REACH THROUGH HOT CARRIER SILICON ELECTROLUMINESCENCE IN STANDARD CMOS

by

Petrus Johannes Venter

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor (Electronic Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology
UNIVERSITY OF PRETORIA

May 2013

SUMMARY

REACH THROUGH HOT CARRIER SILICON ELECTROLUMINESCENCE IN STANDARD CMOS

by

Petrus Johannes Venter

Promoter(s):	Prof Monuko du Plessis and Prof Enrico Bellotti
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Philosophiae Doctor (Electronic Engineering)
Keywords:	silicon, hot carrier, electroluminescence, avalanche, breakdown, integrated circuit, photonics, microdisplay, reach through, CMOS

The elusive art of integrating an efficient light source into standard CMOS ICs would unlock exciting possibilities for inexpensive integrated photonics. By nature, silicon is ill-suited as an optically emitting material. Electroluminescence based on hot carrier emission offers an integration capability where light sources can coexist and interact with other circuitry on chip without altering the process. Hot carrier electroluminescence in reverse biased silicon *pn*-junctions usually occurs under avalanche breakdown conditions with high associated electric fields. The available junctions and doping concentrations in a standard CMOS process are limited by the process implant steps designed to cater for transistors. A circuit designer is confronted by the restrictive ability to define geometries only through mask definitions allowed by the foundry with no means to control the operating voltage of hot carrier silicon light sources. Since avalanche occurs at high voltages, hot carrier light sources therefore operate at high voltages, usually to the detriment of interaction with other circuits on-chip. The work in this thesis starts with investigating and developing a method for integrating light sources with other digital circuitry in order to prove the integration possiblities afforded by an integrated light source. The resulting work manifests in the creation of a 64×8 pixel microdisplay designed with state of the art light sources which operate at the process defined minimum breakdown voltage. A technique is then developed whereby the breakdown voltage of light sources can be lowered through the use of electric field reach through between two highly doped regions available in practically all

standard CMOS processes. By terminating the electric field existing in a reverse biased junction prematurely, the operating voltage of the light source can be reduced and tailored by the circuit designer using geometrical design input while requiring no changes to the process. An additional improvement beyond the state of the art is presented where field oxide channels are formed at the Si/SiO_2 interface which provides more predictable and reliable electrical behaviour while improving the light extraction efficiency. These improved reach through light sources with field oxide channels are subsequently used in a 128$\times$96 pixel microdisplay. The scope of integration is increased with additional functions and video capability added on-chip. The display operates at a lower voltage and improved efficiency. A luminance analysis and comparison between the two microdisplays implemented in this work show that although the photopic output power is improved substantially using the reach through light sources, the luminance remains sensitive to the emission area and nature of the radiation pattern of the light sources used in the microdisplay systems. This has a substantial impact on microdisplay performance. Integrated silicon light sources indeed offer exciting possibilities, and while difficult by nature, it is possible by design.

OPSOMMING

DEURGESPERDE WARMDRAER SILIKON ELEKTROLUMINESSENSIE IN STANDAARD CMOS

deur

Petrus Johannes Venter

Promotor(s):	Prof Monuko du Plessis en Prof Enrico Bellotti
Departement:	Elektriese, Elektroniese en Rekenaar-Ingenieurswese
Universiteit:	Universiteit van Pretoria

Graad:	Philosophiae Doctor (Elektroniese Ingenieurswese)
Sleutelwoorde:	silikon, warmdraer, elektroluminessensie, lawine, deurbraak, geïntegreerde stroombaan, fotonika, mikrovertoon, deursper, CMOS

Die ontduikende vermoë rakende die integrasie van 'n effektiewe ligbron in 'n standaard CMOS geïntegreerde stroombaan skep opwindende moontlikhede vir lae koste geïntegreerde fotonika. Silikon is weens sy aard nie die ligskeppende materiaal van keuse nie. Straling wat berus op elektries-gedrewe warmdraer ligbronne bied egter die vermoë om ligbronne wat saamwerk met ander stroombane op 'n vlokkie te vestig sonder om die proses te verander. Warmdraer straling in trugespande silikon *pn*-voegvlakke vind gewoonlik plaas onder lawine-deurbraak waar hoë elektriese velde ter sprake is. Die voegvlakke en doteerkonsentrasies beskikbaar in 'n standaard CMOS proses word beperk deur die implanteringstappe eie aan die proses en is gemik op transistorimplementering. 'n Stroombaanontwerper beskik slegs oor die vermoë om geometriese maskeruitlegte te definieer. Dit ontneem die ontwerper die moontlikheid om deurbraakspanningsvlakke van silikon ligbronvoegvlakke te wysig deur prosesstappe. Omdat lawine-deurbraak gewoonlik teen hoë spannings plaasvind skep dit ongunstige omstandighede vir ander stroombaankomponente om met die ligbronne te skakel. Die werk in hierdie proefskrif ondersoek die moontlikhede en vereistes rondom die integrering van warmdraer ligbronne met ander stroombane om die integrasievermoëns te bepaal. Die resultaat manifesteer in die implementering van 'n 64$\times$8-piksel mikrovertoon wat gebruik maak van die huidige stand van die tegniek ligbronne. Hierdie ligbronne funksioneer teen die minimum deurbraakspannings soos deur die proses gedefinieer. 'n Tegniek word vervolgens ontwikkel waar die deurbraakspanning van die ligbronne

verlaag kan word sodra die elektriese veld tussen twee hoog-gedoteerde gebiede, soos beskikbaar in feitlik alle CMOS prosesse, deursper. Deur die elektriese veld in die trugespande sperlaag voortydig te beëindig kan die dryfspanning van die ligbronne bepaald verlaag word deur ’n stroombaanontwerper, deur slegs die geometriese instruksies aan die vervaardiger te wysig. Hierdie benadering verg dus geen verandering aan die proses self nie. Verder word die deurspertegniek uitgebrei deur die veldoksied te omskep in kanaal-tipe strukture wat gevorm word deur die Si/SiO_2-intervlak. Hierdie wysiging voorsien die ligbronne met meer voorspelbare en betroubare elektriese eienskappe sowel as ’n verbetering in die ligonttrekkingseffektiwiteit. Vervolgens word die verbeterde ligbronne in deursper-modus met veldoksied kanale in ’n geïntegreerde 128$\times$96-piksel mikrovertoon gebruik. Die vlak van integrasie word aansienlik verhoog met meer funksies, insluitend video vermoë, ingebou op dieselfde mikrovlokkie. Die 128$\times$96-piksel vertoon funksioneer teen ’n laer dryfspanning en verbeterde drywingseffektiwiteit. Luminansie-analise en ’n vergelyking tussen die twee mikrovertone wys dat hoewel die fotopiese drywingseffektiwiteit verbeter is deur van die verbeterde ligbronne gebruik te maak, bly die luminansie sensitief vir die stralingsarea sowel as die ruimtelike hoeke van die betrokke stralingspatrone. Dit het ’n wesenlike impak op die werksverrigting van die mikrovertoon. Geïntegreerde silikon ligbronne bied opwindende moontlikhede en die werklikheid van ’n volledig geïntegreerde bron, hoewel uitdagend, is moontlik deur ontwerp.

LIST OF ABBREVIATIONS

ADC	Analogue to digital converter
BEOL	Back end of line
BER	Bit error rate
BTE	Boltzmann Transport Equation
CAD	Computer assisted design
CE	Chip enable
CEFIM	Carl and Emily Fuchs Institude for Microelectronics
CMOS	Complimentary metal-oxide-semiconductor
CVD	Chemical vapour deposition
DAC	Digital to analogue converter
DC	Direct current
DI	Data input
DIBL	Drain-induced barrier lowering
DRAM	Dynamic random access memory
EPE	External power efficiency
EQE	External quantum efficiency
ESD	Electrostatic discharge
FOX	Field oxide
FPGA	Field programmable gate array
HBT	Heterojunction bipolar transistor
I/O	Input/Output
IC	Integrated circuit
ILD	Inter-metal layer dielectric
LCD	Liquid crystal display
LCoS	Liquid crystal on silicon
LDD	Lightly doped diffusion
LED	Light emitting diode

LiPo	Lithium polymer
LOCOS	Local oxidation of silicon
MC	Monte Carlo
MCLK	Master clock
MCLR	Master clear
MOS	Metal-oxide-semiconductor
MOSFET	Metal-oxide-semiconductor field effect transistor
MPW	Multi-project wafer
MTTF	Mean time to failure
MUX	Multiplexer
NCL	Nitride-clad LOCOS
NIR	Near infrared
OLED	Organic light emitting device
PBL	Polysilicon buffered LOCOS
PERL	Passive emitter, rear locally-diffused
PICA	Picosecond imaging circuit analysis
PMT	Photon multiplier tube
PSi	Porous silicon
PSL	Polysilicon spacer LOCOS
PWM	Pulse width modulation
RF	Radio frequency
RIE	Reactive ion etch
RT	Reach through
RTL	Register transfer level
SOI	Silicon on insulator
SPICE	Simulation Program with Integrated Circuit Emphasis
SPM	Serial-to-parallel memory
STI	Shallow trench isolation
SWAMI	Sidewall masked isolation
TDDB	Time dependent dielectric breakdown
TM	Typical mean
TZDB	Time zero dielectric breakdown
VGA	Video graphics adapter

VHDL	VHSIC hardware description language
VIS	Visible, refers to spectral range covering visible wavelengths
VLSI	Very large scale integration
WP	Worst power
WS	Worst speed

LIST OF SYMBOLS

ψ_{bi}	Built-in junction potential
E_g	Band gap energy
$E_{\Gamma 1}$	Energy gap between valence and first conduction band at $k = 0$
J_n	Current density due to electrons
J_p	Current density due to holes
E	Electric field
q	Magnitude of electron charge
ρ	Charge density
$\mu_{n,p}$	Electron and hole mobility
D	Diffusion constant
ε_0	Vacuum permittivity
K_S	Relative permittivity of silicon
$r_{n,p}$	Electron and hole recombination rates
$g_{n,p}$	Electron and hole generation rates
T	Temperature
T_e	Effective carrier temperature
c_s	Velocity of sound
v_d	Drift velocity
μ_0	Low field carrier mobility
$\hbar$	Reduced Planck's constant
m_0	Carrer mass in free space
m^*	Effective carrier mass
F	Force
$\alpha_{n,p}$	Impact ionisation coefficients for electrons and holes
$M_{n,p}$	Avalanche multiplication factor for electrons and holes
V_{BD}	Breakdown voltage of *pn*-junction
E_m	Maximum electric field strength

W_{Dm}	Maximum depletion region width
N	Background doping concentration
V_R	Reverse-biased voltage applied to *pn*-junction
I_s	Saturation current in *pn*-junction
k	Boltzmann's constant
$\tau_{n,p}$	Electron and hole carrier lifetimes
n_i	Intrinsic carrier concentration
$N_{D,A}$	Donor and acceptor impurity concentration
$\sigma_{n,p}$	Electron and hole capture cross section
v_{th}	Carrier velocity at thermal equilibrium
N_t	Trap density
E_i	Intrinsic Fermi-energy level
E_t	Trap energy level
J_0	Total equilibrium current density
W_D	Depletion region width
J_{ge}	Current density due to carrier generation
τ_g	Carrier generation lifetime
η_{eqe}	External quantum efficiency
η_{epe}	External power efficiency
E_{ph}	Photon energy
h	Planck's constant
c	Velocity of light
λ	Wavelength
P_{opt}	Optical power, usually output power, same as radiant flux
P_e	Electric power, usually input power
I_e	Electrical current, sometimes radiant intensity
E_{crit}	Critical electric field
ψ	Potential
x_{sep}	Separation distance between two highly doped regions
V_{CY}	Cylindrical depletion region breakdown voltage
V_{SP}	Spherical depletion region breakdown voltage
r_j	Junction curvature radius
n_{Si}	Refractive index of silicon
n_{SiO_2}	Refractive index of silicon dioxide

$t_{g,ox}$	Gate oxide thickness
$E_{g,max}$	Maximum electric field across the gate oxide
V_{sup}	Supply voltage
ϕ	Metal ion activation energy
L_v	Luminance in cd/m^2
A	Area
Ω	Solid angle in steradian or electrical resistance in Ohms
Φ_v	Luminous flux
$J(\lambda)$	Spectral power distribution
$\bar{y}(\lambda)$	Luminosity function
$\phi_v(\lambda)$	Spectral distribution of luminous flux
I_v	Luminous intensity
ε_{Si}	Permittivity of silicon
κ	Surface reaction rate
Φ_e	Radiant flux, sometimes called optical power
I_e	Radiant intensity in context of luminance calculations

ACKNOWLEDGEMENT

The work described in this thesis has not happened in isolation. INSiAVA (Pty) Ltd is a company aiming at commercialising silicon light emission technologies. The majority of this work has been funded by INSiAVA (Pty) Ltd in agreement with the University of Pretoria. The Carl and Emily Fuchs Institute for Microelectronics (CEFIM) houses a group of talented researchers working on silicon light emission. I am indebted to all of my colleagues, especially Marius Goosen and Christo Janse van Rensburg, who have been the ears and sounding board in many discussions pertaining to this work, and the rest of the team, Dr Pieter Rademeyer, Niklaas Faure, Antonie Alberts, Paula Grove and Vianca Mocke, with whom it is a pleasure to work. A thanks to Dr Trudi Joubert who has been the chief designer in most of the logic associated with the 128×96 pixel microdisplay, which helped me in proving the integration capabilities of my designs. I would like to extend my gratitude to Ms Tilla Nel who keeps the CEFIM heartbeat ticking and who always has ears to listen and to the rest of my CEFIM colleagues, Professor Saurabh Sinha and Johan Schoeman, who help in creating the world class characteristic of the CEFIM group.

I am also thankful for my co-promotor, Professor Enrico Bellotti, who has introduced me to the field of quantum mechanics and numerical simulation of semiconductor devices and his guidance on a whole range of topics during the number of visits I have made to Boston and his visits to South Africa.

The guidance and mentorship of Gerrie Mostert have been extremely valuable, both in my professional and personal life. His experience in the intersection of science and business, with his insight and knowledge in intellectual property strategy, has given a breadth to my understanding of our world which I dearly appreciate.

Finally I would like to thank Professor Monuko du Plessis, who has been a role model for me since the undergraduate ELK220 class in 2002. The combination of his methodical approach to science, his focus, creativity and intuition has led to some of the most interesting inventions and insights. In addition, he has the ability to impart his knowledge of the complex and difficult concepts in semiconductor science in such a way that a layman would understand. Fortunately, I was able to draw extensively on this ability of his. His guidance in this work has been invaluable.

To my loving wife and children

who simultaneously distracted and supported me

in the most wonderful ways,

my family

and

in memory of my brother

TABLE OF CONTENTS

CHAPTER 1 Introduction 1

1.1 Background . . . 1
1.2 Problem statement . . . 2
1.3 Motivation . . . 3
1.4 Hypothesis and research questions . . . 4
1.4.1 Hypothesis . . . 4
1.5 Research questions . . . 4
1.6 Justification of research . . . 5
1.7 Research methodology . . . 5
1.8 Contribution to body of knowledge . . . 6
1.9 Outline of thesis . . . 7
1.10 List of publications . . . 9
1.10.1 Peer reviewed journal publications . . . 9
1.10.2 Peer reviewed international conference contributions . . . 9
1.10.3 Patents applications filed and granted . . . 10

CHAPTER 2 Literature review of related work 11

2.1 Electroluminescence in semiconductors . . . 11
2.1.1 Light emission from indirect band gap materials . . . 11
2.1.2 Summary . . . 14
2.2 Hot carrier photon emission from indirect band gap materials . . . 15
2.2.1 First observation . . . 15
2.2.2 Radiative mechanism and modelling . . . 15
2.2.3 Intrinsic bandwidth . . . 16
2.2.4 Reliability . . . 17
2.2.5 Improvement techniques . . . 17

2.3 CMOS integration of silicon light sources 18
2.3.1 Characterisation applications 18
2.3.2 Integration of silicon light emitters 19
2.3.3 Optical communications 19
2.3.4 Integrated microdisplays 20
2.4 Conclusion 20

CHAPTER 3 Theory on device operation 22

3.1 Carrier transport 22
3.1.1 Boltzmann Transport Equation 22
3.1.2 Low field transport 23
3.1.3 High field transport 24
3.1.4 Full band approach 24
3.1.5 Density of states 25
3.2 Transport in reverse-biased *pn*-junctions 26
3.2.1 Avalanche multiplication 26
3.2.2 Band-to-band tunnelling and Zener-breakdown 28
3.2.3 Minority carrier drift current in reverse-biased *pn*-junctions 29
3.2.4 Generation and recombination 29
3.3 Hot carrier luminescence 30
3.3.1 Establishing a high electric field 30
3.3.2 Establishing transport 30
3.3.3 Efficiency calculations 31
3.4 Proposed technique for electric field reach through 32
3.4.1 Electric field profile in reach through 32
3.4.2 Reduced potential integral 34
3.4.3 Improving light extraction efficiency 35
3.5 Conclusion 35

CHAPTER 4 A 64 x 8 pixel microdisplay in standard CMOS 36

4.1 Introduction 36
4.1.1 Target CMOS technology and generalisation 36
4.1.2 Light emitting point sources in standard CMOS 37
4.1.3 CMOS integration of point sources 37

4.1.4 CMOS microdisplay based on point sources . . . 37
4.2 Target CMOS technology and generalisation . . . 37
4.2.1 Definition of standard CMOS . . . 38
4.2.2 Target technology for this research . . . 38
4.3 Light emitting point sources in standard CMOS . . . 39
4.3.1 Point source design . . . 39
4.3.2 Processing steps for creating point sources . . . 41
4.3.3 Breakdown localisation . . . 42
4.3.4 Optical path for light emission . . . 44
4.3.5 Light directing structures . . . 47
4.3.6 Arraying of point sources . . . 48
4.3.7 Radiation pattern . . . 48
4.4 CMOS integration of point sources . . . 49
4.4.1 Well structure and available diodes . . . 50
4.4.2 Limitations on operating voltage of MOS devices . . . 50
4.4.3 High voltage integration . . . 52
4.4.4 Current densities and electromigration of metal . . . 54
4.4.5 Series resistance . . . 55
4.5 CMOS microdisplay based on point sources . . . 56
4.5.1 Point source characteristics . . . 56
4.5.2 Pixel design . . . 57
4.5.3 Scanning architecture . . . 60
4.5.4 Controller design . . . 62
4.5.5 Spectral characteristics . . . 63
4.5.6 Radiation pattern . . . 63
4.5.7 Luminance . . . 64
4.5.8 Electrical characteristics . . . 66
4.6 Conclusion . . . 68

CHAPTER 5 Improved light sources in standard CMOS 70

5.1 Approach . . . 71
5.1.1 Breakdown voltage . . . 71
5.1.2 Proposed approach . . . 72

5.1.3 Conceptual CMOS implementation 73
5.2 Processing considerations 73
5.2.1 Scaling theory for CMOS and device density 73
5.2.2 Oxidation of silicon 74
5.2.3 LOCOS and field oxide 75
5.2.4 Shallow trench isolation (STI) 78
5.2.5 Processing limits 80
5.2.6 Ion implantation 81
5.3 Reference device and experimental approach 84
5.3.1 Reference device for individual point sources 85
5.4 Design for electric field reach through in point sources 85
5.4.1 Geometrical adaptations of point sources 85
5.4.2 Localisation due to reduced barrier potential 85
5.4.3 Layout of a point source using drawn layers in a CMOS process 86
5.5 Results for point sources in reach through 87
5.5.1 Electrical characteristics 88
5.5.2 Optical characteristics 92
5.5.3 Efficiency 96
5.5.4 Summary of results 98
5.6 Improved CMOS geometry with field oxide channels 100
5.6.1 Geometrical approach 100
5.6.2 Design of junctions with oxide channels 103
5.6.3 Fabrication and resulting structures 103
5.7 Results for improved sources in reach through 105
5.7.1 Electrical characteristics 105
5.7.2 Optical characteristics 110
5.7.3 Efficiency 112
5.7.4 Summary of results 114
5.8 Conclusion 116

CHAPTER 6 A 128 x 96 pixel CMOS microdisplay using improved light sources 119

6.1 Introduction 119
6.2 Requirements 120

6.2.1 Qualitative objectives 120
6.2.2 Quantitative objectives 120
6.3 Light source 121
6.3.1 Spectral analysis 121
6.3.2 Point source improvement 123
6.3.3 Point source design 129
6.3.4 Polarity comparison 129
6.4 Integrated pixel design 131
6.4.1 Light source array 131
6.4.2 Driving circuitry 131
6.4.3 Layout 132
6.5 Microdisplay architecture 133
6.5.1 Luminous intensity control 134
6.5.2 Scanning architecture for pixel matrix 135
6.5.3 Integrated controller 138
6.5.4 External controller 140
6.6 Results 141
6.6.1 Electrical characteristics 142
6.6.2 Optical characteristics 145
6.6.3 Luminance 148
6.6.4 Summary 152
6.7 Conclusion 154

CHAPTER 7 Conclusion 156

7.1 Summary of work 156
7.2 Critical analysis 157
7.2.1 Hypothesis 157
7.2.2 Research questions 158
7.2.3 Comparison of efficiency in relation to other work 159
7.2.4 Comparison of CMOS integration in relation to other work 160
7.3 Contribution 161
7.4 Future work 163

APPENDIX A Radiometry 175

A.1 Radiometer 175
A.2 Parameter analysis 176
A.3 Spectral analysis 177
A.4 Method for calculating radiant flux 178
A.5 Conclusion 179

CHAPTER 1

INTRODUCTION

Silicon light emission through hot carrier luminescence has the potential to provide integrated circuits with a light source, resulting in a disruptive capability for standard CMOS circuits to address markets that are inaccessible under conventional wisdom. The electro-optical conversion efficiency is naturally low because silicon is an indirect band gap material. This research work describes a novel way of improving the usability and efficiency of silicon light sources in standard CMOS without process alterations.

1.1 BACKGROUND

Silicon is the element that is fundamentally responsible for the most influential paradigm shift in electronics history. Although the first successful transistor was made of germanium, it was silicon that enabled monolithic integration of multiple functions on a single substrate, utilising the field-effect introduced by Lilienfeld in 1925. Native oxide growth enabled almost perfect gate dielectrics, while the nature of the MOS transistor allowed for ease of scaling seen through much of the 20th century, continuing the trend established by Moore even today.

Silicon domination in all things electronic overshadows one inherent disadvantage: silicon is an indirect band gap material. In an age where bandwidth and data density is becoming increasingly important, a shift to the optical domain is often seen as the next necessary step. Furthermore, as scaling in integrated circuits reduces device dimensions at an unsustainable rate, the "More than Moore"-movement leans towards more integrated functional systems on chip and package rather than just more transistors [1]. Although silicon is often used as a photodetecting material its indirect band gap crystal electronic structure cannot compete with the radiative efficiency of III-V materials and other

direct band gap semiconductors. This fact alone leaves silicon to remain an exclusively electronic material until an integrated light source becomes feasible.

There are two ways of generating light from silicon often reported on in literature. The first is through radiative recombination of free minority carriers as a result of diffusion across a forward biased *pn*-junction. The majority of radiative action occurs through indirect, phonon-assisted recombination, where the probability of radiation is adversely affected by the available phonon dispersion statistics. As the lifetime of minority carriers in the quasi-neutral regions are relatively long, the devices are not inherently fast. The body of knowledge includes a reasonable amount of work in this field. The second method for light generation is through avalanche electroluminescence. Literature typically sees this as a curiosity and a limited number of authors have empirically tried to explain certain observances noted in experimental results, usually based on spectral results. There have been a limited number of attempts to explain the origins of the radiative mechanisms and to explain the behaviour of carriers under the conditions in which this phenomenon occur.

This work focuses on the usability of the second phenomenon *supra*. Since the optically active material is silicon, it becomes possible to work towards an integrated light source in silicon-based electronics and very specifically, standard CMOS processes aimed at manufacturing digital and mixed-signal circuits.

1.2 PROBLEM STATEMENT

Conventional electron-hole recombination in silicon is generally non-radiative and slow due to a misalignment of the lowest conduction band valley and the Γ-point in momentum space. In contrast, the conduction valley and valence peak align at the Γ-point in a direct band gap material such as GaAs. This is an inherent material property.

> *The fundamental problem is therefore that the need for an integrated light source cannot be fulfilled by reverting to conventional radiative recombination mechanisms in standard CMOS circuits.*

Hot carrier luminescence is a term referring to radiative mechanisms associated with carrier actions when the carriers are "heated" by using very high electric fields. Carriers are shifted away from the lowest energy valleys and spread to other states of higher energy in the Brillouin-zone. The

exact photon-generating mechanisms are beyond the scope of this work, although the means for establishing radiation gives rise to difficulties addressed in this research. For example, hot carrier luminescence requires carrier transport in regions of high electric field. One way of establishing both at the same time is by reverse-biasing a *pn*-junction to the point of avalanche breakdown. The voltage required is usually fixed by the specific doping implants used for transistors in standard CMOS processes. This value is often high enough to prevent practical integration of light sources with other electronic circuitry on-chip. Although the nature of the radiative mechanism differs from conventional LEDs, the efficiency is still low. Therefore, any improvement in efficiency is conducive to creating an attractive light source for integration in a standard CMOS process.

> *The problem addressed in this work is creating a means for integrating silicon light sources based on hot carrier luminescence suitable for use in standard CMOS while improving the electro-optical conversion efficiency.*

1.3 MOTIVATION

Silicon, and more specifically the CMOS platform, is the most widely used semiconductor in electronic devices, covering an estimated market of over $300 billion in the semiconductor domain. The physical nature of crystalline silicon and the associated crystal electronic structure makes spontaneous emission in forward biased junctions both slow and inefficient due to its indirect band gap nature.

The ability to employ hot carrier luminescence in silicon eliminates the speed constraint of conventional diffusion LEDs, while the quantum and power efficiency, although not good compared to direct band gap semiconductors, may be improved to a point of usability in integrated circuit environments.

The biggest advantage offered by silicon light sources, particularly when CMOS compatible, is the integration opportunities and cost benefits offered by CMOS. Current infrastructure and over 40 years of technological experience and maturity are proof of how ingrained CMOS is in the electronics market and also indicative of a platform that will likely extend well into the future.

In a broad sense, the motivation for this research is therefore driven by the opportunities which will arise when bridging the gap between CMOS and the optical domain, thereby enabling CMOS to

contend in the field of optics, which is usually reserved for the more exotic and expensive direct band gap semiconductors such as GaAs and InP. With a feasible silicon light source, this may be possible.

From a narrow perspective this research is motivated by a need to find practical solutions for integrating light sources in a technology well suited for low voltage digital circuitry, without any optical devices in mind, and using CMOS compatible improvement techniques for enhancing device efficiency and performance.

1.4 HYPOTHESIS AND RESEARCH QUESTIONS

1.4.1 Hypothesis

The operating voltage of a *pn*-junction in avalanche is the result of the integral of the internal electric field. Only a small region within the electric field is responsible for avalanche multiplication. It is therefore envisaged that by using a highly doped field stop it is possible to terminate the electric field without affecting carrier generation due to the avalanching region, thereby reducing the operating voltage of such a light source. This will facilitate integration and coexistence with other on-chip circuitry and may improve the power efficiency of the light sources.

The hypothesis can therefore be stated as

> *If the electric field distribution can be changed through geometrical adaptations which will result in effects such as electric field reach through between two highly doped regions in a pn-junction, the electrical and optical characteristics of silicon hot carrier luminescent light sources may be altered and optimised in such a way as to facilitate the integration of these light sources in standard CMOS as usable components in integrated systems such as microdisplays.*

1.5 RESEARCH QUESTIONS

1. What is the correlation between the length of the electric field in each through on the operating voltages of hot carrier light sources?

2. If correlation exists, what are the bounds and means to improve and optimise devices for integration?

3. How does electric field reach through affect power efficiency in hot carrier luminescent silicon light sources?

4. What are the issues when integrating hot carrier luminescent light sources into standard CMOS while coexisting with other circuitry?

5. What are the considerations when harnessing the visible part of the light emission for use in CMOS microdisplays?

1.6 JUSTIFICATION OF RESEARCH

Junctions in standard CMOS processes are limited to only a few available configurations. The doping levels are fixed according to what is required for transistor implementation. While light emission from silicon has been reported from CMOS junctions in avalanche- and field emission breakdown modes, most CMOS process rules do not afford the circuit designer structures to control the breakdown voltage of these mechanisms. This work aims to introduce a technique which does offer a circuit designer a deterministic way of controlling the breakdown voltage, without affecting the power efficiency of the electro-optical conversion process. Since hot carrier luminescence in silicon emits visible light, the application of this technique in a microdisplay application proves that it is possible to use this technique from a CMOS circuit designer perspective, without process alterations. It now becomes possible to define a CMOS light source as a predictable component in a CMOS foundry's standard cell library.

1.7 RESEARCH METHODOLOGY

1. A literature study of hot carrier luminescence in semiconductors will serve as a starting point for investigating the radiative phenomenon involved and establishing the conditions under which the phenomenon occurs. The review will also include a comparison of the approaches taken to integrate hot carrier light sources with on-chip circuitry in standard CMOS processes.

2. State-of-the-art in silicon hot carrier luminescent light sources will be investigated and techniques will be developed for integration of these light sources in standard silicon, using a mi-

crodisplay application as a test case.

3. The nature of CMOS, in terms of opportunities and constraints will be investigated for a typical CMOS process technology. The mismatch between operating voltages required, optimisation of surface profile at the Si/SiO2 interface and doping related aspects will be considered and evaluated in terms of its compatibility with the proposed efficiency improvement technique.

4. Improved light sources will be developed targeted at certain shortcomings of the state of the art light sources with an improvement of efficiency and ease of integration in mind.

5. The improved light sources will be used in a microdisplay with scaled up performance specifications in order to evaluate the extent to which the research questions are satisfied.

6. Multiple integrated circuit designs in a standard CMOS process will be designed from first principles to evaluate the premises of this work. This will include designs with the intent to investigate and evaluate the following:

   - Electrically characterise the operation of light sources
   - Optically evaluate the emission characteristics of the luminous phenomenon
   - A comparison of the efficiency improvement between state of the art and improved light sources
   - The evaluation of considerations for integration where light sources need to coexist with digital and analogue circuitry on a standard CMOS integrated circuit
   - The proof of the possibility of such integration
   - Experimental and analysis work and its interpretation in proving the hypothesis.

1.8 CONTRIBUTION TO BODY OF KNOWLEDGE

- This work describes the first directly observable microdisplay based on silicon light sources in standard CMOS using state of the art silicon light sources with usable luminance values for direct observation without image intensification.

- This work is the first to prove that, by using electric field reach through, it is possible to modify the operating voltage of a silicon light source according to the application need.
- This work presents a novel way of improving the reliability and extraction efficiency of a silicon light source by using a field oxide LOCOS structure in a channel-like configuration.
- This work presents the first 128×96 pixel microdisplay, with architecture considerations, capable of being produced in a standard CMOS process and suited for use with the naked eye.
- The results of this work show a strong correlation between external quantum efficiency and device potential. This is indicative of a relation between quantum efficiency and the time a carrier spends under a high electric field.

1.9 OUTLINE OF THESIS

Chapter 1 is an introduction to the relevance of this work in relation to the body of knowledge surrounding silicon hot carrier luminescence. A background is given on the luminescent phenomena in semiconductors. The specific problems associated with the need for an integrated light source in silicon are addressed as well as the problem addressed by this research work. A motivation for this research and hypothesis along with its associated research questions is formulated which is given due consideration throughout the remainder of this thesis. The research methodology is discussed along with the contribution to the body of knowledge arising from this work. Journal articles and conference contributions originating from this work are summarised to substantiate the relevance of the approaches developed herein.

Chapter 2 contains an overview of the approaches followed in trying to create useful light emission from indirect band gap semiconductors. A more detailed investigation into the observations on hot carrier luminescence in silicon provides for a solid foundation on which to compare this work. Although there have been a number of publications on hot carrier luminescence and the modelling thereof, few groups have tried to convert this phenomenon into a useful addition to the set of components in standard CMOS. The efficiency performance of hot carrier luminescence is examined from results by various groups. This serves as a relevant figure of merit with which to compare the results obtained in this work. Finally, the attempts by the few groups to integrate a light source as a standard CMOS component are mentioned.

Chapter 3 contains some background knowledge which is useful and/or necessary to understand the design approaches in this work. Some of the observations gained during experimental work can also be understood in the light of the theoretical tools discussed in this chapter. High field carrier transport considerations are mentioned, as well as expectations in terms of observables, while the reach through approach is considered from an abrupt junction approximation.

Chapter 4 introduces the state of the art CMOS light sources developed by our group at the Carl and Emily Fuchs Institute for Microelectronics (CEFIM), University of Pretoria, where aspects such as the layout masks and geometry of the processed structures are discussed. Localisation of the breakdown region is necessary to ensure uniformity of emission. A technique to accomplish this is explained. Light sources are employed in a pixel and integrated with on-chip driver circuitry and a scanning decoder in order to create a 64×8 pixel CMOS microdisplay capable of directly observable luminance levels. The performance of this microdisplay is characterised and some of the difficulties of integration are stated.

Chapter 5 proposes a technique which can be used for creating layout geometries where the reach through condition is established. The first approach follows the obvious route of reducing the spacing between two highly doped regions using conventional diode layout structures mentioned in the design documents supplied by the foundry. Problems arise pertaining to the field oxide profile between junctions. An improved approach is introduced and shown to result in devices which outperforms the previous approach as well as the state of the art light sources. The improved light sources are electrically more predictable, exhibit an improved optical extraction efficiency and provide predictable operating voltage behaviour as a function of the junction separation distance.

Chapter 6 combines the integration aspects identified and developed in **chapter 4** and the light sources developed in **chapter 5** in a design for a much larger, more integrated microdisplay with an array of 128×96 pixels where each pixel utilised the improved light sources. The driving architecture is improved compared to the 64×8 pixel display to enable a single high voltage feed to the optical array. This voltage is substantially reduced compared to the design employing state of the art light source and becomes compatible with power sources often used in mobile environments. Finally, the power efficiency of the microdisplay is improved by a substantial factor, although the luminance results suffer in the light of the radiation pattern of the improved light sources.

Chapter 7 concludes the study with a critical analysis on this work compared to other works of similar nature. The results of this work are summarised and the contribution to the body of knowledge is highlighted. Future gaps in the body of knowledge are identified and suggested as work to evolve into future studies.

1.10 LIST OF PUBLICATIONS

1.10.1 Peer reviewed journal publications

1. M. du Plessis, P. J. Venter and E. Bellotti, "Spectral characteristics of hot electron electroluminescence in silicon avalanching junctions," *IEEE J. Quantum Electron.*, to be published.

2. P. J. Venter, M. du Plessis, A. W. Bogalecki, M. E. Goosen and P. Rademeyer, "An 8x64 pixel dot matrix microdisplay in 0.35-μm complementary metal-oxide semiconductor technology," *Optical Engineering*, vol. 51, no. 1, pp. 014003-1-014003-7, Jan. 2012.

3. M. E. Goosen, M. du Plessis, P. J. Venter, A. W. Bogalecki, A. C. Alberts and P. Rademeyer, "CMOS avalanche electroluminescence applications - microdisplay and high speed data communication," *Trans. SAIEE*, vol. 103, no. 1, pp. 24-28, 2012.

1.10.2 Peer reviewed international conference contributions

1. P. J. Venter, A. C. Alberts, M. du Plessis, T.-H. Joubert, M. E. Goosen, C. Janse van Rensburg, P. Rademeyer and N. M. Fauré, "A CMOS microdisplay with integrated controller utilizing improved silicon hot carrier luminescent light sources," in *Proc. SPIE 8643, Advances in Display Technologies III, 864309*, 2013 ©SPIE Digital Library, doi: 10.1117/12.2004277.

2. P. J. Venter, M. du Plessis, A. W. Bogalecki and P. Rademeyer, "Late-news poster: An all-CMOS microdisplay utilizing integrated novel avalanche light-emitting sources," *SID Symp. Digest of Technical Papers*, vol. 42, no. 1, pp. 1512-1515, 2011.

3. P. J. Venter, A. W. Bogalecki, M. du Plessis, M. E. Goosen, I. J. Nell and P. Rademeyer, "CMOS dot matrix microdisplay," in *Proc. SPIE 7956, Advances in Display Technologies and E-papers and Flexible Displays, 79560Y*, 2011 ©SPIE Digital Library, doi:10.1117/12.875131.

4. P. J. Venter, M. du Plessis, I. J. Nell, M. E. Goosen and A. W. Bogalecki, "Improved efficiency of CMOS light emitters in punch through with field oxide manipulation," in *2010 International Conference on Microelectronics (ICM)*, Cairo, 2010, pp. 36 - 39.

5. A. W. Bogalecki, M. du Plessis, P. J. Venter, I. J. Nell and M. E. Goosen, "Integrated optical light directing structures in CMOS to improve light extraction efficiency," in *2010 International Conference on Microelectronics (ICM)*, Cairo, 2010, pp. 168 - 171.

6. P. J. Venter and M. du Plessis, "Improved silicon light emission for reach- and punch-through devices in standard CMOS," in *Proc. SPIE 7607, Optoelectronic Interconnects and Component Integration IX, 76070Z*, 2010 ©SPIE Digital Library, doi:10.1117/12.841357.

7. M. du Plessis, P. J. Venter and A. W. Bogalecki, "Using reach-through techniques to improve the external power efficiency of silicon CMOS light emitting devices," in *Proc. SPIE 7606, Silicon Photonics V, 760612*, 2010 ©SPIE Digital Library, doi:10.1117/12.840646.

1.10.3 Patents applications filed and granted

1. South African patent granted (RSA no. 2012/04957)

2. China patent application no. 201180006581

3. European patent application no.11705035.1

4. Japanese patent application no. 2012-549458

5. USA patent application no. 13/574,333

CHAPTER 2

LITERATURE REVIEW OF RELATED WORK

2.1 ELECTROLUMINESCENCE IN SEMICONDUCTORS

Electroluminescence from semiconductors was first observed by Howard Round as a glow observed from a silicon carbide diode in 1907 [2] and independently in 1927 by Oleg Losev [3], who went further to create devices utilising the phenomenon and patenting applications around its use. Throughout the 1950s, a number of authors [4], [5] observed radiance from diodes and noted that the emission spectrum coincided with the band gap energy associated with the semiconductor. Observations linked the radiative mechanism to the recombination of electrons and holes, as described by Shockley [6] as an interband process. Note that the mechanism does not involve phonon action if carrier recombination occurs in direct band gap semiconductors, such as GaAs, making the probability of direct recombination much higher. This is the reason direct band gap semiconductors are much more efficient when depending on electron-hole recombination for radiative action. Silicon has an indirect band gap structure which requires phonon-assisted recombination where these transition rates has to compete with other dominating non-radiative processes [7]. Due to this indirect band gap nature of the crystal electronic band structure, silicon is inherently an inefficient, slow light emitting material. Interband recombination processes are the *de facto* standard way for light emission from semiconductors rendering direct band gap materials as the preferred choice for this purpose.

2.1.1 Light emission from indirect band gap materials

Silicon is used in almost all integrated circuits, rendering silicon as a ubiquitous electronic material. Germanium, with a much higher electron and hole mobility [8] compared to silicon [8], is also used in modern devices, for example, to increase the transconductance of bipolar transistors [9]. It is quite

common to see SiGe HBTs coexisting with CMOS in analogue and RF applications today. Both silicon and germanium are indirect band gap materials and are therefore inefficient optical emitters. A lot of time has therefore been invested in investigating techniques for creating efficient optical emitters using indirect band gap materials.

2.1.1.1 Bulk silicon *pn*-junctions

Forward-biased *pn*-junctions emit light with a photon energy close to the band gap energy [4] of the material. This mechanism will always by nature be slow due to the requirement of phonons with the correct momentum to establish electron-hole recombination for photon emission. Interdigitated *pn*-junctions have been shown to emit light with external power efficiencies of around 10^{-4} at a wavelength of 1.16 μm [10] implemented in standard CMOS. Power efficiencies of up to 10^{-2} have been reported by making use of passive emitter, rear locally-diffused (PERL) techniques [11] although the minority carrier lifetimes are of such longevity that its usefulness is limited to slow switching applications. The designs using PERL techniques are inherently incompatible with CMOS integration without process modifications. More recent work [12] shows results similar to what we have observed in a standard CMOS process with power efficiencies of $\approx 1.2\times10^{-6}$, with optical power reaching levels of around 100 nW with interdigitated structures. Since the absorption wavelength of silicon falls off sharply at the band gap wavelength [13], it is impractical to use a forward-biased approach in bulk CMOS in all-silicon communication systems.

2.1.1.2 Dislocation loop engineering

An interesting approach is to utilise the implantation of Boron ions, available in standard CMOS processes, for creating dislocation loops [14]. Dislocation loops modify the band structure of the silicon crystal by introducing a local strain field. This technique has shown to emit 19.8 μW at 100 mA with an external quantum efficiency of $\approx 2\times10^{-4}$. The approach has drawn critical review [15] regarding the origin of light and the effect of lattice damage. This technique has not yet led to any commercially available integrated light source.

2.1.1.3 Rare earth elements

Erbium doping of the gate oxide in MOS transistors has also shown to allow electroluminescent emission at 1.54 μm with operation at room temperature [16]. The same work also reports on various other

rare earth ions and the quantum efficiencies associated with these. An external quantum efficiency of ≈ 2 % proves emission output powers of usable values. Hot electron and Si nanocrystal interaction with the Erbium ions are reported to be the cause of emission. It is mentioned that reliability of these devices is still preventing commercial use, although it has been a decade since this work was published, without a commercial example to date.

2.1.1.4 Porous silicon

It was observed in 1990 that porous silicon emits light when optically pumped [17]. This has led to a number of investigative works. Electrochemical processing proved to enable the creation of silicon nanowires resulting in porous silicon without reverting to lithographical techniques. Although the hurdles towards commercial application did not seem high [18], the reliability and stability, bandwidth and integration with commercial processes were identified as challenges. Although PSi LEDs are electroluminescent and their potential was showcased as a possible display technology [19], [20], to date no commercial product exists which makes use of this technology.

2.1.1.5 Nanostructures and superlattices

Visible emission from porous silicon has resulted in attention to optical gain from silicon nanocrystals [21]. Silicon nanocrystals are formed in SiO_2 usually through implantation and have been seen to obtain quantum efficiencies as high as 60 % [22]. Most of the successful results are, however, due to photoluminescent pumping [23] and cannot easily be achieved by using conventional carrier injection means for electroluminescent stimulation. In other words, it is difficult to use these structures as electronic components. Furthermore, the implantation of silicon is not a standard CMOS processing step. Nanostructures have also been successfully used to create electrically excited surface plasmon polaritons between metal and dielectric interfaces [24]. Since this approach can potentially utilise the back end of line stack, it may be reconcilable with CMOS processes by changing the BEOL stack.

2.1.1.6 Germanium epitaxial techniques and state filling

Germanium exhibits an indirect band gap much like silicon. One aspect of germanium's electronic band gap structure has received a lot of research attention: while the L-valley determines the conduction band edge with $E_g = 0.664$ eV, the Γ-valley is separated by only $E_{\Gamma 1} = 0.8$ eV from the

valence band edge which provides for a valley "almost within direct band gap reach". There are two approaches currently under active investigation which aim to capitalise on this feature.

1. By creating tensile strain in the correct crystallographic direction, it is possible to "convert" the electronic band structure of germanium to a direct band gap material [25]. This occurs due to the physical deformation of the crystal lattice. A tensile strain of around 1.8 % allows for the Γ-valley to determine the conduction band edge.

2. Under high level injection it is possible to fill the states around the L-point and carriers will start to spill into the Γ-region [26] resulting in direct interband transitions. Another similar approach is to introduce large *n*-type dopants in order to lift the Fermi-energy to a point where state filling will saturate the L-valley and carrier population will start to occur at the Γ-point [27].

Although this approach has shown some experimental success, strain mechanisms based on cantilevers are not compatible with standard CMOS while the high injection levels required in state filling make practical device operation improbable.

2.1.1.7 Hot carrier luminescence

This thesis focuses on hot carrier luminescence and the approach is discussed in great detail below. As a mechanism to create light in an indirect band gap material it is stated here as one of the competing (and most promising) techniques for a complete integrated solution towards a light source in standard CMOS.

2.1.2 Summary

Light emission from silicon and indirect band gap materials has seen a vast amount of interest throughout the past few decades. A review of the current body of knowledge distils this truth: complete compatibility and compliance to standard CMOS baseline processes is an essential criterion.

2.2 HOT CARRIER PHOTON EMISSION FROM INDIRECT BAND GAP MATERIALS

2.2.1 First observation

The first published results of visible radiation from a silicon *pn*-junction was given by Newman [28] in 1955. Visible emission was also reported from the forward biased base-emitter junction where the collector-base junction was biased in avalanche breakdown mode under different biasing conditions of the base electrode [29]. These observations are in stark contrast to spectral results, as observed in [4] from silicon emission, in that the band gap of silicon is 1.12 eV at room temperature, which would result in light beyond the visible spectrum. The observations made was that light seemed to be generated within the silicon itself and the method ascribed to the phenomenon was that of intraband relaxation of hot carriers. The same phenomenon was observed in MOSFETs near the drain region of the transistors [30] which has an effect of being reabsorbed and creating carrier pairs [31].

2.2.2 Radiative mechanism and modelling

Not long after Newman, other works started to investigate the origin of radiation in greater detail. In 1956 Chynoweth and McKay [32] concluded that electron-hole recombination also takes part for photons with energies greater than the band gap of silicon, $E_{g,Si}$, while intraband transitions are responsible for the lower energy photons. Localised radiative spots were also observed, although it was stated that the spot brightness did not increase, but rather the number of spots, as a function of increased drive current. Hot carrier radiation was also observed in germanium reverse biased junctions [33], [34], with the mechanism also ascribed to intraband transitions. The work by Figielski and Torun [35] was the first to investigate the possibility of bremsstrahlung as a radiative mechanism and therefore another contributing radiative phenomenon. This has come under critical review [36], [37] and it is held that for impurity assisted scattering to result in photon emission, the doping concentrations need to be $> 10^{20}$ cm^{-3}.

A number of attempts have been made to fit a combination of these phenomena to the photon spectrum, but due to the wide and almost featureless nature, almost any model will “fit” given the correct empirical parameters. In an attempt to quantify the phenomenon, Bude [36] tried to identify the scattering mechanisms possible for contributing towards radiation and, using a quantum mechanical

modelling approach, tried to fit a theoretical prediction to experimental results, with reasonable success. The electric field profile on the test devices, which were based on MOSFETs, makes analysis of the results difficult. It is held that the dominant source of radiation is the result of direct and phonon-assisted intraband transitions [36] and that holes seem not to contribute to the photon generation mechanism although this possibility cannot be conclusively ruled out [37]. The relatively small contribution to emission due to bremsstrahlung is confirmed by independent measurements [38]. It is interesting to note that, although direct and phonon-assisted transitions seem to be responsible for photon emission, a calculation by the group at *Politecnico di Milano* on this premise finds a marked disagreement between calculations and measured results [39]. Furthermore, unpublished experimental results within our own group at the University of Pretoria show a weak dependency of the emission power on lattice temperature, placing the nature of phonon-assisted intraband mechanism in doubt.

The intraband mechanism has also received critique based on spectral results with no emission presence below the band gap energy of silicon [40] where it is concluded that the dominant process is indeed interband recombination. Akil [41] proposed a multimechanism model, with indirect electron-hole, indirect relaxation intraband and direct intraband events all contributing in various amounts according to a fit of the emission spectrum, while Lahbabhi [42] again leans towards a interband model with material absorption taken into account.

To date, there is currently still no consensus amongst researchers on the exact origin of photons where hot carriers are in transport under high electric fields. This is seen from the widely divergent conclusions *supra*. A closed expression model which is able to predict the quantum efficiency of hot carrier luminescence is still found wanting, which makes optimising the quantum efficiency hot carrier luminescence a difficult task. Although the origin of the photon-generating mechanism is beyond scope of this work, certain observations may help to establish a better understanding of this gap in the body of knowledge. What is well established is that the electroluminescent effect is only seen where high electric fields exist while carriers are in transport under the influence thereof, while the same emission characteristics can be observed by optical pumping as a photoluminescent phenomenon.

2.2.3 Intrinsic bandwidth

Unlike the relatively slow mechanism where electron-hole recombination occurs in direct band gap semiconductors (nanosecond minority carrier lifetimes in bulk GaAs) and the very slow rates at which

recombination occur in indirect band gap semiconductors, hot carrier luminescence is an extremely fast process. It has been shown that the intrinsic bandwidth of the mechanism is beyond 10 GHz [43], where temporal streak camera measurements show oscillatory periods of less than a nanosecond. Our group has also shown direct electrical modulation of hot carrier light sources of up to 1 GHz [44] with electrical detection. This allows for fast operation of the devices and alleviates the need for an additional modulator.

2.2.4 Reliability

The reliability of hot carrier light sources has been reviewed on two occasions with reference to the emission degradation over time. Base-emitter emission was studied over time [45] where it was observed that there is a time dependency of light intensity as a function of time. Significant localised changes were observed, where light spots eventually merged into larger emission areas. However, the junction-averaged quantum efficiency and the spectral energy distribution remained relatively constant indicating no substantial macro-scale degradation of the phenomenon. Accelerated stress tests with larger currents on the same structure by another group show no local effects [46] and the overall light intensity emitted from the device remains constant. It was concluded that most of the conventional accelerated ageing techniques failed to show any degradation of the optical emission from the junction. This is a strong indication that the use of hot carrier silicon light sources can be used as a reliable integrated light source. These findings correlate well with the research within our own group, although our approach is usually to localise the light prior to utilisation as a component in standard CMOS. This is discussed in detail in chapter 4.

2.2.5 Improvement techniques

Although the emission of photons is attractive, the lack of efficiency is one of the biggest obstacles towards practical usage of silicon light sources based on hot carrier luminescence. While most work investigated the mechanism, there has been a research drive for improving the quantum efficiency of *pn*-junctions in avalanche.

During the era of first observation, it was noted that the external quantum efficiency, the rate of photons detected versus the rate of carriers crossing the junction, is in the order of 7×10^{-7} photons/carrier [32]. After CMOS was well established as a semiconductor platform, results from a 2 μm and 3 μm CMOS process showed that external quantum efficiencies can range from 2×10^{-8}

to 6.5×10^{-8} for reverse biased junctions, but can reach values of up to 10^{-4} for forward biased junctions [10]. External power efficiencies, however, were in the order of 10^{-9} due to the large breakdown voltages required to establish avalanche multiplication. In more recent work [47], a 0.35 μm process produced an external quantum efficiency of 3.6×10^{-8} and an external power efficiency of 1.36×10^{-8} using interdigitated structures in a ring configuration. It was claimed in 2010 by a group from the Chinese Academy of Sciences that EQE values of up to 2×10^{-7} and EPE values of up to 7×10^{-7} were possible using field emission devices in Zener breakdown. These results are contrary to our own observations as shown later in this work, as well as in other work within our group [48], which shows that tunnelling current does not contribute towards emission of light.

A trend in the increase of device efficiency has been noted as current density increases [49]. The BEOL-stack in a standard CMOS process is tailored for interconnections between components. The metal structures available can be used as light directing structures, thereby improving the amount of light leaving the chip surface [50]. A recent approach to utilise the gate electrode of a *p*-type MOSFET [51], similar to the approach in [52], showed that an improvement can be seen in quantum efficiency, but at the expense of high operating voltages.

2.3 CMOS INTEGRATION OF SILICON LIGHT SOURCES

The problem addressed by creating a silicon light source is the need of an integrated light source that can coexist with other circuitry on an inexpensive CMOS platform. This is indeed the objective in all the light source technologies mentioned so far. So far, hot carrier luminescence has been the only technology able to be produced using standard CMOS foundries, although porous silicon examples of such a capability have been shown [19], [20]. A number of works have shown that devices fabricated in standard CMOS can be used to emit light [10], [47], [51]–[57]. There are a few examples of where such devices have been successfully integrated as part of a CMOS process as well [58], [59] and our own work [60], [61]. Some of these examples are discussed below.

2.3.1 Characterisation applications

High electric fields and the existence of hot carriers are inherent to modern MOS devices. For this reason, the first practical use of hot carrier light emission was for the characterisation of circuits containing MOS transistors. It was proven possible to evaluate the degradation of MOS devices using the detection of hot carrier luminescence [62]. IBM has focused on using the fast intrinsic bandwidth

of the phenomenon to detect pulses of less than 270 ps for monitoring the switching of transistors in VLSI circuits. This approach was formulated into a technique known as picosecond imaging circuit analysis (PICA) which was used for diagnosing failures on VLSI microprocessors [63]. The advantage of this approach is that separate test structures and methodologies need not be developed, saving valuable chip real estate. The technique is non-invasive and as a result will not influence the operation of the circuit under test.

Another very interesting characterisation application aims at providing functional integrated circuit analysis [64]. It was shown to be possible to make use of hot carrier luminescence for the identification of functions on chip. It is interesting to note that by using engineered inputs it is possible to identify key registers and memory on-chip which can result in some useful (and dangerous) practices in security applications.

2.3.2 Integration of silicon light emitters

Although some works show attempts of using hot carrier luminescent devices in CMOS integrated circuits [65], [66], the state of the art shows very little on the opportunities and challenges associated with CMOS as an integration platform. The situation indicates a need for improving and optimising the way silicon hot carrier luminescent light sources can be used and integrated with adjacent circuitry to coexist in a fully integrated environment. The group at the University of Pretoria has shown approaches for designs in standard CMOS/BiCMOS processes [53]–[55] with localisation of light using point source structures. Field emission was also shown to be a technique for lowering the operating voltage of the devices, although it has been shown that transport due to field emission does not result in optical emission [48]. Although some of the works give an indication of the emitted power, it is held that the optical emission power may have been overestimated. All of the devices described have been manufactured using standard CMOS processes, although all published devices have been tested in isolation without being integrated to interact with the rest of the electronics on an integrated circuit. It is the focus of this work to bridge this gap as well as establish light sources which outperform previous designs.

2.3.3 Optical communications

A major advantage of the spectral nature of hot carrier emission is that silicon can act both as the emitting as well as detecting material. Since there is a substantial amount of spectral content in

the region of energies above the band gap, absorption in silicon provides a way for detecting hot carrier luminescence. A related work within our group is to develop an all-silicon, and eventually an all-CMOS, optical transceiver for allowing integrated optocouplers to become a reality. Current published results show that it is possible to communicate at 10 Mbps with a BER of 10^{-12} [44]. Contactless probing has also shown to be a potential application where hot carrier luminescent light sources provide a means of communicating with an integrated circuit without any physical contact [65], [66]. It was shown possible to modulate silicon light sources at 10 kHz and detect the signal using an avalanche photodiode.

2.3.4 Integrated microdisplays

The possibility of creating an alpha-numeric microdisplay with porous silicon was reported by Fauchet's group [19], [67] where PSi LEDs were integrated with bipolar transistors. Porous silicon has again received attention with an article mentioning the advantages of silicon light emitters in a near-to-eye environment, which greatly reduces the luminance constraints held by conventional display technologies [20]. A successful implementation of commercial suitability remains unpublished.

Hot carrier luminescent light sources have been used in a seven segment display [54] as an experimental concept. This approach offers a more accessible opportunity due to its fabrication using standard CMOS processes, although the state of the art was insufficient for practical purposes.

The current, most successful approach towards integrated silicon light sources as a microdisplay technology has been shown to require image intensification [58]. The work of Chen serves as a useful benchmark for our own work on microdisplays as described in this thesis. Silicon light emitters in avalanche served as the initial emissive elements in the display forming a backplane containing a matrix of all the optical elements. This backplane consisted of an array of 360×200 pixels driven by a 10b current mode driver. The resultant image was intensified using a microchannel plate with a gain of between 10^4 and 10^5 photons/photon. The authors of the work did not, however, mention the luminance value obtained by following this approach, which makes a direct comparison difficult.

2.4 CONCLUSION

Silicon light emission remains a very active field of research. While this is true, the practical applications and the sufficiency parameters for different applications seem to be non-existent. Hot carrier luminescence from light emitting silicon junctions seems to be the most likely candidate for near-future applications based on the fact that light sources can be manufactured using standard CMOS processes. The gap in the body of knowledge remains the origin of the radiative mechanism, large scale integration of light sources coexisting with standard digital components and techniques for improving the power efficiency of hot carrier light sources.

CHAPTER 3

THEORY ON DEVICE OPERATION

3.1 CARRIER TRANSPORT

Carrier transport involves the movement of charge carriers in order to establish electrical current. Electrons and holes are often approached as classical particles which affords them the Newtonian characteristics of position $\mathbf{r}(x,y,z)$ and momentum $\mathbf{p}(x,y,z)$. This six dimensional coordinate system, the phase space, along with its time evolution completely describes the system by tracking the ensemble of all particles. The function $f(\mathbf{r},\mathbf{p},t)$, the *distribution function* then describes the entire system as a probability density function. Solving the Boltzmann Transport Equation yields the distribution function of the system [68].

3.1.1 Boltzmann Transport Equation

The BTE (Boltzmann Transport Equation) is derived by considering a region of the phase space and keeping track of the mechanisms which affect the inflow and outflow of a region in both position and momentum space. Equation 3.1 shows the BTE in implicit form [69] and equation 3.2 shows the expanded form [68].

$$\frac{\partial}{\partial t} f(\mathbf{r},\mathbf{p},t) = 0 \tag{3.1}$$

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \nabla_r f + \mathbf{F} \cdot \nabla_p f = \left.\frac{\partial f}{\partial t}\right|_{coll} + s(\mathbf{r},\mathbf{p},t) \tag{3.2}$$

where $\boldsymbol{v}$ is the carrier velocity, $\mathbf{F}$ is the force acting on the particle and the two terms on the right hand side of the equation denote the inter-particle collision term and scattering term respectively.

3.1.2 Low field transport

For a long time it was sufficient to describe the electronic properties of a semiconductor device with the macroscopic continuity equations given in equations 3.3 to 3.7 [70].

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \boldsymbol{J}_n - r_n + g_n \tag{3.3}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \boldsymbol{J}_p + r_p + g_p \tag{3.4}$$

The first moment of the Boltzmann Transport Equation yields a simplified set of equations known as the drift-diffusion equation. The equations of state, determining the electrostatic solution of the system, is given in equations 3.5 to 3.7.

$$\nabla \cdot E = \frac{\rho}{K_S \varepsilon_0} \tag{3.5}$$

$$\boldsymbol{J}_n = q\mu_n E + qD_N \nabla n \tag{3.6}$$

$$\boldsymbol{J}_p = q\mu_p E - qD_P \nabla p \tag{3.7}$$

The equations above can be used to construct the drift-diffusion equation often used in device simulation [69]. For a given doping profile and knowledge of an external electric field, it is possible to construct an electrostatic solution for typical semiconductor applications.

By examining the solution for n-type carriers, equation 3.6 shows the electron current density as consisting of two components, where

- the first term is the drift component where carriers are subjected to an electric field, E, and behave according to the mobility of the carriers μ_n in the semiconductor material, while
- the second term shows a diffusion term where a gradient of carriers exists, with D_N being the diffusion coefficient of the carrier in the material.

For large devices and moderate electric fields, the constants μ_n and D_N have been experimentally determined. Functioning as proportionality constants, results using field independent mobility fit the behaviour of devices under low electric fields very well [68]. These are often used to solve for simple systems under which the assumptions hold and remains important for providing an initial solution for more complex calculations and simulations when considering systems with larger electric fields or abrupt changes in the potential within the device, such as in small dimension structures.

3.1.3 High field transport

When carrier transport is established through avalanche breakdown in reverse biased *pn*-junctions, the local electric field can reach extremely high values. Under such conditions, parameters such as the mobility become dependent on the local electric field and a number of assumptions used in the classical approximation become invalid.

Thermal equilibrium results in a Maxwellian energy distribution of carriers in the conduction band [71]. As the field rises, the carriers experience a net gain in kinetic energy imparted by the field. The Maxwellian distribution starts to shift due to the net loss of acoustic phonons and the modified distribution can be described by an effective carrier temperature, described by equation 3.8, where μ_0 is the low field mobility, c_s is the velocity of sound, T is the lattice temperature and T_e is the effective carrier temperature.

$$\frac{T_e}{T} = \frac{1}{2}\left[1+\sqrt{1+\frac{3\pi}{8}\left(\frac{\mu_0 E}{c_s}\right)^2}\right] \tag{3.8}$$

Equation 3.8 can then be used to modify the low field drift term in equation 3.9 with the adapted drift term in equation 3.10.

$$v_d = \mu E \tag{3.9}$$

$$v_d = \mu_0 E\sqrt{\frac{T}{T_e}} \tag{3.10}$$

Once the field reaches a certain magnitude, the interaction of carriers with optical phonons results in a saturation of the mobility term and carriers reach a constant velocity irrespective of the applied field. This is referred to as velocity saturation and is a prominent effect in short-channel MOS transistors in modern CMOS processes.

3.1.4 Full band approach

A single electron in a system can be described by the single-electron Schrödinger-equation [72].

$$i\hbar\frac{\partial}{\partial t}\Psi_0(\mathbf{r},t) = -\frac{\hbar^2}{2m_0}\nabla^2\Psi_0(\mathbf{t},t) + U(\mathbf{r},t)\Psi_0(\mathbf{r},t) \tag{3.11}$$

The potential energy term $U(\mathbf{r},t)$ can be expanded into different components as shown in equation 3.12, where $U_L(\mathbf{r})$ is the potential energy the particle experiences due to the lattice itself, $U_S(\mathbf{r},t)$ is a scattering potential describing the incremental potential change due to interactions with, for example, phonons and other particles, while $U_E(\mathbf{r},t)$ is the slow varying potential experienced by the particle due to external applied electric fields.

$$U(\mathbf{r},t) = U_L(\mathbf{r}) + U_S(\mathbf{r},t) + U_E(\mathbf{r},t) \tag{3.12}$$

The potential due to the lattice is the foundation of the crystal electronic band structure which dictates the relationship between the particle energy E and the crystal momentum k. Without giving it the attention it deserves, the U_L-term is usually contained in an effective mass approximation associated with the energy of the carrier based on the crystal $E-k$ relationship, or band structure. The effective mass of a carrier subjected to the lattice potential can be calculated using 3.13.

$$m^* = \frac{1}{\frac{1}{\hbar}\frac{d^2E}{dk^2}} \tag{3.13}$$

This approach enables the BTE to be used in a semiclassical manner where the classical motion of a particle can be modelled, with a mass replaced by the effective mass in equation 3.13. Scattering events are modelled using quantum mechanical techniques such as Fermi's Golden Rule in order to calculate the scattering rates. Equation 3.14 describes Newton's second law with the effective mass replacing classical particle mass [70].

$$F = m^*\frac{dv}{dt} \tag{3.14}$$

The reason why it is important to include a full band structure is that the parabolic approach often used in approximations for high field transport becomes inadequate for investigating hot carrier phenomenon, especially related to light emission.

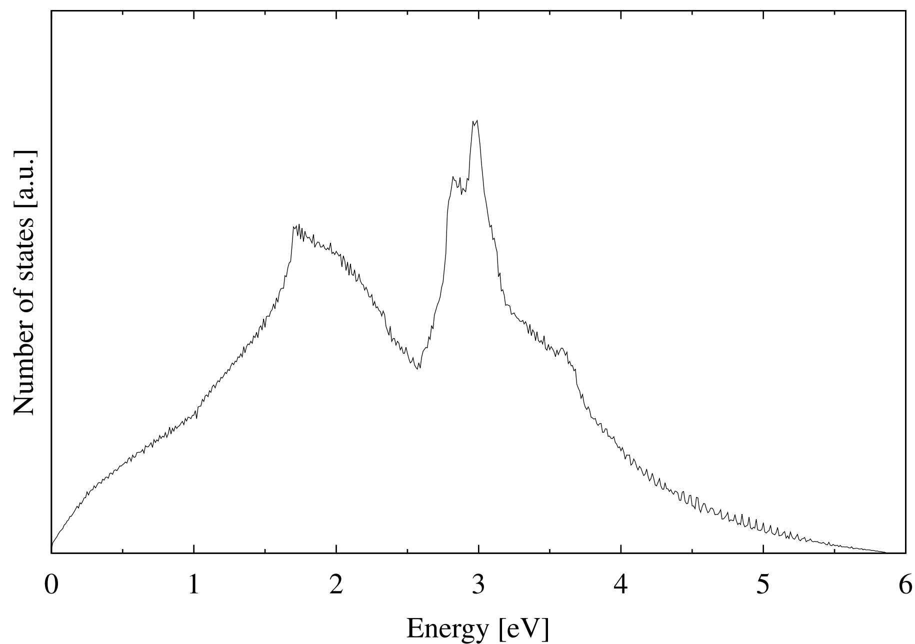


Figure 3.1: Density of states numerically calculated from first two conduction bands in silicon

3.1.5 Density of states

Once a full band approach is adopted, the density of states can be calculated based on a more detailed foundation. The density of states function $g(E)$ is used to calculate the number of available states as a function of energy level. Figure 3.1 shows the relative distribution of available states as a function of energy from the conduction band edge. The plot was numerically constructed using code developed by Professor Bellotti's group at Boston University and produces the number of states for the first two conduction bands using a full band approach.

In order to reorganise figure 3.1 into a more familiar form, figure 3.2 shows the relative number of states as a function of the wavelength corresponding to photon energy. The reason for this representation will become apparent once some of the measured spectra emitted by hot carrier light sources are shown. It is assumed that at least some of the radiative mechanisms rely on intraband relaxation of excited carriers where carriers relax from higher occupied states to the conduction band edge. Note that at $\approx$ 430 nm there is a prominent increase in the density of states corresponding to around 3 eV on figure 3.1.

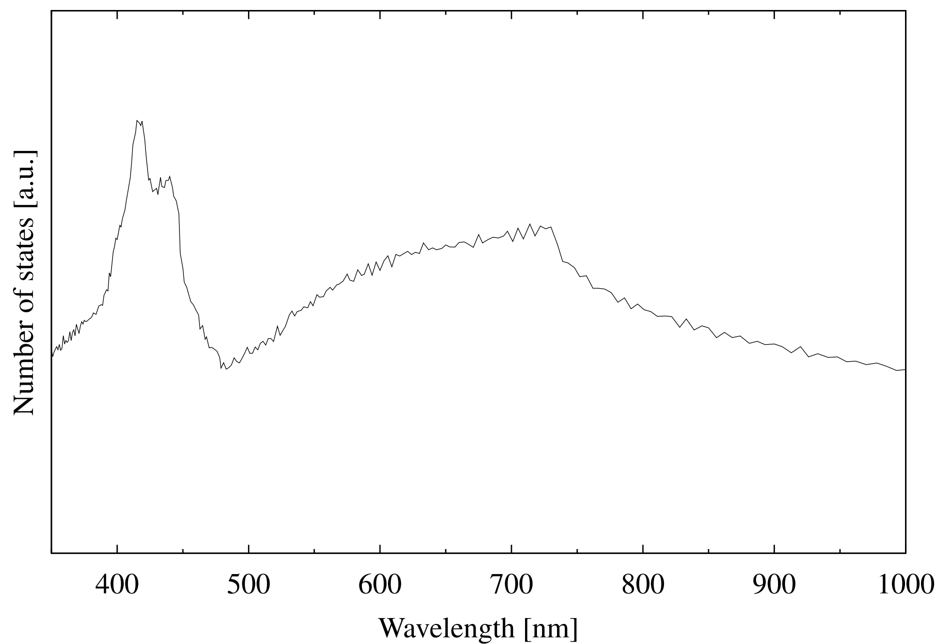


Figure 3.2: Density of states where particle energy is represented as a photon wavelength

3.2 TRANSPORT IN REVERSE-BIASED *PN*-JUNCTIONS

3.2.1 Avalanche multiplication

When a carrier obtains sufficient energy, it becomes possible to lose its energy on collision while exciting another carrier into conduction. This effect is called impact ionisation. It is the main mechanism responsible for avalanche breakdown in *pn*-junctions.

A field-dependent term can be associated with each carrier type:

- $\alpha_n(E)$ is the ionisation rate for electrons and
- $\alpha_p(E)$ is the rate of ionisation for holes.

The quantities M_p and M_n are used to denote the multiplication factor which a carrier experiences on entering the depletion region, resulting in an increased current throughout the device. This term can be quantified by the use of the ionisation rates defined above through equations 3.15 and 3.16 [73].

$$M_n = \frac{1}{1 - \int_0^W \alpha_n \exp\left(-\int_0^x (\alpha_n - \alpha_p)dx\right)} \tag{3.15}$$

$$M_p = \frac{1}{1 - \int_0^W \alpha_p \exp\left(-\int_0^x (\alpha_p - \alpha_n)dx\right)} \tag{3.16}$$

The definition of breakdown voltage is where $M \rightarrow \infty$. Monte Carlo simulations have served as a useful tool in calculating the ionisation rates [74], as well as to perform simulations on actual devices through the semi-classical transport approach. A simple expression for the breakdown voltage in an abrupt junction is shown below [71].

$$V_{BD} = \frac{E_m W_{Dm}}{2} = \frac{\varepsilon_{Si} E_m^2}{2qN}$$

3.2.2 Band-to-band tunnelling and Zener-breakdown

Quantum mechanical tunnelling occurs through a barrier where the wave function of a carrier extends beyond a potential barrier. This means that it is possible for a carrier to "magically appear" on the other side of a potential barrier simply based on its wave function probability. In reverse biased *pn*-junctions, the energy band gap between the conduction and valence band creates such a potential barrier. In a very narrow junction, that is, where the doping on both sides of a *pn*-junction is very high, tunnelling can occur [75]. Depletion region widths in the order of 110 nm were shown to exhibit strong field emission characteristics [76].

An expression for the current generated through tunnelling is given in equation 3.17 where E is the average electric field, m^* the effective mass of the carriers involved and $\hbar$ is the reduced *Planck*-constant.

$$J_t = \frac{\sqrt{2m^*} q^3 E V_{RB}}{4\pi^2 \hbar \sqrt{E_g}} \exp\left(-\frac{4\sqrt{2m^*} E_g^{3/2}}{3qE\hbar}\right) \tag{3.17}$$

The importance of this phenomenon is that avalanche multiplication competes with field emission where the depletion region width is narrow while the electric fields are high, where equations 3.18 to 3.20 describe the different breakdown regimes for magnitudes of the reverse biased breakdown

voltage V_R.

$$V_R \leq 4E_g/q \quad \text{field emission} \tag{3.18}$$

$$4E_g/q < V_R < 6E_g/q \quad \text{mixed mechanism} \tag{3.19}$$

$$V_R \geq 6E_g/q \quad \text{avalanche breakdown} \tag{3.20}$$

Current flow established by field emission exhibits a number of characteristics which include [75]

- softness of the reverse-biased current-voltage relationship,
- higher values for I_s in forward-bias and
- a negative temperature coefficient for the voltage at a constant current.

Since the voltage of an avalanching junction has a positive temperature coefficient [77] due to the increase of phonons affecting the drift mobility term, it can be expected that the temperature coefficients will contend where V_R lies in the region of mixed contributions of field emission and avalanche multiplication.

3.2.3 Minority carrier drift current in reverse-biased *pn*-junctions

Based on the ideal Shockley-equation 3.21, a term J_0 exists where there will be a diffusion-defined drift current flowing even under large reverse-biased conditions.

$$J = J_p + J_n = J_0 \left[\exp\left(\frac{qV}{kT} - 1 \right) \right] \tag{3.21}$$

This diffusion-defined current originates due to the presence of minority carriers at the depletion region edge. Minority carrier charge will lead to drift through the depletion region electric field which renders the concentration right at the depletion region edge zero. The depletion of carriers creates a gradient between the neutral region minority carrier concentration and the region edge, resulting in a drift leakage current. This current can be defined as in equation 3.22.

$$J_0 = q\sqrt{\frac{D_p}{\tau_p}}\frac{n_i^2}{N_D} + q\sqrt{\frac{D_n}{\tau_n}}\frac{n_i^2}{N_A} \tag{3.22}$$

If an abrupt junction approximation is made, the minority carrier concentration on the highly doped side will be negligible and the term can be reduced to the contributing component of the lightly doped side.

3.2.4 Generation and recombination

Another important effect to consider is the generation and recombination of carriers under thermal equilibrium. For the purpose of this work, electrical characterisation of junctions involve analyses of forward- and reverse-biased current-voltage relationships which provide valuable clues about the nature of the junction. In indirect band gap semiconductors, thermal equilibrium is restored through generation and recombination mainly due to bulk traps dependent on electron and hole capture cross sections, σ_n and σ_p respectively [78], resulting in a net transition rate shown in equation 3.23.

$$U = \frac{\sigma_n \sigma_p v_{th} N_t (pn - ni^2)}{\sigma_n \left[n + n_i \exp\left(\frac{E_t - E_i}{kT}\right)\right] + \sigma_p \left[p + n_i \exp\left(\frac{E_i - E_t}{kT}\right)\right]} \tag{3.23}$$

When operating a *pn*-junction in reverse bias, $pn << n_i$ in the depletion region. The low concentration of free carriers will result in a net generation term according to 3.23. The resultant current can be approximated according to equations 3.24 and 3.25 [71].

$$J_{ge} = \frac{q n_i W_D}{\tau_g} \tag{3.24}$$

where

$$\tau_g = \left(1 + \frac{n}{n_i}\right)\tau_p + \left(1 + \frac{p}{n_i}\right)\tau_n \tag{3.25}$$

It is therefore expected that the reverse biased leakage current will exhibit a linear relationship to the depletion region width and the square root of the applied reverse bias voltage V_R according to equation 3.34.

3.3 HOT CARRIER LUMINESCENCE

Conventional light emitting diodes establish photon emission through the recombination of electron-hole pairs. This requires an exciton pair, a localised bounded state of an electron and hole [7]. Unlike conventional LEDs, light emission through hot carrier luminescence requires carriers to traverse through high electrical fields. The easiest way to create extremely high electric fields in semiconductors without excessive current flow is to reverse bias a *pn*-junction, which ramps up the electric field. The increasing depletion region width reduces the diffusion current component. A small leakage current due to presence of minority carriers at the depletion region edge is drifted across the region due

to the presence of the electric field. This small current is however not sufficient for any detectable radiative emission.

3.3.1 Establishing a high electric field

While it is possible to obtain a high electric field within a structure of the same doping kind [38], a reverse-biased *pn*-junction provides for a means of ramping up the electric field without an appreciable amount of free carriers to drift through the region. The only carriers available for drift is the minority carriers at the depletion region edges, as well as thermally generated carriers in the space charge region.

3.3.2 Establishing transport

In order to emit it is necessary to introduce carriers into the high field region. The easiest way of injecting carriers is to increase the electric field to the point where avalanche multiplication starts to occur. This technique establishes increased carrier concentration within the region of high electric field. Carriers are now "injected" by being generated right within the high field region.

3.3.3 Efficiency calculations

Conventional LED efficiency is fundamentally expressed as the internal quantum efficiency, that is, the ratio of photons emitted to carriers injected into the device. Although this metric is the most comprehensive and sets the ultimate limit, it is often difficult to determine. Two very useful figures of merit often used for comparison in literature are

1. the external quantum efficiency (EQE), denoted as η_{eqe} and
2. the external power efficiency, otherwise known as the "wall plug" efficiency, which is the ratio of optical power out (detected) over electrical input power, denoted as η_{epe}.

3.3.3.1 External quantum efficiency

This figure of merit has its origins in conventional LED structures, where it is an indication of the probability of photon production when a majority carrier traverses a junction and recombines as a

minority carrier on the other side.

The energy of a photon is defined as

$$E_{ph} = \frac{hc}{\lambda} \tag{3.26}$$

while emitted power, often called radiant flux, is defined as

$$P_{opt} = \frac{\Delta W}{\Delta t} \tag{3.27}$$

The definition of electrical current is the rate of charge carried through a specific surface, represented as

$$I_e = \frac{\Delta Q}{\Delta t} \tag{3.28}$$

Given the optical power, the external quantum efficiency therefore becomes

$$\eta_{eqe} = \frac{P_{opt}}{E_{ph}} \times \frac{q}{I_e} = \frac{P_{opt}\lambda}{hc} \times \frac{q}{I_e} \tag{3.29}$$

3.3.3.2 External power efficiency

Using external quantum efficiency is misleading when characterising hot carrier luminescent devices since the voltages of these devices are usually high. In fact, this work later shows that it is possible to improve the external quantum efficiency arbitrarily by increasing the voltage of the device. For this reason, the external power efficiency is important to normalise the results for practical use.

The normal way of measuring optical power is by using a detector in conjunction with a radiometer. The detector is usually based on a *pn*-junction of a semiconductor material, which exhibits a certain wavelength dependent responsivity. Optical power absorbed by the detector is converted to current which, with the proper spectral knowledge, can be converted to an optical power representative of equations 3.26 along with 3.27.

External power efficiency is then defined as

$$\eta_{epe} = \frac{P_{opt,out}}{P_{e,in}} \tag{3.30}$$

3.4 PROPOSED TECHNIQUE FOR ELECTRIC FIELD REACH THROUGH

The electrostatic behaviour of an abrupt junction can be used to establish the conditions required for electric field reach through. This section describes an abrupt junction approach in a n^+pp^+-junction

where the highly doped regions act as field stop regions for establishing a condition where the electric field reaches through between the two highly doped regions. Chapter 5 describes the use of the approach developed in this section in junctions available in a standard CMOS process.

3.4.1 Electric field profile in reach through

A one dimensional abrupt junction simplification is used as a starting point for predicting the expected electric field profile. The assumption rests on the fact that highly doped regions used for MOS transistors in a CMOS process form abrupt junctions against the lower background doping of the channel implants and substrate. Figure 3.3a shows an example of the abrupt junction approximation

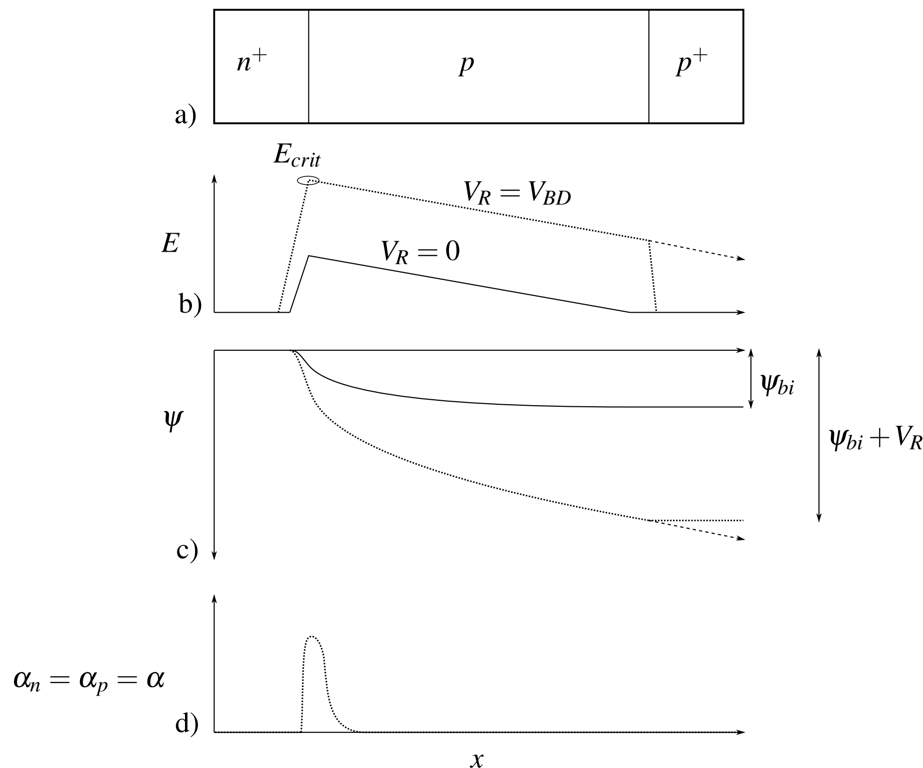


Figure 3.3: Illustration of electric field profile, potential and ionisation coefficient in a n^+pp^+-junction

in one dimension where the electric field, shown in figure 3.3b, without reverse bias follows the predictable triangular shape according to the abrupt junction approximation. As the reverse bias voltage is increased to the point of breakdown, the maximum electric field reaches the critical field strength $E_m = E_{crit}$ and current flows due to impact ionisation. The intent in figure 3.3b is to terminate the

electric field using a highly doped field stop region prior to where the field drops to zero, where the arrowed dashed line shows the behaviour of the field in the absence of the field stop region. Figure 3.3 plots the predicted potential through the device, with ψ_{bi} being the built-in potential of the junction. The dotted line represents the potential under breakdown conditions, where $\psi_{bi} + V_R = \psi_{BD}$. The fact that the region of most impact ionisation occurs where $E = E_{crit}$, shown in figure 3.3d, indicates that the same amount of impact ionisation generated current will flow through the device irrespective of the field stop region.

Figure 3.3b can be approximated in the abrupt junction case by using Poisson's equation 3.31, which relates potential to charge in the space charge region.

$$\nabla^2 \psi = -\frac{\rho}{\varepsilon} \tag{3.31}$$

The built-in potential ψ_{bi} in figure 3.3c can be approximated by regarding the doping concentrations of impurities through equation 3.32.

$$\psi_{bi} \approx \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{3.32}$$

An abrupt junction, although it is an oversimplification of the actual geometric complexity in a CMOS junction, can be used to approximate the maximum electric field E_m using the triangular profile approximation. If the majority of the built-in potential is accumulated on the lightly doped side, equation 3.33 can be used to determine E_m with an applied reverse bias voltage V_R and is mainly determined, except for the ψ_{bi}-term, to the doping concentration on the lightly doped side.

$$E_m \approx \sqrt{\frac{2qN_A(\psi_{bi} + V_R)}{\varepsilon_{si}}} \tag{3.33}$$

The depletion region width for an abrupt junction is approximated in equation 3.34.

$$W_D = \sqrt{\frac{2\varepsilon_s}{qN_A}\left(\psi_{bi} + V_R - \frac{2kT}{q}\right)} \tag{3.34}$$

Note, equation 3.34 depends on the monotonic decrease in electric field existing in the lightly doped region. When the separation distance x_{sep} between the two highly doped regions is reduced to the point where

$$x_{sep} < W_{Dm}$$

(with W_{Dm} the maximum depletion region width) the electric field is terminated "prematurely", resulting in a reduction of the effective depletion region width from $W_D = W_{Dm}$ to $W_D = x_{sep}$. The intended impact of this is discussed next.

3.4.2 Reduced potential integral

The voltage over the terminals of the device is directly related to the potential integral, reduced in equation 3.35 for a one-dimensional case, where the potential is determined by the integral of the electric field as shown in figure 3.3c, assuming that the electric field in the quasi-neutral regions are negligible.

$$\psi = \int_0^x E(x)dx + C \tag{3.35}$$

In the case where a field stop region is used while the electric field is still sloping monotonically towards zero, the integral will result in a smaller potential for the same breakdown conditions around the region of E_m. The effect is that the breakdown voltage can be reduced by reducing W_D through the use of a field stop region, without materially affecting the current generation mechanism. For this thesis, the stated approach can lead to a reduction in breakdown voltage in CMOS junctions for targeting a breakdown regime in the region between the bounds set out in section 3.2.2.

3.4.3 Improving light extraction efficiency

The overall power efficiency can be improved by increasing the amount of light leaving the light source in a usable way. Chapter 5 describes a technique which shapes the surface profile of the Si/SiO_2 interface for improved light extraction efficiency. This directly increases the external power efficiency of the light sources by enhancing the available light leaving the light source.

While it is still unknown from which part of the electric field profile, shown in figure 3.3b, most of the light emission occurs, it may be possible that light emission only occurs where the field is extremely high, in the vicinity of $E = E_{crit}$, which should result in a reduction of voltage without affecting the radiative phenomenon. In theory, this should also result in an improved external power efficiency of the device.

3.5 CONCLUSION

This chapter introduced a number of important concepts to consider when investigating devices which operate at high electric fields. Classical simplified device equations based on the effective mass approximation of parabolic bands will not be able to sufficiently predict carrier behaviour under extremely high electric fields. A full band model is necessary and modelling usually needs an electro-

static solution which serves as a starting point for further full band Monte Carlo analysis. The reach through principle is introduced as a possible means for reducing the breakdown voltage of reverse-biased *pn*-junctions without affecting the magnitude of the current through the device.

CHAPTER 4

A 64 X 8 PIXEL MICRODISPLAY IN STANDARD CMOS

4.1 INTRODUCTION

This chapter discusses the work done to prove the possibility of integrating silicon light sources based on avalanche breakdown and hot carrier luminescence in CMOS integrated circuits. Coexistence of standard CMOS logic circuits and hot carrier light sources are proved by investigating the nature of the light sources, taking into account constraints imposed by the foundry manufacturing process, and implementing methodologies for interaction between digital logic circuits and light emitting devices. The outcome proves that the light emission technology is compatible with CMOS-based circuits, thereby answering one of the research questions forming part of the overall thesis.

The chapter addresses the research question "What are the issues when integrating hot carrier luminescent light sources into standard CMOS while coexisting with other circuitry?" with the implied assumption that it is integrated on the same die. The chapter is broken down into the following sections below.

4.1.1 Target CMOS technology and generalisation

It is necessary to define the definitions and constraints associated with CMOS technology. The definition of CMOS is explained and a target technology is selected on which this work is based.

4.1.2 Light emitting point sources in standard CMOS

The first step in answering the research question relevant to the chapter is to examine how to implement light sources in a standard CMOS process. The design of point sources, light sources generating a light spot at a specific location in silicon, is discussed, with techniques for improving the practical applicability of the light sources given the design freedom afforded to circuit designers making use of a standard CMOS foundry. Aspects relating to the nature and performance of the light sources implemented in CMOS are also investigated.

4.1.3 CMOS integration of point sources

When the light sources have been designed, a method is necessary to allow interaction between the light sources and other circuitry created in an integrated CMOS design. This work looks at the ways that interaction is possible and develops new methods and techniques for allowing such interaction. This is a crucial aspect of the motivation for this research, for if it is possible for light sources to interact and coexist with other CMOS circuits, then the light emission technology becomes a valuable part of a CMOS designer's toolkit.

4.1.4 CMOS microdisplay based on point sources

This section takes state of the art light source technology and the techniques developed in the preceding section and describes how to combine the know-how in order to produce a fully operational CMOS circuit employing light sources. The combination is aimed at realising a fully CMOS microdisplay with optical emission levels sufficient for microdisplay applications without any additional technology required for enabling its use. This is a major step in proving the feasibility of silicon light sources, in CMOS, as a very promising approach for unlocking the potential applications for CMOS as an optical technology in addition to being the dominant electronic technology.

4.2 TARGET CMOS TECHNOLOGY AND GENERALISATION

One of the biggest motivations for silicon light sources is the opportunity afforded to implement light sources in CMOS technologies. Described in section 1.3, CMOS is still today the *status quo* choice for digital integrated circuits and by far the most deployed silicon platform. It is for this reason

that CMOS has a major cost advantage over other semiconductor technologies, an already existing infrastructure and short time to market cycles.

This research is therefore aimed, from the onset, at compliance to the CMOS platform in order to take advantage of the aforementioned advantages. It is subsequently important to define exactly what is meant by a typical, or standard, CMOS process.

4.2.1 Definition of standard CMOS

A good indicator for what a standard CMOS process would be is to limit the target platforms to those suited for the mass manufacturing of microprocessors and FPGAs. The fundamental building blocks of these circuits always centre around transistors in the device layer and interconnects in the back end of line (BEOL) stack.

From a circuit designer's perspective, a process design kit acts as the main interface between circuit and device realisation by the designer, and the foundry's technology capabilities for manufacturing the intended product.

Therefore, in this definition, we restrict ourselves to using what the foundry will offer to a typical digital designer with the aim of mass producing products without altering any of the processing steps.

Any post-processing or other non-standard techniques must be applicable to the end product of such a foundry service, either in processed wafer or individual die format.

4.2.2 Target technology for this research

The target technology used in this work is based on the 0.35 μm CMOS process through **ams AG** (formerly known as *austriamicrosystems* AG), which is a transferred and licensed mixed-signal process developed by TSMC. The devices are targeted for a 0.35 μm single well CMOS technology based in a *p*-substrate. The process uses LOCal Oxidation of Silicon (LOCOS) for lateral device isolation. Interconnects are available through four aluminium metal layers vertically isolated through deposited silicon dioxide. Key features of this process are

- 200 mm (8 inch) baseline,

- p-type wafer,
- polycide gate,
- 0.35 μm minimum drawn gate length,
- 4 metal layers and
- 2.5 V - 3.6 V operating voltage.

The features of CMOS technology in general will be discussed where it impacts designs and structures relevant to this research work, while some mention will be specific on the target process. Due to the non-disclosure agreement between the University of Pretoria and ams AG, details of the target process will generally not be disclosed and assumptions will be based on typical characteristics of a 0.35 μm CMOS process, except where necessary.

4.3 LIGHT EMITTING POINT SOURCES IN STANDARD CMOS

It is known that avalanche breakdown between a highly doped region against a lower doped background can emit light [28]. The same holds true for two regions of high doping where large electric fields are present, although the luminescent phenomenon is influenced by band-to-band tunnelling [48]. In CMOS, the heavily doped source and drain regions against the lighter doping of the channel and substrate will allow breakdown depending on the specific process. Breakdown voltages vary between CMOS processes and depends on factors such as the doping concentrations of the junction regions and their geometries. Breakdown voltages also depend on the lateral isolation techniques used to isolate implanted regions, such as LOCOS and shallow trench isolation (STI), which influence the electric field distribution. For the target technology in this work a minimum breakdown voltage of 9 V is specified and aimed at warning a circuit designer of breakdown expected between the drain-bulk/channel.

4.3.1 Point source design

Although light emission from MOS devices have been studied [30], [31], the presence of a polysilicon gate is not conducive towards light emission from the surface of the die. Furthermore, the formation of granular areas and localised hot spots are observed when creating area and line junctions [28], [32],

[79].

In order to control the breakdown and light emitting region, the breakdown is localised to a single point using a geometry allowable within the typical design rules of a foundry.

4.3.1.1 Process masks

The following typical masks are used for the formation of the point source geometry:

- **ACTIVE** or **DIFF**: This is usually a mask defining where highly doped regions will be implanted in active devices such as transistors. The polysilicon gate will screen implanted ions from entering these regions in the case of self-aligned processes.

- **NPLUS** and/or **PPLUS**: These masks define the species of dopant to be implanted in the active region. Only areas with an active and species definition, for example **ACTIVE** AND **PPLUS**, will receive electrically active dopants. In some (older) processes, the **PPLUS** mask is derived from the **NPLUS** mask as NOT(**NPLUS**) or *vice versa*.

4.3.1.2 Geometrical design

The geometric layout of a point source using masks described in section 4.3.1.1 is shown in figure 4.1 with the important design rules indicated. Deviation from the design rules usually results in variations of the light source performance as features approach the lithographic limits possible through the specific foundry process.

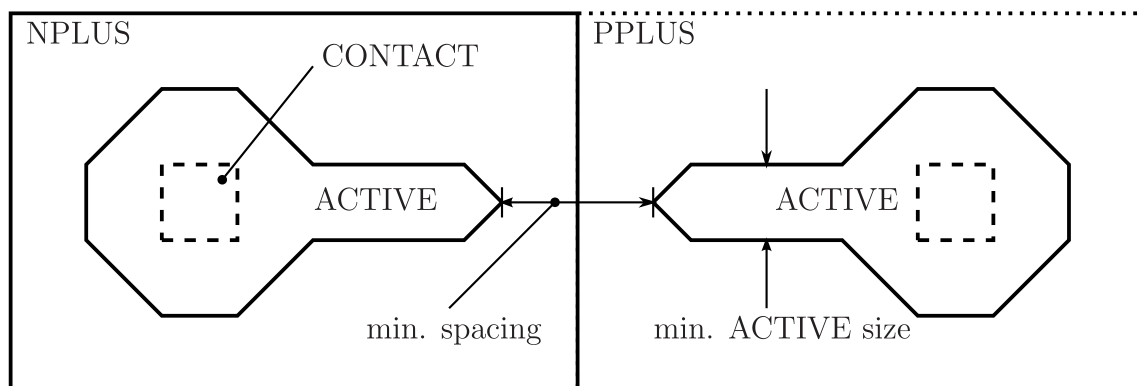


Figure 4.1: Masks used for point source layout

The intended structures are shown in figure 4.2 with a highly doped n^+- and p^+-region against the substrate or well doping in which the structures are formed. The junction between the highly doped region and the substrate of opposite doping kind is then reverse biased to the point of breakdown.

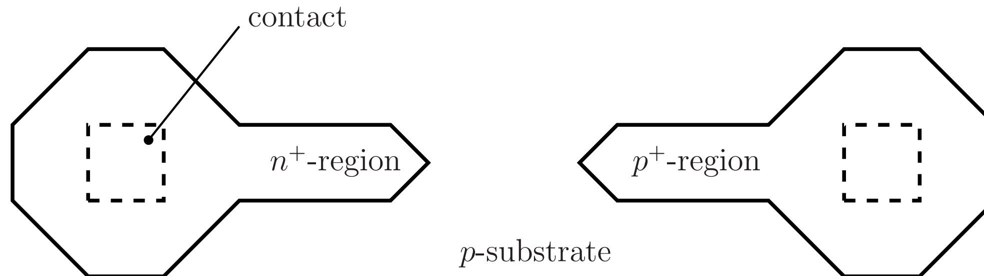


Figure 4.2: Doped structures expected from layout of fig. 4.1

4.3.2 Processing steps for creating point sources

A CMOS process usually starts with a doped wafer of a specific kind, such as a *p*-type in the target process, with substrate of a certain resistivity. An epitaxial layer may be present which serves as the device layer [80]. The ACTIVE or diffusion regions mentioned in section 4.3.1.1 define the highly doped regions for the source and drain diffusions. Although the regions are called diffusion regions, modern processes use ion implantation for better control over the dopant distribution [81].

Figure 4.3 shows simplified steps in a modern CMOS process which is required to create the structures envisaged in section 4.3.1.1 and figure 4.2. Of course, the polarity of the device may differ and be based in an *n*-well, depending on the application. Figure 4.3a starts with a blank wafer, where a *p*-well is formed in 4.3b. In CMOS the intent would be to base nMOSFET-type devices in a *p*-well. Si_3N_4 is used as a mask to block the implantation of Boron dopants in regions which are not intended as *p*-wells. The next definition is that of the active areas, where Si_3N_4 is deposited in areas which will serve as the diffusion areas and the channel region of MOS devices, which will eventually only have a gate oxide between the silicon and the gate. Photoresist may be used, usually derived from the *p*-well structures, to further define windows for field, or channel stop, implants. This will result in regions of slightly higher doping than the background in order to stop depletion regions from spreading laterally and thereby improving lateral device isolation. After stripping of the photoresist, the wafer is oxidised and the Si_3N_4 masks block areas where the growth of LOCOS structures should be inhibited. Channel threshold adjustments are then implanted in the active areas, while blocked by the LOCOS oxidation serving as a mask. This is done to tailor the threshold voltage of MOSFETs.

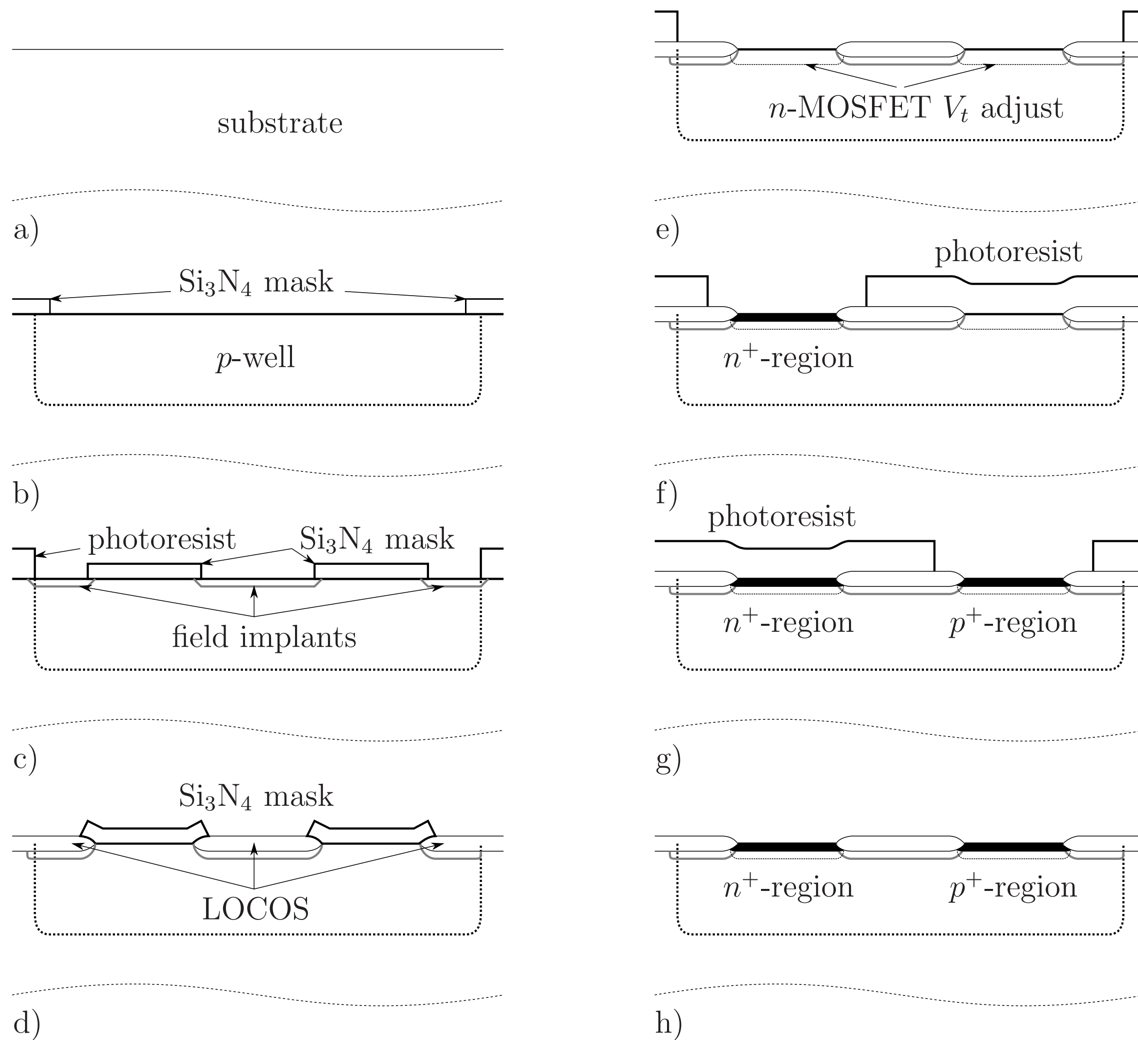


Figure 4.3: Simplified process steps in creating point sources in CMOS

The n^+- and p^+-regions are then defined through steps shown in 4.3f-g, which will result in a cross sectional structure shown in 4.3h. Note that the steps have been simplified, for example, no mention is made of the LDD structures as their importance does not feature in this structure. LOCOS structures are also oversimplified, but elaborated on in section 5.2.3.

4.3.3 Breakdown localisation

It has been shown that localising breakdown yields improved emission intensity, especially if the region of high electric field can be confined to specific light "spots" [52]. Edge effects are prominent where the depletion region experiences curvature [71], [82]. Compared to planar junction breakdown, cylindrical and spherical depletion regions have a reduced breakdown voltage due to an increase of

Table 4.1: Typical minimum specified junction breakdown voltages in CMOS technologies

Technology node	n^+ to p-well	p^+ to n-well
0.8 μm	13 V	13 V
0.35 μm	9 V	9 V
0.18 μm	10 V	7.5 V
0.13 μm	7 V	5.5 V
90 nm	10 V	10 V
65 nm	10 V	10 V

the electric field in the regions. For one-sided abrupt silicon junctions, the reduction in breakdown voltage is given by simple expressions for an ideal case as [83]

$$\frac{V_{CY}}{V_{BD}} = \left[\frac{1}{2}(\eta^2 + 2\eta^{6/7})\ln(1+2\eta^{-8/7}) - \eta^{6/7}\right] \tag{4.1a}$$

$$\frac{V_{SP}}{V_{BD}} = \left[\eta^2 + 2.14\eta^{6/7} - (\eta^3 + 3\eta^{13/7})^{2/3}\right] \tag{4.1b}$$

Equation 4.1 shows the influence of curvature on the breakdown voltage when compared to a planar junction breakdown voltage V_{BD}, with V_{CY} as the cylindrical and V_{SP} as the spherical region breakdown voltages. $\eta = r_j/W_{Dm}$ with r_j being the curvature radius and W_{Dm} the maximum depletion region width of a plane junction at breakdown.

The breakdown voltages typically given in a specific CMOS process as shown in table 4.1 is only indicative and usually represents the minimum point at which breakdown is expected. This will almost always occur where the junction geometry gives rise to a spherical depletion region curvature. In fact, the exact value of junction breakdown is a strong function of the type of geometry employed and this fact can be used to increase the breakdown voltage of transistor drain regions in analogue applications. A curved spherical space charge region is the foundation of the approach used in the design of a pointed structure, shown in figure 4.2, which exploits the fact that breakdown voltage lowers as a function of curvature [82]. The curvature at acute and right angles will be less than obtuse and reflex angles when examining the metallurgical junction. For example, a qualitative examination of figure 4.4 shows that the electric field is compressed for the same potential, resulting in higher fields at the tip of the structure.

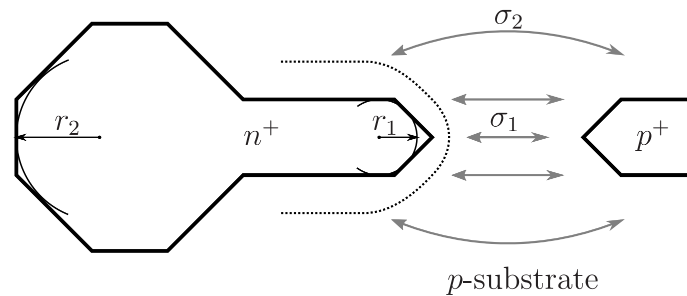


Figure 4.4: Localising avalanche breakdown to tip of n^+-region

Breakdown is then localised to the protruded point of the highly doped region in figure 4.4 for two reasons:

1. The radius of curvature at r_1 is less than r_2 or any other curvature elsewhere in the geometry, and

2. the heavily doped p^+-region provides a least resistance path to ground via the higher conductance in σ_1 than in σ_2 or elsewhere on the geometry.

4.3.4 Optical path for light emission

4.3.4.1 Effect of LOCOS on optical propagation

The region of breakdown as predicted in section 4.3.3 lies at the tip of the drawn structure. The **ACTIVE** definition layer, as described in section 4.3.1.1 is used to block the formation of LOCOS. A Si_3N_4 layer selectively remains in the areas where SiO_2 growth needs to be inhibited [81]. The SiO_2 then also serves as a masking layer for the implantation of dopants forming the highly doped regions.

The highly doped "active" region is bounded by the presence of a vertical SiO_2 isolation feature which inhibits the lateral transport of carriers, where the depth of the vertical extension is the result of local oxidation of silicon which occurs in the unmasked region. This feature has a pronounced effect on the further localisation of breakdown (see section 4.3.3) as well as the optical path associated with photon propagation through the back-end-of-line (BEOL) stack. Snell's Law can be used to estimate the optical path through the field oxide at the Si/SiO_2 interface using equation 4.2.

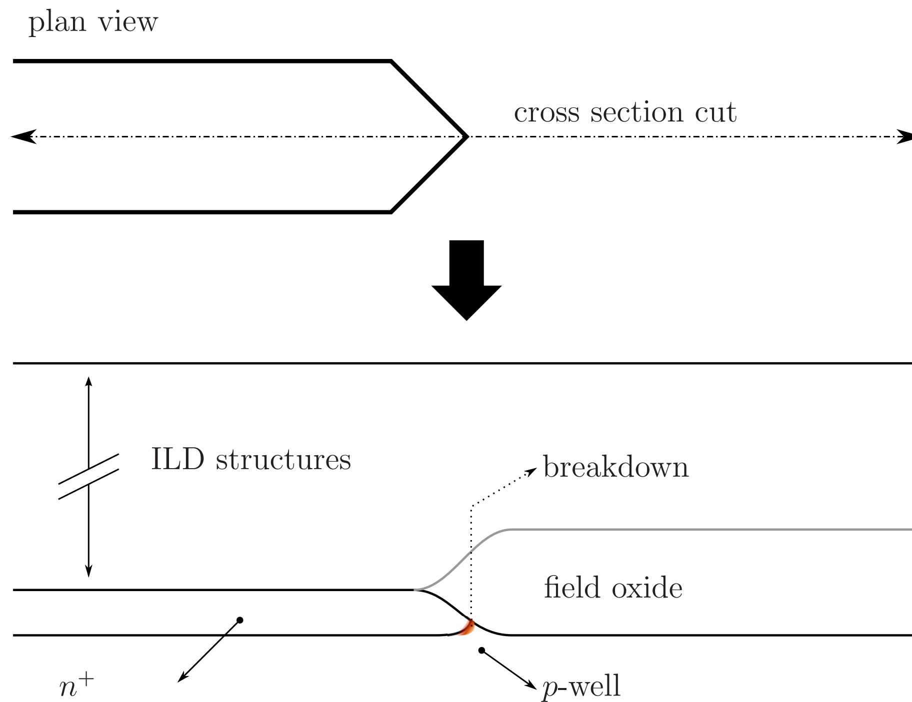


Figure 4.5: A cross section of a n^+ to p-well junction in breakdown showing field oxide profile

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{n_2}{n_1} \tag{4.2}$$

Silicon is absorbent over the wavelength range of interest and the refractive index is complex and a function of frequency. An average refractive index over the wavelength of interest is $n_{si} \approx 4.04$ [13]. The critical angle according to equation 4.2 is then $\theta_{crit} = 21.87°$, while the exit profile is then a cone with an apex angle of $2\theta = 43.74°$ for a point source and simple interface. This can be extended to a 3 dimensional case, where the solid angle of light of the original light point, assumed to be isotropic, is 4π, while the light traversing through the interface can be related to the apex angle 2θ through equation 4.3.

$$\Omega = 2\pi(1 - \cos\theta) \tag{4.3}$$

For the 2 dimensional case, as depicted in figure 4.5, only a fraction of light,

$$\frac{43.74}{360} = 12.15\%$$

of the initial radiant flux is transmitted through the interface. However, the structure cross section in

figure 4.5 is not a simple interface varying in only one dimension and a detailed numerical approach should be employed, as well as taking into account the Fresnel transmission and reflection. From Snell's law it is clear that the light obtains a focused transmission pattern based on equation 4.4 and results in a Lambertian-type emission characteristic.

$$d\theta_1 = \frac{1}{\sqrt{1-(\frac{n_2}{n_1}\sin\theta_2)}} \cdot \frac{n_2}{n_1}\cos\theta_2 d\theta_2 \tag{4.4}$$

Assuming perfect transmission, equation 4.4 show that the power distribution is larger for small values of θ_2 and the beam becomes more focused towards the normal direction of the interface. The light arriving at the interface at an angle of more that the critical angle is simply internally reflected and absorbed. Fresnel-equations can be used to further refine the actual transmitted power distribution and radiation pattern, but transmission through the BEOL-stack still needs to be taken into account.

4.3.4.2 Propagation through the BEOL-stack

The angle of light direction in section 4.3.4.1 depends heavily on the angle of the LOCOS interface with respect to normal surface direction of the silicon substrate. The BEOL-stack consists of layers with dimensions as shown in table 4.2. For the refractive indices of the BEOL-stack materials, $n_{\mathrm{SiO_2}} \approx 1.49$ while $n_{\mathrm{Si_3N_4}}$ varies between 2 and 2.5 over the wavelength region of interest [13]. Where silicon nitride is used, rays travelling through the BEOL-stack and, specifically, the SiO_2 ILD material will simply be bent in slightly and rescattered at the Si_3N_4/air interface. The passivation layers can be removed by defining a pad opening etch using standard design rules. RIE is usually employed for etching, although this depends heavily on the exact passivation materials, as well as the foundry processes itself. For this reason, the difference in refractive indices of SiO_2 and air will be the determining factor in how much optical power leaves the surface.

The critical angle for SiO_2/air is

$$\arcsin\left(\frac{1}{1.49}\right) = 42.15^\circ$$

and 23 % of the optical power emitted by an isotropic point source will leave the surface. From section 4.3.4.1 it is clear that the isotropic assumption does not hold, while the directionality of the light emitted at the LOCOS region has to be taken into account. These limited angular ranges of light transmission severely limits the extraction efficiency in CMOS. A circuit designer is typically not armed with manipulating standard processing steps, but can make use of other BEOL structures for improving extraction efficiency (see section 4.3.5).

Table 4.2: Intermetal dielectric layers and associated thicknesses in the BEOL-stack

ILD layer	Associated metal	$t_{average}$ [nm]	Material	Planarised
Field oxide	-	290	thermal SiO_2	no
ILDFOX	-	645	dep. SiO_2	yes
IMD1	MET1	1000	dep. SiO_2	yes
IMD2	MET2	1000	dep. SiO_2	yes
IMD3	MET3	1000	dep. SiO_2	yes
TPROT1	Top MET	1030	dep. SiO_2	no
TPROT2	-	1000	dep. Si_3N_4	no

4.3.5 Light directing structures

One possible way of improving the optical extraction efficiency, both in the view of the LOCOS structure and the angle of transmission through the BEOL-stack is by using the metal and via structures available to alter the optical path [50]. Ray tracing confirmed the results derived in 4.3.4.1 for light entering the BEOL-stack. By employing the metallisation layers and **VIA** plugs as reflecting surfaces in redirecting the optical power outside the SiO_2/air acceptance angle, it is possible to improve light extraction efficiency by factors up to 3.9.

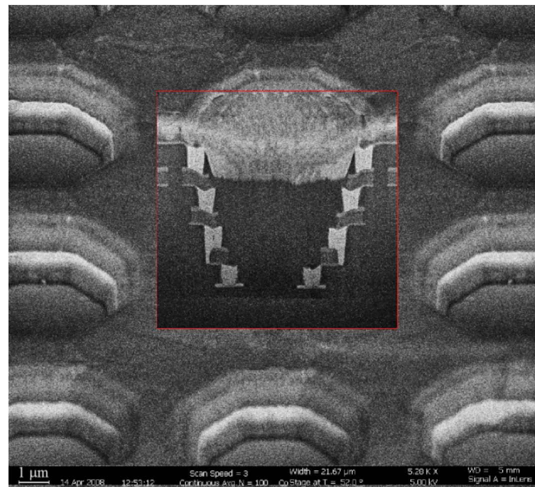


Figure 4.6: A cross section representation of the light directing structure using BEOL-stack metal and via layers

4.3.6 Arraying of point sources

Since the light source transport mechanism is based on impact ionisation and avalanche multiplication, it is possible to use the point sources in arrays connected in parallel. The breakdown voltage determined by avalanche multiplication has a positive temperature coefficient [77], [84]. This is mainly due to the difficulty carriers have in gaining sufficient ionisation energy for impact ionisation due to the increased presence of phonons [71].

This phenomenon is advantageous when connecting point sources in parallel. If local variations exist that cause a spread in voltage for the same current, a device which allows more current at the same potential will heat up locally faster than points with a higher voltage threshold. This in turn will raise the operating voltage in the heated up device, allowing larger currents to flow through the other point sources, thereby creating a feedback behaviour and stabilising the point sources when driven in parallel. This technique then allows

- increasing the optical output of a light source by utilising multiple individual point sources and driving the matrix with larger currents, and
- changing the geometrical area of light emission by placement of individual point sources in arbitrary shapes.

This allows adaptation to fibre cores for short haul communication systems, or for pixel shapes in microdisplay applications.

4.3.7 Radiation pattern

Figure 4.7 shows a cross section of the radiation pattern of light emitted by a typical point source discussed in section 4.3.1 with an angle $\theta = 11.4°$ tilt from the normal of the chip surface. The power in this narrow beam is reduced by the angle of the LOCOS feature against which light emission takes place at the Si/SiO_2 interface and the limited exit angle to to the SiO_2/Si_3N_4/air interface.

The radiation pattern of the beam emitted from the chip surface has been measured in three dimensions as shown in figure 4.8. The majority of the radiant flux is lost at the Si/SiO_2 interface with a limited amount of flux travelling through the BEOL-stack and leaving the chip surface.

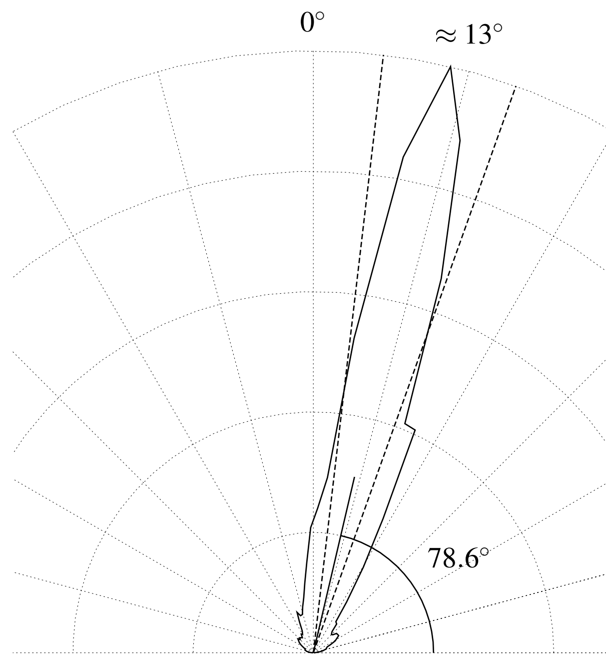


Figure 4.7: Normalised radiation pattern of point source without light directing structure

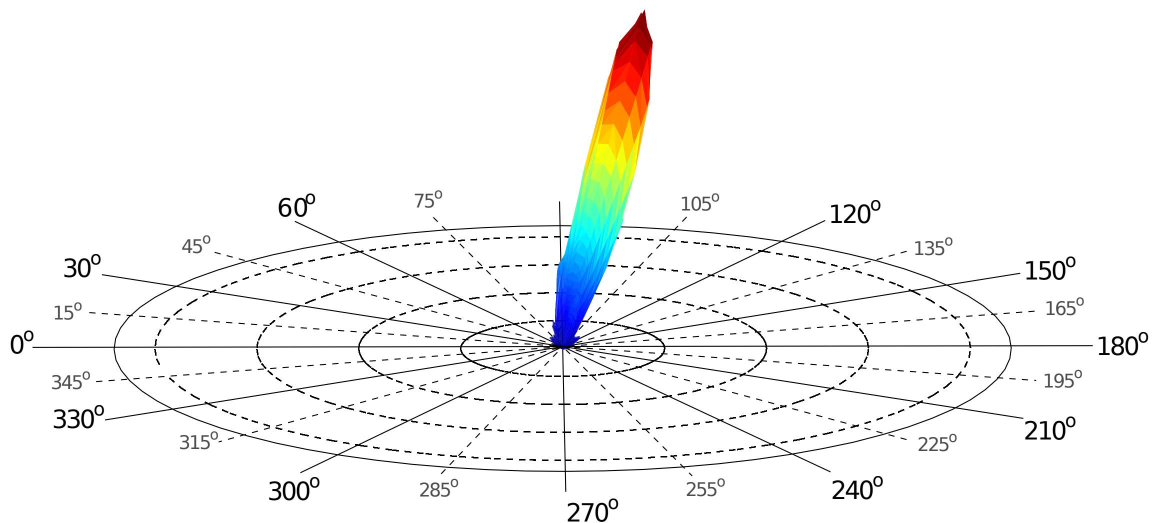


Figure 4.8: Three dimensional radiation pattern of a point source array showing tilt towards the breakdown localisation point

4.4 CMOS INTEGRATION OF POINT SOURCES

One of the research questions and an important part of this research is to prove the possibility of integrating the light sources, discussed in section 4.3, as a functional component and reusable part of

a component library in a standard CMOS process, without the need for process modifications. This work addresses important considerations when moving towards integration of silicon light sources based on hot carrier luminescence with other circuits operating at logic level voltages.

4.4.1 Well structure and available diodes

A highly doped n^+- or p^+-region against a *p*- or *n*-well doping orders of magnitude lower allows an approximation of an abrupt junction when solving the electrostatic system at steady state. Given the typical doping characteristics of a modern CMOS process, this also allows avalanche breakdown at reasonable voltages. However, the combination of *pn*-junctions available is limited by the way the process is constructed.

The target technology, introduced in section 4.2.2 for this work is based on a *p*-type wafer with a *p*-type epitaxial layer forming the substrate. This facilitates the implementation of *n*-type MOS devices directly in the substrate without creating additional wells. The complementary *p*-type MOS devices are realised by introducing an *n*-well implantation step to form the *n*-wells.

This necessarily limits the substrate voltage to the most negative potential when integrating the light sources with other CMOS circuits, such as digital circuitry. The anode of an n^+p-type diode will then typically always be grounded and at the same potential as the bulk of *n*-type MOSFETs used in the rest of the circuit. The only way to establish avalanche breakdown is by applying a high voltage to the cathode. In contrast, devices based on a p^+n-topology is created in an *n*-well, which may float, allowing an extra degree of freedom in terms of the voltage constraints on the diode terminals. As long as the *n*-well potential remains reasonably high, the possibility of creating a diffusion current between the *n*-well and *p*-well can be managed.

4.4.2 Limitations on operating voltage of MOS devices

4.4.2.1 Avalanche breakdown of drain region

The same mechanism allowing light emission is also responsible for device failure when operating transistors at high enough voltage levels. The drain also forms a *pn*-junction between the highly doped region and the bulk. As the voltage on the drain increases with respect to the bulk, avalanche breakdown starts to occur and conduction through the drain-bulk interface effectively renders the

MOS transistor inoperable. The levels in table 4.1 are indicative of when to expect breakdown. These are similar to the point source breakdown discussed, although the specific values depend on geometry. This effect is undesirable as the MOS device fails to operate as a transconductance element.

4.4.2.2 Gate oxide

It is obvious that the quality of a MOS transistor depends heavily on the isolation properties of the gate oxide. Given the higher voltage nature of integrating hot carrier light sources in a CMOS process, it is important to understand how interaction of high voltages with gate oxides may result in device failure.

Two general metrics exist in characterising the breakdown strength of gate oxides: time zero dielectric breakdown (TZDB) and time dependent dielectric breakdown (TDDB). The study of breakdown kinetics is a complicated one, but the instantaneous dielectric breakdown of thermal oxide happens at electric field values of around $E_{g,BD} = 10$ MV/cm [85]. Moreover, the area of the capacitor influences this breakdown [86] as less defects and weak spots exist in smaller exposed areas.

Typical gate oxide thickness of the target process is $t_{g,ox} \approx 7.6$ nm. While standard voltages are applied to the gate, the maximum field over the gate oxide $E_{g,max}$ will remain within normal limits. When driving high voltage regions using the drain of an NMOS as current sink, the drain voltage may exceed normal CMOS operating levels. As there is a small overlap between the drain and gate oxide, this region may be stressed due to the high electric fields present. Given the typical dimensions of gate oxides and the fact that 10 MV/cm = 1 V/nm, the process used allows roughly 7.6 volt of difference between the drain and gate. Given table 4.1, this may lead to reliability problems. The estimate is very conservative, though, due to the small overlap in self-aligned CMOS processes (mainly lateral straggle of implanted ions [81]), but it may be worth investigating the effect of TDDB.

4.4.2.3 DIBL and punch through

Drain-induced barrier lowering (DIBL) happens when the depletion region of the drain extends to the depletion region of the source in a MOSFET [71]. A strong leakage current will then flow between the drain and source of the transistor due to a lowering of the barrier at the source. A localised difference in carrier concentration results in higher injection from the source which is then collected by the drain, causing a drain-voltage-dependent flow of current. As the potential of drain nodes may

be high when interfacing with light sources, this is an important consideration when integrating hot carrier light sources with MOS devices.

4.4.3 High voltage integration

CMOS integration of hot carrier luminescent light sources requires a method to utilise logic level inputs to the driving circuitry of the light sources in avalanche. This task presents a difficulty as the voltage levels required for avalanche breakdown of the junctions under discussion are usually much higher than the designed operating voltages of the transistors used for circuit implementation.

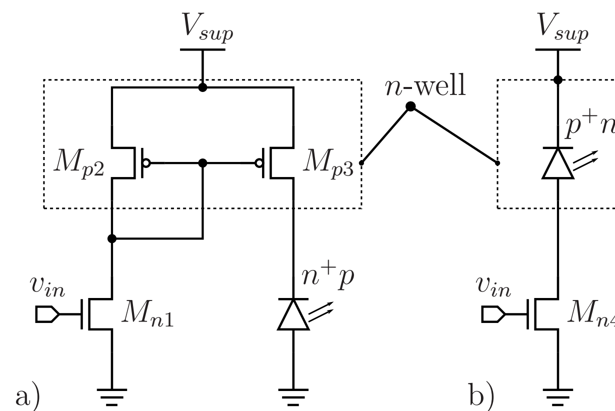


Figure 4.9: Driving configuration for different polarity diodes

4.4.3.1 n^+p-type light sources

Figure 4.9a contains an n^+p-junction which can be reversed biased to establish avalanche breakdown. As mentioned in section 4.4.1, the anode of the diode is fixed to the same substrate as n-type MOSFETs in the circuitry. A mechanism is required to apply a voltage of sufficient magnitude to the diode cathode for breakdown. This is not directly possible using only n-type MOS devices.

One possible topology as shown in figure 4.9a involves "redirecting" the control signal applied on M_{n1} via a PMOS-type current mirror seated in an n-well at sufficient potential for breakdown. The input signal v_{in} is applied to the gate of M_{n1} at levels compatible to the rest of the logic circuitry on chip. In the target process, these values lie between 0 V and 3.3 V. The input signal may be generated through current controlled biasing generating the gate voltage, in which case the I_{D1} will be well defined. Alternatively, logic levels can directly be applied as fed from the output of logic gates, given suitable fanout characteristics. The current is then controlled by device sizing if operated in saturation

or remains a function of V_{DS1} in the triode region of operation. I_{D1}, the drain current of M_{n1}, is shared with M_{p2} and can be scaled depending on the ratio of sizing between M_{p2} and M_{p3} and the overdrive voltages of the devices.

$$V_{sup} = V_{DS1} + V_{SG2} = V_{SD3} + V_{n^+p} \tag{4.5}$$

Upon inspection of equation 4.5 it is clear that the requirement

$$V_{sup} \geq V_{n^+p} + V_{ov3}$$

with $V_{SG2} = V_{SG3}$ as determined by the driving current I_{D1}, $V_{ov3} = V_{SG3} - |V_{tp3}|$ (with $|V_{tp3}|$ as M_{p3}'s threshold voltage magnitude) can easily be calculated for a specific setup, while $V_{n^+p} \approx 9$ V in the target process. This is problematic for the reasons stated below.

1. V_{sup} needs to be greater than the breakdown voltage of the n^+p-junction.
2. The moment M_{n1} is turned off, the voltage V_{SG1} will reduce to zero.
3. This, in turn, will cause V_{D1} to rise to V_{sup} and cause breakdown at the drain of M_{n1}.
4. Although not the full intended current, the branch will settle at steady state with a non-negligible residual current, which will be reflected to the light source's side.

It may be possible to alleviate the problem by trying to adapt the geometric design of the driver NMOS M_{n1} in order to increase the breakdown voltage, although the solution may not be sufficiently robust for practical purposes.

4.4.3.2 p^+n-type light sources

Figure 4.9b shows the complementary configuration with a highly doped p^+-region embedded in an n-well. The main difference is that this structure is floating and affords the designer with an additional degree of freedom. The driving transistor M_{n4} can be driven exactly in the same manner as described in section 4.4.3.1.

While on, M_{n4} will act as a current sink from V_{sup} via the p^+n-junction in breakdown, with a voltage requirement of

Table 4.3: Comparison of junctions for integration suitability

Junction	V_{BD}	Area	Complexity
n-well/p-well	> 30 V	large	low
n^+p	≈ 9 V	large	high
p^+n	≈ 9 V	small	low

$$V_{sup} \geq V_{ov4} + V_{p^+n} \tag{4.6}$$

As long as M_{n4} conducts, V_{D4} will remain within a reasonable value of the required V_{ov4} typical of digital circuit operation. Once M_{n4} turns off though, the nodal voltage V_{D4} will start to rise, thereby preventing conduction through the light source in breakdown. Note that V_{D4} cannot rise to the point of drain breakdown in M_{n4} as this requires the light source junction to conduct. This in turn requires the necessary overhead voltage to support breakdown. Instead, the nodal voltage will settle at a value where the leakage currents of M_{n4} and the light source junction are equal.

4.4.3.3 Comparison of approaches

Based on the analysis in sections 4.4.3.1 and 4.4.3.2 and taking into account that an n-well/p-well structure would probably create an electric field of insufficient magnitude to excite carriers to the appropriate energy levels for radiation, table 4.3 shows a summary of integration suitability of the junctions. Point sources based on a p^+n-junction are preferred, as long as the light emission intensity does not suffer. The light intensities are discussed in section 4.5.1.

4.4.4 Current densities and electromigration of metal

The transport of metal mass under the influence of high current densities was identified as a potential failure mechanism during the 1960s in semiconductor devices where interconnects were made using thin metal layers. Atoms of metal are physically influenced by momentum transfer of electrons to activated metal ions. Black's Rule [87] is shown in equation 4.7, where MTTF is the mean time to failure, A is a constant containing a quantity describing the cross sectional area of the metal, J is the current density, ϕ is the metal ion activation energy in eV, k is Boltzmann's constant and the

Table 4.4: Average sheet resistances of different structures in the target process

Layer	Resistance	Unit
n^+-diffusion	75	$\Omega/\square$
p^+-diffusion	140	$\Omega/\square$
Typ. metal	70	$m\Omega/\square$
Met-n^+ contact	30	Ω/cnt
Met-p^+ contact	60	Ω/cnt
Via	1.2	Ω/via
Top metal	40	$m\Omega/\square$

temperature T is given in Kelvin.

$$\frac{1}{\text{MTTF}} = AJ^2 e^{-\frac{\phi}{kT}} \tag{4.7}$$

Foundries usually have design rules specifying the maximum amount of current per width of the metal track in order to mitigate the effect of metal migration. Disregarding the guidelines may cause breakage of metal tracks and discontinuity in interconnects resulting in failure. For the target technology, a typical value of 1 mA/μm track width is defined. A logical improvement is to stack metal tracks, with sufficient vias, in order to improve the current carrying capability without compromising area. By combining this technique with the light directing structures mentioned in section 4.3.5 it is possible to carry large amounts of currents to the light sources. The combination of light directing structures and providing the power supply path with sufficient conductors is discussed in section 4.5.2.

4.4.5 Series resistance

The interconnect material resistivity impacts the electrical characteristics of the light sources and cannot be neglected when applying large currents to the light sources. As the material thickness is fixed, the length and width can be used to design interconnects and to characterise its resistance. It is also important to factor in the resistance per via, as well as the diffusion region resistances when entering the device layer. Table 4.4 shows a number of sheet resistances with importance to light source design in CMOS.

4.5 CMOS MICRODISPLAY BASED ON POINT SOURCES

The research question posed in section 1.5 is discussed in this section where point sources described in 4.3 are integrated onto a single die, while coexisting and interacting with digital circuitry which form part of the foundry's standard cell library set. As the light sources require sufficient voltages for avalanche breakdown, considerations discussed in section 4.4 are taken into account.

The chosen implementation for proving CMOS integration was a CMOS microdisplay, where controller circuitry are integrated with pixels utilising hot carrier silicon light sources using a standard CMOS MPW run. Modern microdisplay technology is currently dominated by OLED and LCD technologies, but neither of these technologies have the ability for providing the scale of integration offered by CMOS. Although hot carrier luminescent light sources do not match the emitted optical power of OLEDs, it is proved in this work that sufficient light levels exist for naked eye observation. Previous work on integrating microdisplays in silicon has been attempted. Porous silicon light sources can act as active elements in the emissive display [20], while other work based on hot carrier light sources uses MOS-based devices for emission and amplifies the optical signal with an external image intensifier for observation to become possible [58] .

The light sources in this work, as well as the techniques used in integration and control of the active matrix, show that optimised design can result in usable microdisplays with a major cost advantage to the commercial OLED and LCD implementations.

4.5.1 Point source characteristics

A typical point source as described in section 4.3 is limited in the amount of current that can pass through the device before failure occurs. Furthermore, localisation of the point of breakdown suffers as the electric field over the conductive path σ_1 in figure 4.4 becomes non-negligible and breakdown starts to happen at other regions of the structure. This leads to a drop in efficiency [49] as a function of device current. This effect is noticeable in figure 4.10 at around 100μA of current per point source, where the area of efficient operation can roughly be split into two according to equation 4.8:

$$\left.\frac{\partial P_{opt}}{\partial I}\right|_{I<100\mu\mathrm{A}} > \left.\frac{\partial P_{opt}}{\partial I}\right|_{I>100\mu\mathrm{A}} \tag{4.8}$$

Note that this effectively means that it is more efficient to operate the devices at low currents, which results in a trade off around microdisplays which is discussed in section 4.5.2.

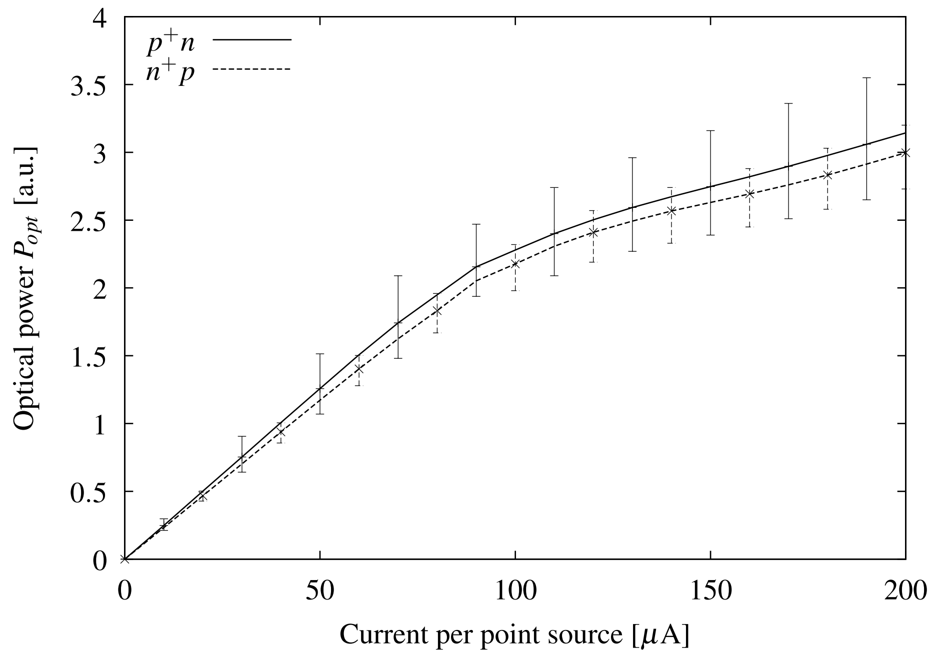


Figure 4.10: A comparison of average light emission from p^+n- vs. n^+p-junction light sources

Figure 4.10 shows the average power emission over a number of samples as detected by a radiometer, along with the associated spread represented by error bars. It can be seen that the p^+n sources perform marginally better than the n^+p sources, although the spread in the former for different ICs is larger. Initially, the transfer characteristic is nearly linear. At around 100 μA, as explained above, a change in slope is seen for both devices, where after the slope decreases but still exhibits a linear trend. This effectively means that current can be increased proportionally to the required optical power P_{opt}.

4.5.2 Pixel design

The pixel pitch was set at 50 μm. Although the power efficiency of point source based light sources are better at lower currents (see section 4.5.1), a trade off exists between the luminance (cd/m^2) of the active display area and the efficiency quiescent operating point of the light source. The former prefers a small area with higher optical power output, while the latter benefits from low driving current per point source, thereby requiring a larger area for the same output power.

The choice of 50 μm, along with integrated light directing structures as in section 4.3.5, allows a packing density of 30 point sources per pixel. The quiescent point for the light sources was set

at ≈ 170 μA for an overall maximum pixel current draw of 5.1 mA per pixel. An n-type MOSFET driver was implemented with the required transconductance capability in order to interact with buffers available in the foundry's standard cell library.

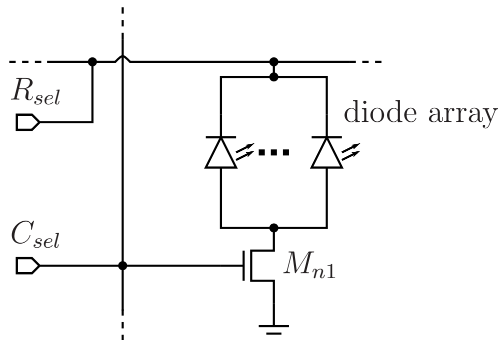


Figure 4.11: Pixel circuit based on an array of p^+n-junction light sources

Metal layers 1 and 2 are used for creating both the high current supply rail and for controlling which rows are energised. For this reason, continuity is required in a horizontal direction when butting the pixels at specification pitch. The metals are both used for conduction but a separation ring is formed between the supply rail metal and the light directing structure. This is due to the nature of the structure which requires both vias and metal layers, effectively connecting the structure as a single node [50].

Metal layers 3 and 4 are used for grounding the structure and to serve as a current return path for the current sunk through the driver. The metals also form part of the light directing structure and the increased coverage allows for lower resistivity of the conduction path. Continuity is established in vertical channels as the pixels are butted together. The column select control signal is also passed through vertical conducting paths adjacent to the ground node, enabling access to the drivers in a column-wise fashion. Figure 4.12 shows the CAD layout of the pixel with a demarcation of the light source array and driver. Each pixel is designed to be driven at a nominal current of 5.1 mA, but may vary as a function of power supply due to the channel length modulation of the driver MOSFETs.

Section 4.4.3.2 states that the off-state nodal voltage of the drain of the driving transistor will depend on the leakage current components acting on the nodal capacitance. Figure 4.13 shows the leakage current components while ignoring the n-well to substrate leakage as the latter does not influence node x. C_x represents the capacitance on the node, while v_x is the voltage.

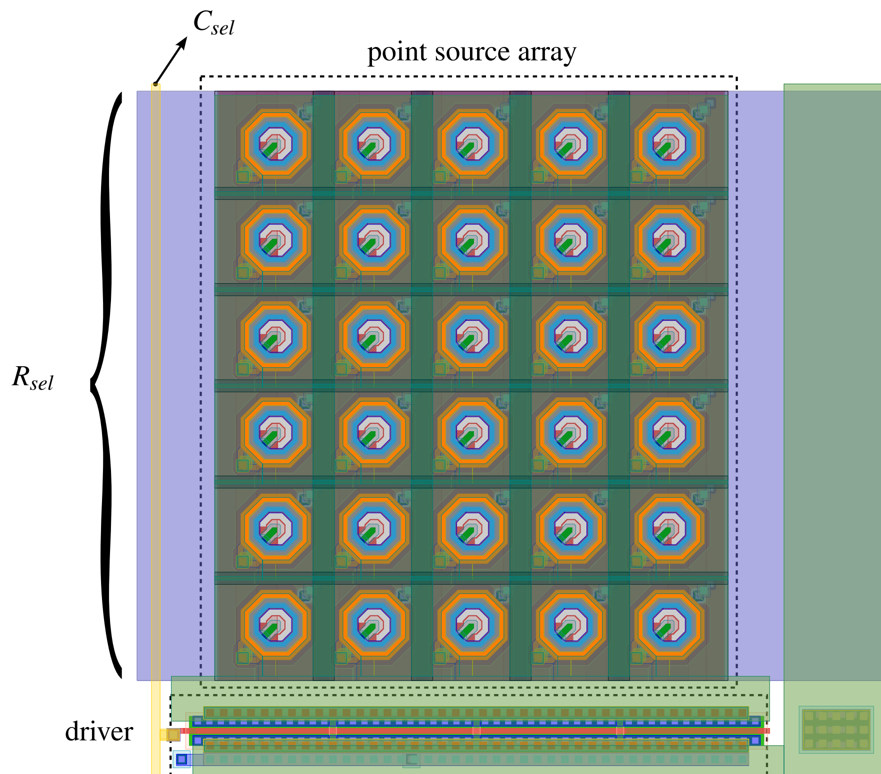


Figure 4.12: Layout of pixel circuit for design in figure 4.11

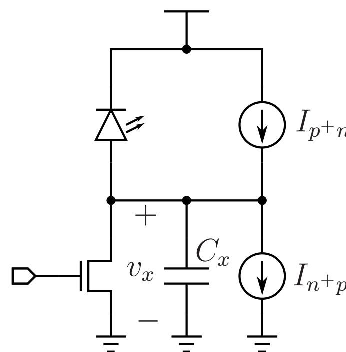


Figure 4.13: Leakage current sources modelled as extrinsic to the junctions

Ideally, the ratio $I_{p^+n}/I_{n^+p} \leq 1$ for a net extraction of charge on C_x in order to keep v_x as low as possible during the transistor off-state. From table 4.5 it is apparent that this ratio ≈ 1.04, which means v_x will slowly rise as a function of time depending on the size of $I_x = I_{p^+n} - I_{n^+p}$. Given that $C_x \approx 150$ fF, the nodal voltage will rise as given in the expression below. As the expression shows that the charge rate is a lot smaller than the typical refresh rate of the display, it is held that v_x will

Table 4.5: Leakage current component as specified in the foundry's process documents

<table>
<tr><th></th><th>Point source</th><th>Array</th><th>n-MOSFET</th></tr>
<tr><td>Perimeter</td><td>5.312 μm</td><td>159.36 μm</td><td>81.6 μm</td></tr>
<tr><td>Area</td><td>1.165 μm^2</td><td>34.95 μm^2</td><td>32 μm^2</td></tr>
<tr><td rowspan="2">I_{p^+n}</td><td colspan="2">0.37 aA/μm</td><td>-</td></tr>
<tr><td colspan="2">0.28 aA/μm^2</td><td>-</td></tr>
<tr><td rowspan="2">I_{n^+p}</td><td colspan="2">-</td><td>0.61 aA/μm</td></tr>
<tr><td colspan="2">-</td><td>0.51 aA/μm^2</td></tr>
<tr><td>I_{total}</td><td>2.292 aA</td><td>68.75 aA</td><td>66.10 aA</td></tr>
</table>

remain within safe operating limits.

$$\frac{\partial v_x}{\partial t} \approx 17.7 \quad \left[\frac{\mu V}{s}\right]$$

4.5.3 Scanning architecture

The planar nature of CMOS technology necessitates sharing of surface area for both light emission as well as supply and control interconnects. This not only limits the fill factor of the active display matrix, but also introduces difficulties in addressing random pixels independently. The display is emissive in nature as it actively produces light only when energising a pixel. Addressing a complete single column is possible while selectively energising the rows as required. For this reason, the scanning architecture was chosen by sequentially stepping through each of the 64 columns, while energising the specific rows where the pixel needs to be turned on. This is done by encoding a 6-bit word to address each of the 64 columns individually. The 6-bit word is encoded in such a way that a word of value 000000_b addresses the first column, 000001_b the second and sequentially through to 111111_b for the final column, as shown in the truth table in table 4.6.

The scanning sequence is designed for ease of interfacing with a digital counter when driven off-chip, which allows a normal incremental counter to scan through the matrix, one column at a time. This sequence can then be synchronised by energising specific rows in order to create a correlated picture or representation. This effectively means that a pixel in a static "on state" will have an energising duty cycle of $\frac{1}{64}^{\text{th}}$ of the period of one complete scan cycle. The scan cycle, or refresh rate, can be set arbitrarily up to the frequency limit of the decoder and column addressing circuitry.

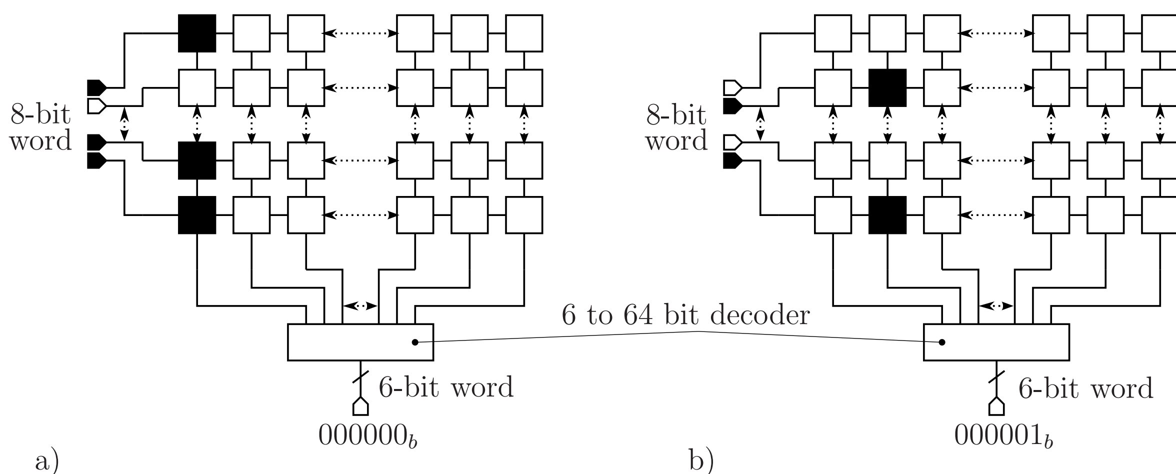


Figure 4.14: An illustration of the scanning sequence with column select and selective row energising

Figure 4.14 shows a representation of the method used for rendering information on the display, where figure 4.14a shows the first column addressed with specific synchronous row inputs, where figure 4.14b shows the subsequent step. This process is repeated for all columns before restarting at column 1. If the refresh rate is sufficient, the observed image creates the impression of being static.

Another important aspect behind the driver circuitry is to extend pixel functionality beyond the described on and off states. Gray scale images and intensities creates the opportunity to augment the display's capability beyond a simple "dot matrix"-type representation of text characters and simple graphics. Two ways of introducing this functionality are by

- varying the supply current proportional to the desired intensity (see figure 4.10), or
- varying the duty cycle on a fixed amplitude driving current.

Exact control over the supply current requires precision analogue circuitry for determining the required current. DAC implementations become cumbersome and geometrically expensive, increasing the pixel complexity and requiring more careful design for desensitising circuit operation from typical process variations. The latter technique is much simpler and more robust to implement as the dependence on intricate analogue circuitry is lessened, the pixel size remains small and allows for direct interaction with the digital domain. This has the added advantage of allowing more flexibility and scalability in the control of the microdisplay.

4.5.4 Controller design

4.5.4.1 Internal controller

bits						column
6	5	4	3	2	1	
0	0	0	0	0	0	1
0	0	0	0	0	1	2
0	0	0	0	1	0	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	1	1	0	1	62
1	1	1	1	1	0	63
1	1	1	1	1	1	64

Table 4.6: Truth table for column addressing in the scanning architecture of the microdisplay

The 6 to 64 bit is mentioned in section 4.5.3 as the means to address individual columns within the active optical matrix of the microdisplay. The design of this subsystem is straightforward, where logic was synthesised using the truth table shown in table 4.6. VHDL code was implemented in a primitive functional form to serve as input to an RTL synthesis engine, which incorporated the foundry's standard cell library, along with fan out constraints, and produced the desired state machine.

Inputs to the internal controller is applied through I/O pads and circuits with the necessary buffering and ESD protection circuitry standard in CMOS designs. The reason for the design approach was to prove that a standard methodology can be used to design digital CMOS circuits without alteration or inclusion of any non-standard procedures. The resulting subsystem can then directly be used, along with the necessary fan out considerations, to drive the optical matrix consisting of avalanche breakdown-based devices.

4.5.4.2 External controller

The architecture described in section 4.5.3 is straightforward in order to minimise the risks associated with prototyping. Debugging is facilitated by implementing the rest of the controlling circuitry as an external subsystem. The external controller is responsible for creating higher level instructions to

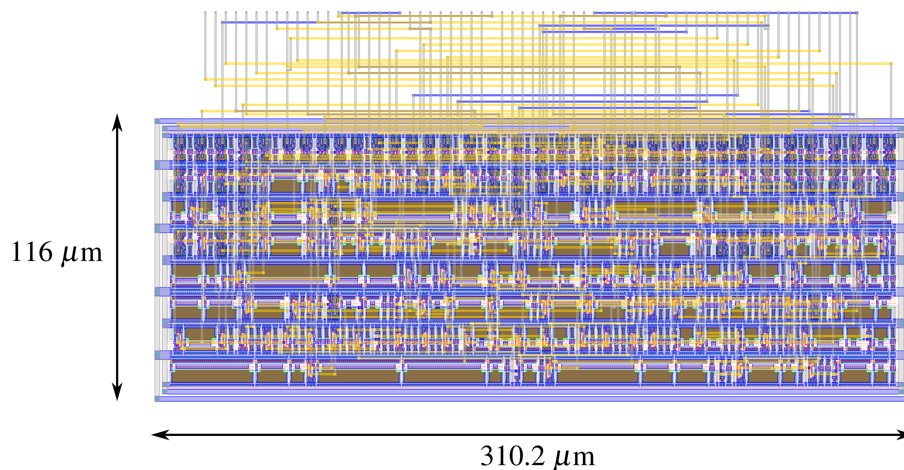


Figure 4.15: Layout of the scanning decoder for microdisplay circuit controlling the column addressing with dimensions indicated

produce text and images, while containing the character set library storage as well. The system is microcontroller based, where the microcontroller synchronises the row energising and column select based on the rendering required on the display. Pass transistors are used to control the high voltage lines as an interface between the external controller logic and the row signals.

4.5.5 Spectral characteristics

Conventional forward biased *pn*-junctions in silicon emit photons mostly centred around the band gap of silicon, which is not perceivable by the human eye. The spectrum of hot carrier emission is much more energetic and the spectral power distribution extends from below 400 nm to beyond 1000 nm with a very broad emission characteristic. This is advantageous for the proof of concept integration in the microdisplay prototype application as some of the spectral content are clearly visible to the naked eye. The spectrum of light is shown in figure 4.19 and is further investigated in section 4.5.7 for microdisplay applications.

4.5.6 Radiation pattern

In section 4.3.5 it was shown that it is possible to utilise the back end stack for creating light directing structures and thereby improving extraction efficiency of light in leaving the surface. The layout in figure 4.12 shows how the metal and via layers are used both as conductors and light directors.

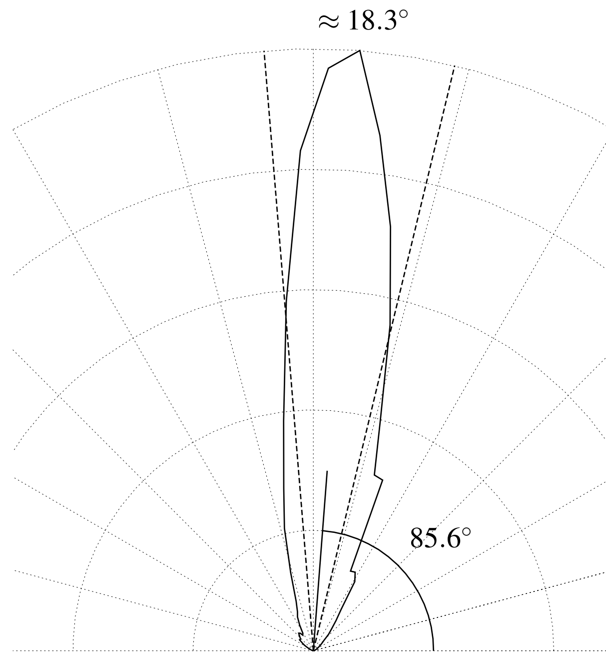


Figure 4.16: Normalised radiation pattern of a pixel including light directing structures

The effect of these light directors is shown in figures 4.16 and 4.8. Figure 4.16 shows a two dimensional cut of the radiation pattern along the axis of tilt in respect to the normal of the chip surface. The tilt of 4.4° is less than without the light directing structures. The apex angle of the half power points is slightly increased to 18.3° through the use of light directing structures when compared to the 13° apex angle subtended by a point source without the light directing structures.The amount of optical power leaving the surface is also increased.

By using the exit angle it is possible to capture almost all light leaving the surface with the appropriate lens design which makes the directionality advantageous in microdisplay applications. The improved tilt also allows better utilisation of light.

4.5.7 Luminance

Microdisplay intensities are characterised by luminance, a measure of luminous flux per solid angle for a given emitting area. This is a measure of the perceived brightness of the active matrix and is given by equation 4.9, with luminance L_v in cd/m^2, Φ_v the luminous flux, Ω the solid angle in sr, A is the area in m^2 and θ is the angle formed between the normal of the emitting surface and the direction

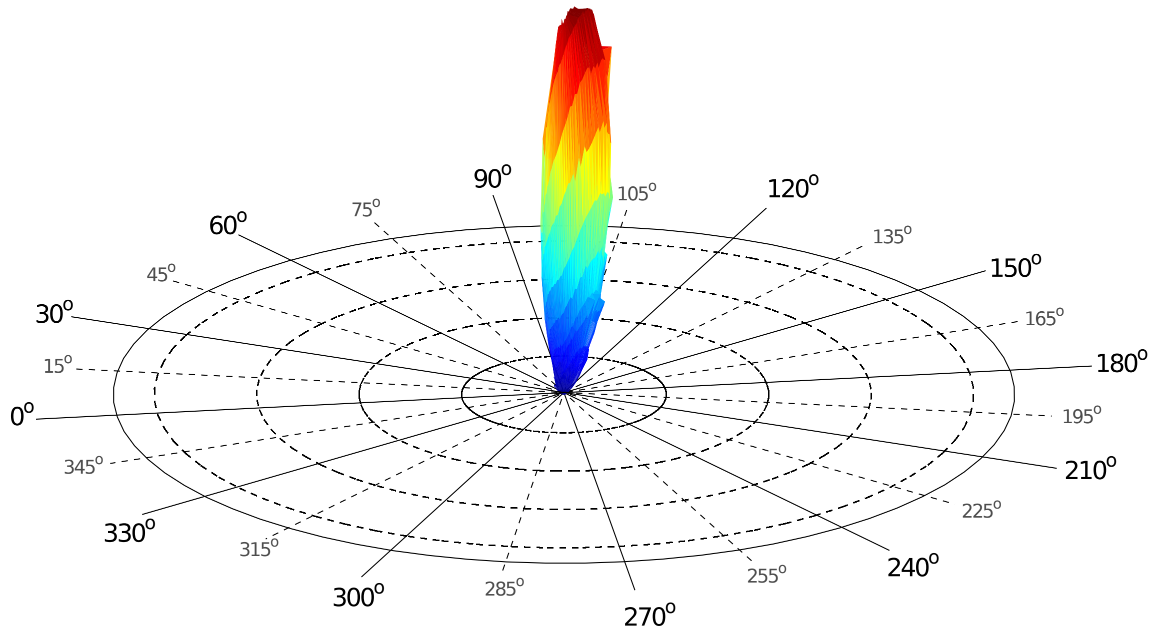


Figure 4.17: Three dimensional radiation pattern of a pixel using light directing structures showing the improved directivity

of interest.

$$L_v = \frac{d^2\Phi_v}{dA d\Omega \cos\theta} \tag{4.9}$$

The most important characteristic of the light source is the spectral power distribution, $J(\lambda)$. The measurement and calculation are documented in appendix A. Not all spectral components are useful for display purposes. Figure 4.18 shows the CIE2008 data on the photopic response of the eye, with the maximum centred around 555 nm. This wavelength is were a normal human eye peaks in responsivity. For microdisplay applications, this becomes important when calculating the luminance of the display.

The spectral content of the source is converted to luminous flux through equation 4.10 where Φ_v is the integrated luminous flux leaving the emitter, $\bar{y}(\lambda)$ is the luminosity function as shown in figure 4.18 for the photopic eye response and $\phi_v(\lambda)$ is the spectral distribution of luminous flux.

$$\begin{aligned} \Phi_v &= \int_0^\infty \phi_v(\lambda) d\lambda \\ &= 683 \cdot \int_0^\infty \bar{y}(\lambda) J(\lambda) d\lambda \end{aligned} \tag{4.10}$$

The solid angle can be obtained by calculating the average solid angle as shown in figure 4.16 as

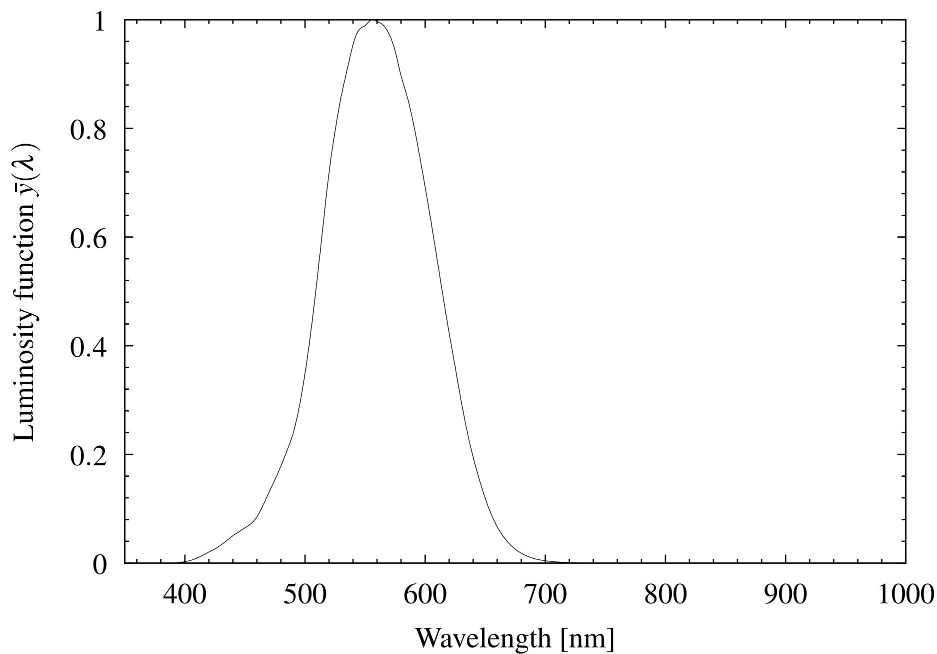


Figure 4.18: Normalised optical spectral power distribution of the pixel emission showing the effect of the photopic luminosity function

the radiation pattern defines in which direction the luminous flux moves. It can be assumed that in microdisplay applications the opportunity exists to harvest the complete flux bundle for use. Using equation 4.3 the solid angle can be calculated using the apex angle from figure 4.16.

Luminous intensity, measured in candela or lm/sr, is determined by the luminous flux passed through a solid angle, each of which is described above. Finally, the display area needs to be taken into account when computing luminance. Section 4.5.3 describes the scanning architecture which, along with the physical size of the display, determines the area over which the optical power is emitted.

The quantities relevant to the calculation of the display luminance are shown in table 4.7.

4.5.8 Electrical characteristics

The hot carrier light sources break down in the region of 9 V for the specific CMOS process. Some overhead voltage were allowed for the driving *n*-MOSFETs to operate in the saturation region. For practical demonstration purposes, a 3-cell Lithium Polymer (LiPo) battery, with a nominal voltage of 11.1 V and full charge voltage of 12.6 V were mostly used to drive the source, except for laboratory

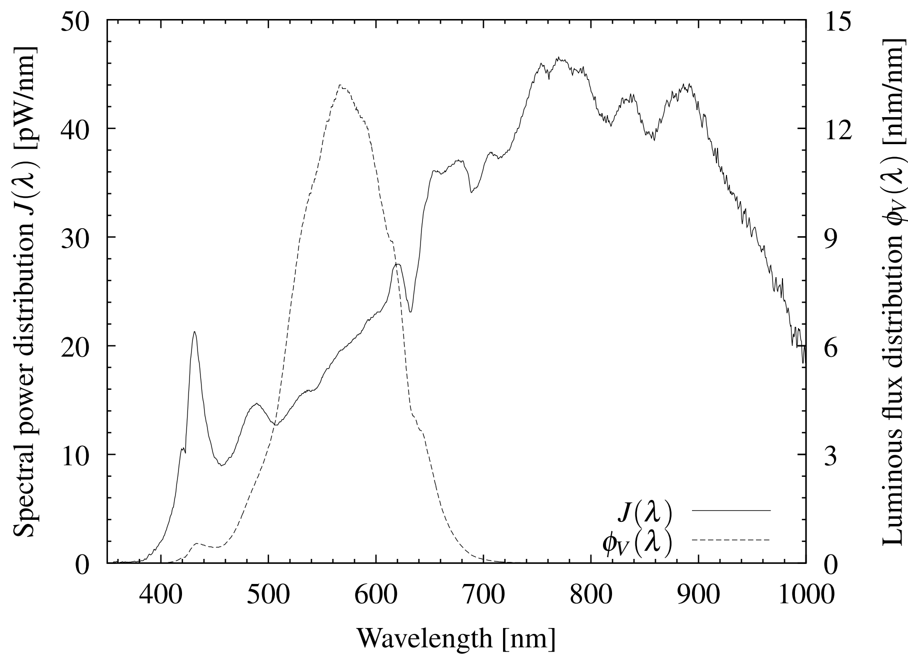


Figure 4.19: Spectral power distribution of the microdisplay output and spectral luminous flux distribution

Table 4.7: Microdisplay optical characteristics for flash light mode monochrome operation at the maximum designed electrical input power

Quantity	Symbol	Units	Value
Operating voltage	V_{sup}	V	12.69
Radiant flux	Φ_e	nW	3.95
Apex angle	2θ	°	18.3
Solid angle	Ω	sr	0.0773
Luminous intensity	I_v	μcd	3.71
Emitting area	A	mm^2	1.28
Luminance	L_v	**cd/m^2**	**2.901**

characterisation. For this reason, the design voltage was set at 12 V $\pm$ 0.6 V. The refresh rate of the microdisplay was limited by the external microprocessor to 42.6 Hz. Table 4.8 shows the objectives aimed for during design and table 4.9 shows the measured operating characteristics of the microdisplay at the limits of the intended design.

Table 4.8: Summary of designed values for the 64×8 pixel display and driver circuitry

p^+n-breakdown voltage	9.2 V
Display supply voltage	12 V ± 0.6 V
Current per point source	170 μA
Current per pixel	5.1 mA
Full column maximum current	41 mA
Refresh time	$<$ 25 ms

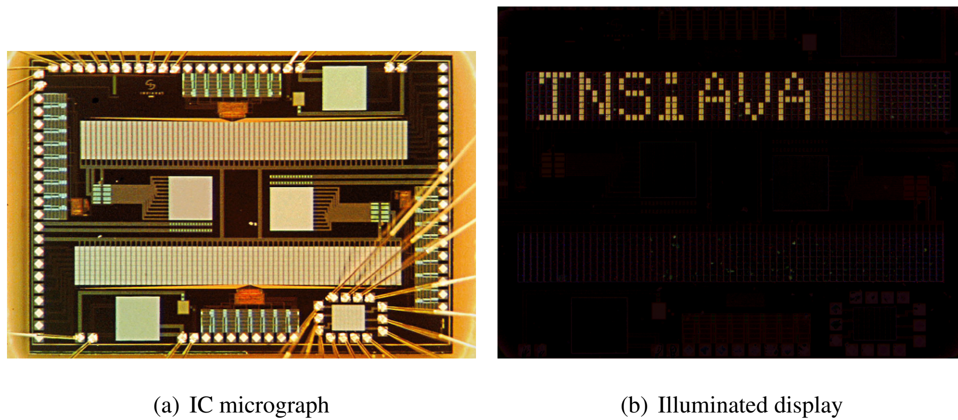
Table 4.9: Summary of measured electrical characteristics at different operating voltages for the 64×8 pixel display and driver circuitry

Operating voltage	12.6 V	11.4 V
Current per point source	339 μA	192 μA
Current per pixel	10.18 mA	5.75 mA
Full column maximum current	81.42 mA	46.01
Overall power consumption	1.85 W	1.20 W
On-chip power consumption	1.03 W	0.52 W
Detected optical power emitted	3.89 nW	2.62 nW
Luminance	2.83 cd/m^2	1.93 cd/m^2

4.6 CONCLUSION

The ability to create light using silicon would be of little value without the capability to integrate the light source with other circuit components. Furthermore, the possibility of using CMOS contributes greatly to the value proposition. This chapter introduced design approaches for creating light sources in a standard CMOS process. Various structures and their impact on interaction with logic circuitry were also shown. Design challenges associated with the high voltage operation were identified. A completely integrated design is presented as a CMOS microdisplay with integrated logic control, proving the feasibility of integrating light sources in CMOS with interaction with other circuitry on a single die. This work was published as a conference proceeding [60] and is further summarised in a peer reviewed journal article [61].

A major drawback is the high operating voltage of the junctions, which may reduce the operating mar-



(a) IC micrograph (b) Illuminated display

Figure 4.20: The 64×8 pixel microdisplay in a 0.35 μm standard CMOS process

gins for reliable circuits. The next chapter introduces novel devices aimed at reducing the operating voltage while improving the external power efficiency.

CHAPTER 5

IMPROVED LIGHT SOURCES IN STANDARD CMOS

Efficiency and reliability will almost certainly remain recurrent objectives in engineering. This chapter documents an approach to address both these qualities in the silicon light sources, introduced in chapter 4, in CMOS with the eventual purpose of improving CMOS integration, to be further elaborated on in chapter 6.

While it remains true that radiation originating from hot carrier action is not nearly as efficient as the mechanisms employed in conventional LEDs, efficiency becomes so much more important when every photon counts. For this reason and the potential each incremental improvement in pure silicon light sources may unlock, the effort spent may be well worthwhile and interesting at the same time.

This chapter describes the work done on

- reducing the operating voltage of CMOS light sources, thereby
  - improving the external power efficiency (EPE) of the devices and
  - easing the integration of light sources along with other electronic devices in CMOS,
- improving the predictability of light source characteristics, and
- improving light extraction efficiency which also impacts the EPE.

5.1 APPROACH

5.1.1 Breakdown voltage

Silicon junctions in avalanche emit light. This requires a reverse biased voltage of sufficient magnitude to initiate avalanche multiplication in order to establish both high electric fields within the junction as well as carrier transport between the terminals. Restating some of the theory discussed in chapter 3, the breakdown voltage V_{BD} of a junction is a function of depletion region width or doping and the critical field E_m. For an abrupt junction [71], equation 5.1 gives an expression for V_{BD} as a function of depletion region width or background doping concentration, with E_m the critical field strength, W_{Dm} the maximum depletion layer width, ε_{Si} the silicon permittivity and N the background doping concentration.

$$V_{BD} = \frac{E_m W_{Dm}}{2} = \frac{\varepsilon_{Si} E_m^2}{2qN} \tag{5.1}$$

One way to reduce V_{BD} through equation 5.1 is to increase the background doping concentration which will necessarily decrease W_{Dm}. Unfortunately, the circuit designer has no control over the doping concentrations in CMOS junctions except for the definitions of the structures itself.

A second approach is based on the fact that the ionisation coefficients in silicon are a strong function of electric field strength. The value of E_m is given in equation 5.2 [71].

$$E_m = \frac{4 \times 10^5}{1 - (1/3)\log_{10}(N/10^{16}\,\text{cm}^{-3})} \tag{5.2}$$

Figure 5.1 shows a plot based on equations 5.1 and 5.2 for the background doping concentration N. The region of interest in this work is where $E_m \approx 600$ kV/cm with a background doping of $N_A \approx 1.5 \times 10^{17}\,\text{cm}^{-3}$. It can be seen that the critical field E_m does not vary a lot as a function of the doping concentration and for all practical purposes can be assumed to remain constant. These are of course first order approximations and a detail analysis on the breakdown characteristics and the electric fields involved will require numerical computation while taking carrier temperatures into account and evaluating the complete distribution function in phase space as a function of position.

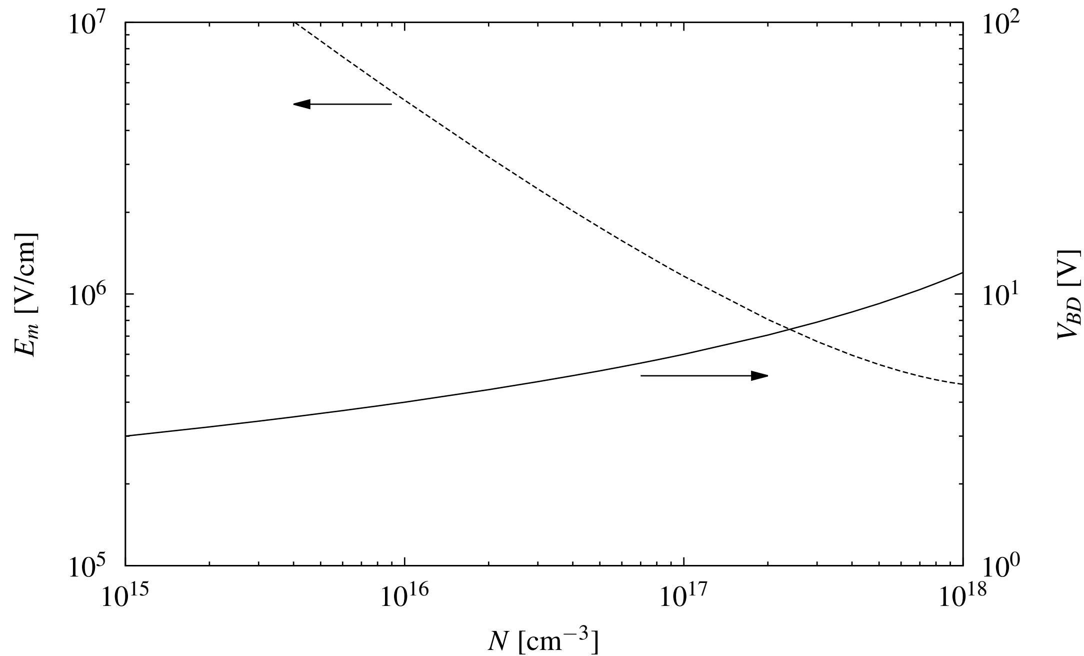


Figure 5.1: Critical electric field and breakdown voltage as a function of background doping for an abrupt junction

5.1.2 Proposed approach

As we do not have control over the doping concentrations, the only parameter we can try to alter is W_D by terminating W_{Dm} in equation 5.1 prematurely by using a highly doped region of the same species as the background, which will effectively create a field stop. Note that Sze [71] uses the term punch through when the electric field extends between two highly doped regions through a lighter doped region. We prefer to use the term "reach through" in order to prevent confusion with the condition in MOS transistors where the depletion regions punch through to each other and results in drain induced barrier lowering (DIBL). As most of the avalanche gain occurs where $E(\mathbf{r}) = E_m$ with $\mathbf{r}$ representing the position vector, the resulting integral of the electric field, yielding potential, will be reduced for the same amount of current flow. This should, in theory, result in avalanching devices operating under a reduced voltage condition *for the same background doping concentration*. This is very important as it allows the designer to control V_{BD} without requiring a change in the relevant CMOS process.

$$\frac{V'_{BD}}{V_{BD}} = \left(\frac{W_D}{W_{Dm}}\right)\left(2 - \frac{W_D}{W_{Dm}}\right) \tag{5.3}$$

The simple expression in equation 5.3 shows the reduced voltage V'_{BD} as a function of the shortened depletion region width W_D [71]. For the curious, an estimate based on the junctions we utilise in conjunction with equations 3.32 and 3.34, given an estimated $N_A \approx 1.4\times10^{17}$, produces a depletion region width of $W_{Dm} \approx 300$ nm.

5.1.3 Conceptual CMOS implementation

In CMOS, the highly doped regions used for transistor source and drain regions can be used for creating highly doped n^+- and p^+-regions. Against the n- and p-well doping, with influence from the channel and field stop implants, a region of lower background doping can be created in between as envisaged in figure 5.2. By using the available definitions it should be possible to construct an $n^+\pi p^+$-diode as depicted in figure 5.2, with the π-region representing a region of lower p-type doping than than compared to the n^+- and p^+-implants. The same could be done for a diode with an n-type background doping to create a $p^+\nu n^+$-diode with the ν-region representing a lower doped n-region compared to the implants. The rest of the chapter will be devoted to exploring the possibilities in CMOS by using this approach. Figure 5.2 shows the layout, electric field and potential profile for such a device with the reduced breakdown voltage indicated.

5.2 PROCESSING CONSIDERATIONS

5.2.1 Scaling theory for CMOS and device density

In the famous paper of Dennard [88] the idea of MOSFET transistor scaling was in theory a simple matter of scaling key transistor parameters by a constant in order to yield smaller devices and the ability to predict their performance improvement in an easy and linear fashion. The initial approach was called constant-field scaling, where by reducing all device dimensions, the threshold and supply voltage while increasing all doping levels by the same factor will yield a smaller device with the performance scaled proportionally.

Supply and threshold voltages, however, did not scale as fast as the reduction in dimensions [89] leading to a constant-voltage approach. The effects, however, were that devices operated under increased field levels and with device dimensions reaching sub-μm dimensions, the effect of short channel effects became more pronounced. With lateral field strengths increasing it was partly necessary to improve lateral device isolation through the use of techniques such as LOCOS and STI. Modern

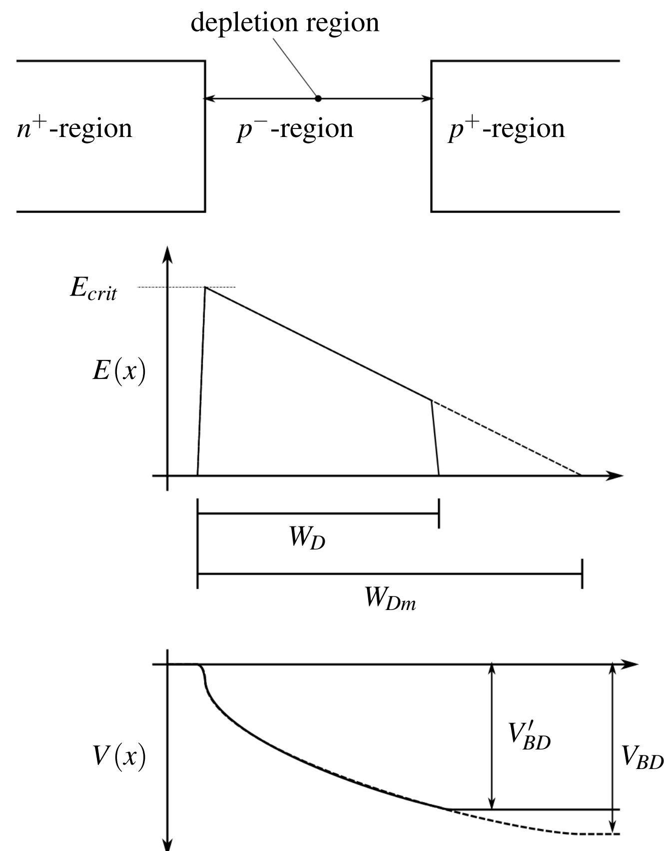


Figure 5.2: Reduced breakdown voltage principle with junctions in CMOS under an abrupt junction approximation

scaling theory is a complex one with a lot of challenges associated with the task.

5.2.2 Oxidation of silicon

One of the most attractive features of silicon for tight integration is the inherent oxidation characteristics of the material and the high quality dielectric and insulating properties thereof resulting from the oxidation process. Silicon dioxide is used for its properties as a gate dielectric, as well as its masking process and passivation properties in CMOS processes. There are usually two ways for creating thin films of SiO_2 in CMOS processes, namely

- thermal growth of SiO_2 by exposing silicon to oxygen at elevated temperatures, and
- deposition of SiO_2 through chemical vapour deposition (CVD).

Thermal oxidation can further be separated into two approaches, namely

- dry thermal oxidation, in which silicon is heated in an environment of pure oxygen, and
- wet thermal oxidation, where the silicon is heated in an environment of water vapour.

Dry oxidation is slow, but creates oxides with extremely attractive electrical characteristics, while wet oxidation results in thermal oxides at a much improved rate. SiO_2 consumes silicon at a thickness rate of $0.44t_{Si}$ per t_{SiO_2} grown. The oxide thickness t_{SiO_2} can be calculated using equation 5.4 [81], with D_{O_2} the diffusion coefficient of the oxidising species, κ is the surface reaction rate, C_0 is the surface concentration of the oxidising species from the environment, C_1 is the number of the oxidising species present in the SiO_2 layer and d_0 is the initial oxide layer thickness, which reduces τ to zero when starting with a new oxidation step.

$$\begin{aligned} t_{SiO_2} &= \frac{D_{O_2}}{\kappa}\left[\sqrt{1+\frac{2C_0\kappa^2(t+\tau)}{D_{O_2}C_1}}-1\right] \\ \text{where} \quad \tau &\equiv \frac{(d_0^2+2D_{O_2}d_0/\kappa)C_1}{2D_{O_2}C_0} \end{aligned} \tag{5.4}$$

5.2.3 LOCOS and field oxide

SiO_2 is often used as a mask in ion implantation steps [81]. The target process for this work is LOCOS-based and uses field oxide for both lateral device isolation as well as for its masking properties. Figure 5.3a shows the design for a pointed light source as described in chapter 4 section 4.3. Combining figures 4.3 and 5.3a results in figure 5.3c which represents a cross-sectional along the direction of intended current of the fabricated device, where the field oxide created by the LOCOS growth step acts as a mask to localise the shallow highly doped implants. In order to form the regions exposed for LOCOS growth, a blocking layer, usually Si_3N_4 as shown in figure 5.3b, acts as the oxidation inhibiting layer and is used to cover the active regions to inhibit oxidation as depicted in figure 5.3c. After field oxide regions have been grown, highly doped ion implantations take place in the regions defined by **NPLUS** and **PPLUS** for the n^+- and p^+-regions respectively. As field oxide growth is prevented where **ACTIVE** is defined, these regions are covered only by a thin oxide layer and does not present great stopping power for implanted ions. In terms of point sources, this means that the region between the n^+-region and the p^+-region (see figure 5.6) remains doped according to the background, with influence from the channel stop implant and dopant redistribution and segregation for subsequent thermal cycles. The field oxide however prevents the highly doped ion implants to reach the silicon underneath.

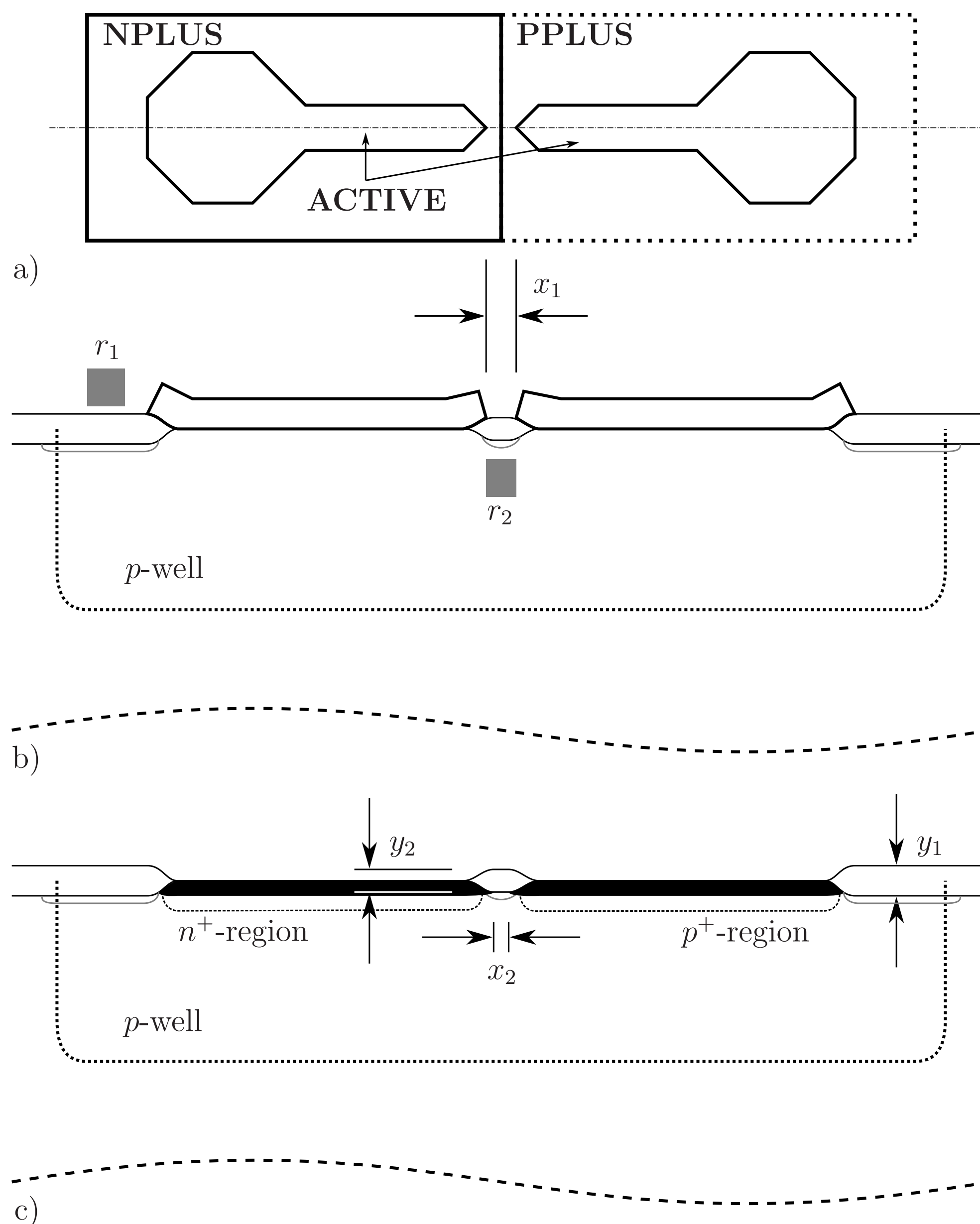


Figure 5.3: Effect of separation distance on field oxide growth with a) showing the design masks, b) the Si_3N_4 field oxide inhibitors and c) the resultant doped regions

Given the isotropic nature of thermal oxidation as well as the presence of a pad oxide layer for stress relief between the silicon and Si_3N_4, a major problem with LOCOS is the lateral oxidation extending into the region underneath the oxidation inhibiting layer. This results in the so called "bird's beak" phenomenon and as a result, the boundaries of the field oxide are almost always ill-defined. This has a negative impact on the packing density of devices such as adjacent transistors and also for light sources, as will be discussed later.

There are a number of variations on the LOCOS techniques as developed through the past few decades in order to accommodate higher device density, some of which are discussed below and most of which are aimed at reducing the “bird’s beak” structure. The work of Peter Smeys [90] summarises the development of LOCOS techniques well, although foundries seldom publish the exact methods and details thereof as used in successful commercial production lines.

5.2.3.1 Semi-recessed LOCOS

LOCOS was introduced as an isolation technique for separating in the early 1970’s when it was realised that Si_3N_4 is hardened to oxidation [91]. Separating the active regions is important for a number of reasons, including to

- reducing leakage currents between junctions,
- reducing the sidewall capacitances associated with highly doped regions,
- inhibiting electric field punch through between doped regions, and
- reduces latch-up when used with retrograde wells and/or other doping profile techniques.

As briefly shown in section 4.3.2, which represents semi-recessed LOCOS growth, Si_3N_4 is used as a block for oxygen diffusion to silicon. This allows growth in selective regions only. Stresses induced by the deposition of Si_3N_4 are relaxed through the growth of a pad oxide, usually in the range of 15 nm thick. This pad oxide, however, creates a diffusion path for oxygen to migrate laterally underneath the mask, resulting in the well-known bird’s beak structure. This severely limits the packing density of active regions and forces large design rules for separating regions.

5.2.3.2 Fully-recessed LOCOS

Fully recessed LOCOS structures can be obtained by first etching into the exposed regions of silicon before starting the major oxidation step [92]. This allows much deeper LOCOS structures and results in better electrical isolation, but the oxidation path through the pad oxide layer is now closer to the surface concentration for much longer, resulting in an increased length of the bird’s beak and worsening the packing density problem.

5.2.3.3 Polysilicon Buffered LOCOS (PBL)

The bird's beak diffusion can be reduced by decreasing the pad oxide layer, albeit at the expense of increased stress. A layer of polysilicon can be deposited on the pad oxide to server as the stress relief layer [93]. This technique allows a decrease in pad oxide layer without damaging the silicon crystal structure. It is possible to reach sub-micron spacing, with reports of limiting the bird's beak extension to 0.1 μm [94].

5.2.3.4 Spacer-based designs

To achieve even better LOCOS characteristics, a number of spacer-based solutions have also seen light in the past, with some successfully employed at very small active region spacings. Sidewall mask isolation (SWAMI) [95] uses a recessed silicon along with multiple nitride depositions and oxide spacers are used to limit lateral oxidation. The process is prohibitively complex, but the idea of employing spacers has resulted in a number of other state of the art techniques used for sub-micron processes.

Polysilicon Spacer LOCOS (PSL) [92] reduces lateral oxidation by plugging the oxidation path with polysilicon spacers. Results have shown good reduction in spacing dimensions with reported values of 0.3 μm separation distances and a 0.6 μm DRAM pitch [96].

The nitride spacer approach is also used, with nitride-clad LOCOS (NCL) [97] showing a lateral encroachment of only 60 nm, making the technique suitable for 0.25 μm CMOS. Although LOCOS is still used at these small feature sizes, deep nanometre designs usually rely on shallow trench isolation with an etch-based approach for better isolation at small feature sizes.

5.2.4 Shallow trench isolation (STI)

Another technique for reducing junction leakage and provide lateral isolation between active devices is shallow trench isolation (STI). A technique similar to LOCOS is used with a Si_3N_4 deposition layer followed by a photoresist mask to define the trench location. A directional etch is done into the silicon resulting in near vertical sidewalls, after which a liner oxide is grown [98]. The trench is then filled with an oxide after which the resulting structure is planarised and the Si_3N_4 removed. This technique provides for much denser device pitches and is heavily employed in the sub-0.5 μm

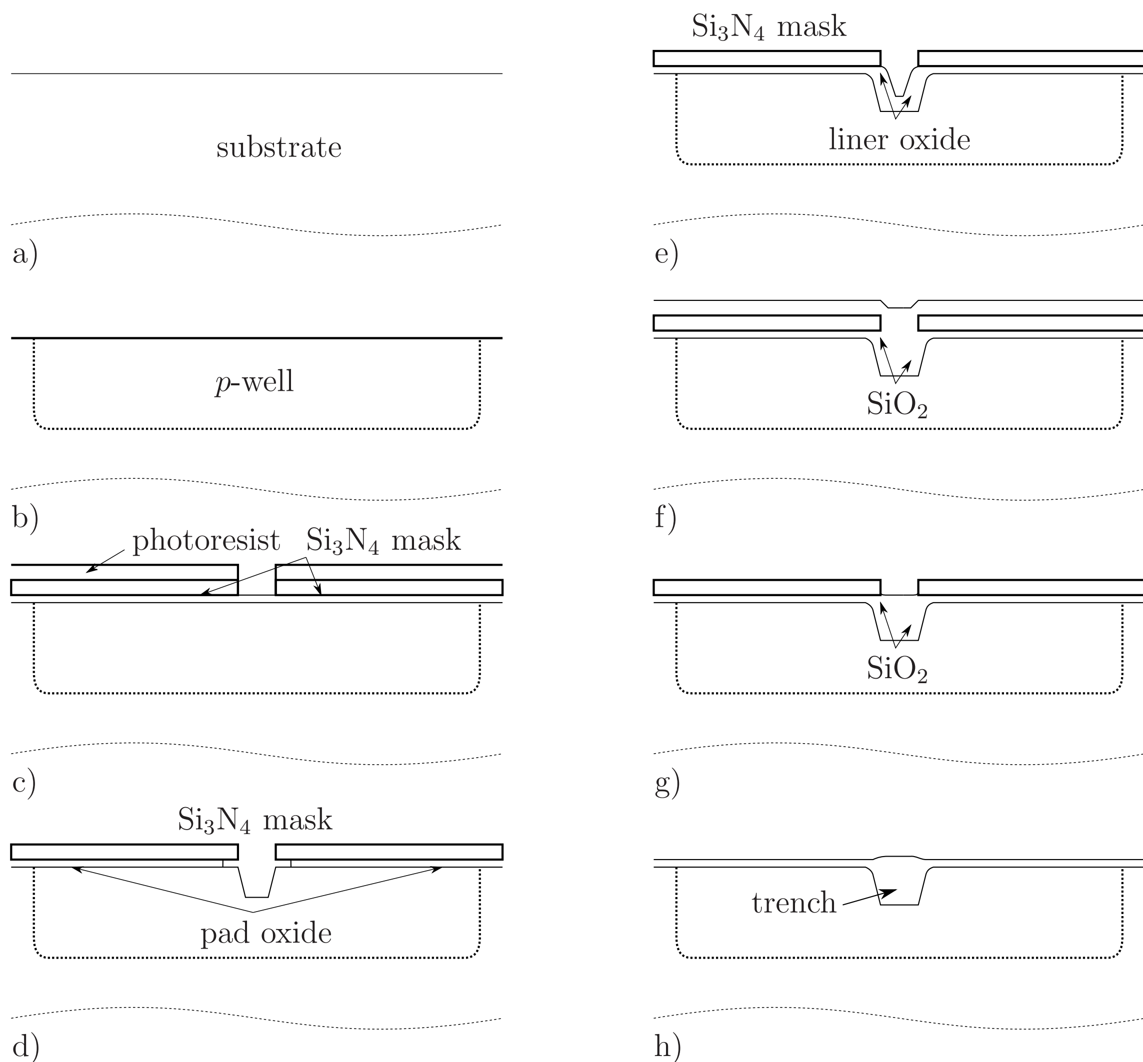


Figure 5.4: An example of a shallow trench isolation process

CMOS technologies. The main feature is the well-defined vertical side wall profile of the isolating structure and the approach alleviates the problems associated with the "bird's beak" formation in LOCOS approaches.

Figure 5.4 shows an example process [98], with figure 5.4d showing the trench etch and pad oxide undercut, 5.4e the growth of the liner oxide, 5.4f-g the planarisation and 5.4h the final result.

5.2.5 Processing limits

5.2.5.1 Lithography

The limit of most CMOS processes lies in the ability to resolve geometrical details during the lithographic stages of processing. At some point a reduction in feature size will result in a wrongful representation of the feature when lithographic exposure happens, which will result in erroneous functioning of the feature. When examining figure 5.3 the value of x_1 is intended to represent a separation of a certain size between the n^+- and p^+-regions. Because of the non-ideal bird's beak discussed *supra*, the actual separation will not be abrupt nor the size of x_1. Furthermore, as x_1 is reduced in design, the measure will reach a point where the lithography will fail.

Equation 5.4 is valid for region r_1 as shown in figure 5.3b, where the direction of oxidation experiences diffusion gradients only in the direction of layer growth. Close to the Si_3N_4 boundaries, where masking of the oxidation growth starts, the situation becomes more complex. Although the mask blocks off the surface exposure to oxygen, a lateral gradient of oxidation species concentration exists and oxygen atoms will migrate sideways through the thin pad oxide used for stress relief in LOCOS processes [90], with some oxidation occurring under the mask. This affects the effective C_0 term in equation 5.4 and will result in slightly different growth characteristics, resulting in the characteristic "bird's beak".

When the gap created by the nitride masks for separating the two highly doped regions becomes comparable to the oxide thickness grown, the effect mentioned above becomes even more pronounced. The same surface concentration now experiences lateral gradients in both directions at a constant surface concentration. The nitride mask also introduces changes to the gas flow across the surface and may cause a further drop in surface concentration. Oxidation taking place in region r_2 will therefore experience a lower growth rate in terms of thickness than in region r_2. This means that $y_2 < y_1$, where y_1 is the typical field oxide thickness and y_2 is the field oxide thickness between the two highly doped regions. More pronounced, however, is the effect of a change in x_2, the separation distance between the two highly doped regions, which is further discussed in the following sections.

5.2.6 Ion implantation

Ion implantation is generally preferred over diffusion doping of impurities as the workhorse technique in the CMOS industry today. Some advantages of ion implantation include

- precise dosage control,
- good doping uniformity,
- doses that are not limited to the solid solubility of the host material,
- smaller lateral doping profile,
- using rapid thermal annealing where very shallow junctions can be obtained,
- the low temperature nature of the process step and
- the ability to allow multiple implants in order to form complex doping profiles suited to specific applications.

Disadvantages include

- the large capital investment for expensive equipment,
- damage to the crystal lattice which requires additional annealing steps to repair and
- the difficulty to dope very thin layers of material, such as the device layers in SOI technology.

5.2.6.1 Range and straggle

Two parameters are important when considering ion implantation for doped regions in a semiconductor, with

- R_p representing the projected range of ions along the direction of implantation, and
- σ_p, the projected straggle of implanted ions around the projected range.

These parameters are used to quantify the approximate Gaussian distribution which an implanted impurity profile assumes [81] and is related to the profile by

$$n(x) = \frac{S}{\sqrt{2\pi}\sigma_{\mathrm{p}}} \exp\left[-\frac{(x-R_{\mathrm{p}})^2}{2\sigma_{\mathrm{p}}^2}\right] \tag{5.5}$$

with S the dose of implanted ions per unit area, usually cm^{-2}. The range and straggle are functions of both the impurity and target and are determined by the way in which the kinetic energy of the accelerated ions are lost as the ions enter and move through the material. The two important energy loss mechanisms are nuclear and electronic. Nuclear energy loss is brought about through energy transfer to the material nuclei, a property which can result in the displacement of a nucleus from its original site within the lattice. Electronic energy loss results from interaction of the implanted ion with the electron cloud associated with the material nuclei, with Coulombic interactions responsible for energy transfer. The latter does not result in damage to the material crystal lattice.

5.2.6.2 Masking properties of SiO_2

To selectively implant ions in specific regions, masks are employed to absorb the ions before they reach the target material. Such masks can include photoresist, SiO_2 and Si_3N_4, each with different masking properties. While Si_3N_4 has the best masking properties for ion implantation of the mentioned thin films readily available in CMOS processing, the field oxide is usually used as a mask for the shallow highly doped regions of a MOS device [81]. The stopping power is determined by the thickness of the mask required to block most of the implanted ions before reaching the target material. Figure 5.5 shows the stopping power of SiO_2 when used as a mask, with E the accelerated ion energy.

5.2.6.3 Channelling and tilt

The range and straggle properties of ion implantation are strong functions of the crystal lattice orientation. Since most device grade wafers are of [100] or [111] surface orientation, implanted ions are channelled through the material. To alleviate this problem, the wafer is tilted by 7° [81] in order to present the implant direction with a [763] plane. An amorphous silicon approximation can then be used to estimate the range with improved agreement between the doping profile and that of a Gaussian distribution.

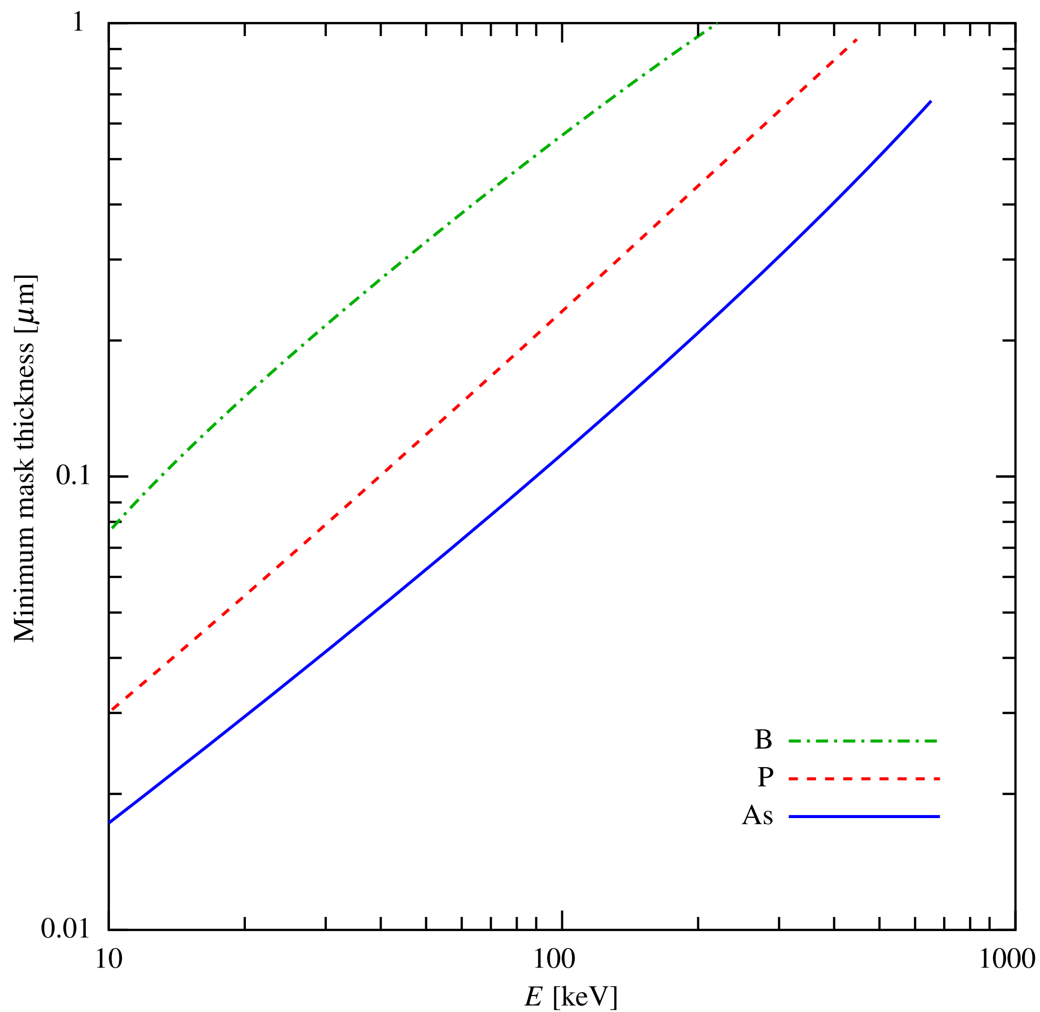


Figure 5.5: Minimum mask thickness of SiO_2 for blocking 99.99 % of accelerated impurities

5.2.6.4 Effect on reduced SiO_2 region separating highly doped regions

In section 5.2.3 in conjunction with figure 5.3 it is shown that the field oxide in the separation region does not experience the same growth environment as typical field oxide where the nitride mask spacing is larger. The reduced thickness impacts the stopping power of SiO_2 as an implantation mask as shown in figure 5.5.

Table 5.1 shows an example CMOS process and the parameters relevant to the field oxide and highly doped region implants. For Boron, combining table 5.1 and figure 5.5 shows that a field oxide thickness of around 200 nm would be a sufficient mask, while for Phosphorous, a slightly heavier ion, a mask thickness of 100 nm seems sufficient.

Table 5.1: Step 23, 41 and 43 from an example CMOS process showing the LOCOS growth and diffusion implant details

Target field oxide thickness	550	nm
For the p^+-implant		
Species	B	-
Energy dose	3×10^{15}	cm^{-2}
Acceleration energy	20	keV
For the n^+-implant		
Species	P	-
Energy dose	3×10^{15}	cm^{-2}
Acceleration energy	40	keV

The typical value of field oxide in the target process (section 4.2.2) is $t_{fox} = 290 \pm 30$nm, while details of the species and implantation dose remain confidential foundry information. It becomes apparent, though, that with a reduction in thickness as expected from the oxide formation between the two closely spaced oxidation masks in figure 5.3 in r_1, ions will start to leak through the mask into the substrate, forming part of the active device and affecting the characteristics of the junctions in the specific region. It is therefore reasonable to expect that the following relationship holds:

$$\begin{aligned} \frac{\partial x_2}{\partial x_1} &= 1 \qquad \text{for large values of } x_1 \\ \frac{\partial x_2}{\partial x_1} &> 1 \qquad \text{when } x_1 \text{ approaches } t_{fox} \end{aligned}$$

The direct effect is that the separation distance between the highly doped regions will experience a sudden reduction when the drawn separation distance reach the region of $x_1 \approx 300$ nm. Furthermore, the variation of field oxide thickness may provide difficulty in the reproduction of junctions with similar characteristics between batches. The effect discussed makes it difficult to design pointed light sources for well controlled electric field reach through conditions.

5.3 REFERENCE DEVICE AND EXPERIMENTAL APPROACH

In order to quantify the performance of devices designed and tested in this section, it is necessary to have a benchmark.

5.3.1 Reference device for individual point sources

The reference used to compare subsequent results will differ for individual point sources and arrayed sources. In all cases where individual point sources are benchmarked, the reference will be the device with the largest separation distance. This approach was chosen because the photon multiplier tube can at best serve as a tool for relative comparison.

5.4 DESIGN FOR ELECTRIC FIELD REACH THROUGH IN POINT SOURCES

A *pn*-junction in avalanche experiences a strong carrier generation term only where the electric field reaches its critical magnitude in silicon. It is theoretically possible to reduce the voltage of silicon light sources operating on the principle of avalanche breakdown by simply terminating the electric field which does not contribute to a high generation term by a highly doped region, thereby allowing the electric field to "reach through". This approach is discussed below.

5.4.1 Geometrical adaptations of point sources

In chapter 3 the effect of a reduced potential integral was discussed in terms of lower voltage for the same current traversing through the high electric field region. The first obvious approach would be to take the structure definition in figure 4.1 and reduce the spacing labelled "min. spacing" in the figure. At some point, the depletion region created by the n^+p-junction will reach through to the p^+-region and start to reduce the potential between the highly doped region, without an immediate drop in optical emission. This is usually against foundry design rules for separating transistor diffusions sufficiently apart. The incremental spacing is restricted by the foundry's minimum design grid and will be influenced by the fracturing of the layout for masking, as well as the resolution of lithography used during photolithography steps.

5.4.2 Localisation due to reduced barrier potential

Based on the assumption of an abrupt junction which for all practical purposes only extends into the lightly doped region, the electric field contributes very little to the potential integral in highly doped

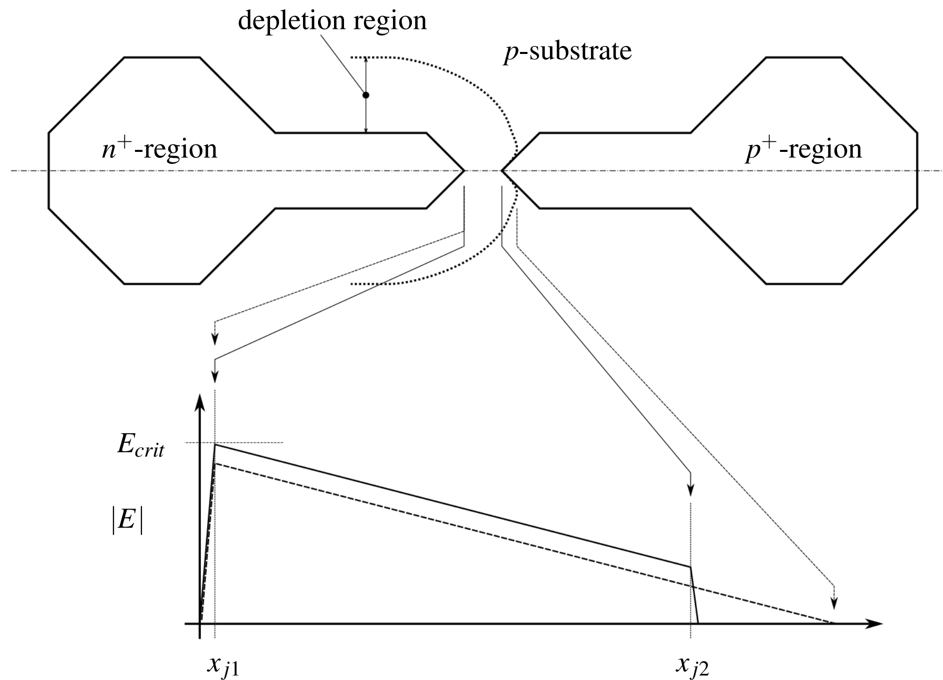


Figure 5.6: Electric field reach through in point sources along with a simplified electric field profile

regions. From the approximation

$$\begin{aligned} E &= -\frac{\partial}{\partial x}\varphi \\ \therefore \psi &\approx -\int_{x_{j1}}^{x_{j2}} E\,\partial x \end{aligned}$$

it can be seen in figure 5.6 that the integral effectively starts at x_{j1} and only extends to x_{j2}, where the p^+-region acts as a field stop. In other parts of the depletion region the integral spans the complete depletion region width W_D in the absence of a highly doped terminating region. As the potential needs to remain constant, the electric field will reach the critical field strength in the compressed depletion region, where the n^+- and p^+-regions are in closest proximity, prior to any other part of the depletion region around the n^+ structure. This local region of higher electric field strength will then become the dominant current generating region and, in addition to the effects discussed in section 4.3.3, will result in an improved localisation of breakdown and current flow.

5.4.3 Layout of a point source using drawn layers in a CMOS process

Figure 5.7 is a representation of the layout in drawn layers for the prototype reach through light source in a CMOS process. The device consists of an n^+-region, a p^+-region and a separation distance defined in the figure as x_{sep}. Standard design rules for the process dictate either a complete

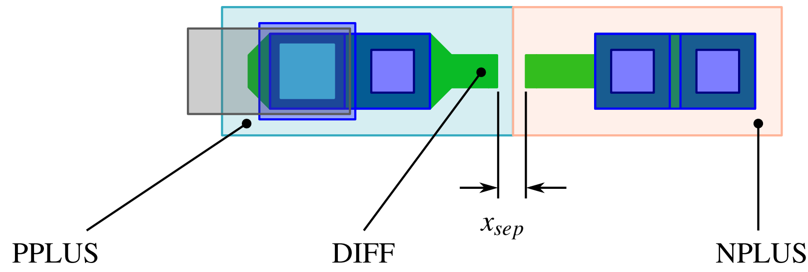


Figure 5.7: Layout of single point source in investigating electric field reach through

abutment, or a separation distance of at least 0.6 μm. The highly doped implants are described by the combination of **DIFF** and **NPLUS** or **PPLUS** respectively, where **DIFF** represents the region where field oxide growth is inhibited. The region of field oxide between the n^+- and p^+-regions masks off the implanted dopants from the region in between, thereby keeping the region at its background doping level. The following values for x_{sep} was designed for the individual point sources and was limited in terms of incremental step by the minimum incremental step of the foundry's design rules.

$$x_{sep} = 275 \text{ nm } + n \cdot 25 \text{ for } n = 0..8 \tag{5.6}$$

The same set of separation distances was used to create arrays of point sources at the same separation distances as described by equation 5.6, where each array consisted of 400 identical elements in parallel. Both the single point sources and arrays were directly connected out to bond pads facilitating the application of DC test signals on the device terminals.

5.5 RESULTS FOR POINT SOURCES IN REACH THROUGH

The point source designs described in section 5.4 is fundamentally a *pn*-junction formed between a highly doped n^+-region and a p-type background determined by the doping of the p-substrate or p-well and the channel stop implants encountered in a lateral direction in parallel to the surface of the wafer. Although simple in design, the layout performs exceptionally well in creating and localising light from the avalanching region of the junction when reverse biased to a sufficient level. This section describes the characterisation process followed in quantifying the electrical and optical characteristics of these point sources as a function of electrical signals applied to the terminals as well as the variation introduced through the separation distance variable x_{sep} shown in figure 5.7 and described by equation 5.6.

5.5.1 Electrical characteristics

The electrical behaviour for the individual point sources and the arrays of point sources are quantified through the application of voltage and current signals applied at DC for these two terminal devices. It is possible to determine the breakdown voltages and electrical junction characteristics of the devices through the application of the correct signals. Large signal behaviour is important for integration purposes while the detailed plots of the logarithmic I-V relationship can give insight into the nature of the electrostatic behaviour of the junction. The tests were conducted on an HP 4155B parameter analyser using a medium integration time setting. Forcing current instead of voltage allowed for much more controlled behaviour as the sensitivity of current on forced voltage is substantially higher than the sensitivity of voltage on a forced current.

5.5.1.1 Large signal current-voltage relationship

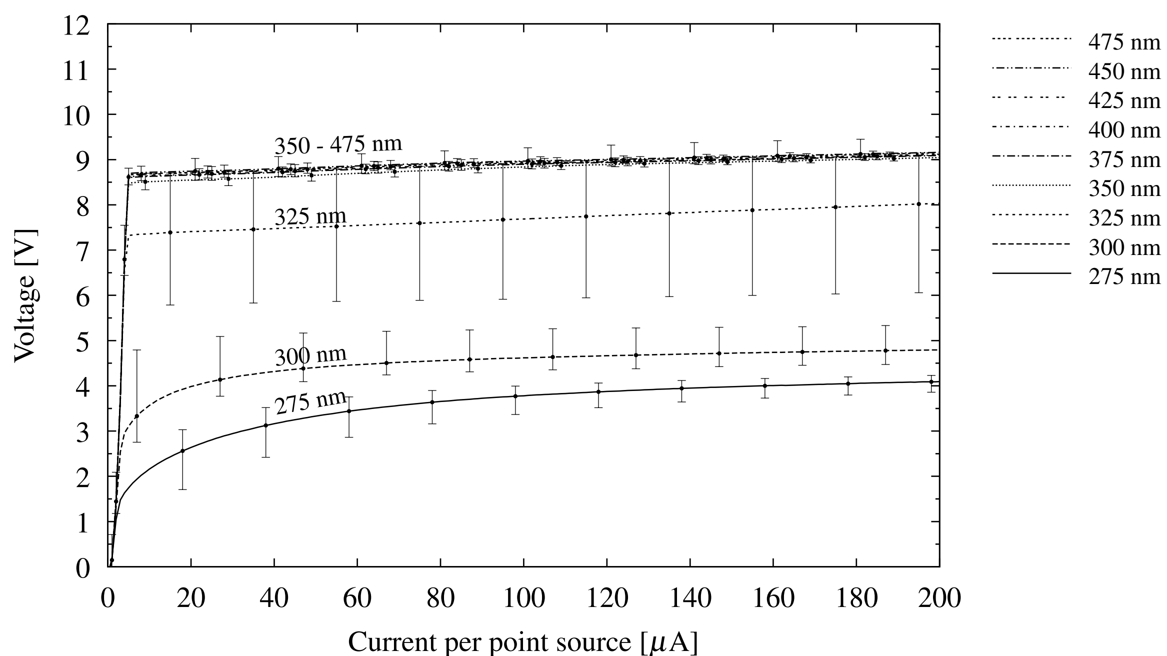


Figure 5.8: Current-voltage relationship of individual point sources as a function of separation distance x_{sep}

An important result is the large signal current-voltage relationship which becomes important for the integration and utilisation of the device when integrated with other CMOS electrical devices. Figure 5.8 shows an averaged result, taken over multiple samples, of individual point sources. A linear

current was applied from 0 to 200 μA in 1 μA steps. The limit of 200 μA was determined to be a safe maximum current that could be applied to an individual point source. Excessive local heating and failure start to occur at currents from 500 μA upward.

Figure 5.8 shows the reverse diode voltage as a function of the driving current. The different curves each represent an average I-V trace of four to eight samples for a specific x_{sep}, with the error bars showing the spread amongst the different samples. The results show a very strong dependency on x_{sep}, with curves for x_{sep} from 350 nm upwards effectively showing no field emission and avalanche breakdown at the foundry specified minimum value, with little spread over sampled device characteristics. At $x_{sep} = 350$ nm the softness of the I-V curve is not prominent as is expected according to the characteristics of field emission described in section 3.2.2. A resistive component in the order of ≈ 2.4 kΩ is seen for devices operating exclusively in avalanche breakdown mode.

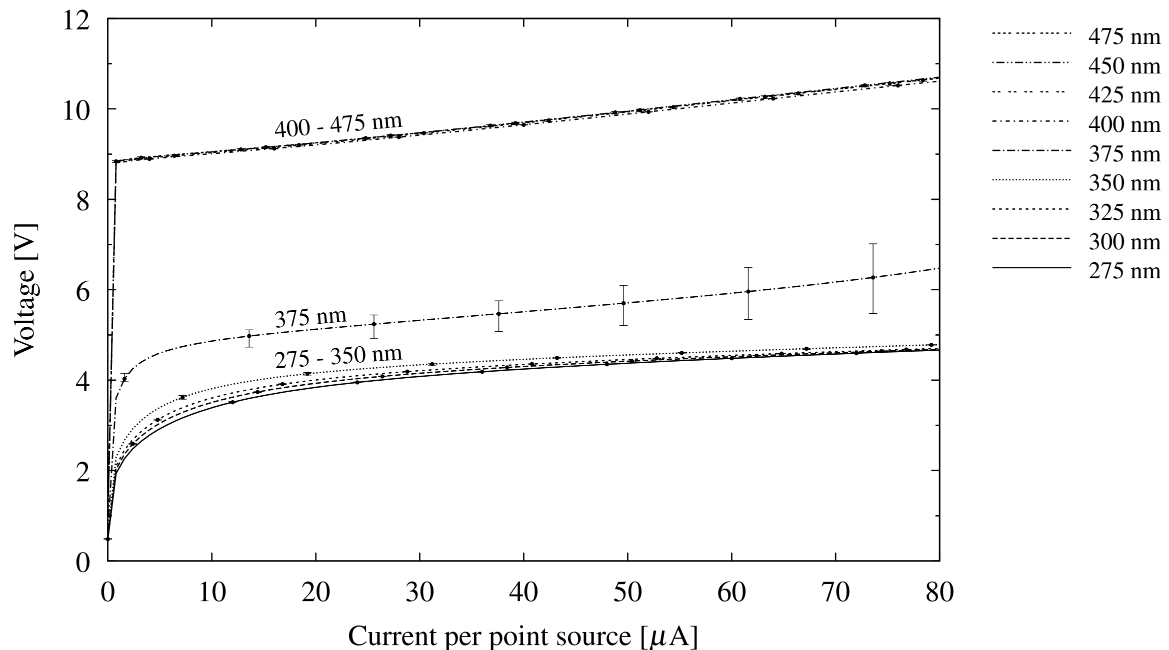


Figure 5.9: Electrical characteristics of point source arrays as a function of separation distance x_{sep}

The arrayed devices were also subjected to a forced current up to a value of 0 - 200 μA per point source which, with 400 elements in parallel, amounted to a maximum driving current of 80 mA. Figure 5.9 shows the average I-V trace for different values of x_{sep}. The resistance characteristics of the arrays where $x_{sep} \geq 400$ nm show a series component of ≈ 23 kΩ or around 9 kΩ per point source. What is also interesting and immediately apparent is the dependency of the I-V curve shape on x_{sep}, which shows that x_{sep} with values from 275 nm - 350 nm does not exhibit the same characteristics as

in figure 5.8 and behaves as expected from Zener-breakdown with a low spread. The 375 nm trace is the only one which departs the Zener-dominant region while still not at the breakdown levels expected from the technology viz ≈ 9 V.

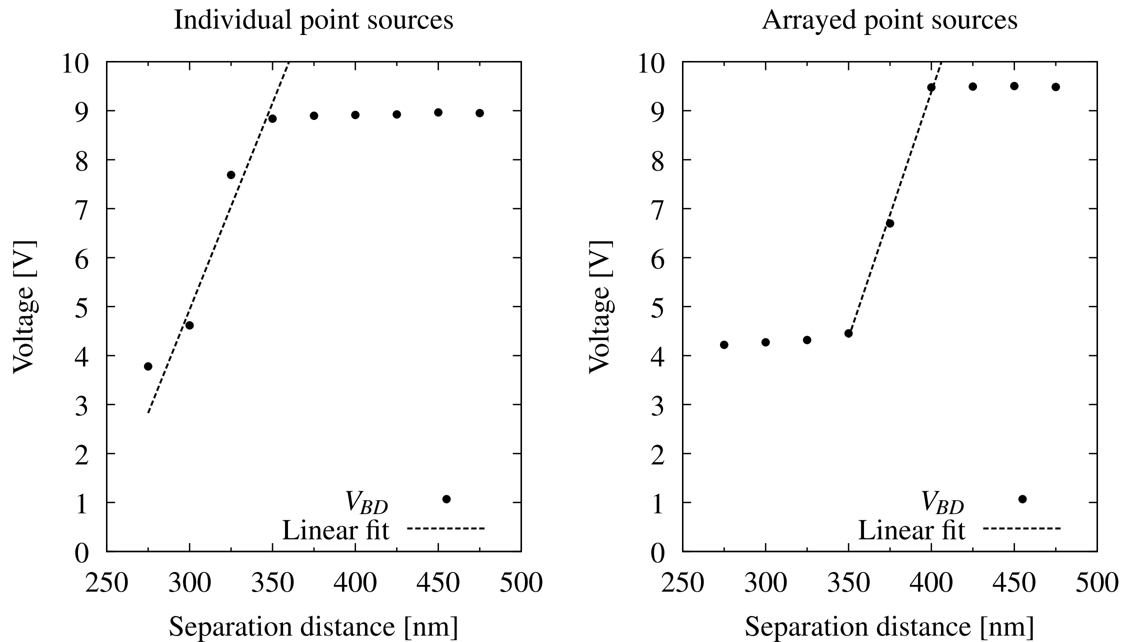


Figure 5.10: Electrical characteristics of individual point sources and arrays with electric field reach through

Although a linear regression approach can be followed over the linear regions of the traces in order to extrapolate the breakdown voltages, the presence of field emission for devices with smaller separation distances makes the breakdown voltage definition somewhat ill-defined. For this reason, it was decided to compare the voltages of the different devices where each individual element is driven at $I_D = 100\ \mu$A. For the arrayed devices, each containing 400 elements, this amounted to a drive current of $I_D = 40$ mA for the 400 element arrays. The results of the breakdown voltage estimations are shown in figure 5.10 for both the individual point sources as well as the arrayed devices. From figure 5.10 it can be shown that the sensitivity of V_{BD} on x_{sep} is described by the following equations when doing a linear approximation over the linear regions as shown in the figure, with equation 5.7 valid for the individual point sources shown in figure 5.8 and equation 5.8 valid for the arrayed devices in figure 5.9.

$$\left.\frac{\Delta V_{BD}}{\Delta x_{sep}}\right|_{indiv} = 84.4 \left[\frac{\text{V}}{\mu\text{m}}\right] \tag{5.7}$$

$$\left.\frac{\Delta V_{BD}}{\Delta x_{sep}}\right|_{array} = 100.4 \left[\frac{\text{V}}{\mu\text{m}}\right] \tag{5.8}$$

5.5.1.2 Junction analysis based on detailed current-voltage results

Although the large signal behaviour can be qualitatively used in determining the breakdown voltages and trying to determine whether the majority current component is due to Zener-transport or avalanche generation, more can be construed from a logarithmic investigation of the current-voltage relationship.

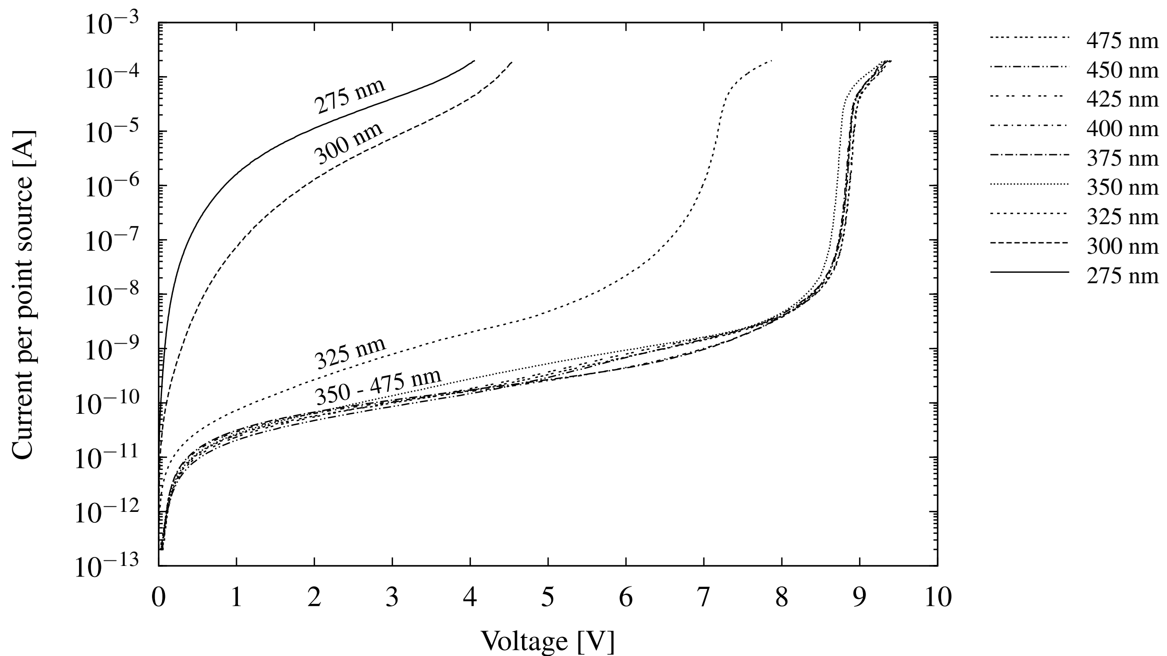


Figure 5.11: Logarithmic plot of average current as a function of voltage for different values of x_{sep} as seen in the individual point source junctions

An inspection of figure 5.11 highlights the fact that appreciable levels of avalanche gain only start to occur in devices with a separation distance of 325 nm which is consistent with the observations from the large signal characteristics. This unfortunately allows us to use only one value of x_{sep} when trying to reduce the voltage without allowing the dominant transport mechanism to be tunnelling. Although

the 325 nm device shows avalanche multiplication from the region of 6 V upwards, it is clear that the leakage component, primarily caused by tunnelling, is still large compared to the devices where $x_{sep} \geq 350$ nm, although orders of magnitude lower than the devices where $x_{sep} \leq 300$ nm.

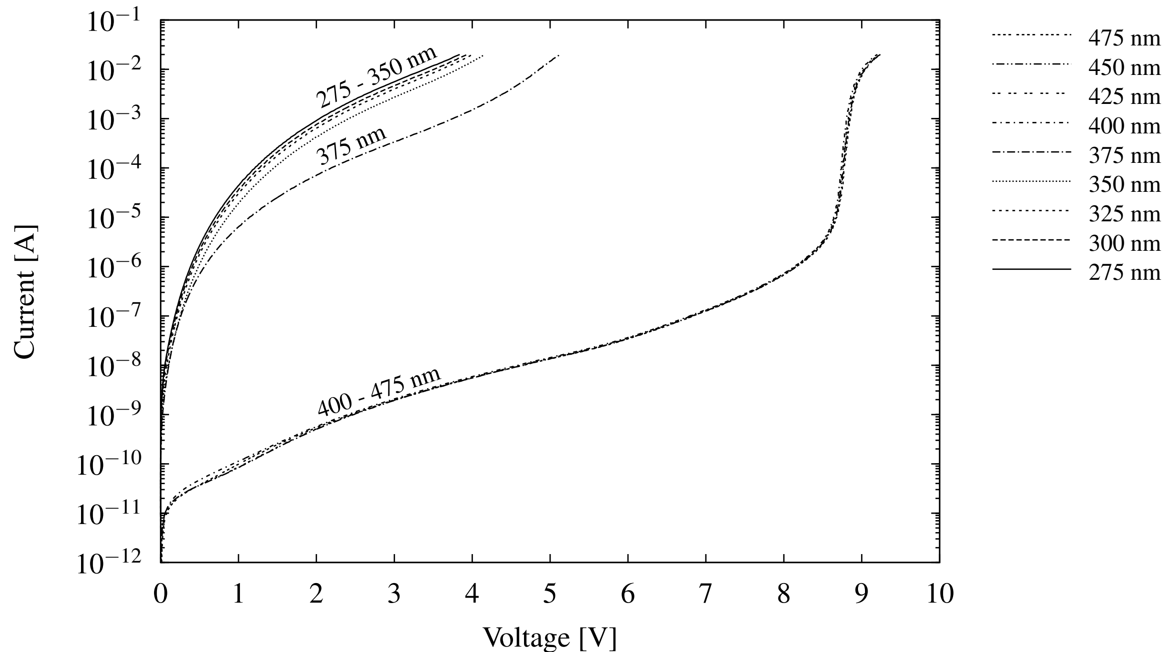


Figure 5.12: Logarithmic plot of average current as a function of voltage for different values of x_{sep} as seen in the arrayed point source junctions

For the arrayed devices, the situation does not look so promising. Figure 5.12 shows that most devices either operate almost exclusively within the Zener-regime or in hard avalanche breakdown, with the characteristics much aligned to that of the reference device. The reference device has been processed in a different batch than that of the test arrays, which may account for the slight deviation. However, all devices where $x_{sep} \geq 400$ nm lies on almost exactly the same trace while devices where $x_{sep} \leq 350$ nm lies practically on top of each other. The only departing trace is that of the junction where $x_{sep} =$ 375 nm, although there is hardly any avalanche multiplication observable.

5.5.2 Optical characteristics

The qualifying metric for determining the performance of a light source is obviously the radiative characteristics of the device. In a similar fashion as for the electrical characterisation of the junctions, measurement of light being generated needs to be investigated for different applied currents as a function of separation distance x_{sep}. In this section, the individual point sources as well as the arrayed

devices are measured to determine the light output characteristics of the sources. A photon multiplier tube (PMT) is used to quantify the individual point sources for a relative evaluation as the output is too low for radiometric or spectral analysis.

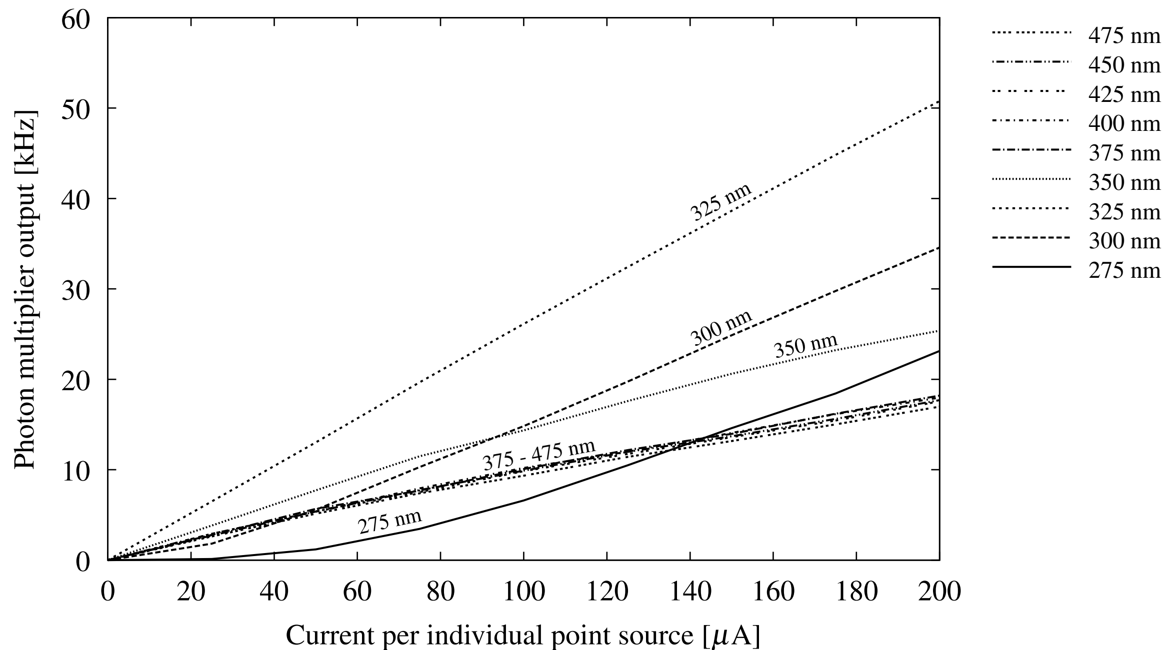


Figure 5.13: Average photon multiplier output measured as a function of forced current and x_{sep} for the individual point source junctions

The results of averaged PMT measurements of multiple samples as a function of current are shown in figure 5.13 for different x_{sep} values, with the dark reading de-embedded from the results. It is very interesting to note the curve characteristics for the 275 nm and 300 nm devices, which both shows a knee at low currents, only to increase to almost linear behaviour at higher currents. Both devices have a substantial initial current contribution due to tunnelling as was observed in the electrical characterisation of the devices. The phenomenon is discussed in the work of du Plessis and Rademeyer [48] and can be used as a technique in order to separate the transport contribution of tunnelling and avalanche multiplication. One of the proofs that the 325 nm device operates at lower voltages due to a lower potential integral lies in the fact that its current to light output is linear. According to results obtained in other work done by our group [48], this shows that the Zener-effect is not the reason for the lower voltage and that the highly doped regions are sufficiently spaced apart for a region of lower background doping to exist in between.

It becomes clear from figure 5.14 that the optical power leaving the chip surface has a strong depend-

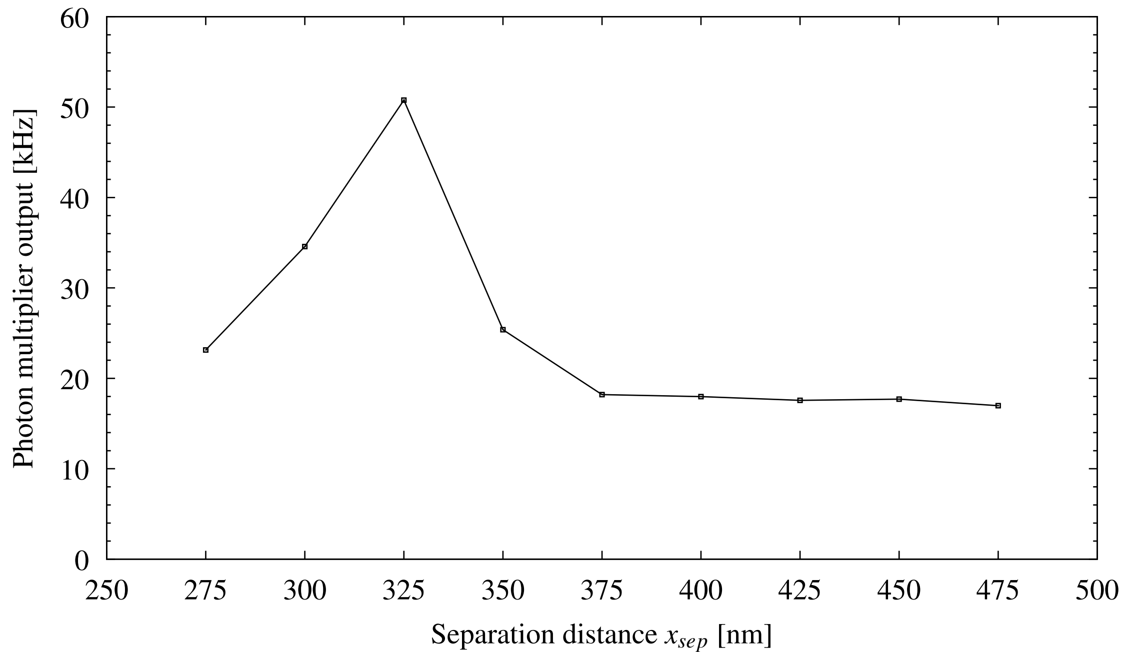


Figure 5.14: Averaged peak output PMT count for each separation distance for the individual point sources

ence on x_{sep}. As the separation distance x_{sep} increases, the light emission initially begins to increase up to 325 nm after which the output begins to drop. One reason, which will be elaborated on in a subsequent section, is that the LOCOS structure starts to affect the light extraction. The complete figure of merit should include the operating voltage as well in order for the comparison to be more representative of the wall plug, or external power efficiency. While the PMT output is not an absolute representation of the optical power leaving the devices, it is useful for a relative comparison. Radiometric and spectral analysis need to be combined for a more accurate description of the optical power leaving the chip surface.

The arrayed devices were measured using a radiometer with a silicon detector. The chip surface was brought into close proximity in order to attempt the capture of all emitted light onto the detection area. Figure 5.15 shows the photon current generated by the detector as a function of input current applied to the array in reverse bias. A trace of the performance of the reference device is also included. Figure 5.15 illustrates that the devices where $x_{sep} \leq 350$ nm, the same devices which show a substantial field emission current from figure 5.12, exhibit non-linear behaviour at the origin of the graph. This is similar to the devices shown in figure 5.13 for the individual point sources in devices where field emission is prominent. The rest of the arrayed light sources in figure 5.15 show a very

Table 5.2: Improvement factors as a function of x_{sep} with $x_{sep} = 475$ nm taken as reference at 200 μA per point source

Separation distance	Improvement factor
275 nm	1.36
300 nm	2.04
325 nm	2.99
350 nm	1.50
375 nm	1.07
400 nm	1.06
425 nm	1.04
450 nm	1.04
475 nm	1

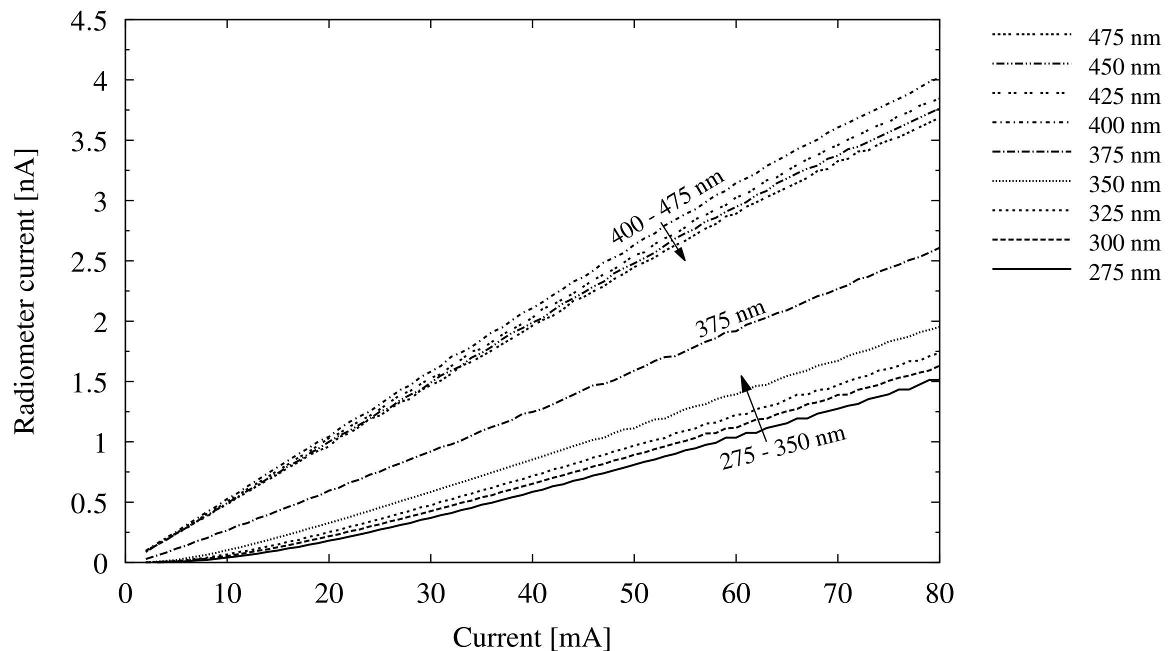


Figure 5.15: Radiometer measurements of point source arrays for light incident on the radiometer photodetector

linear dependence on photon current generated in the detector versus the light source bias current. Contrary to the results in figure 5.13 for the arrayed point sources, almost all devices perform poorer than the device where $x_{sep} = 475$ nm and peak performance is not seen at the same value of x_{sep}. It

is assumed that at $x_{sep} = 475$ nm full LOCOS formation is present and that the electric field does not reach through.

5.5.3 Efficiency

By combining the electrical behaviour of each device with its associated optical performance, it is possible to evaluate the efficiency of the devices based on its optical output signal to input electrical power ratio. Since the idea is to lower the operating voltage, it may well be that the device with the largest optical signal is not necessarily the most efficient. As long as the operating voltage drops faster as a function of x_{sep} than the optical output, the device will become useful thanks to its efficiency.

Figure 5.16 shows the efficiency of point source devices as a function of operating current for different values of x_{sep}. The two devices with tunnel-dominated transport show an increasing efficiency as a function of increased current. The best device is the junction with 325 nm separation distance, although a cross-over occurs between 100 μA and 125 μA, where the 300 nm device overtakes it in terms of efficiency. This is mainly due to its lower voltage operation even though the optical signal is smaller in magnitude. The monotonic decrease in most cases can partly be attributed to an increase of power loss associated with series resistances in the conduction loop.

From figure 5.16 it is clear that the 325 nm device will benefit from low operating currents, with more elements in an arrayed fashion. The trade off is area, which will increase in order to place more elements in parallel for the same effective input current. If area is a constraint, or if more current per element can be employed, the 300 nm device outperforms the rest. The upper limit of operation will of course be the maximum power consumption of the device before failure occurs. The 475 nm device serves as the reference device as it is in hard avalanche breakdown without any electric field reach through or tunnelling. Improvements are shown in table 5.3 for the different values of x_{sep} at both low and high current per point source.

Figure 5.17 shows the output of the photon current generated by the radiometer detector for a given input power as a function of current. The assumption is that the spectral content of the optical signal detected by the radiometer stays constant as a function of array bias current. This plot then represents a metric similar to wall plug efficiency which shows the relationship between input electrical power and output optical power. As can be seen from the plot, the reference device outperforms all devices

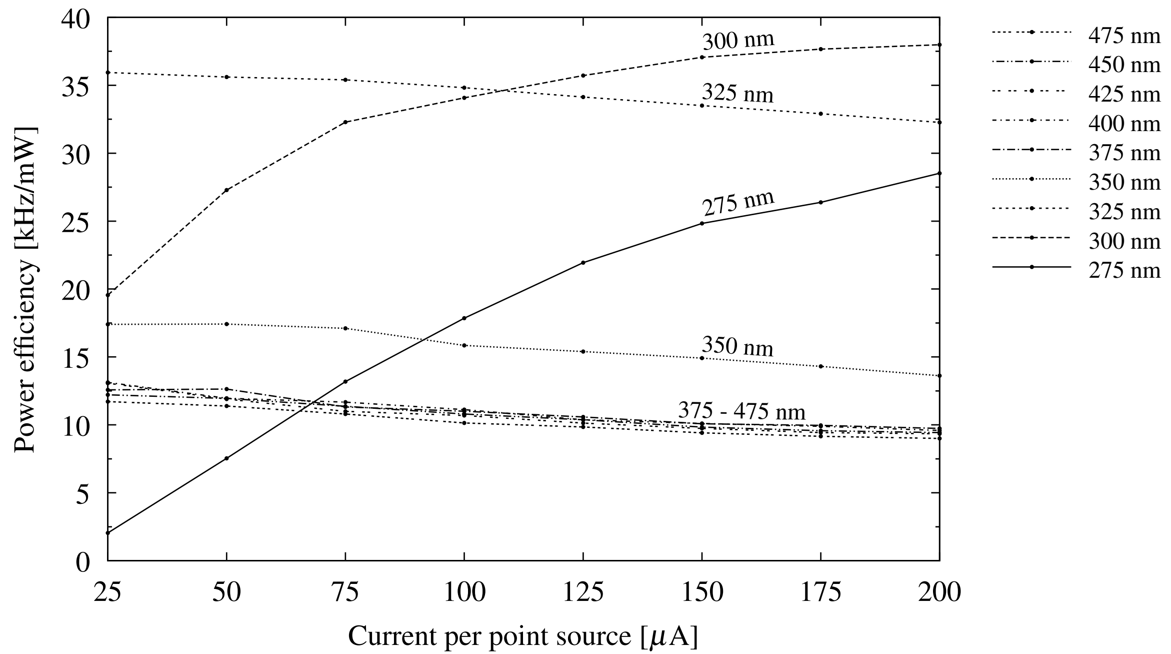


Figure 5.16: Efficiency of point sources as a function of input current for different values of x_{sep}

Table 5.3: Efficiency improvement compared to a reference device

x_{sep}	Low I_{DC}	High I_{DC}
275	0.17	3.17
300	1.67	4.22
325	3.07	3.58
350	1.49	1.51
375	1.07	1.08
400	1.12	1.06
425	1.12	1.04
450	1.04	1.05
475	1.00	1.00

in terms of efficiency, even with an operating voltage of around 9 V. At large currents, the efficiency of the reference device is superseded by the 350 nm and 375 nm device.

Another way of representing the performance of the arrayed light sources is by examining the improvement factors when compared to the reference device. This representation is shown in figure

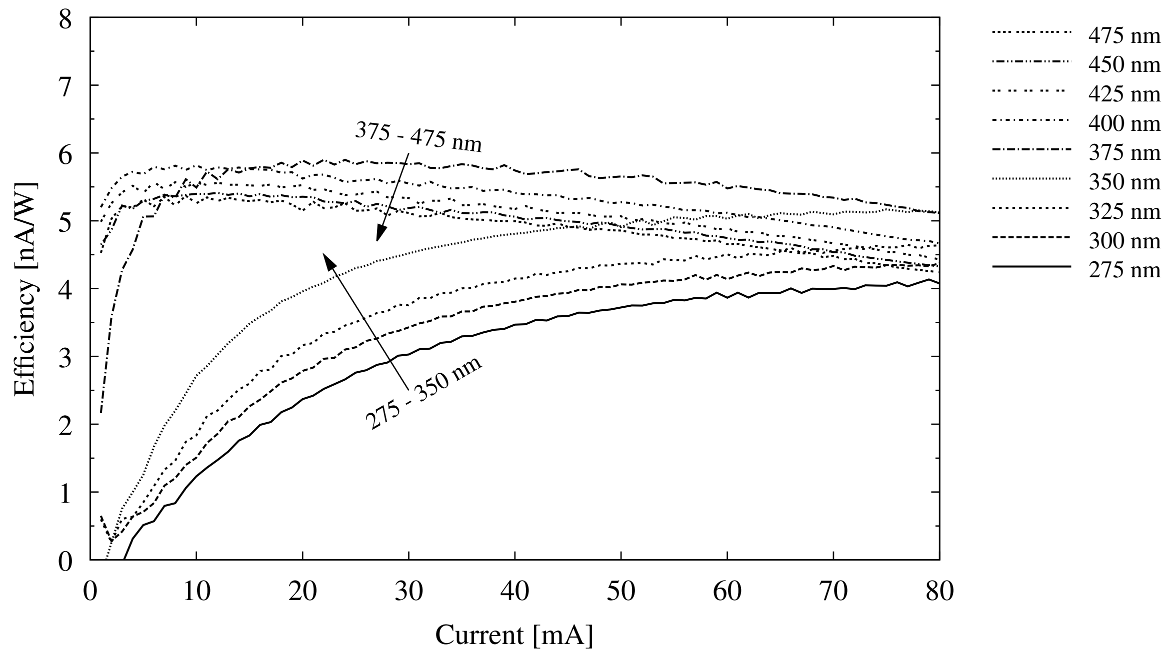


Figure 5.17: Efficiency point source arrays

5.18 where, with the reference located at $y = 1$, the device performance is predominantly poorer than with devices where the separation distance x_{sep} is kept at the design rule minimum, such as in the original design taken from chapter 4.

5.5.4 Summary of results

The electrical results for individual point source elements show the possibility of lowering the breakdown voltage, although the breakdown voltage is extremely sensitive to x_{sep}, the separation distance between the two highly doped regions. There is a discrepancy between the individual point source element results and the arrayed point source results which can partly be ascribed to the sensitivity of the breakdown voltage on x_{sep} and the difference between the layouts of the individual point sources and the arrays. In both cases, it is not possible to predict the breakdown voltage beforehand based on x_{sep} as the spread between samples is too large.

From the optical results it is apparent that there is a great improvement in optical emission for the individual point sources where $x_{sep} = 325$ nm. The reason is probably that the field oxide isolation between the two devices is not as deep between the highly doped regions. Light is therefore generated against a Si/SiO_2 interface more aligned to the chip surface. This results in less internal losses and

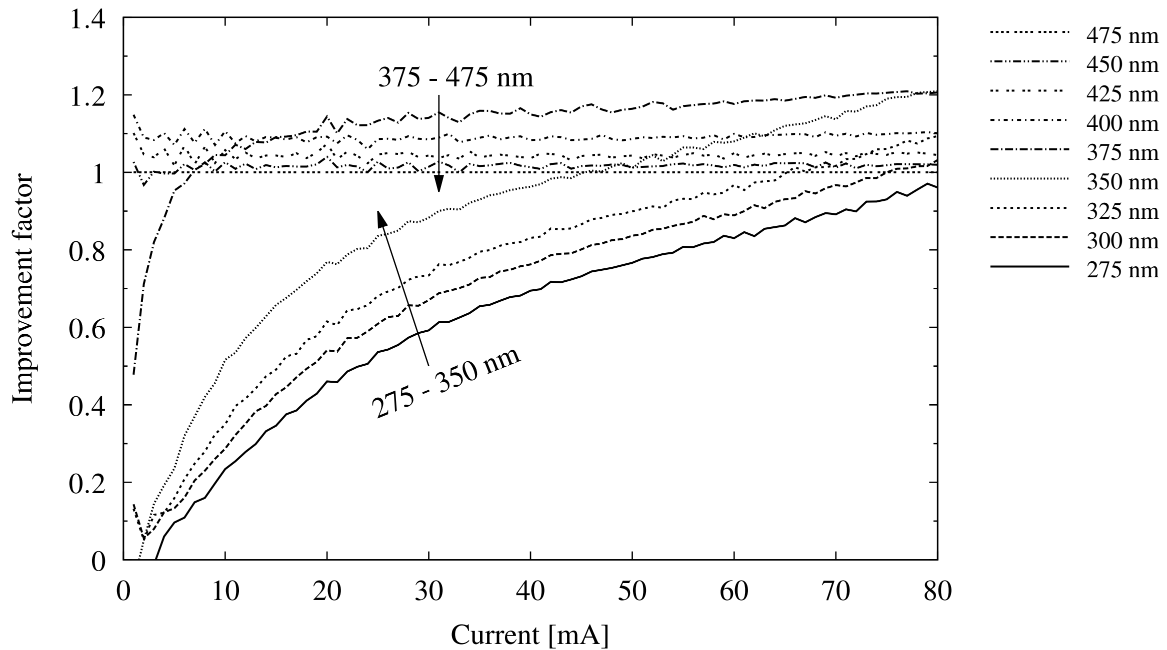
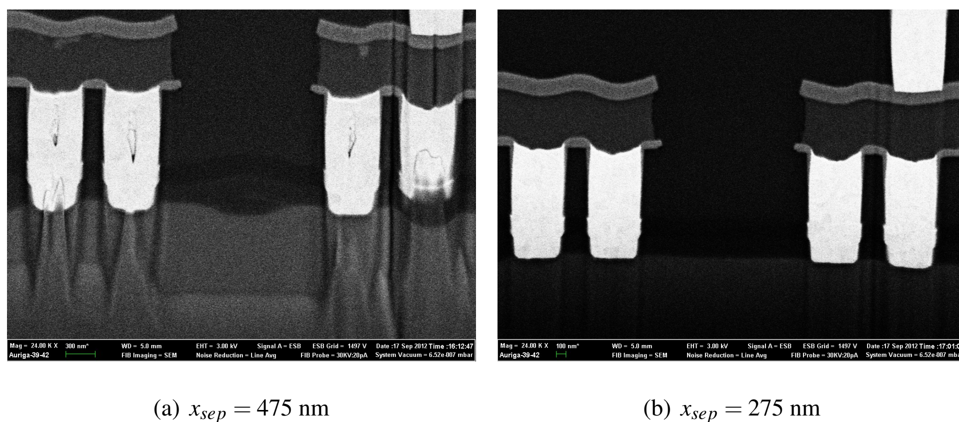


Figure 5.18: Point source array improvement

better external extraction efficiency. The field oxide is used in this design to prevent dopants from entering the silicon by masking the implants. Figure 5.19(a) shows a cross section of the field oxide clearly visible between the two contacts for a separation distance of $x_{sep} = 275$ nm, while figure 5.19(b) shows that for a separation distance of $x_{sep} = 275$ nm none of the field oxide remains and the implants are effectively abutted.



(a) $x_{sep} = 475$ nm

(b) $x_{sep} = 275$ nm

Figure 5.19: SEM photos imaged through the use of back-scatter detector

In summary, the following highlights are derived from the preceding results.

- Lower voltage operation is possible by reducing the distance between two highly doped regions.
- High sensitivity on x_{sep} prevents a systematic and deterministic method for predictable behaviour of breakdown voltage as a function of x_{sep}, while the addressable breakdown voltages between the field emission and avalanche limit are few.
- Substantial differences between individual point source and arrayed device behaviour substantiate the sensitivity to process variations while the field oxide region exists between highly doped regions.
- The reliability of arrayed devices is low in the region between the field emission and avalanche limits.

5.6 IMPROVED CMOS GEOMETRY WITH FIELD OXIDE CHANNELS

The results in section 5.5 present one major drawback in following the mentioned approach: the sensitivity of device performance depends heavily on the LOCOS region between the two highly doped regions. The devices operate either almost exclusively in the field emission regime or in strong avalanche, with only one dimension lying in between. The approach introduced in this section aims at removing the uncertainty introduced by the LOCOS region altogether, while instead using the shallow implant species selection masks in order to define which regions are heavily doped and which regions remain at the background doping concentration.

5.6.1 Geometrical approach

In the CMOS process used in this work, the formation of LOCOS is inhibited by defining an active region through the **DIFF** or **ACTIVE** mask. The design rules specify that **DIFF / ACTIVE** should not be used without defining the shallow implant regions. In the gate region of a MOS transistor, the gate and spacer structures mask off the implants, while where **DIFF / ACTIVE** is not defined, the field oxide simply serves a mask.

It is proposed in this work that, where a CMOS process has separate mask sets for **NPLUS** and **PPLUS**, that is, where one is not derived from the negative of the other, it should be possible to use these masks to define the implant regions. The same holds for a CMOS process employing a mask layer which inhibits the formation of shallow trenches where the isolation technology is based on

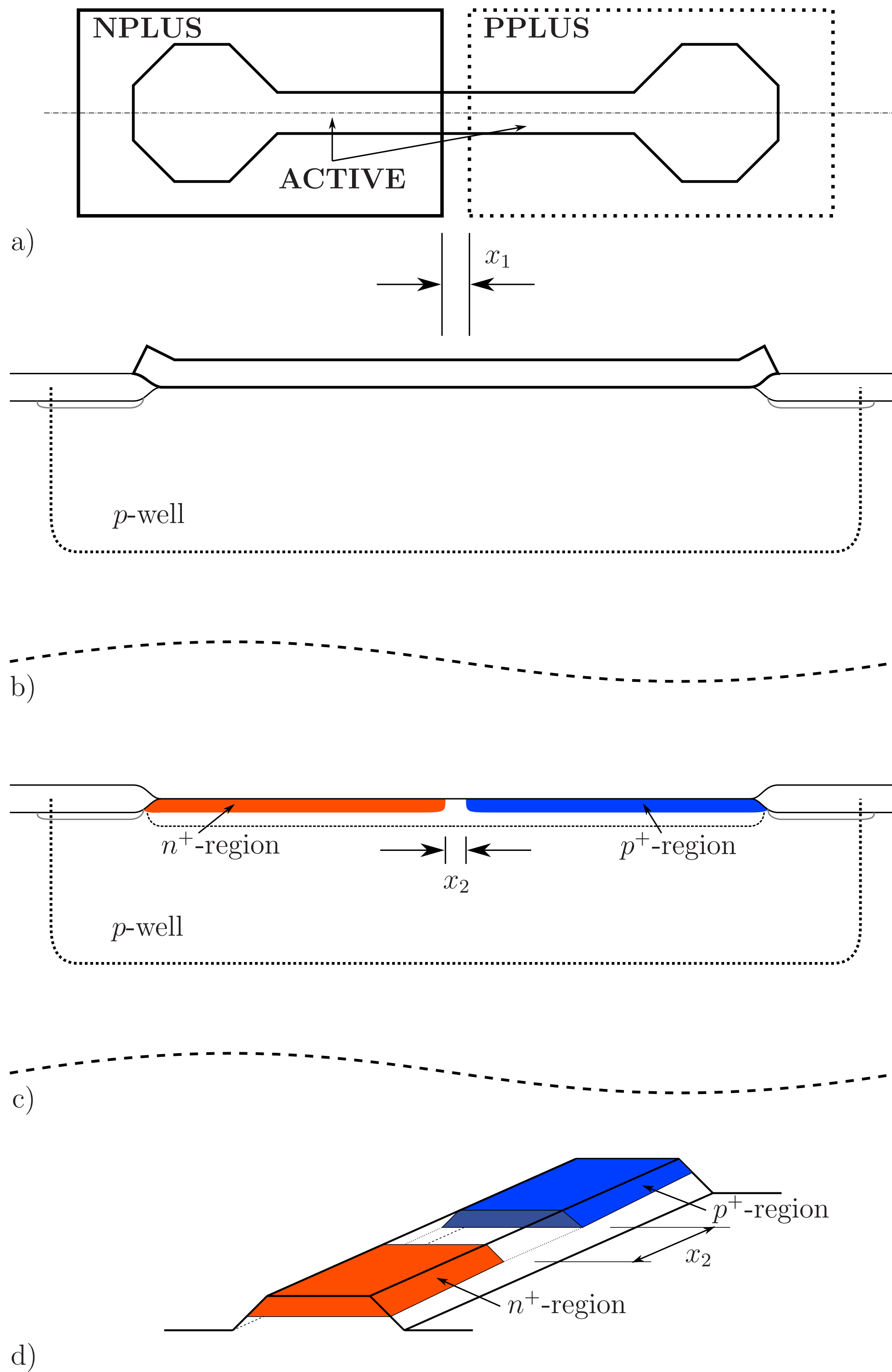


Figure 5.20: Approach to define a *pn*-junction without LOCOS isolation intended to promote electric field reach through in the region where light is generated

shallow trench isolation (STI). Silicide and salicadation may influence both the electrical and optical behaviour of a light source based on a *pn*-junction. Since the CMOS process used in this work is based on a polycide process this is not of a concern since only the polysilicon layer is affected. Figure 5.20 illustrates the approach. The definition in drawn layers uses the **PPLUS** and **NPLUS** masks to set the separation distance shown by x_1 in figure 5.20a. The complete area defined with **ACTIVE** will be masked by Si_3N_4 (figure 5.20b). Due to lateral straggle of the implantation the actual separation distance shown by x_2 in figure 5.20c will be less than x_1. Moreover, the formation of a channel bordered by oxide is formed along the x-direction of the finger-like structure effectively confining carrier action in the direction of intended current flow as illustrated in three dimensions in figure 5.20d.

5.6.1.1 Polycide and salicide processes

As CMOS nodes scale to even smaller structures, the gate, source and drain implanted regions of the devices present an increase in resistance along the length of the gates and between the channel and the contact in the highly doped source- and drain regions. The remedy for improving conduction lies in silicide, a composite material consisting of silicon and usually a metal, followed by a specific process. At present, technology nodes such as the 0.35 μm and larger nodes make use of polycide, while smaller and more densely packed technologies make use of self-aligned salicidation.

The target process for this work utilises silicide on the polysilicon gates of its MOS devices to reduce the resistivity of the material in what is called a polycide process. This process occurs before the etching of the individual polysilicon gates. The wafer is covered in polysilicon. The metal deposition and silicidation are done prior to the lithographic and etching steps, resulting in polycided gates.

In contrast, a salicide process is one which affects both the gate and source/drain implant regions and is effected after the definition and creation of transistor gates. The gate spacer structures do not react in the process as these are not silicon, which prevents short circuits from the gate to the implants. This improves gate conductance as well as the formation of low resistance contacts to the source and drain regions of a MOS transistor.

It is very important to note that, subject to the requirements mentioned in section 5.6.1, a salicidation process would still create a low resistance path between the source and drain regions, irrespective of their implant definitions, as even the region where STI or LOCOS has been inhibited will still be sub-

jected to the formation of silicide. The process should therefore contain at least one lithographic layer where silicidation can be expressly blocked. Fortunately most modern CMOS processes provide for the integration of resistors based on p^+- or n^+-region resistors and allow explicit blocking definitions for resistors which can be used over the background doped region as mentioned in section 5.6.1 to prevent a short between the highly doped regions and maintain a valid *pn*-junction.

5.6.2 Design of junctions with oxide channels

While figure 5.20 shows the approach, the design for a single point source is shown in figure 5.21 using the drawn layers of the target CMOS process used in this work. These individual point sources were used for electrical and optical characterisation, while arrays consisting of 392 elements were used to generate larger optical output signals for radiometry and spectrometry. As the process uses polycide, no blocking layer was required across the region indicated by x_{sep}. The chosen values for x_{sep} in the prototype structures are given by formula 5.9, where the spread was chosen over a much larger value that for the point sources described in section 5.4.

$$x_{sep} = n \cdot 50 \text{ for } n = 0..8, 12 \tag{5.9}$$

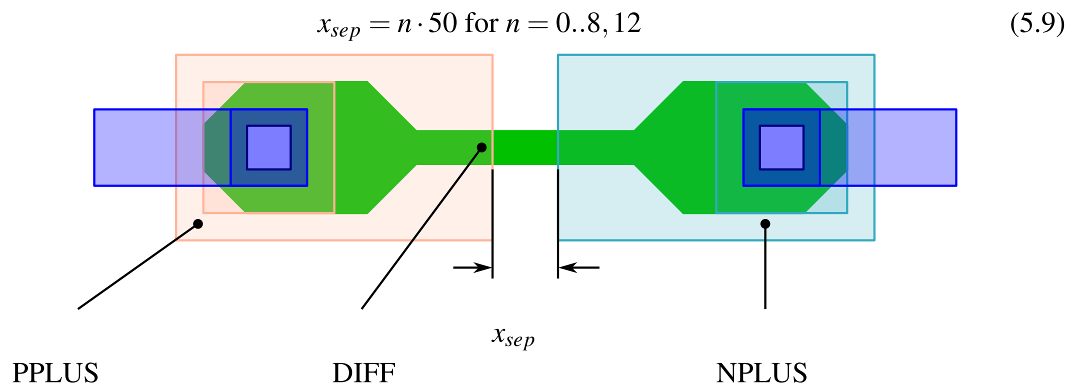


Figure 5.21: Layout of single point source with oxide channel formed by LOCOS or STI

5.6.3 Fabrication and resulting structures

Although CMOS foundries keep their specific process flows strictly confidential and only disclose enough information for the typical circuit designer to accomplish his designs, the effect of the designs portrayed in figures 5.20 and 5.21 is shown in a diagrammatic way in figure 5.22 for an n^+p-device.

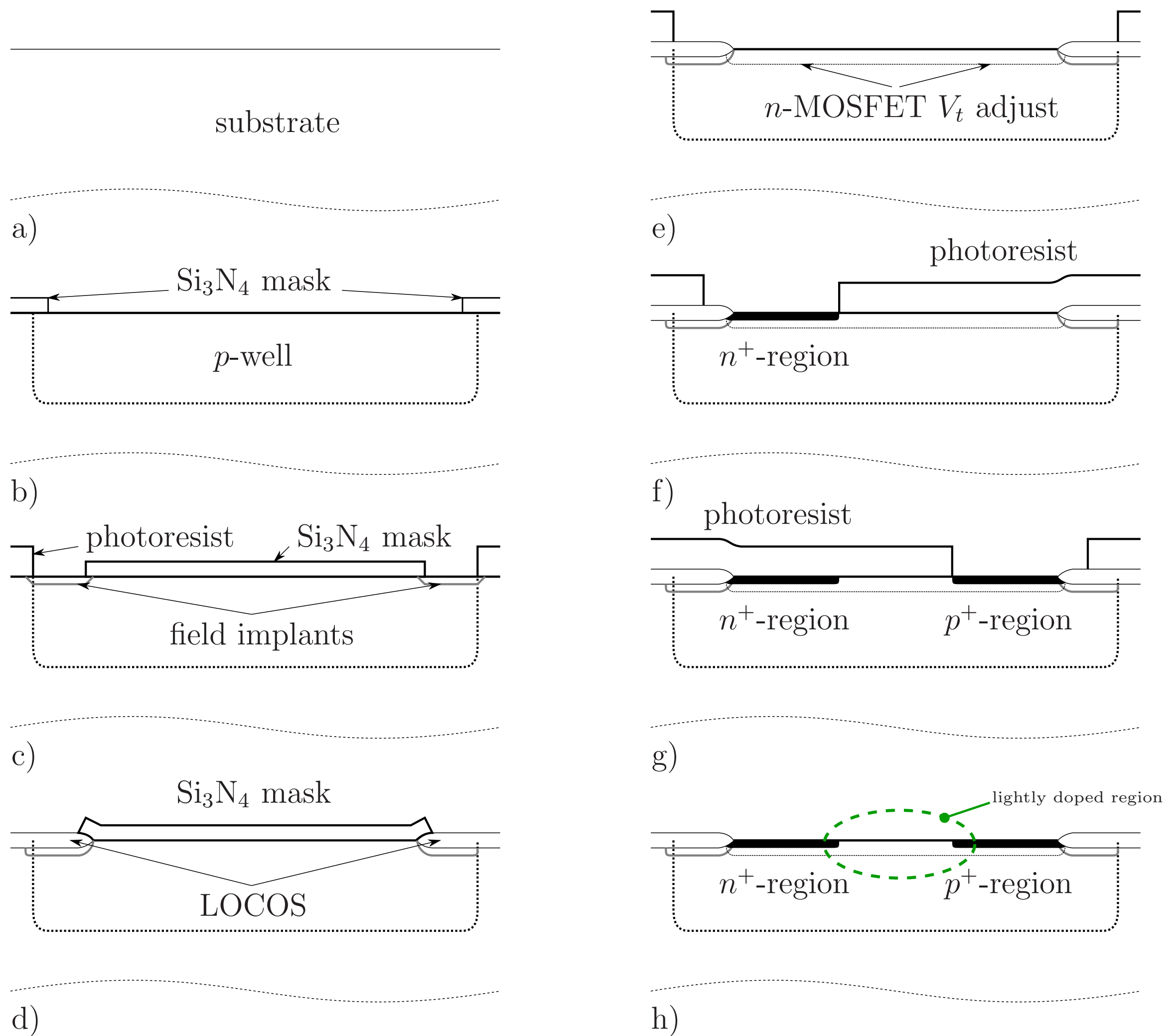


Figure 5.22: The alteration of processing steps allows for the LOCOS isolation to be inhibited while maintaining a region of lower doping in between

In figure 5.22a the wafer starts off serving as a p-type substrate, after which a p-well is implanted and thermally driven in (fig 5.22b). A field implant is then done wherever the active region is not defined, with LOCOS growth occurs in all but the active regions as shown in 5.22d. A channel implant is done in the active region for controlling the threshold voltages of the MOS transistors, after which the n^+- and p^+-regions are defined as described in section 5.6.1. The result should be similar to a structure shown in figure 5.22h with no LOCOS isolation between the two heavily doped regions.

5.7 RESULTS FOR IMPROVED SOURCES IN REACH THROUGH

The same tests performed in section 5.5 were used in characterising the channel-type devices with the LOCOS feature removed described in section 5.6.1. As the doping concentrations of the highly doped implanted regions remain identical, along with the assumption that the LDD regions are derived from the implant definitions where the latter dominates in terms of doping concentration magnitude, it can be expected that the junctions will once again break down close to 9 V when implemented as an n^+p-junction, with the background affected by the channel implant doping. Breakdown is expected to occur close to the surface which is beneficial for light emission when compared to the situation described in section 4.3.4 as there should now exist no LOCOS "bird's beak". The electrical results are first analysed for both the individual channel sources and the arrayed structures, after which the optical results are compared.

5.7.1 Electrical characteristics

As mentioned previously, the large signal characteristics are important from a device integration perspective as the terminal characteristics ultimately determine its usefulness when integrating the light sources with other electrical circuitry on chip. Figure 5.23 shows the current-voltage relationship of individual channel-type devices for the various values of x_{sep}. It is immediately noticeable that, although the incremental separation distance is larger in the channel devices than the point source structures described in section 5.4, that is, equations 5.6 compared to 5.9, the spread of the traces is much more gradual for the different values of x_{sep}. Only from 350 nm upwards the nominal process specific breakdown voltage of 9 V is achieved.

5.7.1.1 Large signal current-voltage relationship

A very interesting observation is that the device where $x_{sep} = 0$ nm has a higher voltage for the same current than the device where $x_{sep} = 50$ nm. The explanation is that the region x_2 in figure 5.20 is not perfectly abutted, but rather, due to the straggle of the ion implantation and possible lithographic misalignment, the dopant species intermix to create an effective regional doping lower than that of the 50 nm device. From 0 nm to 150 nm the devices exhibit a strong tunnelling contribution resulting in current flow, whereas devices from 200 nm upwards exhibits behaviour indicating that avalanche multiplication is the dominant current contributor. Once again, figure 5.23 shows the average current-

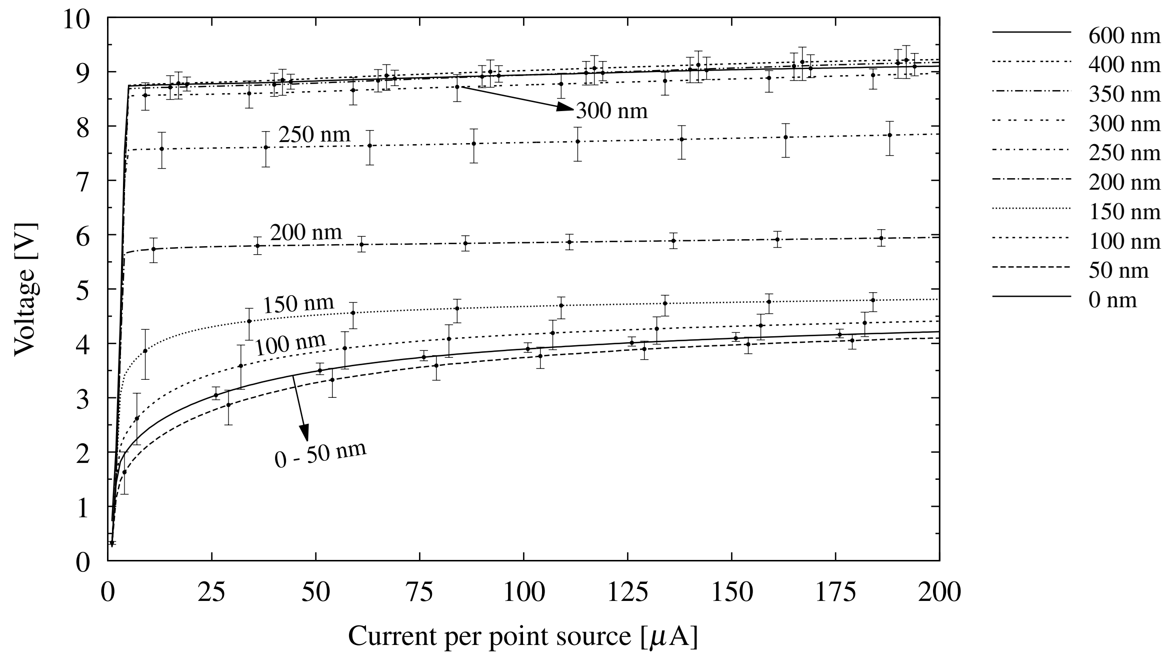


Figure 5.23: Electrical characteristics of point sources with field oxide channels for different values of x_{sep}

voltage trace of multiple samples. The spread is indicated by the error bars and is noticeably smaller than for the devices in figure 5.8 where the field oxide region still exists between the two highly doped regions.

When utilised in an array of parallel light sources, the results look much more promising. The spacing increments for the arrays were once again reduced as the sweet spot from figure 5.23 was determined to be between 150 nm and 250 nm, where formula 5.10 describes the separation distances for the implemented arrays.

$$x_{sep} = 150 \text{ nm } + n \cdot 25 \text{ for } n = 0..4 \tag{5.10}$$

Figure 5.24 shows the traces of arrayed devices with values for x_{sep} described by formula 5.10. None of these devices exhibit strong Zener-like behaviour nor the typical breakdown voltage of the junctions from chapter 4, with an almost linear increase in breakdown voltage as a function of x_{sep}. Moreover, the behaviour of the arrayed devices corresponds with a much higher correlation to that of the individual devices in terms of their breakdown characteristics. For example, the 200 nm device breaks

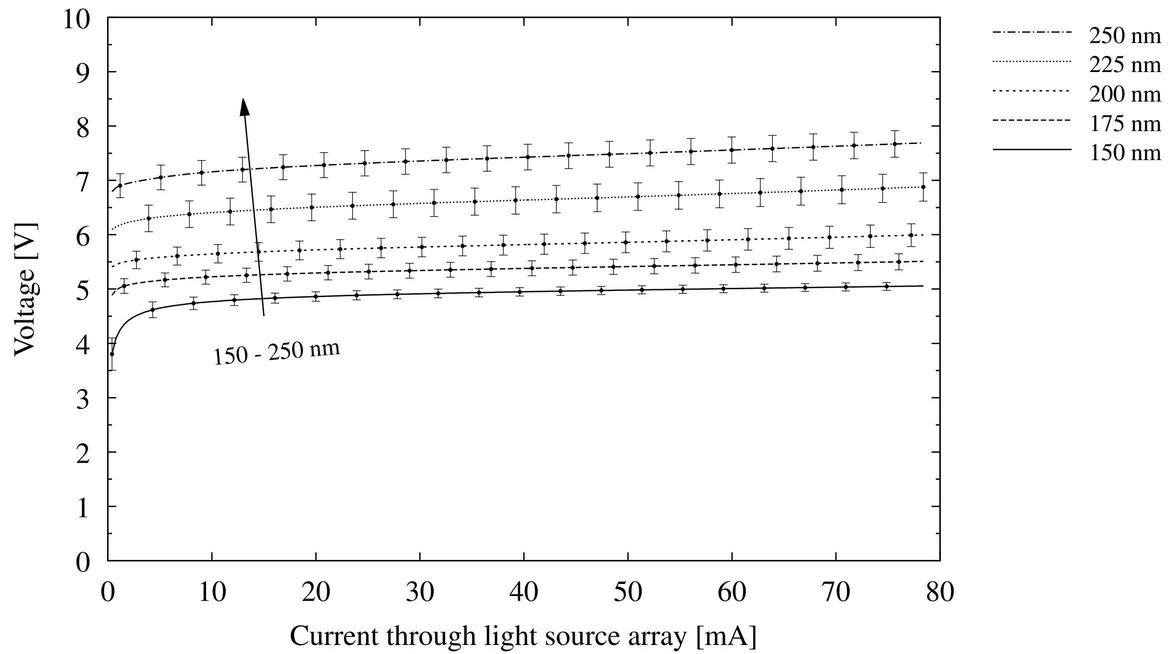


Figure 5.24: Electrical characteristics of point source arrays with field oxide channels for different values of x_{sep}

down just below 6 V in both figures 5.23 and 5.24. The spread of individual devices also remains within reasonable limits, indicating that the proposed technique is conducive towards repeatable array designs.

The breakdown voltages were defined where $I_D = 100\ \mu$A for individual point source elements and $I_D = 39.2$ mA for the 392 element arrays in order to construct figure 5.25. Based on a fit over the linear regions of the plots, the sensitivity of breakdown voltage to separation distance is shown to be much reduced when compared to equations 5.7 and 5.8, with the behaviour of the channel devices shown below.

$$\left.\frac{\Delta V_{BD}}{\Delta x_{sep}}\right|_{indiv} = 28.1 \left[\frac{\text{V}}{\mu\text{m}}\right] \tag{5.11}$$

$$\left.\frac{\Delta V_{BD}}{\Delta x_{sep}}\right|_{array} = 24.8 \left[\frac{\text{V}}{\mu\text{m}}\right] \tag{5.12}$$

Not only is it possible to obtain a much more repeatable breakdown voltage for a given value of

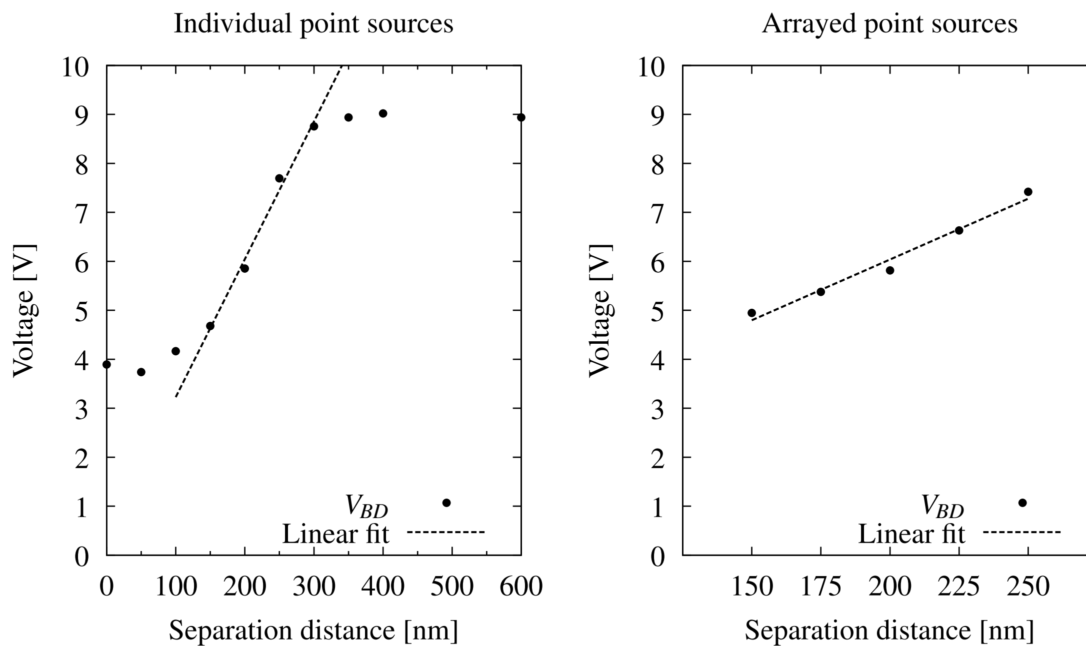


Figure 5.25: Breakdown voltages for both individual point sources as well as arrayed devices as a function of x_{sep}

x_{sep}, but the removal of the field oxide region and the configuration of the junction into a "channel"-like structure reduces the sensitivity of the breakdown voltage dramatically when compared to the previous approach described in section 5.5.

5.7.1.2 Junction analysis based on detailed current-voltage results

Each individual point source was subjected to a forced logarithmic current and a voltage measurement was taken in order to produce the results in figure 5.26. Like the large signal results, it is easy to identify which of the devices operate under conditions where tunnelling is the dominant transport mechanism and which devices exhibit avalanche carrier generation. What is interesting is the spread, even though the incremental $\Delta x_{sep} = 50$ nm is much larger than in the case of the point sources where field oxide exists between the two highly doped regions. Furthermore, from 200 nm upwards the dominant transport mechanism is avalanche generation. At small reverse bias voltages, the junction where $x_{sep} = 250$ nm follows the same trace as junctions with larger x_{sep} values, but departs this characteristic at around 1 V. A more in-depth look at the arrayed structures, where $\Delta x_{sep} = 25$ nm is discussed below.

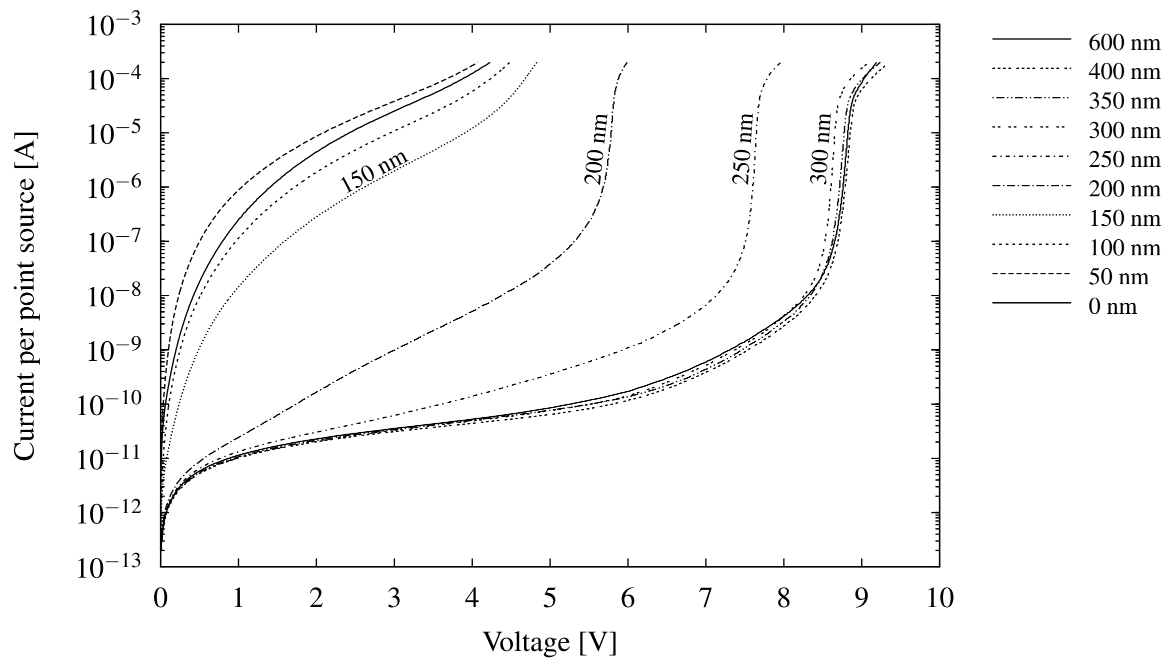


Figure 5.26: Logarithmic current-voltage relationship of individual reverse biased channel-type light sources for different values of x_{sep}

Figure 5.27 contains the logarithmic current-voltage relationship for the devices implemented as arrayed point sources with field oxide inhibited between the two highly doped regions. The first very interesting observation is that of the 425 nm device where the field oxide separating structure is still present. This device is used as reference for the set of results in figure 5.27. Compared to the results without the field oxide region, the initial leakage current is relatively high. One explanation may be that the junction forms against the LOCOS structure interface which contains a higher amount of dangling bonds between the Si/SiO_2 interface. Where the field oxide is absent in the form of a LOCOS structure between the two highly doped regions, the junction is partly bounded by a gate oxide. A dry thermal oxidation is usually used for growing the gate oxide which results in a Si/SiO_2 interface with fewer dangling bonds. Note that this phenomenon needs to be verified in future work. The arrayed junctions where field oxide is not present exhibit the same transition from tunnelling dominated current to avalanche current as x_{sep} is increased as noted in the devices in section 5.5. What is interesting is that from 200 nm upwards, the initial trace behaviour lies on the same curve until it departs at increasing voltages as a function of x_{sep}. Note the extremely low initial leakage currents. It is then also interesting to note the two distinct slopes across the operating regions of devices with $x_{sep} \geq 200$ nm, all of which eventually show strong avalanche multiplication. This departure from

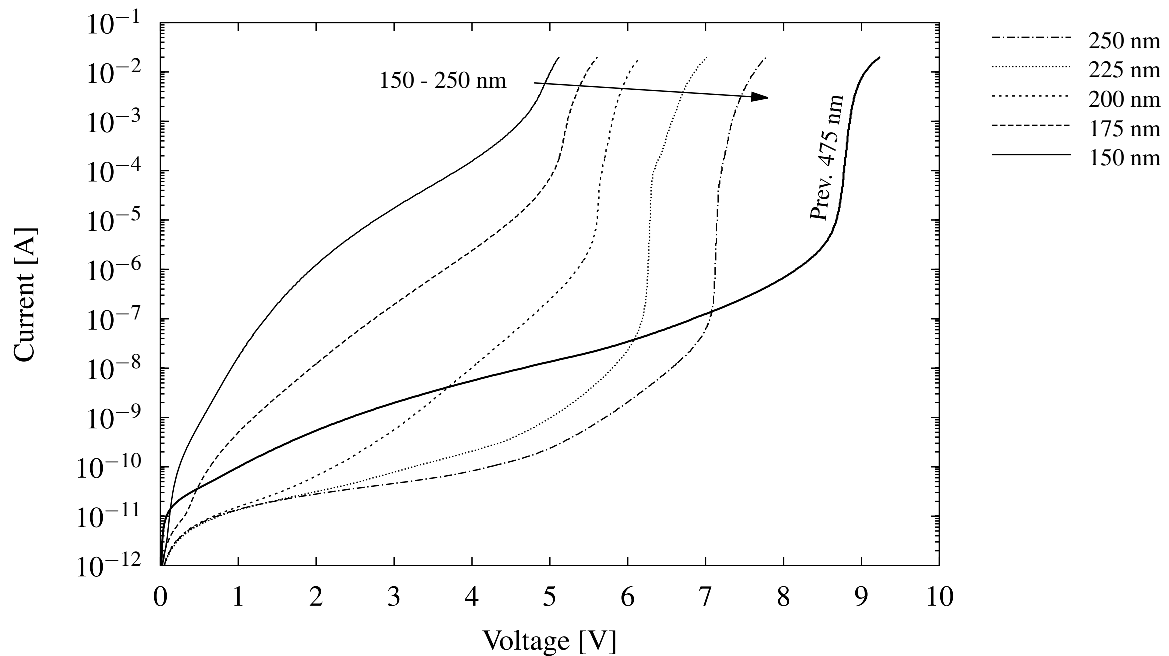


Figure 5.27: Logarithmic current-voltage relationship of arrayed channel type devices for different values of x_{sep}

the shared curve at low current is believed to be the onset of field emission once the electric fields are sufficiently high (see section 3.2.2) causing spatial proximity between the conduction and valence band. At this point, band-to-band tunnelling contributes to the current transport mechanism before appreciable avalanche multiplication starts to occur.

5.7.2 Optical characteristics

A comparison between the PMT output as a function of device current is shown in figure 5.28 as averaged across multiple samples. The traces for $0 \leq x_{sep} \leq 150$ nm show an initial lag behind the device current before transitioning to a linear dependency. The output of the PMT counter shows that where $200 \leq x_{sep} \leq 350$ nm the output is linearly proportional to the input current, with the 400 nm device and 600 nm serving as reference starts to droop for currents larger than 140 μA. The legend on the right hand sign of the graph is ordered such as to directly correlate in sequence with the traces at the maximum forced current point. The emission intensity increases as a function of increasing x_{sep} up to a maximum at $x_{sep} = 300$ nm, where after the emission intensity starts to drop again.

For improved clarity, figure 5.29 shows the peak light output at 200 μA as a function of x_{sep} which

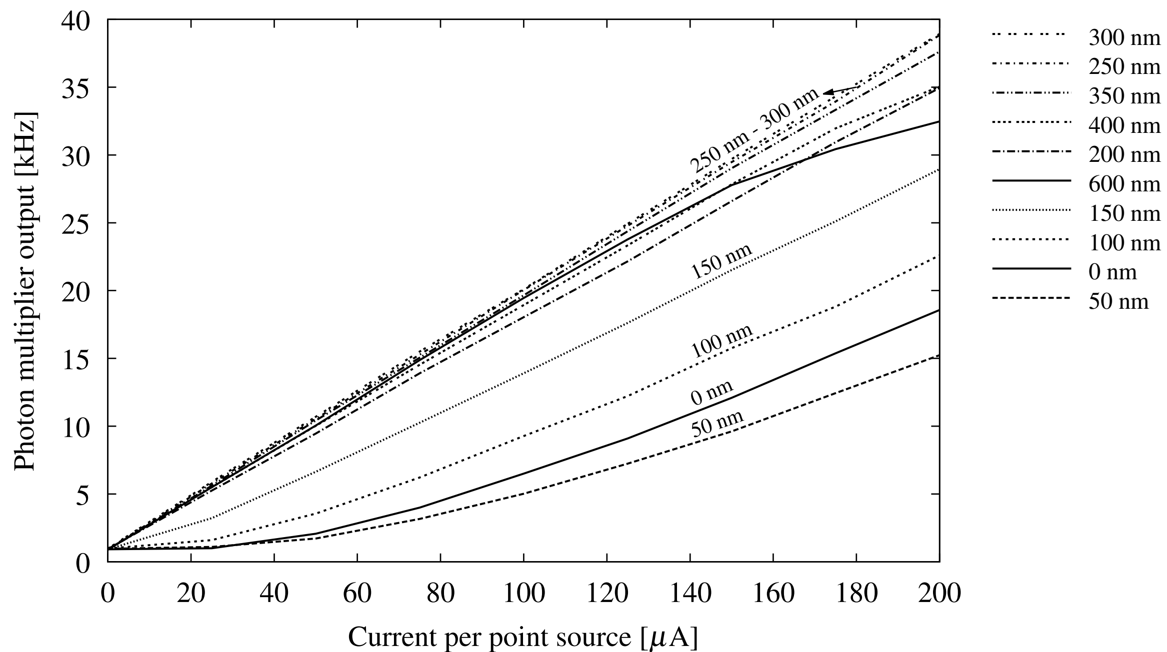


Figure 5.28: Average photon multiplier output for channel devices measured with forced current for different values of x_{sep}

shows a much more gradual curve than in figure 5.13 with a broader spread. The output rate per unit current is also shown as taken across the range from 0 – 200 μA which follows the same trend and it seems as if the peak light output is obtained at values for x_{sep} of 250 - 300 nm. The peak value of 38.9 kHz is less than the 50.8 kHz for the 325 nm source in figure 5.14.

The results for the arrays are just as promising, with figure 5.30 showing the radiometer detector current as a function of current applied to the light source arrays. There is a monotonic increase in the output optical power for each device as the current is increased. There is also a monotonic increase in light output as a function of x_{sep}. The radiometer current relationship with the input current is also linear for every device. Compared to the arrays implemented with the previous light sources in figure 5.15, the array performance in figure 5.30 shows a general trend towards improved emission intensity. To complete the comparison, it is necessary to include the input power in order to quantify the efficiency of the devices as discussed below.

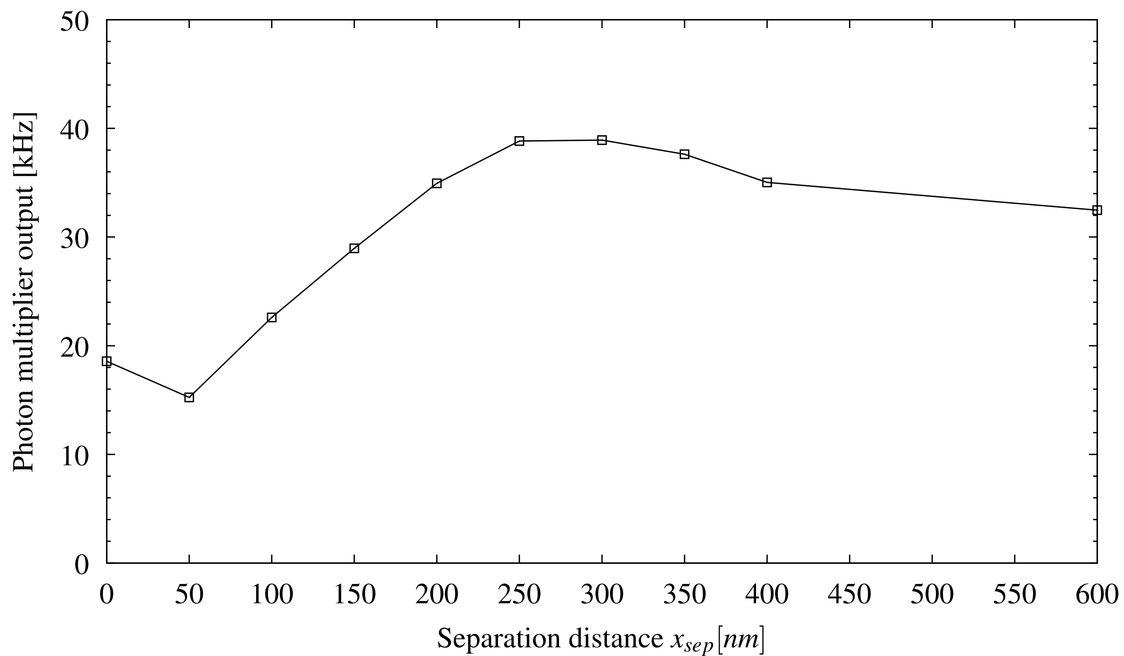


Figure 5.29: Peak PMT output and average count per unit current for channel sources at different values of x_{sep}

5.7.3 Efficiency

By taking both the input voltage and current into account, it is possible to construct figure 5.31 which shows that for $0 \leq x_{sep} \leq 150$ nm the efficiency improves as a function of input current, while for $200 \leq x_{sep} \leq 600$ nm the efficiency starts to drop. The most efficient devices are where $x_{sep} = 200 \pm 50$ nm. It follows that for devices operating in the Zener-breakdown regime, it is more efficient to operate at higher currents per point source, while for devices operating with avalanche generation as the main current contributor it is more efficient to operate at low currents per point source. Notice that at 150 nm the efficiency seems to remain the same irrespective of the drive current.

Figure 5.32 presents the efficiency of the arrayed devices as the radiometer detector current in nA per unit electrical power applied to the arrays. The legend is organised to represent the sequence of traces for different values of x_{sep} at the maximum driven current. The power efficiency of all the arrays is in the region of 10 nA/W which yields a substantial improvement when compared to the efficiency of arrays with the point source designs of section 5.4 as shown in figure 5.17. The improvement is on average a factor 2. With the exception of the 150 nm device, the devices perform marginally better at low currents. The droop, however, is not nearly as bad as for the previous point source designs

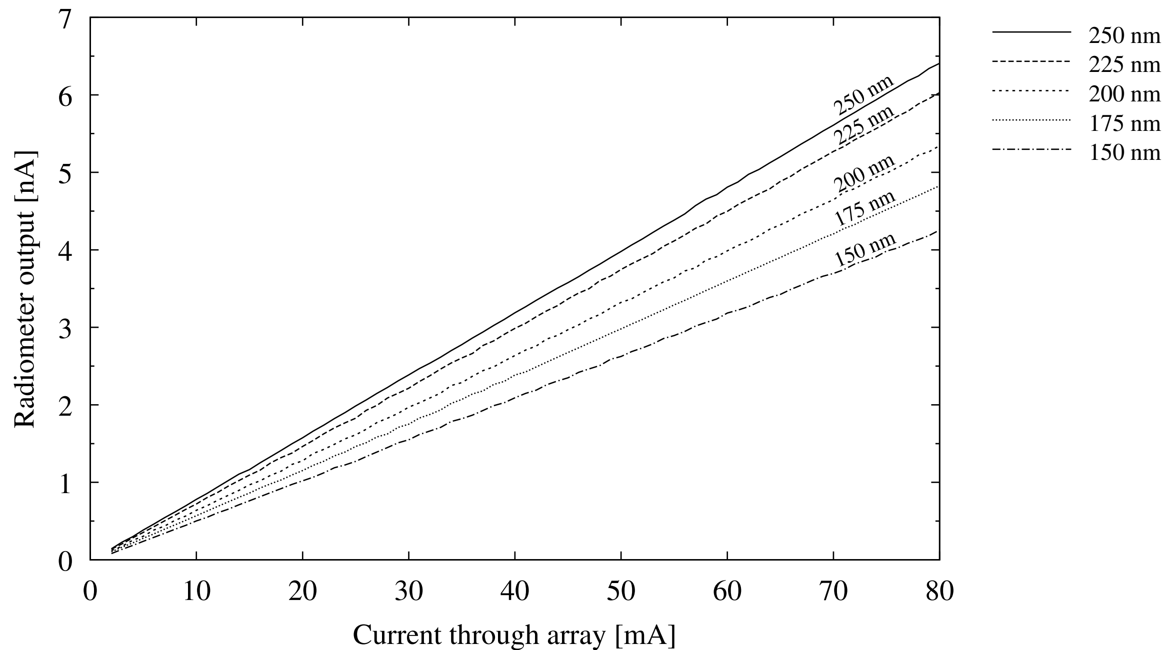


Figure 5.30: Channel-type arrays radiation intensity as a function of device current

from section 5.4. The fact that the efficiency curves of all devices tested lie very close to the 10 nA/W mark implies that it is possible to choose any device based on the voltage requirements of the specific application, with the trade off being that lower voltage devices will require more current for the same optical power output. Efficiency peaks where $x_{sep} = 200$ nm, while it seems futile to reduce $x_{sep} < 150$ nm. For $x_{sep} > 250$ nm the optical output flattens as the voltage continues to climb. This trend is reflected for the individual point source element power efficiencies shown in 5.31. A comparison to previous results is shown in figure 5.33 where the improvement factor is shown as a function of current. The reference array where $x_{sep} = 425$ nm is designed in the fashion of section 5.4 is located at $y = 1$, while for $175 \leq x_{sep} \leq 225$ nm the improvement is on average around 2 times the reference device efficiency and peaks at around 2.25 at high currents. This is a significant improvement in power efficiency when compared to the previous problematic point source design approach.

An interesting observation can be made around the optical emission intensity and the breakdown voltage of the junction. Figure 5.34 shows the detected radiometer current when driving the light source arrays at 100 μA per individual point source, or the entire array at 39.2 mA, when divided by the breakdown voltage. The reader is referred back to the definition of breakdown voltage as defined in section 5.7.1.1 and used to produce the results in figure 5.25 which shows the breakdown

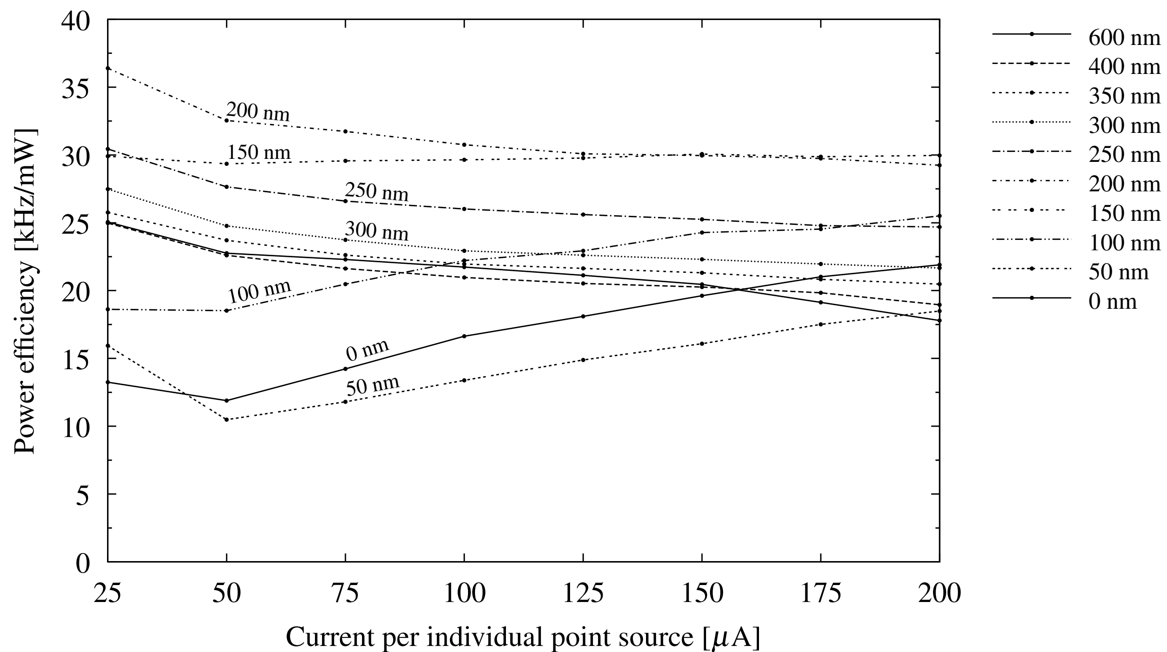


Figure 5.31: Efficiency of individual point source elements with field oxide removed between highly doped regions for different values of x_{sep}

voltage as a function of separation distance. The result in figure 5.34 is significant because of the correlation between the breakdown voltage and the emission intensity. From figure 5.34 it is seen that the detected radiometer current per breakdown voltage remains relatively constant, with an average value of 0.43 nA/V over the range of separation distances shown. There may be two possible reasons for the near linear correlation between breakdown voltage and emission intensity.

- The emission intensity depends on the time of carriers spent in the electric field during which radiative relaxations may occur, or
- the intensity drops as x_{sep} decreases due to the increasing amount of non-radiative tunnelling which occurs at narrower separation distances.

5.7.4 Summary of results

Elimination of the field oxide structure, which is subject to the process lithographic limits on the same mask, and the use of two other independent masks to demarcate the implantation type reduce sensitivity of V_{BD} on x_{sep} substantially. Although the layout violates layout rules tailored to transistors,

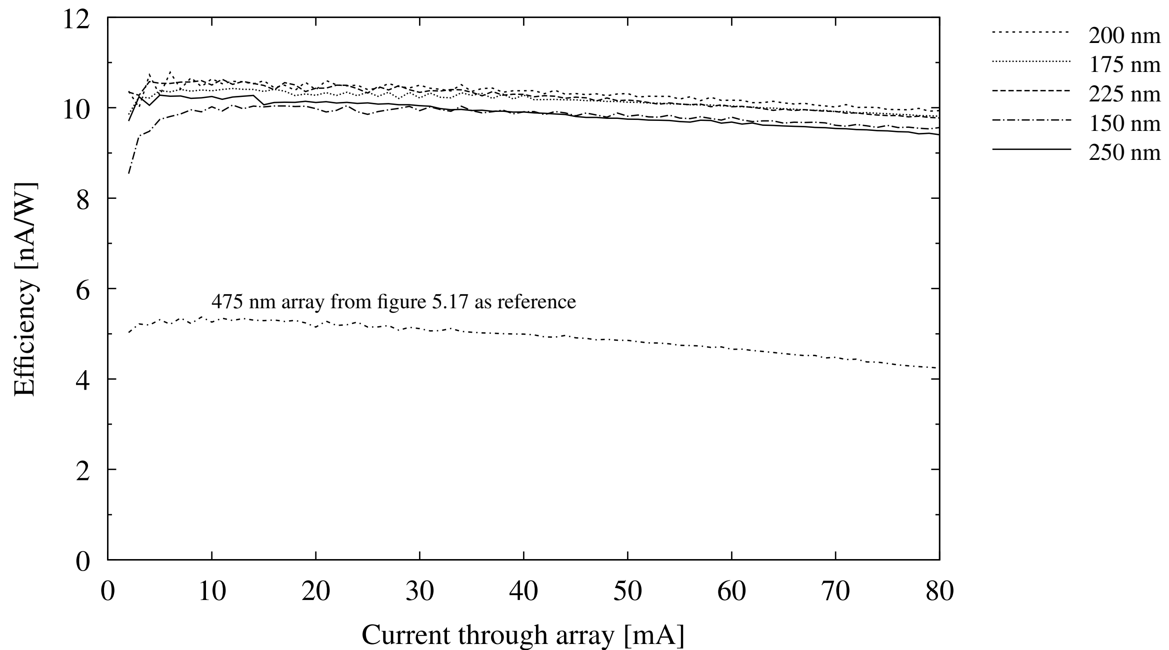


Figure 5.32: Efficiency of arrayed point source elements with field oxide channel-like junctions for different values of x_{sep}

it requires no changes to the CMOS process flow. This means that a circuit designer has the ability to implement this approach. The electrical results show much more repeatable behaviour with lower spread between samples, not only for the individual point source elements, but also for the arrayed structures. Furthermore, a strong correlation exists between $V_{BD}(x_{sep})$ for the individual point source elements and the arrayed structures.

Light generation now happens in a "duct" or channel formed by the Si/SiO_2 interface which means that more of the light spot is bounded by an interface aligned to the normal of the chip surface. This improves light extraction efficiency. To summarise, the following improvements can be seen when compared to the point source design introduced in section 5.4.

- Lower voltage operation is both possible and predictable following this design approach.
- The sensitivity of V_{BD} on x_{sep} is substantially reduced allowing a wider range of values for V_{BD} to be addressed within the limits of the process manufacturing grid.
- The characteristics of individual point sources and arrayed point sources agree well with the design being much more robust against process variations.

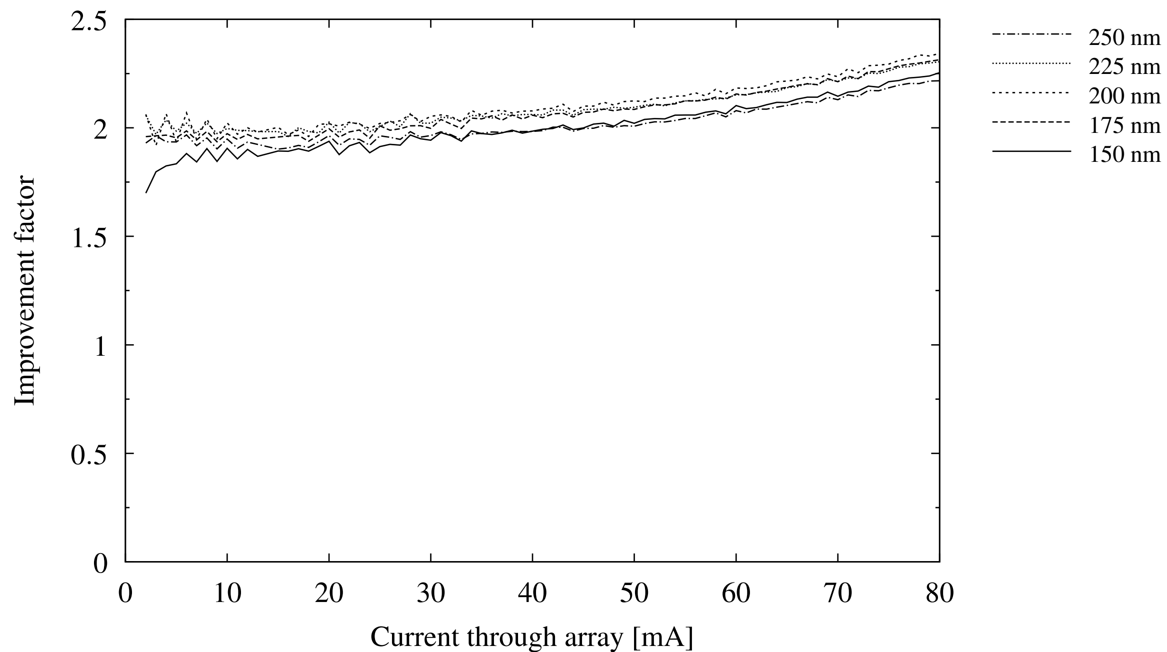


Figure 5.33: Field oxide channel source array improvement factor when compared to the 475 nm device from figure 5.17 used as reference

- From the logarithmic voltage and current relationships it can be seen that electric field reach through occurs, resulting in a reduced breakdown voltages as a function of decreasing values of x_{sep}.
- The failure rate for arrayed point sources is much lower than for the previous design and, although not the objective of this study, all of the experimental samples have been tested multiple times without failure.

5.8 CONCLUSION

The experimental results confirm that it is possible to reduce the breakdown voltage of hot carrier luminescent light sources in a standard CMOS process without modification to the process flow. A new layout technique, which is the subject of a patent filed in a number of countries (see chapter 1 section 1.10), has proven to be a reliable and predictable way of lowering the operating voltage of light sources using electric field reach through. This technique also yields an improved light extraction efficiency.

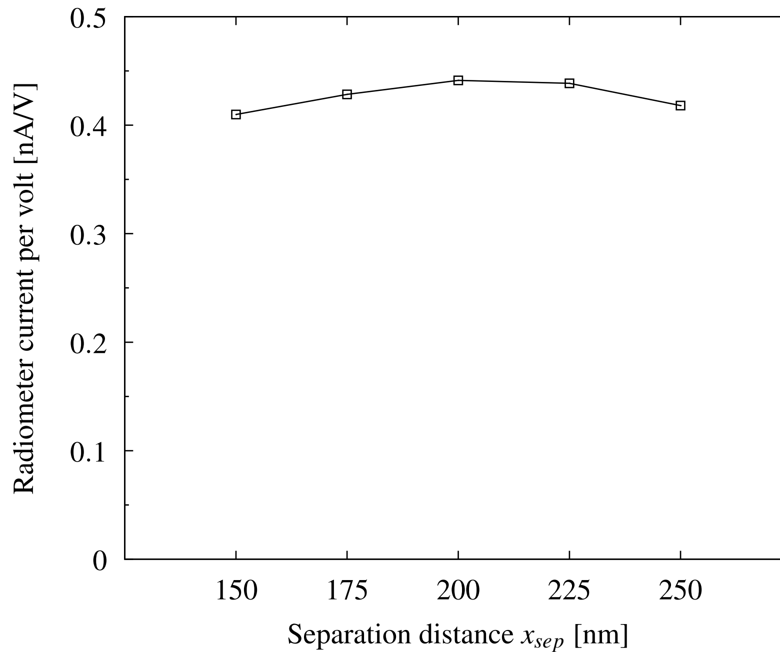


Figure 5.34: Detected radiometer current per volt breakdown voltage at an array current of 39.2 mA (100 μA per individual point source)

An interesting observation is that the power efficiency of the devices utilising electric field reach through remains almost constant, irrespective of x_{sep}, at least to the point where resistive losses start to become appreciable. The radiative flux, or optical output power, increases as x_{sep} increases as seen in figure 5.30. There are two possible explanations for this phenomenon, namely

1. that the number of radiative transitions depends on the time a carrier spends in an electric field and by terminating the field using a field stop, this time is decreased, or

2. that as the value of x_{sep} is reduced, the field emission transport mechanism starts to increase and the luminescent component starts to drop accordingly.

The field oxide technique developed in this chapter is the subject of an invention and the subject of a granted South African patent (RSA no. 2012/04957) and applications pending in China (no. 201180006581), Europe (no. 11705035.1), Japan (no. 2012-549458) and the United States of America (no. 13/574,333).

In conclusion, the light sources developed in this section present a means to enhance a circuit designer's ability to integrate these light sources based on the requirement of the specific application.

For proof of integration suitability, the developed light sources are utilised in the next chapter in a larger microdisplay with improved external power efficiency.

CHAPTER 6

A 128 X 96 PIXEL CMOS MICRODISPLAY USING IMPROVED LIGHT SOURCES

6.1 INTRODUCTION

One of the main objectives in lowering the voltage of silicon light sources based on hot carrier luminescence is that it requires voltages of sufficient levels to induce avalanche breakdown in junctions. The necessary voltage levels are sufficient to cause breakdown in transistor drain regions and may impact the reliability of the gate oxides over a long period. Integration of light sources based on avalanching junctions therefore presents a challenge due to the high operating voltages. An ideal scenario includes having the capability of generating the voltage required for avalanche on-chip. By reducing the operating voltage of silicon light sources based on hot carrier luminescence, it becomes possible to work towards a completely integrated solution, thereby reducing cost and form factor.

This chapter describes the integration of the light sources developed in chapter 5 using reach through techniques and field oxide manipulation. These improved light sources operate at lower voltages than the light sources used in the 64$\times$8 pixel microdisplay discussed in chapter 4. The work in this chapter extends on the principle of CMOS integration with even more CMOS circuitry coexisting with the light sources thereby proving that larger scale integration is possible. In addition, the newly developed light sources are more efficient due to improved light extraction efficiency. The rest of this chapter elaborates on the details of implementation as well as the characterisation and resulting performance of a new 128$\times$96 pixel microdisplay, which substantially broadens the application horizon in which a CMOS microdisplay may find use.

6.2 REQUIREMENTS

Chapter 4 describes a 64×8 pixel microdisplay based on state of the art silicon light sources developed by our group prior to the work described in this thesis. With the reduction in voltage and improvement in efficiency of the light sources introduced and developed in chapter 5 it was decided to improve not only on the pixel structure, but the complete microdisplay architecture.

6.2.1 Qualitative objectives

In order for a light source to serve as a successful CMOS component, integration with the workhorse electronic component, the MOS field effect transistor, is an absolute necessity. For successful integration with CMOS electronic circuitry the envisaged light source must exhibit a few qualitative properties as listed below. The light source must

- be described using only the design input mechanisms available to the CMOS circuit designer,
- connect directly to transistors,
- operate at a voltage which is not harmful to the driving circuitry,
- make use of the available diode structures, and
- be conducive to be packaged as a standard cell with known terminal characteristics.

6.2.2 Quantitative objectives

Table 6.1 breaks down the objectives aimed for in the larger microdisplay. The improved light sources will then become the active light emitting component in the emissive pixels.

Part of the underlying research question, namely that of the possibility of successful integration of a light source, will be answered by qualifying the results against the objectives set out in table 6.1.

Table 6.1: Objectives for the improved microdisplay

Input power P_e (flashlight mode)	$\leq$ 1.4 W
Supply voltage	$\leq$ 9 V
Aspect ratio	4:3
Resolution	128 x 96 pixels
Pixel pitch	25 μm
Refresh rate	$\geq$ 30 Hz

6.3 LIGHT SOURCE

Chapter 5 describes an approach for developing improved light sources when compared to a reference device similar to what was used in the previous microdisplay elaborated on in chapter 4. This section will provide clarity on which device was used based on results obtained in the previous chapter, as well as on the most appealing spectral characteristics which will be quantified in this chapter. The first step is to evaluate the light sources developed in chapter 5 to choose an appropriate design. This design is then compared to the light sources used in the previous 64×8 pixel microdisplay to validate whether the optical power emitted will be sufficiently improved. A polarity comparison is done to ensure that there is no major deviations from assumptions based on the n^+p-device which is invalidated by using a p^+n-type structure. Finally a design is proposed for implementation and integration into the new proposed 128×96 pixel microdisplay.

6.3.1 Spectral analysis

A very important part of the optical spectrum for microdisplays is the photopic region, centred around 555 nm. While the reach through light sources developed in chapter 5 yield an improved optical output at lower voltages, it is necessary to consider the spectral response of the light sources for its suitability in microdisplay applications. For this reason, an experimental setup was constructed where the spectrum was measured by a spectrometer covering wavelengths of 350 nm to 1100 nm, the limits being bound by the silicon detector used in the spectrometer, the grating and the fact that the spectral region covers the visible part of the electromagnetic spectrum.

The two crucial design aspects in the improved reach through sources were the removal of the field

oxide between the highly doped regions, as well as the separation distance between them. Figure 6.1 shows a calibrated power spectral density for an array constructed by using light sources at each of the designed separation distance values x_{sep}, all measured at 80 mA. Following the trend in the raw radiometer detector output, the reference device, constructed by pointed junctions separated by 600 nm has the lowest power content overall. Where x_{sep} is at a minimum of 150 nm, the peak at 430 nm is the largest. Note the correlation between the component at 430 nm and the number of states shown in figures 3.1 and 3.2 in chapter 3 which shows a large number of states available at 3.0 eV above the conduction band edge. The region between 500 nm and 650 nm sees an increase as a function of separation distance up to $x_{sep} = 225$ nm after which it starts to fall off again. In the red to near-infrared (NIR) region, the light source with $x_{sep} = 250$ nm dominates with $x_{sep} = 225$ nm also exhibiting a lot of spectral content, while the performance for the remainder of the devices is similar.

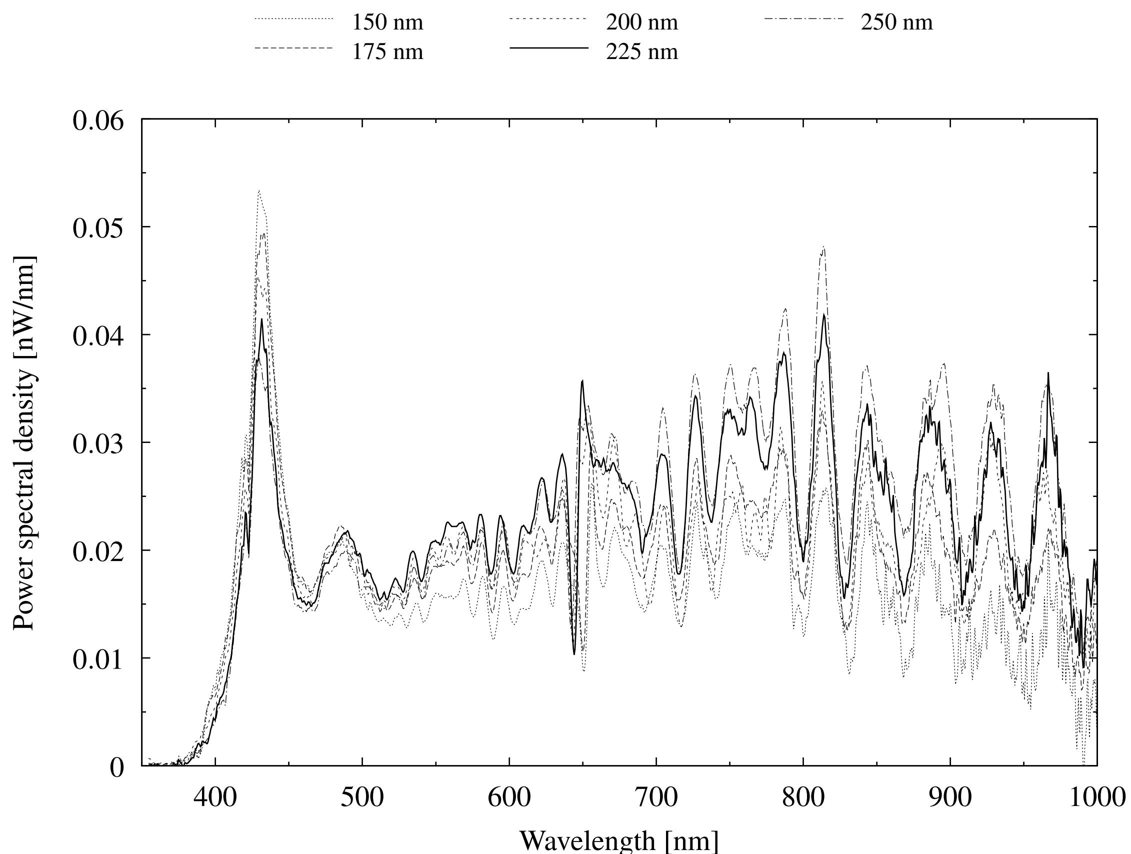


Figure 6.1: Power corrected spectral density of reach through light sources developed in chapter 5 for n^+p-type junctions at 80 mA drive current

By applying the photopic weight function, the power available for observation by the human eye is

extracted from these plots. As seen above, the 225 nm source seems to contain the most power in this region, with figure 6.2 showing the total integrated power over the entire measured spectrum as well as the integrated power weighed by the photopic function. The observation above holds true as the integrated photopic power peaks with the 225 nm device. It was decided to implement the 225 nm device in the optical array utilised in each pixel based on these results and the results from figure 6.9 which yielded very similar results for both n^+p- and p^+n-type light sources both in terms of optical output power and spectral content.

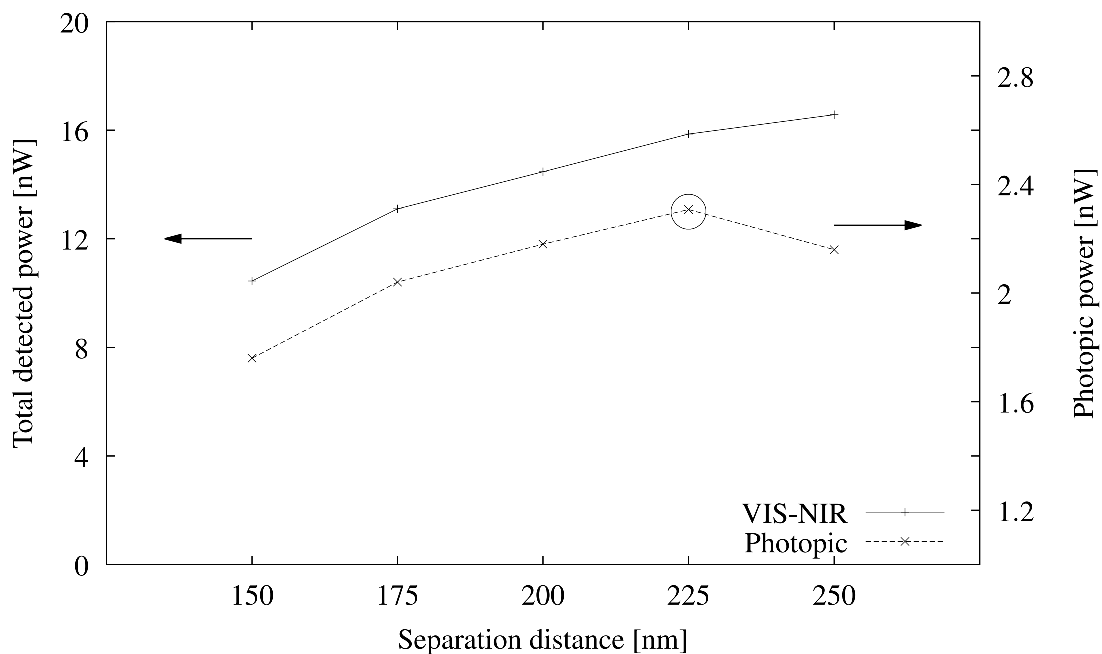


Figure 6.2: Integrated detected power and photopic weighted power as a function of separation distance for n^+p-type junctions

6.3.2 Point source improvement

One of the problems associated with the previous design [61] on which chapter 4 is based was that the operating voltages are very high with breakdown voltages in the region of 9 V, which results in an increased leakage current in the driving transistors while in the off state. The improved light sources employ a technique which lowers the voltage while increasing the optical output power for the same device current. This is a major advantage in the integration of the light sources with other circuitry on the CMOS chip.

Based on the results in developing an improved light source structure operating at lower voltages in

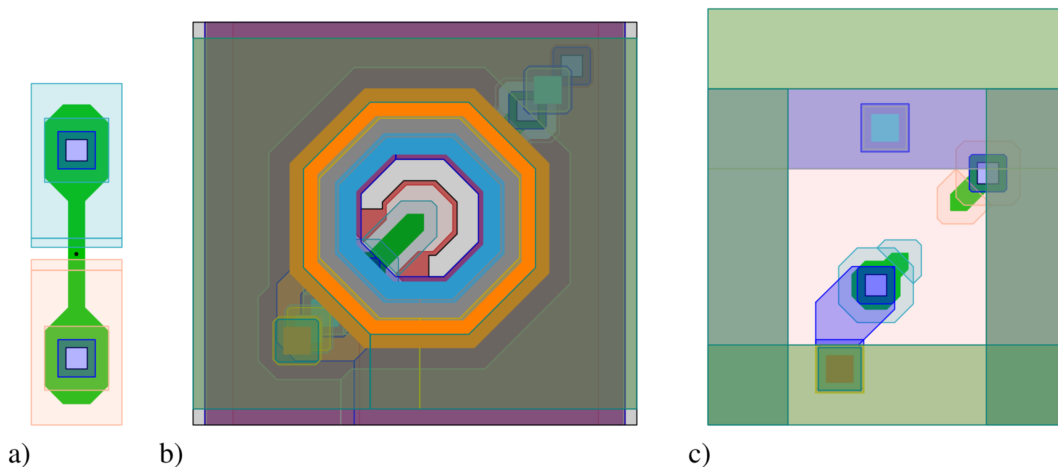


Figure 6.3: Layout of the light sources for comparing the a) RT 225 nm n^+p, b) 64×8 pixel p^+n and c) an avalanching p^+n pointed light source without reflecting structures serving as reference

chapter 5 it is possible to predict the expected improvement that the current display will show over the display employing the pointed sources. As a benchmark, this section compares the voltage vs current and radiometric output vs current between the three devices shown in figure 6.3 implemented as arrays with differing number of individual elements. The figure includes the following devices, where figure

- 6.3a shows the layout of a 225 nm n^+p-device in reach through (referred to as RT 225 nm) developed in the previous chapter, with the specific choice of separation distance based on considerations shown later in this chapter,

- 6.3b shows an individual light source from a pixel of the 64×8 pixel microdisplay with an integrated light directing structure [50], and

- 6.3c is a p^+n-type avalanching junction, shaped as a pointed source without any light directing structures which serves as a reference device for benchmarking.

The light sources were designed as arrays with element counts summarised in the table below. As the array sizes differ, it is necessary to follow two comparative approaches: 1) comparing the macro-characteristics of the entire array as a function of the terminal signals, and 2) comparing the individual element performance inside each array.

Light source	Layout reference	Element count
225 nm RT n^+p	Figure 6.3a	392
64×8 pixel p^+n	Figure 6.3b	4032
Reference p^+n	Figure 6.3c	400

6.3.2.1 Comparison of element array performance

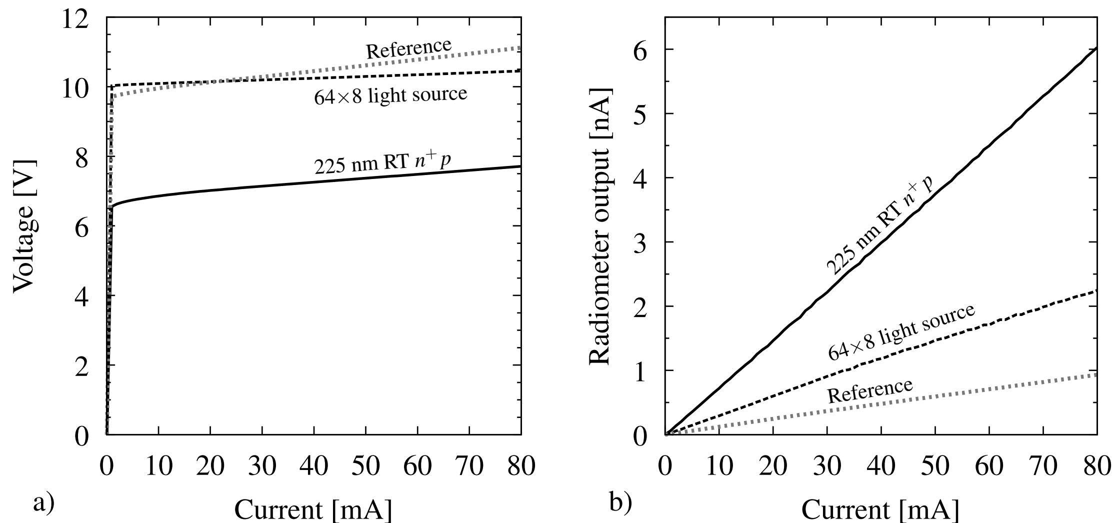


Figure 6.4: The voltage and radiometer output of the devices used in the original microdisplay and reference junction compared to the improved light sources

Figure 6.4a compares the device voltages of three arrays implemented with the light sources shown in figure 6.3. As expected, the 225 nm RT device breaks down at a much lower voltage of around 6.5 V compared to the pixel and reference light sources. The reduction of around 3.5 V has a profound impact on the wall plug efficiency of the light source. It is also clear that the series resistance component of the 225 nm RT and reference light sources are higher than the array with a high element count which explains the differences in slope.

Using the same independent parameter viz forced current, figure 6.4 compares the radiometer detector current generated by the three respective devices. A linear fit produces the results shown in table 6.2 and shows that the device developed in chapter 5 offers the best performance, with an improvement factor of 6.35 above the reference device.

Efficiency can be measured by relating the optical output power as detected by the radiometer to the

Table 6.2: Improvement factors of optical power emitted by arrays of light sources in figure 6.3

	Reference	64×8 light source	225 nm RT n^+p
Reference	1	2.44	6.35
64×8 light source	0.41	1	2.60
225 nm RT n^+p	0.16	0.39	1

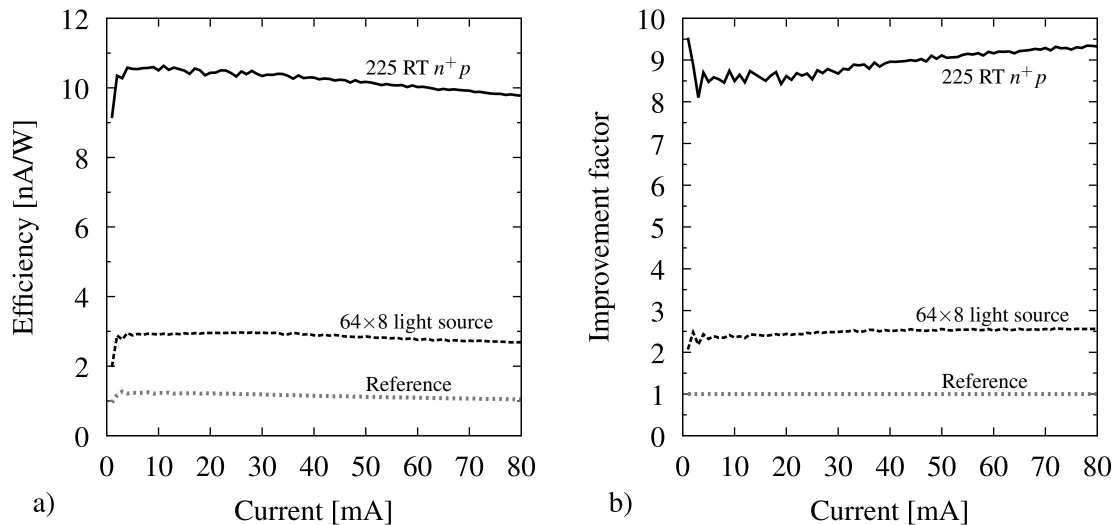


Figure 6.5: Power efficiency and improvement factor comparison

input electrical power of the devices. Since the output of the radiometer detector is a current, it is valid to compare this metric as a function of input power as long as the spectrum of the light sources does not differ substantially. Figure 6.5a shows the efficiency, the ratio of radiometer detector current over input electrical power, of the three light sources as a function of input current. The average efficiency of the light sources is shown in table 6.3.

A key figure of merit is the efficiency improvement factor when comparing the different light sources.

Table 6.3: Average efficiency of light source arrays

225 RT n^+p	10.21 nA/W
64×8 light source	2.85 nA/W
Reference	1.15 nA/W

Table 6.4: Improvement factors of array efficiency of figure 6.3 light sources

	Reference	64×8 light source	225 nm RT n^+p
Reference	1	2.48	8.91
64×8 light source	0.40	1	3.59
225 nm RT n^+p	0.11	0.28	1

Table 6.4 summarises the efficiency improvement of the light sources under review. The combination of improved emitted optical power and a reduction in supply voltage gives the 225 nm RT light source a major advantage compared to the other light sources, with an average improvement factor in efficiency of 8.91.

6.3.2.2 Comparison of individual element performance

Comparing arrays of different size may yield results which is interpreted out of context. In order to improve the comparative platform for comparing the light sources, the element count was taken into account and the results were translated to a performance metric per element inside the array. This allows a fair comparison between individual light source elements.

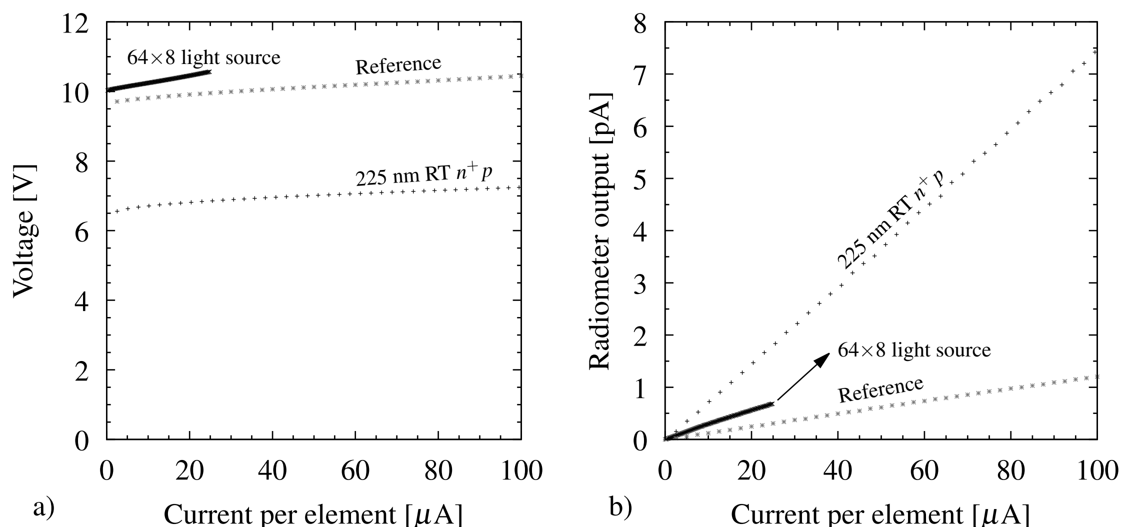


Figure 6.6: Comparison of current per each element in the array of light sources

As with the array measurements, the voltage and radiometer detector current were examined as a function of input current, but in this case, current per individual light source element. Figure 6.6a

shows the voltage characteristics of the three light source elements as a function of current per element. The series resistance for the arrays with low element counts is now much less pronounced. Unfortunately, the range over which the large array could be measured is very small. There are no discrepancies between the element and the array results.

Figure 6.6b shows the current generated by the radiometer detector as a function of input current per element of each light source. The trend remains similar to the array results and is summarised in table 6.5.

Table 6.5: Improvement factors of optical power emitted by light sources elements of figure 6.3

	Reference	64×8 light source	225 nm RT n^+p
Reference	1	2.42	6.40
64×8 light source	0.41	1	2.65
225 nm RT n^+p	0.16	0.38	1

The efficiency of each individual light source can also be quantified in the same way as for the arrays. Figure 6.7a shows the efficiency of each element. The efficiency is again defined as the radiometer detector current that will be generated as a function of input electrical power into the light source. The results correspond almost exactly with those of the arrayed devices, with the average efficiencies shown in table 6.6.

Table 6.6: Average efficiency of single element light sources

225 RT n^+p	10.21 nA/W
64×8 light source	2.80 nA/W
Reference	1.12 nA/W

The improvement factors are shown as a plot in figure 6.7b and are summarised in table 6.7. This is an important result because it enables a proper selection of light source for integration into the larger microdisplay.

Based on the results above, the choice of the 225 nm reach through n^+p source seems like an ideal candidate for a larger resolution display with a reduced supply voltage. Although the polarity of the device is not suitable for integration using the CMOS process in this work, it is expected that the p^+n-type light source will yield similar results.

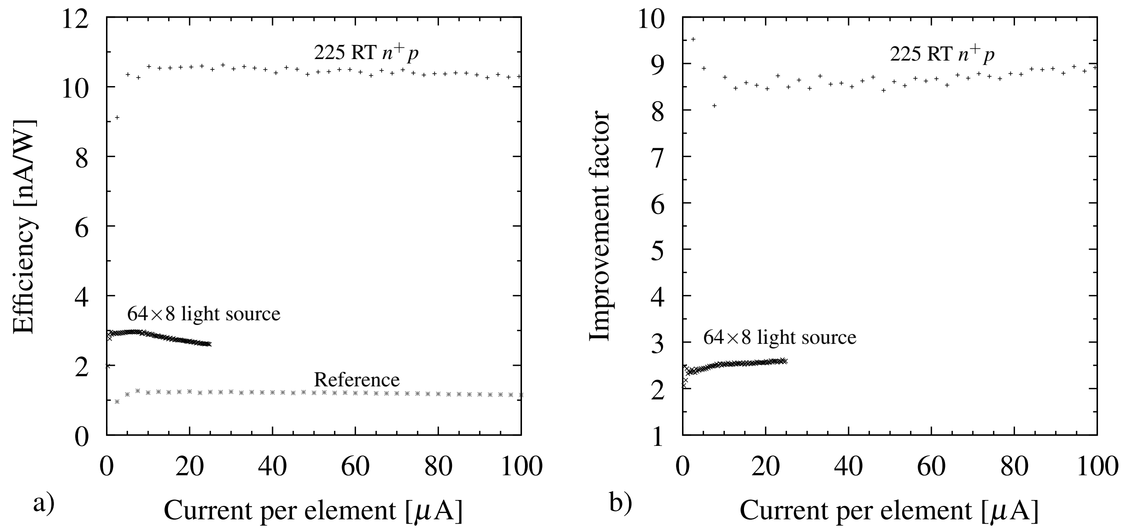


Figure 6.7: Comparison of current per each element of array

Table 6.7: Improvement factors of individual element efficiencies of figure 6.3 light sources

	Reference	64×8 light source	225 nm RT n^+p
Reference	1	2.50	9.10
64×8 light source	0.40	1	3.64
225 nm RT n^+p	0.11	0.27	1

6.3.3 Point source design

The light sources used for the improved microdisplay are based on the design given in section 5.6.2 where the field oxide is absent between the two highly doped regions. Figure 5.21 is restated below in figure 6.8 which shows the conceptual layout of the light sources used in the microdisplay pixel array. The light sources are adapted for use in the design of the pixel as explained in the sections below.

6.3.4 Polarity comparison

The design in figure 6.8 shows a light source designed with a diode polarity based on a n^+p-junction. This design does not allow direct integration with CMOS logic circuitry because the anode shares a low impedance connection to the same substrate as the logic transistors. The only way the light

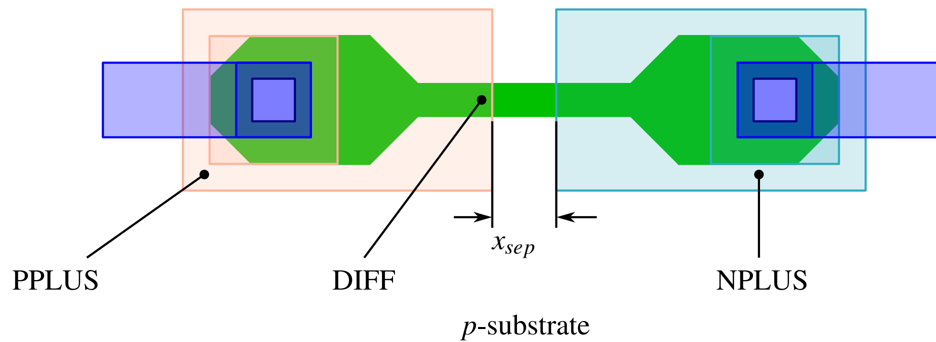


Figure 6.8: Layout of single point source with oxide channel formed by LOCOS or STI

source can be energised is by pulling the cathode high enough for breakdown, with the anode tied to ground. This problem has been discussed in section 4.4.3. The preferred way of energising the pixel light sources is by using a NMOS-based driver circuit pulling the anode voltage to ground, with the cathode remaining fixed at a higher potential. For this reason, it is worth investigating the effect of reversing the diode polarity.

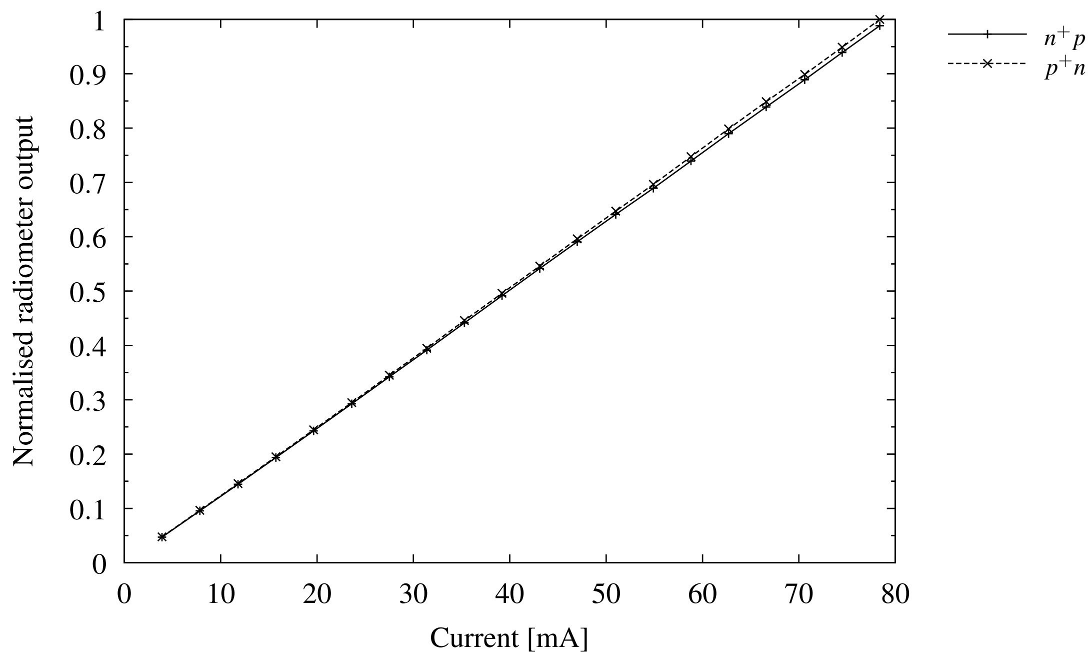


Figure 6.9: A comparison between averaged n^+p- and p^+n-junctions radiometer detector current

Figure 6.9 shows a normalised plot of the average optical output versus input current for n^+p- and p^+n-junction light sources. There is no appreciable difference between the output characteristics of the two junctions. The average difference is less than 1 %. The significance of the results lies in the

fact that both polarities essentially perform the same. This allows the use of the $p^{+}n$-type light source with a floating anode, which makes it possible to energise the light source by pulling the anode low. It is therefore the preferred approach for integrated pixel designs.

6.4 INTEGRATED PIXEL DESIGN

6.4.1 Light source array

A point source as individual element will be unable to accommodate enough current to be used in isolation. For this reason it is necessary to use point sources in an arrayed fashion. Moreover, figure 5.32 shows that it is almost invariably more efficient to drive the light sources at lower currents, with more elements in parallel, than to drive fewer elements at higher current values. The only exception is when the arrays are operating where band-to-band tunnelling contributes a substantial proportion of the total current. With this in mind, the maximum amount of elements were fitted into a pixel, with sufficient metal implemented to support conduction without approaching the current density limitations of the metal layers. A total of 63 elements were incorporated into the active light emission part of the array.

6.4.2 Driving circuitry

The previous microdisplay architecture was designed to switch the 8 rows directly using high voltage lines. The drawback with this approach is that it requires external pass transistors to act as switches for energising the row-addressed pixels, while the columns were addressed using CMOS logic. The current design improves on the previous approach by requiring a single high voltage input feeding the whole active matrix at a single time, while both the row and column selectors can be manipulated through the use of CMOS buffer drivers.

Figure 6.10 shows the implementation of this approach by utilising two NMOSFETs in series acting as the switches for determining the potential of the light source anode. The light source array shares a common cathode connection to the high voltage supply. M_{n1} and M_{n2} are two NMOS transistors connected in series, where the gate of M_{n1} is connected to a logic signal C_{sel} selecting the column to be energised, while M_{n2} has a gate connection to the signal selecting the row to be energised, R_{sel}. Only when both signals are high will the transistors conduct current.

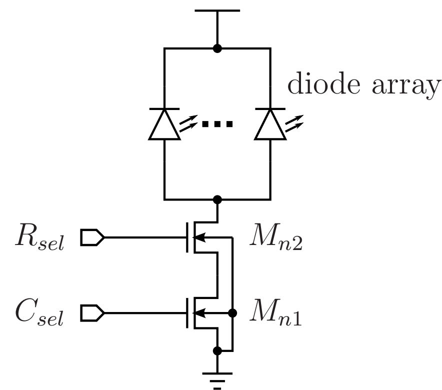


Figure 6.10: Circuit diagram of on-chip pixel with row and column selectors

M_{n1} and M_{n2} are sized to allow operation in the triode region, which will still create a strong voltage dependency of the current based on the high voltage supply line. Transistor scaling was set to accommodate the maximum allowable current, based on the power constraint given in table 6.1. Both of the transistors are scaled to the same W/L ratio in order to facilitate layout of the pixel array. M_{n2} will have a reduced transconductance due to the body effect of the raised source potential, which has been taken into account.

$$\left(\frac{W}{L}\right)_{M_{n1}} = \left(\frac{W}{L}\right)_{M_{n2}} = \frac{10}{0.35} \tag{6.1}$$

The power constraint is shown in figure 6.11 along with a SPICE simulation of the current-voltage characteristics of the cathode terminal with reference to ground, which represents the complete V_{sup} value for the entire on-chip pixel. The IV-curves include the behaviour of both the light source as well as the driving transistors. The region to the left of the $P_e = 1.4$ W trace is the designed operating region when driving the display in flashlight mode, while maintaining the power consumption of the overall optical matrix with all pixels driven at maximum duty cycle within the intended limits. When displaying video, the average power consumption will be much lower, allowing overdrive of the display up to the point where the transistors saturate and operate in a constant current regime. For the typical mean (TM) corner, this is between 3.0 mA and 3.5 mA. This approach, with the transistors operating in triode, will allow for near-linear overall brightness control.

6.4.3 Layout

Each pixel was designed with a pitch of 25 μm in mind, where the design is constructed to allow pixels to be exactly abutting each other. The high voltage and ground supplies are fed to each pixel

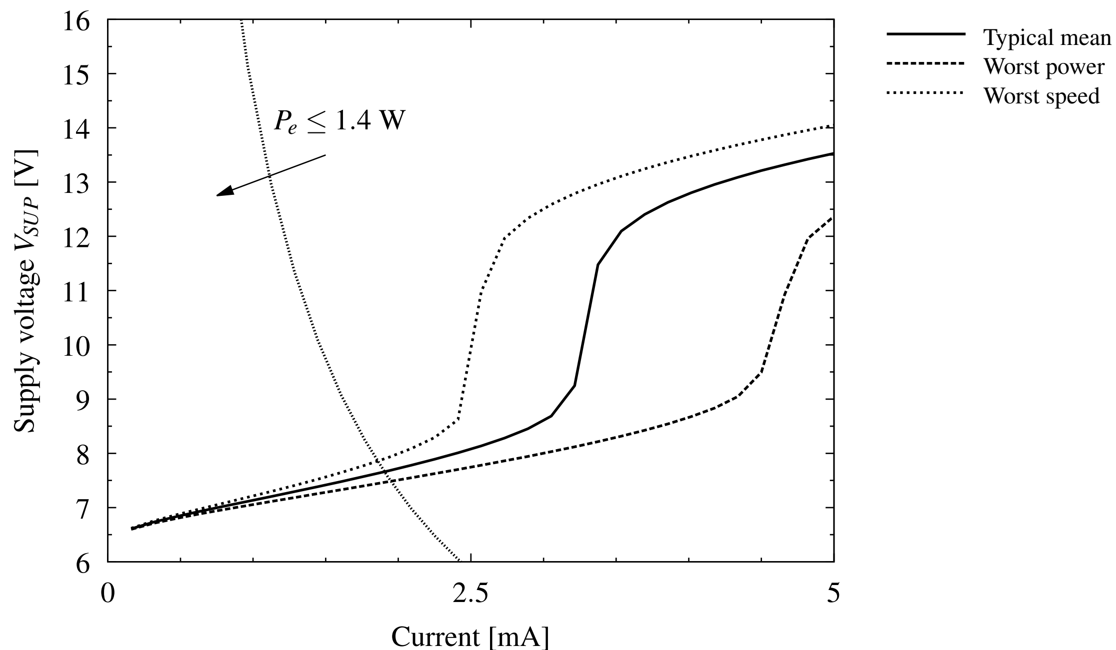


Figure 6.11: Simulated electrical characteristics of one pixel when energised with R_{sel} and C_{sel} high

in a meshed fashion to decrease resistance and support the current requirement of the array. The row address lines are shared for all 128 pixels in a row, while the same holds true for the 96 pixels in a column. Only when both the row and column signals R_{sel} and C_{sel} respectively are logic high, the pixel will emit light. Figure 6.12 shows the layout of the pixel with the driving transistors and light emitting area as indicated.

6.5 MICRODISPLAY ARCHITECTURE

This work focuses on the integration of a light source in a pixel structure which is directly compatible to be interfaced with other electronic circuitry on-chip. Although the digital circuitry requiring successful operation of the microdisplay is beyond the detailed scope of this work, it is important to mention the methodology used to drive the microdisplay. This section will review the methods chosen for pixel luminous intensity control as these directly impact the way the active emission matrix is controlled. The scanning architecture is also subject to the nature of the light source and pixel behaviour. Internal and external control is elaborated on in order to showcase the capability of the light source as part of a functional CMOS microdisplay and proves with substantial evidence that the light source can indeed form a drop-in part of a CMOS standard cell library.

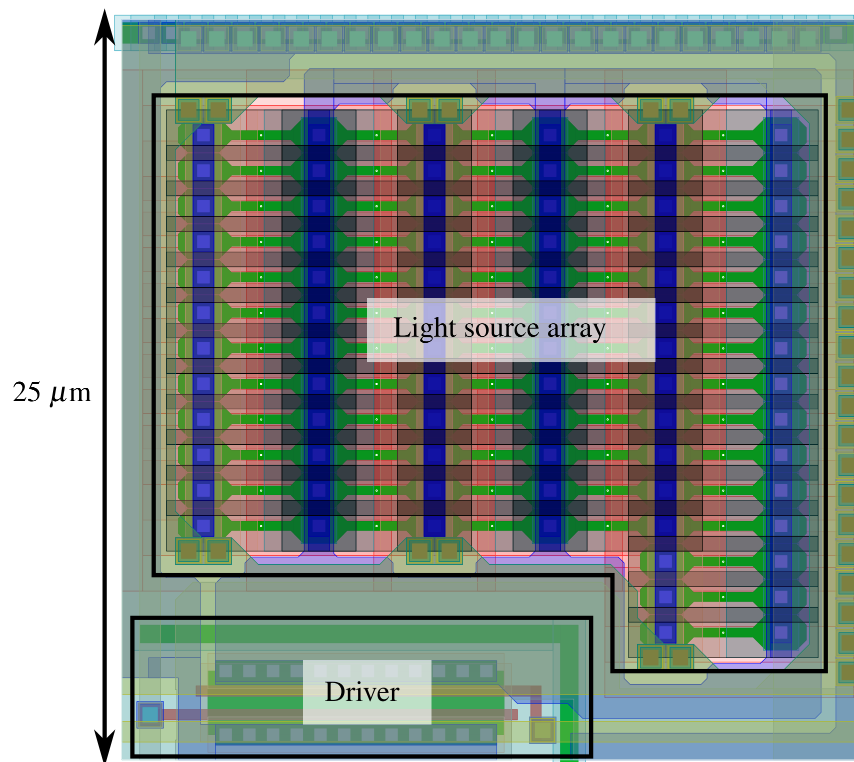


Figure 6.12: Single pixel layout utilising 63 reach through light sources with oxide channels in parallel

6.5.1 Luminous intensity control

There are two ways in which the luminous intensity of a pixel can be varied, viz.

1. changing the current through the diode array in reverse breakdown, and

2. using a fixed "on" current but varying the duty cycle of the energising signals applied to the pixel's diode array.

Chen [58] uses a complex circuit involving digital-to-analogue conversion of a digital word to a current which is supplied to the diode array in avalanche. It was deemed more appropriate in this work to utilise a pulse-width modulation scheme applied to the row select signal R_{sel} in figure 6.10 as it provided a much more stable intensity control without the need to take analogue mismatches in the driving circuitry into account. The timing reference controlling the digital driving circuitry can then be used to accurately control the row select signal duty cycle for each column being energised.

Since the pixel driver and light source combination voltage-current relationship is shown in figure 6.11 with the intended region of operation where both transistors operate in the triode region, a near-linear intensity control, controlling the “on” current of the pixel, can also be varied according to the supply voltage within this region supplied to the optically active matrix. This allows display-wide control over the entire array.

6.5.2 Scanning architecture for pixel matrix

Unlike liquid crystal displays (LCDs) or liquid crystal on silicon (LCoS), light sources based on hot carrier luminescence creates a display of an emissive nature. A pixel will only emit light if energised. Ideally, given the results in chapter 5 figure 5.32, a low current should be used to make use of increased efficiencies at low currents. This would mean that it would be more efficient to address each individual pixel and energise each one continuously. This presents a few problems, namely

- the routing and light emission area of a pixel share the same area, and
- the controlling circuitry will have to increase in complexity.

6.5.2.1 Single pixel scanning

Cathode ray tubes use a single electron beam projected onto a fluorescent screen onto a spot which will emit light. This effectively means that only one pixel is energised at any given time. The electron beam is swept horizontally and vertically across the screen which, along with the fluorescent decay, relies on the human eye for an averaging effect. If done fast and bright enough, the result creates an impression that the entire screen is illuminated.

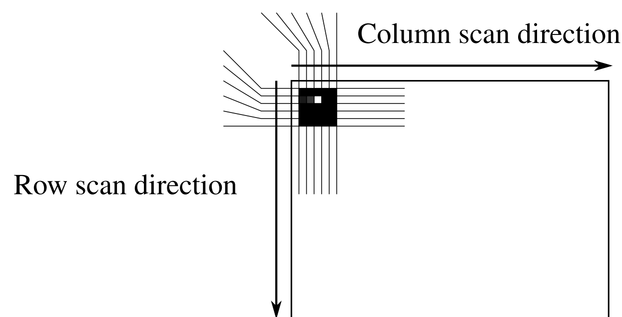


Figure 6.13: Single pixel scanning addressing one pixel at a time

The same method can be used for energising pixels in a microdisplay if each pixel described in section 6.4 is abutted and shares the column and row address lines per column and per row respectively. Figure 6.13 represents this methodology, with a single pixel being addressed at a time, while completely scanning through all columns before advancing to the next row. This technique requires the least amount of logic circuitry as a single pulse-width modulator can be employed together with a decoder-type addressing circuit of a sequential nature. The topology above is not suitable for the integrated light source as the brightness cannot be arbitrarily increased and the light source does not have a decay as in the case of a fluorescent screen. For an input power of $P_e = 1.4$ W, this method will require an instantaneous application of almost 175 mA for a single pixel, albeit for a very short duration. Another limitation lies in the extreme frequency of the scanning clock. For PWM, the clock needs to be further subdivided into a granularity supporting the levels of grey scale supported by the display.

6.5.2.2 Segmented scanning

Segmenting the display area in an attempt to approach the ideal situation is another possibility. If the architecture of the single pixel method above is used while the display is segmented into four parts, it is possible to address four pixels at a time. The limitation lies in access to the row and column lines. Since the display area is rectangular, there are four sides available for row and column selector lines. If these lines are severed as shown in figure 6.14, it is possible to simultaneously address a pixel in each of the four segments without interrupting neighbouring segments. This will reduce the current requirement for each pixel by a factor of four. This value is still above the safe operating current per pixel.

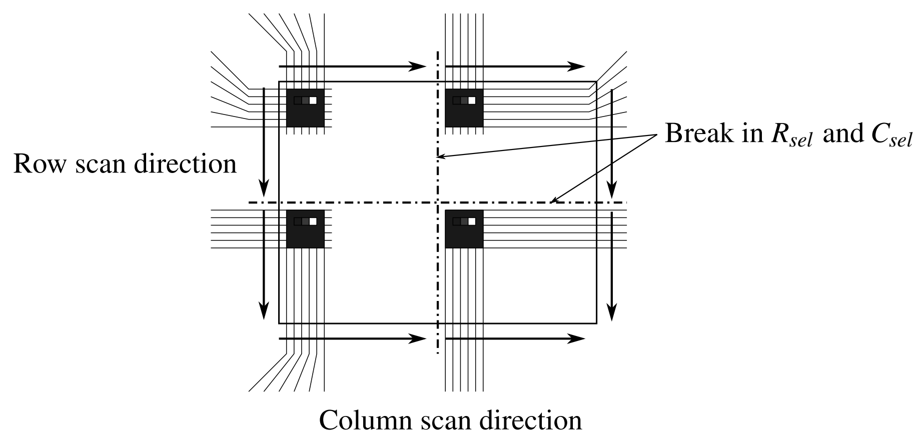


Figure 6.14: A segmented version of the single pixel addressing method

6.5.2.3 Column scanning

Even with the improvement above, the requirements on the clock and pixel current capability remain high. In the microdisplay described in chapter 4 a row parallel method was used to address the pixels. All of the rows are simultaneously addressed, while a column selector scans through each column line individually. The signals on the row select lines are then synchronised with the column scanning in order to produce the desired image.

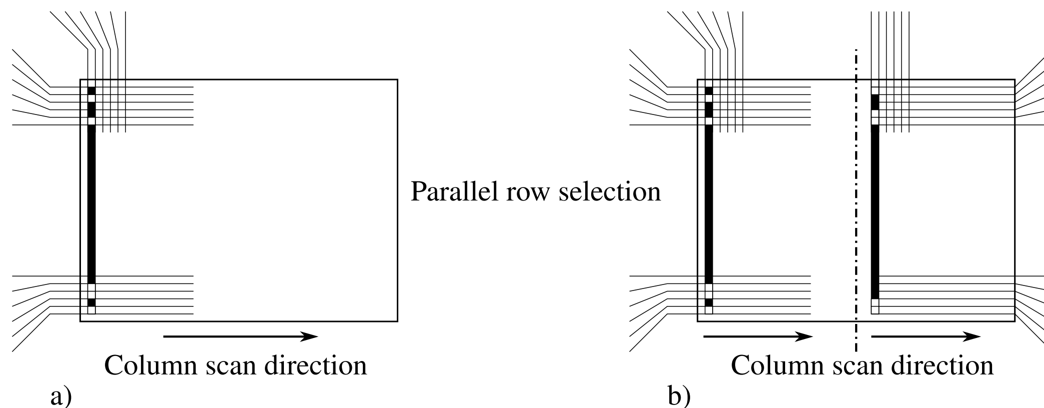


Figure 6.15: Row parallel scanning technique for addressing pixels

Figure 6.15 shows the concept, with figure 6.15a showing the method while addressing rows only from one side of the matrix, and figure 6.15b showing a hybridised approach between row parallel scanning and the segmented technique from above. By adopting this method, it is possible to create a display much closer to the idealised situation of addressing all pixels simultaneously

6.5.2.4 Summary and preferred approach

Table 6.8: Summary of scanning approaches

	Single pixel	Segmented	**Row parallel**	Row parallel segmented
Clock frequency	188.7 MHz	47.19 MHz	**1.966 MHz**	983 kHz
Peak pixel current	185.28 mA	46.32 mA	**1.93 mA**	965 μA
Circuit complexity	low	medium	**medium**	high

Table 6.8 shows a summary of the requirements for the different scanning approaches, given flash light operation of a 128×96 pixel display operating at $P_e = 1.4$ W on a typical mean curve according

to figure 6.11. Although it is possible to support clock speeds of $f_c \geq 188$ MHz on a 0.35 μm CMOS process, the design becomes extremely difficult. The pixel current requirement is also large and although the effective value for I_{rms} remains the same for each of the choices above, failure is more likely for high peak values. The approach chosen in this work is the best compromise between circuit complexity and peak pixel current, namely the row parallel method.

6.5.3 Integrated controller

Although the scope of the research reported on in this thesis does not include the controller logic design itself, the circuitry with which the optical display matrix interacts is fundamental to the operation of the microdisplay. The nature of an emissive display is such that light can only be emitted if the pixel is energised. As mentioned above, ideally all pixels in the active area matrix should be powered the whole time, with the specific brightness for each pixel controlled continuously as a function of the image it is intended to display. Of course, this approach requires dedicated access to each individual pixel which becomes impractical. Since a CMOS microdisplay needs to share area between the light source, driver and electrical connections, a practical solution is to scan the active area at a specific rate, while leaving it to the eye to do the time-averaging.

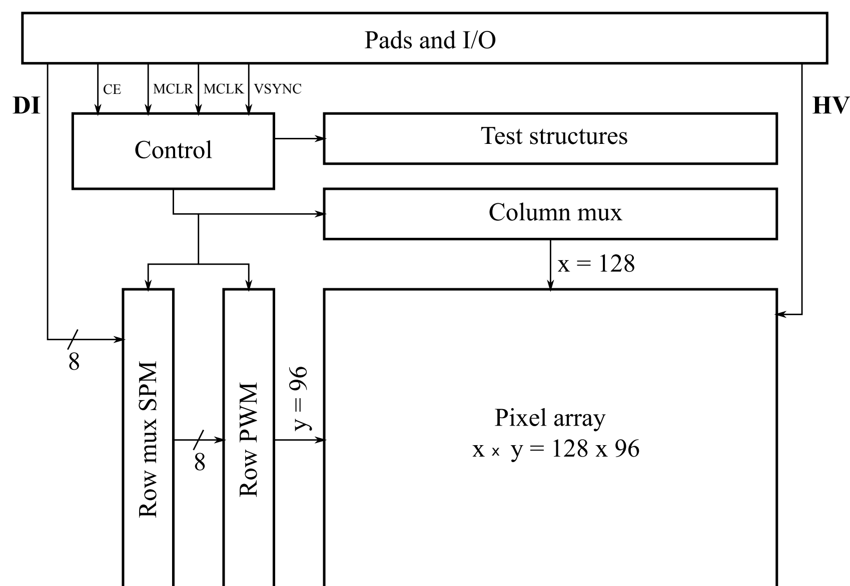


Figure 6.16: Block diagram of on-chip internal controller

Hot carrier luminescence seen from avalanching junctions is known to be extremely fast [44], which

means that the limit to the rate at which the pixels can be toggled is limited by other factors such as the electronics and architecture. A refresh rate of 60 Hz was set as a design specification while the screen is repeatedly scanned, allowing the eye to perceive the image across the whole display as smooth and uninterrupted. As mentioned above, the pixels require a scanning methodology to be accessed. Unlike a cathode ray tube where a single beam is swept across the entire screen multiple times per second, this design allows for a row parallel architecture. In this way, the columns are selected and swept linearly from one side to the other, while the rows can be energised selectively in parallel. For the 128 x 96 pixel array implemented in this work and a refresh rate of 60 Hz, each pixel needs to be addressed for a time duration of 130.2 μs. Brightness is controlled through an 8-bit pulse width modulation (PWM) scheme, resulting in 255 levels of brightness. A block diagram of the digital driver circuitry is shown in figure 6.16. The control block globally directs the timing of all scanning functions, as well as other design features. Sweeping across the columns with the column multiplexer, the 8-bit digital row data for each column is read in sequentially from an external source, loaded and then displayed in parallel. The serial-to-parallel memory (SPM) in the "Row mux SPM"-block converts the serial input data on the data input (DI) pads to parallel data, and loads it into a memory register. While the current data is transformed to PWM signals that drive a column in the display in parallel, the sequential data for the next column is entered into the SPM. The serially communicated bitmap data for the microdisplay is transposed from the standard row-wise sweep of columns, and requires external manipulation of images.

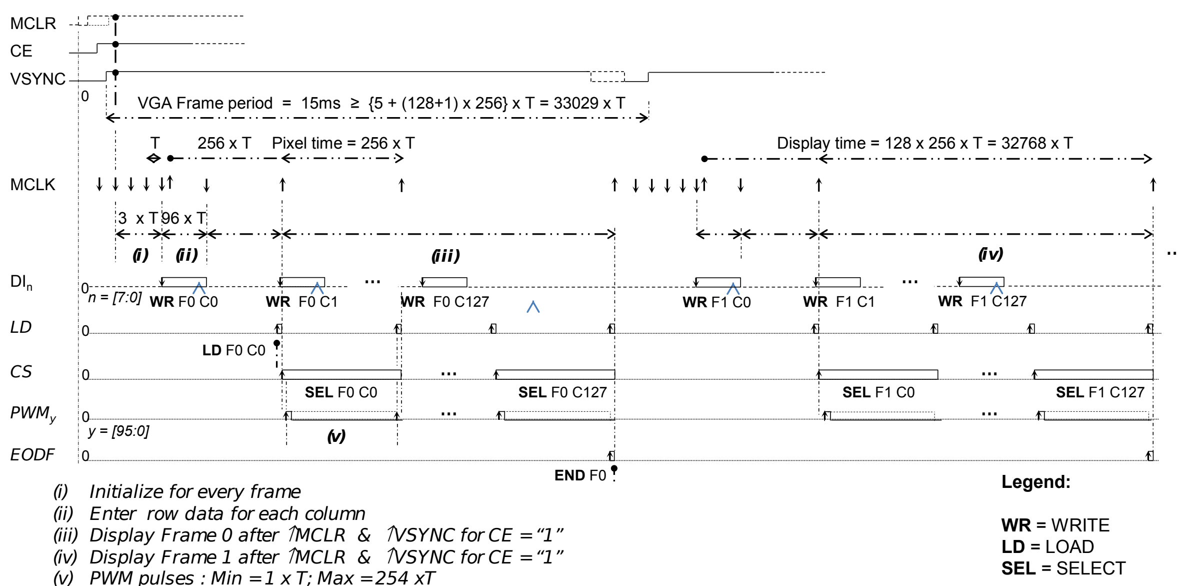


Figure 6.17: Timing diagram of the on-chip display controller

A simple interface is defined for the microdisplay driver circuitry. All logic “1” signals are represented by a 3.3 V voltage level while a 0 V level is interpreted as a logic “0”. Frames are synchronised with a signal VSYNC, which is equivalent to the standard VGA vertical synchronisation signal. Within the frame, timing of all operations is defined by a master clock signal (MCLK). The interface provides a master clear (MCLR) signal that globally initialises the driver circuitry to a known state. Another interface signal is the chip enable (CE), which can put the microdisplay in a low power idle state. Figure 6.17 presents a timing diagram describing the operation of the microdisplay driver.

6.5.4 External controller

One of the main design objectives of the display is to allow a video stream to be presented on-screen. While the internal controller was designed with a generic interface in mind, the video capability of the display was extended to accepting a VGA input. This was given effect through the use of an external controller with figure 6.18 showing a block diagram of the custom FPGA solution. The image or video displayed on the microdisplay is obtained from the VGA port of a personal computer. The VGA protocol is selected to supply the video data to the microdisplay due to its large proliferation in display technology. The external controller was implemented on a custom FPGA platform as shown in figure 6.18 enabling video data to be displayed. The VGA protocol is an analogue technology and a conversion to a digital format is required to enable the video data to be displayed on the microdisplay. A Cyclone IV E from Altera was used to generate control signals and sample ADC data from the VGA port. Three primary colours (red, green, and blue) are sampled individually from the VGA port using an 8 bit ADC and are subsequently stored in RAM. The microdisplay is monochrome and the three colours are averaged to produce a grey scale value of 8 bits. Most graphics cards still support a resolution of 640 x 480. For this reason the external controller accepts a VGA signal of 640 x 480 while the controller allows the microdisplay to display a small portion of the VGA screen in its native resolution of 128 x 96 pixels. Scrolling and panning through the larger captured image are supported by the controller through the use of buttons.

The microdisplay uses a synchronisation signal to signify the start of a screen. The video data stored in RAM is sequentially loaded into the microdisplay controller in the correct order to enable video to be displayed. The refresh rate of the microdisplay is controlled by a display clock and is independent of the image capturing clock. The VGA port is sampled at a rate of at least 24.175 MHz, by the image capturing clock, to ensure the full 640 x 480 screen is captured into memory. Over or under sampling

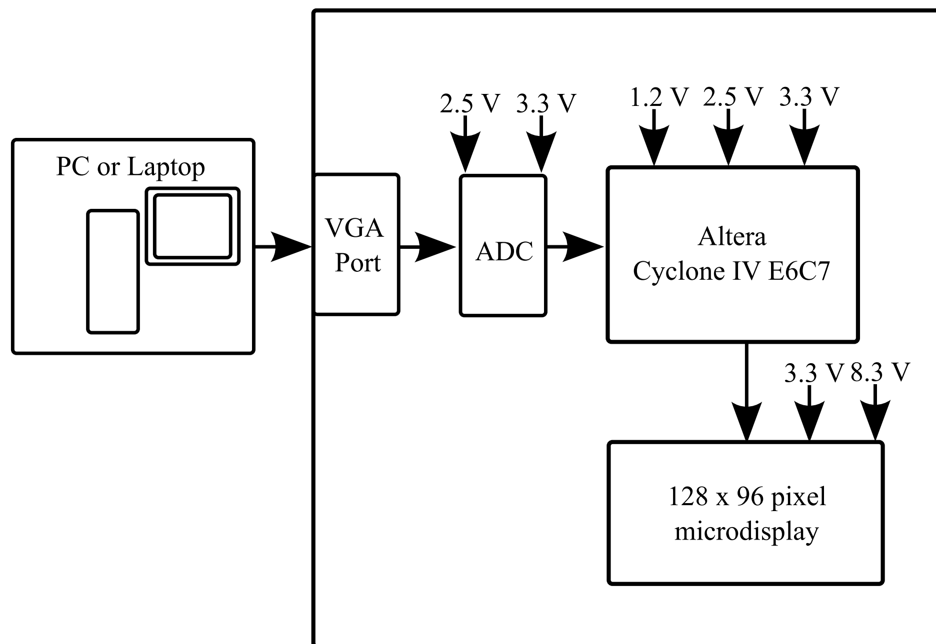


Figure 6.18: A block diagram of the external controller based on a custom FPGA platform

would result in the image becoming blurry. The refresh rate of the microdisplay is set to 60 Hz to match the VGA signal and is not limited by the driving technology or the light sources but is rather selected for a practical reason. The maximum refresh rate is limited by the driving technology and the current microdisplay can be refreshed at rates higher than 1 kHz. Refresh rate is therefore not the limiting factor. The microdisplay is synchronised to the VGA screen in order to prevent image tearing.

6.6 RESULTS

The microdisplay optical pixel array was designed with a direct power supply line to feed the pixel array which makes it possible to directly discriminate the electrical and optical performance of the emissive region of the microdisplay. Both electrical and optical results were characterised for the device at various levels of input power in order to quantify device performance. For all subsequent results, the microdisplay was set up in "flash light" mode, which means the pulse width modulators were set up at full duty cycle and the display was driven as completely on, with applying logic "1" to all row select ports R_{sel} while sweeping the column select ports C_{sel} as per the designed scanning methodology discussed in section 6.5.2.3.

The qualitative and quantitative objectives described at the start of the current chapter translate into

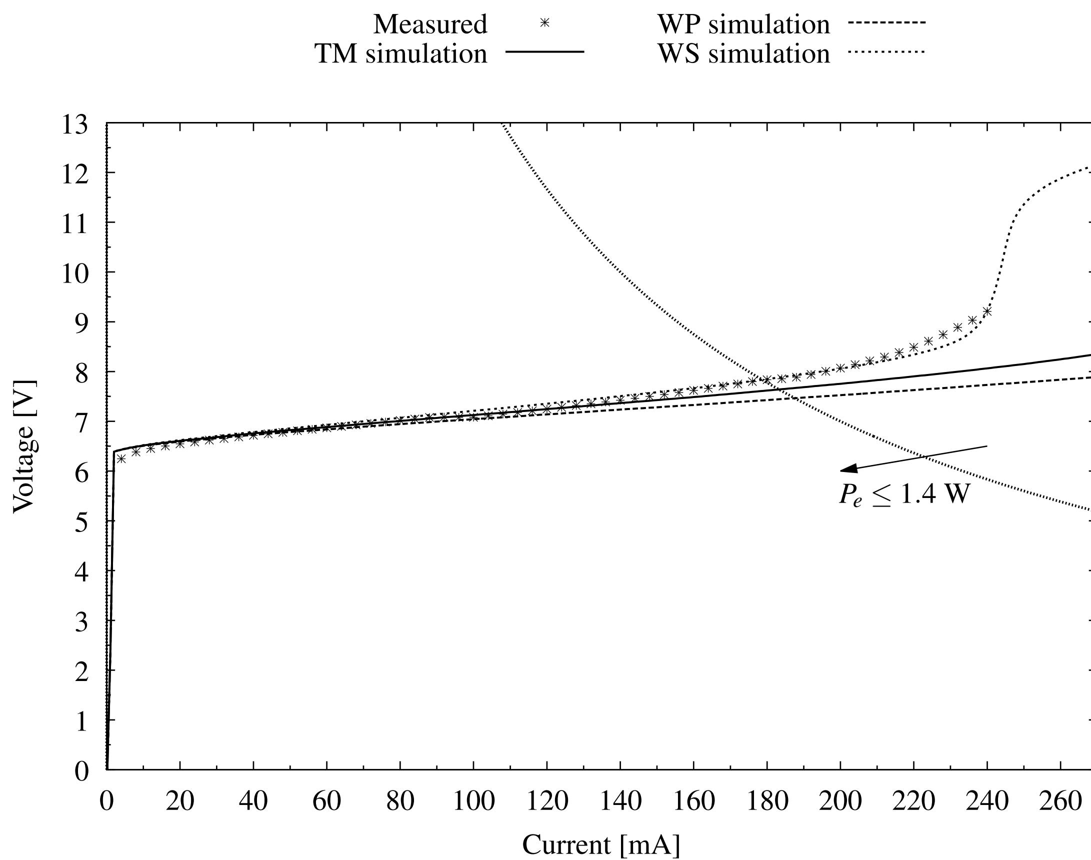


Figure 6.19: Measured voltage and current relationship for the pixel array with previously simulated value indicated

a need to "improve" on the previous 64×8 pixel microdisplay. Lowering the voltage and improving the amount of optical output power emitted by the microdisplay can be used to quantify such an improvement. A comparison is therefore used in the electrical and optical characterisation subsections to understand the improvement gained by utilising reach through light sources.

6.6.1 Electrical characteristics

After the preliminary expectation given in figure 6.11 which was based on the assumption that the n^+p-type light source would behave the same as a p^+n-type junction in both its electrical and optical characteristics, a 392 element p^+n-type array of reach through sources was constructed with a separation distance of 225 nm. The current-voltage characteristics were scaled according to the expected current flow in a pixel containing 63 individual elements. Using this as a comparative platform, the driving transistors were simulated at the typical mean (TM) process corner, the worst speed (WS)

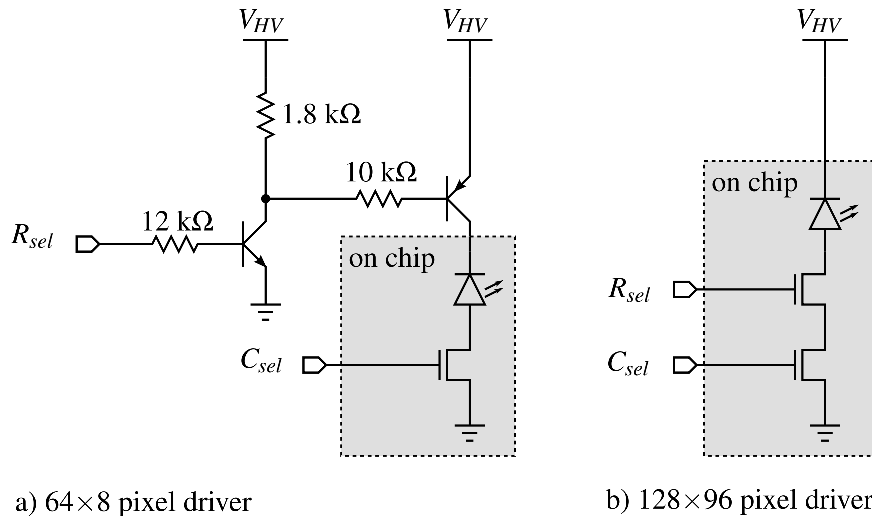


Figure 6.20: A comparison between the 64×8 and 128×96 pixel microdisplays showing the complete internal and external driving circuitry

corner which accounts for weak NMOS and PMOS devices and the worst power (WP) corner covering strong NMOS and PMOS devices. Of course, the PMOS devices are irrelevant in this case and the worst zero and worst one corners can effectively be used as well.

Measurements on the actual microdisplay were done in order to compare the actual performance with the simulated expectations. The optical pixel matrix terminals were then forced currents while operating in full flash light mode from 0 to 240 mA and the voltages were monitored using a parameter analyser. The results were recorded to produce the measurements shown in figure 6.19 which shows the expected light source results at the different process corners as well as the actual measured results. At lower currents it seems that the display follows the TM curve well, while at higher currents it starts to deviate.

As the parameter analyser was only able to force currents up to 100 mA, parallel current sources were applied to the node to extend its range, although the confidence level of the results is reduced. Furthermore, the pixel matrix was scanning while the measurements were taken which may also affect the results. To counter these effects and to increase the reliability of the measurements, the integration time on the parameter was set to maximum in order to average out any oscillatory behaviour on the terminals. Overall, the results are in good agreement with the predictions.

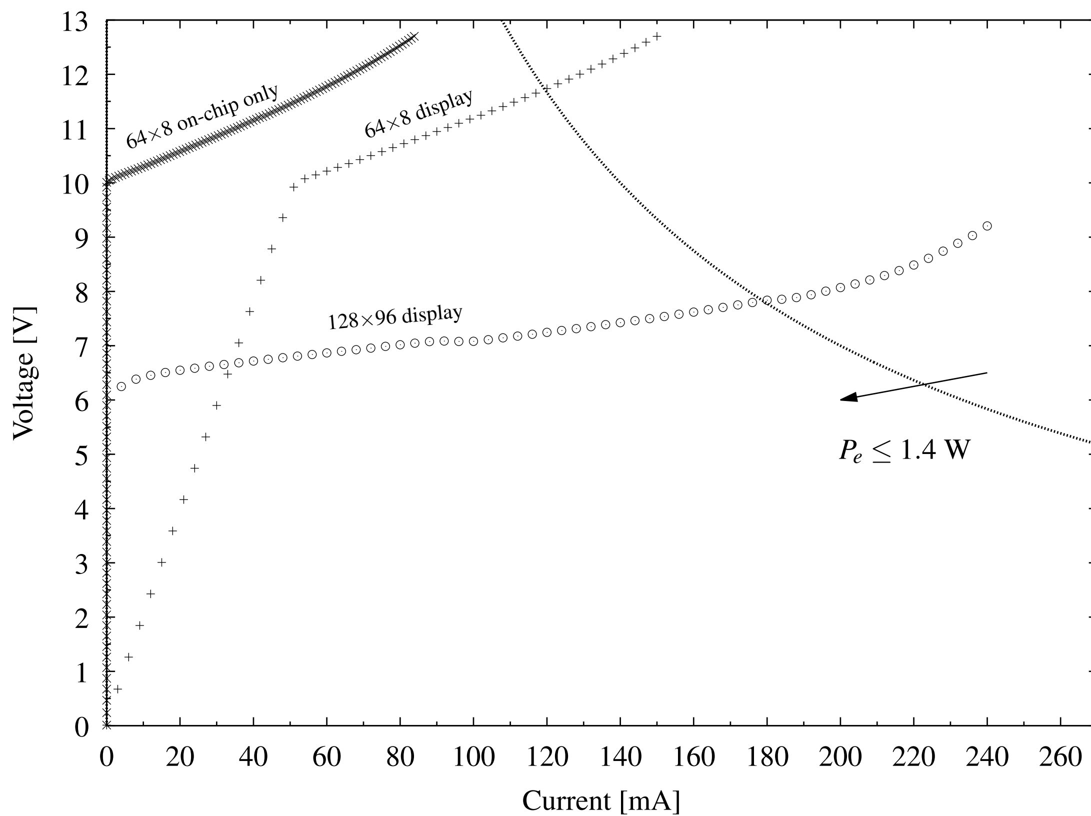


Figure 6.21: Comparison of voltage and current characteristics with previous display

Note that in figure 6.19 the power constraint region is also shown, which indicates that the display operates under $P_e \leq 1.4$ W while the pixel matrix is driven below 180 mA. In practical applications, this constraint can be relaxed as images and video will seldom operate at full scale brightness for all pixels simultaneously.

In order to compare the result from the 128×96 pixel display with the 64×8 pixel display, it is important to understand the difference between the circuit topologies used. Figure 6.20 shows the two driver topologies. Figure 6.20a shows that external circuitry was required in order to drive the microdisplay's optical matrix. In contrast, the 128×96 pixel display (figure 6.20b) was designed to have no requirement on external circuitry by integrating all of the driver circuitry on-chip. The approach was to try and determine the electrical characteristics of the on-chip component, shown as grey area in figure 6.20, which required de-embedding the effects of the external circuitry when evaluating the measured results.

Figure 6.21 compares the results of the 64×8 pixel microdisplay to the 128×96 pixel display. The two

traces present on the graph for the 64×8 pixel display represent the electrical characteristics with and without the external circuitry taken into account. What is clear is that, before the light sources turn on, the external driver bias circuitry presents an effective series load of 192 Ω while the calculated value is

$$R_{ext} \approx \frac{(1.8\ \text{k}\Omega \,||\, 10\ \text{k}\Omega)}{8} \approx 190.6\,\Omega \tag{6.2}$$

The operating voltage of the 64×8 pixel display is significantly higher than the 128×96 pixel display, where the former reaches the $P_e = 1.4$ W mark in flash light mode at around 120 mA when including the external circuitry, while the latter reaches the point at around 180 mA. The slope of the voltage-current relationship is steeper for the previous design.

6.6.2 Optical characteristics

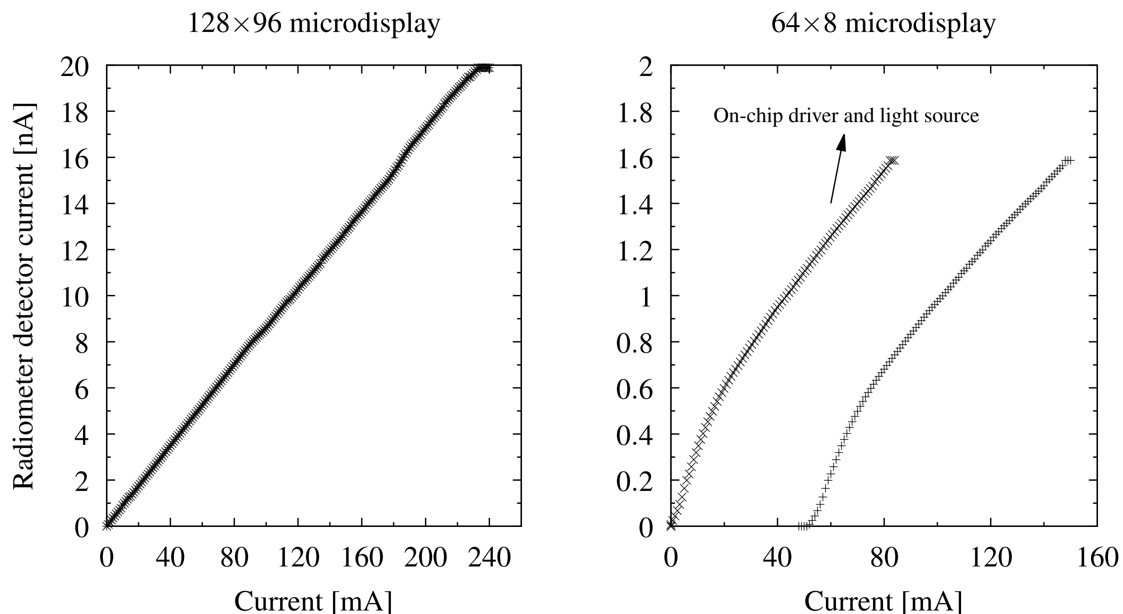


Figure 6.22: Radiometer output of the microdisplay emission using p^+n-type reach through sources

The microdisplay optical power was measured using a radiometer detector which produces a detector current proportional to the optical power detected. A detector was placed in proximity to the microdisplays in order to measure its optical output as a function of input current. This was done in the same fashion for both 64×8 and 128×96 pixel microdisplays in order to provide a sensible comparison. Figure 6.22 shows the results for both displays. As expected from the results obtained in chapter 5, the 128×96 pixel display, which utilises reach through light sources, shows a linear relationship

between the current and optical output power. However, this is not true for the 64×8 pixel microdisplay. Keeping the explanation of the external drive circuitry in mind, there are again two traces on the plot; one representing the complete display including external drivers and one pertaining only to the on-chip components. From the magnitude of the radiometer reading it is clear that the 128×96 pixel display performs according to expectations based on the results in section 6.3. At 150 mA, the light emitted by the previous microdisplay generates 1.56 nA at 12.6 V, including external circuitry, while the improved display registers 12.8 nA at 7.52 V.

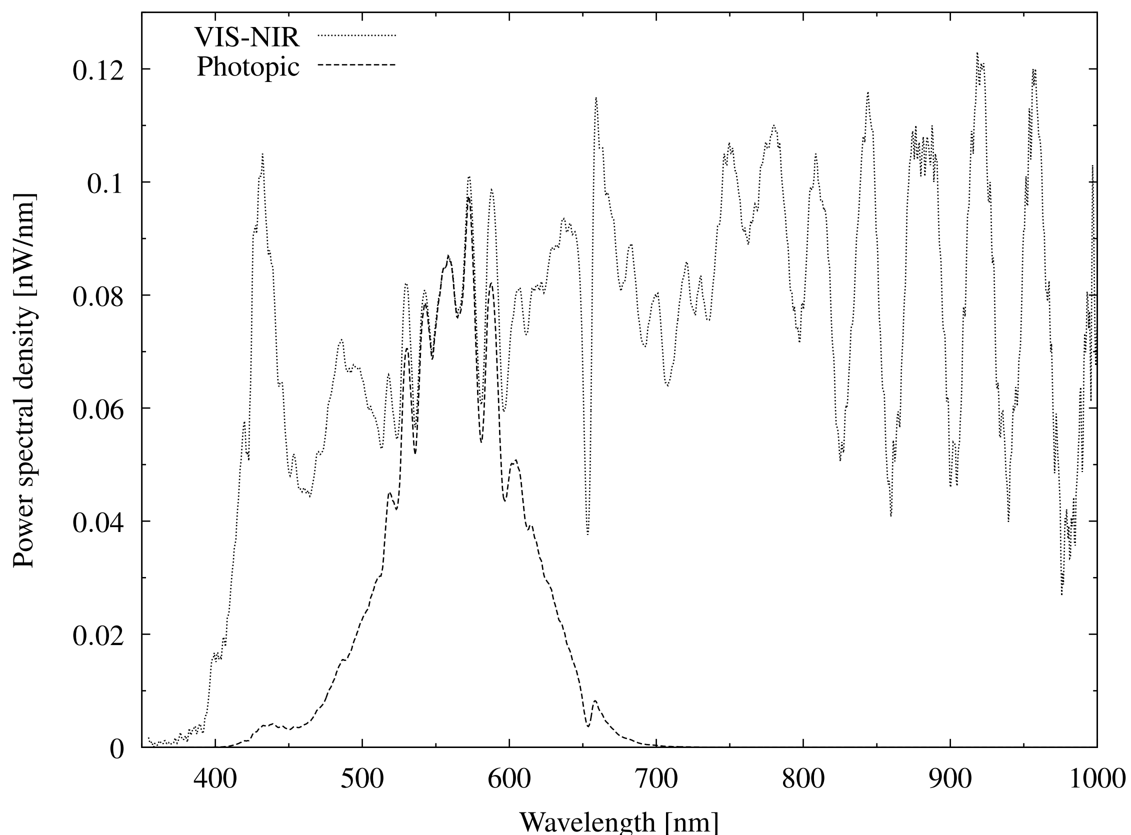


Figure 6.23: Power spectral density and photopic weighted spectral density of light source used in the microdisplay at 230 mA

The power spectral density of the p^+n-type light source used in the 128×96 pixel microdisplay, including the photopic weighted spectral density which is used in luminance calculations, is shown in figure 6.23 . Since the spectrum remains constant in a relative sense at different currents, the relative spectrum can easily be adapted according to the radiometer detector output. Unlike the 64×8 pixel display light sources, there was no nitride etch defined over the light sources. This left the complete SiO_2 layer intact in addition to the Si_3N_4 passivation layer, resulting in a thick back end layer through

Table 6.9: Electrical input and optical output characteristics of the 128×96 pixel display compared to the 64×8 pixel display at maximum designed input current

Parameter		128×96	64×8	64×8 on-chip only
Supply current	I_{sup}	230 mA	150 mA	83 mA
Supply voltage	V_{sup}	8.81 V	12.7 V	12.7 V
Electrical input power	P_e	2.03 W	1.905 W	1.054 W
Radiometer detector current	I_{det}	19.80 nA	1.59 nA	1.59 nA
Detected optical power	P_{opt}	52.57 nW	3.95 nW	3.95 nW
Photopic weighted power	$P_{photopic}$	8.51 nA	0.44 nW	0.44 nW

which light needs to travel. The interference patterns are prominent on the spectrum. The optical output power emitted by both displays as registered by the radiometer photodetector is shown in table 6.9 along with the input electrical parameters where the displays are operating at its designed maximum input power point.

The conversion efficiency of both displays is compared as a continuous function of electrical input power and detected optical power as shown in figure 6.24. These results show that the 128×96 pixel display yields a great improvement in efficiency when compared to the 64×8 pixel display. The improvement is largely attributable to

- the reduced voltage of the reach through light sources utilised in the 128×96 pixel display and
- the improved optical extraction efficiency due to the field oxide enhancement developed in chapter 5.

Although it is expected that the results in figure 6.24 will result in much improved display characteristics, it should be noted that luminance depends on two additional factors, namely

- the area of emission and
- the solid angle over which radiation is emitted by the light sources.

In order to understand these effects, the luminance, which ultimately determines the perceived brightness of the display, is calculated in the next section.

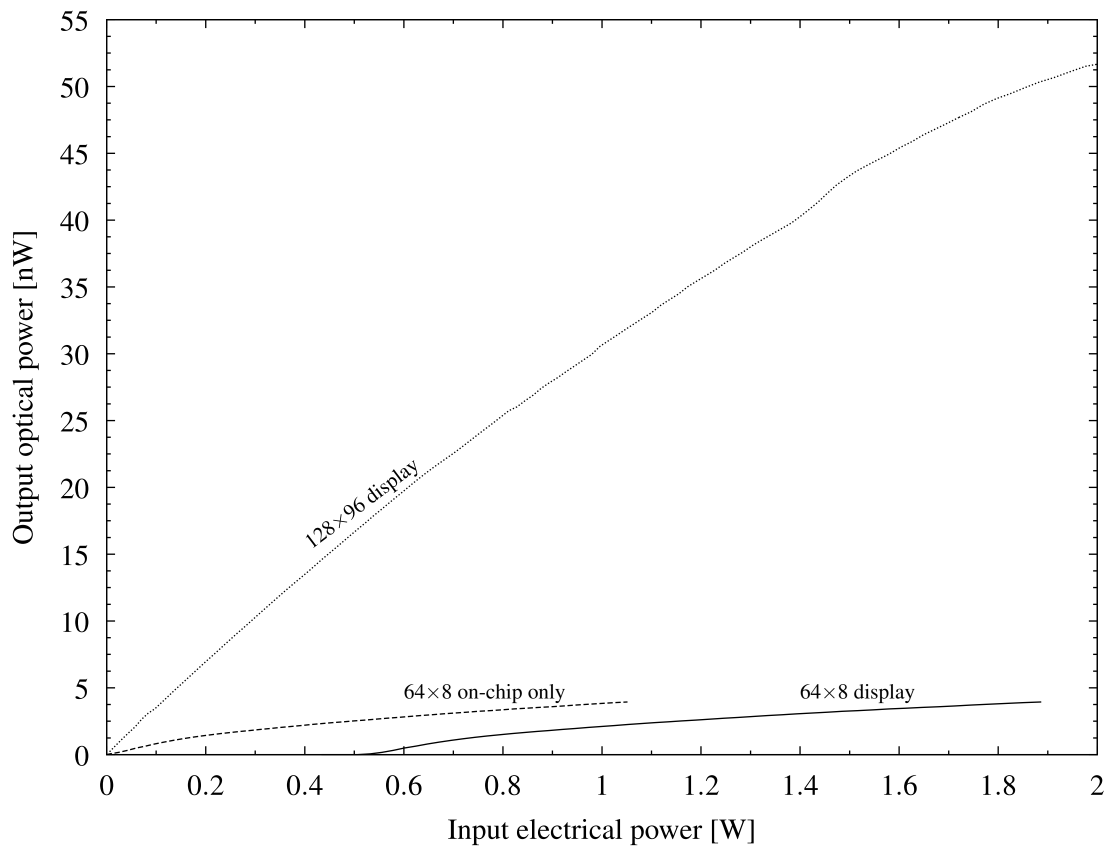


Figure 6.24: Comparison of power efficiency between the 64×8 and 128×8 pixel displays

6.6.3 Luminance

If the power as weighted by the photopic luminosity function is known, it is possible to calculate the luminance values as described by equation 4.9 explained in section 4.5.7 if the following values are known:

- Φ_e radiant flux (optical power) in watt,
- I_e radiant intensity in watt per solid angle in steradian,
- I_v luminous intensity in candela requiring radiant intensity weighted by the luminosity function and
- the area of emission in square meters.

The luminance value can then be computed through equation 4.9.

Based on the information required, it is important to have knowledge of the radiation pattern emitted by the surface for which the luminance calculation is required. There are two determining factors when measuring a small surface such as a point source in a microdisplay, viz

1. the solid angle subtended by the detector element, such as a radiometer photodetector, and

2. the angle of emission by the light source itself, where the limits can be set as the half-power angle of the radiation pattern,

whichever one is the smallest. For practical reasons, it is assumed that the detector is in sufficient proximity to the light source in order to use the same basis for comparison as in chapter 4.

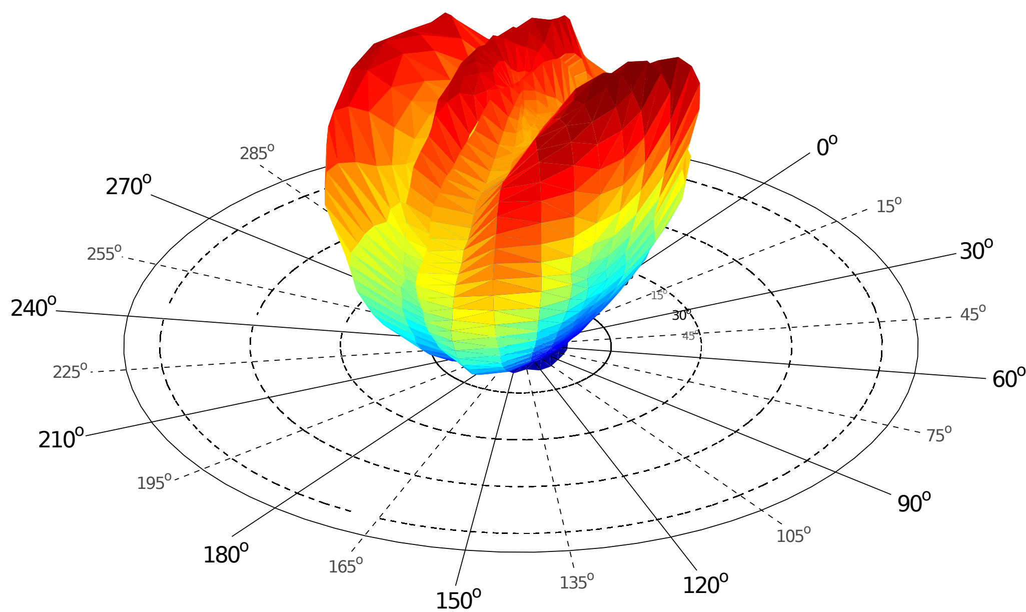


Figure 6.25: Three-dimensional radiation pattern of the reach through device used in the improved pixel design

Figure 6.25 shows a relative representation of the three-dimensional radiation pattern as emitted by the sources which are employed in the 128×96 pixel microdisplay. The following interesting observations are worth noting.

1. The radiation pattern subtends a much larger solid angle than for the 64×8 pixel light sources

from chapter 4.

2. There are three very distinct peaks visible which runs in parallel to the device in one direction. This will be explained below.

3. In the sweep direction orthogonal to the peaks the radiation pattern exhibits behaviour similar to a Lambertian cosine distribution.

Figure 6.26a contains a cross section representation when taken along the direction of current flow in the light source. The p^+-region is shown on the left. The centre region where light generation occurs is an n^--region followed then by an n^+-region in order to stop the field and make decent ohmic contact. Figure 6.26a shows that light originating in the silicon bulk experiences a flat Si/SiO_2 interface. This explains the shape of emission from this cross-sectional profile as shown in figure 6.27.

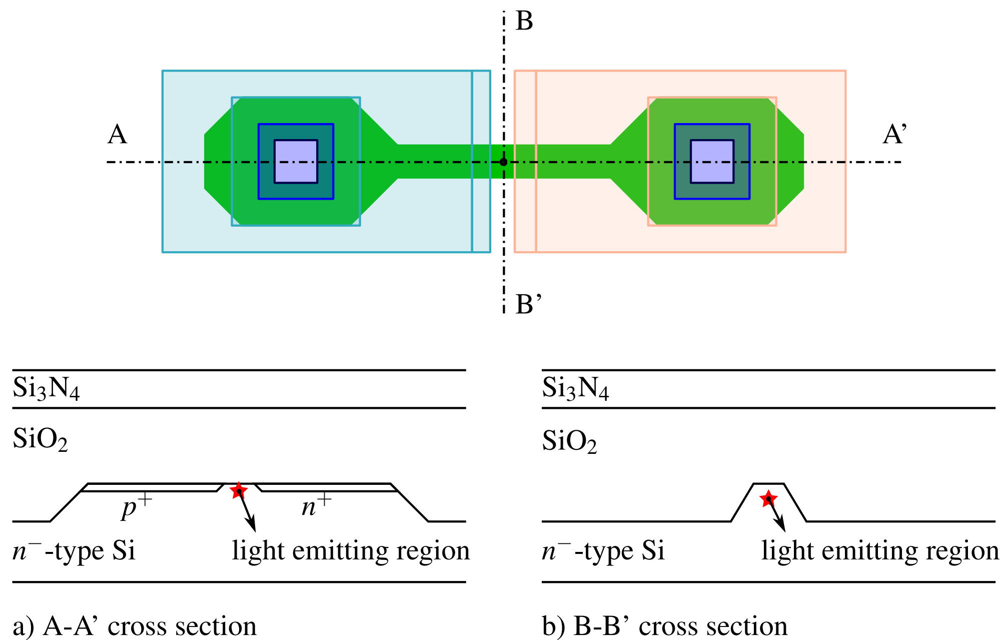


Figure 6.26: Cross sections of light source showing LOCOS surface features

The three peaks can be explained when examining a cross section of the device where the LOCOS profile of the device is shown in figure 6.26b. One of the peaks is located in a normal direction to the chip surface, while the other two peaks are symmetrical and lie at an angle of 27.5° from the normal. The rotational axis of the measurement is around A-A' while the radiation pattern is viewed as originating from the cut B-B'. Figure 6.27 shows the two dimensional radiation pattern originating

from the cross section profile along B-B' and three peaks are clearly visible. The origin of this can be inferred from the approximation in figure 6.26 where the region on top of the light source is flat, while the light source is also bounded on both sides with a Si/SiO_2 interface formed at an angle due to LOCOS growth.

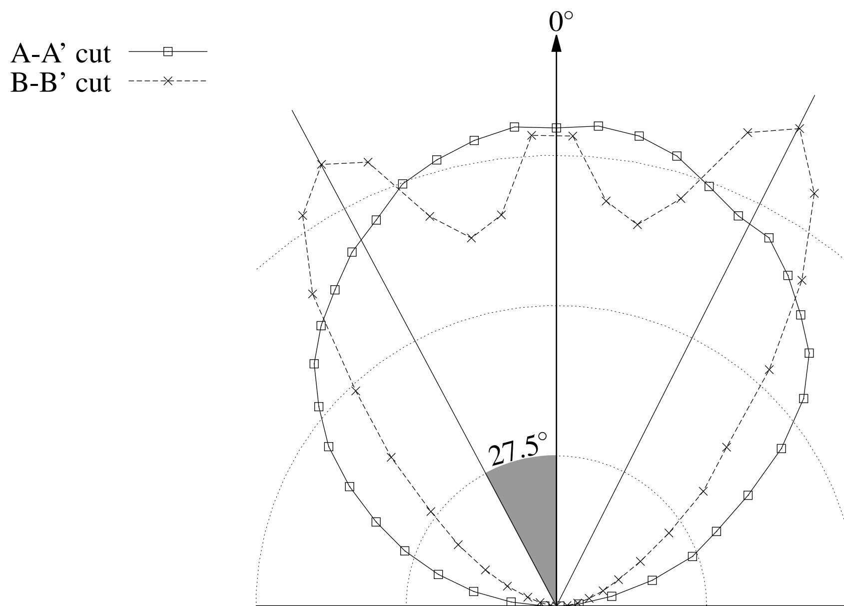


Figure 6.27: Radiation pattern across two orthogonal cross-sectional cuts with reference to figure 6.26

A qualitative explanation is that light emission normal to the Si/SiO_2 interface will experience maximum transmission, while the critical angle of this interface is in the region of 22° before total internal reflection occurs. The exact nature of the LOCOS profile with respect to the light source origin will determine how much light is emitted, but due to the LOCOS profile will exhibit peaks in the regions where maximum transmission occurs. An in-depth study of the complete effect of the LOCOS interface is beyond the scope of this work and will be investigated in future work.

From figures 6.25 and 6.27 an approximation to the half power angles can be made according to the following philosophy while understanding that the approach provides, at best, only an indication of the solid angle.

1. The half power point is calculated for each cross section of the radiation pattern.

2. An average is taken between the two half power points.

Table 6.10: Half power angle derivation results

	A-A' cut	B-B' cut
Half power (averaged) angle 1 from normal	56.95°	46.2°
Half power (averaged) angle 2 from normal	59.05°	46.3°
Average angle from normal for each cut	58.00°	46.25°
Average angle from normal	52.13°	
Average apex angle 2θ	104.26°	

3. The angle that is associated with the averaged half power point is extracted from the radiation pattern.

4. According to the average value of the radius of an ellipsoid it is valid to take the arithmetic mean in order to compute an average half power angle of both radiation patterns.

This approach is tabulated in table 6.10 in order to provide an effective apex angle derived by taking the angles from the normal direction to the chip surface to be used in the luminance calculation for an approximate result.

Under the following key assumptions that 1) the spectrum does not materially change as a function of light source current, 2) the light spot does not change position which would result in current dependent radiation pattern, it is possible to compare the luminance values for the two displays under discussion. The results are shown in figure 6.28. The surprising result is that the 128×96 pixel display exhibits a low luminance compared to the 64×8 pixel microdisplay. Proper consideration of the difference in area and solid angle clarifies the reason for these seemingly unintuitive results.

6.6.4 Summary

An overview of the performance of the 128×96 pixel microdisplay highlights a number of distinct advantages, especially compared to the previous 64×8 pixel implementation. Some of the electrical improvements include

- complete integration of the driver circuitry on-chip with successful operational results,
- lower operating voltage of the light source and driving circuitry of between 6.3 V and 8.5 V

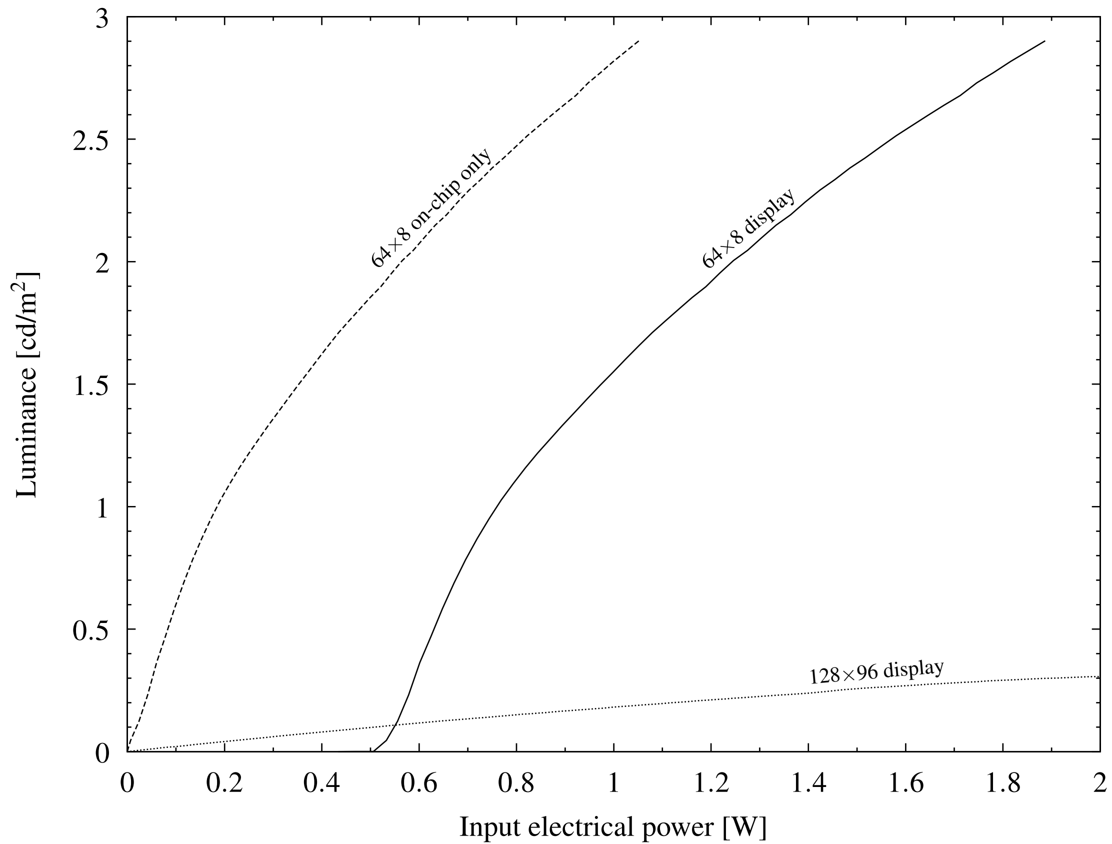


Figure 6.28: Comparison of luminance between displays

Table 6.11: Luminance comparison of the 128×96 and 64×8 pixel microdisplays

Parameter		128×96	64×8
Radiant flux [nW]	Φ_e	52.57	3.95
Apex angle [°]	2θ	104.26	18.3
Solid angle [sr]	Ω	2.43	0.08
Luminous intensity [cd]	I_v	2.40×10^{-6}	3.71×10^{-6}
Area [m^2]	A	7.68×10^{-6}	1.28×10^{-6}
Luminance [cd/m^2]	L_v	0.312	2.901

over the operating range of interest,

- higher current capacity of the pixels than in the previous design and

- proof of the possibility to create an operational integrated circuit where digital logic can interact

directly with reach through light sources.

The optical results yield proof of an increase in optical power emitted by the microdisplay and a wide solid angle of light emitted from the surface with the direct influence of the Si/SiO_2 interface affecting the radiation pattern.

Power efficiency improvements of above a factor 10 were obtained by the reach through light sources developed in chapter 5 compared to the performance of the state of the art light sources used in the 64$\times$8 pixel microdisplay. This is the result of better extraction efficiency and a reduction in operating voltage of the light sources.

Although the power efficiency is much better using the improved light sources, the luminance value dropped substantially as implemented in the 128$\times$96 pixel display for two reasons, namely

- the area of the display is much larger and
- although the power emitted is much more, the solid angle through which it emits is much broader.

The result is that the luminance performance of the 64$\times$8 pixel microdisplay outperforms the 128$\times$96 pixel display by up to a factor 15 when considering the on-chip power consumption.

6.7 CONCLUSION

Electric field reach through offers a practical degree of freedom to a circuit designer who wishes to reduce the operating voltage of hot carrier luminescent CMOS light sources. This chapter proves beyond doubt that the technique not only facilitates the direct interaction between digital logic circuits and integrated light sources, but also proves that such integration can be scaled up depending on the application, such as where higher resolution displays may be attractive. The lower voltage also opens the opportunity to make use of a number of mobile power sources, such as batteries, for use in portable systems. Figure 6.29 shows the beautiful 128$\times$96 pixel microdisplay as implemented in a standard 0.35 μm CMOS process which is described in a conference proceeding [99] and will form the subject of a journal publication.

Although the luminance of the 128$\times$96 pixel display is much lower, a VGA signal displaying video

created a usable, miniature display which is perceptible with the human eye given the correct background environment. The major improvement in power efficiency also promises a substantial improvement in luminance *if* it is possible to use light directing structures to reduce the solid angle without a loss in optical power.

In conclusion, the improved reach through light sources and the integration techniques discussed in this and prior chapters enable any standard CMOS foundry to become a microdisplay manufacturer.

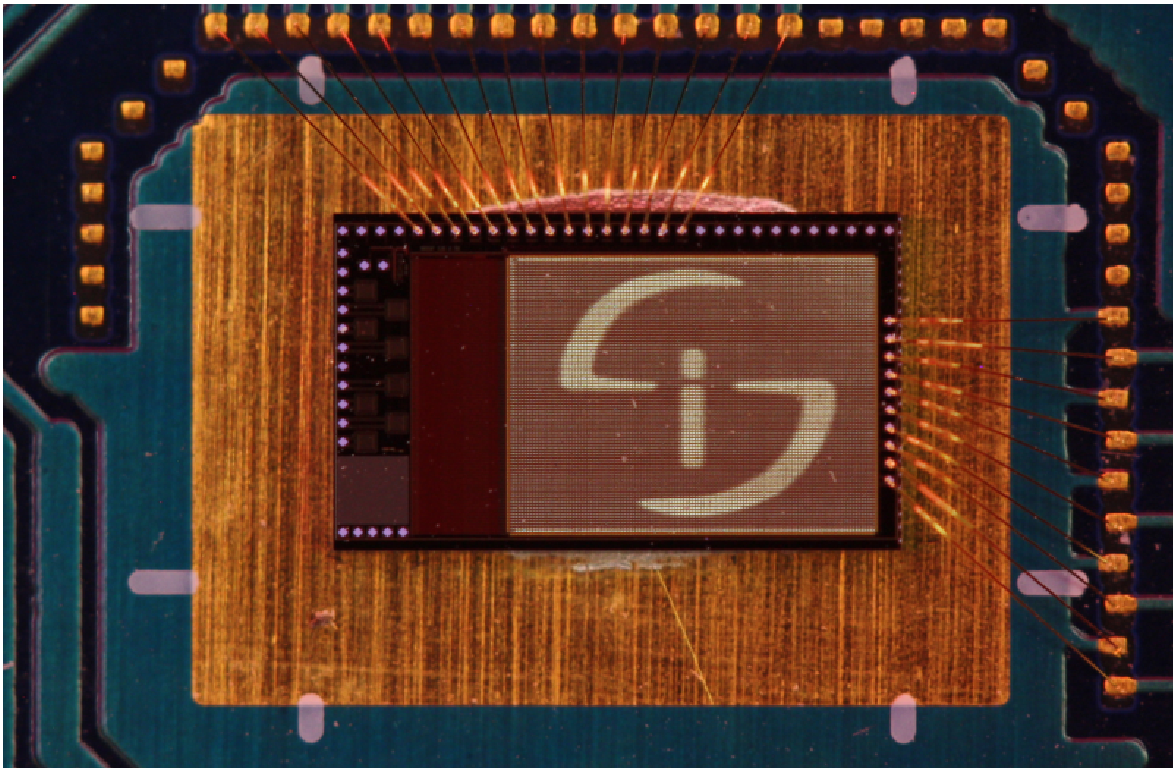


Figure 6.29: Micrograph of the 128×96 pixel microdisplay displaying an image

CHAPTER 7

CONCLUSION

7.1 SUMMARY OF WORK

This work is the first to address the problematic aspects of integrating hot carrier luminescent light sources in standard CMOS with the purpose of coexistence and interaction with other digital circuitry on-chip. State of the art light sources, previously developed by our research group, were incorporated into the design of a 64×8 pixel dot matrix microdisplay. Techniques for facilitating integration and overcoming the difficulties associated with the high voltages of the light sources were developed. However, the need for customisable voltage levels became apparent.

The hypothesis states that electric field reach through between the two highly doped regions of a reverse-biased junction in breakdown, can be used for facilitating integration. Since the approach discussed in chapter 3 section 3.4 proposes lowering the voltage using the electric field in reach through conditions, the results from chapter 5 prove that electric field reach through can be used to facilitate integration by lowering the operating voltage of hot carrier luminescent light sources based on *pn*-junctions in breakdown. A novel contribution, through which the removal of field oxide between the highly doped region is effected and carrier transport is directed in a channel-like structure, provides a robust tool for an integrated circuit designer. This technique improves the electrical predictability of the breakdown voltage as well as the external extraction efficiency, allowing more light to leave the chip surface. New light sources were developed using this technique and proved to be fully compliant with standard CMOS without any required process modification.

The improved light sources were implemented in a scaled-up version of a CMOS microdisplay, resulting in a 128×96 pixel microdisplay with more complicated digital logic integrated on the same chip

as the emissive area. This proved not only that large scale integration is possible, but also that the newer light sources reduced the supply voltage of the entire display substantially. The electrical and optical characteristics were rigorously compared to the previous microdisplay and it was found that the new light sources developed in this work had EPE improvements of up to a factor 10. However, the wide emission angle visible in the radiation pattern and larger emissive area had a detrimental impact on the luminance of the larger display since the optical power, although improved, is spread over a much larger solid angle.

This work proves the hypothesis true and successfully addresses the research questions as set in chapter 1.

7.2 CRITICAL ANALYSIS

7.2.1 Hypothesis

The hypothesis put forth in chapter 1 is restated for convenience.

> *If the electric field distribution can be changed through geometrical adaptations which will result in effects such as electric field reach through between two highly doped regions in a pn-junction, the electrical and optical characteristics of silicon hot carrier luminescent light sources may be altered and optimised in such a way as to facilitate the integration of these light sources in standard CMOS as usable components in integrated systems such as microdisplays.*

Reach through has been shown to exist in the devices developed in chapter 5, with the effect of lowering the operating voltage of the light sources as a function of separation distance between the highly doped regions of a *pn*-junction. The reduced voltage allows for easier integration, relaxing the extreme operating conditions of other circuit components which need to interact with the light sources.

Chapters 4 and 6 both proved that integration of silicon light sources is possible by utilising the light sources in microdisplay implementations. Since the spectral content of the light sources affords sufficient optical power in the visible part of the electromagnetic spectrum, the microdisplays presented excellent platforms for showcasing both the integration aspects and evolution of light source improvements throughout this work.

7.2.2 Research questions

By directly addressing the research questions, the value and contribution of this work become apparent. With reference to chapter 1 section 1.5, the questions are elaborated on below.

1. Electric field reach through can be used as a means for lowering the voltage of silicon light sources based on hot carrier luminescence. Since the breakdown voltage can be modified by terminating the electric field using a field stop region, a circuit designer can tailor the voltage according to the application requirements.

2. The bounds between which optimisation and design will yield usefull devices are limited on one end by the region where the electric field length is short enough for field emission to prevail as dominant transport mechanism, which limits the minimum operating voltage of hot carrier light sources, and on the other end when the electric field falls to zero before reaching the field stop region, whereafter resistive losses degrades the performance of a device.

3. Efficiency improvements are somewhat complicated as the exact radiative mechanism is not yet fully understood and modelled. From the results of this work, it seems that impact ionisation and field emission start to compete at the separation distances of interest. Although the voltage is reduced, the radiative flux is also reduced. However, in a junction where the electric field does not reach through, there exists more ohmic losses than in reach through conditions. Carriers are not subject to a drift force in the quasi-neutral region of the lightly doped side, resulting in larger series resistance. Reach through therefore improves power efficiency but not necessarily quantum efficiency. The specific geometrical configuration of reach through light sources developed in this work does improve the extraction efficiency, although this is not directly ascribable to the reach through condition itself.

4. Hot carrier light sources have proven, on first silicon run iteration for two microdisplays, that they can coexist and furthermore interact with other on-chip circuitry. This work proves it true especially for digital circuitry since the standard cell library is where the power of a foundry's scalability lies. The high operating voltage of hot carrier light sources provides challenges which can be alleviated by either controlling the operating voltage, as proven in this work, or by using appropriate circuit design techniques as was also used in this work.

5. Both microdisplays utilised hot carrier luminescent silicon light sources and both of the mi-

Table 7.1: Non-exhaustive summary of results on efficiency and operating voltage of hot carrier luminescence from silicon

Year	Ref.	IF	5-y IF	Technology	Operating mode	Voltage	η_{eqe}	η_{epe}
2013	This work	-	-	CMOS	Reach through n^+p	5.56	6.83×10^{-8}	2.35×10^{-8}
2013		-	-	CMOS	Reach through n^+p	6.15	9.41×10^{-8}	2.66×10^{-8}
2013		-	-	CMOS	Reach through n^+p	6.72	1.08×10^{-7}	2.69×10^{-8}
2013		-	-	CMOS	Reach through n^+p	7.71	1.19×10^{-7}	2.57×10^{-8}
2013		-	-	CMOS	Reach through n^+p	8.52	1.26×10^{-7}	2.43×10^{-8}
2013		-	-	CMOS	Avalanching n^+p	10.56	8.13×10^{-8}	1.22×10^{-8}
2013	[51]	2.168	2.169	CMOS	PMOSFET/diode Ava.	-	4.30×10^{-8}	-
					PMOSFET/3-term Ava.	35	1.50×10^{-7}	8.26×10^{-9}
2010	[56]	Proc.	Proc.	CMOS	Zener p^+n^+	4.8	2.0×10^{-7}	7.0×10^{-8}
					Forw. bias p^+n^+	1.2	6.4×10^{-7}	7.0×10^{-7}
2010	[57]	0.731	0.656	CMOS	Avalanche	8.7	7.0×10^{-8}	1.40×10^{-8}
					Zener	5	2.0×10^{-7}	7.0×10^{-8}
2006	[47]	Proc.	Proc.	CMOS	Avalanching p^+n	10	3.60×10^{-8}	1.36×10^{-8}
1993	[10]	1.82	2.009	CMOS	Forw. bias	1.2	1×10^{-4}	1×10^{-4}
					Avalanching p^+n	52	3.40×10^{-8}	1.10×10^{-9}
					Avalanching n^+p	31.2	6.50×10^{-8}	3.90×10^{-9}
					Avalanching p^+n^+	6.5	5.40×10^{-8}	2.00×10^{-8}
					Avalanching p^+n	33.2	2.00×10^{-8}	1.10×10^{-9}
1956	[32]	-	-	Silicon crystal	Avalanche	8 - 156	7.00×10^{-9}	-

crodisplays offered luminance levels discernible by the naked eye. Although the 64×8 pixel microdisplay had integrated on-chip logic, the 128×96 pixel display had a much higher degree of integration, including memory registers and PWM controllers. This work therefore proves it possible to create large scale integrated microdisplays in standard CMOS without the need for process modifications or post processing. Aspects around photopic spectral content, emission solid angle and luminance were investigated in detail.

7.2.3 Comparison of efficiency in relation to other work

It is important and insightful to position this work in relation to other work considering the efficiency of the light sources in converting electrical power to optical power. Table 7.1 shows the results developed in chapter 5 as improved devices in context of the work by other groups.

Two types of efficiencies are most often found in literature and are used in table 7.1 for comparison. These are

- η_{eqe}, the external quantum efficiency in photons emitted per charge carrier through the device and
- η_{epe}, the external power efficiency in optical power emitted versus electrical input power.

This work shows that the light sources developed in chapter 5 where the reach through effect occurs have an average EPE of $\eta_{epe} \approx 2.5\times10^{-8}$. This value improves on the EPE of an avalanching junction based on the same design, where the reach through effect ceases, by a factor 2. The work described in this thesis also shows a dramatic improvement of the external quantum efficiency where the reach through light sources offer $\eta_{eqe} \approx 10^{-7}$ while most other works report results in the order of $\eta_{eqe} \approx 10^{-8}$.

The efficiencies obtained in this thesis represent some of the highest reported values in literature. In most cases the developed devices are an order of magnitude more efficient than other reported results. An exception is results reported by the group from the Chinese Academy of Science [56], [57]. The trends observed in our experimental work are in conflict with the results in their reported findings, although their designs are based on geometries our group has explored more than a decade ago. The following possibilities may be able to explain the discrepancy.

1. Features of their designs which form the foundation of the improvements remain undisclosed.
2. The measured results originate from an experimental measurement setup which does not capture the complete extent and complexity of characterising the emitted light.
3. The results may have been exaggerated without scientific grounds.

7.2.4 Comparison of CMOS integration in relation to other work

By nature, the CMOS integration of silicon light sources by other groups vary significantly in nature. The work from this thesis is the first to successfully develop a method whereby the breakdown voltage of the devices can be tailored by the circuit designer for different requirements. However, there are two groups who have experimented with integrating hot carrier luminescent light sources in CMOS ICs with a microdisplay application in mind. The work by Chen [58] is a detailed work on integrating circuitry with light emitters and overlaps on what was developed independently in chapter 4 on a number of aspects of integration. However, this work differs on three fundamental issues:

1. this work needs no image intensification and the measured luminance levels originate directly from the light sources in CMOS,

2. this work aims at providing a solution where the display integrates directly with digital circuitry, which alleviates the need for precision analogue design and provides a much more robust approach for integrating the light sources with other on-chip circuitry, and

3. this work targets the improvement of the light source through techniques which allow lower operating voltages for better integration and field oxide channels for improved extraction efficiency.

Another attempt at integrating hot carrier silicon light sources in a microdisplay application was attempted by the group at the Chinese Academy of Science [59] using light sources previously developed by their group [47], [56], [57]. The work describes integrating light sources using a matrix of 16×16 pixels. However, the 64×8 pixel microdisplay [60], [61] from chapter 4 enjoys prior publication. The work of the Chinese group also does not dwell on the light sources itself and the scale of integration is not comparable to what was done in this thesis.

7.3 CONTRIBUTION

The contribution of the work described in this thesis is summarised below by breaking the contributions down into topic related aspects.

1. The work in this thesis is the first to investigate the integration of hot carrier light sources in standard CMOS processes with the intent of creating seamless interaction between light sources and other digital circuitry from a foundry's standard cell library. A few specific aspects that were investigated include

    - an analysis of the available junctions suitable for use as light sources,
    - the leakage components of junctions and the impact of these on driving circuitry,
    - the effect of lateral device isolating structures, such as LOCOS, on light propagation through the BEOL optical path,
    - localisation of light spot regions through geometrical designs for improved homogeneity

and controllability of light emission,

- polarity considerations based on the doping species of the junctions for selecting light sources suitable for interaction with driving circuits and
- different driving topologies for compact design without compromising efficiency.

2. Integration of light sources was proven by implementation in microdisplay prototypes to an integration level representing products. This included the development of contributions directly related to microdisplay architectures, including

   - the type of scanning architecture suitable for the light source used,
   - all-digital intensity control which improves the robustness of the design,
   - sharing optical surface where light is emitted with optimum routing and power supply schemes as CMOS is a planar technology,
   - electrical and optical characterisation and comparison of the two microdisplay systems developed in this thesis, each of which uses a different light source and pixel design,
   - radiation pattern analysis of pixels in order to assess the impact of the optical beam on microdisplay performance and
   - luminance calculations which quantify the perceptible performance of the microdisplays according to the human eye.

3. Improved light sources utilising electric field reach through were developed and properly characterised in electrical and optical terms. This part of the work is truly novel with highlights including

   - the use of masks available in standard CMOS processes which means that no postprocessing is required,
   - a novel field oxide manipulation approach whereby the electrical characteristics of reach through light sources are stabilised and the extraction efficiency of photons is improved simultaneously,

- the spectral content of reach through light sources proven to be useful for microdisplay applications as there is a substantial overlap between the luminosity function and the light source spectra and

- lower voltage operation was made possible without the need to alter the process, allowing a circuit designer to pick a sweet spot suitable to a specific application based solely on geometrical design input.

7.4 FUTURE WORK

Although this thesis addresses the hypothesis and associated research questions, the work opens a number of subsequent gaps in the body of knowledge which will need to be investigated as future research topics. This section provides an overview on what will become the basis of new work to be investigated. A non-exhaustive summary of the aspects for future work is given below.

1. The dependency of emission characteristics, both in terms of quantum efficiency and the spectral nature of emission, on electric field magnitude needs to be investigated. While the critical field changes slowly once breakdown starts to occur, the exact magnitude of the average electric field used to excite carriers are still open for investigation. From chapter 6 section 6.3.1 shows that a blue shift occurs as the separation distance x_{sep} is decreased in reach through sources, this effect is not explained in this thesis and requires intensive device simulation and more measurements.

2. Reliability and the time dependent dielectric breakdown gate oxide are not exhaustively investigated in this work. Products suited for mass production will require further reliability studies of both the light sources and the circuits directly interacting with these light sources.

3. It was seen in chapter 6 that the radiation pattern of reach through sources utilising the field oxide channel approach differs fundamentally from the sources used in the 64×8 pixel microdisplay in chapter 4. Optical propagation through Si/SiO_2 interface and the subsequent BEOL-stack needs to be intensely investigated for understanding

   - the exact location of the light spot and

   - for designing improved light directing structures in order to enhance the extraction effi-

ciency even more.

4. For display applications, reducing the solid angle of reach through sources is the main reason why the 128×96 pixel microdisplay suffers in luminance compared to the 64×8 pixel display, although the emitted optical power is almost a factor 10 more efficient.

5. Chapter 5 shows that the reach through devices operate in a mixed mode between avalanche and field emission breakdown. Temperature dependence on breakdown voltage needs to be investigated to confirm mechanism of dominant current transport in order to improve our understanding of the correlation between light emission and the breakdown phenomenon.

6. It does seem odd that, suppose emission only results from electrons and therefore depends on intraband carrier relaxation, the prominent blue peak around the wavelength 430 nm (see figure 6.1) remains in both n^+p- and p^+n-type devices while there is still a linear relationship to light emission vs input current. This means that the quantum efficiency does not depend on the time a carrier spends in the electric field, otherwise the results between n^+p- and p^+n-type would have differed. The loss mechanism resulting in an almost linear relationship in the decrease of optical output as the voltage decreases revolves around the field emission vs avalanche mechanism for transport. A thorough investigation in the above is therefore warranted.

7. Band-to-band tunnelling may be the cause of the second slope in the improved reach through devices, notably in figure 5.27. Temperature tests are needed to confirm the origin of the second slope in the logarithmic IV characteristics of the reach through arrays.

The list above is not exhaustive and indeed provides an exciting need for more research in the field of hot carrier light sources in standard CMOS processes. The opportunity presented by a light source with usable output levels remains sufficiently attractive for any improvements to be worthwhile.

REFERENCES

[1] R. Tummala, "Moore's law meets its match (system-on-package)," *IEEE Spectr.*, vol. 43, no. 6, pp. 44 – 49, 2006.

[2] H. J. Round, "A note on carborundum," *Electrical World*, vol. 49, p. 308, 1907.

[3] O. V. Losev, "Luminous carborundum detector and detection with crystals," *Telegrafiya i Telefoniya bez Provodo*, vol. 44, pp. 489 – 494, 1927.

[4] J. R. Haynes and H. B. Briggs, "Radiation produced in germanium and silicon by electron-hole recombination," *Phys. Rev.*, vol. 86, no. 4, p. 647, 1952.

[5] R. Braunstein, "Radiative transitions in semiconductors," *Phys. Rev.*, vol. 99, no. 6, pp. 1982 – 1983, 1955.

[6] W. Shockley, *Electrons and holes in semiconductors with applications to transistor electronics*. Toronto: Van Nostrand, 1950.

[7] J. I. Pankove, *Optical processes in semiconductors*, ser. Solid state physical electronics series, N. Holonyak, Ed. Prentice-Hall, 1971.

[8] C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, "A review of some charge transport properties of silicon," *Solid-State Electron.*, vol. 20, no. 2, pp. 77–89, 1977.

[9] G. Patton, S. Iyer, S. Delage, S. Tiwari, and J. Stork, "Silicon-germanium base heterojunction bipolar transistors by molecular beam epitaxy," *IEEE Trans. Electron Devices*, vol. 9, no. 4, pp. 165–167, 1988.

[10] J. Kramer, P. Seitz, E. Steigmeier, H. Auderset, B. Delley, and H. Baltes, "Light-emitting devices

in industrial CMOS technology," *Sens. Actuators, A*, vol. 37-38, no. 0, pp. 527–533, 1993.

[11] J. Zhao, M. A. Green, and A. Wang, "High-efficiency optical emission, detection, and coupling using silicon diodes," *J. Appl. Phys.*, vol. 92, no. 6, pp. 2977–2979, 2002.

[12] H.-C. Lee and C.-K. Liu, "Si-based current-density-enhanced light emission and low-operating-voltage light-emitting/receiving designs," *Solid-State Electron.*, vol. 49, no. 7, pp. 1172–1178, 2005.

[13] E. D. Palik, *Handbook of Optical Constants of Solids*. Elsevier, 1998.

[14] W. L. Ng, "An efficient room-temperature silicon-based light-emitting diode," *Nature*, vol. 410, no. 6825, pp. 192–194, 2001.

[15] T. Hoang, P. LeMinh, J. Holleman, and J. Schmitz, "The effect of dislocation loops on the light emission of silicon LEDs," *IEEE Electron Device Lett.*, vol. 27, pp. 105–107, 2005.

[16] M. E. Castagna, S. Coffa, M. Monaco, A. Muscara, L. Caristia, S. Lorenti, and A. Messina, "High efficiency light emitting devices in silicon," *Mater. Sci. Eng., B*, vol. 105, no. 1-3, pp. 83–90, 2003.

[17] L. T. Canham, "Silicon quantum wire array fabrication by electrochemical and chemical dissolution of wafers," *Appl. Phys. Lett.*, vol. 57, no. 10, p. 1046, 1990.

[18] P. Fauchet, L. Tsybeskov, C. Peng, S. P. Duttagupta, J. Von Behren, Y. Kostoulas, J. M. V. Vandyshev, and K. Hirschman, "Light-emitting porous silicon: materials science, properties, and device applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 1, no. 4, pp. 1123–1139, 1995.

[19] P. Fauchet, "Progress toward nanoscale silicon light emitters," *IEEE J. Sel. Topics Quantum Electron.*, vol. 4, no. 6, pp. 1020–1028, 1998.

[20] P. Jaguiro, P. Katsuba, S. Lazarouk, and A. Smirnov, "Porous silicon avalanche LEDs and their applications in optoelectronics and information displays," *Acta Phys. Pol., A*, vol. 112, no. 5, pp. 1031–1036, 2007.

[21] L. Pavesi, L. Dal Negro, C. Mazzoleni, G. Franzo, and F. Priolo, "Optical gain in silicon nano-

crystals," *Nature*, vol. 408, no. 6811, pp. 440–444, 2000.

[22] R. J. Walters, J. Kalkman, A. Polman, H. A. Atwater, and M. J. A. de Dood, "Photoluminescence quantum efficiency of dense silicon nanocrystal ensembles in SiO_2," *Phys. Rev. B*, vol. 73, no. 13, p. 132302, 2006.

[23] Y. Kanemitsu, "Efficient light emission from crystalline and amorphous silicon nanostructures," *J. Lumin.*, vol. 100, no. 1-4, pp. 209–217, 2002.

[24] R. J. Walters, R. V. A. van Loon, I. Brunets, J. Schmitz, and A. Polman, "A silicon-based electrical source of surface plasmon polaritons," *Nature Materials*, vol. 9, no. 1, pp. 21–25, 2010.

[25] P. H. Lim, S. Park, Y. Ishikawa, and K. Wada, "Enhanced direct bandgap emission in germanium by micromechanical strain engineering," *Optics Express*, vol. 17, no. 18, pp. 16 358–16 365, 2009.

[26] T.-H. Cheng, C.-Y. Ko, C.-Y. Chen, K.-L. Peng, G.-L. Luo, C. W. Liu, and H.-H. Tseng, "Competitiveness between direct and indirect radiative transitions of Ge," *Appl. Phys. Lett.*, vol. 96, no. 9, p. 091105, 2010.

[27] X. Sun, J. Liu, L. Kimerling, and J. Michel, "Toward a Germanium laser for integrated silicon photonics," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 124–131, 2010.

[28] R. Newman, "Visible light from a silicon p-n junction," *Phys. Rev.*, vol. 100, no. 2, pp. 700–703, 1955.

[29] J. D. van Wyk, "Visible electroluminescence from the base-emitter p-n junction of an avalanching silicon transistor," *Solid-State Electron.*, vol. 8, no. 10, pp. 803–805, 1965.

[30] M. Herzog and F. Koch, "Hot-carrier light emission from silicon metal-oxide-semiconductor devices," *Appl. Phys. Lett.*, vol. 53, no. 26, pp. 2620–2622, December 1988.

[31] S. Tam and C. Hu, "Hot-electron-induced photon and photocarrier generation in silicon MOSFET's," *IEEE Trans. Electron Devices*, vol. 31, no. 9, pp. 1264–1273, September 1984.

[32] A. G. Chynoweth and K. G. McKay, "Photon emission from avalanche breakdown in silicon,"

Phys. Rev., vol. 102, no. 2, pp. 369–376, 1956.

[33] M. Kikuchi and K. Tachikawa, "Visible light emission from germanium diffused p-n junction," *J. Phys. Soc. Jpn*, vol. 14, no. 12, p. 1830, 1959.

[34] A. G. Chynoweth and H. K. Gummel, "Photon emission from avalanche breakdown in Germanium p-n junctions," *J. Phys. Chem. Solids*, vol. 16, no. 3-4, pp. 191–197, 1960.

[35] T. Figielski and A. Torun, "On the origin of light emitted from reverse-biased p-n junctions," in *Proc. 6th International Conference on the Physics of Semiconductors*. Exeter: Institute of Physics and The Physical Society, London (1962), 1962, pp. 863 – 868.

[36] J. Bude, N. Sano, and A. Yoshii, "Hot-carrier luminescence in Si," *Phys. Rev. B*, vol. 45, no. 11, pp. 5848–5856, 1992.

[37] A. Lacaita, F. Zappa, S. Bigliardi, and M. Manfredi, "On the bremsstrahlung origin of hot-carrier-induced photons in silicon devices," *IEEE Trans. Electron Devices*, vol. 40, no. 3, p. 577, 1993.

[38] L. Selmi, M. Mastrapasqua, D. Boulin, D. Bude, M. Manfredi, E. Sangiorgi, and M. Pinto, "Characterization and modeling of hot-carrier luminescence in silicon $n^{+}/n/n^{-}$ devices," in *International Electron Devices Meeting*, 1995.

[39] S. Villa, L. Lacaita, and A. Pacelli, "Photon emission from hot electrons in silicon," *Phys. Rev. B*, vol. 52, no. 15, pp. 10 993–10 999, 1995.

[40] A. T. Obeidat, Z. Kalayjian, A. G. Andreou, and J. B. Khurgin, "A model for visible photon emission from reverse-biased silicon p-n junctions," *Appl. Phys. Lett.*, vol. 70, no. 4, pp. 470–471, 1997.

[41] N. Akil, S. Kerns, D. K. Jr., A. Hoffmann, and J.-P. Charles, "A multimechanism model for photon generation by silicon junctions in avalanche breakdown," *IEEE Trans. Electron Devices*, vol. 46, no. 5, pp. 1022–1028, 1999.

[42] M. Lahbabi, A. Ahaitouf, E. Abarkan, M. Fliyou, A. Hoffmann, J.-P. Charles, B. L. Bhuva, S. E. Kerns, and D. V. Kerns, "Analyses of electroluminescence spectra of silicon junctions

in avalanche breakdown using an indirect interband recombination model," *Appl. Phys. Lett.*, vol. 77, no. 20, pp. 3182–3184, 2000.

[43] A. Chatterjee, B. Bhuva, and R. Schrimpf, "High-speed light modulation in avalanche breakdown mode for Si diodes," *IEEE Electron Device Lett.*, vol. 25, no. 9, pp. 628–630, 2004.

[44] M. E. Goosen, P. J. Venter, M. du Plessis, A. W.Bogaleck, A. C. Alberts, and P. Rademeyer, "A high-speed 0.35 μm CMOS optical communication link," in *Proc. SPIE 8267, Optoelectronic Interconnects XII, 826716*. SPIE Digital Library, 2012.

[45] M. de la Bardonnie, D. Jiang, S. Kerns, J. Kerns, D.V., P. Mialhe, J. P. Charles, and A. Hoffman, "On the aging of avalanche light emission from silicon junctions," *IEEE Trans. Electron Devices*, vol. 46, no. 6, pp. 1234–1239, 1999.

[46] A. Chatterjee and B. Bhuva, "Accelerated stressing and degradation mechanisms for Si-based photoemitters," *IEEE Trans. Dev. Mat. Rel.*, vol. 2, no. 3, pp. 60–64, 2002.

[47] H.-D. Chen, H.-J. Liu, B.-J. Huang, M. Gu, and J.-B. Liu, "Silicon optical emitter in CMOS technology," in *8th International Conference on Solid-State and Integrated Circuit Technology*, 2006, pp. 278–280.

[48] M. du Plessis and P. Rademeyer, "Novel electroluminescence technique to analyze mixed reverse breakdown phenomena in silicon diodes," *Solid-State Electron.*, vol. 54, no. 4, pp. 433 – 438, 2010.

[49] L. W. Snyman, H. Aharoni, and M. du Plessis, "A dependency of quantum efficiency of silicon CMOS n+pp+ LEDs on current density," *IEEE Photonics Technol. Lett.*, vol. 17, no. 10, pp. 2041–2043, 2005.

[50] A. W. Bogalecki, M. du Plessis, P. J. Venter, M. E. Goosen, and I. J. Nell, "Integrated optical light directing structures in CMOS to improve light extraction efficiency," in *Proc. 22nd IEEE International Conference on Microelectronics*. Egypt, Cairo: IEEE, December 2010.

[51] K. Xu and G. P. Li, "A novel way to improve the quantum efficiency of silicon light-emitting diode in a standard silicon complementary metal–oxide–semiconductor technology," *J. Appl. Phys.*, vol. 113, no. 10, p. 103106, 2013.

[52] M. du Plessis, H. Aharoni, and L. W. Snyman, "A silicon transconductance light emitting device (TRANSLED)," *Sens. Actuators, A*, vol. 80, no. 3, pp. 242–248, 2000.

[53] M. du Plessis, H. Aharoni, and L. W. Snyman, "Silicon light emitting devices in standard CMOS technology," in *Semiconductor Conference, 2001. CAS 2001 Proceedings. International*, vol. 1, 2001, pp. 231–238.

[54] M. du Plessis, H. Aharoni, and L. W. Snyman, "Silicon LEDs fabricated in standard VLSI technology as components for all silicon monolithic integrated optoelectronic systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 8, no. 6, pp. 1412–1419, 2002.

[55] M. du Plessis, H. Aharoni, and L. W. Snyman, "Two- and multi-terminal CMOS/BiCMOS Si LED's," *Opt. Mater.*, vol. 27, no. 5, pp. 1059–1063, 2005.

[56] W. Wang, B.-J. Huang, Z. Dong, X. Zhang, N. Guan, J. Chen, Y. Gui, and H.-D. Chen, "A low-voltage two-wavelength light emitter in standard CMOS technology," in *Proc. 7th IEEE International Conference on Group IV Photonics*, September 2010, pp. 111–113.

[57] W. Wang, B.-J. Huang, Z. Dong, H.-J. Liu, X. Zhang, N. Guan, J. Chen, W.-L. Guo, P.-J. Niu, and H.-D. Chen, "A low-voltage silicon light emitting device in standard salicide CMOS technology," *Chinese Physics Letters*, vol. 27, no. 4, p. 048501, 2010. [Online]. Available: http://stacks.iop.org/0256-307X/27/i=4/a=048501

[58] A. R. Chen, A. I. Akinwande, and H.-S. Lee, "CMOS-based microdisplay with calibrated backplane," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2746–2755, December 2005.

[59] Z. Dong, W. Wang, B. Huang, X. Zhang, N. Guan, and H. Chen, "A 630dpi dynamic LED display array in standard Si-based CMOS technology," *Science China Information Sciences*, vol. 55, no. 10, pp. 2409–2416, 2012.

[60] P. J. Venter, A. W. Bogalecki, M. du Plessis, M. E. Goosen, I. J. Nell, and P. Rademeyer, "CMOS dot matrix microdisplay," in *Proc. SPIE 7956, Advances in Display Technologies and E-papers and Flexible Displays, 79560Y*. SPIE Digital Library, 2011.

[61] P. J. Venter, M. du Plessis, A. W. Bogalecki, M. E. Goosen, and P. Rademeyer, "An 8 x 64 pixel dot matrix microdisplay in 0.35-μm complementary metal-oxide semiconductor technology,"

Optical Engineering, vol. 51, p. 014003, February 2012.

[62] L. Selmi, "Silicon luminescence techniques for the characterization of hot-carrier and degradation phenomena in MOS devices," *Microelectron. Eng.*, vol. 28, no. 1-4, pp. 249–256, 1995.

[63] M. M. Manus, J. Kash, S. Steen, S. Polonsky, J. Tsang, D. Knebel, and W. Huott, "PICA: Backside failure analysis of CMOS circuits using Picosecond Imaging Circuit Analysis," *Microelectronics Reliability*, vol. 40, no. 8-10, pp. 1353–1358, 2000.

[64] D. Nedospasov, J.-P. Seifert, A. Schlosser, and S. Orlic, "Functional integrated circuit analysis," in *IEEE International Symposium on Hardware-Oriented Security and Trust*, 2012, pp. 102–107.

[65] S. Sayil, D. Kerns, and S. Kerns, "All-silicon optical contactless testing of integrated circuits," *International Journal of Electronics*, vol. 89, no. 7, pp. 537–547, 2002.

[66] S. Sayil, "Optical contactless probing: an all-silicon, fully optical approach," *IEEE Des. Test Comput.*, vol. 23, no. 2, pp. 138–146, 2006.

[67] S. P. Duttagupta, K. D. Hirschman, L. Tsybeskov, and P. M. Fauchet, "Silicon-based visible light-emitting devices integrated into microelectronic circuits," *Nature*, vol. 384, no. 6607, pp. 338–341, 1996.

[68] M. Lundstrom, *Fundamentals of carrier transport*, 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2000.

[69] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, 1984.

[70] R. F. Pierret, *Advanced semiconductor fundamentals*, ser. Modular series on solid state devices. Addison-Wesley, 1987.

[71] S. M. Sze and K. N. Kwok, *Physics of semiconductor devices*, 3rd ed. New Jersey, United States of America: John Wiley & Sons, Inc., 2006.

[72] S. Datta, *Quantum phenomena*, ser. Modular series on solid state devices, R. F. Pierret and G. W. Neudeck, Eds. Addison-Wesley, 1989, vol. VIII.

[73] W. Maes, K. D. Meyer, and R. V. Overstraeten, “Impact ionization in silicon: A review and update,” *Solid-State Electron.*, vol. 33, no. 6, pp. 705–718, 1990.

[74] J. Bude and K. Hess, “Thresholds of impact ionization in semiconductors,” *J. Appl. Phys.*, vol. 72, no. 8, pp. 3554–3561, 1992.

[75] A. G. Chynoweth and K. G. McKay, “Internal field emission in silicon p-n junctions,” *Phys. Rev.*, vol. 106, no. 3, pp. 418–426, 1957.

[76] A. G. Chynoweth, W. L. Feldmann, C. A. Lee, R. A. Logan, G. L. Pearson, and P. Aigrain, “Internal field emission at narrow silicon and germanium $p-n$ junctions,” *Phys. Rev.*, vol. 118, no. 2, pp. 425–434, 1960.

[77] C. R. Crowell and S. M. Sze, “Temperature dependence of avalanche multiplication in semiconductors,” *Appl. Phys. Lett.*, vol. 9, p. 242, 1966.

[78] W. Shockley and W. T. Read, “Statistics of the recombinations of holes and electrons,” *Phys. Rev.*, vol. 87, no. 5, pp. 835–842, 1952.

[79] J. Shewchun and L. Y. Wei, “Mechanism for reverse-biased breakdown radiation in p-n junctions,” *Solid-State Electron.*, vol. 8, no. 5, pp. 485–493, 1965.

[80] H. Veendrick, *Nanometer CMOS ICs: from Basics to ASICs*. Springer, 2008.

[81] G. S. May and S. M. Sze, *Fundamentals of semiconductor fabrication*. Wiley, 2004.

[82] S. M. Sze and G. Gibbons, “Effect of junction curvature on breakdown voltage in semiconductors,” *Solid-State Electron.*, vol. 9, no. 9, pp. 831–845, 1966.

[83] S. K. Ghandhi, *Semiconductor Power Devices*, S. K. Ghandhi, Ed. Wiley, New York, 1977.

[84] C. Y. Chang, S. S. Chiu, and L. P. Hsu, “Temperature dependence of breakdown voltage in silicon abrupt p-n junctions,” *IEEE Trans. on Electron Devices*, vol. 18, no. 6, pp. 391–393, June 1971.

[85] S. Lombardo, J. H. Stathis, B. P. Linder, K. L. Pey, F. Palumbo, and C. H. Tung, “Dielectric breakdown mechanisms in gate oxides,” *J. Appl. Phys.*, vol. 98, no. 12, pp. 121 301–12 301–36,

2005.

[86] J. Verweij and J. Klootwijk, “Dielectric breakdown i: A review of oxide breakdown,” *Microelectron. Eng.*, vol. 27, no. 7, pp. 611–622, October 1996.

[87] J. R. Black, “Electromigration - a brief survey and some recent results,” *IEEE Trans. Electron Devices*, vol. 16, no. 4, pp. 338–347, April 1969.

[88] R. Dennard, F. Gaensslen, V. Rideout, E. Bassous, and A. LeBlanc, “Design of ion-implanted MOSFET’s with very small physical dimensions,” *IEEE J. of Solid-State Circuits*, vol. 9, no. 5, pp. 256 – 268, October 1974.

[89] B. Razavi, *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, 2000.

[90] P. Smeys, “Geometry and stress effects in scaled integrated circuit isolation technologies,” Ph.D. dissertation, Stanford University, 1996.

[91] J. Appels, E. Kooi, M. M. Paffen, J. J. H. Schatorje, and W. H. C. G. Verkuylen, “Local oxidation of silicon and its application in semiconductor-device technology,” *Philips Research Reports*, vol. 25, no. 2, pp. 118–132, 1970.

[92] G. Burton, P. Tuntasood, F. Chien, R. Kovacs, and M. Vora, “New techniques for elimination of the bird’s head and bird’s beak,” in *International Electron Devices Meeting*, vol. 30, 1984, pp. 582 – 585.

[93] Y. P. Han and B. Ma, “Isolation process using polysilicon buffer layer for scaled MOS/VLSI,” *J. Electrochem. Soc.*, vol. 131, no. 3, p. 98, 1984.

[94] J.-P. Mieville, R. Rooyackers, and L. Deferm, “An optimized poly-buffered LOCOS process for a 0.35 μm CMOS technology,” in *Proc. 24th European Solid State Device Research Conference*, Scotland, Sept. 1994, pp. 199 – 202.

[95] K. Chiu, J. Moll, K. Cham, J. Lin, C. Lage, S. Angelos, and R. Tillman, “The sloped-wall SWAMI - a defect-free zero bird’s-beak local oxidation process for scaled vlsi technology,” *IEEE Trans. Electron Devices*, vol. 30, no. 11, pp. 1506 – 1511, nov 1983.

[96] D. Ahn, S. Ahn, P. Griffin, M. Hwang, W. Lee, S. Ahn, C. Hwang, and M. Lee, “A highly prac-

tical modified LOCOS isolation technology for the 256 Mbit DRAM," in *International Electron Devices Meeting*, dec 1994, pp. 679 –682.

[97] J. Pfiester, P. Kenkare, R. Subrahmanyan, J.-H. Lin, and P. Crabtree, "Nitride-clad LOCOS isolation for 0.25 μm CMOS," in *Symposium on VLSI Technology*, 1993, pp. 139 –140.

[98] M. Nandakumar, A. Chatterjee, S. Sridhar, K. Joyner, M. Rodder, and I.-C. Chen, "Shallow trench isolation for advanced ULSI CMOS technologies," in *International Electron Devices Meeting*, December 1998, pp. 133 –136.

[99] P. J. Venter, A. C. Alberts, M. du Plessis, T.-H. Joubert, M. E. Goosen, C. Janse van Rensburg, P. Rademeyer, and N. M. Fauré, "A CMOS microdisplay with integrated controller utilizing improved silicon hot carrier luminescent light sources," in *Proc. SPIE 8643, Advances in Display Technologies III, 864309*. ©SPIE Digital Library, 2013.

APPENDIX A

RADIOMETRY

A.1 RADIOMETER

The radiometer system that was used in characterising the light emission from the chip surface uses a silicon detector (International Light SED100) in conjunction with an integrating measurement and display unit (International Light IL1700 research radiometer). The measurement unit outputs the detector current exactly as is if the correction factor is set to zero. Since the majority of radiometer measurements are tailored for narrow wavelength lasers with a calibration factor at a specific wavelength, all measurements above were done with a calibration factor of zero for raw detector current output. The SED100 detector response is shown in figure A.2.

Figure A.1: International light IL1700 radiometer

Although the radiometer detector gives an output as a current, it is important to have a power density spectrum of the light as detected by the detector in order to be able to quantify optical power. The

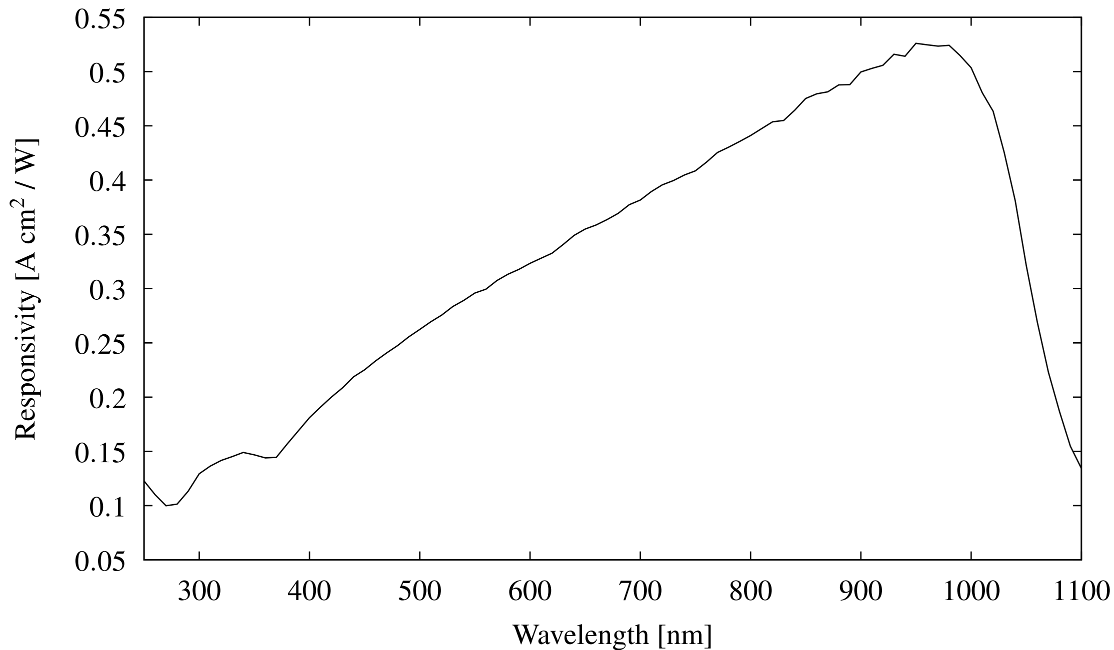


Figure A.2: SED100 detector responsivity as a function of wavelength

radiometer system was calibrated recently in 2013 and the detector results in figure A.2 is representative.

The IL1700 allows for an analogue output at the back of the device to be used as being representative of the detected current. This enables a direct connection to a parameter analyser where the device current can be forced, the voltage over the device's terminals can be measured and a representation of the radiometer detector current can be captured for each bias point of the light emitting device. This is usually the first characterisation step as both the electrical as well as optical characterisation are completed simultaneously.

A.2 PARAMETER ANALYSIS

The parameter analyser used for all measurements was calibrated in 2013 and is used for analysing the electrical behaviour of devices by forcing currents and sensing voltages on the terminals where the signals are applied. The analyser used in this thesis is the Agilent / HP4155B Parameter Analyser with 4 SMUs and current capabilities of up to 100 mA per SMU.

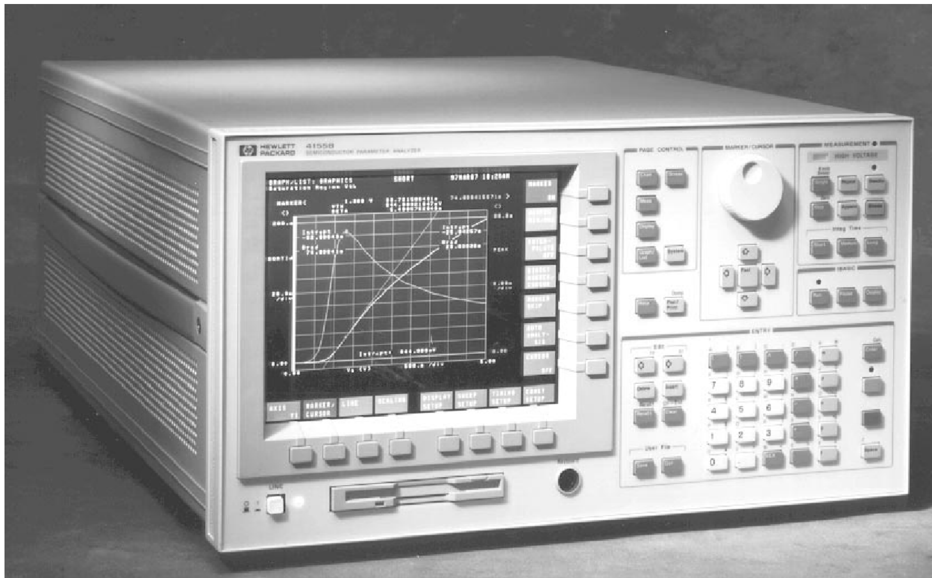


Figure A.3: Agilent / HP 4155B parameter analyser

A.3 SPECTRAL ANALYSIS

Figure A.4: Avantes AvaSpec-2048TEC

Spectral analysis was done on an Avantes AvaSpec-2048TEC spectrometer with a silicon-based CCD array containing 2048 pixels. The device is temperature compensated for reduced noise. A set of calibration data was constructed using a monochromator and a calibrated radiometer in order to quantify the spectrometer response. Although this dataset is calibrated in a relative way, positioning of the light

source in front of the slit makes an absolute measurement of radiated optical power impossible.

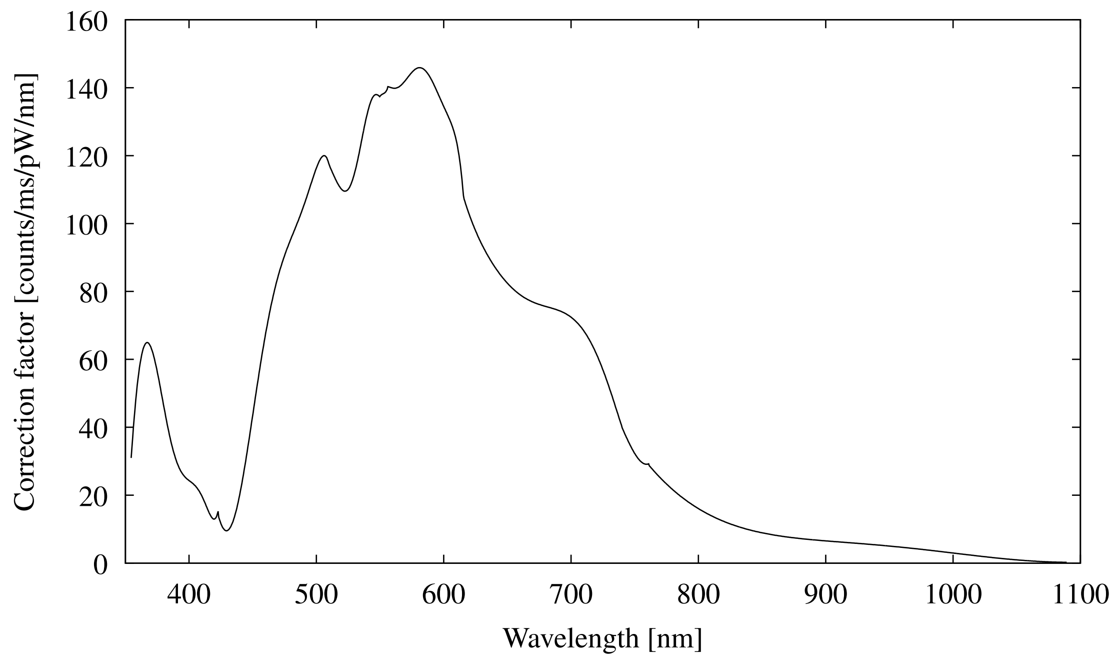


Figure A.5: Correction factors used for calibrating the spectrometer response

The spectrometer output consists of a number of integrated counts over a specific integration time for a bin centred at a wavelength. The power spectral density is obtained by weighting the recorded spectrum with the correction factors to result in a relative power spectral density.

A.4 METHOD FOR CALCULATING RADIANT FLUX

In order to obtain the power emitted by a light source, the radiometer current is recorded at a specific current, while the spectrum of the emitted light is taken at the same current. The assumption is that the spectral response does not vary significantly as a function of different angles in the radiation pattern. This assumption is necessary since the radiometer detector captures practically the entire radiation pattern of radiant flux, while the spectrometer slit is limited to a very narrow part of the radiated flux directly normal to the chip surface. The following procedure is used to calculate the absolute optical power emitted by a light source.

1. The radiometer detector current is recorded, along with the electrical voltage and current, at a specific bias point. This results in a detector current I_d.

2. The spectrum is taken at the same bias point by placing the IC directly in front of the spectrometer slit. A raw spectrum is therefore generated.

3. The raw spectrum is corrected using the correction data in figure A.5 in order to produce a relative power density spectrum.

4. The relative power density spectrum is weighted by the calibrated SED100 response as shown in figure A.2 and integrated over all wavelengths in order to produce a scaled detector current that will be produced if a light source with the relative spectrum was to be measured by the detector.

5. By taking the actual current measurement from the radiometer detector and the integrated current produced by the previous step, a scaling factor is derived.

6. The scaling factor is applied to the relative power density spectrum in order to produce an absolute power density spectrum for the light source.

7. Radiant flux, or optical power emitted, can then be found by integrating the absolute power spectral density over all wavelengths to produce a quantity in Watts of optical power.

A.5 CONCLUSION

Measurement of light in the nanowatt range is a complicated task and, while the above procedure produces a good estimate, the technique needs to be refined by including the knowledge of the spectrum at different angles of emission with respect to the chip surface. This task is non-trivial and offers an opportunity for further investigation on how to measure the optical power of wide spectrum, low level light sources. The procedure above, however, is a best-effort approach which, according to the author, seems to be the most extensive yet when compared to published methods by other groups in order to quantify light source performance metrics such as external quantum efficiency and external power efficiency.