South African Conference on Semi and Superconductor Technology

Spier Wine Estate, Stellenbosch, South Africa

8-9 April 2009

National Research Foundation

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Cape Peninsula University of Technology

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ISBN: 978-0-620-43865-0

FOREWORD

It is my pleasure to welcome you to the first South African Conference on Semi- and Superconductor Technology (SACSST 2009). The aim of SACSST 2009 is to establish a forum for the exchange of in-depth ideas for South African researchers working in the relatively small fields of semiconductor and superconductor technologies.

The conference intends to provide a platform for industry, universities and research institutions to report on their research activities. It is envisaged that the conference will help to establish cooperation amongst researchers. It is therefore our desire that you, the participants, will get acquainted with the research activities in your field of research, and share technical information and problems with your peers.

I would like to thank the host institutions, the Cape Peninsula University of Technology, and the Universities of Stellenbosch and Pretoria, for making this event possible. I also wish to acknowledge the NRF for their financial support.

- Prof. Willem Perold, Conference Chair

ORGANISING COMMITTEE

Conference Chair:	Prof. Willem Perold
Technical Chair (Superconductor Technology):	Prof. Willem Perold
Technical Chair (Semiconductor Technology):	Prof. Monuko du Plessis
Programme Chair:	Prof. Robert van Zyl
Local Arrangements:	Dr Coenrad Fourie
Secretariat:	Mrs Hannelie van Wyk , Mr Ian van Zyl

TECHNICAL PROGRAMME COMMITTEE

CHAIRS

Prof. WJ Perold	University of Stellenbosch, Stellenbosch, South Africa
Prof. M du Plessis	University of Pretoria, Pretoria, South Africa

REVIEWERS

Prof. Lukas Snyman	Tshwane University of Technology, Pretoria, South Africa
Dr Trudi Joubert	Solid State Technology cc, Pretoria, South Africa
Prof. J.R. Botha	Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
Prof. Japie Engelbrecht	Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
Dr Saurabh Sinha	University of Pretoria, South Africa
Dr Walter Meyer	University of Pretoria, South Africa
Dr Christo Schutte	Detek, Denel, Pretoria, South Africa
Dr Pieter Rademeyer	InSiAva (Pty) Ltd, Pretoria, South Africa
Mr S.J. de Beer	SAMES, Pretoria, South Africa
Prof. Igor Krylov	University of the Western Cape, Cape Town, South Africa
Dr Sylvain Halindintwali	University of the Western Cape, Cape Town, South Africa
Mr Retief Gerber	University of Stellenbosch, Stellenbosch, South Africa
Dr Coenrad Fourie	University of Stellenbosch, Stellenbosch, South Africa
Mr Graham Hardie	University of Stellenbosch, Stellenbosch, South Africa
Dr Srinivasu Vallabhapurapu	National Centre for Nano-structured Materials, CSIR, Pretoria, South Africa
Dr Leigh Jarvis	University of KwaZulu-Natal, Durban, South Africa

REVIEW PROCESS

All refereed papers were refereed by a double blind reviewing process according to the Department of Education's guidelines. (In a double-blind review process the identities of both the authors and the reviewers are concealed. Each paper has been reviewed by two independent reviewers.) A double-blind review by peers of full papers has been followed. Papers were reviewed according to the following criteria: relevance to conference themes, relevance to audience, contribution to scholarship, standard of writing, originality and critical analysis.

CONFERENCE PROGRAMME

SOUTH AFRICAN CONFERENCE ON SEMI- AND SUPERCONDUCTOR TECHNOLOGY 2009

Spier Wine Estate, Stellenbosch
8 – 9 April 2009

Programme

	Tuesday, 7 April 2009	
	Neethlingshof Wine Estate	
19:00	**IEEE's 125th anniversary celebration**	
	Hosted by Prof Monuko du Plessis and Dr Saurabh Sinha	
	Wednesday, 8 April 2009	
	Spier Wine Estate	
8:00	**Registration, tea and coffee**	
9:00	**Opening, welcome and house rules**	**Plenaries**
	Prof Willem Perold (Univ Stellenbosch) - Conference Chair	
	Dr Thandi Mgwebi (NRF) - Programme Director: Human Capacity Programme	
	Prof Robert van Zyl (CPUT) - Programnme Chair	
9:20	**A South African Perspective**	
	Prof Monuko du Plessis (semiconductor technology)	
	Prof Willem Perold (superconductor technology)	Chair: Prof Robert van Zyl
9:40	[35] *The growth mechanism and photonic properties of 1D and 2D ZnO nanostructures by low temperature method* - Sithole MJ	**Session 1: Posters in the foyer**
	[38] *The capability at Stellenbosch University for the fabrication of micron and submicron structures* - Buttner U	
	[22] *A SONET OC-48 CMOS based integrated optical transceiver* - Kalombo TY	
	[03] *Effects of grain size on mechanical properties of nanostructured copper alloy by Severe Plastic Deformation (SPD) Process*	
	[11] *3D CMOS Photodiode modelling and simulator design using the Finite Element Method* - Ellinghaus P	
	[16] *Electrochromic and photochromic behaviour of electrospun WO3 nanofibres deposited on TCO substrates* - Sone BT	Chair: Prof Robert van Zyl
10:40	**Tea**	
11:00	[19] *A CMOS based multiple-access DSSS transceiver* - Sinha S	**Session 2: HF Microelectronics**
11:20	[12] *Impact of future CMOS scaling on power consumption: Electrical versus optical clock distribution networks* - Venter PJ	
11:40	[2] *Design methodology for SiGe-based Class E power amplifier* - Božanić M	
12:00	[20] *Adaptive FIR filter pre-emphasis for high speed serial links* - Goosen ME	
12:20	[24] *Phase noise reduction of a 0.35 µm BiCMOS SiGe 5 GHz voltage controlled oscillator* - Lambrechts JW	
12:40	[23] *On-chip impedance tuning for mm-wave applications* - Mabuza BC	Chair: Dr Trudi Joubert
13:00	**Lunch**	
14:00	[25] *A novel electroluminescence technique to analyse mixed field emission and impact ionisation reverse breakdown* - du Pless	**Session 3: Electron Devices / Nanostructures**
14:20	[invited] *Optoelectronic Multifunctionality in VO2 based nano-systems* - Maaza M	
14:40	[10] *Microbolometer model and characterisation* - Swart MJ	
15:00	[37] *A CMOS based multiplexer design of a 16 x 16 read-out array for infrared microbolometer sensors* - Visser PJ	
15:20	[8] *Nanostructured photovoltaic cells: Using AFM and QCM to produce accurate film thicknesses* - Botha AF	
15:40	[41] *Optimisation strategies for GaAs mm-Wave Gunn diodes* - van Zyl RR	Chair: Prof Willie Perold
16:00	**Tea**	
16:20	[04] *Overview of superconducting digital electronics* - Taylor J	**Session 4: Superconductor Technology**
16:40	[06] *FEM solution using vector elements for a superconducting microstrip line* - Young A	
17:00	[07] *AFM plough YBCO microbridges: substrate effects* - Elkaseh AAO	
17:20	[13] *Superconducting sensors and imagers* - Northeast D	
17:40	[17] *In situ superconducting MgB2 thin films grown by combined pulsed laser deposition and thermal evaporation* - Hardie GL	
18:00	[29] *PBCO/YBCO bilayer growth and optimization for the fabrication of buffered step-edge Josephson Junctions* - Buttner U	Chair: Prof Willie Perold
19:00	**Conference Dinner at Moyo**	

	Thursday, 9 April 2009	
	Spier Wine Estate	
8:00	**Tea, coffee**	
8:30	**Plenary Session: Prof Cor Claeys**	**Plenary**
	Trends and Challenges in More Moore and More than Moore Research	Chair: Prof Monuko du Plessis
9:30	**Short break**	
9:40	[1] *A 384x288 pixel CMOS image readout chip* - Joubert T	**Session 5: IC Simulation, Layout**
10:00	[30] *A CMOS implementation of Jakes Fading Channel* - Buckle JP	
10:20	[36] *High speed, low power CMOS optical receiver front-end* - Janse van Rensburg C	
10:40	[26] *Expanded adaptive LMS algorithm for the removal of interference in capacitive sensing devices* - van der Merwe SJ	
11:00	[21] *Mathematical analysis of input matching techniques for application in wide-band LNA design* - Weststrate M	
11:20	[28] *Development of a next-generation commercial computer aided design (CAD) tool* - Pool J	Chair: Prof Lukas Snyman
11:40	**Tea**	
12:00	[09] *Design and manufacture of nanometre-scale SOI light sources* - Bogalecki AW	**Session 6: Photonics / Optics / Physics**
12:20	[39] *Photonic transitions (1.4eV-2.8eV) in Silicon p+np+ injection-avalanche CMOS LEDs as function of depletion layer profiling and defect engineering* - Snyman LW	
12:40	[40] *Simulation of Si LED (450nm - 750nm) light propagation phenomena in CMOS integrated circuitry for MOEMS applications*	
13:00	[15] *MOVPE growth and characterisation of III-V Sb-based quantum well structures* - Wagener V	
13:20	[27] *Hall and thermoelectric evaluation of narrow gap semiconductors* - Wagener MC	Chair: Dr Saurabh Sinha
13:40	**Lunch / prizegiving / Closing** - Prof Monuko du Plessis	

TABLE OF CONTENTS

Session 6

Trends and Challenges in More Moore and More than Moore Research

Cor Claeys, *Fellow, IEEE*

***Abstract*—The Strategic Research Agenda for micro- and nano-electronics contains three main domains, i.e., More Moore, Beyond CMOS and More than Moore. For each of them some general trends and challenges are addressed. The convergence of the top-down technology, with bottom-up methods derived from fundamental disciplines such as materials physics, chemistry and biotechnology is opening a totally new world of applications. The diversity of the 'More than Moore' domain, includes applications related to ambient intelligence, automotives, molecular electronics, nano-biotechnology, polymer electronics, sustainable energy based on photovoltaic cells and human healthcare.**

***Index Terms*— Scaling, Moore's law, heterogeneous integration, packaging, healthcare, automotives**

I. Introduction

THE discovery of the semiconductor transistor in 1948 by Shockley, Bardeen and Brattain [1], followed by the invention of the planar technology [2] leading to the fabrication of integrated circuits, form the cornerstones of the present-day microelectronics industry with a turnover of more than 320 billion US$, not including the nearly 100 billion US$ associated with semiconductor equipment and materials. One of the driving forces has been the device scaling in order to achieve a higher packing density and an improved device performance at a lower cost/function. Today, 45 nm technology nodes are manufactured on 300 mm silicon wafers, while research is focusing already on 32 nm and below. Several emerging technologies are being explored.

Microelectronics products have enabled the internet revolution, provide global communication and are at the basis of most of the consumer products. Typical examples are smartphone, mobile internet devices, netbook, MEMS applications, healthcare, etc. Semiconductors will remain playing an essential role in the different application fields related to societal needs, as illustrated in Fig. 1. This figure indicates the strong need of electronic-based solutions. It can even be stated that there are no innovations or innovative solutions without micro- and nano-electronics.

Key fields such as healthcare, transport, security, energy, communication and infotainment are gaining more and more market and becoming a part of everyday life. The micro–electronic content and the type of devices and systems needed, are depending on the envisaged application field. In many cases the required functionality will be higher than what can be offered by CMOS alone, so that one has to enter the field of heterogeneous integration.

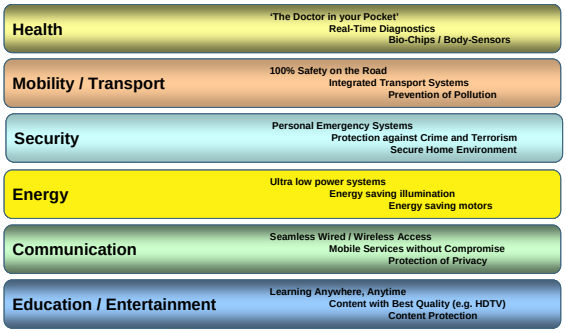


Fig. 1. Illustration of the importance of microelectronics in the key domains related to present and future societal needs.

The Strategic Research Agenda is generally divided into three main domains, i.e., More Moore, Beyond CMOS, and More than Moore (MtM), as shown in Fig. 2. The first one is dominated by device scaling, requiring the implementation of new materials and advanced process modules, and the use of alternative device architectures. Therefore, collaborative research programs involving many partners with large human resources are a prerequisite in order to reduce the total cost and to ensure to be ready on time. The digital content is important and may lead to System-on-Chip solutions.

More Moore needs a further scaling of the device dimensions, which leads to the transition from micro- to nano-electronics, and is since the 1960-ties performed according to Moore's law [4], as will be further discussed in section II. Physical, technological and also economical limitations will determine how far scaling can be pushed in a manufacturing environment. Only looking at physical limitations, devices below 5 nm will be the ultimate limit as we are entering the atomic scale of the Si atoms. For the so-called 'Beyond CMOS' domain, device operation is no longer only based on electron transport, but rather magnetic spin, flux or quanta. This domain covers topics such as spintronics, quantum computing, molecular electronics, carbon based devices, etc.

Manuscript received March 9, 2009.

C. Claeys is with the Interuniversity Microelectronics Center (IMEC), 3001 Leuven, Belgium, and also with the Electrical Engineering Department, Katholieke Universiteit Leuven, 3001 Leuven, Belgium.

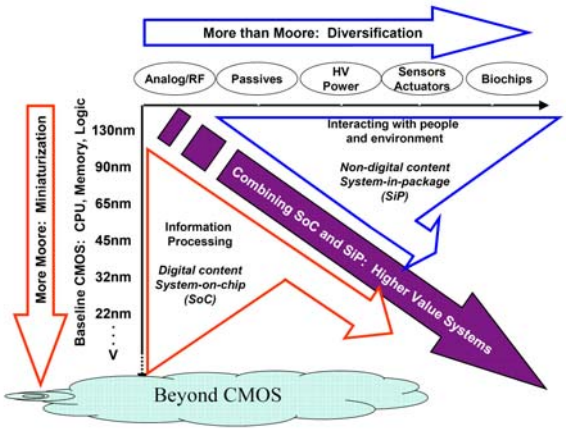


Fig. 2. The three domains of the Strategic Research Agenda [3].

For the More than Moore domain the digital content is less important and the effort is towards increasing the functionality enabling going over to System-in-Package (SiP) applications. Of key importance is the interaction of the device with the environment and with people. This domain needs research efforts related to heterogeneous integration, 3D packaging, etc and has numerous applications in fields such as automotive, ambient intelligence, healthcare, domotics, etc. There is a large diversity of technologies under study, such as e.g. analog/RF, on-chip integration of passive elements, sensors and actuators, biochips, etc.

Trends and research challenges for these three domains are addressed and will be illustrated by typical examples related to research going on at IMEC

II. DEVICE SCALING - MORE MOORE

The device scaling trend is performed in agreement with the well known Moore's law, first postulated in middle of the 60-ties and illustrated in Fig. 3. It states that the packing density is doubling every two years, necessitating the introduction of a new technology node. The additional challenge is that the cost/function has to decrease.

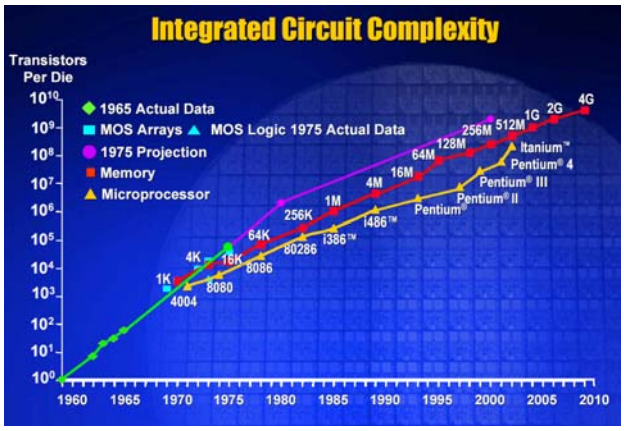


Fig. 3. Moore's law showing the evolution of the transistor packing density as a function of time, requiring the reduction of the device geometries and the introduction of advanced process modules.

The technological research is performed according to the International Technology Roadmap for Semiconductors (ITRS) [5], which determines the technological innovations that are needed in order to be ready on time to introduce in manufacturing the next technology generation. During the last decade extensive research has been carried out to introduce new processing modules such as e.g. silicides, ultra shallow junctions, alternative metallization systems, high-κ gate dielectrics, low-κ dielectrics for intermetal isolation, metal gates to replace polysilicon, Cu-based metallization systems, strain engineering, etc. Today, 45 nm technologies are in manufacturing, while industrial research is oriented towards 32 and 22 nm. The translation of the roadmap towards new processing modules and the introduction of new materials is shown in Fig. 4.

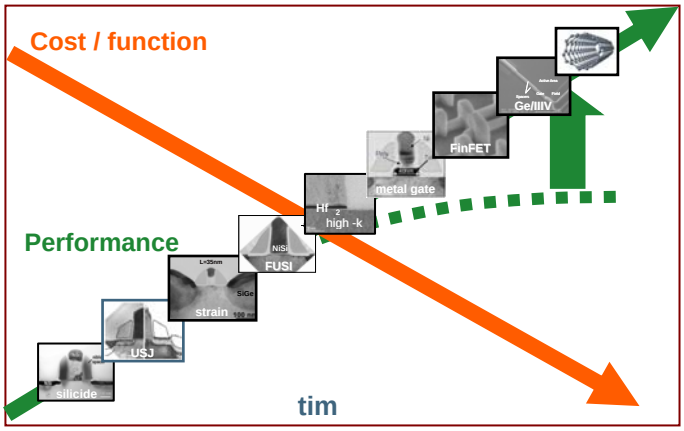


Fig. 4. To follow Moore's law it is essential to introduce new materials and new process modules (courtesy IMEC).

The feasibility of MOSFETs with 5 nn gate length has already been demonstrated as early as 2003 [6]. The thickness of the dielectric layers involved is reaching the atomic levels as illustrated in Fig. 5 in the case of ultra-thin gate oxides [7]. A 1.2 nm SiO_2 layer corresponds with the distance of only 4 Si atoms. To overcome problems with gate tunneling for SiO_2 layers smaller than 3 nm, which would strongly increase the power consumption, extensive research has been devoted to introduce gate materials with a high dielectric constant [8]. In present-day industrial processes a preference is given to Hf-based dielectrics. However, the gate stack has to include an interfacial SiO_2 layer in order to optimize the interface quality. For Ge devices the gate stack even consists of a $Si/SiO_2/HfO_2$ sandwich. TEM images of these systems are also shown in Fig. 5. In case of Ge, it is essential to optimize the thickness of the Si layer, which is in the range of 5 to 10 monolayers.

The introduction of high-κ dielectrics reduces the carrier mobility compared to the SiO_2/polysilicon system due to remote phonon scattering effects, which can partially be overcome by replacing the polysilicon electrode by a fully silicided (FUSI) or metal gate.

To further boost up the carrier mobility a variety of strain engineering approaches can be used, based on either global wafer-level or local, device-level techniques. The global approach generally results in biaxial stress in the transport

plane, while local stressors generate uniaxial stress in the channel direction. The latter is usually more effective in translating strain into performance enhancement for short channel transistors [9].

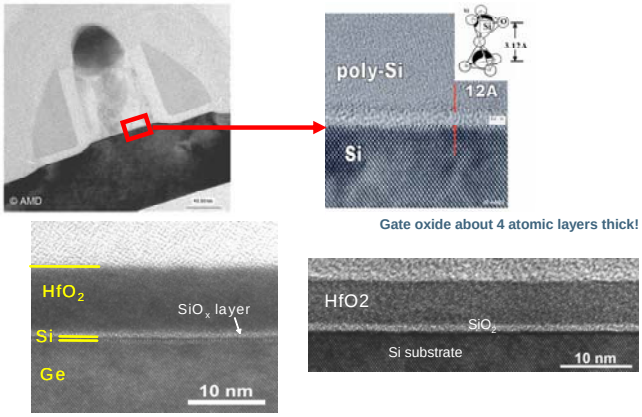


Fig. 5. Illustration of thin gate dielectrics, (a) SEM of a transistor cross-section, (b) 1.2 nm SiO_2 with a poly-Si gate, (c) HfO_2 on a SiO_2 interface layer and (d) gate sandwich of $Si/SiO_2/HfO_2$ on a Ge substrate.

The global strain approach is using so-called high-mobility substrates based on strain-engineering. A variety of substrates such as SiGe channels, strained Si (sSi) on strain relaxed buffer layers, Ge, and even strained Ge (sGe) are being studied and used already. The state of the art in Ge processing has been reviewed [10], pointing out that it is possible to fabricate good performing p-channel MOSFETs. Recently, a record on/off current ratio (I_{on}=478 µA/µm and I_{off}=37 nA/µm) has been reported for 65 nm Ge nMOSFETs [11]. This is illustrated in Fig. 6.

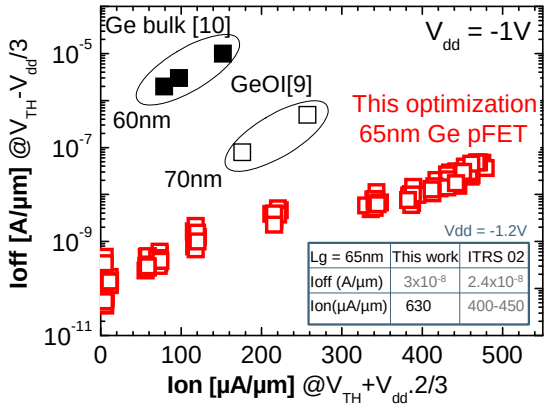


Fig. 6. Record performance of a 65 nm pMOSFET in a Ge technology [11]

The Ge n-channel counterparts are more troublesome because of the interface quality and the difficult control of n-type shallow junctions with the required sheet resistivity. A possible way out is to switch over to the on-chip integration of pMOSFETs in Ge with nMOSFETs in III-V layers, which are grown by selective epitaxial deposition on a silicon substrate. This approach, illustrated in Fig. 7, has a good potential for system-on-chip (SOC) applications [12].

A variety of new gate concepts have been developed for enhancing the device performance. Figure 8a shows schematically the FinFET, the double gate, the tri-gate, and the nanowire concept. A TEM photograph of a fabricated Si bulk FinFET and a Si nanowire [13] are given in Fig. 8b and 8c, respectively. Although there is a world-wide research effort on optimizing these device structures, some of the basic concepts were already reported and experimentally tested in the early 90-ties such as e.g. the gate-all-around device [14]. For FinFETs both bulk and SOI approaches are investigated.

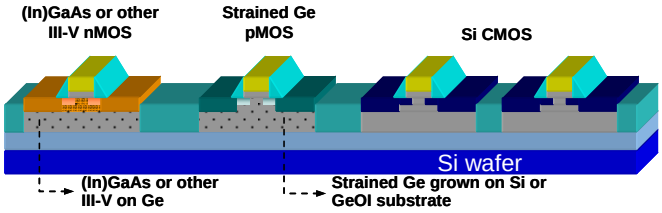
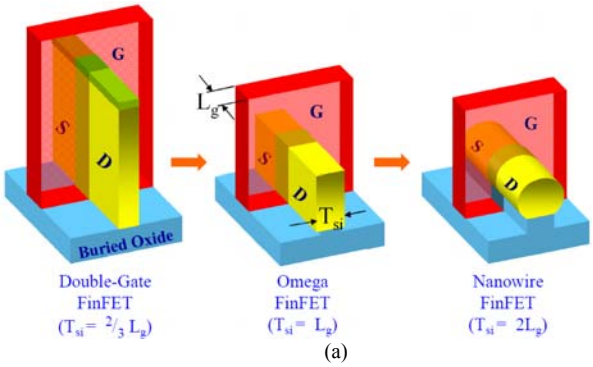
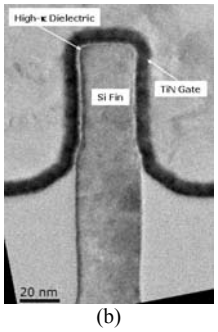


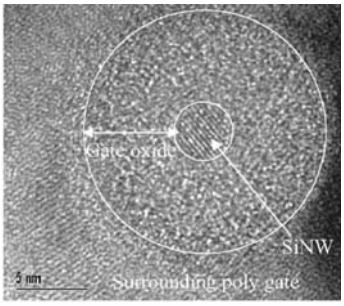
Fig. 7. Monolithic integration of Ge and III-V devices on Si (courtesy M. Heyns).



(a)



(b)



(c)

Fig. 8. (a) Schematic illustration of a double gate, a triple gate and a nanowire FinFET. TEM photograph of (b) a bulk FinFET and (c) a silicon nanowire [13].

The FinFETS and in the limit the silicon nanowires are forming a bridge between the standard scaling and the domain called "Beyond CMOS". A special type of device in this category is the tunnel-FET, which is schematically shown in Fig. 9 [15]. The tunnel FET can be all Si or the p-region can be SiGe, Ge or even an III-V compound. The devices have a subthreshold slope < 60 mV/dec, reduced short channel effects and are ideal for 3D stacking. The fabrication can be done either bottom up (i.e., growth of the Si nanowire using a

catalyst such as Au) or top down (i.e., based on the etching of the Si. Instead of using Au, it is also possible to work with In as a catalyst, which is enhancing the compatibility with a Si process line [16-17]. Different types of gate structures (overlapping, non-overlapping, short, etc) are being studied.

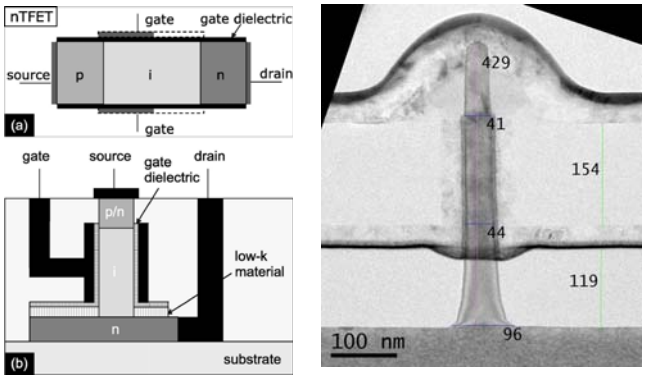


Fig. 9. (a) Conventional nTFET (highly-doped *p*-type source region, intrinsic channel region, highly-doped *n*-type drain region). (b) A vertical nanowire-based FET -TFET or MOSFET depending on the source doping type [15]. (c) TEM photograph of a 40 nm vertical FET.

Single or multi-wall carbon nanotubes (CNT) technologies have made much progress and both doping types can be achieved. An early discussion of different type of CNT designs has been published by Appenzeller *et al.* [18]. A good transistor performance has already been demonstrated. CNT also has a very strong potential to be used as interconnect material whereby it may even replace the Cu metallization. Much effort is going on to develop the appropriate catalysts to grow both lateral and vertical CNTs in a well-controlled manner. An illustration of a CNT array is shown in Fig. 10. There is also a strong interest to use CNTs for the back-end processing. CNTs have a 15 times higher thermal conductivity than Cu (6000 W/m.K compared to 400 W/m.K for Cu), in addition to their 1000 times higher current capacity 10^9 A/cm^2 compared to 10^6 A/cm^2 for Cu).

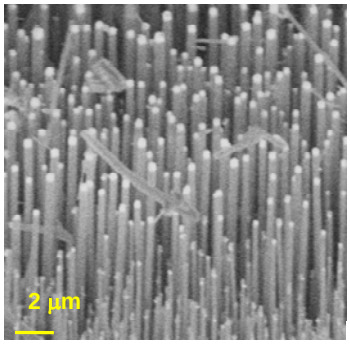


Fig. 10. Illustration of a CNT array.

Although the graphene technology is still in an early research phase, very promising electrical device performance has been reported with at a field of 0.4 MV/cm a hole and electron mobility of about 4780 and 4780 cm^2/Vs, respectively [19], for the uncovered devices. After deposition of the top gate electrode the mobility reduces to 710 and 530 cm^2/Vs. This topic has recently been reviewed by Lemme *et al.* [20]. A SEM photograph of a graphene transistor is shown in Fig. 11.

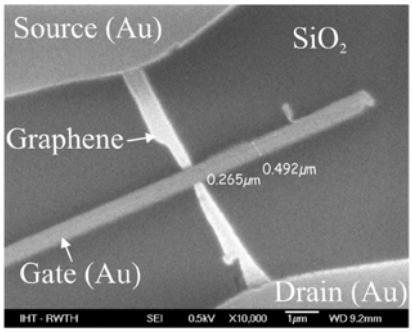


Fig. 11. SEM photograph of a grapheme transistor [19].

It is essential to remark that these alternative or emerging type of devices are already taken into account in the ITRS roadmap, indicating that they might have a good potential for the future so that extensive research is fully justified in order to be ready for implementation in manufacturing by the time that Si based CMOS is reaching its ultimate scaling. In Europe, the Network of Excellence NANOSIL is investigating a large variety of these Beyond CMOS topics [21].

III. MORE THAN MOORE

The societal needs will for a large extend be relying on the More than Moore domain, whereby in first instance scaling is not essential but the main goal is to increase the functionality of the circuits. The field is mainly application driven and puts stringent demands on the package related aspects. System-in-Package (SiP) and System-on-Chip (SoC) are surely becoming important issues. Besides packaging, another key challenge to face is the interface with the outside world, which is different for the different application fields. Two examples of extreme requirements are automotive and biomedical applications.

An important application field is automotive since the amount of electronics in cars is not only strongly increasing but also becoming an important cost element of the total cost of the present-day car. However, not only environmental concerns (e.g. CO_2 emission, fuel consumption) but also safety (ABS, EBS, airbag sensors, cruise control, etc) and comfort (GPS, board computer, local networks, etc) lead to an increased need of electronic components. In this field MEMS devices are finding their way. Only as an example, it can be mentioned that a poly SiGe MEMS technology is very suitable for the development of integrated gyroscopes used to measure the angular rate. An illustration of a designed circuit for this purpose, fabricated in a poly SiGe MEMS module on top of a 0.35 µm CMOS technology is shown in Fig. 12 [22]. Process optimization lead to a resolution of 0.07°/s within at least 50 Hz bandwidth [23].

Other applications of a poly SiGe technology are accelerometers, optical mirrors, resonators, etc [24-25]. Some of these devices are shown in Fig. 13.

There is also a lot of interest in after standard CMOS processing to use the layer above for the integration of passive components, such as highly-linear capacitors and integrated

inductors. An example of an application is a shunt inductor realized in a 90 nm above-IC technology, to give a better ESD projection as compared to a back-end inductor realization for a LNA and VCO [26]. Also other RF MEMS modules can be processed above-IC, as was demonstrated for a bulk acoustic wave resonator [27].

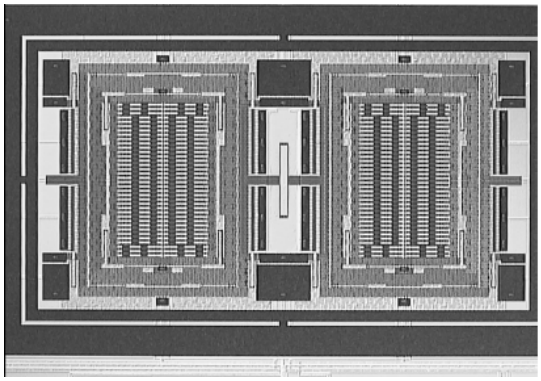


Fig. 12. Illustration of an integrated gyroscope, designed in poly SiGe on top of a 0.35 μm CMOS technology [23]

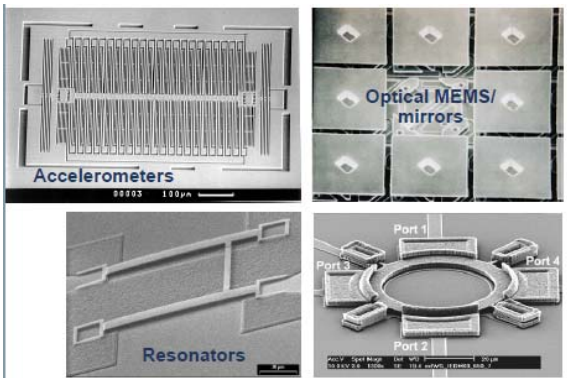


Fig. 13. Illustration of a poly SiGe technology for a variety of applications.

This approach, leading to a saving of silicon real estate, may replace some the presently System-in-Package (SiP) concepts used. The MEMS modules have to be tuned for the specific applications.

The More than Moore domain is surely requiring 3D packaging to achieve a higher packing density, to shorten the interconnects and to lower the interconnect density. Beside technological challenges an important aspect is the control of the thermal budget and to enable a cost effective processing. Nowadays, there is strong interest in 3D wafer level packaging (3D-WLP), whereby through silicon vias (TSV) are needed, Systems-on-Chip (SoC), and 3D SiP approaches. An illustration of the latter is shown in Fig. 14, giving both a schematic e-Cube and a practical realization for a medical application [28]. The roadmap of the 3 dimensional packaging research going on at IMEC, including 3D-SiP, 3D-WLP and 3D-SiC is given in Fig. 15 [28].

Semiconductor expertise can also be used to boost the sustainable energy generation. Photovoltaic research is covering a large variety of semiconductor materials such as Si, Ge, III-V, II-VI, etc. Depending on the application different requirements are put forward. The light-weight high efficient GaAs/Ge tandem solar cells used in space applications have different requirements than the technologies used for terrestrial applications, for which the main driving force is the cost/kw-peak. For the latter research is on going to develop thin layer high efficient and low-cost solar cells. Future approaches may be based on organic materials such as e.g. the low band gap polymers (thiphenes) [29-31].

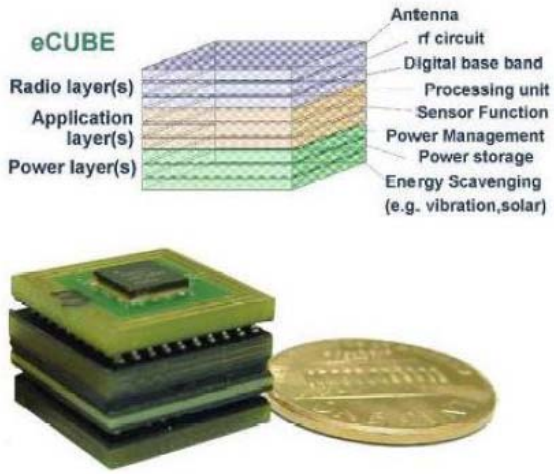


Fig. 14. Schematic representation of a 3D-SiP concept called eCube for the realization of a distributed fully autonomous ambient intelligent system. Bottom: Photograph of a 2 cm^3 eCube developed for a medical application, integrating a rf-SiP with integrated antenna, low-power DSP SiP, 19 channel EEG/ECG sensor die and a power SiP [28]

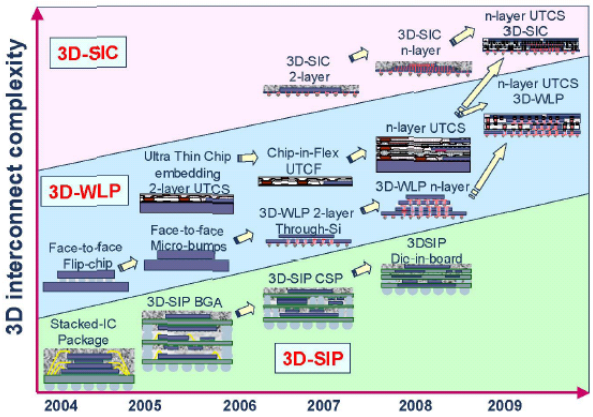


Fig. 15. IMEC's packaging roadmap for three important 3D approaches which will become important for More than Moore applications in different fields [28].

Another field with a huge potential is related to healthcare. The population is becoming older thanks to the great progress in medical sciences. However, it is essential not only to increase the lifetime but to ensure the quality of life. This necessitates to diagnose and monitor and whenever possible to cure the deceases. Typical examples of important deceases are Alzheimer and Parkinson, which are frequently associated with the elderly society. The world of bio-electronics is making large progress and opens new application fields. The nanometer geometry of the state-of-the-art devices is very similar to the size of a virus or an anti-body. Using the so-

called neurons on a chip or artificial synapse approach, enables the bi-directional communication between neurons and an integrated circuit. The interaction can be stimulated either electrically or chemically.

The bio-electronics applications require a strongly multi-disciplinary approach with a research team formed by experts in microelectronics, chemistry, biology and medicine in order to fabricate the different system components needed: sensors, tranducers, the on-chip cellular microenvironment, the in-vitro hybrid devices, the in-vivo neuroprobes, the modeling and signal processing, and the appropriate packaging and integration aspects [32-33]. This will have a strong impact on the curriculum of our future engineers.

ACKNOWLEDGMENT

The author has based this manuscript mainly on research carried out at IMEC and would therefore like to thank K. Baert, E. Beynen, C. Bartic, S. De Gendt, M. Heyns, A. Mercha, R. Mertens, J. Poortmans, P. Vereecken, A. Verhulst and C. Van Hoof for many discussions and the use of some of their material.

REFERENCES

[1] Bardeen J, Brattain WH, "Physical principles involved in transistor action", *Phys. Rev*, 75, pp. 1208-1225, 1949

[2] Kilby JS, "Invention of the integrated circuit", *IEEE Trans. Electron Dev*., 23, pp. 648-654, 1976

[3] European Nanoelectronics Initiative Advisory Council (ENIAC), Strategic Research Agenda, November 2007, http://www.eniac.eu

[4] Moore GE, "Cramming more components onto integrated circuits", *Electronics Mag*., 38, pp. 114-117, 1965

[5] International Technology Roadmap for Semiconductors, http://www.itrs.net

[6] Wakabayashi H, Yamagami Y, Atsushi I, Ogira A, Narihiro M, Arai KI, Ochiai Y, Takeuchi K, Yamamoto T, Mogami T, "Sub 10-nm planar-bulk-CMOS devices using lateral junction control", *IEDM Techn. Dig*., pp. 389-391, 2003

[7] Thompson SE, Chou RS, Ghani T, Mistry K, Tyagi S, Bohr MT, "In search of 'Forever' continued transistor scaling – One new material at a time", *IEEE Trans Semiconductor Mft*., 1. pp. 26-36, 2005

[8] Huff HR, and Gilmer DC, "High Dielectric Constant Materials: VLSI MOSFET Applications", Springer Verlag, 2005

[9] Claeys C, Simoen E, Put S, Giusi G, and Crupi F, "Impact of strain engineering on gate stack quality and reliability", Solid-State Electron., 52, pp. 1115-1126, 2008.

[10] Claeys C, Simoen E, "Germanium-Based Technologies" From Materials to Devices", Elsevier, 2007

[11] Mitard J, De Jaeger B, Leys FE, Hellings G, Martens K, Eneman G, Brunco DP, Loo R, Lin JC, Shamiryan D, Vanderweyer T, Winderickx G, Vrancken E, Yu CH, De Meyer K, Caymax M, Pantisano L, Meuris M, and Heyns MM, "Record Ion/Ioff performance for 65-nm pMOSFET and novel Si passivation scheme for improved EOT scalability", *Technical Digest of IEDM*, pp. 873-875, 2008

[12] DUALLOGIC, "Next Generation Nanoelectronics Components & Electronics Integration", Framework 7 project financially supported by the European Commission, www.ims.demokritos.gr/DUALLOGIC

[13] Singh N, Agarwal A, Bera LK, Liow TY, Yang R, Tustagi SC, Tung CH, Kumar R, Lo GQ, Balasubramanian N, Kwong DL, "High-performance fully depleted silicon nanowire (diameter < 5 nm) gate-all-around CMOS devices", *IEEE Electron Dev. Lett*., 27, pp. 383-386,

[14] Colinge JP, Gao MH, Romano A, Maes H, Claeys C, "Silicon-on-insulator gate-all-around device" *Technical Digest of IEDM*, pp. 595-598, 1990

[15] Verhulst AS, Vandenberghe G, Maex K, and Groeseneken G, "Tunnel field effect transistor without gate-drain overlap", *Appl. Phys. Lett*., 91, 053102, 2007

[16] Iacopi F, Richard O, Eichhammer Y, Bender H, Vereecken PM, De Gendt, and Heyns MM, "Size-dependent characteristics of Indium-seeded Si nanowire growth", *Electrochem. Soc. Lett*., 11, pp, K98-K100, 2008

[17] Iacopi F, Vereecken, PM, Schaekers, M, Caymax M, Moelans N, Balnpain B, Richard O, Detavernier C, and Griffiths H, "Plasma-enhanced chemical vapour deposition growth of Si nanowires with low melting point metal catalysts: an effective alternative to Au-mediated growth", *Nanotechnol*., 18, pp. 505307/1-7, 2008

[18] Appenzeller J, Lin YM, Knoch J, Chen Z, and Avouris P, "Comparing carbon nanotube transsitors – The choice: A novel tunnel device devising", *IEEE trans. Electron Dev*., 52, pp. 2568-2576, 2005

[19] Lemme MC, Echtermeyer TJ, Baus M, Kurz H, "A graphene field effect transistor" *IEEE Electron Dev. Lett*., 28, pp. 282-284, 2007

[20] Lemme MC, Echtermeyer JT, Baus M, Szafranek BN, Schmidt M, Kurz H, "Towards graphene field effect transistors", *Electrochem. Soc, Trans*., 11(6), pp. 413-419, 2007

[21] NANOSIL, "Silicon-based nanostructures and nanodevices", Framework 7 Network of Excellence, www.nanosil-noe.eu

[22] Witvrouw A, Nehta A, Verbist A, Du Bois B, Van Aerde S, Ramos-Martos J, Ceballos J, Ragel A, Mora JM, Lagos MA, Arias A, Hinijosa J, Spengler J, Leinenbach C, Fuchs T, Kronmuller S, "Processing of MEMS Gyroscopes on top of CMOS ICs", *Proc. Int. Solid-State Circuit Conf*., pp. 88-89, 2005

[23] Scheurle A, Fuchs T, Kehr K, Leinenbach C, Kronmueller S, Arias A, Ceballos J, Lagos M, Mora J, Munoz J, Ragel A, Ramos J, Van Aerde S, Spengler J, Mehta A, Verbist A, Du Bois B, Witvrouw A, "A 10 µm thick poly-SiGe gyroscope processed above 0.35 µm CMOS", *Proc. IEEE MEMS*, pp. 39-42, 2007

[24] Van Barel G, Van Hoof R, Du Bois B, Verbist A, Witvrouw A, Haspeslagh L, "Fabrication scheme for the in-situ mechanical characterization of freestanding MEMS structural layers, applied to Poly-SiGe", *Proc. NAM Conf*., pp. 1-2, 2007

[25] Wang Z, Leonov V, Fiorini P, Van Hoof C, "Micromachined polycrystalline SiGe-based thermopiles for micropower generation on human body", *Proc. Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS*, pp. 2840289, 2007

[26] Linten D, Sun X, Thijs S, Mahadeva Iyer N, Mercha A, Carchon G, Wambacq P, Nakaie T, Decoutere S, "Low-power low-noise highly ESD robust LNA and VCO design using above IC inductors". *Proc. Custom Integrated Circuits Conference*, pp. 492-495, 2005

[27] Dubois MA, Carpentier JF, Vincent P, Billard C, Parat G, Muller C, Ancey P, Conti P, "Monolithic above-IC resonator technology for integrated architectures in mobile and wireless communication", *IEEE J. Solid State Circuits*, 41, pp. 7-17, 2006

[28] Beyne E, Swinnen B, "3D system integration technologies", *Proc. IEEE Integrated circuit Design and Technology Conf*., pp. 1-3, 2007

[29] Aernouts T, Vanlaeke P, Haeldermans I, D'Haen J, Heremans P, Poortmans J, Manca J, "P3HT:PCBM bulk heterojunction solar cells: morphological and electrical characterization and performance optimization", *Organic and Nanoparticle Hybrid Photovoltaic Devices*, pp. 1013-Z01-02, 2007

[30] Aernouts T, Girotto C, Gommans H, Cheyns D, Rolin C, Genoe J, Poortmans J, "Processing technologies for organic solar cells", *Proc. 22nd European Photovoltaic Solar Energy Conference and Exhibition*, pp. 560-565, 2007

[31] Cheyns D, Gommans H, Odijk M, Poortmans J, Heremans P, "Stacked organic solar cells based on pentacene and C60", *Solar Energy Materials and Solar Cells*, 91, pp. 399-404, 2007

[32] Spira M, Kamber D, Dormann A, Cohen A, Bartic C, Borghs G, Langendijk J, Yitzschaik S, Shabthai K, Shappir J, "Improved neuronal adhesion to the surface of electronic device by engulfment of protruding micro-nails fabricated on the chip surface", *Transducers & Eurosensors '07: The 14th International Conference on Solid-State Sensors, Actuators and Microsystems*, pp. 1247-1270, 2007

[33] Van Meerbergen B, Raemaekers T, Winters K, Braeken D, Bartic C, Spira M, Engelborghs Y, Annaert W, Borghs G., "Improving neuronal adhesion on chip using a phagocytosis-like event", *J. Experiment. Nanoscience*, 2, pp. 101-114, 2007

A SONET OC-48 CMOS Based Integrated Optical Transceiver

T. Y. Kalombo*, *Student Member, IEEE* and S. Sinha†, *MIEEE*
*†Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa
*toni@tuks.co.za, †ssinha@ieee.org

***Abstract* – Fibre optic communication provides a platform for high speed and high density data transmission over networks. The increasing demand for such high capacity systems has prompted a migration away from the commonly used bipolar transistor devices towards cheaper complementary metal-oxide semi-conductor (CMOS) based devices. A major advantage of CMOS is that their gates only dissipate power during switching, and their sizes can be scaled to much smaller proportions, making them faster and more efficient. An integrated solution has the added advantage of breaking down practical complexities inherent with cascading multiple discrete subsystems. This paper describes the design of an integrated optical transceiver to be used in applications for Synchronous Optical Network OC-48 (optical carrier at 48 times base frequency). At the receiver a broadband transimpedance amplifier performs current to voltage conversion on incoming signals. The voltage signal at the output of the transimpedance amplifier is further amplified using a limiting amplifier. The limiter drives the clock recovery circuit which recovers the clock from the data signal. A quadrature voltage controlled oscillator is used in the clock recovery loop to generate in-phase and quadrature beat signals at 2.488 Gbps. The transmitter consists of a laser diode driver which drives an externally connected laser diode**.

I. Introduction

AMONGST various means, high density data transfer, optical fibre remains a favourable medium. Optical networks however rely on hardware that can carry the data and maintain the required switching speed. Advances in complementary metal-oxide semi-conductor (CMOS) technology have made CMOS transistors the device of choice for high data rate applications. CMOS devices not only perform at high speed, they are also cost efficient and more so efficient in their design and their manufacturing process. This paper presents the use of CMOS technology within a front-end transceiver for optical communication devices.

II. Design Specifications

The system was designed according to three primary specifications. In order to be compatible with SONET OC-48 devices, the system was designed to operate at a frequency of 2.488 GHz. The other two specifications that were considered were a static power consumption of less than 200 mW and a transimpedance gain of approximately 60 dBΩ.

III. Conceptual Design

The optical transceiver is comprised of a transmitter and a receiver. At the transmitter, a laser diode driver is sufficient to drive the laser diode. Conceptually, this can simply be represented by a voltage and a current amplifier. The receiver accounts for much of the complexity involved in designing the optical transceiver. The system block diagram is shown in Figure 1. The optical receiver employs a quadricorrelator architecture [1] to perform clock recovery on random non-return to zero (NRZ) signals. This topology is ideal for high speed applications where random binary data is being received. Figure 2 shows the mathematical model of the CRC. Random binary data contains no even-order harmonics. For a data rate of 2.488 Gbps, no linear time invariant (LTI) operation can extract a 2.488 GHz period clock from NRZ data. This can be remedied by performing edge detection – a non-linear operation – on the data [2] .

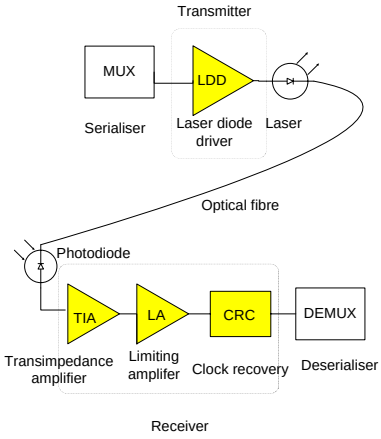


Fig. 1. Block diagram of the optical transceiver. The shaded blocks are the subsystems that have been implemented in this paper.

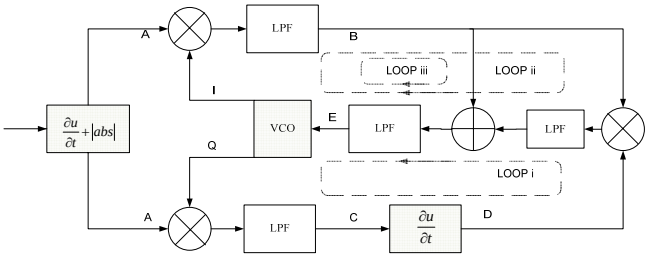


Fig. 2. Simplified mathematical model of the clock recovery circuit.

Edge detection is performed by differentiating and full-wave rectifying the NRZ signal. Differentiating a square pulse yields positive impulses for rising edges and negative impulses for falling edges. A differentiator has the following transfer function: $H(s) = s$. Where

$$|H(j\omega)|^2 = \omega^2 \tag{1}$$

For an input power spectral density $S_{in}(f)$, the output spectrum is represented by $S_{out}(f)$:

$$S_{out}(f) = S_{in}(f) \cdot |H(j\omega)|^2 \tag{2}$$

$$= \left[\frac{2\sin(\pi f T_b)}{T_b} \right]^2 \tag{3}$$

The differentiated signal still yields zero energy at the spectral line. This brings about the need for full-wave rectification on the differentiated data. In order to obtain a spectral line at a certain frequency, the differentiated signal can be correlated with a cosine and its time average can be computed. A zero time average means that there is no spectral line [2]. This indeed is the case for a differentiated binary NRZ sequence, since it contains an even number of opposite impulses for each bit and their time average for an arbitrary phase is zero.

Rectification allows these impulses to have the same polarity, hence a non-zero average. Edge detection is mathematically performed by the functional block labeled "$\frac{\partial u}{\partial t} + |ABS|$" in Figure 2. The edge detected signal can be assumed to have a spectral component at an arbitrary frequency ω_1. Let us represent this signal at node A by a sinusoid

$$x_1(t) = \sin(\omega_1 t) \tag{4}$$

Multiplication is performed on equation 4 by the in-phase/quadrature (I/Q) signals generated by the VCO, which oscillates at an arbitrary frequency ω_2. These are low-pass filtered to suppress higher order harmonics and produce quadrature beat signals

$$\sin\left[(\omega_1 - \omega_2)t\right] \tag{5}$$

and

$$\cos\left[(\omega_1 - \omega_2)t\right] \tag{6}$$

at nodes B and C, respectively. The signal at node C is differentiated in time to produce

$$-(\omega_1 - \omega_2)\sin\left[(\omega_1 - \omega_2)t\right] \tag{7}$$

at node D. The signals at node D and at node B are multiplied and low-pass filtered, the result is an averaged error signal with amplitude proportional to

$$\frac{1}{2}(\omega_1 - \omega_2) \tag{8}$$

This error is reduced as the feedback loop brings the VCO's frequency; ω_2 closer to ω_1. This is a frequency locking operation which occurs in the two external loops in the quadrature branch. Once the frequency is in lock, the third loop begins to dominate and the error signal at node E is further minimized to obtain phase lock.

The quadricorrelator architecture is therefore a quadrature branch with two loops that perform frequency locking and a third loop that performs phase locking.

IV. CMOS Implementation

A. Transimpedance amplifier

In order to convert the electrical current generated by the photodiode to a voltage current, a feedback TIA has been implemented [2]. Using a feedback topology not only increases the amplifier's bandwidth, it also lowers the output resistance which yields a better drive capability. Figure 3 shows the topology used to realise the broadband TIA.

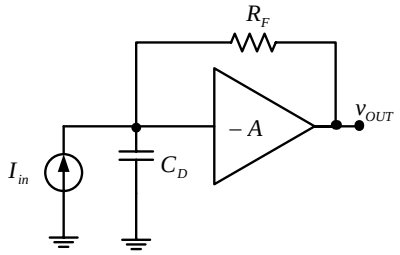


Fig. 3. Feedback TIA topology.

R_F provides a feedback path around A, the ideal voltage amplifier. C_D is the capacitance seen at the output of the photodiode and I_{in} is the input current that the TIA converts to a corresponding output voltage V_{out}. The CMOS implementation of the TIA is a second order circuit with a transfer function as follows

$$\frac{v_{OUT}(s)}{i_{IN}(s)} = -\frac{A(s)R_F}{A(s) + 1 + R_F C_D s} \tag{9}$$

$$= -\frac{A_o R_F}{\frac{R_F C_D}{\omega_o} s^2 + (R_F C_D + \frac{1}{\omega_o})s + A_o + 1} \tag{10}$$

Where A(s) is the open loop amplifier transfer function, A_o is the amplifier's gain and ω_o is the open loop pole.

The TIA has a single-ended output. In order to suppress common-mode noise, reduce data feedthrough and improve the power source rejection ratio (PSRR), a differential topology has been adopted for the entire system. The differential topology also has a larger voltage swing than the single-ended topology and it provides better component matching when manufacturing. The TIA single-ended output is converted to a differential output [2] with a single-ended to differential converter whose topology acts as a high-pass filter (HPF). Its cut-off frequency is calculated as

$$f_{HPF} = \frac{1}{2\pi R_1 C_1} \tag{11}$$

B. Limiting amplifier

The limiting amplifier boosts the signal at the output of the

TIA and drives the CRC. A limiter with an inverse scaling topology was adopted [3], [4]. This topology allows for bandwidth extension to be achieved by scaling the driven stage to be smaller than the driving stage. The gain-bandwidth product (GBWP) can be maintained by scaling for each stage, all MOS transistors and their parameters as well as all current sources by the same factor. A differential topology was adopted and inductive peaking was implemented for bandwidth enhancement. Figure 4 shows the circuit topology implemented for each gain stage.

PMOS transistors P1 and P2 are active resistors, and combined with NMOS transistors M3 and M4, they respectively form active inductors. The inductor impedance can be approximated by [3]

$$Z_L(s) \approx \frac{1+sR(C_{gs3}+C_{gd3})}{g_{m3}+s[g_{m3}RC_{gd3}+C_{gs3}+C_L]} \qquad (12)$$

Where C_{gs3}, C_{gd3} and C_L are gate-source, gate-drain and load capacitances, respectively. R is the resistance of PMOS transistor P1, which is identical to P2, and g_{m3} is the transconductance of transistor M3.

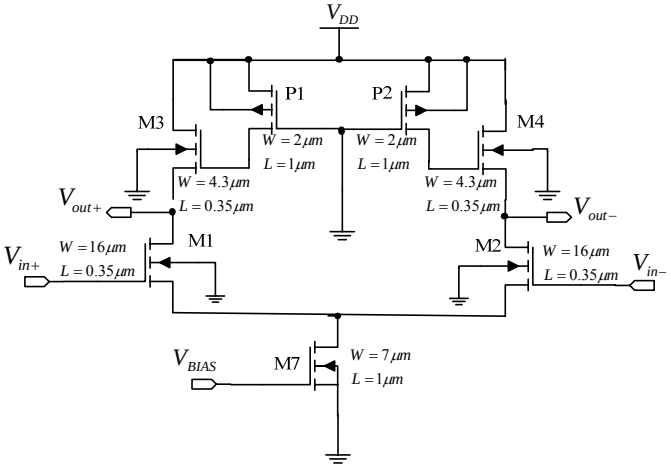


Fig.4. Schematic of an individual gain stage of the limiting amplifier.

A six-stage limiting amplifier was adopted for post-amplification after current-to-voltage (I/V) conversion, where each stage was multiplied by a scaling factor of $\alpha = 1.2$ [3].

C. The integrated differentiator, rectifier, mixer and low-pass filter subsystem (DRML)

The DRML presented in [1] is adapted and designed for use in the 0.35 μm CMOS process. It is an integration of a mixer, a differentiator, a rectifier and a low-pass filter into one functional subunit. Not only does this topology reduce the power consumption, it also reduces loading effects that would have occurred had all four subsystems been implemented discretely then integrated.

1. The differentiator

A differential pair with capacitive degeneration was used in the implementation of the differentiator. The capacitor C_D, as seen in Figure 5 influences the pulse-width of the differentiated NRZ signal. A square wave that is differentiated reproduces bipolar impulses whose polarities depend on whether the wave is at a rising or at a falling edge. The aim of edge detection is to generate narrow unipolar pulses corresponding to the data transition [2]. At the bit-rate specified by the OC-48 standard, half a period is 0.2 ns long. An impulse in the order of a few pico-seconds requires a value for C_D as small as 12 fF. Transistors M3 and M4 are source followers that are used to ensure that M1 and M2 remain in saturation. When D_{in+} goes high, the charge retained on C_D causes both I_{D1} and I_{D2} to momentarily bias NMOS transistor M1 while M2 remains unbiased. The quick discharge of C_D results in impulses to be generated at the drain of M1. The biasing role is reversed for a rising transition at the gate of M2.

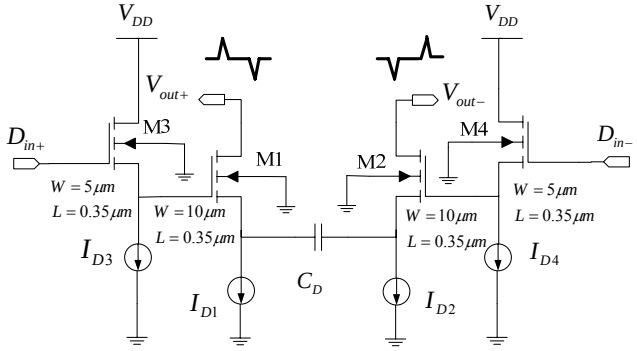


Fig.5. Circuit diagram of the differentiator.

2. The full wave rectifier

Figure 6 shows the CMOS implementation of the full-wave rectifier using a Gilbert cell topology, in cascade with the differentiator. The differentiated signal is full-wave rectified by multiplying the NRZ data, at the output of the limiter, with the bipolar impulses produced by the differentiator. This happens as D_{in+} goes high, M1 becomes biased and a positive current impulse flows through both transistors M5 and M8, while a negative current impulse flows through M6 and M7.

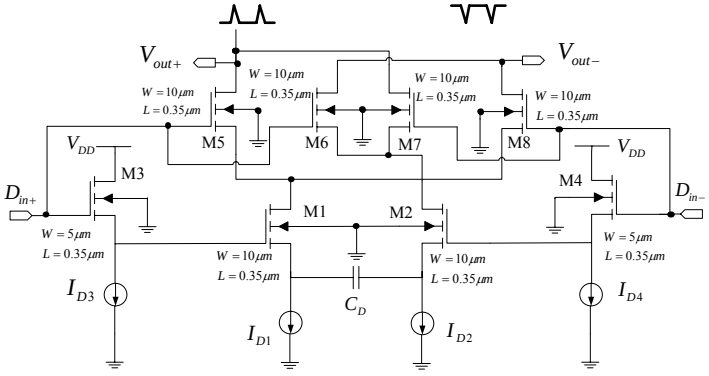


Fig.6. Circuit diagram of the rectifier cascaded with the differentiator.

During this instant, the signals at the gate of M5 and M6 are at a rising edge, while conducting current impulses of opposite polarities on each respective transistor. This positive signal is therefore multiplied by the positive impulse at M5 and the negative impulse at M6. The product is as follows.

- A positive impulse results at the drain of M5, and
- A negative impulse results at the drain of M6.

With M1 still conducting, at the falling edge, the current impulse at M5 becomes negative, while the impulse at M6 becomes positive. This results in a negative signal on both gates of M5 and M6 to be multiplied by the following signals.

- A negative impulse at M5, yielding a positive impulse at the drain of M5, and
- A positive impulse at M6, resulting in a negative impulse at its drain.

This cycle repeats when D_{in-} goes high and M2 switches on while M1 goes off. It can be seen from Figure 6 that M5 and M7 both produce positive unipolar impulses and that M6 and M8 both produce negative unipolar impulses. The Gilbert cell therefore produces at its outputs, a full-wave rectification of the differentiated NRZ data. M5 – M8 are identical NMOS transistors.

3. The mixer

Once edge detection has been performed, the unipolar impulse train is multiplied by a sinusoidal signal originating from the VCO. A second identical Gilbert cell, cascaded on top of the one used for rectification is used to perform this multiplicative operation. Figure 7 shows the circuit diagram of the differentiator in cascade with the rectifier, the multiplier and low-pass filter.

4. The low-pass filter

The final subcircuit of the DRML circuit is the low-pass filter (LPF). The differential output is taken over the terminals of C_1. A passive LPF is suitable for suppressing higher order harmonics introduced due to mixing of signals. Its transfer function can be expressed as

$$F(s) = \frac{1}{1+s\tau} \tag{13}$$

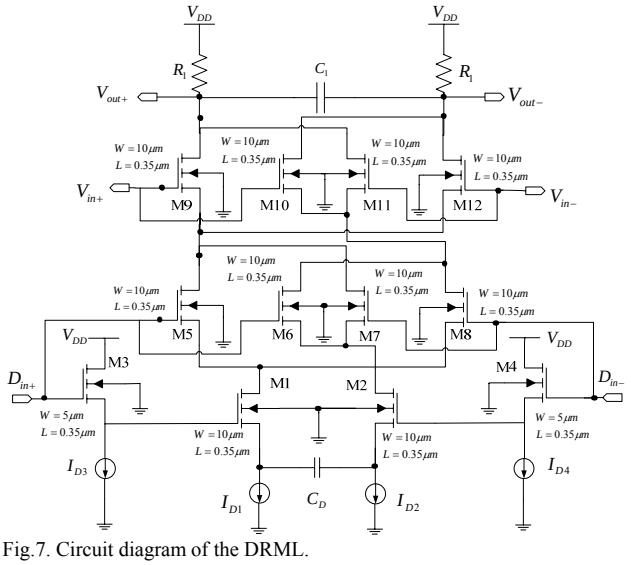


Fig.7. Circuit diagram of the DRML.

Where τ represents the time constant determined by

$$\tau = R_1 C_1 \tag{14}$$

where $R_1 = R_2$. The -3 dB cut-off frequency of the LPF is given by

$$f_{-3dB} = \frac{1}{2\pi R_1 C_1} \tag{15}$$

Low-pass filtering after mixing the signal with a sinusoid is approximately equivalent to band-pass filtering before down-conversion [2]. The harmonics also introduce asymmetry to the circuit. A 530 MHz cut-off enables these asymmetries to be suppressed, while enough energy is kept available to allow frequency and phase locking to take place. For $\tau = 3ns$, choosing $C_1 = 100\ fF$ yielded $R_1 = 30\ k\Omega$.

D. The integrated differentiator, mixer and low-pass filter (DML)

In this subsystem the differentiated signal is multiplied by the sinusoidal signal at the output of the DRML. This is low pass filtered to generate a DC control signal. The same cascade approach that has been applied to design the DRML is adopted in implementing the DML. For the low-pass filter, a time constant of 3 ns was selected, resulting in resistor values of $R_1 = R_2 = 30\ k\Omega$ for a capacitance of $C_1 = 100$ fF.

E. The adder

To incorporate the third loop in the system and perform phase-locking, an adder is used. This is cascaded to a low-pass filter that has the same characteristics as the LPF used in the DML. The adder is implemented with the differential pair topology shown in Figure 8. Source followers were used to maintain the desired operation of the transistors in saturation.

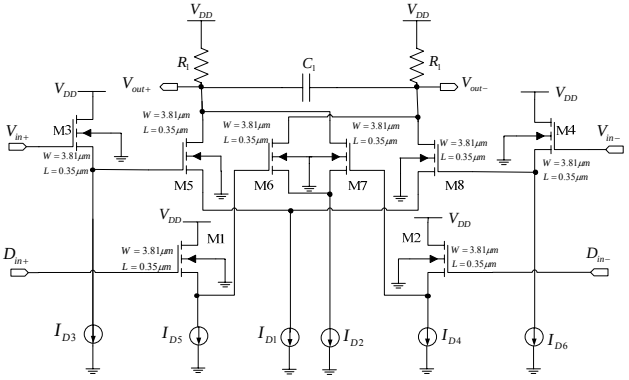


Fig.8. Circuit diagram of the adder cascaded to a LPF.

F. The voltage controlled oscillator

The VCO used in the optical transceiver is a quadrature voltage controlled oscillator adapted from the quadrature oscillator by [5]. The oscillator is an I/Q cross-coupled oscillator that employs direct and cross coupling techniques between two identical LC oscillators. This configuration makes it possible to generate differential in-phase and quadrature signals. The LC oscillator can be viewed as a feedback system or a negative resistance in parallel with a lossy tank.

The oscillator is converted to a VCO by applying a controllable capacitive load to its outputs. The LC tanks of the VCO are implemented with spiral inductors in parallel with the tank's capacitive element. The capacitive load can be used to control the resonant frequency of the LC tank by altering the net capacitance of the tank. The AMS variable capacitance model CVAR was used as a controllable capacitive load. This dictates the VCO's tuning range.

The LC tank resonant frequency is given by

$$f_{res} = \frac{1}{2\pi\sqrt{LC}} \tag{16}$$

C is the sum of the LC tank capacitance and the load capacitance.

$$C_{TOTAL} = C_{VAR} + C_P \tag{17}$$

Figure 9 shows the circuit diagram of the VCO.

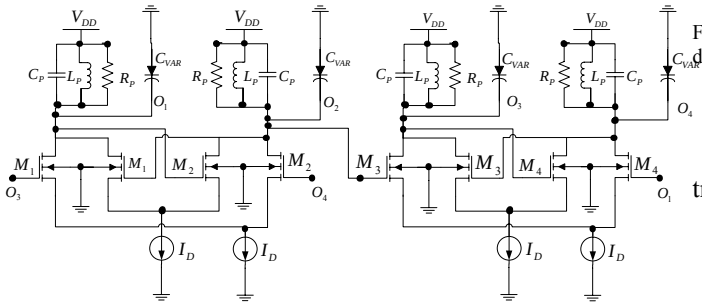


Fig. 9. Schematic of the quadrature VCO.

The VCO uses a single-ended input. A difference amplifier can be used as a buffer between the differential output at the adder and the LPF and the input at the VCO. This VCO topology has a voltage swing of 1 $V_{p\text{-}p}$, with a 3.3 V offset. A voltage divider circuit is used to bring the signal to a 1.65 V offset. This enables the driven stage to operate effectively without breaking down the transistors.

G. *The transmitter*

A laser diode driver is in essence a current controlled switch that turns the laser on and off. The output power of laser diodes is more accurately defined as a function of current than of voltage [2]. For the SONET OC-48 bit-rate, major requirements for the design of the laser diode driver (LDD) are high switching speed, an output current high enough to generate optical power and a large voltage swing tolerance from the laser diode (LD). Due to the low power requirement of the system, the LDD presented here is intended for low-current LD's with large voltage swing capabilities. The topology and the design principles adopted in the design of the limiting amplifier also apply for the LDD. Of the three functional units that make up the LDD, the first two are voltage amplification stages and the last functional unit employs a differential pair that steers its tail current to the LD. Though the output is single-ended, a differential driver is implemented to provide common-mode rejection while keeping the output current swing high.

The first stage of the LDD is equivalent to the six-stage cascade that makes up the limiting amplifier. In order to minimise jitter and provide further voltage gain, while minimising Miller capacitances, a common-source, common-gate differential amplifier is introduced in the second stage. This is followed by an output buffer with an open-drain structure. These two stages are presented in the circuit diagram of Figure 10.

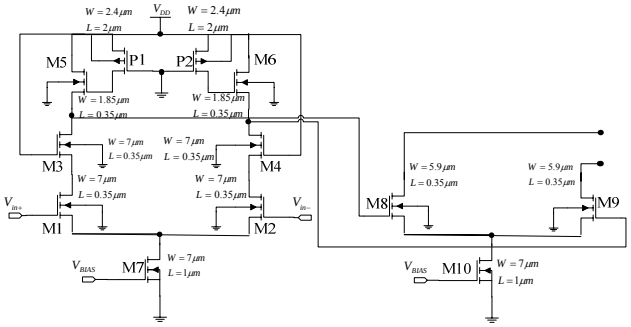


Fig.10. Circuit diagram of the second stage (common-source common gate differential amplifier) and output buffer of the LDD.

V. LAYOUT OF THE CMOS BASED OPTICAL TRANSCEIVER

After the schematics were simulated, a layout for the optical transceiver was developed. This is shown in Figure 11.

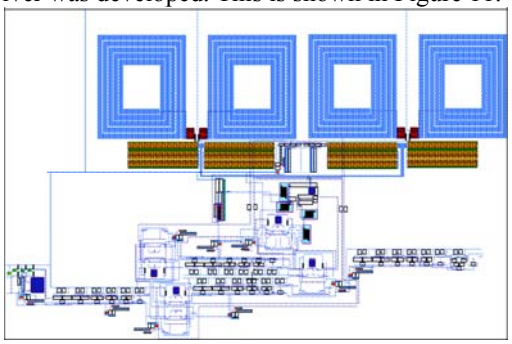


Fig. 11. Layout of the optical transceiver. Dimensions: 672.14 µm × 327.6 µm.

VI. SIMULATION RESULTS

Simulations were performed on the concept design in MATLAB and on the CMOS implementation in SPICE. The SPICE results were compared to the MATLAB results in order to verify the operation of the CMOS subcircuits.

A. *Subsystem outputs*

Figure 12 shows the SPICE output of the edge detected signal that results from an input binary NRZ sequence. This signal was modelled as an ideal current source representing a laser diode in MATLAB and its output was exported to SPICE. This serves as input to the TIA, whose output feeds the edge detector. Unipolar spikes are generated at the output of the edge detector as expected from the theoretical illustration in Figure 6. The rectification performed incorporates a Gilbert Cell that actually multiplies a signal with itself.

Figure 13 shows that low-pass filtering after mixing results in the signal being down-converted and subsequently band-pass filtered. This creates a spectral line at the frequency of interest.

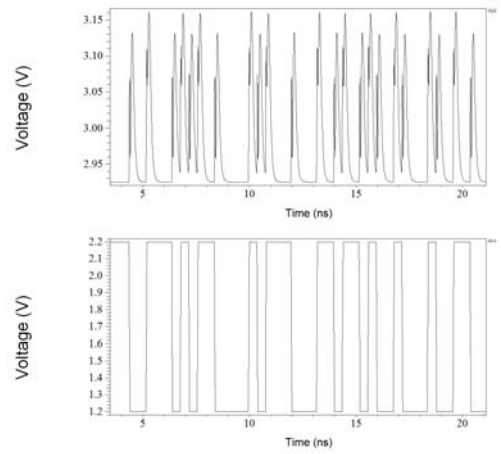


Fig. 12. Output impulse train and the input binary NRZ signal to the CMOS based edge detector in SPICE.

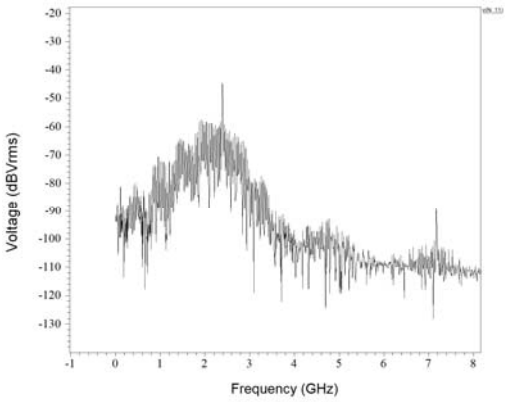


Fig. 13. Output spectrum of the DRML in SPICE.

B. Simulated system specifications

1) Operating frequency

The clock was recovered on 2.488 GHz random binary data. Figure 14 shows the spectrum of the recovered clock.

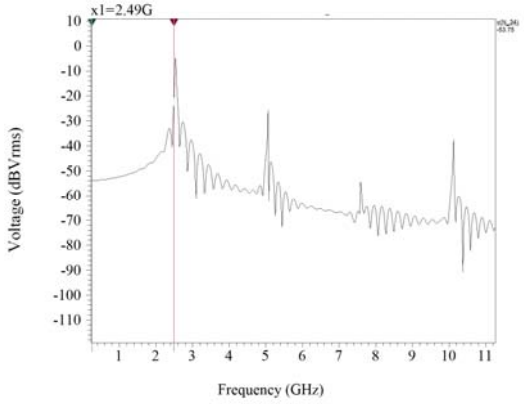


Fig. 14. The spectrum of the recovered clock in SPICE.

2) Static power consumption

The static power was derived by measuring the total current drawn from the supply. With a rail voltage of 3.3 V, a current of 35.86 mA was measured. This results in a total power consumption of 118.35 mW.

3) Transimpedance gain

From the magnitude response of the TIA. The -3 dB bandwidth was measured at 3.2 GHz and the TIA gain at 66 dBΩ.

VII. SUMMARY OF IC SPECIFICATIONS

The CMOS implementation specifications are summarized in Table II.

TABLE II
SUMMARY OF SPECIFICATIONS

Parameter	Value
Process	AMS 0.35 µm
Single rail supply voltage	3.3 V
Transimpedance gain	66 dBΩ
Limiter bandwidth	1.77 GHz
Operating frequency	2.488 GHz
LDD output current	10 mA
Power consumption	118 mW
Tracking range	± 10 MHz

VIII. CONCLUSION AND DISCUSSION

NRZ binary data is a preferred choice for optical transmission. This signal however does not carry a spectral line at its transmission bit-rate. It is shown that the edge detector extracts the frequency information from the NRZ data. Since the rectification performed incorporates a Gilbert cell that actually multiplies a signal with itself, it results in up-conversion occurring at twice the frequency. In the process of generating the VCO control signal, further mixing and low-pass filtering also takes place. It was illustrated that the effect of low-pass filtering after mixing is a result in the signal being down-converted and subsequently band-pass filtered. This creates a spectral line at the frequency of interest and also gives the VCO a reference for frequency and phase detection through error reduction in a feedback loop. The use of quadrature beat signals helps to further reduce the error while facilitating phase and frequency locking.

IX. REFERENCES

[1] B. Razavi, "*A 2.5-Gb/s 15 mW clock recovery circuit", IEEE Journal of Solid-State Circuits*, pp. 472-480, Vol. 31, No. 3, Mar. 1996.

[2] B. Razavi, Design of Integrated Circuits for Optical Communication, 1st edition, McGraw-Hill Education, 2003.

[3] C.G. Chen, W. Chen, and R. Luo, "*A 2.5 Gbps CMOS Laser Diode Driver with Preemphasis Technique". Proc. of IEEE Asia-Pacific conference* , Taipei, pp. 65-68, 6-8 Aug. 2002.

[4] *E. Sackinger and W.C. Fischer "A 3-GHz 32-dB CMOS limiting amplifier for SONET OC-48 receivers", IEEE Journal of Solid-State Circuits*, pp. 1884-1888, Vol. 35, No. 12, Dec. 2000.

[5] N. Naudé, M. Božanić and S. Sinha, "Analogue CMOS direct sequence spread spectrum transceiver with carrier recovery employing complex spreading sequences," in *Proc. IEEE Mediterranean Electrotechnical Conference,* 2006, pp. 1227-1230.

3D CMOS Photodiode Modeling and Simulation using the Finite Element Method

Paul Ellinghaus* and Petrus J. Venter†, *MIEEE*
University of Pretoria, Dept. Electrical, Electronics and Computer Engineering, CEFIM, South Africa
*pellinghaus@tuks.co.za, †jannes.venter@up.ac.za

***Abstract*— This paper investigates the simulation of three-dimensional CMOS photodiode structures (n-well/p-substrate and p+/n-well/p-substrate). The simulation of the transient response, frequency response and spectral responsivity is considered. The approach followed divides the structures into bulk regions and depletion regions and calculates the generated photocurrent within each. The carrier continuity equations are solved within bulk regions for minority carriers using the finite element method (FEM).**

I. INTRODUCTION

The ever decreasing dimensions of semiconductor devices in integrated circuits dictate smaller and more densely integrated interconnects. This increases the interconnects' capacitance and resistance, leading to reduced bandwidth and excessive power consumption. The mentioned problems – collectively known as the interconnect dilemma in industry parlance – makes electrical interconnects unviable for future high-speed clock and data distribution on, and between, chips.

Optical interconnects are envisaged to provide a solution to the present interconnect limitations and is strongly pursued by the semiconductor industry. Despite the potential optical nature of future interconnects, signal processing will mostly still operate electrically. This requires light to be converted into electrical energy at several instances throughout a circuit, making the design of monolithically integrated CMOS photodetectors, in an optimal manner, a necessity.

As a result of the complex semiconductor device physics involved, a computer aided design approach is preferred, when designing these photodetecting devices. The discerning use of a full-fledged semiconductor device simulation package, like Silvaco/ATLAS, would yield the best results. However, such packages are not financially viable for the resource-constrained CMOS designer. This paper discusses the design of a device simulator (implemented in MATLAB), customized to CMOS photodiodes, to assist in this cause.

II. SIMULATION MODEL

A. Overview

In essence, a photodiode is a *pn*-junction. Incident photons with sufficient energy excite electron-hole carrier pairs throughout the device that are swept up by the electric field present in the depletion region, thereby generating the so-called photocurrent. Electron-hole pairs generated within the depletion region are immediately subjected to the drift transport mechanism (electric field) active within it. Carrier pairs generated outside of the depletion region (i.e. in the bulk region) first have to diffuse to the depletion region before they are swept up by the electric field present there. The model employed in this paper divides the photodiode structure into bulk regions and depletions regions, considering either exclusively diffusion or drift carrier transport in isolation, within each region, respectively.

The simulation model considers only the minority carrier perturbations from equilibrium, caused by incident photons; all equations henceforth use p and n as the excess minority carrier concentrations from their respective equilibrium values. The optical carrier generation rate within the photodiode is influenced by the photon flux and the optical wavelength (influencing the photon energy and the light's absorption lengths within the silicon substrate). The spatial distribution of light's carrier generating capability, at a depth x, is described by Lambert-Beer's law [1] and is stated as:

$$G_{opt}(x) = \frac{P_{in}}{h\nu}\alpha e^{-\alpha x} = \Phi_0 \alpha e^{-\alpha x}\,, \qquad (1)$$

where P_{in} , h, ν, α and Φ_0 denote the incident optical power intensity (on the device surface), Planck's constant, the light's frequency, the light's absorption coefficient in silicon and the surface photon flux, respectively. All other generation effects are assumed to remain constant and consequently may be ignored.

The carrier continuity equations form part of the basic semiconductor transport equations along with Poisson's equation [2]. The carrier continuity equation for the electron's case is given as:

$$\frac{\partial n}{\partial t} = R_n + \nabla \cdot (\mu_n n\bar{\boldsymbol{\varepsilon}} + D_n \nabla n) \qquad (2)$$

where n, R_n, μ_n, $\bar{\boldsymbol{\varepsilon}}$ and D_n denote the electron carrier concentration, the net recombination-generation rate, electron mobility, electric field and electron diffusivity constant, respectively. The carrier continuity equations are solved in the bulk regions using finite element analysis (FEA).

All the assumptions made in the photodiode simulation model may be summarized as follows:

- Zero electric field in bulk regions.
- Infinite electric field in depletion regions.
- No recombination within depletion region.
- Zero surface recombination velocity.
- 100% internal quantum efficiency.

Under the mentioned assumptions the generated photocurrent may be calculated and the required simulator outputs derived from these results.

B. Spatial Discretization

The calculation of the photocurrents in the various photodiode regions requires the device structure, firstly, to be subdivided into bulk regions and depletion regions. Secondly, each of these regions is further subdivided in rectangular slabs. This process is illustrated for an n-well/p-substrate structure in Fig. 1. It is important to notice that the bulk regions are subdivided in such a manner that the (at least partial) contact of a slab surface to the depletion region is ensured. This requirement stems from the boundary conditions that are enforced when solving the carrier continuity equations in the bulk regions, using FEA. The depletion region may be subdivided in any manner that facilitates the easy integration of the optical carrier generation term in this region.

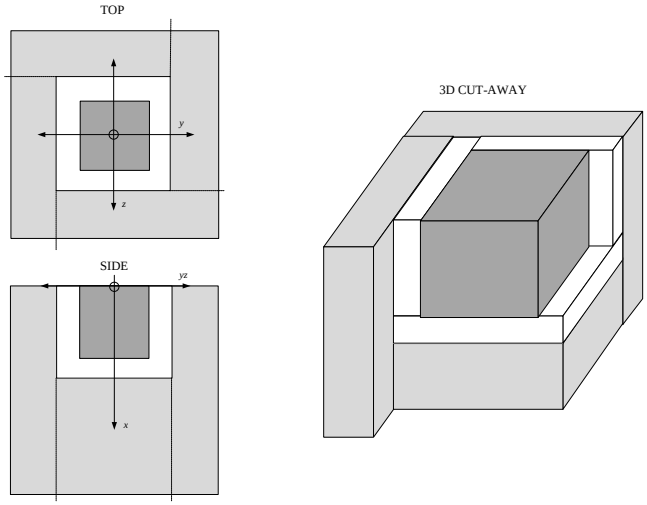


Fig. 1. Spatial discretization of an n-well/p-substrate photodiode structure using rectangular slabs to approximate the bulk regions (grey) and depletion region (white).

C. Depletion Region Photocurrent

The assumptions of an infinite electric field within the depletion regions, along with negligible recombination processes, are valid when considering the typical junction widths encountered in modern manufacturing processes [3]. Consequences of these assumptions are that each and every electron-hole pair excited within the depletion region contributes to the generated photocurrent, furthermore, this happens instantaneously. The implied infinite bandwidth of this current is a valid assumption when comparing the comparatively negligible transit times of carriers crossing the depletion region to the time minority carriers in the bulk regions require to diffuse to the depletion region. Consequently, the carrier continuity equations need not be evaluated within the depletion region.

The stated assumptions allow the photocurrent, generated within the depletion regions, to be calculated by integrating the optical carrier generation term (1) over the volume consumed by these regions. With reference to Fig. 1, an analytical evaluation of the integral for each slab (3) within the depletion region is a straight-forward affair.

$$I_{SCR} = \int_{\Omega} \Phi_0 \alpha e^{-\alpha x} d\Omega = l_y l_z \int_{x_1}^{x_2} \Phi_0 \alpha e^{-\alpha x} dx \qquad (3)$$

The newly introduced variables q, l_y and l_z denote the elemental charge quantity and the slab's width and length, respectively.

D. Bulk Region Photocurrent

The finite element method is applied to solve the carrier continuity equations for minority carriers in the bulk regions – thereby allowing calculation of the photocurrent. The analytic solution techniques used in [4, 5] limits the geometry description of the photodiode structures to two dimensions. This paper's approach of using the FEM – a numeric solution technique – allows photodiode geometry descriptions to be easily extended to three dimensions. The FEM's fundamental theory and techniques are extensively discussed in the, by now classic, text of [6].

The equation to be solved (4) – a parabolic partial differential equation – is derived from the complete continuity equation (2). The drift component is omitted and the net recombination-generation rate is expanded by a simplified Shockley-Reed-Hall model and the optical carrier generation term (1), respectively.

$$\frac{\partial n}{\partial t} = D_n \nabla^2 n - \frac{n}{\tau_n} + \Phi_0(t) \alpha e^{-\alpha x} \qquad (4)$$

$$\int_{\Omega} \phi_i \frac{\partial n}{\partial t} d\Omega = \oint_{\Gamma} \phi_i D_n \nabla n \cdot \hat{n} d\Gamma - \int_{\Omega} \nabla \phi_i \cdot D_n \nabla n d\Omega - \int_{\Omega} \phi_i \frac{n}{\tau_n} d\Omega + \int_{\Omega} \phi_i \Phi_0(t) \alpha e^{-\alpha x} d\Omega \qquad (5)$$

This equation is transformed into a form suitable for FEA (5), using Galerkin's weighted residual method. A complete discussion and development of parabolic partial differential equations into a weak formulation suitable for FEA is given in [6, 7]. Notice should be taken of the weighting function ϕ_i and the boundary integral (over surface Γ) to allow later discussion.

Equation 5 needs to be solved for n, the electron carrier density. As the name suggests, FEA entails dividing the bulk region into a finite number of smaller elements – a process known as meshing – wherein the solution to the equation is approximated using simple mathematical functions (mostly linear polynomials). Tetrahedral elements (the 3D analogue to triangles) are formed within each slab of the bulk region using Delaunay tessellation. This process entails placing nodes throughout the domain to be meshed and then associating each node with the three nodes closest to it (its natural neighbors) thereby describing the formed tetrahedrons using a nodal connectivity matrix. As is customary, the vertices of the tetrahedron are used as nodal points.

The solution to the equation under investigation may be approximated over a single element as follows.

$$n \approx \sum_{j=1}^{4} L_j n_j \qquad (6)$$

where n_j and L_j denote the carrier concentration at node j and the shape function associated with this node, respectively. The shape function weighs the contribution of node j to the total solution over the element. The shape functions are chosen to be linear polynomial functions which are expressed as follows.

$$L_j = a_j + b_j x + c_j y + d_j z \quad (7)$$

Shape functions assume a value of unity at their associated node and zero at the other nodes (as dictated by FEM theory [6, 8]). Exploiting this fact, along with the known Cartesian coordinates of the nodes, the coefficients of the shape function (a_j to d_j) may be determined by forming a matrix from the resulting four, independent linear relations and then inverting it as shown.

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{bmatrix}^{-1}$$

By substituting the approximation of (6), (5) is expressed for a single element as:

$$\sum_{j=1}^{4} \int_\Omega \phi_i \frac{\partial}{\partial t} L_j n_j d\Omega + D_n \int_\Omega \nabla\phi_i \cdot \nabla L_j n_j d\Omega + \frac{1}{\tau_n} \int_\Omega \phi_i L_j n_j d\Omega = \Phi_0(t) \int_\Omega \phi_i \alpha e^{-\alpha x} d\Omega \quad (8)$$

The boundary integral may be omitted here since inter-element boundaries' contributions 'cancel out' [8]. The time derivative may be approximated by the backward-difference scheme:

$$\frac{\partial}{\partial t} n_j \approx \frac{n_j^k - n_j^{k-1}}{\Delta t}, \quad (9)$$

where n_j^k and n_j^{k-1} denote the values of n_j at time-steps k and k-1, respectively; Δt denotes the size of the time-step.

The Galerkin FEM formulation (5) of a differential equation implies that the shape functions are also employed as weighting functions. By substituting (9) into (8) and now also using the shape functions as weight functions, (8) may be expressed as:

$$([A] + \Delta t[B])\{n\}^k = [A]\{n\}^{k-1} + \Delta t\{P\}^k \quad (10)$$

where

$$A_{ij} = \int_\Omega L_i L_j d\Omega \quad (11)$$

$$B_{ij} = D_n \int_\Omega \nabla L_i \cdot \nabla L_j d\Omega + \frac{1}{\tau_n} \int_\Omega L_i L_j d\Omega \quad (12)$$

$$P_i = \int_\Omega L_i \alpha e^{-\alpha x} d\Omega \quad (13)$$

Note, *i* and *j*, are the indices for the weight function and the shape function, respectively. By appropriate addition and multiplication (since $\{n\}^{k-1}$ and {P} are known) the system can be simplified by taking n_j as a common factor in (10). By considering all possible combinations of weight functions in the set $\{\phi\}$, the equation may be expressed as a linear system of the form:

$$[K]\{n\} = \{F\}, \quad (14)$$

where [K] is the stiffness matrix (4x4), {n} denotes a row vector (4x1) of the nodal carrier concentration values and {F} is the force vector (4x1). The integrals encountered in (11) to (13) may be solved using numerical integration techniques like Gaussian quadrature or, if volume coordinates are used (as described in [6]), integration formula.

Each element makes a contribution to the total solution of the global system. By assembling the local element matrices and vectors into a global system, the total solution to *n* may be found by enforcing specific boundary conditions. Each element (tetrahedron) is formed by a set of four nodes. If these nodes carry global node numbers of 1, 3, 4 and 7 (say), they will contribute to the elements of [K] (global) where the correspondingly numbered rows and columns intersect; the contribution to the force vector is similarly limited to rows 1, 3, 4 and 7.

Two types of boundary conditions are prevalent for this particular case: Dirichlet and Neumann type boundaries. Dirichlet boundary conditions specify the value of the variable to be solved (*n*) on a domain boundary (Γ_1) whereas with Neumann boundary conditions the variable's gradient (∇n) in the outward normal direction on the boundary (Γ_2) is specified ($\Gamma_1 \cup \Gamma_2 = \Gamma$).

Any part of a bulk region's boundary surface that is in direct contact with the depletion region should be classified as a Dirichlet boundary, with a boundary value of zero. This is justified by the fact that any minority carrier that diffuses up to this boundary is immediately swept up by the electric field present within the depletion region. The boundaries at the bottom of the device are assigned Dirichlet boundary conditions with a value of zero. This is justified by the physical consideration that any carriers present deep in the substrate will recombine there, as they are too far from the depletion region to successfully contribute to current generation by diffusing to the depletion region.

The boundaries that lie on the top surface of the device ($x = 0$) are assigned Neumann boundary conditions with a zero value. This is justified by the fact that the surface recombination taking place is a comparatively slow process compared to the other carrier movement and recombination processes at the frequencies of interest [9]. The remaining lateral boundaries are of the Neumann type, also assigned a value of zero. No lateral carrier gradient is evident, as equal amounts of light (implying generated carriers) are received at a given depth.

All the discussed Neumann boundaries are zero, consequently the boundary-integral introduced in (5) never needs to be evaluated and can be omitted. The Dirichlet boundary conditions are enforced by appropriately modifying [K] and {F} of the global system. By inverting the stiffness matrix, the solution of {n} follows:

$$\{n\} = [K]^{-1}\{F\} \quad (15)$$

With the system solved for the primary variable, *n*, the flux on the Dirichlet boundaries may be calculated using the unmodified stiffness matrix and force vectors (i.e. before the boundary conditions were enforced – modifying [K] and {F} in the process).

$$\{flux\} = \{n\}[K] - \{F\} \text{ and} \quad (16)$$

$$\{flux\} = \int_{\Gamma_1} \phi_i D_n \nabla n \cdot \hat{n} d\Gamma_1, \quad (17)$$

where Γ_1 denotes the Dirichlet boundary surface.

When considering the diffusion current density equation, defined as part of (2), a strong resemblance to the flux integral (17) is immediately evident. In fact, by simply multiplying the calculated flux value with q (elemental charge quantity), the total charge flowing through the Dirichlet boundary in time Δt can be calculated by adding all the boundary nodes' contributions. From this result the current may be calculated by dividing the charge amount with Δt to obtain an answer in units of Amperes. Note that the current density term is already implicitly integrated over the Γ_1 boundary surface, thus the current and not the current density is obtained. The total current contributed by the entire bulk region can be determined in this manner by only considering the nodes on the surfaces in contact with the depletion region.

The same principle holds for all bulk regions and for the hole carrier case.

III. SIMULATOR OUTPUTS

The preceding discussions allow the photocurrent to be calculated and, consequently, the desired simulator outputs to be devised.

A. Transient Response

The transient response of the photodiode is obtained by solving for the photocurrent over various time steps under a time-varying optical excitation in the form of a clock pulse train. An example of obtained results is shown in Fig. 2.

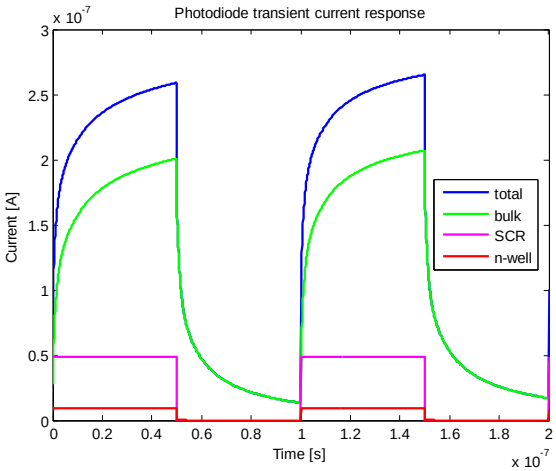


Fig. 2. Simulated transient response of an n-well/p-substrate structure.

B. Frequency Response

The frequency response of a photodiode is limited by both its physical (intrinsic) bandwidth and its electrical (extrinsic) bandwidth. The intrinsic bandwidth relates to the finite speed with which the carriers within the device react to an optical excitation, whereas the electrical bandwidth is determined by the RC time constant formed by the photodiode's junction capacitance and the input impedance of the subsequent transimpedance amplifier.

The junction capacitance of a specific photodiode structure may be estimated from process parameters supplied by the foundry. This approach fails to account for the reverse bias conditions under which photodiodes are mostly operated. This can be successfully done by approximating the junction as a parallel plate capacitor using (18) as given in [10].

$$C_j = \frac{A\epsilon}{W} \tag{18}$$

where A, W and ϵ denote the effective area of the junction (plate), the junction width (plate separation) and the permittivity of Silicon.

The intrinsic frequency response of the photodiode is obtained by exciting the device with a Dirac impulse. By performing a Laplace transform on (8) (assuming zero initial conditions) one obtains (19). This replaces the time dependent terms with their appropriate Laplace transforms, thereby allowing a direct evaluation of the equation in the frequency domain. The spatial discretization of the equation is retained. This allows (19) to be solved using the same FEM approach as for the transient case.

$$\int_\Omega \phi_i s n \, d\Omega + \int_\Omega \phi_i \Phi_0 \alpha e^{-\alpha x} d\Omega = \int_\Omega \phi_i D_n \nabla^2 n d\Omega - \int_\Omega \phi_i \frac{n}{\tau_n} d\Omega \tag{19}$$

By the substitution $s = j\omega$, the frequency response of the photodiode may be calculated by solving for the photocurrent over the frequency range of interest. The photocurrent generated by the depletion region is assumed to have infinite bandwidth and consequently its value will not vary with frequency.

C. Spectral responsivity

The spectral responsivity of a photodiode relates the steady-state current generated to the wavelength of the incident light. The same approach as for the frequency response case is used; the frequency is held constant at 0 Hz while the wavelength is varied over the spectral range of interest (250 nm to 1110 nm). The wavelength is implied in (1) since it is related to the light's frequency. The responsivity values may be normalized to the maximum value to allow easy comparison between different photodiode structures and/or geometries.

IV. RESULTS

A. Transient Response

The magnitude of the photocurrent predicted by simulation of transient responses was validated by devising a spectral responsivity graph (by simulation at various wavelengths) and comparing this to the results measured and simulated in [11].

The results shown in Fig. 3 illustrate that the simulated photocurrent magnitude is underestimated, especially in the 600 nm to 850 nm spectral range. The general trend, however, is correctly predicted.

B. Frequency Response

The junction capacitance values predicted by the simulator are in good agreement with the values expected from the process parameters (for the zero bias case).

Published results on the subject of the intrinsic frequency response of CMOS photodiodes are sparse as CMOS compatible photodetectors were not of particular interest until quite recently. The frequency response of photodiodes is

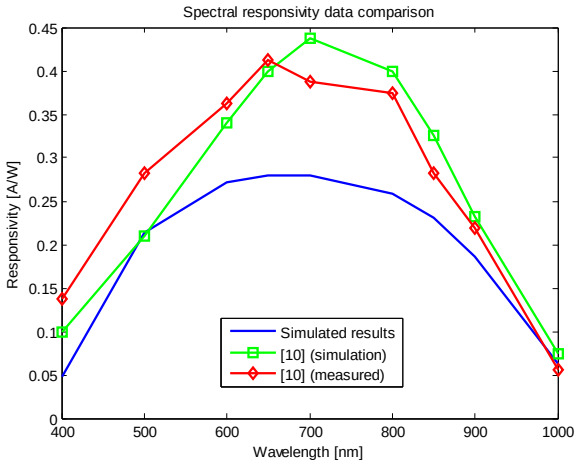


Fig. 3. Spectral responsivity plot comparing results from [11] to those obtained from transient response simulation for an n-well/p-substrate structure.

highly dependent on the device structure and process technology used. The simulated results shown here are based on a 0.35 µm process. The simulation results obtained in [4, 9] for a 0.18 µm process serve as a comparison to the attained simulation results.

A comparison between the bandwidth values simulated to those reported in [4, 9] is made in table 1. The results, generally, fall within the same order of magnitude and show agreement in the predicted behavior of the various regions' current components. Since the simulation model used in [4] works on the assumption that the lateral depletion regions are negligible for the twin-tub process considered, the large deviations at short wavelengths compared to the simulation results presented in this paper are partly explained. This has a significant effect at short wavelengths, since the carrier generation capacity of the incident light is restricted very close to the surface. As a result of the high absorption at wavelengths below 500 nm the contribution of the lateral depletion regions are significant for the 0.35 µm process considered here.

The p+/n-well/p-substrate structure shows increased intrinsic bandwidth compared to the n-well/p-substrate structure as is evident from Fig. 4 and Fig. 5. This may be explained by the fact that the added p-diffusion forms an extra depletion layer resulting in a higher percentage of the total photocurrent being contributed by the depletion regions (and its associated 'infinite bandwidth').

C. *Spectral Responsivity*

The normalized spectral responsivity simulated is compared to the normalized values of [11] as used for comparison in Fig. 3. The results are shown in Fig. 6 for an n-well/p-substrate structure. The wavelength of peak responsivity is correctly predicted – which is the most important information to be gathered from such a graph.

V. CONCLUSION

This paper has presented a CMOS photodiode model suitable for the purpose of device simulation to aid in the design of optimal photodetectors. The simulator is capable of predicting the critical performance parameters of a three dimensional

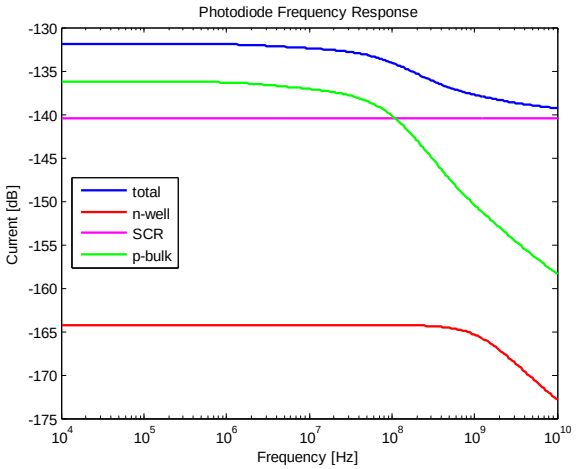


Fig. 4. Intrinsic frequency response of an n-well/p-substrate structure (650 nm, -0.8 V bias). Total bandwidth of 168 MHz.

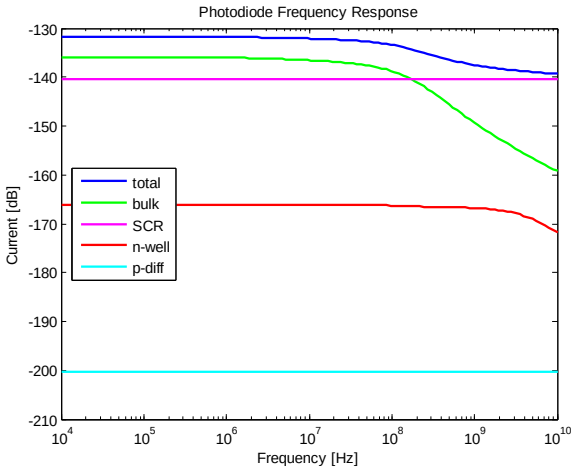


Fig. 5. Intrinsic frequency response of an p+/n-well/p-substrate structure at 650 nm and a -0.8 V bias. Total bandwidth of 217 MHz.

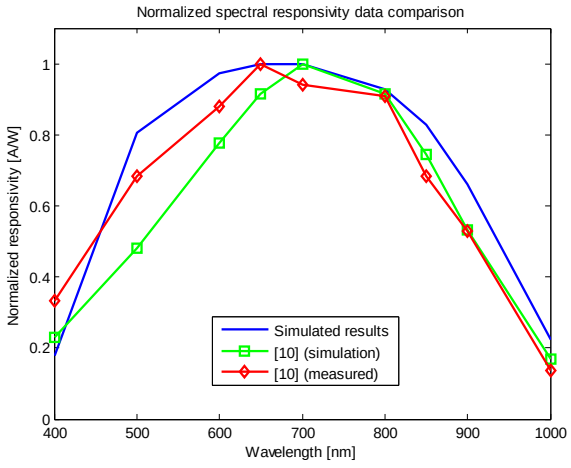


Fig. 6. Normalized spectral responsivity simulation results compared to values from [11]

CMOS photodiode structure. A simplified form of the carrier continuity equation was solved for the bulk regions using the finite element method opposed to the generally adopted analytic solution techniques.

The assumptions made for the presented simulation model prove to be problematic under circumstances when the total generated photocurrent is dominated by the depletion region's

contribution. This may arise when short optical wavelengths (<500 nm) and/or high reverse bias voltages are applied. Furthermore, a notable underestimate of the generated photocurrent is made within the previously mentioned spectral range.

The overall results obtained by the designed simulator proved to be acceptable. Correct behavioral trends were obtained for all the desired outputs. The absolute values of simulated results fall within the same order of magnitude, compared to the results available in literature. This allows the developed simulator to be used effectively in the computer aided design of optimal CMOS photodiode structures.

ACKNOWLEDGMENT

A sincere thanks is extended to the Carl and Emily Fuchs Institute of Microelectronics (University of Pretoria, South Africa) and its staff for the continued support in completing the project on which this paper is based.

REFERENCES

[1] T. Naeve, "A quasi-2-dimensional photodiode model for high-speed short-wavelength applications," *Electron Devices for Microwave and Optoelectronic Applications, 2004. EDMO 2004. 12th International Symposium on,* pp. 66-70, 2004.

[2] S. Selberherr, *Analysis and Simulation of Semiconductor Devices.* Springer, 1984, pp. 293.

[3] S. Radovanovic, "On optimal structure and geometry of high-speed integrated photodiodes in a standard CMOS technology," *Lasers and Electro-Optics, 2003. CLEO/Pacific Rim 2003. the 5th Pacific Rim Conference on,* vol. 1, pp. 87 Vol.1, 2003.

[4] S. Radovanović, A. Annema and B. Nauta, "Bandwidth of integrated photodiodes in standard CMOS for CD/DVD applications," *Microelectronics and Reliability,* vol. 45, pp. 705-710, 0. 2005.

[5] T. Naeve, M. Hohenbild and P. Seegebrecht, "A new analytical compact model for two-dimensional finger photodiodes," *Solid-State Electronics,* vol. 52, pp. 299-304, 2. 2008.

[6] O. C. Zienkiewicz, R. L. Taylor, J. Z. Zhu, O. C. Zienkiewicz and O. C. Zienkiewicz, *The Finite Element Method : Its Basis and Fundamentals.* ,6th ed.Oxford ; Boston: Elsevier Butterworth-Heinemann, 2005, pp. 733.

[7] Y. W. Kwon and H. Bang, *The Finite Element Method using MATLAB.* ,2nd ed.Boca Raton, Fla.: CRC Press, 2000, pp. 607.

[8] J. N. Reddy, *An Introduction to the Finite Element Method.* New York: McGraw-Hill, 1984, pp. 495.

[9] S. Radovanović, A. Annema and B. Nauta, *High-Speed Photodiodes in Standard CMOS Technology.* ,1st ed., vol. 869, Dordrecht, The Netherlands: Springer, 2006, pp. 160.

[10] G. W. Neudeck, *The PN Junction Diode.* ,2nd ed., vol. 2, Reading, Mass; 1993: Addison-Wesley; Milton Keynes, 1989, pp. 180.

[11] Wei-Jean Liu, "A CMOS photodiode model," *Behavioral Modeling and Simulation, 2001. BMAS 2001. Proceedings of the Fifth IEEE International Workshop on,* pp. 102-105, 2001.

TABLE I

SIMULATED RESULTS COMPARED TO [4, 9]

	Bandwidth [MHz]					
	Total		**Bulk**		**n-well**	
Wavelength [nm]	*Simulated*	[4, 9]	*Simulated*	[4, 9]	*Simulated*	[4, 9]
400	10000	800	105	-	1831	700
650	164	160	75	28	2078	2000
780	33	8	5	4.5	2136	1200
850	8	5	5	3	2136	850

A CMOS Based Multiple-Access DSSS Transceiver

S. Sinha[1], *Member, IEEE,* M Božanić[2], *Student Member, IEEE,* J. Schoeman[3], *Member, IEEE,*
M. du Plessis[4], *Senior Member, IEEE,* and L.P. Linde[5], *Senior Member, IEEE*
[1, 3, 4, 5]Department of Electrical, Electronic & Computer Engineering, Carl & Emily Fuchs Institute for Microelectronics, University of Pretoria, Pretoria, 0002, South Africa
[2]Azoteq (Pty) Ltd, P O Box 16767, Lyttelton 0140, South Africa
[1]ssinha@ieee.org, [2]mbozanic@ieee.org, [3]johan.schoeman@eng.up.ac.za,
[4]monuko@up.ac.za, [5]louis.linde@up.ac.za

***Abstract*— This paper presents a fully-integrated CMOS system that uses complex spreading sequences as spreading codes in a direct sequence spread spectrum transceiver. The system can operate either as a receiver or as transmitter, in balanced or unbalanced configurations, to accommodate for higher noise immunity or higher data rate. Simulated operation has been achieved for the 2.4 GHz industrial, scientific and medical band with maximum possible data rate of 6.17 Mbits/s. A 10 dBm onchip power amplifier for the transmitter and low-noise amplifier for the receiver are included. The system provides a CMOS based replacement to the discrete-component systems previously proposed. The contribution of the paper lies in the circuit level modifications done at sub-system level aimed towards eventual integration. For multiple-access communication systems, where a number of independent users are required to share a common channel, the transceiver proposed in this paper, can contribute towards improved data rate or bit error rate. The design is completed for fabrication in a standard 0.35-µm CMOS process with minimal external components. With an active chip area of about 5 mm^2, the simulated transmitter consumes about 125 mW and the receiver consumes about 97 mW.**

I. Introduction

LARGE numbers of new transmission techniques have been proposed in more recent telecommunication history, with direct sequence spread spectrum (DSSS) being one of them. In DSSS, before a signal is transmitted in the conventional way by means of a digital modulation technique, it must be spread in the frequency domain by a spreading code. In this way the bandwidth of a signal increases, but the transmitted signal is superimposed onto an orthogonal base, so that transmission of more signals in one frequency band does not call for frequency division multiplexing, but they can simply be transmitted and extracted.

Conventional DSSS transceivers use binary spreading sequences as spreading codes. Spreading codes are chosen in such a way that different transmitters operating in the same band use mutually orthogonal codes for spreading. If complex spreading sequences (CSS) is used instead of binary spreading sequences, a greater number of these orthogonal bases can be created for smaller increase in bandwidth [1], giving a faster and highly noise immune system.

DSSS transceivers that implement CSS have been reported before [1-2], but the size of these devices prevents any other use but in the laboratory. To decrease the size and cost, an IC design was proposed.

II. Design Specifications

Most of the design specifications are influenced by the bandwidth of the transmission channel and the CSS length. Therefore, the two specifications will be treated separately, and resulting additional specifications will also be given.

A. Transmission Bandwidth

The transmission bandwidth is determined by the channel bandwidth. The industry, scientific and medical (ISM) band allows for maximum bandwidth of 22 MHz. To accommodate for small filtering errors, the proposed maximum available transmission bandwidth chosen is 20 MHz.

B. Complex Spreading Sequences

The types of CSS used are generalized chirp-like sequence families [3] and specifically, Barker sequence family [2]. These are root-of-unity sequences, with perfectly flat spectrum and excellent auto- and cross-correlation properties. The sequences are created digitally, with the number of chips or the length of the sequence equal to number of discrete voltage values of one sequence repetition. This length can be chosen arbitrarily as any prime number between 11 (since shorter sequences do not posses enough information for successful data recovery) and 511 (as the transmission data rate decreases with the increase in sequence length and systems with a longer sequence would be too slow). Longer sequences however have the advantage of giving the noise immunity to the system and allowing larger number of users within the band.

For this paper, a sequence length of 13 was chosen to

accommodate higher data rates, and the sequences are sampled at the maximum rate of 160 Msamples/s (eight samples/chip).

C. Additional Specifications

The following is a list of other specifications for the system.

- Number of orthogonal bases is four (CSS real part and quadrature carrier, CSS real part and in-phase carrier, CSS imaginary part and quadrature carrier and CSS imaginary part and in-phase carrier).
- Four orthogonal bases dictate the use of QPSK (quadrature phase shift keying) as the modulation technique.
- Transmitter and receiver can operate in both balanced and unbalanced configurations. Balanced configuration provides for higher noise immunity, whereas the data rate can be increased up to four times the nominal if unbalanced configuration is used.
- Maximum data rate prior to transmission is equal to bandwidth divided by the chosen sequence length, or 20 MHz ÷ 13 = 1.54 Mbits/s. Lower data rate (such as 1 Mbits/s) can be used if the sequence sampling frequency is decreased accordingly.
- Transmitter power is 10 dBm and receiver sensitivity is -71 dBm [4].
- Bit-error rate is controllable.
- Single power supply with voltage level of 0 V for ground and 3.3 V for V_{DD} is used, and power dissipation is limited to typical figure of 120 mW.

III. Design Methodology

Detailed MATLAB/Simulink model was developed that allowed for full mathematical simulation of the system (both the transmitter and receiver at the same time) together with the transmission channel. Except for allowing for these simulations, this ideal model allowed for comparison of the simulation results with the non-ideal (including various parasitic effects) results later obtained in SPICE.

Each subsystem was further developed and simulated in SPICE where each Simulink block was replaced by a MOS transistor based circuit. To complete the design, a layout level implementation was achieved.

IV. Conceptual Design

Conceptual design was done prior to performing the simulations in Simulink and SPICE. The design was done around the core blocks of the system, which are the spreading and modulation for the transmitter and despreading and demodulation for the receiver. Mathematical equations that describe the transmitter are given in [2].

If the transmitter operates in balanced mode, four orthogonal signals are created by multiplying the data signal with all four combinations of sine and cosine carriers and CSS real and imaginary parts. These four signal are added, resulting in

$$u(t) = a(t)C_r \cos\omega_c t + a(t)C_i \cos\omega_c t + a(t)C_r \sin\omega_c t + a(t)C_i \sin\omega_c t \quad (1)$$

where $u(t)$ denotes the transmitted signal, $a(t)$ is the digital transmitted signal in NRZ (non-return-to-zero) form (values of -1 and 1), C_r and C_i CSS real and imaginary parts and ω_c the carrier frequency. For the transmitter to operate in unbalanced mode, $a(t)$ is replaced with four different signals, $a_1(t)$ through $a_4(t)$.

A set of equations similar to (1) can be developed for the receiver, which describes how the balanced signal or four unbalanced signals can be recovered from the transmitted signal. As an example, $a(t)$ can be recovered from the quadrature branch spread by CSS real part as in (2):

$$a_1(t) = \begin{cases} 1, T_b < t < 2T_b \text{ iff } \left\{ \int_0^{T_b} \left[u(t)C_r \cos\omega_c t \otimes \Im^{-1} \prod\left(\frac{f}{BW}\right) \right] \right\}\Big|_{T_b} > 0, \\ -1, T_b < t < 2T_b, \text{otherwise} \end{cases} \quad (2)$$

For successful recovery, bandwidth, BW, and bit period, T_b must also be known.

Fig. 1 depicts the complete conceptual block diagram, where the core blocks are shown together with the supporting blocks that add to the functionality of the system.

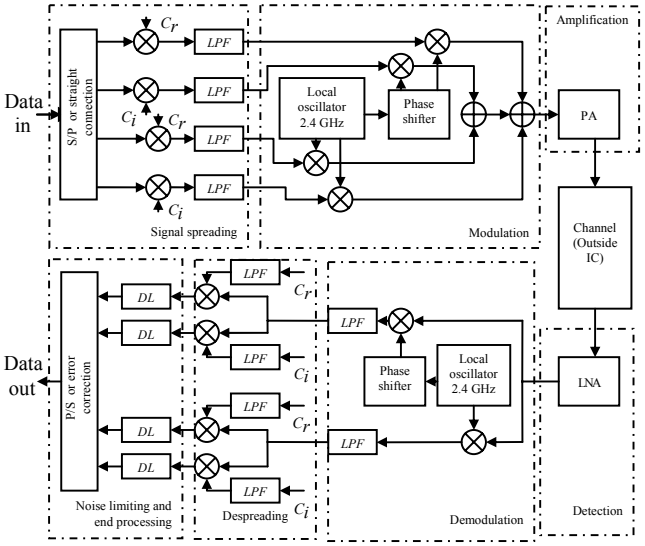


Fig. 1: Block diagram of the full transceiver system (including the channel) showing the transmission and recovery processes

In the transmitter, the data signal that has to be transmitted is first taken through the digital processing block that distributes the data along the four branches of the system (depending on whether the user has chosen the balanced or unbalanced configuration). From this point onwards, signal processing in the transmitter is fully analogue. Each of the four signals is spread with either the CSS real or imaginary part. Spreading is followed by modulation, in which the four spread signals are modulated onto quadrature or in-phase carriers, creating orthogonal signals. Finally, the signals are added together to form one signal. Power amplification is the final step of the transmission process, raising the power level to one suitable to feed the antenna (10 dBm).

In the receiver, a reversal of the signal processing functions takes place. Weak signals (sometimes with the power as low as -71 dBm) detected from the channel are amplified once again, with the aid of a LNA (low-noise amplifier). The LNA is followed by demodulation and despreading. These are conceptually similar to the spreading and modulation sub-systems of the transmitter.

Recovery of the data from the demodulated and despread signals needs to be done by means of filtering. Various data recovery techniques are available [5], with the most sophisticated one being integrate and dump. If the bit period is known, integration can be performed to create ramps in the direction of the transmitted bit. Sampling at the end of those bit periods can be used to recover the data in parallel branches. The parallel-to-serial conversion or error correction serves to finally re-generate the transmitted signal.

V. MIXING

Mixing is the method used to spread, modulate, demodulate and despread the signals. Two types of mixers can be identified: one that multiplies a digital signal (signal that can take a value of either 0 or 1) with an analogue signal, and the other that multiplies any two analogue signals with each other. Analogue mixing is conventionally achieved by analogue multiplication, whereas digital mixing can be simplified, by using switching logic. Use of the switching technique is made for spreading: if 0 is to be transmitted, an inverted version of the sequence is chosen, while if 1 was transmitted, a non-inverted version is chosen. This simplified the design of the transmitter considerably, eliminating the need for changing the digital logic voltage levels to the ones suitable for analogue mixing.

VI. CMOS IMPLEMENTATION

Six distinct sections of the system are visible. These are the digital, low-frequency analogue and RF sections of the transmitter and receiver.

A. Digital Section of the Transmitter

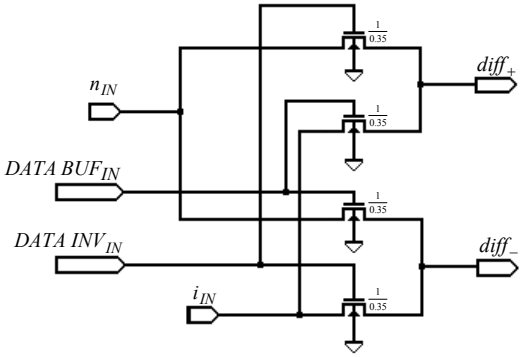
Digital switching is done to activate the serial-to-parallel converter. The serial-to-parallel converter is driven by the clock that is recovered from the data signal. The recovery is digital, and consists of edge detection and counter. Whenever the edge in the signal is detected, the counter resets. If there is no edge, the counter runs in free mode, by means of the external clock. This external clock needs to run approximately eight times faster than the recovered clock, as it is just a reference for the counter running in free mode, but small drift from the ideal speed is allowed [6].

B. Analogue Section of the Transmitter

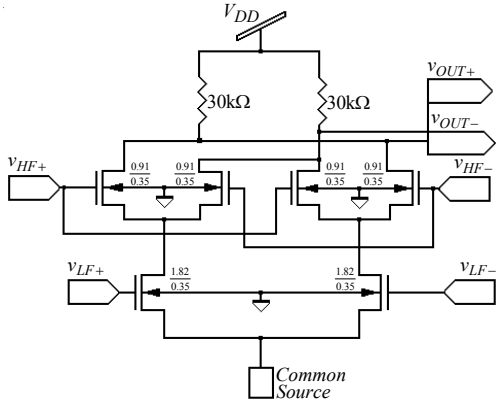
The analogue section entails all sub-systems that accomplish spreading and modulation. The internal oscillator generates the carrier signals for modulation. All the mixers used are differential, so two pairs of carriers are required, two 90 degrees out-of-phase and inverted versions of the two. An oscillator configuration with phases of 0, 90, 180 and 270 degrees is presented in [7] and adapted for the abovementioned purpose. For the layout implementation of this sub-system, polysilicon capacitors and spiral inductors were used for design simplicity.

The transmitter utilizes the types of mixing techniques (in a differential configuration) described in V. Two sets of signals serve as inputs, and the output is the mixing result. Fig. 2a) shows the mixer that multiplies analogue and digital signals and Fig. 2b) shows the one mixing two analogue signals. The mixer of Fig. 2a) uses a CMOS switch that lets signals pass through if there is digital logic of 1 (3.3 V) at the gate of the transistors. The mixer of Fig. 2b) is the differential configuration of a Gilbert mixer that uses the second order effects of i_D vs. v_{GS} curve [8]:

$$i_D = k_n' \frac{W}{L}(v_{GS} - V_t)v_{GS} + \frac{1}{2}k'_n v_{GS}^2 \qquad (3)$$



(a)



(b)

Fig 2: (a) Mixer with a pair of digital and a pair of analogue inputs, and (b) differential Gilbert mixer

Real and imaginary parts of sequences [2] are input externally to the chip with no DC offset, and processing circuitry is included to add the DC offset suitable for processing (1.65 V) and to generate the inverted versions for use in differential configurations.

Summing of four orthogonal signals created is done by utilisation of Kirchhoff's current laws.

C. RF Section of the Transmitter

Power amplifier is the RF sub-system of the transmitter. Power amplification is done in three stages. The first two stages are necessary for voltage amplification. The final stage achieves the required current amplification, which effectively amplifies the power (assuming no voltage loss). Single ended Class-E stage power amplifier is used. The design equations for this stage are given in [9] and [10]. If these are followed, optimum power will be delivered to the load of

$$R_L = 0.577 \frac{V_{DD}^{\ 2}}{P_O} \tag{4}$$

where P_O is the desired output power. Since this value was not equal to 50 Ω, the modeled impedance of the antenna, impedance matching was performed. A resulting power of 10 dBm was achieved.

D. RF Section of the Receiver

LNA is used for signal amplification, since very low power signals (sensitivity of -71 dBm to -21 dBm) are detected from the ideal channel. A differential [11] implementation of the LNA assists to achieve the maximum noise rejection.

E. Analogue Section of the Receiver

The analogue part of the receiver consists of two mixers for demodulation and despreading. Both mixers are differential Gilbert mixers. During demodulation, no filtering is required as all double frequency components are attenuated later in the circuit.

Integrate and dump is done after another stage of amplification. Simple transistor differential amplifier configuration was used. Full integrate-and-dump circuitry is shown in Fig. 3a). Fig. 3b) shows the circuit diagram of the operational amplifier designed for integration [12]. A Deboo integrator [13] configuration is used. This integrator is reset by shorting its output to reference voltage (1.65 V) by means of a transistor that switches on for a small period of time when a clock edge is detected. A simple comparator is used to saturate the voltages towards one of the two power rails and allow for the sampling into digital signal. The shape of the ramps will be shown later in Fig 9.

Layout of the analogue part of the receiver is shown in Fig. 4.

F. Digital Section of the Receiver

The digital sub-systems in the receiver are responsible for data recovery and their interpretation depending on the mode of operation. For the unbalanced case, parallel-to-serial converter is activated, and error correction circuitry in the case of a balanced configuration. In the balanced configuration, all four branches of recovery should have identical data; if this is not the case, error recovery assumes that the correct bit is the one that was detected in more branches. In the case when two 0s and two 1s are detected, 0 is assumed as there is no way to guess the value.

An external clock is required for the parallel-to-serial converter to operate. The same clock is divided and used for integrate and dump in the analogue part.

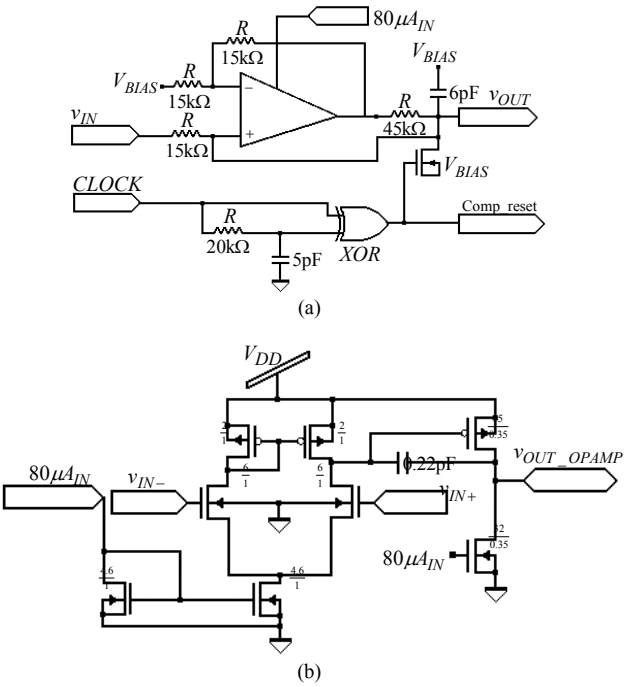


Fig. 3: (a) Integrate and dump circuit, and (b) operational amplifier

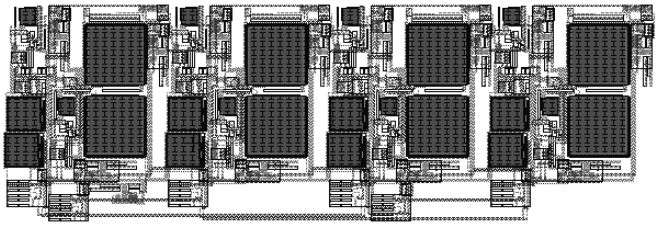


Fig. 4: Layout of the analogue part of the receiver. All four branches are shown.

G. A Note on Synchronisation

Except for the clock recovery described in section VI, no synchronisation is performed. Therefore, all signals should be perfectly aligned for the system to operate. It is the intention of this system to eventually include synchronization [2], [13].

VII. Simulated Results

Full transmitter and receiver simulation of the system was done in SPICE. An unbalanced mode of operation was used. Simulation was done until 10 μs. Figure 5 shows a CSS of length 13, sampled at frequency of 104 Msamples/s [2]. Data rate was set accordingly to 1 Mbits/s, and clock periods to 0.125 μs for transmitter and 1 μs for the receiver.

Figure 6 shows the digital data signal that has been spread with the CSS real part and filtered.

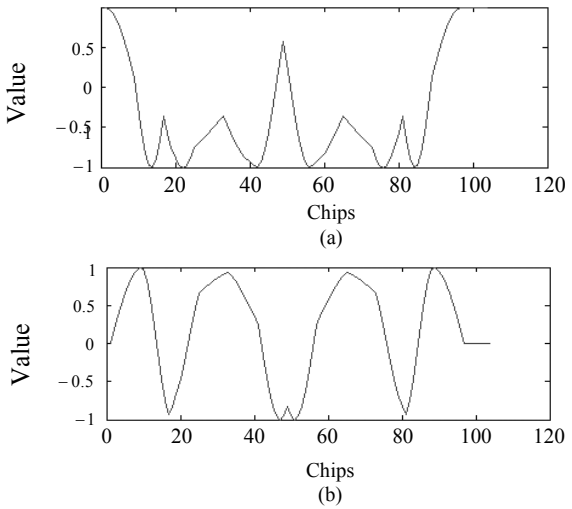


Fig. 5: Complex Spreading Sequence used in the simulation: (a) real part, and (b) imaginary part

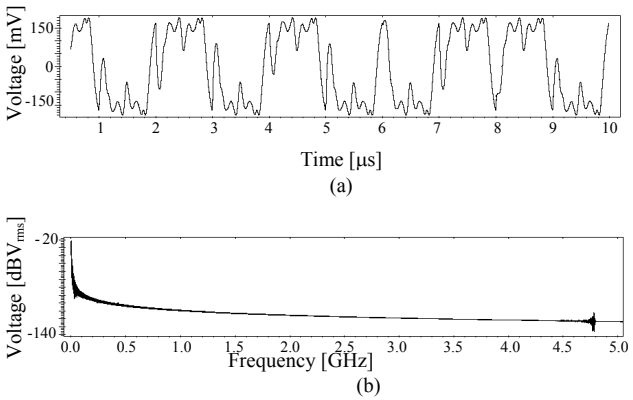


Fig. 6: Signal in one of the branches of the transmitter after it has been spread by the CSS and filtered: (a) time domain, and (b) frequency domain

Fig 7 shows the final transmitted signal, with maximum power of 10 dBm. This is the signal resulting after four digital signals were spread by different combinations of CSS real and imaginary parts and carriers, and added (as shown in Fig. 1). A carrier frequency of 2.4 GHz is achieved.

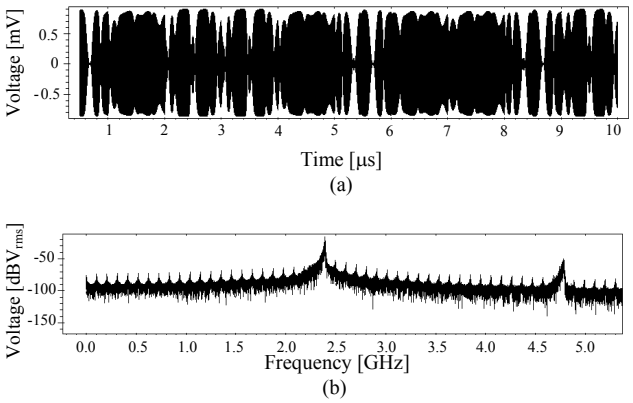


Fig. 7: Final transmitted signal in (a) time domain, and (b) frequency domain

A signal with power of -71 dBm was used to simulate the receiver. Figure 8 shows the signal in one of the four branches of the receiver after demodulation and despreading. This figure shows that the signal will be closer to V_{DD} if bit 1 was transmitted and closer to ground if bit 0 was transmitted.

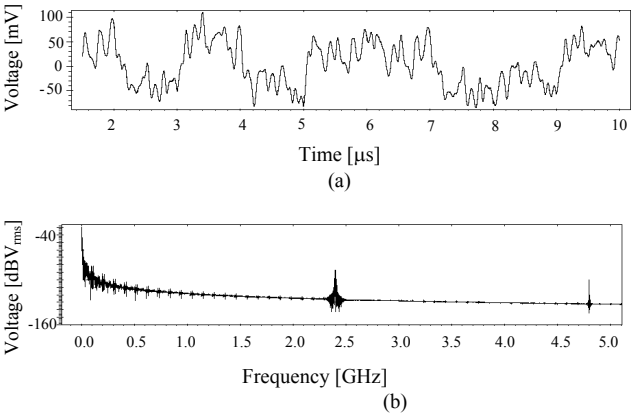


Fig. 8: Signal after demodulation and despreading in (a) time domain, and (b) frequency domain

Figure 9 shows the shape of the signal if it has been processed by the integrate-anddump circuitry. This signal is sampled at the end of bit periods, where the distance of ramps is the furthest away from the reference voltage.

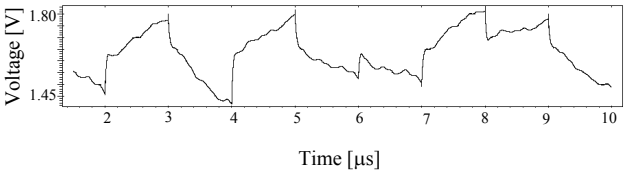


Fig. 9: Signal after the integration block. Note that the signal ramps towards the correct bit value (at the end of the bit period).

The total power consumption for the receiver was 97 mW and 125 mW for the transmitter. Table I shows the details of the signals used and recovered in simulation.

TABLE I
SIMULATION PARAMETERS AND RESULTS

Parameter	Value
CSS length	13
CSS Sampling frequency	104 Msamples/s
Mode of operation	Unbalanced
Data rate (after serial-to-parallel conversion)	1 Mbits/s
Oscillator frequency	2.4 GHz
Transmission power	10 dBm
Power of the pseudo-received signal	-71 dBm
Clock period (transmitter)	0.125 μs
Clock period (receiver)	1 μs
Power consumption (transmitter)	125 mW
Power consumption (receiver)	97 mW
Duration of simulation	10 μs
BER	0

VIII. CONCLUSION

The device presented in this paper is a highly innovative system for communication in a wireless LAN (local area network) link. Data rates of up to 4 Mbit/s are accomplished over 20 MHz bandwidth. The system can be expanded for even faster operation if more orthogonal CSS are added and expanding the size of the parallel-to-serial converter. Whereas 2.4 GHz operation is demonstrated, operation at any other frequency in ISM band is possible. Orthogonal sequences can be used to accommodate larger number of users. The data rate flexibility also allows for various data speed demands. Full integration and small power consumption allows for the chip to be installed in portable devices or even powered by batteries, such as in environments where mobility is essential (e.g. in hospitals). As the power amplifier is integrated, no external PA IC is required. The adjustable sensitivity of the receiver and ability of the error correction circuitry (for balanced operation) serves to compensate for fading.

REFERENCES

[1] F.E. Marx and L.P. Linde, "Theoretical analysis and practical implementation of a balanced DSSS transmitter and receiver employing complex spreading sequences", In *Proc. of 4th IEEE Africon: Vol. 1,* Stellenbosch, 1996, pp. 402-407.

[2] F.E. Marx, "DSSS communication link employing complex spreading sequences", Master dissertation, School of Engineering, University of Pretoria, Pretoria, South Africa, 2005.

[3] S.H. Kratzet, "A comparison of two direct sequence spread spectrum (DSSS) systems". App. note AN116B, 2000.

[4] B.M. Popović, "Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE trans. IT,* 38, pp. 1406-1409, July 1992.

[5] B. Razavi, Design of Analog CMOS Integrated Circuits, New York: McGraw-Hill, 2001, pp. 558-582.

[6] H. Twyman, (2009, February, 26), Digital clock recovery [online]. Available: http://www.twyman.org.uk/clock_recovery.

[7] A. Rofougaran, J. Rael, M. Rofougaran, A. Abidi, "A 900 MHz CMOS LC-Oscillator with Quadrature Outputs," Elect. Eng. Dept., University of California, Los Angeles, CA., 1996.

[8] M.T. Terrovitis, and R.G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," *IEEE Journal of Solid-State Circuits,* Vol. 35, No. 10, pp. 1461-1472, October 2000,

[9] M. Božanić and S. Sinha, "Software Aided Design of a CMOS Based Power Amplifier Deploying a Passive Inductor," *SAIEE Africa Research Journal,* vol. 99, no. 1, pp. 18-24, 2008..

[10] N.O. Sokal, and A.D. Sokal, "Class E - A new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE Journal of Solid-State Circuits,* Vol. SC-10, No 3, pp. 168-176, June 1975.

[11] W.F. Swart, "Design and implementation of a CMOS low-noise amplifier for monolithic modems," Final year report: Project EPR400, Dept. Electrical, Electronic and Comp. Eng., Univ. of Pretoria, Pretoria, South Africa, 2004.

[12] P.E. Allen, D.R. Holberg, *CMOS Analog circuit design*. Oxford: Oxford University Press, 2002, pp. 267-281.

[13] S. Franco, *Design with operational amplifiers and analog integrated circuits.* New York: McGraw-Hill, 2002, p. 117.

[14] N. Naudé, M. Božanić and S. Sinha, "Analogue CMOS direct sequence spread spectrum transceiver with carrier recovery employing complex spreading sequences," in *Proc. of IEEE Mediterranean Electrotechnical Conference,* 2006, pp. 1227-1230.

Impact of future CMOS scaling on power consumption: Electrical versus optical clock distribution networks

Petrus J. Venter[†], *MIEEE* and Monuko du Plessis[*], *SMIEEE*

***Abstract*—Scaling of CMOS circuits is typically associated with an increase in performance and a decrease in cost. The nanometer regime, however, is seeing a change between the conventional transistor dominated performance constraints to that of interconnect domination. As function density increases, the capacitive component follows exponentially, leading to heavy increases in power consumption and skew effects. This work investigates the future prospects of scaling based on ITRS 2008 predictions [1]. The utilisation of predictive SPICE modelling further enhances the work in relation to current predictive texts as well as the extension of the predictions down to 11nm. A design methodology is also introduced to assist in optimising the repeater sizing in electrical clock networks, while a comparison is drawn between the power consumption of electrical and optical networks.**

***Index Terms*—Integrated circuit, optical clock, power consumption, scaling**

I. Introduction

The modern success of CMOS was greatly influenced by the ease of scaling. Ironically, as the devices are entering the deep nanometer domain, practical scaling will probably not be limited by the devices, but rather the feeds that carry its signals: the interconnects.

With the aggressive scaling of channel gate lengths and future prospects of FinFET and other SOI devices, the density of active devices have increased more rapidly than some of the supporting technology. This forces modern MPU designers to make careful architectural decisions in order to maintain the associated performance. Unfortunately there are some fundamental limits regarding interconnects which presents some serious fundamental challenges to modern CMOS technologies as a whole, especially when venturing into the nanometer regime.

Dynamic power consumption of MPUs is mainly affected by the capacitances involved. As density increases, so does the amount of wiring. Increasing clock frequencies expected with scaling also worsen the problem, while degrading resistivity due to the skin [2] and grain boundary effects [3] require larger buffering components to maintain signal integrity. Some enhancements, such as a return to copper and utilisation of low-κ dielectric between conductors have slowed the pending red brick wall. Active mechanisms can also be effectively employed to reduce much of the local activity, but global signals, such as clock and synchronisation signals are becoming more of a power concern. This will also be the main aspect looked at in this work.

Some interesting alternatives are under speculation for replacing conventional copper interconnects. Amongst others, the possibility of employing optical means to distribute global signals is worth investigating. As such, the global clock network makes for a suitable comparative platform for electrical versus optical clock networks [4], [5].

The Carl and Emily Fuchs Institute for Microelectronics (CEFIM) in the Department of Electrical, Electronic and Computer Engineering at the University of Pretoria (UP), researches this work with the financial assistance of INSiAVA (Pty) Ltd.

[†]Petrus J. Venter is a lecturer and research engineer at CEFIM, Dept. EER/C, University of Pretoria. (Tel: +27 (0) 12 420 3648, email: jannes.venter@up.ac.za)

[*]Monuko du Plessis is a full-time professor and director of CEFIM, Dept. EER/C, University of Pretoria. (Tel: +27 (0) 12 420 3794, email: monuko@up.ac.za).

II. Clock network architecture

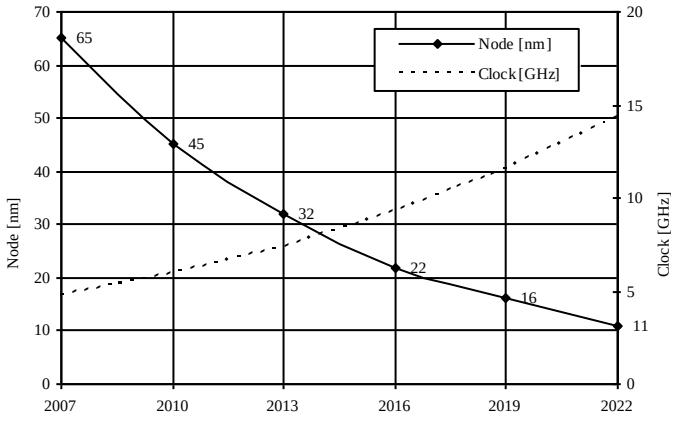


Fig. 1. Future scaling prediction and associated clock frequencies

A. Future CMOS for MPUs

Fig. 1 shows the expected scaling as a function of time [1], along with the local clock frequency for each relevant node. The predictions in fig. 1 play an integral part in the predictions on power consumption for future nodes, as clock frequency is linearly related to dynamic power consumption.

B. Balanced H-Tree

The balanced H-tree is a distribution architecture aimed at minimal skew from the clock source to the terminating end points. Each end point sees exactly the same path as any other, making skew depend only on process and environmental variations. For the purpose of this work, a square die is assumed, sized according to the predicted logic portion of a

typical modern MPU. Fig. 2 shows a die partitioning using an H-tree, with n representing the tree depth. Note that n for this work is not necessarily defined the same as compared to other

TABLE I

$\frac{3}{4}L\sum_{k=1}^{n}2^k$	Total H-tree length
$3\times L\times 2^{n-2}$	Length contributed at level n
$L\times 2^{-(n+1)}$	Segment length at level n
$3\times 2^{2n-1}$	Number of segments at level n
$L^2/2^n$	Area of local region at level n
$l_{die}/2^n$	Local area sidewall dimension
2^{2n}	Number terminating end points

works [4], [5]. For each increment of n, four new terminations, or end points will be instanced per (n - *1*) level end point. A summary of characteristics in a typical H-tree network is shown in table I.

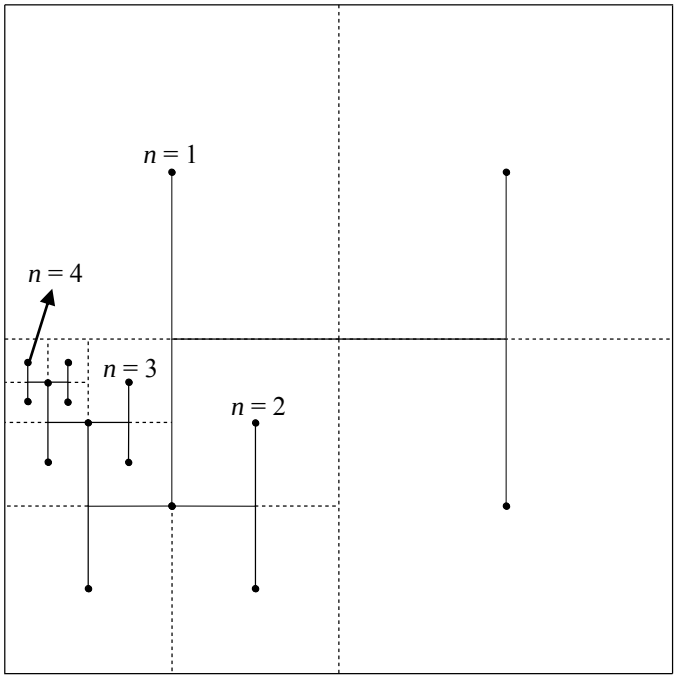


Fig. 2. Partitioning of an H-tree with tree depth indicated, bird's eye view of IC

The depth of the tree, n, is determined by the maximum allowable skew for a local region, defined in [4]. As technology scales and the skew reach a certain critical limit, the depth is incremented. The introduction of repeaters between tree splits along the segment becomes important to maintain signal flanks of an acceptable level. The metric chosen is that the 20% - 80% transition times should be maintained below 10% of the relevant technology node's clock period, T_{clk}. The expressions developed in [6] can be modified to determine an expression for the interconnect length at which the transition time does not meet the stated criteria (see section IV.A).

C. Implementation

The H-tree starts at a central point. When n = 1, the point first splits into two at the starting point, then splits into two again, resulting in four end points forming. An interconnect segment, driven by a buffer at the current split, runs to the next split which terminates into a low capacitance point consisting of either a split buffer with a fan out of two repeaters, or into an equivalent low capacitance amplifier driving a local region buffer. Depending on the segment length, repeaters are introduced to maintain the transition time specification. Each local region buffer then supplies the clock signal into a local mesh, which delivers timing information to the relevant circuitry. The local mesh depends on predicted register and transistor density [1].

Compared to an optical network, the local region model remains the same, while the optical front end is introduced right before the local region buffer.

D. Predictive SPICE modelling

Assumptions based on future technology nodes can only be made if the interconnect characteristics can be predicted, along with the associated active device behaviour. For this reason, predictive transistor modelling (PTM) is used [7], where predictions of future technology trends of front end processes are incorporated into SPICE models and as such can present a useful platform for future design exploration. As the model can predict behaviour difficult for hand analysis, a hybrid approach is used.

III. IMPORTANT TECHNOLOGY TRENDS

The ITRS publishes data on an annual basis wherein technology requirements are stipulated, based on past technology trends and future market interests. Combining this information with physical modelling of, for instance, the interconnects and CMOS active devices, key technology parameters affecting the performance of clock networks can be derived. Table II summarises the most important characteristics to be expected in the near future.

TABLE II

Parameters	Units	65nm	45nm	32nm	22nm	16nm	11nm
System characteristics							
V_{DD}	[V]	1.1	1.1	1	0.9	0.9	0.8
f_{clk}	[GHz]	4.7	5.9	7.3	9.2	11.5	14.3
Logic area	[cm^2]	2.86	2.87	2.87	2.88	2.88	2.88
D_{xtor}	[M/cm^2]	357	714	1427	2854	5708	11416
Device characteristics							
I_{on}	[µA/µm]	1006	1370	1948	1943	2344	2533
$L_{physical}$	[nm]	32	24	18	14	10.7	8.1
T_{oxp}	[nm]	1.2	0.95	0.7	0.7	0.6	0.55
T_{oxe}	[nm]	1.85	1.27	1.1	1.1	1	0.95
V_{th}	[mV]	225	175	103	105	109	109
Global interconnect characteristics							
r_{int}	[Ω/µm]	0.39	0.91	1.74	3.53	6.20	12.67
c_{int}	[F/µm]	2E-16	1.8E-16	1.7E-16	1.5E-16	1.5E-16	1.3E-16
Local interconnect characteristics							
r_{int}	[Ω/µm]	1.20	2.74	5.14	10.33	19.53	39.35
c_{int}	[F/µm]	1.8E-16	1.6E-16	1.5E-16	1.3E-16	1.3E-16	1.1E-16

Logic area is predicted to maintain a relatively constant portion of chip area, as opposed to the great reduction of logic portion in [4], since both logic and SRAM functions are

expected to increase at the same rate [1]. Intel's 45nm Penryn [8] reinforces this trend where the logic area portion remained roughly the same as the previous generation MPU, although the core size is smaller than predicted for high performance MPUs [1]. The current trend is roughly double the functions per area from one technology generation to the next.

Where channel depths become small enough, it is important to include the quantisation effects as an electrical equivalent oxide thickness. The physical thickness represents the effective thickness for SiO_2, while newer high-κ solutions might utilise thicker gates to minimise gate tunnelling [8]. I_{on} shows the strong inversion saturation current for NMOS devices at logic levels.

The interconnect resistance and capacitance terms were calculated assuming full shielding, with minimum dimensions used for maximum density. Resistivity calculations consider effective resistance increases as dimensions decrease, including the grain boundary component.

IV. TREE DESIGN

For a valid comparison, the design of the electrical tree should ideally relate to the optical tree. Using an H-tree is to an advantage as the end points of an electrical tree can easily be replaced by optical detectors for a good comparative study, explained in section II.C. The two circuit elements that play a critical role in the overall performance will then be the repeaters used for boosting the signal along the electrical interconnects, and the optical receiver front-end, used to convert the optical signals into electrical pulses.

A. Maximum allowable segment length

Given values for C_B, R_B, $c_{int}$and r_{int}, a maximum allowable segment length is found through solving equation (1)

$$\begin{aligned} a &= 0.5545 r_{int} c_{int} \\ b &= 1.386\ r_{int} C_B + c_{int} R_B \\ c &= 1.386 R_B C_B - T_{10\%} \\ l_{seg} &\le \frac{-b + \sqrt{b^2 - 4ac}}{2a} \end{aligned} \tag{1}$$

B. Repeater design

Section IV.A refers to the metric of $T_{20\text{-}80}$ which is used to determine the length of a driven segment, before another repeater needs to be inserted. In order to ease the design and optimisation of repeaters for a specific technology, a width dependent technology related parameter for output resistance and input capacitance was developed.

Suppose a quantity, *A*, exists where *A* is dependent on the width of the NMOS device in an inverter CMOS pair, and *A* is measured in Ω·µm. Thus, for a specific (NMOS) width, an output resistance, R_B can be attributed to that inverter by taking *A* divided by *W*, the width associated with the inverter NMOS device. The same can apply for input capacitance if, say a term *B* represents F/µm for that specific inverter design. The inverter input capacitance, C_B can then be found by multiplying *B* with *W*.

In reality, such two parameters do not ideally exist for all *W* while *A* and *B* will also be somewhat operating point dependent. Good approximations, where the results are valid over a wide range of *W* for logic transition purposes, provide a solid estimate for designing suitable buffers. The method for obtaining *A* and *B* require either physical devices, or at least a SPICE model for each relevant technology. A classical approach in larger technologies (> 1µm) would have been to approximate the driving device in triode region as a linear resistor using

$$R_B = \frac{(V_{DD} - V_{th})}{2 \cdot I_{on} \cdot W} \tag{2}$$

This obviously assumes a small V_{DS} value, which is invalid across the whole output swing range. For nanometer technologies, short channel effects basically voids equation 2, and solving analytically for the exact R_B value for all output states is too complex to use in an optimisation scheme for repeater sizing. Linear regression techniques are therefore used to fit the actual R_B data as a straight-line approximation, resulting in a single *A* value specific only to a technology and the relevant NMOS to PMOS sizing ratio. The same method is used to extract the input capacitance quantity *B*. A wide range of widths is used as control for the extraction procedure.

Fig. 3 shows an example of the procedure used to determine *A* for a specific technology, where the I_D versus V_{DS} curves are averaged for a number of different *W* values, and approximated as a linear curve. The inverse slope is then R_B which can be scaled with *W* to find a value for *A*. The method is the same for the input capacitance term, with the difference that a transient RC_{in} behaviour on the input node is used to estimate C_{in} and then to determine *B*.

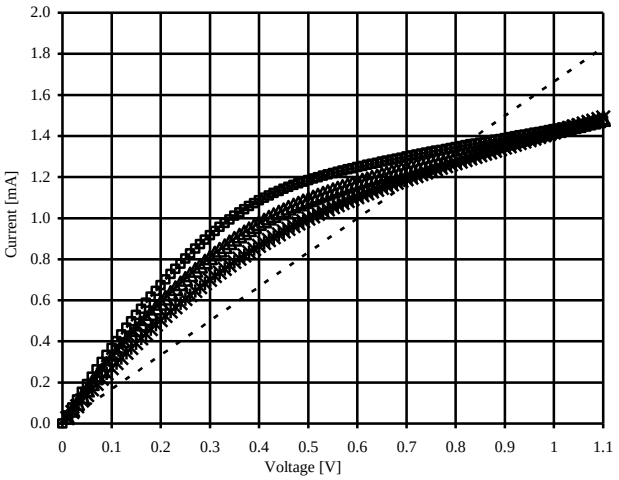


Fig. 3. An example of inverter output curves used to estimate a linear R_B which is scaled with *W* to obtain *A*.

C. Repeater optimisation

The repeaters are sized with the technology related parameters in section IV.A taken into account. A limit of two inverter pairs per repeater is set to maintain practicality. The

input inverter pair is sized smaller compared to the output pair by a ratio *R*. This has the advantage of reducing the input capacitance, while maintaining a stronger driving capability. A lower limit exists on the value of *R* to maintain the requirement of a 20% - 80% transition time between inverter pairs. Equation 3 quantifies the lower limit.

$$R = \frac{1.386AB}{T_{10\%}} \times 1.1 \tag{3}$$

$T_{10\%}$ is one tenth of a clock period and the factor of 1.1 is inserted as a safety margin to ensure that the inter-stage transition time is not limiting. Although equation 3 states a limit, an optimal value for *R* can be found if the total capacitance, that is the sum of the interconnect and repeater components, are normalised to a per unit length metric.

$$\frac{\partial C_{total/length}}{\partial R} = \frac{\partial}{\partial R}\left(c_{int} + \frac{B \times W \times (1+R)}{l_{seg}} \right) = 0 \tag{4}$$

Solving equation 4, with l_{seg} as the maximum allowable segment length and c_{int} as the interconnect capacitance per unit length, yields a solution for an optimal *R* value.

The width *W* is determined through maximising the segment length in equation 1, which presents the elegant solution of equation 5.

$$W = \sqrt{\frac{c_{int} A}{r_{int} RB}} \tag{5}$$

D. Summary of tree design

Table III shows the design parameters for a tree, based on the specific technology node at hand. Note the total number of repeaters increases exponentially.

TABLE III

Parameters		65nm	45nm	32nm	22nm	16nm	11nm
Tree design							
Max local l_{seg}		464.2	296.4	201.9	138.0	90.3	62.2
Tree depth *n*		6	6	7	7	8	9
# of end points		4096	4096	16384	16384	65536	262144
Max global l_{seg}		412.73	260.93	193.62	126.23	84.17	56.03
# of repeaters		4850	7070	19532	28484	80642	311798
# of split buffers		6142	6142	24574	24574	98302	393214
Repeater design parameters							
A	[Ω·µm]	695.8	575.65	395.4	366.15	315.65	273.25
B	[F/µm]	5.7E-15	4.8E-15	2.9E-15	3.0E-15	3.0E-15	2.9E-15
R		0.35	0.37	0.47	0.44	0.43	0.42

V. Optical End Point

A. Choice of photodiode

A designer has freedom to create a CMOS photodiode with any *pn*-junction forming structure available. Typical structures include n-well to p-substrate, n+ diffusion to p-substrate, p+ to n-well and p+ to n-well to p-substrate sandwiches. The choice of structure depends on the technology as well as the light source available. Should the designer have control of the light source, the wavelength can be chosen to maximise absorption in a depleted region, thereby decreasing the diffusion current tail on transient optical to electrical conversions. Unfortunately, photodiode responsivity decreases linearly as the source wavelength decreases, due to higher energy photons. This means bluer light will require more optical power to maintain a constant current. A trade-off exists between the intrinsic bandwidth of the structure and the wavelength of light.

B. Optical receiver design

Suppose a n+ to p-substrate photodiode is directly connected to a minimum sized inverter pair, the diode can charge or discharge the nodal capacitance as a function of incident light. A similar approach is used in [4], but with constant current biasing. The inverter then becomes a high-impedance optical receiver. Noise performance and power consumption for such a configuration have advantages above some of the low-impedance approaches involving TIAs, while the size and reduced complexity are beneficial for integration into typical CMOS VLSI designs.

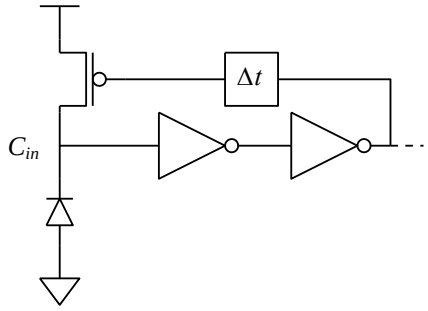


Fig. 4. An optical end point amplifier replacing their electrical equivalents.

The photocurrent will be sufficient if the total nodal capacitance can charge at $T_{10\%}$, where the PMOS shown in fig. 4 recharges the node back to a logic level after a specific time delay. Of course, this configuration assumes that signals are only sensitive to rising clock edges, but further signal conditioning is possible. The PMOS interchangeable with an NMOS device if the diode is realised through a p+ implant in an n-well. The complete amplifier output resistance needs to be exactly enough to drive the end point buffers, thus effectively replacing the electrical clock network up to that specific point. For comparative purposes for capacitance calculations, it is assumed that the minimum *W* has to be twice the minimum drawn gate length.

The photodiode has a parallel capacitance associated with the area used by the element. Leakage current also increases with area, and for practicality, the total photodiode should at least occupy less than 1% of the total local area to be comparable to the electrical solution. These three aspects limit the upper bound of photodiode area, where the lower limit is

TABLE IV

Parameters		65nm	45nm	32nm	22nm	16nm	11nm
Optical receiver							
C_{in}	[F]	7.4E-16	4.3E-16	1.9E-16	1.3E-16	9.5E-17	6.6E-17
$C_{electrical}$	[F]	4.8E-14	4.0E-14	1.4E-15	1.2E-14	7.6E-15	4.5E-15
Responsivity	[A/W]	0.35	0.33	0.30	0.27	0.24	0.21
W	[µm]	0.13	0.09	0.064	0.044	0.032	0.022
I_{ph}	[µA]	38.0	28.0	13.9	11.0	9.8	7.5
P_{opt}per EP	[µW]	217.02	169.39	92.63	81.73	81.93	71.70

dependent on the possible types of optical waveguides and the light source optical power per unit area available at the terminating ends. For this, it is assumed that the photodiode is designed with a capacitance equivalent to the inverter input capacitance.

VI. POWER CONSUMPTION

A. Common components to electrical and optical networks

Since logic will remain electrically clocked for the foreseeable future, there will always need to be a local region grid to supply clock signals to registers and other synchronous blocks. Therefore, the local grid power consumption will be the same for both approaches. The same argument holds for the buffer device driving the local mesh. The exact specifications of these will differ with implementation, but the increase in logic density will surely require an increase in the amount of local interconnects [1].

B. Global clock network components

For practicality, the short circuit current of the repeaters are neglected in dynamic power calculations, while only the capacitive component are examined. The analytical model for electrical power consumption then reduces to the simple expression in equation 6.

$$P_{dyn} = C \times f_{clk} \times V_{DD}^2 \quad (6)$$

For the electrical network, all interconnect capacitances are added together, along with the input and inter-stage capacitances of all the repeaters and split buffers used. The resulting power can then be calculated using equation 6.

The optical network approach allows for the receiver capacitances on all but the input node to be added and the power to be calculated as for the electrical case. The required input as shown in table IV does not include the losses associated with coupling and splitting of interconnects.

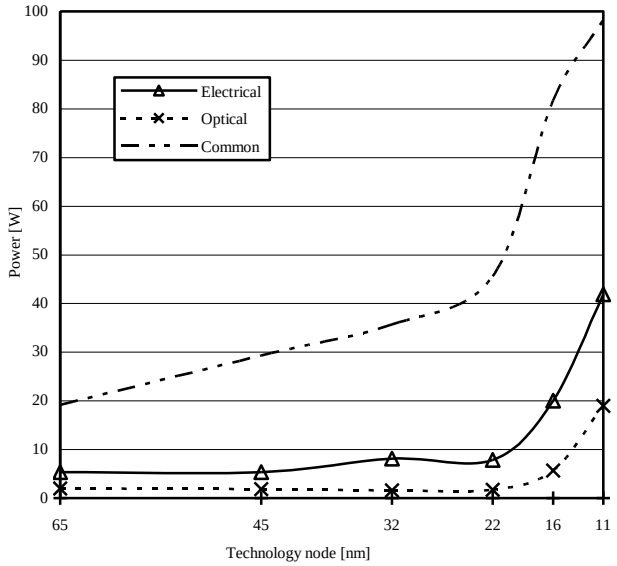


Fig. 5. A comparison of global power consumption

Fig. 5 shows the trend for future scaling. The difference between global optical and electrical power consumption becomes especially prominent in the last two nodes shown. The common, or local mesh component, is also shown in relation to the global components. Table III shows the input capacitance term *A* remaining constant for the last three nodes. This results in a linear decrease in the amount of optical power, while the amount of end points increases exponentially, which will eventually limit the possibility of an optical clock for nodes below 11nm.

C. Losses associated with the optical path

The implementation of a viable interconnect system as a replacement for an electrical H-tree has yet to be proven. Some predictions can be made regarding the three loss components[9]-[11]:

- Coupling losses due to insertion and interface effects
- Travel associated losses as a function of interconnect length
- Bending and splitting losses, associated with an increase in tree depth

TABLE V

Parameters		65nm	45nm	32nm	22nm	16nm	11nm
Optical receiver							
$P_{available}$	[dB]	5.21	5.41	8.14	7.89	11.56	13.60
$P_{coupling}$	[dB]	0.46	0.46	0.46	0.46	0.46	0.46
P_{length}	[dB]	0.9	0.9	0.9	0.9	0.9	0.9
P_{split}	[dB]	1.2	1.2	1.4	1.4	1.6	1.8
$P_{overhead}$	[dB]	2.65	2.85	5.38	5.13	8.60	10.44
$P_{overhead}$	[W]	1.84	1.93	3.45	3.26	7.25	11.06
ε_{source}	[%]	23	18	30	27	42	63

Table V summarises the expected losses, under the condition that bending losses are negligible.

VII. CONCLUSION

It seems that optical clock networks might present a solution for the foreseeable future, but the requirement of a power efficient light source becomes more important as tree depth increases. A suggestion by [4] is to use hybrid technique, at the expense of overall power consumption. At the local level, gating and asynchronous data transfer can alleviate much of the power consumption. For electrical networks, much can be done to optimise the repeaters for minimum power usage. For optical networks, the receiver design can be optimised by trying to minimise the input capacitance without compromising the gain of the system.

As an extension to this work, analysis of the skew differences in the global network might yield benefits to the optical approach not stipulated in this work.

REFERENCES

[1] The International Roadmap for Semiconductors (ITRS), 2008. Online: http://www.itrs.net

[2] H. Ymeri *et al*. A physics-based VLSI interconnect model including substrate and conductor skin effects. *Semiconductor Science and Technology*, 19(3):516-518, 2004.

[3] S. Yokogawa *et al*. Tradeoff characteristics between resistivity and reliability for scaled-down Cu-based interconnects. *IEEE Trans. on Electron Devices*, 55(1):350-357, 2008.

[4] B. Ackland, B. Razavi, and L. West. A comparison of electrical and optical networks in nanometer technologies. In *Proc. of the IEEE 2005 Custom Integrated Circuits Conference*, pages 779-782.

[5] E. Friedman. Clock distribution networks in synchronous digital integrated circuits. *Proceedings of the IEEE*, 89(5):665-692, May 2001.

[6] T. Sakurai. Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSIs. *IEEE Trans. on Electron Devices*, 40(1):118-124.

[7] W. Zhao and Y. Cao. A new generation of predictive technology model for sub-45nm design exploration. In *Proc. of the 7th International Symposium of Quality Electronic Design*, 2006, pages 585-590.

[8] G. Varghese *et al*. Penryn: 45-nm next generation Intel® core™ 2 processor. In *Proc. of the Asian Solid-State Circuits Conference*, 2007, pages 14-17.

[9] J. Gan *et al*. Two-dimensional 45° surface-normal microcoupler array for guided-wave opical clock distribution. *IEEE Photonic Technology Letters*, 11(11):1452-1454.

[10] R.T. Chen *et al*. Guided-wave Si CMOS process-compatible optical interconnects. In *CAS '99 Proceedings, International Semiconductor Conference,* 1999, vol. 2, pages 467-471.

[11] G. Tosik *et al*. Optical versus electrical clock systems in future VLSI technologies. In *Proc. of the IEEE International Systems-on-Chip Conference,* 2003, pages 261-262.

Design Methodology for SiGe-Based Class-E Power Amplifier

M. Božanić[*], *Student member, IEEE*, and S. Sinha[†], *Member, IEEE*
[*]Azoteq (Pty) Ltd, P O Box 16767, Lyttelton 0140, South Africa
[†]Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa
[*]mbozanic@ieee.org, [†]ssinha@ieee.org

***Abstract*—Power amplifier technology has matured rapidly over recent years and has become highly integrated onto several process technologies including SiGe BiCMOS, CMOS, and GaAs. The main passive component of a power amplifier is an inductor, which has always presented a bottleneck towards accomplishing a fully integrated power amplifier. Semiconductor manufacturing plants tend to offer inductors with precalculated inductances and Q-factors for discrete frequency values, complicating amplifier designs at any other frequencies. This paper proposes a new software routine for SiGe-based Class-E amplifier design. The software routine is used to calculate the necessary passive component values, and in the case of inductors, generate their layouts with Q-factors optimised at a desired frequency. The results for inductors obtained using this routine were compared against the results tabulated for existing inductors at discrete frequencies and it was found that the two sets of results corresponded well. The design and usage of high-Q integrated inductors are the main step in achieving high efficiency requirement for power amplifiers, since they allow for better output waveform filtering, or high-Q power amplifiers. For a typical power amplifier design case where several amplifiers are designed for application over different channels, the routine presented in this paper contributes by streamlining the design flow.**

I. Introduction

TRANSMITTER devices which employ various modulation schemes all demand a power amplifier (PA) for their successful operation. The PA technology has matured rapidly over recent years and has become highly integrated into several process technologies including SiGe BiCMOS, CMOS, and GaAs [1]. In many RF modulation schemes, such as the direct sequence spread spectrum (DSSS) technique, it is necessary to design several PAs in order to ensure operation over different channels of the same band. For example, for a system based on DSSS [2], transmission is possible over a number of channels in Industrial, Scientific and Medical (ISM) band; their centre frequencies ranging from 2.4000 GHz to 2.4835 GHz.

Basic PA is designed around one or more active devices, either metal-oxide semiconductor field-effect transistors (MOSFETs), bipolar junction transistors (BJTs) or heterojunction bipolar transistors (HBTs). Several PA output stages are commonly used in telecommunications, with Class E and Class F most widely used [3].

Additional to active devices, a number of passive components (inductors and capacitors) must be included. At layout level, where layout refers to drawing of mask layers on a silicon wafer, designing of inductors poses a special problem, because of a great number of possible inductor implementations [4], including spiral inductors, bond wires and active inductors. Even with a proper integrated inductor topology selected, the more affordable electronic design automation (EDA) software packages do not have built-in procedures for integrated inductor netlist extraction. This drawback requires manual modelling of inductors for the post-layout simulations. Although a few models of integrated inductors exist [5-7], they are often not used by designers who are either satisfied by simulations at the schematic level or ignore inductor modelling in order to proceed with the post-layout simulations. Downside of such an approach might not be apparent at low frequencies, but inductor modelling becomes particularly important in the design of PAs at RF, due to the fact that even small differences between actual and designed values of inductance can strongly affect the centre frequency, gain or efficiency of the amplifier. Often, these mismatches can only be seen after the fabrication of the chip is completed, thus introducing additional unnecessary chip fabrication iterations.

In this paper, a new design methodology for a BiCMOS Class-E PA is proposed. For a given set of specifications such as PA bandwidth, centre frequency and class of operation, the best possible PA is found and designed. This method is coined as a software routine. The same routine determines geometry of a spiral inductor that gives the highest possible quality factor, using process parameters for a particular process. Extracted layout and netlist can be exported by the routine and imported into layout design software for correct layout-level modelling. To verify this software routine, a Class-E PA has been designed and simulated in the SiGe S35 (0.35 µm BiCMOS) process from Austriamicrosystems (AMS).

II. Spiral Inductor

A. Background

Traditionally, capacitor and resistor implementations are easily accomplished in CMOS and are almost exclusively fabricated on-chip. In case of inductors, this is not always the case. Various factors, such as inductor size and low Q-factor of integrated passive inductors often result in alternative implementations: external inductors, active integrated inductors, microelectromechanical systems (MEMS) inductors

or bond wires. Although these implementations are widely used because of their advantages over passive integrated inductors, they are normally too complex to implement or make the PA devices too bulky and expensive. This leaves the passive spiral inductors as a reasonable choice for integration with PAs.

There are several spiral inductor geometries commonly used on chip. These include square and circular inductors as well as various polygons [8]. The square spiral has traditionally been more popular since some IC processes constrain all angles to 90° [4], but it generally has lower Q-factor than the circular spiral, which most closely resembles the common off-chip solenoidal inductors, but it is difficult to lay out. A polygon spiral is a compromise between the two. The square spiral, shown in Fig. 1, is fully specified by the number of turns (n), the turn width (w) and two of the following: inner, outer or average diameter (d_{in}, d_{out} or d_{avg} = (d_{in}+ d_{out})/2). The parameter, s, is the pitch between the turns of the spiral. Total length of the spiral is

$$l = 2(d_{in} + w) + 2n(2n-1)(s+w)\,. \tag{1}$$

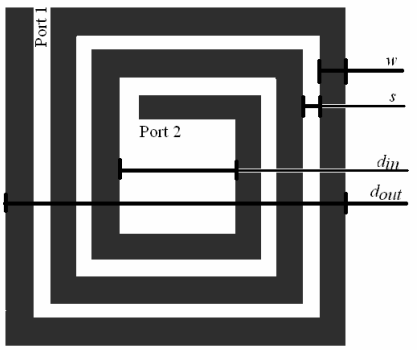


Fig. 1: Square inductor spiral with geometry parameters

B. Inductor Model

A lumped single-π nine-component inductor model of Fig. 2 is sufficient to accurately model spiral inductors for frequencies below resonance [8]. In this model, L_S is the inductance at the given frequency, R_S is the parasitic resistance and C_S is the parasitic capacitance of the spiral inductor structure. C_{ox} is the parasitic capacitance due to oxide layers directly under the metal inductor spiral. Finally, C_{Si} and R_{Si} represent the parasitic capacitance and resistance due to the silicon substrate, respectively. This topology does not model for the distributive capacitive effects, but it models correctly for parasitic effects of the metal spiral and the oxide below the spiral, as well as for the substrate effects.

The series inductance is calculated by data-fitted monomial expression that results in an error typically smaller than 3%. It has been developed by curve fitting over a family of 19000 inductors [8]. Inductance in nanohenries (nH) is calculated as

$$L_S = \beta d_{out}^{\alpha_1} w^{\alpha_2} d_{avg}^{\alpha_3} n^{\alpha_4} s^{\alpha_5}\,, \tag{2}$$

where coefficients β, α_1, α_2, α_3, α_4 and α_5 are in Table I for square geometry.

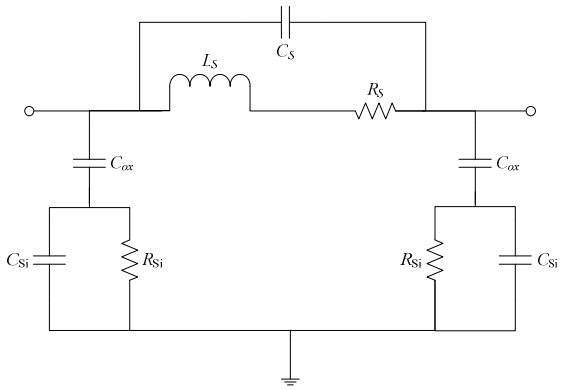


Fig. 2: Nine-component spiral inductor model [8]

TABLE I
COEFFICIENTS FOR THE SQUARE SPIRAL INDUCTOR INDUCTANCE CALCULATION [8]

β	α_1 (d_{out})	α_2 (w)	α_3 (d_{avg})	α_4 (n)	α_5 (s)
$1.62\cdot10^{-3}$	-1.21	-0.147	2.40	1.78	-0.030

Parasitic resistance is dependent on the frequency of operation. At DC, this value is mostly dependent on the sheet resistance of the material from which the wire is made. At high frequencies, the sheet resistance is overshadowed by the resistance that arises due to formation of eddy currents. Parasitic resistance depends on resistivity of metal layer in which the inductor is laid out (ρ), total length of all inductor segments, as well as on the width and effective thickness (t_{eff}) of the inductor [5]

$$R_S = \frac{\rho l}{w t_{eff}}\,. \tag{3}$$

Effective thickness, t_{eff}, is dependent on the actual thickness of the metal layer, t:

$$t_{eff} = \delta(1 - e^{-t/\delta})\,, \tag{4}$$

where δ is skin depth dependent on frequency f via relation

$$\delta = \sqrt{\frac{\rho}{\pi \mu_r \mu_0 f}}\,. \tag{5}$$

Parameter μ_r is the relative permeability of the metal layer and $\mu_0 = 4\pi\cdot10^{-7}$ H/m is the permeability of the free space. Parasitic capacitance is the sum of all overlap capacitances created between the spiral and underpass. If there is only one underpass and it has the same width as the spiral, then the capacitance is equal to [5]

$$C_S = n w^2 \frac{\varepsilon_{ox}}{t_{oxM1-M2}}\,, \tag{6}$$

where $t_{oxM1\text{-}M2}$ is oxide thickness between the spiral and the underpass and ε_{ox} is dielectric constant of the oxide layer between the two metals.

The oxide and substrate parasitics are approximately proportional to the area of the inductor spiral ($l\cdot w$), but are also highly dependent on conductivity of the substrate and

operating frequency. In order to calculate the oxide capacitance C_{ox} and substrate capacitance C_{Si}, effective thickness (t_{eff}) and effective dielectric constant (ε_{eff}) of either oxide or substrate must be determined. Effective thickness is calculated as [9]

$$t_{eff} = w\left[\frac{w}{t} + 2.42 - 0.44\frac{t}{w} + \left(1-\frac{t}{w}\right)^6\right]^{-1}, \text{ for } \frac{t}{w} \le 1 \quad (7)$$

or

$$t_{eff} = \frac{w}{2\pi}\ln\left(\frac{8t}{w} + \frac{4w}{t}\right), \text{ for } \frac{t}{w} \ge 1 \quad (8)$$

for both oxide and substrate. Effective dielectric constant is determined as

$$\varepsilon_{eff} = \frac{1+\varepsilon}{2} + \frac{\varepsilon - 1}{2}\left(1+\frac{10t}{w}\right)^{-1/2}. \quad (9)$$

Then,

$$C_{ox} = \frac{wl\varepsilon_0\varepsilon_{effox}}{t_{effox}} \quad (10)$$

and

$$C_{\mathrm{Si}} = \frac{wl\varepsilon_0\varepsilon_{eff\mathrm{Si}}}{t_{effox\mathrm{Si}}}. \quad (11)$$

Similarly, to calculate R_{Si}, effective thickness (t_{eff}) and effective conductivity (σ_{eff}) of substrate are needed. As for the capacitance, effective thickness is given by Eq. (8), and effective conductivity can be obtained from [9]

$$\sigma_{eff} = \sigma\left[\frac{1}{2} + \frac{1}{2}\left(1+\frac{10t}{w}\right)^{-1/2}\right], \quad (12)$$

where σ represents the substrate conductivity. Therefore,

$$R_{\mathrm{Si}} = \frac{t_{eff\mathrm{Si}}}{\sigma_{eff} wl}. \quad (13)$$

C. Quality Factor

The quality factor is the basic characterisation technique for inductors. For the single-π model, Q-factor can be calculated as [10]

$$Q = \frac{\omega L_S}{R_S}\cdot\frac{R_P}{R_P + \left[\left(\omega L_S / R_S\right)^2 + 1\right]R_S}\cdot\left[1-\left(C_P + C_S\right)\cdot\left(\omega^2 L_S + \frac{{R_S}^2}{L_S}\right)\right], \quad (14)$$

where

$$R_P = \frac{1}{\omega^2 {C_{ox}}^2 R_{\mathrm{Si}}} + \frac{R_{\mathrm{Si}}\left(C_{ox} + C_{\mathrm{Si}}\right)^2}{{C_{ox}}^2}, \quad (15)$$

$$C_P = C_{ox}\cdot\frac{1+\omega^2\left(C_{ox}+C_{\mathrm{Si}}\right)C_{\mathrm{Si}}{R_{\mathrm{Si}}}^2}{1+\omega^2\left(C_{ox}+C_{\mathrm{Si}}\right)^2{R_{\mathrm{Si}}}^2} \quad (16)$$

and $\omega = 2\pi f$. Low Q-factors of spiral inductors are attributed to the losses of the inductor spiral, substrate loss in the semiconducting silicon substrate and self resonance loss due to total capacitance $C_S + C_P$.

D. Inductance Search Algorithm

Although spiral inductors are a good choice for exclusively on-chip PAs, their use is not as straight forward. They occupy large areas on the chip, suffer from the low quality factors, and are difficult to design for the low tolerance. Capacitors and resistors, on the other hand, do not suffer from these problems.

As opposed to the common sense, one cannot just simply increase the number of turns, width of a single turn or another parameter to change the inductance. Complicated inductance relationship given in (2) can illustrate this interdependency. This complexity of spiral inductor models is one of the reasons why it is a common practice to choose an inductor from a library of inductors supplied with the process rather than to design an inductor for the needed inductance value. If a standard inductor cannot be used in the application of interest, then the iterative process is needed, such as one detailed in [11]. As part of this process, one guesses the geometry parameters that could result in the required inductance (and Q-factor) value, calculates inductance and other relevant parameters given the guessed parameters, thereafter repeating this process until satisfied with the performance of the inductor.

In this section, a new non-iterative routine is proposed that will result in an inductor of the specified value, with the highest possible Q-factor, occupying a limited area, and using predetermined technology layers. The intention behind this routine is to find a square inductor geometry resulting in the highest Q-factor for specified inductance given some design constraints. For accurate inductor modelling, the process parameters and frequency of operation of the inductor must be known. For the highest Q-factor, higher metal layers in the process should be used for the inductor spiral. Geometry needs to be constrained by the minimum input diameter, maximum output diameter, and minimum turn spacing and turn width. The search algorithm looks into a range of geometries and identifies a geometry that results in the required inductance within certain tolerance by using (2). More than one geometry will result in the correct inductance at a given frequency, but each of these geometries will have a different Q-factor (14). Geometry that gives the highest Q-factor is chosen by the algorithm as its output. Accuracy of the algorithm depends on the tolerance for the required inductance values and on the search grid resolution. Higher tolerance on the inductance value will result in less accurate inductance values, but there is greater probability that an inductor geometry that will result in that inductance will be found with lower grid resolution. A simplified flow diagram of the inductance search algorithm is shown in Fig. 3.

E. Verification

In order to verify the correctness of the inductance search algorithm, eleven inductors were designed using parameters for the S35 BiCMOS process from AMS. The results were compared against the measured results for the same spiral inductors provided by this foundry. Comparison of Q factors of these inductors is shown in Fig 4. As shown in this figure, the Q-factor values predicted by the inductance search algorithm and the Q-factor values measured by the foundry at the inductor centre frequency correspond with the average relative error almost equal to zero.

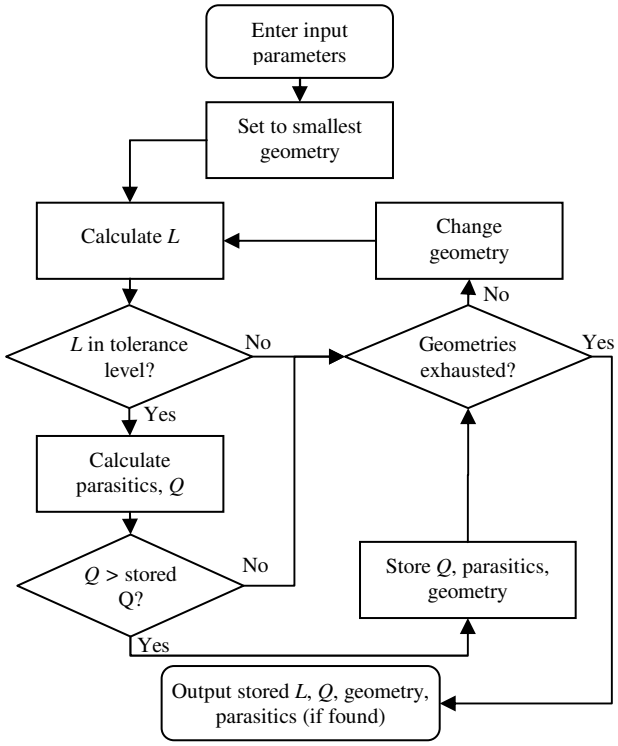


Fig. 3: Flow diagram of the inductance search algorithm

III. CLASS-E AMPLIFIER

A. Background

Class-E amplifiers were proposed by Sokal and Sokal in 1975 [12] and together with the Class-F amplifiers [13] have been used in communication ever since. They are classified as switching amplifiers and as such they can exhibit efficiencies close to 100%. A single ended Class-E PA with matching is shown in Fig. 5 [12]. Simple PA analysis can be performed if the following is assumed: inductance of the RF choke (L_1) is very high; output capacitance of the transistor is independent of the switching voltage; and transistor is an ideal switch with zero resistance and zero switching time, open for half of the signal period.

B. Equations

From [12], the value of the optimum load resistance to deliver the highest power to the load $P_{out\max}$ with peak voltage equal to supply voltage ($v_{peak} = V_{CC}$) and non-zero collector-emitter saturation voltage V_{CEsat} is

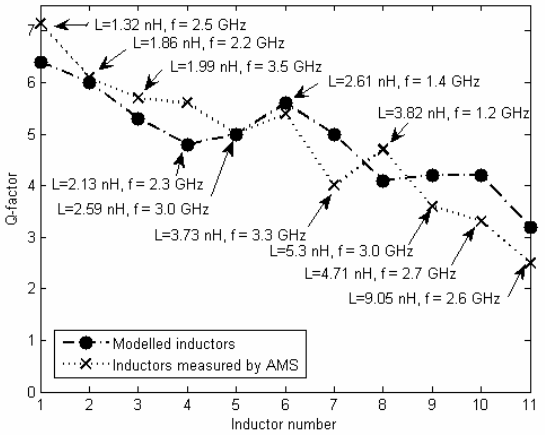


Fig. 4: Comparison between inductors designed using inductance search algorithm and inductors designed and measured by AMS pertaining to Q-factor

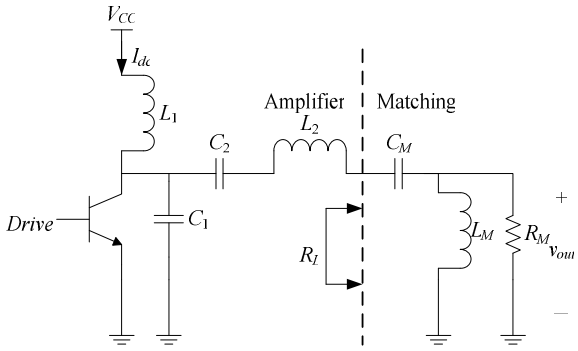


Fig. 5: Complete PA schematic

$$R_L = \frac{2}{\pi^2/4+1}\frac{(V_{CC}-V_{CEsat})^2}{P_{out\max}} = 0.577\frac{(V_{CC}-V_{CEsat})^2}{P_{out\max}}. \quad (17)$$

For the desired Q-factor of the output resonant tank Q_L, inductance L_2 can be calculated:

$$L_2 = \frac{Q_L R_L}{2\pi f}. \quad (18)$$

Shunt capacitance C_1 is given by

$$C_1 = \frac{1}{2\pi f R_L(\pi^2/4+1)(\pi/2)} = \frac{1}{5.447(2\pi f R_L)} \quad (19)$$

and resonant capacitance C_2 is given by

$$C_2 = \frac{1}{(2\pi f)^2 L_2}\cdot\left(1+\frac{1.42}{Q_L-2.08}\right) = C_1\left(\frac{5.447}{Q_L}\right)\cdot\left(1+\frac{1.42}{Q_L-2.08}\right). \quad (20)$$

Capacitance C_1 includes any output parasitic capacitances of the transistor.

C. Design Algorithm

The PA is designed with an HBT as an active device. Several input parameters are needed for the Class-E design algorithm to complete successfully. These include the centre frequency

of the channel (f_o), loaded quality factor (Q_L), output power (P_{out}) supply voltage (V_{CC}) and several optional parameters.

Centre frequency is determined by the specifications of the transmitter system of which the PA is a part. Quality factor is the Q-factor of the series resonator created by the inductance L_2 and capacitance C_2. This Q-factor differs from the Q-factor for the inductor given in (14), but the two are related, since the high-Q inductors will contribute to the design of the high-Q PAs. PA Q-factor can be chosen freely by the PA designer, but there is a tradeoff between high efficiency and power (low Q_L) on one side, and total harmonic distortion (THD) of the output signal on the other (high Q_L), which needs to be considered. Plausible Q factor is in the range of 5 to 10 [14]. If the efficiency is important, a lower Q_L can be chosen and harmonics can be removed by additional filters at the output of the amplifier [12]. A narrowband output matching network, described in IV, can serve for this purpose. Output power (P_{out}) is also determined by the transmitter system specifications. The higher the output power requirement, the smaller the load (R_L) will need to be, which will pose higher demands on the output impedance matching. High output power can also result in the need for a higher supply voltage. Supply voltage can be chosen by the designer, but attention must be paid that it does not result in a transistor entering collector-emitter breakdown.

The design algorithm utilizes (17) through (20) to calculate R_L, L_2, C_2 and C_1. DC current (I_{dc}), peak transistor voltage and current are also presented as the outputs of the program. The gain of the amplifier will depend on the type of the HBT available for specific design. In principle, HBTs with several long emitters should be chosen for a high gain, since they exhibit better performance at higher frequencies. It is the responsibility of the designer to support the input of the PA with correct impedance matching and biasing depending on the available gain.

IV. Complete PA System Integration

The inductor search algorithm and PA design algorithm described earlier in this paper are combined into a software routine to be used to perform complete PA system integration.

At the beginning of the execution of the routine, the user is asked to enter parameters for the Class-E PA design. Following this, the Class-E subroutine, described in III is executed. After this, the user has a choice to perform output impedance matching to the standard impedance of 50 Ω. At microwave frequencies, where wavelengths are correspondingly small, this matching can be accomplished with microstrip lines [15]. At low gigahertz frequencies, the microstrip lines are impractically long to be used on a chip, so the software routine described here employs impedance matching using discrete components. Three impedance networks are available: a wideband two-component network (L network, Fig. 5) and two narrowband three-component networks (T and Π networks). Finally, the user can choose to utilize the inductor search algorithm to design spiral inductors for all the inductors (including the matching ones), needed for the full PA design. For the AMS BiCMOS process, both 3-metal and thick-metal inductors are designed for each inductor; this allows users to choose geometry they consider better for the particular application. Furthermore, the routine can also export T-SPICE netlist of each inductor structure, complete with the inductance value and the parasitics, and a Caltech intermediate form (CIF) file, which contains the mask geometry information of any inductor [16]. The netlist can be used in SPICE simulations without the need for one to draw the schematic of the inductor with its parasitics in the schematic editor. The CIF file can be imported into the layout software to eliminate the need for one to draw any inductor layout structures.

V. Simulation

In this section, a set of simulations of a single Class-E PA system is used to illustrate the operation of the PA design routine. Simulations were performed in S35 process from AMS. A high speed HBT with 4 emitters, each 24 μm in length, was used for maximum gain. Supply voltage of 0.5 V was chosen in order to avoid collector-emitter breakdown. Design was done at 2.4 GHz to enable operation in ISM band. Conservative values for output power (25 mW or 14 dBm) and quality factor of the resonant tank (Q_L = 5) were chosen. A simple L network was used for the matching. Table II lists the calculated values for all required PA quantities, all shown in Fig. 5.

TABLE II
CALCULATED QUANTITIES FOR CLASS-E PA

Quantity	Value	Unit	Quantity	Value	Unit
R_L	5.77	Ω	C_2	3.4	pF
L_1	RFC	-	I_{dc}	50.0	mA
L_2	1.91	nH	L_M	1.2	nH
C_1	2.1	pF	C_M	4.15	pF
			R_M	50	Ω

TABLE III
GEOMETRY PARAMETERS AND PARASITICS OF INDUCTORS L_2 AND L_M

Parameter	Value (L_2)	Value (L_M)	Unit
L_S	1.92	1.20	nH
L_{self}	1.79	1.13	nH
Q	10.1	12.3	-
w	25	44	μm
s	1	1	μm
d_{in}	79	112	μm
d_{out}	233	290	μm
n	3	2	-
R_S	1.66	0.84	Ω
C_S	66.4	137	fF
C_{Si}	69.4	70.0	fF
C_{ox}	203	288	fF
R_{Si}	305	301	Ω

The output waveform v_{out} of the simulated system is shown in Fig. 6(a) for the design using ideal inductors and in Fig. 6(b) for the design using spiral inductors. Table III shows geometries, parasitics and Q-factor of thick-metal inductors L_2 and L_M, determined by the inductance search algorithm and used in simulation in Fig. 6(b). The layout of inductor L_2 created by importing of the CIF by layout design software is shown in Fig. 7 [11]. It has been assumed that the RFC inductor L_1 was external to the integrated system, and has a very high inductance value.

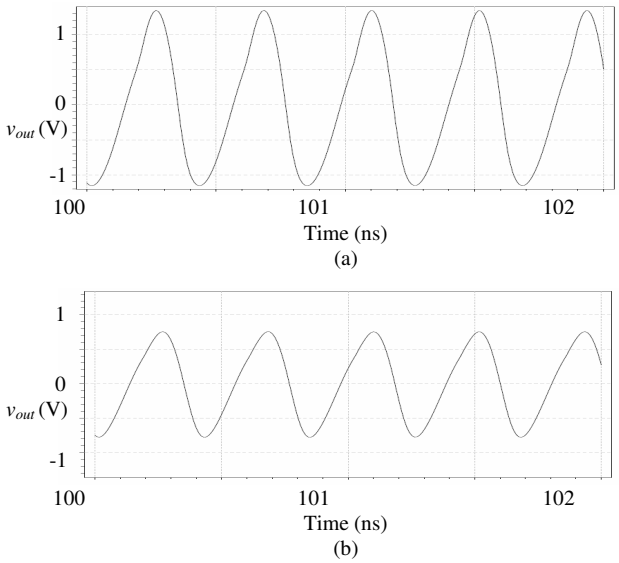


Fig. 6: Output voltage waveform for PA using (a) ideal inductors, and (b) spiral inductors

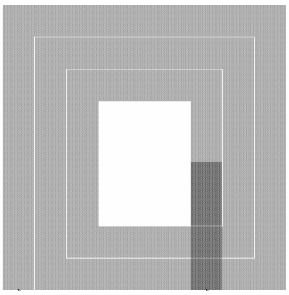


Fig. 7: Layout of 1.92 nH inductor used in Class-E PA design

The first simulation shows that the output voltage has the peak of about 1.3 V, which is equivalent to the power of about 17 mW, or 12.3 dB. This is a 1.7 dBm decrease from the value the design was intended for. This is attributed to the fact that the transistor has been assumed to be an ideal switch. Also, a slight amount of distortion is present, which is attributed to the low Q-factor of the resonant tank.

The waveform in the second simulation has the peak of about 0.7 V, which results in power of about 5 mW, or 6.9 dBm. This is further 5.4 dBm decrease in power, which is attributed to the presence of spiral inductors. Although the search algorithm finds inductors with the highest Q-factors possible within the space and technology constraints, the Q-factors are still fairly low and highly influential on the total performance of the integrated PA.

In both simulations, the DC current drawn from the supply is 43 mA, which is slightly less than the predicted value. This is equivalent to the supply power of 21.4 mW, resulting in collector efficiencies of 79 % and 23 % respectively for the two simulations. Evidently, a loss in efficiency of the Class-E stage is present in the design involving spiral inductors.

VI. Conclusion

In this paper, the software routine for the design of SiGe PAs was presented. Apart from determining the optimum values of passives needed for correct waveform filtering, the routine also includes spiral inductor design and output impedance matching. Inductors are designed with the aid of the inductor search algorithm, resulting in spiral inductors with Q-factors that are the highest within geometry and process constraints. This is of particular importance for the PA designs for which the use of alternative inductor implementations, such as active or MEMS inductors, is not feasible. Both filtering and matching inductors are designed by this routine. The streamlined use of the method described in this paper was demonstrated by designing a 2.4 GHz PA, which was subsequently simulated in the AMS S35 process.

References

[1] K. Nellis and P. Zampardi, "A Comparison of Linear Handset Power Amplifiers in Different Bipolar Technologies," *IEEE Journal of Solid-State Circuits,* vol. 39, no. 10, pp. 1746-1754, 2004.

[2] N. Naudé, M. Božanić and S. Sinha, "Analogue CMOS direct sequence spread spectrum transceiver with carrier recovery employing complex spreading sequences," in *Proc. IEEE Mediterranean Electrotechnical Conference,* Málaga, 16-19 May 2006, pp. 1227-1230.

[3] J. Rogers and C. Plett, *Radio Frequency Integrated Circuit Design*, 3rd ed. Boston: Artech House, 2003.

[4] A. M. Niknejad and R. G. Meyer, *Design, Simulation and Application of Inductors and Transformers for Si RF IC*, 2nd ed. Boston: Kluwer Academic Publishers, 2002.

[5] C. P. Yue and S. S. Wong, "Physical Modeling of Spiral Inductors on Silicon," *IEEE Transactions on Electron Devices,* vol. 47, no. 3, pp. 560-568, 2000.

[6] W. Gao and Z. Yu, "Scalable Compact Circuit Model and Synthesis for RF CMOS Spiral Inductors," *IEEE Transactions on Microwave Theory and Techniques,* vol. 54, no. 3, pp. 1055-1064, 2006.

[7] K. Y. Lee, S. Mohammadi, P. K. Bhattacharya and L. P. B. Katehi, "Compact Models Based on Transmission-Line Concept for Integrated Capacitors and Inductors," *IEEE Transactions on Microwave Theory and Techniques,* vol. 54, no. 12, pp. 4141-4148, 2006.

[8] S. S. Mohan, M. del Mar Hershenson, S. P. Boyd and T. H. Lee, "Simple Accurate Expressions for Planar Spiral Inductances," *IEEE Journal of Solid-State Circuits,* vol. 34, no. 10, pp. 1419-1424, 1999.

[9] X. Huo, P. C. H. Chan, K. J. Chen and H. C. Luong, "A Physical Model for On-Chip Spiral Inductors with Accurate Substrate Modeling," *IEEE Transactions on Electron Devices,* vol. 53, no. 12, pp. 2942-2949, 2006.

[10] C. Y. Lee, T. S. Chen, J. D. S. Deng and C. H. Kao, "A Simple Systematic Spiral Inductor Design with Perfected Q Improvement for CMOS RFIC Application," *Transactions on Microwave Theory and Techniques,* vol. 53, no. 2, pp. 523-528, 2005.

[11] M. Božanić and S. Sinha, "Software Aided Design of a CMOS Based Power Amplifier Deploying a Passive Inductor," *SAIEE Africa Research Journal,* vol. 99, no. 1, pp. 18-24, 2008.

[12] N. A. Sokal and A. D. Sokal, "Class E-A New Class of High-Efficiency Tuned Single-Ended Switching Power Amplifiers," *IEEE Journal of Solid-State Circuits,* vol. 10, no 3, pp. 168-176, 1975.

[13] P. Colantonio, F. Giannini, G. Leuzzi and E. Limiti, "On the Class-F Power Amplifier Design," *International Journal of RF and Microwave Computer-Aided Engineering,* vol. 9, no. 2, pp. 129-149, 1999.

[14] S. H. L. Tu and C. Toumazou, "Effect of the Loaded Quality Factor on Power Efficiency for CMOS Class-E RF Tuned Power Amplifiers," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications,* vol. 46, no. 5, pp. 628-634, 1999.

[15] R. Ludwig and P. Bretchko, *RF Circuit Design: Theory and Applications*, 1st ed. Upper Saddle River: Prentice Hall, 2000.

[16] R. Sproull, R. Lyon and S. l. Trimberger, "The caltech intermediate form for LSI layout description," Addison-Wesley Publishing Company, Inc, Reading, 1980.

Adaptive FIR Filter Pre-emphasis for High Speed Serial Links

Marius E. Goosen[*], *Student member*, IEEE. Saurabh Sinha[†], *Member*, IEEE.
[*†] Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa.
[*]marius.goosen@up.ac.za [†]ssinha@ieee.org

***Abstract*— Due to the advances of multimedia applications in recent years, the requirement for data bandwidth has increased significantly. Jitter requirements have become more stringent with higher speed serial communication links. Jitter degrades the performance of such a high speed serial link by limiting the maximum achievable data rate. A proposal for reducing jitter, with the main focus on reducing data dependent jitter, is presented by employing adaptive finite impulse response filter pre-emphasis. The channel bandwidth limitation, mainly caused by the skin effect of the conductor, and the parasitic components added by the integrated circuit package cause a non-flat frequency response and hence data dependent jitter. The pre-emphasis finite impulse response filter thus compensates for the frequency dependent chip package parasitic and the communication channel by altering the signal waveform before transmission. The signal waveform is altered in such a way that ideally a signal with a flat frequency response is received at the far end of the serial communication link. This paper presents the mathematical analysis and evaluation of adaptive finite impulse response filter pre-emphasis to reduce data dependent jitter and to ultimately achieve higher data rates. Typical copper cable and backplane channels will be modelled and analysed using passive components.**

I. INTRODUCTION

THE INCREASE in user end bandwidth in recent times, for applications such as high definition multimedia streaming, brought forth a need for higher speed serial data links. High speed serial links are preferred over parallel links for its immunity against clock skew [1]. Serial communication links further provide a lower cost solution to high bandwidth point to point communication [2].

Serial communication links, however, suffer from its own set of non-idealities, one of which is jitter. Jitter is defined as the deviation of a timing event of a signal from its intended or ideal occurrence in time. A deviation in time from the ideal sampling point could cause erroneous bits to be received.

Jitter can be subdivided into two main categories: random jitter and deterministic jitter. Random jitter is mainly caused by thermal vibrations, semiconductor doping and process variations which include thermal noise, flicker- or 1/*f* noise and shot noise [3], [4]. Random jitter is most profound in the clock signal used to serialise various data lines in the transmitter. Using a clock containing random jitter transfers this random jitter to the data signal.

Deterministic jitter is mainly caused by crosstalk, switching noise, insufficient power delivery, electromagnetic interference, duty cycle distortion, inter symbol interference and discontinuities in the transmission path [3]. These interferences on the transmission path can be compensated for by applying the inverse effect of the interferences on the transmitted signal to ultimately receive a high fidelity signal.

Deterministic jitter in high speed serial links is mainly caused by frequency dependent distortion. This frequency dependent distortion cause data dependent distortion or jitter since different data sequences will have a different spectral content. This will be elaborated upon in the next section.

One way to apply the inverse effect of the interferences is to apply pre-emphasis in the transmitter to alter the transmitting signal before transmission in order to compensate. Pre-emphasis has been widely implemented in high speed serial links as a means to overcome the channel bandwidth limitation causing data dependent jitter [1], [2]. Fig. 1 illustrates a typical serialiser-deserialiser (SerDes) serial communication link employing pre-emphasis. The signal is transmitted using low voltage differential signalling for improved noise immunity.

Adaptive pre-emphasis provides the ability to adapt for time varying conditions such as supply voltage drift and temperature fluctuations.

The next section will discuss the channel bandwidth limitation, followed by a detailed mathematical discussion of jitter. The adaptive FIR pre-emphasis analysis follows with a detailed analysis of different adaptation techniques.

II. FREQUENCY DEPENDENT DISTORTION

Frequency dependent distortion is mainly caused by two effects, namely, the package parasitic and the channel bandwidth limitation. Both of these effects are inherent to the respective technologies and the effect of each will be analysed next.

A. Package parasitic

The package parasitic influencing the transmitted signal is strongly dependent on the type of package implemented. Typical high frequency packages available include ball grid arrays (BGAs) and quad flat no-lead packages (QFNs). Wire bond QFN packages are preferred for high frequency applications below 40 GHz where cost is a consideration [5]. Wire bonding is preferred over flip-chip packaging for its higher yield and robustness at varying temperatures [5].

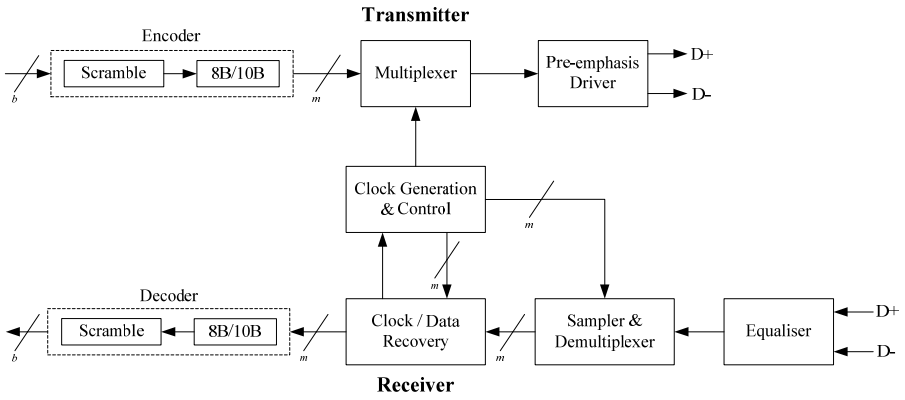


Fig. 1: Typical SerDes serial link employing pre-emphasis in the transmitter (adapted from [1], [6])

The bonding wire, I/O pad capacitance and package capacitance cause frequency dependent distortion. A typical model for determining the package parasitic is shown in Fig 2. Typical inductance for 1 mil thick gold bond wire is 1 nH/mm [5]. The capacitance to ground associated with the chip I/O pad was calculated using a typical 95 x 95 μm^2 I/O pad with the substrate capacitance properties of the IBM 8HP process. The chip to PCB capacitance was estimated to be the same as the chip substrate to ground capacitance for simplicity [5].

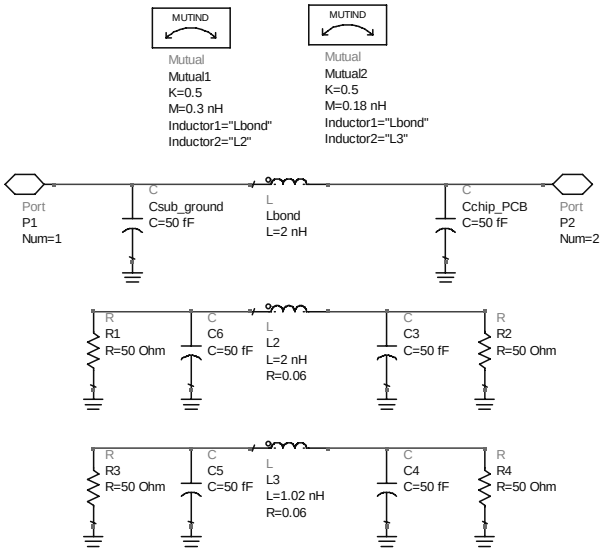


Fig. 2: Package parasitic modelling (adapted from [5])

The package parasitic modelling includes the modelling of mutual inductance caused by adjacent bonding wires. The adjacent bonding wires also cause crosstalk which introduces another form of distortion affecting the integrity of the transmitted signal.

The insertion loss of the package parasitic utilising an on-chip 50 Ω termination and a 50 Ω transmission line is presented in Fig. 3. It is important to remember that the package parasitic response shown is only for one end of the serial link and that the effect will be worse when the receiver chip parasitic is taken into account in establishing the complete channel response.

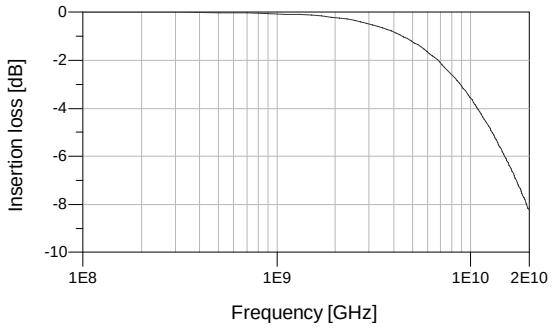


Fig. 3: Insertion loss of the modelled QFN package. Simulation using Agilent ADS.

B. Backplane channel bandwidth limitation

The backplane channel bandwidth limitation is mainly caused by the skin effect but dielectric losses also play a part at frequencies approaching 10 GHz and higher. Total transmission line attenuation can be written as [7]

$$\alpha_{tot} = \alpha_c + \alpha_d \tag{1}$$

where α_c is the conductor loss caused by the skin effect and α_d is the dielectric loss. The skin effect in essence creates a frequency dependent resistance attenuating the signal. The frequency dependent resistor can be expressed as [7]

$$R_{skin} = \frac{\sqrt{\pi f \mu \sigma}}{\sigma} \tag{2}$$

where σ is the conductivity and μ is the magnetic permeability of the conducting material. For copper, the conductivity is 56 x10^6 S/m and the magnetic permeability is 4π x10^7 H/m. Conductors can also be coated with gold to improve the conductivity and reduce the frequency dependent resistor.

The dielectric loss is due to the delay of polarisation in the dielectric material when subjected to changing electric fields [7]. The dielectric loss can be expressed as [7]

$$\alpha_d = \frac{\pi f}{c}\frac{\left(\varepsilon_{eff}-1\right)}{\sqrt{\varepsilon_{eff}}}\frac{\varepsilon_r}{\left(\varepsilon_r-1\right)}\tan\delta \qquad (3)$$

where ε_r is the dielectric permittivity (4.34 @ 1 GHz for FR4), ε_{eff} is the effective dielectric permittivity of the implemented substrate, δ is the effective skin depth and c is the speed of light (300 x10^6 m/s).

Thus it is seen that the conductor loss, which is proportional to the skin effect resistance, is proportional to the square root of frequency and the dielectric loss is directly proportional to frequency.

A typical frequency response of a copper backplane channel can be modelled as a low pass filter with a -3 dB cut-off frequency of approximately 400 MHz [8]. The complete channel model can now be evaluated by combining the package parasitic response and the copper backplane channel response.

C. Complete channel model

The complete channel model combines the effect of the package parasitic and the copper backplane channel frequency response. At low frequencies the copper backplane channel frequency response dominates the complete channel response but at higher frequencies the frequency dependent package parasitic becomes more significant. The complete channel model is shown in Fig. 4.

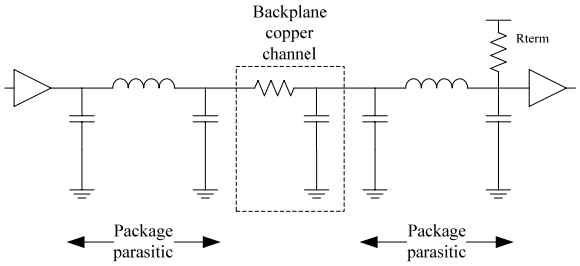


Fig. 4: Complete channel model derived using lumped components

Fig. 5 illustrates the total channel frequency response compared to the package parasitic and backplane channel frequency responses. The complete modelled channel response corresponds well to measured channel responses [8], [9].

III. Jitter Definition

Total system jitter can be subdivided into two main categories, random jitter and deterministic jitter. These two categories can be further subdivided by the cause of the two types of jitter. Fig. 6 illustrates the different types of jitter.

Data dependent jitter can be subdivided into two dominant types, inter symbol interference (ISI) and duty cycle distortion (DCD). DCD is caused by a current limited output stage which is unable to properly drive the load on both rising and falling edges.

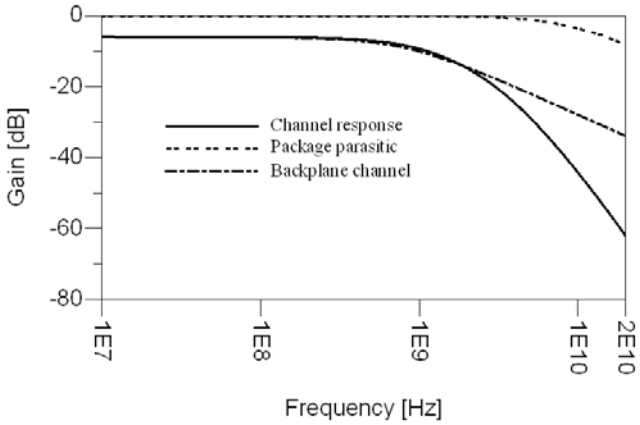


Fig. 5: Channel frequency response. Also shown are the package parasitic frequency response and the backplane channel frequency response.

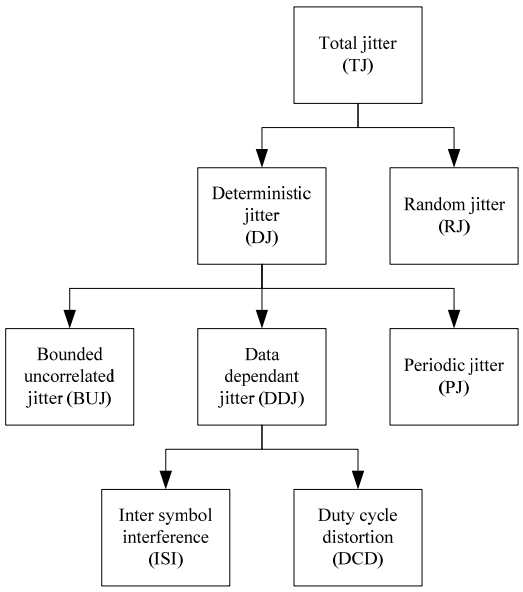


Fig. 6: Different types of jitter contributing to total system jitter (adapted from [4])

ISI on the other hand is caused by a bandwidth limitation in the channel used for data transmission. The focus of this work is on reducing data dependent jitter caused by a bandwidth limitation while keeping the random jitter at a minimum. It will be assumed that there is sufficient output power to drive the load.

A. Statistical analysis

The total system jitter probability density function (PDF) can be evaluated by convoluting the random jitter PDF with the deterministic jitter PDF [14]. Thus both the random jitter and the deterministic jitter should be characterised by their respective PDFs. Random jitter is characterised as having a Gaussian distribution described by the well-known equation

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad (4)$$

Deterministic jitter on the other hand is characterised as having a double delta-dirac distribution evaluated using the direct de-convolution method [10]. The double delta-dirac distribution can be expressed as

$$f_{DJ}(t) = \frac{1}{2}\left[\delta\left(t - \frac{D}{2}\right) + \delta\left(t + \frac{D}{2}\right)\right] \quad (5)$$

where D is the distance between the two delta-dirac functions. The convolution of the random jitter and deterministic jitter produces the result shown in Fig. 7. The distribution is situated around the pulse edges of the transmitted signal distorting the received signal at the receiver.

Whenever the distributions situated at the two pulse edges cross, an error region will be produced defining a situation where an error bit could be received.

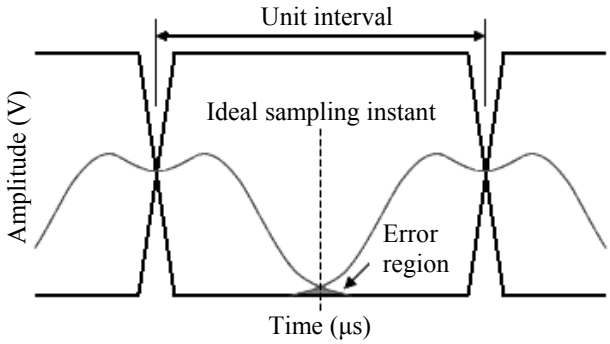


Fig. 7: Total jitter PDF situated around the pulse edges of the transmitted signal

B. Overcoming data dependent jitter

Overcoming data dependent jitter requires that the bandwidth limitation imposed by the channel be alleviated by extending the -3 dB point. Two ways of extending the effective -3 dB point is by either implementing equalisation in the transmitter or in the receiver.

The criterion for error free transmission, namely a flat frequency response and a linear phase response are mathematically expressed as [11]

$$\begin{aligned} \left|H_P(\omega)H_C(\omega)\right| &= \mathrm{C} \\ \theta_P(\omega) + \theta_C(\omega) &= -\omega t_d \end{aligned} \quad (4)$$

where $H_P(\omega)$ is the FIR pre-emphasis filter transfer function, $H_C(\omega)$ is the channel transfer function and C is a constant depicting a flat magnitude frequency response. $\theta_P(\omega)$ is the FIR pre-emphasis filter phase response, which can be designed to have a linear phase response [12], and $\theta_C(\omega)$ is the channel phase response.

IV. FIR Filter Pre-Emphasis

The data dependent jitter completely closes the eye diagram at the receiver making clock and data recovery almost impossible. Fig. 8 illustrates such a closed eye diagram at the receiver. The data rate of the signal was specified as 5 Gb/s and was transmitted across the modelled channel as presented in Fig. 5. From Fig. 8, it is seen that any means to try and recover the clock and data at the receiver will not be straightforward.

The FIR filter response used to correct the data dependent jitter is ideally the inverse of the complete channel response in order to accomplish a flat frequency response at the receiver. A flat frequency response at the receiver will ensure a perfect open eye diagram for sampling. However a perfect inverse channel response is not possible with a finite amount of FIR filter taps [12], thus the actual frequency response achieved will flatten of at higher frequencies to only extend the -3 dB point of the overall frequency response [13].

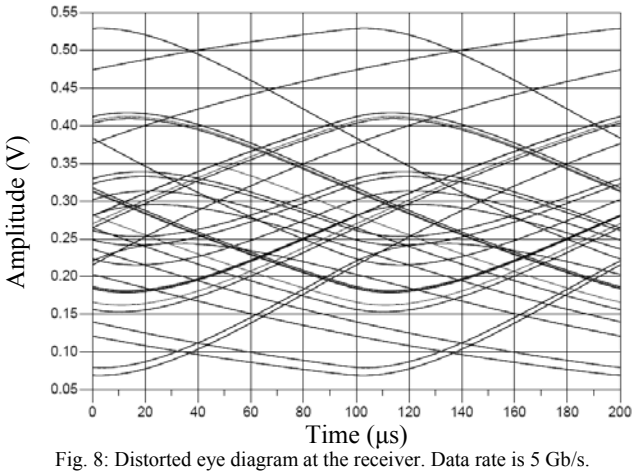


Fig. 8: Distorted eye diagram at the receiver. Data rate is 5 Gb/s.

The extension of the -3 dB point will ensure that the bandwidth limitations have a less profound effect on the transmitted signal. A FIR filter is characterised by its impulse response and frequency response and is mathematically expressed as [12]

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (5)$$

$$H(z) = \sum_{k=0}^{N-1} h(k)z^{-k} \quad (6)$$

where $h(k)$, $k=0,1,\ldots,N-1$ are the impulse response weight coefficients. From the mathematical expression presented for FIR filters it is clear that the implemented filter only requires knowledge of the current and previous transmitted bits. This makes FIR filters very attractive to implement using digital circuits. The filter response is further dependent on the filter coefficients and the frequency response can be completely altered by changing only the filter tap coefficients.

V. Adaptive Pre-Emphasis

Two methods of adaptive filtering can be applied to adaptive pre-emphasis namely, a least mean squares (LMS) convergence engine, or by using training sequences to train the filter tap coefficients.

A. LMS convergence engine

A popular way of implementing adaptive FIR filters is by means of a LMS convergence engine. The filter tap coefficients are updated after every sample in order to minimise the mean square error (MSE). The mean square error is calculated between the desired signal and the distorted input signal, hence knowledge of the desired and distorted signal is a prerequisite. Adaptive equalisers have been widely implemented in the receiver of high speed serial links by emulating the desired signal by means of clock and data recovery.

Implementing a LMS convergence engine in the transmitter is a more complicated task since the distorted signal is not present. One way of implementing a LMS convergence engine is by connecting two adaptive receiver equalisation transceivers in a master slave connection and using the LMS convergence engine of the second receiver to update the filter tap coefficient of the first pre-emphasis filter [6], [15].

The sign-sign LMS convergence engine is popular for its ease of implementation using digital circuits. The tap update algorithm for a sign-sign LMS engine can be expressed as

$$h_k(j+1) = h_k(j) + \mathrm{sgn}\left(\sum_{i=0}^{L-1} \mathrm{sgn}\left(e_{jL-1}\right)\mathrm{sgn}\left(d_{jL-i-k}\right)\right) \quad (7)$$

where $h_k(j)$ is the k_{th} tap coefficient for block j using a block length of L, e is the error signal and d is the data signal. The sign-sign LMS engine is easily implemented using a digital round robin adaptation engine [6], [15]. The LMS convergence engine continues to update the filter tap coefficients and continues to converge around the optimal tap coefficients.

B. Pilot signalling and peak detection

Filter tap coefficients can also be determined by applying simple combinations of pilot signalling and peak detection [16]. By expressing the channel impulse response as

$$y_{chn}(n) = \sum_{k=0}^{N-1} c_k \delta(n-k) \quad (8)$$

where c_k is the equivalent taps of the channel impulse response, the received signal amplitude at sample times at the far end can now be approximated as the convolution of the channel impulse response (8) and the FIR filter impulse response (5). The determined received signal usually exhibits a long tail causing ISI. The filter tap coefficients can now be determined by minimising the tail of the received signal [16].

Each tap coefficient is updated by applying different training sequences to the input. Such a tap coefficient training method requires however that data transfer be temporarily stopped to allow for optimal tap weight determination.

VI. Conclusion

High speed serial links are the preferred way of transferring large amounts of data in a point to point environment, but certain limitations restrict the overall data rate. Pre-emphasis as presented is one means of overcoming data dependent jitter.

A typical channel model has been derived using lumped components. The modelled backplane copper channel and package parasitic component frequency responses compare well with measured responses [8], [9].

Two types of adaptive pre-emphasis were presented and traded off against each other. Adaptive pre-emphasis utilising training sequences requires only a low frequency return path for filter coefficient updates whereas LMS adaptive pre-emphasis requires a high frequency return path. LMS adaptive pre-emphasis on the other hand will converge to optimal tap coefficients quicker and will also keep converging to optimal tap coefficients without interrupting data transfer. Piloting and peak detection on the other hand is easier to implement and does not require multiple transceivers or accurate sampling to calculate optimal tap coefficients. Filter updates can also happen by means of a lower frequency return path since the update frequency is solely dependent on the pilot signal which is user definable.

Acknowledgement

The authors would like to thank ARMSCOR, the Armaments Corporation of South Africa Ltd, (Act 51 of 2003) for sponsoring this study.

References

[1] M. Li, T. Kwasniewski, S. Wang and Y. Tao, "A 10 Gb/s transmitter with multi-tap FIR pre-emphasis in 0.18 μm CMOS technology", *IEEE Proc. of the 2005 Asia South Pacific Design Automation Conf.*, Shanghai, pp 679-682, 18-21 Jan. 2005.

[2] M. Bichan and A.C. Carusone,"A 6.5 Gb/s backplane transmitter with 6-tap FIR Equaliser and variable tap spacing", *Proc. of IEEE Custom Integrated Circuits Conf.*, pp. 611-614, San Jose, 13-16 Sept. 2008.

[3] M. Cases, D.N. de Araujo and E. Matoglu, "Electrical design and specification challenges for high speed serial links", *Proc. of IEEE Electronics Packaging Tech. Conf.*, Vol. 1, Singapore, pp. 29-33, 7-9 Dec. 2005.

[4] A. Kuo, T. Farahmand, N. Ou, S. Tabatabaei and A. Ivanov, "Jitter models and measurement methods for high-speed serial interconnects", *Proc. of Int. Test Conf. (ITC)*, Charlotte, pp. 1295-1302, 26-28 Oct. 2004.

[5] Unchwaniwala K.B. and Caggiano M.F., "Electrical analysis of IC packaging with emphasis on different ball grid array packages", *Proc. of IEEE Electronic Comp. and Tech. Conf.*, Orlando, pp 1496-1501. 29 May – 1 Jun. 2001.

[6] D. Tonietto, J. Hogeboon, E. Bensoudane, S. Sadeghi, H. Khor, P. Krotnev, "A 7.5Gb/s transmitter with self-adaptive FIR", *Proc. of IEEE Symp. on VLSI Circuits Digest of Technical Papers*, Honolulu, pp. 198-199, 18-22 Jun. 2008.

[7] S.M. Wentworth, *Fundamentals of Electromagnetics with Engineering Applications*, 1st edition, John Wiley & Sons, 2004.

[8] L. Lin, P. Noel and T. Kwasniewski, "Implementing a digitally synthesized adaptive pre-emphasis algorithm fur use in a high-speed backplane interconnection", *Proc. of Canadian Conf. on Electrical and Computer Engineering*, Ontario, Vol. 3, pp. 1221-1224, 2-5 May 2004.

[9] K. Yoo and G. Han, "An adaptation method for FIR pre-emphasis filter on backplane channel", *Proc. of IEEE Int. Symp. on Circuits and Systems*, Island of Kos, pp. 5151-5154, 21-24 May 2006.

[10] P.R. Trischitta and E.L. Varma, *Jitter in Digital Transmission Systems*, Artech House, Inc., 1958.

[11] B.P Lathi, *Modern Digital and Analog Communication Systems*, 3rd edition, Oxford University Press , New York, pp. 567-572, 1998.

[12] E.C. Ifeachor and B.W. Jervis, "Finite impulse response filter design" in *Digital Signal Processing: A Practical Approach*, 2nd edition, Prentice Hall, Essex, England, 2002.

[13] K. Yoo, G. Han and S. Park, *A 10 Gbps Analog Adaptive Equaliser and Pulse Shaping Circuit for Backplane Interface*, [Online]. Available: http://cad.yonsei.ac.kr/conference/international/IC_29.pdf.

[14] J. Sun, M. Li and J. Wilstrup, "A demonstration of deterministic jitter (DJ) deconvolution", *Proc. of IEEE Instrumentation and Measurement Conf.*, Anchorage, Vol. 1, pp. 293-298, 21-23 May 2002.

[15] J.T. Stonick, G.Y. Wei, J.L. Sonntag and D.K. Weinlader, "An adaptive PAM-4 5 Gb/s backplane transceiver in 25 μm CMOS", *IEEE Journal of Solid-State Circuits*, pp. 436-443, Vol. 38, No. 3, Mar. 2003.

[16] K. Yoo and G. Han, "An adaptation method for FIR pre-emphasis filter on backplane channel", *Proc. of IEEE Int. Symp. on Circuits and Systems*, Island of Kos, pp. 5151-5154, 21-24 May 2006.

Phase Noise Reduction of a 0.35 µm BiCMOS SiGe 5 GHz Voltage Controlled Oscillator

Johannes W. Lambrechts*, *Student member, IEEE*, and Saurabh Sinha†, *Member, IEEE.*
*†Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa
*wynand.lambrechts@ieee.org, †ssinha@ieee.org

***Abstract*— Voltage controlled oscillators are essential components in radio frequency systems and are used in narrow band receivers for communication, tracking, and radar applications. Noise in these oscillators limits the channel capacity of communication systems, degrades the resolution of radar, and restricts the resolution of spectrum analyzers. Focus on the research issues of a low phase noise VCO for use in RF narrow band receivers, using SiGe BiCMOS technology is being proposed in this paper. Modern communication's increasing bandwidth requirements as a result of low-temperature growth of Silicon epitaxy has rekindled research in the BiCMOS field. Combining bipolar with CMOS transistors enhance system performance when operating in the GHz range (especially from 5 GHz), as heterojunction bipolar transistors with SiGe doped bases can easily have switching delays of less than 10 ps. More specifically, this paper proposes a method to achieve low phase noise employing a tail-current noise suppression technique, adding no additional power to the system. This technique prevents the low-frequency tail-current noise from being converted into phase noise. Simulated phase noise improvements of 3.2 dBc/Hz at a 1 MHz offset from a 5 GHz carrier has been achieved using this technique, implemented with Austriamicrosystem's 0.35 µm S35D4M5 SiGe BiCMOS process. This paper contributes by providing mathematical models relating to circuit parameters providing insight towards SiGe based oscillator performance in terms of phase noise.**

I. INTRODUCTION

PHASE NOISE PERFORMANCE in modern radio frequency (RF) communication applications has become more stringent as the noise in the sidebands at high frequencies result in spurious responses of the receiver to nearby interfering channels. This contributes to the degradation of the modulation accuracy of the transmitter [1]. Voltage controlled oscillators (VCOs) are an essential functional block in these communication systems, used for clock generation and data recovery. The phase noise in the VCO should be limited to ensure effective operation of these systems.

The term phase noise is widely used for describing short term random frequency fluctuations of a signal, whereas frequency stability is a measure of the degree to which an oscillator maintains the same value of frequency over time [2].

Several techniques exist to improve phase noise performance in VCOs. These techniques range from VCO topology considerations, inductor dimensioning and placement for improved quality factor (Q factor), semiconductor process parameter considerations, and reducing low-frequency tail-current noise sources. The latter option is employed in this paper to study the overall effect on VCO phase noise performance, as it has been recognized that the tail transistor may have a large impact on the generation of phase noise, often being the largest contributor [1]. A challenge that arises when designing an oscillator for low phase noise, is maintaining a reasonable tuning range to allow operation in the frequency bands adjacent to the centre frequency [3]. A large tuning range and a low available control voltage (limited by the supply voltage and device breakdown voltage) results in a large oscillator gain, which increases both intrinsic device noise and coupled spurious tones up-conversion around the centre frequency [1], [3], and [4].

Tail current noise suppression in RF BiCMOS VCOs prevents the low-frequency tail current noise from being converted into phase noise during normal operation of the oscillator [5]. The tail current is made large when the oscillator output voltage reaches its maximum or minimum value and when the sensitivity of the output phase to injected noise is the smallest [5]. The tail current is made small during the zero crossings of the output voltage when the noise sensitivity is large. No additional power is added to the system, ensuring low power operation at low noise levels. Tail current shaping techniques reduce phase noise with three separate, but simultaneous mechanisms [6]. The increased oscillation amplitude, narrower drain current pulses, and finally the shunt capacitor that acts as a noise filter for the tail current, all contribute to lowering the phase noise.

Implementation of a VCO in a BiCMOS process entails the utilization of Silicon Germanium (SiGe) heterojunction bipolar transistors (HBTs). These transistors are capable of operating at very high cutoff frequencies (up to 250 GHz [7]) with remarkable noise performance. This is achieved through doping of the base-emitter region with SiGe which increases the electron mobility while maintaining gain [7].

II. PHASE NOISE IN OSCILLATORS

The unit for phase noise is dBc/Hz. For instance, -150 dBc/Hz means that for 1 Hz bandwidth, at a specified offset frequency from the carrier frequency, the single sideband noise power is 10^{-15} of the total signal power.

A. Background (Leeson's Model)

One of the most well-known models for predicting phase noise in feedback oscillators is Leeson's model. This model is shown in Fig. 1 [10].

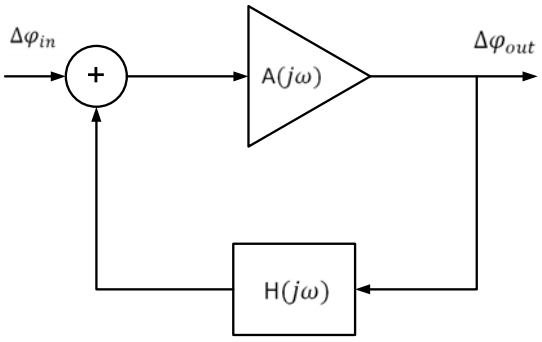


Fig. 1: Representation of a feedback oscillator using Leeson's model

Leeson [10] has derived the following expression by using a single resonator feedback network. This expression is called the Leeson formula and is given by

$$L\{\omega_m\} = 10\log\left[\left[\left(\frac{\omega_o}{2Q_L\omega_m}\right)^2 + 1\right] \times \frac{FkT}{P} \times \left(\frac{\omega_c}{\omega_m} + 1\right)\right] \quad (1)$$

where ω_m is the offset frequency (measured at 1 MHz) from the 5 GHz carrier (ω_c), $L\{\omega_m\}$ is the noise level at ω_m in dBc/Hz, Q_L is the loaded Q of the tank circuit, k is Boltzmann's constant (8.6173 x 10^{-15} eV/K), F is the noise factor, T is absolute temperature (0 K), P is the carrier power in dBm, and ω_c is the corner frequency for flicker noise. The loaded Q of a resonator circuit describes the passband characteristics of the circuit under loaded or operating conditions. It is dependent on three factors, namely the source resistance (R_S), the load resistance (R_L), and the Q of the inductor [28]. This can be put in equation form, represented by (2).

$$Q_L = \frac{R_P}{X_P} \quad (2)$$

where R_p is the equivalent parallel resistance of R_S and R_L, and X_p is either the inductive or capacitive reactance (equal at resonance). From (1) it is evident [10] that phase noise (at a given offset frequency) improves with increasing carrier power (which is considered a trade-off in a low power design strategy as followed in this paper) as the ratio to the fixed thermal noise is increased, and increasing Q_L. At large frequency separations, the flat noise floor in dBc/Hz simply becomes the difference between the power delivered into the amplifier and the noise floor of the amplifier in dBm/Hz. Also, at frequencies close to the carrier frequency, the bandwidth of the filter causes the noise that is produced at the amplifier output to be amplified with a positive feedback that depends on the frequency separation. Finally, it can be noted that at some frequency separation, flicker noise will cause phase modulation [9].

The flicker noise slopes at 10 dB/decade and changes the slope of the amplifier phase shift from 20 dB/decade to 30 dB/decade. In [10] it is shown that by directly applying Leeson's formula, a maximum (flattened) noise floor of -174 dBc/Hz can be achieved at 100 kHz from the offset frequency.

As phase noise is the ratio of power in the carrier compared to the power present at the specified offset, it is evident that this ratio will increase as the offset frequency decreases. This relationship is depicted in Fig. 2.

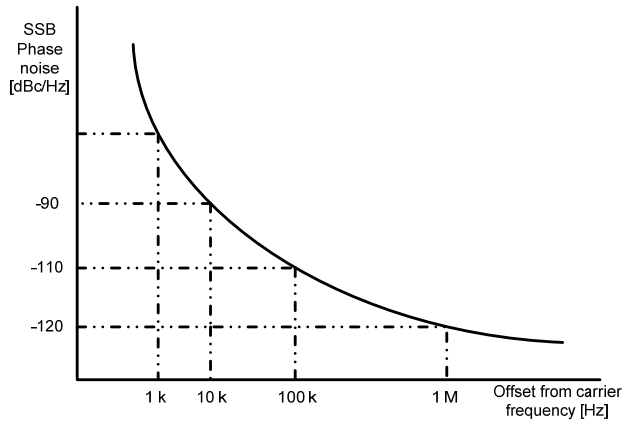


Fig. 2: Single sideband phase noise relative to offset frequency

B. Oscillator Figure of Merit

To evaluate the overall performance of a VCO, the figure of merit (FOM) can be determined by [11]

$$FOM = L\{\Delta\omega\} + 10\log P_{DC} - 20\log\left(\frac{\omega_o}{\Delta\omega}\right) \quad (3)$$

where $L\{\Delta\omega\}$ is the single sideband phase noise measured at a frequency $\Delta\omega$ from the carrier, and P_{DC} is the direct current (DC) power consumption in milliwatt (mW). From [5], this equation can be rewritten as

$$FOM = 10\log\left(\left(\frac{\omega_o}{\Delta\omega}\right)^2 \frac{1}{L\{\Delta\omega\}P_{DC}}\right) \quad (4)$$

where the inverse proportionality between the FOM and the power in the carrier is evident.

III. Considerations to Reduce Phase Noise

A. Process Technology

Some very good oscillator specifications have been achieved to date using CMOS, BiCMOS, and InGaAs/GaAs technologies. It is therefore important to weigh performance versus cost-effectiveness and ease of implementation as trade-offs when choosing the process technology.

A 5 GHz VCO designed in a 0.13 µm partially depleted silicon-on-insulator (SOI) CMOS has achieved a phase noise of -116 dBc/Hz at an offset of 1 MHz from the carrier. This design is however expensive due to the high resistivity SOI

substrate and needs additional digital circuitry for the tuning varactors [12].

Recently [13], GaAs metamorphic high electron mobility transistor (mHEMT) based VCOs have achieved phase noise ratings of -120 dBc/Hz at 1 MHz offset (with a 7 GHz carrier), proving this technology to be more than sufficient. GaAs transistors exhibit extremely low noise figures and high f_m, but have low breakdown and operating voltages. Positive and negative voltage supplies are also needed and foundries are only now commercializing their mHEMT processes, making it expensive.

A 1.8 – 6 GHz multi-band (achieved using switched resonators) SiGe BiCMOS based VCO has been presented in [14]. This design operates at a supply voltage as low as 0.29 V due to the low knee-voltage provided by the technology, and consumes 580 µW of power. The design exhibits low power operation, good phase noise performance (-112.2 dBc/Hz at 1 MHz offset from a 5.8 GHz carrier) and a wide tuning range. To achieve the high tuning range, switched resonators are used that consist of MOS transistors, which are lossy and degrade the resonator Q.

From the above examples, it is evident that all these technologies, combined with some enhancement technique, display very good phase noise performance. Considering cost-effectiveness versus performance and ease of implementation, SiGe BiCMOS was chosen. An advantage of this process (without SOI) compared to GaAs, is its compatibility with silicon VLSI processes and scalability to higher current densities. This technology will be implemented using Austriamicrosystem's (AMS) 0.35 µm S35D4M5 SiGe BiCMOS process.

B. Topology consideration

Different topology implementations in designing a high-frequency VCO exist, and these topologies were considered to determine which would display the lowest phase noise performance whilst maintaining cost-effective implementation and modularity.

Crystal oscillators were not considered for this design. These oscillators present very high quality factors (50,000 in some cases [15]) in the MHz range, but implementation in the GHz frequency spectrum are very expensive and tuning range is limited.

Ring oscillators can be used to implement a GHz VCO, however these oscillators cannot present phase noise performance comparable to that of LC oscillators. Phase noise ratings of -96 dBc/Hz at a 1 MHz offset from the 4.47 GHz carrier have been reported in [16]. They present a large tuning range, but also occupy large areas on chip.

Relatively, LC oscillators display the most attractive characteristics for high-frequency, low phase noise VCO design. These oscillators consist of a tank circuit with a quality factor that can be improved by improving the Q of the inductor in the tank as capacitors generally have a much higher Q and are not considered the limiting factor. Two switching transistors enable oscillation at the outputs and a current source supplies the current to sustain oscillation. All of these factors can be modified in the design to ultimately increase phase noise performance, making LC oscillators the chosen topology for this design. Phase noise ratings of between -110 and -130 dBc/Hz at an offset of 1 MHz from their respective 5 GHz carriers have been reported in [1], [2], [12], [17], [18], and [19].

Fig. 3 shows the typical configuration of a LC VCO. This configuration uses inductors L_1 and L_2 together with capacitors C_1 and C_2 to obtain the oscillation frequency (f_c) using the following equation.

$$f_c = \frac{1}{2\pi\sqrt{LC}} \tag{5}$$

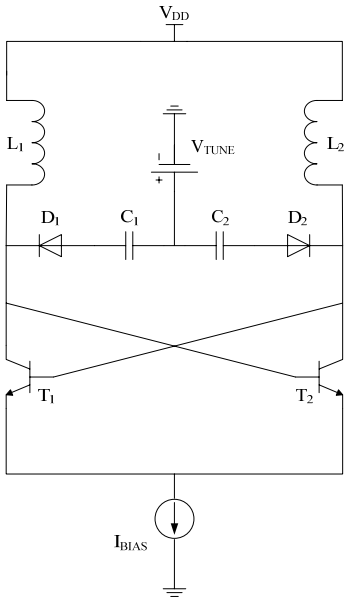


Fig. 3: Typical configuration of a LC VCO using bipolar transistors

V_{TUNE} is applied at the cathode of diodes D_1 and D_2 to tune the centre frequency to a certain range. Transistors T_1 and T_2 are the switching transistors and the outputs are measured at the collectors of these transistors.

C. Inductor dimensioning

For this paper, the standard thick metal inductors supplied by the AMS 0.35 µm process were used. There are however different inductor implementations that can alternatively be used to improve inductor Q and ultimately the quality of the tank [27]. The quality factor of the tank is one of the main contributors to phase noise and it is therefore important to design with a high Q in mind.

External inductors that present very good quality factors can be used. These inductors are not integrated on-chip, and therefore may suffer from parasitic effects induced by long bondwires between the package pins and the inductor connections, altering the inductance of the inductors and have detrimental effects on the design. Toroid inductors are commonly used due to their relatively small physical size and low inductance values. The inductance of a toroid can be calculated using (5).

$$L \cong 0.01257N^2\left(R-\sqrt{R-a^2}\right) \quad (6)$$

In (5), L is the inductance (in μH), R is the mean diameter of the windings to the core, N is the number of turns (windings), and a is the diameter of the windings.

Microelectromechanical systems (MEMS) inductors minimize ohmic and substrate-induced losses by suspending the inductor 40 – 50 μm above the substrate. These inductors can display Q factors of over 25 in the 1 – 4 GHz range [8]. Implementing inductors using this method requires modifications to standard fabrication processes, making these designs less cost-effective.

The characteristics of a bond wire (inductance and quality factor) are determined by the diameter of the bond wire, the height above the ground plane and the spacing between adjacent bond wires. By changing these dimensions, it is possible to control the inductance and Q factor of the bond wires for use in integrated circuits. These inductors can display quality factors of 25 at 1 GHz operating frequency [20] but their inductance is limited to generally above 1 nH making them ineffective at very high frequencies (5 GHz and above).

Spiral inductors are commonly used for on-chip integration and can be constructed in a number of configurations (square, hexagonal, octagonal, and circular). Square spiral inductors presented by AMS were used in this design as these inductors present reasonable Q factors (maximum of 11.8 at 5 GHz for a 1.05 nH inductor) due to the use of thick metal layers and careful placement of NTUB masked layers to reduce eddy currents. Fig. 4 shows the basic layout of a square inductor.

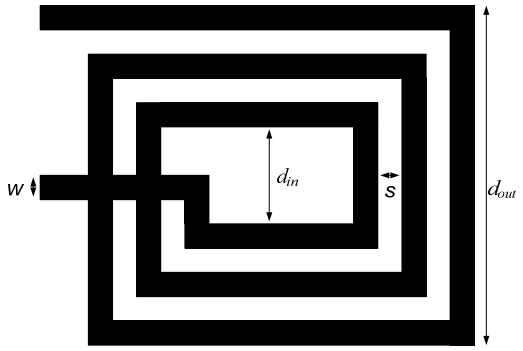


Fig. 4: Basic layout of a square spiral inductor used for on-chip integration

From Fig. 4, w is the conductor width, s is the spacing between inductors and d_{in} and d_{out} are the inner and outer diameters of the spiral respectively.

D. Tail-current shaping

Tail-current noise suppression in RF BiCMOS VCOs prevents the low-frequency tail-current noise from being converted into phase noise during normal operation of the oscillator [5]. High-frequency tail-current noise at twice the oscillation frequency is down-converted into phase noise by the hard switching oscillator. Subsequently, it has been shown that, to a first approximation, phase noise contribution from the switches are independent of their transconductance, and a closed-form expression has been given for phase noise generated by high-frequency noise sources in the VCO [4]. Fig. 5 is a representation of the tail-current shaping technique.

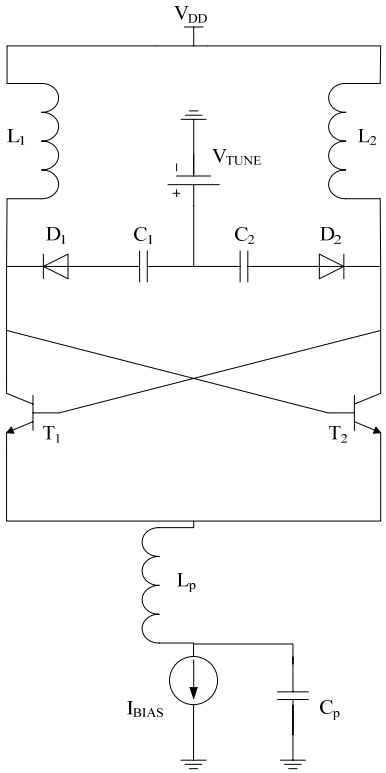


Fig. 5: Basic layout of a square spiral inductor used for on-chip integration

This technique entails the use of an additional inductor (L_p) in series between the switching transistors and the MOS current source to increase the impedance path between these sections at the oscillation frequency. It should be noted that MOS transistors are used in the tail-current source as its low frequency white noise which is up-converted to phase noise can be reduced by reducing its gain (g_m) through sizing of its aspect ratio (W/L). The power of the MOS transistor in the current source is reduced by the factor $\left|1+jg_m\omega L_p\right|$ [14] where g_m is the transconductance of the transistor and ω is the operating frequency. Capacitor C_p creates a small reactance path at the second harmonic, effectively shorting the noise effects at 2ω towards ground. From [21], by using the effective noise current power, the following equation for phase noise was derived.

$$L\{\Delta\omega\} = 10\log\frac{\overline{v_{noise}^2}}{v_{signal}^2} = 10\log\left[\frac{2FkT}{P_{SBC}}\left(\frac{\omega_o}{2Q\Delta\omega}\right)^2\right] \quad (7)$$

where F is the known device excess noise number in dB, k is Boltzmann's constant (8.6173 x 10^{-15} eV/K), T is the absolute temperature in Kelvin (0 K), P_{SBC} is the single sideband power (dBc) present in the carrier frequency (ω_o), $\Delta\omega$ is the offset

frequency from the carrier, and Q is the effective quality factor of the tank circuit. P_{SBC} is given by

$$P_{SBC}\left\{\Delta\omega\right\} = 10\log\left(\frac{I_n c_n}{4q_{\max}\Delta\omega}\right)^2 \tag{8}$$

where I_n is the current amplitude (in A) at the n^{th} harmonic, and q_{max} is the maximum value of the total injected charge due to the current impulse, and c_n are real-valued coefficients of the Fourier series.

Rewriting (7) as

$$L\left\{\Delta\omega\right\} = 10\log\left(\frac{\frac{i_n^2}{\Delta f}\sum_{n=0}^{\infty} c_n{}^2}{8q_{\max}^2\left(\Delta\omega\right)^2}\right) \tag{9}$$

and applying Parseval's theorem

$$\sum_{n=0}^{\infty} c_n{}^2 = \frac{1}{\pi}\int_0^{2\pi}\Gamma(x)^2\,dx = 2\Gamma_{eff,rms}^2, \tag{10}$$

and realizing $C = \Delta q / \Delta V$, the phase noise in the tail-current filter can be defined as

$$L\left\{\Delta\omega\right\} = 10\log\left(\frac{\frac{i_n^2}{\Delta f}\Gamma_{eff,rms}^2}{8C_p^2A^2(\Delta\omega)^2}\right) \tag{11}$$

where $i_n^2/\Delta f$ is the current source power spectral density of the tank, A is the signal amplitude and C_p is the filtering capacitor. It is now possible to observe that the phase noise decreases with increasing the filter capacitor.

Very few documented results were found using tail-current shaping through inductor and capacitor filtering for VCOs operating in the 5 GHz frequency range. In [22] inductive filtering was used and a phase noise figure of -115 dBc/Hz at a 1 MHz offset from a 4.52 GHz centre frequency using TSMC 0.18 µm technology was presented. In [23] a phase noise of -90 dBc/Hz was achieved but at an offset frequency specified at 100 kHz from the 5.5 GHz carrier, using 0.18 µm CMOS technology.

The work discussed in this paper achieved through simulation a phase noise rating of about -110 dBc/Hz at a 1 MHz offset from a 5 GHz carrier and proving that tail-current noise suppression does have a noticeable effect on overall phase noise performance.

IV. Simulation Results

Following the design criteria specified in III to reduce phase noise, the following simulation results were achieved. These results verify that tail current shaping does in fact have improved on phase noise performance of a LC VCO.

Two circuits are compared. The first circuit is a normal LC VCO without tail-current shaping, and the second is the same circuit, but with tail-current shaping.

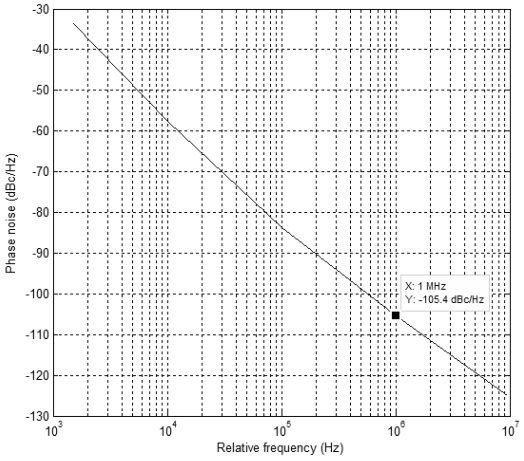


Fig. 6: Phase noise simulation results without tail-current shaping

The oscillation frequency for the first circuit (simulated with Cadence® Virtuoso® and the Spectre® engine) is 4.968 GHz, with a tail current of 850 µA. From Fig. 6, a phase noise rating of -105.4 dBc/Hz was simulated at a 1 MHz offset frequency, and -83.74 dBc/Hz at a 100 kHz offset. The second circuit is identical to the first, except that a tail-current filter inductor and capacitors is introduced. Fig. 7 displays the results of this circuit. The oscillation frequency of the circuit has now changed somewhat, to 5.0034 GHz due to the addition of these components to the emitters of the switching transistors. It can however be noted that the phase noise performance has increased significantly to -108.6 dBc/Hz at a 1 MHz offset (3.2 dBc/Hz improvement) and -88.48 dBc/Hz at a 100 kHz offset (4.74 dBc/Hz improvement). Both circuits employ the same current sources and the same tank circuits. The inductors used in the tank circuits have an inductance of 1.58 nH at 5 GHz oscillation, with a Q of 10.2 at this frequency.

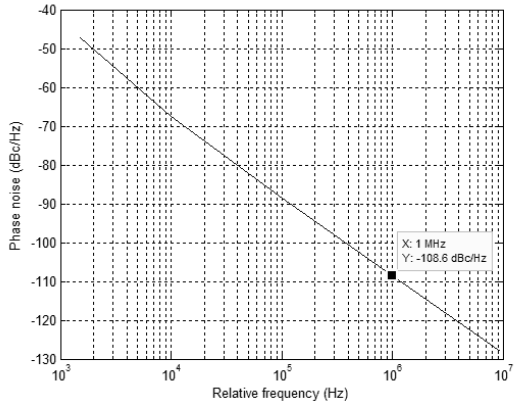


Fig. 7: Phase noise simulation results with tail-current shaping

Simulation results therefore confirm that tail-current shaping has enhanced the phase noise performance of a LC BiCMOS VCO.

V. Conclusion

From [5], [24], [25], and [26] an expected improvement of between 3 and 15 dBc/Hz is expected when implementing tail current filtering. Designing the VCO using the differential LC topology, should provide a phase noise rating of between -100 and -110 dBc/Hz at 1 MHz offset without any noise improvement techniques [9], [26], and [27], offering an estimate of 7 dB phase noise improvement from conventional quadrature VCO (QVCO) implementations [18]. Simulation results confirmed this hypothesis, with a 3.2 dBc/Hz improvement at a 1 MHz offset from the 5 GHz carrier. Phase noise ratings are however not as high as expected (best figure obtainable -108.6 dBc/Hz at a 1 MHz offset). This can be traced back to the fact that a relatively cost-effective design was employed in a 0.35 µm process, with limited inductor *Q*. Phase noise improvement has however been confirmed.

Acknowledgment

The author would like to thank Grintek Ewation (Pty) Ltd (www.gew.co.za) for funding the research project.

References

[1] T. K. Johansen and L. E. Larson, "Optimization of SiGe HBT VCOs for Wireless Applications," *Proc. Of IEEE RF IC Symposium, Philadelphia*, pp. 273-276, 8-10 June. 2003.

[2] P. Kinget, "Integrated GHz Voltage Controlled Oscillators," Bell Labs, Aug. 1998.

[3] A. Maxim, "A Multi-Rate 9.953–12.5-GHz 0.2-µm SiGe BiCMOS LC Oscillator Using a Resistor-Tuned Varactor and a Supply Pushing Cancellation Circuit," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 918-934, April 2006.

[4] S. Saeedi, S. Mehrmanesh, A. Tajalli and M. Atarodi, "A Technique to Suppress Tail Current Flicker Noise in CMOS LC VCOs," *Proc. Of IEEE RF IC Symposium*, San Francisco, pp. 3229-3233, Jun. 2006.

[5] P. R. Kinget and B. Soltanian, "A Tail Current-Shaping Technique to Reduce Phase Noise in LC VCOs," *Proc. Of the IEEE 2005 Custom IC Conf,* San Jose, pp. 579-582, 18-21 Sept. 2005..

[6] S. Saeedi, S. Mehrmanesh, A. Tajalli and M. Atarodi, "A Technique to Suppress Tail Current Flicker Noise in CMOS LC VCOs," *Proc. Of IEEE RF IC Symposium*, San Francisco, pp. 3229-3233, June 2006.

[7] T. H. Ning. (2002). IBM Journal of Research and Development. Why BiCMOS and SOI BiCMOS?. [Online]. Available: http://www.research.ibm.com/journal/rd/462/ning.html

[8] E. Park, S. Baek and J. Yoon, "Fully Integrated Low Phase-Noise VCOs With On-Chip MEMS Inductors," IEEE Tans. On Micro. Theory and Techniques, vol. 51, no. 1, Part 2, pp. 289-296, Jan. 2003.

[9] M. Rudolph, R. Doerner, L. Klapproth and P. Heymann. "An HBT Noise Model Valid Up to Transit Frequency," *IEEE Electron Device Letters*, vol. 20, no. 1, pp. 24-26, Jan. 1999.

[10] X. Huang, F. Tan, W. Wei and W. Fu, "A Revisit To Phase Noise Model Of Leeson," *Proc. Of Frequency Control Symposium*, Miami, pp. 238-241, 29 May-1 June 2007.

[11] A. Hajimiri and T. H. Lee, "Oscillator Phase Noise: A tutorial," *IEEE J. Solid-State Circuits*, vol. 35, no. 3, pp. 326-336, Mar. 2000.

[12] P. Delatte, G. Picun, L. Demeus and P. Simon, "A low-power 5 GHz CMOS LC-VCO Optimized for High Resistivity SOI Substrates," *Proc. Of ESSCIRC,* Grenoble, vol. 4, pp. 395-399, 12-16 Sept. 2005.

[13] M. Ferndahl and H. Zirath, "Residual and oscillator phase noise in GaAs metamorphic HEMTs," *Proc. Of Asia-Pacific Micro. Conf.*, Yokohama, pp. 472-475, 12-15 Dec. 2006.

[14] R. Mukhopadhyay, C. Lee, J. Laskar, "A 580 µW 1.8-6 GHz Multiband Switched-Resonator SiGe VCO with 0.3 V Supply Voltage," *IEEE & Wireless Components Letter*, vol. 17, no.11, pp. 793-795, Nov. 2007.

[15] K. K. Clarke and D. T. Hess, Communication circuits: Analysis and Design, London, Addison-Wesley Publishing Company, 1974.

[16] A. Hajimiri, S. Limotyrakis and T. H. Lee, "Phase Noise in Multi-Gigahertz CMOS Ring Oscillators," *Proc. Of IEEE Custom IC Conf.*, Santa Clara, pp. 49-52, 11-14 May 1998.

[17] J. Zhan, J. Conan, J. Duster and K. T. Kornegay, "Analysis and Design of Negative Impedance LC Oscillators using Bipolar Transistors," *IEEE Trans. On Circuits and Systems*, vol. 50, no. 11, pp. 1461-1465, Nov. 2003.

[18] O. Mazouffre, H. Lapuyade, J. B. Begeuret, A. Cathelin, D. Belot and Y. Deval, "A 5 GHz Low-Power Quadrature SiGe VCO with Automatic Amplitude Control," *IEEE J. Solid-State Circuits*, pp. 57- 60, 28-30 Sept. 2003.

[19] D. I. Sanderson and S. Raman, "5-6 GHz SiGe VCO with Tunable Polyphase Output for Analog Image Rejection and I/Q Mismatch Compensation," *Proc. Of IEEE RF IC Symposium*, San Jose, pp. 683-686, 8-10 June 2003.

[20] K. Kim, K. H. Ahn and T. H. Lin, "Low Phase Noise Bond Wire VCO for DVB-H," *Proc. Of 4th IEEE Int. Symp. For Electronic Design*, Hong Kong, pp. 103-106, 23-25 Jan. 2008.

[21] A. Hajimiri and T. H. Lee, "A General Theory of Phase Noise in Electrical Oscillators," *IEEE J. Solid-State Circuits*, vol. 33, no. 2, pp. 179-194, Feb. 1998.

[22] W. Liou, T. Chen, M. Yeh and J. Ho, "A low-noise low-power 4.6 GHz CMOS VCO with coupling effect reduced inductors," *Intl. J. Of Electronics*, vol. 94, no. 2, pp. 107-121, Feb. 2007.

[23] O. Seung, J. Byeong, N. Yoo and K. Taek, "A 5.5 GHz, 8 mW, LC Tank CMOS VCO with Optimal phase noise performance," *Proc. Of Asia-Pacific Micro. Conf.*, Yokohama, pp. 989-992, 12-15 Dec. 2006.

[24] S. Saeedi, S. Mehrmanesh, A. Tajalli and M. Atarodi, "A Technique to Suppress Tail Current Flicker Noise in CMOS LC VCOs," *Proc. Of IEEE RF IC Symp.*, San Francisco, pp. 3229-3233, 11-13 June 2006.

[25] N. Itoh, "A Study of Capacitor-Coupled Varactor VCO to Investigate Flicker Noise Up-Conversion Mechanism," *Proc. Of 2005 IEEE Workshop on Sensor Networks*, pp. 157-160, 30 Nov.-2 Dec. 2005.

[26] L. Jia, Y. B. Choi and W. G. Yeoh, "A 5.8 GHz VCO with Precision Gain Control," *Proc. Of IEEE RF IC Symposium*, Honolulu, pp. 701-705, 3-5 June 2007.

[27] M. Božanić and S. Sinha, "Design methodology for a CMOS based power amplifier deploying a passive inductor," *Proceedings of IEEE Africon 2007,* Windhoek, pp. 1-7, 26-28 Sept. 2007.

[28] C. Bowick, RF Circuit Design, Indianapolis, SAMS, 1991.

On-chip Impedance Tuning for mm-wave Applications

B. C. Mabuza[1], *Student Member, IEEE*, S. Sinha[2], *Member, IEEE*, and D. Foty[3], *Senior Member, IEEE*
[1, 2]University of Pretoria, Department of Electrical, Electronic and Computer Engineering, South Africa
[3]Gilgamesh Associates LLC, 452 Black Mountain Rd., Fletcher VT USA
[1]bongani.mabuza@up.ac.za, [2]ssinha@ieee.org, [3]dfoty@sover.net

***Abstract*— On-chip impedance tuning is used to overcome IC perturbations caused by packaging stress. Tuning is even more important for matching networks of RF systems. Possible package resonance and fabrication process variations may cause instability and this is a major problem in RF systems. Thus precautions need to be taken in order to maintain overall stability of components and the final system itself. This paper proposes tuning by using EEPROM switches for mm-wave applications. EEPROM switches occupy less die area and this is the main advantage when compared to e-fuses and MEMS switches. It is anticipated that EEPROM switches will not only occupy less area, but will also enable for multi-time programming, adding a further advantage. With the use of more switches, fine tuning can also be achieved. The main focus of this paper is to qualitatively characterise the suitability of EEPROM switches for tuning with systems operating in the mm-wave spectrum. In this paper, the insertion (closed switch) and isolation (open switch) losses will be discussed.**

I. Introduction

THE demand for bandwidth for wireless transmission, especially for high volume data applications like video streaming and local area networks, is ever increasing. In complementary metal-oxide semiconductor (CMOS) design higher bandwidths have been achieved through device scaling. However, scaling has reached its limit and for over-scaled devices, a compromise is made between performance and the cost for advanced fabrication processes. Furthermore, the bandwidth gains that are achieved through device scaling are still low and limited [1].

New technologies have been proposed and are promising to provide better performance for high frequency applications. Some of these technologies are targeting the available 5-7 GHz unlicensed bandwidth around the 60 GHz region. Thus there is a great interest in millimetre-wave (mm-wave) applications. BiCMOS is such an enabling technology, especially for analogue and radio frequency (RF) applications, while CMOS remains dominant for digital applications.

The frequency characteristics of transistors are dominated by the intrinsic parameters; capacitance and resistance, *RC* time constant. The values of these parameters can differ from one fabrication process to another, operating temperatures, packaging stress and other controllable parameters like supply voltage. Since designs are based on initial intrinsic values, these variations will also affect the system performance or specifications. In order to overcome this, Post-Assembly tuning is desired to compensate for such process variations and reduce packaging stress. Thus reliable on-chip tuning circuits are required to vary the chip reactance accordingly. Testing against process variations can be expensive, thus by using programmable or reprogrammable tuning, testing cost can be greatly reduced. As an example, tuning can be used for the following applications; switch mode regulators, oscillators, radio frequency tags, RFID transponders and for transmission line matching networks.

Plastic packaging is known to mechanically stress integrated circuits (ICs), which can cause voltage threshold changes [2]. Thus in-package trim is important, in order to achieve high accuracy. Tuning is also required for chip identification (ID), RF frequency tuning and also for trimming of impedance matching networks. One method of adjusting IC perturbations is to have on-chip circuitry, modules or components that can be used to adjust the reactance of the IC. This is done by switching mechanisms that either connect or disconnect reactance within an IC. Tuning switches can be implemented by using fuses, microelectromechanical systems (MEMS) or electrically erasable programmable read-only memory (EEPROM).

The mm-wave applications have brought new challenges into the design of new systems [3]; the materials and operational behaviour have changed. Modelling of transistors is becoming more challenging and new models are necessary for analysis and circuit design. Thus performance evaluation needs to be done on current systems based on new developing technologies and a new design approach will need to be derived or followed based on these new challenges.

The main problem addressed in this paper is to characterise the performance of EEPROM switches for mm-wave applications. Section II discusses the tuning methods that can be used. Sub-systems and the design of EEPROM are discussed in section III. The performance characteristics are described in section IV. The paper is concluded with current and future trends discussion of using EEPROM switches for tuning and other applications, section V.

II. Tuning Methods

A. E-Fuse

The types of fuses that are available include metal fuses,

polysilicon fuses and lasercut fuses. All these fuses, before advanced polysilicon fuses, had the disadvantages of size, high blowing currents and a large separation distance required between fuses and all were limited to on-chip programming [4]. Another preferred requirement for tuning is the ability to tune on-chip and on-package. In addressing this, electrically programmable fuses (e-fuse) have been developed [4]-[7]. The electrical characterization of the e-fuse is based on four things: the blowing current, the blowing voltage, the blowing time and the resistance before and after blowing. Typical resistance values before and after blowing are in the range of <200 Ω and >1 kΩ, respectively. In 0.13 µm SiGe BiCMOS technology, values of 200 Ω and 5 kΩ have been reported [5]. Programming currents can range between 10 mA and 40 mA and; voltages ranging between 3.3 V and 6 V. The fuses can be blown to achieve both open/short circuit and the process is irreversible. FETs or PIN diodes can be used, but the turn-on voltage and poor isolation and insertion loss makes the use of these devices unfavourable.

B. MEMS

MEMS can also be used for tuning. There is considerable interest in the development of meso-scale MEMS devices fabricated using printed circuit board (PCB) processing techniques. PCB-MEMS technology is compatible with multi-chip module lamination technology [8]. Therefore monolithic integration of embedded passives and surface mount electronics could be accomplished using well established assembly and packaging techniques. Another advantage of MEMS is that RF-MEMS switches show better performance such as low loss, low power consumption, and no measurable inter-modulation. The key features of MEMS are described in [8]. One of the main disadvantages of MEMS is the size of these devices when compared to EEPROMs (discussed below) and also designing for repeatability or production [1]. A challenge is to achieve fast switching and low actuation voltage.

C. EEPROMs

Another alternative method for tuning is by using EEPROM switches. The advantages of EEPROM are that it offers multi-time programming and it is easy to trim once the die is packaged. It occupies a small area and thus more cells can be included, this will also enable fine tuning which is more essential for RF and matching networks. One of the concerns for e-fuses is the programming current, but EEPROMs require low current for trimming. A disadvantage that is inherent with EEPROMs is their process cost. Previous EEPROMs where fabricated in multi-polysilicon processes and multi-oxidation for thin SiO_2 layers [9]. Thus, many masks were required which resulted in longer process turnaround time, lower yield, higher cost and lower reliability.

Over the years single poly EEPROMs have been investigated and developed for standard CMOS processes [9]-[11]. This technology advancement enabled on-chip EEPROM to be easily realized in digital and analogue circuits, and there are also merits in lower cost, shorter turnaround time and higher reliability circuits. Another disadvantage of multi-poly EEPROMs is the high voltage programming requirement which is unfavourable for embedded applications. A novel single poly EEPROM has been proposed [11] and the voltage levels for single poly have been lowered from 10 V to 6.5 V.

III. EEPROM SWITCHES

A. Memory Cells

There are different structures by which memory cells can be implemented and the cells are characterized by three operations; writing, erasing and reading operations. A channel hot electron injection and Fowler Nordheim (FN) tunneling can be used for writing the cell. FN tunneling is done between the gate and n+ diffusion in the NMOS. Erasing the cell can be achieved by tunneling between the gate and p+ of the PMOS or between the n+ and gate of the NMOS. The mode of operation is determined by the CPG / CGN ratio [9]. The general structure for floating gate transistor [12] is shown in Fig. 1.

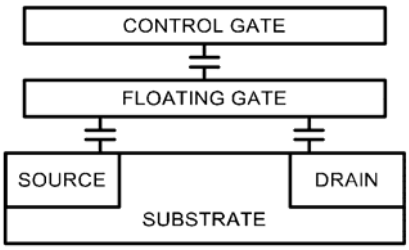


Fig. 1: Cross section of a floating gate transistor [12]

The capacitance at the drain and source changes as the charge is accumulated or removed from the floating gate. Since the source and drain are doped differently, both the drain and source capacitance would differ. Some EEPROM cells are programmed and erased via FN tunneling and composed of two transistors. In flash memory, the cell is programmed and erased electrically by a single transistor. A dual-control gate EEPROM cell can be used to eliminate the select transistor. Another advantage of dual-control gate is that it reduces hot-hole injection degradation induced by band-to-band tunneling [13].

B. Charge Pumps

For low power supply devices, as is the case with new technologies and mm-wave systems, the programming voltage is higher than the supplied voltage. Thus charge pump circuits are used to generate higher programming voltage [14]. The charge pump circuits must be designed to minimize the current drawn from the supply, this is important since the current drawn by the charge pump can be several times the current used by the cell. It is based on this that most power can be consumed by the charge pump circuit. The charge pump circuit is designed with *N*-stages which can be

controlled by either using switches (Fig. 2) or diodes (Fig. 3) [15].

The diodes must have very low turn-on voltage and the switches must have good isolation and insertion loss. The structure of the circuit that has been used mostly is Dickson charge pump [16], Fig. 4, but other structures which can offer other advantages have been proposed in the literature [15]. Charge pumps are used for writing operation and can use up to 90 % of the supply current.

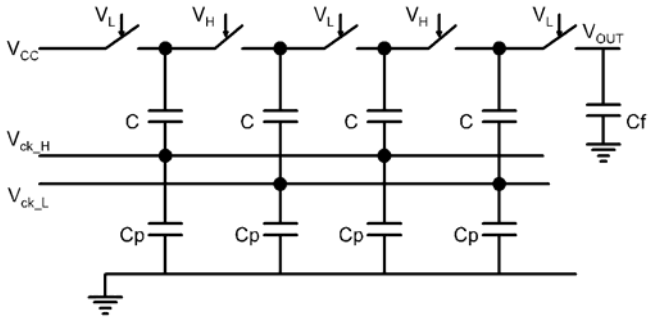


Fig. 2: Charge pump with ideal switches [15]

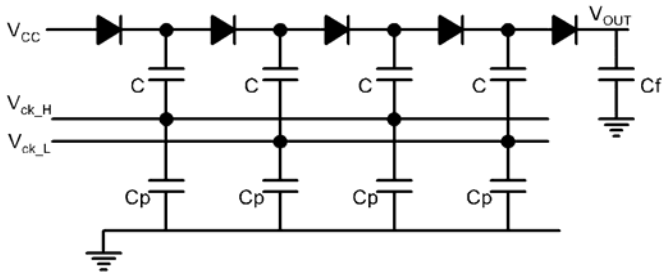


Fig. 3: Charge pump with diodes [15]

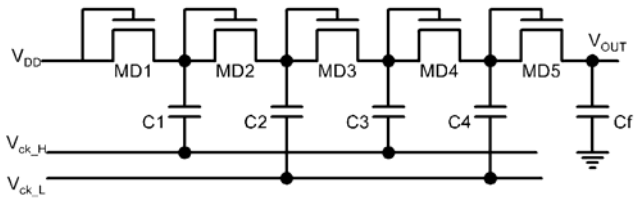


Fig. 4: A four stage Dickson charge pump [16]

C. *Sense Amplifier*

The read operation is based on sense amplifiers which can be implemented as either current- and voltage-mode sensors. Current-mode sensors have been a choice for most applications with conventional sense amplifiers [17] (Fig. 5) having a disadvantage of large reference current and a small sense-line swing. In minimizing this problem, bit line direct sensing circuits as shown in Fig. 6 has been proposed as it offers better sense-line swing for low supply-voltage operations [17].An advantage of current-mode is that sensing of memory cells does not require large voltage swing.

A voltage-mode is used to overcome degeneration of the floating gate transistor which makes the reference current to fluctuate with the average readout currents [18]. The sensing speed of voltage-mode is limited by high-speed large-size memories, where there is very large bit-line capacitance. Fig. 7 shows a proposed structure for the voltage sense amplifier.

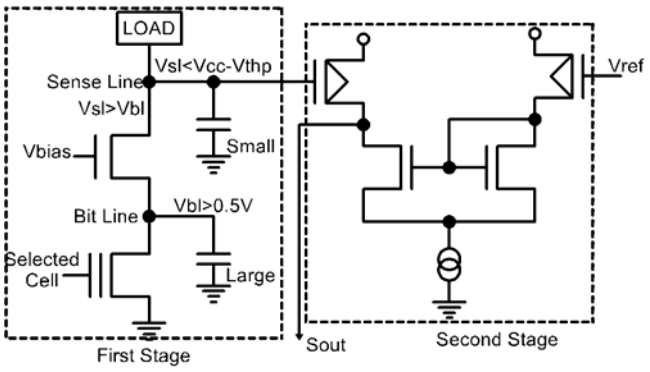


Fig. 5: Conventional sense-amplifier circuit [17]

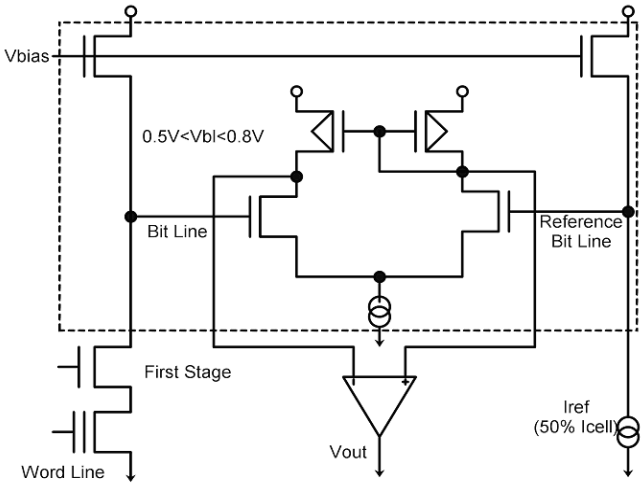


Fig. 6: Bit line direct sensing circuit [17]

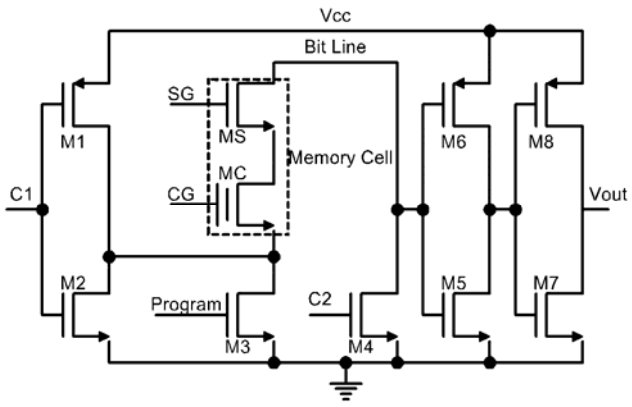


Fig. 7: Voltage sense-amplifier proposed in [18]

IV. CHARACTERIZATION

A. Reliability

The measure of reliability is determined based on retention and endurance tests. Endurance is a measure of the cell's performance based on the number of cycles for programming or erasing the memory cell. The stored information must be maintained with time, retention is a measure of this. A parameter that is also important for reliability concerns is the oxide thickness and its controllability to avoid defects. The oxide thickness thus is chosen to trade off between the performance and reliability concerns. For example, thin oxide poses problems because it is close to break down between the substrate and source junction. In addressing this problem, asymmetric structure is implemented between the source-substrate and the drain-substrate junction. The source is more doped than the drain, thus the diffusion for both source and the drain will be different [12],[17]. The dependency of the tunnel current on the oxide-electric field also causes process control problems.

Endurance and retention are used to determine the effects and causes of charge gain/loss which may change voltage thresholds. As the cell is programmed or erased, some charges are trapped or more charges can be drawn from the floating gate, thus this leads to threshold changes and also affects the durability of the cells.

B. Insertion and Isolation Losses

MEMS are known for their good insulation and isolation losses when compared to either EEPROMs or e-fuses. Even though this is the case, there are other reliability issues which make MEMS unfavorable. It is however expected that MEMS will in future be improved and that some of these issues will be addressed. Currently there is a great interest in MEMS and a lot is anticipated in future. The challenges of MEMS development are due to the multitude of failure mechanisms of such devices, i.e. cracking, creep, fatigue, deposition and growth on the contact surfaces, and static friction [19]. Other improvements that are still under investigation is fast switching and low actuating voltage.

Switches can be implemented as capacitive or resistive switches, but capacitive switches are mostly used in high frequency applications and they avoid large resistive losses which contribute to high power consumption. This is the same reason diode switches are not preferred; high power consumption and large resistive losses. Diodes are nonlinear, but have an advantage of switching speed.

Since the switches are capacitive, the insertion and isolation loss can be characterized by using either current or capacitive ratios in the ON and OFF states. The current of an OFF state can be in the range of 10 nA and this may be a concern for some of the measuring instruments. This value may be at threshold or below minimum sensible current and thus it may bring challenges during the testing process. The ON state current is in the range of 100 µA. The operation of the memory cell is based on a floating gate. The cell is written/erased by either injecting or removing a charge from the floating gate. Thus the current flow is controlled by the amount of charge present in the floating gate. The capacitance at the drain and source is also changed by the amount of charge. The capacitive ratio is thus different for the ON and OFF state and it can be used to determine the state of the switch. The isolation and insertion loss can therefore be measured using either current or capacitive ratios. An advantage of the floating gate is that it can be used at the extreme ends for an ON/OFF switch or it can be used for other characteristics which operate between the extreme states by controlling the charge in the floating gate [20].

C. Performance

The memory cell can be implemented by using either an n- or p-channel. N-channel has higher electron mobility and it suffers less drain disturbance than p-channel, but it has disadvantages of current degradation and closure occurs with increasing cycles. The p-channel achieves high speed, better reliability and lower power; it is fast in programming but slow in erasing [21]. Injection efficiency is one of the performance parameters which measures the ratio of the electrons collected at the floating gate to those generated in the channel. P-channel offers better performance than N-channel and it also has high injection efficiency.

Table I provides a comparison of MEMS switches against PIN and MESFET diode switches [22]. MEMS offer a number of advantages, such as small series resistance, low loss, good isolation and small size. But, it has slow switching speed and can require a high control voltage. PIN diodes are also slow and require a high control current. MESFET have high series resistance, higher insertion loss and the isolation can be lower by 20 dB. The advantages of MESFET are low control current and fast switching speed. EEPROMs have advantages of small size, fast switching and lower control currents. Control voltage and series resistance are expected to be well comparable to that of MESFET. The parasitics of switches have an influence on the performance of the switch. The effects include parasitic capacitance, resistance and transistor leakage currents [20].

TABLE I
RF MEMS VS. PIN AND MESFET SWITCH COMPARISON [22]

	MESFET	PIN Diode	MEMS
Series resistance (Ω)	3 to 5	1	< 1
Loss at 1 GHz (dB)	0.5 - 1	0.5 - 1	0.1
Isolation at 1 GHz (dB)	20 - 40	40	> 40
IP3 (dBm)	40 – 60	30 – 45	> 66
1 dB compression (dBm)	20 – 35	25 – 30	> 33
Size (mm^2)	1 -5	0.1	< 0.1
Switching speed	~ ns	~ µs	~ µs
Control voltage (V)	8	3 – 5	3 – 30
Control current	< 10 µA	10 mA	10 µA

These are the factors which affect the switches not to function ideally. For short circuit, insertion loss is due to capacitive and resistive loss along the path. While for open circuit, current leakages flowing through the switch contributes to isolation losses. Unlike for low frequency applications, for mm-wave applications the capacitive parasitics have a critical influence which cannot be ignored. All factors which contribute to either insertion or isolation loss of EEPROM switches need to be determined and evaluated for mm-wave applications.

V. CONCLUSION

There are different methods by which tuning can be achieved and these are not limited to after package trimming. RF circuits and matching networks are other examples where tuning is desired. The tuning circuits can be implemented by using either e-fuses, MEMS or EEPROMs. A full application of these devices for mm-wave is still under development and the main determining factor will depend on how these devices will perform in mm-wave spectrum.

This paper brings forward different tuning methods and qualitatively compares their performance parameters. The paper proposes EEPROM switches for tuning and qualitatively identifies the use of EEPROMs to offer advantages of speed, size and low-power consumption. These are the most important parameters for mm-wave applications. Although insertion and isolation loss will be poor compared to MEMS, but EEPROM switches can still achieve acceptable values for normal or desired operations.

REFERENCES

[1] D. Foty, "Prospects for Nanoscale Electron Devices: Some Little-Recognized Problems," *Proceedings: International Semiconductor Conference,* Sinaia, Romania, 13-15 October 2008, pp. 3-13.

[2] E. G. Gebreselasie, S. H. Volda, Z. X. He, D. Coolbaugh, R. M. Rassel, T. Kirihata, A. Paganini, C. G. Cox, S. A. S. Onge, J. S. Dunn, R. E. Halbach and J. M. Lukaitis, "Electrically Programmable Fuses for Analog and Mixed Signal Applications in Silicon Germanium BiCMOS Technologies," *Proceedings: IEEE Bipolar/BiCMOS Circuits and Technology Meeting,* Boston, USA, 30 September – 2 October 2007, pp. 238-241.

[3] A. Hajimiri, "mm-wave Silicon ICs: Challenges and Opportunities," *Proceedings: IEEE Custom Integrated Circuit Conference,* San Jose, USA, 16-19 September 2007, pp. 741-747.

[4] C. Kothandaraman, S. K. Iyer and S. S. Iyer, "Electrically Programmable Fuse (eFuse) Using Electromigration in Silicides," *IEEE Electron Device Letters,* vol. 23, no. 9, pp. 523-525, September 2002.

[5] M. Alavi, M. Bohr, J. Hicks, M. Denham, A. Cassens, D. Douglas and M. Tsai, "A PROM Element Based on Salicide Agglomeration of Poly Fuses in a CMOS Logic Process," *Proceedings: IEEE International Electron Devices Meeting,* Washington, USA, 7-12 December 1997, pp. 855-858.

[6] D. Maier-Schneider, S. Kolb, B. Winkler and W. M. Werner, "Novel Surface-Micromachined Low-Power Fuses for On-Chip Calibration," *Sensors and Actuators. A, Physical,* vol. 97-98, pp. 173-178, April 2002.

[7] W. R. Tonti, J. A. Fifield, J. Higgins, W. H. Guthrie, W. Berry and C. Narayan, "Product Specific Sub-Micron e-Fuse Reliability and Design Qualification," *IEEE International Integrated Reliability Workshop Final Report,* pp. 36-40, October 2003.

[8] R. Ramadoss, S. Lee, Y. C. Lee, V. M. Bright and K. C. Gupta, "RF-MEMS Capacitive Switches Fabricated Using Printed Circuit Processing Techniques," *Journal of Microelectromechanical Systems,* vol. 15, no. 6, pp. 1595-1604, December 2006.

[9] K. Ohsaki, N. Asamoto and S. Takagaki, "A Single Poly EEPROM Cell Structure for Use in Standard CMOS Processes," *IEEE Journal Solid-State Circuits,* vol. 29, no. 3, pp. 311-316, March 1994.

[10] C. Lin and C. Sun, "A Single-Poly EEPROM Cell Structure Compatible to Standard CMOS Process," *Solid-State Electronics,* vol. 51, no. 6, pp. 888-893, May 2007.

[11] H. Sung, T. F. Lei, T. Hsu, S. W. Wang, Y. Kao, Y. Lin and C. S. Wang, "Novel Single-Poly EEPROM with Damascene Control-Gate Structure," *IEEE Electron Device Letters,* vol. 26, no. 10, pp. 770-772, October 2005.

[12] P. Pavan, R. Bez, P. Olovo and E. Zanoni, "Flash Memory Cells – an Overview," *Proceedings of the IEEE,* vol. 85, no. 8, pp. 1248-1271, August 1997.

[13] A. Regnier, R. Laffont, R. Bouchakour and J. M. Mirabel, "A New Architecture of EEPROM for High Density Reliability Application," *Proceedings: IEEE Non-Volatile Memory Technology Symposium,* Marseille, France, 15-17 November 2004, pp. 143-148.

[14] R. E. Barnett and L. Jin, "An EEPROM Programming Controller for Passive UHF RFID Transponders With Gate Clock Regulation Loop and Current Surge Control," *IEEE Journal Solid-State Circuits,* vol. 43, no. 8, pp. 1808-1815, August 2008.

[15] G. Palumbo, D. Pappalardo and M. Gaibotti, "Charge-Pump Circuits: Power-Consumption Optimization," *IEEE Transaction on Circuits and Systems I: Fundamental Theory and Applications,* vol. 49, no. 11, pp. 1535-1542, November 2002.

[16] J. Wu and K. Chang, "MOS Charge Pumps for Low-Voltage Operation," *IEEE Journal of Solid-State Circuits,* vol. 33, no. 4, pp. 592-597, April 1998.

[17] S. Atsumi, A. Umezawa, T. Tanzawa, T. Taura, H. Shiga, Y. Takano, T. Miyaba, M. Matsui, H. Watanabe, K. Isobe, S. Kitamura, S. Yamada, M. Saito, S. Mori and T. Watanabe, "A Channel-Erasing 1.8-V-Only 32-Mb NOR Flash EEPROM with a Bitline Direct Sensing Scheme," *IEEE Journal of Solid-State Circuit,* vol. 35, no. 11, pp. 1648-1654, November 2000.

[18] D. Lui, X. Zou, F. Zhang and M. Deng, "Circuits and Implementation of a Low-Power Embedded EEPROM Memory," *Proceedings: 8th International Conference on Solid-State and Integrated Circuits Technology,* Shanghai, China, 23-26 October 2006, pp. 836-838.

[19] H. A. C. Tilmans, "MEMS Components for Wireless Communications," *Proceedings: EURO Sensors,* Prague, Czech Republic, 15-18 September 2002, pp. 1-34 (on CD-ROM only).

[20] T. S. Hall, C. M. Twingg, J. D. Gray, P. Hasler and D. V. Anderson, "Large-Scale Field-Programmable Analog Arrays for Analog Signal Processing," *IEEE Transaction on Circuits and Systems-I: Regular Paper,* vol. 52, no. 11, pp. 2298-2307, November 2005.

[21] S. S. Chung, S. T. Liaw, C. M. Yih, Z. H. Ho, C. J. Lin, D. S. Kuo and M. S. Liang, "N-channel Versus P-channel Flash EEPRM- Which One has Better Reliabilities," *Proceedings: IEEE International Reliability Physics Symposium,* Orlando, USA, 30 April – 3 May 2001, pp. 67-72.

[22] R. J. Richards and H. J. De Los Santos, "MEMS for RF/Microwave Wireless Applications: The Next Wave," *Microwave Journal,* vol. 44, no. 3, pp. 20+24+28+32-41, March 2001.

A Novel Electroluminescence Technique to Analyse Mixed Field Emission and Impact Ionisation Reverse Breakdown

Monuko du Plessis and Pieter Rademeyer

Abstract— **A novel technique to analyse the low voltage breakdown regime of silicon diodes is presented. It is shown that the field emission tunnel current component of the reverse current does not cause energy transitions of carriers, and therefore will not emit photons. All photons being emitted from the pn junction are due to avalanche electroluminescence as a result of hot carrier energy relaxation processes. Measuring the light intensity output as a function of reverse current, we are able to extract the two current components (field emission and impact ionisation) as a function of reverse voltage. This technique gives results that correlate well with photo-induced and excess noise characterisations performed previously by other researchers.**

Index Terms—**Avalanche breakdown, electroluminescence, light-emitting diodes, tunneling.**

I. Introduction

THE reverse breakdown characteristics of silicon p-n junctions have been investigated intensively over many years [1]. However, even today, the low-voltage breakdown phenomenon of mixed field emission and impact ionisation has not been fully understood, or the low voltage reverse breakdown current analysed in terms of the two different physical mechanisms [2].

One of the very first publications on mixed breakdown regime was published in 1957 [3]. The reverse current properties and the carrier multiplication characteristics of low voltage reverse breakdown diodes were investigated by the photo-induced injection of carriers into the junction. The multiplication factor M was defined as the ratio of the photocurrent at a given voltage to its constant value at small reverse bias voltages. The multiplication factor M was determined to be less than 2 for reverse bias voltages up to 3.6V [3]. The break over voltage of 3.6V is the average value of reverse voltage required for field emission carriers to acquire sufficient energy to produce impact ionisation electron-hole pairs before leaving the junction, and the onset of an increasing current represents multiplication of field emission carriers.

The field emission to avalanche transition region was also determined by measuring the differential dynamic impedance of the reverse biased diodes [4]. It was concluded that in the transition region the current is primarily caused by multiplied zener current. The starting point for multiplication in the transitional breakdown voltage region was determined as a reverse bias voltage ranging between 3.3V and 4.5V.

A more detailed analysis of breakdown in silicon [5] also established that the hardness in the reverse characteristics at a voltage slightly in excess of 3V is caused by the avalanche multiplication of field emission carriers. Since the field emission carriers travel on average a shorter distance through the depletion region than the thermally generated carriers (dominated in silicon by the recombination-generation current of carriers in the depletion region), the multiplication gains for the two current components will be different. For low voltage devices (breakdown voltage less than 6V) the current in the narrow junction devices are almost entirely caused by internal field emission [5].

A significant contribution was made in 1976 [6] to distinguish the tunneling current from the impact ionisation multiplied current, using noise as diagnostic tool. If the randomly emitted tunneling electrons undergo multiplication due to impact ionisation, the randomness of the multiplication process generates additional noise, and the noise ratio becomes greater than unity. The results show that the departure of the reverse current from the tunneling relation is due to the impact ionisation multiplication of the tunneling current. The simulation of low voltage breakdown devices [6] resulted in the principal conclusion that the transition from tunneling to impact ionisation occurs when the energy of the tunneling electrons exceeds the electron threshold energy of ionisation. A transition voltage of approximately 4V was observed [6] in the noise characteristics, and attributed to transitions from a maximum of two to a maximum of three ionising collisions per electron transit across the diode space charge layer.

Recently, photo-induced avalanche multiplication studies to measure the avalanche multiplication factors, even in the presence of large tunneling currents, were performed [8], [9]. The two experimental studies were similar, but more

This work was sponsored and financially supported by InSiAva (Pty) Ltd.

Monuko du Plessis is with the Carl and Emily Fuchs Institute for Microelectronics (CEFIM) at the University of Pretoria, Pretoria 0002, South Africa. (phone: +27 12 420-2952; fax: +27 12 362-5115; e-mail: monuko@up.ac.za).

Pieter Rademeyer is a consultant to InSiAva (Pty) Ltd, P.O. Box 14679, Hatfield 0028, Pretoria, South Africa.

advanced, than the original work reported in 1957 [1]. The data in [8] is for diode widths down to 100 nm, while [9] extends the measurements to diode widths as small as 30 nm. Both reported avalanche gains very similar to the results previously published [3].

In this paper we present a new and novel technique to distinguish between the field emission and avalanche currents in low voltage silicon diodes, using the light emission properties from these devices [10], [11]. It was noted very early on that very faint light is emitted from low voltage field emission diodes compared to avalanche diodes, almost a factor 100 smaller in intensity [3]. It was also reported recently that tunneling currents do not produce hot carriers and the resultant light emission, and can be viewed as a parasitic current, which decreases quantum efficiency of silicon light emitting devices [12].

II. Experimental Results

A. Low voltage silicon devices

The p^+n^+ low voltage silicon LED devices were manufactured at Austriamicrosystems (AMS), Austria, in a 0.8 µm BiCMOS process through the Europractice program. The structure of the diodes is shown in Fig. 1, and the reverse current vs. voltage characteristic in Fig. 2.

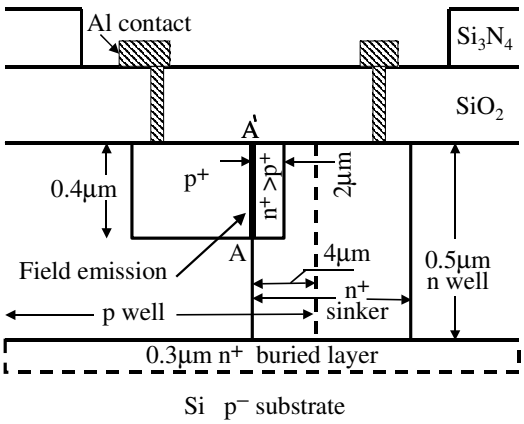


Fig. 1. Structure of the low voltage n^+p^+ diodes in a 0.8 µm BiCMOS process

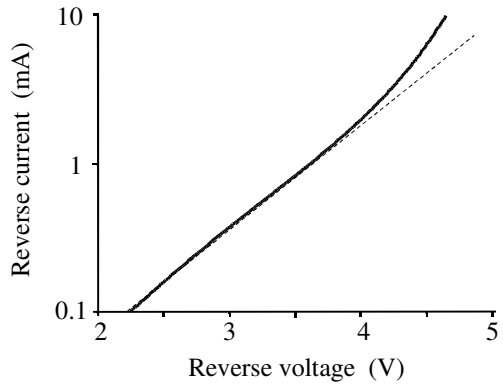


Fig. 2. Reverse current vs. voltage characteristic of the n^+p^+ diode.

In Fig. 2 the deviation from the tunnel current behaviour can be observed as an additional increase in reverse current at a voltage just below 4V.

B. Light emission

The visible light emission from the diode was in the form of a 30 micron long line segment where the n^+ and p^+ regions overlap. The linear segment light pattern is presented in Fig. 3.

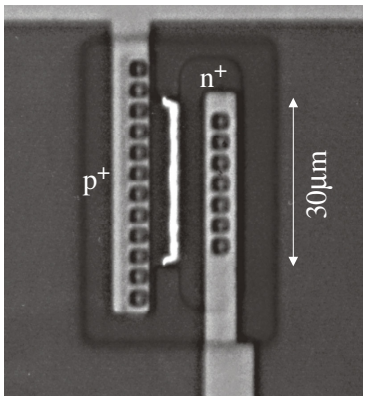


Fig. 3. Light output from the linear segment n^+p^+ diode

Optical light intensity measurements were performed using a R955 photomultiplier tube. A discrepancy was observed in the light emission versus reverse current characteristics of the low voltage silicon p^+n^+ diodes [10], [11]. Previously, diodes operating in the avalanche mode of operation with higher breakdown voltages, exhibited a linear optical output power versus reverse current relationship, however, the low voltage devices exhibit a non-linear behaviour at low currents levels, as seen in Fig. 4 [11].

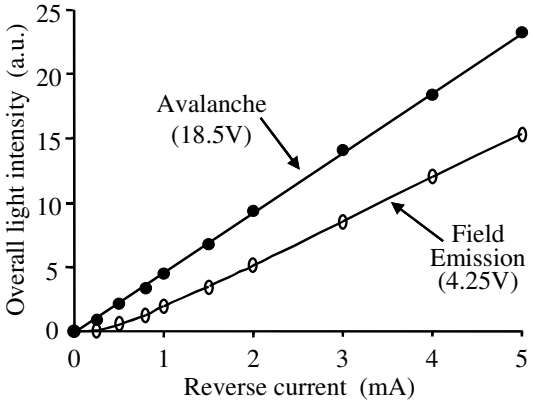


Fig. 4. Light output intensity as a function of the reverse current for low voltage field emission n^+p^+ diodes and high voltage avalanche devices [11].

The non-linear effect is attributed to the co-existence of a field emission (or tunneling current) and an avalanche current component as discussed in the introduction above. The tunnel current does not contribute to the optical power generation process, but only the avalanche component of current. The

tunnel current can thus be extracted from the optical power versus reverse current transfer characteristics.

The assumption that the tunnel current does not contribute to optical power output can be attributed to the effect that the tunnel current does not undergo band to band energy shifts, but only indirect electron tunneling takes place from the valence band on the p-side to the to available empty ionised sites in the conduction band on the n-side, with phonon and momentum energy dissipation in the lattice [13], [14].

C. Extraction of the different current components

The optical output power versus diode current curve was measured, with optical power integrated over the 400 nm to 900 nm spectrum, using a R955 photomultiplier tube, and the results are shown in Fig 5. In Fig. 5 the non-linear behaviour at low current levels can be observed clearly, as well as the linearity of the characteristic at larger current levels.

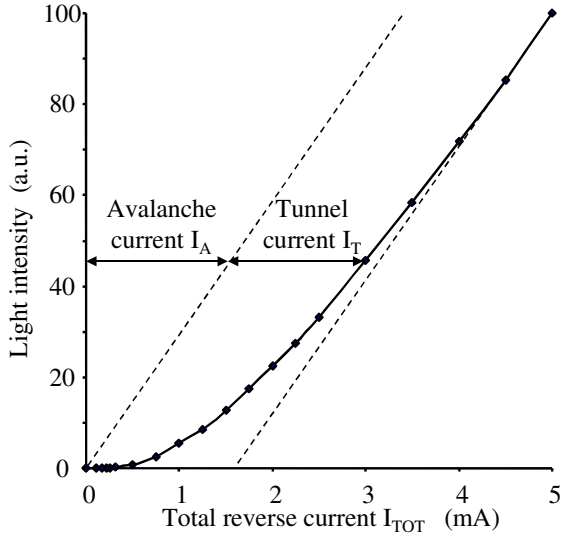


Fig. 5. Extraction of the tunnel current component from the light output intensity vs. reverse current characteristic.

High voltage avalanche current devices have been observed to result in linear light emission intensity vs. reverse current characteristics [11]. Under the assumption that the tunnel current does not contribute to the light emission process, the linear avalanche current component I_A can be shifted to the origin of the curve. Subtracting the avalanche component I_A from the total current I_{TOT} will result in the tunnel current I_T. Following this procedure, the two current components I_A and I_T can be plotted as a function of the total reverse current I_{TOT}, as shown in Fig. 6. From this figure it follows that at low current levels the tunnel current I_T is the dominant current flow mechanism, with almost no avalanching taking place. As the reverse current is increased the tunnel current saturates eventually and the avalanche current increases sharply to be the dominant current component mechanism at higher reverse current levels.

Using the current vs. voltage relationship of Fig. 2, the two current components can also be plotted as a function of the applied reverse bias voltage, as shown in Fig. 7.

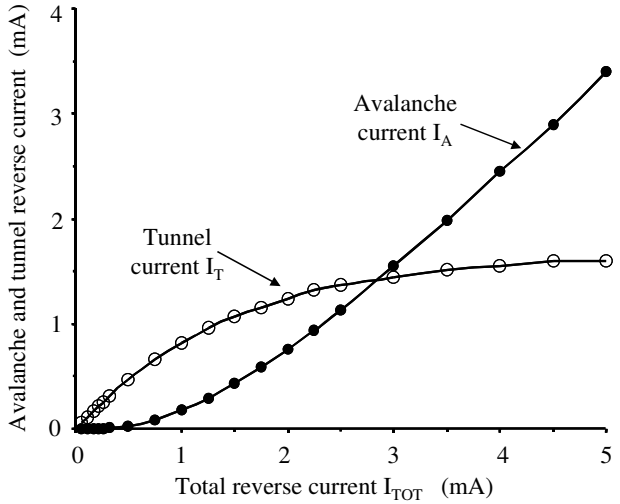


Fig. 6. The avalanche current component I_A and field emission tunnel current component I_T as a function of total reverse current I_{TOT}.

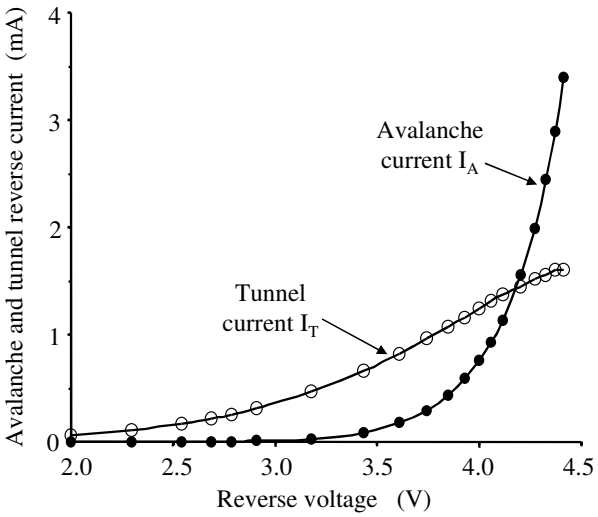


Fig. 7. The avalanche current component I_A and field emission tunnel current component I_T as a function of reverse bias voltage.

From Fig. 7 it is evident that the avalanche current only starts to appear and then increase at reverse voltages of about 3.5V, which correlate well with previous observations using other techniques [3], [4], [5], [6].

The avalanche gain M_A can be calculated as the ratio of total reverse current I_{TOT} to the tunnel current component I_T. At low current levels there is basically no avalanche multiplication due to impact ionisation, and the avalanche gain will then be given by $M_A = 1$. As the reverse voltage increases, the depletion region of the pn junction will widen, and the probability of the tunnel current carriers to initiate avalanche impact ionisation will increase. The avalanche gain M_A is expected to increase with an increase in reverse bias voltage. This effect is illustrated in Fig. 8. The avalanche gain measured using our light emission technique correlates very well with the avalanche gains reported by authors using different methods [3], [4], [5], [6].

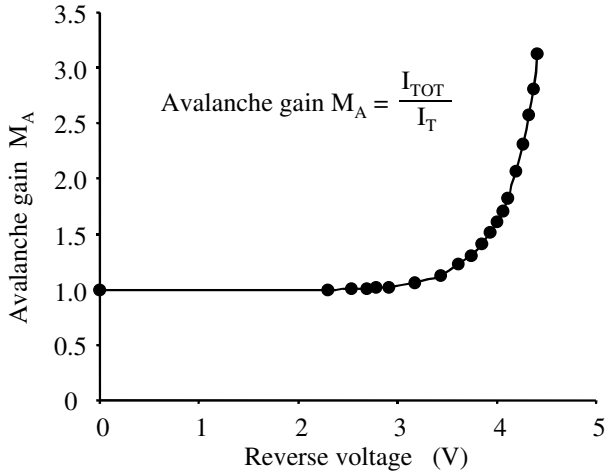


Fig. 8. The avalanche gain $M_A = I_{TOT}/I_T$ as a function of reverse bias voltage.

The avalanche gain shown in Fig. 8 starts to increase at reverse bias voltages in the vicinity of 3.5V, indicating that the impact ionisation of the tunneling carriers only start to happen at reverse bias voltages in excess of approximately 3.5V.

III. Theoretical Interpretation

Our theoretical interpretation of the observed behaviour is based on the theoretical models proposed earlier [3], [5]. The energy band diagram of the pn junction is showed in Fig. 9 to illustrate the tunneling principle. Fig. 9 shows the tunneling of an electron from the valence to conduction band at the metallurgical interface of the junction where the electric field is at its highest.

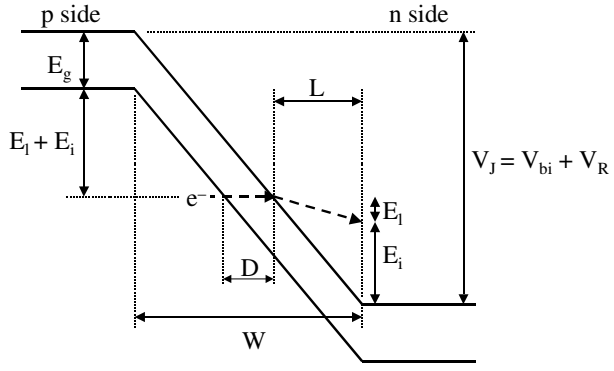


Fig. 9. The energy band diagram of the pn junction illustrating an electron tunneling from the valence to the conduction band at the position where the electric field is a maximum.

In Fig. 9 the following symbols are used to illustrate the tunneling and avalanche processes:

- E_g: Energy band gap of silicon
- E_l: Energy losses due to inelastic phonon scattering
- E_i: Impact ionisation threshold energy of electrons
- D: Barrier distance electrons tunnel through
- W: Total depletion region width
- L: Distance electrons will drift in the depletion region
- V_J: Total voltage across the pn junction
- V_{bi}: Built in potential of the pn junction
- V_R: Applied reverse bias voltage

An electron tunneling through the barrier width D of the junction will be accelerated over a distance L in the depletion region. During this acceleration it will lose energy E_l due to inelastic phonon scattering. If the energy of the carrier before leaving the depletion region is equal to the impact ionisation threshold energy E_i, there is a probability that the carrier may produce an electron-hole pair through impact ionisation, and contribute to the avalanche multiplication.

Assuming the mean free path for optical phonon scattering to be 7 nm and the energy loss is 0.063 eV per collision [5], the total energy loss E_l will be given by

$$E_l = 0.063 \times \frac{L}{7}\ eV \quad \text{with [L] = [nm]} \tag{1}$$

From Fig. 9 it is evident that the total junction potential V_J required for at least one instance of electron-hole pair production (L = 7 nm) by a field emitted carrier is given by

$$V_J = E_g + 2(E_i + E_l) = 1.1 + 2(1.8 + 0.063) \approx 4.8\,V \tag{2}$$

with E_i = 1.8 eV [5]. In our case the built-in potential is estimated as V_{bi} = 1.1V, giving an applied reverse bias voltage of approximately V_R = 3.7V needed to initiate avalanche multiplication. This value agrees well with the experimental results of Fig. 7 and Fig. 8.

IV. Conclusion

A new optical emission technique has been described to characterise the low voltage n^+p^+ silicon diode breakdown region of operation. The tunnel and avalanche components of the reverse breakdown current could be determined from the light intensity vs. reverse current characteristic. It was shown that at voltages less than 3.5 V the tunnel current dominates, while at larger values of reverse voltage the avalanche multiplication of the tunnel current will dominate. It has been shown theoretically that the threshold voltage for impact ionisation of the tunnel-injected carriers is in the order of 3.7 V, which correlates well with the experimental observation.

Acknowledgement

The authors thank InSiAva (Pty) Ltd for financially supporting the research.

References

[1] W. Maes, K. de Meyer, and R. van Overstraeten, "Impact ionization in silicon: A review and update", *Solid State Electronics*, Vol. 33, No. 6, pp. 705-718, 1990.

[2] C. I. Lee, V. H. Ngo, and D. S. Pan, "New phenomena of mixed breakdown in silicon", *Phys. Stat. Sol. (b)*, Vol. 243, No. 4, pp. R25-R27, 2006.

[3] A. G. Chynoweth and K. G. MacKay, "Internal field emission in silicon p-n junctions", *Physical Review*, Vol. 106, No. 3, pp.418-427, May 1957.

[4] H. Wienerth, “Silicon diode breakdown in the transition range between avalanche effect and field emission”, *Solid State Electronics*, Vol. 10, No. 11, pp. 1053-1062, 1967.

[5] M S. Tyagi, “Zener and avalanche breakdown in silicon alloyed p-n junctions – I. Analysis of reverse characteristics”, *Solid State Electronics*, Vol. 11, No. 1, pp. 99-115, 1968.

[6] W. A. Lukaszek, A. Van der Ziel, and E. R. Chenette, “Investigation of the transition from tunnelling to impact ionization multiplication in silicon p-n junctions”, *Solid State Electronics*, Vol. 19, No. 1, pp. 57-71, 1976.

[7] M. E. Eta and G. I. Haddad, “Mixed tunneling and avalanche mechanisms in p-n junctions and their effects on microwave transit-time devices”, *IEEE Transactions on Electron Devices*, Vol. ED-25, No. 6, pp.694-702, June 1978.

[8] D. J. Massey, J. P. R. David, C. H. Tan, B. K. Ng, G. J. Rees, D. J. Robbins, and D. C. Herbert, “Impact ionization in submicron silicon devices,” *J. Appl. Phys.*, Vol. 95, No. 10, pp. 5931–5933, 2004.

[9] P. Agarwal, M. J. Goossens, V. Zieren, E. Aksen, and J. W. Slotboom, “Impact ionization in thin silicon diodes”, IEEE Electron Device Letters, Vol. 25, No. 12, pp. 807-809, December 2004.

[10] H. Aharoni and M. du Plessis, “Low operating voltage integrated silicon light emitting devices.” *IEEE Journal of Quantum Electronics*, Vol. 40, Issue 5, pp.557-563, May 2004.

[11] M. du Plessis, H. Aharoni and L. W. Snyman, “Two- and multi-terminal silicon light emitting devices in standard CMOS/BiCMOS IC technology”, *Physica Status Solidi (a)*, Vol. 201, No. 10, pp. 2225-2233, July 2004.

[12] P. Jaguiro, P. Katsuba, S. Lazarouk and A. Smirnov, “Porous silicon avalanche LEDs and their application in optoelectronics and information displays”, *Acta Physica Polonica A*, Vol. 112, No. 5, pp. 1031-1036, 2007.

[13] S. M. Sze, *Physics of Semiconductor Devices*, New York: John Wiley & Sons, 1969, pp. 156-169.

[14] E. S. Yang, *Fundamentals of Semiconductor Devices*, New York: McGraw-Hill Inc., 1978, pp. 100-103

Microbolometer model and characterization

M.J. Swart and M. du Plessis
Carl and Emily Fuchs Institute for Microelectronics (CEFIM)
University of Pretoria, Pretoria

***Abstract*— Based on a theoretical electrical model of a bolometer element, the methods used to extract the different parameters required to model the bolometer element, are discussed. The operation of a bolometer is based on the infrared induced temperature change. Electrical resistance, thermal conductance and thermal capacitance are the three basic parameters of a bolometer element. The theoretical temperature dependencies of these parameters are formulated and form the basis for the characterisation and electrical modeling of the bolometer element. A number of prototype surface machined metal film bolometer elements were manufactured with different configurations. Bolometer element parameters were evaluated for different configurations. Using a Semiconductor Parameter Analyzer, the electrical parameters of the bolometer element are measured as a function of substrate temperatures and bolometer bias current. From these measurements the electrical resistance and the thermal conductance can be derived at different temperatures. Based on these values the thermal capacitance is derived by measuring the frequency response of the bolometer element. Based on the theoretical model and measured parameters a bolometer model was developed to analyze the dynamic operation of the bolometer element in terms of electrical parameters, as well as thermal parameters represented as electrical equivalences in the model..**

***Index Terms*—microbolometer, modeling, characterization**

I. Introduction

RESISTANCE based bolometer theory has been addressed by several authors [1]-[5] over the last 50 plus years.

Uncooled resistance based microbolometer theory is a subset of these. The bolometer is constructed by an infrared (IR) absorbent, temperature sensitive resistor, based on a thermally isolated IR detector pad without a cooling system to reduce the bolometer's temperature. The resistance of the element is then measured as an indication of the IR energy received.

Characterization of bolometers over the last 50 years was done with the tools at hand. With current computer based numeric simulation it was necessary to go back to the fundamental operation of the bolometer and reformulate the operation in terms of parameters for these new tools. The work focuses on modeling the bolometer operation rather than modeling the manufacture process or the physical geometry of the device. Bolometers' operation is based on changes in temperature and a number of system parameters are sensitive to temperature changes. This indicates that the modeling of an uncooled bolometer in terms of temperature changes is called for.

Acknowledgement:
The authors thank the Advanced Manufacturing Technology Strategy (AMTS) of the Department of Science and Technology, South Africa for financially supporting the research.

II. Structure of Prototype Microbolometers

A number of prototype CMOS based microbolometers were manufactured using titanium as the resistance element and aluminium as the sacrificial layer (Fig 1). Three units with different device dimensions were selected for measurement.

Table 1. Dimensions of the different microbolometers manufactured. See Fig 1. for the microbolometer structure.

Device	Dimension (D)	Supporting arm length (L)	Bolometer element membrane area	Resistor squares
MA1	60µm	22µm	$(29\mu m)^2$	23.8
MB1	72µm	22µm	$(41\mu m)^2$	23.8
MC1	99µm	41µm	$(41\mu m)^2$	31.4

In the literature a thin resistivity of 120 micro-ohm cm is mentioned [6] for a 100nm Ti thin firm. The sheet resistance should then be 12 ohm/square.

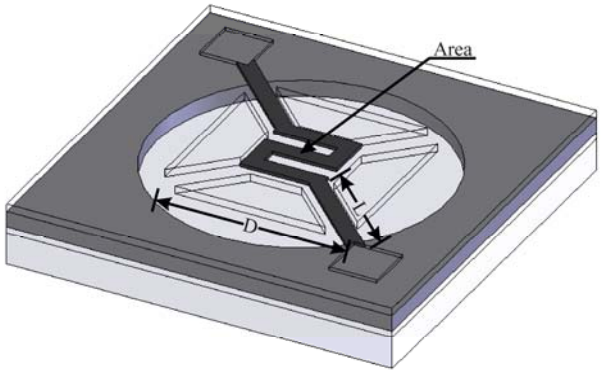


Fig 1. The microbolometers manufactured with different sizes, leg length and bolometer element or membrane areas as defined in table 1.

III. MICROBOLOMETER OPERATIONAL PARAMETERS

The temperature of the bolometer element is determined by the fraction of infrared power absorbed by the bolometer and the heat generated by any electrical current flowing through the bolometer resistor. It must be noted that no distinction exists between the effects on the bolometer variables as a result of these two sources of power input. Subsequently there will be concentrated on the Joule power input with the microbolometer shielded from infrared power.

The bolometer consists of two systems, a thermal system where the heat from the bolometer flows to its surroundings (or a heat sink) and an electrical system used to bias the bolometer and/or measure the bolometer element's resistance.

1. The thermal system is characterized by the thermal conductance between the bolometer element and its surroundings; Conductance being the power absorbed by the element divided by the temperature difference between the element and its surroundings. With this thermal conductance known energy balance calculations can be done giving the relationship between the bolometer power absorbed and the temperature rise in the bolometer element [7].

$$H\frac{dT(t)}{dt}+G\left[T(t)-T_a\right]=P(t) \quad (1)$$

with

H = the bolometer element's thermal capacitance
G = the thermal conductance between the bolometer element and its surroundings.
$T(t)$ = the bolometer element's temperature at time *t*
T_a = the ambient or heat sink temperature
$P(t)$ = the infrared and electrical power absorbed by the bolometer element.

In the steady state this can be simplified to:

$$P=G\bullet\Delta T=G\left(T-T_a\right) \quad (2)$$

With

T is the bolometer element's temperature

To summarize: the thermal system is characterized by the thermal conductance (G) and the thermal capacitance (H) of the microbolometer.

2. The electrical system is characterized by the bolometer element's electrical resistance. The bolometer resistance is formulated differently for a metallic bolometer resistance element and a semiconductor based bolometer element [1]. For metallic materials the electrical resistance

$$R(T)=R_0\left(1+\alpha\left[T-T_0\right]\right) \quad (3)$$

With

R_0 = Bolometer resistance at T_0 - constant
T_0 = Bolometer head sink temperature - reference point
α = Bolometer thermal coefficient of resistance (TCR)

$$\alpha=\frac{1}{R_0}\frac{R(T)-R_0}{T-T_0} \quad (4)$$

This α is a positive value for metallic resistance materials with a constant *R-T* slope.

To summarize: the electrical system is characterized by the microbolometer elements electrical resistance (R_0) at temperature T_0 and the Thermal Coefficient of Resistance (TCR) of the microbolometer element α .

For semiconductor resistance elements, the TCR is negative and the resistance change with temperature is based on an e function.

IV. THEORETICAL ESTIMATES OF OPERATIONAL PARAMETERS

Based on [8] for the material and process used in the microbolometer manufacturing the following calculations can be made:

In general the heat capacity *H* is given by

$$H=V\rho c \quad (5)$$

with

V = volume of the detector membrane,
ρ = the density of the membrane material
c = the specific heat of the membrane material

It was found that the heat capacity is independent of temperature [9].

The total thermal conductance *G* of a detector is the sum of three components [10]:

$$G=G_{leg}+G_{gas}+G_{rad} \quad (6)$$

G_{leg} is the conductance through the legs supporting the membrane, G_{gas} is the conductance through the surrounding gas, and G_{rad} is due to radiation. G_{rad} is the ultimate low limit of G given by:

$$G_{rad}=4\sigma\varepsilon A_{mem}T^3 \quad (7)$$

with

σ = Boltzmann constant = $5.67\text{x}10^{-8}$ W/m^{2}K
ε = effective emissivity = 0.88 for metal film bolometers with black coating [11].
A_{mem} = membrane area
T = membrane temperature

G_{gas} is the thermal loss through conduction. Only heat transfer from the membrane to the substrate is considered due to the small separation d between the membrane and substrate and is given by:

$$\frac{1}{G_{gas}} = \frac{d}{\lambda_{hp} A_{mem}} + \frac{d}{\gamma_{lp} P A_{mem}} \quad (8)$$

with

d = membrane to substrate separation.

λ_{hp} = pressure-independent thermal conductivity in the high pressure regime. = 0.026 W/mK for nitrogen at 300K.

λ_{lp} = thermal conductivity per unit pressure and length in the low pressure regime. = 1.9 m/sK for nitrogen at 300K.

Under low pressure conditions the leg conductance (G_{leg}) will be the dominant thermal conductance [9]. In calculating G_{leg}, it is assumed that conduction is completely due to the conduction through the supporting legs, and conduction within the membrane plate is negligible. The leg conductance of one leg G_{leg} is given by:

$$G_{leg} = \lambda_{leg} \frac{W_{leg} d_{leg}}{l_{leg}} \quad (9)$$

with

W_{leg} = width of supporting leg

d_{leg} = thickness of supporting leg

l_{leg} = length of supporting leg

λ_{leg} = thermal conductivity of supporting leg material

The supporting legs may consist of several layers of different materials. In this case the conductances can be added as a result of parallel conducting paths, and the total leg conductance G_{leg} will be given by:

$$G_{leg} = \sum_i \lambda_i \frac{W_i d_i}{l_i} \quad (10)$$

The thermal properties of the 1.2 mm AMS (Austriamikrosysteme) double poly double metal CMOS process was investigated in detail by researchers in Switzerland [12]. The local technology is a derivative of the AMS process, making the values reported by the investigators applicable to this device design.

Table 2. Thermal parameters used in prototype microbolometer calculations.

Material [12]	Thermal conductivity λ W/mK	Specific heat c J/gK	Density g/cm^3
CVD Silicon dioxide	0.76	1	2.22
CVD Silicon nitride	1.65	1.2	3.10
Titanium	16.3	0.5	4.51

Using the above values and equations, the thermal and resistance parameters of each prototype bolometer could be calculated:

Table 3. Thermal and electrical properties of prototype microbolometers.

Device	G	G	H	R_0	TCR
	μW/K atmosph	μW/K vacuum	nJ/K	ohm	α
MA1	12.9	4.3	4.3	285	0.001
MB1	21.6	4.3	8.6	285	0.001
MC1	19.6	2.3	8.6	376	0.001

V. Measurements of Operational Parameters

The following method was used to determine the operational parameters of the different prototype microbolometers using only electrical means. Using a HP4155 Semiconductor Parameter Analyser the resistance of the different prototype microbolometers was measured to verify the theoretical calculations. It was found that the resistors were around 600 ohms rather than 300. This was contributed to a process problem where the titanium also etched when the sacrificial aluminium was removed.

To measure R_0, α and G the microbolometer is heated in steps in a temperature controlled oven with the temperature recorded. At a steady state temperature it is assumed that the bolometer element is at the measured oven temperature. It is now necessary to measure the resistance of the bolometer element; (a) without heating the bolometer element as a result of the measurement and (b) with electrical power input that will heat the element to within it normal operating temperature. This is done by applying a current sweep across the bolometer element and plot V as a function of I. The high power measurement is only taken after a delay of a second after applying the current in order to make sure the temperature has stabilized. The following measurement currents were used:

Table 4. Measurement currents

		Current (micro-amps)	Estimated Temperature Increase (K)
Vacuum	Low Power	10	0.01
	High power	100	60
Atmospheric	Low Power	20	0.01
	High power	1000	60

By using the low power measurements a series of resistance values, at different temperatures, are obtained. It can be assumed that the microbolometer element's temperature is equal to the oven temperature as almost no temperature change is introduced by the low measurement current. Using these low power measurements the following resistance and TCR values were calculated using equation (5):

Table 5. Measured electrical properties of prototype microbolometers

Device	Base Temperature T_0 K	Resistance at T_0 R_0 ohm	TCR α
MA1	297	468	0.00101
MB1	297	519	0.00123
MC1	297	660	0.00112

Using these measured resistance and TCR values the microbolometer element is now used as a Joule heater, as well as a measuring probe at the high power values. At each point the input Joule heat is calculated using the measured I and V values and the microbolometer element's temperature is calculated using the resistance based on the measured I and V values and the measured T_0, R_0 and TCR obtained from the low power measurements.

It can be assumed that the microbolometer element's temperature is at the calculated temperature and the environment or heat sink at the measured oven temperature. From the difference between these two temperatures and the input power the thermal system's conductance can be calculated using equation (3) at each microbolometer element temperature.

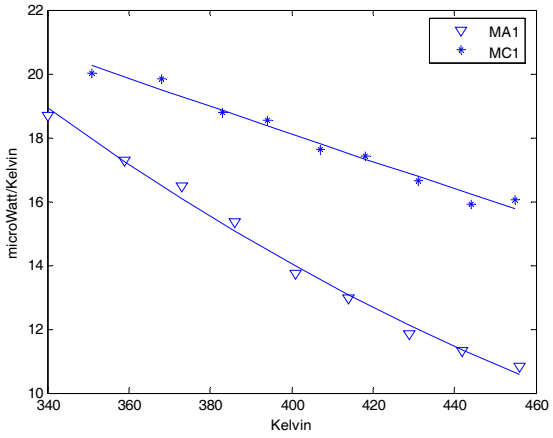


Fig 2. Measured thermal conductance of prototype microbolometers at atmospheric pressure.

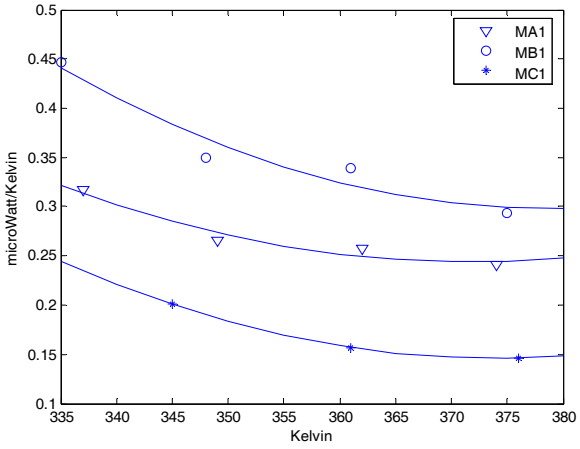


Fig 3. Measured thermal conductance of prototype microbolometer under a vacuum

At atmospheric pressure Fig 2 indicate that the measured conductance is in line with the calculated values, where MC1 has a higher conductance because of its larger area. The large almost linear dependence of the conductance with temperature must be noted. Contrary to other bolometer publications [7], [15]; the temperature coefficient is negative. Contrasting Fig 2 with Fig 3 the position of MA1 and MC1 has inverted because now the conductance trough the legs dominate and MC1 has longer legs. The conductance values under a vacuum are lower than the calculated values, again as result of the aluminium etching.

Measuring the microbolometer thermal capacitance can not easily be done with static electrical measurements. An AC electrical measurement method is used.

When AC power is applied to the microbolometer resistor the Joule temperature change and the associated resistance change of the microbolometer follows the AC power applied at low frequencies. At high frequencies this resistance change is filtered out by the thermal capacitance of the microbolometer.

Measurements are done by applying a high power AC signal to the microbolometer. AC signals of 1 milli-amps on top of a DC bias of 500 micro-amps are used. A high power input AC signal is used as the filtering effect is dependant on the Joule heating of the microbolometer element. Measurements were done with a HP 3585A Spectrum Analyzer. The following current-mirror circuit was used:

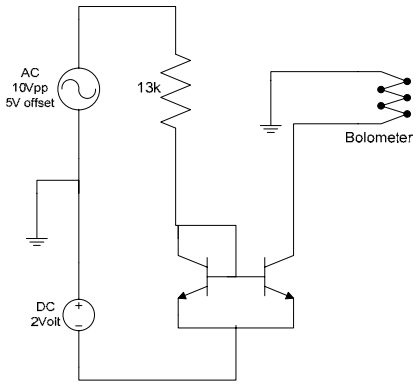


Fig 4. Test circuit used to measure the microbolometer's frequency spectrum with an AC stimulus.

The results of this experimental measurement for the prototype microbolometers at atmospheric pressure are given in Fig 7. The voltage values have been normalized to the same values at low frequency obtained in the simulation generated in the following section.

VI. Simulation Models for Microbolometers

By using computer simulation tools such as PSPICE complex systems can be modeled. As indicated by both [13] and [14] a thermal Bolometer consists of two interlink systems: the thermal system and the electrical system. The following is a summary of a combined model of these two systems:

1a. The microbolometer thermal system is defined by equation (2). Let the absorbed power vary in accordance with the complex time factor $e^{j\omega t}$. The power absorb is present as heat in the bolometer element. The bolometer element has a specific heat capacity. The power absorbed

will be stored in the bolometer heat capacity as well as flow out as heat via thermal conductance. In a steady state the element's heat capacity is filled but for a dynamic operation the input power will be absorbed or discarded from the element's heat capacity until a steady state is again obtained. Similar to electric circuits the following can be an equivalent system for the bolometer element's thermal system:

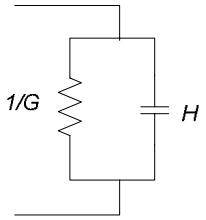


Fig 5. Electrical equivalent circuit for a bolometer's thermal system.

$$G(\omega) = G + i\omega H \tag{11}$$

The following thermal values will be presented in the model by the corresponding electrical terms:

Table 6. Electrical equivalent parameters for simulation a thermal system.

Thermal			Electrical		
Capacitance	Joule/Kevin	*H*	Capacitance	Amps/Volt	*C*
Conductance	Watt/Kelvin	*G*	Conductance	1/Ohm	*1/R*
Temperature	Kelvin	*T*	Potential	Volts	*V*
Power	Watt	*W*	Current	Amps	*I*

1b. As outlined above the microbolometer is only a resistor seen from the electrical side and it behaves as a complex electrical element with a frequency response. An equivalent electrical circuit was developed by [1]. The thermal system and electrical system that define the bolometer element is interlinked in that a change in input power will change the bolometer element's temperature (the thermal system) which then changes the bolometer element's electrical characteristics (resistance or impedance). As outlined above the change in thermal conductance is a complex variable and with this thermal-electrical interdependence the bolometer's element electrical impedance will be a complex variable as well. The following diagram defines the bolometer element's electrical impedance for a metallic bolometer element where the thermal capacitance manifests itself as an electrical capacitance in parallel with a resistor in the electrical system:

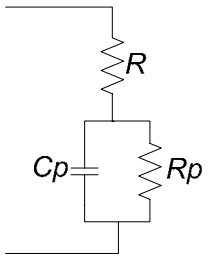


Fig 6. Electrical equivalent circuit for a bolometer's thermal feedback.

$$Z(\omega) = R + \frac{\dfrac{R_P}{j\omega C_P}}{R_P + \dfrac{1}{j\omega C_P}} \tag{12}$$

with

R = Bolometer static resistance at temperature T

ω = Angular frequency of dynamic input power (IR radiation or electrical bias)

R_P = Impedance parameter given by

$$\frac{1}{R_P} = \frac{1}{2R}\left(\frac{G}{\alpha P_X} - 1\right) \tag{13}$$

α = Bolometer thermal coefficient of resistance

G = Bolometer element thermal conductance

P_X = Input electrical power (V*I)

Impedance parameter $C_P = \dfrac{1}{2R}\dfrac{H}{\alpha P_X}$ (14)

The above model can be implemented using a simulation system such as PSPICE by simulating the temperature variables *R* and *G* using a PSPICE analog-behavior-modeling where the variable is presented by a variable current source define by temperature of the microbolometer element's temperature in the thermal model. This complication results in a loss of the simplicity normally associated with a SPICE modeling.

2. It was found to be more effective to build a MATLAB model based on the following two fundamental equations, again with no infrared power input considered.

The thermal power equation from (1) above:

$$H\frac{dT(t)}{dt} + G\left[T(t) - T_a\right] = P(t) \tag{1}$$

The electrical power input equation from (3) above:

$$P(t) = \left[I_0 + i_m \cos\omega t\right]^2 \bullet R_0\left[1 + \alpha\left(T(t) - T_0\right)\right] \tag{15}$$

With

I_0 = the DC bias current

i_m = the AC bias current

Using the measured and calculated values for the prototype microbolometers manufactured the following is the simulated frequency response of the microbolometer seen from the electrical side:

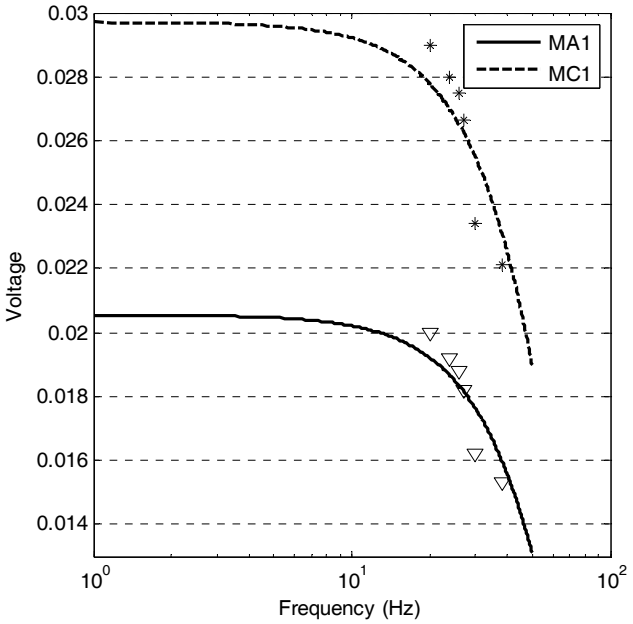


Fig 7. Simulated and measured frequency response of the prototype microbolometers at atmospheric pressure

The simulated frequency response and the measured values for the prototype microbolometer is a reasonable fit. The theoretical thermal capacitance as per Table 3 has been used in the simulation, confirming the thermal capacitance value in the manufactured units.

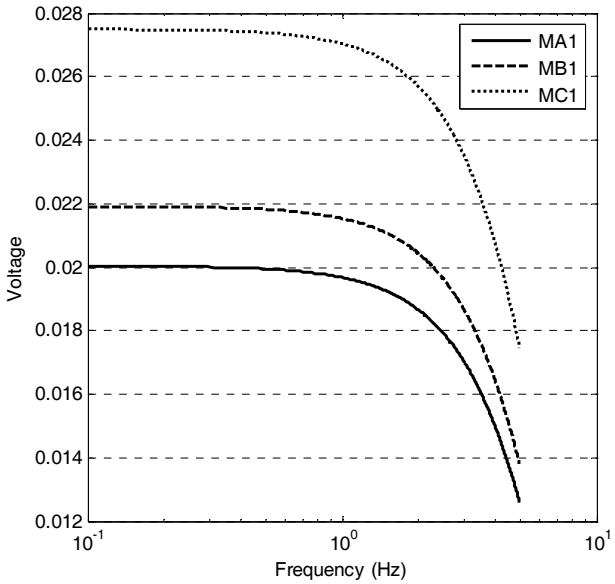


Fig 8. Simulated frequency response of the prototype microbolometer under a vacuum

Measurement for the prototype microbolometer under a vacuum could not be done at frequencies of less than 10Hz.

VII. CONCLUSION

By investigating the fundamental system parameters of bolometers, viewed from an electrical point, it was possible to develop simulation models that provide insight into the operation of a bolometer and could be verified by measurements of prototype microbolometers manufactured.

REFERENCES

[1] R. Clark Jones "*The General Theory of Bolometer Performance"* Journal of the Optical Scociety of America Volume 43 Number, 1 January 1953

[2] John C. Mather *"Bolometer Noise: nonequilibrium theory"* Applied Optics volume 21 number 6, 15 March 1982

[3] John C. Mather *"Electrical self-calibration of nonideal bolometers"* Applied Optics Volume 23 number 18, 15 September 1984

[4] M. J. Griffin and W. S. Holland *"The influence of background power on the performance of an ideal bolometer"* International Journal of Infrared and Millimeter Waves, Volume 9 number 10, 1988

[5] R.V. Sudiwala, M.J. Griffin and A.L. Woodcraft *"Thermal Modelling and Characterisation of Semiconductor Bolometers"* International Journal of Infrared and Millimeter Waves, Volume 23 number 4, April 2002

[6] A. Tanaka et al, "*Infrared focal plane array incorporating silicon IC process compatible bolometer*", IEEE Transactions on Electron Devices, Vol. ED-43, No. 11, pp.1844-1850, November 1996.

[7] Mang Ou-Yang, Chin-Shown Sheen and Jin-Shown Shie "*Parameter Extraction of Resistive Thermal Microsensors by AC Electrical Method*" IEEE April 1998

[8 M. von Arx, O.Paul and H. Baltes, "*Process-dependent thin-film thermal conductivities for thermal CMOS MEMS*", Journal of Microelectromechanical systems, Vol. 9, No. 1, pp. 136-145, March 2000.

[9] X. He, G. Karunasiri, T.Mei, W.J. Zeng, P. Neuzil and U. Sridhar, "*Performance of microbolometer focal plane arrays under varying pressure*", IEEE Electron Device Letters, Vol. 21, No. 5, pp. 233-235, May 2000.

[10] P. Eriksson, J.Y. Andersson and G. Stemme, "*Thermal characterization of surface-micromachined silicon nitride membranes for thermal infrared detectors*", Journal of Microelectromechanical systems, Vol. 6, No. 1, pp. 55-61, March 1997.

[11] J-S. Shie, Y-M.Chen, M. Ou-Yang and B.C.S. Chou, "*Characterization and modeling of metal-film microbolometer*", Journal of Microelectromechanical systems, Vol. 5, No. 4, pp.298-306, December 1996.

[12] M. von Arx, O.Paul and H. Baltes, "*Process-dependent thin-film thermal conductivities for thermal CMOS MEMS*", Journal of Microelectromechanical systems, Vol. 9, No. 1, pp. 136-145, March 2000.

[13] Holger Vogt " *SPICE modeling of resistive, diode and pyroelectric bolometer cells*" Infrared Technology and Applications Proc of SPIE Vol 6206, 2006

[14] Hollis H Jones, Shahid Aslam and Brook Lakew "*Bolometer Simulation using SPICE*" http://ssed.gsfc.nasa.gov/code693/tdw03/proceedings/docs/session_3/Jones.pdf 19 June 2003

[15] Qinglin Song, Zheng Cui, Shanhong Xia, Zhaofeng Chen, Jiangang Zhang *"Measurement of SiNx thin film thermal property with suspended membrane structure"* Sensors and Actuators A 112 (2004) 122-126, 16 December 2003

A CMOS Based Multiplexer Design of a 16 × 16 Read-Out Array for Infrared Microbolometer Sensors

Pieter Johannes Visser, *Student Member, IEEE*, Johan Schoeman, *Member, IEEE* and Monuko du Plessis, *Senior member, IEEE*
Department of Electrical, Electronic and Computer Engineering
University of Pretoria, South Africa

***Abstract*—Infrared imaging detectors require highly temperature sensative sensors placed in a focal plane array. Due to the nature of these sensors, they can not simply by biased continuously, as this causes unwanted self heating that distorts the desired measurement. This is typically overcome by the use of pulse based biasing systems to reduce the time during which a sensor can measure incoming radiation, but this in turn requires a careful timing analysis and synchronisation in the design. This paper presents the design of a fully-integrated CMOS time-division analogue multiplexer circuit serving both as a biasing and a read-out circuit for a 16 × 16 array of uncooled microbolometers. The contribution and the emphasis of the design is to focus on an architecture aimed towards eventual integration specifically to minimise self heating by ensuring that each sensor is pulse-biased and measured for the shortest possible time. The initial design is completed for fabrication in a standard 0.35-μm CMOS process with minimal external components, but changes may still be proposed depending on sensor characteristics and parameter variation.**

I. Introduction

The interest in uncooled infrared imaging arrays have increased steadilty since the nineties due to the decrease in price of these devices. This has sparked commercial interest in many forms, including night vision for automotive applications [1] and infrared imaging based person detection [2]. Due to the complexity of the overall system, focal plane array (FPA) sensor units are typically divided into various subsystems. Hsieh *et al* did this by seperating the FPA into two parts, the detector array consisting of the sensor elements and the read-out circuitry [3]. The latter functions to bias the sensor elements and to measure the signal generated due to the incoming stimulus. Further processing on these values include signal conditioning in the form of amplification, analogue to digital conversion and output processing [4]. It is often also desired (esp. high vacuum scenarios) to reduce the number of wires connected to the FPA, and multiplexing offers a possible solution [5].

One of the largest problems faced with infrared/temperature sensitive sensors lies with the biasing of these elements. As these elements behave with a resistive nature, biasing will result in power dissipation within the element, resulting in self heating that causes an undesired component in the output signal. In an effort to combat self heating, the sensor elements are pulse biased with a time division principle so that only the sensors currently being addressed by the read-out circuitry are active [6].

The remainder of this paper will include conceptual design presented in *Section* II, detailing the choice for the method of biasing (*Section* II-A) and an architectual overview of the sensor access multiplexer circuit (*Section* II-B). *Section* III details the implemented circuit designs, followed by simulation results of the multiplexer circuits as presented in *Section* IV. The final section (*Section* V) draws some conclusions and after which some references are presented.

II. Conceptual design

The conceptual design covers two very important aspects, namely how the sensors should be biased and how the output of each sensor should be accessed and presented to the processing circuitry. The following subsections present various design methodologies that may be followed to address this.

A. Sensor biasing

It will not be possible to measure any value from the sensor elements without adequate biasing. Two different approaches that may be considered is either a constant voltage biasing network or a constant current biasing network. Constant voltage biasing will require a resistor in series with the sensor element that will result in a variable voltage across the sensor element that is dependent on the incident infrared radiation, since the resistance of the sensor element will change [7]. Constant current biasing requires a constant current source that drives the sensor element. Again, a variable voltage will appear over the element caused by the change in resistance that is dependent on the incident infrared radiation [7]. The latter approach was selected for this design.

Biasing, however, causes a current to flow through a resistive element, resulting in power dissipation and eventually an increase in temperature in the device. This undesired effect of biasing needs to be minimised when the selfheating of the element is larger than the desired component of the measured signal. A typical approach is the implementation of a pulsed biasing source.

A number of alternative methods where considered to bias the sensor elements. The first method is to use a biasing source for each sensor, resulting in an unacceptably large component

count. A second possible method entails connecting all the elements in a column in series and then driving each column with a source. This reduces the 256 current sources to only 16. A third method proposes a modification to the second method, whereby each column is driven with the same current source at different times (*Fig.* 1). The single constant current source can be switched through all the columns by means of an analogue multiplexer. This reduced compenent count significantly and has the added advantage that sensors are biased for a shorted time, thereby reducing selfheating.

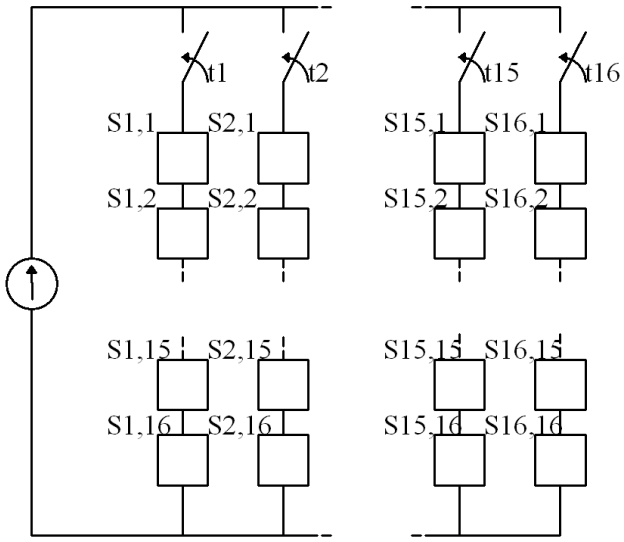


Fig. 1. Concept design of the pulsed biasing circuit

B. Sensor accessing architecture

It is necessary to measure the voltage values across each of the biased sensors and to store these values for further processing [8]. This may be accomplished with analogue multiplexers that are switched between the sensor elements to read the voltage values. Based on the specific choice of biasing, only two possible switching methods were considered.

The first method is to read the values of the sixteen biased sensors sequentially down each of the biased columns. This method was discarded as all the sensors in a column are continously biased. This will cause a sensor that is read later in a column to yield a higher output value due to longer biasing and larger selfheating. The second method considered and implemented solves this problem by reading all the output values of a single column at the same time and then storing these values with sample-and-hold (S/H) circuits. This approach thus requires a very short biasing time, seeing as the elements only need to be biased during the sampling stage, effectively decreasing the selfheating in each element again. All processing may be done on the output of the S/H circuits during the hold time. Reading of the values was performed with differential analogue multiplexers due to the serial connection between the sensor elements in a column. A conceptual diagram of the sensor elements and various levels of multiplexers (biasing and sensor accessing) is illustrated in *Fig.* 2. The highlighted multiplexer group represents the main switching circuit and S/H circuits that allows for sensor access, the top multiplexer represents the final switching circuit and the bottom multiplexer represents the switched biasing circuit.

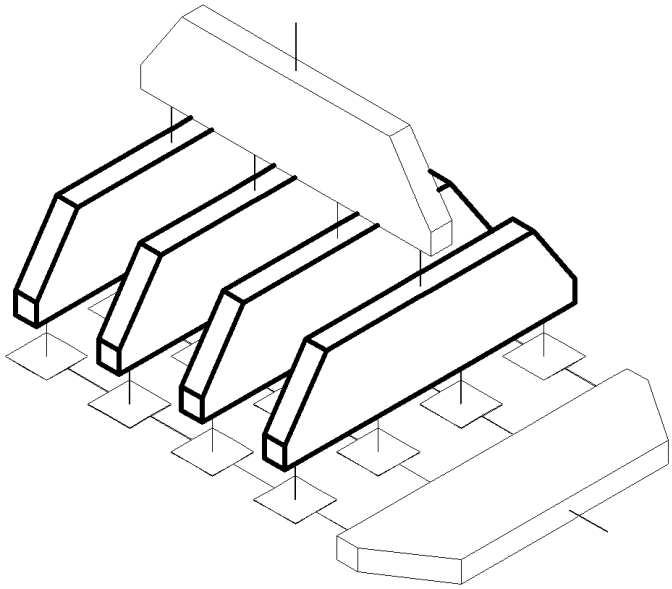


Fig. 2. A simplified conceptual architecture of the multiplexers for a 4×4 sensor.

III. Circuit design

A. One-to-sixteen (1:16) multiplexer

This device is required to switch one input signal to multiple outputs. The 1:16 multiplexer was implemented by using sixteen CMOS switches as illustrated in *Fig.* 3. The inputs of the switches are wired together to comply with the proposed architectures of *Fig.* 1 and *Fig.* 2. The outputs of the switches form the 16 outputs of the multiplexer and each output has a control input that is driven with one of sixteen address lines.

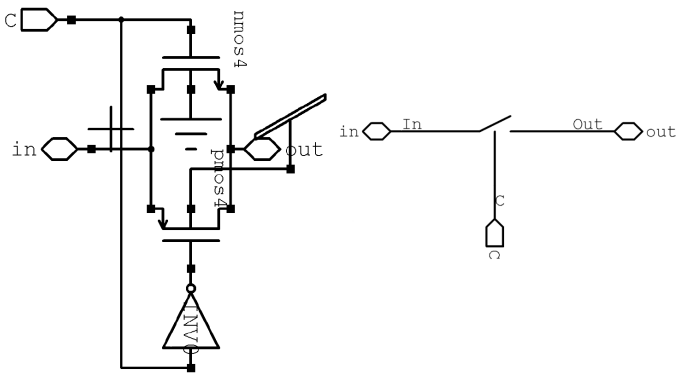


Fig. 3. Circuit diagram of the CMOS switch and symbol used

The input will appear on any output that has its address line asserted, even potentially multiple outputs. This may not necessarily be desired and the design overcomes this potential problem by using a four to sixteen decoder to drive the addressing of the multiplexer. This also reduces the 16 bit addressing value to a 4 bit binary sequence. The capabilities of

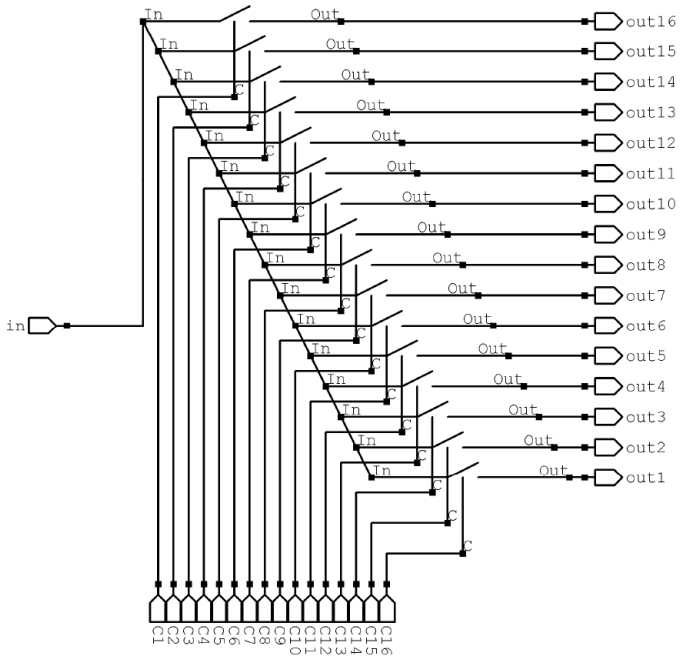


Fig. 4. Circuit diagram and symbol of the 1:16 multiplexer

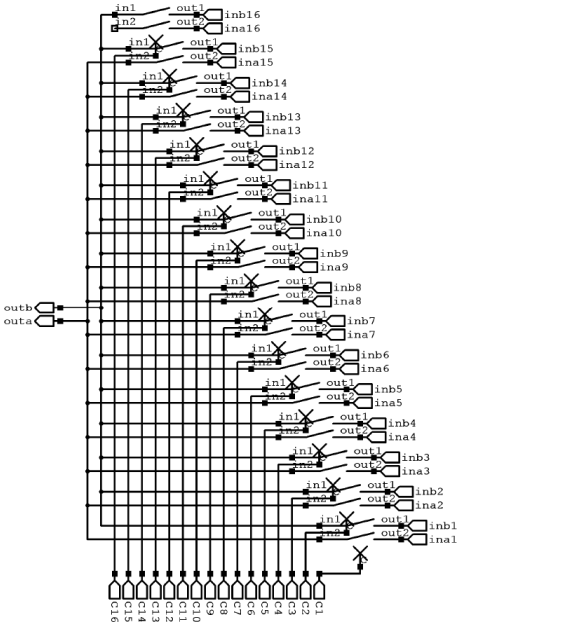


Fig. 5. Circuit diagram and symbol of the extended sixteen to one differential multiplexer

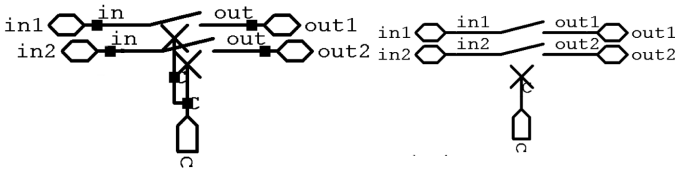


Fig. 6. Circuit diagram and symbol of the extended differential CMOS switch

the multiplexer can be further extended by driving the decoder with a four bit counter. This enables the multiplexer to switch the input from one output to the next on each clock cycle and keep it there for the duration of the clock cycle.

B. Extended sixteen to one (16:1) differential multiplexer

The extended 16:1 differential multiplexer (*Fig.* 5) is constructed and functions similarly to the 1:16 multiplexer. The first notable exception is that a combination of two extended CMOS switches (*Fig.* 7) are used to create an extended differential CMOS switch (*Fig.* 6). Secondly, the direction of forward signal flow is reversed. Furthermore, the capabilities of the extended 16:1 differential multiplexer are also extended by means of a decoder and counter.

C. S/H circuit

The sample and hold circuits are implemented by means of an extended differential CMOS switch that drives two capacitors. In this configuration the voltage across the capacitors will track the input voltage when the switches are closed and voltages will remain stored on the capacitors when the switches are opened.

D. Main switching circuit

The goal of the main switching circuit is to switch the voltages that are generated across a single column of sensor elements to the rest of the circuitry. Since the sensor elements are biased one column at a time, the voltages across the sensors need to be switched differentially. This is due to the fact that not all voltages across the series elements are measured relative to ground. The differential nature of the voltage being switched from the sensors thus lead to the requirement for differential multiplexers (as presented earlier).

The switching functionality as discussed in *Section* II-B may be realised by 16 16:1 differential multiplexers. Each of the sixteen multiplexers are connected to a row of the sensor elements (refer to *Fig.* 2), where each of the inputs connects differentially across a single element in a different row. Notice that only one column of input elements will ever be biased at a specific time, thus only the values of this column will be held in the S/H circuits. All these multiplexers are addressed with the same control signal that is synchronised to the biasing multiplexer, thereby ensuring that all the multiplexers are addressing the same (and biased) column.

Initially the address lines of each of the sixteen multiplexers were driven synchronously by its own decoder and counter combination. However, deeper insight in the design revealed that the outputs of all the decoders are the same. Efficiency could, therefore, be improved by driving the multiplexers on a common address bus with a signal generated by a single counter and decoder combination. This resulted in a device reduction of more than 60%.

E. Biasing switching circuit

The goal of the biasing switching circuit is to ensure that the column of sensors currently addressed in the readout circuit is biased for as short a time as possible. Synchronisation is ensured when the multiplexer that switches the biasing is driven by the same signal addressing the main switching

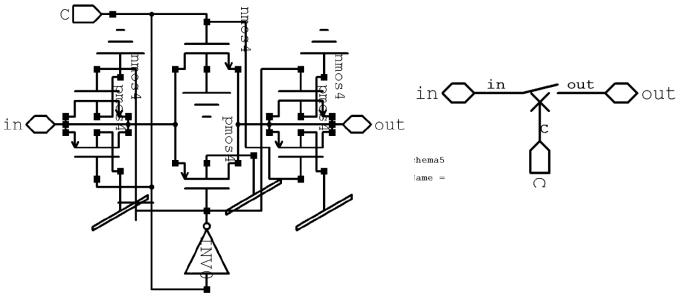


Fig. 7. Circuit diagram and symvol of the extended CMOS switch

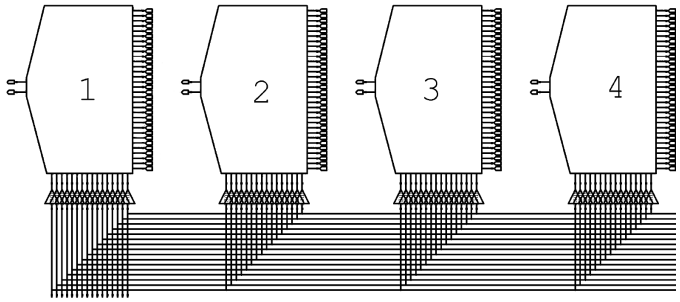


Fig. 8. Circuit diagram of part of the main switching circuit that shows how the multiplexers were driven with the address lines in parallel

circuit.

The biasing switching circuit was implemented with a single 1:16 multiplexer. Each of the sixteen outputs of the multiplexer drives one of the columns of sensor elements and the input of the multiplexer is driven by a constant current source (*Fig.* 2). For the constant current source a single MOSFET biased in the saturation region was used, this is possible because the effects of loading doesn't have a significant effect on the biasing of the transistor.

However, this leads to a longer biasing time of an entire clock cycle, which negates the inclusion of the S/H amplifier. This problem was overcome by driving the multiplexer with its own decoder running from the same counter as the main switching circuit. However, this decoder has an additional input that reset the decoder outputs to zero when asserted, thus effectively removing the biasing from the sensor elements. The result is a circuit that can be used to switch a constant current bias source through all the sensor elements one column at a time and where the biasing can effectively be turned off after the sensor outputs have been stored in and the S/H amplifier is performing a hold function.

F. Final switching circuit

The final switching circuit is required to convert the parallel sensor outputs of each column, as stored in the capacitors of the S/H amplifer, to a serial output format that may be used for further processing. This was accomplished with a single 16:1 differential multiplexer, where each input was differentially connected to one of the 16 main switching circuit multiplexer outputs, or rather the S/H amplifier storing the value of those multiplexers (refer again to *Fig.* 2).

Since this multiplexer has to switch the voltage values at the outputs of the main switching circuit to the rest of the circuit one at a time, the final switching circuit has to be driven with a clock that runs at a frequency sixteen times higher than that of the main multiplexer circuit. Also, due to the fairly large values of the capacitors in the S/H cuircuits, the extended differential multiplexer was selected due to the reduced input and output parasitic capacitances of the device.

IV. Results

The following section provides simulation results of the proposed system. Transistor models for a standard 0.35-μm CMOS process were used.

A. Main switching circuit

The use of a constant current source to drive the series connected bolometers will result in a voltage generated across each element. This was emulated by using an array of DC voltage sources, where two voltage sources were used to simulate the differential output voltage of each sensor. Furthermore, these voltage values were scaled to facilitate unique output voltages for easy identification.This was achieved by setting the voltage values representing sensor one to 1mV and 1.5mV, the voltage values of the sources representing the second sensor to 2mV and 2.5mV, and incrementally increasing the values to the last sensor at 256mV and 256.5mV. Scaling the test voltages in this fashion allows testing of the correct differential circuit operation over a wide voltage range. It is clear from the results that the switching occurs in the correct sequence by the staircase like output that is generated (*Fig.* 9).

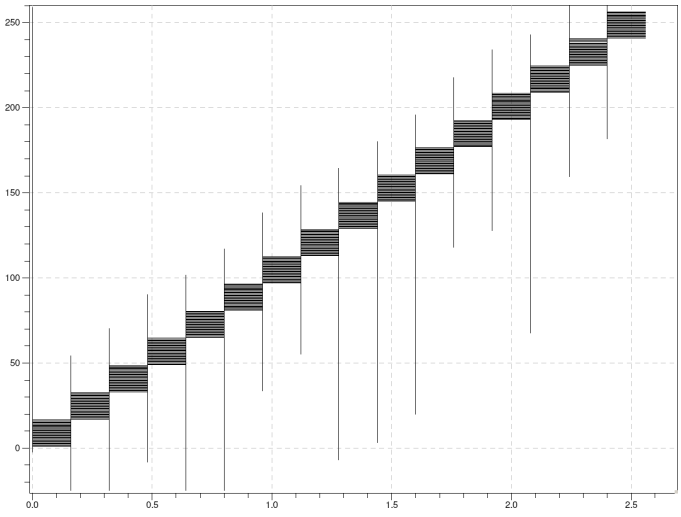


Fig. 9. Differential outputs of the sixteen multiplexers

B. Biasing switching circuit

The biasing switching circuit was tested by using it to drive different resistor values (17kΩ for column one, 16kΩ for column two, etc.). This results in a unique output voltage to be switched to the circuit output, which indicates whether the correct array column is biased at the correct time. Furthermore, the multiplexer is also turned off during the hold times of the S/H circuitry (*Fig.* 12 (a)), resulting in a considerable reduction in the biasing duty cycle (*Fig.* 10).

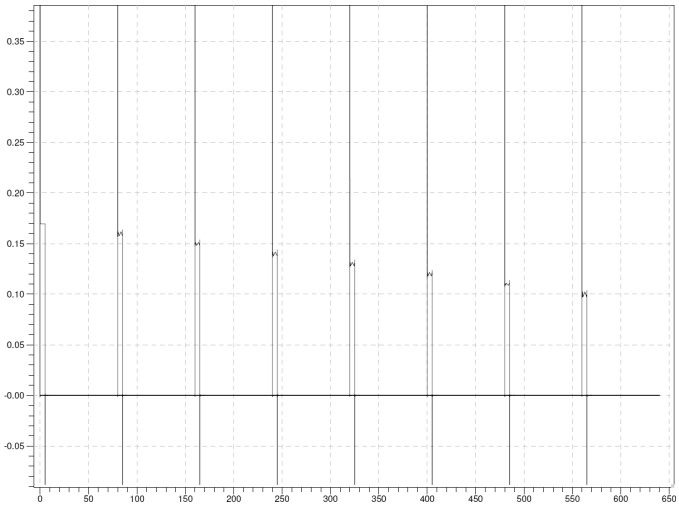


Fig. 10. Output voltages over the resistors when the transistor current source was used to supply the bias and the set0 input is used to shorten the pulse widths

C. Final switching circuit

The multiplexer of the final switching circuit is connected to the outputs of the sample and hold circuits and switches the outputs of multiplexers one through sixteen to the output in the time that the voltages are stored in the S/H circuit. The output of the final switching circuit will be a fast analog series stream of the various test voltages applied to the sensors (*Fig.* 11). The fast, series nature of the analogue levels is clear when compared with the slow, parallel nature of the analogue levels in *Fig.* 9. As with any mixed signals application some glitches are introduced in the analogue signals. The glitches can be reduced by using an amplifier with a low pass response.

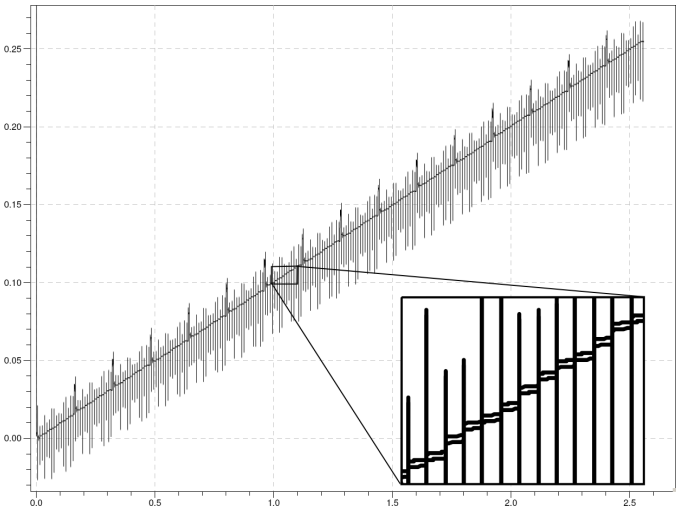


Fig. 11. Output of the final switching circuit

D. Timing and synchronisation

In *Fig.* 12 the clock signals used to synchronise the multiplexers are shown. *Fig.* 12 (a) shows the control signal that is used to turn the biasing multiplexer on/off. An asserted signal turns the biasing multiplexer off (refer to *Section* IV-B). *Fig.* 12 (b) and (c) shows the clock signals for the main switching circuit and the final switching circuit respectively. The biasing multiplexer is driven from the same clock as the main switching multiplexer and the S/H circuits are controlled by the inverse of *Fig.* 12 (a).

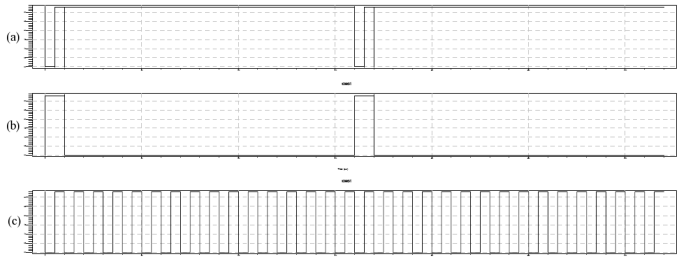


Fig. 12. Timing diagrams for the switching circuitry

V. Conclusion

It was established that it is possible to perform a read-out operation from an array of sensors by using analogue multiplexers operating in both voltage and current modes for switching signals and biasing. It was shown that it is possible to reduce the biasing duty cycle significantly, reducing the biasing time of a single sensor to less than 0.2% of the read-out cycle for a entire frame.

VI. Acknowledgement

The authors thank the Advanced Manufacturing Technology Strategy (AMTS) of the Department of Science and Technology, South Africa for the financial support of the research.

References

[1] M. Hirota, Y. Nakajima, M. Saito, M. Uchiyama, 'Thermoelectric infrared imaging sensors for automotive applications, *Proc SPIE 2004*, Vol. 5359, pp. 111-125, San Jose, USA.

[2] J.-E. Kllhammer, H. Pettersson, D. Eriksson, S. Junique, S. Savage, C. Vieider, J.Y. Andersson, F. Niklaus, G. Stemme, 'Fulfilling the pedestrian protection directive using a long-wavelength infrared camera designed to meet the performance and cost targets, *Proc. SPIE 2006*, Vol.6198, pp.74-84, Strasbourg, France.

[3] C. Hsieh, C. Wu, F. Jih, T. Sun, 'Focal-Plane-arrays and CMOS Readout Techniques of Infrared Imaging Systems', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 4.

[4] R. A. Wood, 'High-Performance Infrared Thermal Imaging with Monolithic Silicon Focal Planes Operating at Room Temperature', *IEDM '93*, Technical Digest.

[5] A. Benoit, S.J.C. Yates, E. Grmion, P. Camus, T. Durand, C. Hoffmann, S. Leclercq, A. Monfardini, A. Cavanna, B. Etienne, U. Gennser, Y. Jin, 'A Time Domain Multiplexer for Large Arrays of High Impedance Low Temperature Bolometers ', *Journal of Low Temperature Physics*, Volume 151, Numbers 3-4, May 2008, pp. 940-945.

[6] X. Gu, G. Karunasiri, J. Yu, G. Chen, U. Sridhar, W.J. Zeng, 'On-chip compensation of self-heating effects in microbolometer infrared detector arrays', *Sensors and Actuators*, 1998.

[7] T. Hsin, C. Wu, Y. Chin, P. Chen, F. Chi, J. Luo, C. Chiang, Y. Cherng, 'A New CMOS Readout Circuit for Uncooled Bolometric Infrared Focal Plane Arrays', *ISCAS*, 2000.

[8] D. Jakonis, C. Svensson, C. Jansson, 'Readout architectures for uncooled IR detector arrays', *Sensors and Actuators*, 2000.

Nanostructured photovoltaic cells: Using AFM and QCM to produce accurate film thicknesses

Authors: A.F. Botha, U. Büttner, and W. J. Perold, Senior member, IEEE.
Department of Electrical and Electronic Engineering, Stellenbosch University, Stellenbosch 7600, South Africa.

***Abstract*—In thermal evaporation techniques QCM (Quartz Crystal Measurement) systems are used to measure the film thickness and deposition rates of materials, for example the deposition of silver for wire-bonding pads. In organic solar cells, film thickness plays an integral roll in the working of the device. This is due to diffusion lengths in the nanometer range of the organic semiconductors. To accurately measure film thicknesses with a QCM sensor, three values are important in the setup: Tooling factor (relative position of the QCM sensor to substrate), Z-factor (ratio of acoustic impedance of deposition material to that of quartz) and material density. An iterative method with AFM (Atomic Force Microscope) thickness measurements is shown here for the calibration of these values. The organic semiconductors CuPc (Copper Phthalocyanine) and C_{60} (Buckminister Fullerenes or Bucky Balls) are used as sublimation materials to construct working photovoltaic cells.**

***Index Terms*—Acoustic impedance, AFM, photovoltaic, QCM.**

I. Introduction

THE use of organic semiconductors for solar cell applications was realised in the early 1970's with materials including anthracene, tetracene and chlorophyll [1]. Initially organic materials did not receive much attention due to very low quantum efficiencies of Schottky Junction type structures. Sam-Shajing Sun [2] reports on four generations of organic photovoltaic cells some thirty years later with a low photoelectric power conversion efficiency of less than 6%, from cells made of purely organic and polymeric materials. The first generation is the Schottky Junction (semiconductor layer sandwiched between electrodes). The second generation is the bi-layer heterojunction which uses multiple materials to create a p-n interface between electrodes. The third is a bulk heterojunction which is a mechanical mixture of materials in a single layer to increase the absorption length, without increasing the thickness of the materials. The difficulty with the mixture device is that it is difficult to control the current path, with one type of semiconductor (n or p) sometimes connecting to both electrodes or not connecting at all. The fourth generation [2], being a nanostructured configuration, is still only a theoretical device, which can be best understood as a honeycomb structure with p-type and n-type pillars, to increase the absorption thickness without increasing the thickness of the donor/acceptor (D/A) interface. This honeycomb structure will improve the absorption length a photon has to travel to micrometers, while the distance to the closest D/A interface can still remain in the nanometer range (diffusion length range). A collection of well known organic semiconductors (not including the metal-Phthalocyanine range) are given in Table I.

TABLE I
MATERIALS THAT CAN BE USED AS ELECTRON DONOR/ACCEPTOR SEMICONDUCTORS [2].

P-Type Semiconductor	N-Type Semiconductor
PBZT(D)	Perylenes
PPV	C60
MEH-PPV	PCBM
MDMO-PPV	CN-PPV
RO-PPV-10	SF-PPV
PTh	BBL

To fabricate organic bi-layer heterojunction solar cells, CuPc and C_{60} (sublimation grade) were obtained from Luminescence Technology Corp. ITO (Indium tin oxide) coated glass with an ITO thickness of $1200\,\text{Å}$ to $1600\,\text{Å}$ and a sheet resistance of $9 - 15\,\Omega/\text{sq}$ were used as substrates, also obtained from Luminescence Technology Corp.

II. Experimental Details

Since the exciton diffusion lengths of CuPc and C_{60} are $10\,\text{nm}$ and $40\,\text{nm}$ [3] respectively, a QCM sensor was used to ensure that the correct film thicknesses could be deposited. The correct QCM setup values first needed to be determined for each material. This was done with an iterative method of material deposition and thickness measurement with an AFM. A pattern with square blocks was created on polished MgO samples using AZ5214 as a positive photoresist, as can be seen in Figure 1. One drop of the AZ 5214 was dripped onto each of the MgO substrates, which were cleaned with acetone in an ultrasonic bath, and spun for $30\,\text{sec}$ at $6000\,\text{rpm}$ to ensure even, thin and smooth films. The substrates were then prebaked for $1\,\text{min}$ on a hot plate at $95\,°\text{C}$. The samples were then exposed for $25\,\text{sec}$ to ultraviolet light under a mask to form the pattern in Figure 1. The samples were then placed in a developer for $1\,\text{min}$ and blow dried with Nitrogen.

The initial values for the QCM sensor were set as: Tooling - $110\,\%$ as the sensor is mounted higher than the sample holder. Density - $1620\,\text{kg/m}^3$ for CuPc [4] and $1720\,\text{kg/m}^3$ for C_{60} [5], which will stay fixed for the duration of the iteration. The initial Z-factors were calculated as 15 for CuPc and 3.45 for C_{60} from

$$z = \sqrt{\frac{D_q \times U_q}{D_m \times U_m}} \quad (1)$$

where D_q and D_m respectively represents the densities of quarts ($2643\,\text{kg/m}^3$) and the deposited material. U_q and U_m

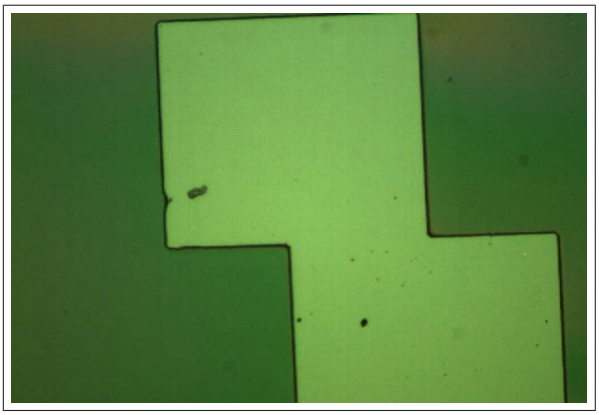


Fig. 1. Photoresist blocks on a polished MgO sample.

respectively represents the shear modulus of quartz (32 GPa) and the deposited material. For the C_{60} Equation 1 could be applied directly since the values for the density ($1720\,\mathrm{kg/m^3}$ [5]) and shear modulus ($4.14\,\mathrm{GPa}$ [6]) could be obtained. For CuPc the density is known, but the shear modulus was approximated to be $3.4\,\mathrm{GPa}$ with from

$$G = \frac{E}{2(1+v)} \quad (2)$$

In Equation 2 the Poisson's ratio (v) was set to that of Cu (0.36). The Young's modulus, E, was taken as $9.29\,\mathrm{GPa}$ [7].

All the physical vapour deposition of the sublimation materials were started and maintained at a vacuum pressure of more than $2.6 \times 10^{-5}\,\mathrm{mbar}$. The materials where then deposited onto the MgO samples until the QCM sensor read $50\,\mathrm{nm}$. The materials were evaporated by resistively heating a tungsten boat. The temperature was slowly increased over a 20 min period before the shutter was opened, in order to allow any impurities (moisture) to dissipate. The optimal temperatures for a steady evaporation rate was found to be 580 °C for the CuPc and 650 °C for C_{60}. After deposition the samples were then dipped into an acetone bath for approximately $1\,\mathrm{min}$ until all the photoresist lifted off. The samples, left with only small squares of deposited material on, were then taken to an AFM. The tip of the AFM was positioned over the edge of each square, as can be seen in Figure 2. The area around the step is then scanned to obtain figures such as Figure 3. A cross section is taken with the software to obtain Figure 4 from which the thickness of the deposited material can be read. Several scans should be taken of various steps on the sample, and measurements averaged to ensure that an even film thickness was deposited and the correct reading is obtained, removing faults resulting from poor manufacturing (lift-off).

The actual measured thicknesses for each material can then be fed into Equations 3 and 4 to obtain the new parameters for the QCM sensor. In Equations 3 and 4 T_1 is the thickness indicated by the QCM sensor and T_2 is the thickness measured by the AFM. This proses is then repeated until the values from the QCM sensor and the AFM are equal.

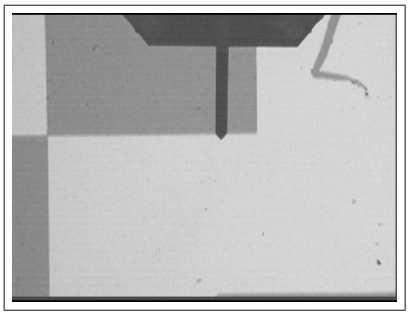


Fig. 2. Tip of the AFM positioned over the step formed by the deposited material on MgO sample.

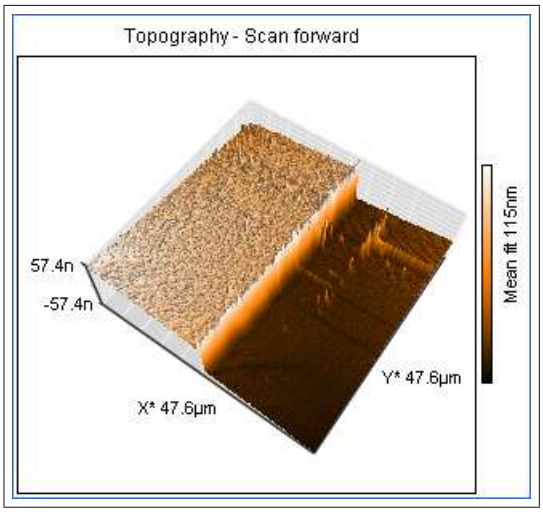


Fig. 3. A 3-dimensional view of a scan.

$$Tooling_{new} = Tooling_{old} \times \frac{T_2}{T_1} \quad (3)$$

$$Z_{new} = Z_{old} \times \frac{T_1}{T_2} \quad (4)$$

III. Results and Discussion

After completing several iterations, the Z-factors for CuPc and C_{60} were obtained to be 0.437 and 3.45 respectively. The

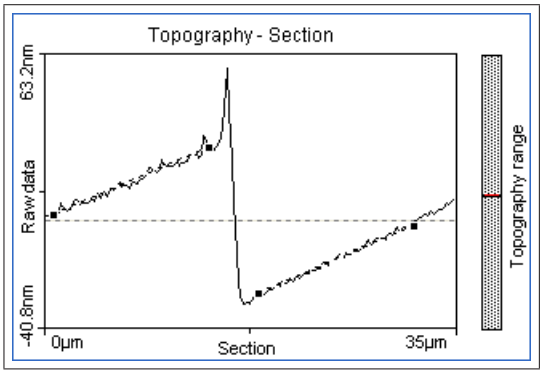


Fig. 4. Software cross section to determine material thickness.

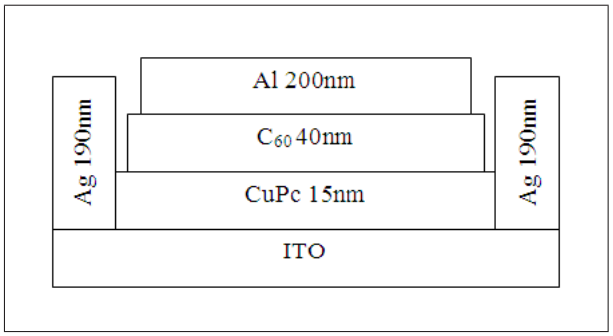


Fig. 5. Device layout of a organic photovoltaic cell.

respective acoustic impedances are then calculated from

$$Z = \frac{Z_q}{Z_m} \tag{5}$$

where Z_q is the acoustic impedance of quartz (8.83) and Z_m of the material used. The acoustic impedances of CuPc and C_{60} were calculated as 20.21 and 2.56 respectively.

Once the QCM system was set up, a working solar cell was constructed. The device layout is shown in Figure 5, and has an open circuit voltage (V_{oc}) of $400\,\mathrm{mV}$ in direct sunlight, which is on par with other reports [3].

To improve the efficiency of organic photovoltaic cells, it is important to implement the fourth generation organic photovoltaic device [2], since it is the most promising device so far to improve photon absorption without losses coupled with exceeding exciton diffusion lengths. Before such a nano-structured device will be realised, the current techniques, material properties and device shortcomings should be well understood.

IV. Acknowledgements

The financial assistance of the South African National Energy Research Institute towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to SANERI.

References

[1] H. J. Hovel, *Solar Cells*, ser. Semiconductors and Semimetals, R. K. Willardson and Albert C. Beer, Ed. Academic Press, 1975, vol. Volume 2.

[2] Sam-Shajing Sun, *Advances in Solar Energy: An annual Review of research and Development*, D. Y. Goswami, Ed. American Solar Energy Society, 2007, vol. 17, Chapter 3.

[3] Sung Woo Hur et al, "Organic photovoltaic effects using CuPc and C_{60} depending on layer thickness," *Synthetic Metals*, vol. 154, pp. 49–52, 2005.

[4] "Copper phthalocyanine," IRPTC Data Profile.

[5] M. S. Dresselhaus, P. C. Ekluna, and G. Dresselhaus, *Carbon Materials for Advanced Technologies*, T. D. Burchell, Ed. Elsevier, 1999.

[6] O. F. Yagafarov, E. L. Gromnitskaya, A. G. Lyapin, V. V. Brazhkin, and M. V. Kondrin, "Ultrasonic study of monomeric fullerite C_{60} under pressure," *Journal of Physics: Conference Series*, vol. 121, 2008.

[7] Kanari Moriyasu et al, "Nanoindentation test for evaluating intermolecular elastic and plastic characteeristics of copper phthalocyanine thin films," *Japanese Journal of Applied Physics*, vol. 44, pp. 8249–8255, 2005.

Optimisation strategies for W Band GaAs Gunn diodes

Robert R. van Zyl[1], Smita Francis[2], Willem J. Perold[3], Reinhardt Botha[4]

[1] Cape Peninsula University of Technology, [2] Polytechnic of Namibia, [3] University of Stellenbosch,
[4] Nelson Mandela Metropolitan University

***Abstract*—This paper proposes an effective approach to the optimisation of Gunn diodes operating at mm-wave frequencies, specifically, 94GHz. A novel multi-domain diode structure, with multiple hot launchers is investigated. Rigorous Monte Carlo particle simulations suggest that 160mW may be obtained, which far exceeds the current state-of-the-art diode performance of about 80mW in this frequency range.**

***Index Terms*—Gunn diode, Monte Carlo, mm-Wave, multi-domain, optimisation.**

I. Introduction

Gunn diodes have been around since the 1960's, following the almost chance discovery of the Gunn effect by JB (Ian) Gunn on 19 February 1962 [1]. Through the years both the output power and efficiency of these diode oscillators have steadily increased through judicious research and innovation. Current state-of-the-art GaAs diodes yield about 80mW at 2% efficiency in this frequency range [2]. A goal of 100mW at 94GHz has therefore been set to achieve high power GaAs Gunn diode oscillators. (This is in agreement with a similar venture under the auspices of the South African Council for Scientific and Industrial Research (CSIR) [3].) This has been accomplished by incorporating a two-domain diode, with a hot-electron launcher preceding each domain. This diode structure has been proposed by the author in previous publications [4], [5], but now for the first time it has been subjected to rigorous Monte Carlo particle simulations including a thermal model of the diode.

The Monte Carlo method was chosen because it closely resembles a physical experiment. It is therefore a very powerful tool, but does come at a high computational cost. A novel parallel implementation of the algorithm has been developed by the author [6-8] and implemented on a cluster of personal computers. A discussion on the Monte Carlo method, and the incorporated thermal model, falls outside the scope of this work. Numerous references in the literature can be consulted on this subject, e.g. [9-11].

There are several avenues of optimisation at the disposal of the researcher. These will be discussed in subsequent sections. The general approach to the optimisation problem is to firstly establish an optimised single-domain diode benchmark, and then to extend this to a two-domain diode structure for increased output power.

The paper follows the optimisation procedure. As background to the problem, an overview of the Gunn effect is given in the next section. This is followed by a description of the optimisation avenues in Section III, after which the simulation results will be presented. The paper concludes with a discussion of the results in Section V.

II. The Gunn effect in the strict sense

A. The transferred electron mechanism and Negative Differential Resistance

When no bias is applied to a semiconductor, almost all the electrons occupy the central Γ-valley. If the sample is biassed, the electrons are accelerated by the applied electric field and may gain sufficient energy to transfer to the satellite valleys. This phenomenon is verified by Monte Carlo simulations and illustrated by the graphs in Figure 1 [12]. (These simulations are based on a simple two valley energy band model with band gap $\Delta = 0.36$ eV.)

It is clear from the graphs in Figure 1 that the mean electron energy increases for increasing field bias. This results in an ever increasing number of electrons gaining enough energy (Δ) to bridge the gap between the central and satellite valleys, and transfer from the lower Γ-valley to the upper satellites.

The electrons that have transferred from the Γ-valley to the satellite valleys will immediately move slower due to the increase in their effective mass. The average drift velocity of the electrons, and consequently the current, will therefore decrease with an increase in the applied electric field. This manifests as a region of negative differential resistance (NDR) for applied fields exceeding about 0.4 MVm^{-1}.

B. The formation of Gunn domains

The question of exactly how the NDR phenomenon in GaAs results in Gunn oscillations can now be answered with the aid of Figure 2 [12].

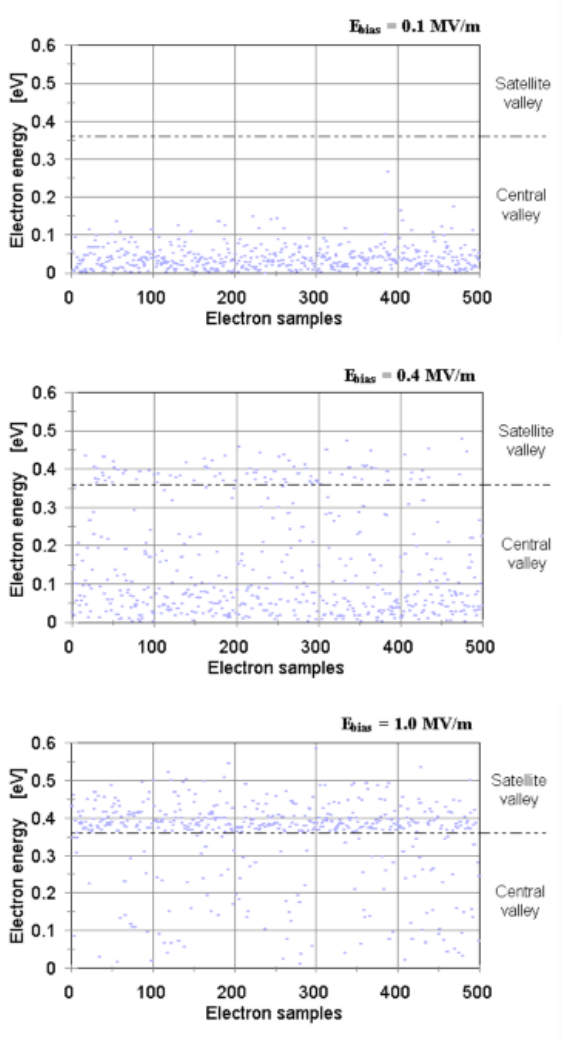


Fig. 1. Valley occupation of electrons in bulk GaAs for three applied electrical fields E_{bias}. The mean energy of the ensemble of electrons increases with stronger applied fields. Significant population of the satellite valleys takes place at fields exceeding 0.4MVm^{-1}.

A sample of uniformly doped n-type GaAs of length L is biased with a constant voltage source V_0. The initial electrical field is therefore constant and its magnitude given by $\xi_0=V_0/L$. From the bottom graph in Figure 2 it is clear that the electrons flow from cathode to anode with constant velocity v_3.

It is now assumed that a small local perturbation in the net charge arises at $t = t_0$, indicated by the solid curve in Figure 2. This non-uniformity can, for example, be the result of local thermal drift of electrons. The resulting electrical field distribution is also shown (solid curve). The electrons at point A, experiencing an electric field ξ_{L1}, will now travel to the anode with velocity v_4. The electrons at point B are subjected to an electrical field ξ_{H1}. They will therefore drift towards the anode with velocity v_2 which is smaller than v_4. Consequently, a pile-up of electrons will occur between points A and B, increasing the net negative charge in that region. The region immediately to the right of point B will become progressively more depleted of electrons, due to their higher drift velocity towards the anode than those at point B.

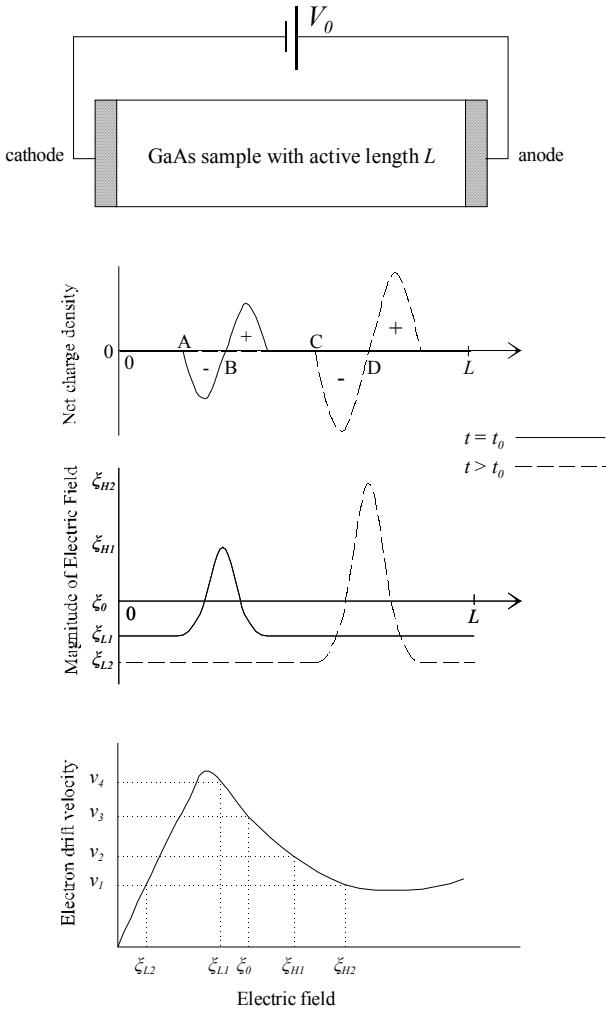


Fig. 2. An illustration of the formation of Gunn domains in a uniformaly doped GaAs sample.

The initial charge perturbation will therefore grow into a dipole domain, commonly known as a Gunn domain. Gunn domains will grow while propagating towards the anode until a stable domain has been formed. A stable Gunn domain is shown at a time instance $t > t_0$, indicated by the dashed curve. At this point in time, the domain has grown sufficiently to ensure that electrons at both points C and D move at the same velocity, v_1, as is clear from the bottom graph in Figure 2.

It is important to note that the sample had to be biased in the NDR region to produce a Gunn domain. Once a domain has formed, the electric field in the rest of the sample falls below the NDR region and will therefore inhibit the formation of a second Gunn domain.

As soon as the domain is absorbed by the anode contact region, the average electric field in the sample rises and domain formation can again take place. The successive formation and drift of Gunn domains through the sample leads to a.c. current oscillations observed at the contacts.

C. Transit-time devices

The mode of operation described above is referred to as the Gunn mode.

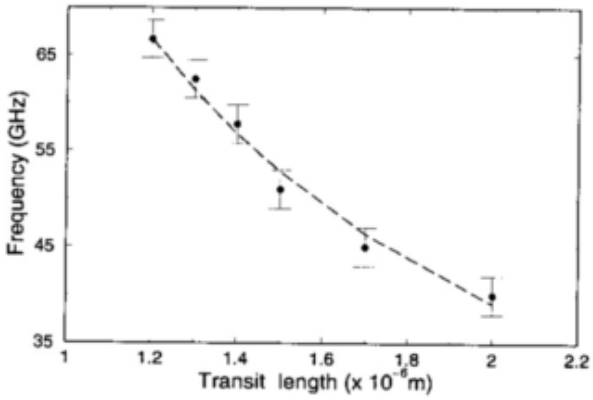


Fig. 3. Variation of transit frequency with active layer length (from [12]).

In this mode the frequency of the oscillations is determined primarily by the distance the domains have to travel before being annihilated at the anode. This distance is roughly the length of the active region, *L*, of the diode. These diodes are therefore also referred to as "transit-time" devices. The approximate relationship between the transit length of the diode and the fundamental harmonic component of the output power is given in Figure 3.

III. GUNN DIODE OPTIMSATION

Each of the avenues of optmisation will now be discussed briefly.

A. Hot-injection launcher

The effect of high-energy electron injection at the cathode has been discussed by Greenwald [13], [14]. Hot-injection entails the launching of electrons from the cathode at elevated energies, close to the threshold for significant electron transfer from the central to satellite valleys. Neylon et al [15] points out numerous advantages of using hot-electron injectors in Gunn diodes:

- In the first place it reduces the dead zone, since domain formation and nucleation are greatly enhanced close to the cathode. As already mentioned, this enhances efficiency.
- Hot-injector Gunn diodes also display much improved turn-on characteristics, compared to conventional diodes. This allows coherent oscillations around peak power over the full military specification temperature range.
- The diode's performance is much more independent of operating temperature than its conventional counterpart. This is due to the fact that hot-injected electrons have temperatures of the order 2000K, much more than the nominal operating temperature of Gunn diodes (about 450K at ambient room temperature).
- The diode's output power and position of domain nucleation can be made much less sensitive to bias variation with an appropriate launcher structure (see also [16]). The latter would imply greater frequency stability.

An appropriate launcher is proposed in [17] and [18]. The heterostructure injector consists of an undoped linearly graded AlxGa1-xAs layer, followed by a very narrow n+ doping spike. This spike serves as a non-equilibrium connector to prevent depletion, set up by the forward biased injector, from extending into the active region. The structure is illustrated in the next section.

B. Notch doping

The incorporation of a nominally undoped notch at the cathode, preceding the active layer, also reduces the dead zone. The doping notch encourages high electric fields, above the critical value for Gunn domain formation, at the cathode. An early mentioning of a doping notch is by Tully [19]. Doping notches are frequently used in conjunction with hot-electron launchers, e.g [20].

C. Grading of the active region doping profile

For optimum device efficiency, a uniform resistivity profile across the active layer is required [21], [22]. This has erroneously been interpreted by Hasegawa [23] to imply a uniform doping profile, and many has followed suit. However, it was subsequently shown that increasing the doping density from cathode to anode enhances diode efficiency and output power [24], [25], [26], [27]. This can only be explained by taking the varying temperature profile across the active layer into account (Hasegawa assumed a constant temperature profile) [22]. In the work presented here, the temperature profile is determined throughout the device, and not assumed constant.

D. Multi-domain operation

Intuitively, the output power of a Gunn diode can be enhanced by merely increasing the cross-sectional area of the diode. This will increase the current for a given terminal voltage. This is, however, unrealistic because the admittance that the external circuit must present to the diode for optimal matching purposes, also decreases. Circuit losses will increase as a result [28]. This, together with an increase in operating temperature due to enhanced d.c. power dissipation, will place a limit on the minimum cross-sectional area of the diode.

A solution to this problem has been the combination of series-connected diodes to improve output power, as originally proposed Thim [29] in 1968. Thim stacked several wafers on top of each other. In subsequent work by Slater and Harrison [30], a horizontal diode was implemented where multi-domain nucleation centres were forced by literally scratching the surface of the active layer. Subsequently, in 1979, Talwar [31] reported on a dual-diode 73GHz Gunn oscillator that produced double the output power of a single diode. (Two discrete diodes were used.)

The basis of these approaches is that, with series connected Gunn diodes, the impedance of the grouped diodes (whether discrete or integrated into a single device) increases. Therefore, the area of each individual diode may be increased

while the group will still present a favourable impedance level to the external circuitry.

Tsay *et al* [32] incorporated this approach into a single diode with multiple domain nucleation and quenching regions. Their diode essentially consists of multiple active layers, separated by highly doped regions where the electric field is quenched. They showed that, with an N-domain diode, an increase of N^2 in output power could be obtained in favourable frequency ranges. Teo and Dunn [33] verified this approach with Monte Carlo simulations. They have found no obvious limit to the number of domains that can successfully be incorporated, which is unrealistic if thermal effects are properly accounted for.

E. Multi-domain operation with multiple hot-electron launchers

The two main optimisation routes mentioned in the foregoing discussions, namely multi-domain operation and the incorporation of hot-electron launchers, was proposed by the author [34], [35]. These multi-domain Gunn diodes with multiple hot-electron launchers will, intuitively, benefit from the advantages associated with both hot-injection and multi-domain operation. This premise forms the basis of the work presented here. Furthermore, the incorporation of notch doping and graded doping profiles are investigated and incorporated in the overall optimisation approach.

Previous work on multi-domain diodes mentioned above ([32], [33]) ignored the very important thermal aspects of Gunn operation, by assuming unrealistic nominal operational temperatures, as well as constant temperature profiles in the active layers. Multi-domain diodes operate at highly elevated temperatures, due to the fact that the power dissipation also increases with the square of the number of domains. In this work, the temperature is determined throughout the device with high grid resolution, and not assumed constant. Thermal effects are therefore addressed accurately in the optimisation of the Gunn diode. It is worth mentioning here that this treatment points to the necessity of using diamond heat sinks for efficient multiple-domain operation.

IV. SIMULATION RESULTS

A benchmark, single domain, diode will be simulated firstly to reflect current state-of-the-art diodes. An optimized double-domain diode, based on the benchmark model, will then be simulated.

A. Single-domain benchmark diode

1) Diode structure description:

The diode structure is illustrated in Figure 4.

- Hot launcher

The Al concentration is graded from 0 to 0.3 across 50nm, followed by a 10nm wide n+ spike with doping 1×10^{24} m^{-3}.

- Active region doping profile

A doping notch, preceding the active region on the cathode side, has been implemented. It consists of a 0.2µm undoped region. From an extrapolation of Figure 3 the active region should be ≈1.6µm long to coincide with a fundamental frequency of 47GHz. The required active region length has been shortened accordingly to 1.4µm to compensate for the inclusion of the notch.

For efficient Gunn operation, the minimum required doping density is $0.71\times10^{22}m^{-3}$. Choosing the doping concentration close to the minimum will degrade the diode efficiency, as will a value that is much higher. The latter will cause a marked increase in the d.c.bias current which will lead to excessive thermal losses. This is especially the case for double-domain operation where the d.c. power dissipation increases quadratically with an increase in d.c. current. A value of $1.5\times10^{22}m^{-3}$ has been found to be optimum.

Grading of the active region is incorporated in anticipation of higher efficiencies compared to flat doping profiles. The doping concentration is increased over the last 25% of the active region (toward anode) by a factor 1.67.

- Diode diameter

A diameter of 75µm has been chosen. This yields favourable admittance levels for matching the diode to external circuitry.

- Heat sinking

A copper heat sink is used. As will be discussed later, the use of a diamond heat sink is proposed for double-domain operation.

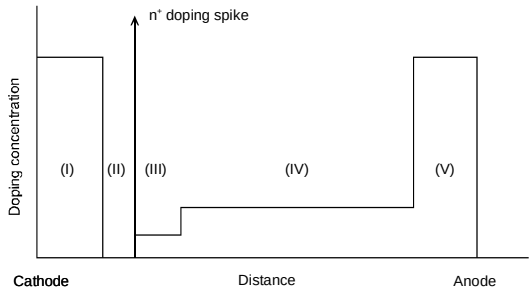


Fig. 4. Doping profile of the benchmark single-domain Gunn diode with hot-injection. Numerical values are given in the text.

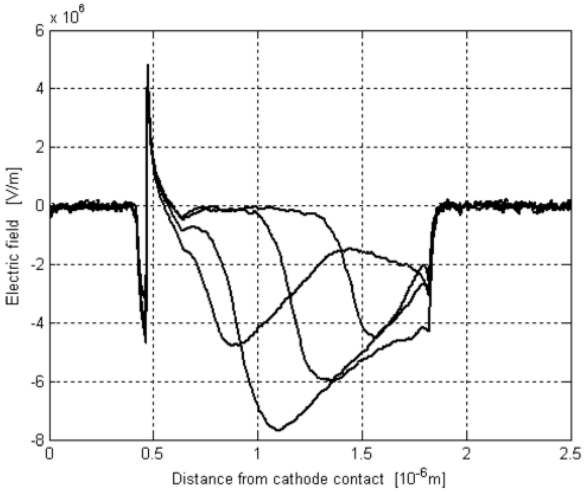


Fig. 5: Internal electric field distribution of the benchmark diode at successive time intervals.

2) Microwave performance:

Simulation results indicate a maximum output power of 60mW at 1.65% efficiency. The diode admittance at 94HGz is -0.22 + j0.33 S, which is favourable for matching purposes. The bias voltage and current are 3.5V and 0.9A respectively. These values correspond fully with current state-of-the-art single domain Gunn diodes. The crystal temperature at the cathode is 420K.

The internal electric field distributions for successive time intervals are shown in Figure 5. The absence of the dead zone at the cathode is as a result of the hot-injection launchers.

B. *Optimised double-domain diode*

1) Diode structure description:

The doping profile of a typical double-domain Gunn diode with hot launcher is shown in Figure 6. It consists of two series-coupled, hot-injection single-domain diodes. The two diodes are separated by a highly doped buffer region.

The buffer region can be thought of as a low resistance connector between the two single-domain diodes. The width of the buffer region must be chosen large enough to quench the domains that enter it from the cathode. However, choosing it excessively long, will remove the first active region further away from the heatsink. This will result in increased operating temperatures in this domain. Furthermore, the buffer regions introduce additional series (and positive) resistance which translates into lower efficiency. This mechanism is of lesser importance due to the negligible resistance of the highly doped buffer regions [32]. A minimum buffer width is proposed by Tsay *et al* [32], based on simplified domain models. An extrapolation of their results suggests a minimum of ≈0.2µm for doping densities exceeding $2\times10^{22}m^{-3}$. In the following sections, an initial buffer width of 0.5µm, with a doping density of $2\times10^{23}m^{-3}$, is assumed. It has been found that this is adequate for effective domain quenching.

Each of the single-domain diodes is identical to the single domain benchmark diode. A nominal active region doping density of $1.5\times10^{22}m^{-3}$ has been found to be optimal. The doping density is graded identical to the benchmark diode.

- Diode diameter

A diameter of 100µm has been chosen, which results in a cross-section of twice that of the benchmark diode to compensate for the doubling of the diode's active length. This yields favourable admittance levels for matching the diode to external circuitry.

- Heat sinking

A *diamond* heat sink is applied to the anode side of the diode. A copper heatsink has been found insufficient and resulted in highly elevated temperatures, which in turn negated any expected increase in output power.

2) Microwave performance:

Simulation results indicate a maximum output power of 160mW at 2% efficiency. The diode admittance at 94HGz is -0.21 + j0.27 S, which is favourable for matching purposes. The bias voltage and current are 6V and 1.37A respectively. These values correspond fully with current state-of-the-art single domain Gunn diodes.

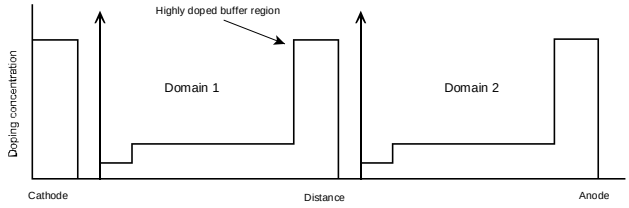


Fig. 6: Doping profile of a typical double-domain Gunn diode with hot injection. Numerical values are given in the text.

The internal electric field distributions for successive time intervals are shown in Figure 7. It is notable that the field distributions in each domain are similar, proving that the diodes act as two series coupled devices.

The active region temperature profile at optimum bias is given in Figure 8. It is clear that the operating temperatures in the two active regions are quite dissimilar. Also noted is the 5K linear temperature variation across the buffer region. This underscores the requirement to keep the buffer as short as possible.

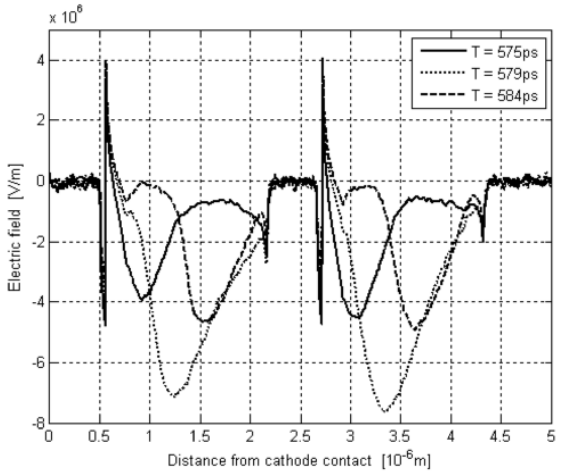


Fig. 7: Internal electric field distributions at successive time intervals for the optimised double-domain diode.

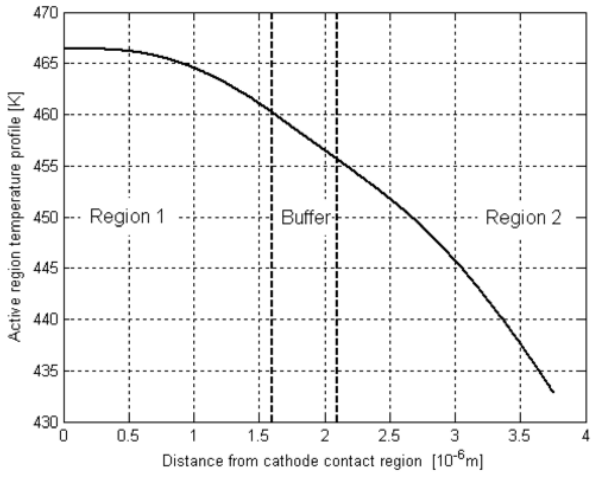


Fig. 8: Active region temperature profile for the optimised double-domain diode at optimum bias.

V. CONCLUSION

The optimization of GaAs Gunn diodes at 94GHz has been described. Each of the avenues of optimization has been discussed, followed by rigorous simulations of a benchmark single-domain diode, as well as that of the optimised double-domain diode. A novel double-domain diode structure with two hot-injection launchers has previously been proposed by the author. In this work, the proposed diode was subjected to rigorous Monte Carlo particle simulations, including thermal effects.

The benchmark diode performed in accordance with current state-of-the-art Gunn diodes, with an expected 60mW of power at 2% efficiency.

The optimised double-domain diode with two hot-injection launchers outperforms the benchmark by a factor of 2.7, producing a remarkable 160mW at similar efficiencies than that of the single diode. The expected fourfold increase in output power has not materialised, due to thermal effects that become more severe for the double-domain diode. However, the results reported here, prove that the proposed diode indeed shows much promise.

Further optimisation of the diode will inevitably revolve around addressing the thermal performance of the diode. Thermal effects limit the output power in multi-domain diodes. Measures to curtail the effects of thermal dissipation may include novel heatsinking structures, other heatsinking material, and further optimisation of the doping grading in the active region.

REFERENCES

[1] J.B. Gunn, "Microwave Oscillations of Current in Group III-V Semiconductors", Solid State Consm., vol. 1, pp. 88 - 91, Sept. 1963.

[2] E2V Technologies: "Gunn diode application notes", A1A-Gunn Diodes AN1, Issue 4, September 2002. [Online]. Available: http://e2v.com/applications/diodes.htm

[3] F.F. Mundell, "Development of a high yield fabrication process for millimeter wave Gunn diodes", M. Sc. Thesis, University of Natal, Natal, South Africa, December 1998.

[4] R.R van Zyl and W.J Perold, "Optimization of millimeter wave Gunn oscillators with the aid of the Monte Carlo simulation technique", presented at the 6th International Conference on Simulation of Devices and Technologies (ICSDT '98), Cape Town, pp. 93-96, 14-16 October 1998.

[5] R.R van Zyl, W.J Perold and R. Botha, "Multi-domain Gunn diodes with multiple hot electron launchers: a new approach to mm-wave GaAs Gunn oscillator design", presented at the IEEE Africon 1999, Cape Town, pp. 1193-1196, 1999.

[6] R.R van Zyl, W.J Perold and H. Grobler, "An efficient parallel implementation of the Monte Carlo particle simulation technique on a network of personal computers", Intl. Jnl. Numerical Modelling: Electronics Networks, Devices and Fields, John Wiley and Sons, vol. 13, pp. 369 -380, 1996.

[7] R.R. van Zyl, W.J. Perold and H. Grobler, "Parallel implementation of the Monte Carlo particle simulation technique on a network of personal computers", presented at the 6th International Conference on Simulation of Devices and Technologies (ICSDT '98), Cape Town, pp. 81-86, 2000.

[8] L.P. Wicomb and R.R van Zyl, "A budget supercomputer: implementation and application", presented at IEEE AFRICON 2004, Gaborone (Botswana), pp. 15-17, Sept. 2004.

[9] C. Moglestue, Monte Carlo simulation of semiconductor devices, Chapman & Hall, London, 1993.

[10] C. Jacoboni and P. Lugli, The Monte Carlo method for semiconductor device simulation, Springer-Verlag, New York, 1989.

[11] P. Lugli, "The Monte Carlo method for semiconductor device and process modeling", IEEE Trans. Computer-aided Design, vol. 9, no. 10, pp. 1164-1176, 1990.

[12] R.R. van Zyl and W.J. Perold, "The Gunn-diode: Fundamentals and fabrication", presented at IEEE AP/MTTS-98, University of Cape Town, Cape Town, pp. 121-124, 1998.

[13] Z. Greenwald, "The effect of a high energy electron injection cathode on the performance of the Gunn oscillator", Ph.D. dissertation, Cornell University, New York, August 1986.

[14] Z Greenwald et al, "The effect of a high energy injection on the performance of mm wave Gunn Oscillators", Solid-State Electronics, vol. 31, no. 7, pp. 1211-1214, 1988.

[15] S. Neylon et al, "State-of-the-art performance millimetre wave Gallium Arsenide Gunn diodes using ballistically hot electron injectors", IEEE MTT-S Digest, L-47, pp. 519-522, 1989.

[16] H. Spooner and N.R. Couch, "Advances in hot electron injector Gunn diodes", GEC Jnl. Research, vol. 7, no. 1, pp. 34-45, 1989.

[17] N.R Couch et al, "High-performance, graded AlGaAs injector, GaAs Gunn diodes at 94GHz", IEEE ED Lett., vol. 10, no. 7, pp. 288-290, July 1989.

[18] P.H. Beton et al, "Use of n+ spike doping regions as nonequilibrium connectors", Electronic Letters, vol, 24, no. 7, pp. 434-435, March 1988.

[19] J.W. Tully, "Monte Carlo simulation of a millimeter-wave Gunn-effect relaxation oscillator", IEEE Trans. ED., vol. ED-30, no. 6, June 1983.

[20] C. Lee and U. Ravaioli, "Monte Carlo comparison of heterojunction cathode Gunn oscillators", Electron. Lett., vol. 26, no.7, March 1990.

[21] A.R Batchelor, "Thermal design considerations for GaAs transferred-electron devices", PhD. Dissertation, University of Leeds, Leeds, 1990.

[22] A.R Batchelor, "On the design of doping concentration profiles for GaAs transferred-electron device layers", Solid-State Electronics, vol. 35, no. 5, pp. 735-741, 1992.

[23] F. Hasgawa and M. Suga, IEEE Trans. ED, vol. ED-19, pp. 26-37, 1972.

[24] R. Vaidyanathan and R.P Joshi, "Simulations for improved heterostructure Gunn oscillator based on transit region doping variation", Electronic Letters, vol. 27, no. 17, pp. 1555-1557, August 1991.

[25] A. Paolella, R.L Ross and J. Ondria, Microwave Journal, vol. 29, p. 149, 1986.

[26] J. Ondria and R.L Ross, presented at the Proc. 17th Eur. Microwaves Conf., Rome, Italy, p.673, 1987.

[27] J.M Szubert et al, "W-Band GaAs Gunn diodes with high output power", Solid-State Electronics, vol. 33, no. 8, pp. 1035-1037, 1990.

[28] H. Kroemer, Proc. IEEE, vol. 54, pp. 1980-1981, 1966.

[29] H.W Thim, "Series-connected bulk GaAs amplifiers and oscillators", Proc. IEEE, vol. 56, p. 1244, 1968.A.

[30] M. Slater and R. Harrison, “An investigation of multiple domain Gunn effect oscillators”, IEEE Trans. ED, vol. ED-23, no.6, pp. 560-567, June 1976.
[31] Talwar, “A dual-diode 73-GHz Gunn oscillator”, IEEE Trans. MTT, vol. MTT-27, no.5, pp. 510-512, May 1979.
[32] J. Tsay et al, “Multidomain Gunn diodes”, Microwave and Optical Technology Letters, vol. 3, no. 2, pp. 54-60, February 1990.
[33] Y.P. Teoh and G.M Dunn, “Monte Carlo modelling of multiple transit regions Gunn diodes”, Electronic Letters, vol. 38, no. 15, pp. 830-831, July 2002.
[34] R.R van Zyl and W.J Perold, “Optimization of millimeter wave Gunn oscillators with the aid of the Monte Carlo simulation technique”, presented at the 6th International Conference on Simulation of Devices and Technologies (ICSDT ‘98), Cape Town, pp. 93-96, 14-16 October 1998.
[35] R.R van Zyl, W.J Perold and R. Botha, “Multi-domain Gunn diodes with multiple hot electron launchers: a new approach to mm-wave GaAs Gunn oscillator design”, presented at the IEEE Africon 1999, Cape Town, pp. 1193-1196, 1999.

Overview of Superconducting Digital Electronics

J. Taylor*, C. J. Fourie†
*†University of Stellenbosch, Department of Electrical and Electronic Engineering, South Africa
*jpt@sun.ac.za, †coenrad@sun.ac.za

***Abstract*—The purpose of this paper is to provide an overview of the field of superconducting electronics.**

An overview of the phenomenon of superconductivity and some physical properties of superconductors are presented first. This is followed by a discussion of the Josephson Junction and the relevant equations. Application examples are given next, with a brief explanation of each.

RSFQ circuits are discussed in more detail. Example circuits of a DRO, JTL and AND gate are presented. Simulation results for a small RSFQ circuit is shown and a typical physical layout of an RSFQ logic gate is given.

The paper ends with a brief discussion on the integration of superconductive digital electronics into commercial and industrial electronic systems. Some popular room temperature interfaces are presented as well as practical difficulties in the implementation of a functional system.

I. Introduction

Superconductivity was discovered in 1913 by Heike Kamerlingh-Onnes, director of the Low-Temperature Laboratory at the University of Leiden [1]. He investigated the electrical resistance of mercury at low temperatures when he discovered a 'resistance step' at 4.2 K.

Ever since then the field of superconductivity and its applications have been growing. New materials have been found that exhibit a transition point at higher temperatures. An active component – the Josephson junction – has been discovered and mathematically modelled. Very fast digital electronics have also been developed.

The field of superconductivity is still in a developmental stage. This paper provides an overview of the basic physical properties of superconducting materials as well as a few applications. Practical difficulties are also discussed and plausible solutions are presented.

II. Superconductivity

A. Overview

The most obvious and striking property of superconductors is the absence of resistance. When a superconducting material is cooled below its critical temperature T_c, electrical resistance becomes negligible.

This property does not, however, hold for any condition. There is also a critical current density J_c as well as magnetic flux density B_c. When either of these increase beyond their respective critical values, the superconducting state is lost [2].

A familiar physical phenomenon is the different phases of water at certain temperatures. Superconductivity can be explained in a similar way. Above the transition temperature T_c, the electrons (or charge-carriers) exhibit the properties of fermions, i.e. no two electrons occupy the same quantum state and therefore behave as separate particles [1].

When the temperature drops below T_c, the electrons condense to Cooper pairs [1]. This can be seen as a different physical phase (such as steam condensing to water). Cooper pairs exhibit the properties of bosons, i.e. the Cooper pairs can all occupy the same quantum state. They can therefore be modelled by a macroscopic coherent matter-wave [1].

The analogy between superconductivity and the phases of water can be taken further. When the pressure is decreased, water boils at a lower temperature. Similarly, when the magnetic field is increased, Cooper pairs break to form electrons at a lower temperature [1]. Each superconducting material has a different 'phase diagram'.

Superconductors also exhibit perfect diamagnetism. This is described by the Meissner effect [2]. A superconductor, when cooled below T_c, will expel any applied magnetic field. It even expels a field that was present before the superconducting state was reached. It the latter case, the energy released when the electrons condense to Cooper pairs provide the energy required to expel the magnetic field.

A magnetic field can be trapped indefinitely within a superconducting ring. The magnetic flux of the trapped field is quantised to multiples of a fluxon Φ_0 [1].

B. Classical Mathematical Model

Fritz London developed a model from a classical electromagnetic point of view. London I and London II are given by (1) and (2) respectively [2].

$$\mathbf{E} = \frac{\partial}{\partial t}(\Lambda \mathbf{J}) \tag{1}$$

$$\mathbf{B} = \nabla \times (\Lambda \mathbf{J}) \tag{2}$$

London I describes perfect conduction. It states that the electric field strength $\mathbf{E}$ is proportional to the time-derivative of the current density. The proportionality constant Λ is called the isotropic London coefficient and given by (3) [2]. Here m^* is the mass of a Cooper pair, n^* the density of Cooper pairs and q^* the charge of a Cooper pair. It is also a function of temperature, since n^* is temperature dependant. London II describes the Meissner effect and is used mostly by implication only.

$$\Lambda = \frac{m^*}{n * (q^*)^2} \tag{3}$$

These equations also imply that current flows only on the surface of a superconductor, similar to the skin-effect [3] in high-frequency circuits. The London penetration depth λ is constant for all frequencies and is given by (4) [2].

$$\lambda = \sqrt{\frac{\Lambda}{\mu_0}} \tag{4}$$

C. Quantum Mechanical Model

After further experimental investigation it was found that the above model is not accurate for all cases [1]. It does not, for instance, describe the quantisation of flux. Nor does it explain the formation of Cooper pairs. A microscopic quantum-mechanical model was derived by Bardeen, Cooper and Schrieffer – the BCS theory [1].

The macroscopic quantum model (MQM) expands the BCS theory by stating that all the Cooper pairs within a superconducting medium can be described by a single wave-function Ψ, given by (5) [2]. After application of the Schrödinger equation, the super current is given by (6) [2].

$$\Psi(\mathbf{r},t) = \left(\sqrt{n^*(\mathbf{r},t)}\right) e^{i\theta(\mathbf{r},t)} \tag{5}$$

$$\mathbf{J}_s = -\frac{1}{\Lambda}\left(\mathbf{A}(\mathbf{r},t) - \frac{\hbar}{q^*}\nabla\theta(\mathbf{r},t)\right) \tag{6}$$

From this new equation for the super current, the flux inside a superconducting ring may be derived to yield (7) [2]. A single flux quantum Φ_0 is given by (8) [2]. This was verified experimentally in 1961 by two groups: Doll and Näbauer in Munich as well as Deaver and Fairbank in Stanford [1].

$$\int_s \mathbf{B}\cdot d\mathbf{s} = n\Phi_0 \quad , n \in \mathbb{Z} \tag{7}$$

$$\Phi_0 = \frac{h}{2e} = 2.07{\cdot}10^{-15}\ \mathrm{T{\cdot}m^2} \tag{8}$$

III. Josephson Junctions

A. Overview

The MQM also predicts a phenomenon called 'tunnelling'. It is possible for a current to pass through a thin insulating layer without any resistance. The governing equations for a small Josephson tunnel junction when the current i is smaller that the critical current I_c are given by (9) and (10) [1]. The voltage v is proportional to the time-derivative of the gauge-invariant phase difference φ.

$$i = I_c \sin\varphi \tag{9}$$

$$v = \frac{\Phi_0}{2\pi}\frac{d\varphi}{dt} \tag{10}$$

This gauge-invariant phase difference is the phase difference in the macroscopic matter-wave across the Josephson junction. It results from the diminished Cooper pair density in the thin insulating region. Fig. 1 shows the effect graphically. In the figure, φ is represented by γ and n_s is the Cooper pair density.

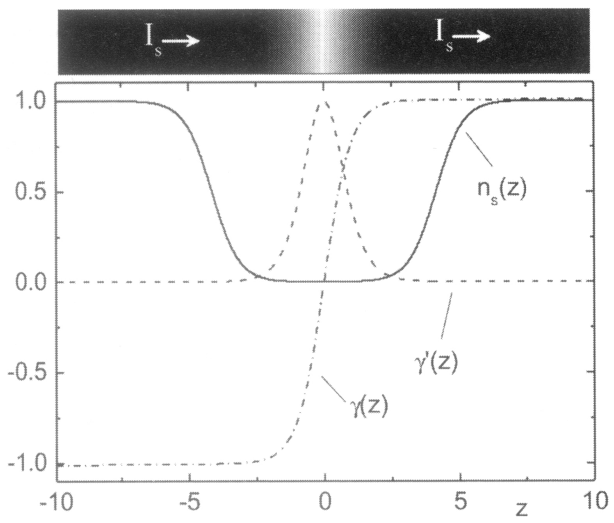


Fig. 1: Phase difference across a Josephson junction (after, [1])

When the junction is placed within a superconducting ring, (10) may by expanded to yield (11), where L is the inductance of the ring.

$$v = \frac{\Phi_0}{2\pi}\frac{d\varphi}{dt} = L\frac{di}{dt} \tag{11}$$

B. RCSJ Model

When the current through the Josephson junction goes above I_c, the junction becomes resistive [1]. An underdamped junction also shows hysteresis. These effects can be modelled by the resistor-capacitor shunted junction (RCSJ) model [1].

Cooper pairs are not the only charge carriers that can tunnel through a junction. It is also possible for a Cooper pair to break up into electron quasi-particles [1], which then tunnel through the junction and recombine on the other side. When these electrons tunnel through a junction, the process is resistive with a resistance R_n [1]. Due to the physical structure of the junction, it is also capacitive.

Fig. 2 shows two analogous systems [1]. Fig. 2(a) shows the RCSJ model for a Josephson junction. It consists of a resistor, ideal tunnel junction and capacitor. The current i can be expressed in terms of φ and the voltage U by (12) [1]. This can be simplified to (13) and (14) [1].

$$i = I_c \sin\varphi + \frac{U}{R_n} + C\dot{U} \tag{12}$$

$$i = \sin\varphi + \dot{\varphi} + \beta_c\ddot{\varphi} \tag{13}$$

$$\beta_c = \frac{2\pi I_c R_n^2 C}{\Phi_0} \tag{14}$$

The constant β_c is the damping factor, or McCumber-Stewart parameter [4]. For $\beta_c > 1$, the junction is underdamped. For $\beta_c < 1$, the junction is overdamped.

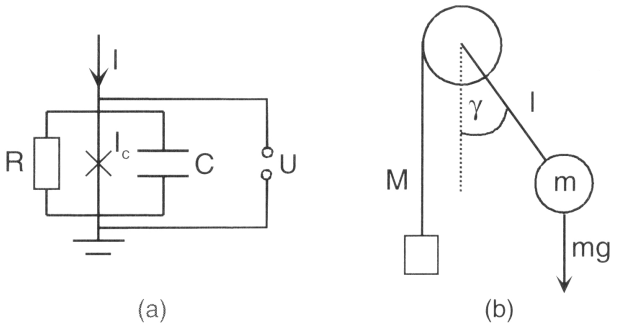


Fig. 2: RCSJ Model for a Josephson junction (after, [1])

Fig. 2(b) shows an insightful equivalent mechanical model for the RCSJ model. The mass M represents the current whereas the time-derivative of the angle $\dot{\gamma}$ represents the voltage [1].

It must be noted that the RCSJ model is a simplification. The current has small higher-order harmonics that are ignored. Also, the resistance is only linear for small voltages [1].

IV. Digital Circuits

A. Overview

There are two popular logic families in digital superconducting electronics. These are complimentary output switching logic (COSL) [5] and rapid single flux quantum (RSFQ) [4]. RSFQ is the more popular of the two because of the simplicity of its layout and speed. A 20 GHz RSFQ microprocessor has been implemented [1].

With such fast digital circuits, many applications are possible. Signal processing is the most widely used – especially in the telecommunications industry using software defined radio (SDR) [6]. The advantages of SDR are well documented and include superior filters, noise-immunity, modulation techniques and bandwidth efficiency.

The signal output of RSFQ is very small (in the order of 100 μV). It is therefore challenging to interface with room-temperature electronics. COSL is a voltage state logic family, making interfacing much easier. It is possible to convert from RSFQ to COSL in order to make use of the best of both worlds [7].

A superconducting version of a field programmable gate array (FPGA) has also been designed and is called the superconducting programmable gate array (SPGA) [8], [9].

B. ADC

Very high bandwidth analogue to digital converters (ADCs) have been developed [10]. A bandwidth of 20 GHz was obtainable in 1999 [6], but this could be much higher in future technologies. This enables SDR systems to be implemented at microwave frequencies [6].

Another advantage of superconducting ADCs is that they are quantum-mechanically quantised and therefore exactly linear. This results in an effective reduction in noise.

C. RSFQ

RSFQ circuitry rely on the pulses (voltage and current) produced when a Josephson junction switches between the superconductive and resistive states. The process of producing these pulses is analogous to forcing the pendulum to swing through 360°. Mass M is pulled briefly, representing the incoming current pulse (see Fig. 4). This in turn swings the pendulum through 360°. The angular velocity represents the voltage pulse over the junction. This voltage pulse then induces the outgoing current pulse. If the incoming current pulse is not large enough, the junction remains superconductive and does not induce the outgoing pulse.

If a pulse is present within a certain clock period, it represents a '1'. If the pulse is absent during the entire clock period, it represents a '0'.

The clock period synchronisation is achieved by means of a destructive read-out (DRO) latch [4]. When a pulse enters the DRO, its energy is stored in the form of a circulating current. When a 'reset' signal (also an RSFQ pulse) is applied, the stored energy is released and it continues to the next logic gate. When the reset signal is applied to a DRO without stored energy, the output remains zero. Fig. 3 shows the schematic of a DRO. Parasitic inductances are not included.

Junction B_2 receives the incoming current pulse, which then induces approximately one fluxon in L_2, as described above. It is not exactly one fluxon because the current through B_2 is different after the process than before, resulting in a different φ.

If this circulating current is too small to switch junction B_4, the energy is stored. When the circulating current is large enough to switch junction B_4, the pulse is propagated, as is the case with the Josephson transmission-line (JTL) [4]. A 250 μA JTL is shown in Fig. 4. As a rule-of-thumb, the pulse is trapped when $L \approx \frac{\Phi_0}{I_c}$ and propagated when $L \approx \frac{1}{2}\frac{\Phi_0}{I_c}$.

Fig. 5 and 6 show simulation results for the JTL and DRO respectively. The signal label is on zero and the scale is indicated above the label.

A basic OR-gate can be built by using a pulse combiner followed by a DRO [11]. The presence of a pulse on any of the inputs results in the DRO being set. The 'reset' signal then reads the result of the OR function and resets the latch.

Similarly, by changing the critical currents of the junctions in the pulse combiner, an AND-gate can be constructed

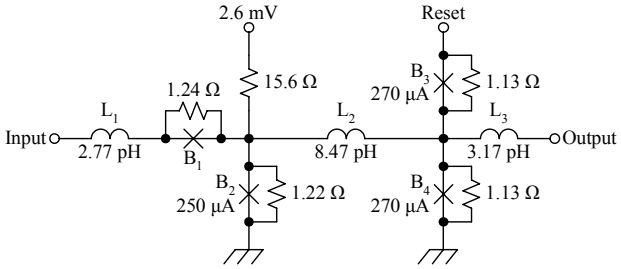


Fig. 3: DRO schematic [11]

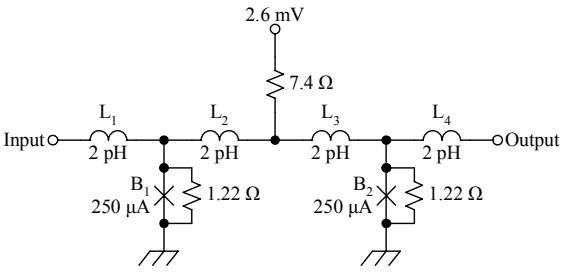


Fig. 4: JTL schematic [11]

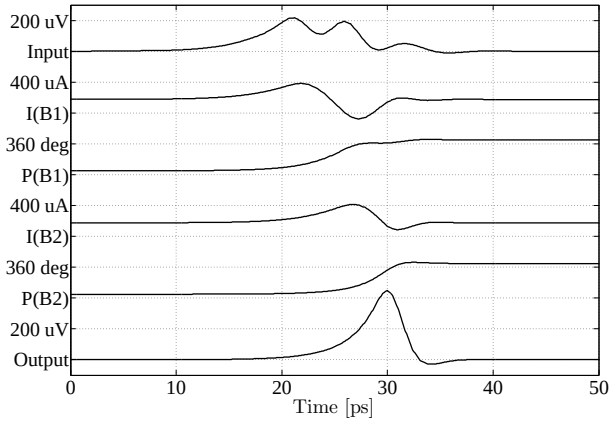


Fig. 5: JTL simulation results

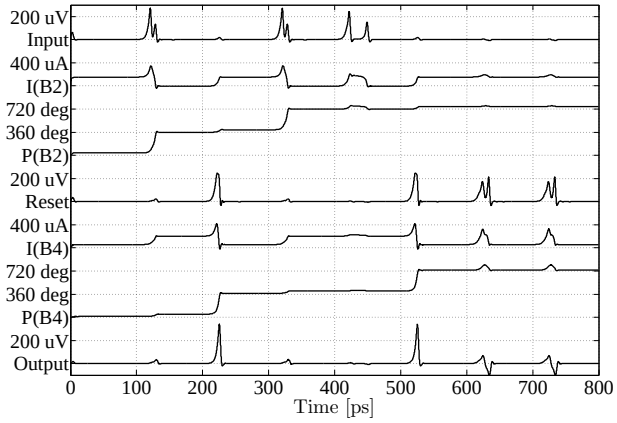


Fig. 6: DRO simulation results

(see Fig. 7). A single RSFQ pulse is not enough to switch junction B_3. The analogy with the pendulum is a jerk on mass M not large enough to make the pendulum swing past 180°. When two pulses reach the gate at the same time, however, the junction switches and passes the pulse on to the next gate.

Junctions B_1 and B_2 are buffers (245 μA), which allow pulses to travel in one direction only. Pulses on the one input therefore do not return on the other input, but rather continue to the output. Since the two input pulses must reach the inputs of the AND gate at exactly the same time, each input has a DRO. It is then not required to have a DRO on the output.

In order to connect gates that are physically far apart, a JTL can be used. Connecting gates directly results in inductances too large for the circuit to function correctly. A JTL can also be used as a standard termination, similar to the 50 Ω impedance in radio frequency (RF) circuits.

One might notice that each junction in the circuits presented has a parallel resistance. This is to critically dampen the junctions. The value of the resistor is such that the parallel combination of it, the effective normal resistance of the junction and the active impedance of the electrodynamic environment as seen by the junction results in $\beta_c = 1$ [4].

D. Physical Layout

Although YBCO (yttrium barium copper-oxide) has a much higher T_c than Nb (niobium), Nb is much easier to use in integrated circuit (IC) manufacturing. YBCO is a ceramic that must remain oxygenated to function properly, whereas Nb is a pure metal. The Hypres process [12] is a popular Nb based IC manufacturing process. The AND gate in Fig. 7 could be laid out as in Fig. 8, including a pulse splitter and two DROs.

Inductor dimensions were calculated using a program called 'SLine'. The properties of the layers are given and the resulting micro-stripline impedance, capacitance per μm and inductance per μm are calculated. Inductors must as far as possible be implemented as straight lines. Bending changes the properties of the line and a much more complicated calculation must be used to calculate the inductance [13].

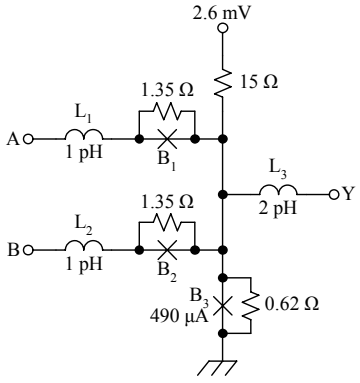


Fig. 7: AND gate schematic

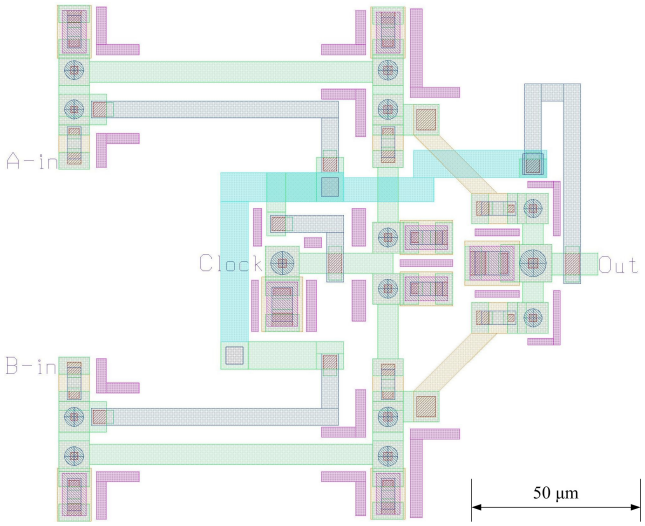


Fig. 8: The physical layout of an AND-gate

Ground plane moats are included to trap magnetic fields around the junctions. This helps to reduce parasitic magnetic fields within the junctions and therefore ensures the correct function of those junctions.

E. Simulation Results

The OR-gate and AND-gate can be combined to form a circuit similar to that in Fig. 9. In order to create accurate single flux quantum (SFQ) input pulses, a direct current (DC) to SFQ converter was used. The circuit for such a DC to SFQ converter was taken from [11]. The output of the AND gate is buffered with a JTL and then terminated with a 5 Ω resistor to ground.

The circuit was simulated using WRSpice. Fig. 10 shows the resulting voltage waveforms of the circuit in Fig. 9. The maximum of the correct output appears 55 ps after the start of the relevant clock pulse. This delay is largely due to the JTL.

V. Room Temperature Interface

A. Data Input

It is near impossible to generate RSFQ pulses using room temperature electronics. SFQ pulses are quantised to one fluxon, which is difficult to achieve with a circuit that is not inherently quantised.

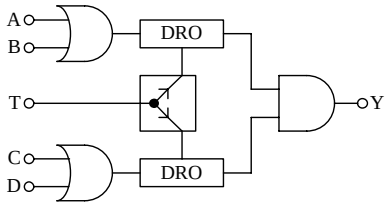


Fig. 9: Block diagram of the OR-AND gate

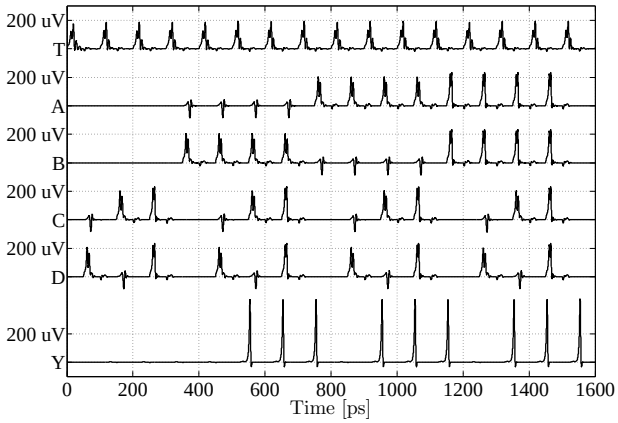


Fig. 10: Functional simulation results

One of the more popular techniques to interface room temperature digital circuits with superconducting circuits is the DC to SFQ converter [11]. This device generates an SFQ pulse on every rising-edge of the input signal. One can therefore encode the desired input signal to one that produces a '10' waveform for a 'high' and a '00' waveform for a 'low'. The clock signal could then be a series of '01' waveforms.

The idea would then be to input data relatively slowly, process it quickly using the fast RSFQ circuitry, and then clock the result out slowly.

B. Data Output

The RSFQ pulses are very small. A typical pulse is 441 μV high and 10 ps wide into a 5 Ω load (see trace Y in Fig. 10). The area under the pulse is equal to one fluxon. This means that the energy dissipated into the 5 Ω load is $114{\cdot}10^{-21}$ J (or 0.710 eV) per pulse, calculated using signal Y in Fig. 10 as well as (15), where R is 5 Ω.

$$E = \int \frac{v^2}{R} dt \tag{15}$$

When considering a pulse repetition rate of 1 GHz, the power in the resulting signal is 114 pW (or -69.4 dBm) over the entire spectrum. Due to the bandwidth of each pulse, the 1 GHz component is much smaller.

Popular converter circuits include the hybrid unlatching flip-flop logic element (HUFFLE) [11], SFQ to COSL converter [11], SFQ to DC converter [11] and Suzuki stack amplifier [14].

The Susuki stack amplifier has similar functionality to the DRO in that it changes state upon receiving an SFQ pulse and requires an external reset signal to revert back to the default state. In the 'reset' state the output is zero, whereas in the 'set' state the output is a few mV [15].

C. Biasing

All the logic gates in the RSFQ circuit must be biased with a specific current. This may be achieved my means of an on-

chip resistor and external voltage. It may also be achieved with an external current-source.

Current source biassing is preferred, because the resistance of the DC feed-lines are temperature-dependant in a cryogenic assembly. Bias current to a circuit is set at the combined requirement of all gates and/or cells and distributed with resistors on the chip.

VI. Cooling

A. Cryostat

For a niobium IC to function properly, it must be kept well below the T_c of niobium (9.2 K [1]). Fabrication processes are specified for 4.5 K [12]. This can be achieved by means of a cryostat.

A cryostat is a vessel, usually similar in structure to a vacuum flask, in which liquid helium is poured. It usually also has an outer container filled with liquid nitrogen. The IC is then lowered into the liquid helium (boiling point of 4.2 K at atmospheric pressure [1]).

One advantage of the cryostat is the noise-free environment. Since RSFQ circuitry function on fluxons being moved around, it is very susceptible to outside magnetic interference.

Disadvantages of the cryostat include the ease with which the liquid helium spontaneously evaporates (quenches), the cost of liquid helium and the time available for each experiment. The cryostat is therefore not a practical option for the industry.

B. Cryo-Cooler

Another option is the use of a cryogenic refrigeration unit (cryo-cooler). It is a refrigerator that uses helium as refrigerant rather than the freon gas usually found in household refrigerators. It is relatively inexpensive to operate continuously long term and is therefore the superior option for industrial use. Fig. 11 shows a typical block diagram for a superconducting electronics installation assembly.

One major disadvantage of industrial cryo-coolers is the electromagnetic noise they generate. The sensitive RSFQ IC must therefore be shielded. The use of a lead shield is popular, but it has a T_c of 7.2 K [1]. It will therefore become superconductive only after the niobium IC. One potential problem with this is that niobium is a type 2 superconductor [1], meaning that it can trap magnetic field in so-called vortices [1]. This trapped magnetic field lowers the I_c of the Josephson junctions and therefore inhibits the circuit from functioning properly. It may even prevent the niobium from becoming superconductive.

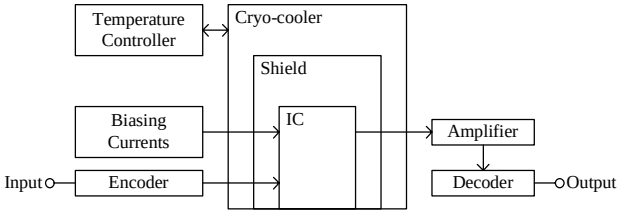


Fig. 11: Equipment Schematic

Although the cryo-cooler is the superior option with regards to industrial use of superconducting ICs, much research remains before it becomes more economical than the room-temperature equivalent.

VII. Conclusion

Superconducting technology is young and much research opportunities remain. Although most of the mathematical model is complete, there are some aspects of superconductivity that remain unexplained.

Many applications have been found in various aspects of physics and electronics. Very sensitive and accurate sensors are possible as well as very fast analogue to digital converters. Also, very fast digital circuits can be implemented, enabling the production of microwave-band digital signal processors.

There is much to be discovered and explored in this vast and mysterious world of superconductivity.

References

[1] W. Buckel and R. Kleiner, *Superconductivity: Fundamentals and Applications*, 2nd ed. Wiley-VCH, 2004, ISBN 3-527-40349-3.

[2] T. P. Orlando and K. A. Delin, *Foundations of Applied Superconductivity*. Massachusetts Institute of Technology: Addison-Wesley Publishing Company, 1991, ISBN 0-201-18323-4.

[3] H. A. Haus and J. R. Melcher, *Electromagnetic Fields and Energy*. Prentice Hall, 1989, ISBN 0-13-249020-X.

[4] K. K. Likharev and V. K. Semenov, "RSFQ logic / memory family: A new josephson-junction technology for sub-terahertz-clock-frequency digital systems," *IEEE Transactions on Applied Superconductivity*, vol. 1, no. 1, pp. 3 – 28, Mar. 1991.

[5] W. J. Perold, M. Jeffery, Z. Wang, and T. van Duzer, "Complementary output switching logic – a new superconducting voltage-state logic family," *IEEE Transactions on Applied Superconductivity*, vol. 6, no. 3, pp. 125 – 131, Sep. 1996.

[6] E. B. Wikborg, V. K. Semenov, and K. K. Likharev, "RSFQ front-end for a software radio receiver," *IEEE Transactions on Applied Superconductivity*, vol. 9, no. 2, pp. 3615 – 3618, Jun. 1999.

[7] C. J. Fourie and W. J. Perold, "A single-clock asynchronous input COSL set-reset flip-flop and SFQ to voltage state interface," *IEEE Transactions on Applied Superconductivity*, vol. 15, no. 2, pp. 263 – 266, Jun. 2005.

[8] ——, "An RSFQ DC-resettable latch for building memory and reprogrammable circuits," *IEEE Transactions on Applied Superconductivity*, vol. 15, no. 2, pp. 348 – 351, Jun. 2005.

[9] H. van Heerden, "The design and testing of a superconducting programmable gate array," Master's thesis, University of Stellenbosch, Dec. 2005.

[10] O. A. Mukhanov, V. K. Semenov, W. Li, T. V. Filippov, D. Gupta, A. M. Kadin, D. K. Brock, A. F. Kirichenko, Y. A. Polyakov, and I. V. Vernik, "A superconductor high-resolution ADC," *IEEE Transactions on Applied Superconductivity*, vol. 11, no. 1, pp. 601 – 606, Mar. 2001.

[11] C. J. Fourie, "A tool kit for the design of superconducting programmable gate arrays," Ph.D. dissertation, University of Stellenbosch, Dec. 2003.

[12] Hypres, "Niobium Integrated Circuit Fabrication Design Rules," Available at www.hypress.com, Jul. 2006.

[13] C. J. Fourie, "Quantum Electronics 833 – Applied Superconductivity: Digital and Mixed-Signal Electronics," Lecture Notes, 2007.

[14] J. X. Przybysz and P. Hills, "Superconducting Digital Logic Amplifier," United States Patent, 1989.

[15] L. Badenhorst, "Cryogenic Integrated Circuit CMOS Devices for RAM and Interface Applications," Master's thesis, Electronic Engineering, University of Stellenbosch, Stellenbosch, South Africa, 2008.

FEM Solution using Vector Elements for a Superconducting Microstrip Line

André Young , David B. Davidson and Coenrad J. Fourie
Department of Electrical and Electronic Engineering
University of Stellenbosch, South Africa
{ayoung, davidson, coenrad}@sun.ac.za

***Abstract*—As superconducting integrated circuits become smaller and operating frequencies increase, electromagnetic coupling effects become important and full-wave analysis of circuits is required. Although the development of full-wave methods and their application to electromagnetic structures are widely published, relatively little literature is found on the application of such methods to structures containing superconducting materials. The finite element method is one of the most-used methods in the computational electromagnetic community due to its computational efficiency and its ability to handle complicated geometries and material compositions. Here a mixed vector-nodal finite element method is formulated that incorporates superconducting materials through the two-fluid model. The method is applied to a superconducting microstrip and results are compared to analytic and numeric solutions found in literature.**

I. Introduction

The Finite Element Method (FEM) is well-established within the computational electromagnetic community and has been applied extensively to waveguide and other transmission line problems. The analysis of these problems are usually concerned with finding the propagation characteristics of a structure. For the FEM such an analysis generally leads to an eigenvalue problem of which the solutions (eigenvalue and eigenvector pairs) represent different propagation modes (propagation constants and field distributions) [1]. Owing to its origin in structural mechanics many of the first applications of the FEM in electromagnetics used nodal elements. However the nodal element approach was soon realised to contain many shortcomings when used to model electromagnetic fields. Of these the most prominent are the difficult treatment of sharp conducting corners where the field becomes singular and the inconvenience of added boundary conditions between material interfaces [1]. This led to the development of what is now the well-known vector finite element. In addition to alleviating the mentioned problems associated with nodal elements, the vector element approach also offers better approximation of the null space of the curl operator, which means spurious modes are more easily identified [2].

Propagation modes that exist in homogeneously filled waveguide can be divided into pure TE or TM modes, in which case the FEM is formulated using the appropriate transverse field components (electric field for TE and magnetic field for TM) as working variable. For inhomogeneously filled waveguide and microstrip the propagation modes are coupled TE-TM modes which means both the electric and magnetic field have transverse components. In [3] a mixed vector-nodal FEM formulation that can be used to solve for these propagation modes is presented. This formulation models only the electric or magnetic field and uses vector elements to model the transverse component and nodal elements to model the axial component of that field.

A FEM solution for solving propagation characteristics of a superconducting microstrip has been found in literature [4]. Although the method presented there eliminates spurious modes completely, nodal elements are used which presents the difficulties already mentioned. Furthermore all three vector components of both the electric and magnetic fields are modeled which degrades the computational efficiency of the solution.

Here the FEM solution presented in [3] is adapted to incorporate superconducting materials through the two-fluid model [5]. Advantages of this approach include the benefits associated with the use of vector elements over nodal elements as well as the increased efficiency resulting from modeling only the electric field.

II. Formulation

Formulation of the full-wave FEM solution starts with the derivation of an appropriate vector wave equation. Using this equation a functional is defined which is then discretised and rendered stationary to obtain a linear system in the form of an eigenvalue problem. Solving the linear system results in the required eigenvalues (propagation constants).

The Maxwell equations governing the propagation of electromagnetic fields are well-known. Through combination of these equations the vector wave equation can be obtained

$$\nabla \times \frac{1}{\mu_r} \nabla \times \bar{\mathbf{E}} - k_0^2 \epsilon_r \bar{\mathbf{E}} + j\omega\mu_0 \bar{\mathbf{J}} = 0. \tag{1}$$

From the two-fluid model of superconductivity the total current density $\bar{\mathbf{J}}$ is simply the superposition of the normal conduction current density $\bar{\mathbf{J}}_n$ and the supercurrent density $\bar{\mathbf{J}}_s$. The first London equation relates the supercurrent and electric field

$$\bar{\mathbf{E}} = \frac{\partial}{\partial t}(\Lambda \bar{\mathbf{J}}_s). \tag{2}$$

Introducing the London penetration depth λ_L and transforming to the frequency domain gives the form preferred here

$$\bar{\mathbf{J}}_s = \frac{1}{j\omega\mu_0\lambda_L^2}. \tag{3}$$

The normal conduction current is related to the electric field through the usual $\bar{\mathbf{J}}_n = \sigma_n \bar{\mathbf{E}}$ where σ_n is the normal conductivity. Note that σ_n and λ_L are temperature dependent

$$\sigma_n(T) = \sigma_0 \left(\frac{T}{T_c}\right)^4 \quad (4)$$

$$\lambda_L(T) = \frac{\lambda_0}{\sqrt{1-\left(\frac{T}{T_c}\right)^4}}. \quad (5)$$

Substituting the total current density as the sum of the normal and supercurrent densities the vector wave equation for the electric field equation becomes

$$\nabla \times \frac{1}{\mu_r} \nabla \times \bar{\mathbf{E}} - (k_0^2 \epsilon_r - j\omega\mu_0\sigma_n - \frac{1}{\lambda_L^2 \mu_r})\bar{\mathbf{E}} = 0. \quad (6)$$

The FEM finds a solution in a closed domain and requires adequate specification of the boundary conditions on the perimeter of this domain. To this end the microstrip structure is enclosed in a perfectly electrical conducting (PEC) box. The floor of this box forms the ground plane while the side walls and roof are placed sufficiently far from the strip as to not have a significant effect on the field distribution. Since the enclosing walls are PEC the boundary condition here simply forces the electric field component tangential to these boundaries to zero. The field distribution of the dominant propagation mode (desired) also allows the substitution of a magnetic wall in place of the plane of symmetry passing vertically through the center of the microstrip. In effect this halves the spatial domain on which a solution is sought and the resulting system is computationally less costly. Thus the boundary for the closed domain comprises a magnetic wall Γ_2 and three electric walls Γ_1 and the boundary conditions on these boundaries are

$$\hat{n} \times \bar{\mathbf{E}} = 0 \quad \text{on} \quad \Gamma_1 \quad (7)$$

$$\hat{n} \times \nabla \times \bar{\mathbf{E}} = 0 \quad \text{on} \quad \Gamma_2. \quad (8)$$

Obtaining a functional from (6) is standard procedure in finite element formulations and the details will not be given here. In brief it involves multiplying with a testing function, integrating over the entire problem region and finally using Green's theorem to transfer one of the curl operators to the testing function. Here a Galerkin method is used and the testing and approximation functions are the same. Also, to seperate the field into transverse $\bar{\mathbf{E}}_t$ and axial E_z components, the z-dependence of the electric field is assumed as in [3], $\bar{\mathbf{E}}(x,y,z) = \bar{\mathbf{E}}(x,y)e^{-jk_z z}$. Note that the propagation constant k_z is generally complex valued, where the real and imaginary parts are the phase and attenuation constants, respectively. The resulting functional is now

$$F(\bar{\mathbf{E}}) = \frac{1}{2}\iint_\Omega \left[\frac{1}{\mu_r}(\nabla_t \times \bar{\mathbf{E}}_t)\cdot(\nabla_t \times \bar{\mathbf{E}}_t) - \left(\frac{1}{\lambda_L^2} + j\omega\mu_0\sigma_n - k_0^2\epsilon_r\right)\bar{\mathbf{E}}_t \cdot \bar{\mathbf{E}}_t + \frac{1}{\mu_r}(\nabla_t E_z + jk_z\bar{\mathbf{E}}_t)\cdot(\nabla_t E_z + jk_z\bar{\mathbf{E}}_t)\right] d\Omega. \quad (9)$$

Since the operating frequency (and therefore both ω and k_0) is generally known and the desired unknown is the propagation constant k_z, the above functional has to be manipulated into a more suitable form. This is done through the scaling [3]

$$\bar{\mathbf{e}}_t = k_z \bar{\mathbf{E}}_t \quad (10)$$

$$e_z = -jE_z. \quad (11)$$

Substituting these expressions into the functional (9) gives

$$F(\bar{\mathbf{e}}) = \frac{1}{2}\iint_\Omega \left[\frac{1}{\mu_r}(\nabla_t \times \bar{\mathbf{e}}_t)\cdot(\nabla_t \times \bar{\mathbf{e}}_t) + \left(\frac{1}{\lambda_L^2} + j\omega\mu_0\sigma_n - k_0^2\epsilon_r\right)\bar{\mathbf{e}}\cdot\bar{\mathbf{e}} + \frac{1}{\mu_r}(\nabla_t e_z + jk_z\bar{\mathbf{e}}_t)\cdot(\nabla_t e_z + jk_z\bar{\mathbf{e}}_t)\right] d\Omega. \quad (12)$$

The next step in the FEM is discretisation which involves the subdivision of the domain into smaller elements and definition of approximation (and testing) functions on these elements. Here triangular elements are used to enable a graded mesh for finer resolution near the microstrip. The well-known CT/LN (first order) and LT/QN (second order) vector basis functions [2] are used to approximate the tangential field within these elements, while first and second order scalar basis functions are used to approximate the axial fields. With these approximation functions the field within the k^{th} element can then be expressed as the weighted sum of the vector and nodal basis functions

$$\bar{\mathbf{e}}_t^k = \sum_i^{N_k} a_i^k \bar{\mathbf{v}}_i^k \quad (13)$$

$$e_z^k = \sum_i^{N_k} b_i^k \phi_i^k. \quad (14)$$

In the above v_i^k and ϕ_i^k are the vector and nodal basis functions, respectively; and a_i^k and b_i^k are the unknown weights associated with each basis function.

The above expansions are substituted into (12) and the variation of the functional with respect to the unkown weights is equated to zero. The result is a linear system

$$\begin{bmatrix} A_{tt} & 0 \\ 0 & 0 \end{bmatrix}\begin{Bmatrix} a \\ b \end{Bmatrix} = -k_z^2 \begin{bmatrix} B_{tt} & B_{tz} \\ B_{zt} & B_{zz} \end{bmatrix}\begin{Bmatrix} a \\ b \end{Bmatrix} \quad (15)$$

which is a generalised eigenvalue problem. Finally the scaling suggested in [3] is applied to obtain

$$\begin{bmatrix} B_{tt} & B_{tz} \\ B_{zt} & B_{zz} \end{bmatrix}\begin{Bmatrix} a \\ b \end{Bmatrix} = \frac{\theta^2}{\theta^2 - k_z^2}\begin{bmatrix} B_{tt} + \frac{A_{tt}}{\theta^2} & B_{tz} \\ B_{zt} & B_{zz} \end{bmatrix}\begin{Bmatrix} a \\ b \end{Bmatrix} \quad (16)$$

where $\theta^2 = k_0^2 \epsilon_{r(\max)} \mu_{r(\max)}$ and $\epsilon_{r(\max)}$ and $\mu_{r(\max)}$ are the highest relative permittivity and highest relative permeability, respectively, within the structure analysed. The scaling aims to ensure that the eigenvalue associated with the desired propagation mode becomes the dominant eigenvalue. The advantages of this are two-fold: firstly, it allows for easier separation of the obtained physical and spurious modes; and

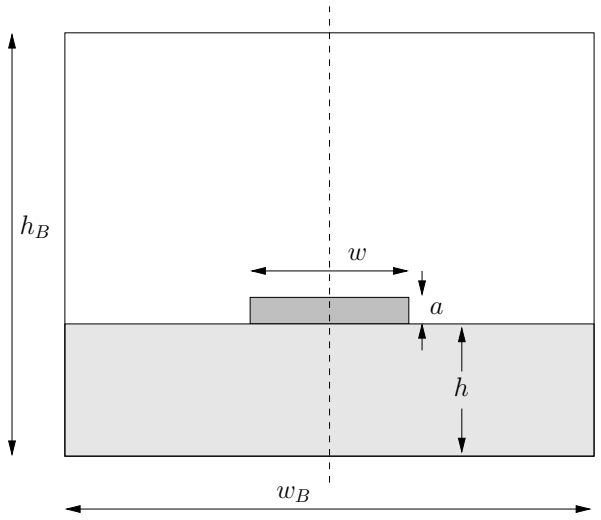


Fig. 1. Dimensions describing the geometry of the boxed-in microstrip structure. The dashed line indicates the plane of symmetry (magnetic wall).

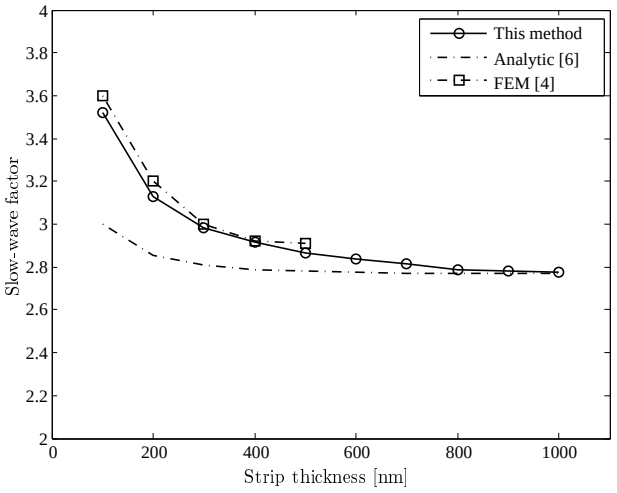


Fig. 2. Effect of strip thickness on the slow-wave factor of the superconducting microstrip.

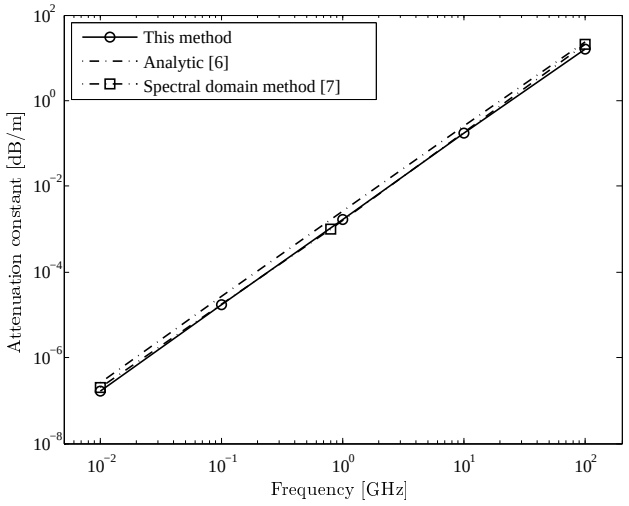


Fig. 3. Variation of attenuation constant with increasing frequency.

secondly, a larger eigenvalue means faster convergance when using sparse methods to obtain the solutions of (16).

Solving the system (16) and descaling gives the propagation constant $k_z = \beta - j\alpha$. From this result the attenuation constant α and slow-wave factor $\sqrt{\epsilon_{\text{eff}}} = \frac{\omega}{c\beta}$ are calculated.

III. Numerical Results

The FEM solution is implemented and used to solve for the propagation mode of superconducting microstrips of various dimensions. Numerical results obtained with this method are compared to results obtained using the analytic solution presented in [6] as well as other numerical results found in literature [4] [7]. Fig.1 defines the geometrical parameters that desribe the microstrip structure.

The first microstrip structure considered is as in [4], with w = 5 μm, h = 4.5 μm, h_B = 50 μm, w_B = 50 μm and the strip thickness a is increased from 100 nm to 1000 nm. Material parameters are λ_0 = 320 nm, $\sigma_0 = 10^4$ S/m, ϵ_r = 10.5, $\tan\delta$ = 0 and T_c = 12.15 K. The solution is obtained for a temperature of 4 K and an operating frequency of 10 GHz . The results for the slow-wave-factor are shown in Fig.2 and compared to numerical results in [4] and analytic results obtained using the method in [6]. For a thin strip the two finite element solutions give very similar results, while the analytic solution predicts a much lower slow-wave factor. The discrepancy between the numerical and analytic solutions is ascribed to the assumption that $a >> \lambda_L$ inherent in the analytic solution [6]. As the strip thickness increases the slow-wave factor is seen to decrease, more so for the numerical results than for the analytic solution. Above a thickness of about 700 nm the results for the current method and the analytic solution converge, affirming the suggested reason for the discrepancy in results for a thin strip.

Next a configuration from [7] is used, with w = 70 μm, h = 100 μm, h_B = 1000 μm, w_B = 1000 μm and a = 3 μm. Material parameters are λ_0 = 180 nm, $\sigma_0 = 7.46 \times 10^6$ S/m, ϵ_r = 12.6, $\tan\delta$ = 0 and T_c = 90 K. The solution is obtained for a temperature of 77 K and the frequency is increased from 0.01 GHz to 100 GHz. Results for the attenuation constant are shown in Fig.3 and compared to numerical results from [7], which employs a spectral domain method, and again analytic results using [6]. The conduction losses predicted by each of the three methods give are very similar and show a steady increase with frequency.

Now for the same configuration the frequency is kept constant at 7 GHz and the ratio T/T_c is increased from 0.2 to 0.95. Fig.4 shows the results for the attenuation constant. Again the results from all three solutions are found to be very similar and show a drastic increase in losses as the temperature approaches the critical temperature of the superconductor.

IV. Conclusion

A mixed vector-nodal FEM solution has been formulated and implemented to solve for the propagation modes of a

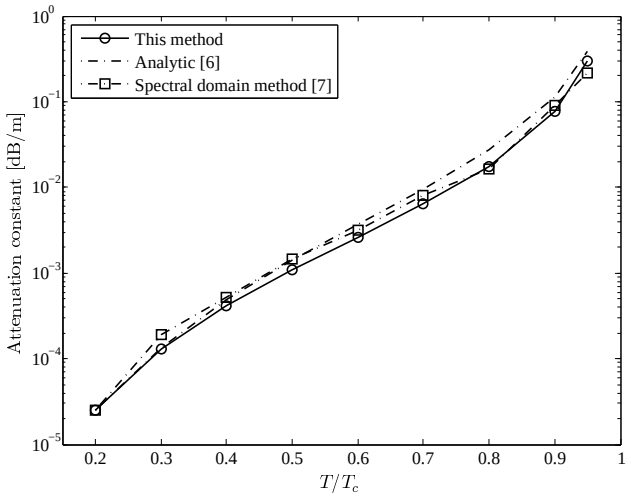


Fig. 4. Variation of attenuation constant with increasing temperature.

superconducting microstrip. Results have been found to compare well with other numerical solutions as well as an analytic solution found in literature. Advantages of the current method are the use of vector elements, along with the mentioned benefits associated with these as opposed to nodal elements, as well as the fact that the formulation requires approximation of only the electric field and is therefore computationally efficient. Using appropriate scaling along with vector elements also results in easy separation of spurious modes present in the solution.

V. Acknowledgements

The first author wishes to thank the NRF for funding in the form of a bursary.

References

[1] J.-M. Jin. *The Finite Element Method in Electromagnetics*. Wiley, New York, NY, 2nd edition, 2002.

[2] D. B. Davidson. *Computational Electromagnetics for RF and Microwave Engineering*. Cambridge University Press, 2005.

[3] J.-F. Lee, D.-K. Sun, and Z. J. Cendes. Full-wave analysis of dielectric waveguides using tangential vector finite elements. *IEEE Trans. MTT*, 39(8):1262–1271, August 1991.

[4] D. Kinowski, P. Pribetich, and P. Kennis. Finite-element analysis of superconducting microstrip line for MMIC applications. *Microwave and Optical Technology Letters*, 5(8):369–371, July 1992.

[5] T. P. Orlando and K. A. Delin. *Foundations of applied superconductivity*. Addison-Wesley, Reading, Mass., 1991.

[6] M. S. Boutboul, H. Kokabi, and M. Pyée. Modeling of microstrip quasi-TEM superconducting transmission lines, comparison with experimental results. *Physica C: Superconductivity*, 309(1-2):71–78, December 1998.

[7] S.-G. Mao, J.-Y. Ke, and C. H. Chen. Propagation characteristics of superconducting microstrip lines. *IEEE Trans. MTT*, 44(1):33–40, January 1996.

AFM plough YBCO microbridges: Substrate effects

A. Elkaseh, U. Büttner, V.V. Srinivasu, and WJ. Perold, *Senior Member, IEEE*

***Abstract*— Atomic force microscope (AFM) nanolithography was used as a novel cutting technique to defined micro-size constrictions on $YBa_2Cu_3O_{7-x}$ striplines. $YBa_2Cu_3O_{7-x}$ (YBCO) thin films are deposited using an inverted cylindrical magnetron (ICM) sputtering technique. The films are then patterned into 8-10 micron width strips, using photolithography and dry etching. In order to understand the effects of substrates, we fabricated YBCO planar microbridges on MgO and STO substrates. We studied the substrate effects on Current-Voltage (I-V) characteristics and Shapiro-steps. We show that the observed Shapiro steps from the bridges on STO substrates are of poor quality, when compared to those on MgO substrates. This is because the STO substrate is 'microwave dirty' and it is well known that the substrate has resonant microwave losses at low temperatures.**

***Index Terms*—AFM lithography, Josephson junction, SrTiO3, YBCO**

I. Introduction

One of the essential factors to obtain epitaxial HTS thin films and to fabricate good quality thin-film electronic devices is the choice of substrate material. The substrates should have a clean smooth surface and a close lattice match with the films to ensure epitaxial growth. The thermal expansion coefficients of the substrate and the HTS film must be comparable over a large temperature range to avoid excess strain relaxation - significant differences in the thermal expansion coefficients will lead to cracking of the film [1]. Chemical compatibility between the substrate and the film is important and there should not be any chemical reaction, because such a reaction between the substrate and the film will inhibit good epitaxy and may prevent the formation of the superconductive phase.

There are many materials that have been tested as substrates for HTS cuprate thin-films, especially $YBa_2Cu_3O_{7-x}$ thin films. Generally, the substrates that are suitable for growth of YBCO can be categorized into two groups: Perovskite structures such as $SrTiO_3$, $LaAlO_3$, and $NdGaO_3$, and non-Perovskite structures such as MgO and Al_2O_3.

$SrTiO_3$ has a very small lattice mismatch with YBCO (2%), however, the substrate is unsuitable for high frequency applications due to its large dielectric constant (277 at room temperature). MgO substrates have a modest dielectric constant (9.65), it is readily available, and has a 9% lattice mismatch with YBCO [2]. A very low dielectric constant and the low cost make MgO a more suitable substrate for junctions for high frequency applications.

Planar Josephson junctions such as micron and nanobridges were successfully fabricated on high-temperature superconductors using focused ion beam (FIB) and electron beam lithography (EBL) [3, 4].

Recently we have demonstrated AFM ploughing as a novel way of fabricating YBCO planar constriction type micro-bridges [5]. In this paper, we report on the substrate effects on the I-V characteristics and Shapiro-steps in these recently fabricated micron size bridges, on MgO and STO substrates. Understanding the substrate effects on the performance of these microbridges is very important. For example, STO substrates exhibit severe microwave resonances in a range of temperatures from 60 K to 90K [6]. These resonances can interfere with the observation of Shapiro-steps. Our measurements show that microbridges fabricated on STO substrates show poor quality Shapiro-steps, as compared to those on the MgO substrates.

II. Experimental procedure

A. Thin film preparation

YBCO thin films were grown on (100) oriented MgO and $SrTiO_3$ substrates by Inverted Cylindrical Magnetron sputtering. Before the deposition, the thicknesses of the films were between 120 nm and 150 nm. The substrates were cleaned with acetone in an ultrasonic bath for 10 minutes, and then blown dry with nitrogen. The substrates were glued on a heater plate with silver paste. The YBCO ceramic target–substrate distance was 30mm. Pre-sputtering was applied for 15 minutes to eliminate any contamination on the surface of the target. The substrate temperature was 740° C, the total pressure of the 1:1 argon/oxygen gas mixture was 320 mTorr, and the dc sputtering power 72 W with a current of 400 mA and a voltage of 180V. The deposition rate was approximately 2.7 nm/min. After the deposition, the films were cooled and

Manuscript received 06 March 2009. This work was supported in part by the Centre for Macromolecular Chemistry and Technology in Libya.

V. V. Srinivasu is with Superconducting Materials and Technology Programme, Materials Science & Manufacturing Unit , CSIR, Pretoria, South Africa 0001 (e-mail: svallabhapurapu@csir.co.za).

W. J. Perold is with the Department of Electrical & Electronic Engineering, Stellenbosch University, Western Cape, South Africa (e-mail: wjperold@sun.ac.za).

annealed at 460° C for 30 min in an oxygen environment, before it was allowed to cool down to room temperature.

The surface morphology of the films was characterized by AFM.

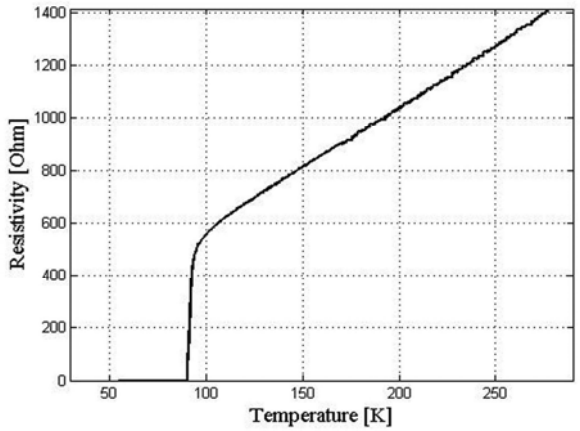


Fig. 1. Resistance versus temperature characteristic for 100 nm YBCO film.

The resistive behaviour of the films was measured. Silver (Ag) contact pads with thicknesses of 400 nm were evaporated onto the films surface and subsequently annealed at 470° C in a 1 atm oxygen environment for 30 min, in order to obtain low-resistivity contacts.

B. Microbridge fabrication

The YBCO film was patterned into microstrips with 8–10 μm widths, using standard photolithography and argon ion milling.

Photoresist (ma-P 1225) was spun onto the film surface at 4000 rpm to a thickness of 2 μm, and soft-baked on a hot-plate at 100° C for 5 min. The resist was exposed to 360 nm UV through a chrome contact mask for 25 sec, and then developed for 50 sec in ma-D 331 developer. The resist was then hard-baked at 110° C for 15 min.

Argon ion milling was performed using a Kaufman-type ion source with an incidence angle of 45° to the film surface. The rf power was 50 W, the argon gas pressure 0.25 mTorr, and the voltages of the electrodes were 750 V.

In order to minimize the loss of oxygen from the superconducting layers during the milling process, the film was mounted on a water-cooled copper sample holder with thermal paste.

The etching was terminated when an open circuit on the substrate at the sides of the YBCO striplines was measured.

Photoresist residue was removed by immersing the film in mr-Rem 660 photoresist remover for 10 min, and then immersing it in acetone for 10 min in an ultrasonic bath.

We used a diamond coated tip as a cutting tool to define the plough. The AFM was operated in contact mode, with the tip vertically displaced toward the YBCO surface with a loading force of 11 μN, sufficient to completely remove the YBCO layers.

The YBCO stripline is imaged first, and then the middle of the stripline is moved to the centre of the image. At the start of the nanolithography process the tip is placed at the centre of the image at the centre of the stripline width. The constriction width W is controlled by displacing the tip by

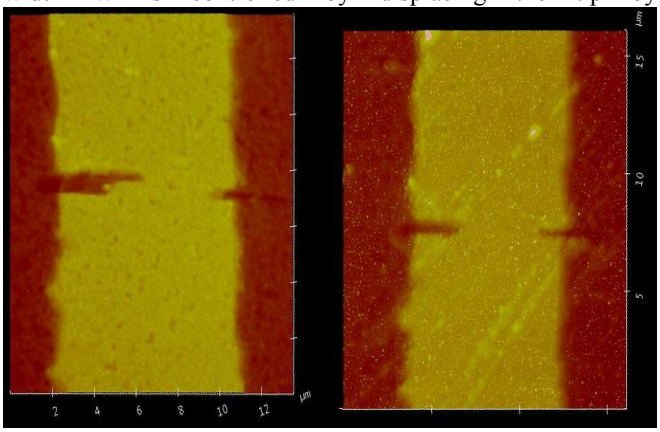


Fig. 2. Microbridges junction on MgO and STO substrates made by AFM.

W/2 to the left side on the stripline. The tip is then driven into the YBCO surface, and displaced on the same line for a few hundred cycles. The tip velocity was 4 μm/s. The constriction is completed by applying the same nanolithography on the right side of the stripline. Constrictions with fully controlled width and depth were achieved by adjusting the tip movement and the scan speed of the plowing operation. The YBCO residuals from the cuts were cleaned by dipping the film in acetone in an ultrasonic bath for a few minutes. The drawback of this technique is the wear and resolution degradation of the tips after it has been used for a few times in the lithography.

III. RESULTS AND DISCUSSION

A. Thin film characterization

The surface morphology of the YBCO thin films was investigated using AFM. The films had smooth surfaces, with surface roughness values between 4 to 6 nm. No droplets or outgrowths were observed on the surface of the films.

Fig. 1 shows the temperature dependence of the electrical resistance. The films show metallic-like resistive behaviour in the normal state and have sharp superconducting transitions.

The residual resistance ratio R (300)/R (100) was 2.83, the zero resistance temperature (T_{C0}) was 90.2 K, and the width of the superconducting transition about 1 K. These results confirmed that the films had good superconducting properties.

B. Josephson-effect measurements on microbridges

After the successful fabrication of the microbridges with AFM nanolithography (see Fig. 2), the film was mounted on a PC board. Gold wires were bonded from the silver contact pads to the copper striplines on the PCB. The sample was then positioned inside the cold finger cryocooler unit for testing.

We used the Mr. SQUID Electronics unit [7] as excitation source. It generates a triangular waveform to test the device, and we monitored the I-V response on the oscilloscope window.

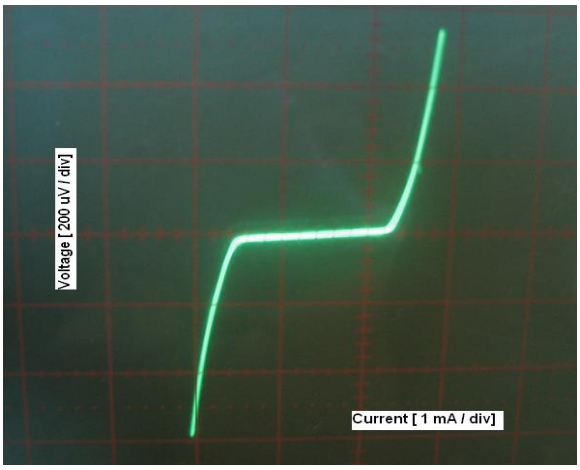


Fig. 3. I-V curve of a 3.6 μm microbridge junction on MgO substrate.

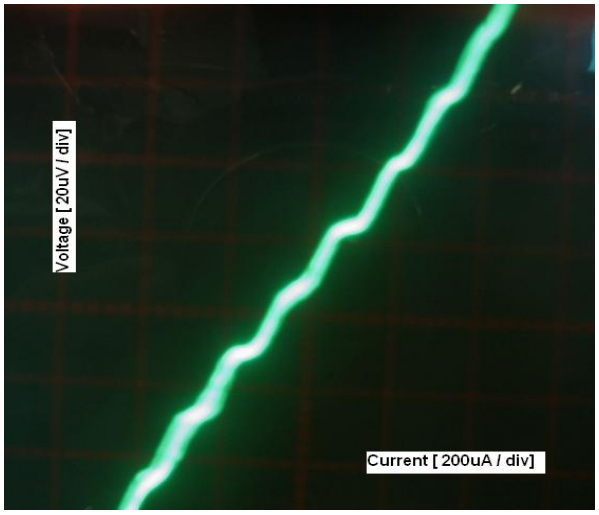


Fig. 4.a. Shapiro-steps of the same constriction junction (Fig. 3) exposed to 8.1522 GHz MW power.

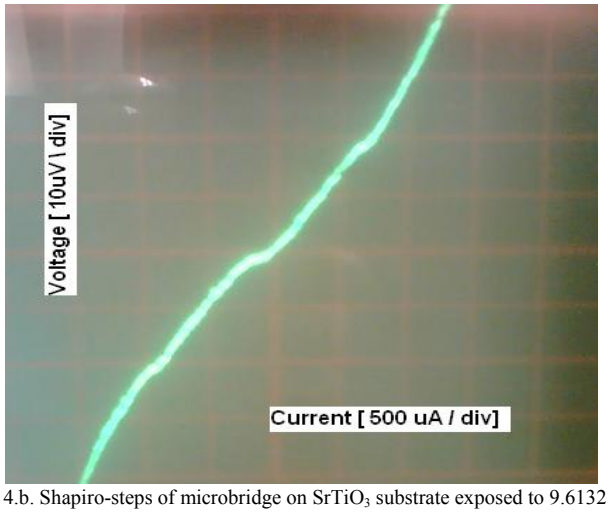


Fig. 4.b. Shapiro-steps of microbridge on $SrTiO_3$ substrate exposed to 9.6132 GHz.

The current-voltage (I-V) characteristics of a 3.6 μm width microbridge on a MgO substrate, in the absence of microwave power, at 57 K are shown in Fig. 3. The critical current was measured to be 1.58 mA at 47 K. This temperature was the lowest that the specific crycooler could reach. The normal resistance of the junction was $1.2\,\Omega$, and the I_cR_n-product was 90 μV at 77 K.

As a result of thermal fluctuations or flux flow a rounding effect can be present in the I-V characteristics of a junction [8]. However, our I-V curves exhibit sharp knees, possibly ruling out any thermal fluctuations or flux flow effects. This means that the I-V curves are most probably representing true Josephson behaviour.

However, a mere observation of I-V characteristics with dc currents may not necessarily mean the demonstration of the Josephson-effect. In order to establish the Josephson-effect, one should demonstrate ac and dc Josephson-effects, and the magnetic modulation of critical currents. In this paper we adopt the method of the observation of Shapiro-steps, which is a direct consequence of the Josephson-effect. Accordingly, measurements on the I-V characteristics of the microbridge junctions have been performed in the presence of an external microwave power source. Here we should emphasize that the substrate can have an influence. For example STO has microwave surface impedance resonances in the temperature range 60K to 90K [5]. This can interfere with the Shapiro-step quality in the microbridges fabricated on STO. MgO is supposed to be relatively microwave silent in this temperature range. In fact, we did observe poor quality Shapiro-steps in the case of microbridges fabricated on STO substrates, as compared to that of MgO substrate, as shown in Fig. 4 (a) and (b).

We applied 3 dBm microwave power in the 2 – 18 GHz range via a coaxial cable terminated with an antenna above the device. We observed well defined Shapiro-steps on the I-V curve of the microbridge on a MgO substrate at 8.1522 GHz (see Figure 4(a)). The response of the constriction to microwave radiations (Shapiro-steps) clearly gives positive evidence of the Josephson-effect. This is because the observed step size satisfies (1), which can only be derived from the fundamental Josephson-effect.

The step sizes can be used to calculate the theoretical constant $\frac{e}{h}$.

The voltage step is expressed as

$$V_0 = n\left(\frac{\Phi_0}{2\pi}\right)\omega_s \qquad (1)$$

which is equivalent to

$$V_0 = n\left(\frac{h\omega_s}{4\pi e}\right) = n\frac{hf_s}{2e} \qquad (2)$$

where $\Phi_0 = \frac{h}{2e}$. This equation can be rewritten to yield the theoretical constant $\frac{e}{h} = 2.41796\times10^{14}\,Hz\,/\,V$ as:

$$\frac{e}{h} = n\left(\frac{f_s}{2V_0}\right). \qquad (3)$$

The voltage V_0 is approximately $17\,\mu V$ with some degree of uncertainty and the frequency f_s =8.1522 GHz. Substitution of these values into (3) results in $\frac{e}{h} = 2.3977 \times 10^{14}\ Hz/V$. This value corresponds quite well with the theoretically predicted value, constituting positive evidence for the Josephson-effect.

Fig. 4(b) shows Shapiro-steps on the I-V curve of a microbridge fabricated on a $SrTiO_3$ substrate. The voltage step height is about $20\,\mu V$ at a frequency f_s = 9.6132 GHz, resulting in $\frac{e}{h} = 2.403 \times 10^{14}\ Hz/V$.

The Shapiro-steps of our junction on a $SrTiO_3$ substrate are not clear compared to the steps on the MgO substrate. Nevertheless, the critical current is larger in the microbridge made on the $SrTiO_3$ substrate.

IV. Conclusion

In conclusion, we have produced planar micron-size bridge type junctions on YBCO thin films by AFM lithography. The junctions were tested for the Josephson-effect, and we have observed I-V characteristics and Shapiro-steps with external microwave irradiation of the microbridges on both MgO and STO substrates. The voltage steps are well defined on the MgO substrate, which demonstrate the Josephson-effect in these junctions. The voltage step value corresponds well with the predicted theoretical constant for Shapiro-steps.

References

[1] R. Wordenweber, "Growth of high-tc thin films," *Superconuct. Sci. Technol*, vol. 12,1999, pp. R86-R102.

[2] J. Philips, "Substrate selection for high-temperature superconducting thin films," *Jour. Appl. Phys.*, vol. 79, 1996, pp. 1829–1848.

[3] S. Lee, S Oh, C. Kang, and S. Kim, "Superconducting nanobridge made from YBCO film by using focused ion beam," *Physica C*, vol. 460, 2007, pp. 1468–1469.

[4] J. Wendt, J. Martens, C. Ashby, T. Plut, V. M. Hietala, C.Tigges, and D. Ginley, "YBCO nanobridges fabricated by direct-write electron beam lithography," *Appl. Phys. Lett.*, vol. 61, 1992, pp. 1597–1599.

[5] AAO Elkaseh, V.V. Srinivasu and WJ Perold, IEEE Trans. Appl. Supercond. 2009 (in press)

[6] N. Pompeo, L. Muzzi, V. Galluzzi, R. Marcon and E. Silva, Supercond. Sci. Technol. Vol. 20 , 2007, pp. 1002-1008.

[7] Conductus, *Mr.SQUID User's Guide.* Conductus Inc, 969 West Maude Avenue, Sunnyvale, California, USA, 2 ed., 1992.

[8] A. Barone, and G. Paterno, *Physics and Applications of the Josephson effect,* John Wiley & sons, New York, USA, 1982.

Superconducting sensors and imagers

D. Northeast*, C. J. Fourie†
*†University of Stellenbosch, Department of Electrical and Electronic Engineering, South Africa
*northeast@sun.ac.za, †coenrad@sun.ac.za

***Abstract*—Superconductor-based sensors provide improvements over traditional types of sensors (such as semiconductor-based), with improvements in both sensitivity, lower noise levels, and detection speed. Popular detectors are outlined, as well as common applications. Typical manufacturing, experimentation and measurement methods are also discussed.**

I. Introduction

Materials enter the superconducting state with their resistance rapidly transitioning to zero. Small continuous or impulse signals can create a large change in the properties of a superconductor, and so these materials can make excellent sensor devices. In scientific fields where technological deficiencies in efficient detectors exist (such as the terahertz regime[1] [1]), superconductors provide a new avenue for research and the creation of useful devices for industry.

Described first are the photon detectors. These are frequently used as counting devices similar to photomultipliers or photodiodes. They have applications ranging from quantum computing [2], [3] to telecommunications [2], [3]. Subsequently two kinds of antenna-coupled detectors are characterized. These are often used in astronomy, imaging, and spectroscopy.

Superconducting sensors often rely on changes in properties when superconductivity is broken. The steep change to a resistive mode provides an easily measurable output signal. Devices built on this principle can detect single photons usually of optical wavelengths, but also radiation at higher and lower frequencies. Other detectors use the change in resistance (see Fig. 1) to allow for frequency down-mixing. If a system is biased in the resistive state, then any signal passing through a transitioning superconducting element will cause a nonlinear change in the element's resistance.

[1]Often defined from 0.1 THz to 10 THz.

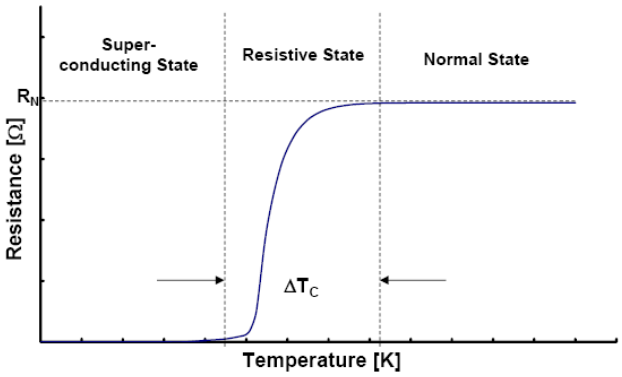


Fig. 1: The sharp transition from the superconducting to the normal state is a sensitive mechanism for the detection of incident energy.

The technologies which rival superconductor-based systems are typically created with semiconductors. For instance, Schottky diodes are used in terahertz (THz) imaging [1], along with hot-electron bolometers and superconductor-insulator-superconductor junctions. The diodes can be operated at room temperature, and systems capable of generating images of solid objects have been made. The benefits of superconductor-based sensors lie in the inherently low noise of the systems [4] and the small characteristic times related to detection and electron recombination. This speed, sensitivity, and low noise are requirements for passive detectors used in astronomy and terrestrial imaging devices.

II. Photon Detectors

Ideal pulse-based detectors will have an easily measurable output that has an infinitesimally small response time and no dead time[2]. Real pulsed-based detectors strive to mimic these properties as closely as possible. If a sequence of photon events (a few nanoseconds apart) are to be detected, a good detector would have a low dark count rate[3], a high quantum efficiency[4], a short dead time, and a low timing jitter[5].

Thin films (usually less than 10 nm) etched in long meanders (often several micrometers by some fraction of a micrometer, as seen in Fig. 2) exhibit great sensitivity, low noise, and fast pulse rates [5]. It has been shown that when most superconductors are made too thin superconductivity can be broken [6]. Kinetic inductance of the wires also increases as wires are made narrower [7], slowing the reset time of the detectors [7]. This theoretically sets bounds to a sensitivity versus reset time problem, and optimization would make use of a superconductor with a relatively low kinetic inductance and small crystal size.

Currently, niobium and niobium nitride are the most commonly used superconducting photon detectors material.

[2]The time after a detection event until the detector is able to detect once more.

[3]The number of counts when a detector is under no illumination.

[4]The number of detected events divided by the number of actual events impinging the detector, assuming the detector is fast enough to theoretically detect all events.

[5]The time difference between when an event is recorded and when it actually occurred.

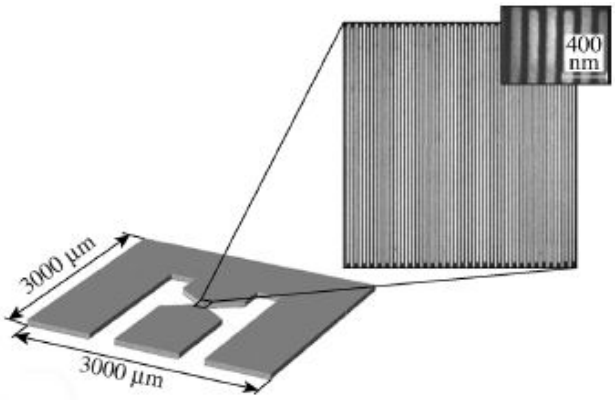


Fig. 2: A nanowire meander structure. [2]

Superconductors with higher critical temperatures, such as magnesium diboride [8], are potential replacements. Higher critical temperatures (T_c) often allows for greater biasing currents (since $0.5 * T_c$ is a common operating point), which will increase the output signal and boost sensitivity. Magnesium diboride and other high-T_c superconductors have a lower kinetic inductance than niobium and niobium nitride, and this allows for a greater counting rate for the same wire dimensions.

The physical process that leads to the pulse generation can be modeled as the generation of hot spots [5]. When a photon impinges on a superconductor, it has the possibility of breaking Cooper pairs if the photon energy exceeds the Cooper binding energy. Photon energy exceeding the binding energy will give the electrons some kinetic energy. If this kinetic energy is greater than the Cooper binding energy, further Cooper pairs can be broken by cascading electrons [5]. An avalanche scenario can create a region of normal resistance. If the system is biased near the critical current, J_c, the current that bypasses the resistive region will cause a breakdown of superconductivity across the entire path of the wire, generating a voltage pulse. These pulses can be read out using SQUID electronics [9].

The recombination of the electrons occurs in one of two scenarios: phonon-cooling or diffusion-cooling. Electrons give their energy to the crystal lattice in phonon-cooled systems, and recombine before leaving the meander or bridge [4]. The characteristic recombination time is less than the time the electrons spend in the structure. Diffusion-cooled devices are precisely the opposite: the electrons do not have the time to recombine in the sensor area, and are diffused out in to the larger system. The recombination characteristic time is a material property, and obviously superconductors with lower recombination times are desired. For photon counting, phonon-cooled systems generally have faster response times, meaning thinner films can improve performance [10].

Superconducting photon counters can used in fibre optic telecommunications in the near future [3]. Such sensitive detectors are also required in quantum computing and quantum cryptography. The high counting speeds, low dark counts, and low littering time allow for faster data transmission. The effective detection area is a cause for concern [11], and great care must be made to calibrate the detectors to account for photons events which go undetected due to the quantum efficiency or space between the wire bends.

Area-based photon counters area also used to detect photons. These devices are often square patches of thin metal (often tungsten, molybdenum, or titanium) cooled down to the order of hundreds of mK. They are biased so that they are just below the T_c of the detector metal. When a photon impinges on the surface, a pulse is created, just as for the meander detectors. The lower T_c (and system temperature) allows for detection of photon energies on the order of tenths of an eV. These highly sensitive devices are used to detect photons in the infrared to ultraviolet spectrum [12], and even at X-ray energies [13].

III. Antenna-coupled Detectors

A. Superconductor-Insulator-Superconductor Mixers

Superconductor-insulator-superconductor junction detectors (hereafter called SIS mixers) use the nonlinearities of the Josephson effect [14] to mix two signal frequencies together. Photons interact with the electrons in the junction and allow them to tunnel through the insulating barrier [15]. Antennas can be coupled with the niobium junctions, allowing detection of signals limited by:

$$F_{gap,Nb} = 2\Delta_{Nb}/h \tag{1}$$

where Δ is the superconducting gap energy and h is Planck's constant. This sets a physical upper frequency limit, based on the superconductor used in the junction, which must be considered before any detector is designed. For niobium-based SIS junctions, the upper frequency threshold is approximately 1.4 THz. Frequencies greater than those calculated by 1 will break the Cooper pairs inside a junction and lead to signal attenuation at the desired intermediate frequency.

Continued interest in SIS junctions is maintained by their most impressive quality: the sensitivity of SIS systems is quantum limited [4], making them currently the most sensitive terahertz detectors design. Fabrication of junctions based on high temperature superconductor (HTS) materials may increase use of SIS mixers due to many having higher superconducting gap energies than niobium [4].

B. Hot-electron Bolometer Mixers

Hot-electron bolometers (HEBs) operate in a similar weak link manner as do nanowire photon detectors. The difference is that meander lines are typically used to count photons by the voltage pulses they create, while HEBs use the nonlinear resistance of materials while transitioning between the normal and superconducting state. Accordingly, the output is a frequency spectrum instead of pulses signifying photon

impingement. The electron recombination physics described for photon detectors apply to HEBs too.

Hot-electron bolometers are used currently in astronomy to detect faint high frequency astronomical signals, or in frequency bands undetectable by previous technology. In theory, any two wavelengths can be heterodyned by an HEB mixer, provided the correct dimensions and system qualities (such as signal-to-noise ratio) are satisfied [4]. A mixed intermediate frequency can then be amplified and measured, with the frequency spectrum being of prime importance.

The data gathered from HEBs can be used to generate images based on material density or even chemical composition [1]. If a frequency spectrum is analyzed, absorption lines can be indicators of certain molecules absorbing radiation. In the THz spectrum, there are many such absorption lines, including many explosives. Thus objects can be identified by density and material composition. This can be done using passive radiation or by actively scanning an object with radiation either at a constant frequency or sweeping through some frequency band.

The physics of the hot electron bolometer is similar to single photon detectors in that the transition from superconductivity to normal conductivity is used to observe the phenomenon of interest. The difference is that photon detection is interested in the generation of pulses (a sign of a photon collision event) whereas hot electron bolometers use the non-linear change of resistance that occurs during the transition between normal and superconducting states [4]. This can be seen in Fig. 3 when looking at the bridge region. This non-linear transition allows the mixing of the source signal and local oscillator. In the case where the source and local oscillator are driving the bolometer at frequencies in the THz regime, often only the difference[6] can be measured or is present at all[7]. All the required information is contained in the intermediate frequency, provided there is a spectrally stable local oscillator at a frequency known to be close to the source frequency.

.

Hot electron bolometers can be coupled with narrow band (Fig. 4) or broadband antennas [1], like the equiangular spiral antenna, seen in Fig. 5. The application dictates the required bandwidth while the source dictates the optimal polarization of the antenna. Linearly polarized, narrow band antennas are used most often for imaging, when transmission of a single-frequency signal through an object is all that is needed to form an image. Circularly polarized, broadband antennas are used when the source is polarized in a random or unknown manner and the spectral absorption of a wide frequency band is of

[6]Source frequency minus local oscillator frequency.

[7]The driving frequencies can be so high that the electron vibrations cannot keep up [4], effectively filtering the system.

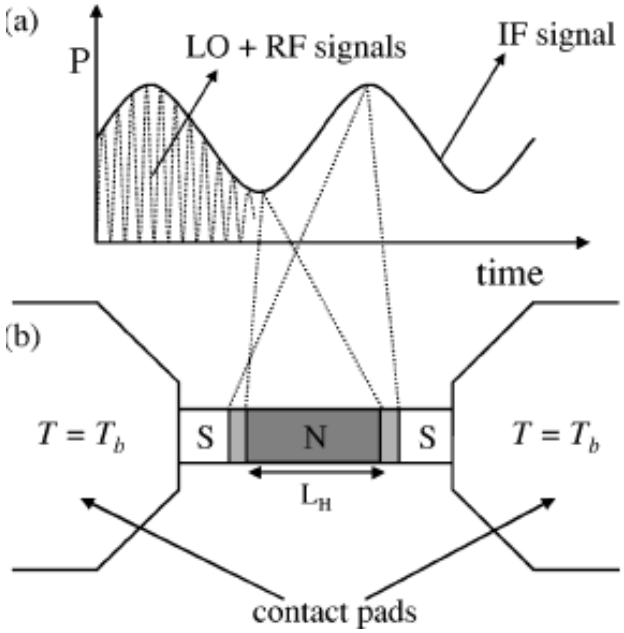


Fig. 3: The image in (a) shows how the power from two signals can be mixed by the action of nonlinear resistance of the bridge in (b). One model (from [5]) has the area of bridge that is normal metal (N) oscillating at the intermediate frequency. The section labeled S remains in the superconducting state.

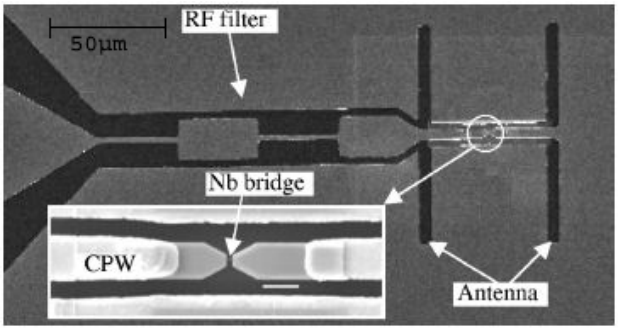


Fig. 4: A twin-slot antenna. The inset shows the superconducting bridge in greater detail.

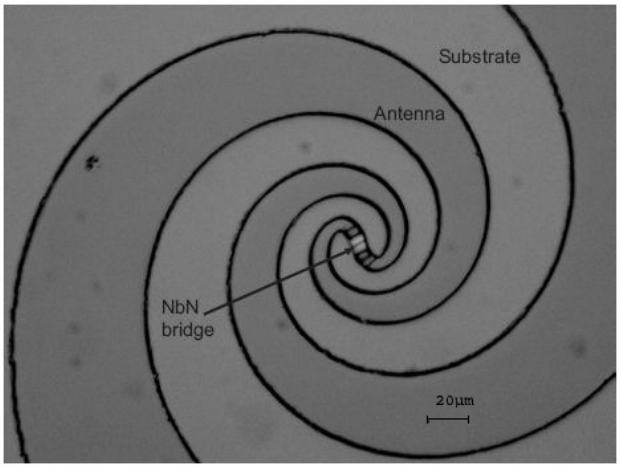


Fig. 5: An equiangular spiral antenna. The superconducting bridge connects the two antenna arms. Photo from [4]

interest.

HEBs are currently being made in arrays for both astronomy and astrophysics research and security imaging [16]. Since most of the energy radiated in the universe is theorized to be in the THz regime[8], further development will help us learn more about the universe.

IV. Manufacturing and Experimentation

A. Common Manufacturing Procedures

High quality thin film superconducting layers are required for both the photon detectors and the antenna-coupled mixers. Typically, Type I superconductors are used since most current applications are in space and laboratory-based physics experiments [18], where cooling to 10 K is not an issue. At the time these devices were first developed, Type I superconductor films could be created reliably and helium cooling environments were readily available.

Films are patterned using standard photolithography methods as well as electron beam lithography [9]. The latter is often required to create structures on the order of 10 to 100 nanometers[9]. The complexity of construction is one area where improvements are being made [19], a necessity if these devices are ever to leave the laboratory setting. The manufacturing processes that can optimize the sensitivity of bolometers, nanowires, and SIS mixers, such as annealing [20], are still being developed.

Antenna structures are often made from deposited gold that is shaped by a lift off procedure [18]. Care must be taken to design proper wave guide structures that connect an antenna to the bridge or junction. The lift off techniques and gold deposition methods should create smooth gold structures with sharp, well-defined features [17]. The contact between the gold and superconductor benefits greatly with an intermediate material, often a thin film of niobium-titanium nitride (if the device is niobium-based) or pure titanium [17], which increases the contact and signal strength from the antenna to the superconductor.

For the creation of basic images, narrow band THz radiation sources are used, such as Gunn diodes, Backward Wave Oscillators, and Quantum Cascade Lasers. Gunn diodes can be used with frequency multipliers produce signals oscillating at frequencies approaching 1 THz [1], but the transmitted power is often low relative to other sources. Gunn diodes are used as both local oscillators [21] and active illumination sources. Backward Wave Oscillators can be used [1], but such systems are currently large and expensive. Quantum Cascade Lasers (QCL) are an excellent technology for tunable local oscillator sources of terahertz imaging systems [4]. The QCL is a relatively new technology currently used only in laboratory experiments [22]. Heated elements can be used as a broadband source [1].

[8]The Cosmic Microwave Background [17].

[9]This refers to wire width. Wire lengths can be several microns long.

B. Experimental Procedures

With all superconducting EM detectors, great care must be taken to ensure the sensitive devices are not saturated by unwanted environmental noise, including magnetic flux. Depending on the energy of the radiation of interest, either one of the following are currently required for most superconducting EM radiation detector: either cooling to below liquid helium temperatures or providing an illuminating source of known frequency and intensity [1]. Thermal and current stability or predictability is of utmost importance. These devices are operated near T_c or J_c; any inability to control the system bath temperature or biasing current will result in high noise levels and perhaps cessation of operation.

For optical photon counters, the detector areas must be enclosed in a dark environment and often excited by attenuated signals. Optical fibre is often used to help isolate the system from unwanted light sources [3]. Attenuators can also be used to mask a window into the cold system and detector. In order to increase the quantum efficiency and partially account for the gaps between the meander wires, current designs place a quarter wavelength resonator cavity (seen in fig. 6) over the detectors and have impinging radiation reach the device through the substrate [2].

Meander-based detectors can be biased to be on the edge of the superconducting T_c or J_c. When either of these values are exceeded, a signal pulse is created. Better performance

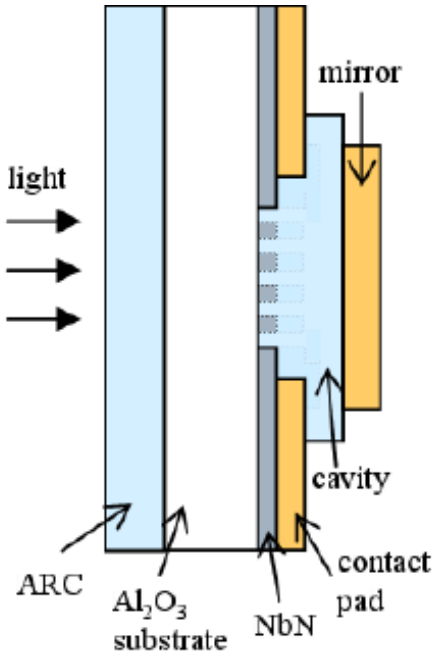


Fig. 6: A cross-sectional depiction of a single photon counter. A quarter wavelength resonator cavity captures more of the signal radiation within a dielectric between the meander structure and a metal with high reflectance at a wavelength of interest [11]. The anti-reflection coating (ARC) allows more light to pass into the cavity.

has been noted on devices biased as close to J_c as possible without the system noise fluctuations triggering a pulse. Area-based photon detectors are cooled down to just below the T_c of the metal used. Such detectors are metals such as titanium, so cooling the system down to the mK range is necessary. Meander lines are most commonly created with niobium or niobium nitride, and so are cooled between 5 to 10 K.

Both types of photon counters use low-noise amplifiers (LNAs) to increase the signals to be fed into a pulse counting circuit. other circuit elements like DC blocks and signal discriminators are also used to improve performance and signal-to-noise ratio [11].

SIS mixers and hot-electron bolometer mixers use the same experimental procedures (only different in the way the detector is biased). Since many objects will emit terahertz EM radiation even below 10 K, it is often required to cool an operating device to helium temperatures or to illuminate an object with some known and characterized source and ignore the passive radiation entirely [1]. A schematic showing the operation of a superconducting mixing detector is shown in fig. 7. The IF signal from the HEB is fed through the chip and impedance matched to a coaxial cable (usually 50 Ω) and passed through several stages of amplifiers. In fig. 7, a yttrium iron garnet (YIG) filter is used to isolate the IF signal to be measured by a power meter or spectrum analyzer.

V. Conclusions

Superconductor-based sensors offer many advantages to many fields, in both research and commercial applications [1], [3]. Material and system properties allow for sensitive devices with low noise levels. Nanowire and nanobridge (or microbridge) structures operating near T_c or J_c can be used to detect electromagnetic radiation or as a nonlinear resistance to perform heterodyning on two input waveforms. SIS junctions can be used to detect electromagnetic radiation with quantum-limited sensitivity. All of these technologies are being utilized in space-based and terrestrial astrological observatories as well as small-scale imaging devices for security and medical science [23] (see Fig. 8).

Improvements will come not only from the reduction of device sizes, but from the materials used. Systems constructed using HTS materials [25] or cryocoolers will allow more usage outside laboratories and commercialization of superconducting electromagnetic sensors and imagers.

At Stellenbosch University, it is possible to create and test HEB devices using available equipment and processes. Without a narrow band THz source, antennas must be made to resonate to signals available from high-speed semiconductor electronics for imaging to be practical. An HEB system designed to resonate at 60 GHz is currently being designed and manufactured. Using a heated element as a blackbody radiator (such as a heated silicon carbide rod), it is possible to use an HEB as part of a THz spectrometer [26]. Meander-based photon counters can also be manufactured, but a reliable electron beam lithography technique has not been demonstrated with the electron microscope in the laboratory. Any meander will be defined by photolithography techniques and thus will have wider gaps than the wires described in this paper. Area-based photon counters can be created, but the cryocooler systems used for testing cannot cool to the mK temperatures required for optimal performance and sensitivity. Josephson junctions have been fabricated at the university, but these were weak-link type devices. It may be of interest to experiment if this type of Josephson junction can reproduce the performance of an SIS device. Theoretically, it is possible

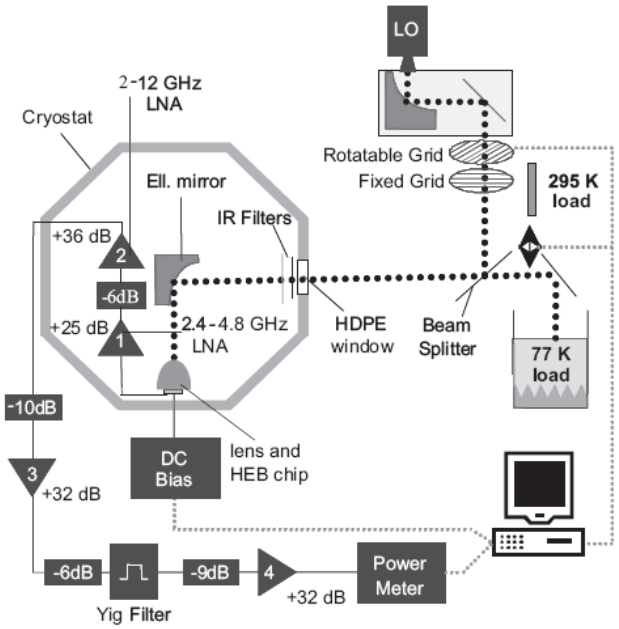


Fig. 7: A schematic of an antenna-coupled mixer experiment used in [4].

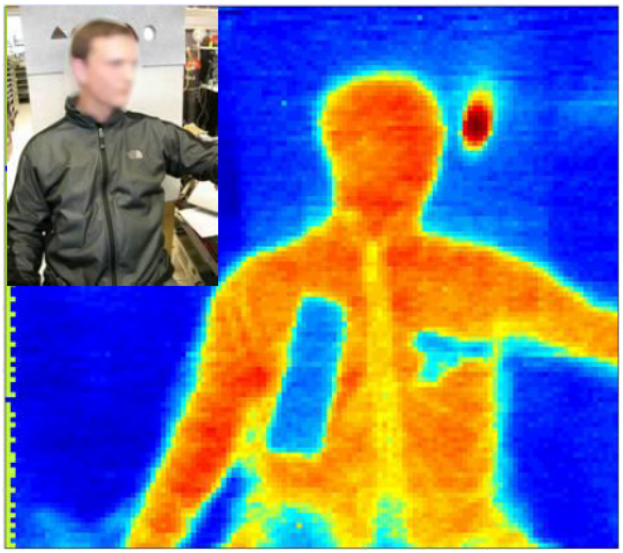


Fig. 8: An image taken of a man (inset) carrying a gun. A single pixel HEB mixer was used to create the image [24].

to manufacture an SIS Josephson junction, and subsequently an SIS heterodyne mixer, with the equipment and materials available at the university.

REFERENCES

[1] H. Liu et al, "Terahertz spectroscopy and imaging for defense and security applications," in *Proceedings of the IEEE*, vol. 95, no. 8, Aug. 2007, pp. 1514 – 1527.

[2] G. Gol'tsman et al, "Middle-infrared to visible-light ultrafast superconducting single-photon detectors," vol. 17, no. 2, pp. 246 – 251, Jun. 2007.

[3] I. Milostnaya et al, "Superconducting single-photon detectors designed for operation at 1.55-μm telecommunication wavelength," in *Journal of Physics: Conference Series; 7th European Conference on Applied Superconductivity*, vol. 43, no. 1, Jun. 2006, pp. 1334 – 1337.

[4] M. Hajenius, *Terahertz heterodyne mixing with a hot electron bolometer and a quantum cascade laser*. Delft University of Technology: Ponsen & Looijen, 2006, ISBN-10: 90-8593-024-3.

[5] D. Floet et al, "Hotspot mixing: A framework for heterodyne mixing in puperconducting hot-electron bolometers," vol. 74, no. 3, pp. 432 – 435, Jan. 1999.

[6] M. Zgirski et al, "Size dependent breakdown of superconductivity in ultranarrow nanowires," vol. 5, no. 6, pp. 1029 – 1033, May 2005.

[7] H. Merkel et al, "A two-dimensional hot-spot mixer model for phonon-cooled hot electron bolometers," vol. 11, no. 1, pp. 179 – 182, Mar. 2001.

[8] M. Khafizov et al, "Mechanism of light detection in current-biased superconducting mgb2 microbridges," vol. 17, no. 2, pp. 2867 – 2870, Jun. 2007.

[9] A.D. Semenov et al, "An energy-resolving superconducting nanowire photon counter," vol. 20, no. 1, pp. 919 – 924, Aug. 2007.

[10] M. Bell et al, "Photon number-resolved detection with sequentially connected nanowires," vol. 17, no. 2, pp. 289 – 292, Jun. 2007.

[11] K. Rosfjord et al, "Nanowire single-photon detector with an integrated optical cavity and anti-reflection coating," vol. 14, no. 2, pp. 527 – 534, Jan. 2006.

[12] B. Cabrera et al, "Detection of single infrared, optical, and ultraviolet photons using superconducting transition edge sensors," vol. 73, p. 735, Aug. 1998.

[13] K.D. Irwin et al, "X-ray detection using a superconducting transition-edge sensor microcalorimeter with electrothermal feedback," vol. 69, p. 1945, Sep. 1996.

[14] T. P. Orlando and K. A. Delin, *Foundations of Applied Superconductivity*. Massachusetts Institute of Technology: Addison-Wesley Publishing Company, 1991, ISBN 0-201-18323-4.

[15] J. Zmuidzinas, "Coherent detection and sis mixers," in *Proceedings of the Far-IR, sub-MM, and MM Detector Technology Workshop*, Montery, CA, Apr. 2002, pp. 149 – 154.

[16] M.O. Reese et al, "Niobium hot electron bolometer development for a submillimeter heterodyne array camera," vol. 17, no. 2, pp. 403 – 406, Jun. 2007.

[17] M.I. Finkel et al, "Terahertz heterodyne receivers based on superconductive hot-electron bolometer mixers," vol. 48, no. 10-11, pp. 859 – 864, Jan. 2005.

[18] M. Hajenius et al, "Low noise nbn superconducting hot electron bolometer mixers at 1.9 and 2.5 thz," vol. 17, no. 1, pp. S224 – S228, Mar. 2004.

[19] J.K.W. Yang et al, "Fabrication development for nanowire ghz-counting-rate single-photon detectors," vol. 15, no. 2, pp. 626 – 630, Jun. 2005.

[20] M. Hajenius et al, "Optimized sensitivity of nbn hot electron bolometer mixers by annealing," vol. 17, no. 2, pp. 399 – 402, Jun. 2007.

[21] J. Kawamura et al, "Terahertz-frequency waveguide nbn hot-electron bolometer mixer," vol. 11, no. 1, pp. 952 – 954, Mar. 2001.

[22] J.R. Gao et al, "Terahertz heterodyne receiver based on a quantum cascade laser and a superconducting bolometer," vol. 86, no. 1, pp. 1 – 3, Jun. 2005.

[23] W. Withayachumnankul et al, "T-ray sensing and imaging," in *Proceedings of the IEEE*, vol. 95, no. 8, Aug. 2007, pp. 1528 – 1558.

[24] A. Luukanen, "Bolometers and terahertz imaging," in *Microsensing Seminar*, Mar. 2006, p. 25.

[25] T. Uchida et al, "If properties of ybco hot-electron bolometer mixers for millimeter-wave radiation," vol. 372, no. 1, pp. 387 – 390, Aug. 2002.

[26] D. Santavicca et al, "A far-infrared fourier transform spectrometer with an antenna-coupled niobium bolometer," in *Superconductor Science and Technology*, vol. 20, no. 11, Nov. 2007, pp. S398 – S402.

Page left blank intentionally.

PBCO/YBCO bilayer growth and optimization for the fabrication of buffered step-edge Josephson Junctions

W.F. van Staden, U. Büttner, C.J. Fourie, V.V. Srinivasu, G. L Hardie and W.J. Perold

Abstract— **Bilayers of $PrBa_2Cu_3O_{7-\delta}$ and $YBa_2Cu_3O_{7-\delta}$ are grown epitaxially on MgO substrates using Pulsed Laser Deposition. We discuss the entire optimization process in detail, giving quantitative parameter values. Film characterization included X-ray diffraction (XRD), Atomic Force Microscopy (AFM) and susceptance tests. The optimal process yielded bilayer structures which can be utilized in the fabrication of novel buffered step-edge Josephson junctions.**

Index Terms—**bilayer, PLD, step-edge junction, XRD, AFM**

I. Introduction

$YBa_2Cu_3O_{7-\delta}$ (YBCO) thin film based Josephson junctions are now routinely used in various superconducting electronic devices [1-3]. The deposition techniques for YBCO thin film growth are very well optimized, enabling such junctions to be operated at liquid nitrogen temperatures [4-6]. Of the several junction topologies available, step-edge junctions (SEJ's) are very promising because they can be easily integrated into small and large scale electronic circuits [7]. Optimization of the step-fabrication parameters remains an intriguing research field, which in turn can produce more easily manufacturable SEJ's with improved quality.

The manufacturing process of a step-edge junction involves etching of a step-edge into a suitable substrate (like MgO) before the epitaxial growth of YBCO over the step-edge template. Etch rates of oxide substrates like MgO are very slow as compared to that of typical photoresists used. To overcome this substrate's slow etch rate problem, we proposed a buffered step-edge [8] structure. Essentially a buffer layer which is crystallographically compatible with the YBCO, and which possesses a faster etch rate than the substrates used for film deposition, was used. An ideal buffer layer was found by us in $PrBa_2Cu_3O_{7-\delta}$ (PBCO) [8].

PBCO has a crystallographic structure almost identical to $YBa_2Cu_3O_{7-\delta}$ (YBCO) with comparable values of the lattice parameters. Furthermore, PBCO is not metallic; its resistivity is high at room temperature and increases at low temperature (semiconducting behaviour) without any trace of a transition to a superconducting state. These properties make this material ideal for a buffer layer. The etch rate of PBCO should also be quite similar to that of YBCO, which is much higher than that of oxide-substrates.

Possible problems that can occur when a PBCO buffer layer is used, are Pr/Y interdiffusion, and charge transfer from the YBCO to the PBCO layer. Interdiffusion of Pr may lead to doping at Y-site and depression of T_c. For example a 35% and 50% doping of Pr at Y-site can depress T_c to ~ 50K and 20K respectively [9,10] for the resulting alloy. Charge transfer from the YBCO to the PBCO layer would also result in a decrease of the transition temperature [11]. Fortunately, these effects should not have a significant effect when a reasonably thick YBCO layer is deposited over the PBCO. However, one may encounter several problems during the course of the growth of these bilayers. The thickness of the PBCO/YBCO superlattice could possibly approach critical values where microcracks can start to become a factor. The added surface roughness introduced by using a buffer layer will possibly lead to structural defects in the superconducting YBCO layer, reducing its current-carrying abilities and critical temperature. For this reason, it was necessary to pay special attention to the epitaxy and surface quality of the deposited PBCO layer. In this paper we give complete optimization details of the thin film growth of PBCO/YBCO bilayer structures.

II. PBCO Fabrication

A. Pulsed Laser Deposition

A Pulsed Laser Deposition (PLD) system equipped with a pulsed oxygen valve was used to deposit the PBCO layers onto the MgO substrates. For this purpose, a planar $PrBa_2Cu_3O_{7-\delta}$ target was created using the standard solid state route [8]. A comprehensive study on the deposition conditions of PBCO on MgO substrates could not be found in literature. Consequently, it seemed reasonable to use the deposition parameters similar to those used in a standard YBCO deposition. Optimization of the deposition parameters was restricted to temperature variation due to its dominating effect on surface quality. The temperature variation was over the temperature range 700-800°C. Each deposition was performed for 15 minutes, corresponding to a layer thickness of around 250 nm. Characterization of six buffer layers was done through XRD and AFM analysis.

B. Film Optimization

XRD results confirmed that c-axis PBCO favours deposition at lower temperatures than YBCO. The layer deposited at 700°C showed excellent c-axis crystallinity. From Fig. 1 we can clearly see that prominent (00l) peaks

W. F. van Staden was with the University of Stellenbosch, South Africa. He is now with Azoteq, 109 Main Street, Paarl, South Africa (telephone: +27-21-863-0033).

U. Büttner, C. J. Fourie, G. L. Hardie and W.J. Perold are with the Department of Electrical and Electronic Engineering, University of Stellenobosch, Private Bag X1, Matieland, 7602, South Africa (telephone: +27-21-808-4990; fax: +27-21-808-4981).

V. V. Srinivasu is with the National Center for Nanostructured Materials and Mechatronics and MicroManufacturing, CSIR, Pretoria, South Africa; e-mail: svallabhapurapu@csir.co.za.

without any prominent secondary phases or a-axis growth being present.

As the deposition temperature was raised, the introduction of unwanted peaks was observed. This included a prominent (110) peak at 32.7° , others at 60° and 68.12° (see Fig.2).

AFM analysis confirmed the presence of droplets characteristic to the PLD process. A definite difference in growth modes was observed for higher temperatures. It would seem that the presence of edge dislocations are more prominent at higher temperature, which coincides with the rougher surface obtained. This is clearly illustrated in Fig. 3.

The results for the measured AFM roughness are summarized in Table I.

A surface roughness of only 7 nm was obtained for the film deposited at 700°C. This deposition temperature was therefore chosen as optimal from both a crystallinity and surface quality perspective (see Table II for optimal parameter set). A 25° step edge is then milled on the PBCO film and annealed as discussed elsewhere [8]. Over this PBCO step-edge template an YBCO thin film is deposited.

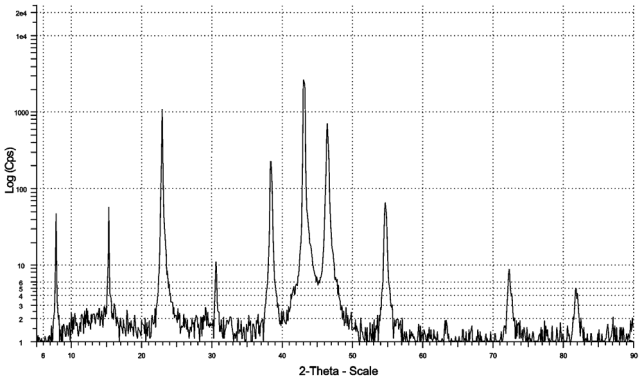


Fig. 1. XRD analysis of a PBCO thin film deposited at 700°C.

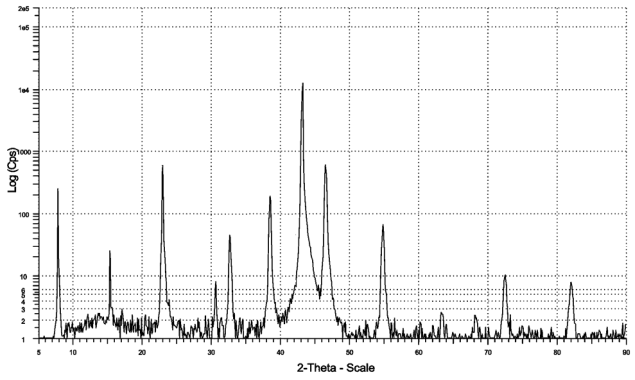
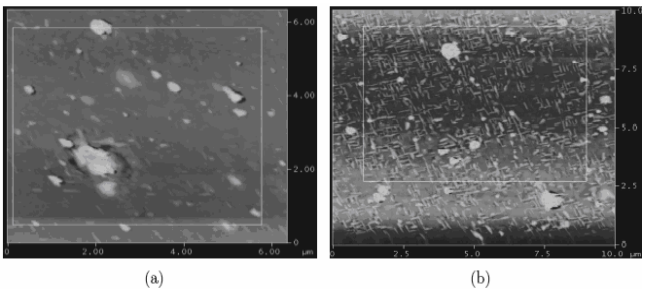


Fig. 2. XRD analysis of a PBCO thin film deposited at 720°C.



(a) (b)

Fig. 3. AFM analysis of PBCO thin films deposited at (a) 700°C and (b) 720°C.

III. YBCO Thin Film Deposition

After the step was annealed, a superconducting YBCO layer was deposited by pulsed reactive crossed beam laser ablation [8] [12]. This deposited thin film has to possess certain desirable qualities. These include a T_c as close to 90 K
as possible, a small transition temperature range (ΔT), good surface roughness and pure c-axis growth. To ensure proper step coverage and to minimize chip-to-chip parameter drifts, the film thickness should also be uniform over the substrate. PLD is, however, renowned for producing small-area uniformity, necessitating very good alignment of the plume and substrate. Magnetron sputtering was not used for step-edge film coverage due to reports that the growth mechanisms can be unpredictable and not suited for producing step-edge junctions [13]. The unavailability of Transmission Electron Microscopy (TEM) analysis would also have limited characterizing such a process.

To realize an operational Josephson junction, the thickness of the deposited film should also be monitored very carefully. If the film is too thick, the possibility exists that the step-edge would become too rounded. In such a case, the grain boundary close to the step-edge is shunted by a superconducting channel which dominates current transport properties. For this reason, all films thicknesses were about 70% of the step height. The deposition

TABLE I
MEASURED SURFACE ROUGHNESS FOR 6 PBCO DEPOSITIONS

T_{DEP} [°C]	Optimal Value [nm]
700	7.88
720	11.85
740	12.27
760	26.52
780	33.12
800	31.34

TABLE II
OPTIMAL PARAMETER SET FOR PBCO DEPOSITION

Process Parameter	Optimal Value
T_{dep}	700 °C
Laser frequency	16 Hz
Laser fluence	$5x10^{-2}$ mbar
P_{O2}	4.09 J/cm^2
Working distance	60 mm
Focal Distance	15 mm
Deposition Time	15 min
T_{anneal}	30 min

TABLE III
OPTIMAL PARAMETER SET FOR YBCO DEPOSITION

Process Parameter	Optimal Value
T_{dep}	740 °C
Laser frequency	16 Hz
Laser fluence	$5x10^{-2}$ mbar
P_{O2}	4.09 J/cm^2
Working distance	60 mm
Focal Distance	15 mm
Deposition Time	11 min
T_{anneal}	30 min

conditions for YBCO (as optimized for deposition on MgO substrates) are listed in Table III.

Using the optimization conditions as shown in Table 3, an YBCO film was deposited on the PBCO step-edge template. AC susceptance of this structure was then measured.

IV. Susceptibility Tests

To characterize the superconducting properties of the PBCO/YBCO bilayer structure, a susceptibility test setup was used. Such a test yields the critical temperature, T_C, and illustrate the transition into superconductivity (ΔT) of the thin film. This test is commonly replaced with 4-point probe measurements in literature. The 4-point probe measurement gives the dc resistance of the thin film but requires the deposition of gold pads and wire bonding. The susceptibility test setup does not require prior sample preparation and yields results that are more representative of the entire film. The test setup consists of a cold finger, a two-stage (20 K) cryocooler and a temperature controller. The sample is mounted between two planar coils situated on the cold finger. The primary coil is driven by a 1 MHz sinusoidal waveform, which induces a magnetic field perpendicular to film surface. The secondary, or pickup coil, returns a signal to the controller indicating the percentage signal that is sensed. Below the critical temperature of the sample, it expels magnetic fields according to the Meissner principle. During the process of cooling down the cold finger, the temperature and corresponding percentage susceptibility measurements are sampled and sent serially to a computer where the results are displayed.

The susceptibility measurement on this PBCO/YBCO structure, as shown in Fig. 4, revealed a T_C of about 85 K, but a very poor transient profile with ΔT = 20 K. This is an indication of moderate uniformity in the film quality. Despite this, the film quality is still sufficient to manufacture a Josephson junction.

Such a superconducting PBCO buffered YBCO step-edge junction exhibiting Josephson characteristics has been successfully fabricated using the optimized bilayer structures discussed here [8]. Though there are reports on the usage of PBCO layer along with YBCO to make Josephson junctions [14], they are different class of junctions, as compared to our actual step-edge with PBCO as a buffer step template.

V. Conclusion

The novel idea to use an insulative PBCO buffer layer with a faster etch rate as compared to that of MgO, as a template for a step-edge, was attempted. This required the growth of a PBCO thin film and establishing appropriate PLD deposition parameters for this new process. First we have optimized the PLD growth of PBCO layer. After fabricating the PBCO step-edge, YBCO is grown epitaxially on it. Optimization of the entire process was discussed in detail, and we presented a superconducting YBCO layer growth on the PBCO step-edge template which is suitable for actual fabrication of Josephson Junctions.

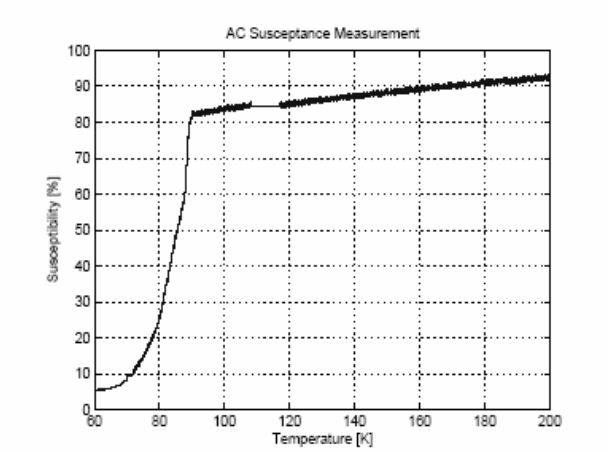


Fig. 4. Susceptance measurement of YBCO film on PBCO step.

References

[1] E. Sarnelli, G. Testa, D. Crimaldi, A. Monaco and M.A. Navacerrada, Supercond.Sci. Technol., vol.18, L35, 2005.

[2] J.H. Park, J.H. Kim, G.Y. Sung, C.H. Kim, K.R. Jung, J.H. Kang and T.S. Hahn, IEEE Tran.Appl. Supercond. VolII, 932, 2001.

[3] D.He, M.Yoshizawa, H. Yashiro and M. Nakamura, IEEE Trans.Appl. Supercond. Vol.15, 41, 2005.

[4] Alex. I. Braginski, Supercond. Sci. Technol. Vol16, 1315, 2003.

[5] M. Tonouchi, H. Shimakage, Z. Wang, S. Tomozawa, M. Hangyo and Y. Murakami, Supercond. Sci. Technol. Vol.9, A161, 1996.

[6] W. Prusseit, B. Utz, P Berberichi and H. Kinder, J. Supercond.vol.7, 231, 1994.

[7] W. Reuter, M. Siegel and K. Herrmann, Appl. Phys. Lett. Vol.62, 2280, (1993).

[8] W. van Staden, "The Fabrication of PBCO Buffered Step-edge Josephson Junctions" *MScEng thesis*, Stellenbosch University, South Africa, 2007.

[9] M.S. Hegde, Boben Thomas, N.Y. Vasanthacharya, S.V. Bhat, V.V. Srinivasu and N. Kumar, Phys.Rev. B. vol.48, 6465, 1993.

[10] J.-M. Triscone, O. Fischer, O. Brunner, L. Antognazza, A.D. Kent, M.G. Karkut, Phys.Rev. Lett. Vol.64, 804, 1990.

[11] H.H. Wang, R.P. Wang, and Y.L. Zhou, "Growth and characterisation of PBCO thin film used as template layer on STO substrate," *Journal of Superconductivity*, vol. 13, no. 3, pp. 447–451, 2000.

[12] A. Gupta and B.W. Hussey, "Laser deposition of YBCO film using a pulsed oxygen source," *Applied Physics Letters*, vol. 58, no. 11, pp. 1211–1213, 1991.

[13] N. Khare, *Handbook of high-temperature superconductor electronics.* Marcel Dekker, Inc., 2003.

[14] Y. Sawada, H. Terai, A. Fujimaki, Y. Takai and H. Hayakawa, IEEE Tran. Appl. Supercond. Vol.5, 2099, 1995.

A 384x288 pixel CMOS image readout chip

Trudi-Heleen Joubert, Monuko du Plessis, *Senior Member, IEEE*

***Abstract*— In the design of imaging arrays, high resolution has become a major requirement in many applications, leading to large arrays and small pixel sizes, and also constrains power dissipation. The design of a characterized image readout chip for an array of 384x288 image pixel currents with a pixel pitch of 25µm is presented here. In this readout chip, a simple pixel circuit is used to satisfy both pixel size and power dissipation limitations. Additionally, a novel power saving scheme is discussed. The integrated circuit, with more than 2 million transistors and a maximum power dissipation of 65mW, was manufactured in a four-metal double-poly 0.35µm CMOS technology.**

***Index Terms*— CMOS readout circuit, focal plane processor, infrared imaging, power saving.**

I. Introduction

Historically electronic imaging systems have been manufactured in various technologies, starting in 1926 as imaging cathode ray tubes. After the advent of silicon planar technology in 1960, integrated photodiodes were combined with integrated amplification circuitry. The first self-scanned solid-state MOS photodiode array was introduced in 1967. In 1970 a variant of silicon MOS technology emerged that was especially suitable for image sensing: the well-known CCD technology. The dramatic quality improvement compared to the early MOS arrays made CCD the technology of choice for nearly 3 decades [1].

Recently however, the advances in the ubiquitous microelectronics CMOS process technology, and also detector fabrication technology, are leading to the development of large format arrays of infrared (IR) imaging detector systems of acceptable quality at lower cost, pushing CCD out of all low-end markets and increasingly pushing toward the high-end market. This enables applications such as driver aid, aircraft aid, industrial process monitoring, community services, firefighting, law enforcement, search and rescue, portable mine detection, border surveillance, night vision, as well as various military applications [2].

In this work the interest is in circuit techniques based on silicon CMOS VLSI technology for the inherent advantages of packing density, low cost, flexibility, and good feasibility of system integration, although BiCMOS technology would more readily achieve low noise requirements.

The term focal plane array (FPA) refers to an assemblage of individual picture elements (pixels) located at the focal plane of an imaging system. The optics focus the image onto the detector array, and in a staring array these detectors are scanned electronically with readout integrated circuits (ROIC's), often bonded to the array using hybrid integration techniques such as Indium bump-bonding, or even monolithically integrated with the array [2], [3]. The readout circuitry is also known as the focal plane processor (FPP) in the context of FPA design.

The design of readout circuit electronics is a major part of the development of IR detector FPA's. They must support a good interface between the IR detectors and the following signal processing stage. Functions include the biasing of the detectors, extraction of the detector signal (or injection of the signal to the readout circuit), signal integration and amplification, multiplexing signals from an array, and driving the chip output. Smart sensors may include yet other functions such as pixel substitution, pixel averaging, edge detection or image recognition [3]. Operational requirements include parameters like array size and pitch, bias control, injection efficiency, noise, dynamic range, readout rate, integration time, storage capacity and power dissipation, which all impact the design trade-offs.

In most readout configurations the detector current is linearly extracted, amplified and integrated to an output voltage. High injection efficiency and wide bandwidth lead to good responsivity and readout performance. To achieve these, the input impedance of the interface circuit should be lower than the shunt resistance of the detector. The value of the integrating capacitor and the levels of background and dark current determine the charge storage capacity. The dynamic range is determined by the noise level, storage capacity, and the linearity. Readout rate is limited by circuit operation speed and chip power dissipation. Frame rate is defined by the readout rate and the maximum integration time.

Some simple readout structures like source-follower per detector (SFD), direct injection (DI) and gate-modulation input (GMI) are still commonly used in large staring arrays because of the small pixel size and low power consumption obtainable. On the other hand, more complex circuit techniques like buffered direct injection (BDI) and capacitive transimpedance amplifier (CTIA) have been developed to provide excellent bias control, high injection efficiency,

Manuscript received March 2, 2009. This work was supported by Denel Aerospace Sysytems.

T-H. Joubert is with Solid State Technology (SST), Pretoria, South Africa (corresponding author - phone: 082 449 6209; e-mail: tjoubert@solidstatetechnology.co.za).

M. du Plessis is with the Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa (e-mail: mplessis@postino.up.ac.za).

linearity and noise performance. Recently, readout structures like shared buffered direct injection (SBDI), switched-current integration (SCI), and buffered gate modulation input (BGMI) have been proposed to achieve better performance under pixel size limitation [3].

In typical readout circuits, the background current is amplified and integrated directly along with the signal current, requiring a large integration capacitance to avoid saturation and restricting the dynamic range. Gain variations, along with noise, also determine the dynamic range achievable without gain compensation.

Temporal and pattern noise are two main types of noise in IR imaging systems. The temporal noise sources include shot, thermal, flicker (1/f), generation-recombination, and switching (KTC) noise which are contributed by detectors and readout circuit. Usually the temporal noise can be reduced by detector technology, operational conditions, circuit techniques and system arrangement, whereas fixed pattern noise (FPN) can be reduced by calibration techniques. Some general strategies of noise reduction for readout circuits are correlated double sampling (CDS), multiple correlated sampling (MCS), and chopper stabilized input (CSI) [3].

Non-uniformity in detector arrays is an important matter since offsets caused by detector variation lead to fixed-pattern noise (FPN). Offsets in the readout electronics also contribute to the FPN. Choppers and shutters operate on system level to reduce the FPN, while electronic offset correction and compensation methods are used also often employed.

In summary, the pixel pitch of the IR FPA is reduced with increasing array size and resolution. Moreover, the total power dissipation of IR FPA's is limited by typical imaging systems. These two forces usually constrain the pixel design size and complexity, requiring a trade-off with circuit performance. Non-uniformity across a large array also requires consideration. This paper contains a discussion of how these issues are addressed in the design of a 384x288 pixel CMOS image readout chip.

II. Design of the Readout Chip

A. Architecture of the readout chip

The floorplan of the chip is shown in Fig. 1 and directly depicts the architecture of the readout chip. The FPP is a mixed-mode design with an analog pixel array, analog integrators, and analog amplifiers, all controlled by the digital circuitry.

The detectors in the array are biased with a common substrate voltage VSUB. The detector currents are integrated into an array of pixel cells and then multiplexed out. The integration period for the array is defined by the relative timing of the reset cycle and an externally provided integration logic signal. This allows for great flexibility in integration time, with only clock period as resolution limitation, since the integration signal is synchronized to the master clock.

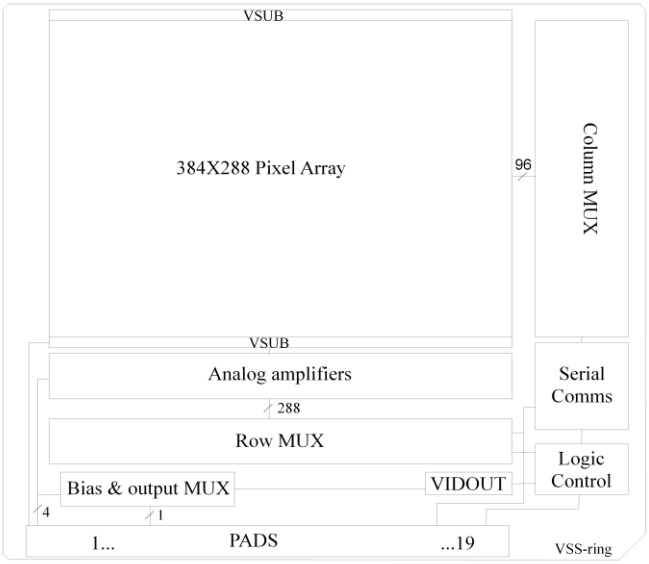


Fig. 1. Floorplan of the readout chip.

Output multiplexing is done row-by-row, with two output modes, namely four pixels at a time and a single pixel at a time. The chip also has a windowing mode, where smaller windows of 4nx1n (n x four pixel group by n x one row) can be read out.

The row multiplex clock will sequentially switch all the pixel output voltages in a row onto a group of four column buses, as the chip allows for four video outputs in parallel. The column group bus voltages are fed to unity gain buffers.

The column multiplex clock sequentially transfers the buffered voltages to internal buses. The voltages on the internal buses can now be multiplexed to four buffered video outputs, or alternatively the four bus lines can be further multiplexed to a single buffered video output at a higher readout rate.

The logic control signals for the chip are generated digitally. Some control parameters, as well as performance parameters, are communicated to the chip via a serial communications word.

B. Pixel design

In the pixel a differential buffered direct injection (BDI) readout scheme is employed to satisfy both pixel size and power dissipation limitations, while permitting high gain and stable detector bias. In direct injection a common-gate transistor is used for controlling the detector bias and to sense the current of the detector. The circuit structure of the buffered direct injection (BDI) configuration of Fig. 2 connects a buffer with inverted gain of –A between the gate node of the common-gate transistor and the input node. The input impedance is decreased by a factor A due this feedback structure, and the injection efficiency is near unity. Due to the virtual-short property of the gain stage, the BDI detector bias control is more stable, releasing strict requirements on the external voltage bias source. Advantages of the BDI are

reasonable immunity to transistor threshold voltage variations, input referred noise and operational bandwidth.

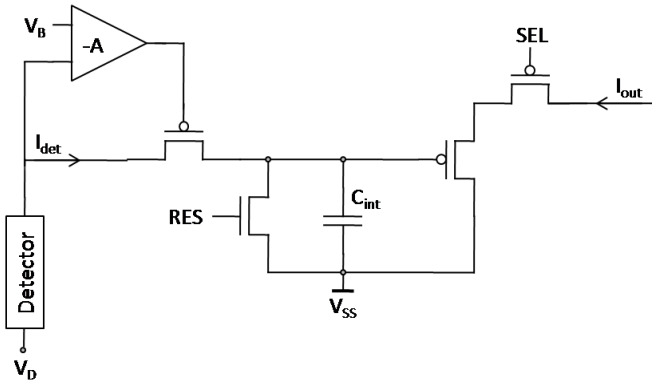


Fig. 2. Principle of buffered direct injection (BDI) readout configuration. V_B is the detector bias voltage, C_{int} the integration capacitance, RES the integration reset signal, and SEL a pixel multiplexing signal.

Pixel features also include the injection of a programmable test current at the input for evaluation and calibration purposes, and the subtraction of an externally controlled background current after the BDI stage. Both features are controlled with array-wide logic voltages. The output current from the bias circuit is given by

$$I_{OUT} = I_{DET} + x_T I_{TEST} - x_S I_{SUB}, \quad (1)$$

with both x_T and x_S logic signals with values 1 or 0, as programmed by a serial communication word. The odd and even test currents can be activated independently to set up an off-off test pattern across the array. In the case where $x_T = 0$ and I_{SUB} is larger than I_{DET}, a detector current path to the supply voltage V_{SS} is provided.

A current mirror with gain transfers I_{OUT} to the integration stage, which consists of an integration timing switch and a resettable integration capacitance.

Correlated double sampling (CDS) is performed by connecting both sides of the CDS capacitance to V_{DD} during the reset cycle. After reset, and during charge integration, the voltage on the gate of an output source follower device will follow the integration signal voltage, with the DC offset across the integration capacitance eliminated. This technique will eliminate DC and low frequency offsets originating from the detector, detector bias circuit, and resetting of the large integration capacitance.

After the integration period and before the reset cycle, the output voltage of the CDS source follower is multiplexed by the relevant row multiplexing signal to the correct column bus.

C. Off-array analog circuitry

A schematic depiction of the pixel array (group of "P"-blocks) and the off-FPA readout functions is given in Fig. 3.

Each column bus is driven by an off-FPA switched current source, which is biased with a common bias voltage. Any column group array of 4 X 288 is enabled by a column select signal CPOW. When CPOW is enabled, the current sources

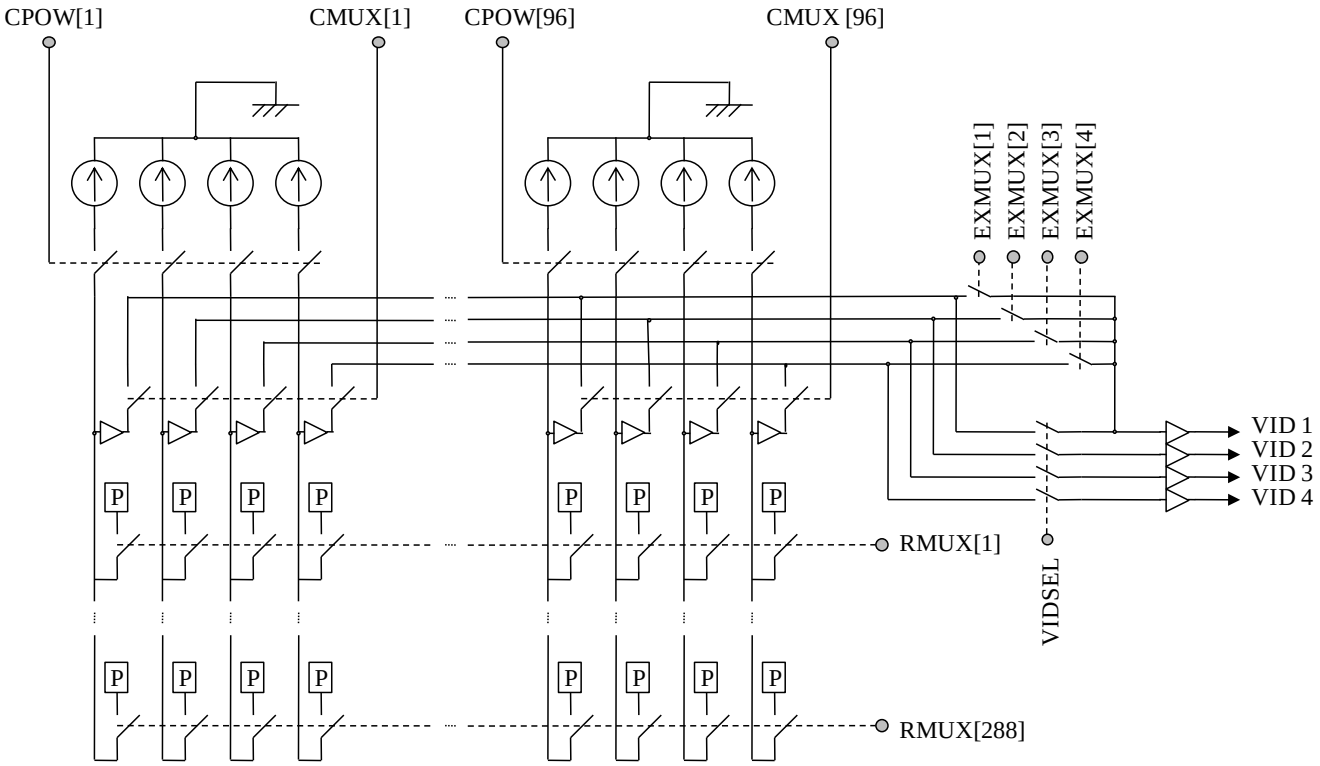


Fig. 3. Schematic depiction of the pixel array and FPA readout circuits. Pixels are designated by the blocks labeled "P".

are switched to the column group buses. The unity gain buffers for the column bus voltages are implemented off-FPA using operational amplifiers to maximize gain and to avoid level shifting problems.

A unity gain two stage operational amplifier voltage follower is used as output driver, and there is one for each of the four video lines. The signal VIDSEL selects between the four output and single output modes. In the single output mode, three of the output buffers are disconnected, and the four internal bus voltages are multiplexed sequentially to the output with the EXMUX logic signals.

The bias signals for the array circuitry and amplifiers are also generated outside the array area. The amplifier bias generator accepts an externally supplied bias current and all internal bias currents are scaled from it. For biasing uniformity across the large array, bias distribution is done in the current domain. A group of 24 columns of pixels is biased by a single local bias circuit to further improve bias uniformity across the array. The column bias distribution scheme is shown in Fig. 4.

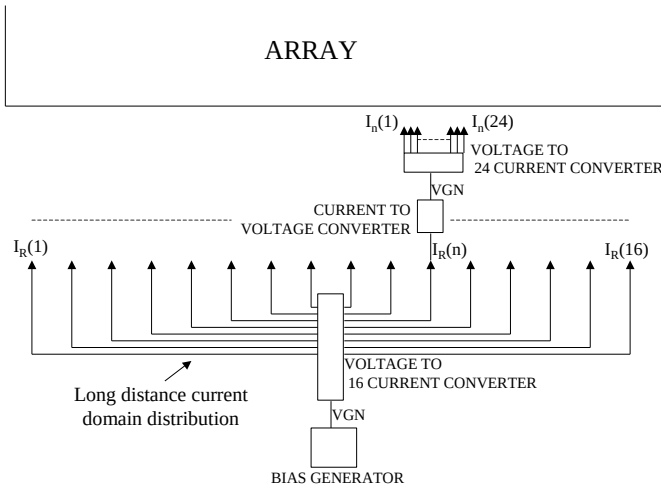


Fig. 4. Distribution scheme for bias currents in column buffers.

D. Power saving

In order to save power dissipation on the readout chip, the major sources of power must be decreased as far as possible. The biasing of the unity gain column group buffers dissipates most power. Switching the buffers on only when a specific column group is addressed, will save significant power.

Any 4 X 288 column group array in Fig. 3 can be disabled by the column select signal CPOW, disconnecting the biasing current sources from the four column group buses. This means that power will be saved since the current sources will not dissipate any power. Switching the column bus lines to the supply voltage V_{DD} causes the unity gain buffers to switch off, and also not dissipate power.

To prevent switching transients corrupting signal voltage values, the next column group will be switched on while multiplexing the current column group. The previous column group will only switch off when the current column group has been addressed. In this scheme three column groups will be on at any time.

The signals to implement the power saving scheme are shown in Fig. 5, where CS[n] is the column select signal to define the window to be multiplexed out, and CPOW[n] the signal to define when power is dissipated by the column group buffers and current sources.

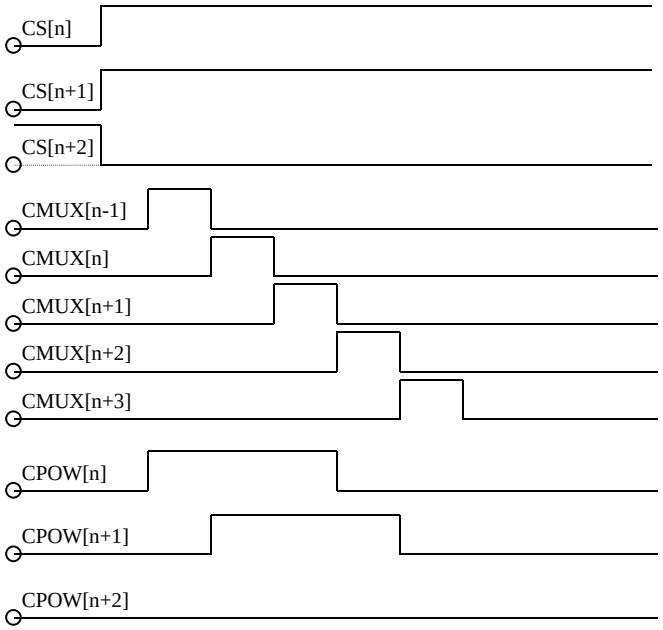


Fig. 5. Signals for the power savings scheme.

In addition to the power saving scheme, all amplifier bias currents are 2-bit programmable through the serial communications word to allow manipulation of the bias power dissipation, while still complying with output video speed and voltage level requirements. Each amplifier bias current can vary from 0 μA, which switches the specific amplifier off, to the maximum designed value.

E. Digital circuitry

A custom-made library of digital standard cells was designed to fit into the 25μm pixel pitch, based on the static design methodology. All digital operations are synchronized by a master clock signal, which directly defines the operational frequency. The nominal frequency of operation is specified as 10MHz, but in implementation the frequency is limited by the analog circuits rather than the digital. An asynchronous master clear signal is used to clear the FPP to a default state.

The control logic implements the following functions:

- Output multiplexing is done row-by-row, with two output modes, namely four pixels at a time and a single pixel at a time.
- Analog multiplexers for readout and buffers for driving the output(s) with power-down mode and the power-saving switching scheme.
- Timing control of the integration and readout multiplexing. During the integration period there are

no clocks applied to the circuit to avoid clock noise coupling into the integration capacitance.

- Control of the windowing mode, where smaller windows of 4nx1n (n x four pixel group by n x one row) can be read out. The window is defined by two opposing corner coordinates.
- Selection of various array modes, including discrete DC bias current values to different analog sections, background current subtraction, and the test currents.

A serial communication word of 55 bits is used to set up the analog array control modes and the windowing parameters. If serial communication is selected, the data word is sent at the beginning of the frame readout period. When no serial communication occurs in a frame, the operation continues with the previously stored data.

The storage of the window and control data can be activated separately, as the pixel array control may be more static than windowing data. Activation is controlled through the serial communications word.

An 8-bit CRC is performed on the communicated data with an external signal to indicate whether an error has been identified. If there has been a communication error, no storage occurs the previously stored data is retained. In determining communication reliability, bit inversion errors, burst errors, and double clocking errors were used in a simulation experiment, using over-conservative error rates. The conclusions of the experiment are as follows:

- All bitstreams with an uneven number of faulty bits will be identified.
- All bitstreams with 2 faulty bits will be identified.
- Approximately 1% of bitstreams with an even number of faulty bits and that number being larger than 2 will not be identified.

These results indicate an acceptable compromise between the number of bits in the CRC and communication reliability.

III. Characterization

A chip photograph of the readout circuit is shown in Fig. 6. The FPA chips were bonded in 84-pin PLCC packages for characterization purposes (test structures on the chip were also bonded).

A custom measurement setup was designed using an FPGA development board to generate the required digital control signals. An interface PCB connects, but with electrical

Fig. 6. Photograph of the readout chip. The die size is 10.46 mm x 9.24 mm.

isolation, the FPGA board to the parallel port of a computer, which is used to configure the operation of the FPP. A measurement setup is shown in Fig. 7. The chip was evaluated for all operational conditions, and only some interesting results will be given here.

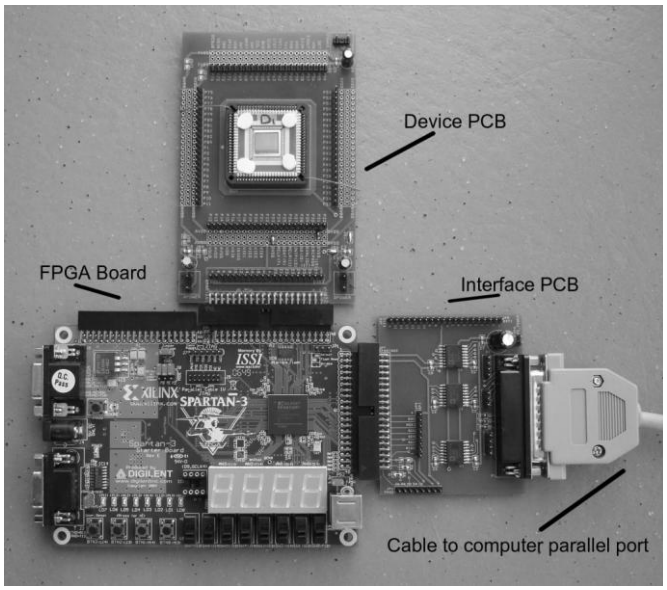


Fig. 7. Measurement setup, showing the readout chip being measured in an 84-pin PLCC package on the device PCB, the FPGA board that generates the control signals, and an electrically isolated interface board for connection to the measurement management PC.

By programming the bias of the amplifiers in the readout chip, the analog power dissipation could be varied between 11.32mW and 40.76mW, delivering video output settling times between 120ns and 50ns respectively. The digital circuitry dissipates 5.46mW dynamic power for the worst case when all four video output are used. The total power dissipation of the readout chip under the worst case conditions is 65mW.

The background current subtraction feature was shown to be operational and linear with respect to the control voltage, except for a slight non-linearity near the origin.

The test current feature is operational, but the values of test current measured were lower than the simulated values. This can be expected due to inaccurate current mirroring for small currents. Furthermore, the specific values of test currents do not affect the functionality of the feature.

By tracking pixel patterns in the array, it was clearly demonstrated that the windowing mode is functional.

The experimental results indicate full functionality of the readout chip.

IV. Conclusion

This paper presented a characterized readout circuit for an array of 384 X 288 infrared detectors. The main performance characteristics of the chip are summarized in Table I.

TABLE I
SUMMARY OF CHIP PERFORMANCE CHARACTERISTICS

Characteristic	Performance
Technology	0.35 μm 4-metal CMOS
Array size	384 x 288 pixels
Pixel size	25 μm x 25 μm
Die size	10.46 mm x 9.24 mm (almost 1cm^2)
Power dissipation	65 mW
Power supply voltage	3.6 V
Maximum video signal	2 V
Maximum settling time	80 ns
Frequency of operation	10 MHz
Transistors/pixel	18
Transistors /array	1 990 656
Transistors/chip	2 088 503 > 2 million

References

[1] H. E. Maes, C. Claeys, R. Mertens, A. Campitelli, C. Van Hoof, J. De Boeck, "Trends in microelectronics, optical detectors and biosensors", Advanced Engineering Materials 2001, Vol. 3(10), pp. 781-7, 2001.

[2] A. Rogalski, "Optical detectors for focal plane arrays", Opto-electronics Review Vol. 12(1), pp. 221-245, 2004.

[3] C-C. Hsieh, C-Y. Wu, F-W Jih, T-P Sun, "Focal-Plane-Arrays and CMOS readout techniques of infrared imaging systems", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7(4), pp. 594-605, August 1997.

A CMOS Implementation of a Jakes Fading Channel

Johannes Petrus Buckle and Johan Schoeman, *Member, IEEE*
Department of Electrical, Electronic and Computer Engineering
University of Pretoria, South Africa

Abstract – Communication engineers frequently use either a filter based or a Jakes fading channel emulator during the performance evaluation of proposed telecommunication systems. These are typically implemented with complex baseband simulation models, as system performance can easily be measured in simulation without any RF difficulties related to a similar discrete implementation. The authors propose a fully-integrated CMOS implementation of the Jakes fading channel model based on the sum-of-sinusoids method, capable of performing single and multipath fading on the envelope of the transmitted RF signal. The contribution of the paper lies in the overall system implementation, as well as the circuit level modifications done at sub-system level aimed towards eventual integration. A hardware prototype was designed and implemented to validate the Jakes model in an analogue milieu, rather than the typical digital simulation. The integrated design is completed for fabrication in a standard 0.35-µm CMOS process with minimal external components. The simulated fading channel is aimed at a GSM900 system and emulates a receiver speed of up to 150km/h, while consuming approximately 15.4 mW per path.

I. INTRODUCTION

Jakes method is an efficient way to implement a simulation of the Raleigh fading multi-path channel with sum-of-sinusoids (SOS). A number of sinusoids, each with a random Doppler shift, are generated by oscillators and added by superposition to give the output of the simulated channel.

Figure 1 illustrates the nature of a wireless channel where a signal propagates to the receiver via multiple paths. The transmitted signal is reflected and refracted, thus the received signal is made up of a superposition of waves. These waves may add constructively or destructively, giving rise to fading of the envelope of the received signal. The random Doppler effect that shifts each of the sinusoids is due to the difference in velocity of the receiver or sender.

To emuate this type of multipath channel a set of oscillators are implemented on CMOS. The outputs of these offset oscillators are then modulated onto the transmitted signal to emulate the effect of a single path. This single path fading generator may be considered as a single module. This module may then be duplicated to form a multipath fading generator.

The main engineering challenge lies in the implementation of the random sinusoidal oscillators, since these oscillators need to operate at different frequencies and phase shifts. Careful consideration has been taken regarding the type of oscillators that would be used, since this has a large effect on the size of the CMOS implementation and the power consumption of the system.

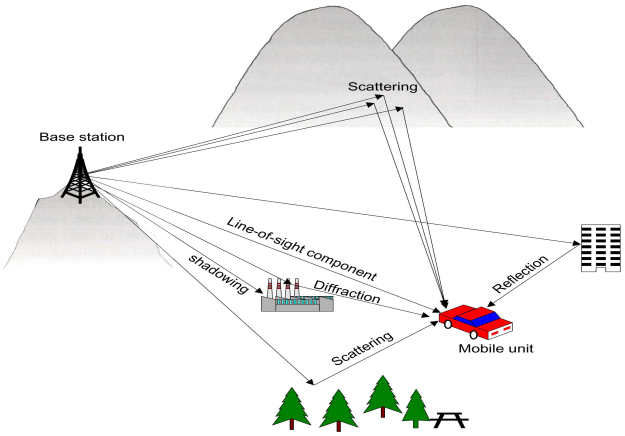


Fig. 1: Typical mobile radio scenario illustrating multipath propagation

The fact that the Doppler frequencies are extremely low causes a challenge in the CMOS implementation of large capacitance values for the oscillators. Further challenges addressed include parasitic effects present in any radio frequency (RF) design when a microelectronic implementation and the integration of subsystems are attempted.

II. CONCEPTUAL DESIGN

The most important aspects of the system include the design and implementation of the oscillators, mixers and capacitive multipliers. The design approach of these components will be addressed in the following subsections.

A. Offset oscillator

It is required for this application that it is possible to control the frequency, amplitude and distortion level of the oscillator. This is to ensure that the output of the oscillators in the circuit design and the sine wave generators in the ideal mathematical model correlate. A number of analogue techniques are available for generating sinusoidal waves. Some of the oscillator architectures that were considered included phase shift, Wein-bridge, LC negative resistance and tuning fork oscillators. Each individual circuit approach has inherent strengths and weaknesses, which were compared and considered as design alternatives for this implementation.

The frequencies of the offset oscillators are determined by the Doppler shift experienced by the received signal, which should be variable between 34 Hz and 125 Hz. This Doppler

shift range is for nine offset oscillators that corresponds to a carrier frequency of 900 MHz and a mobile velocity of 150 km/h. As a result of these very low frequencies, the oscillators need to have large capacitance values, which is very difficult to implement in a microelectronic design.

There are two possible approaches to solve the problem of large capacitor values. The first is to use external capacitors, while the second requires the implementation of capacitance multipliers. The system implementation requires the use of nine offset oscillators per path and three unique propagation paths, all of which leads to 27 oscillators. This results in a challenge concerning the amount of external capacitors and output pins to connect these capacitors. External capacitors are also known to have a high variance in capacitance values; this could increase the randomness of the process which will be advantageous in terms of the statistical stationary of the system. However, this random change in oscillator frequency will cause the system to deviate from the required power spectral density of a typical Jakes fading channel emulator.

With the capacitance multiplication approach a small capacitor is implemented on the silicon die and then multiplied with the aid of a specially designed capacitor multiplier configuration. This makes the design more difficult and affects the choice of the oscillator type, since the need arises to ground some of the nodes in the capacitance multiplier [1]. Due to this grounding property of impedance multipliers, an oscillator with a floating capacitor may not be considered.

An additional system requirement is that the Doppler shift should be controllable with an external reference voltage. This again affects the choice of oscillator type. It is possible to implement a CMOS transistor as a varactor and in this way change the frequency of the oscillator, effectively resulting in a voltage controlled oscillator (VCO). The gate to source capacitance is used for this implementation and has an even smaller value than those capacitors implemented between metal layers in the AMS process. This does, however, lead to a capacitor multiplier with an even higher gain factor.

After all these alternatives were considered a specific triangular VCO was designed. This VCO uses the linear part of the charge and discharge curves of a capacitor to generate a very low frequency triangular wave. The frequency of oscillation can be changed by varying the charge current through the capacitor, thus eliminating the need to vary the capacitor value and allowing for the use of larger capacitors.

The triangular wave is converted to a sinusoidal wave using the square law characteristics of the MOS transistor differential pair amplifier. This can be described as passing the triangular wave through a circuit exhibiting a sinusoidal voltage transfer curve (VTC), as shown in Figure 2. The result is a sinusoidal wave [2]. Since nonlinear shaping is independent of frequency, this form of sinusoidal wave generation is particularly convenient when used in conjunction with the triangular VCO described above, as this configuration gives a wider tuning range than the oscillators considered above.

Extensive research and design was done to achieve 10 000 times multiplication of a capacitor. The design could not be stabilized for the charge and discharge curves that the triangle VCO needs, but the AC-sweep simulation shows that the multiplier can be used for filter application. Although it was eventually decided not to use the circuit in the emulator, the

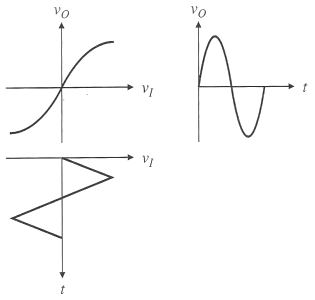


Fig. 2: VTC of a triangle to sinusoidal wave converter [2]

design is still presented in subsection C to provide guidance to alternative future implementations.

Due to time constraints it was decided to solve the problem with a mixing approach. The capacitor of the VCO was minimised and the offset oscillators generated the Doppler spread between 1011Hz and 1125Hz and then mixed these down to 11Hz to 125Hz. In this way the system can still be implemented without the use of the capacitor multiplier.

B.Mixer

The mixer is an important part of the QAM scheme, which is used to multiply the incoming signal to the fading envelope and to mix the higher Doppler spectrum (due to the capacitive problem discussed earlier) down to the desired values. One of the implementations considered was an analogue current multiplier. This multiplier uses the square law characteristics of transistors to multiply two currents and divide it by a third. This configuration needs dual supplies to be a balanced mixer and thus has higher power dissipation.

For this reason it was decided to use the Gilbert cell for this design. It is very well suited for implementation on CMOS and is often used in communication applications. The cell has a differential input and output to realise a balanced mixer. Two single to differential ended converters are used to convert the single ended input signals to a differential signal. A differential to single ended converter is also used to convert the output back to single ended.

C.Capacitance multiplier

The capacitor that the triangular VCO uses to generate the charge and discharge curves is very large in comparison to the rest of the implementation. Large capacitor values can not be implemented in CMOS.

It is possible to implement high capacitance values from smaller ones by performing capacitance multiplication [3]. This can be obtained using a current gain second generation current conveyor (CGCCII) which shows a current gain

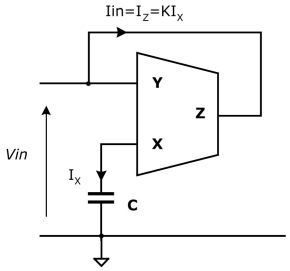


Fig. 3: CGCCII-based capacitor multiplier [3]

between the X and the Z terminals, as in the circuit illustrated in Figure 3. A second generation current conveyor (CCII) is essentially a current mode operation amplifier. The voltage on the Y input is duplicated to the Z output and the current through X is then multiplied by a gain factor and sent to the Z output.

If the current flowing from the Z node is K times greater than that flowing from the X node, the impedance at the X node is K times higher than that seen from the Y node, as clarified by the following formulas:

$$Z_{IN} = \frac{V_{IN}}{I_{IN}} = \frac{V_Y}{I_Z} = \frac{V_Y}{KI_X} = \frac{V_Y}{K\frac{V_X}{Z_C}} = \frac{V_Y}{K\frac{V_Y}{Z_C}} = \frac{Z_c}{K} \quad (1)$$

from which

$$Z_{IN} = \frac{1}{sCK} \quad (2)$$

being $Z_c = 1/sC$.

This results in an equivalent impedance whose value is K times lower than X_C at the Y node, which may be interpreted as a capacitance K times higher than C. Capacitance multiplication is, therefore, achieved. If a real CCII is used as a capacitance multiplier the equivalent impedance at Y node Z_{IN} is given by:

$$Z_{IN} = \frac{R_Z\left(1+sCR_x\right)}{1+s\left(C_Z R_Z + CR_Z + KCR_Z + C_Y R_Z\right) + s^2 CR_X R_Z\left(C_Z + C_Y\right)} \quad (3)$$

From (3), if $R_X = C_Z = C_Y = 0$ and $R_Z = \infty$ (the ideal case), (2) is recovered.

Figure 4 shows the ideal and non-ideal impedance behavior (neglecting the high frequency pole, yielding $C_Z = C_Y = 0$) at CGCCII based capacitance multiplier input.

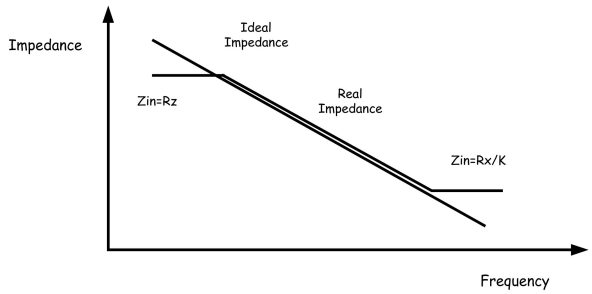


Fig. 4: Ideal and real (or non ideal) impedance behaviour (for $C_Z = C_Y = 0$) at CGCCII based capacitance multiplier [3]

Simulation results suggest that an acceptable agreement between theoretical and obtained behavior is not possible for more than three or four frequency decades [3]. From (3), it can be found, neglecting the second order term, that for low frequencies the equivalent impedance is equal to R_Z, while at high frequencies it tends to R_X/K (considering $C_Z = C_Y = 0$). Thus, the frequency range is limited by these two values and can be adjusted by modifying the R_X and R_Z values with an opportune modification of CCII output stages biasing currents.

III. Circuit design

A. Triangular VCO

The offset oscillators use the linear part of the charge and discharge curve of a capacitor to generate the low frequency triangular waves. The schematic diagram of the circuit (from [4]) is shown in Figure 5. The BiCMOS components were changed to CMOS and the voltage reference circuit was changed so that the output is between 669.68mV and 1.23V. For stable operation a capacitor (±20pF) is connected between ground and the output port. This capacitor is charged and discharged through the switching network which is controlled by the comparator circuit. The capacitor charging current alternately comes from the programming current directly, and then a mirror of the programming current. This programming current is set by the input voltage.

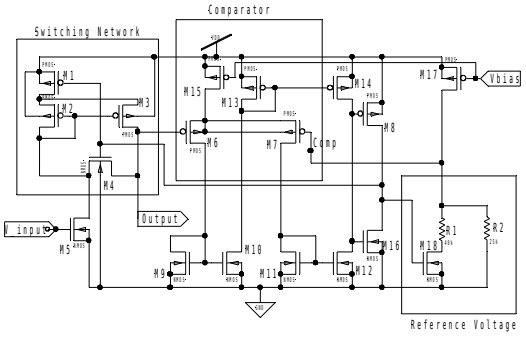


Fig. 5: Schematic diagram of the triangular VCO

The comparator compares the voltage across the capacitor and the reference circuit. If transistor M18 is off and the capacitor voltage (also the output voltage) is less than the reference voltage (1.23V for M1 off), the switching network charges the capacitor. As soon as the capacitor is charged to 1.23V the comparator turns M18 on (the reference voltage rises to 669.68mV) and sets the switching network to discharge. The comparator effectively controls the amplitude of oscillation by adding a second 40kΩ resistor parallel to the 25kΩ resistor by means of controlling M18 as a switch.

The frequency of oscillation is determined by the charge and discharge times given by

$$f_o = \frac{i_{in}}{0.8C} \quad . \quad (4)$$

where i_{in} can be determined by

$$i_{in} = \frac{K}{2}\frac{W}{L}\left(v_{GS} - V_{TR}\right)^2 \quad (5)$$

The values for K and v_{gs} are typically 170μA/V² and 0.5V respectively.

B. Triangle to sine converter

The differential pair in Figure 6 is used to convert the triangular wave from the VCO to a sinusoidal wave suitable for the Doppler spread generation. The current mirror is designed to sink 4μA from the differential pair. Active loads were considered, but it caused the VTC to have sharp corners, thus the output sine wave was distorted. For this reason the load resistors (R_d) are set to 100kΩ, this is for an output DC-level of 3.1V and an output swing of 200mV.

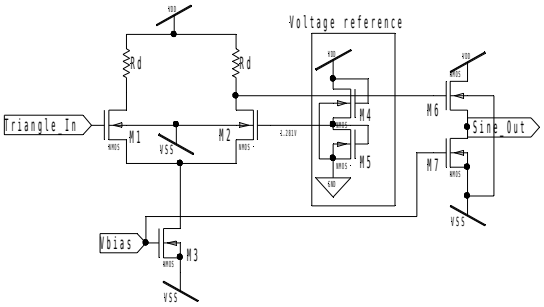


Fig. 6: Schematic diagram of the triangle to sine converter

For each transistor in the differential pair,

$$g_m = \frac{I}{V_{gs} - V_t} \tag{6}$$

and the drain current through each transistor can then be given by

$$i_d = g_m \left(v_{id} / 2 \right). \tag{7}$$

If the aspect ratios (W/L) of the transistors are decreased the corners of the VTC become rounder. This is due to the fact that the aspect ratios affect the gain of the differential pair and thus the slope of the linear part of the VTC. The voltage reference circuit is set to provide a DC reference of 949.35mV to the differential pair. This is the mean value of the triangle wave from the VCO, which has an amplitude of 27.9mV from this mean. The differential pair was designed for

$$|vid|_{\max} = \sqrt{2}\left(V_{gs} - V_t\right) = 27.9\,mV \ . \tag{8}$$

This is done to ensure that the peaks of the triangle wave only just reach the saturation point of the VTC. If $|v_{id}|$ is too small, the output sine wave will have flat tops and, if it is too large, the output sine wave will have sharp peaks.

The output DC-level of the converter must be 1.6V for successful interfacing with other modules. This was accomplished by the use of a source follower to provide the necessary level shifting [5].

C. Gilbert mixer

The mixer is also one of the main building blocks of the system and is used for QAM and down mixing. A double balanced Gilbert cell [6] was selected and implemented.

The Gilbert cell needs differential inputs and outputs, thus the input in converted to differential-ended and the output is converted back to single-ended. For this special modules were designed.

D. Capacitor multiplier

Capacitor multiplication, could be a very effective way of implementing capacitors onto CMOS. With ideal CCII's the capacitor multiplication works perfectly. However a CCII could not be designed to achieve the desired charge and discharge curves needed for the triangle VCO or for the filtering of the sum term after mixing. The closest proposed capacitor multiplier is shown in Figure 8.

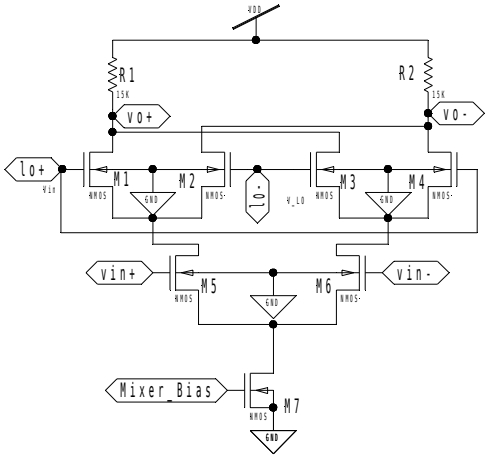


Fig. 7: Schematic diagram of the implemented Gilbert cell mixer

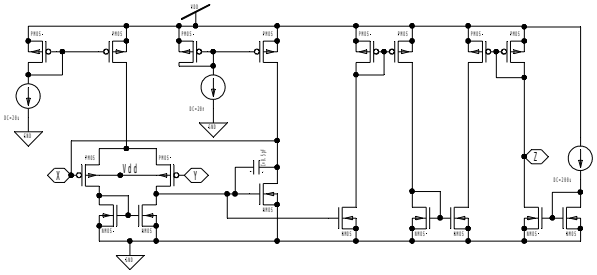


Fig. 8: Schematic diagram of the capacitor multiplier [3]

IV. Results

Due to the fact that the implementation was aimed to be an ASIC, the simulations and verification of the system functionality was done mainly in software. A hardware demonstration module was built to illustrate that the concept could indeed be implemented and operates similarly to the mathematical model.

A. Offset oscillators

The sinusoidal wave output and FFT from the offset oscillators can be seen in Figure 9.

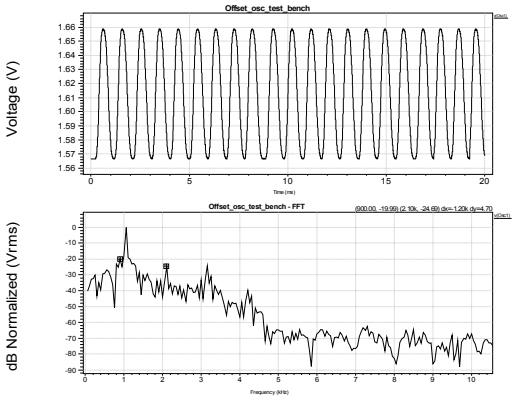


Fig. 9: Simulation results of the offset oscillator. The transient output is at the top and the FFT at the bottom.

The oscillation starts after 200µs and is at full amplitude from the first cycle. This start-up time is good considering that the period of oscillation is 950µs, thus the start-up time is less than one period. There was some distortion in the sinusoidal wave, but it occurs randomly and does not effect the FFT to a great extend. This causes the phase of the oscillator to be random, which is a desirable property for the implementation of the simulator, because the phase of the offset oscillators in the mathematical model was also set to be random. The THD of the output signal determined from this experiment was 3.1%.

B.Multipath characteristics

The S-Edit results are shown in Figure 10. The LCR and AFD could not be determined because the threshold level was unknown. This could be a subject for further study. The Simulink and S-edit results correlate and the fading of the signal envelope can easily be seen.

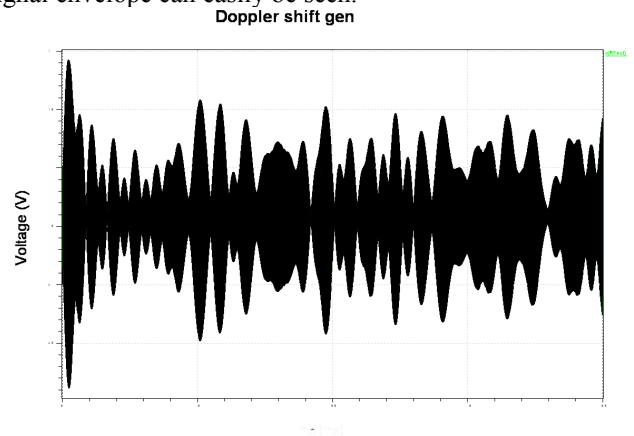


Fig 10. S-Edit time domain output of the system.

C.The power dissipation of the system

The current drawn from the positive and negative rails is shown in Figure 11. This was used to compute the power dissipation of a single path generator, which was determined to be 15.39mW. This leads to a total power dissipation of 46.18mW. The maximum specified current for the simulation was 100mW. This indicates that during the design of the subsystems, too much emphasis was laid on low power consumption. If there was designed with higher biasing currents, the resistance values could have been lower. This results in smaller die area.

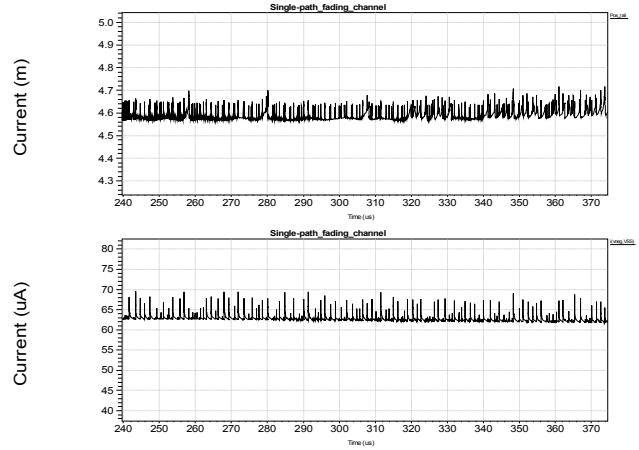


Fig. 11: The current consumption of a single path fading simulator. The positive (top) and negative (bottom) rails are shown.

V.Conclusion

The offset oscillators' capacitor charging and discharging curves were successfully used for the generation of a low frequency triangular wave. This triangle wave was converted to a sine wave using the square law characteristics of a differential amplifier pair. This resulted in a low frequency VCO using only one grounded capacitor. The grounded capacitor could be implemented with a capacitor multiplier. This concept was investigated but not implemented. The capacitor multiplier could reduce the silicon area of the design.

A multipath fading channel was implemented in Simulink and this was the basis on which the CMOS implemented system was tested. To minimize capacitor values the Doppler spread was generated at 1kHz. The frequency components were then mixed down to 100Hz and the sum term was filtered out with an external capacitor. The fully implemented system was also simulated with Simulink and the subsystems were designed accordingly.

Capacitor multiplier implementations can solve many challenges faced in microelectronic design. Further work needs to be done on the transient and frequency response of capacitor multipliers as current research outputs has mainly been done in high frequency filtering applications. It would be beneficial to study and implement a well designed CCII, which also allows for a floating inductor implementation.

This design may also be improved by overcoming some limitations. The Doppler spread in this simulator is not directly dependent on the input frequency. It is predetermined and the offset oscillators are then set accordingly and can only be changed by the reference voltage. This reference voltage should only be used for the mobile speed variations.

References

[1] Horowitz, P. and Hill, W., 1995. *The art of electronics*. 2nd ed. USA. Cambridge University Press

[2] Franco, S., 2002. *Design with Operational Amplifiers and Analog Integrated Circuits.* 3rd ed. New York: McGraw-Hill.

[3] Ferri, G. and Guerrini, N., 2003. *Low voltage, low power CMOS current conveyors.* 1st ed. Springer.

[4] Eaton, H. and Lesho, J., 1993. *A low voltage, low power, BiCMOS, Audio-frequency voltage controlled oscillator with sinusoidal output*. Circuits and systems, p1433-1466 vol 2.

[5] Horenstein, M.N., 1995. *Microelectronic circuits and devices*. 2nd ed. New Jersey. Prentice Hall

[6] Greyling R.M.M., 2004. *Design of an integrated direct sequence spread spectrum (DSSS) transceiver for cellular systems.* (Final Report: Supplementary exam). Pretoria, South Africa. University of Pretoria.

High Speed, Low Power CMOS Optical Receiver Front-End

Christo Janse van Rensburg[1] and Petrus J. Venter[2], *MIEEE*
[1]s24119505@tuks.co.za, [2]jannes.venter@up.ac.za

***Abstract*—This paper presents an optical receiver analog front-end that has been realized in the 0.35 µm CMOS technology for low noise, low power applications. The approach includes small geometric, high field effects to explain the negative impact of operating short channels MOSFETs at high levels of inversion.**

***Index Terms*—CMOS, optical receiver, transimpedance amplifier, regulated cascode, limiting amplifier, gain enhancement, negative resistance, frequency compensation.**

I. Introduction

THE exponential increase in modern microprocessor unit (MPU) transistor density has been accomplished by the decrease in transistor and interconnect dimensions. Unlike transistors for which performance improves with scaling, the delay of interconnects increases with scaling. This has led to the interconnect replacing the transistor as the dominant determinant of chip performance. Optical interconnects has been proposed as a viable alternative to electrical interconnects, and this will require light to be converted to electrical energy to provide functionality on-chip. The analogue front-end is a critical subsystem of optical receivers which demands a low noise input current, high transimpedance gain, wide bandwidth and low power consumption. Unfortunately these requirements present a trade-off and will therefore require careful optimisation. Fig. 1 shows a conventional optical receiver front-end comprising a photodiode (PD), a transimpedance amplifier (TIA) and a limiting amplifier (LA). The photodiode transforms the incident light intensity into a proportional current. The TIA will convert this small current signal to a voltage signal characterized by its transimpedance gain. The LA will further amplify the small voltage signal from the TIA to logic level signals.

A number of CMOS TIAs have been reported. The common-source (CS) amplifier implementing shunt feedback at the input is a popular topology but proves to be incapable of isolating the large parasitic capacitance of the photodiode, thereby limiting the bandwidth. Several common-gate (CG) topologies also exist providing wide bandwidth and low input impedance, but the small transconductance of MOSFETs deteriorates the noise performance of the TIA. The regulated cascode (RGC) is a topology that has been used as the input stage of optical receivers, [1]-[5]. The RGC behaves as a CG stage with a large transconductance as a result of the local feedback stage. This inherently improves the bandwidth by suppressing the parasitic capacitance of the photodiode.

The fundamental trade-off in the design of the LA is gain versus bandwidth. LAs usually consist of a number of cascading gain stages. Due to the cascading drawback, a reduction in bandwidth will result as the number of cascading stages is increased to improve the gain. Several broadband techniques have been developed to allow each individual LA stage to operate at higher speeds. Some of these techniques include inductive peaking, capacitance degeneration and f_T doublers [6]. While these techniques are usually intuitive in the design of LAs, characteristics such as power dissipation and chip area is usually sacrificed to achieve the required gain and bandwidth.

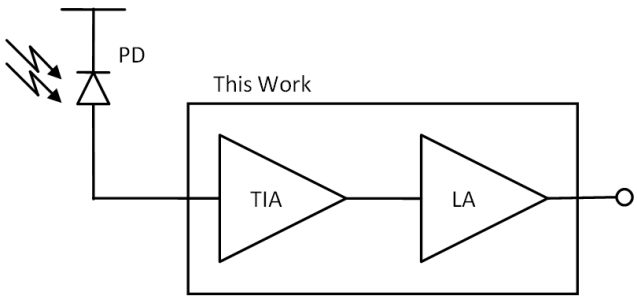


Fig. 1. Block diagram of a conventional clock recovery circuit.

This paper presents an optical receiver front-end, realized with 0.35µm CMOS technology for optical clock recovery applications. Small geometric, high field effects like velocity saturation have also been included in the analysis of the optical receiver front-end. Section II describes the RGC as the input stage while Section III describes the LA which utilizes negative resistance and a frequency compensation element for simultaneous gain and bandwidth improvement. Section IV explains the simulation results and the layout of the front-end including the specifications achieved.

II. Proposed Transimpedance Amplifier

A schematic diagram of the RGC configuration is shown in Fig. 2. The photodiode is connected to the input terminal of the RGC. Transistor M_2 and resistor R_2 act as the local feedback stage, biasing transistor M_1 and lowering the input

impedance to produce a virtual ground at the input. This isolates the input more effectively from the large parasitic capacitance of the photodiode when compared to a normal CG stage.

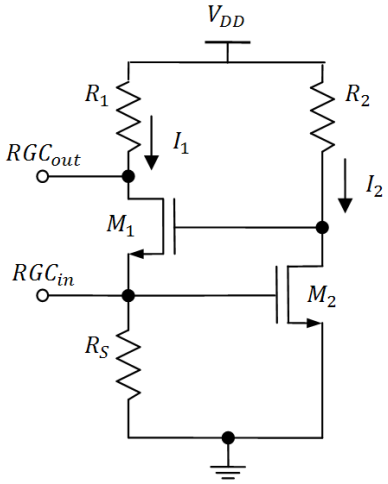


Fig. 2. Schematic diagram of the RGC configuration.

The small signal analysis of the RGC configuration reveals that the input impedance is given by

$$Z_{in}(0) = \frac{1}{g_{m1}(1+g_{m2}R_2)} \tag{1}$$

where g_{m1} and g_{m2} denote the transconductance of transistors M_1 and M_2, respectively. Equation (1) indicates that the input impedance is reduced by the voltage gain of the local feedback stage. Therefore the RGC configuration is characterized as a CG input stage with a large transconductance (G_m) of

$$G_m = g_{m1}(1 + g_{m2}R_2). \tag{2}$$

Calculation of the DC transimpedance gain reveals that the gain can be approximated as

$$Z(0) = \frac{g_{m1}R_S(1+g_{m2}R_2)}{1+g_{m1}R_S(1+g_{m2}R_2)}R_1 \approx R_1. \tag{3}$$

The RGC contains two poles and one zero. The input pole is given by

$$f_{in} = \frac{g_{m1}(1+g_{m2}R_2)}{2\pi(C_{PD}+C_{gs2})} \tag{4}$$

were C_{PD} denotes the photodiode capacitance and C_{gs2} denotes the gate-source capacitance of transistor M_2. The output pole is situated at

$$f_{out} = \frac{1}{2\pi R_1(C_{out}+C_{gd1})} \tag{5}$$

where C_{out} denotes the load capacitance at the output node and C_{gd1} denotes the gate-drain capacitance of transistor M_1. A comparison between (4) and (5) shows that the output node has now become the dominant pole and that f_{in} has shifted to a much higher frequency due to the larger transconductance of transistor M_1. Care should also be taken as a result of to the addition of a zero in the frequency response due to the local feedback stage. The zero can cause peaking in the frequency response at

$$f_Z = \frac{1}{2\pi R_2(C_{gs1}+C_{gd2})} \tag{6}$$

where C_{gd2} denotes the gate-drain capacitance of transistor M_2. The zero can be shifted to a much higher frequency by either lowering resistor R_2 or reducing the channel width of transistor M_1, thus reducing C_{gs1}. The appropriate choice would be to decrease the resistance of R_2 at a cost of an increase in its thermal noise contribution. This will increase the drain current through transistor M_2 which will increase the transconductance of M_2, thereby negating the effect of a reduction in the total transconductance (G_m).

The total input referred noise spectral density of the RGC input stage is given by

$$\begin{aligned}\overline{i_{eq}^2} = {} & \frac{\omega^2(C_{gs1}+C_{gd2})^2}{g_{m1}^2}\left(\frac{4kT\lambda}{R_1} + 4kT\lambda g_{d0,1} + K\frac{I_1^a}{f}\right) \\ & + \frac{\omega^2(C_{PD}+C_{gs2})^2}{(g_{m2}+{}^1/_{R_2})^2}\left(\frac{4kT\lambda}{R_2} + 4kT\lambda g_{d0,2} + K\frac{I_2^a}{f}\right) \\ & + \frac{4kT}{R_1} + \frac{4kT}{R_S}\end{aligned} \tag{7}$$

where g_{d0} is the zero-bias drain conductance of the MOS transistor, k is the Boltzmann constant, T is the absolute temperature and K and a are flicker noise constants. From (7) it is seen that the resistance values should be maximized to reduce the thermal noise contribution. To this effect, increasing resistor R_1 also increases the transimpedance gain of the RGC. This will unfortunately reduce the current through transistor M_1, decreasing its transconductance. The largest noise contributor at high frequencies will be the second term in (7) as it contains the large parasitic capacitance of the photodiode. Therefore the transconductance of transistor M_2 is chosen larger than that of transistor M_1.

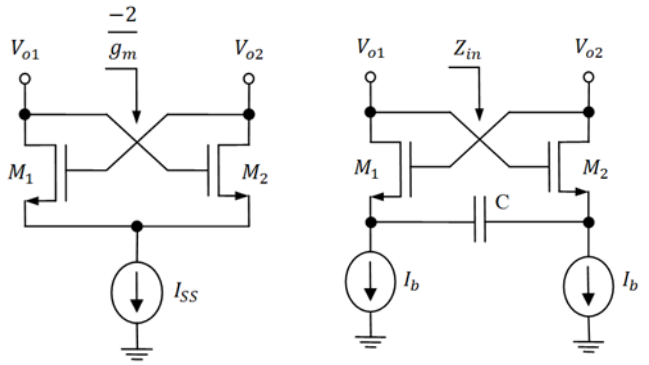
Designing the RGC at minimum transistor gate-length leads to several factors negatively impacting some key performance characteristics. Small geometric high field effects like velocity saturation is associated with the loss of effective carrier mobility caused by a high horizontal electric field between the pinched-off drain and the source. This causes the drain-current of the transistor to drop below the value predicted by the square-law. This deviation from the ideal square-law relation leads to a decline in the transconductance of the transistors. Under velocity saturation conditions, the transconductance is reduced compared to the equivalent transconductance under negligible velocity saturation. Equations (1), (2) and (4) clearly show the effect of a reduced transconductance on various circuit parameters. This may even cause the pole at the input of the RGC to revert back to being the dominant pole, negating the advantage of using the RGC configuration. Another critical factor is the additional thermal noise due to the smaller gate-length transistors. According to [7], thermal

noise up to 5 times larger has been measured using transistor lengths of less than 1 µm. This increase in thermal noise may be attributed to the large electric fields, producing hot electrons, which may be responsible for the increase in the intrinsic device noise compared to long channel transistors.

III. Proposed Limiting Amplifier

A cascade of CS stages usually proves inadequate to operate as a broadband amplifier since the total parasitic capacitance increases with each cascading stage. Consequently, LAs are designed using minimum transistor gate-lengths. Even though velocity saturation is a disadvantage of operating short-channel MOSFETS at high levels of inversion, the small gate areas lead to minimum gate capacitances which lead to optimum circuit bandwidth. The advantages of a decrease in capacitance generally outweigh the disadvantage of a reduced transconductance. With the voltage gain of the CS stage equal to g_mR_D, it becomes increasingly more difficult for each LA stage to achieve a high gain due to the various trade-offs in the design with the addition of small geometric high field effects.

The gain of the CS stage can be increased by increasing the output resistance, R_D. Instead of increasing the resistance of R_D, a negative resistance circuit is placed in parallel with the CS stage, thereby increasing the effective output resistance without affecting the drain current of the CS stage. Unfortunately, the load resistance cannot be increased indefinitely as the dominant pole is directly related to its value resulting in a decrease in bandwidth.



(a) (b)

Fig. 4. Schematic diagram of (a) shows the configuration used to realize a negative resistance for gain boosting and (b) shows the frequency compensation element used to introduce a zero in the LA frequency response.

A. Negative Resistance Circuit

The realization of a negative resistance circuit is shown in Fig. 4(a). The negative resistance is achieved when noting that feedback multiplies or divides the input and the output impedances by a factor of one plus the loop gain. Thus, if the loop gain is sufficiently negative, a negative resistance is achieved. With the negative resistance configuration of Fig. 4(a) applied to a differential CS stage, the effective output resistance is given by

$$R_L = R_D || \frac{-1}{g_{m1}} = \frac{R_D}{1 - g_{m1}R_D} \tag{8}$$

where R_D denotes the drain resistor of the CS stage. In order to ensure a positive resistance,

$$g_{m1} < {}^{1}/_{R_D}. \tag{9}$$

B. Frequency Compensation Element

The severe trade-off between gain and bandwidth can be lessened by using pole-zero placements and cancelling the dominant output pole of the CS stage. A frequency compensation element is therefore included in the output of the differential CS stage. The purpose of the frequency compensation element is to introduce a zero in the amplifier frequency response which, with careful placement, can cancel the dominant output pole, leading to an increase in bandwidth. The frequency compensation element is shown in Fig. 4(b).

The small signal analysis reveals that the input impedance is given by

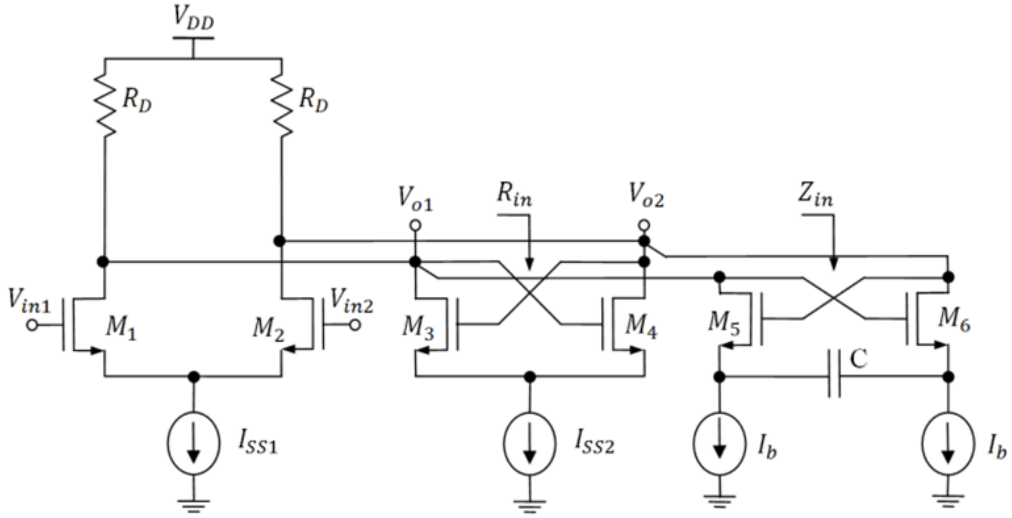


Fig. 5. The complete LA gain stage.

$$Z_{in}(s) = \frac{1}{sC_{gs1}}\left[\frac{1+\frac{g_{m1}+sC_{gs1}}{sC}+\frac{g_{m1}+sC_{gs1}}{g_{m2}+sC_{gs2}}}{1-\frac{g_{m2}}{sC_{gs1}}\left(\frac{g_{m1}+sC_{gs1}}{g_{m2}+sC_{gs2}}\right)}\right]. \quad (10)$$

Under the assumption that $g_{m1} = g_{m2}$ and $C_{gs1} = C_{gs2}$, (10) reduces to

$$Z_{in}(s) = \frac{g_{m1}+s(C_{gs1}+2C)}{sC(sC_{gs1}-g_{m2})}, \quad (11)$$

therefore, introducing a zero at

$$f_Z = \frac{g_{m1}}{2\pi(C_{gs1}+2C)}. \quad (12)$$

The value of *C* can be chosen as to ensure that the zero is placed at the dominant output pole of the LA gain stage, thereby increasing the bandwidth.

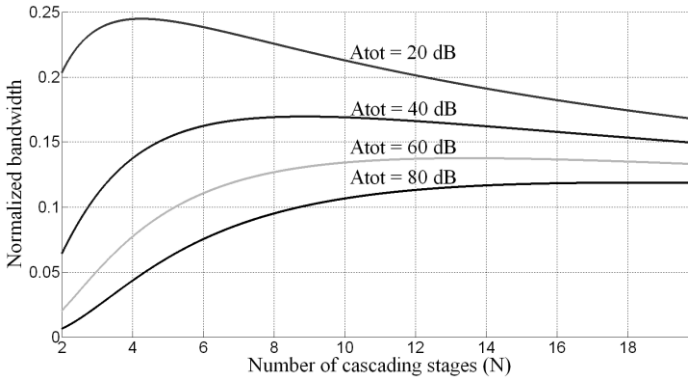


Fig. 6. The normalized bandwidth over the number of cascading gain stages for a one pole system. The normalized bandwidth is equal to BW_{-3dB}/B.

C. Limiting Amplifier Gain Stage

The LA gain stage is shown in Fig. 5, [8]. It shows that the negative resistance circuit and the frequency compensation element are placed in parallel with the differential CS stage. The output of the LA gain stage is taken at V_{o1} and V_{o2}. The small signal analysis of Fig. 5 reveals the total output resistance equal to

$$R_L = R_D||R_{in}||Z_{in} \quad (13)$$

and the open-loop voltage gain equal to

$$A_v = -g_{m1}R_L\frac{g_{m5}+s(C_{gs5}+2C)}{s^2CR_LC_{gs5}+s(g_{m5}R_LC+C_{gs5}+2C)+g_{m5}}. \quad (14)$$

The natural frequency and quality factor can be determined as

$$\omega_n = \sqrt{\frac{g_{m5}}{R_LC_{gs5}C}} \quad (15)$$

$$Q = \frac{\sqrt{g_{m5}R_LC_{gs5}C}}{g_{m5}R_LC+C_{gs5}+2C} \quad (16)$$

The number of cascading gain stages used in the LA determines the achievable bandwidth and gain. Assuming that each LA gain stage of Fig. 5 is approximated as a one pole system, the bandwidth of each LA gain stage is related to the number of cascading stages by

$$BW_{-3dB} = BW_0\sqrt{\sqrt[N]{2}-1} \quad (17)$$

where BW_{-3dB} is the bandwidth of the LA, BW_0 is the bandwidth of each individual stage and N is the number of stages. Equation (17) can be modified to include the gain-bandwidth product,

$$BW_{-3dB} = B\frac{\sqrt{\sqrt[N]{2}-1}}{\sqrt[N]{A_{tot}}} \quad (18)$$

where *B* and A_{tot} denotes the gain-bandwidth product and the total gain of the LA, respectively. Equation (18) is plotted with various values of A_{tot} in Fig. 6. These plots reveal that for A_{tot} = 60dB, there is only an 8% bandwidth improvement when increasing the number of stages from 8 to 14. Therefore LAs usually employ no more than 7 cascading gain stages.

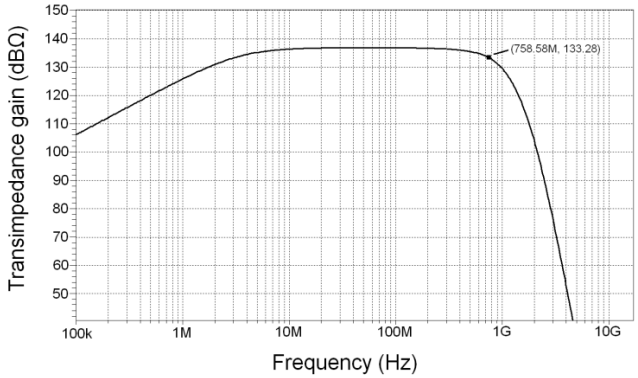


Fig. 7. Frequency response of the proposed optical receiver front-end.

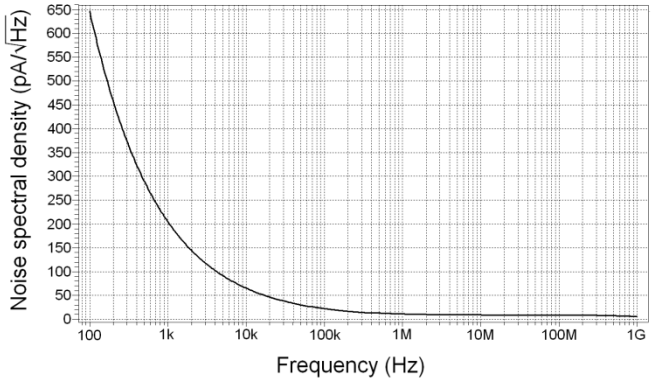


Fig. 8. Input referred noise current spectral density.

IV. Layout and Simulation Results

Simulations were conducted with the proposed optical receiver front-end using 0.35µm CMOS technology. The frequency response of the front-end is shown in Fig. 7 for frequencies between 100 kHz and 10 GHz. The circuit achieved a bandwidth of 750 MHz and a single-ended transimpedance gain of 136 dBΩ.

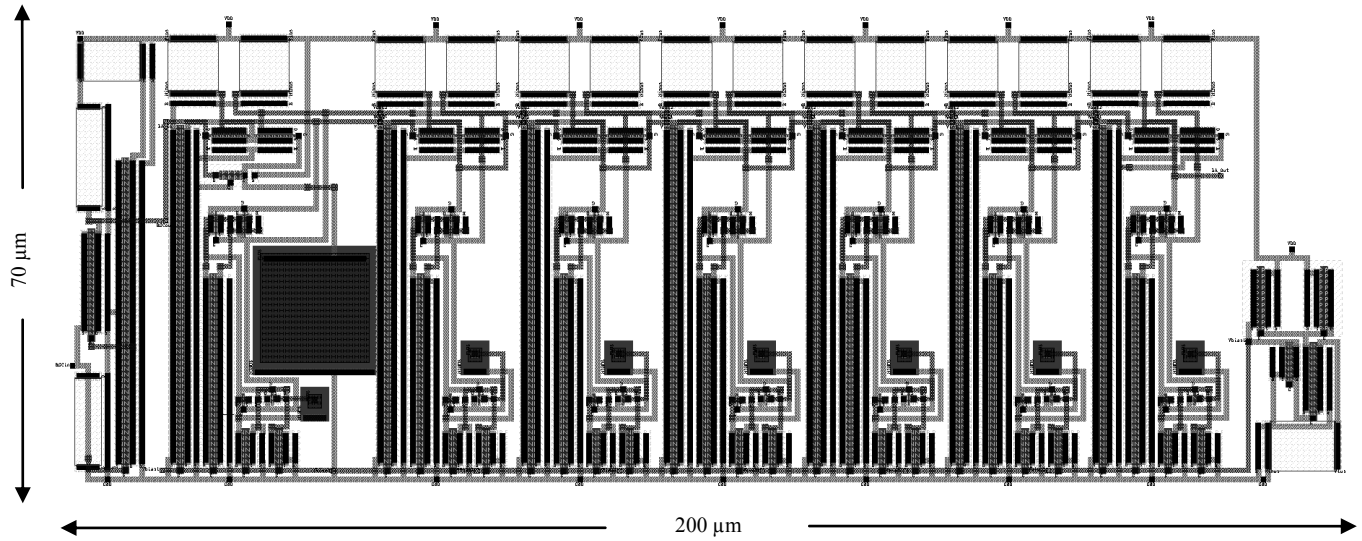


Fig. 9. The layout of the optical receiver front-end.

The simulated input referred noise current spectral density is shown in Fig 8. At 300 MHz the noise current spectral density is measured equal to 8.1 pA/√Hz.

The RGC draws 750 µA from the 3.3 V supply while each stage of the LA draws 1.9 mA. The total circuit power dissipation is therefore 46.2 mW. The layout of the proposed front-end is depicted in Fig 9. The amplifier core occupies an area of 0.014 mm^2.

V. Conclusion

An optical receiver front-end has been realised in 0.35 µm CMOS technology using the RGC topology for suppressing the large parasitic capacitance of the photodiode, together with a LA exploiting the negative resistance configuration and a frequency compensation element for simultaneous gain and bandwidth improvement. Small geometric, high field effects have been considered in the design approach as it may negatively impinge on critical circuit parameters or even prohibit proper circuit operation. In most cases however, simply increasing the power dissipation of the RGC may ensure correct circuit operation under velocity saturation effects. Although this is not an optimal solution, it does ensure a virtual ground at the RGC input, suppressing the photodiode capacitance.

References

[1] Sung Min Park and Hoi-Jun Yoo, "1.25-Gb/s regulated cascode CMOS transimpedance amplifier for Gigabit Ethernet applications," *Solid-State Circuits, IEEE Journal of,* vol. 39, pp. 112-121, 2004.

[2] Han Liang, Ye Yizheng and Bai Tao, "2.5Gb/s CMOS RGC transimpedance amplifier," *Microelectronics, 2007. ICM 2007. Internatonal Conference on,* pp. 19-22, 2007.

[3] Cheng-Ta Chan and O. T. -. Chen, "A 10 Gb/s CMOS optical receiver using modified regulated cascade scheme," *Circuits and Systems, 2005. 48th Midwest Symposium on,* pp. 171-174 Vol. 1, 2005.

[4] Sung Min Park and C. Toumazou, "A packaged low-noise high-speed regulated cascode transimpedance amplifier using a 0.6µm N-well CMOS technology," *Solid-State Circuits Conference, 2000. ESSCIRC '00. Proceedings of the 26th European,* pp. 431-434, 2000.

[5] Zhenghao Lu, Kiat Seng Yeo, Jianguo Ma, Manh Anh Do, Wei Meng Lim and Xueying Chen, "Broad-Band Design Techniques for Transimpedance Amplifiers," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 54, pp. 590-600, 2007.

[6] B. Razavi, Design of Integrated Circuits for Optical Communications. Boston: McGraw-Hill, 2003, pp. 370.

[7] A. A. Abidi. "High-Frequency Noise Measurements on FETs with small dimensions," *IEEE Transactions on Electron Devices*, Vol. 33, no. 11, pp. 1801-1805, November 1986.

[8] Kwisung Yoo, Gunhee Han and Sung Min Park, "A 5.2-mW, 2.5-Gb/s Limiting Amplifer for OC-48 SONET Applications," *Electronics, Circuits and Systems, 2006. ICECS '06. 13th IEEE International Conference on,* pp. 537-540, 2006.

Expanded Adaptive LMS Algorithm for the Removal of Interference in Capacitive Sensing Devices

S.J. van der Merwe, *Student member, IEEE.* J.P. de Villiers, *Member, IEEE*

Abstract — **Several advances have recently been made in capacitive sensing technology. Noise coupling into these systems, especially 50/60 Hz power-line interference, degrades system sensitivity and performance. This interference is not static in mean or frequency and component mismatches as well as environmental changes can cause these quantities to drift. A solution for removing this interference by expansion of the LMS algorithm is presented [1]. Taking the main purpose of the capacitive sensing system into account, a modified version of the LMS algorithm must be able to pass the needed data unchanged. The expanded LMS algorithm presented in this paper passes important data without filtering and re-engages once steady state in the desired signal has been reached. This is achieved by using a slope detection algorithm which is incorporated in the LMS algorithm. A third weight, which influences the sensitivity of the other weights in the system, is added to the LMS algorithm to track the mean of the signal and achieve improved attenuation of the interference component. This third weight can also be utilised as output of the system. This paper presents the mathematical analysis, simulation and evaluation of the expanded LMS algorithm for use in capacitive sensing systems. The algorithm is tested on data from an existing capacitive sensing system.**

***Index Terms*— capacitive sensing, interference, least mean square, slope detection**

I. Introduction

THE recent increase in applications for capacitive sensing devices has stimulated major growth, research and development in the capacitive sensing industry. Devices using this technology include mobile phones, multimedia players, proximity detection sensors, management systems and household appliances. Modern capacitive sensing integrated circuits can detect capacitance changes in the femto Farad range, which makes it ideal for devices used in high sensitivity applications. Owing to their high sensitivity, these capacitive sensing devices have a very high vulnerability to noise and interference. Noise and interference which is coupled into the system affects accuracy as well as sensitivity to small changes in the input. This could lead to unexpected results and false detection.

Power line interference in the 50/60 Hz range is especially common in these systems. Often these components have a higher power than other noise components due to its strong presence in the environment. These noise frequencies are also dynamic in nature and will typically be either 50 Hz or 60 Hz depending on the country in which the device operates. By strategically choosing a sampling frequency of 55 Hz, the difference frequency will always be around 5 Hz, thereby allowing simple implementation of an adaptive interference canceling device.

Several methods exist to remove noise centered on a specific frequency. These include a static notch filter with a wide bandwidth, the LMS algorithm, neural networks and the RLS algorithm.

This paper focuses on the standard LMS algorithm developed by Widrow and Sterns [1] and how the LMS algorithm performance can be improved for this application. An expanded LMS algorithm scheme is presented for 50/60 Hz noise in capacitive based sensors. Fig. 1 shows actual input data at steady state to a capacitive based sensing device, with the power spectrum for this data shown in Fig. 2. This device employs a charge transfer method of capacitive sensing. Charge is continuously transferred from the sensor capacitance into a charge collection capacitor and with the use of processing circuitry, proximity is detected.

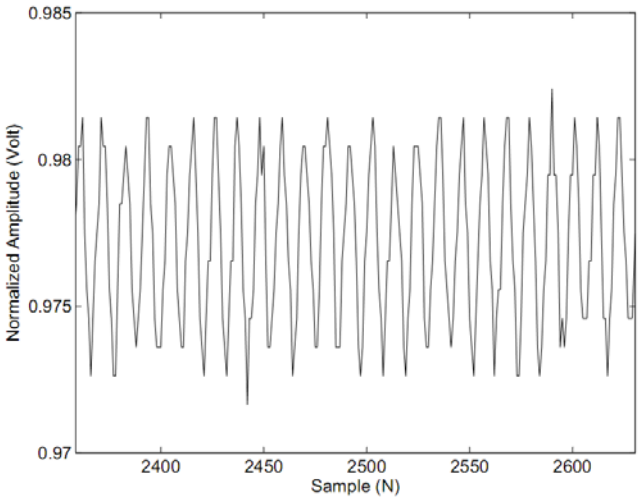


Fig. 1: Input steady state data from a capacitive based sensing system.

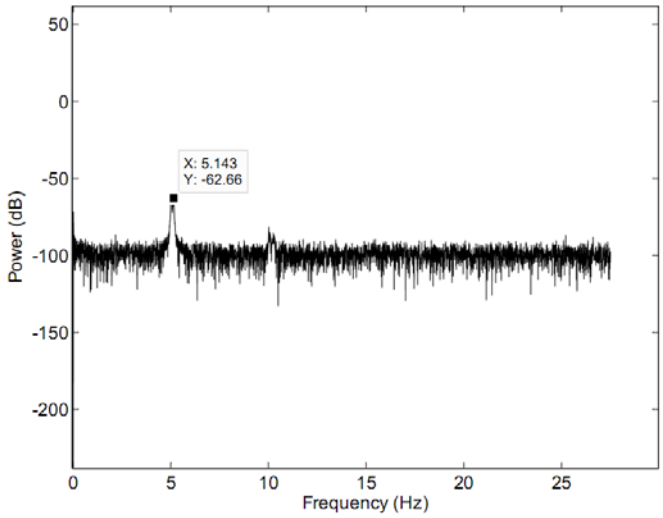


Fig. 2: Power spectrum of the steady state input data in fig. 1.

II. Noise Removal Methods

A. LMS Algorithm

Adaptive signal processing has become more popular owing to the recent progress in microcircuit design for compact, economical and reliable signal processors that rival biological nervous systems in size and are clearly superior to biological system in speed [1].

The LMS (least mean square) algorithm is derived from Wiener filter theory and uses an estimate of the gradient of the performance surface to descend on this surface towards an optimal solution [2]. A widespread application for the LMS algorithm is to use it as an adaptive notch filter acting as an interference canceller. By using a reference signal, which has some correlation with the noise present in the input signal, an adaptive filter adjusts the weights of a digital filter in order to match the reference input after the filter with the corresponding noise in the input [1],[3]. Subtraction of this filtered signal from the input is performed and, ideally, the noise is removed from the input. Fig. 3 shows the general structure of the LMS algorithm.

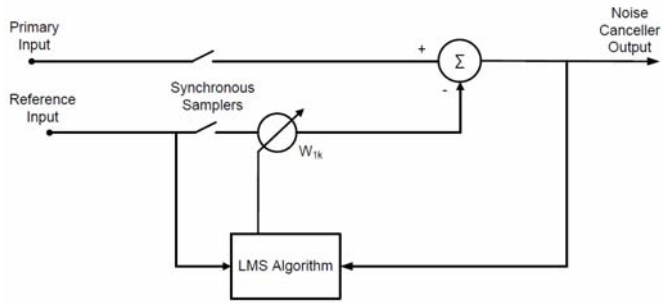


Fig. 3: Basic structure of the LMS algorithm. Figure adapted from [1].

The initial filter weights are updated at each sampling instant. The new weight, after each iteration, is given by

$$W_{k+1} = W_k - \mu\nabla_k = W_k + 2\mu\varepsilon_k X_k \qquad (1)$$

where W_k is the previous weight, μ is a conversion factor, ∇_k is the gradient estimation of the expected value of the error squared, ε_k is the error and X_k is the input from the reference signal [1],[2],[3].

To ensure that the weights reach an optimum solution, the convergence factor is bounded by

$$\frac{1}{\lambda_{\max}} > \mu > 0 \qquad (2)$$

where λ_{max} is defined as the sum of the diagonal elements of the input correlation matrix, represented by

$$R = E[X_k X_k^T] = E\begin{pmatrix} x_{0k}^2 & x_{0k}x_{1k} & \mathrm{L} \\ x_{1k}x_{0k} & x_{1k}^2 & \mathrm{L} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} \end{pmatrix} \qquad (3)$$

X_k represents the reference inputs of the algorithm [1]. For this specific application for removing a sinusoidal interference at a specific frequency, Widrow and Sterns [1] devised an alternative form of the adaptive canceller using the LMS algorithm. To be able to lock onto the phase of the noise, two reference inputs are used, a sinusoid and a cosine signal. The filter weights adapt to the ratio required to obtain phase lock.

III. Expanded LMS Algorithm

There are two problems with the standard LMS algorithm shown in Fig. 3 for application in capacitive sensing devices. The standard LMS algorithm does not make any provision for proximity/contact with the sensor. As a result, the algorithm can either saturate due to a large change in input or important data may be removed if its frequency is in range of the LMS algorithm but falls outside of the noise range. The LMS algorithm also does not contain any provision for varying input data with a DC offset. Both these problems are addressed in the expanded LMS algorithm.

A. DC Weight

The steady state input data in this application is biased by a DC carrier. Thus a method had to be devised to account for this DC component in the error generated by the LMS algorithm. Even if the algorithm is locked onto the signal, there will still be a non-zero error. This DC component may be large in comparison with the amplitude of the noise and may cause overshoot and even saturation in the LMS algorithm. The proposed scheme utilises a third DC weight that is added to the basic LMS algorithm; this weight only influences the error sent back to the LMS algorithm and not the signal. This scheme is shown in fig. 4.

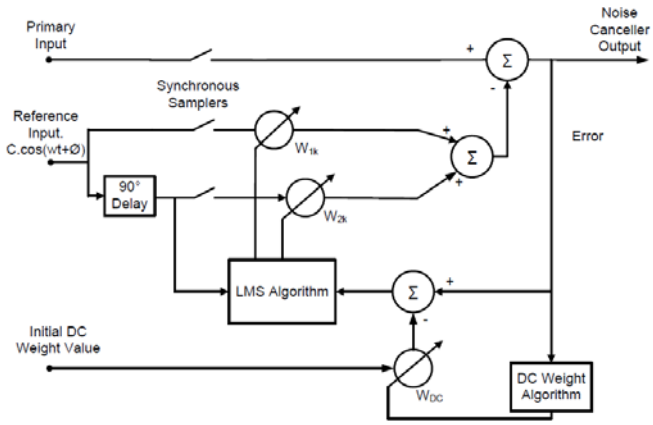


Fig. 4: Expanded LMS algorithm with DC weight. Partially adapted from [1].

The DC weight algorithm is represented by

$$W_{DC(k+1)} = W_{DC(K)} + \alpha\varepsilon_k \qquad (4)$$

where α represents the weight conversion factor and ε_k is the error of the standard LMS algorithm. The DC weight is then subtracted from the LMS error to eliminate offset errors.

The factor α determines the DC weight's sensitivity to errors in the LMS algorithm and also determines the speed at which the DC weights reaches the approximate offset value of the data, as well as initial oscillation and overshoot. Fig. 5 shows DC weight convergence with a starting value of W_{DC} = 0 and a convergence factor of α = 0.01.

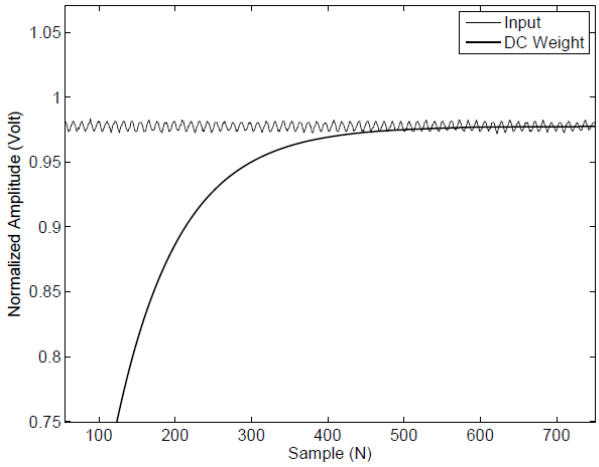


Fig. 5: LMS Algorithm with a DC weight – $W_{DC(1)}$ = 0 and α = 0.01.

It can be seen that the weight takes a long time to converge because of the small convergence factor. Fig. 6 shows the DC weight with an initial weight of W_{DC} = 0 and a large convergence factor of α = 0.5. A result of the large α is the overshoot and oscillation occurring before the weight converges. A possible solution to this problem can be to set the initial weight value equal to the first input value. Fig. 7 shows convergence with α = 0.05 and W_{DC} = *input(1)*. Convergence speed is remarkably improved. The performance of the initial convergence depends on the combination of initial value and the convergence factor.

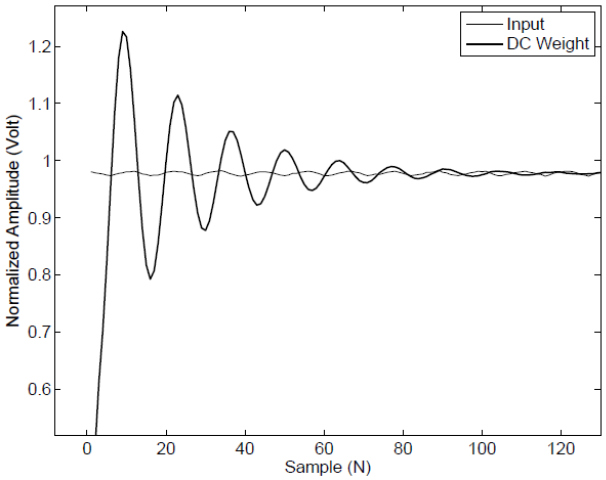


Fig. 6: LMS Algorithm with a DC weight – $W_{DC(1)}$ = 0 and α = 0.5.

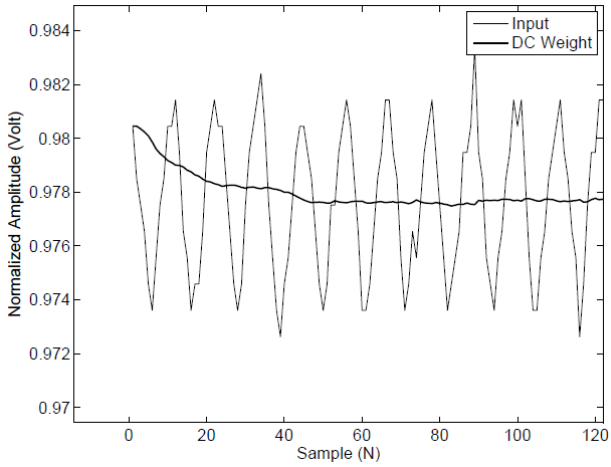


Fig. 7: LMS Algorithm with a DC weight – $W_{DC(1)}$ = input(1) and α = 0.05.

It was noted that the DC weight generated in this scheme, as shown in Fig. 4, was a better approximation of the desired signal and posed some improvements in attenuation of the 50/60 Hz noise. Fig 8 shows the proposed algorithm. The DC weight is used as output of the system. This is equivalent to also subtracting the error signal from the input.

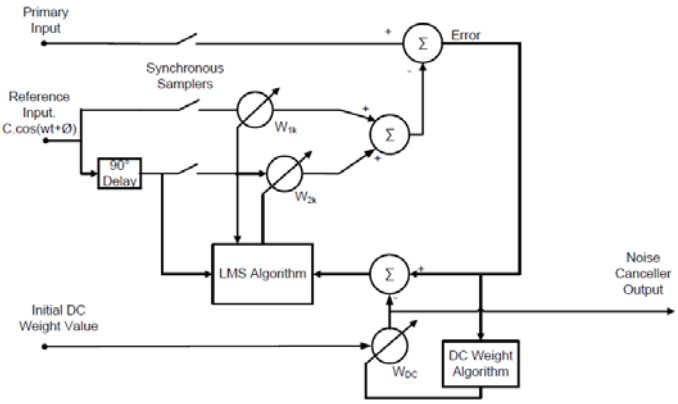


Fig. 8: Expanded LMS algorithm with the DC weight utilized as output. Partially adapted from [1].

B. Slope proximity detection criteria

Typical applications in a capacitive sensing system will be to detect proximity or contact to the sensor. Real data from a capacitive sensing IC is shown in Fig. 9.

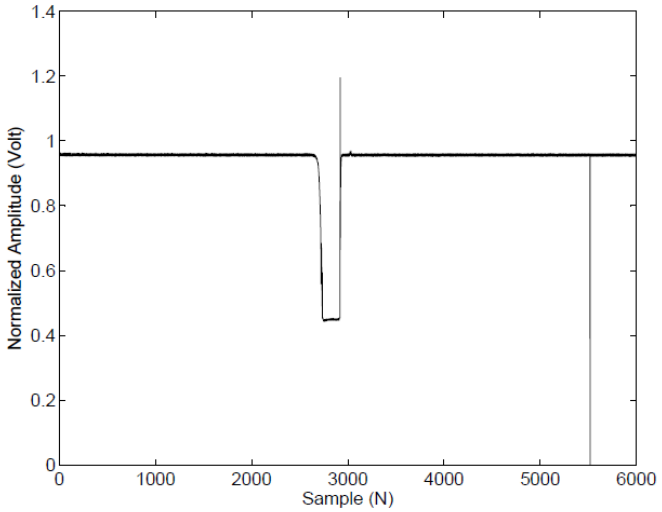


Fig. 9: Input data from a capacitive based sensing system with proximity/contact.

It can be seen in an active (proximity or contact to the sensor) state, input levels decrease rapidly and reach steady state if the sensor is kept active. When contact is removed, the input returns to its original steady state levels.

To detect the active state of the input, a scheme based on the slope of the input signal was devised. The slope of the input signal is defined as the difference between two samples divided by the sample period. A slope threshold β is defined. This threshold determines if the expanded LMS algorithm is active or if input data is passed to the output without any changes. In essence, the expanded algorithm is enabled when the slope of the input signal falls below β and the input is passed without changes to the output if the slope falls above β. In the case of 50/60 Hz sinusoidal interference, the maximum possible slope must be determined. The formula for an analogue sine wave is given by:

$$y = M\sin(2\pi ft) \tag{5}$$

where f represents the frequency in Hertz and t time in seconds. The slope, defined as units per second, is:

$$\frac{\Delta y}{\Delta t} = M2\pi f \quad \text{at} \quad t = \frac{T}{4} \quad \text{or} \quad t = \frac{3T}{4} \tag{6}$$

To account for the sampling frequency, which has an effect on the slope, we determine our slope threshold β as:

$$\beta = \frac{M2\pi f}{f_s} = M2\pi fT_s \tag{7}$$

β determines the sensitivity of the proximity detection criteria to noise in the input signal. From the nature of the input data, β also determines the delay for detection of the active state after proximity or contact with the sensor. There is a trade off between speed of contact/proximity detection and undesired suspension of the expanded LMS algorithm at steady state.

In the case of 50/60 Hz interference, the sampling frequency is 55 Hz. Thus the maximum frequency of the interference will be at 5 Hz. This is the result of frequency shift when sampling.

The amplitude of our data is found to be approximately 6 units and can easily be estimated using a digital peak detection algorithm. Thus, β would be:

$$\beta = M2\pi fT_s = (6)(2\pi)(5)(0.018) = 3.42 \tag{8}$$

To account for additional white noise and frequency shift in the signal we increase this factor to 5. Fig. 10 shows a close up view of input data at proximity/contact to the sensor. The '+' shows where the data is passed without filtering based on the slope proximity detection criteria. It can be seen that the filter re-enables when proximity is kept and steady state is reached.

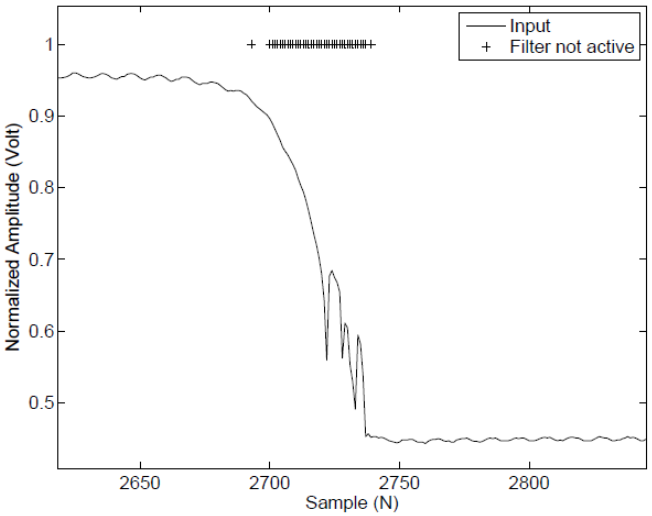


Fig. 10: Input data processed with the slope proximity detection criteria.

IV. Results

A data set from the capacitive sensing system with proximity/contact data was used to obtain results.

A. Standard notch filter

To be able to compare the performance of the expanded LMS algorithm to classical methods, a classic notch filter was implemented. Fig. 11 shows the input data power spectrum. Fig 12. shows the output data power spectrum of the classical notch filter centered at 5 Hz. Attenuation of the 5 Hz component (50/60 Hz signal sampled at 55 Hz) is given by

$$51.77dB - 24.55dB = 27.22dB \tag{9}$$

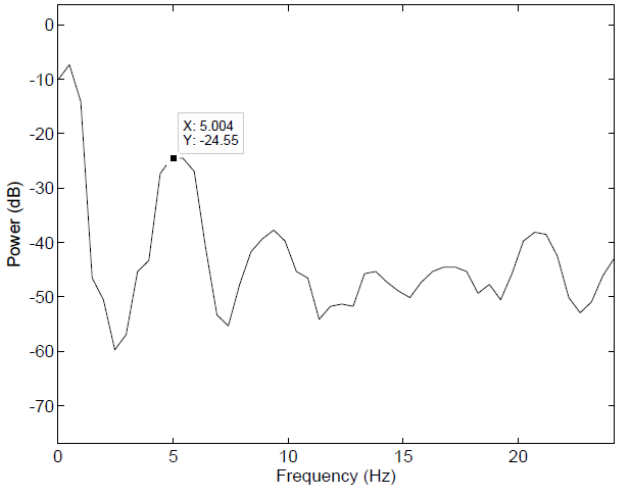


Fig. 11: Input data power spectrum .

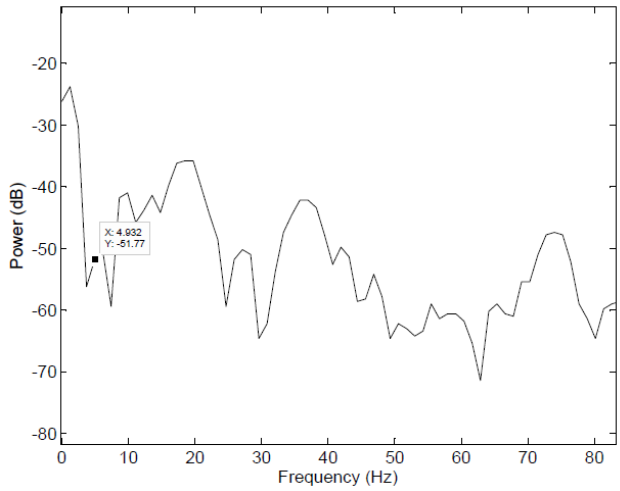


Fig. 12: Output data power spectrum for a classic notch filter centered at 5 Hz.

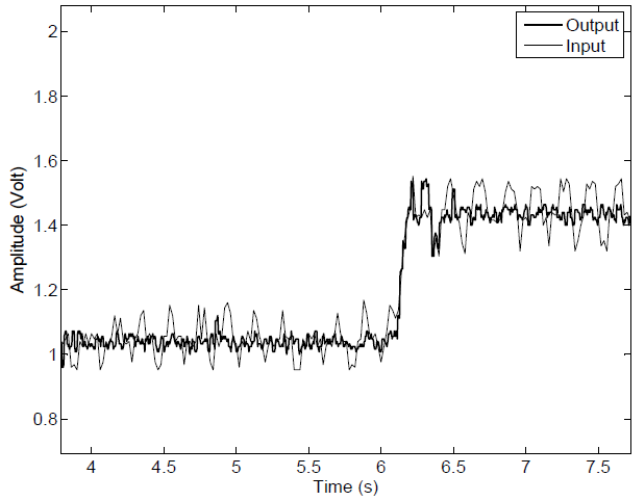


Fig. 13: Input and output data for the expanded LMS algorithm shown in Fig 4.

B. Expanded LMS Algorithm with DC weight

Results for the method in Fig. 4 where the DC weight is subtracted from the error are shown. These results were obtained from the system implemented on a DSP chip. Fig. 13 shows the input and the output for this method. It can be seen that at proximity/contact, the output follows the input.

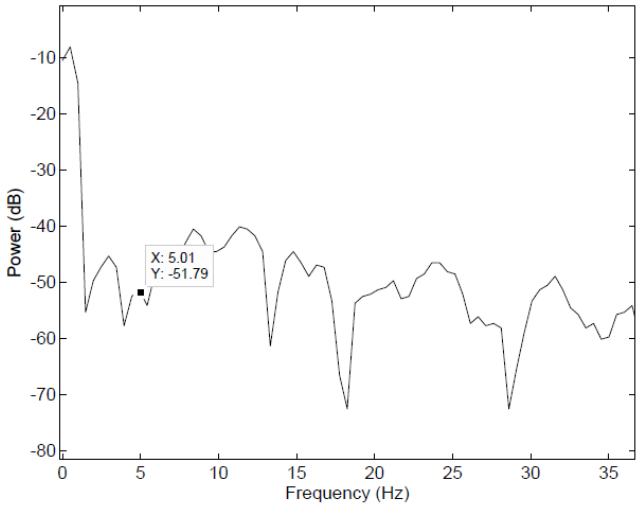


Fig. 14: Output data power spectrum for the expanded LMS algorithm shown in fig. 4.

Fig. 11 and 14 show the power spectrum for the input and output signals respectively. Attenuation of the 5 Hz component (50/60 Hz signal sampled at 55 Hz) is given by

$$51.79dB - 24.55dB = 27.24dB \qquad (10)$$

C. Expanded LMS Algorithm with DC weight utilized as output

Results for the scheme in Fig. 8 where the DC weight is used as output are shown. These results were obtained from the system implemented on a DSP chip. The input data spectrum is shown in Fig. 11. Fig. 15 and Fig. 16 show the input and the output for this scheme. Fig. 17 shows the power spectrum of the output

The slope proximity detection criterion causes the input and output to be equal at proximity/contact. Attenuation of the 5 Hz component is given by

$$69.48dB - 24.55dB = 44.93dB \qquad (11)$$

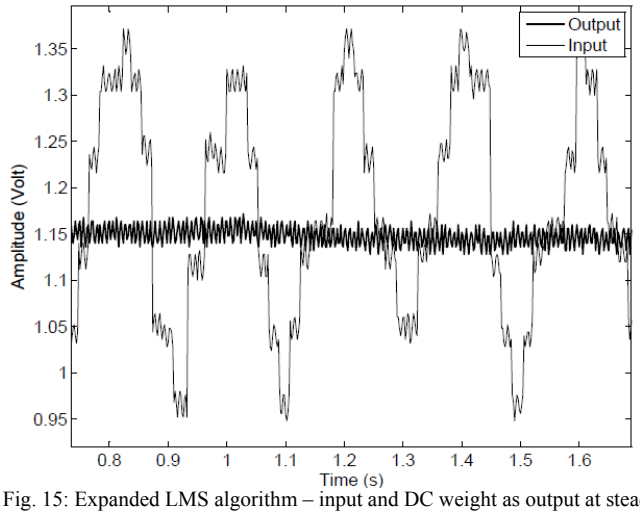


Fig. 15: Expanded LMS algorithm – input and DC weight as output at steady state.

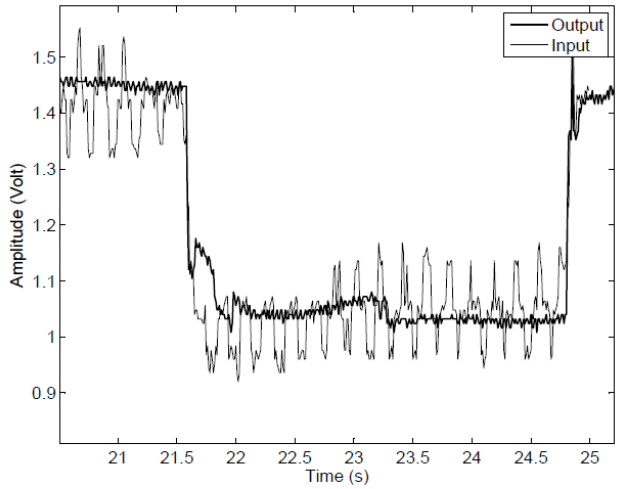


Fig. 16: Expanded LMS algorithm – input and DC weight as output at proximity.

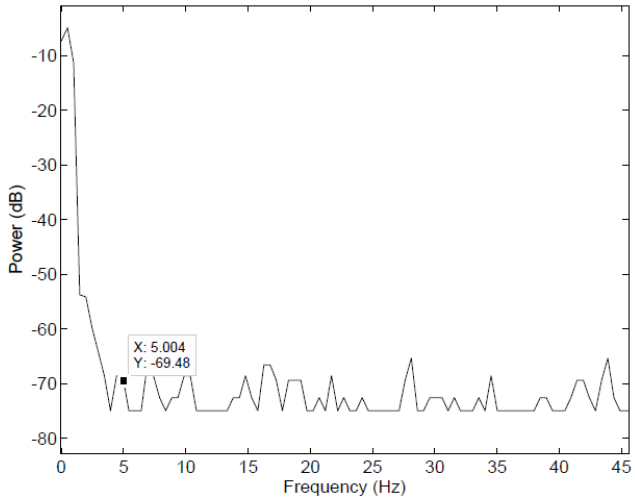


Fig. 17: DC weight power spectrum – output.

This shows an increase in the attenuation of the 5Hz component of 17.69 dB in comparison with the expanded LMS algorithm shown in fig.4. This method also shows an increase of 17.71 dB over classical notch filter methods.

V. Conclusion

Capacitive based sensors have widespread applications in modern technology. Environmental 50/60 Hz power line interference causes a decrease in performance in these devices. Adaptive noise filters, such as the LMS algorithm, does not make provision for data with a DC offset or proximity/contact detection. A method for taking all these factors into account for the specific application in capacitive based sensor devices was proposed. A summary of results obtained is shown in Table 1.

Table 1. Summary of Results

Method	Attenuation (dB)
Classic Notch Filter	27.22
Expanded LMS algorithm – DC subtracted from error	27.24
Expanded LMS algorithm – DC used weight as output	44.93

Although the expanded LMS algorithm where the DC weight is not utilised as output does not offer significant improvement over classic notch filter methods, it provides a basis method for further expansion for this specific application.

The expanded LMS algorithm, where the DC weight is utilised as output, provides improved proximity/contact detection and interference rejection in ultra-sensitive capacitive sensing devices. The expanded LMS algorithm achieves a significant improvement in performance compared to conventional notch filter methods.

References

[1] Bernard Widrow and Samuel D. Steams, *Adaptive signal processing*, New Jersey: Prentice-Hall, 1985.

[2] Bernard Widrow, "Thinking about thinking: The Discovery of the LMS Algorithm," *IEEE Signal Processing Magazine*, pp. 100-106, January 2005.

[3] B. Widrow et al., "Adaptive noise cancelling: Principles and applications," *IEEE Proceedings*, vol. 63, pp. 1692 - 1716 , December 1975

[4] M.L. Ahlstrom and W.J. Tompkins, "Digital filters for real-time ECG signal processing using microprocessors," *IEEE Trans. Biomed. Eng.* vol. BME-32, pp. 708-713, Sept. 1985.

[5] R.E. Barr and E.K.Y. Chan, "Design and implementation of digital filters for biomedical signal processing," *J. Electrophysiol. Tech.,* vol. 13, pp. 73-83, 1986.

[6] E.C. Lfeachor and B.W. Jervis, "*Digital signal processing: A practical approach*," Second ed., USA: Pearson Educated Limited, 2002.

Page left blank intentionally.

Mathematical Analysis of Input Matching Techniques for Application in Wide-band LNA Design

Marnus Weststrate*, *Student member, IEEE*, and Saurabh Sinha†, *Member, IEEE.*
*†Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa
*marnusw@ieee.org, †ssinha@ieee.org

***Abstract*—The objective in low noise amplifier design is to achieve sufficient gain over a required frequency band while maintaining a very low noise figure since the first amplification stage dominates the noise figure of an entire system. Although a steady improvement in transistor speed and device scaling has served to improve noise performance over the past years, the noise figure ultimately depends on the input matching network. Not only is an optimal noise match desirable to minimize noise figure but also a conjugate match to achieve maximum power transfer. This paper presents a comparison of various state of the art input matching techniques that achieve good noise figure and in some cases simultaneous optimal noise and conjugate input matching as well. The linearity of the amplifiers are also taken into account. Finally the possibility of combining certain approaches to achieve a greater degree of freedom in the design as well as improving the important performance measures is discussed. The analysis is aimed at designing a wideband amplifier for the Ku-frequency band using the IBM 0.13 µm 8HP SiGe BiCMOS process which has a f_T of 200 GHz.**

I. Introduction

Due to the vital role of the low noise amplifier (LNA) in any wireless receiver it has remained an active area of research for many years. Even with the vast improvements allowed by the coding schemes employed in wireless communications today the signal-to-noise ratio remains the fundamental limiting factor of data throughput. Since it is well known that the LNA is the determining factor in the noise figure (NF) of a system [1] any improvement in noise figure is of great importance.

Apart from the NF there are four figures of merit commonly used to evaluate LNA topologies. These are amplifier gain specified as S_{21}, the input reflection coefficient (S_{11}), linearity of the amplifier and power consumption.

Noise performance is fundamentally limited by the integrated circuit process used. Since bipolar transistor noise comprise of base and collector current shot noise and base resistance voltage noise it is clear that for a given collector current, a higher current gain (β) helps in reducing the base current and hence the power spectral density (PSD) of the input noise current. A higher unity gain frequency (f_T) helps to increase h_{21} which further reduces the current noise PSD due to the collector current. Finally a lower base series resistance (r_b) results in a smaller PSD of the input noise voltage [2].

Although a steady improvement in transistor speed and device scaling has served to improve noise performance over the past years, the NF ultimately depends on the input matching network [3].

Many amplifier topologies have been demonstrated at various frequency ranges with varying degrees of success [4], [5], [6], [7], [8], [9]. Most LNA designs are aimed at achieving S_{21} of 20 dB and S_{11} < -10 dB. Power consumption generally ranges from 10 mW to 30 mW and IIP3 above 0 dBm is deemed to be high. At present, state of the art NF results vary between 1.5 dB and 3.5 dB, with low noise figures favouring narrow band applications.

Comparing these amplifier topologies however is difficult since they are implemented using different integrated circuit processes and at various frequency ranges. This work provides an analytical discussion of various state of the art LNA topologies for the purpose of selecting the best configuration for a wideband amplifier over the 1 to 20 GHz band. This includes both the ultra wide band (UWB) and the Ku-band. The transistor process to be used in the final design is the IBM 8HP SiGe BiCMOS process with f_T = 200 GHz and β_0 = 600, placing the β cut-off frequency, f_β, at 333 MHz which is below the frequency band of interest.

Since the LNA NF is inversely proportional to β and f_T and proportional to r_b, SiGe HBTs are inherently low noise devices well suited for LNA design.

II. Impedance Matching in the Super High Frequency Range

When designing amplifiers using the common-emitter or cascode configuration in the super high frequency (SHF) range (3 GHz – 30 GHz), also referred to as the centimetre wave range, the transistor is operated far beyond the beta cut-off frequency f_β defined as

$$2\pi f_\beta = \frac{\omega_T}{\beta_0}, \tag{1}$$

where ω_T is the transistor unity gain frequency and β_0 the DC current gain.

Since the β rolls of at approximately -20 dB/decade above f_β it can be shown that β_{RF} is given by

$$\beta_{RF} = -j\frac{\omega_T}{\omega}, \tag{2}$$

for sufficiently high ω [10].

For the case where $\omega >> \omega_T/\beta$ one also finds that the impedance of the parasitic base-emitter capacitance (C_π) becomes small compared to the equivalent input resistance (r_π). This results in the base-emitter impedance (Z_π) being dominated by C_π and thus $Z_\pi \approx 1/(j\omega C_\pi)$.

Due to the finite transport time of carriers from the emitter to the collector some correlation between the base and collector current noise sources exist [2]. However this correlation can be neglected at frequencies much lower than $f_T/2$ [11]. In such cases the correlation admittance seen by the equivalent input noise voltage and current generators is only due to the collector current noise component and equal to Y_{11} of the amplifier two-port.

Input matching is performed by generating an equivalent input resistance using either inductive series-series feedback in the emitter (L_E) of the transistor or shunt-shunt feedback between the collector and base terminals. This, together with a series inductor at the base (L_B) results in an equivalent RLC circuit seen at the input of the amplifier. The input impedance of such a RLC circuit is

$$Z_{IN} \approx s(L_E + L_B) + R_{IN} + \frac{1}{sC_\pi}. \tag{3}$$

The frequency response of S_{11} is then [5]

$$\begin{aligned} S_{11} &= \frac{Z_{IN} - R_S}{Z_{IN} + R_S} \\ &\approx \frac{s^2 + \dfrac{1}{C_{IN}L_{IN}}}{s^2 + s\dfrac{R_{IN} + R_S}{L_{IN}} + \dfrac{1}{C_{IN}L_{IN}}} \\ &\equiv \frac{s^2 + \omega_0^2}{s^2 + s\frac{\omega_0}{Q_{IN}} + \omega_0^2} \end{aligned} \tag{4}$$

where $R_{IN} = R_s$ has been assumed, ω_0 is the resonant frequency of the RLC circuit and Q_{IN} the quality factor (Q-factor). Since S_{11} has a standard notch transfer function and it is required that $|S_{11}| < -10$ dB over the operating frequency range the Q-factor of the circuit is governed by [5]

$$\Delta f_{10dB} = \frac{\omega_0}{6\pi Q_{IN}} \approx \frac{R_{IN} + R_s}{6\pi L_{IN}}. \tag{5}$$

It follows that there is a limit on the maximum value of L_{IN} where $R_{IN} = R_s$ is fixed.

Using this matching technique it is also possible to achieve a simultaneous optimal noise and near conjugate input match at the resonant frequency of the RLC circuit by proper selection of the transistor emitter length, l_e [12].

III. Narrowband Matching Techniques

A. Inductive emitter degeneration technique

Of all traditional input matching techniques inductive emitter degeneration achieves the best noise performance [13] and was first introduced in [14] to generate the real part of the input impedance required for matching. At high frequencies β-multiplication of the emitter inductor (L_E) results in the equivalent impedance

$$Z_{eq} = j\omega L_E(1 - j\frac{\omega_T}{\omega}) = j\omega L_E + \omega_T L_E \tag{6}$$

at the base of the transistor. With an additional series inductor at the base L_B+L_E can be chosen to resonate with C_π at the required centre frequency, providing a real input impedance of $\omega_T L_E$ which can be chosen as 50 Ω providing conjugate matching for a 50 Ω source impedance.

It has further been shown that simultaneous optimal noise and conjugate matching is possible [12] since $R_{s,opt}$ scales inversely with the transistor emitter length according to [10]

$$R_{s,opt} = \frac{\omega_T}{\omega}\frac{1}{l_e}\sqrt{\frac{2(r_b \cdot l_e)}{J_C \cdot w_e}V_T}, \tag{7}$$

where w_e is the emitter width, J_C is the collector current density and V_T the thermal voltage. Thus with J_C selected to set h_{21} the emitter length can be selected to set $R_{s,opt}$ to 50 Ω which is equal to the source resistance. Since the base inductor reduces the imaginary part of $Z_{s,opt}$ to zero an optimal noise match is achieved.

Although this technique is very effective and widely used in narrow band LNAs, it is not well suited for wide band amplifier implementations since the low Q-factor required for the S_{11} specification severely degrades the NF as shown later. It has further been shown [4] that the collector-base feedback capacitance and Miller effect modifies the input impedance and thus in reality a conjugate match is not achieved.

B. Capacitive shunt-shunt feedback technique

A drawback of the emitter degeneration technique is that it requires two area consuming inductors. An alternative method for generating an equivalent input RLC circuit is using capacitive shunt-shunt feedback between the collector and base terminals as shown in Fig. 1 [5].

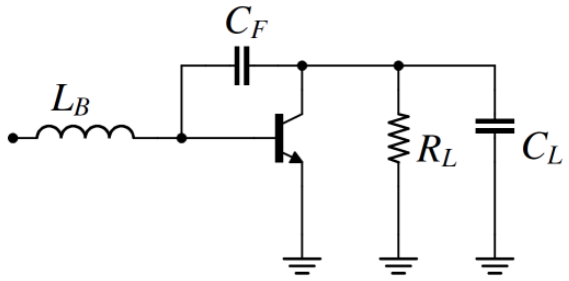


Fig. 1: Circuit diagram of a capacitive shunt-shunt feedback circuit used to produce an equivalent input RLC circuit [5]

Through straight forward small signal analysis it can be shown that the input capacitance and resistance is respectively given by

$$C_{IN} = C_\pi + (1 + g_m R_L)(C_\mu + C_F) \tag{8a}$$

$$R_{IN} = R_M \left(\frac{C_M}{C_\pi + C_M} \right)^2 \approx 50\,\Omega \tag{8b}$$

$$R_M = \frac{R_L}{(1 + g_m R_L)} \left(1 + \frac{C_L}{C_{BC}} \right). \tag{8c}$$

In these equations g_m is the transistor transconductance, C_μ the base-collector capacitance and C_F an intentionally added capacitance in parallel with C_μ to increase the total base-collector capacitance (C_{BC}) where necessary. R_L and C_L are the respective parallel connected load resistance and capacitance.

Compared to the inductive degeneration technique this approach is more area efficient since no area consuming emitter inductor is needed to generate the equivalent 50 Ω resistance.

A combination of inductive degeneration and capacitive feedback has also been used in [8] with a cascode configuration. The use of C_F provided another degree of freedom allowing improved linearity without degrading the NF.

IV. Wideband Matching Techniques

A. LC-ladder input matching network

As an alternative to using a low Q-factor series RLC circuit for input matching a fourth-order doubly terminated bandpass filter can produce a uniform input impedance over an arbitrary bandwidth through proper selection of the reactive elements [15].

A comparison of the filter circuit reveals that the series RLC circuit required for the filter can be generated using the equivalent input circuit of the transistor, and thus a wideband 50 Ω input match can be achieved. This is illustrated in Fig. 2.

A possible disadvantage of this technique is that C_2 (as in Fig. 2) is the constant base-emitter capacitance of the

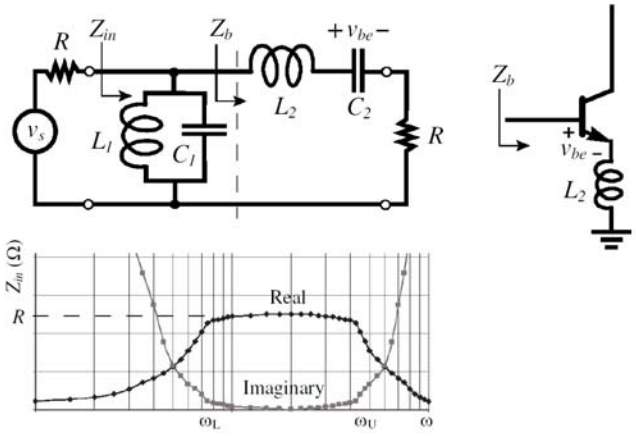


Fig. 2: Fourth-order bandpass ladder filter used to create a wideband 50 Ω input impedance. The series RLC circuit is realized by the input impedance of the transistor [15]

transistor and its dependence on ω_L through

$$C_\pi = C_2 = \frac{1}{\omega_L Z_0} \tag{9}$$

restricts the bias current to

$$I_C = \frac{\omega_T V_T}{\omega_L Z_0} \tag{10}$$

when using the configuration in [15], prohibiting the arbitrary selection of collector current to optimize NF or power consumption.

B. Frequency controlled feedback

In [9], rather than using capacitive feedback, a combination of resistive and inductive shunt-shunt feedback was employed to provide a wideband flat gain.

The configuration uses negative feedback over most of the operating frequency range, however at the high end of the range where the gain starts to roll off with frequency the feedback inductor introduces a controlled amount of positive feedback to compensate for this roll-off, in turn resulting in a flat response over the entire bandwidth. The circuit configuration is shown in Fig. 3.

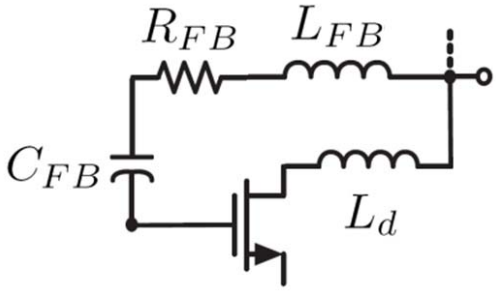


Fig. 3: Inductive and resistive shunt-shunt feedback network achieving a wideband flat gain response [9]

Through proper modelling of the Miller-multiplied feedback network, including the effect of gain roll-off, the effective input impedance was determined. In this case a third-order ladder structure was used to provide conjugate matching, and through proper design of the feedback network a simultaneous optimal noise match was achieved.

V. Performance Measure Trade-offs

A. Noise figure

As mentioned earlier the matched bandwidth of the input impedance when using a narrowband matching technique is extended by lowering the Q-factor of the equivalent input RLC circuit. This however is in conflict with achieving high gain at ω_0 and it can also be shown that the collector current contribution to the NF is inversely proportional to the square of the Q-factor [15].

For a low-Q input RLC circuit $s^2 L_B C_{IN} \approx 0$ holds over the band of interest [5] and hence the NF contribution due to the collector current is given by

$$\begin{aligned} n_{ic} &\approx \frac{g_m}{2} Z_0 \left(\frac{\omega_0}{\omega_T} \right)^2 \\ &\approx \frac{R_S}{2g_m} \cdot (C_\pi + C_\mu)^2 \cdot \omega_0^2 \end{aligned} \tag{11}$$

which, when noting that $Q_{IN} = 1/\omega C_\pi R_S$, neglecting C_μ becomes

$$n_{ic} \approx \frac{1}{2 g_m R_S Q_{IN}^2}. \tag{12}$$

This clearly indicates the trade-off between NF and the input impedance matching bandwidth.

In reality the base current also contributes to the NF and its contribution increases with Q_{IN} [5]. This means that an optimal Q_{IN} exists for achieving a minimum NF and can be set through proper transistor sizing. In addition it is well known that an I_C exists which results in minimum NF since collector current initially decreases with I_C while NF increases with I_B.

This allows use of the LNA optimization technique first formulated in [13] which has been adapted for use with bipolar transistor amplifiers in [5]; however the optimization is still restricted by the Q-factor requirement of the input matching.

Clearly it is desirable to decouple the input matching bandwidth from the input Q-factor to allow the Q-factor to be optimized for NF. The wideband LC-ladder matching network achieves this. The only remaining constraint on the input capacitance now results from ω_L through (9), however since the input capacitance resulting from C_{BC} is Miller-multiplied, the additional C_F included between the base and collector terminals can also be used to modify C_2. This can potentially decouple I_C and ω_L completely. Since the NF is only affected by the actual value of C_{BC} and not the Miller-multiplied value, a large change can be effected in ω_L with only minor impact on the NF. A similar result was found when deriving the NF with C_F included in [8].

The above indicates a possible advantage of using the capacitive feedback method [5] to generate the input RLC circuit compared to the inductive emitter degeneration method used in [15].

The inductive degeneration technique is also impacted negatively by the parasitic feedback capacitance [4] which is usually neglected in calculations. The capacitive feedback technique includes the parasitic capacitance in the design equations and thus no such performance degradation is present.

B. Linearity considerations

While NF places a lower limit on the dynamic range of a LNA the IIP3 determines the upper limit. IIP3 is especially important in wideband LNAs where many signals stronger than the signal of interest may exist at surrounding frequencies compared to the narrowband case where out of band interference is filtered.

In the narrowband case it can be shown that the IIP3 is inversely proportional to the square of f_T [15]. This shows that maximizing f_T in order to minimize NF lowers (worsens) the IIP3 of the amplifier indicating a design trade-off.

In the wideband case, such as in Fig. 2, the IIP3 is defined as [15]

$$V_{IIP3_{LNA}} \approx 4\sqrt{2} V_T \cdot \left(1 + \left(\frac{I_C \omega L_E}{V_T} \right)^2 \right)^{3/4} \tag{13}$$

showing that the IIP3 improves with frequency in contrast to the NF, and the linearity can be improved by increasing the feedback through L_E.

It was shown in [8] that feedback through an additional C_F also improves the IIP3. However, as discussed above C_F has a small effect on NF compared to the much larger effect of L_E. This makes the technique of using C_F as a feedback element preferable.

The question arises as to whether the cascode configuration should still be used with the feedback capacitance which effectively negates the advantage of high reverse isolation, compared to simply adding the additional feedback capacitance between the base and collector of a common-emitter stage.

Regarding linearity the cascode configuration reduces the voltage drop on the non-linear parasitic C_μ, as well as the avalanche multiplication which both serve to improve linearity [10]; and since the added C_F is bias independent (linear) improved linearity should be observed when comparing the cascode to the common-emitter case.

C. Amplifier gain comparison

Although the first stage of a LNA typically focuses on achieving a low NF since a further amplifier stage can provide additional gain, the gain of the first stage remains important as it serves to reduce the noise of a second stage when referred to the input. In addition it is desirable to use a single stage LNA wherever possible to reduce physical size and power consumption.

As discussed, the magnitude of the current gain at the frequency range of interest for the applicable transistor process is given by $\beta = \omega_T/\omega$ and subsequently the power gain also increases with ω_T [10]. This satisfies intuitively why IIP3 would decrease with f_T, which is a result of the higher gain. Increasing f_T however requires increased J_C which degrades the NF and thus NF must often be sacrificed to achieve a required gain specification.

Since the transistor β and thus the gain rolls off with frequency, an inductive load is usually required to equalize the voltage gain.

A disadvantage of the techniques using capacitive feedback is that the gain is reduced further by the additional feedback. This will result in less reduction of the NF of a second amplification stage, which may be required due to this reduction in initial gain, when referred to the input. The resulting higher input referred noise will worsen the NF. In severe cases this may negate the other improvements achieved through this technique as discussed earlier. This indicates the importance of considering all relevant design constraints in selecting the best amplifier topology.

VI. Conclusion

Due to low Q-factor standard emitter degeneration configurations are not ideal for achieving a wideband low NF as well as the S_{11} specification. This suggests that the LC-ladder input matching technique is the optimal choice in wideband LNA design.

The feedback offered by C_F has proven to be an effective means of both generating the equivalent input RLC circuit, as well as improving linearity without degrading the NF and it also requires one less inductor reducing the physical size of the amplifier.

Therefore the optimal amplifier configuration which should be chosen for the implementation appears to be a LC-ladder matched amplifier with capacitive shunt-shunt feedback. The use of an emitter inductor will be considered if suitable linearity is not attained at the cost of NF.

Acknowledgment

The authors would like to thank ARMSCOR, the Armaments Corporation of South Africa Ltd, (Act 51 of 2003) for sponsoring this study.

References

[1] H. T. Friis, "Noise figure of radio receivers," *Proceedings of the IRE*, vol. 32, no. 7, pp. 419-422, July 1944.

[2] G. Niu, "Noise in SiGe HBT RF Technology: Physics, Modelling, and Circuit Implications," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1583-1597, September 2005.

[3] H. A. Haus et al., "Representation of noise in linear two-ports," *Proceedings of the IRE*, vol. 48, no. 8, pp. 69-74, January 1960.

[4] B. Min and G. M. Rabeiz, "Ka-band SiGe HBT Low Noise Amplifier Design for Simultaneous Noise and Input Power Matching," *IEEE Microwave and Wireless Components Letters*, vol. 17, no. 12, pp. 891-893, December 2007.

[5] Y. Lin, H. Chen, T. Wang, Y. Lin and S. Lu, "3-10-GHz Ultra-Wideband Low-Noise Amplifier Utilizing Miller Effect and Inductive Shunt-Shunt Feedback Technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 9, pp. 1832 - 1843, September 2007.

[6] G. D. Nguyen, K. Cimino and M. Feng, "A RF CMOS amplifier with optimized gain, noise, linearity and return losses for UWB applications," *Proceedings of the IEEE Radio Frequency Integrated Circuits Symposium*, 15-17 June 2008, Atlanta, USA, pp. 505-508.

[7] A. Chen, H. Liang, Y. Baeyens, Y. Chen and Y. Lin, "A broadband millimeter-wave low-noise amplifier in SiGe BiCMOS technology," *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, 23-25 January 2008, pp. 86-89.

[8] B. Kang, S. Yang, J. Yu, W. Choo and B. Park, "Design and analysis of a high-performance cascode bipolar low noise amplifier with shunt feedback capacitor," *IEEE Radio Frequency Integrated Circuits Symposium*, 15-17 June 2008, Atlanta, USA, pp. 613-616.

[9] Y. Soliman, L. MacEachern and L. Roy, "A CMOS Ultra-wideband LNA Utilizing a Frequency-Controlled Feedback Technique," *Proceedings of the IEEE International Conference on Ultra-Wideband*, 5-8 September 2005, Zurich, Switzerland, pp. 530-535.

[10] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors,* 1st ed., Artech House, Inc., 2002.

[11] J. P. Roux, L. Escotte, R. Plana, J. Graffeuil, S. L. Delage and H. Blanck, "Small-signal and noise model extraction technique for heterojunction bipolar transistor at microwave frequencies," *IEEE Transactions on Microwave Theory Technology*, vol. 43, no. 2, pp. 293-297, February 1995.

[12] S. P. Voinigescu et al., "A Scalable High-frequency Noise Model for Bipolar Transistors with Application to Optimal Transistor Sizing for Low-Noise Amplifier Design," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 9, pp. 1430-1439, September 1997.

[13] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS low noise amplifier," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, pp. 745-759, May 1997.

[14] A. van der Ziel, "Noise in solid-state devices and lasers," *Proceedings of the IEEE*, vol. 58, no. 8, pp. 1178-1206, August 1970.

[15] A. Ismail and A. A. Abidi, "A 3-10-GHz low-noise amplifier with wideband LC-ladder matching network," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 12, pp. 2269-2277, December 2004.

[16] Q. Liang, G. Niu, J. D. Cressler, S. Taylor and D. L. Harame, "Geometry and bias current optimization for SiGe HBT cascode low-noise amplifiers," *IEEE International Microwave Symposium Digest,* 2-7 June 2002, Seattle, USA, vol. 1, pp. 517-520.

Development of a Next-generation Commercial Computer Aided Design (CAD) Tool

J. Pool, H. R. Gerber, J. J. D. Bull, C. J. Fourie, *Member, IEEE*, H. G. Marais, C. W. C. Oliver,
T. A. J. van der Walt and W. J. Perold, *Senior Member, IEEE,*
Department of Electrical & Electronic Engineering, University of Stellenbosch, Stellenbosch, South Africa

***Abstract*—We introduce NioCAD, a next-generation commercial Computer Aided Design (CAD) tool for the superconductive Electronic Design Automation (EDA) industry, and discuss the key development aspects needed to produce a tool that addresses all the requirements of a new technology.**

NioCAD is developed to be generic, so that it can be adapted rapidly to the changing requirements of a new technology. Cross-platform operation, currently on Windows, Linux and Mac OS X systems, results in a modern platform with a modular, completely scriptable architecture that makes it possible and easy for advanced end-users to implement custom extensions to the suite. NioCAD's data model allows modules to operate on the same set of data, stored in a database with full version control, and enables users to traverse designs backwards and forwards without loss of consistency.

NioCAD is the first fully integrated suite of tools that enables superconductive electronic engineers to rapidly design, analyse, optimise, layout and verify Rapid Single Flux Quantum (RSFQ) electronic circuits that are ready for fabrication. The suite's capabilities include custom component creation; support for multiple SPICE engines; version controlled component, circuit, sub-circuit & cell library management; intuitive schematic circuit design, analysis and yield optimisation; guided physical mask layout with interactive electrical rule checking, design rule checking, extraction of estimated component values and extraction of parasitic components. All these capabilities are tightly integrated to ensure a fast round-trip design cycle that produces a manufacturable physical integrated circuit that corresponds to the high-level design.

***Index Terms*—CAD, RSFQ, superconductive electronics, layout, simulation, physical implementation, inductance extraction, rapid design, verification**

I. Introduction

NIOCAD is the first integrated suite of Electronic Design Automation (EDA) tools specifically designed for Superconducting Electronics (SCE). Currently, SCE designers have to use custom and modified semi-conductor EDA tools [1]. These solutions typically consist of a number of stand-alone tools that are not interoperable or offer only limited integration support for superconductive technologies.

NioCAD is a next-generation Computer Aided Design (CAD) tool that offers a modern Graphical User Interface (GUI), advanced graphical editing, tight module integration, scripting, interactive design feedback and versioning to users, while providing a modular and extendable architecture with a rich set of frameworks to developers.

NioCAD supports superconductive analogue and digital circuits, such as Rapid Single Flux Quantum (RSFQ) [2]. RSFQ differs from conventional semi-conductor transistor electronics in a number of important ways:

- Circuit operation is described by quantum mechanical equations, rather than electromagnetic equations.
- RSFQ uses the Josephson Junction (JJ) [3] as the active component instead of transistors.
- Digital signals are represented by ultra-short electrical pulses instead of voltage levels.
- RSFQ is highly sensitive to parasitic inductance and capacitance in circuit layouts and requires more accurate parameter extraction.
- Signal propagation between logic elements requires carefully engineered transmission lines, and not mere metallic interconnects. Propagation distance is limited, which places serious restrictions on circuit layout.

These and other differences make it extremely difficult, unproductive, expensive and error-prone to develop SCE circuits using the tools available for semiconducting circuit design.

NioCAD allows the creation of small, medium and, eventually, large scale circuit designs [4], [5]. Additionally, the product also aims to provide the following:

- Quick turnaround time for designs that work the first time.
- Closed-cycle design process.
- High levels of integration and automation.
- Ease of use.
- Easily extensible architecture.

NioCAD, funded by the South African National Research Foundation's Innovation Fund, has been under development at the University of Stellenbosch since February 2007 by a dedicated team of software developers and electronic engineers. The development team works closely with SCE researchers and designers at the university to ensure that the product addresses the SCE designer's needs.

NioCAD's first public release will include support for multiple Simulation Program with Integrated Circuit Emphasis

Manuscript received March 3, 2009. This work was supported by the National Research Foundation's Innovation Fund.

(SPICE) engines, SPICE component design, schematic circuit design, circuit simulation and analysis, circuit optimisation, physical layout design (with design rule checking, electrical rule checking, schematic versus layout comparison and parameter extraction) and logic-level circuit design and simulation.

In the following sections we provide an introductory overview of *NioCAD*'s architecture, the frameworks it is build on, the current functionality available in the application suite and the future direction of *NioCAD*.

II. Platform: Architecture & Frameworks

A. Modular Architecture

NioCAD was designed and implemented from scratch. The development team created a dynamic, modular architecture that can easily be extended by the development team, third party tool developers and end-users.

The Model-View-Controller (MVC) design pattern [6] is used throughout the system. The pattern separates the data model from the way the data is viewed, while the controller manipulates the data model. This makes it easier to modify the data model, visual representation and functional logic in isolation.

NioCAD's core platform is build using the Java [7] programming language and the Java Swing GUI Application Programming Interface (API). This makes it possible for *NioCAD* to run with minimal additional effort on Windows, Linux and Mac OS X systems, while providing a native look and feel to users accustomed to these systems.

The basic capabilities of the core platform include a generic extensible data model, data persistence, entity (components, circuit, sub-circuit, physical layouts, etc.) versioning, graphical editing, user manageable views (docking), collections of views (perspectives), management of design projects, integrated scripting, undo/redo support, configuration and customisation, internationalisation and localisation, built-in help and copy protection.

Besides the functionality provided by the core platform, everything else is implemented as modules. Modules can stand alone or may build on each other. A module extends the data model; adds controllers to manage the data; adds views to visualise the data; may define view collections; adds, extends or replaces GUI actions; adds, extends or replaces menu items and can define services or extend existing services.

Services are global collection points for common requests required by modules in the system. There are certain predefined services such as:

- Communications service: used for message passing between components.
- User interface service: a common interface that ensures that the MVC paradigm is enforced.
- Persistence service: an interface between the actual data storage mechanism and the various data controllers.

Services can also be added dynamically to the run time environment which allows for an extendable environment: a newly defined service required by some module can simply be registered when the module is loaded. This new service is then also available to all other loaded modules in the system.

B. Data Model, Persistence and Version Control

The basic data entity, called a Graph Base Item (GBI), is used as a single node in a multi-leaf tree structure. The node (GBI) consists of a list of children and an attribute map. The list of children nodes allows for the recursive definition of a tree structure while the attribute map is used for data association. Any available data type is stored in the attribute map including references to GBI nodes. This model allows for a directed graph with flexible data association for each node.

In order to persist (store and retrieve) data, it is only needed to know how to read and write a GBI; its attributes and all of its children. After that the framework relies on Java's serialisation to persist objects. A complete hierarchy of GBIs can be constructed from its parent object.

The application's data is currently persisted to a database, but abstraction of the data layer makes it possible to support other data stores such as Extensible Markup Language (XML) [8]. By using a database for persistence, it is possible to support a centralised data repository for workgroups.

This generic data scheme makes it easy to extend the data model and it significantly simplifies the management and persistence of data objects. A disadvantage is the additional overhead required to present data.

NioCAD supports multiple versions of all stored entities. This makes it possible to create a new, possibly incompatible, version of an entity, while allowing existing designs to use the previous version. This ensures that older designs remain functional while the user implements improved versions of existing components and circuits.

C. Graphical Editing Framework

NioCAD's Graphical Editing Framework (GEF), which was forked from the Tigris GEF [9], is a central part of the application framework. It provides functionality to create, manipulate and manage graphical objects on a canvas.
The GEF is build on top of Java 2D and offers the following functionality:

- A canvas with vertical and horizontal rulers marked in real-world units.
- Pan and zoom capabilities.
- Ordered drawing layers and intra-layer draw order (z-ordering [10]).
- Configurable grid with custom settings for each layer along with optional grid snapping.
- Drawing of primitive shapes such as ellipses, arcs, rectangles, rounded rectangles, lines, polygons, paths, splines and text figures.
- Formatting of figures include line colour, type, width, join type and transparency; fill colour, pattern and transparency. Additional text formatting options include font type and size; bold, italic and underline emphasis and text colour.

- Affine transformations such as translation (move), rotation, flipping and scaling.
- Alignment and space distribution of figures.
- Recursive grouping of shapes.
- Powerful selection capabilities, including selections within groups and selections of figures lying on top of each other.
- Intersection, subtraction, unification and exclusive-or operations on shapes.
- Conversion of any shape to a polygon.

D. Scripting

NioCAD provides scripting interfaces to most of the functionality offered. This makes it possible and easy to automate repetitive tasks.

Scripts can be executed from within the GUI and can manipulate designs in the current workspace or can be used without the GUI for purposes such as batch processing.

The Java virtual machine provides integrated support for many scripting languages, such as JavaScript, Python, Ruby and Groovy. This makes it considerably easier to add scripting support to applications. Scripts have full access to all classes in the class path and objects instantiated in the virtual machine.

Most of the functionality provided by *NioCAD* is wrapped by simple tools which provide a common interface to the GUI components and the scripting interfaces. Besides simplifying the application's design, it also ensures that the data models and all views are synchronised.

E. Parallel Processing

In EDA tools there are many operations that can be processor intensive. In *NioCAD* this includes margin analysis, yield analysis, yield roll-off analysis, circuit optimisation, physical parameter extraction and design rule checking.

For operations that are parallelizable, *NioCAD* makes full use of multi-processors and multi-core systems.

III. NIOCAD SUITE

The *NioCAD* suite provides a number of tightly integrated tools to assist superconductive electronic engineers to design, analyse, optimise, layout and verify SCE circuits that are ready for fabrication. Users can start from scratch, but typically they will make use of existing cell libraries.

The design team aims to make the application easy to use, increase the speed of circuit design while reducing the failure rate of fabricated designs. This is done by providing an intuitive, modern and familiar user interface; automating repetitive tasks; providing assistance and guidance where possible and reporting errors to the user.

The suite currently consists of a SPICE component designer, an analysis designer, a model designer, a schematic designer with analysis capabilities specifically aimed at superconductive electronics, a circuit optimisation framework and a physical mask layout designer.

A. Entity Designers

The entity designers are used to create new components, analysis types and models for one or more SPICE engines. With these tools a user can add support for additional SPICE engines by creating new entities (when a component is unique to a specific SPICE) or by updating existing entities (components which are similar, but have different grammars). For each entity, support for one or more SPICE engines can be added by simply defining a translation grammar. *NioCAD* currently supports SPICE3F5 [11], JSim [12], [13], JSpice [14] and its commercial successor WRSpice [15]. The component designer is shown in Fig. 1.

From a data and SPICE grammar point of view the component, analysis and model entities are similar. The only difference is that components require a visual representation for the schematic designer and connection points.

A new entity is defined by specifying a number of properties as well as how those properties will be translated for a SPICE engine. *NioCAD* supports text, numerical, constants lists, model, property reference, node reference and source reference properties. The SPICE translation grammar supports required (logical and), optional, logical or and constant fields. The grammar and property information are used by the schematic designer to dynamically create an editing dialog for an entity. In order to aid the user with the translation process, a graphical grammar builder is provided along with a preview of the grammar.

For components, the user must create a symbol to represent the component in the schematic designer. The symbol contains a figure, the connection nodes and labels (component name, value, etc.). The user has full control over the look of the symbol (line thickness, line type, line colour, fill colour, fill pattern and transparency), the labels to show, as well as where nodes and labels are placed.

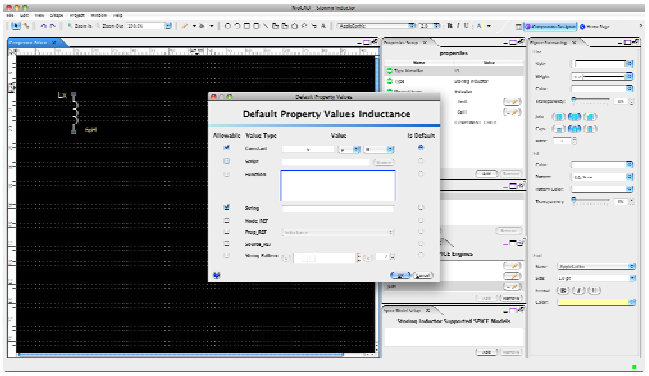


Fig. 1. Modifying the "Inductor" component with the component designer. The component's "Inductance" property is being edited in the dialog [16].

B. Schematic Designer

The schematic designer is used to design circuits and sub-circuits. The user can select which of the supported SPICE engines to use for simulation. This choice determines what components and sub-circuits are available to the user. Fig. 2 shows the design of a Destructive Read-out (DRO) circuit in the schematic designer.

Circuits are designed graphically by placing components on the canvas, editing their values and connecting them together. The connection logic ensures that the user cannot make illegal connections.

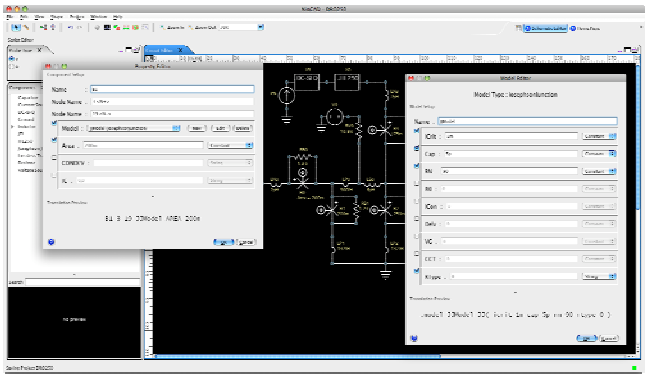


Fig. 2. Designing a DRO circuit with the schematic designer. The property editors of one of the JJs as well as a JJ model are shown [16].

The schematic designer is also used to create sub-circuits. A sub-circuit is a circuit that can be used as a component in another circuit. A sub-circuit therefore needs a symbolic representation and connection nodes. The component designer views are used to define these.

NioCAD provides a generic method to support parameterised sub-circuits that are independent of the SPICE engine used. This makes it possible to create parameterised sub-circuits even if the SPICE engine does not support it. The functionality is achieved by generating a unique sub-circuit definition for each combination of parameters before simulation.

C. *Analysis Capabilities*

NioCAD provides a number of circuit analysis methods. The most fundamental analysis is a single SPICE simulation. SPICE analysis results are presented graphically.

Once the user is satisfied that the basic operation of a circuit is correct, margin, yield and yield roll-off analyses can be performed on the circuit [17]. These analyses work by modifying the circuit's component values in some way, simulating the modified circuit, followed by determining if the circuit operates correctly.

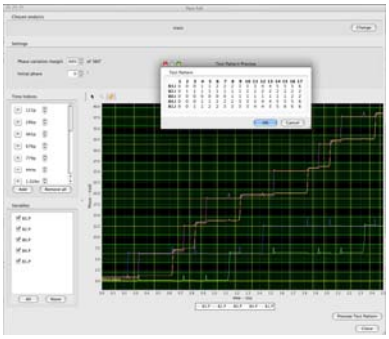


Fig. 3. Setting up the PFC. The resulting test pattern is also shown [16].

NioCAD implements a simple Pass-Fail Criteria (PFC) check for RSFQ circuits. Each time a JJ switches the phase difference across its terminals jumps by 2π radians. Therefore, by looking at the phase of a JJ over time, one can determine if the JJ switched. To configure a PFC, Fig. 3, the user selects the important JJs and specifies the times when to compare the phase values. From this a matrix comprising of JJs and their phases are constructed. The resulting matrix is then used as the template for comparison in future analyses.

Fig. 4. Margin analysis configuration together with the results [16].

Margin analysis, Fig. 4, is used to determine the operating margin of the circuit if a single component's value is varied around its nominal value. The user can select the components, models or group of components to vary as well as specifying the percentage to vary each of the selected values. The result is presented to the user as a set of horizontal bar graphs.

The yield analysis is used to determine the percentage of circuits that will operate correctly if all the component values vary randomly (determined using process tolerances) around their nominal value [18]. The user can select the components, models or group of components to vary, specify the percentage to vary each of the values and specify a variance that will be applied to all the components. The result is presented to the user as a percentage of working circuits and a bar graph that indicates the outcome for each circuit.

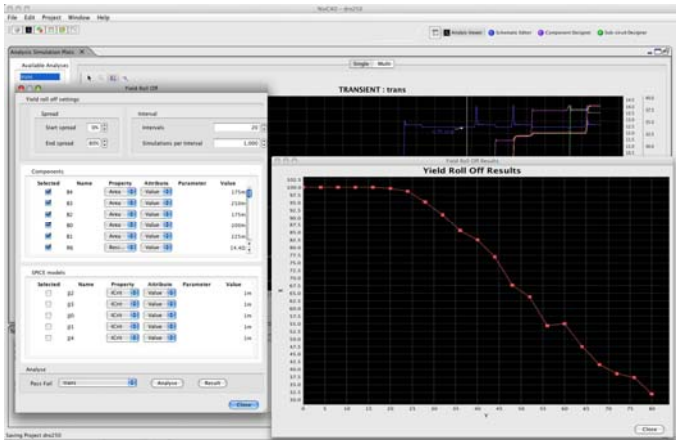


Fig. 5. Yield roll-off configuration together with the results [16].

The yield roll-off, Fig. 5, is used to determine what percentage of circuits will operate for different variances. A common indicator of circuit quality is to determine what the variance is for a 50% yield. A larger variance indicates a better circuit.

For the margin, yield and yield roll-off analyses, the user can inspect the SPICE deck and results, Fig. 6, of each of the generated circuits individually to determine why it failed or succeeded.

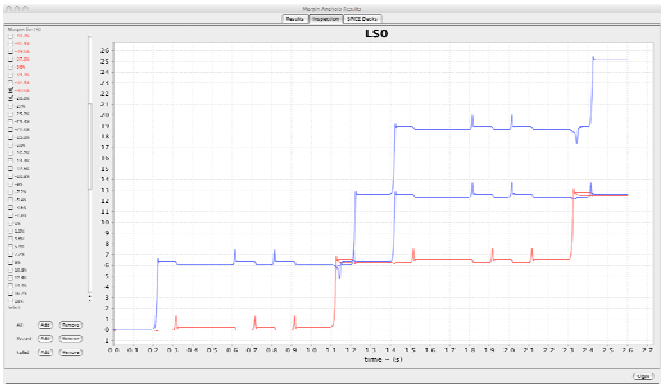


Fig. 6. Inspecting the margin analysis results. A correct and incorrect transient analysis is shown [16].

D. Optimisation

A generic schematic circuit optimisation framework is provided. The framework supports multiple fitness measures, with methods to combine the fitness values into a single fitness measure, as well as multiple optimisation methods [19], [20]. The optimiser allows the user to select the components to optimise, select the fitness measures to use, configure the selected fitness measures as well as selecting and configuring the optimisation method to use.

During the optimisation iterations, the best fitness and overall fitness variance is displayed as a real-time graph. The iterations will terminate if the best fitness value does not change more than a specified value over a number of iterations. Alternatively, the user can terminate the optimisation.

After completion, a margin, yield and yield roll-off analysis is performed to compare the old and new component values. Along with the analysis results, the old and new component values are displayed to the user. If the new values are acceptable to the user, it can be applied to the circuit.

NioCAD provides an API to enable third parties to develop their own fitness measures and optimisers.

E. Physical Designer

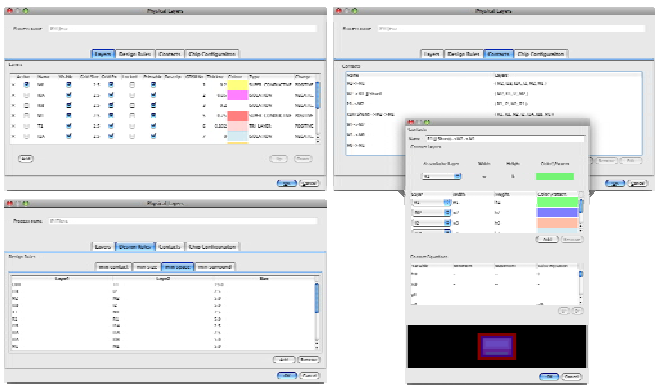


Fig. 7. Various dialogs for configuring the IPHT process' layers, design rules and contacts [16].

The physical layout designer is used to create the fabrication mask for the integrated circuit. *NioCAD* currently supports the Hypres [21] and the Institute of Photonic Technology (IPHT) [22] processes. However, support for additional processes can easily be added by the user, see Fig. 7.

A physical layout is associated with a schematic circuit or sub-circuit. For a single circuit, one can associate multiple physical layouts. For example one can create a layout for more than one process or create different versions using the same process.

The schematic and physical layouts use the same data model. This tight integration makes it possible to perform schematic-versus-layout verification and electrical rule checking continuously while the designer creates the layout. Furthermore, the user can only create physical realisations of components that are present in the schematic, while enforcing legitimate connections between components.

The list of available schematic components is displayed in a panel during a physical design. The user can then either drag a component to the canvas (predefined components or sub-circuits/cells) or define custom dimensions by drawing them (resistors and inductors). A component is drawn as a rectangle, path or polygon on one or more process layers. An icon in the component panel will indicate if a component is correctly defined. A component's expected value (as defined in the schematic) and extracted physical value is also displayed in the component panel. Fig. 8 shows the layout of a Josephson Transmission Line (JTL) in the physical designer.

A component knows where other components may connect to them. The user is graphically informed of connections that are missing (rubber bands are drawn between components) or illegal (red highlighting over illegal connections).

Contacts between layers (vias) are predefined for each process. These predefined contacts adhere to the process design rules. Having predefined contacts makes it much easier to create connections between layers. A contact is associated with a component. The contact resizes along with the component and is removed with the component. The user can define custom contacts for a process, see Fig. 7.

A generic design rule checker, which can check for minimum size, contact, spacing and surround rule violations, is available. The user has the option to add, remove and modify the design rules. Design rule violations are listed in a special issue view and also displayed on the canvas. From the issue view the user can click on a violation to jump to its location on the canvas. The user can also specify that the design rule checker ignores a design rule for a specific figure.

NioCAD provides integrated extraction of component values and identification and extraction of parasitic components. FastHenry [23], [24] and InductEx [25], [26] are used to extract inductance values. InductEx makes it possible to extract a network of inductor values simultaneously. Once an extracted value is available for a component it is compared to the required schematic value. If the expected and extracted

values differ by more than a specified amount, the user is informed through an icon on the component list as well as through an item in the issue view.

A modified schematic circuit, that includes the parasitic components and uses the extracted component values, are generated for the physical layout. This allows the user to compare the original and extracted circuits using all the available analysis tools.

The final layout can be exported to the Graphic Data System II (GDSII) format [27], which is currently used by the two supported fabrication processes. During exportation it is ensured that all shapes are converted to polygons that are not enclosed by other shapes, as required by the GDSII formats.

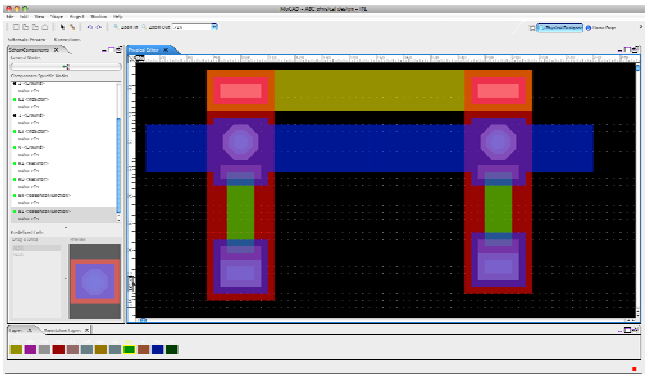


Fig. 8. Partial layout of a JTL in the physical designer [16].

IV. Future Work

The development team must implement a logic designer and simulator with support for a Hardware Description Language (HDL) [28], provide a test-bench framework to simplify circuit testing, provide real-time estimation of physical component values while they are created, perform real-time design rule checking, allow importing and exporting of designs and provide printing of designs.

In the long term *NioCAD* will add support for digital timing analysis and optimisation, mixed signal and radio frequency (RF) design, automated cell characterisation and automated layout.

V. Summary

We presented *NioCAD*; a next-generation EDA platform and SCE application suite, developed using modern software engineering paradigms.

By developing the platform from scratch it was possible to create a dynamic, modular architecture along with a rich set of generic frameworks. This allows developers to rapidly extend the platform and grow the application suite.

The application's ease of use, familiar and intuitive GUI and tight integration between modules allow users to rapidly design analogue and digital circuits of varying complexity.

References

[1] K. Gaj *et al.*, "Tools for the Computer-Aided Design of Multigigahertz Superconducting Digital Circuits," *IEEE Transactions on Applied Superconductivity*, vol. 9, pp. 18-38, Mar. 1999.

[2] K. K. Likharev and V. K. Semenov, "RSFQ Logic/Memory Family: A new Josephson-Junction Technology for Sub-Terahertz-Clock-Frequency Digital Systems," *IEEE Transactions on Applied Superconductivity*, vol. 1, pp. 3-28, Mar. 1991.

[3] T. Van Duzer and C. W. Turner, *Principles of Superconducting Circuits.* New York: Elsevier, 1981.

[4] K. Gaj *et al.*, "Towards a Systematic Design Methodology for Large Multigigahertz Rapid Single Flux Quantum Circuits," *IEEE Transactions on Applied Superconductivity*, vol. 9, pp. 45914606, Sep. 1999.

[5] H. R. Gerber, C. J. Fourie and W. J. Perold, "Investigating the requirements on CAD tools for the development of VLSI RSFQ circuits," in *ASC 2004 Technical Program, 3EH03*, 2004.

[6] MVC design pattern. Available: http://en.wikipedia.org/wiki/Model-view-controller

[7] Java, Sun Microsystems. Available: http://java.com/en/

[8] http://www.w3.org/XML/

[9] Tigris GEF. Available: http://gef.tigris.org/

[10] http://en.wikipedia.org/wiki/Z-order

[11] Berkeley SPICE3F5. Available: http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE/

[12] E. S. Fang and T. van Duzer, "A Josephson Integrated Circuit Simulator (JSim) for Superconductive Electronic Application," *Ext. Abstract of Int. Superconducting Electronics Conf.*, pp. 407-410, 1989.

[13] JSim. Available: http://www-cryo.eecs.berkeley.edu/CADtools.html

[14] R. E. Jewett, "Josephson Junction in SPICE2G5," Electronics Research Laboratory Internal Report, University of California, Berkley, December 1982.

[15] WRSpice, Whiteley Research Inc. Available: http://www.wrcad.com/wrspice.html

[16] High resolution screen captures of *NioCAD* is available at http://www.niocad.co.za/products_info.php

[17] C. A. Hamilton and K. C. Gilbert, "Margin and Yield in Single Flux Quantum Logic," *IEEE Transactions on Applied Superconductivity*, vol. 1, pp. 157-163, Dec. 1991.

[18] C. J. Fourie, W. J. Perold and H. R. Gerber, "Complete Monte Carlo model description of lumped-element RSFQ logic circuits," *IEEE Transactions on Applied Superconductivity*, vol. 15, pp. 384-387, Jun. 2005.

[19] Q. P. Herr and M. J. Feldman, "Multiparameter Optimisation of RSFQ Circuits Using the Method of Inscribed Hypersheres," *IEEE Transactions on Applied Superconductivity*, vol. 5, pp. 3337-3340, Jun. 1995.

[20] T. Harnisch *et al.*, "Design Centering Methods for Yield Optimization of Cryoelectronic Circuits," *IEEE Transactions on Applied Superconductivity*, vol. 7, pp. 3434-3437, Jun. 1997.

[21] Hypres niobium integrated circuit fabrication process #03-10-45 Revision #24, Jan 11, 2008. Available: http://hypres.com/pages/foundry/bfoundnio.htm

[22] Rapid Single Flux Quantum – Design Rules for Nb/Al_2O_3-Al/Nb Process Version 22.06.2007 at IPHT Jena. Available: http://www.ipht-jena.de/?id=158

[23] M. Kamon, M. J. Tsuk and J. K. White, "FastHenry: A Multi-pole-Accelerated 3-D Inductance Extraction Program," *IEEE Trans. Microwave Theory and Techniques*, vol. 42, pp. 1750-1758, Sep. 1994.

[24] B. Guan *et al.*, "Inductance Estimation for Complicated Superconducting Thin Film Structures with a Finite Segment Method," *IEEE Transactions on Applied Superconductivity*, vol. 7, pp. 2776-2779, Jun. 1997.

[25] C. J. Fourie and W. J. Perold, "On using finite segment methods and images to establish the effect of gate structures on inter-junction inductances in RSFQ circuits," *IEEE Transactions on Applied Superconductivity*, vol. 13, pp. 539-542, Jun. 2003

[26] C. J. Fourie and W. J. Perold, "Simulated inductance variations in RSFQ circuit structures," *IEEE Transactions on Applied Superconductivity*, vol. 15, pp. 300-303, Jun. 2005.

[27] S. M. Rubin, "Computer Aids for VLSI Design," 1994, Appendix C, Available: http://www.rulabinsky.com/cavd/text/chapc.html

[28] K. Gaj *et al.*, "Functional Modeling of RSFQ Circuits Using Verilog HDL," *IEEE Transactions on Applied Superconductivity*, vol. 7, pp 3151-3154, Jun. 1997.

Design and Manufacture of Nanometer-scale SOI Light Sources

Alfons W. Bogalecki and Monuko du Plessis, *Senior Member, IEEE*

***Abstract*—In an effort to investigate quantum confinement effects on silicon (Si) electroluminescence (EL) properties like quantum efficiency and spectral emission, Si wire junctions with thicknesses less than 50 nm were designed and manufactured in a fully custom designed silicon-on-insulator (SOI) technology.**

Since commonly available photolithography is unusable to consistently define nanometre-scale line-widths accurately and electron-beam lithography (EBL) by itself is too time-consuming and expensive to expose complete wafers, the wafer manufacturing process employed a selective combination of photolithography and EBL.

The device design and mask layout had to incorporate various effects caused by processing, material limitations and physical phenomena like impurity redistribution, impurity diffusion and self-limiting oxidation occurring during the physical manufacturing process.

The wafer manufacturing process was performed in clean-rooms of both the Carl and Emily Fuchs Institute for Microelectronics (CEFIM) at the University of Pretoria (UP) and the Georgia Institute of Technology's Microelectronic Research Centre (MiRC), which made an JEOL JBX-9300FS electron-beam pattern generator (EPG) available.

Since no standard process recipe was employable, the complete manufacturing process design is based on self-obtained equipment characterization data and material properties.

***Index Terms*—Nanometre-scale SOI, quantum confinement, silicon electroluminescence, silicon light source**

I. Introduction

The Carl and Emily Fuchs Institute for Microelectronics (CEFIM) at the University of Pretoria (UP) has been developing silicon (Si) light sources since 1992, [1] - [5].

Si electroluminescence (EL) improvement research within the INSiVA project added further promising light source configurations.

Without explaining their functional principles here, the three Si light source candidates selected for silicon-on-insulator (SOI) miniaturization in this research work are:

- Avalanche (n^+p or p^+n junctions),
- Punch-through (n^+pn^+ or p^+np^+ junctions) and
- Carrier-injection light sources.

The Carl and Emily Fuchs Institute for Microelectronics (CEFIM) in the Department of Electrical, Electronic and Computer Engineering at the University of Pretoria (UP) in South Africa researches this technology with financial support from INSiAVA (Pty) Ltd.

A. W. Bogalecki is a senior researcher at the CEFIM, University of Pretoria, 0002, South Africa (phone: +27124202983; fax: +27123625115; e-mail: alfons.bogalecki@up.ac.za).

M. du Plessis, Prof., is the director of the CEFIM, University of Pretoria, 0002, South Africa (e-mail: mplessis@postino.up.ac.za).

A. Quantum Confinement

Enhancement factors of up to 30 in EL due to quantum mechanical confinement in ultra-thin single crystal Si compared to bulk devices were reported in [6] - [9].

References [8] and [9] reported a strong efficiency improvement in forward-biased SOI light-emitting diodes (LEDs) on a buried oxide (BOX) when the thickness of the regions shown in Fig. 1 was reduced.

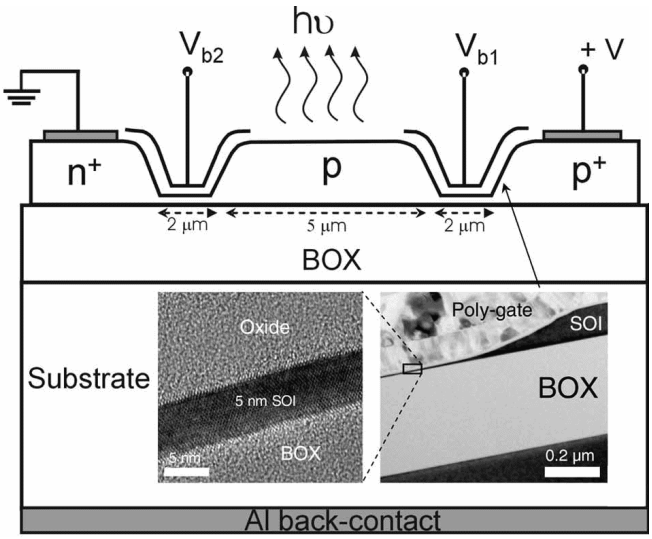


Fig. 1. Cross-section of the SOI LED manufactured in [8] and [9].

The dramatic increase in integrated EL with reducing device layer thickness (Fig. 2) is attributed to the suppression of non-radiative recombination.

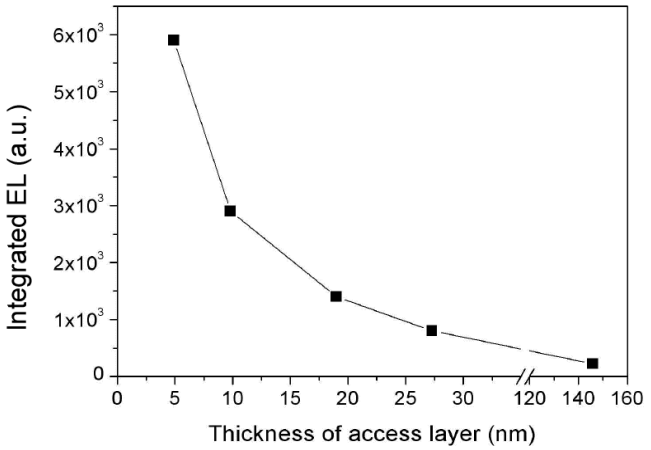


Fig. 2. Integrated EL intensity against access layer thickness [9].

Abovementioned SOI light sources are only thin in one dimension, their planar thickness, with device widths ranging between 20 and 60 µm. Their *pn*-junctions are also located outside the thinned area.

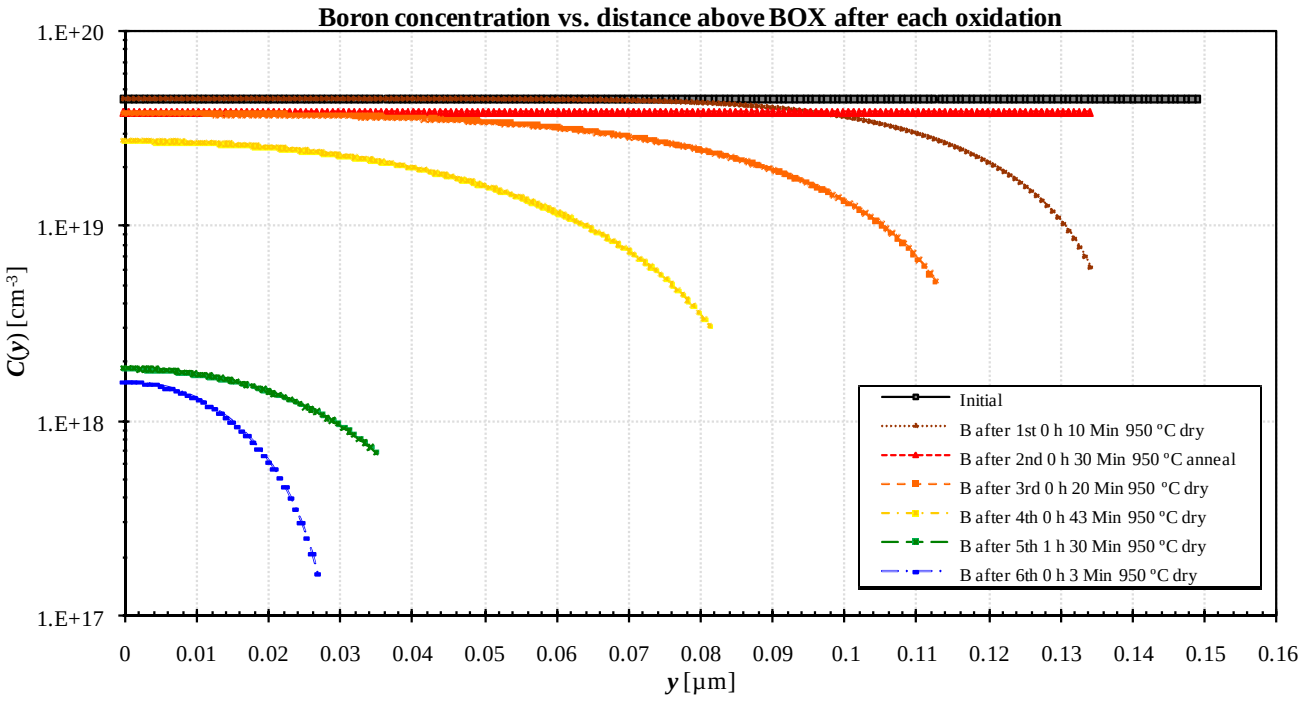


Fig. 4. Boron redistribution during the thermal oxidation and anneal steps.

B. *Objectives and Approach*

The main purpose of the INSiAVA6 test-chip is to investigate the effect of quantum confinement in avalanche, punch-through and carrier-injection Si light sources on electroluminescence characteristics like quantum efficiency and spectral emission.

Instead of just creating planar thin devices, the technical objective is to design and manufacture SOI light sources that are smaller in two dimensions.

With reference to the definitions in Fig. 3, the following finger junction dimensions are aimed at:

$$5\text{ nm} \le \boldsymbol{t} \le 100\text{ nm}, \quad (1)$$

$$10\text{ nm} \le \boldsymbol{w} \le 100\text{ nm and} \quad (2)$$

$$200\text{ nm} \le \boldsymbol{l} \le 400\text{ nm}. \quad (3)$$

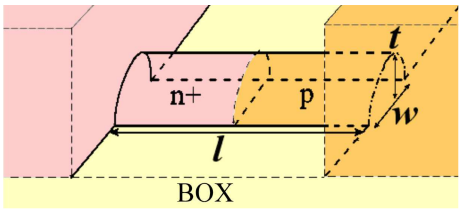


Fig. 3. SOI finger junction dimension definitions.

Manufacturing larger rectangular Si structures and selectively oxidizing these created the desired thinner Si fingers. Nanometre-scale Si wires have already been manufactured through oxidation [10], but *pn*-junctions have, to our knowledge, never been implemented inside such thin Si wires.

II. Physical and Optical Simulation

Simulations predict the physical and optical properties achievable at the end of the manufacturing process.

A. *Carrier Redistribution during Oxidation*

Arsenic's (As) low diffusivity and solubility in silicon dioxide (SiO_2) results in the "snow-shovel" effect that causes As to pile up against the moving SiO_2 boundary during thermal oxidation, [11] - [13].

Boron (B) with its higher diffusivity and solubility in SiO_2 is absorbed into the SiO_2 during oxidation, therefore decreasing its concentration in the Si, [14] - [18].

Since the lower B background doping concentration of n^+p junctions plays a larger role in determining junction characteristics like depletion region width $\boldsymbol{w_d}$ than the higher As concentration, only the spatial B concentration variation with thermal oxidation was further considered.

The SOI B concentration $\boldsymbol{C(y,t)}$ as a function of distance $\boldsymbol{y}$ from the BOX and oxidation time $\boldsymbol{t}$ during thermal oxidation can be expressed as [14]

$$\frac{C(y,t)}{N_B} = 1 - \frac{\frac{k-m}{2}\sqrt{\frac{B\pi}{D}}\left\{\operatorname{erfc}\left(\frac{l-m\sqrt{Bt}-y}{2\sqrt{Dt}}\right)+\operatorname{erfc}\left(\frac{l-m\sqrt{Bt}+y}{2\sqrt{Dt}}\right)\right\}}{1-e^{-\left(\frac{l-m\sqrt{Bt}}{\sqrt{Dt}}\right)^2}+\frac{k-m}{2}\sqrt{\frac{B\pi}{D}}\left\{1+\operatorname{erfc}\left(\frac{l-m\sqrt{Bt}}{2\sqrt{Dt}}\right)\right\}}, \quad (4)$$

where $\boldsymbol{m} \approx 0.45$ is the ratio between Si thickness removed and SiO_2 thickness grown during oxidation, $\boldsymbol{B}$ the SiO_2 growth-rate constant, $\boldsymbol{D}$ the B diffusion constant, $\boldsymbol{l}$ the initial SOI active layer thickness, $\boldsymbol{k} = \boldsymbol{C_{SiO2}}/\boldsymbol{C_{Si}}$ is the segregation coefficient defined by $\boldsymbol{C_{SiO2}}$, the impurity concentration on the SiO_2 side of the Si/SiO_2 boundary, and $\boldsymbol{C_{Si}}$, the impurity concentration on the Si side of the same interface.

Fig. 4 shows the simulated B concentration decrease from the initial implanted concentration during successive oxidations.

Prior knowledge of oxidation steps therefore allowed specifying the initial B implantation dose so that the desired final average finger background B concentration of $\approx 10^{18}$ cm^{-3} was achievable.

In reality, some B segregation also occurs at the BOX interface [19], but was ignored in the simulation shown in Fig. 4. The B re-diffusion from the BOX into the Si will be modelled in the next design revision.

B. *Optical Radiation Simulation*

Employing ray-tracing software enabled the simulation of spatial light radiation characteristics of the Si fingers with varying finger geometries and light source locations.

Fig. 5 for example shows how light generated at the centre of a finger with rounded corners focuses into dominant lobes emanating from the corners.

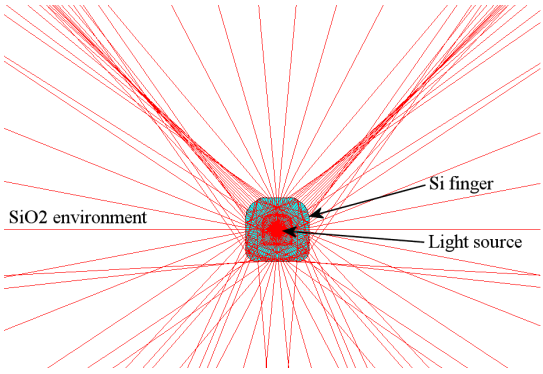


Fig. 5. SOI finger spatial light radiation pattern simulation.

It was *inter alia* determined that the maximum useful light emission directed away from the chip surface is achieved with a hemispherical round top finger surface to minimize internal reflection beyond the critical interface incidence angle and a flat bottom surface to maximize reflection back to the top surface.

III. DESIGN

To achieve the previously stated objective of creating Si wire junctions with diameters less than 50 nm, the SOI test-devices could not be manufactured in a standard process with supplied design rules. Instead, the layout and process design had to consider various equipment limitations and characteristics as well as physical effects occurring during the self-processing.

A. *Lithographic Patterning*

As indicated in Table I, the small geometries and precise pattern alignment of the current work required electron-beam lithography (EBL), but the single beam exposure scan of EBL is too slow to write all features across complete wafers.

TABLE I
PHOTOLITHOGRAPHY AND EBL COMPARISON

Aspect	Photolithography	EBL
Wavelength/Spot-size	≈ 300 nm	≈ 2 nm
Minimum feature size	≈ 0.5 µm	≈ 6 nm
Alignment accuracy	> 1 µm	> 6 nm
Exposure speed per wafer	Whole wafer at once Fast: Minutes	Serial scanning beam Very slow: hours

For above reason, as Table II shows, different processing steps and geometry areas of the wafer selectively employed photolithography or EBL.

The "rough" photolithographic *Si Island* mask defines where Si islands remain on the BOX after reactive ion etching (RIE) removes the superfluous Si.

The thin finger structures are created by employing the "fine" *Finger Spacing* EBL mask to RIE groove spacings into the Si islands. The *Oxidation* mask is then used to selectively oxidize the Si between the spaces into thin SOI fingers.

TABLE II
PHOTOLITHOGRAPHIC AND EBL MASKS

Processing Step	GDS No.	Layout Rendering	Mask Name	Mask Type	Mask Polarity
3d	1		*Si Island*	Photo	Positive
4d	2		*Arsenic (EBL)*	EBL	Positive Tone
4j	3		*Arsenic (Photo)*	Photo	Negative
5b	4		*Finger Spacing*	EBL	Positive Tone
6b	5		*Oxidation (EBL)*	EBL	Positive Tone
6c	6		*Oxidation (Photo)*	Photo	Negative
7a	7		*Contact*	Photo	Negative
7e	8		*Metal*	Photo	Positive

To achieve the required alignment accuracy when positioning the *Arsenic*, *Finger Spacing* and *Oxidation* masks relative to each other, EBL alignment marker crosses were initially etched into the SOI active layer and refined with the first EBL and etch step.

B. *CAD Layout*

Fig. 6 shows the layouts of the avalanche and punch-through SOI finger junction light sources.

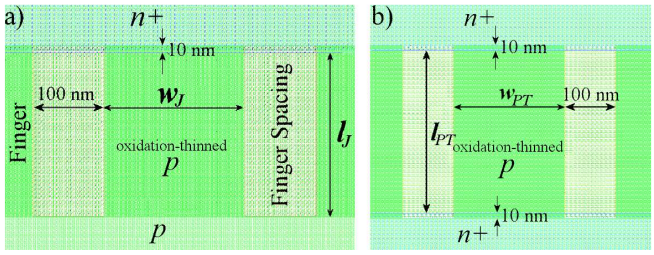


Fig. 6. a) Avalanche and b) punch-through finger layout dimensions.

The Si island regions above and below the fingers remain thick while an EBL-written oxidation mask opening across the fingers allows the selective thinning through oxidation of the SOI finger junctions.

Fig. 7 shows how, for example, 100 punch-through fingers shown in Fig. 6 b) are placed in parallel between thicker and larger Si islands that allow electrical biasing through the interconnect metallization.

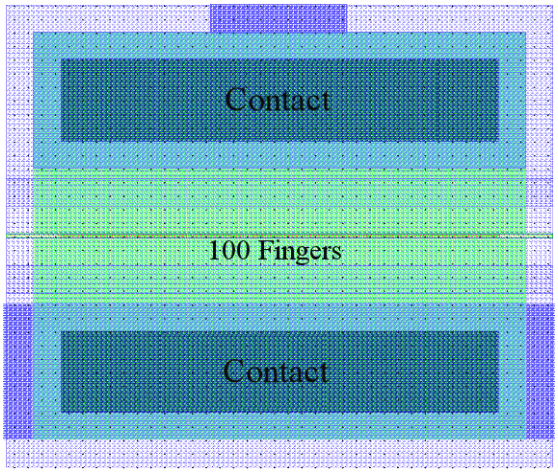


Fig. 7. Finger device layout incorporating 100 parallel punch-through fingers shown in Fig. 6 b).

As shown in Fig. 8 two different carrier-injection SOI light sources were designed and implemented.

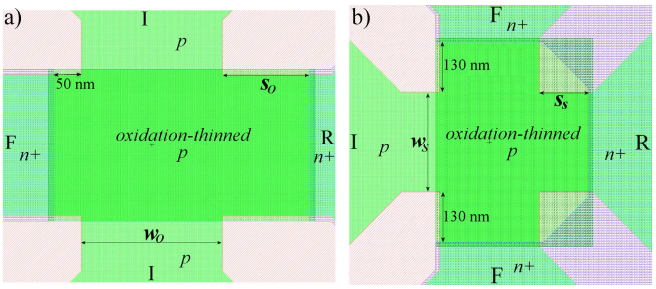


Fig. 8. a) Opposite- and b) side-injection SOI device layouts.

In the opposite-injection device the forward-biased n^+p junction (located at the interface between the *F* and *I* regions) is placed opposite the reverse-biased n^+p junction (located at the boundary between the *R* and *I* regions). In the side-injection device, two injecting junctions are adjacent to the reverse-biased junction.

The injection devices in Fig. 8 are implemented in cells as shown in Fig. 9, which allowed laying them out in interconnected arrays of 19 devices in parallel.

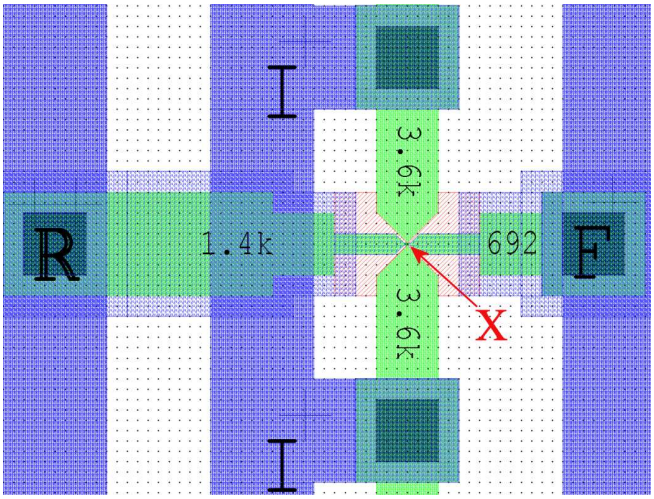


Fig. 9. SOI injector element layout on a 1-µm grid. X demarcates the location of one of the injector types shown in Fig. 8.

To facilitate the successful manufacture of the SOI light sources, self-made design rules had to be set-up. These design rules had considered possible mask alignment errors, limitations of the processing equipment and physical phenomena occurring during the wafer manufacturing process.

The dimensioning of the devices shown in Fig. 6 and Fig. 8 for example had to obey the following specifications to ensure proper functionality of the devices:

$$l_J \geq w_d + d_{Ox} + 3\Delta R_{\perp As}, \quad (5)$$

$$w_J = w + 2\Delta Si, \quad (6)$$

$$l_{PT} \approx w_d + 2\left(d_{Ox} + 3\Delta R_{\perp As}\right), \quad (7)$$

$$w_{PT} = w + 2\Delta Si, \quad (8)$$

$$s_S \approx d_{Ox} + 3\Delta R_{\perp As}, \quad (9)$$

$$w_S = w + 2\Delta Si, \quad (10)$$

$$s_O \approx d_{Ox} + 3\Delta R_{\perp As} \text{ and} \quad (11)$$

$$w_O > w_d + \Delta Si; \quad (12)$$

where w_d is the reverse bias depletion region width shortly before breakdown, d_{Ox} the worst-case As diffusion distance during all thermal oxidation steps after ion implantation, w the desired final finger width, ΔSi the amount of Si removed during the thinning oxidations and $\Delta R_{\perp As}$ the transverse As implantation straggle.

Table III shows the laid-out and final measured SOI finger dimensions.

TABLE III
ACHIEVED SOI FINGER DEVICE DIMENSIONS

Dimension	Layout [nm]	Final [nm]
l_J	230	296 - 359
w_J	200, 220, 240, 260, 280, 300	38 - 101
l_{PT}	320, 330, 340, 350, 360, 370, 380, 390, 400	348 - 502
w_{PT}	220, 260, 300	30 - 210
s_S	50, 90, 130	Not measured yet
w_S	200, 260, 300	Not measured yet
s_O	120, 160, 200	Not measured yet
w_O	200, 260, 300	Not measured yet

Due to the SOI active layer thickness variation of ± 100 nm a large range of final finger dimension were achieved.

Fig. 10 depicts the four-chip cluster CAD layout that was used to manufacture the photolithographic masks and generate the EBL data files required by the JEOL JBX-9300FS electron-beam pattern generator EPG.

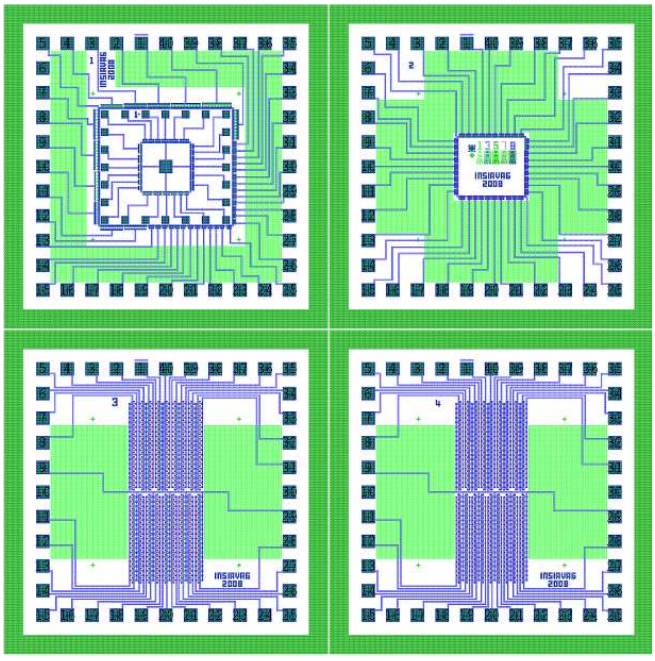


Fig. 10. CAD layout showing the four chips in a 5.1 mm x 5.1 mm cluster.

Chip 1 in the top-left quadrant of the cluster contains 18 avalanche and 19 punch-through 100-finger SOI light source arrays, *Chip 2* (top right) contains 39 100-finger punch-through device arrays while *Chip 3* and *Chip 4* contain 36 19-element arrays of the two different injection light source types.

IV. Physical Wafer Manufacturing Process

Fig. 11 shows the supplier specifications of the SOI starting material wafers.

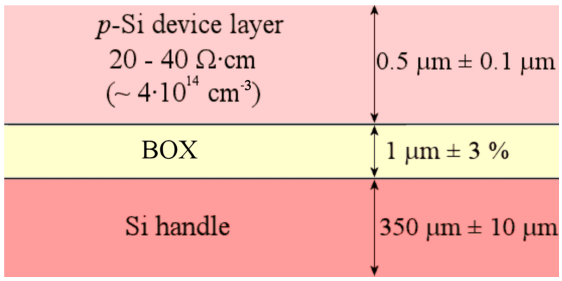


Fig. 11. Initial SOI wafer specifications.

The 500 nm thick SOI active device layer above the BOX is the region where the light sources were implemented after its thickness was reduced to about 150 nm through thermal oxidation.

The 100 nm tolerance in SOI active device layer thickness might seem excessive, but is useful since a large variation in SOI light source thicknesses can be achieved with relative ease.

Table IV gives a brief summary of the processing steps and equipment involved in the manufacture of the SOI light sources.

TABLE IV
Process Flow Summary

Step	Action	Equipment employed	Facilities
1	Si thinning	Furnace, Reflective spectrometer	CEFIM, MiRC
2	Blanket B implant	Ion implanter	Core Systems
3	Si island definition	PECVD, Mask aligner, RIE	MiRC
4	As implant	PECVD, EPG, Mask aligner, RIE, Ion implanter	MiRC, Core Systems
5	Finger definition	EPG, RIE	MiRC
6	Finger oxidation	PECVD, EPG, Furnace	MiRC, CEFIM
7	Metallization	PECVD, Mask aligner, Sputterer	CEFIM, Elume

Since no standard process recipe was employable, the complete manufacturing process had to be designed on self-obtained equipment characterization data, material and chemical properties.

The importance of characterizing the processing equipment to determine limitations and adapt the process accordingly can be illustrated by comparing the etch profiles of two different RIE systems. Fig. 12 a) shows that while the Vision Oxide RIE has an excellent etch selectivity towards Si when the ZEP EBL resist is used as a mask, its excessive horizontal etch spread makes it unsuitable for etching the 100 nm wide finger spacing slits. Fig. 12 b) shows that the PlasmaTherm ICP, which was eventually used for the finger spacing etching, exhibits a very narrow etch profile, but suffers from a relatively poor etch selectivity of only 0.5.

To ensure consistent results and confirm acceptable manufacturing performance, process control and monitoring was employed. This involved the simultaneous processing of monitor wafer pieces and the measurement of on-chip test-structures that could be analyzed with a scanning electron microscope (SEM), a reflective spectrometer (to measure thin film thicknesses) and a profilometer.

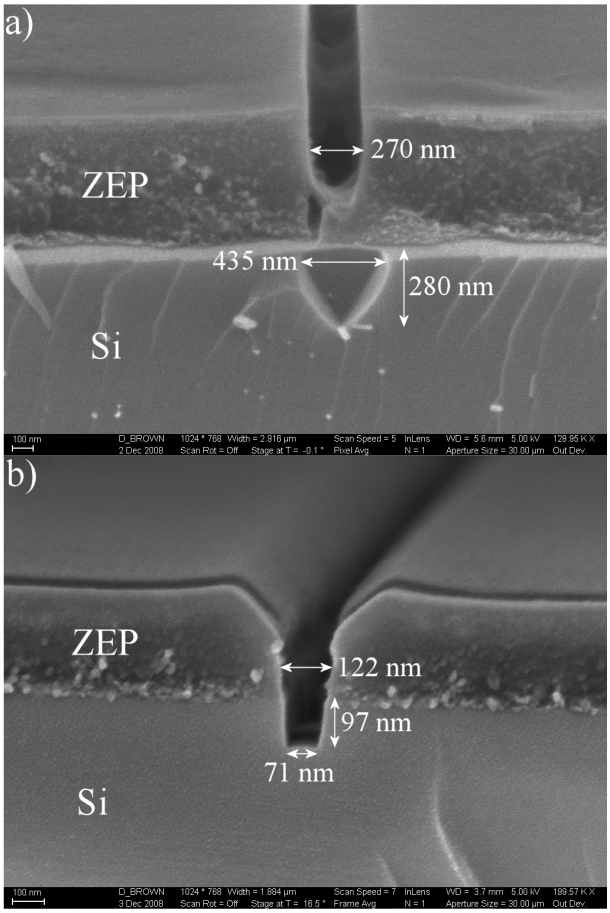


Fig. 12. a) Vision Oxide RIE and b) PlasmaTherm ICP 100 nm groove Si etch test.

The processing equipment was controlled by adapting machine-specific "recipe" variables like: processing time, pressure, DC & AC (plasma) power, gas flows, etc.

The SEM image in Fig. 13 shows the finger spacing holes in the Si that resulted once the ZEP EBL resist was de-scummed after RIE (before thinning oxidation).

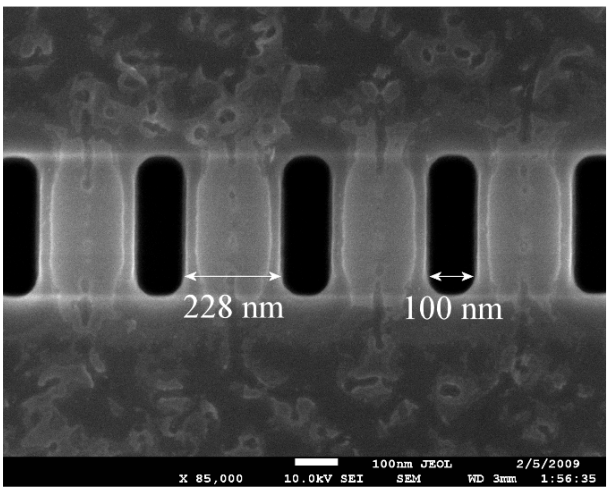


Fig. 13. SEM image showing some finger definition spacing holes in the Si after EBL and RIE when the Si is still about 117 nm thick.

Although the EBL exposure and PlasmaTherm ICP etch resulted in rounded corners of the holes, the dimensions of the holes are acceptably close to the layout.

V. CURRENT STATUS

Fig. 14 depicts examples of the thin SOI finger junctions achieved after the thinning oxidation.

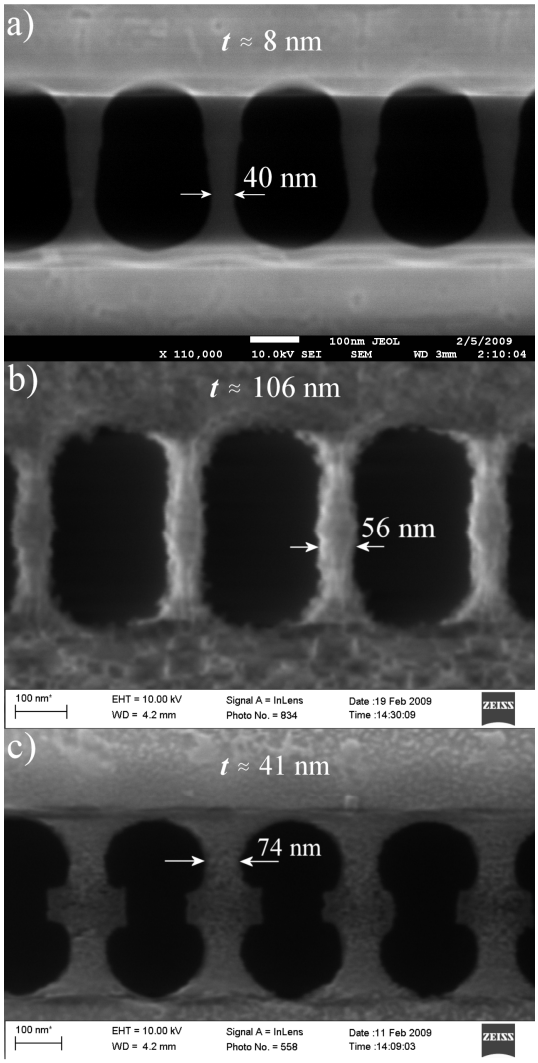


Fig. 14. a) and b) avalanche and c) punch-through SOI finger final dimension examples after thinning oxidation with interpolated finger thickness estimates.

In contrast to the broad, but thin fingers in Fig. 14 a), the fingers in Fig. 14 b) are more thick than narrow.

The punch-through SOI light source in Fig. 14 c) clearly shows how the heavily As doped Si oxidized faster than the lightly doped background.

Above oxidation effect should result in higher useful light emission from the chip surface since less light is lost due to internal reflection along the finger axis.

VI. FUTURE WORK

After completing the metallization, wafer inspection and electrical probing allows device selection for chip dicing and packaging. Optical characterization with a photo-multiplier tube, radiometer and spectrometer then allows the calculation of quantum and external power efficiency.

ACKNOWLEDGMENT

The authors would like to thank the Georgia Institute of Technology's Microelectronic Research Centre (MiRC) for help and usage of their clean-room and Devin Brown for operating their JEOL JBX-9300FS EPG.

The authors would also like to thank Dr. A. Botha at the UP Microscopy and Microanalysis laboratory for making the Zeiss Ultra Plus SEM available.

REFERENCES

[1] 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 Journal of Selected Topics in Quantum Electronics, Vol. 8, Issue 6, pp. 1412-1419, November 2002.

[2] L. W. Snyman, H. Aharoni, M. du Plessis and R. B. J. Gouws, "Increased efficiency of silicon light emitting diodes in a standard 1.2 micron complimentary metal oxide semiconductor technology", Optical Engineering, Vol. 37, No. 7, pp. 2133-2141, July 1998.

[3] L. W. Snyman, H. Aharoni, M. du Plessis, J. F. K. Marais, D. van Niekerk and A. Biber, "Planar light-emitting electro-optical interfaces in standard silicon complimentary metal oxide semiconductor integrated circuitry", Optical Engineering, Vol. 41, No. 12, pp. 3230-3240, December 2002.

[4] M. du Plessis, H. Aharoni and L. W. Snyman, "Two- and multi-terminal CMOS/BiCMOS Si LEDs", Optical Materials, June 2004.

[5] L. W. Snyman, J. M. Matjila, H. Aharoni and M. du Plessis, "Increasing the efficiency of a three-terminal silicon CMOS LED through current density and carrier injection techniques", Proceedings of the 12th IEEE International Symposium on Electron Devices for Microwave and Optoelectronic Applications, EDMO 2004, Kruger National Park, South Africa, pp. 71-82, 8-9 November 2004.

[6] S. Saito, D. Hisamoto, H. Shimizu, H. Hamamura, R. Tsuchiya, Y. Matsui, T. Mine, T. Arai, N. Sugii, K. Torii, S. Kimura and T. Onai, "Electro-luminescence from Ultra-thin silicon", Japanese Journal of Applied Physics, Vol. 45, No. 27, 2006, pp. L679-L682.

[7] A. Karsenty, A. Sa'ar, N. Ben-Yosef and J. Shappir, "Enhanced electroluminescence in silicon-on-insulator metal-oxide-semiconductor transistor with thin silicon, Applied Physics Letters, Vol. 82, No. 26, June 2003, pg. 4830.

[8] T. Hoang, P. LeMinh, J. Holleman and J. Schmitz, "Strong efficiency improvement of SOI-LEDs through carrier confinement", IEEE Electron Device Letters, Vol. 28, No. 5, May 2007, pg. 383.

[9] H. Tu, "High efficient infrared light emission from silicon LEDs", Doctor degree dissertation at the University of Twente, ISBN 978-90-365-2557-2, Wöhrmann print service, Zutphen, The Netherlands, 2007, pg. 105.

[10] X. Tang, X. Baie, J. P. Colinge, A. Crahay, B. Katschmarsyj, V. Scheuren, D. Spôte, N. Reckinger, F. van de Wiele and V. Bayot, "Self-aligned silicon-on-insulator nano flas memory device", Solid-state Electronics 44, 2000, pp. 2259-2264.

[11] G. A. Tarnavskii, S. I. Shpak and M. S. Obrecht, "Features of segregation of doping impurities of V(a) subgroup elements on angular configurations of the silicon/silicon dioxide oxidation boundary", Journal of Engineering Physics and Thermophysics, Vol. 75, No. 1, 2002, pp. 190.

[12] S. S. Choi, "Temperature-dependant redistribution of arsenic impurities at the Si/SiO_2 interface", Journal of the Korean Physical Society, Vol. 28, No. 6, December 1995, pp. 763 – 767.

[13] S. S. Choi and M. J. Park, "Redistribution of arsenic during thermal oxidation of arsenic ion-implanted silicon", Journal of the Korean Physical Society, Vol. 24, No. 6, December 1991, pp. 499 - 505.

[14] Z. Đurić, M. M. Smiljanić, K. Radulović and Z. Lazić, "Boron redistribution during SOI wafer thermal oxidation", Proc. 25th International Conference on Microelectronics (MIEL 2006), Belgrade, Serbia and Montenegro, 14 – 17 May 2006.

[15] K. Suzuki and T. Miyashita, "Models of Boron redistribution during thermal oxidation with general oxidation rate", IEEE Transactions on Electron Devices, Vol. 47, No. 3, March 2000, pp. 523.

[16] P. H. Ladbrooke, "Boron segregation data for D.M.O.S. devices", Electronics Letter, Vol. 14, No. 5, 2 March 1978, pp. 128.

[17] C. P. Wu, E. C. Douglas and C. W. Mueller, "Redistribution of ion-implanted impurities in Silicon during diffusion in oxidizing ambients", IEEE Transactions on Electron Devices, Correspondence, September 1976, pp. 1095.

[18] S. Margalit, A. Neugroschel and A. Bar-Lev, "Redistribution of Boron and Phosphorous in Silicon after two oxidation steps used in MOST fabrication", IEEE Transactions on Electron Devices, Vol. ED-19, No. 7, July 1972, pp. 861.

[19] A. F. Saavedra, K. S. Jones, M. E. Law, K. K. Chan and E. C. Jones, "Electrical activation in silicon-on-insulator after low energy implantation", Journal of Applied Physics, Vol. 96, No. 4, 15 August 2004, pp. 1891.

Photonic transitions (1.4 eV- 2.8 eV) in Silicon p^+np^+ injection-avalanche CMOS LEDs as function of depletion layer profiling and punch through techniques

Lukas W. Snyman (MIEEE), Monuko du Plessis (SMIEEE) and Enrico Bellotti

***Abstract*— p^+np^+ CMOS Si LED structures were modeled in order to investigate the effect of various depletion layer profiles and defect engineering on the photonic transitions in the 1.4 to 2.8 eV, 450-750nm regime. Particularly, modeling shows that by utilizing a short linear increasing E-field in the p^+n reverse biased junction with a gradient of approximately 5 x 10^5 $V.cm^{-1}$. μm^{-1}, and injecting carriers from an adjacent p^+n junction, has the potential to increase optical yield. A number of device designs were subsequently realized using CMOS 0.35µm technology. Areas in the devices show marked increases in emission efficiency of factors of up to 50 – 100 as compared to previous realizations utilizing no such techniques. The current devices operated in the 6-8V, 1uA - 2mA regime and yield emission intensities of up to 100 $nW.\mu m^{-2}$. The current emission levels are about three orders higher than the low frequency detectability limit of Si CMOS p-i-n detectors of corresponding area.**

***Index Terms*— Electroluminescence, light emitting diodes, silicon, CMOS integrated circuitry, physical modeling.**

Manuscript received March 10, 2009. This work was supported in part by the (TUT), the Carl Emily Fuchs Institute for Microelectronics(CEFIM), The S.A. National Research Foundation Focus Area Programme and the National Innovation Fund from the S.A . National Research Foundation.

L.W. Snyman is at the Laboratory for Innovative Electronic Systems, Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa (Email: snymanlw@tut.ac.za).

M. du Plessis is from Carl and Emily Fuchs Institute for Microelectronics, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa (Email: mplessis@postino.up.ac.za).

H. Belotti is from the Department of Electrical Engineering , Boston University , USA. (Email: bellotti@bu.ac.usa). He is currently a regular visiting professor at the Tshwane University of Technology.

Some of the content in this article forms the subject of South African Patents 2008/00593, 98/9305, 96/2528, 96/2478 and of USA Patents No. 6,111,271 of 29th August 2000 and No. 5,944,720 of 30 November 1999.

I. Introduction

VARIOUS researchers [1 - 5] have highlighted the need for small-dimension, efficient light emitters which are compatible with mainstream silicon CMOS integrated circuit technology. The realization of sufficiently efficient light-emitters are currently a major technological challenge and will considerably boost opto-electronic applications and integration at the integrated circuit level. Although longer wavelength light emitters have application in integrating data-telecommunication systems on the chip level, shorter wavelength Si light emitters can find wide-scale applications as on chip opto-couplers, short-range wave-guiding based systems, and on-chip mechanical-optical sensor (MOEMS) applications. . Recently, Wada et al [7] proposed the utilization of 450 nm based LED's for clock-pulsing in large diameter, next-generation silicon microprocessor circuits. The main advantages of the avalanche Si LED technology are its ease of integration into standard CMOS technology with no variation to design and processing procedures as well as its very high modulation speed possibilities (> 10 GHz) [6]. Furthermore, the current emission levels of CMOS Si avalanche LEDs are about three to four orders of magnitude higher than the detectivity of standard CMOS detectors of comparable dimension. However, this technology seemed not very viable because of the low external power and quantum conversion efficiencies which were measured for the early devices (10^{-8} to 10^{-7}) [11].

Light emission from silicon devices has been realized in reverse-biased p-n avalanche structures [8-14]. Various theories have been put forward in order to explain the phenomenon. These include phonon assisted intra-band relaxation phenomena [10 - 13], as well as phonon inter-band recombination processes [14]. In depth theoretical modelling and experimental evidence indicated that the highest optical yield results from intra-band phonon assisted relaxation processes mainly in the conduction band [9].

Snyman, Du Plessis and Aharoni have over some time realized a series of *practical* and *utilizable* light emitting devices (Si LED's) in *standard CMOS* technology using the avalanche light-emitting phenomenon (See Fig. 1) [15 - 26]. This was mainly achieved by using novel surface engineering, current density modeling and dynamic carrier density engineering techniques. The developed devices show about three orders of increase in optical output as compared with previous similar work [19, 20]. Particularly promising results have recently been obtained with regard to further increasing the efficiency, as well as the emitted intensity [22 - 26], with what we call the InSiAva (Injection-enhanced Silicon in Avalanche) technology. This technology utilizes the hypothesis that the light emission in these structures can be increased due to interaction of high energy (hot) electrons, as excited in the avalanching junction, with low energy (cool) energy holes as injected into the avalanching junction by a nearby forward-biased pn junction.

In this paper, we specifically report on further promising results which were obtained through modeling and experimental realizations of p^+np^+, two junction, diagonal avalanche control and injection-based devices. Specifically we report on the observed effects of depletion layer profiling and defect engineering on the optical emission phenomena. We further make important derivations regarding the internal optical yield as derived from the results.

II. STIMULATING HIGHER THAN BANDGAP OPTICAL TRANSITIONS IN SILCON

Fig. 2 illustrates some of the conceptual design aspects of our p^+np^+ injection-avalanche technology, referring to specific device zones, electric field distribution, possible photonic transitions in the silicon energy band gap. Fig. 2 (a) shows three clear activity regions of the device, viz a high field excitation zone generating energetic carriers, a relaxation and recombination zone for excited electrons as they traverse the depletion region of the first junction and a so called injection zone where low energy carriers is injected into depletion region of the first junction by forward biased second junction. The following provides more detail of the hypothesized operation of the device:

(1) Upon reverse voltage biasing the device, a high linear increasing electric field is created with its maximum at the one p+n interface (Fig 2 (a)). At some bias voltage, the E-field at the interface region attains high enough values such that charge multiplication (avalanche) processes occur in a narrow zone surrounding the interface. Energetic ("hot") electrons are transferred towards the n-side of the junction and energetic ("hot)" holes are transferred towards the p^+ side of the junction.

(2) Since the E-field decays linearly with distance away from the p^+n interface (Fig 2 (b)), the transferred

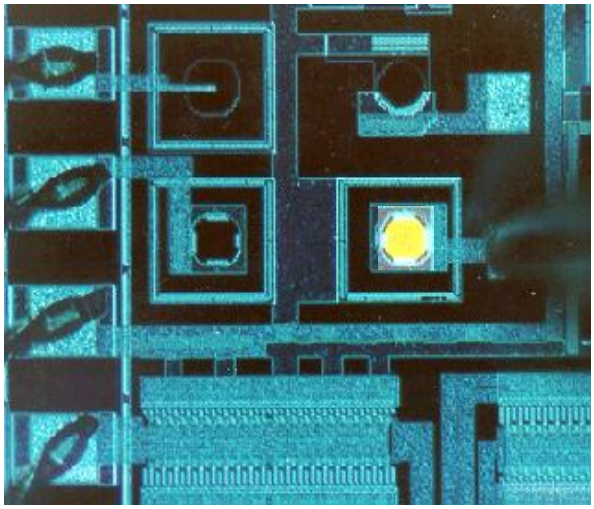


Fig. 1 60 μm diameter Si avalanche-based LED and optical interface as realized in 1.2 μm Si CMOS technology with no alterations to the design and processing procedures.

electrons soon reach regions where the E-field is not high enough in order to sustain ionization and carrier multiplication processes, and they are transferred away from the junction in the linearly decreasing E-field . This region is referred to the depletion zone or drift zone of the device.

(3) If a second np^+ junction is placed near the depletion or drift zone (Fig. 2 (a)), and this junction is forward biased, low energy holes may be injected into the drift zone. They will drift in a opposite direction and interact with the excited energetic electrons as well as with defect centers in the depletion region.

Early investigations into the origin of light emission processes in avalanching np silicon junction, suggested that the main light production processes are related to host silicon atom ionization processes in the high field avalanching conditions, followed by subsequent intra-band relaxation processes and phonon assisted indirect band-to-band transitions [9, 10, 14].

We have particularly realized that some high energy inter–band optical transitions may be stimulated in the silicon band structure by utilizing the recombination behavior between excited carriers and lower energy carriers, and by utilizing some novel device designs. Fig. 2 (c) hence demonstrates some of the photonic transition processes that may be stimulated in the above scenario

(1) During the avalanche process the carriers are accelerated in the high electric field region of the depletion layer but lose most of their energy through lattice interaction and phonon generation. As the electric field value in a particular zone increases, carriers eventually reach enough energy to ionize host Si atoms and break

valence bonds. The threshold for such processes has been reported to be of the order of 1.8 eV for electrons and 2.4 eV for holes, respectively [27]. It can, therefore, be expected that the highest possible energy levels for the excited electrons and holes will follow a distribution function around these values. Subsequent intra-band relaxation processes may occur, for example relaxation of energetic electrons to lower energy levels in the conduction band e.g transitions of Type A in Fig. 2 (c) ; and relaxation of energetic holes to lower energy levels in the valence band (Transitions of Type B in Fig. 2 (c)).

(2) Energetic electrons may also recombine directly with hot holes inside as well as outside the excitation zone and cause direct inter-band transitions of Type C, of approximately 2.8 to 3.2 eV, as demonstrated in Fig. 2 (c)).

(3) Low energy holes that are injected into the avalanching junction at the top of the valance band from a nearby junction (G in Fig 2 (c)), may interact with energetic (hot) electrons generated in the avalanching (excitation) region of the junction through phonon assisted recombination processes (F in Fig 2 (c)), resulting in transitions of Type D as illustrated in Fig 2 (c). The energy of these transitions are calculated as approximately 2.8 eV.

(4) If a large number of mid bandgap defects states are present in the depletion region, phonon assisted transitions of Type E may also be stimulated, stimulating transitions of respectively 2.3 eV and 0.5 eV . Slight variations of the above may occur, for example the transitions are greatly dependent on the position of the defect energy states within the energy gap

A further hypothesis is that the respective transition probabilities may be enhanced by means of engineering carrier densities to populate specific regions in the energy band diagram e.g. stimulating high energy populations of at F and G in the energy band diagram in Fig 2 (c)); increasing the dynamic carrier densities through the junction; and by introducing mid band gap defect state densities in the junction.

The respective optical emission rates can be related by the following quantitative relationships : For example, for the relaxation of excited electrons in the conduction band (intra-band transitions):

$$L_p = \alpha n' N_c \quad \ldots\ldots (1)$$

where L_p is the relaxation rate leading to photon emissions, n' is the density of excited electrons , N_c is the density of available states at lower energy levels, and α is a relaxation probability which is dependent on various secondary parameters. A similar process holds for excited holes in the valence band.

For inter-band direct transitions, the relationship may be defined as

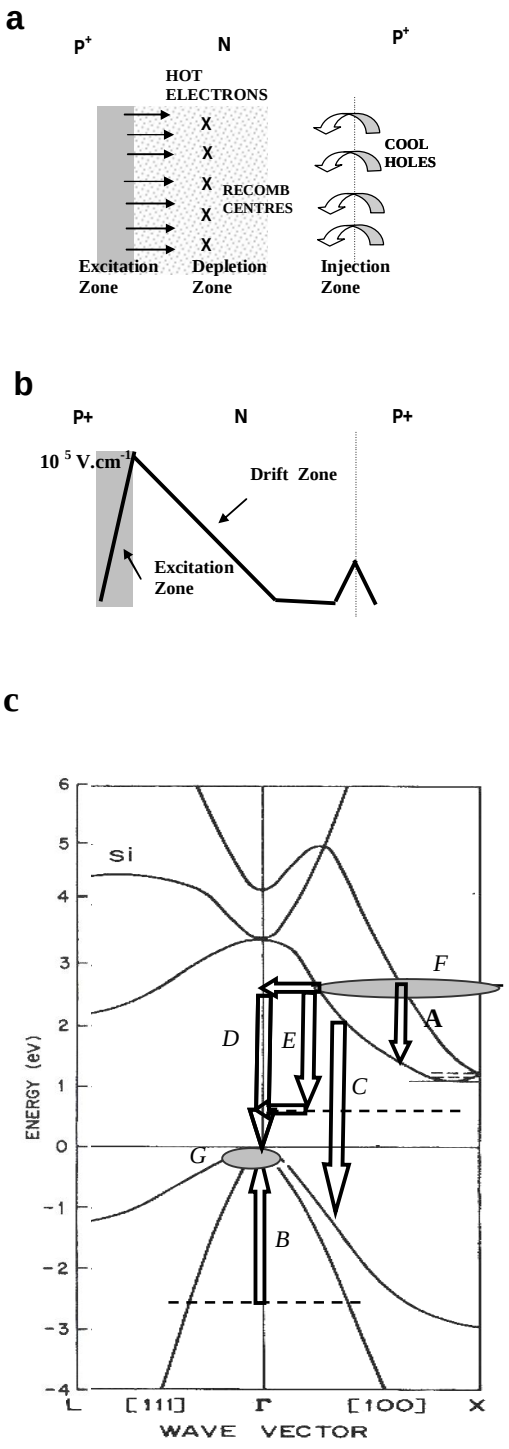


Fig. 2 Schematic representations of the design concepts for a p^+np^+ injection-avalanche CMOS Si LED. (a) Structure of the device (b) Electric field profile through the device (c) Possible photonic energy transitions for the device in the Si band diagram.

$$R_p = \beta n'p' \quad \text{.....(2)}$$

where R_p is the so-called photonic recombination rate, n' and p' are the densities of energetic electrons and holes respectively and β is a recombination probability constant.

For injection of secondary holes into the main avalanching junction ,

$$R_p' = \gamma n' \Delta p \quad \text{…….. ..... (3)}$$

where n' is the density of energetic hot electrons, Δp is the density of additionally injected holes from the second junction, and γ is a corresponding recombination probability constant. The recombination probability constant depends on various secondary parameters such as density of available states, phonon energies available and defect energy states.

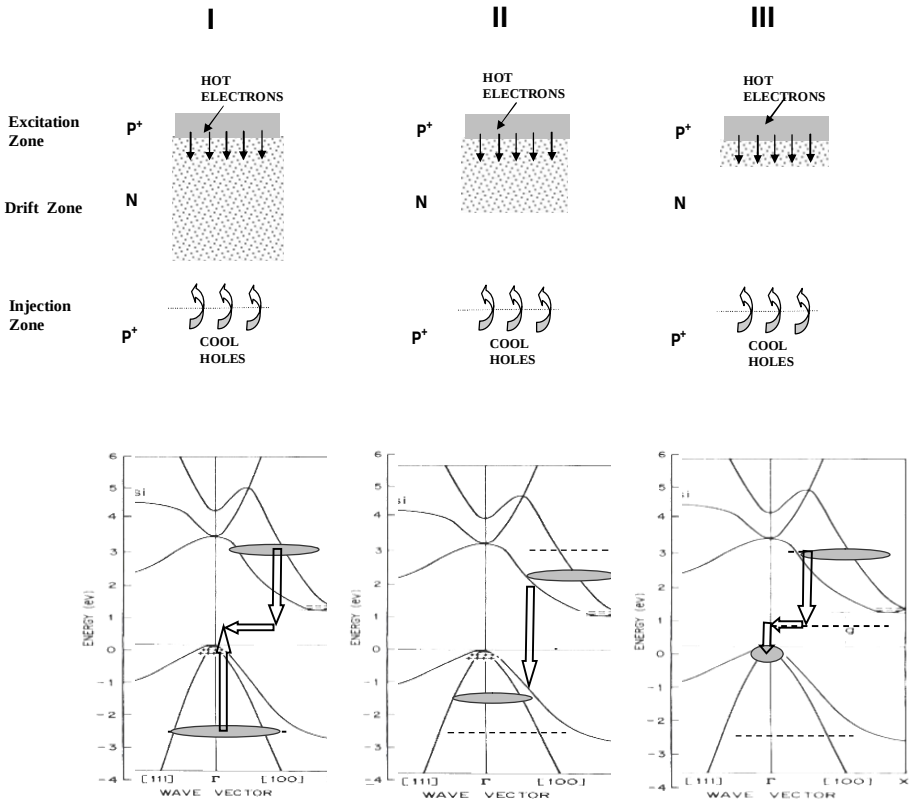


Fig. 3 Schematic presentation of projected dominant higher- than- bandgap photonic transitions in silicon band structure for different depletion layer profiling configurations in a p^+np^+ CMOS LED : (I) Long depletion layer design ; (II) medium range depletion layer design ; and (III) very short depletion layer design.

Consideration of equation (3) clearly reveals that much higher photon emission rates and subsequently increase of orders of magnitude in light emission from the device, may potentially be achieved if the injected hole density, Δp , could be increased several orders of magnitude. These will greatly stimulate transitions of Type C and D and E. Such conditions can particularly be attained by means of the two junction device design as illustrated in Fig. 2. In avalanching junctions the current density through the excitation zone can reach 10^4 A.cm^{-2} at 1 mA, with corresponding average dynamic carrier density of 10^{18} cm^{-3} to 10^{19} cm^{-3} for electrons exiting the junction [20]. The availability of minority carrier holes on the electron-emitting side of an avalanching p^+n junction for this scenario is normally very low and of the order of 10^{1} to 10^{3} cm^{-3}. It follows that if the dynamic cool hole concentration in the neutral regions in the depletion region or in the regions adjacently to the depletion layer, could be raised by five to six orders of magnitude, the recombination rate leading to optical emissions may be correspondingly raised by about six to seven orders of magnitude.

The final optical yield is of course a function of recombination probability constants for the various direct and indirect transition processes. These processes are indeed be the limiting parameter for the light emission processes in Si. However, it is generally accepted that the recombination probability, γ in eq. (3), associated with direct band- to- band transitions is the highest, and, therefore, these transitions (Type C in Fig 2 (c)) should as far as possible be stimulated by means of appropriate device design. Similarly, the indirect transitions of Type D may be stimulated by means of introducing appropriate defect states within the energy band structure for these devices. Both the latter two goals can easily by realized with injection-avalanche Si LED technology.

III. MODELING OF INTRABAND PHOTONIC TRANSITIONS IN SILICON BY MEANS OF DEPLETION LAYER PROFILING

Consideration of various doping profiles and design layouts as hypothesized in Fig. 2, leads to different possibilities with regard to choice of dimensions of the excitation zone, depletion layer thickness and size of the drift zone. On the other hand, the high doping density associated with the p^+n combination, both the excitation zone and hot electron mean free path length

(approximately 0.1 µm) [28] are very narrow, and are not really variable.

Fig. 3 gives a few common scenarios which will each favor inter band photonic transitions. Intra band transitions are assumed to be present but are not shown.

In Configuration I, a long depletion layer (typically 2-3 µm) and subsequent low gradient profile of the E –field are chosen. Injected holes travel subsequently long distances and are heated to high energy levels as they traverse through the depletion layer. This results in both the holes and emitted electrons being high in energy when they interact near the excitation zone, with subsequent little change of direct inter-band recombination.

In Configuration II, the depletion layer is designed much shorter, resulting in a much "steeper" heating of injected holes, before interacting with the emitted energetic electrons. *If the heating profile is roughly of the same dimension of the 0.15 µm free path length of exiting high energy electrons, electrons and holes can interact with each other over a much wider energy range*. Energetic (hot) electrons are losing their energy fast in the depletion layer, while the injected holes are gaining energy fast. It is therefore projected that inter-band direct transitions may be the dominant transition processes if both the emitted carrier densities as well as the injected carrier densities are high. The maximum transitions will occur at the minimum band-to-band edge of approximately 2.8 to 3.2 eV.

In Configuration III, Fig. 3, the depletion layer width is of approximately the same length as the mean free path of the emitted excited (hot) electrons of 0.1 µm, and the electric field profile is very steep. Injected holes will therefore be still in a very low energy state when reaching the emitted hot carrier front. Both carrier densities may be high. Because of this specific energy distribution, it is projected that inter band recombination will dominate, i.e from the maximum of the hot electron carriers (1.8 eV) to the minimum of hole energies (~ 0 eV) , mainly stimulating defect assisted inter-band transitions of approximately 2.8 eV (transitions of Type E in Fig 2 (c)).

Monte Carlo carrier transport simulations in both k space and in E-space as a function of electric field strength in silicon

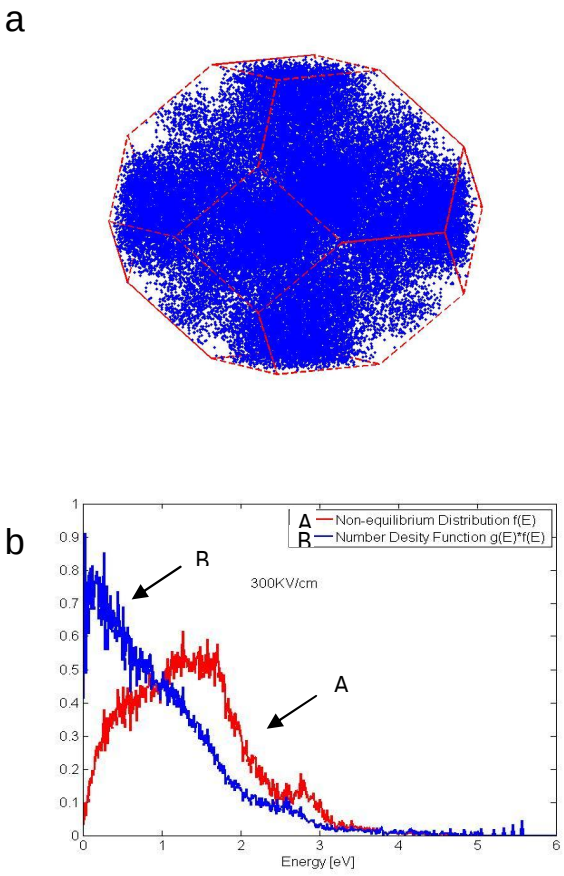


Fig. 4. Monte Carlo simulation of the electron distributions in Silicon at 300kV.cm^{-1}. (a) K-space distribution , (b) Energy distribution in the conduction band .

indeed confirm many aspects of the above modeling. In Fig. 4 (a), the electron distributions were simulated in k-space in the presence of a high electric field of 300kV.cm^{-1} . The carrier distribution in k-space clearly indicate that the

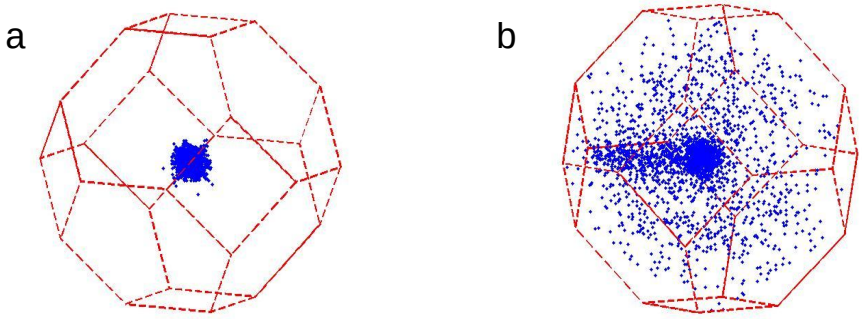


Fig. 5. Monte Carlo simulations of the hole distributions in k-space at (a) 10 kV.cm^{-1} and at (b) 500kV.cm^{-1}.

carriers assume a wide variety of momentum states, quite evenly spread out over the whole range as available. This clearly supports transitions from the whole lateral k space around 1.8eV in the conduction band. Fig 4 (b) demonstrates a favored higher distribution of electron energies around 1.8 eV in the conduction band. An indication of the equilibrium density of states is also indicated.

For holes similar distributions were derived. Particularly significant, were that the hole distributions as function of increasing E-field particularly populates more and more towards higher hole energies and also towards heavy hole bands (See Fig . 5 (a) and (b)). These modeling observations therefore particularly confirms that direct band- to-band transitions may indeed be stimulated under injection-avalanche conditions as modeled above (transition Type C in Fig. 2 (c)).

Fig. 6 attempts to couple some quantitative relationships to the above modeling. Fig. 6 (a) gives projected physically modeled values for the respective carrier densities in the device under dynamic current carrying conditions as a function of distance in the device. The majority of electrons is generated in the excitation zone through multiplication processes. Since the ionization rate for electrons in the prevailing electric field at the n^+p interface during avalanche conditions, is roughly twice that of holes [28] , the excited electron density exiting the p^+n junction interface towards the n-side of the interface, is expected to be twice that of holes, resulting in mainly electrons being populated in the drift zone. If holes are injected into the drift zone from the forward biased p^+n junction, it is expected that interaction between the respective carriers may occur through both energy relaxation and recombination processes in a so-called recombination zone as indicated, resulting in an eventual dynamic carrier density profile through the device as indicated. The estimated hot electron mean path length and profile are also indicated, which is of the order of 150nm (0.15 μm) [28].

Fig. 6 (b) gives projected energy gain and energy loss profiles for the respective charge carriers as a function of distance in the device. Profiles for the energy of electrons (solid lines) and for the energy of holes (dotted lines) are indicated for the three different depletion layer scenarios.

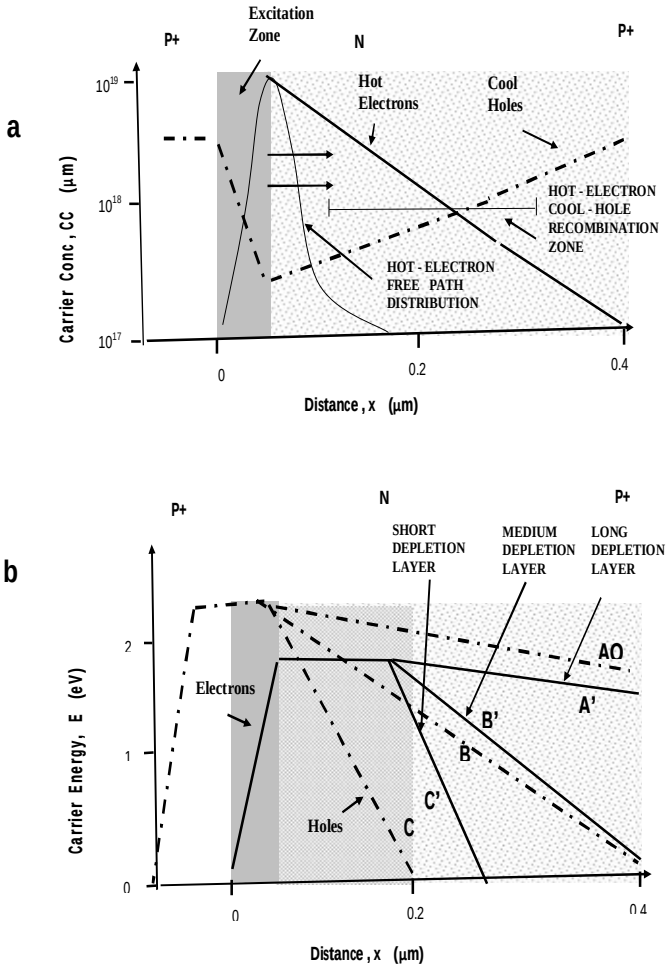


Fig. 6. Schematic presentation of the projected (a) carrier density versus distance in p^+np^+ injection- avalanche Si LED for a medium range depletion layer; and , (b), the associated energy versus distance profiles for the carriers for short, medium and long depletion layer designs, respectively.

From this analysis, it can be derived that a medium-sized depletion layer of approximately 0.4 μm (Configuration II in Fig. 4) with its associated projected energies of the carriers during diffusion is the most favourable device configuration for stimulating direct inter-band photonic transitions in the injection -avalanche device of Type C (Fig 2 (c)). Both the energy values for the exited electrons and holes in the conduction and valance bands have moderate values of approximately 0.8 eV, respectively (Curves B and B' in Fig. 3 (b)), leading to possible direct inter-band transition of approximately 2.8 – 3.2 eV.

Secondly, it can be derived from these profiles that a short depletion layer configuration will result in electron and hole energy profiles as indicated by Curves C and C' respectively. This energy profile is the most favorable for stimulating transitions of Type D and E in Fig 2. (c) , since it favors recombination of high densities of high energy electrons with high densities of low energy holes, as in Configuration III, Fig. 3. The optimum depletion layer thickness for this scenario is projected to be in the order of 0.3 µm. The current densities for both electrons and holes at this distance are also both high (Fig. 6 (b), and therefore favours a very high recombination probability according to equation 3. The E-field gradient for this configuration is calculated as $5x10^5$ $V.cm^{-1}\mu m^{-1}$.

The presence of mid-band defects states may enhance indirect recombination processes, (transitions of Type D in Fig. 2 (c)) and in Configurations II and III, Fig. 3), since both the energy values and high densities of carriers present may favor these transitions according to Equation 3. A variety of methods are currently available in order to introduce specific type of defects at specific locations and depths in silicon devices. These include, implantation of specific atom species followed by post annealing procedures in order achieve specific defect levels and active defect sate densities [29,30,31], direct wafer bonding in inert atmospheres [32], as well as current density stressing techniques [30,33].

IV. ENHANCING EFFICIENCY BY MEANS OF PUNCH THROUGH CONFIGURATIONS

The heating profiles of the injected holes in the structures may be further modified by insertion of a p^+ finger into the depletion layer region, i.e when the p^+n junction touches or is inside the depletion region during voltage bias conditions. Although the E-field profile is predominantly a function of the doping levels in the n-side, the gradient of the heating profile for holes will now be function of the distance, shorter distances leading to higher heating profiles and that could particularly stimulate inter band indirect transitions of Type D and Type E as in Fig. 3, Configuration III. Since the decay in E field in the depletion region is mainly a function of doping profile, it follows furthermore that the electron energy may be selected for preferred interaction with the low energy holes. This may even favour modulation of the emitted wavelength of the radiation.

The punch trough configuration is particularly favorable since it injects low energy holes of very high density directly into the depletion region, and hence avoid minority carrier current losses in the neutral n-regions that are not in depletion mode. These regions can act as additional traversing routes for injected carriers and will not contribute towards light emission processes. This aspect has been identified as a limitation in our previous injection based designs [24,25].

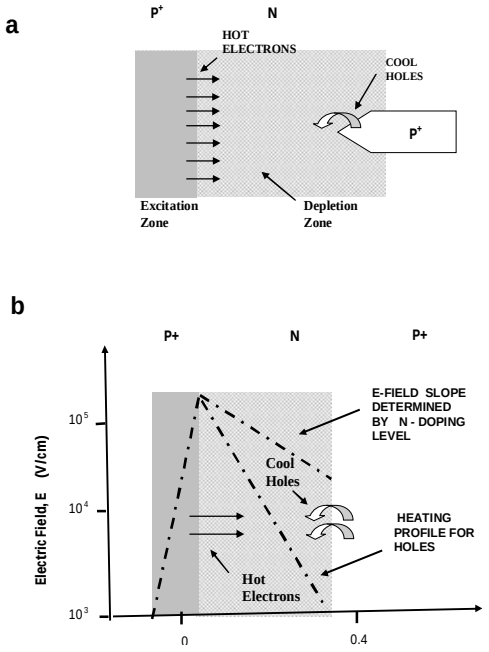


Fig. 7 Schematic presentation of (a) of the inclusion of a p^+ finger that penetrates the depletion layer under dynamic current carrying conditions corresponding to punch through conditions; and (b) the corresponding electric field and hole heating profile in such conditions.

The design will also greatly reduce energy dissipation of the device since carriers do not have to traverse long neutral regions under low E-field conditions.

V. PROTOTYPE DEVICE DESIGN

We have tested some of the above modeling by designing and realizing some prototype test structures.

Fig 8 (a) demonstrates a five bodied test structure (TS1), viz a p^+ needle shaped body at terminal T1, facing a pentagon shaped p^+ body at terminal T3, and, diagonally opposite, two n^+ bodies, all embedded in a n-well substrate. If T2 is made the common (grounding) element and a negative bias voltage is applied to T1, a p^+n abrupt avalanching junction is formed at the needle like tip with its depletion layer extending away towards the diagonal bodies T2. If appropriate positive bias voltage is simultaneously placed on T3 relative to T1, the p^+n body at T3 (J2) is forward biased and injects holes towards T1. The extend of the avalanche process in the main p^+n junction, as well as the extent of forward bias by T3 was controlled by means of an external bias resistor, R. Conditions were chosen such that for bias between T3 and T1, at low

currents, the main current predominantly flowed between the diagonal n^+ (avalanche control) bodies, while at higher currents the main current flows between the injection body T3 and T1. The cross-section through the device along the dotted curve in Fig. 7 (a) is shown adjacently. The distance from the injection point to the p^+n interface was 2.0 µm, therefore creating a medium heating profile for injected holes in a projected E-field gradient of 2×10^{-4} V.cm^{-1} µm^{-1}.

Fig. 8 (b) shows a test structure (TS2), comprising a small p^+ body pointed body embedded in a n-well and again embedded in a larger p-well. The p^+n junction (J1) was reverse biased through voltage biasing between T1 and T2. By voltage biasing between T1 and diagonal terminals T3, the larger n-well p-well junction J2 was slightly forward biased and hole carriers were injected into the depletion region at J1. By choosing the dimension and position of the n-well tub accurately, the extension of the depletion layer towards the T3 terminals was limited and a punch-through condition was reached. This created a medium sized depletion layer of about 1 micron into which hole carriers were injected. Hence a device was created of which the depletion region width could be dimensioned to be in accordance with Configuration 2 or Configuration 3 in Fig. 3. A trapezium shaped poly-Si layer and CMOS gate oxide layer was also placed on the structure (T4). This enabled a slight overlap of oxide over the reversed biased junction J1 for specific investigation purposes, and allowing modification of the depletion region to only 0.5 micron, also corresponding to a punch through condition and a heating profile of 5×10^5 V.cm^{-1}. µm^{-1}

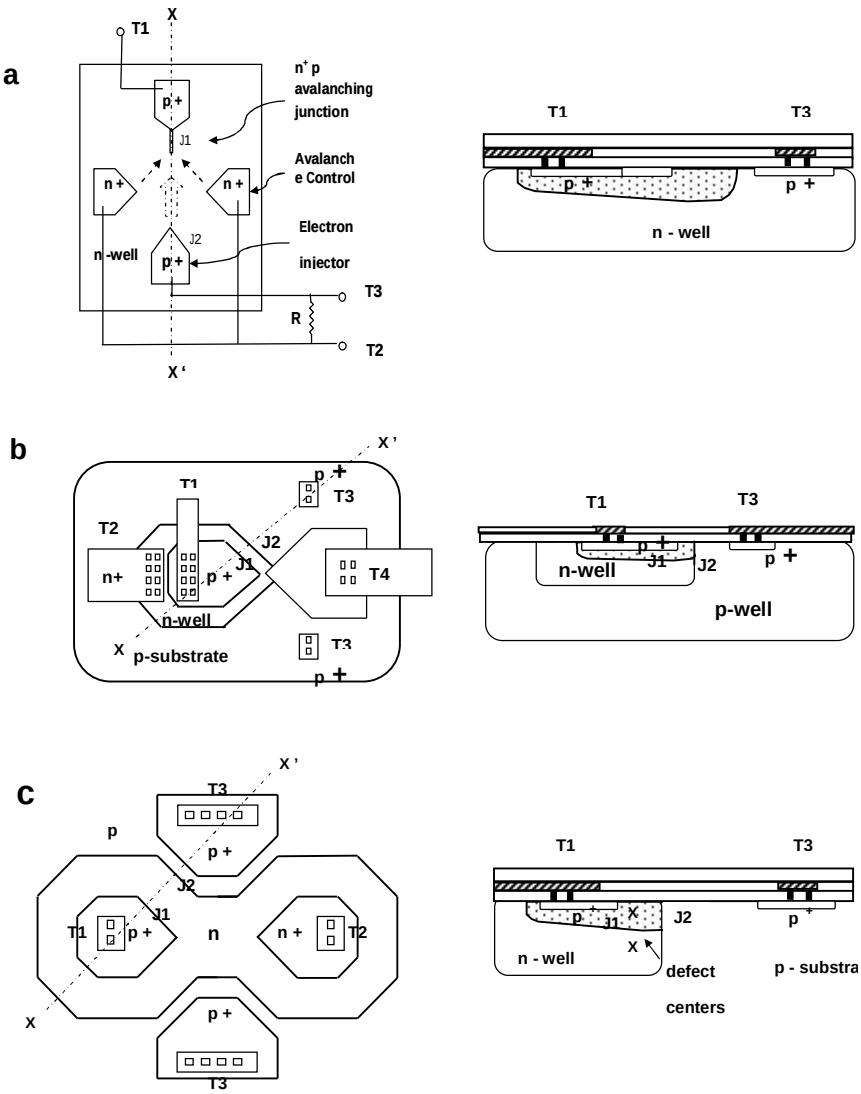


Fig. 8. Test structure designs for (a), of p^+np^+ Injection-Avalanche Si LED with a long depletion layer design (TS1) ; (b) limited depletion layer design (TS2); and (c) medium depletion layer design plus defect engineering applied. The respective body and layer designations are applicable to 0.35 µm CMOS design and processing procedures.

A further test structure, TS3, is demonstrated in Fig. 8 (c)). The p^+ body was place in a n-tub well again. Biasing between T1 and T2 enables an elongation of the depletion layer towards T3. The width of the depletion region could again be varied in order to limit the depletion region width by choosing the dimensions and positioning of the larger n-well carefully. The main device current eventually flows predominantly between terminals T1 and T3, and creates a reverse biased p^+n junction in series with a forward pn junction. The extent of the avalanche process at J1 was controlled by the voltage bias between T1 and T2. By stressing the junction at J2 beforehand according to a predetermined current density vs lattice stress calibration curve, some densities of defects could be introduced at the p^+n interface before operating the device. Hence the optical yield at J1 could be investigated when defect centers were present at J1.

All the above devices was realized using 0.35 µm state-of-the- art CMOS technology design and processing procedures [34].

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

Fig . 9 demonstrates some of the most important results as observed. Fig. 9 (a) (b) and (c) shows bright field optical micrographs of the test structures 1 to 3 as seen through the optical microscope with visible light and at very high magnification. Adjacently in Fig 9 , (d) and (e) , (f) , (g) and (h), corresponding optical emission performances are illustrated as taken under dark field conditions with a CoolPix Model 4500 Nikon CCD Digital Camera with recording facilities. A progressive increase in total optical emission as well as in localised intensities are demonstrated in Fig. (d) to (h) at roughly similar operating voltages and currents.

Fig. 9 (d) shows the TS1 version of the device under operation for pure avalanching conditions at J1 with no current injected into the junction J1. The extent of avalanching and light emission around the peripheries of the p^+n junction can clearly be observed. At the tip of the body a much brighter zone can clearly be identified. This phenomenon is attributed to higher optical emission intensity as a result of current density confinement according to previous interpretations [20].

Quantitative analyses of the optical emission versus electrical power input to the device as in Fig 10 (a), reveal an interesting two sloped curve which clearly indicate that a mechanism change occurred for the device at about 10 mW. This mechanism is attributed to a partial collapse of the depletion layer as they occur in three terminal transistor devices for injection conditions in traditional emitter to base configurations [35]. The optical yield for only avalanche operation at low current levels is about 2 photomultiplier units (PMU's) per mW of electrical input power, while at higher current levels (injection conditions) , the slope is higher and is of the order of about 4 PMU's per mW .

Fig. 9 (e) show the characteristics for the TS2 version device for roughly similar operating conditions as in TS1. In this case the total p^+n junction length and pn^+ junction area were much larger (See dimensions in Fig. 9 (b). Particularly noticeable, is the much higher brightness along the right hand side periphery of the device, where according to our analyses, much higher injection current of holes occurred into the junction, as compared to peripheries on the left hand side which is closer to the avalanche control terminal T2 and which represents light

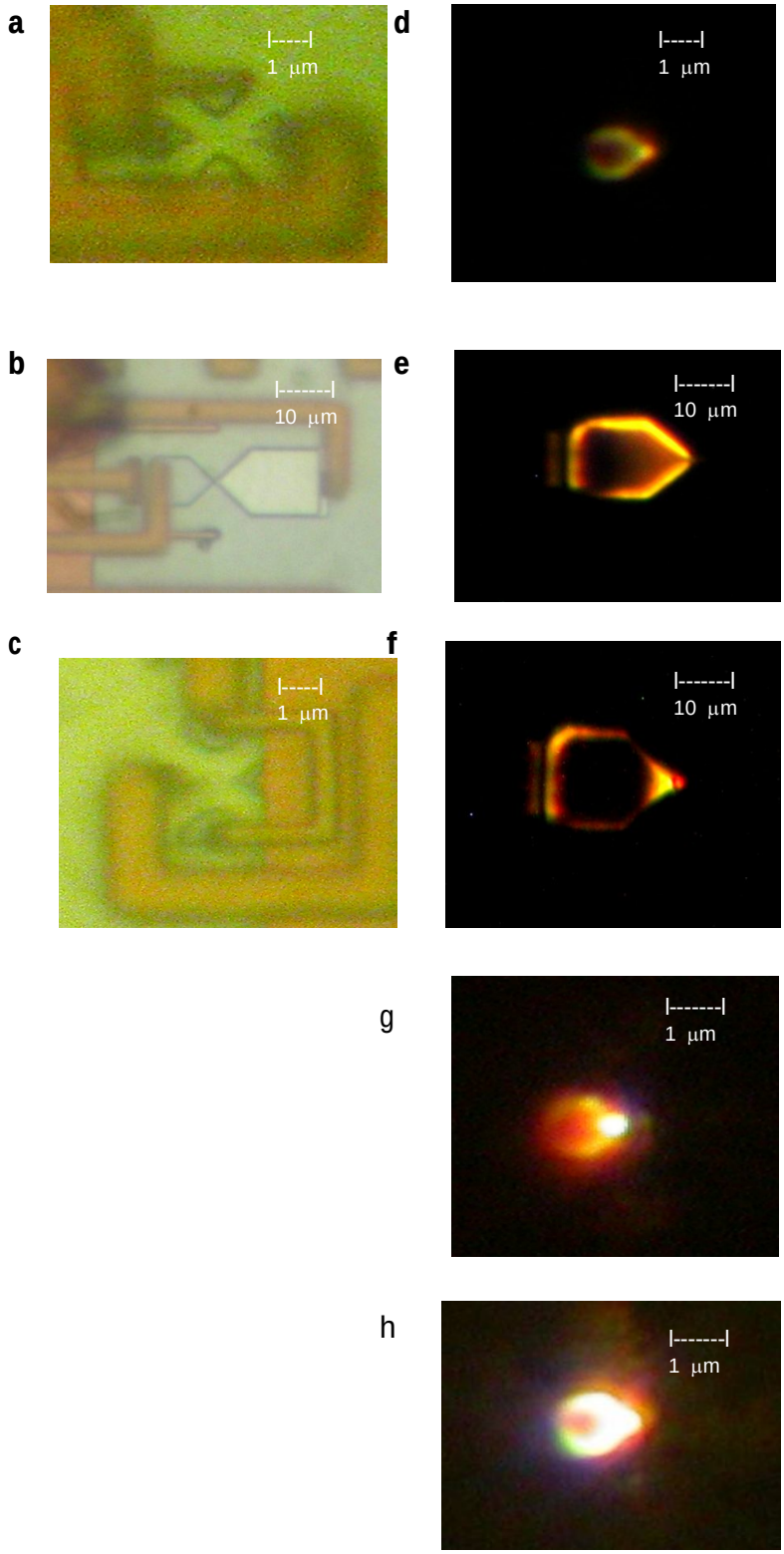


Fig. 9: Bright field and dark field photo micrographs for Si injection-avalanche CMOS Si LED's: (a) Test Structure 1 (TS1) (b) TS2 with ploy Si layer added. (c) TS3 structure (d) TS1 for pure avalanche operating conditions for 15V and 1 mA. (d) TS2 for 8V and 2mA operation. (f) TS2 plus Poly Si test structure at 15V and 1mA operation. (g) TS3 for 8V and 1mA operation under mixed avalanche-injection operation (80-20%). (h) TS3 for 8V and 3 mA operation under full injection conditions.

emission corresponding to pure avalanching conditions. Quantitative analyses as in Fig 10 (b) show a clear changeover in mechanism with a generated higher optical yield of approximately 20 PMU's per mW of input electrical power for this device at an estimated 50% hole injection conditions. It is important to note that because of the larger dimension of the p^+ layer in this device, a large injection current component was injected towards the bottom part of the p^+ layer embedded in the n-well (Fig. 8 (b)) cross-section layout and did not contribute towards light emission at the p^+n peripheries. Considering these conditions, as well as the fact that the current density at the p^+n peripheries was much lower during injection conditions as during pure avalanche, the inherent optical yield during injection-avalanche for smaller depletion layer configuration is derived to be much higher. From the data curves as derived in Fig. 10 (b), the projected internal optical yield for the 1 micron limited depletion layer conditions as in TS2, Fig. 9 (b), is anticipated to be of the order of 400 PMU's per mW of input electrical power (dotted curve C, higher end). This curve was derived by means of the following approximated expression:

$$Npp = \frac{\Delta P MU\ (O)}{Vd\ \Delta Id} \times \frac{At}{Ac} \qquad \text{..... (3)}$$

where

Npp = the projected internal power conversion efficiency;

$\Delta P MU\ (O)$ = incremental optical power output in PMU units ;

Vd = bias voltage for the device at the change in slope ;

ΔId = incremental current change at change in slope

At = total depletion layer cross-section area that absorbs injected current; and

Ac = depletion layer area that contribute towards light emission .

Further experiments were conducted with the same device structure but with a poly-Si gate overlapping the n-well as in Fig. 9 (b). When a leakage path was set up through the oxide layer, current could be injected into the depletion layer only 0.5 micron from the p^+n junction interface. Particularly noticeable was a very bright (whitish in color) but very small spot (Fig. 9 (f)). Also noticeable is a fainter (more reddish in color) second spot at the right hand edge, a very short distance away. The magnitude of brightness of the first spot is at least one order of magnitude higher than the periphery emissions when the intensity profiles are

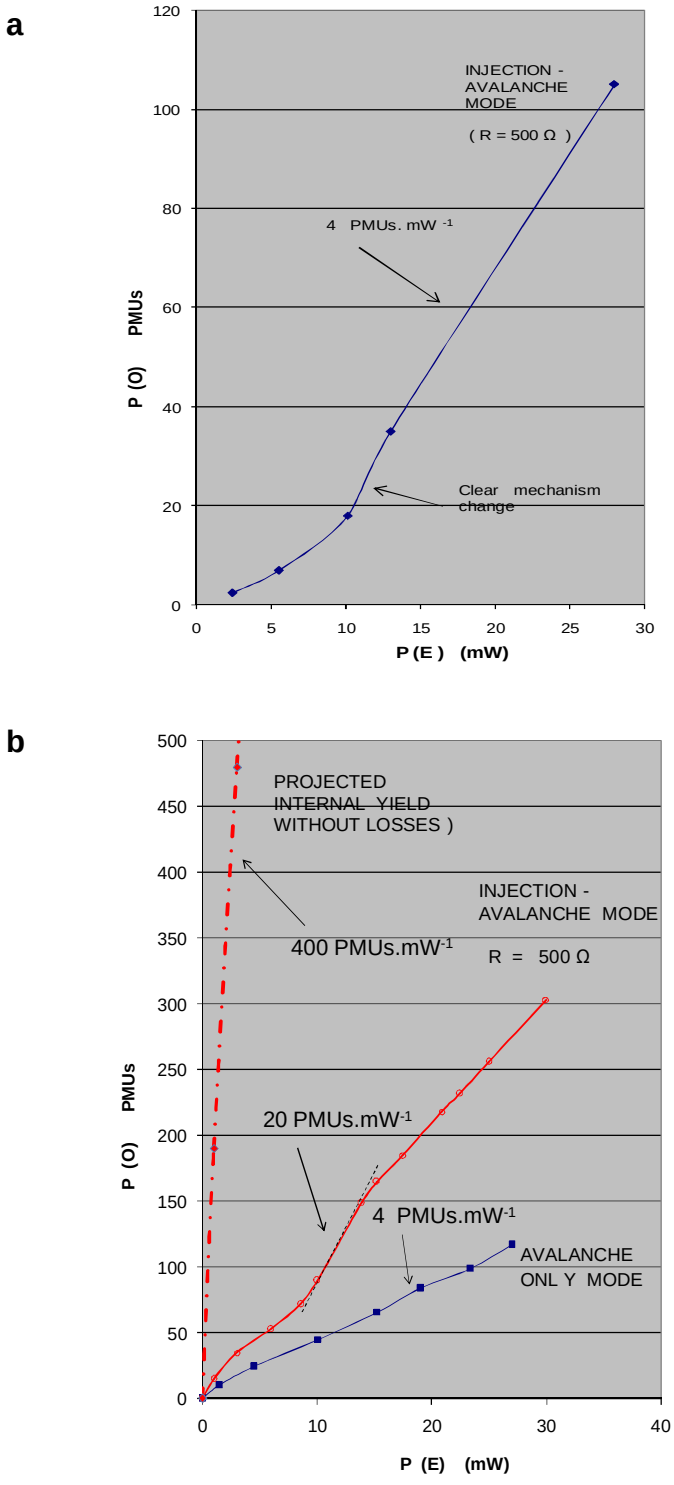


Fig. 10 (a) Optical power versus electrical input power performance for TS1 in injection avalanche mode of operation and depletion layer width 2 µm.. (b) Optical power versus electrical input power performance for TS2 with depletion layer width , 1 µm. The projected internal optical yield without losses d due to electron-hole interaction in the device is indicated by means of the dotted curve.

considered. It is our opinion that the spot emission originates from the interaction of exciting high energy electrons at the p^+n interface with injected holes into the depletion region only 0.5 µm away from a large sea of electrons as generated by the avalanching junction at the p^+n interface. The curves in Fig. 11 (a) as well as the observation of the formation of secondary light spots with increase in bias voltage, indicated that some of the light emission processes in this device was much more efficient than in previous devices. The projected contribution of the brightest spot to the total emitted power is indicated by the dotted curve. Investigation of the total emitted intensities versus area revealed that there was a high leakage current present in the device due to avalanche conditions in the rest of the p^+n periphery. The device configuration and operation is therefore not optimized yet. After calculations, the inherent optical yield for the bright spot was projected to be of the order of 1000 PMU's per mW of input electrical power, if it is assumed that only about 50 µA contributed to the spot emission. The following approximated expression was used in this calculation:

$$Np\ (Local) = \frac{\Delta P\,MU\ (O)}{Vd\ \Delta Id} \times \frac{A\ t}{A\ (Spot)} \qquad \text{.....(4)}$$

where

$Np\ (Local)$ = the local power conversion efficiency;
$\Delta P\,MU\ (O)$ = incremental optical power output in PMU units ;
Vd = bias voltage for the device at the change in slope ;
ΔId = incremental current at change in slope
$A\ t$ = total area that contribute to light emission; and
$A\ (Spot)$ = light emitting area for the spot causing the change in light emission .

Fig . 9 (g) and Fig. 11 (b) show emission phenomena that was observed for Test Structure 3, Fig. 8 (c), into which defect densities have been introduced near the p^+n interface. Again the presence of a bright spot is observed which is much brighter than the areas which are in pure avalanche light emission mode of operation (seen towards the left and right along the p^+n periphery). Particularly noticeable, was the growth of the spot intensity (almost as a flare explosion) as current was increased. This again indicated a further change of light producing mechanism as the net increase in detected optical power now comes from only a very small area in the device of approximately 0.1 µm x 0.1 µm. Quantitative investigations of optical yield as a function of incremental electrical power increase and as a function of contributing area to total light emitting area yields an optical yield of approximately 50 PMU's per mW external and projected 1000 PMU's per mW of electrical input power inherently per defect generating area if losses are ignored (Several cases

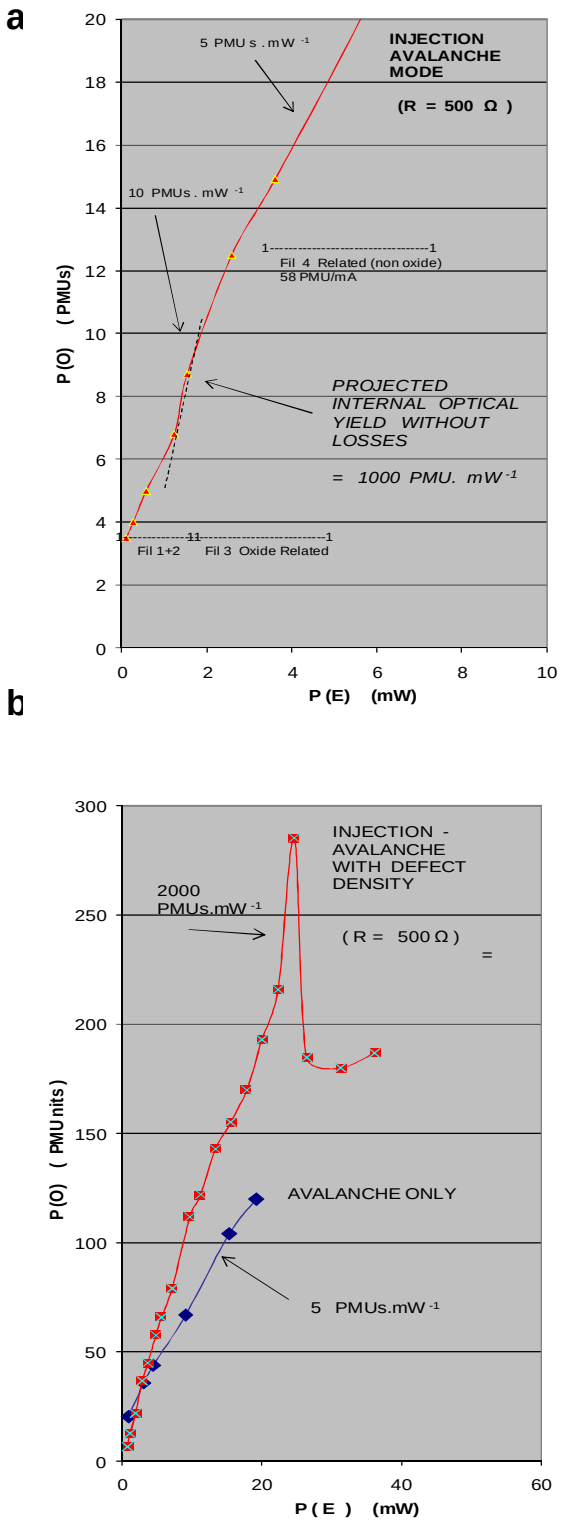


Fig. 11 . (a) Optical power versus electrical input power performance for TS2-Poly Si Injection Test Structure with depletion layer width of 0.5 µm . (b) Optical power versus electrical input power performance for TS3 for injection –avalanche mode of operation and with defect densities introduced into the depletion layer. The localized optical yield as emitted due to electron-hole interaction and defect interaction is indicated .

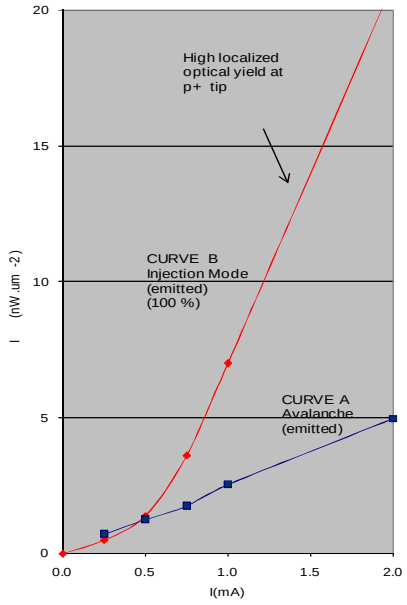


Fig. 12. Optical intensity versus electrical current performance for TS3 for 20% avalanche current and 80% injection current.

were investigated). In this particular case the particular injection conditions into the depletion layer has not been optimized which predicts that the eventual emission could still be much improved if the device is further optimized.

Fig. 9 (h) shows an optimized version of the operation of the TS3 device with the p^+n junction, needle tip device in strong avalanche and with very high hole current injected (approximately 2 mA) into the avalanching junction, which we attribute to a culmination and a combination of all the effects as discussed above. The clear higher intensity at the needle tip region is clearly apparent, as well as the definite much more blue-whitish color of emission as compared with pure avalanching light emitting junctions. The optical emission per area (intensity) of the light emission as emitted is very promising , and is of approximately 10 nW . μm^{-2} at 1mA of current at 8V .

SOME SPECTROGRAPHIC OBSERVATIONS

Fig. 13 shows typical spectrographic information as measured for our devices. Fig 12 (a) shows the wavelength spectrum of the emitted light as observed for a pure avalanching junction of type TS1, while Fig 12 (b) shows the observed change observed in the spectra as a function of increase in injection current. The spectra has been corrected for absorption effects as they occur in the silicon subsurface as a function of wavelength. Spectroscopic measurements of the emitted light reveals clear emission peaks at 1.8eV, 2.2 eV 2.4 eV and 2.8 eV in the 450 - 750 nm wavelength region. Very significant is the a prominent

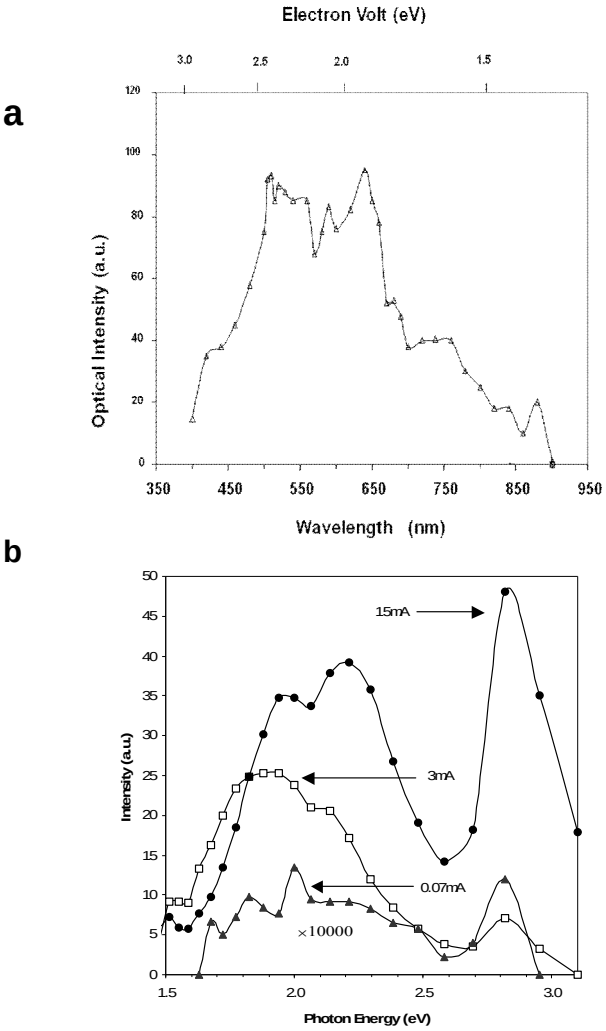


Fig. 13 : Intensity spectra for optical emissions for , (a), CMOS LED for pure avalanche mode of operation. (b) Intensity versus photon energy for operation for a CMOS LED in the injection-avalanche mode of operation for various injection current levels . The wavelength emissions have been corrected to present optical yields at the generation sites.

and significant growth of the 2.2 and 2.8 eV peaks when the injection current was increased. The spectroscopic information therefore confirm some interesting phenomena:

(1) The observed 1.8 eV photonic emissions as observed in the spectra may correspond with intra-band relaxation mechanisms within the conduction band (Transition Type A, Fig. 2 (c) as a result with interaction of host Si atoms. The value of 1.8 eV corresponds with energy required for ionization of host silicon atoms by electrons [27].

(2) The observed 2.4 - 2.5 eV photonic emissions as observed in the spectra may correspond with intra-band relaxation mechanisms (Transition Type B, Fig. 2 (c)) within the conduction band as a result of interaction with host Si atoms.
(3) The observed increase in the 2.8eV energy peak in Fig. 12 (b) as a function of injection current may indicate that

band-to-band transitions of Type C and Type D s described in Fig. 2 (c) and Fig 3 Configuration III may indeed be present. Since the inter-band transition probability is proportional to the both the carrier densities n' and Δp , as predicted by equation (3), the 2.8 eV peak should definitely grow with increase in hole injection current. This is indeed observed.

(4) The presence of a clear emission peak at 2.2 eV indicate that the phonon-assisted transitions from the 1.8 eV energetic position for electrons, to defect states to the middle of the band gap are highly feasible and that it may indeed be stimulated by inter-band acoustical phonon assisted transitions of Type E, as in Fig. 2 (c). According to equation (3), these transitions will also be dependent on the product of the densities of the excited electrons as well the density of injected low energy holes. A non linear growth of this peak is therefore expected. From the results shown in Fig 12, this seems to be observed. This observation also correlates with the non-linear increase in the optically emitted power as function of current at the higher current levels when predominantly injection current as observed in both Fig. 9 (a) and Fig. 11 further confirms this phenomenon. Both the latter two physical trends thus support the 2.2 eV defect and phonon dependant optical emissions.

VII. SOME QUANTITATIVE DERIVATIONS

Using calibrated external illumination sources, both the CCD camera as well as the photo-multiplier could eventually be calibrated. This enabled conversion of the PMU photonic yields units into more quantitative values. The photo-multiplier tube was calibrated as approximately 10nW per 100 PMU's for the particular measuring conditions used. Using photographic white levels and objective acceptance angles as a measure, the optically emitted power per unit area on the devices (i.e intensity) could also be derived from the CCD camera photo-micrographs. Fig. 12 gives typical characteristics as derived. An intensity values of up to 100 nW. μm^{-2}, could be derived at 8V and 1 mA operation for a 0.1 μm^{-2} emitting area of the TS2 Poly Silicon device .

Considering the optical power outputs per unit area as in section 3 wrt current and voltage bias conditions, (1 nW μm^{-2} at 1 mW for single junction mode of operation and 10 nW μm^{-2} at 1 mW for injection mode of operation), estimations of the respective external and internal power conversion efficiencies could be made. This was achieved by taking the 2.2 eV (650nm) peak as main reference peak. Calculations using the refractive indices of n = 3.48 for Si, an absorption factor for Si of 0.59 at 650nm, and internal reflection considerations as the light moves from higher refractive indices to lower refractive indices surface layers, reveals an emission factor of 0.21 for emission from Si to SiO_2. Calculations using the refractive indices of n = 1.4 for SiO_2, 2.5 for Si_3N_4 an absorption factor of 0.3 for Si_3N_4 layer, and considering 1 μm thickness layers, reveal emission factors of 0.25 for transmission through the SiO_2 and Si_3N_4 surface layer to ambient air. The Si_3N_4 layer contributes to an absorption loss of about 0.3. The total emission factor from the subsurface n^+p-avalanching junction to air is therefore $1.5x10^{-2}$. The implies that the internal power conversion efficiency for a single avalanching junction is of the order of 10^{-6} and of the order 10^{-5} for two junction injection mode of operation. However, the injection mode of operation suffers from current contribution as well as current density reduction loss through the active avalanche–injection light regions of part of the device of the order of 10^{-2} (The injection current does not conduct only through the needle tip at J1 but through the whole n^+p reverse bias junction including the bottom surface area of the n^+p embedded region as well, Fig. 7 (a)). This implies that the maximum achievable power conversion efficiency for injection mode of operation alone may be as high as 10^{-3} to 10^{-2} . This is much higher than the previous determined values of approximately 10^{-6} for single junction avalanche only devices [10, 19].

The internal power to optical quantum conversion efficiency can be derived from the relationship:

$$N_Q = \frac{(e \lambda V)}{h c} N_p \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where N_p denotes the power conversion efficiency, e denotes the elementary charge, λ the wavelength , V is the applied voltage, h is the Planck constant and c is the speed of light. [11]. This relationship represents the ratio of the number of photons generated at the junction to the number of charge carriers that pass through the junction. Using this relationship, values for the internal quantum conversion efficiency of the order of 10^{-5} is derived for our single junction avalanching junctions and of the order of 10^{-2} is derived for injection-avalanche light emitting process .

In summary, the experimentally realized as well as the projected optical yield as a results of injection-avalanche mechanisms are presented in Table 1. Particularly promising is the optical yield that can be realized with a combination of depletion layer and defect engineering with TS3 if the device designs .

VIII. CONCLUSIONS

(1) It was demonstrated that the optical emissions for two junction injection-avalanche based silicon light emitting devices are dependent on depletion layer profiling as well as on the introduction of defect centers in the depletion layer.

(2) The obtained experimental observations sufficiently supports inter band photonic transitions according to the l modelling and hypotheses that are based on the potential increase in optical emission due to interaction of injected low energy holes with higher energy electrons in an electric field gradient of approximately

4×10^5 $V.cm^{-1}$. μm^{-1} as created by depletion layer and defect state profiling techniques.

(3) Particularly promising is that a much higher emission intensity of up to 100 nW μm^{-2} was observed for the injection-avalanche mode of operation as compared with previous results using single junction devices operating in the avalanche light emitting mode and which only typically yielded 0.1 – 1 nWμm^{-2} [20]. This is a very important characteristic, since the concentration of optical power in small micro dimensions and also the direction of optical power in certain directions in future microstructures will be of extreme importance

(4) It that it was observed that the introduction of defects, as caused by junction current stressing, remarkable increase optical emissions.

(5) Indications have been obtained that the inherent and internal yield due to these mechanisms are much higher. Therefore the external optical power yield as associated with these devices could be much higher if further optimization is implemented with regard to device design. The absorption and emission losses from the surface layers in the current structure still offer major challenges

(6) To our opinion a much better understanding of the basic mechanisms that is responsible for higher than band-gap light emission mechanisms in silicon has been achieved.

ACKNOWLEDGEMENTS

The utilization of facilities at the Carl and Emily Fuchs Institute for Microelectronics is great fully acknowledged.

REFERENCES

[1] E.A. Fitzgerald, L.C. Kimerling, "Silicon-Based Technology for Integrated Opto-electronics", *MRS Bulletin*, . 39-47, (1998).

[2] R. Soref, "Applications of Silicon-Based Optoelectronics", *MRS Bulletin*, . 20-47, (1998).

[3] D.J. Robbins, Editorial comments on volume "Silicon Opto-electronics" at Photonic West annual conference at San Jose, California, U.S.A, *Proc. of SPIE* (3953), (2000).

[4] P. M. Fauchet ,"Progress Toward Nanoscale Silicon Light Emitters,", *IEEE Jour. Selected Topics in Quantum Electron.* 4, 1020-1028 (1998) (Invited).

[5] N. Savage , "Linking with Light", *IEEE Spectrum*, (39) , pp. 32–36, (2002).

[6] A. Chatterjee, B. Bhuva and R. Schrimpf, "High-speed light modulation in avalanche breakdown mode for Si diodes", *IEEE Electron Device Letters,* Vol. 25, No. 9, pp.628-630, September 2004.

[7] K. Wada, "Electronics and Photonics convergence on Silicon CMOS Platforms" *Proc. of SPIE* (5357) , 16 (2004).

[8] W.G. Ghynoweth, K.G. McKay, *Physical Rev.* (102) , 369-376 (1956) .

[9] R. Newman, "Visible light emission from a silicon p-n junction" *Phys.R ev.* (100), 700 – 703 (1955).

[10] J.Bude, N Sano, A Yoshii, "Hot carrier Luminescence in Silicon" *Physical Review B,* Vol 45, No 11, pp 5848 – 5856, 1992.

[11] J. Kramer, P. Seitz, E. F. Steigmeier, H. Auderset, B. Delley, *Sensors and Actuators* (A37-38), 527-533 (1993),

[12] L. Carbone, R. Brunetti, C. Jacoboni, A. Lacaita, M. Fischetti, "Polarization analysis of hot-carrier emission in silicon, " *Semicond.Sci. Technol.*(9), 647-676 (1994).

[13] Lacaita, F. Zappa, S. Bigliasrdi, M. Manfredi, "On the Brehmstrahlung origin of hot-carrier-induced photons in silicon devices", *IEEE Trans. Electron Devices (ED40),* 577-582 (1993).

[14] N. Akil, S. E. Kerns, D. V. Kerns, Jr., A. Hoffmann, J-P Charles, "A multi-dimensional model for photon generation in silicon junctions in avalanche breakdown", *IEEE Trans. Electron Devices* (46) No5, 1022-1027 (1999).

[15] L.W. Snyman, M.du Plessis, E. Seevinck, and H. Aharoni, "An efficient , low voltage, high frequency silicon CMOS light emitting device and electro-optical interface" , *IEEE Electron Device Letters (U.S.A.)* (20), 614-617 (1999).

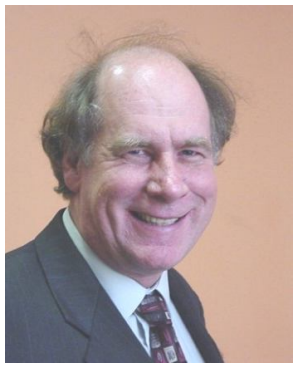
[16] M. du Plessis, H. Aharoni and L. W. Snyman , " Silicon Light Emitting

Table 1 : Optical yields realized, optical intensities realized and internally projected optical yields for the various Si CMOS LED test structure designs under specific operating conditions.

DESIGN	**TOTAL EXTERNAL OPTICAL YIELD** (nW mW^{-1})	**PROJECTED LOCAL INTERNAL OPTICAL YIELD (INJECTION RELATED)** (nW mW^{-1})	**EMITTED OPTICAL INTENSITY (Maximum as in 0.1 μm^2 areas)** (nW μm^{-2})
TS1 p^+np^+ injection-avalanche Only avalanche mode of operation	0.2		1
TS1 p^+np^+ injection-avalanche mode (100 % injection current)	0.4	10	10
TS2 p^+np^+ injection-avalanche mode (100% injection current) (depletion layer limitation = 1 μm)	1	40	15
TS2 p^+np^+ injection-avalanche (PolySi avalanche control) (80% injection current) (depletion layer limitation = 0.5 μm)	0.4	100	100
TS3 p^+np^+ injection-avalanche (100% injection current) (depletion layer limitation = 2 μm) (defect engineering introduced by means of current stressing beforehand , 10^5 A . cm^{-2}, 1 minute)	2	200	100

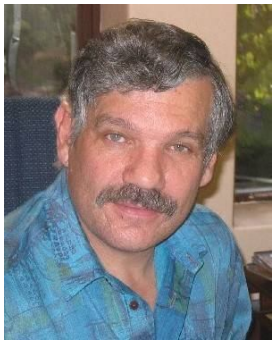
Devices in Standard CMOS Technology" , *Proceedings of the IEEE Conferenta Anuala de Semiconductoare (*CAS) , 231- 238 (2001) (invited paper).
[17] L.W. Snyman, H. Aharoni, M. du Plessis, R.B.J. Gouws, "Increased efficiency of silicon light emitting diodes in a standard 1.2 micron complementary metal oxide semiconductor technology" , *Optical Engineering (U.S.A.)* (37), 2133 – 2141(1998).
[18] L.W. Snyman, H Aharoni, M. du Plessis, J.F.K Marais, D van Niekerk and A Biber," Planar light emitting electro-optical interfaces in standard silicon complementary metal oxide semiconductor integrated circuitry",*Optical Engineering (SPIE-USA)* (41), 3230 – 3240 (2002).
[19] M. du Plessis, H. Aharoni and L.W. Snyman: "Two- and multi-terminal CMOS/BiCMOS Si LED's." *Optical Materials* (27), 1059-1063 (2005).
[20] L. W. Snyman, H. Aharoni and M du Plessis, "Two order increase in the quantum efficiency of silicon CMOS n+pn avalanche-based light emitting devices as a function of current density" *IEEE Photonic Technology Letters* (17), 2041-2043 (2005).
[21] M. du Plessis, H. Aharoni and L.W. Snyman: "Spatial and intensity modulation of light emission from a silicon LED matrix.", *IEEE Photonics Technology Letters*, Vol. 14, Issue 6, pp. 768-770, June 2002.
[22] L.W. Snyman, M. du Plessis and H. Aharoni, *Proc. of the IEEE International Symposium on Industrial Electronics (ISIE* 2005)(3), 1159 - 1169) (2005) (Special Session Paper).
[23] L . W. Snyman, M. du Plessis, H . Aharoni, "Three terminal optical sources (450nm - 750nm) for next-generation silicon CMOS OEIC's" ,*12th International Conference, Mixed Design of Integrated Circuits and Systems (MIXDES'2005)* Krakov, Poland, 737 – 747 (2005) (Invited Plenary Session Paper).
[24] L. W. Snyman, M. du Plessis, and H. Aharoni, " Two order increase in the optical emission intensity of CMOS integrated circuit LED's (450 – 750 nm) Comparison of n^+pn and p^+np designs " , ***Proc. of the SPIE (5730) ,*** 59-72 **(**2006).
[25] L. W. Snyman, M. Du Plessis and H. Aharoni, "Injection-avalanche based n+pn Si CMOS LED's (450nm . 750nm) with two order increase in light emission intensity - Applications for next generation silicon-based optoelectronics" *Jpn. J. Appl. Physics* , vol. 46, No 4B, pp 2474-2480, 2007.
[26] L. W. Snyman and M. du Plessis , "Increasing The emission intensity of p+np+ CMOS LED's (450 – 750 nm) by means of depletion layer profiling and reach-through techniques" ; in *Silicon Photonics III* , edited by Joel A Kubby, Graham T. Reed, Proceedings of SPIE Vol 6898 , (SPIE, Bellingham, W.A. 2008) , OE 1- 12, May 2008. (ISSN: 0277-786X, ISBN : 9780819470737)
[27] M.S. Tyagi, "Zener and avalanche breakdown in silicon alloyed p-n junctions", *Solid State Electronics* (11) , pp 577- 582. (1968).
[28] S.M. Sze , *Semiconductor Devices, Physics and Technology*, Wiley and Sons, 1985, Chapter 2.
[29] N.A. Drozdov, A.A. Patrin, V.D. Tkachev, *JEPTP Lett*. 23, 597 (1976).
[30] M. Reiche, K. Scheerschmidt, D. Conrad, R Scholz, A Plössl, U. Gösela, K.N. Tu, *Inst. Phys. Conf. Ser.* 157,447 (1997).
[31] M.A. Green, J Zhao, A. Wang, P.J Reece, M. Gal, *Nature* 412, 805, (2001).
[32] M.Kittler, M. Reiche, T. Mchedlidze, T. Arguirov, G. Jia, W. Sifert, S. Suckov, T Wilhelm, "Stark effect at dislocations in silcon for modulation of a 1.5 µm light emitter", in *Silicon Photonics III* , edited by Joel A Kubby, Graham T. Reed, Proceedings of SPIE Vol 6898 , (SPIE, Bellingham, W.A. 2008) , OG- 1- 12, May 2008. (ISSN: 0277-786X, ISBN : 9780819470737)
[33] Snyman et al, 2008, unpublished research.
[34] 0,35 µm BiCMOS design and processing procedures, *Austrian Mikro Systeme* , Graz, Austria.
[35] S.M. Sze , *Semiconductor Devices, Physics and Technology*, Wiley and Sons, 1985, Chapter 4.

Lukas W. Snyman (M'92-) received the B.Sc., B.Sc Hons., M.Sc. and PhD degrees, all in semiconductor physics, from the University of Port Elizabeth, South Africa in 1974, 1975, 1978 and 1987 respectively. During the period 1988 to 1999, he was with Carl and Emily Fuchs Institute for microelectronics, University of Pretoria. South Africa, as an Associate Professor. Since October of 2000, he has been a Professor at Tshwane (Technical) University of Technology (TUT). His current research interest includes optoelectronic devices, CMOS IC design, and micro-wave and high-frequency electronics. He has to date published about 70 scientific articles, mostly in international conference proceedings and scientific journals. He has given several guest lecturer on CMOS light emitting device design in the United States, Switzerland, Japan and Finland. He is the main author of two U.S. patents, co-author of ten South Africa patents on the topic of CMOS light emitting devices.



Prof. Snyman won the WIRSAM scientific awards in South Africa for the lectures at scientific conference in 1983 and 1985. In 1987, he received the South African microelectronics Industries (SAMES) award for best contribution in semiconductor physics and microelectronics in South Africa. In 2004, and 2008 respectively, Professor Snyman received an Academic Excellence Awards for Researcher of the year (runner up) in 2004 and also Innovator of the Year (runner up) in 2008, at the Tshwane University of Technology in South Africa.

Monuko du Plessis (M'88-SM'00) received the B, Eng., M.Eng. and D.Eng. Degree in electrical and electronic engineering from the University of Pretoria, Pretoria, South Africa, in 1972, 1978, and 1984, respectively. He obtained a B.A degree in Psychology, the B.Com. (Hons) degree from the University of South Africa, South Africa, in 1988, 1992 and 1998 respectively. Since 1973, he has been with the Department of Electrical, Electronics and Computer Engineering, University of Pretoria, where he is director of Carl and Emily Fuchs Institute for Microelectronics (CEFIM).



Professor Du Plessis specializes in CMOS digital and analog design, and in optoelectronic devices, mainly detector devices. He is currently a senior member of the IEEE as well as a distinguished researcher of the University of Pretoria. He has published 55 articles, mostly in international conference proceedings and international journals. He is also author and co-author of two U.S. patents, two PCT patents and ten South Africa patents on the topic of CMOS light emitting devices.

Enrico Belotti graduated in Politecnico di Milano, Italy in 1989 with laure a in technology and with Ph.D from the Georgia Institute of Technology in 1999. Since 200 he associate Professor at the Department of Electrical Engineering at the Boston University where he heads a group on advanced simulation of optoelectronic devices, mainly Ga N devices.

Professor Belotti has published about 55 articles in international conference proceedings and international journals , mostly on silicon and III-V optoelectronic devices. He has given about 8 invited talks to distinguished groups in the USA on the topic of GaN devices and device design and accumulated more than 1million $ grant money from US agencies. During 2003 he received the distinguished 2003 ONR Young Investigator Award in the USA and in 2005 the NSF CAREER Award in the USA.

Simulation of Si LED (450nm-750nm) light propagation phenomena in CMOS integrated circuitry for MOEMS applications

Kingsley A. Ogudo, Lukas W. Snyman (MIEE), Monuko du Plessis (SMIEE) and Gustave Udahemuka, Student Member of IEEE.

***Abstract*—This paper discusses the development and application of a dedicated software evaluation tool for evaluation of optical propagation mechanisms in pre-specified complementary metal oxide semiconductors (CMOS) integrated circuit structures in the wavelength range of 450nm -750nm. A MONTE CARLO simulation technique was developed in which the optical wave propagation phenomena as relevant in CMOS structures were continuously updated as the optical ray progresses through the structure. Refractive index of the material, layers thickness and structure curvatures were all incorporated as ray propagation parameters. By using a multi-ray simulation approach , the overall propagation phenomena wrt refraction, reflection, scattering, and intensities could be evaluated in globular context in any complex CMOS integrated circuit structure in a progressive way. MATLAB software was used as a mathematical capable and programmable language to develop the dedicated software evaluation tool. Subsequently, some correlations are made with regard to experimental results obtained. Some, conceptual, applications of MOEMS structures as can be implemented in Si CMOS integrated circuitry, are also discussed.**

***Index Terms*—Silicon, Light emitting device, Opto-electronic integrated circuits, CMOS technology, Si CMOS structure, Optical simulation, Monte Carlo, objects based modeling.**

Manuscript received March 10, 2009. This work was supported in part by the Tshwane University of Technology (TUT), the Carl Emily Fuchs Institute for Microelectronics(CEFIM) and the National Innovation Fund from the S.A . National Research Foundation.

K. Ogudo and L.W Snyman are from the Laboratory for Innovative Electronic Systems, Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa (Email: snymanlw@tut.ac.za).

M. du Plessis is from Carl and Emily Fuchs Institute for Microelectronics, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa (Email: mplessis@postino.up.ac.za).

Gustave Udahemuka is currently at French South Africa Technical Institute in Electronic (F'SATIE), Tshwane University of Technology, Pretoria, South Africa (Email: gustaves2000@yahoo.fr)

The technical content as contained in this article forms the subject of the following patents: USA Patent No. 6,111,271 of 29th August 2000. Other patents also relevant are USA Patent No. 5,944,720 of 30 November 1999 as well as of South African Patents 96/2528, 96/2478 and SA Provisional Patent Application 2009/00632.

I. Introduction

The need for optical signal transmission, has been spurred by ever-increasing information demand in computer and data processor networks, covering shorter distances. High speed, high bandwidth, noise interference immunity and electrical isolation are just some facets which suggest the superiority of optical data transmission and processing over more conventional electrical implementations. Optical communications and opto-electronic integrated circuits are currently mainly driven by so-called 'III-V' semiconductors, like Gallium Arsenide (GaAs) and associated compounds, and form the basis of a quite mature and inexpensive technology. These technologies are, however, very advanced and cannot be integrated in mainstream silicon technology. The realization of large scale opto-electronic integrated circuits and optical data "highway" in standard CMOS integrated circuitry utilizing only Si CMOS compounds has been envisioned and hold much promise [1-6].

Substantial development work has been done in the field of silicon pn-junctions emitting visible light when operated in reverse breakdown avalanche mode [7-13] . Improvements have recently been made with regard to developing Silicon
CMOS light emitting sources of much higher efficiencies by utilizing an Si CMOS avalanche light emitting diode technology [14 -24]. Further advances have recently been made with regard to both total emission powers as well as increased intensity levels by developing a so-called Injection Avalanche (SiInAv) technology [25, 26]. These devices emit optical radiation in the wavelength region of 450nm-750nm. They can also be integrated with ease in standard state-of-the-art CMOS integrated circuitry. Various practical prototype devices have recently been realized with respect to these technologies, offering low operating voltage (8V), and low current (80µA – 1mA) of operation with emitted emission intensities of 10 to 100nW in area as

small as 1x1 micron. The internal light emission levels are much higher.

The purpose of this paper is to (1) demonstrate some first iteration simulation results that have been achieved with regard to the nature of optical propagation phenomena in prototype 0.35 and 1.2 micron CMOS integrated structures the 450nm to 750nm wavelength regime; and , (2), to demonstrate some potential applications of these in MOEMS.

II. OPTICAL CHARACTERISTICS OF CMOS TECHNOLOGY

Fig. 1 shows the major composition and topography as associated with CMOS technology. Particularly noticeable is the many optically transparent components in the structure in the 450 to 750nm wavelength range such as the field oside layer, inter-metallic oxide layers and even the silicon nitride passivation layer. Further optically transparent layers such as silicon oxi-nitride layers and silicon dioxide layers can be added as overlayers [27] . The optical transparency of the
various components vary slightly. The detailed dimensions and optical characteristics as associated with 1.2 micron CMOS technology are described in Table 1 [27, 28, 29].

The characteristics for 0.35 micron CMOS technology are essentially the same, except that the silicon field oxide is thinner (250nm) and that a process of planarization is used after transistor fabrication in order to accommodate a higher number of metal over-layers [30]. Silicon 1.2 micron technology follows no such procedures and the subsequent over-layers follow curvatures and contours as a function of the lower lying base structures [31]. Whereas the transparency of silicon is very low, the transparency of the field oxide layer is excellent. The transparency of silicon nitride and silicon oxi-nitride is essentially transparent for radiation for radiation higher than 600nm [28, 29]. This is substantially lower than the absorption edge of silicon which lies at approximately 850-950 nm. If Si LEDs are used as optical sources, silicon CMOS technology therefore offer the scope for accommodating both optical sources and silicon pn detectors without the use of III-V or Si –Ge technology.

III. Si CMOS LEDS AND SI CMOS DETECTORS

Special avalanche based Si LED's have been designed by us in order to optimally couple emitted light into CMOS based opto-coupler with wave-guiding structures [25, 26]. Fig .2 shows a schematic layout of a simple n^+p well Si avalanche based CMOS LED that could feed light into an optically transparent field oxide layer which is lying adjacently to the LED. The light emission in this case occurred slightly below the apex of SiO_2 "field oxide" region Figure .3 (d), a structure which is commonly known as the "bird's peak" in Si 1.2 micron CMOS technology. The positioning of the light emitting region here ensured an optimum coupling of light into the 1 micron thick "field oxide"layer..

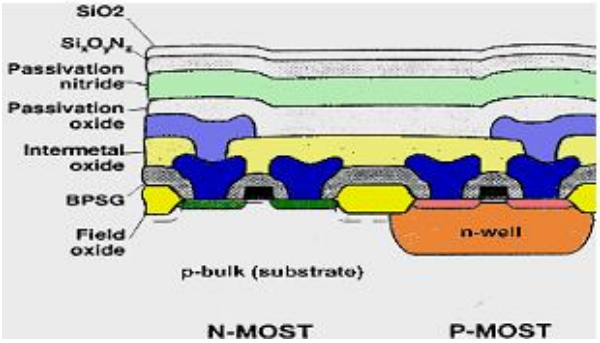


Figure 1: Schematic diagram showing the typical layers used in CMOS integrated circuit technology. The optically transparent layers are SiO_2 field oxide layer, silicon nitride (Si_3N_4) passivation layer and inter-metallic plasma deposited oxide layers . Additional layers can be deposited.

Fig. 3 shows a more detailed design in terms of top-down and plan layout. The device consists of an n^+ bar placed opposite two p^+ bodies placed diagonally away from the n^+ bar. All were embedded in a p-well of $2x10^{-16}$. The definition of the n^+ and p^+ regions also defines the peripheries of the SiO_2 field oxide in the CMOS process and these regions

TABLE 1: OPTICAL PROPERTIES OF 1.2µm CMOS INTERATED CIRCUIT TECHNOLOGY FOR 450-750 nm OPTICAL RADIATION

Component	Thickness (µm)	Refractive index	Assumed Absorption constants	State
Silicon substrate	1 µm	3.5	0.7 - 0.3	Crystalline solid
Silicon field oxide	1 µm	1.4	0.2 - 0.1	Fused solid
Silicon plasma oxide	0.9 µm	1.5	0.4 - 0.2	RF Deposited plasma
Silicon nitride	0.9µm	2.4	0.3 - 0.05	RF Deposited plasma
Air	3 µm	1	0.00	Gas phase

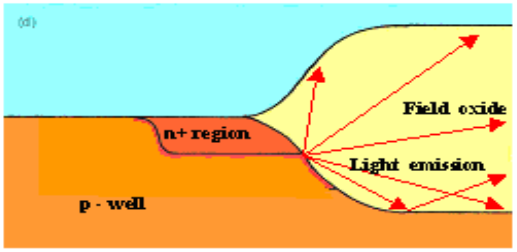


Figure. 2: Schematic diagram illustrating the design of an initial line source Si Light emitting diode device positioned at the "Bird's peak" in a Si CMOS structure and which coupled light mainly into the silicon field oxide layer in 1.2 micron CMOS technology.

could hence be used to define a centre SiO_2 field oxide region. Upon biasing the structure with +V on the n^+ and -V on the p^+ regions, a depletion region extends on the n^+ edge facing the p^+ regions SiO_2 field oxide region. The light emission occurred with a luminescence intensity of approximately 1nW per μm line length at operating voltage of about 15V and at currents from 80μA–1mA.The wavelength spectrum of the emitted light showed a broadband emission in the 450nm – 750nm region with a tail-off running deep into the infrared region (Fig. 3 (b)).

Si avalanche-based light emitting devices that can be integrated into CMOS integrated circuitry offer the following advantages:

1) Good optical signal detection

Silicon technology offers the option to fabricate small micro detector structures in the 450nm to 750nm range with good detection efficiency. The predicted leakage currents at room temperature for small 10x10 micron detector are of the order of pico Amp and the low frequency floor power detection levels are in the order of pW's [32, 33]. As pointed out above the current power emission levels as emitted
from Si CMOS InAv LEDs are of the order of 10 -100 nW's which is nearly three to four orders higher than the detectivity limits of these CMOS pn detectors.

2) High potential bandwidth speed of operation greater than 1 Gb/sec

The small size In Av CMOS LED structures offers high modulation speeds because the main component of the device operate in reverse bias mode of operation, and the small dimensions of the device offers low parasitic capacitances, enabling very high modulation speeds [5, 34]. The eventual speed of integrated Si LED optoelectronic components in CMOS integrated

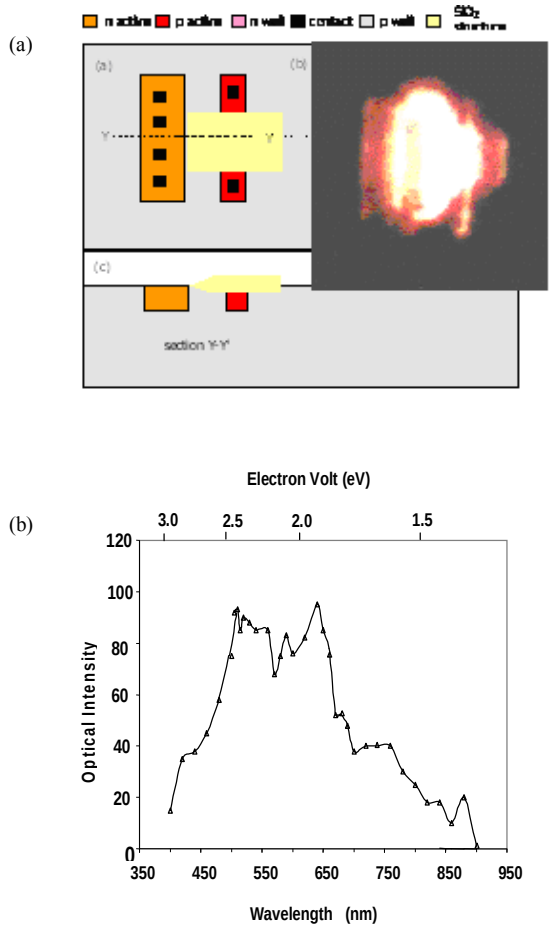


Figure. 3: (a) Schematic diagram showing detail of the design of a line source Si avalanche based LED which couple light laterally into the CMOS field oxide layer (b) Spectral detail of the emitted light as externally measured with a spectrometer.

circuitry is determined by the surrounding driving and signal processing circuitry. Substantial progresses have been made in this regard leading to state of the at GHz signal processing speeds [1, 35].

3) Good electrical isolation

Because of low leakage currents at room temperature, silicon CMOS technology offers very high electrical isolation between circuitry components, immunity to interference and good signal security.

4) System reliability

Si avalanche diodes operate at high reliability levels in silicon technology and have a proven track record in this regard.

Special detectors have also been designed in order to ensure optimum detection efficiency for laterally incident

radiation exiting at the end of the structures. Lately, 1-100nW optical powers could be obtained from our CMOS LED devices that typically operate in the 5-15 V and 0.1 to 1mA range. This suggests that the optical radiation as emitted by these devices could indeed be effectively utilised in opto-coupling and wave-guiding applications in later 0.35 µm silicon CMOS integrated circuitry.

Fig. 4 (a) and (b) demonstrates successful first iteration optical coupling between a custom-designed optical sources and distantly located, 100 micron away, Si CMOS optical detectors in standard state of the art, silicon 1.2 micron CMOS integrated circuit technology. Special Si LED optical sources and detectors were developed in order to optimise optical coupling into and from the opto-coupling structures. Signal levels ranging from 6nA to 1µA were detected at the detectors [36, 37].

IV. POTENTIAL APPLICATION OF SI LEDS AND CMOS INTEGRATED CIRCUITRY AS MOEMS

The above technology especially offer the possibility of realizing various MOEMS structures which can either be integrated in hybrid form or monolithic form at micro dimensions levels in standard CMOS technology. Si LEDs silicon detectors and associated electronic processing circuitry can be integrated at ease into standard CMOS integrated circuitry at mass production and sat extremely low cost. Appropriate optical interface modules can be integrated into the CMOS integrated circuitry in hybrid form utilizing RF recess technology. Alternatively if waveguide technology is realized, whole MOEM structures can be monolithically integrated into the CMOS integrated circuitry. MOEMS structures could range from simple vibration, motion and accelerometer structures, to more advanced chemical and physical parameter detection structures with applications into the emerging bio-technology sensor and nano-technology fields. Application of Si injection avalanche (SiInAv) CMOS LED's in miniature CMOS based MOEMS applications are therefore currently very a promising technology.

Current challenges with regard to first prototype generation with the above technologies are the following:

1) *Confinement of the emitted optical radiation*
Radiation patterns need to be accurately determined in terms of both direction and irradiation patterns. This is essential to eliminate excessive losses during emission and transmission and to avoid spill over into adjacent structures which will degrade optical isolation.

2) *Directional coupling into nearby structures*
The nature at which the emitted light from SiInAv LED's propagate through silicon CMOS structures, oxide , inter-metallic and passivation layer have not been accurately determined yet.

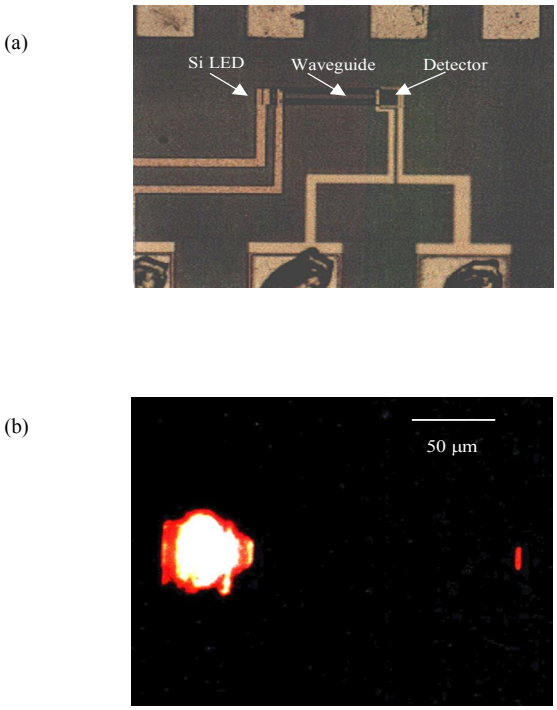


Figure. 4 Photomicrographs of a CMOS opto-coupler arrangement consisting of a CMOS Av-based light source, an optically wave-guiding arrangement and a CMOS lateral incident optimized CMOS based photo-detector.

3) *Efficient wave-guiding of optical radiation in CMOS structures.*
The establishment of efficient wave-guiding structures in CMOS integrated circuitry in the wavelength region 450nm–750nm will be a great benefit for CMOS optoelectronic circuit and CMOS MOEMS development. If such structures are established, it will enable optical signal propagation to adjacent structures further away, say up to 100 micron distances. This capability will offer various on board optical signal processing capabilities such as filtering, optical splitting and the generation of resonance cavity structures. A particular good candidate for such a propagation medium is silicon oxide-nitride and silicon nitride which are essential optically transparent above 500nm and 700nm, respectively [27, 28] .

V. GENERATION OF AN INTIAL CMOS STRUCTURE WITH MATLAB

A. Requirement for domain definition, function definition, matrix generation

MATLAB v7.1 was chosen as a mathematical capable programmable tool to develop the dedicated program algorithm and for the CMOS model design [38]. MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation and graphical transformed images. Typical uses include mathematical and computation algorithm development, data acquisition modeling, simulation, and prototyping data analysis, exploration, scientific and engineering graphics application development and graphical user interface building.

The most basic data structure in MATLAB is the matrix: a two-dimensional, rectangular shaped data structure capable of storing multiple elements of data in an easily accessible format. These data elements can be numbers, characters, and logical states of true or false, or even other MATLAB structure types. MATLAB uses these two-dimensional matrices to store single numbers and linear series of numbers as well. In these cases, the dimensions are 1-by-1 and 1-by-n respectively, where n is the length of the numeric series. MATLAB also supports data structures that have more or equal to two dimensions. These data structures are referred to as arrays in the MATLAB documentation.

B. Domain definition

The domain of the modeled silicon CMOS structure was defined by geometrical algebraic statements generated with MATLAB in the code which we call "*K-Dedicated Code A"*. Matrix geometrical algebraic statements were used to generate the graphical preliminary matrix elements look-up table for a typical CMOS structure. Two methods have been used to study the optical propagation in Si LED's integrated with CMOS structure:

Initially a non-versatile and fixed silicon CMOS structure was generated from "*K-Dedicated Code A"* preliminary domain functions is as showed in Fig. 5 below

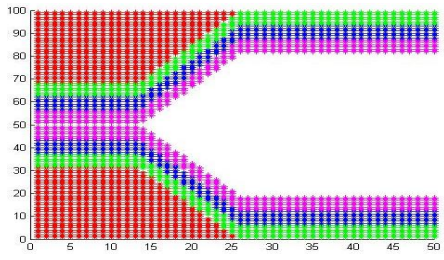


Figure. 5: First iteration graphical matrix generation .output.

```
==================================================
//Pseudo-code/block of flow diagram for domain definition//
maxX = 160;
sizeX = maxX*10;
maxY = 116.4;%Fixed
sizeY = maxY*10;

x = 0:sizeX;
y = 171-50 + (x-50).*(x>50) - (x-100).*(x>100) ;
z = 134-50 + (x-50).*(x>50) - (x-100).*(x>100) ;
q = 98-50 + (x-50).*(x>50) - (x-100).*(x>100) ;
s = 53-50 + (x-50).*(x>50) - (x-100).*(x>100) ;
yn = -y;
zn = -z ;
qn = -q ;
sn = -s ;
refractive_indices = [1 2.4 1.5 1.4 3.5 3.5 3.5];

for j = 1:sizeX %Column
   m = 1;
   for k = (-201+0.345:0.345:201-0.345)
        if k>y(j)
           Body1(j,m)= refractive_indices(1);
        else if (k > z(j))&&(k < y(j))
           Body1(j,m)= refractive_indices(2);
        else if (k > q(j))&&(k < z(j))
           Body1(j,m)= refractive_indices(3);
        else if (k > s(j))&&(k < q(j))
           Body1(j,m)= refractive_indices(4);


        else if (k < sn(j))&&(k > qn(j))
           Body1(j,m)= refractive_indices(6);
        elseif (k < qn(j))&&(k > zn(j))
           Body1(j,m)= refractive_indices(7);
        elseif (k < zn(j))&&(k > yn(j))
           Body1(j,m)= refractive_indices(8);
        elseif k<yn(j)
           Body1 (j,m)= refractive_indices (9);

       else
           Body1(j,m)= refractive_indices(1);
         end
          m = m + 1;
   end
end

for f = 1:rays_number
       af = a(f);
==================================================
```

and elucidated by the software example adjacently. This figure shows some first iteration matrix generation and graphical outputs for our CMOS structures. A four dimensional array was generated containing the matrix elements position [x, y] as first two elements followed by secondary elements in each [n, color] index information and color information for graphical presentation. Matrix generation was created by appropriate stepping algorithms generating different x and y values. The secondary elements generation was created by means of globular mathematical

equations which coupled different refractive index values and display color to different x and y values. The resolution of the first look-up tables were low in order to facilitate mastering of the basic techniques. Subsequently appropriate mathematical algorithms were used to generate different micro domains which contained different height, length, width and refractive index.

This analysis has been performed at a fixed 2D structure dimension with a specific scaling factor generated from the code. A total of 2500 matrix elements were generated for the first structure. This was eventually expanded to matrices containing 15000 elements which considerably increased the resolution capabilities of the structure.

The stimulated structure correspond to silicon LED's 450m-750nm light propagation phenomena in CMOS 0.35 micron integrated circuitry. The refractive indices for the various silicon materials were: 1.00 for air, 2.4 for silicon nitride, 1.5 for silicon plasma, and 1.4 for silicon oxide and 3.5 for silicon substrate (as shown in Table 1).

VI. DEVELOPMENT OF THE MONTE CARLO OPTICAL PROPAGATION STEPPING FUNCTION

A Monte Carlo optical ray propagation and stepping algorithm was developed as a subroutine within the *"K-dedicated code B"* as defined in MATLAB and applied to multiple rays each with a different initial launch angle. Snells and Fresnel laws were applied as basic physical laws governing reflection, refraction, and even for determining polarization characteristics [33, 39] :

$n_1 \sin\theta_1 = n_2 \sin\theta_2$ .....(1)

$\sin\theta_1 / \sin\theta_2 = n_2/n_1 = n$ ……. (2)

where n_1 = refractive Index of the first medium

n_2 = refractive index of the second medium.

n = ratio index

θ_1 = initial angle of incident ray

θ_2 =refractive angle of the ray

For the total internal reflection of the rays within the structure, the critical angle for internal reflection was used as

$\sin\theta_c = n_2/n_1$ .....(3)

where θ_c = critical angle

For detailed reflectance calculations, the following Fresnel equations were used [35]:

```
============================================================
=
//Fresnel formula for refraction and Snell's law of reflection//

 bb = angle_next_point(f);
 if
abs(BODY2(yy(step,f),xx(step,f))/BODY2(yy(step+1,f),xx(step+1,f))*sind(90
+ inc(f) - angle_next_point(f))) <= 1
            angle_next_point(f) = 90+inc(f) -
asind(BODY2(yy(step,f),xx(step,f))/BODY2(yy(step+1,f),xx(step+1,f))*sind(9
0 + inc(f) - angle_next_point(f)));
            signR = 1;
            FLAG = 0;
 else
 disp('REFLECTION')
if angle_next_point(f) - inc(f) > 90 & angle_next_point(f)>90
            angle_next_point(f) = - angle_next_point(f) + 2*inc(f);
  elseif
angle_next_point(f) - inc(f) < 90 & angle_next_point(f)>90
            angle_next_point(f) = angle_next_point(f) - 2*inc(f);
  elseif
angle_next_point(f) - inc(f) > 90 & angle_next_point(f)<90
            angle_next_point(f) = angle_next_point(f) - 2*inc(f);
 elseif
angle_next_point(f) - inc(f) < 90 & angle_next_point(f)<90
            angle_next_point (f) = -angle_next_point (f) + 2*inc(f);
 end            angle_next_point(f)

xx(step+1,f)=xx(step,f);            yy(step+1,f)=yy(step,f);
 a(f) = BODY2(yy(step,f),xx(step,f));
    FLAG = 1;
 End
============================================================
```

$$E'/E = \frac{\cos\theta - (n^2 - \sin^2\theta)^{1/2}}{\cos\theta + (n^2 - \sin^2\theta)^{1/2}} \quad \text{.....(4)}$$

for Transverse Electric (TE) polarization , and

$$E'/E = \frac{-n^2\cos\theta - (n^2 - \sin^2\theta)^{1/2}}{N^2\cos\theta + (n^2 - \sin^2\theta)^{1/2}} \quad \text{.....(5)}$$

for TM (Transverse Magnetic) polarization.

The concepts of Monte Carlo simulation [40] were incorporated in the stepping algorithm. This implies that the initial launch position and launch angle is defined within a specific defined matrix which contain information such as refractive index, thickness of layer and lateral dimension. The algorithm progressively performs a stepping function in which a new propagation position and new launch angle is defined based on the local matrix values. If a change of refractive index did occur, the incident angle to the interface was determined, Snell law for refraction and reflection was applied and new rays with new propagation positions and launch angles was defined within the larger matrix . Recursive formula by Hamersley and Handscomb [41]

was applied which ensures that the next optical ray launch angle in degree makes use of the previous one as initial parameters. This recursive formula plus Snell's law of reflection and refraction of optical ray within the medium was incorporated into the *"K-dedicated code B"* to simulate the ray propagation phenomena using refractive index parameters, layer dimensions, and reflective surfaces all contained in a four dimensional lookup table containing the elements and as defined in the previous section. The stepping program was developed as a subroutine within the *"K-dedicated code A"* and applied to multiple rays at different initial launch angles using Handscomb principles [41] .

$$x'_{x+1} = a_1 x'_1 + c_1 \quad (\text{modulo } n) \qquad (6)$$

Table 2 shows some extracted incident and refracted angles as were derived for some of the simulated rays and some of the code used in order to generate the Monte Carlo based stepping function is shown below.

The initial propagation phenomena for the defined silicon CMOS structure are shown in Figure 6. where optical ray propagation was launched as a point source at the apex of the Si field oxide . The light rays step from one medium to another medium of different refractive indexes, the incident and refractive angles values are calculated automatically and displayed at the MATLAB command window . This spread sheet values obtained from the simulation were analyzed and compared with experimental quantitative calculations.

==========

Creating complete structure

Begin

- fill each block with allocated refractive index
- Create matrix of inclinations
- Re-sampling for matrix distribution

end

==========

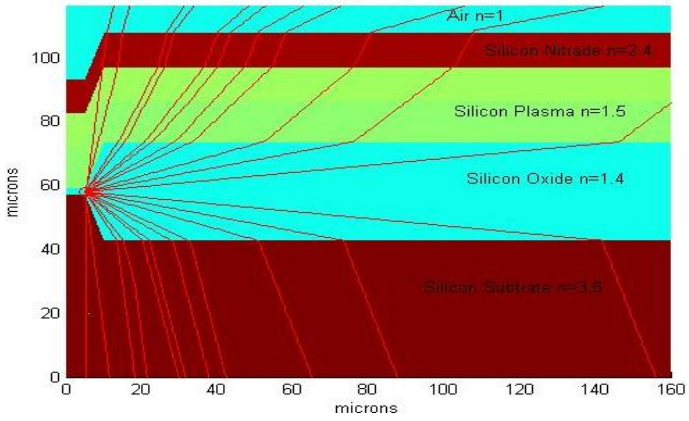


Figure 6: First iteration Monte Carlo Simulation of light scattering processes in a CMOS structure

A separate subroutine were also incorporated in order to compensate for any tilt angle of a new interface that may occur when a particular ray hits a new interface.

Table 2: illustration of results for multiple initial lance of incident rays according as generated by the stepping algorithm. .

CMOS structure	Silicon substrate	Silicon field oxide	Silicon plasma oxide	Silicon nitride	Air
Incident angles in degrees	45	50	55	60	70
Refracted angles in degrees	39	47	40	55	64
Reflected angles in degrees	61	43	60	45	36

VII. GENERATION OF A MORE VERSATILE CMOS STRUCTURE

The main limitation of the first matrix look-up table was that the structure generated was not very versatile and it could not generate a new and adaptable structures very quickly. This lead us to a further research to develop an algorithm that could model and simulate CMOS structures which had the following characteristics-

- Fast generation of the final matrix look up table
- Versatility with respect to the graphical user interface to quickly change the structure dimensions, composition and refractive index
- Ability to generate many new structures with different geometrical design fast and efficiently.

A. Graphical built-up of individual block structures:

The structure generation process was divided into three stages:

- Individual block generation
- Initializing and unifying the whole block body matrix

The geometric blocks used to build up the structure were limited to basic triangular and rectangular shapes, because CMOS structures can be generated and model by these shapes effectively. As for the triangles two sides (TSIDE_1 and TSIDE_2), one angle (TANGLE12) in between the sides and one vertex (TVERTEX1) was the parameters used to construct the various triangles. Likewise for the rectangles one, vertex (RVERTEX1) and two sides (RSIDE1 and RSIDE2) was used to construct the various rectangular shapes involved in the generation of the CMOS structure.

```
============================================================
//Pseudo code or flow diagram for structure shape generation//

RT =1;
figure(1)
axis equal
INCLINATION_CHECK = 45;%67.5;
ANGLE_CHECK = 90;%45;
ANGLE_CHECK_HALF = 45;%22.5;
VERTEX_MATRIX_X =  eros(4,1);
VERTEX_MATRIX_Y =  eros(4,1);
BLOCKSHAPE=[];
BLOCKNUMBER = 1;
BLOCKSHAPE(BLOCKNUMBER)=0; %0 for triangle

TVERTEX1 = [0;0];
TSIDE1 = RT;
TSIDE2 = RT;
TANGLE12 = ANGLE_CHECK;
TROTATION = INCLINATION_CHECK;
TCOLOR = 'y';
REFRACTIVE_INDEX(BLOCKNUMBER) = 1.4;
[a b]=trblock (TVERTEX1, TSIDE1, TSIDE2, TANGLE12, TROTATION, TCOLOR,
REFRACTIVE_INDEX);
VERTEX_MATRIX_X (:, BLOCKNUMBER) = [a;0];
VERTEX_MATRIX_Y(:,BLOCKNUMBER) = [b;0];
fill (a, b,TCOLOR)
hold on

BLOCKNUMBER = BLOCKNUMBER + 1;
BLOCKSHAPE (BLOCKNUMBER)=1;%1 for rectangle
[xs,ys]=getstart (a, b,2);
RVERTEX1 = [xs;ys];
RSIDE1 = RT*3;
RSIDE2 = RT*cosd(INCLINATION_CHECK)*2;
RROTATION = 90;
RCOLOR = 'y';
REFRACTIVE_INDEX(BLOCKNUMBER) = 1.4;
[c,d] = rectblock (RVERTEX1, RSIDE1, RSIDE2, RROTATION, RCOLOR,
REFRACTIVE_INDEX)
VERTEX_MATRIX_X (:, BLOCKNUMBER) = c;
VERTEX_MATRIX_Y (:, BLOCKNUMBER) = d;
fill(c, d, RCOLOR)
hold off
```

Fig. 7 (a) and (b) illustrate how a complex

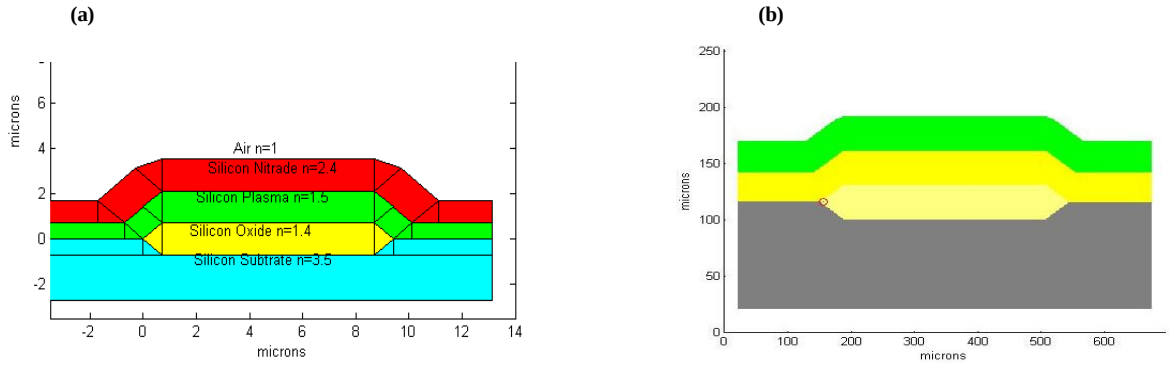


Figure 7. (a) Developing **g**eometrical building blocks for generating a CMOS test structure and (b) generation of a unified matrix that could be used for Monte Carlo Simulation (1st layer (from top) = silicon nitride; 2nd layer = inter-metallic plasma oxide ; 3rd layer =field oxide; substrate=silicon.

new CMOS structure was generated using a number of pre-defined sub block structures, each defined with particular dimensions, refractive index and color information. A sample of the line code from *"K-dedicated code B"* for this purpose for generating the dynamic shapes is also outlined below:

VIII. SOME OPTICAL SIMULATION RESULTS FOR 0.35 AND 1.2 MICRON CMOS INTEGRATED CIRCUITRY.

The developed optical simulation tool was subsequently applied for testing various propagation phenomenon in 0.35 and 1.2 micron CMOS silicon integrated circuitry. Refractive indexes and layer thicknesses and other parameters as in Table 1 were implemented.

Fig. 8 show a first iteration analysis as observed when the optical source was placed at the bottom of the first Si - SiO_2 interface and a certain amount of "bending" of the overlaying plasma oxide and passivation nitride was assumed, due to the absence of any metal contact layers in the structure. These are common phenomena in 1.2 micron CMOS integrated circuit technology. Some pertinent observations were the following:

1. The algorithm seems to work very efficiently and clear propagation paths were portrayed for each ray. In simulation mode, all rays progressed in parallel in real time such that progressive reflection and refraction phenomena could be clearly followed.

2. The propagation analysis using the multiple -ray approach makes the overall optical performance of the structure very powerful - Definitely much more powerful than any single ray analysis and hand computer analysis approaches .

3. For initial simulations, only primary refection and refraction mechanisms were considered and secondary reflection and refraction rays when intensities was lower than 25 % were avoided.

4. The propagation analyses show clearly that light penetrates quite well vertically upwards in the structure through the field oxide, the plasma deposited as well as the passivation silicon nitride layer.

5. The silicon nitride layer, being the highest refractive index layer of all the over-layers, has a dominating effect on bending of the rays . A very high dispersion of the vertically emitted rays is observed towards the right hand side of the structure in Fig. 7 . This effect will be especially be more pronounced with the shorter wavelength radiation experience much more bending and dispersion. When bending or curvature of the silicon nitride layer (towards the left hand side of the structure in Fig. 8), no such dispersion effects are present and the rays are emitted from the structure into air in quite undisturbed manner.

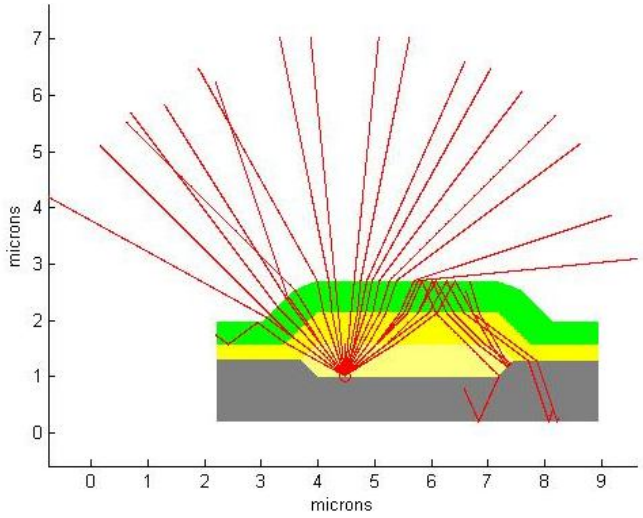


Figure 8. First iteration observed optical propagation phenomena as observed in a scaled unified CMOS Si LED test structure

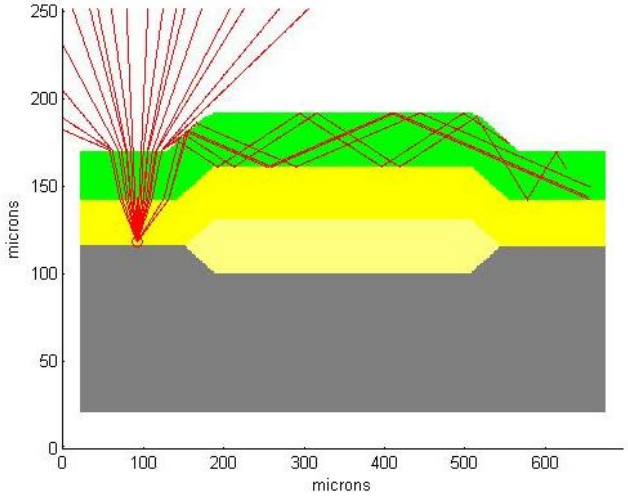


Figure 9. Effect of changing of optical emission point source in a Si LED CMOS structure.

6. When the angle of incidence into the silicon nitride layer gets bigger, some internal reflection occurs at the top silicon nitride interface leading to primary backwards refection of rays which are eventually scattered back into the silicon substrate.

7. Some interesting internal reflection phenomena occur in the silicon nitride layer (left hand side of the structure in Fig. 8) with rays apparently waveguided laterally along the silicon nitride layer. . The silicon nitride layer is for all practical purposes transparent for wavelengths higher than 600nm since it has a wide free spectral range, ranging from 600nm to over 1500nm [31]. The effects of absorption as function of wavelength of the radiation on the lateral propagation has to be further investigated

Fig. 9, in globular context, shows the effects observed when the main optical emission point (or the optical source is shifted to the left along the silicon - plasma oxide interface. A much more pronounced vertical emission of the radiation is observed with high, but even, dispersion occurring to both sides . A much more pronounced coupling of a certain percentage of the radiation into the silicon nitride layer is also observed leading to a much more pronounced "waveguiding" of the optical rays along the silicon nitride layer.

Fig. 10 shows the effects observed when the thicknesses of some of the field oxide and plasma deposited oxide layers were reduced as typically occurring in 0.35 micron CMOS technology. In general, a better coupling of radiation into the silicon nitride layer seems to occur.

IX. DEVELOPMENT OF DEDICATED STRUCTURES FOR CMOS MOEMS APPLICATIONS

Having developed the optical propagation evaluation tool, we considered various structures by arbitrarily changing lateral and even vertical dimensions wrt thickness layers in order to generate optimized applications wrt optical emission and propagation phenomena for MOEMS applications. A few achievements are presented here. Because of the versatile nature of both the structure building tool, as well as the propagation evaluation tool, many more investigations and developments can be performed.

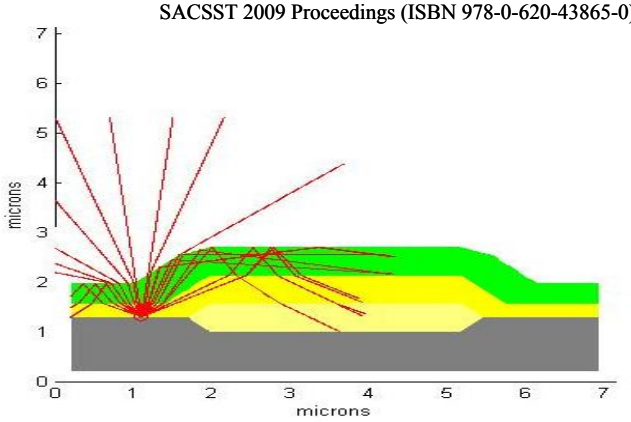


Figure 10. Change of layer thicknesses and its effect on the propagation phenomena in a Si LED CMOS structure.

A. Optimized vertical outwards optical emission from A SiLED CMOS structures

In Fig. 11, the lateral length of the structure was reduced and the position of the optical source was placed at an optimized point at the silicon substrate –field oxide interface in the centre of the structure. The optical launch angles for the rays were chosen in the range, 30^0- 150^0. The design can be implemented to include a circular nature of the centre field oxide region.

The result is a clear semi-isotropic radiation pattern with almost 120 degrees of solid angle emission to the chip external environment. This lead to a total (external) optical emission factor approximately 0.40 of all the optical radiation as generated at the Si-SiO_2

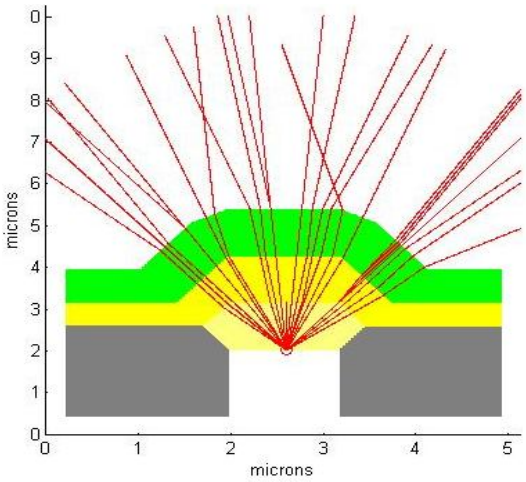


Figure 11. Dedicated structure for maximizing vertical optical emission from a CMOS Si LED structure.

interface source. This is up to twenty times higher than has been observed for any of our previous designs where emission factors of 0.02 to 0.04 were recorded. Some minor scattering is observed due to surface bending irregularities at the silicon nitride air interface and some absorption loss will occur in the silicon nitride layer for shorter wavelength radiation.

If our latest InSiAva technology is implemented [25, 26] , very high intensity values could be achieved per smaller area of optical radiation . This will substantially further increase the total optical emission levels as emitted from the structure.

B. Optical focusing and directional emission of optical radiation in CMOS Si LED structures.

Fig. 12 demonstrates the same basic spherical silicon nitride structure as has been used before, but the source emitter positions have been shifted. When the source is placed at position A, a clear focusing of almost all rays emitting vertically from the structure is observed. When the source is placed at position B, a clear directional emission of the radiation towards a slanted 45 degree angle is observed, also with some focusing or converging of optical rays present.

C. Optical splitting of optical radiation in CMOS structures.

Fig. 13 demonstrates a generated structure that can be used to create optical splitting of the optically radiated power in more or less equal percentages in two different paths. In this case a 0.3 micron field oxide layer and 1micron silicon passivation and plasma deposited layers were assumed as commonly encountered in 0.35 micron CMOS technology. One portion of the radiation emitted out of the structure into air, and the other radiation was emitted laterally into the silicon nitride layer. Optical splitters of this kind has wide and diverse applications in MOEMS applications and can be used to generate interference and phase contrast in certain applications.

D.Optical wave-guiding longitudinally along a Si LED CMOS structure

Fig. 14 demonstrates optimized lateral wave-guiding longitudinally along the silicon nitride layer for a very high proportion of the totally emitted Si LED radiation. The plasma oxide layer thickness has been reduced locally and the optical emission point in the generated structure was positioned near the "bird peak" point in the silicon

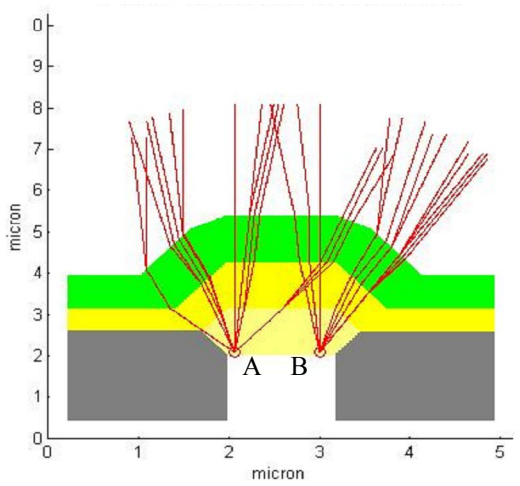


Figure 12. Dedicated structure for facilitating directional focusing from predetermined sources.

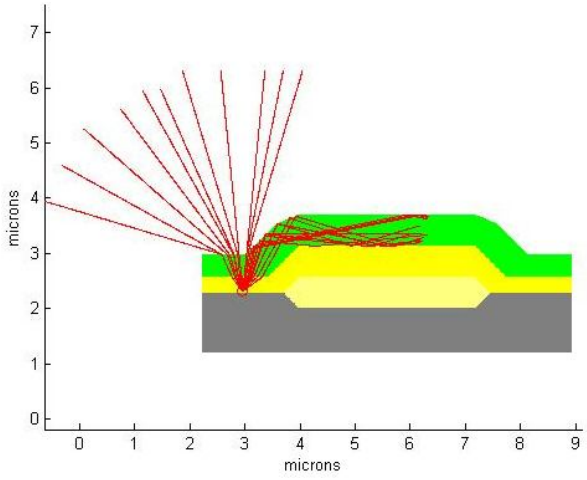


Figure 13. Dedicated structure for optical splitting of radiation from a CMOS Si LED structure

oxide layer. Analysis show that initial rays launch angles ranging from 34^0-76^0 couple effectively into the silicon nitride layer. Due to the lower refractive indexes of both the plasma and field oxide below the layer and the refractive index of air above the layer, initial indications are that the radiation can be quite effectively guided along the silicon nitride layer. Appropriate phase and mode contrast analysis still has to be performed. The silicon nitride is effectively transparent for the longer wavelengths above 600nm . Special layers utilizing silicon oxi-nitride (SiO_xN_x) compositions offer transparency at lower wavelengths [

27]. In both cases the waveguided or transmitted radiation are still much lower than the absorption edge wavelength for Si detectors (approximately 950 nm) . The combination of Si LED light source, optical waveguiding or optical transmission, together with good Si detectors that can be incorporated into the silicon CMOS structure makes diverse optical processing and MOEMS applications and realizations a viable option. The current total optical emission levels as can be extracted from our Si CMOS LEDs are of the order of 1 - 100 nW in micron area dimensions (with compatible CMOS operating voltages and currents). These emission levels are much higher than the low frequency detectivity level for Si pn detectors of comparable dimensions (with typical floor levels of pW) . This leave nearly a three order range in the power link budget, which can accommodate various losses and spiltting of optical radiation. The power link budget even allows for various cavity resonant structures to be fabricated in silicon CMOS structures that may even allow lasing or improved phase contrast characteristics of the emitted radiation.

Our current understanding of the light generation processes in our Si CMOS Av LED and InAv Si CMOS LED technology even allows for tuning of the emitted radiation towards certain dedicated wavelength ranges, which offers further diverse application possibilities.

.

X.CORRELATION WITH EXPERIMENTAL RESULTS

Fig. 15 (a) and (b) shows an optical micrograph of a section of CMOS structure as fabricated with 0.35 micron CMOS Technology which closely resemble the cross-section structure as shown in Fig. 7. The tip of the light emission pattern is from a sharp tip of penetrating n^+p junction which is surrounded by field oxide layers. Emission occurs mainly at the n^+p junction interface adjacently to the field oxide layers. Very interesting features as apparent from the experimental emission pattern are the following:

Towards the n^+p silicon substrate side, a very bright emission characteristics is observed with even a slight blue-ish tint. This observation correlates with the vertical up bending of some of the rays as observed towards the left hand side of the CMOS structure in Fig. 7 and also with the vertical upward focusing effect of emitted radiation as observed in the structure of Fig. 8.

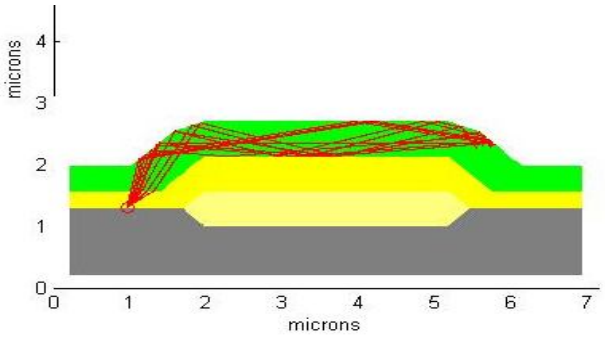
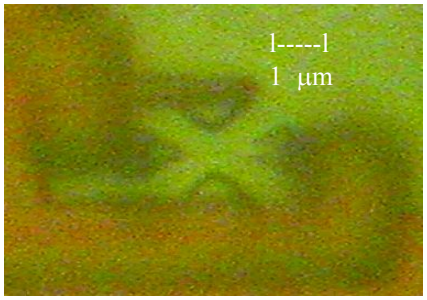


Figure 14. Dedicated structure for optimized lateral wave-guiding in a Si LED CMOS structure

a



b

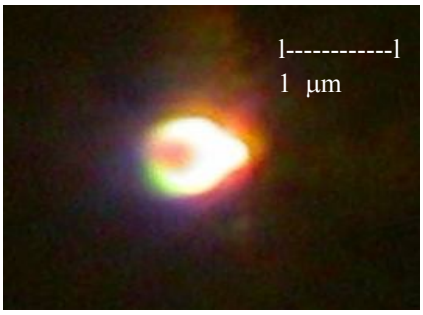


Figure 15. Experimental observations from a n^+p oxide structure confirming some of the simulation predictions as derived from Fig . 11. (a) represents a bright field optical micrograph of the structure as taken through an optical microscope, while (b) is the corresponding optical emission pattern showing definite blue-ish and red-dish dispersion tints in the optical emission pattern.

Towards the regions of surrounding field oxide as portrayed towards the right hand of structure in Fig. 11, a definite reddish tint is observed in the emission light pattern. This behavior surely correlates with the derivation that excessive dispersion (downwards bending) of the emitted rays occur , especially for the shorter wavelengths. More red light were refracted into the acceptance angle of the objective of the microscope, while the blue component of the light were scattered , dispersed at much higher angles towards the surface of the structure and were not taken up into the objective of the microscope and did not participate in the image formation process on the CCD of the microscope camera.

The above two experimental observations greatly support the validity of some of our simulation results and the derivations made.

XI. CONCLUSIONS AND FINAL REMARKS

The following can be regarded as main summarized conclusions with regard to this work:

1. A dedicated software evaluation tool was developed for evaluation of optical propagation phenomena in pre-specified complementary metal oxide semiconductors (CMOS) silicon integrated structures. Si LEDs using the avalanche light emission and the avalanche –injection (InSiAva) technology were incorporated in the structures and was used a light sources in the wavelength range of 450nm -750nm. Refractive index of the material, layers thickness and structure curvatures were all incorporated as ray propagation parameters. A MONTE CARLO propagation tool was incorporated in which optical wave propagation phenomena was continuously updated in terms of refraction and reflection principles as the optical ray progresses through the structure.. Using a multiple ray approach, the overall propagation phenomena could be followed through any complex CMOS integrated circuit structure in a progressive and globular context.

2. The simulation results supports the development of tailor - designed structures that can increase the vertical emission of light out of the structures (from factors of 0.02 to 0.21); the focusing of emitted emission rays in parallel format from the structures ; the splitting of optical radiation from a single source into two propagation paths; as well as dedicated transmission and waveguiding of the light along the silicon nitride passivation layers.

3. All of the above has very high potential for developing new optoelectronic device development utilizing Si LED CMOS structures. Particularly it offers the development of a wide variety of new generation Micro Optical Mechanical Sensors (MOEMS) sensors. These sensors will each incorporate a Si LED Av or Si InAva LED , a small optical module (either in hybrid or monolithic form) , a small mechanical module as well as small silicon detector module. Diverse designs are possible and a wide variety of such sensors are possible. The competitive advantage as offered by this technology in generating such devices is that the all components including the optical source, the optical paths, the mechanical module as well as the detectors can be incorporated at micro-dimension level into standard silicon CMOS integrated circuit technology. This could enable the development of a whole new range of smart optical –mechanical based CMOS sensor chips as well as a whole new range of supporting smart optical processing and optical digital and analog based interfacing CMOS integrated circuitry chips. Optical intensity, polarization, propagation modes as well as the development of more efficient waveguiding structures will all be the topics of future research using possible upgraded versions of our dedicated software, or in conjunction with existing other software.

4. In this article the utilization of existing CMOS structures and technology were mainly exploited, However, for the purpose of academic investigations in optimizing device performance, certain dimensions such as vertical layer thicknesses, were also varied in some cases , and some of the Si nitride layer curvatures may also have been violated. However it is our believe that all modern CMOS processes may be able to accommodate such violations , by employing slight modifications of RF etching procedures, local field oxidation procedures as well changing some of the plasma oxide planarization procedures on a localized basis during the CMOS fabrication process.

ACKNOWLEDGEMENTS

This work has been made possible with the financial support of Tshwane University of Technology (TUT) and the Laboratory for Innovative Systems in Electronics of the Faculty of Engineering and the Built environment Electrical department at TUT. The assistance of Mr Kingsley Ogudo, post graduate student at the Department of Electrical in Tshwane University of Technology in Pretoria, South Africa. The collaboration of Prof Monuko du Plessis, of the Carl and Emily Fuchs Institute for Microelectronics at the University of Pretoria is gratefully acknowledged. The assistance of Mr. Gustave Udahemuka with respect to MATLAB programming is great fully acknowledged.

REFERENCES

[1] C. Zhao, X. Zhang, H. Lui, and D. Huang, "Tunable all -optical NOR Gate at 10Gb/s based on SOA fiber ring laser", *Optics Express* 13, 2793-1798 (2005).
[2] D.A. Miller, "Silicon integrated circuits shine", *Nature* 384, 307 (1996).
[3] E.A. Fitzgerald, L.C. Kimberling, "Silicon-Based Technology for Integrated Optoelectronics", *Materials Research Society (MRS) Bulletin* 23, No.4, 39-47, (1998).
[4] D.J. Robbins, "Silicon Opto-electronics" ,*Proc. SPIE* 3953, i5 (2000).
[5] L.W. Snyman , H.Aharoni, A Biber, A Bogalecki, L.Canning, M.Du Plessis and P.A. Maree. "Optical souces, integrated optical detectors and optical waveguides in standard CMOS integrated circuitry" , *Proc. SPIE* 3953, 20-35 (2000).

[6] K. Wada, "Electronics and Photonics convergence on Silicon CMOS Platforms" *Proc. SPIE* 5357, 16 (2004).
[7] W.G. Ghynoweth, K.G. McKay, "Visible light emission from a silicon p-n junction", *Phys. Rev.* 102, 369-376 (1956).
[8] R. Newman, "Visible light emission from a silicon p-n junction" *Phys. Rev*. 100, 700 - 703 (1955).
[9] J.Bude, N Sano, A Yoshii, "Hot carrier Luminescence in Silicon" *Physical Review B*, 45 (11), 5848 - 5856 (1992).
[10] J. Kramer, P. Seitz, E. F. Steigmeier, H. Auderset, B. Delley, "Light emitting diodes in industrial CMOS technology", *Sens. and Act*. A37, 527-533 (1993).
[11] L. Carbone, R. Brunetti, C. Jacoboni, A. Lacaita, M. Fischetti, "Polarization analysis of hot-carrier emission in silicon, " *Semicond.Sci. Technol.* 9, 647-676 (1994).
[12] Lacaita, F. Zappa, S. Bigliasrdi, M. Manfredi, "On the Brehmstrahlung origin of hot-carrier-induced photons in silicon devices", *IEEE Trans. Electron Devices* ED40, 577-582 (1993).
[13] N. Akil, S. E. Kerns, D. V. Kerns, Jr., A. Hoffmann, J-P Charles, "A multi-dimensional model for photon generation in silicon junctions in avalanche breakdown", *IEEE Trans. Electron Devices* 46 (5), 1022-1027 (1999).
[14] L.W. Snyman, M.du Plessis, E. Seevinck, and H. Aharoni, "An efficient , low voltage, high frequency silicon CMOS light emitting device and electro-optical interface" , *IEEE Electron Device Letters* (U.S.A.) 20, 614-617 (1999).
[15] M. du Plessis, H. Aharoni and L. W. Snyman , " Silicon Light Emitting Devices in Standard CMOS Technology" , *Proc. CAS 2001*, 231- 238 (2001).(Invited paper).
[16] L.W. Snyman, H. Aharoni, M. du Plessis, R.B.J. Gouws, "Increased efficiency of silicon light emitting diodes in a standard 1.2 micron complementary metal oxide semiconductor technology" , *Optical Engineering (U.S.A.)* 37, 2133 - 2141(1998).
[17] L.W. Snyman, H Aharoni, M. du Plessis, J.F.K Marais, D van Niekerk and A Biber," Planar light emitting electro-optical interfaces in standard silicon complementary metal oxide semiconductor integrated circuitry", *Optical Engineering (SPIE-USA)* 41, 3230 - 3240 (2002).
[18] M. du Plessis, H. Aharoni and L.W. Snyman: "Two- and multi-terminal CMOS/BiCMOS Si LED's." *Optical Materials* 27, 1059-1063 (2005).
[19] L. W. Snyman, H. Aharoni and M du Plessis, "Two order increase in the quantum efficiency of silicon CMOS n+pn avalanche-based light emitting devices as a function of current density" IEEE *Photon. Techn. Lett*. 17, 2041-2043 (2005).
[20] M. du Plessis, H. Aharoni and L.W. Snyman: "Spatial and intensity modulation of light emission from a silicon LED matrix.", *IEEE Photon. Techn. Lett.*14, Issue 6, 768-770 (2002).
[21] L.W. Snyman, M. du Plessis and H. Aharoni, "Efficiency analysis of Avalanche Si LEDs", *Proc. IEEE ISIE 3*, 1159 - 1169(2005). (Special Session Paper).
[22] L . W. Snyman, M. du Plessis, H . Aharoni, "Three terminal optical sources (450nm - 750nm) for next-generation silicon CMOS OEIC's" ,*MIXDES' 5*,Europe , 737-747 (2005) (Invited Plenary Session Paper).
[23] L. W. Snyman, M. du Plessis, and H. Aharoni, " Two order increase in the optical emission intensity of CMOS integrated circuit LED's (450 - 750 nm) Comparison of n+pn and p+np designs " , *Proc. SPIE* 5730, 59-72 (2006).
[24] L. W. Snyman, M. Du Plessis and H. Aharoni, "Injection-avalanche based n+pn Si CMOS LED's (450nm - 750nm) with two order increase in light emission intensity - Applications for next generation silicon-based optoelectronics" *Jpn. J. Appl. Physics* 46, 2474-2480 (2007).
[25] L. W. Snyman and M. du Plessis , "Increasing The emission intensity of p+np+ CMOS LED's (450 - 750 nm) by means of depletion layer profiling and reach-through techniques", *Proc. SPIE* , Vol 6898, OE 1-12 (2008) (Invited paper).
[26] Lukas W. Snyman, Monuko du Plessis and Enrico Bellotti, "Photonic transitions (1.4 eV- 2.8 eV) in Silicon p+np+ injection-avalanche CMOS LEDs as function of depletion layer profiling and defect engineering", *Optical Engineering*, Submitted (2008).
[27] E. Fullin, G. Voirin, A. Lagos, J.M. Moret, "MOS-Based Technology for Integrated Opto-electronic Circuits", *Scientific and Technical Report CSEM* 26, ,(1993).
[28] M.Beals, J . Micheal, J.F. Liu, D.H. Ahn, D Sparacin, R. Sun, C.Y. Hong and L.C. Kimerling. et al , " Process flow innovations for photonic device integration in CMOS", *Proc. SPIE,* Vol. 6898, 689804 (2008).
[29] L.K. Rowe, M. Elsey, E, Post, N. Tarr, A.P. Knights , " A CMOS compatible rib waveguide with local oxidation of silicon isolation", *Proc. SPIE* 6477, 64770L, (2007).
[30] Austrian Mikro Systeme, [0,35 µm BiCMOS design and processing procedures], Graz, Austria, (2002).
[31] ORBIT Semiconductor Inc, [1.2 micron design rules and process parameters],, USA, (2002).
[32] S.M. Sze , "*Semiconductor Devices:Physics and Technology*", Wiley and Sons, (1985).
[33] J. M. Senior, "*Optical Fiber Communications: Principles and Practice*", Prentice Hall, USA, Chapter4, (1985).
[34] A. Chatterjee, B. Bhuva and R. Schrimpf, "High-speed light modulation in avalanche breakdown mode for Si diodes", *IEEE Electron. Device Lett.* 25 (9), 628-630 (2004).
[35] H. Zimmermann,"Improved CMOS-integrated Photodiodes and their Application in OEICs",*IEEE EDMO*, 346-351 (1997).
[36] L.W. Snyman , H.Aharoni, A Biber, A Bogalecki, L.Canning, M.Du Plessis and P.A. Maree. "Optical souces, integrated optical detectors and optical waveguides in standard CMOS integrated circuitry" , *Proc. SPIE, Vol* 3953, 20 - 35 (2000).
[37] L.W. Snyman, L M Canning, A Bogaleki, A Aharoni, M. Du Plessis, "200-Mbps optical integrated circuit design and first iteration realizations in 0.8 micron bi-CMOS silicon integrated circuitry", *Proc SPIE*, Vol 5357-6, (2004).
[38] Math Works, "MATLAB software version 7.1", *Complete documentation and simulation tool box,* (2005).
[39] F.L. Pedrotti, "*Introduction to Optics*", Prentice Hall, Europe, 407-415 (1993).
[40] C. Moglestue, "*Monte Carlo Simulation of Semiconductor Devices*" , Springer, USA, 336 (1993).
[41] Hammersley and Handscomb, "*Monte Carlo Methods*", Methuen, USA, (1964).

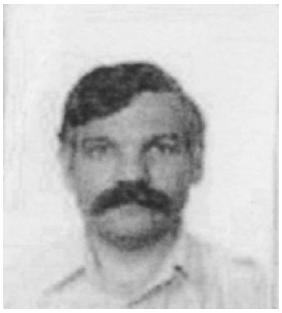
K.A.Ogudo- received the N.Dip, NHD and B. Tech degree, all in electrical and electronics engineering from Federal Polytechnic, Nigeria and Tshwane University of Technology (TUT), South Africa in 1998, 2000 and 2005 respectively. During the period 2003 to 2005 he was with French South African Technical institute in Electronics (F'SATIE) completing his B. Tech courses and styding in opto-electronics and telecommunications technology. He currently lectures as a part-time lecturer in Academy of Business and Computer Studies, Pretoria in as well as research assistant at the Department of Electrical Engineering Department at Tshwane University of Technology (TUT), South Africa. He has also offer students assistantship in various Engineering N.Dip and B-Tech subjects. His research interest includes electronic, optoelectronic devices, telecommunication engineering high-frequency electronics, physics and mathematics. He is currently at completion of his M.Tech studies on "Simulation of Si LED's (450nm-750nm) Light propagation phenomena in CMOS 0.35 micron integrated circuitry-MOEMS applications" at the Department of Electrical Engineering at the Tshwane University of Technology, under the supervision of Professors Snyman and Du Plessis.

Lukas W. Snyman (M'92-) received the B.Sc., B.Sc Hons., M.Sc. and PhD degrees, all in semiconductor physics, from the University of Port Elizabeth, South Africa in 1974, 1975, 1978 and 1987 respectively. During the period 1988 to 1999, he was with Carl and Emily Fuchs Institute for microelectronics, University of Pretoria. South Africa, as an Associate Professor. Since October of 2000, he has been a Professor at Tshwane University of Technology (TUT). The Tshwane University of Technology is currently the largest technical university in

Africa that specializes in the application of technology. His research interest includes optoelectronic devices, CMOS IC design, and micro-wave and high-frequency electronics.

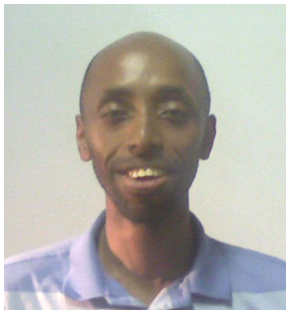
He has to date published about 70 scientific articles, mostly in international conference proceedings and scientific journals. He has given several guest lecturer on CMOS light emitting device design in the United States, Switzerland, Japan and Finland. He is the main author of two U.S. patents, co-author of two PCT patents and ten South Africa patents on the topic of CMOS light emitting devices.

Prof. Snyman won the WIRSAM scientific awards in South Africa for the lectures at scientific conference in 1983 and 1985. In 1987, he received the South African microelectronics Industries (SAMES) award for best contribution in semiconductor physics and microelectronics in South Africa. In 2004, and 2008 respectively, Professor Snyman received an Academic Excellence Awards for Researcher of the year (runner up) in 2004 and also Innovator of the Year (runner up) in 2008, at the Tshwane University of Technology in South Africa.



Monuko du Plessis (M'88-SM'00) received the B, Eng., M.Eng. and D.Eng. Degree in electrical and electronic engineering from the University of Pretoria, Pretoria, South Africa, in 1972, 1978, and 1984, respectively. He obtained a B.A degree in Psychology, the B.Com. (Hons) degree from the University of South Africa, South Africa, in 1988, 1992 and 1998 respectively. Since 1973, he has been with the Department of Electrical, Electronics and Computer Engineering, University of Pretoria, where he is director of Carl and Emily Fuchs Institute for Microelectronics (CEFIM).

Professor Du Plessis specializes in CMOS digital and analog design, and in optoelectronic devices, mainly detector devices. He is currently a senior member of the IEEE as well as a distinguished researcher of the University of Pretoria. He has published 55 articles, mostly in international conference proceedings and international journals. He is also author and co-author of two U.S. patents, two PCT patents and ten South Africa patents on the topic of CMOS light emitting devices.



Gustave Udahemuka (Student member IEEE 06 -) -received the B-Tech in Telecommunication Technology from Tshwane University of Technology in 2004, M-Tech and M-Sc (with distinctions) in Electronic Engineering from Tshwane University of Technology and Ecole Superieure d' Ingenieur en Electronique et Electrotechnique (ESIEE) in 2006 respectively. He is currently a PhD candidate at Tshwane University of Technology. His research interest includes satellite remote sensing, software development and optoelectronics devices.

Hall and Thermoelectric Evaluation of Narrow Gap Semiconductors

Magnus C. Wagener, Viera Wagener, and J. R. Botha

Abstract— **This paper will report on some of the methods used to extract the electrical characteristics of narrow gap semiconductors, as well as demonstrate strategies for characterizing materials with more complex conduction mechanisms. In particular, the influence of the surface accumulation layer characteristic to InAs is discussed and incorporated into the analysis of the temperature dependent Hall and thermoelectric measurements.**

Index Terms—**Thermoelectricity, Hall Effect, Semiconductor materials, Doping**

I. Introduction

HALL effect measurements have been the preferred method for determining the conduction characteristics of semiconductor materials since its first conception. However, difficulties associated with more complex conduction systems, such as contributions from additional surface and/or interface conduction paths have necessitated more complex analysis to extract the basic doping and conduction properties. The additional variables associated with these degenerate conduction pathways also require more detailed Hall measurements, considering the dependence on temperature, layer thickness, magnetic field strength, *etc.* to separate the various contributions to the measured Hall voltage [1, 2]. The success of these methods is further hampered for materials with low mobilities and high intrinsic carrier densities. This has been apparent for studies of p-type InAs, where a surface accumulation layer [3] completely conceals the bulk characteristics. This limitation of Hall measurements results from its sensitivity to degenerate conduction pathways. As a result, Hall effect measurements typically rely on results obtained for highly doped material in order to extrapolate the doping characteristics of lightly doped material [4]. We will present a detailed analysis of Hall measurements performed on lightly doped InAs, and then introduce an alternative approach where the thermoelectric properties of small bandgap materials are used to extract the doping characteristics.

Manuscript received March 2, 2009. This work is based upon research supported by the SA Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation. The financial support of NMMU and the PRISM-M (DPSS CSIR) project is gratefully acknowledged. Authors are with the Department of Physics, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa (corresponding author phone: +27-041-4573; fax: +27-041-4579; e-mail: Magnus.Wagener@nmmu.ac.za).

II. Theory

A. Hall Effect

Mobile carriers within a conductor placed in a magnetic field will experience a Lorentz force proportional to the magnitude of the electric and magnetic fields applied. The resulting accumulation of charge within one side of the conductor produces and electric field that eventually counteracts any further charge accumulation. Since the strength of the induced electric field is related to the charge distribution along the conduction path, the associated potential difference, referred to as the Hall voltage, V_H, can be used to determine the free carrier density within the conductor. In order to present the Hall measurements in a form independent of the drift current and magnetic field used, the Hall coefficient R_H is defined as

$$R_H = \frac{tV_H}{BI} \quad (cm^3/C) \quad , \tag{1}$$

where t is the thickness of the conductor, B the magnetic field strength and I the drift current. The sign of the Hall coefficient is negative for electrons and positive for holes, with the carrier density given by [5]

$$n, p = \mp \frac{r_H}{eR_H} \quad (cm^{-3}), \tag{2}$$

where e is the elementary charge and r_H the Hall scattering factor. When both electrons and holes are present, as is the case for small bandgap semiconductors, the Hall coefficient becomes [5]

$$R_H = \frac{p - n\left(\frac{\mu_n}{\mu_p}\right)^2}{e\left[p + n\left(\frac{\mu_n}{\mu_p}\right)\right]^2} \quad , \tag{3}$$

for a weak magnetic field, where μ_n and μ_p are the electron and hole mobilities, respectively. The mobility ratio μ_n/μ_p was assumed to be temperature independent and determined by the effective electron and hole masses. This is a valid assumption within the intrinsic ($n \sim p$) temperature range if optical-phonon scattering dominates for both carrier types.

If an additional conduction path is present, the measured Hall coefficient will be a weighted average of the contribution from the bulk and the second conduction layer. Considering a two layer model composed of a layer of thickness t_1 and

conductivity σ_1, and a second of thickness t_2 and conductivity σ_2, the Hall coefficient is instead given by [5]

$$R_H = \frac{(t_1 + t_2)(R_{H1}\sigma_1^2 t_1 + R_{H2}\sigma_2^2 t_2)}{(\sigma_1 t_1 + \sigma_2 t_2)^2} \quad , \qquad (4)$$

in the weak magnetic field limit, where R_{H1} and R_{H2} are the Hall coefficients of the two layers, respectively. In the case of a degenerate second layer, R_{H2} and σ_2 are assumed to be temperature independent.

B. Seebeck Effect

For a non-degenerate conductor, the Seebeck coefficient is determined from the formula [6]:

$$\alpha_{n,p} = \mp\frac{k}{e}\left[r + 2 + ln\frac{2\left(2\pi m^*_{n,p} kT\right)^{3/2}}{h^3 n,p}\right] \quad (V/K), \qquad (5)$$

where n and p are the concentrations, m^* the effective masses, and r the scattering parameter for electrons and holes, respectively, and k and h are Boltzmann's and Planck's constants. Considering mixed conduction, the contributions of both carriers are weighted by the conductivities of the conduction and valence bands, such that the total Seebeck coefficient is given by [6]:

$$\alpha_{bulk} = \frac{\alpha_p \mu_p p + \alpha_n \mu_n n}{\mu_p p + \mu_n n} \quad . \qquad (6)$$

When considering small bandgap p-type semiconductors, the Seebeck coefficient changes sign from positive to negative as the temperature is increased from the extrinsic ($p >> n$) to intrinsic ($p \sim n$) temperature range. This transition temperature is directly related to the mobility ratio of the carriers and the net hole concentration, N_A-N_D. In the absence of any additional conduction paths, this transition occurs when the intrinsic carrier density, $n_i = (N_A\text{-}N_D)(\alpha_p\mu_p/\alpha_n\mu_n)$. For degenerate layers, the Seebeck coefficient needs to be analysed using Fermi-Dirac statistics, in which case [7]

$$\alpha_{n,p} = \mp\frac{k}{e}\left[\frac{r+2}{r+1}\frac{F_{r+1}(\eta)}{F_r(\eta)} - \eta\right] \qquad (7)$$

and

$$n,p = 2(2\pi m^*_{n,p} kT/h^2)^{3/2} F_{1/2}(\eta)\,, \qquad (8)$$

where η is the reduced Fermi level (= E_F/kT). The Fermi integrals $F_r(\eta)$ of integer and half-integer order have been tabulated by Rhodes [8] and Mc Dougall and Stoner [9]. The scattering parameter, $r = 0$ for lattice phonon scattering and $r = 2$ for scattering by impurity ions. For Seebeck measurements performed above 300K, phonon scattering is assumed to dominate. Some polar semiconductors, such as InAs, are known to form a degenerate charge accumulation layer at the surface [3]. A surface layer with typically an electron density of 10^{18} cm^{-3} would then have a Seebeck coefficient of 0.1 mV/K at 600 K, decreasing to 0.05 mV/K at 300 K. Since these values are relatively low compared to that of the bulk, combined with the increased bulk conductivity at elevated temperatures, a degenerate electron conduction layer is expected to make a negligible contribution to the overall thermoelectric emf. The inclusion of a degenerate conduction layer is therefore anticipated to merely lower the measured thermoelectric emf, decreasing the observed Seebeck coefficient to

$$\alpha_{apparent} = \frac{\sigma_b}{\sigma_b + \sigma_s}\,\alpha_{bulk} \quad , \qquad (9)$$

where σ_s and σ_b are the sheet conductivities of the bulk and degenerate layer, respectively. The sign of the Seebeck coefficient is thus solely determined by the bulk conduction properties, leaving the transition temperature *unaffected* by the addition of a degenerate conduction layer. The accuracy by which the doping density can be determined is then only limited by knowledge of the intrinsic carrier density and mobility ratio of the carriers within the bulk material at the transition temperature. Although the transition temperature typically exceeds the temperature range of Varshni equations quoted for most semiconductor materials, it generally remains a valid extrapolation due to the linearity of the bandgap energy above room temperature. The Seebeck coefficient is also accurately described by the non-degenerate approximation (5), since the transition temperature increases with increased doping, thereby assuring that the Boltzmann approximation remains valid throughout the doping range investigated.

III. Experimental Details

The InAs layers were grown by metal-organic vapour phase epitaxy on semi-insulating GaAs substrate. The layers were all specular and were roughly 4 µm thick. The van der Pauw contact geometry (Fig. 1), with four indium ohmic contacts fabricated on each corner of the epilayer, was used to measure the Hall voltage and conductivity. The Hall measurements were performed in a 1 kG magnetic field, with the sample placed in a closed system helium cryostat for the low temperature measurements.

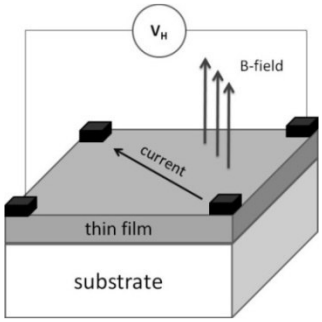


Fig. 1. Van der Pauw contact configuration used for the Hall effect measurements. The sample is placed in a magnetic field perpendicular to the surface of the layer, with the Hall voltage measured diagonal to the current flow.

The Seebeck measurements were performed on 5mm×2mm×0.5mm samples mounted between two copper blocks, with indium ohmic contacts fabricated on both ends of the epilayer (Fig. 2). The Seebeck coefficient was obtained by

measuring the Seebeck voltage, ΔV_S, for different temperature gradients $\Delta T = T_{HOT} - T_{COLD}$ and then calculating $\alpha_{apparent}$ from the slope of $\Delta V_S(\Delta T)$. The differential temperature between the contacts was monitored using a type-T thermocouple mounted directly to each of the two ohmic contacts. The temperature difference was typically less than two degrees Kelvin, while the average sample temperature varied between 350 K and 10 K.

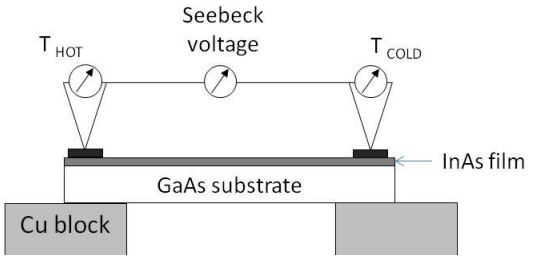


Fig. 2. Experimental setup used for the temperature dependent Seebeck measurement. The temperature was measured at the ohmic contacts using type-T (Cu *vs* Cu-Ni) thermocouples, whereas the thermoelectric emf was measured between the two Cu elements.

IV. Results and Discussion

A. Hall Effect

Figure 3 depicts the temperature dependence of the Hall coefficient measured for a cadmium doped p-type InAs epilayer. The Hall coefficient is negative at 300 K, changing sign twice as it is cooled down to 10 K. The temperature dependence can however be well represented by the two-layer model described in section II.A. The parameters used to generate the theoretical curves are listed in Table I. The surface conduction parameters were obtained from the low temperature (<20 K) Hall measurements where carrier freeze-out reduces the bulk contribution. The acceptor density, acceptor ionisation energy and the temperature dependence of the hole mobility were optimised using a least-square routine. The remainder of the parameters are from literature [10].

TABLE I
Parameters used to fit the temperature dependence of the Hall coefficient in Fig. 3.

Symbol	Description	Value
m_n^*	Hole effective mass	0.023
m_p^*	Electron effective mass	0.41
E_0	Varshni parameters	0.415 eV
β		210 K
α		3.4×10^{-4} eV/K
t	Layer thickness	4.2 µm
μ_n/μ_p	Mobility ratio	72
r	Hall scattering factor	1
N_A	Acceptor concentration	2.3×10^{16} cm^{-3}
E_A	Acceptor ionisation energy	(20 ± 2) meV
n_s	Surface sheet density	3.4×10^{12} cm^{-2}
μ_s	Surface electron mobility	400 cm^2/V·s
μ_L	Lattice scattering mobility	$8\times10^5\times T^{-3/2}$ cm^2/V·s
μ_I	Ionised impurity scattering	$33\times T^{3/2}$ cm^2/V·s

A comparison of the measured Hall coefficient and the calculated bulk contribution reveals that the Hall effect measurements drastically underestimate the bulk values throughout most of the temperature range studied. The only reasonable correlation exists near 50 K, where the bulk mobility is sufficiently high to suppress the contribution by the surface accumulation layer. The degree to which the measured Hall coefficient represents the bulk value is therefore highly sensitive to the mobility, and hence the quality of the epilayer investigated.

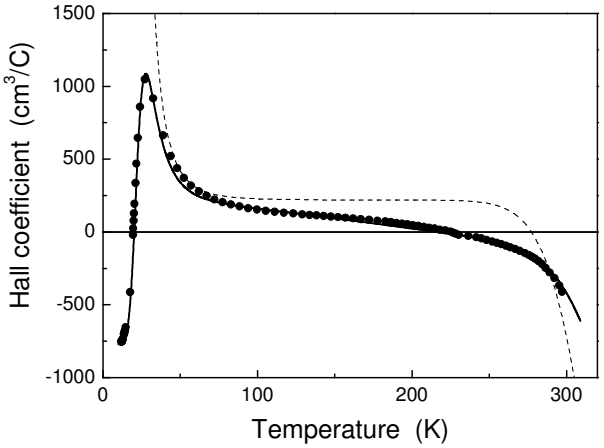


Fig. 3. Temperature dependence of the Hall coefficient of cadmium doped InAs (solid circles). The solid line represents the theoretical curve calculated for a two-layer structure composed of a bulk layer with an acceptor concentration of $N_A = 2.3\times10^{16}$cm^{-3} and a negative accumulation layer, with the dashed line representing the bulk component.

The ionisation energy of cadmium E_{Cd} = (20±2) meV was obtained from the low temperature measurements (<50 K). This is slightly higher than the value reported by Guseva *et al.* [11], who obtained an ionisation energy of 15 meV.

It is important to note that the large temperature range used for the Hall measurements is essential for the extraction of the basic doping characteristics of the material. The measurements obtained within the intrinsic temperature range are most sensitive to the doping density, whereas the contribution from surface conduction dominates at low temperatures. The accurate analysis of material with larger doping densities therefore requires a progressively larger temperature range for the Hall measurements.

B. Seebeck Effect

In order to justify the assumption made for (9), where the presence of a degenerate conduction channel solely reduces the overall thermoelectric voltage without affecting the sign, the temperature dependence of the Seebeck coefficient was measured for the cadmium doped InAs film used in the Hall effect measurements. The experimentally obtained Seebeck values are shown in figure 4. The sign reversal observed at 355 K corresponds to an acceptor concentration of 3.6×10^{16} cm^{-3} using (6) and (9). It is apparent that the Seebeck coefficient displays an anomalous temperature dependence compared to the expected bulk behavior (represented by the dashed curve). The incorporation of a surface conductivity term into the Seebeck model, however, provides a very good representation of the experimental trend. The bulk mobility and surface conductivity values used for the theoretical curves

also correlate well with the Hall analysis of Sec. IV.A, thereby further confirming the validity of (9).

Although a large increase in the Seebeck coefficient at low temperatures (dashed curve) is often ascribed to thermal drag [12], here it is simply related to carrier freeze-out, as observed in the low temperature Hall effect measurements.

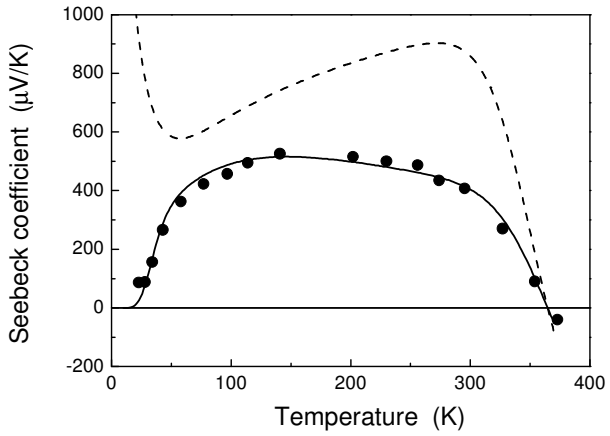


Fig. 4. Temperature dependent Seebeck coefficient of cadmium doped InAs (solid circles). The solid line represents the theoretical curve calculated for a two-layer structure with a bulk acceptor concentration of $N_A = 3.5\times10^{16}$cm^{-3}, with the dashed line representing the bulk contribution.

It is instructive to note that contrary to Hall effect measurements, Seebeck measurements allow for the hole density to be easily determined by only noting the temperature at which the Seebeck coefficient reverses sign. Instead of the detailed measurements presented in Fig. 4, the sample is instead slowly heated, whilst maintaining a small temperature gradient across the length of the sample. It then remains to monitor the potential between the two ohmic contacts until it changes sign. This approach was followed for a series of p-type InAs layers grown with different levels of cadmium doping. The Seebeck transition temperature increased from 374 K to 535 K as the cadmium mole fraction was increased from 2×10^{-5} to 3×10^{-4}. The corresponding acceptor concentrations (assuming negligible compensation) presented in Fig. 5 display an excellent one-to-one relation to the cadmium mole fraction. This simple approach is the first to give direct information on the incorporation efficiency of lightly doped p-type InAs. The Seebeck measurements are therefore clearly a more attractive approach, requiring no thickness information and displaying an apparent insensitivity to the layer quality.

V. Conclusion

We have demonstrated that the temperature dependence of the Hall coefficient of p-type InAs can be well described by a two-layer model composed of a degenerate electron accumulation layer and a dual carrier bulk layer. The analysis required Hall measurements extending from temperatures sufficiently low to observed carrier freeze-out, up to intrinsic temperatures. The complexity associated with the accumulation layer was overcome by revealing the direct relationship between the acceptor concentration in the bulk and the temperature at which the Seebeck coefficient reverses sign. This allowed for the accurate determination of the acceptor concentration of cadmium doped InAs down to a concentration of 3.5×10^{16} cm^{-3}.

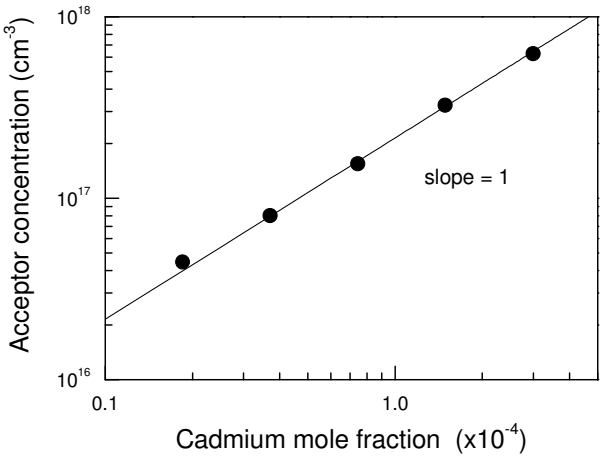


Fig. 5. Acceptor concentration of p-type InAs as a function of cadmium mole fraction introduced during thin film growth.

Acknowledgment

This work is based upon research supported by the SA Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation. The financial support of NMMU and the PRISM-M (DPSS CSIR) project is gratefully acknowledged.

References

[1] D. C. Look and R. J. Molnar, "Degenerate layer at GaN/sapphire interface: Influence on Hall-effect measurments"Appl. Phys. Lett., vol. 70, no. 25, p. 3377, 1997.

[2] Y. Lin, A. R. Arehart, A. M. Carlin, and S. A. Ringel, "Separation of bulk and surface electron transport in metamorphic InAs layers using quantitative mobility spectrum analysis" Appl. Phys. Lett., vol. 93, p. 062109, 2008.

[3] L. Ö. Olsson, C. B. M. Andersson, M. C. Håkansson, J. Kanski, L. Ilver, and U. O. Karlsson, "Charge accumulation at InAs surfaces" Phys. Rev. Lett., vol. 76, p. 3626, 1996.

[4] O. J. Pitts, D. Lackner,Y. T.Cherng, S. P.Watkins, "Growth of InAsSb/InPSb heterojunctions for mid-IR detector applications" J. Crystal Growth, vol. 310, p. 4858, 2008.

[5] D. K. Schroder, *Semiconductor material and device characterization* New York: Wiley-Interscience Publications, 1990, ch. 5

[6] V. A. Johnson and K. Lark-Horovitz, "Theory of thermoelectric power in semiconductors with applications to germanium" Phys. Rev., vol. 92, p. 226, 1953.

[7] C. Wood, V. Harrap, and W. M. Kane, "Degeneracy in Ag_2Te" Phys. Rev., vol. 121, no. 4, p.978, 1961.

[8] P. Rhodes, Proc. R. Soc., London A, vol. 204, p. 396, 1950.

[9] J. McDougall and E. C. Stoner, Philos. Trans. R. Soc. London A, vol. 237, p. 67, 1983.

[10] S. M. Sze, *Physics of Semiconductor Devices*, 2nd Ed., New York: Wiley-Interscience Publications, 1981, p. 849.

[11] M. I. Guseva, N. V. Zotova, A. V. Koval, and D. N. Nasledov, Sov. Phys. Semicond., vol. 8, no. 10, p. 1323, 1975.

[12] A. V. Inyushkin, A. N. Taldenkov, V. I. Ozhogin, K. M. Itoh, and E. E. Haller, "Isotope effect on the phonon-drag component of the thermoelectric power of germanium", Phys. Rev. B., vol. 68, p. 153203, 2003.