therefore the master equation becomes

$$k_o \nabla^2 U + g = 0. \tag{C.8}$$

The solution is best found by conversion to cylindrical coordinates:

$$U = \sum_{n=1}^{\infty} [\frac{g_n}{k_n} + C I_o(kr)] \cos(k_n z) \tag{C.9}$$

where the coefficient C is given in terms of modified Bessel functions

$$C = \frac{g}{k^2} \frac{K_0(kb) I_1(ka) - I_0(kb) K_1(ka)}{I_0(kb)[I_0(ka) K_1(ka) - I_1(ka) K_o(ka)]} \tag{C.10}$$

and the coefficients k_n are given by

$$k_n = \frac{\pi(n - \frac{1}{2})}{h} \tag{C.11}$$

and the coefficients g_n are given by

$$g_n = \frac{2}{h}\frac{\alpha^2}{\alpha^2 + k_n^2}[\frac{k_n e^{\alpha h}(-1)^{n+1}}{\alpha^2} + \frac{1}{\alpha}]. \tag{C.12}$$

Finally temperature is obtained by invoking the Kirchhoff transformation and then computed using a MATLAB program. For example with $P = 0.25$ mW, a laser spot of radius $a = 1$ μm, a wafer thickness of $h = 150$ μm, a chip size of $b = 1$ cm and an absorption coefficient of 0.3×10^5 m^{-1}, the resulting local temperature increase comes to 100°C! This is rather large. Also it took MATLAB over 24 hrs run time to compute this! However, it should be noted the result is somewhat extreme as reflections have not been considered and heat is assumed to be all generated within the cylinder. Given that our main results are obtained with a laser power of $P = 1.4$ μW, we can safely assume, therefore, that there are no significant heating effects. Another factor leading to insignificant heating, is that the laser is automatically scanned across the surface of the chip, whereas the above calculations are for steady-state. The original contribution is that this analysis considers non-constant heat generation. In conclusion, we have established that laser heating effects can be ignored when considering mechanisms to explain the photovoltaic self-biasing edge-effect in GaAs MESFETs, provided the laser power is in the order of micro Watts.

Appendix D

Complex Integration

"The scientific mind does not so much provide the right answer as the right questions."
Claude Lévi-Strauss (1908-)
FRENCH SOCIAL ANTHROPOLOGIST

"Calculus is the art of numbering and measuring exactly a thing whose existence cannot be conceived."
François Marie Arouet de Voltaire (1694-1778)
FRENCH AUTHOR

D.1 General Complex Integral for Capacitor Problem

We need to solve the integral of the general form:

$$I = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \frac{a_0 + a_1 s + a_2 s^2}{(b_0 + b_1 s + b_2 s^2)(b_0 - b_1 s + b_2 s^2)} ds \tag{D.1}$$

Let $b_0 + b_1 s + b_2 s^2 = b_2(s - s_1)(s - s_2)$, so by factorising the denominator and taking a contour integral we have,

$$I = \frac{1}{2\pi j} \oint_C \frac{a_0 + a_1 s + a_2 s^2}{b_2^2 (s - s_1)(s - s_2)(s + s_1)(s + s_2)} ds. \tag{D.2}$$

Taking the sum of the residues,

$$I = \frac{a_2 s_1 s_2 (s_1 - s_2) - a_0 (s_1 - s_2)}{2 s_1 s_2 (s_1 - s_2)(s_1 + s_2) b_2^2} \tag{D.3}$$

and using $s_1 s_2 = b_0/b_2$ with $s_1 + s_2 = -b_1/b_2$, finally gives

$$I = \frac{a_0 b_2 - a_2 b_0}{2 b_0 b_1 b_2}. \tag{D.4}$$

This result is quite fascinating as the a_1 term has totally dropped out giving a purely real result. This can be explained because $a_1 s$ is an *odd* function of s.

Another curious matter is that because of the conjugation on the denominator of the original integral, the Cauchy-Riemann equations are *not* satisfied. However, the method of residuals happens to nevertheless work because the integral is essentially real given that the a_1 term drops out. As a precaution, we integrated the real part of the integral the long hand way and found that it did indeed reduce to the same result provided by the method of residues. Due to the great length of this procedure, this was accomplished using the MAPLE math editor software.

The fact that the method of residues is found to work on a non-analytic integral, is not discussed in the complex analysis literature. Hence this discovery and discussion is an original contribution to complex analysis. There is scope for future work to formally define a class of such integrals.

D.2 Noise Analysis of 2-Stage RC Ladder

From nodal analysis of the circuit we find that,

$$v_1 = \frac{e_1(1 + sC_2R_2) - e_2sC_2R_1}{1 + s(C_1R_1 + C_2R_2 + C_2R_1) + s^2C_1C_2R_1R_2} \tag{D.5}$$

and

$$v_2 = \frac{e_1 + e_2(1 + sC_1R_1)}{1 + s(C_1R_1 + C_2R_2 + C_2R_1) + s^2C_1C_2R_1R_2}. \tag{D.6}$$

Using $e_1 = 2kTR_1$ and $e_2 = 2kTR_2$ and multiplying by complex conjugates, gives us the spectral densities,

$$S_{11} = 2kT\frac{R_1|1 + sC_2R_2|^2 + R_2|sC_2R_1|^2}{|1 + s(C_1R_1 + C_2R_2 + C_2R_1) + s^2C_1C_2R_1R_2|^2} \tag{D.7}$$

$$S_{22} = 2kT\frac{R_1 + R_2|1 + sC_1R_1|^2}{|1 + s(C_1R_1 + C_2R_2 + C_2R_1) + s^2C_1C_2R_1R_2|^2} \tag{D.8}$$

$$S_{12} = 2kT\frac{R_1(1 + sC_2R_2) - sC_2R_1R_2(1 - sC_1R_1)}{|1 + s(C_1R_1 + C_2R_2 + C_2R_1) + s^2C_1C_2R_1R_2|^2}. \tag{D.9}$$

These spectral densities are now integrated using the general solution given in the last section. This yields the noise voltages in Volts squared per Hertz, and the integrals simply reduce to

$$\overline{v_1^2} = \frac{kT}{C_1}, \quad \overline{v_2^2} = \frac{kT}{C_2}, \quad \overline{v_1v_2} = 0$$

but if $R_2 \to 0$,

$$S_{11} = S_{22} = S_{21} = \frac{2kTR_1}{|1 + s(C_1R_1 + c_2R_1)|^2}$$

therefore,

$$\overline{v_1^2} = \overline{v_2^2} = \overline{v_1v_2} = \frac{kT}{(C_1 + C_2)}.$$

If $R_1 \to \infty$,

$$S_{11} = \frac{2kTR_2C_2^2}{|C_1 + C_2 + sC_1C_2R_2|^2}, \quad S_{22} = \frac{2kTR_2C_1^2}{|C_1 + C_2 + sC_1C_2R_2|^2}, \quad S_{12} = \frac{-2kTR_2C_1C_2}{|C_1 + C_2 + sC_1C_2R_2|^2},$$

therefore,

$$\overline{v_1^2} = \frac{kTC_2}{C_1(C_1 + C_2)}, \quad \overline{v_2^2} = \frac{kTC_1}{C_1(C_1 + C_2)}, \quad \overline{v_1 v_2} = -\frac{kT}{C_1 + C_2}.$$

Appendix E

Polynomial Equations

"I am ill at these numbers: I have not art to reckon my groans."
Hamlet (act II, scene 2)

The general form of a polynomial is as follows:

$$y(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0$$

This may be re-written in the nested form like so:

$$y(x) = ((...(a_n x + a_{n-1})x + a_{n-2})...)...)x + a_0$$

In this form, in order to evaluate points on the general curve, n multiplications and n additions are necessary. The multiply operation is the most important of these to consider, since multipliers in general occupy an area (of the chosen substrate, be it Si or GaAs) proportional to the product of the bit sizes of the operands. The representation of the polynomial is to be decomposed and transformed into a form which is equivalent to the original form of the polynomial, but is more suitable for a recursive algorithm to evaluate.

Beginning with the general form of the polynomial, as before:

$$y(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0 \tag{E.1}$$

This may be expressed in the Taylor series form:

$$y(x_0 + h) = y(x_0) + h\dot{y}(x_0) + \frac{h^2}{2}\ddot{y}(x_0) + ... + \frac{h^n}{n!}y^{(n)}(x_0) \tag{E.2}$$

where:

x_0 = some value of x for which the derivatives $\dot{y}, \ddot{y}, ..., y^{(n)}$ exist,

$h = x - x_0$

and

x = the point at which evaluation of y is required.

Equation E.2 may be re-written as:

$$y(x_0 + h) = y(x_0) + d_1(x_0) \tag{E.3}$$

where:

$$d_1(x_0) = h\dot{y}(x_0) + \frac{h^2}{2}\ddot{y}(x_0) + ... + \frac{h^{n-1}}{(n-1)!}d_1^{(n-1)}(x_0) \tag{E.4}$$

This polynomial may be reduced to a recursive set options by iteration of this process, ie.

$$d_1(x_0 + h) = d_1(x_0) + h\dot{d}_1(x_0) + \frac{h^2}{2}\ddot{d}_1(x_0) + ... + \frac{h^{n-1}}{(n-1)!}d_1^{(n-1)}(x_0) \tag{E.5}$$

$$d_2(x_0) = h\dot{d}_1(x_0) + \frac{h^2}{2}\ddot{d}_1(x_0) + ... + \frac{h^{n-1}}{(n-1)!}d_1^{(n-1)}(x_0) \tag{E.6}$$

Equation E.6 corresponds to:

$$d_1(x_0 + h) = d_1(x_0) + d_2(x_0) \tag{E.7}$$

This process may be continued until the first derivative is zero.

The polynomial may then be re-written in the new form, ie.

$$\begin{aligned} y(x_0 + h) &= y(x_0) + d_1(x_0) \\ d_1(x_0 + h) &= d_1(x_0) + d_2(x_0) \\ d_2(x_0 + h) &= d_2(x_0) + d_3(x_0) \\ &\vdots \\ d_{n-1}(x_0 + h) &= d_{n-1}(x_0) + d_n(x_0) \\ d_n(x_0 + h) &= d_n(x_0) \end{aligned}$$

where, in the most general terms:

$$d_j(x_0 + h) = d_j(x_0) + h\dot{d}_j(x_0) + \frac{h^2}{2}\ddot{d}_j(x_0) + ... + \frac{h^{n-j}}{(n-j)!}d_j^{(n-j)}(x_0) \tag{E.8}$$

and

$$d_j(x_0) = h\dot{d}_{j-1}(x_0) + \frac{h^2}{2}\ddot{d}_{j-1}(x_0) + ... + \frac{h^{n-j+1}}{(n-j+1)!}d_{j-1}^{(n-j+1)}(x_0), d_0 = y \qquad \text{(E.9)}$$

If

$$x_1 = x_0 + h,$$

then given that

$$d_j(x_0), j = 1, I, n, y(x_0),$$

it becomes possible to calculate exactly

$$y(x_1), d_j(x_1), j = 1, I, n.$$

Hence, any point at an integral number of h away from x may be evaluated by the following recursion:

$$y(x_{i+1}) = y(x_i) + d_1(x_i)$$

$$d_1(x_{i+1}) = d_1(x_i) + d_2(x_i)$$

$$d_2(x_{i+1}) = d_2(x_i) + d_3(x_i)$$

$$\vdots$$

$$d_{n-1}(x_{i+1}) = d_{n-1}(x_i) + d_n(x_i)$$

$$d_n(x_{i+1}) = d_n(x_i)$$

where

$$x_{i+1} = x_i + h$$

The system of equations above are significant because they require no multiplications. There is a trade-off though, since the initial conditions expressed in E.9 are required to be calculated.

If $x_0 = 0$, the problem is significantly reduced, because only the highest order derivative is non-zero, ie.

$$d_j(x_0) = \frac{h^{n-j+1}}{(n-j+1)!}d_{j-1}^{(nj+1)}(x_0) \qquad \text{(E.10)}$$

this system may be calculated quickly or possible stored for retrieval when required.