

# FEASIBILITY OF CMOS OPTICAL CLOCK DISTRIBUTION NETWORKS

By

**Petrus Johannes Venter**

Studyleader: Professor M. du Plessis

Submitted in partial fulfillment of the requirements for the degree

**Master of Engineering (Microelectronic Engineering)**

in the

Department of Electrical, Electronic & Computer Engineering

in the

Faculty of Engineering, Built Environment & Information Technology

UNIVERSITY OF PRETORIA

November 2009

# SUMMARY

---

FEASIBILITY OF CMOS OPTICAL CLOCK DISTRIBUTION NETWORKS

by

Petrus Johannes Venter

Studyleader: Professor M. du Plessis

Department of Electrical, Electronic & Computer Engineering

Master of Engineering (Microelectronic Engineering)

---

CMOS is well known for its ability to scale. This fact is reflected in the aggressive scaling on a continual basis from the invention of CMOS up to date. As devices are scaled, device performance improves due to shorter channel lengths and more densely packed functions for the same amount of area. In recent years, however, the performance gain obtained through scaling has begun to suffer under the degradation of the associate interconnect performance. As devices become smaller, interconnects need to follow. Unlike transistors, the scaling of interconnects results in higher capacitances and resistances, thereby limiting overall system performance. Trying to alleviate the delay effects results in increased power consumption, especially in global structures such as clock distribution networks. A possible solution to this problem is the use of optical interconnects, which are fast and much less lossy than the electrical equivalents. This dissertation describes an investigation on what future technology nodes will entail in terms of power consumption of clock networks, and what is required for an optical alternative to become feasible. A common clock configuration is used as a basis for comparison, where both electrical and optical networks are designed to component level. Optimisation is done on both to ensure a reasonable comparison, and the results of the respective power consumption components are then compared in order to find the criteria for a feasible optical clock distribution scheme.

**Keywords:**

**CMOS, optical clock, optical interconnect, repeater, H-tree, detection, power consumption, clock distribution.**

# OPSOMMING

---

FEASIBILITY OF CMOS OPTICAL CLOCK DISTRIBUTION NETWORKS

deur

Petrus Johannes Venter

Studieleier: Professor M. du Plessis

Departement Elektriese-, Elektroniese- & Rekenaar Ingenieurswese

Meester in Ingenieurswese (Mikroëlektronika Ingenieurswese)

---

CMOS is bekend vir die gemaklike skalering van toestelle. Die feit word beklemtoon deur die agressiewe skalering wat plaasgevind het vanaf die uitvinding van CMOS tot op hede. Soos toestelle geskaleer word neem die werkverrigting van die toestelle toe, met meer funksionaliteit op 'n kleiner area. Daar word egter onlangs gesien dat die wins in werkverrigting ly onder skalering weens die nadelige effek wat skalering het op die interverbinding werkverrigting. Soos toestelle afneem in grootte volg die interverbindinge ook. Anders as transistors beteken die afname in grootte groter weerstand en kapasitansie vir interverbindings, met 'n gepaardgaande afname in werkverrigting. Om hierdie probleem te omseil word 'n toename in drywing vereis, veral met globale interverbinding-netwerke, soos klokverspreiding. Een moontlike oplossing is om die interverbindings te vervang met optiese golfleiers, wat vinniger is en minder verliese toon as die elektriese ekwivalente. Hierdie verhandeling ondersoek toekomstige tegnologienodes in terme van klokdrywingsverbruik en op watter stadium dit lewensvatbaar sal word om te spring vanaf die elektriese na optiese domein. 'n Algemene kloknetwerkskema word gebruik as 'n vergelykingsbasis vir die vergelyking van drywingsverbruik in elektriese en optiese kloknetwerke. Die werk behels ook die optimering van albei om te verseker dat 'n sinvolle vergelyking getref kan word. Die drywingsverbruik word dan as 'n raamwerk gebruik waarmee die kriteria vir die wesenskap van 'n optiese netwerk bepaal kan word.

**Sleutelwoorde:**

**CMOS, optiese klok, optiese interverbinding, herhaler, H-boom, deteksie, drywingsverbruik, klokverspreiding.**



*In memory of my brother*

*Marnus Venter*

# ACKNOWLEDGEMENT

---

I would like to thank the Carl and Emily Fuchs Institute for Microelectronics (CEFIM) as well as the INSiAVA project, as well as all people involved therein for supporting this work, with an extension of gratitude to my wife and family members.

# CONTENTS

CHAPTER ONE - INTRODUCTION	<b>1</b>
1.1 Context of this work . . . . .	2
1.2 Motivation . . . . .	3
1.3 Organisation of dissertation . . . . .	3
1.3.1 Chapter 1: Introduction . . . . .	3
1.3.2 Chapter 2: Photonic principles in standard CMOS . . . . .	3
1.3.3 Chapter 3: Clock distribution networks . . . . .	4
1.3.4 Chapter 4: Power consumption in future electrical networks . . . . .	4
1.3.5 Chapter 5: Optical receiver and front end design . . . . .	4
1.3.6 Chapter 6: Comparative results . . . . .	4
1.3.7 Chapter 7: Conclusion . . . . .	5
CHAPTER TWO - PHOTONIC PRINCIPLES IN STANDARD CMOS	<b>6</b>
2.1 Photon properties and optical power . . . . .	6
2.2 Semiconductor photodetectors . . . . .	7
2.2.1 Photomultipliers . . . . .	8
2.2.2 Photoconductors . . . . .	8
2.2.3 Photodiodes . . . . .	9
2.3 The optical source and signal path . . . . .	10
2.4 Optical interference at interfaces . . . . .	11
2.4.1 Interference patterns in insulator stack . . . . .	11
2.4.2 Locating the minima and maxima for a specific wavelength . . . . .	14
2.5 Modelling of a <i>pn</i> -junction photodiode . . . . .	15
2.5.1 Mechanisms involved in the physics of a photodiode . . . . .	16
2.5.2 The doping profile and <i>pn</i> -junction characteristics . . . . .	19
2.5.3 Analytical model of a photodiode . . . . .	22
2.6 On-chip interconnects . . . . .	23

<b>CHAPTER THREE - CLOCK DISTRIBUTION NETWORKS</b>	<b>25</b>
3.1 Clock distribution architectures . . . . .	25
3.1.1 Grids . . . . .	25
3.1.2 Trees . . . . .	26
3.1.3 Serpentine . . . . .	26
3.2 Balanced H-tree . . . . .	26
3.2.1 Global network . . . . .	27
3.2.2 Local region network . . . . .	28
3.3 The interconnect . . . . .	29
3.3.1 Electrical interconnect model . . . . .	30
3.3.2 Optical interconnects . . . . .	34
3.4 Skew in clock networks . . . . .	35
3.4.1 Sources of skew in electrical clock networks . . . . .	36
3.4.2 Sources of skew in optical clock networks . . . . .	37
3.5 Signal propagation over an interconnect . . . . .	38
3.5.1 Propagation across an electrical interconnect . . . . .	38
3.5.2 Propagation through an optical waveguide . . . . .	39
3.6 Power dissipation in electrical clock distribution networks . . . . .	40
3.6.1 Power dissipation mechanisms . . . . .	40
<b>CHAPTER FOUR - POWER CONSUMPTION IN FUTURE ELECTRICAL NETWORKS</b>	<b>41</b>
4.1 Summary of future technology nodes . . . . .	41
4.1.1 Specific device models . . . . .	41
4.1.2 Important technology defining parameters . . . . .	42
4.1.3 Summary of device parameters . . . . .	44
4.1.4 Model SPICE curves . . . . .	44
4.1.5 Interconnect parameters . . . . .	45
4.2 Chip size model . . . . .	46
4.3 Determining tree depth . . . . .	47
4.4 Modelling local network load . . . . .	49
4.5 Repeater design . . . . .	50
4.5.1 Extraction of repeater parameters <i>A</i> and <i>B</i> . . . . .	51
4.5.2 Repeater optimisation . . . . .	54
4.5.3 Improved repeater optimisation . . . . .	55
4.5.4 Optimisation in terms of capacitance . . . . .	56

4.5.5	End point buffer design . . . . .	60
4.5.6	Split buffer design . . . . .	61
<b>CHAPTER FIVE - OPTICAL RECEIVER AND FRONT-END DESIGN</b>		<b>63</b>
5.1	Photodiode as circuit element . . . . .	63
5.1.1	Photodiode intrinsic bandwidth . . . . .	64
5.1.2	Photodiode responsivity . . . . .	65
5.1.3	Photodiode capacitance . . . . .	65
5.2	Receiver considerations . . . . .	65
5.2.1	Bandwidth limitations . . . . .	65
5.2.2	Noise limitations . . . . .	66
5.2.3	Drive capability . . . . .	66
5.3	Overview of typical receivers . . . . .	66
5.3.1	Common source switching amplifier . . . . .	66
5.3.2	Digital buffer amplifier . . . . .	70
5.3.3	Common gate amplifier . . . . .	72
5.3.4	Common source feedback transimpedance amplifier . . . . .	75
5.3.5	Regulated cascode amplifier . . . . .	77
5.3.6	Complementary transimpedance amplifier . . . . .	80
5.4	Choice of amplifier . . . . .	82
5.4.1	High impedance receiver . . . . .	83
5.4.2	Low impedance receiver . . . . .	83
5.4.3	High versus low impedance approaches . . . . .	83
5.5	Design of high impedance amplifier . . . . .	84
5.5.1	Design requirements . . . . .	84
5.5.2	Designed values . . . . .	84
5.5.3	Power consumption per amplifier . . . . .	84
<b>CHAPTER SIX - COMPARISON OF NETWORK POWER CONSUMPTION</b>		<b>88</b>
6.1	Overview . . . . .	88
6.2	Explanation of results . . . . .	88
6.3	65 nm node . . . . .	89
6.3.1	Tree design . . . . .	89
6.3.2	Power dissipation components at different levels . . . . .	89
6.4	45 nm node . . . . .	92

6.4.1	Tree design . . . . .	92
6.4.2	Power dissipation components at different levels . . . . .	92
6.5	32 nm node . . . . .	93
6.5.1	Tree design . . . . .	93
6.5.2	Power dissipation components at different levels . . . . .	94
6.6	22 nm node . . . . .	97
6.6.1	Tree design . . . . .	97
6.6.2	Power dissipation components at different levels . . . . .	97
6.7	16 nm node . . . . .	99
6.7.1	Tree design . . . . .	99
6.7.2	Power dissipation components at different levels . . . . .	99
6.8	11 nm node . . . . .	100
6.8.1	Tree design . . . . .	100
6.8.2	Power dissipation components at different levels . . . . .	100
6.9	Hybrid approach . . . . .	104
6.10	Power consumption across nodes . . . . .	104
<b>CHAPTER SEVEN - CONCLUSION</b>		<b>106</b>
7.1	Interpretation of results . . . . .	106
7.1.1	Overall power consumption performance . . . . .	106
7.1.2	Requirements on the source . . . . .	106
7.1.3	Dependence on operating frequency . . . . .	107
7.2	Limitations and shortcomings of the work . . . . .	107
7.2.1	Predictive work . . . . .	107
7.2.2	Photodiode limitations . . . . .	107
7.2.3	H-tree branching . . . . .	108
7.2.4	Interconnect model . . . . .	108
7.2.5	Skew and jitter . . . . .	108
7.2.6	Optimisation of electrical tree . . . . .	108
7.2.7	Global clock frequency model . . . . .	109
7.3	Contribution of this work . . . . .	109
7.4	Future research possibilities identified in this work . . . . .	109
7.5	Conclusion . . . . .	110
<b>REFERENCES</b>		<b>111</b>

APPENDIX A - ANALYTICAL PHOTODIODE PARTIAL DIFFERENTIAL EQUATION	<b>115</b>
A.1 The problem . . . . .	115
A.2 X(x) and terms . . . . .	116
A.3 Y(y) and terms . . . . .	116
A.4 Z(z) and terms . . . . .	117
A.5 Time and more... . . . .	117
A.6 Complete solution for $p(x, y, z, t)$ . . . . .	117
A.7 The function $f(x, y, z, t)$ . . . . .	117
A.8 Substitution and solution . . . . .	119
A.9 Modelling $\Phi(t)$ as a square wave . . . . .	120
A.10 Summary . . . . .	120
A.11 Determining current densities . . . . .	121
A.12 $J_{px}$ and $I_x$ solution . . . . .	121
A.13 $J_{py}, J_{pz}, I_y$ and $I_z$ solution . . . . .	122

# LIST OF FIGURES

2.1	A semiconductor under illumination . . . . .	8
2.2	Energy band model for different photoelectric effects . . . . .	9
2.3	Thin-film interference affecting the optical transmission power through multiple layers . . . . .	13
2.4	Effect of slight deviation of layer thickness on transmission coefficient . . . . .	13
2.5	Transmission coefficient as a function of layer thicknesses . . . . .	14
2.6	An illustration of the gradual transition between n-well and p-substrate doping . . . . .	20
2.7	An illustration of the abrupt nature of a p+ to n-well transition in doping concentration . . . . .	20
2.8	Geometric representation of analytical photodiode model . . . . .	22
3.1	A grid topology of a global clock network . . . . .	26
3.2	Balanced H-tree clock network representation . . . . .	27
3.3	A typical implementation of the local region grid . . . . .	29
3.4	A reduced modern interconnect stack adapted from [1] . . . . .	30
3.5	Cross section of the interconnect physical model . . . . .	31
3.6	Interconnect model for top global metal . . . . .	32
3.7	Interconnect model with top and bottom plates . . . . .	33
3.8	Illustration of current density and skin depth . . . . .	34
3.9	An example of skew . . . . .	36
4.1	Future scaling in CMOS technology as predicted by the ITRS 2008 [2] . . . . .	42
4.2	Effective gate length as compared to a drawn transistor . . . . .	44
4.3	NMOS $V - I$ characteristics for future technologies, gate at $V_{DD}$ , $1 \mu\text{m}$ width . . . . .	45
4.4	Local region representation for skew calculations . . . . .	47
4.5	Extracting the resistance data for the parameter $A$ . . . . .	52
4.6	Example of linearised buffer resistance extraction on 32 nm node . . . . .	52
4.7	Extracting the capacitance data for the parameter $B$ . . . . .	53

4.8	SPICE simulation of optimal and mismatched buffer widths at the 45 nm node	55
4.9	Optimal width for repeater transistors . . . . .	57
4.10	Predicted reduction of segment length for future technology nodes . . . . .	58
4.11	A representation of a terminating clock network . . . . .	61
5.1	A simple electrical equivalent of an n-well photodiode . . . . .	64
5.2	Common source amplifier configuration as an optical end point . . . . .	66
5.3	Front end utilising buffer high impedance input node for charge accumulation	70
5.4	Common gate amplifier configuration as an optical end point . . . . .	72
5.5	Small signal model of common gate amplifier . . . . .	72
5.6	Common source feedback transimpedance amplifier . . . . .	75
5.7	Small signal model of the common source feedback TIA . . . . .	76
5.8	Regulated cascode circuit as an optical end point . . . . .	78
5.9	Small signal model for the regulated cascode . . . . .	78
5.10	A transimpedance feedback amplifier based on a complementary CMOS pair	80
5.11	Small signal model of complementary feedback TIA . . . . .	81
5.12	Simplified small signal model for complementary feedback amplifier . . . . .	81
6.1	Combination of electrical and optical tree for the 65 nm node . . . . .	91
6.2	Combination of electrical and optical tree for the 45 nm node . . . . .	95
6.3	Combination of electrical and optical tree for the 32 nm node . . . . .	96
6.4	Combination of electrical and optical tree for the 22 nm node . . . . .	98
6.5	Combination of electrical and optical tree for the 16 nm node . . . . .	102
6.6	Combination of electrical and optical tree for the 11 nm node . . . . .	103
6.7	Results of power consumption comparison across nodes . . . . .	105

# LIST OF TABLES

3.1	H-tree equations regarding lengths and segments . . . . .	28
4.1	ITRS 2008 requirements for high performance logic devices . . . . .	44
4.2	Comparison between SPICE and ITRS predicted values . . . . .	45
4.3	Intermediate interconnect electrical characteristics . . . . .	46
4.4	Global interconnect electrical characteristics . . . . .	46
4.5	Information for determining effective logic area . . . . .	47
4.6	Requirements for tree depth calculation . . . . .	48
4.7	Tree depth metrics associated with technology nodes . . . . .	49
4.8	Estimation of local region capacitances for future technology nodes . . . . .	49
4.9	Parameters required for repeater design . . . . .	50
4.10	Extracted $A$ and $B$ values for future nodes based on predictive models . . . . .	53
4.11	Repeater design summary . . . . .	55
4.12	Improved optimised repeater characteristics . . . . .	56
4.13	Capacitance optimised repeater characteristics . . . . .	59
4.14	Theoretical power consumption per repeater . . . . .	60
4.15	Simulated power components per repeater . . . . .	60
4.16	Repeater configuration split buffer design parameters . . . . .	62
4.17	Theoretical and simulated power consumption per split buffer . . . . .	62
5.1	Requirements for high impedance optical receiver front end . . . . .	84
5.2	Design parameters for high impedance amplifier across technology nodes . . . . .	85
5.3	High impedance amplifier capacitive components for power calculations . . . . .	86
5.4	Theoretical power consumption per amplifier . . . . .	86
5.5	Simulated power components per amplifier . . . . .	87
6.1	Summary of attributes for the 65 nm technology node . . . . .	90
6.2	Summary of attributes for the 45 nm technology node . . . . .	92

LIST OF TABLES

---

6.3	Summary of attributes for the 32 nm technology node . . . . .	93
6.4	Summary of attributes for the 22 nm technology node . . . . .	97
6.5	Summary of attributes for the 16 nm technology node . . . . .	99
6.6	Summary of attributes for the 11 nm technology node . . . . .	100

## ABBREVIATIONS

---

AC	alternating current
ARC	anti-reflective coating
ASIC	application specific IC
CMOS	complementary MOS
DC	direct current
EP	end point
EPE	external power efficiency
FET	field effect transistor
IC	integrated circuit
ILD	inter-metal dielectric
ITRS	International Technology Roadmap for Semiconductors
LOCOS	local oxidation of silicon
MOS	metal-oxide semiconductor
MPU	microprocessor unit
NMOS	n-type MOS transistor
PMOS	p-type MOS transistor
PTM	predictive technology model
RGC	regulated cascode
SOI	silicon-on-insulator
SPICE	simulation program with integrated circuit emphasis
TIA	transimpedance amplifier
VLSI	very large scale integration

# CHAPTER ONE

## INTRODUCTION

---

**I**F the success of an invention is measured by the extent of commercialisation opportunities, the impact reach on individuals and just the brute number of manufactured instances, the transistor is arguably the most successful invention in human history. From the concept of J. E. Lilienfeld's field-effect device, first patented in 1925, through the first operational bipolar injection device by W. Shockley, J. Bardeen and W. Brattain at Bell Laboratories in 1947, up to the advanced MPUs in computers and cellular telephones today, the transistor has changed the front of technology continuously for more than half a century.

Although the transistor is electrically similar to the vacuum tube preceding it as a gain element, one of the strongest drivers for development is the fact that transistors are much smaller and hence integrate extremely well. It is possible to place multiple individual devices on a single semiconductor substrate. G. E. Moore, co-founder of Fairchild Semiconductor and Intel, wrote an article in 1965 [3] stating that the complexity of such an integrated circuit would double annually for a given minimum component cost. This became known as the famous Moore's Law, and has been very accurate even 40 years later.

Today a number of materials are used in the manufacturing of different semiconductor devices. However, one material far surpasses the rest in terms of usage as a semiconductor. Silicon, the second most abundant material in the earth crust at 27.7 % by mass, is also by far the most abundant material used in ICs today. The success thereof lies in the fact that silicon-dioxide, which can be easily grown on a pure silicon substrate, is a very good electrical insulator. This fact paved the way for the creation of tightly integrated, highly economical manufacturing of circuits consisting of field effect devices. The metal-oxide-semiconductor, or MOS structure is therefore much easier to manufacture than bipolar injection-type devices.

CMOS (complimentary metal-oxide-semiconductor) is the dominant technology in the

semiconductor industry today. There are a number of reasons why CMOS took up the dominant seat when it comes to integrated circuits.

Although the first MOS integrated circuits only consisted of one type of transistor, F. Wanlass not only refined the fabrication of MOS transistors on a silicon substrate; he also recognized that complimentary devices can be used to create circuits with extremely low static power [4]. This discovery reshaped the foundation for digital circuits for the future. Added to the power benefit, CMOS circuits can have signals on the output node traversing the full range of the supply voltage. This means full utilisation of signal range, as well as a benefit to the noise margins of CMOS digital circuits.

While CMOS hold the edge over other logic technologies in terms of power consumption, two other key features of CMOS devices shifted CMOS fully into the focus of VLSI designers and manufacturers. The first is the fact that CMOS devices are planar, hence the designer holds some degree of freedom in terms of the device geometry [5]. Moreover, this ability can be generalised and are not bound by a specific technology. The second feature is the ease of scaling CMOS devices. While modifying a number of key parameters [6] in appropriate relation, devices can be shrunk, and circuits designed can easily be “transferred” to a new, smaller and faster technology node. While it was predicted early in the evolution of CMOS that the “ease-of-scaling” attribute cannot continue indefinitely [7], strong emphasis has been placed on developing solutions to problems as they appeared, with great success so far. Moore’s law still seem valid up to date.

## 1.1 CONTEXT OF THIS WORK

As scaling continues, device performance increase. Along with this decrease, the metal tracks, or interconnects, conveying information from one device to another also need to scale. However, this comes at the cost of degraded performance, because

1. as dimensions decrease, the cross-sectional area of a conductor decreases, thereby increasing its resistance, and
2. more densely packed components means more capacitance and crosstalk.

This is commonly known as the *interconnect dilemma* [8], which represents a so called *red brick wall*, aptly named because of the red bricks allocated in the ITRS predictions [1]. Within this context, this work focuses on utilising the information available from a number

of sources in order to predict whether a feasible solution to the *interconnect dilemma* might exist.

## 1.2 MOTIVATION

As candidate solutions to the interconnect problem become available, it becomes necessary to create a set of metrics whereby the feasibility of these solutions can be judged. The specific subject of this work emphasises the possibility of optical clock networks. The outcome of this work is aimed at proving

1. whether optical clock distribution will improve or at least equal the performance of electrical clock networks, and
2. under which circumstances will a light source be sufficient for optical distribution of clock signals.

If the possibility of optical clock networks can be proven, the individual aspects arising from this work will serve as a good starting point for further research efforts in trying to accomplish an operational substitute for electrical clock networks in future technology nodes.

## 1.3 ORGANISATION OF DISSERTATION

### 1.3.1 Chapter 1: Introduction

This chapter serves as an overview to the importance of CMOS in the modern semiconductor domain. This includes the motivation for this research, as well as the context, contribution and relevancy to work actively done on trying to solve the red brick wall called the *interconnect dilemma*.

### 1.3.2 Chapter 2: Photonic principles in standard CMOS

The chapter introduces basic photonic principles common to semiconductors and then extends the explanation to apply to standard CMOS technology. An explanation is given on how optical energy interacts with semiconductors, how to detect and convert optical energy into electrical energy, and what influences can be expected in standard CMOS. The chapter concludes with a mathematical model of a photodiode detector, which will be used to identify

certain characteristics and limitations in practically implementable photodiodes on a CMOS chip.

### **1.3.3 Chapter 3: Clock distribution networks**

This chapter covers the reasoning behind the necessity for clock networks. Some architectures are mentioned and one is chosen on which to base the comparison between optical and electrical networks. Important performance specifications are identified and explained, which serve as key points in determining the relative performance of the two approaches to clock networks.

### **1.3.4 Chapter 4: Power consumption in future electrical networks**

Based on ITRS predictions on current and future trends in CMOS, electrical clock networks are designed for technology nodes from 65 nm down to 11 nm. The techniques are extended to make use of predictive SPICE models used to future design exploration. Optimisation methods are also developed in order to reduce power and maximise performance. The philosophy is that there is room for a number of optimisations in electrical networks, thus it will be more sensible to compare an optical network to an efficient electrical network than a suboptimal choice.

### **1.3.5 Chapter 5: Optical receiver and front end design**

An optical clock network cannot utilise digital standard cell library cells alone in the optoelectronic conversion process. A number of receiver architectures are investigated in terms of their suitability for use in optical clock networks. The different designs result in different trade-offs. The most suitable is chosen, designed and analysed in terms of power consumption.

### **1.3.6 Chapter 6: Comparative results**

This chapter entails the cumulative work of the previous chapters as a set of results comparing the power consumption of the different approaches. The results of each node is shown individually, with the emergence of the possibility of a hybrid approach at smaller technologies. Finally, a comparison in terms of power consumption is shown for the full optical, full electrical and hybrid approaches.

### **1.3.7 Chapter 7: Conclusion**

The chapter consists of a continuation of interpretation of the results obtained in the previous chapters. The requirements on the optical source as well as some limitations and potential for future research is discussed herein.

# CHAPTER TWO

## PHOTONIC PRINCIPLES IN STANDARD CMOS

---

### 2.1 PHOTON PROPERTIES AND OPTICAL POWER

THE advantages of using light as a transport mechanism for information can only be utilised when an optical signal can be effectively generated, transmitted through a compatible medium and detected on the receiving side, with minimal loss of information. Data are most often handled by electronic circuits on both the transmit and receive side of the communication system, which means that electrical power must be converted to optical power, transmitted, received and converted back into an electrical signal. In the development of an optical clock distribution network, both mechanisms are equally important. The emphasis of this research, however, will briefly lie in the detection process, since it will be assumed that the electrical to optical conversion will happen off-chip. More in depth analysis of photodetection is beyond the scope of this work, since it is only required to gain enough insight for a reasonable circuit equivalent for the photodetection principle.

The detection process is fundamentally limited by the available optical power reaching the detector. This requires an understanding of light itself, behaving as both a wave and particle in terms of quantised energy. It is these quanta that are most useful in analysing photodetection, since a light wave can only have energies as integer multiples of these quanta, called photons. Ideally, each photon will be part of the conversion process, meaning each photon will act on the electrical signal being generated. Of course, the former only describes the ideal scenario. The energy of a single photon can be described by equation 2.1,

$$E = h \frac{c_0}{\lambda} \quad [\text{J}] \quad (2.1)$$

with Planck's constant  $h = 6.626 \times 10^{-34}$  J.s,  $c_0$  being the speed of light in a vacuum as  $c_0 = 2.99792 \times 10^8$  m/s and  $\lambda$  the photon wavelength in m.

The photon energy can also be referred to in eV rather than Joule, with the conversion described by equation 2.2,

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J} \quad (2.2)$$

Optical power can therefore be defined by equation 2.3,

$$P_{opt} = rh \frac{c_0}{\lambda} \quad [\text{W}] \quad (2.3)$$

where  $r = \text{photons/second}$ . Equation 2.3 can further be modified as per equation 2.4 in a more appropriate form for optical detectors in terms of a unit area of active optical detective material.

$$P_{area} = \frac{rhc_0}{\lambda A} \quad \left[ \frac{\text{W}}{\text{cm}^2} \right] \quad (2.4)$$

where  $A$  is the unit area typically in  $\text{cm}^2$ .

The photon flux, a term used to describe the amount of photons passing through a plane of unit area per time unit, can be expressed as in equation 2.5, where the term is often described in photons per second per  $\text{cm}^2$ .

$$\Phi = \frac{r}{A} \quad (2.5)$$

Using equations 2.1 to 2.5, it is possible to determine how many photons will ideally arrive at a photodetector surface and, in the ideal case, generate a photocurrent. As per [9] we have thin film behaviour, which influences the photon flux reaching the silicon substrate dramatically. Section 2.4 contains an analysis on the effects thereof, where an analytical model is introduced in order to predict the losses at optical interfaces.

## 2.2 SEMICONDUCTOR PHOTODETECTORS

When a photon with energy greater than the semiconductor bandgap energy enters the material, it can be absorbed. Upon absorption, an electron-hole pair is generated. Depending on the type of detector used, this carrier generation action can be used to detect optical signals applied to the material. This section will focus on different types of detectors making use of the photoelectric generation of carriers for optical detection. An investigation of

different types of optical detectors will yield information on the feasibility of employing such detectors in a standard CMOS processes. Although a number of detectors will not apply to CMOS integrability, a short discussion will be given for the purpose of relevancy.

### 2.2.1 Photomultipliers

A photomultiplier uses two electrodes in a vacuum to generate a spatial electric field. The cathode is usually made of a material prone to photo-emission, such as a low work-function metal or a semiconductor [10]. A photon is absorbed in the material if the photon energy is large enough, and the probability exists of emission into the electric field. The electron accelerates through the field, where additional electrodes are set up such that the energetic electron hits these and causes the emission of multiple electrons. Very similar to the avalanche effect in  $pn$ -junctions, large current gains are achievable as a function of an optical input power.

Although a semiconductor might be utilised in photomultipliers, the way of operation does not permit standard CMOS ICs to take advantage of this effect. Therefore, this type of detection method is not compatible with CMOS.

### 2.2.2 Photoconductors

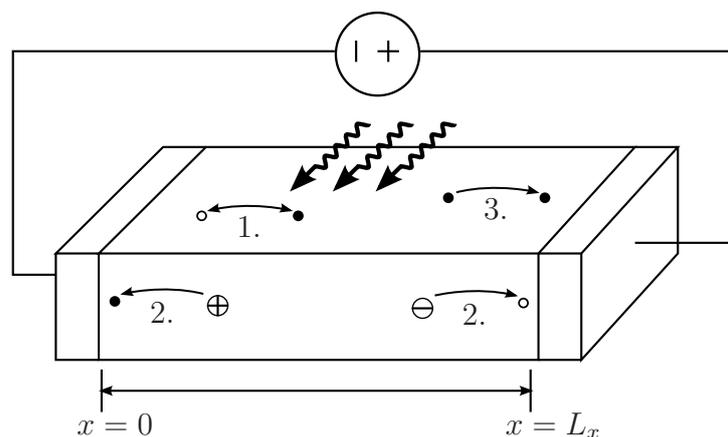


FIGURE 2.1: A semiconductor under illumination

Photoconductors typically refer to a piece of material between contacts, where the material resistivity changes as a function of light. Semiconductors are very suitable for this, since the generation of carrier pairs upon the absorption of photons can result in carriers

available for conduction. This obviously implies the application of an electric field between the contacts, otherwise generated carriers will simply diffuse in random order and eventually recombine, with no net current flowing between the terminals.

A number of mechanisms for carrier excitation exist in semiconductors, thereby allowing the material to be utilised for the photoconduction effect. According to [10], three generation mechanisms can be responsible for photoelectric action (refer to figure 2.2):

1. Interband, or band-to-band absorption, where an electron-hole pair is generated upon photon absorption,
2. impurity absorption, where an impurity can absorb a photon with low energy and free a carrier, and
3. intraband, or free-carrier absorption, where an already conducting carrier is excited to a higher energy state.

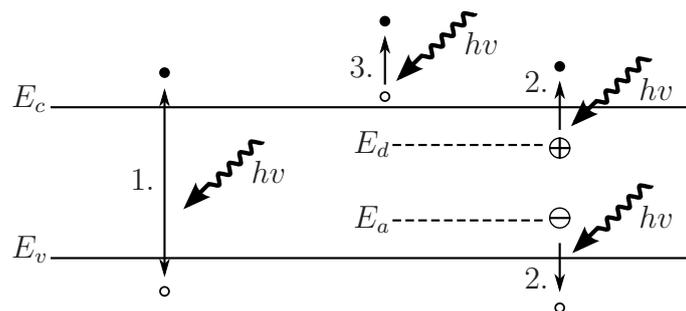


FIGURE 2.2: Energy band model for different photoelectric effects

A photoconductor setup is shown in figure 2.1 indicating the listed carrier generation mechanisms. Figure 2.2 shows the band diagram with the associated mechanisms.  $E_c$ ,  $E_v$ ,  $E_d$  and  $E_a$  represent the conduction, valence, donor and acceptor energy levels respectively, while  $h\nu$  is the energy of the absorbed photon. Only mechanism 1 is usually contributory to optical detection at room temperature.

### 2.2.3 Photodiodes

While photoconductors rely on an applied electric field in order to create carrier drift, another type of semiconductor detector makes use of the electric field induced by a  $pn$ -junction in order to generate a photo current. This type of detector is called a photodiode. Ideally, all

incident photons should knock out carrier pairs in the depletion region formed when two opposite types of semiconductors are abutted. The electric field caused by the junction potential will then cause the carriers to drift with respect to the electric field, thereby generating a photocurrent irrespective of an externally applied voltage. By applying an external voltage, actual junction potential can be modified to obtain some specific effects.

Practically implementable photodiodes have regions of bulk material around the depletion region, wherein the electric field magnitude is usually small enough to be negligible in terms of influencing the photocurrent generated in the depletion region. But since absorption can often not be confined to the depletion region alone, another photocurrent mechanism originates in these bulk regions. Where the carrier pairs in the depletion region are subject to an electric field and therefore drift, the excess carriers created in the bulk regions are subject to diffusion. When a majority carrier reaches the depletion region through diffusion, the electric field sets a barrier for transport in that direction. However, when a minority carrier reaches the edge of the depletion region, the electric field can sweep the carrier into and through the depletion region, resulting in a charge imbalance that corrects itself through the flow of current.

Photodiodes are probably the most suited for CMOS integration, as photoconductors require an isolated piece of semiconductor. Unlike photoconductors where all carrier action happens under the influence of an electric field, photodiodes have two sources for the limitation of intrinsic device bandwidth:

1. Carriers generated in the depletion region, subjected to an electric field.
2. Carriers generated in bulk regions, subjected to the much slower process of diffusion.

The diffusion currents are usually the main limitation in the frequency response of photodiodes and will often manifest as long tailing currents on the application of short optical pulses. A more detailed mathematical analysis on the modelling of a photodiode in standard CMOS will be investigated in section 2.5.

## 2.3 THE OPTICAL SOURCE AND SIGNAL PATH

Investigating the origin of photon generation and the path of an optical signal in the detection process is critical in understanding the mechanisms that influences the optical to electrical conversion of energy. Since the amount of optical power reaching the silicon is directly related to the amount of photons available to generate carrier pairs, it is necessary to model

the interface losses, which can be quite substantial where the layers' thickness is comparable to the operating wavelength. Hence it becomes worthwhile to investigate these losses in detail, as done in sections 2.4 and 2.6.

## 2.4 OPTICAL INTERFERENCE AT INTERFACES

Standard CMOS processing techniques are usually focused more towards digital design optimisation. As a result, the SiO<sub>2</sub> insulation layers and the Si<sub>3</sub>N<sub>4</sub> passivation layers are not tailored for optical design. These layers are usually comparable to the interested wavelength of light in terms of layer thickness and will, in behaviour, act as a thin film optical filter. The result is that light transmission through these layers are subject to optical interference patterns, as a function of the layer material refractive indices, layer thickness and the amount of interfaces present in the path of light.

As light is emitted from the ideal source, photons travel towards the photodiode through various media. In a free space system, the first medium of travel will be air, with a refractive index  $n_{air} = 1$ . In a standard LOCOS CMOS technology, the next medium will probably be Si<sub>3</sub>N<sub>4</sub>, used as a passivation layer to prohibit the migration of harmful ions from reaching the lower insulators and causing time-varying capacitance fluctuations. SiO<sub>2</sub>, which forms the insulation between stacked conductive elements, forms the layer between the Si substrate and Si<sub>3</sub>N<sub>4</sub> passivation layer. In modern processes, with complex planarisation methods and low- $\kappa$  dielectrics, this back end layer stack might become even more complex. Note that the passivation layer might be removed and even replaced with an anti-reflective coating (ARC) as a post-processing step for more beneficial optical characteristics.

### 2.4.1 Interference patterns in insulator stack

The reflection from the surface of a multilayer stack of materials is described in [11] by equation 2.6.

$$\mathcal{R}_j = \frac{r_{j+1} + \mathcal{R}_{j-1}e^{-j\frac{4\pi}{\lambda}n_jd_j}}{1 + r_{j+1}\mathcal{R}_{j-1}e^{-j\frac{4\pi}{\lambda}n_jd_j}}, \quad \mathcal{R}_0 = r_1 \quad (2.6)$$

where equation 2.7 describes the Fresnel reflection component for wave amplitudes and  $n_j$  and  $d_j$  are the refractive index and thickness of layer  $j$  respectively.

$$r_j = \frac{n_{j-1} - n_j}{n_{j-1} + n_j} \quad (2.7)$$

Note that the refractive index  $n_j$  may be complex for the case of an absorptive material, where the quantity simply then becomes  $\mathbf{n}_j = n_j - \mathbf{j} \cdot k_j$ , where  $n_j$  is the real refractive index and  $k_j$  is known as the extinction coefficient. For the standard CMOS case, the band gap for  $\text{SiO}_2$  is around 8.0 eV and above [12], and for  $\text{Si}_3\text{N}_4$  the band gap is above 5.0 eV [13]. Although these values may vary depending on the deposition method and amount of accidental impurities, the values fall well beyond the photon energies for detection purposes. With this in mind, it can be assumed that the materials are non-absorbent with refractive values of  $n_{\text{SiO}_2} = 1.458$  and  $n_{\text{Si}_3\text{N}_4} = 2.035$ . For Si, the required values was taken from [14] over a wavelength range  $\lambda$  between 200 nm and 1000 nm. Including all known values and keeping  $\mathbf{n}_{\text{Si}}$ ,  $d_{\text{SiO}_2}$  and  $d_{\text{Si}_3\text{N}_4}$  as variables, it can be shown that the amplitude reflection for the standard CMOS case is represented by equation 2.8.

$$\mathcal{R} = \frac{0.3410 + \left[ \frac{-0.1652 + \left( \frac{\mathbf{n}_{\text{Si}} - 1.458}{\mathbf{n}_{\text{Si}} + 1.458} \right) e^{-j1.458 \frac{4\pi}{\lambda} d_{\text{SiO}_2}}}{1 - 0.1652 \cdot \left( \frac{\mathbf{n}_{\text{Si}} - 1.458}{\mathbf{n}_{\text{Si}} + 1.458} \right) e^{-j1.458 \frac{4\pi}{\lambda} d_{\text{SiO}_2}}} \right] e^{-j2.035 \frac{4\pi}{\lambda} d_{\text{Si}_3\text{N}_4}}}{1 + 0.3410 \cdot \left[ \frac{-0.1652 + \left( \frac{\mathbf{n}_{\text{Si}} - 1.458}{\mathbf{n}_{\text{Si}} + 1.458} \right) e^{-j1.458 \frac{4\pi}{\lambda} d_{\text{SiO}_2}}}{1 - 0.1652 \cdot \left( \frac{\mathbf{n}_{\text{Si}} - 1.458}{\mathbf{n}_{\text{Si}} + 1.458} \right) e^{-j1.458 \frac{4\pi}{\lambda} d_{\text{SiO}_2}}} \right] e^{-j2.035 \frac{4\pi}{\lambda} d_{\text{Si}_3\text{N}_4}}} \quad (2.8)$$

While equation 2.8 represents the amplitude reflection coefficient, it is necessary to determine the amount of optical power reflected from the surface of the layered stack, as well as the optical power available to the photodiode in the Si substrate. The amplitude and power of a wave can be related by a square law and, since both the  $\text{SiO}_2$  and the  $\text{Si}_3\text{N}_4$  layers are non-absorbent, it can be assumed that no power is lost through the layers. The amount of transmitted optical power can then be determined by equation 2.9.

$$T(\mathbf{n}_{\text{Si}}, d_{\text{SiO}_2}, d_{\text{Si}_3\text{N}_4}) = 1 - |\mathcal{R}|^2 \quad (2.9)$$

A typical CMOS detector with both  $\text{SiO}_2$  and  $\text{Si}_3\text{N}_4$  layers present, will typically have a transmission coefficient as shown in figure 2.3.

Although the interference can be precisely determined under the assumptions of orthogonal incidence and completely flat surfaces, the thickness of actual fabricated layers may differ greatly. For the most part, the back end layer stack is created using deposition techniques lacking fine control. Figure 2.4 shows the difference in optical power transmission for even a small change in layer thicknesses. The vicinity of the plot is chosen in the region of interest for photodetection.

It is therefore necessary to take the extremities of layer deviations into account and

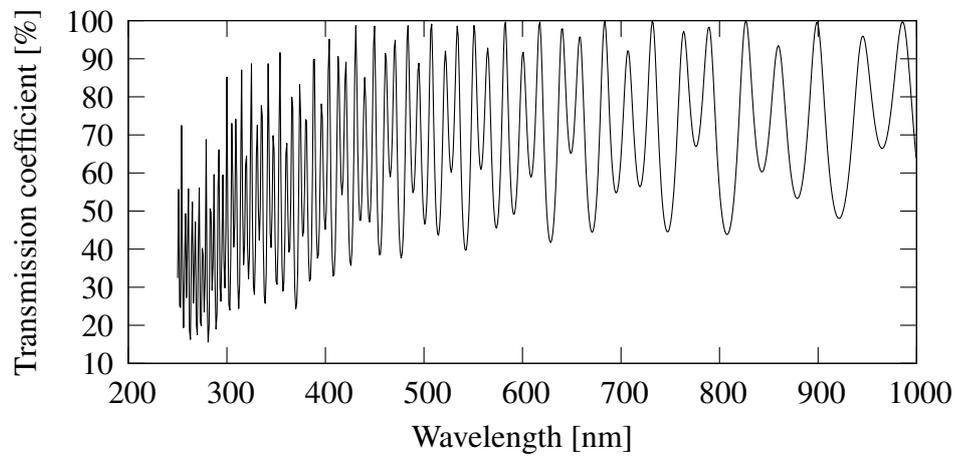


FIGURE 2.3: Thin-film interference affecting the optical transmission power through multiple layers

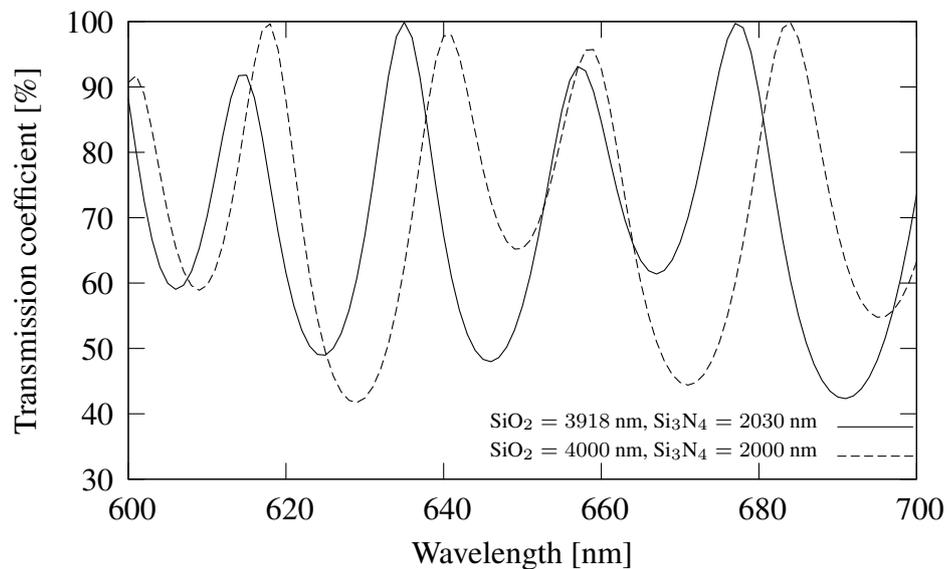


FIGURE 2.4: Effect of slight deviation of layer thickness on transmission coefficient

determine the minimum expected  $T$  as a worst case value. The best case scenario can, as shown, lead to values as high as 100 % transmission.

## 2.4.2 Locating the minima and maxima for a specific wavelength

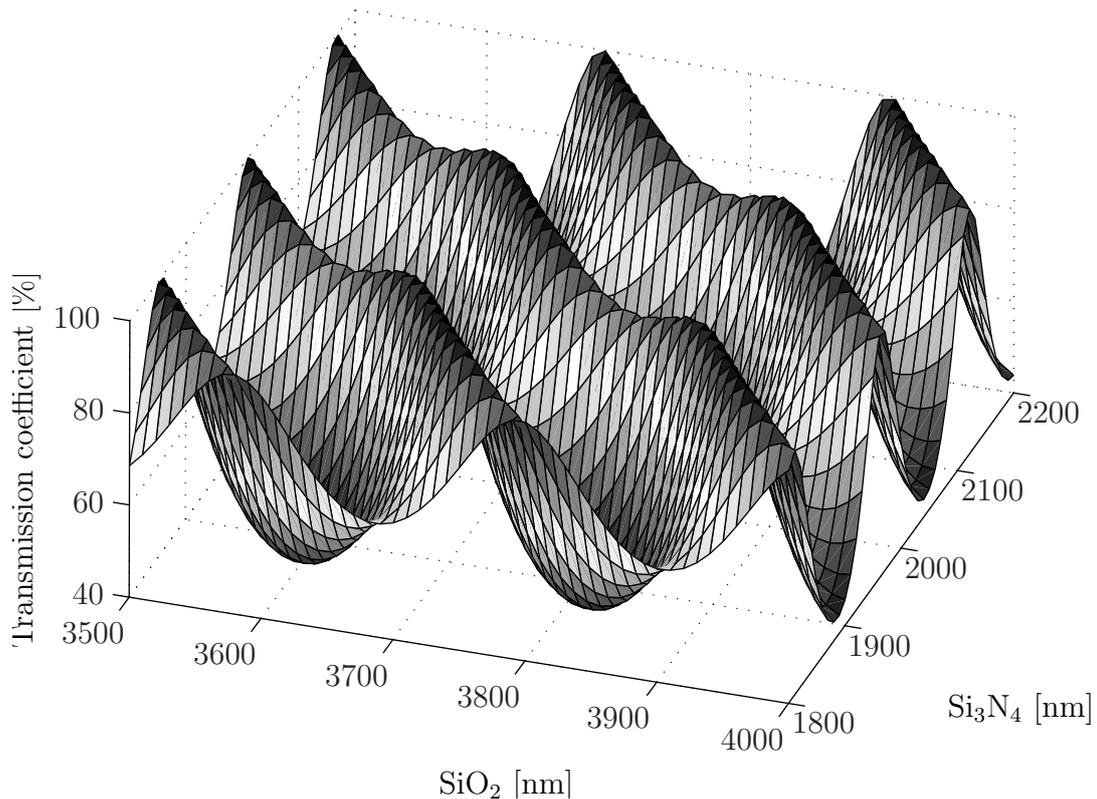


FIGURE 2.5: Transmission coefficient as a function of layer thicknesses

Figure 2.5 shows that the transmission coefficient is a periodic function of the layer thicknesses, where a maximum and minimum can be determined by taking equation 2.10 and determining the critical points, while substituting the results back into equation 2.9.

$$\frac{\partial T}{\partial d_{SiO_2}} = 0 \text{ and } \frac{\partial T}{\partial d_{Si_3N_4}} = 0 \quad (2.10)$$

In this example,  $T_{min} = 41.35\%$  and  $T_{max} = 100\%$ . This is based on a typical  $0.35\ \mu\text{m}$  LOCOS CMOS process.

## 2.5 MODELLING OF A *pn*-JUNCTION PHOTODIODE

The most practical way of implementing a photodetection element is to make use of *pn*-junctions and the surrounding bulk to capture and convert photons to charge carriers. Although, for a given piece of silicon, having an entirely depleted illumination region in order to avoid the slow diffusion and recombinative effects of electron-hole pair would be ideal, the latter effects often has a strong dependency on wavelength and these structures can potentially be effective as detection elements at high pulsing frequencies.

Although a designer has limited choice in a CMOS process in terms of available structures, it is worthwhile modelling the physical behaviour of a detector both for the optimisation of the device, as well as the subsequent electrical design involving the device. A good model should be able to predict the following:

- the generated current for a certain optical input power,
- the generated current for different wavelengths, given a fixed optical input power,
- a good estimate of the device's electrical bandwidth, a term to take into account when designing with the device, and
- the limit on intrinsic bandwidth, often the limiting factor in the ultimate speed obtainable by the device.

It is worth elaborating on the two different types of bandwidth referred to above. Electrical bandwidth is usually only determinable when viewing the device in circuit context. The electrical bandwidth will typically model the leakage conductance, or dark current, as a resistive or conductive element, as well as the capacitive part of the detector as a voltage dependent capacitor. A typical scenario would then include a transimpedance amplifier, with an input also represented by a resistive and possibly a capacitive element. The combination of these elements will relate to a specific electrical bandwidth. It is also important to understand that this is usually not the fundamental limit of the device's operating frequency.

The intrinsic bandwidth refers to the limit of the physical mechanisms inside the device responsible for generating current. As mentioned, diffusion of minority carriers in the different bulk areas is usually responsible for this characteristic and no manner of electrical design will change this. More often, careful device design and geometric optimisations will alter and improve this limitation.

### 2.5.1 Mechanisms involved in the physics of a photodiode

In modelling the photodiode, it is important to identify the mechanisms which will be important enough to influence device behaviour, while trying not to increase the complexity beyond a reasonable level. Some important parameters in analytically describing a photodiode are:

- the diffusivity of minority carriers in different bulk areas,
- the lifetime of minority carriers in bulk regions,
- the absorption coefficient of light through the material at different wavelengths, and
- a description of the physical make-up of the diode and the details on the bulk and depletion regions.

Each of the items above will be discussed and used in the subsequent modelling procedure.

#### 2.5.1.1 Diffusivity in doped silicon

Diffusivity of minority carries has a strong dependency on the carrier mobility and temperature through the thermal voltage. Diffusivity can be related to carrier mobility using equation 2.11, the famous Einstein relation.

$$\frac{D}{\mu} = \frac{kT}{q} \quad (2.11)$$

Mobility can be determined using fitting parameters and experimental data, built around equation 2.12 as done in [15].

$$\mu = \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + \left(\frac{N}{N_{REF}}\right)^2} \quad (2.12)$$

In equation 2.12,  $N$  is the impurity doping concentration of the minority carriers of either type,  $N_{REF}$ ,  $\mu_{min}$  and  $\mu_{max}$  are experimental fitting parameters.

Some typical values for silicon that can be used are also given by [15], which can be used in estimating values that can be expected in a typical CMOS process. Although a lot of experimental extractions of these will be necessary to ensure a good practical versus theoretical fit, it is still key to understanding which parameters will eventually determine the fundamental operating limits of a device. An improved set of parameters and fitting

equations have been derived by [16], where regions well into the range above  $10^{19}$  are also fitted for different dopants.

### 2.5.1.2 Minority carrier lifetimes in doped silicon

Section 2.5.1.1 enables the understanding of the limit on the rate of movement of minority carriers in doped silicon. It is, however, inevitable for carriers, especially minority carriers, to recombine with the abundance of majority carriers present. In the light of device performance, this can both positively and negatively influence the operation of a device. If carriers recombine, the photon responsible for separating the initial electron-hole pair is wasted and does not contribute to the responsivity of the photodiode. While this is true, it also means that if light of a longer wavelength penetrated too deep into a certain bulk region, carriers beyond a certain point, usually expressed in some form of the diffusion length, will never reach the *pn*-junction boundary and thus never contribute to the tailing currents for which diffusion is so well known for. This in turn increases the intrinsic bandwidth of the device.

Since both the carrier lifetime and the diffusion length are thus of importance in device design, it is necessary to relate these quantities as shown in equation 2.13, where  $\tau$  is the average minority carrier lifetime,  $D$  is the diffusivity minority carriers within the bulk region and  $L_D$  is the average length travelled by a minority carrier in doped silicon.

$$L_D = \sqrt{\tau D} \quad (2.13)$$

It can also be shown that carriers of effectively 3 diffusion lengths away from a depletion region boundary will make no contribution to the generated current of a photodiode. This is also the reason why *p-i-n*-diodes make use of an extremely wide depleted piece of silicon, while minimising the dimensions of the bulk regions.

### 2.5.1.3 Photon absorption and carrier generation

The fundamental driver behind photodiodes is the generation term depending on the amount of photons being able to generate carrier pairs. This causes current flow and is the mechanism responsible for converting optical energy to electrical energy. In working towards modelling a photodiode, it is necessary to mathematically quantify this mechanism. Assuming 100 % quantum efficiency, that is, all of the photons absorbed in the silicon will be converted to electron-hole pairs, the absorption coefficient holds the key to this expression. It can be shown that the absorption of light in silicon at a certain depth can be described by equation

2.14, where the depth into the substrate is represented by  $x$  and the absorption coefficient, a fundamental property of the material itself, is  $\alpha$ .

$$\Phi(x) = \Phi_0 e^{-\alpha x} \quad (2.14)$$

Absorption at any point  $x$  will obviously then become  $1 - \Phi(x)$ . The rate of carrier generation will also be the rate of absorption in the direction  $x$ , therefore, by indirectly taking the derivative of 2.14, the generation term per unit volume becomes expression 2.15.

$$G(x) = \Phi_0 \alpha e^{-\alpha x} \quad (2.15)$$

The absorption coefficient has a strong dependency on wavelength, and an approximate expression for the quantity  $\alpha$  has been fitted in the form of a polynomial equation [17] as equation 2.16.

$$\alpha = 10^{13.2131} - 36.7985\lambda + 48.1893\lambda^2 - 22.5562\lambda^3 \quad (2.16)$$

#### 2.5.1.4 *pn*-Junction and depletion region calculations

The *pn*-junctions are responsible for much of a photodiode's characteristics. Although the generation term is independent of region within a piece of semiconductor, the behaviour of generated carriers differs a lot depending on the region within the diode where generation occurs. For carriers generated in bulk regions, that is, regions which are not depleted of carriers and where no electric field exists, the dominant transport mechanism is diffusion, and the carriers responsible for a photocurrent is the minority carriers. If an electron-hole pair is created within a bulk region, the majority carriers will contribute to the majority free carriers available for conduction. The majority carriers will locally diffuse in order to settle in an equilibrium state, but majority carriers reaching a depletion region boundary will be rejected due to the electric field inside the depletion region. The minority carriers, also under the influence of diffusion, will reach the boundary and be swept into the depletion region due to the polarity of the electric field. This will cause charge imbalance in the bulk region, resulting in a unit current flow in order to restore the imbalance.

Inside the depletion regions itself, an electron-hole pair will immediately be separated and will start to drift under the influence of the electric field. This will also result in a unit current flow in the series loop of the photodiode. Therefore, the two contributing transport mechanisms have to be determined based on geometric input on how the diode is physically constructed, as well as the location and width of the depletion regions in the

series photodiode loop. Crucial to characterising the diode in terms of bulk and depletion regions, as well as the lifetimes and diffusivity of minority carriers is the knowledge of the doping profile.

## 2.5.2 The doping profile and *pn*-junction characteristics

An understanding of the doping profile of a CMOS process is a necessary part of designing photodiodes, since physical attributes influence the operational characteristics directly. Unfortunately, it is not common practice for foundries to include manufacturing data in their design manuals. A method to extract the necessary information is therefore needed for accurate designs and predictions on the behaviour of photodiodes and the peripheral circuitry. The topic of this section is the development of “tools” that will facilitate the extraction of the information necessary for photodiode design.

A standard CMOS process implemented on a p-type wafer will typically have a lighter doped epitaxial layer on a low resistive substrate, an n-well implantation step and n+ and p+ diffusion deposition/implantation to form the required structures necessary for the creation of both NMOS and PMOS transistors. This creates a possible three interfaces for the formation of *pn*-junctions:

- n+ to p-substrate
- p+ to n-well
- n-well to p-substrate

Notice that the p-substrate might also be defined as regions of epitaxial p-type material, or the more heavily doped deep p-substrate. This probably requires different assumptions for the n+ to p-substrate and n-well to p-substrate interfaces.

### 2.5.2.1 Doped n-well region

For PMOS devices, an n-type substrate is necessary. The n-well is created for this purpose by implanting impurities to a certain depth within the substrate. After implanting the dopants, the wafer is exposed to a thermal drive-in cycle, where the diffusion of the impurities allows the n-well to extend both deeper into the substrate as well as fully up to the silicon surface. The effect is a distribution of impurities with a Gaussian distribution [18] around the point of implantation, where this is obviously interrupted on the surface side. Due to the drive-in process, impurities also diffuse laterally in an isotropic piece of silicon. The gradual

transformation of the material from an n-type to a p-type results in a junction that closely resembles a linear junction.

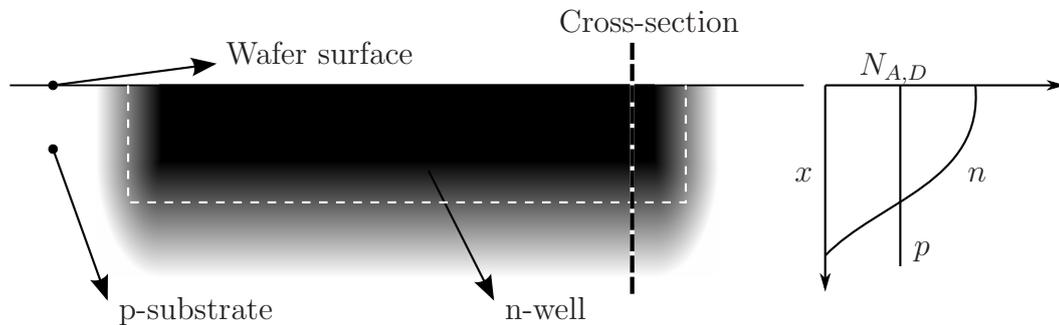


FIGURE 2.6: An illustration of the gradual transition between n-well and p-substrate doping

### 2.5.2.2 Doped n+ and p+ regions

The process for creating the n+ and p+ regions are usually done after well implantations. These diffusions can either be deposited on the silicon surface, or be implanted at a shallow depth. There is no explicit drive-in, diffusion happens through the thermal cycles used in other processing steps. This results in relatively abrupt junctions, or if linear, junctions with a high doping gradient. Because of the shallow nature of these diffusions, the resistivity information given in most design manuals should be sufficient to calculate the type of concentration of free carriers in these regions. This can be used as a starting point for calculating the rest of the profile.

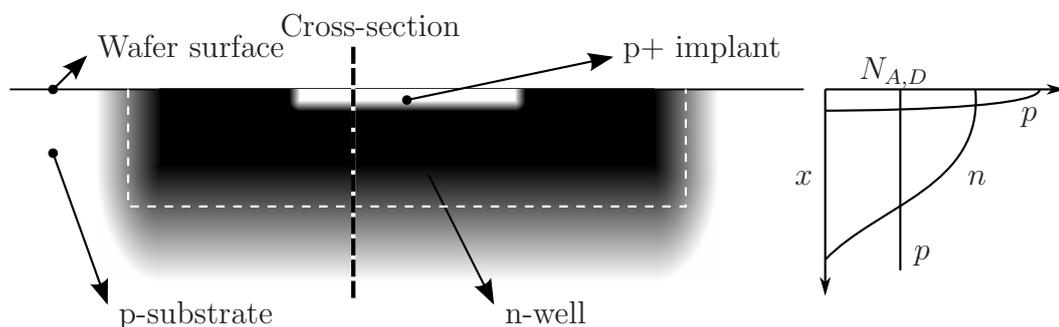


FIGURE 2.7: An illustration of the abrupt nature of a p+ to n-well transition in doping concentration

### 2.5.2.3 Epitaxial layer and substrate regions

The epitaxial layer is often used to create a lighter doping region where the NMOS devices can operate more efficiently, while the rest of the substrate is of higher doping concentration and therefore of lower resistivity, which often helps alleviating problems with latch-up and other substrate feedback and noise effects. Having a lower resistivity provides a low impedance path for stray signals to be conducted to the grounding node. Care has to be taken in identifying this feature, since this might influence the bottom wall capacitance of the n-well to p-substrate interface should the bottom of the well reach the low resistive substrate. It also complicates the profiling procedure.

### 2.5.2.4 Calculation doping quantities through resistivity

One possible way of calculating doping concentration is through known resistivity. This is especially useful in shallow highly doped diffusion areas, since the conduction would be relatively uniform in terms of depth. It is commonly known that the resistivity of a semiconductor is related to its carrier mobility by equation 2.17 [19], which can be used with equation 2.12 to derive the doping concentration of the material based on a given resistivity.

$$\rho = \frac{1}{qN\mu} \quad (2.17)$$

By algebraic manipulation, equation 2.17 can be used to determine  $N$  as the doping concentration of n- or p-type, where  $\mu$  refers to majority carrier mobility,  $q$  is the magnitude of electron charge and  $\rho$  represents the resistivity of the material.

### 2.5.2.5 Junction capacitance and doping relationships

Using parameters commonly given in design manuals regarding junction capacitances, it is possible to extract a lot of information on the junctions characteristics and the doping levels on the sides of the junction. Often included in the process parameters is the depth of the metallurgical junctions, which can give a photodiode designer an idea on where the carrier concentrations are at the intrinsic level.

The first parameter of interest is the grading coefficient. This describes the type of transition, linear versus abrupt, between the n- and p-type region. Although these values can range anywhere between 0.33 and 0.5 for linear and abrupt junctions respectively, it is usually close enough to one of the two extremes to make an assumption. These derivations,

with information on how the regions are usually created during wafer processing allow the designer to do educated guess work on the quantities of interest.

With the zero biased junction capacitance known, as usually given in process documentation, and one impurity doping quantity known, it is possible to determine the value of the other using reverse biased junction capacitance equations.

### 2.5.3 Analytical model of a photodiode

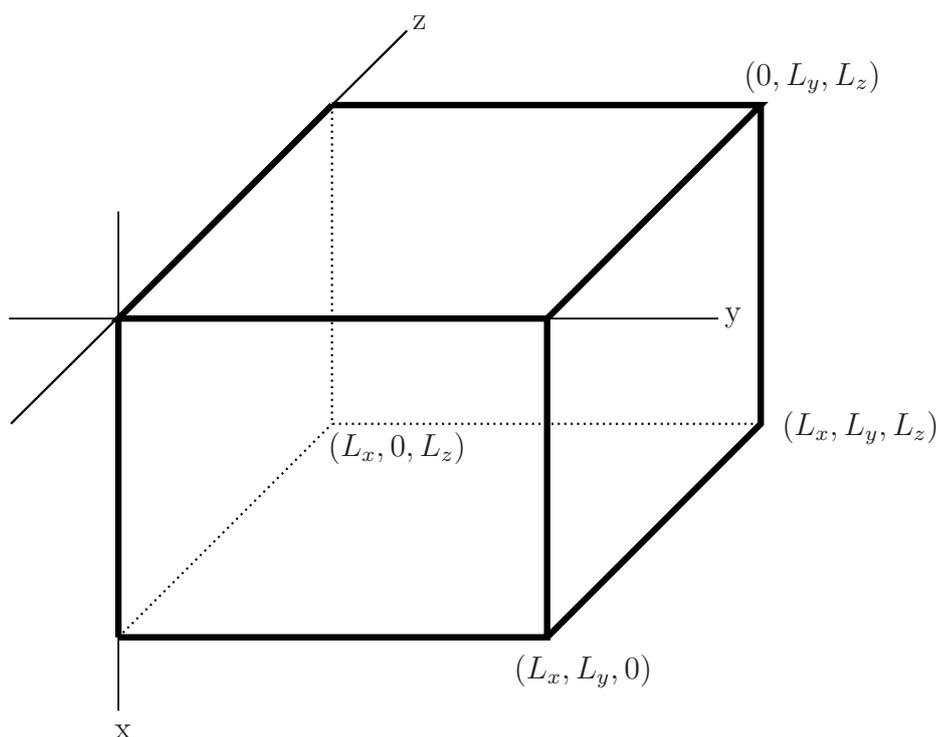


FIGURE 2.8: Geometric representation of analytical photodiode model

An analytical model for a simple geometry can be determined by solving the partial differential equation as created when the continuity equation terms are taken along with a carrier generation term representing the light generated carriers [20]. Equation 2.18 shows the continuity equations that need to remain valid across all semiconductor regions. Note that, for a semiconductor in equilibrium, the perturbations from equilibrium also adhere to equation 2.18. Thus, the effect of light absorbed by the semiconductor, generating carriers, can be superimposed on the equilibrium state, with the assumption that it does not have a

pronounced effect on the potential distribution in the region of interest.

$$\frac{\partial \Delta p}{\partial t} = D_p \left( \frac{\partial^2 \Delta p}{\partial x^2} + \frac{\partial^2 \Delta p}{\partial y^2} + \frac{\partial^2 \Delta p}{\partial z^2} \right) - \frac{\Delta p}{\tau_p} + \Phi_0(t) \alpha e^{-\alpha x} \quad (2.18)$$

A total solution for the current over the region of interest is shown in equation 2.19.

$$I_{n-well}(t) = p(x, y, z, t) \Big|_{V(x,y,z)=qD_pV_{xyz}} \quad (2.19)$$

where

$$V_{xyz} = \left( \frac{(2m-1)L_y L_z}{2\pi L_x q n} + \frac{2q L_y L_z}{\pi(2m-1)n L_x} + \frac{2n L_x L_z}{\pi(2m-1)q L_y} \right) \times \sin\left(\frac{2m-1}{2}\right) \pi (\cos(n\pi) - 1)(\cos(q\pi) - 1) \quad (2.20)$$

and

$$p(x, y, z, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \sum_{q=1}^{\infty} \frac{1}{n\pi} (1 - \cos(n\pi)) \frac{1}{q\pi} (1 - \cos(q\pi)) \times 4\alpha \left[ \frac{2\alpha L_x e^{\alpha L_x} + \pi(1-2m)\cos(\pi m)}{4\alpha^2 L_x^2 + 4\pi^2 m(m-1) + \pi^2} \right] \times e^{-\alpha L_x} \times \frac{\Phi_0}{2} \left( \frac{1 - \frac{1}{e^{D_p \gamma t}}}{D_p \gamma} + \sum_{j=1}^{\infty} \frac{2T}{\pi j} (1 - \cos(\pi j)) \times \left[ \frac{\frac{2\pi j}{e^{D_p \gamma t}} - 2\pi j \cos\left(\frac{2\pi j}{T}t\right) + T D_p \gamma \sin\left(\frac{2\pi j}{T}t\right)}{T^2 D_p^2 \gamma^2 + 4\pi^2 j^2} \right] \right) \times V(x, y, z)$$

where

$$V(x, y, z) = \sin\left(\frac{q\pi}{L_z} z\right) \sin\left(\frac{n\pi}{L_y} y\right) \cos\left(\frac{2m-1}{2} \frac{\pi}{L_x} x\right) \quad (2.21)$$

Appendix A shows an analytical approach to solving the continuity equations along with the method for determining current generated by a photodiode of a simple geometry. More complex geometries will require the use of computer software for solutions.

## 2.6 ON-CHIP INTERCONNECTS

Optical interconnects fully compatible with CMOS is still in its infancy. One option is that of post-processing techniques involving the deposition of polyimide as a final step [21], [22].

Another would be to utilise the existing nitride layer [23], isolated silicon dioxide layer or even SOI channels to serve as optical interconnects. The following losses are associated with the optical signal path:

- Coupling losses due to interfaces at the source into the waveguide
- Waveguide losses per length of optical propagation
- Bending losses associated with geometrical paths
- Splitting losses where the interconnects branch
- Coupling losses due to the interface from the interconnect into the photodiode

If external power efficiency (EPE) is defined as the ratio of optical output power over total electrical input power of a source as in equation 2.22, then the required EPE on the detector side needs to take into account the losses due to interface and interconnect effects.

$$EPE = \frac{P_{O_{out}}}{P_{E_{in}}} \times 100 \quad [\%] \quad (2.22)$$

As reported by [21], very efficient couplers are possible with polyimide interconnects, while propagation losses of 1.04 dB/cm are possible at 633 nm. As bending losses and branch losses depend too much on geometry, a sensible prediction is difficult.

$$\begin{aligned} P_{O_{source}} &= L_{coupling} + L_{branch} + L_{propagation} + L_{bending} + P_{O_{detector}} & [\text{dB}] \\ EPE_{req} &= P_{O_{source}} - P_{E_{in}} & [\text{dB}] \end{aligned} \quad (2.23)$$

Equation 2.23 shows the required source EPE as a function of losses and the required optical power at the detector in decibels. Given a fixed electrical input power, the equation can be used as an estimate for determining the type of optical source required for the given application.

# CHAPTER THREE

## CLOCK DISTRIBUTION NETWORKS

---

### 3.1 CLOCK DISTRIBUTION ARCHITECTURES

THE electrical distribution of clock signals is important in synchronous design, as events must happen simultaneously across the whole chip. This usually means that a single timing reference must be generated and distributed to the various regions on chip without a large temporal difference between the terminating points.

A number of techniques are used to distribute these signals. Three main types of topologies are often used, each with its own advantages and drawbacks [24], namely

1. grids,
2. trees, and
3. length matched serpentes.

#### 3.1.1 Grids

A well known example of a grid implementation is used in the DEC Alpha series of MPUs. The topology connects global interconnects in a grid-like fashion, where the clock signal is fed from the chip periphery.

Figure 3.1 shows a grid topology with the crosses representing the terminations into a local clock region. Notable of this configuration is the fact that clock skew is prominent, since the timing signal will arrive at a different time for each cross.

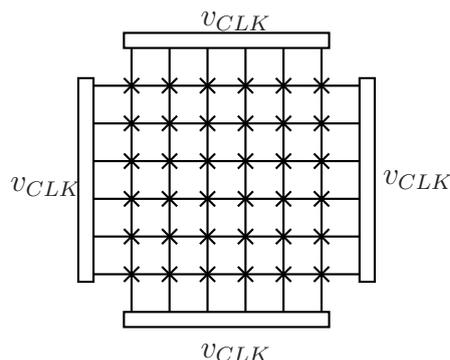


FIGURE 3.1: A grid topology of a global clock network

### 3.1.2 Trees

The skew problem can be alleviated to a large extent by using symmetrical tree structures such as the X-tree and H-tree topologies. The philosophy behind a symmetrical tree is that the length of travel to each terminating point is equal for all terminating points. This results in skew being only dependent on process variations, as briefly discussed in section 3.4. An example of an H-tree topology is shown in figure 3.2.

### 3.1.3 Serpentine

The Intel Itanium processor family utilises a scheme called the length matched serpentine topology [25], where the length of all clock feeding interconnects are once again equal. A large amount of interconnect is required, which increases the power consumption of the tree.

## 3.2 BALANCED H-TREE

The clock network architecture chosen for comparison is the balanced H-tree, primarily since it enables a direct comparison to be made between electrical and optical equivalents, and the skew component in the global network can be dramatically suppressed. The overall tree capacitance is also smaller compared to other topologies, as mentioned in [24]. Other works comparing optical networks with conventional electrical networks have been done [26], where a direct comparison with these works can also be made.

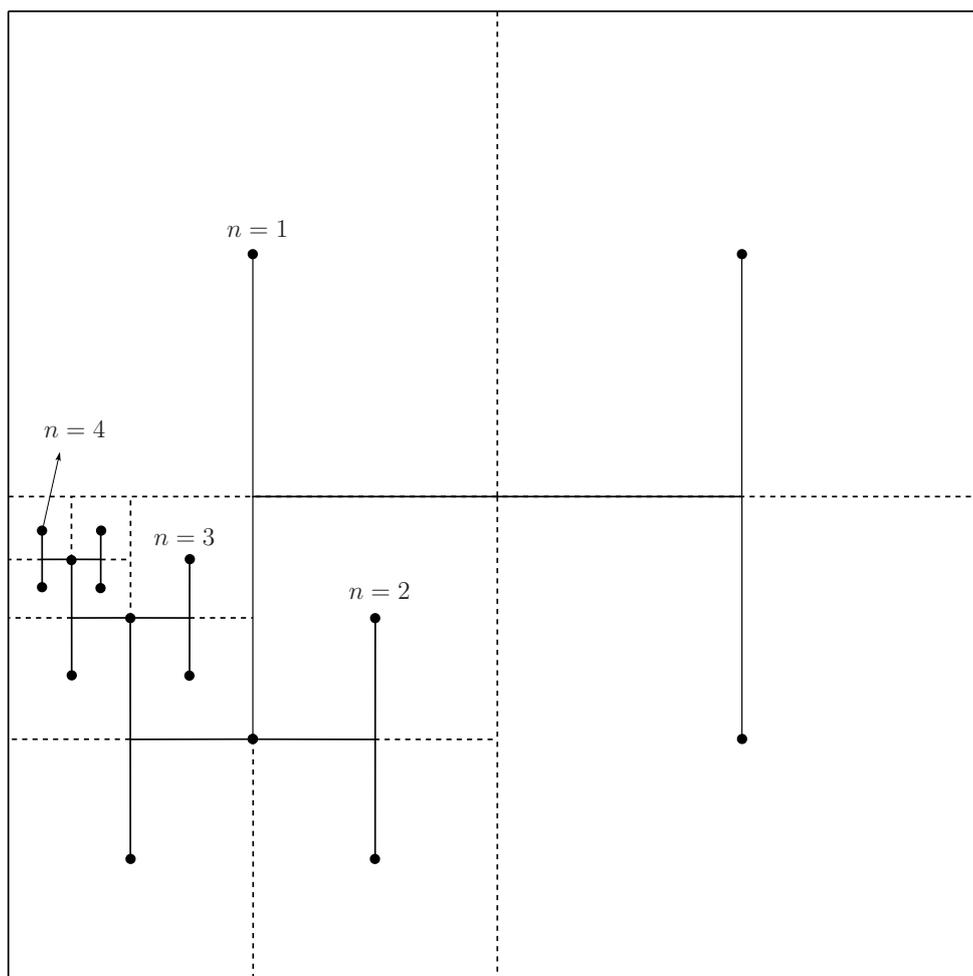


FIGURE 3.2: Balanced H-tree clock network representation

### 3.2.1 Global network

Assuming that the die is square, the following definitions are used to describe the H-tree geometrically (refer to figure 3.2):

1.  $L$  is the length of the square die side.
2.  $n$  is the clock tree expansion number, such that an end point always covers a square area.
3. An end point refers to the termination of the clock tree, where a signal is then supplied to a local clock grid.  $EP$  refers to the number of end points.
4.  $D_H$  refers to the classical depth of an H-tree. This is used in other works as the branching level. Typically, if  $D_H = 0$ , only one end point exists. When  $D_H = 1$ , the

previous end point serves two new end points and when  $D_H = 2$  there are four end points. Note that  $EP = 2^{D_H}$  but  $EP = 2^{2n}$ .

Using the above definitions as guidelines, it is easy to derive the equations in table 3.2.1 with reference to figure 3.2.

TABLE 3.1: H-tree equations regarding lengths and segments

$L_{total} = \frac{3}{4}L \sum_{k=1}^n 2^k$	Total H-tree length
$L_n = \frac{3}{4}L \times 2^n$	Length contributed at level $n$
$\frac{L}{2^{(n+1)}}$	Segment length at level $n$
$\frac{3}{2} \times 2^{2n}$	Number of segments at level $n$
$\frac{L^2}{2^{2n}}$	Area of local region at level $n$

When utilising repeaters, each section is driven by a buffer with a fan-out of branches, and terminated in another buffer with the same fan-out. The exception is the end-points, which terminates in a buffer designed to drive the local clock grid. If the segment length reaches a critical length as determined in section 3.5.1, a repeater must be inserted. Alternatively, the segments can be designed in terms of width to control this critical length by sizing the wires in order to produce a rise- or fall time flank to suit the desired specification.

### 3.2.2 Local region network

A global network, such as described in 3.2.1, is terminated in end points which feeds a local region of registers. The depth of an H-tree can be determined through the constraints on local clock skew and the slope of clock signal transition. Figure 3.3 shows a possible implementation of a local network, where each register is connected to the local buffer through an intermediate layer interconnect.

It is possible to derive the amount of intermediate wiring necessary for this and most other local grid implementations if the register density is known. Suppose the local area square in figure 3.3 has a side dimension of  $l$  and the amount of registers is  $N_{reg}$ , then, under the assumption of uniform register distribution, the length of the local wiring will be

$$L_{local} = \left( \sqrt{N_{reg}} + 1 \right) \times l \quad (3.1)$$

If the transistor density,  $D_{xtor}$  is known, a model of the number of local registers can be represented with the following assumptions:

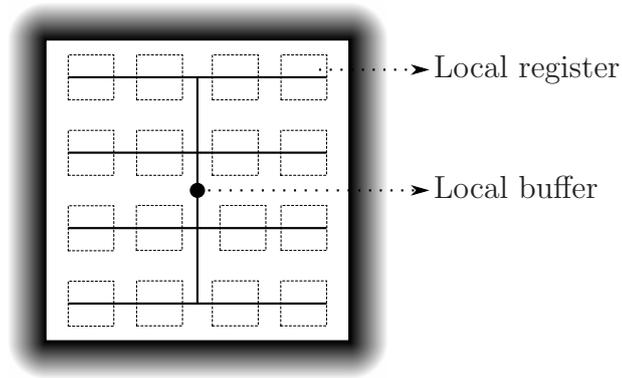


FIGURE 3.3: A typical implementation of the local region grid

- There are 25 transistors per gate, typical to D-type flip flop configurations.
- There are 64 gates per register, representing a register in modern 64-bit architectures.
- The clock signal is fed to one minimum sized inverter per gate, representing the load capacitance  $C_{load_{reg}}$

Therefore, the number of transistors can be computed as

$$\#xtor = 25 \times 64 = 1600 \quad (3.2)$$

It follows then that, using equation 3.3, the amount of registers per local area side length,  $l$ , can be determined.

$$N_{reg} = \frac{D_{xtor}}{1600} \times l^2 \quad (3.3)$$

### 3.3 THE INTERCONNECT

Modern ASIC and MPU integrated circuits based on synchronous logic require a common timing component, mostly in the form of a global clock signal. Data signals also need to be routed to nearby circuitry and bus connections carry information between different parts of the IC. Since logic density is increasing as per Moore's law, the interconnect density also need to increase. Modern VLSI processes can have more than 10 vertical levels of interconnects, whereas the wiring pitch for the lowest level metal is approaching 130 nm [1]. Expanding in the vertical direction leads to mechanical stability problems and an increase in process complexity, whereas reducing the pitch and minimum line width impacts the coupling capacitance and cross-talk.

### 3.3.1 Electrical interconnect model

Signals travelling across interconnects are affected by four electrical characteristics: resistance, capacitance, inductance and inter level conduction along the interconnect. Usually the conduction leakage is neglected, since the dielectric between metal layers is usually made of  $\text{SiO}_2$  or another good isolating alternative. As the clock frequency increases, the inductance becomes more prominent. For the calculation of delay, skew and power consumption, it is assumed that the mutual inductance is negligible, although this might not be true for technology nodes approaching 11 nm. Self inductance will also not be considered since the segment lengths are sufficiently smaller than the operating wavelength [27]. For the purpose of this study, the capacitive and resistive components will be investigated as the main determining parameters of signal delay and slope degradation.

Interconnect density also affects the cross talk behaviour of adjacent lines. One way to reduce this is to shield the active routing layers vertically using ground plane grids, while also reducing lateral cross talk by placing grounded wires as current return paths surrounding the active wires. This will also be the assumption on the models used for interconnects in this study.

#### 3.3.1.1 Capacitance

On the assumption of a rectangular cross section for the interconnect segment, electric field lines will form between the interconnect and the ground plane underneath the segment when there exists a potential difference. The field lines originate orthogonally from all boundaries of the interconnect, and terminate orthogonally in the ground plane. This means that only the bottom boundary of the line will adhere to standard parallel plate capacitance. The field lines

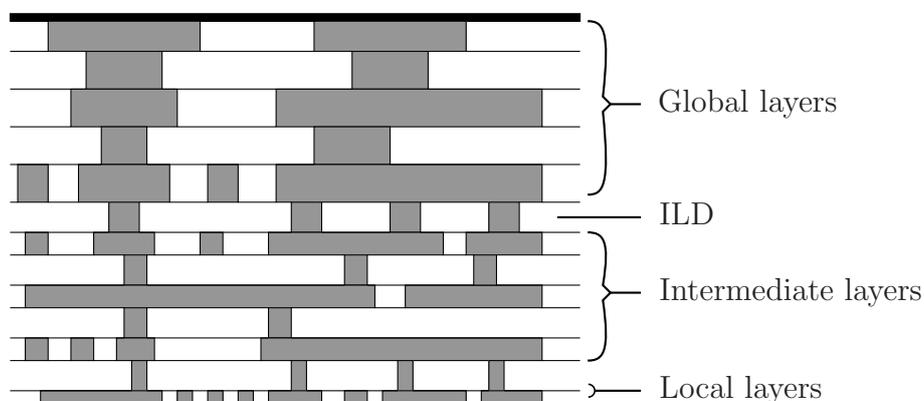


FIGURE 3.4: A reduced modern interconnect stack adapted from [1]

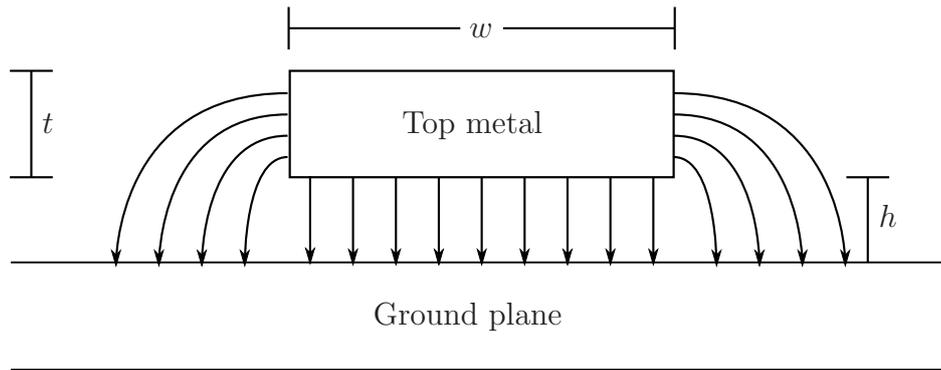


FIGURE 3.5: Cross section of the interconnect physical model

originating in the side wall and the top boundary will contribute to a fringing capacitance. In larger technologies, the latter component is usually lumped into one term irrespective of line width, under the assumption that the top boundary contribution is minimal. This can be done when the width of the line is much larger than the line thickness. In more modern technologies, this assumption tends to fall away, complicating the analysis. A general equation is presented in [27] whereby an approximation can be made to incorporate the fringing effects into one capacitance term. This requires that the line be isolated above a ground plane, and that the permittivity remains constant, which might not be the case for low- $\kappa$  processes.

$$C = \epsilon \left[ \frac{w - \frac{t}{2}}{h} + \frac{2\pi}{\ln \left( 1 + \frac{2h}{t} \sqrt{\frac{2h}{t} \left( \frac{2h}{t} + 2 \right)} \right)} \right] \text{ for } w \geq \frac{t}{2} \quad (3.4)$$

$$C = \epsilon \left[ \frac{w}{h} + \frac{\pi \left( 1 - 0.0543 \times \frac{t}{2h} \right)}{\ln \left( 1 + \frac{2h}{t} \sqrt{\frac{2h}{t} \left( \frac{2h}{t} + 2 \right)} \right)} + 1.47 \right] \text{ for } w < \frac{t}{2} \quad (3.5)$$

Where multi-layered dielectrics are used between conductors, an effective permittivity can be derived based upon the individual layer thicknesses.

Although the model by [27] approximates the capacitance of a top metal layer without surrounding conductors, more densely packed interconnects can often have dominant capacitive component due to the adjacent ground lines. A better model for both the topmost and lower level interconnects is described by [28].

The capacitance components relevant to figure 3.6 are shown by [28] to be as described in equations 3.6, 3.7 and 3.7, where  $\epsilon = 3.9$  is the effective relative permittivity of the

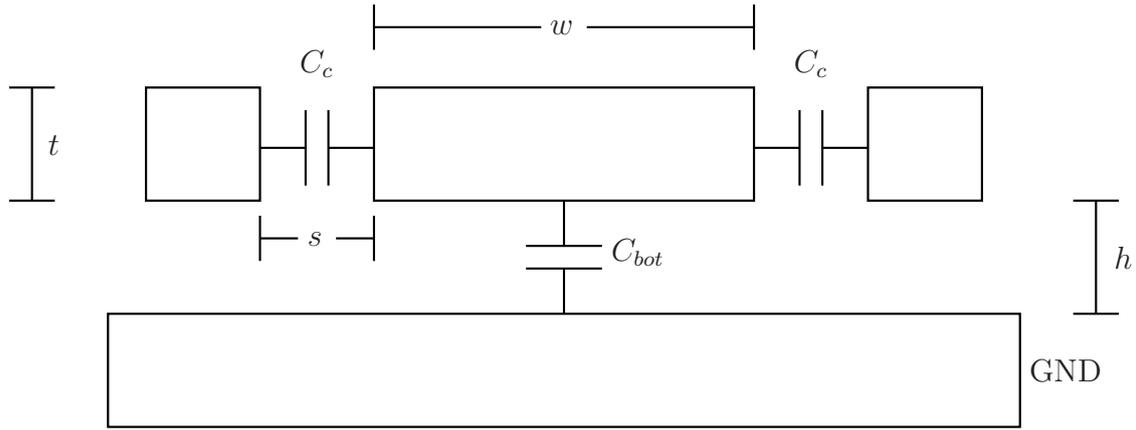


FIGURE 3.6: Interconnect model for top global metal

surrounding interlayer dielectric of  $\text{SiO}_2$  and lower for modern low- $\kappa$  materials.

$$C_{bot} = \epsilon \left[ \frac{w}{h} + 2.217 \left( \frac{s}{s+0.702h} \right)^{3.139} + 1.171 \left( \frac{s}{s+1.510h} \right)^{0.7642} \cdot \left( \frac{t}{t+4.532h} \right)^{0.1204} \right] \quad (3.6)$$

$$C_c = \epsilon \left[ 1.144 \frac{t}{s} \left( \frac{h}{h+2.059s} \right)^{0.0944} + 0.7428 \left( \frac{w}{w+1.592s} \right)^{1.144} + 1.158 \left( \frac{w}{w+1.874s} \right)^{0.1612} \cdot \left( \frac{h}{h+0.9081s} \right)^{1.179} \right]$$

$$C_{total} = 2 \times C_c + C_{bot} \quad (3.7)$$

Lower level global layers, intermediate and local routing layers need to be modelled as shown in figure 3.7, where the capacitance equations can be approximated as shown in equations 3.8, 3.9 and 3.10.

$$C_{top} = C_{bot} = \epsilon \left[ \frac{w}{h} + 2.04 \left( \frac{t}{t+4.5311h} \right)^{0.071} \cdot \left( \frac{s}{s+0.5355h} \right)^{1.773} \right] \quad (3.8)$$

assuming  $h_1 = h_2$

$$C_c = \epsilon \left[ 1.4116 \frac{t}{s} e^{-\frac{4s}{s+8.014h}} + 2.37 \left( \frac{w}{w+0.3078s} \right)^{0.257} \cdot \left( \frac{h}{h+8.961s} \right)^{0.757} \cdot e^{-\frac{2s}{s+6h}} \right] \quad (3.9)$$

assuming  $h_1 = h_2$

$$C_{total} = 2 \times C_c + 2 \times C_{top/bot} \quad (3.10)$$

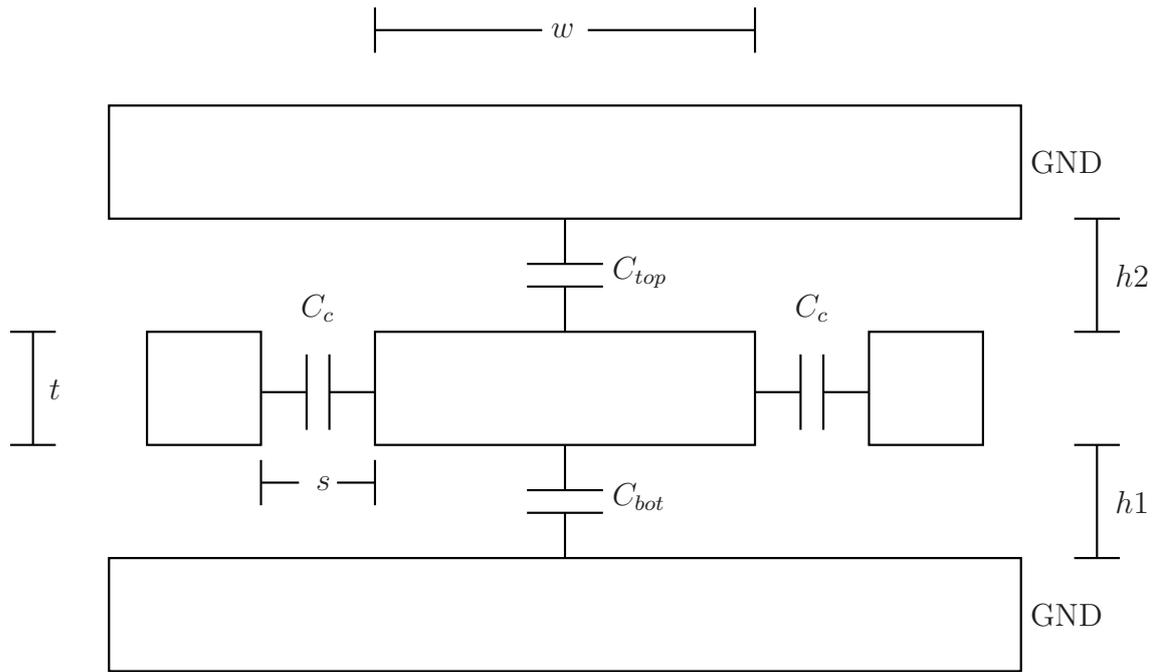


FIGURE 3.7: Interconnect model with top and bottom plates

The assumption that  $h1 = h2$  will mostly be valid, since it is clear that the value of the total capacitance mainly dependent on the wire pitch. One again,  $\epsilon$  refers to the effective surrounding dielectric constant to the ILD.

### 3.3.1.2 Skin effect and effective resistance

The skin effect is a well-known phenomenon, where alternating current flowing in a conductor tends to increase in density near the surface, or skin, of the conductor. This results in a reduced effective cross-sectional area, depending on the skin depth. A shallower skin depth, resulting from an increase in frequency, gives rise to an increase in resistance, where [29] shows that equation 3.11 can be used to calculate the frequency dependent increase.

$$R_{int}(\omega) = \begin{cases} R_{ac} = \frac{1}{\sigma \delta (1 - \exp(-\frac{t}{\delta})) (w+t)}, & \text{if } R_{ac} > R_{dc} \\ R_{dc} = \frac{1}{\sigma w t}, & \text{otherwise} \end{cases} \quad (3.11)$$

The skin depth is a function of frequency and can be calculated [30] using equation 3.12.

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \quad (3.12)$$

As technology nodes scale down and device operating frequencies increase, the skin effect will definitely affect the behaviour of high speed interconnects. An upper limit is also placed on the width and thickness of interconnect wires for reducing series resistance, since increasing the wire dimensions well beyond the skin depth will not decrease the effective series resistance. Figure 3.8 shows the skin depth, where current density is roughly  $e^{-1}$  of the surface density.



FIGURE 3.8: Illustration of current density and skin depth

### 3.3.2 Optical interconnects

As with electrical interconnects, optical interconnects will need to carry timing information to different parts of an IC. The difference: it has to carry light, not electrical current. The manner of implementation can vary and, since optical clock distribution is not a mainstream technology (yet), a proven, compatible cost effective implementation has yet to be developed. Common, however, to all waveguides are

- a medium for guiding the light with associated material properties, and
- the optical power losses involved in coupling and propagation through the waveguide.

Light propagation through a medium is dependent on the medium refractive index, defined in equation 3.13, where  $c$  is the speed of light in vacuum and  $v$  is the speed of light in the material under investigation.

$$n = \frac{c_0}{v} \quad (3.13)$$

Thus, if a material's refractive index is known, the speed of optical propagation can be calculated.

### 3.3.2.1 Total internal reflection

The basics of waveguides revolve around a reflection principle explained by Snell's Law in equation 3.14.

$$\begin{aligned}\frac{\sin \theta_1}{\sin \theta_2} &= \frac{n_2}{n_1} \\ \theta_{crit} &= \arcsin \left( \frac{n_2}{n_1} \right)\end{aligned}\quad (3.14)$$

### 3.3.2.2 Absorption

Absorption is one kind of loss encountered in optical waveguides. It is usually expressed in terms of a loss quantity as dB/cm, which means a certain amount of optical power is lost per unit length of waveguide.

### 3.3.2.3 Dispersion

Dispersion, more specifically chromatic dispersion, can also impact the performance of an optical waveguide. Different wavelengths travel at different phase velocities through an optical medium. If the signal is not of a narrow band nature, this might result in spread, effectively distorting the transition flank of a pulse. In optical clock networks, this might affect the performance if the signal slope becomes comparable to the clock period.

### 3.3.2.4 Coupling losses

As briefly discussed in section 2.6, coupling loss is another loss mechanism important in the design of optical clock networks. If excessive power is lost due to coupling, the optical signal might not be detectable by the optical end point amplifiers.

## 3.4 SKEW IN CLOCK NETWORKS

Synchronous digital systems are based on the principle that all circuits change state at exactly the same time, at certain discrete points in time. As previously stated, this common timing signal is called a clock signal. In modern VLSI circuits, secondary timing signal might be derived from the clock, but synchronisation is key.

A problem arises in systems when the clock signal in different parts of a circuit becomes unsynchronised, that is, when the same clock signal is not aligned.

Figure 3.9 shows an example of skew, where  $l_1$  and  $l_2$  might represent different length lines or lines of different electrical characteristics. Note that any mismatch in terms of

components and component characteristics on the respective lines can also contribute to skew. The misalignment is represented by  $\Delta t$ , which is a time difference between the arrival of the signal at different circuit locations.

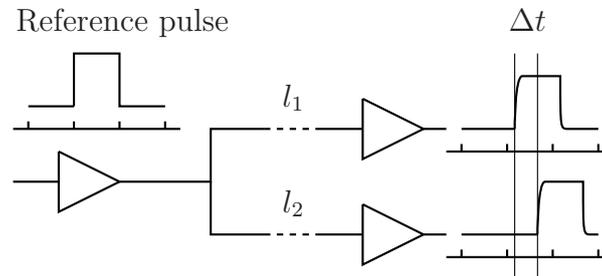


FIGURE 3.9: An example of skew

Note that this is the case for both optical and electrical clock networks, although the interconnect types and the boosting circuitry might differ.

### 3.4.1 Sources of skew in electrical clock networks

The reason the balanced H-tree is an attractive architecture for global clock distribution is the theoretical symmetric propagation of the signals from a common point. Ideally, all signals generated by the clock source will travel through exactly the same interconnect length, the same number of buffers and therefore the signal arrival at the end points will coincide. Unfortunately process variations and environmental fluctuations will cause skew along the individual paths.

To summarise, skew in an electrical clock network revolves around differences in

- interconnects,
- the buffers boosting the signal, and
- the loads whereupon the clock signals terminate.

A study by [31] identifies the variation of the following main parameters as contributing to skew, along with compact models for the individual components:

Interconnect effects	
ILD thickness	$\left  \frac{\partial T}{\partial h_{ILD}} \right $
Wire thickness	$\left  \frac{\partial T}{\partial t_{int}} \right $
Component effects	
Threshold voltage	$\left  \frac{\partial T}{\partial V_{th}} \right $
Channel length	$\left  \frac{\partial T}{\partial l_{ch}} \right $
Gate oxide thickness	$\left  \frac{\partial T}{\partial C_{OX}} \right $
System-wide effects	
Power supply fluctuations	$\left  \frac{\partial T}{\partial V_{DD}} \right $
Non-uniform register distribution	$\left  \frac{\partial T}{\partial C_L} \right $

### 3.4.2 Sources of skew in optical clock networks

Optical networks make use of a light source for the generation of the clock signal. The signal must then be distributed across chip by means of an optical transmission medium. At some point, since digital circuits still operate electrically, the signal needs to be converted into an electrical signal containing the timing information to serve as a clock.

Unlike electrical systems, the field of optical clock distribution is at an infancy level, and research on the appropriate light source and wave-guiding structures are still inconclusive. Common to all optical networks will be the waveguide's refractive index and losses, as well as the receiver's ability to accurately convert the optical signal into an electrical representation.

To summarize, skew in an optical clock network will therefore be dependent on

- the optical transmission media, or waveguides,
- the optical receiver converting and boosting the signal for local region use, and
- the loads whereupon the clock signals terminate.

Since optical clock networks are not a matured technology, no standard method of implementation exist. However, all optical media does have certain common characteristics, also in contributing to skew. If driven from single common optical source and the local regions remain identical to the electrical variant, the differences are in the interconnects and the specific receiving circuitry.

Interconnect effects	
Refractive index	$\left  \frac{\partial T}{\partial n_{wg}} \right $
Component effects	
Threshold voltage	$\left  \frac{\partial T}{\partial V_{th}} \right $
Channel length	$\left  \frac{\partial T}{\partial l_{ch}} \right $
Gate oxide thickness	$\left  \frac{\partial T}{\partial C_{OX}} \right $
System-wide effects	
Power supply fluctuations	$\left  \frac{\partial T}{\partial V_{DD}} \right $
Non-uniform register distribution	$\left  \frac{\partial T}{\partial C_L} \right $

### 3.5 SIGNAL PROPAGATION OVER AN INTERCONNECT

As a signal propagates across an interconnect, signal fidelity can be affected in a number of ways. This depends on the medium of transmission, as well as the conditions on the terminating point. Most often, the signal integrity worsens as line length increases, therefore, signal fidelity can be used for line length selection.

#### 3.5.1 Propagation across an electrical interconnect

If taking into account only the resistive and capacitive interconnect components, the signal will degrade in terms of timing according to [32]. This is especially important regarding the rise and fall times of the clock signal. As in [26], the rise and fall time is defined as the 20 % - 80 % rise time, or the 80 % - 20 % fall time, depending on the worst case scenario. [32] gives the following formula to determine signal transition degradation as a function of time:

$$\frac{t_v}{RC} = 0.1 + \ln\left(\frac{1}{1-v}\right)(R_T C_T + R_T + C_T + 0.4) \quad (3.15)$$

where the parameters are described in the following table.

$r_{int}$	Interconnect resistance per unit length
$c_{int}$	Interconnect capacitance per unit length
$l$	Length in $\mu\text{m}$
$R = r_{int}l$	Segment resistance
$C = c_{int}l$	Segment capacitance
$R_B$	Buffer output resistance driving the segment
$C_B$	Capacitance terminating the segment
$R_T = \frac{R_B}{r_{int}l}$	Normalized buffer output resistance
$C_T = \frac{C_B}{c_{int}l}$	Normalized buffer input capacitance

Equation 3.15 can be rewritten more specifically for  $t_{28}$  as

$$t_{28} = (r_{int}c_{int}l^2)(0.5545 + 1.386(\frac{R_B C_B}{r_{int}c_{int}l^2} + \frac{R_B}{r_{int}l} + \frac{C_B}{c_{int}l})) \quad (3.16)$$

If the requirement is set on  $t_{28}$  such that it remains ten times smaller than the clock period, it can be stated that

$$t_{28} < \frac{1}{10 \times f_{clk}} \quad (3.17)$$

Combining 3.16 and 3.17 results in equation 3.18.

$$\begin{aligned} a &= 0.5545r_{int}c_{int} \\ b &= 1.386(r_{int}C_B + c_{int}R_B) \\ c &= 1.386R_B C_B - \frac{1}{10 \times f_{clk}} \\ l &< \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \end{aligned} \quad (3.18)$$

Equation 3.18 is symmetrical in terms of rise and fall times, assuming that the buffers are designed for symmetrical operation. It must be said that  $r_{int}$  will be a function of frequency as investigated in section 3.3.1.2.

### 3.5.2 Propagation through an optical waveguide

Analogous to section 3.5.1, an optical signal flank might also suffer degradation due to dispersion (see section 3.3.2.3). The assumption for this work is that in an optical system, the flank degradation is negligible, which might not be true for a wide band optical signal.

## 3.6 POWER DISSIPATION IN ELECTRICAL CLOCK DISTRIBUTION NETWORKS

### 3.6.1 Power dissipation mechanisms

Power losses occurring in the distribution of clock signals can amount to more than 50 % [33], [34] of overall power consumed in modern day microprocessors. The main component of power loss arises from the need to switch extremely large capacitances between the supply rails when distributing a global clock signal. As more logic is packed onto a single die, clock lines increase rapidly in terms of length. Scaling also affects the capacitance as the oxide thickness between metal layers decreases. Low- $\kappa$  dielectrics, transmission line methods [35] and the use of unscaled interconnect metal layers help to alleviate increased power consumption due to scaling. Three mechanisms are directly involved and consumes power when distributing clock signals:

1.  $C \times V_{dd}^2 \times f_{clk}$  capacitive power losses
2. Leakage currents from devices in steady state
3. Short circuit currents in repeaters and buffers along the signal paths when transitions occur

The assumption is that the first term is the dominant term, while the second term is negligible. Of course, the third term is difficult to model analytically, and will be accounted for by using SPICE simulations, as discussed in chapters 4 and 5.

# CHAPTER FOUR

## POWER CONSUMPTION IN FUTURE ELECTRICAL NETWORKS

---

### 4.1 SUMMARY OF FUTURE TECHNOLOGY NODES

As a source for an assessment on what the requirements would be for future CMOS technology nodes, the predictions given by the International Technology Roadmap for Semiconductors 2008 updated edition [2] are used to construct a predictive platform. The predictions are the product of an international collaborative effort for strategic planning. Some predictions are borrowed from older sources such as [36] when such information is required. Figure 4.1 is taken from [2] and shows in general the trend in terms of local clock frequency as a function of time, along with predicted technology node scaling.

It is important to note that the ITRS reports on the requirements of future technology, while predictive models used, as discussed in section 4.1.1, aim at predicting actual device behaviour on a physical level. Precedence is given to the latter, where care was taken to adapt the models according to the predictions made by the ITRS.

#### 4.1.1 Specific device models

SPICE modelling forms an integral part in predicting both sub-system as well as overall system behaviour. For this reason, it is important to incorporate current trends with predictive modelling, while necessary to apply this knowledge in the form of operational SPICE models for active devices such as transistors.

Models are based on work done by [37] and are obtained from [38]. Under the assumptions in [39] and [40] regarding new emerging structures for MOS transistors, the

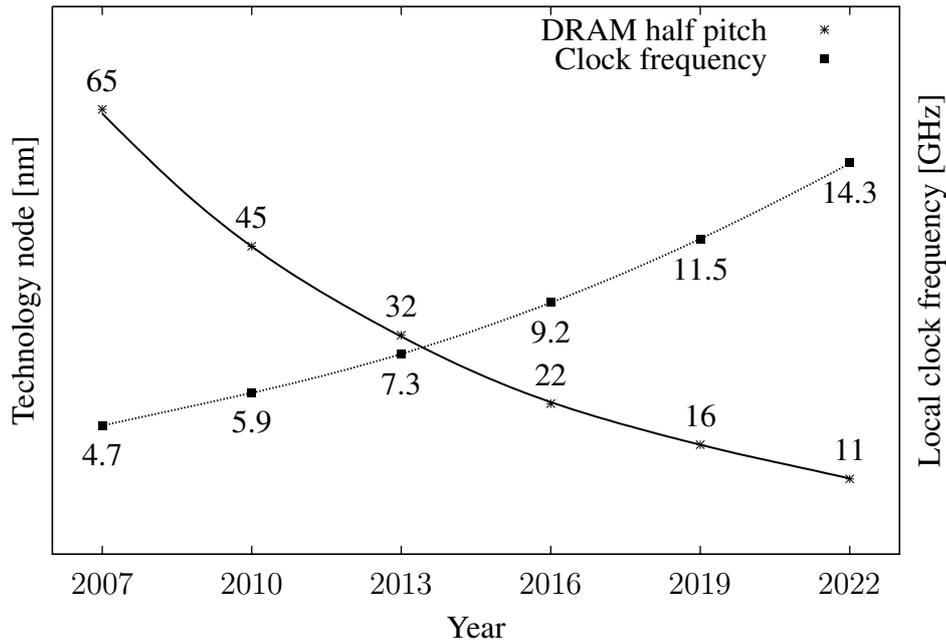


FIGURE 4.1: Future scaling in CMOS technology as predicted by the ITRS 2008 [2]

following two MOS structures are assumed:

- Standard bulk MOS devices for nodes preceding and up to 45 nm
- Fully depleted thin bodied and double gate transistors, such as FinFET, for 32 nm and downward

The predictive transistor models (PTM) also take into account future advances in strained silicon devices for mobility alteration and enhancement, high- $\kappa$  gate dielectrics in order to reduce the effect of tunnelled carriers across the isolator, as well as features such as reintroduction of metal gates.

### 4.1.2 Important technology defining parameters

As pointed out in [37], only about ten parameters are responsible for the dominant changes from one technology to the next. The models are taken from [38], where the relevant parameters are modified in such a way that the V-I characteristics closely match the predictions as per the current ITRS report. This ensures more accurately adapted SPICE models for this work.

The following parameters were matched to data obtained from the ITRS 2008 PIDS report:  $V_{dd}$ ,  $T_{ox}$ ,  $L_{eff}$ ,  $R_{dsw}$  and  $V_{th}$

The scaling of  $V_{dd}$  and  $V_{th}$  has been historically present throughout CMOS history, altering (usually increasing) the channel doping,  $N_{ch}$  and the corresponding drain and source implant doping from one technology to the next.

Due to the decrease in size of both the gate oxide and the channel length, the two parameters that need extra attention in nanometer technologies are  $T_{ox}$  and  $L_{eff}$ .

#### 4.1.2.1 $T_{ox}$ in nanometer devices

Quantisation effects are known to form a finite spatial charge distribution below the gate dielectric interface to the channel [41]. In compact models, this can be contained as a capacitance in series with the conventional gate dielectric capacitance, resulting in equation 4.1. Note that modern gate dielectric stacks [42], such as Hf-based solutions, are modelled as an effective  $\text{SiO}_2$  material for the capacitive component. The gate leakage is handled separately.

$$C_{oxeff} = \frac{C_{ox}C_{charge}}{C_{ox} + C_{charge}} \quad (4.1)$$

This can be implemented in simulators, such as the BSIM4 model as used in this research, by defining two of the following: an equivalent physical, electrical or difference gate oxide thickness. The BSIM4 model uses TOXP, TOXE and DTOX for the above three quantities respectively.

The ITRS specifies both the physical and electrical equivalent oxide thicknesses, therefore enough information is available to include as augmentation to the predictive SPICE models.

#### 4.1.2.2 $L_{eff}$ in nanometer devices

A circuit designer is responsible for drawing a certain gate dimension,  $L_{drawn}$ , as part of structuring a device for a specified behaviour. Manufacturing aspects often tend to reduce this by a certain offset, resulting in an effective gate length smaller than what the circuit designer intended. Figure 4.2 shows the discrepancy between the drawn and effective dimension.

This offset is critical when modelling device behaviour. As devices get smaller, the offset becomes larger in proportion to the drawn length, and including this effect results in a more pronounced influence on the devices' electrical characteristics.

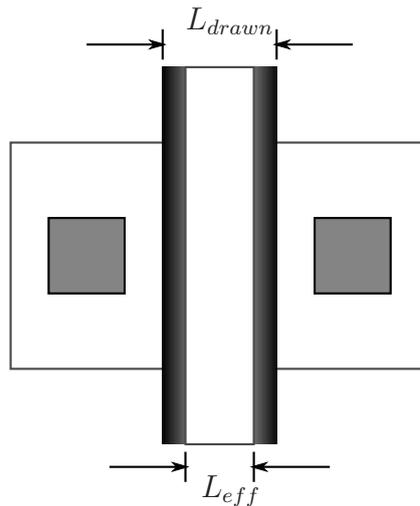


FIGURE 4.2: Effective gate length as compared to a drawn transistor

### 4.1.3 Summary of device parameters

TABLE 4.1: ITRS 2008 requirements for high performance logic devices

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$V_{DD}$ [V]	1.1	1.1	1.0	0.9	0.9	0.8
$V_{th}$ [mV]	225	175	103	105	109	109
$L_{eff}$ [nm]	32	24	18	14	10.7	8.1
$R_{dsw}$ [ $\Omega$ ]	200	200	180	180	160	150
$T_{oxp}$ [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

A summary of the parameters taken from the ITRS 2008 Process integration, Devices and Structures (PIDS) report is summarised in table 4.1 from [43]. These values are directly used in the predictive SPICE models used for this work.

### 4.1.4 Model SPICE curves

To indicate the correlation between predictions from ITRS and the custom developed SPICE models, the simulated results are compared and tabulated in table 4.2. Figure 4.3 shows the drain current, with the gate at the technology's  $V_{DD}$ , as a function of the drain-source voltage.

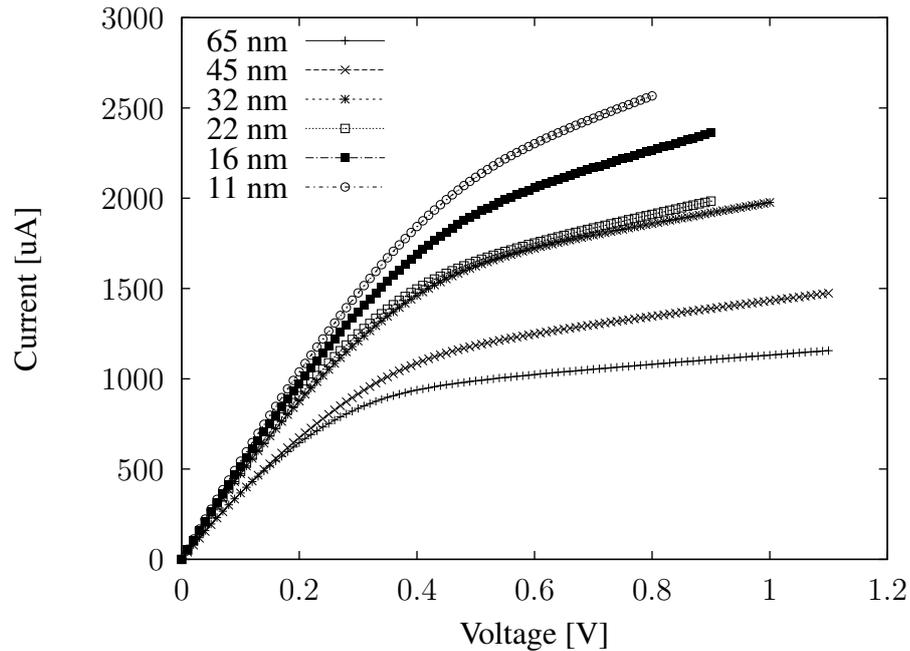


FIGURE 4.3: NMOS  $V - I$  characteristics for future technologies, gate at  $V_{DD}$ ,  $1 \mu\text{m}$  width

TABLE 4.2: Comparison between SPICE and ITRS predicted values

Node	SPICE [ $\mu\text{A}/\mu\text{m}$ ]	ITRS 2008 [ $\mu\text{A}/\mu\text{m}$ ]	% Error	Device type
65 nm	1156	1006	+14.9	Bulk
45 nm	1471	1370	+7.37	Bulk
32 nm	1973	1948	+1.28	UTB FD/SOI
22 nm	1980	1943	+1.90	DG FinFET
16 nm	2372	2344	+1.19	DG FinFET
11 nm	2574	2533	+1.62	DG FinFET

With the exception of the current 65 nm node, the future predicted SPICE models all fit the ITRS values for saturation current per unit width well.  $CV/I$  modelling is dominantly dependant on the gate oxide thickness, which serves as input to the SPICE models from ITRS data, thus the capacitance modelling are assumed accurate.

#### 4.1.5 Interconnect parameters

According to the ITRS 2008 Interconnect report [44], interconnect dimensions in future technology nodes will need to continue scaling down due to the increasing active device density. Derivations of the resistive and capacitive interconnect components used in this

TABLE 4.3: Intermediate interconnect electrical characteristics

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
Pitch [nm]	136	90	64	44	32	22
Aspect ratio	1.8	1.8	1.9	2	2	2.1
Capacitance [pF/cm]	1.8	1.6	1.5	1.3	1.3	1.1
Resistivity [ $\mu\Omega\cdot\text{cm}$ ]	3.43	4.08	4.83	6.01	7.34	9.84
Width [nm]	68	45	32	22	16	11
Height [nm]	122	81	61	44	32	23
Cap. per length [F/ $\mu\text{m}$ ]	1.8E-16	1.6E-16	1.5E-16	1.3E-16	1.3E-16	1.1E-16
Res. per length [ $\Omega/\mu\text{m}$ ]	1.2	2.74	5.14	10.33	19.53	39.35

TABLE 4.4: Global interconnect electrical characteristics

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
Pitch [nm]	210	135	96	66	48	33
Aspect ratio	2.3	2.4	2.5	2.6	2.8	2.9
Capacitance [pF/cm]	2	1.8	1.7	1.5	1.5	1.3
Resistivity [ $\mu\Omega\cdot\text{cm}$ ]	2.73	3.10	3.52	4.20	4.92	6.30
Width [nm]	105	68	48	33	24	17
Height [nm]	242	162	120	86	67	48
Cap. per length [F/ $\mu\text{m}$ ]	2E-16	1.8E-16	1.7E-16	1.5E-16	1.5E-16	1.3E-16
Res. per length [ $\Omega/\mu\text{m}$ ]	0.39	0.91	1.74	3.53	6.20	12.67

work are based on predictions made by the ITRS.

## 4.2 CHIP SIZE MODEL

It is important to incorporate the fraction of chip area consumed by logic. Modern high performance MPUs typically contain level 2 cache in the form of SRAM as well, which is not necessarily clocked continuously. Since the scope of this work emphasised the possibility of optical clocks on *high performance* MPUs and ASICs, a model for the typical amount of on chip memory is necessary. If the typical amount of L2 cache in a 65 nm MPU is 4096 kBytes, extrapolation can be done to cover future nodes assuming that the amount of memory roughly doubles between one technology generation to the next.

TABLE 4.5: Information for determining effective logic area

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
Chip size [cm <sup>2</sup> ]	3.1	3.1	3.1	3.1	3.1	3.1
SRAM memory [kBytes]	4096	8192	16384	32768	65536	131072
SRAM transistors [ $M_{xtor}$ ]	201.33	402.65	805.31	1610.61	3221.23	6442.45
SRAM density [ $M_{xtor}/\text{cm}^2$ ]	827	1718	3532	7208	14625	29588
Fraction SRAM [%]	7.85	7.65	7.35	7.21	7.11	7.02
Effective logic area [cm <sup>2</sup> ]	2.857	2.866	2.872	2.877	2.880	2.882

### 4.3 DETERMINING TREE DEPTH

The depth of the H-tree is critical in determining to what extent the tree can be optical or electrical. The key performance parameter placing a lower limit on this component is the local region clock skew. Since the local regions are electrical in both electrical and optical networks, the minimum tree depth will be equivalent. The local region model is based on the design in section 3.2.2.

With referral to table 3.2.1, the maximum assumed Manhattan-type line length will be  $\frac{L}{2^n}$  set in the intermediate metal layers. The shortest length can be assumed to be close to zero, which represents a driven cell very close to the end point and local buffer. Skew can then be determined by taking the difference between the two delays [31]. Using a distributed RC delay model as given by [31], the 50 % time of flight between point zero and the predicted maximum can be calculated using equation 4.2, where  $l_{line}$  is the interconnect length,  $\epsilon_r$  is the relative permittivity of the ILD and  $c_0$  the speed of light in vacuum.

$$T_{local} = 0.4(r_{int}c_{int}) \cdot l_{line}^2 + \frac{\sqrt{\epsilon_r}}{c_0} \cdot l_{line} \quad (4.2)$$

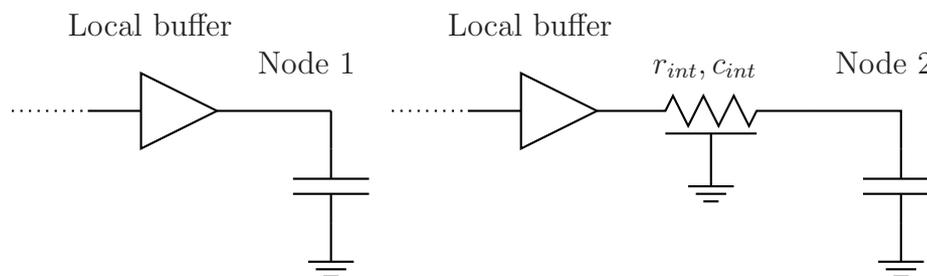


FIGURE 4.4: Local region representation for skew calculations

TABLE 4.6: Requirements for tree depth calculation

$r_{int}$ and $c_{int}$	Interconnect resistance and capacitance
$\epsilon_r$	Intermediate layer dielectric permittivity
$f_{clk}$	Local region clock frequency
$L_{die}$	Complete effective die area consumed by logic

Node 1 in figure 4.4 represents the point where both the closest local registers and the local region clock buffer capacitances are lumped, while node 2 represents the furthest point of interconnect reach where this is terminated in another group of local registers. Since the distribution of registers is random at best, an approximation can be made that roughly half of the local region registers capacitances is lumped in node 1, while the other half can be summed as a capacitance at node 2.

The tree depth can then be designed to allow a local skew of a tenth of the total clock period, to which most digital circuits should be tolerant.

$$\Delta t_{skew} < \frac{1}{10} \times T_{clk} \quad (4.3)$$

For calculating a specific technology tree depth, table 4.6 shows the parameters required. Based on equations 4.2 and 4.3 and table 4.6, an expression for the maximum local region dimension, assuming a square, is shown in equation

$$\begin{aligned} l &= \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a} \\ a &= 0.4(r_{int}c_{int}) \\ b &= \frac{\sqrt{\epsilon_r}}{c_0} \\ c &= -\frac{1}{10 \cdot f_{clk}} \end{aligned} \quad (4.4)$$

The number of end points can be determined using equation 4.5.

$$\#EP = \left( \frac{l_{die}}{l_{local}} \right)^2 \quad (4.5)$$

Using the number of end points, the tree depth can be calculated using the definitions as given in 3.2.1. Since classical tree depth  $D_H$  is an integer quantity and the expansion depth  $n$  is always twice  $D_H$ , the minimum depth will be determined by the first integer  $n$ . Up to this point, both electrical and optical clock networks is effectively equal, where the local region represents a lumped capacitance to a buffer or driver circuit. Note that, due to the base 2 logarithmic nature of the H-tree partitioning, there is a discrepancy between the required end points and the actual implementable number of end points.

TABLE 4.7: Tree depth metrics associated with technology nodes

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$10 \% \times T_{clk}$ [ps]	21.3	17.0	13.6	10.9	8.7	6.97
Allowable $l_{local}$ [ $\mu\text{m}$ ]	464.2	296.4	201.9	138.0	90.3	62.2
Required end points	1326	3262	7044	15110	35286	74473
$n$	6	6	7	7	8	9
$D_H$	12	12	14	14	16	18
Actual $l_{local}$ [ $\mu\text{m}$ ]	264	265	132	133	66	33
Actual end points	4096	4096	16384	16384	65536	262144

#### 4.4 MODELLING LOCAL NETWORK LOAD

Both intermediate interconnects, as well as the nodes on which the clock conductors terminate, contribute towards clock power consumption. The local network can essentially be modelled as a lumped capacitance, since the charged and discharged values must swing between ground and  $V_{DD}$ . Local region area is determined by taking the maximum local segment length from equation 4.4 as the side dimension of a local region,  $l_{local}$ . Only the intermediate layer interconnect capacitance is taken into account, since the terminating register capacitance is a strong function of the circuit itself. If this is the case, then the transistor density  $D_{xtor}$  along with the local region dimensions can be used to estimate a load capacitance. Table 4.8 summarises the local area dimensions, as obtained through equation 4.4, along with the expected transistor density for a high performance MPU as per [45]. Since equations 3.1 and 3.3 combine to give the required intermediate interconnect length, table 4.8 also shows the resulting interconnect length  $L_{local}$  and the amount of registers per local area,  $N_{reg}$ .

TABLE 4.8: Estimation of local region capacitances for future technology nodes

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$l_{local}$ [ $\mu\text{m}$ ]	264	265	132	133	66	33
$D_{xtor}$ [ $\mu\text{m}^{-2}$ ]	3.57	7.14	14.27	28.54	57.08	114.16
$N_{reg}$	156	313	155	316	155	78
$L_{local}$ [ $\mu\text{m}$ ]	3.56E3	4.95E3	1.78E3	2.50E3	8.88E2	3.24E2
Capacitance [F/ $\mu\text{m}$ ]	See section 4.3					
$C_{local}$ [F]	6.41E-13	7.93E-13	2.66E-13	3.25E-13	1.15E-13	3.57E-14

## 4.5 REPEATER DESIGN

This power dissipation model is based on utilising a driving buffer at least at each point where the H-tree branches, as well at the start of each interconnect line segment. Line capacitances and resistances are used as given in section 4.1. As the line segment increases, the 20 % - 80 % rise time metric is kept below 1/10 of the clock period in order to keep signal fidelity intact, as explained in section 3.5.1. As the line length increases, repeaters are inserted in order to compensate for the propagation distortion.

Table 4.9 shows the required technology parameters used in the design of repeater circuits. At all times, the repeater circuits are constrained to a number of two stages, or four transistors. It is possible, based on approximation, to assign two key parameters to each

TABLE 4.9: Parameters required for repeater design

$I_{Dsat}$	NMOS saturation (on) current per unit width
$C_g$	MOS gate capacitance per unit width
$V_t$	Threshold voltage
$V_{DD}$	Power supply applicable to technology
$R_{int}$	Interconnect resistance per unit length
$C_{int}$	Interconnect capacitance per unit length

specific technology. Although both the NMOS and PMOS transistors traverse through the linear and saturated regions, linearising

1. the average repeater input capacitance, and
2. the average repeater output resistance,

as a function of transistor width, results in these two parameters to be used in a wide range of design spaces. The approach is to reference the width value to the NMOS devices, where the PMOS devices are sized accordingly for each technology. Therefore,

$$W_P = R \times W_N \quad (4.6)$$

where  $R$  is a ratio specific to each technology, indicating the relative driving strength of the PMOS devices to that of the NMOS devices.

The ITRS provides required values for  $I_{Dsat}$ , the saturation current per unit width. An exact output resistance calculation would require more information than provided in the

ITRS predictions [46], although a first iteration design can be based on equation 4.7 [31]. The SPICE models as modified in 4.1.1 are in fact based on the ITRS predictions, but does not correlate exactly with the simplified predictions of the ITRS. These models will take precedence over the ITRS data since the models are more inclusive of various predictions, and with the models, simulations are possible.

$$R_B = \frac{1}{\mu C_{OX} \frac{W}{L} (V_{DD} - V_T)} \quad (4.7)$$

If the assumption is made that the output resistance of a repeater reduces (as approximated by equation 4.7) linearly with the width of the transistors, and the input capacitance increases linearly with transistor width, then the two technology related parameters  $A$  and  $B$  can be defined as in equation 4.8.

$$\begin{aligned} R_B &= \frac{A}{W} & \text{where } A & \text{ in } [\Omega \cdot \mu m] \\ C_B &= B \cdot W & \text{where } B & \text{ in } \left[ \frac{F}{\mu m} \right] \end{aligned} \quad (4.8)$$

Equation 4.7 can be rewritten as showed in equation 4.9 in order to estimate the value of  $R_B$ , although this expression will tend to overestimate the repeater's drive capability due to the simplification assumption (that is,  $V_{DS}$  is small) which is not the case across the entire output voltage range.

$$R_B = \frac{(V_{DD} - V_t)}{2 \cdot I_{Dsat} \cdot W} \quad (4.9)$$

### 4.5.1 Extraction of repeater parameters $A$ and $B$

A SPICE-based extraction method can be used to determine the parameters  $A$  and  $B$  for a specific technology.

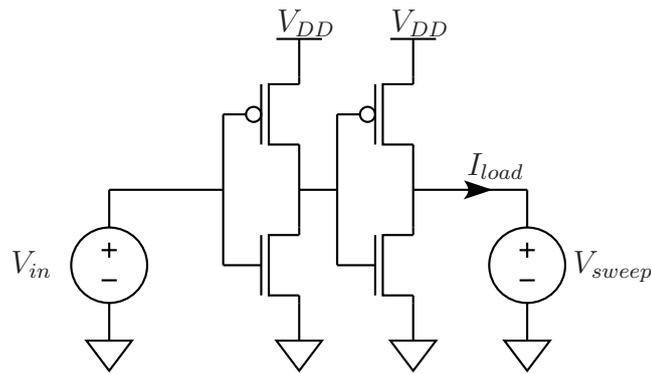
Linear regression is a technique used to fit experimental data to a linear expression. For a given set of  $n$  points  $(x_i, y_i)$ , equation 4.10 shows the calculation for the slope of the linear function  $y = mx + b$ .

$$m = \frac{\sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (4.10)$$

#### 4.5.1.1 Extracting the resistance term

The output resistance data was generated from the DC  $V - I$  characteristics on the output node for a 2-stage buffer. Figure 4.5 shows the set up using a SPICE simulator.

The input voltage, labelled  $V_{in}$ , is set to both 0 and  $V_{DD}$  for each technology. For each input voltage state, the DC source  $V_{sweep}$  is swept from 0 to  $V_{DD}$  particular to

FIGURE 4.5: Extracting the resistance data for the parameter  $A$ 

each technology. With the values of  $V_{sweep}$  and  $I_{load}$  known, the effective average output resistance is simply determined by linearising the results through regression, and obtaining the resistance from the slope of the voltage/current relationship through

$$R = \frac{\partial V_{sweep}}{\partial I_{load}} \quad (4.11)$$

The simulation is done over a wide range of  $W$  in order to obtain a scalable value for  $A$ . An example is shown in figure 4.6.

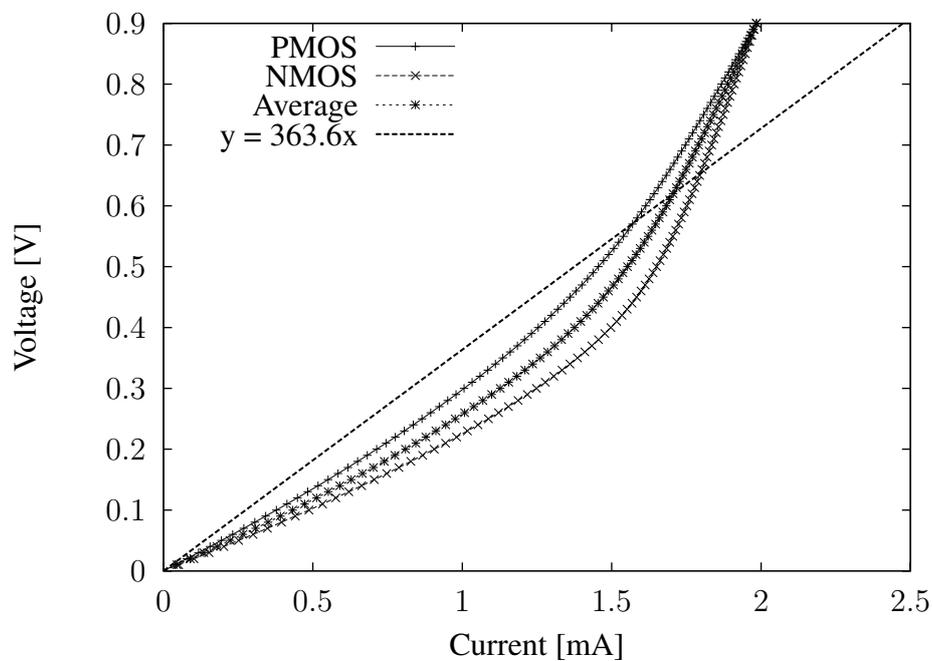


FIGURE 4.6: Example of linearised buffer resistance extraction on 32 nm node

### 4.5.1.2 Extracting the capacitance term

Extracting the capacitance from a SPICE based repeater required the use of a transient analysis. Figure 4.7 shows the set up used, where the method takes the  $RC$  time constant with a known  $R$  value and derives the  $C$  value.

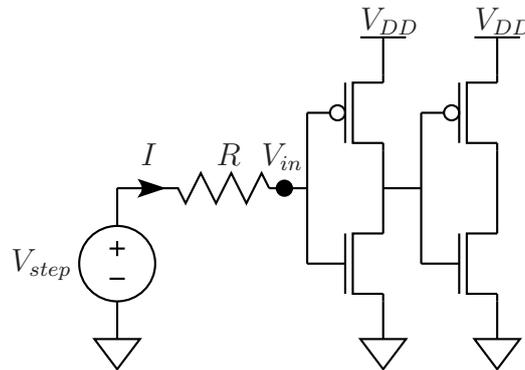


FIGURE 4.7: Extracting the capacitance data for the parameter  $B$

A step is applied through  $V_{step}$  and the resulting repeater input voltage  $V_{in}$  transient values are recorded. A first order  $RC$  equation is linearised (see 4.12) and used in conjunction with the linear regression technique to solve for  $C$ , the input capacitance of the repeater.

$$t = C \times \left( -R \cdot \ln\left(1 - \frac{V_{in}}{V_{DD}}\right) \right) \quad (4.12)$$

### 4.5.1.3 Summary of extracted values

Table 4.10 shows the results of the extraction procedure covering a wide range of values for  $W$ . The results are consistent even up to small values for  $W$ .

TABLE 4.10: Extracted  $A$  and  $B$  values for future nodes based on predictive models

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$A$ [ $\Omega \cdot \mu\text{m}$ ]	695.8	575.65	395.4	362.85	315.65	273.25
$B$ [ $\frac{\text{F}}{\mu\text{m}}$ ]	5.65E-15	4.81E-15	2.88E-15	2.96E-15	2.95E-15	2.91E-15

### 4.5.2 Repeater optimisation

An interesting observation can be made when examining equation 3.18 with the substituted values of  $R_B$  and  $C_B$  as from equation 4.8. The maximum allowed segment length increases as  $W$  increases, up to a certain point, and then starts to decrease again with an increase in  $W$ . This is due to the dependence of the subsequent stage input capacitance on  $W$ . It therefore follows that the maximum segment length can be optimised in terms of the repeater transistor widths, for a given transistor technology and technology interconnect characteristics.

$$\begin{aligned} 0.5545r_{int}c_{int} &= X \\ 1.386(r_{int}B \cdot W + c_{int}\frac{A}{W}) &= Y \cdot W + \frac{Z}{W} \\ 1.386AB - \frac{1}{10 \times f_{clk}} &= Q \end{aligned} \quad (4.13)$$

When taking equation 3.18 and making the substitutions as shown in equation 4.13, the maximum drivable segment length can be rewritten as in equation 4.14.

$$l_{max} = -\frac{Y \cdot W}{2X} - \frac{Z}{2X \cdot W} + \frac{\sqrt{Y^2W^2 + 2YZ + \frac{Z^2}{W^2} - 4XQ}}{2X} \quad (4.14)$$

Upon differentiating equation 4.14, the result shown in 4.15, a local maximum for the value of  $W$  can be obtained.

$$\frac{\partial l_{max}}{\partial W} = -\frac{Y}{2X} + \frac{Z}{2XW^2} \pm \frac{1}{4X} (Y^2W^2 + 2YZ + \frac{Z^2}{W^2} - 4XQ)^{-\frac{1}{2}} \cdot (2Y^2W - \frac{2Z^2}{W^3}) \quad (4.15)$$

By taking

$$\frac{\partial l_{seg}}{\partial W} = 0$$

the interesting result shown in equation 4.16 is obtained.

$$W = \pm \sqrt{\frac{Z}{Y}} \quad (4.16)$$

After substituting the technology parameters back into 4.16 as per 4.13, an elegant optimisation relationship is presented in equation 4.17.

$$W = \pm \sqrt{\frac{c_{int}A}{r_{int}B}} \quad (4.17)$$

This proves that, for a specific front end and interconnect technology, there is an optimum width for a symmetrical two stage buffer which maximises the possible drivable interconnect segment length. Figure 4.8 shows three repeater designs, with the output NMOS device scaled according to the optimal device width.

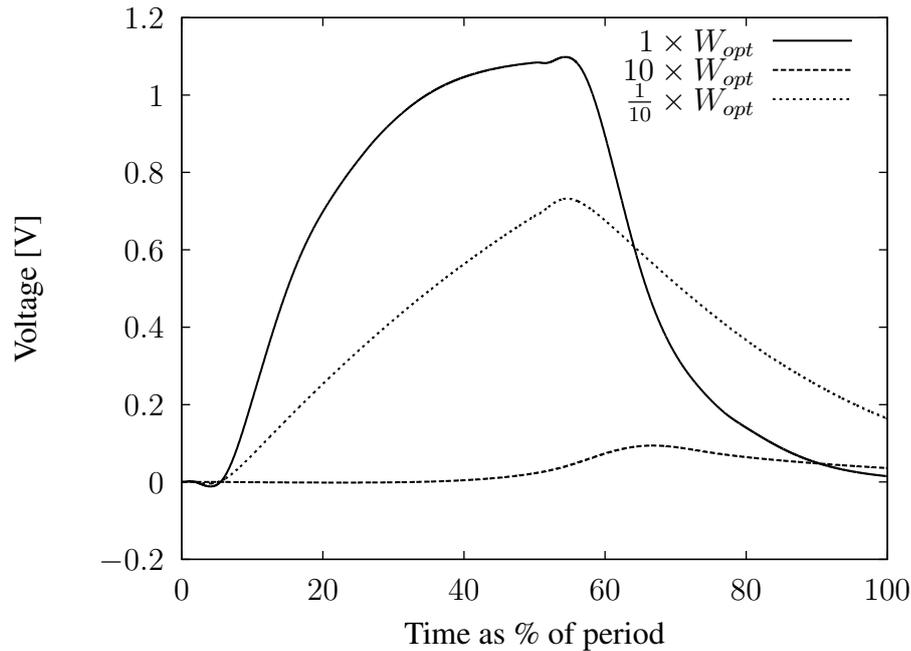


FIGURE 4.8: SPICE simulation of optimal and mismatched buffer widths at the 45 nm node

TABLE 4.11: Repeater design summary

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$W$ [ $\mu\text{m}$ ]	8.25	5.08	3.67	2.31	1.64	0.99
$C_B$ [F]	4.28E-14	2.23E-14	1.06E-14	6.73E-15	4.66E-15	2.82E-15
$R_B$ [ $\Omega$ ]	84	113	108	158	193	275
$l_{seg}$ [ $\mu\text{m}$ ]	276.5	184.4	157.3	99.1	65.6	42.7
NMOS width [ $\mu\text{m}$ ]	8.25	5.08	3.67	2.31	1.64	0.99
PMOS width [ $\mu\text{m}$ ]	17.65	10.67	4.51	3.14	2.29	1.38

Table 4.11 summarises the values used for the design of the repeaters. Since the repeater input MOS inverter pair is replicated in the output inverter pair, only one value is given for the NMOS and PMOS devices.

### 4.5.3 Improved repeater optimisation

While the repeaters in section 4.5.2 show that it is possible to optimise for a specific technology, the repeaters themselves also determine much of the network capability. For instance, by relaxing the criteria of an equivalent input and output MOS inverter pair, the input capacitance can be minimised for a fixed output resistance, as long as the inter-stage signal also maintains the 20 % - 80 % restriction. This allows for less repeaters to be

used. Equation 4.14 shows that, when the width  $W$  in equation 4.13 is split into the capacitance and resistance contribution widths, the maximum segment length obtainable becomes infinite. This, however, implies infinite buffer stages, which in turn means infinite repeater area. When limited to two MOS inverter pairs, a reasonable optimisation is made. By taking equation 4.13 and substituting the capacitance  $W$  term with a scaled version,  $r_{int}B \cdot W = r_{int}B \cdot R \cdot W$ , equation 4.17 becomes equation 4.18.

$$W = \pm \sqrt{\frac{c_{int}A}{r_{int}RB}} \quad (4.18)$$

$R$  now represents the allowed scaling factor of the first stage, in comparison to that of the second stage.  $R$  can be determined by limiting the allowed 20 % - 80 % time to 10 % of the clock period, as shown in equation 4.19. The factor of 1.1 in 4.19 is a safety margin to ensure that the inter-stage delay is not dominant.

$$R = \frac{1.386AB}{T_{10\%}} \times 1.1 \quad (4.19)$$

For the various technology nodes, table 4.12 shows the respective ratio  $R$  as in equation 4.19, the optimum reference width  $W$ , the linear input capacitance  $C_B$ , the linear output resistance  $R_B$  and the theoretical segment length to be driven,  $l_{seg}$ .

TABLE 4.12: Improved optimised repeater characteristics

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$R$	0.282	0.248	0.127	0.151	0.163	0.174
$W$ [ $\mu\text{m}$ ]	4.38	2.53	1.33	0.91	0.66	0.42
$C_B$ [F]	2.27E-14	1.11E-14	3.82E-15	2.65E-15	1.89E-15	1.19E-15
$R_B$ [ $\Omega$ ]	45	56	39	62	78	116
$l_{seg}$ [ $\mu\text{m}$ ]	447	294	236	151	101	66
NMOS <sub>stage1</sub> [ $\mu\text{m}$ ]	4.38	2.53	1.33	0.91	0.66	0.42
PMOS <sub>stage1</sub> [ $\mu\text{m}$ ]	9.37	5.31	1.63	1.24	0.93	0.58
NMOS <sub>stage2</sub> [ $\mu\text{m}$ ]	15.54	10.21	10.14	5.86	4.04	2.35
PMOS <sub>stage2</sub> [ $\mu\text{m}$ ]	33.25	21.44	12.48	7.97	5.66	3.27

#### 4.5.4 Optimisation in terms of capacitance

Although the approach in section 4.5.3 maximises the drivable segment lengths, the capacitance of the repeaters themselves increases as well. Using the total capacitance, that is,

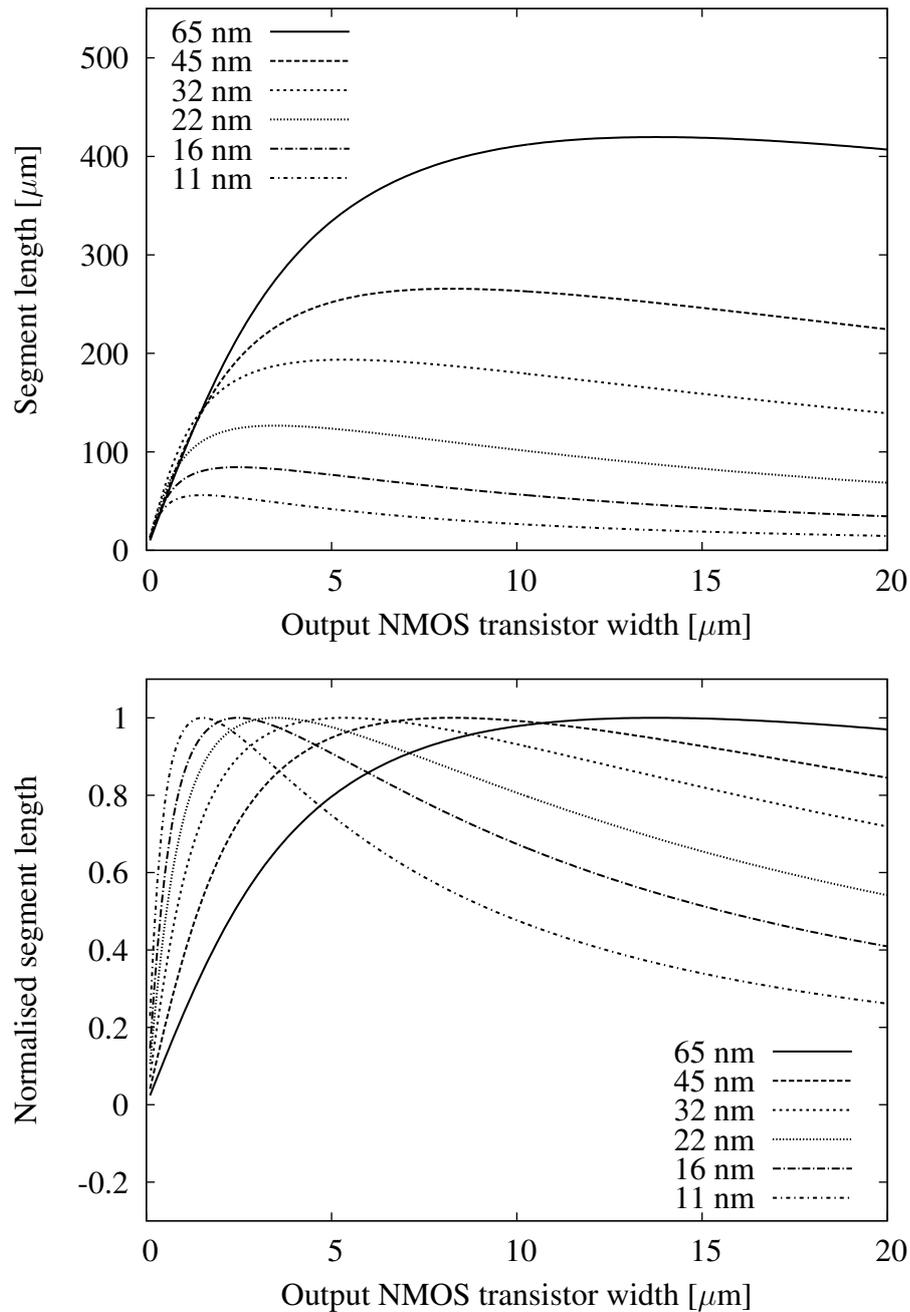


FIGURE 4.9: Optimal width for repeater transistors

repeater plus interconnect, and normalising the quantity as a per unit length metric, another, more sensible optimisation can be made: to minimise the effective combined capacitance per unit length.

The capacitance of a repeater can be written as

$$C_{\text{repeater}} = B \times W \times (1 + R) \quad (4.20)$$

with  $B$  and  $W$  from section 4.5.1 and  $R$  from section 4.5.3. The associated segment length

can be written as

$$C_{segment} = l_{seg} \times c_{int} \quad (4.21)$$

with  $l_{seg}$  as the maximum segment length and  $c_{int}$  as the interconnect capacitance per unit length.

#### 4.5.4.1 Design

Using these above mentioned relationships, the total capacitance per unit length is shown in equation 4.22 with  $l_{seg}$  explained in section 3.5.1.

$$C_{total/length} = c_{int} + \frac{B \times W \times (1 + R)}{l_{seg}} \quad (4.22)$$

By taking the derivative of equation 4.22 and determining

$$\frac{\partial C_{total/length}}{\partial R} = 0 \quad (4.23)$$

the optimum value for  $R$  is found. Of course, this value has a lower bound as described in section 4.5.3 in order to maintain interstage signal swing speeds. The result for each node is shown in figure 4.9, which shows the maximum drivable interconnect segment length as a function of the output NMOS transistor width, which is in good agreement with the SPICE results in figure 4.10.

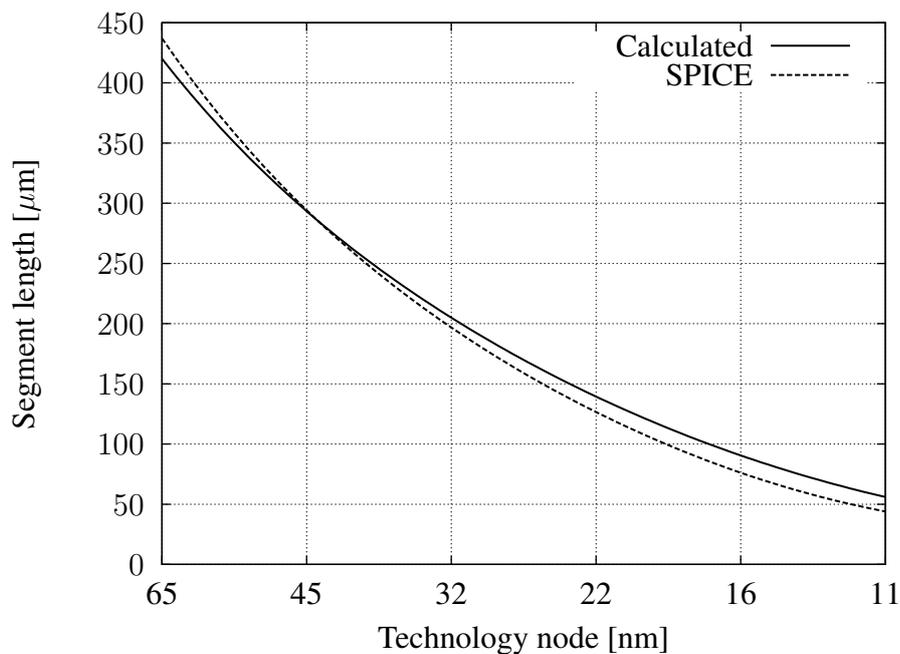


FIGURE 4.10: Predicted reduction of segment length for future technology nodes

Although this technique can be utilised to find the optimum buffer transistor dimensions, equation 3.16 by [32] does not take into account the output capacitance of the buffer driver transistors. This might lead to a difference in the predicted maximum segment lengths and the actual obtainable lengths. SPICE simulations were done in order to verify this optimisation theory, where figure 4.10 shows the predicted and obtained segment lengths when trying to maintain a 20 % - 80 % transition time. Of course, the approximation in terms of a linear buffer resistance and capacitance also introduces an error in the ultimate DC and transient behaviour.

TABLE 4.13: Capacitance optimised repeater characteristics

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$R$	0.36	0.38	0.47	0.44	0.44	0.43
$W$ [ $\mu\text{m}$ ]	13.75	8.24	5.35	3.48	2.47	1.52
$C_B$ [F]	2.57E-14	1.37E-14	7.23E-15	4.46E-15	3.09E-15	1.85E-15
$R_B$ [ $\Omega$ ]	51	70	74	105	128	180
$l_{seg}$ [ $\mu\text{m}$ ]	420	266	194	127	84	56
NMOS <sub>stage1</sub> [ $\mu\text{m}$ ]	4.95	3.13	2.52	1.53	1.09	0.65
PMOS <sub>stage1</sub> [ $\mu\text{m}$ ]	10.59	6.58	3.09	2.08	1.52	0.91
NMOS <sub>stage2</sub> [ $\mu\text{m}$ ]	13.75	8.24	5.35	3.48	2.47	1.52
PMOS <sub>stage2</sub> [ $\mu\text{m}$ ]	29.42	17.31	6.58	4.74	3.46	2.11

#### 4.5.4.2 Power consumption per repeater

Given the results in table 4.13, a theoretical computation of power consumption is possible. However, since short circuit currents and the transistor drain capacitances have been neglected, this value should be checked with simulation results.

The underestimated theoretical power consumption of the repeaters becomes

$$\begin{aligned}
 C_{rep} &= C_{in} + C_{interstage} \\
 P'_{rep} &= C_{rep} V_{DD}^2 \quad \left[ \frac{\text{W}}{\text{Hz}} \right]
 \end{aligned}
 \tag{4.24}$$

Equation 4.24 is can then be used to populate table 4.14, with  $P'_{rep}$  as the power consumption per unit frequency (also translated as Joule, and indication of the energy consumed per clock cycle).

TABLE 4.14: Theoretical power consumption per repeater

	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$P'_{rep} \left[ \frac{\text{fW}}{\text{Hz}} \right]$	117.3	60.4	22.6	11.8	8.2	3.9
At clock frequency (see figure 4.1)						
$P_{rep} [\mu\text{W}]$	551.5	354.8	166.2	108.6	94.0	56.5

The simulated power consumption for each repeater operating at the given clock frequency is shown in table 4.15, with  $P_{rep}$  as the power consumption at the specified clock frequency and  $\Delta P_{rep}$  representing the difference between the predicted and simulated power consumption value.

TABLE 4.15: Simulated power components per repeater

Power [ $\mu\text{W}$ ]	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$f_{clk}$ [GHz]	4.7	5.9	7.3	9.2	11.5	14.3
$P_{rep}$	1011.0	787.7	396.6	218.6	220.9	213.7
$\Delta P_{rep}$	459.5	432.9	230.4	110.0	126.9	157.2

## 4.5.5 End point buffer design

An end point denotes the point where the global clock network terminates into a buffer design to drive the local region capacitive load. Each endpoint terminates with a repeater driving an end point buffer, designed to carry the local region load. Figure 4.11 shows a schematic representation of what the terminating clock network looks like.

### 4.5.5.1 Design

Since the capacitive load is calculated in section 4.4 and the characteristic buffer parameters for relevant technology are known through table 4.10, the end point buffer can be designed with the requirement to swing the load capacitance to  $V_{DD}$  and ground in less than 10 % of the clock cycle. Using the ratio calculated with equation 4.19, it can easily be proved that a single buffer viz. two inverter pairs are sufficient to provide the gain necessary for driving the local region.

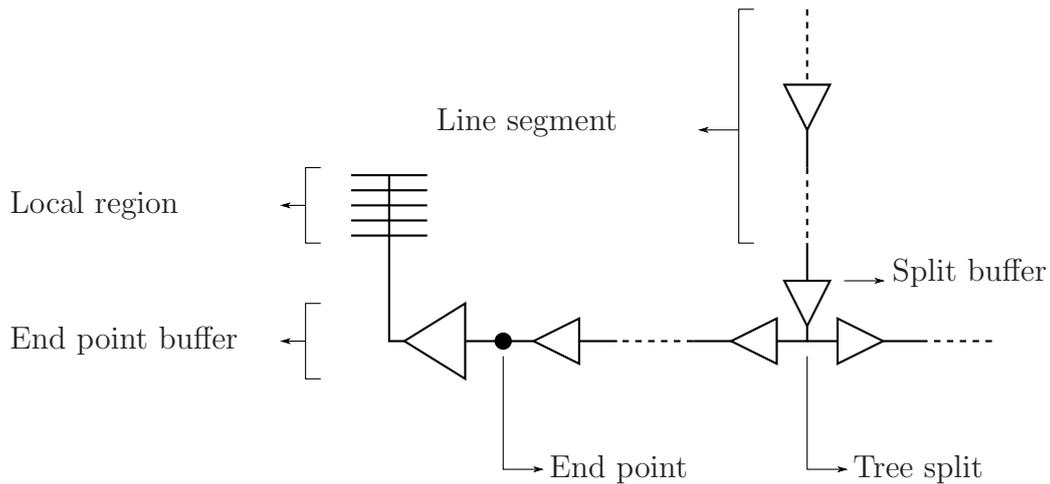


FIGURE 4.11: A representation of a terminating clock network

### 4.5.6 Split buffer design

The split buffer is a low input capacitance buffer with a fan-out of two repeaters, or tapered repeaters. The reason for this unit is simply to reduce the excess capacitance of using repeaters to terminate the line segments before a split. Theoretically, when the number of splits and end points increase, a noticeable power saving can be made. Equation 4.25 shows the requirement on the output resistance  $R_B$  of the split buffer in order to drive the subsequent two repeaters, with  $C_B$  representing the total input capacitance of the load.

$$R_B = \frac{T_{10\%}}{1.386C_B \times 1.1} \quad (4.25)$$

#### 4.5.6.1 Design

The input capacitance of a repeater as in section 4.5.4 is multiplied by two to present the load that the split buffer has to drive. Once again, equation 4.19 can be modified (refer to equation 4.25) to state the required drive capability as used for the interface between the split buffer and the load, as well as within the split buffer itself.

#### 4.5.6.2 Power consumption per split buffer

Equation 4.24 can also be used to calculate the theoretical power consumption for the split buffers.

Again, the drain capacitances and short circuit currents are not theoretically modelled. Simulation results, as shown in table 4.17, accounts for these power losses.

TABLE 4.16: Repeater configuration split buffer design parameters

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
2 fanout load [F]	5.13E-14	2.75E-14	1.45E-14	8.93E-15	6.18E-15	3.70E-15
Ratio $R$	See table 4.13					
$R_B$ required [ $\Omega$ ]	272	406	617	800	924	1236
$C_{in}$ [F]	4.78E-15	2.36E-15	8.66E-16	5.86E-16	4.27E-16	2.70E-16
$C_{interstage}$ [F]	1.33E-14	6.22E-15	1.84E-15	1.33E-15	9.71E-16	6.28E-16
NMOS <sub>stage1</sub> [ $\mu\text{m}$ ]	0.922	0.539	0.301	0.201	0.150	0.095
PMOS <sub>stage1</sub> [ $\mu\text{m}$ ]	1.972	1.131	0.370	0.274	0.210	0.132
NMOS <sub>stage2</sub> [ $\mu\text{m}$ ]	2.560	1.417	0.641	0.457	0.341	0.221
PMOS <sub>stage2</sub> [ $\mu\text{m}$ ]	5.478	2.976	0.788	0.622	0.478	0.307

TABLE 4.17: Theoretical and simulated power consumption per split buffer

	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$P'_{split}$ [ $\frac{\text{fW}}{\text{Hz}}$ ]	21.85	10.38	2.71	1.55	1.13	0.57
At clock frequency (see figure 4.1)						
$P_{split}$ [ $\mu\text{W}$ ]	102.7	61.0	19.9	14.3	13.0	8.2
Simulated results						
$P_{split}$ [ $\mu\text{W}$ ]	176.4	130.6	46.3	28.0	29.7	30.0
$\Delta P_{split}$ [ $\mu\text{W}$ ]	73.7	69.6	26.4	13.7	16.6	21.7

# CHAPTER FIVE

## OPTICAL RECEIVER AND FRONT-END DESIGN

---

**O**PTICAL interconnects may pose a very attractive solution to the problem of clock distribution in VLSI circuits. But while the signal may contain timing information, it will be futile to the rest of the electronic circuitry if a method of sensibly converting this information is not introduced.

### 5.1 PHOTODIODE AS CIRCUIT ELEMENT

In order to sensibly use a diode as a circuit element, it is necessary to model it as a component true to its electrical behaviour. Ideally, the photodiode should behave like an ideal current source, with the output current a function of the incident light. As described in section 2.2.3, a photodiode has two bandwidth limiting mechanisms: the electrical and intrinsic bandwidth of the device. The latter is a function of the geometry, process parameters and depletion region width of the device, while the former is, along with its circuit implementation, a function of the capacitance between the device terminals and the impedance of the subsequent stages.

Since the intrinsic bandwidth cannot be sensibly changed by electrical circuits (except by applying higher reverse bias), this intrinsic bandwidth can be included, along with responsivity, as a frequency dependent current source. With either a n-well and p-substrate diode, or a p+ addition to the n-well, the circuit can still be reduced to an equivalent where the diode capacitance is modelled as a capacitor, the bulk and contact resistances as a resistor and the leakage current as a conductance in parallel to the current source. The leakage and capacitive effects can be contained in one diode element based on the appropriate process

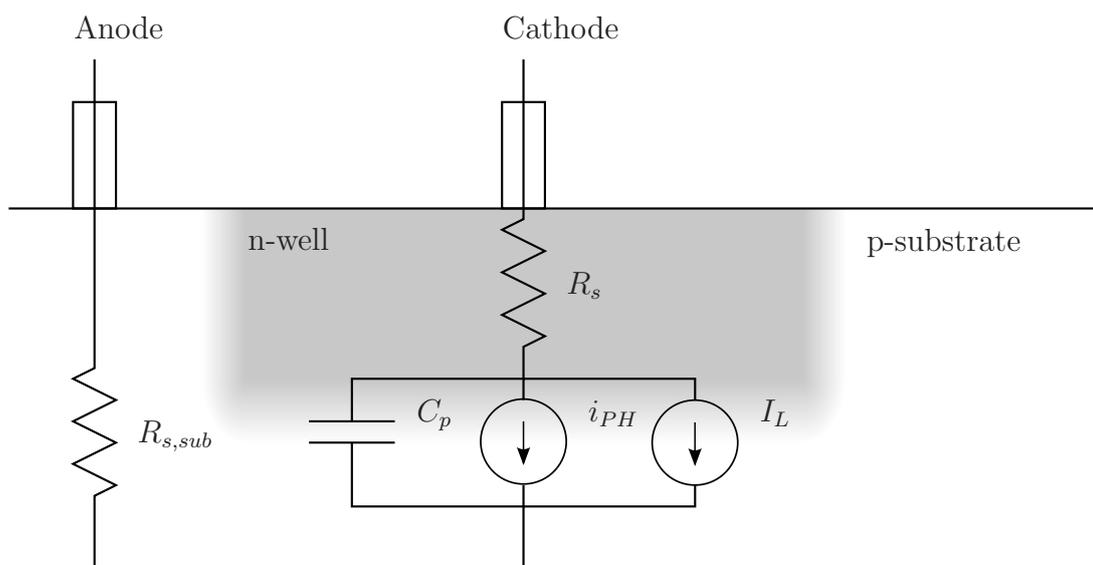


FIGURE 5.1: A simple electrical equivalent of an n-well photodiode

models. Figure 5.1 shows an electrical equivalent of a photodiode consisting of an n-well. Often, the leakage  $I_L$  current can simply be ignored if the photocurrent  $i_{PH}$  is large enough. The resistive component  $R_{s,sub}$  can also be neglected since the substrate is usually well grounded with guard rings and through the back of the wafer through conducting epoxy.

Most SPICE-based simulators can also easily accept data in the form of points being read in from a data file. Using the information in section 2.5, such a datafile can be created to account for both the photodiode intrinsic bandwidth and any light source signal distortion or degradation.

### 5.1.1 Photodiode intrinsic bandwidth

Analysis of the response of photodiodes in section 2.5 shows that there are some methods for increasing the intrinsic bandwidth. Some of these are:

- Reduce the light wavelength such that penetration remains within the depletion region.
- Reduce the size of the areas not under electrical field influence (in other words, maximising the depletion region).

Of course, the physical size also affects the electrical capacitance, which is crucial to the subsequent electrical designs. Fast photodiodes are possible in standard CMOS. As an example, [17] shows designs exceeding 6 GHz in a 0.18  $\mu\text{m}$  process.

### 5.1.2 Photodiode responsivity

Given the difficulty in scaling photodiodes, the elements need at least be dimensioned to the wavelength of detectable light, an estimate as in equation 5.1 is taken as an assumed responsivity for all subsequent designs, which is in accordance with results from [17].

$$R = 0.3 \left[ \frac{\text{A}}{\text{W}} \right] \quad (5.1)$$

### 5.1.3 Photodiode capacitance

Likewise, the electrical capacitance of devices in [17] is taken and scaled according to a photodiode of approximately  $4 \mu\text{m}^2$ , which gives a capacitance of

$$C_{PD} = 5 \text{ [fF]} \quad (5.2)$$

The value for the photodiode capacitance in 5.2 will be used as an estimate for the subsequent amplifier designs.

## 5.2 RECEIVER CONSIDERATIONS

### 5.2.1 Bandwidth limitations

Throughout this work, the 20 % - 80 % transition time metric was used as a foundation for a switching speed constraint. It follows that the bandwidth of the optical receiver needs to be such that this transition time can be maintained. It can be shown, through a similar analysis as in section 3.5.1, that  $\Delta t_{20-80} = 1.386\tau$ , where  $\tau$  is the amplifier dominant pole time constant. If  $\Delta t_{20-80}$  is to be kept under 10 % of the clock period, equation 5.3 shows the constraint on the amplifier bandwidth for the criterion to be met.  $f_c$  refers to the amplifier -3dB bandwidth.

$$f_c > 2.21f_{clk} \quad (5.3)$$

Equation 5.3 is less of a concern where the limitation on signal switching is due to large signal constraint, such as in the high impedance approaches. Nevertheless, the subsequent amplification stages still need to ensure adequate bandwidth operation for the detection system to be operational.

Throughout the rest of this chapter, bandwidth is assumed to be the most important design parameter. It is important that the optical end point amplifier meets at least the technology

node clock frequency, before sensible power comparisons can be done between the optical and electrical equivalents.

### 5.2.2 Noise limitations

Noise presents the lower fundamental limit of an input signal. Since noise generators exist in both active devices and resistors, it is important to choose an architecture capable of minimising the noise referred to the input node, while still supplying enough gain and bandwidth for the signal to be accurately amplified.

### 5.2.3 Drive capability

As the first electrical point in an optical clock network, the amplifiers will need to serve both the purpose of converting from optical to electrical signals, as well as drive the required load. For the purpose of comparison, the end point buffer presents a capacitive load to the optical amplifier front end. It is therefore necessary for the front end to maintain the necessary drive capability to drive an end point buffer, as explained in section 4.5.5.

## 5.3 OVERVIEW OF TYPICAL RECEIVERS

### 5.3.1 Common source switching amplifier

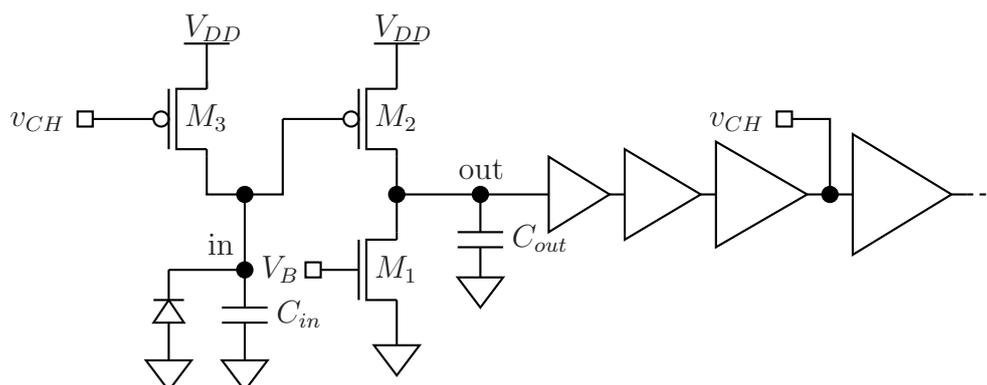


FIGURE 5.2: Common source amplifier configuration as an optical end point

### 5.3.1.1 Operational overview

The common source amplifier can serve as a high impedance optical amplifier, where the photocurrent is responsible for charging the capacitance on the high impedance input node, labelled  $in$  in figure 5.2. The resulting drop in  $v_{OUT}$  will then trigger the inverters to follow. After a certain time delay, which can be derived from the series cascade of inverters, the charge transistor resets the voltage on the input node to  $V_{DD}$ , thereby generating a periodic signal based on the input photocurrent. The pull-up load is also used to set the bias current  $I_D$ , although  $I_D$  will not remain constant as  $M_1$  and  $M_2$  operate in switching mode between triode and saturation regions. This is similar to the approach given in [26], except that the PMOS serves as gain transistor, as its threshold is closer to  $V_{DD}$  than to ground.

### 5.3.1.2 Bandwidth considerations

If  $C_{in}$  and  $C_{out}$  represent the total nodal input and output capacitance respectively, it follows that the photodiode needs to discharge  $C_{in}$  at a sufficient rate such that the voltage  $v_{IN}$  maintains the  $\Delta t_{20-80}$  transition time. The voltage only needs to drop from  $V_{DD}$  to the point where the output node  $out$  makes the transition between low and high. This is strongly dependent on the sizing of the amplification and current source transistors, where both the gain curve and the transition point can be covered when  $v_{IN}$  is discharged to below  $V_{DD}/3$ . The choice of using a PMOS amplifier assists with shifting the transition point to a higher value, but at the cost of a slight increase in input capacitance and a relaxed DC voltage transfer curve. With the above constraints taken into account, a voltage of  $0.25V_{DD}$  needs to be extracted within  $1/10 \times T_{CLK}$ . It follows that, for each clock period equation 5.4 needs to hold true for the pulse duration and optically generated current.

$$i_{PD}(t) = \begin{cases} 2.5 \times C_{in} V_{DD} f_{CLK} & \text{for } 0 \leq t < \frac{4}{15} T_{CLK} \\ 0 & \text{otherwise} \end{cases} \quad (5.4)$$

On the output node, a nodal capacitance  $C_{out}$  needs to be charged and discharged in a similar fashion. The bias current set via  $v_B$  needs to support the large signal slew requirements for the voltage swing to support the  $\Delta t_{20-80}$  timing constraint. This voltage should, however, support logic levels for the subsequent buffer stages.

$$I_{D1} = I_{D2} = 6 \times C_{out} V_{DD} f_{CLK} \quad (5.5)$$

Of course, the small signal bandwidth is limited by the pole formed as shown in equation 5.6, which places a limit on the gain of the amplifier.

$$f_{p_{out}} = \frac{1}{2\pi C_{out} r_{o1||o2}} \quad (5.6)$$

### 5.3.1.3 Noise considerations

Both  $M_1$  and  $M_2$  will be responsible for noise on the output node, which will result in temporal jitter due to the hard limiting nature of the amplifier. Noise analysis is not as straightforward as in the case of a small signal model, since the voltage swings are large and the transistors operate in multiple regions in a clock period. Nevertheless, an approximate model can be described using the bias current as computed in 5.5. Note that, by referring the output node noise back to the input as a current, this amplifier choice results in a dramatical drop in contributed noise when compared to some other configurations.

The noise on the output simply reduces to a term

$$\overline{v_{no}} = (\overline{i_{n1}} + \overline{i_{n2}}) \times r_{o1||o2} \left[ \frac{V}{\sqrt{\text{Hz}}} \right] \quad (5.7)$$

### 5.3.1.4 Device sizing

Device  $M_3$  is responsible for recharging the discharged node, where the length is chosen at minimum size and the width scaled as to provide roughly the same charge rate as the discharge rate of the photodiode. For maximum gain,  $r_{o1}$  and  $r_{o2}$  need to be maximized. However, increasing the length of  $M_2$  results in a larger input capacitance, which increases the input current requirement as per equation 5.4. Since the amplification term is roughly  $g_{m2} r_{o1||o2}$ , increasing the width of  $M_2$  will increase gain, at the cost of an increase in  $C_{out}$ . Increasing the bias current above the set value of equation 5.5 will result in a linear increase in power consumption, with a square rooted increase in gain. The length of  $M_1$  can be increased at the expense of headroom. Assuming a 10 % increase in  $C_{out}$  above the minimum transistor dimension values is allowable, with a 10 % reduction in headroom for the output node, a reasonable optimisation can be done.

### 5.3.1.5 Input signal magnitude

Section 5.3.1.2 shows that the input signal has a direct influence on the speed attainable with this configuration. The magnitude of the input signal required can be computed with equation 5.4, which can be used to determine the particulars of the light source and photodiode combination.

### 5.3.1.6 Power consumption per amplifier

The average bias current is set via  $V_B$  as  $I_D$ . Along with this quantity, the switching of the inverter pairs and the discharge transistor capacitance creates a frequency dependent power consumption component. Suppose  $C_{sw}$  represents the capacitances of the inverters excluding  $C_{out}$ , which is covered by the bias current, the electrical power can be computed as a quantity per unit operating frequency as in equation 5.8. Note that this expression does not include the effects of the short circuit current originating in the inverters.

$$P'_E = V_{DD} \left( \frac{I_D}{f_{CLK}} + (C_{sw} + 0.5C_{in})V_{DD} \right) \quad \left[ \frac{\text{W}}{\text{Hz}} \right] \quad (5.8)$$

If  $R$  is the responsivity in A/W then the instantaneous optical power required for the photocurrent in 5.4 is  $p_O(t)$ , shown in equation 5.9.

$$p_O(t) = \begin{cases} 2.5 \times \frac{C_{in}V_{DD}f_{CLK}}{R} & \text{for } 0 \leq t < \frac{4}{15}T_{CLK} \\ 0 & \text{otherwise} \end{cases} \quad [\text{W}] \quad (5.9)$$

From the above, the average power expression is found to be equation 5.10.

$$P_O = \frac{1}{T_{CLK}} \int_0^{T_{CLK}} p_O(t) dt \quad [\text{W}] \quad (5.10)$$

$$P'_O = \frac{1}{3} \frac{C_{in}V_{DD}}{R} \quad \left[ \frac{\text{W}}{\text{Hz}} \right]$$

### 5.3.1.7 Optical signal characteristics

Based on equations 5.9 and 5.10, it is clear that power consumption can be reduced by decreasing the duty cycle. Since only the temporal edge information is necessary for clock purposes, this can be to an advantage in reducing the power requirements. The clock will however not be symmetrical. Losses between the source and the photodiode also need to be included in the power calculation, as well as whether the source is a modulated continuous wave or a switched source.

The above mentioned equations assume the optical power is available to the photodiode, without any losses between the light source and photodiode. It is also worth mentioning that the light intensity can become excessive when working with small photodiodes. Even a  $1 \mu\text{W}$  requirement on a  $1 \mu\text{m}^2$  photodiode gives light intensities of  $100 \frac{\text{W}}{\text{cm}^2}$ , which is a thousand times the intensity of the sun!

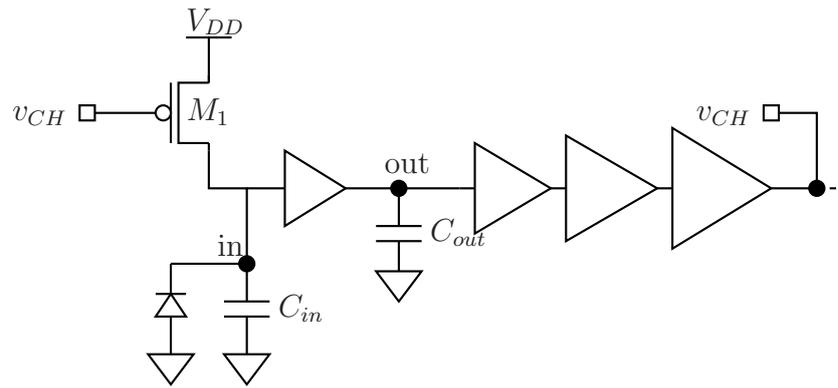


FIGURE 5.3: Front end utilising buffer high impedance input node for charge accumulation

## 5.3.2 Digital buffer amplifier

### 5.3.2.1 Operational overview

Sizing an NMOS to absolute minimum dimensions along with the matching PMOS in an inverter pair results in the smallest input capacitance possible on a digital gate. An optical current pulse can directly play a role on the voltage of this input node. As in section 5.3.1, a photocurrent is used to insert or withdraw charge from the high impedance node, which results in charge that will subsequently need replacement to sustain a periodic signal. Again, a PMOS device is used in conjunction with a grounded anode photodiode (one n+/n-well type in p-substrate) to replace the charge extracted by the photodiode after a small time period. The cycle can then repeat on the next optical pulse.

Of course, the idea of the amplifier is to utilise the voltage on the high impedance node, modulated by the current of the photodiode, to generate an amplified signal on its output node. The idea behind this approach as opposed to that of section 5.3.1 is that one can do away with a NMOS/PMOS pair, at the expense of an increase in input capacitance.

### 5.3.2.2 Bandwidth considerations

As shown in section 4.5.3, the interstage transition time is maintained according to the  $\Delta t_{20-80}$  metric. The input signal strength thus determines the switching speed of the system, as per equation 5.4.

### 5.3.2.3 Noise considerations

As with the common source switching amplifier, the noise will translate to an amount of jitter due to the hard limited nature of the series cascade of inverters. Techniques similar to

the common source amplifier can be used to estimate the noise on transitions, where both transistors are on.

#### 5.3.2.4 Device sizing

The input devices are kept at a minimum, as a reduction in input capacitance results in improved sensitivities. The method described in section 4.5.1 is used to taper the design to the point where the final inverter can drive a buffer designed for a local network, as described in section 4.5.5.

Therefore, there are no degrees of freedom for the designer in this implementation, except for the strength of  $M_1$ , which simply needs to restore the node to  $V_{DD}$  in time.  $M_1$  will have a certain amount of influence on the short circuit current, as a higher recharge time will shorten the duration of a short circuit condition. This requires a larger device, which adds to the switched capacitance, thus an optimum should be found.

#### 5.3.2.5 Input signal magnitude

The voltage on node  $in$  needs to swing to full logic levels, since the designer has no control over the exact threshold voltage of the first amplifier. This increases the requirement on the input signal magnitude as shown in equation 5.11.

$$i_{PD}(t) = \begin{cases} 6 \times C_{in} V_{DD} f_{CLK} & \text{for } 0 \leq t < \frac{1}{6} T_{CLK} \\ 0 & \text{otherwise} \end{cases} \quad (5.11)$$

#### 5.3.2.6 Power consumption per amplifier

Neglecting the short circuit current effects, the total capacitance  $C_{sw}$  being switched, the supply voltage and the operating frequency determine the power consumption. Following the approach in formulating equation 5.8, the average electrical power consumption per unit frequency of the digital buffer amplifier is simply

$$P'_E = V_{DD}^2 (0.5 C_{in} + C_{sw}) \left[ \frac{W}{Hz} \right] \quad (5.12)$$

The optical power requirements become a bit more stringent since the full  $V_{DD}$  needs to be swung on the input node. Equation 5.10 simply changes to equation

$$P'_O = \frac{C_{in} V_{DD}}{R} \left[ \frac{W}{Hz} \right] \quad (5.13)$$

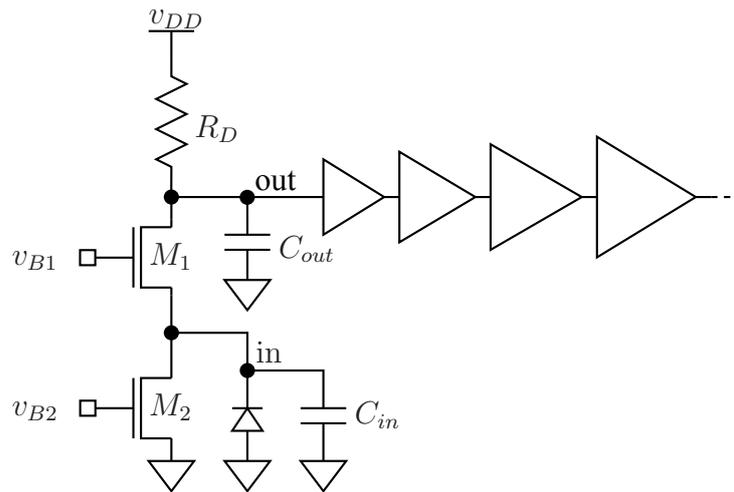


FIGURE 5.4: Common gate amplifier configuration as an optical end point

### 5.3.3 Common gate amplifier

#### 5.3.3.1 Operational overview

Figure 5.4 shows the source terminal of a MOS device when used as an input terminal, where the input impedance is then roughly set as  $1/g_{m1}$ . Voltages  $v_{B1}$  and  $v_{B2}$  set the bias points for the amplifying and current mirror transistors respectively, where  $I_{D1} = I_{D2} = I_{RD}$  typically sets the DC output voltage on node *out*. Of course, depending on the input photocurrent, this bias should be located such that  $v_{OUT}$  swings around the threshold voltage of the subsequent inverter stages. Since the source of  $M_1$  is not at bulk potential, an extra current generator,  $g_{mb1}$  can be added to the  $g_{m1}$  term, since they are in parallel. A resistive pull-up load can then be used to create transconductance gain of the input photocurrent. Figure 5.5 shows the

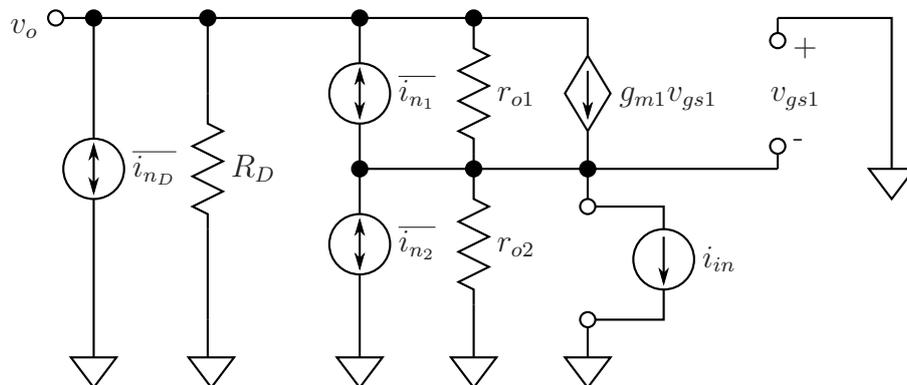


FIGURE 5.5: Small signal model of common gate amplifier

small signal equivalent for the amplifier with channel length modulation and noise current

generators. Note that  $g_{m1}$  includes the body effect transconductance.

### 5.3.3.2 Bandwidth considerations

$C_{out}$  and the amplifier output impedance determines the response on the output node, while  $C_{in}$  and the amplifier input impedance limits the response on the input node. The input resistance is shown in equation 5.14.

$$R_{in} = \frac{r_{o1} + R_D}{g_{m1}r_{o1} + 1} \quad (5.14)$$

Equation 5.14 shows that the input resistance is approximately  $1/(g_{m1} + g_{mb1})$  plus a reduction of  $R_D$  of roughly  $(g_{m1} + g_{mb1}) \times r_{o1}$ , assuming the parallel resistance  $r_{o2}$  is much larger and can be neglected. This means that higher frequencies can be obtained when compared to a simple resistance as a pull up load directly on the photodiode. The output resistance is calculated as in equation 5.15.

$$\begin{aligned} R_{out} &= (r_{o1} + (1 + g_{m1}r_{o1})r_{o2}) || R_D \\ &\approx R_D \end{aligned} \quad (5.15)$$

Note that, because of the cascode nature of the amplifier, the output pole is largely determined by the pull up load. The two poles associated with this amplifier can then be shown as in equations 5.16 and 5.17. Note that, as stated in section 5.2.1, the dominant pole needs to be at least at  $2.21f_{CLK}$ .

$$f_{pin} = \frac{1}{2\pi R_{in} C_{in}} \quad (5.16)$$

$$f_{pout} = \frac{1}{2\pi R_D C_{out}} \quad (5.17)$$

The choice of  $R_D$  therefore depends on  $C_{out}$  and effectively determines the gain of the amplifier. The bias current has to be chosen accordingly, since the output node voltage swing will depend directly on the choice of  $I_D$ . Taking the input and output resistances into account, the transimpedance gain of this amplifier then becomes

$$R_{CG} = \frac{v_{out}}{i_{in}} = -\frac{r_{o2}}{R_{in} + r_{o2}} \times R_D \approx -R_D \quad (5.18)$$

### 5.3.3.3 Noise considerations

Common gate amplifiers are typically not used for low noise applications for the very simple reason that the noise of the pull up load and current mirror adds directly as input noise.

$$\begin{aligned}\overline{v_{n,out}} &= \overline{i_{n,D}} \times R_D \\ &+ \overline{i_{n,M1}} \times \frac{1}{\frac{1}{R_D} + \frac{1}{r_{o1}} + \frac{r_{o2}}{R_D} \left( \frac{1}{r_{o1}} + g_{m1} \right)} \\ &+ \overline{i_{n,M2}} \times \frac{r_{o2}}{R_{in} + r_{o2}} R_D\end{aligned}\quad (5.19)$$

Dividing equation 5.19 by the gain in 5.18, it is clear that the input referred noise is not reduced substantially, since the current gain  $\approx 1$ .

### 5.3.3.4 Device sizing

From equation 5.17, the maximum value of  $R_D$  can be derived. In order for the subsequent inverters to sensibly utilise the output signal, the DC steady state value for  $V_{OUT}$  must lie above the threshold of the gate, while the term  $i_{IN} \times R_D$  must pass to well below the threshold for the inverter to sense a zero. This means that both  $M_1$  and  $M_2$  need to be scaled such that the minimum value of  $v_{OUT}$  is still above the saturation voltages of both. This requires an increase in  $g_{m1}$  and  $g_{m2}$ , which in turn increases  $C_{in}$ . Ideally,  $g_{m1}$  and  $g_{m2}$  should be set to exactly the point that the overdrive voltages  $V_{ov,1} + V_{ov,2} < V_{ILmax}$ , where  $V_{ILmax}$  is the load inverter's maximum guaranteed input voltage that will be recognised as a low.

### 5.3.3.5 Input signal magnitude

Since the value of  $R_D$  is effectively set, the gain dictates the input current required for operation. With  $I_D$  set such that the output node is kept at  $V_{IHmin}$ , the required input signal can be expressed as in 5.20.

$$i_{in}(t) = \begin{cases} \frac{V_{IHmin} - V_{ILmax}}{R_D} & \text{for } 0 \leq t < D \cdot T_{CLK} \\ 0 & \text{otherwise} \end{cases}\quad (5.20)$$

$D$  is a parameter introduced to represent the duty cycle of the applied optical pulse, where  $0 < D \leq 1$ .

### 5.3.3.6 Power consumption per amplifier

By taking  $C_{sw}$  to represent the sum of the switched capacitances for the inverters minus  $C_{out}$ , the electrical component of the power consumption per amplifier is

$$P'_E = V_{DD} \left( \frac{I_D}{f_{CLK}} + C_{sw} V_{DD} \right) \left[ \frac{\text{W}}{\text{Hz}} \right] \quad (5.21)$$

The optical power consumption is shown in equation 5.22. Notice that the power consumption is not dependent on the clock frequency, only on the duty cycle of the applied pulse.

$$P'_O = \frac{D \cdot i_{in,max}}{f_{CLK} R} \left[ \frac{\text{W}}{\text{Hz}} \right] \quad (5.22)$$

### 5.3.4 Common source feedback transimpedance amplifier

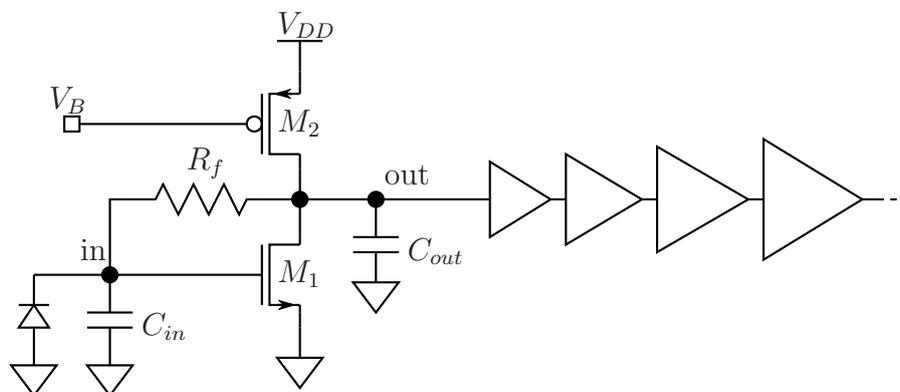


FIGURE 5.6: Common source feedback transimpedance amplifier

#### 5.3.4.1 Operational overview

A common source amplifier based on 5.2 in section 5.3.1 can be changed to a linear transimpedance amplifier by introducing a resistor between the drain and gate of the amplifier. The conversion of the configuration in section 5.3.1 to a shunt-shunt feedback topology results in a transimpedance amplifier with gain based on the feedback resistor and amplification factor on the drain node. Ideally, the resistor acts as a gain element while the dominant pole frequency increases. Because of its simplicity, component count and large bandwidth, this topology serves well as an optical front end.

Figure 5.7 shows the small signal equivalent of the amplifier, with  $r_o = r_{o1} + r_{o2}$  and  $\overline{i_{n,M}} = \overline{i_{n,M1}} + \overline{i_{n,M2}}$ . The gain of the circuit, including channel length modulation, can easily be determined as

$$R_{CSF} = \frac{r_o (g_{m1} R_f - 1)}{g_{m1} r_o + 1} \approx R_f \quad (5.23)$$

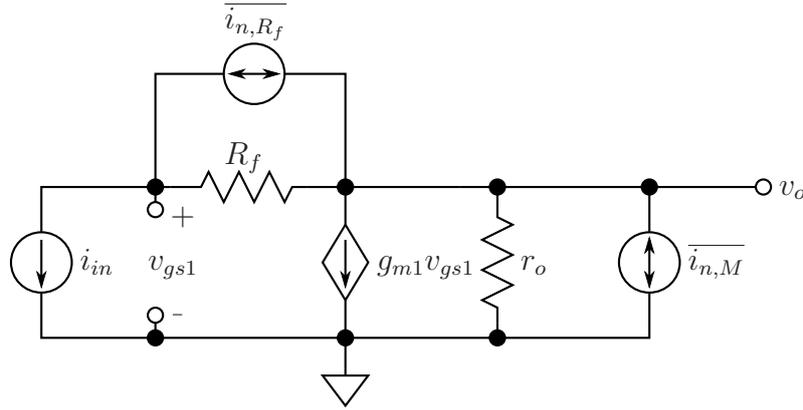


FIGURE 5.7: Small signal model of the common source feedback TIA

### 5.3.4.2 Bandwidth considerations

The input resistance as seen by the photodiode is expressed in equation

$$R_{in} = \frac{r_o + R_f}{1 + g_{m1}r_o} \approx \frac{R_f}{g_{m1}} \quad (5.24)$$

The output resistance, shown in equation 5.25, seen by the node *out* sets the limit on frequency of operation on the output node. Note that the transistor output resistance is reduced substantially by the feedback, favouring this configuration above that of the common gate approaches.

$$R_{out} = \frac{r_o}{1 + g_{m1}r_o} \approx \frac{1}{g_{m1}} \quad (5.25)$$

### 5.3.4.3 Noise considerations

While the noise contribution from the feedback resistor is directly referred to the input node, the transistor noise on the output node gets reduced by  $1/g_{m1}$  and again by  $R_f$  when referred back to the input (see equation 5.26).

$$\begin{aligned} \overline{v_{n,out}} &= \overline{i_{n,R_f}} \times \frac{g_m R_f r_o}{g_{m1} r_o + 1} \\ &+ \overline{i_{n,M}} \times \frac{r_o}{g_{m1} r_o + 1} \\ &\approx \overline{i_{n,R_f}} \times R_f + \overline{i_{n,M}} \times \frac{1}{g_{m1}} \end{aligned} \quad (5.26)$$

If no biasing current transistor is present, the noise performance of the common source feedback amplifier exceeds that of the common gate topology for the same gain.

### 5.3.4.4 Device sizing

Feedback in this configuration depends on the assumption that  $g_{m1}r_o \gg 1$ . Therefore, it is a necessary requirement to ensure this through device sizing. Normally,  $g_{m1}$  can be increased

in a close to square root dependency on the transistor width,  $W_1$ . However, both the input and output nodal capacitances increase linearly. The quantity  $r_o$  can be increased by increasing the length of either of the devices  $M_1$  or  $M_2$ . Unfortunately, an increase in  $L_1$  would result in a drop of  $g_{m1}$  unless  $W_1/L_1$  is kept constant.

This means that only  $W_1$  and  $L_2$  can be changed from minimum dimensioned values. These two parameters, along with  $V_B$ , need to set the bias point such that the output node should reside in the vicinity of  $V_{ILmax}$  of the subsequent stage. Notice the non-inverted operation for the chosen input current direction, as opposed to the common gate amplifier.

#### 5.3.4.5 Input signal magnitude

The input current should, when peaked, shift the output node voltage to just above  $V_{IHmin}$  of the load inverter, while still remaining below  $V_{DD} - V_{ov,2}$  for  $M_2$  to remain saturated. Thus, depending on the bandwidth requirements and the choice of  $R_f$ , the input signal is defined in equation

$$i_{in}(t) = \begin{cases} \frac{V_{IHmin} - V_{ILmax}}{R_f} & \text{for } 0 < t \leq D \cdot T_{CLK} \\ 0 & \text{otherwise} \end{cases} \quad (5.27)$$

The duty cycle is defined the same as in equation 5.20.

#### 5.3.4.6 Power consumption per amplifier

As with the power consumption terms in section 5.3.3, the electrical power consumption depends on the switched capacitances and the bias current component. Both the optical and electrical power components are identical to the expressions in section 5.3.3.6, with the exception that  $i_{in,max}$  refers to equation 5.27.

### 5.3.5 Regulated cascode amplifier

#### 5.3.5.1 Operational overview

A technique used to improve the output impedance of current mirrors [47] is that of active or regulated cascoding, which can also be employed as a transimpedance amplifier [48].  $M_2$  in figure 5.8 attempts to maintain the voltage  $v_{IN}$  as a constant. When  $v_{IN}$  climbs,  $M_2$  amplifies the signal and lowers the gate of  $M_1$  accordingly, thereby serving as a very low impedance input port on node  $in$ . This configuration, although relaxing the requirements on  $C_{in}$ , is still

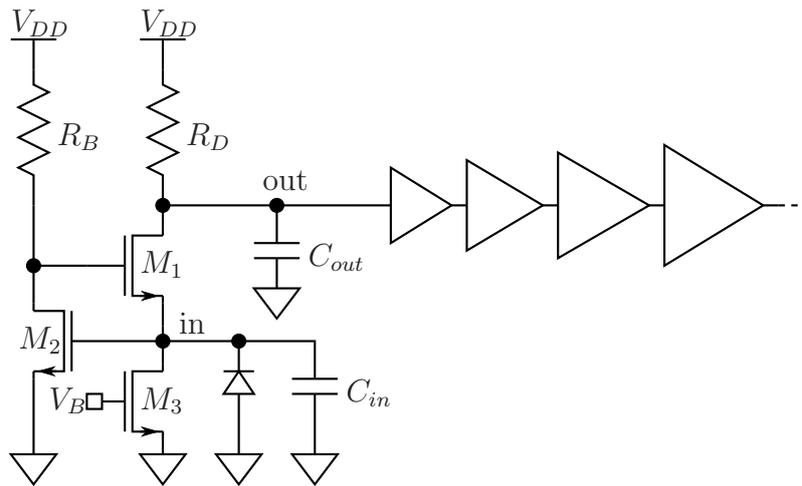


FIGURE 5.8: Regulated cascode circuit as an optical end point

limited in gain by the pole created by  $C_{out}$ . Therefore, the same biasing restrictions apply as for the common gate amplifier. Note that a body effect current generator  $g_{mb1}$  should be added to  $g_{m1}$  to take the non-grounded potential of the source of  $M_1$  into account.

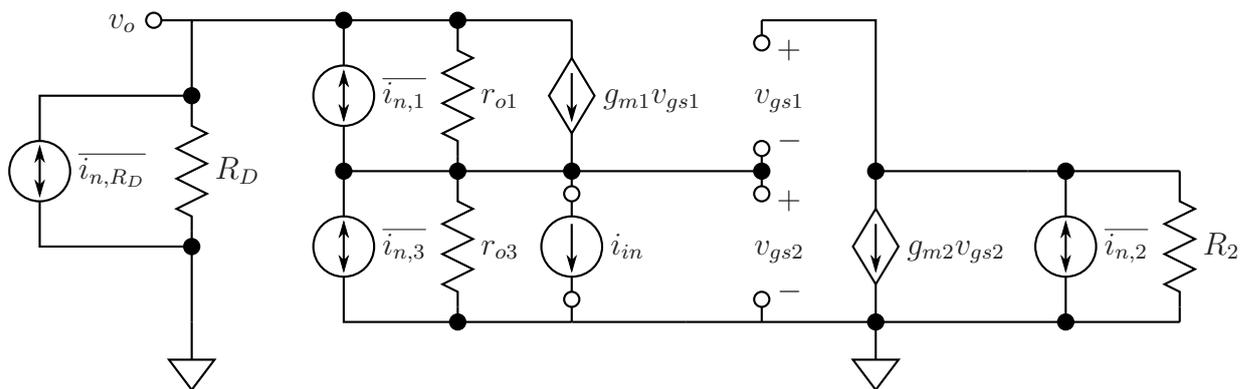


FIGURE 5.9: Small signal model for the regulated cascode

### 5.3.5.2 Bandwidth considerations

If  $R_2 = r_{o2} || R_B$  and assuming  $r_{o3}$  is large, the input resistance of the amplifier operating in the pass band can be expressed as

$$R_{in} = \frac{r_{o1} + R_D}{(1 + g_{m1}r_{o1}(1 + g_{m2}R_2))} \quad (5.28)$$

This expression is similar to equation 5.14, with a further reduction of roughly  $g_{m2}R_2$ , which means that this configuration is well suited to large input diode capacitances. The

output resistance of the amplifier as seen into the drain of  $M_1$  is also multiplied by the  $(1 + g_{m2}R_2)$  term, which results in an extremely high output impedance. As shown in equation 5.29, the output impedance seen by  $C_{out}$  in figure 5.8 can simply be approximated as  $R_D$ , which then determines the frequency pole due to the output node.

$$R_{out} = [r_{o1} + r_{o3} + g_{m1}(1 + g_{m2}R_2)r_{o1}r_{o3}] || R_D \approx R_D \quad (5.29)$$

Although this configuration presents a fast alternative for large diode capacitances, the gain of the system is again limited depending on the required bandwidth, as  $R_D$  functions once again as the gain element.

### 5.3.5.3 Noise considerations

If  $\overline{i_{n,2}}$  is defined as the noise of both  $M_2$  and  $R_B$ , then

$$\begin{aligned} \overline{v_{n,out}} &= \overline{i_{n,R_D}} \times R_D \\ &+ \overline{i_{n,3}} \times \frac{r_{o3}}{R_{in} + r_{o3}} \\ &+ \overline{i_{n,2}} \times \frac{R_2}{\frac{1}{R_D} + \frac{1}{r_{o1}} + \frac{r_{o3}}{R_D} \left( \frac{1}{r_{o1}} + g_{m1}(g_{m2}R_2 + 1) \right)} \\ &+ \overline{i_{n,1}} \times \frac{1}{\frac{1}{R_D} + \frac{1}{r_{o1}} + \frac{r_{o3}}{R_D} \left( \frac{1}{r_{o1}} + g_{m1}(g_{m2}R_2 + 1) \right)} \end{aligned} \quad (5.30)$$

The dominant two components contributing to input referred noise still remain terms  $\overline{i_{n,R}}$  and  $\overline{i_{n,3}}$ , with noise performance similar to the common gate approach.

### 5.3.5.4 Device sizing

As long as the  $(1 + g_{m2}R_2)$  term from the feedback transistor  $M_2$  is valid, the sizing of devices can be done according to section 5.3.3.4 with little loss of generality.  $M_2$  simply needs to keep the gate bias of  $M_1$  such that  $M_3$  has sufficient overdrive voltage. The choice of  $R_B$  depends on  $r_{o2}$ , where the combination of the resistors should generate a pole well beyond the dominant pole set by  $C_{in}$  or  $C_{out}$ . Thus,  $r_{o2}$ ,  $R_B$  and  $g_{m2}$  should be maximised until the frequency response of the feedback starts to interfere with the rest of the amplifier.

### 5.3.5.5 Input signal magnitude

The input signal calculation is done exactly as in section 5.3.3.5.

### 5.3.5.6 Power consumption per amplifier

Power consumption of a regulated cascode can be determined in a similar fashion to the power consumed in section 5.3.3.6, with the addition of the power resulting from the biasing of  $M_2$ .

$$P'_E = V_{DD} \left( \frac{I_{D1} + I_{D3}}{f_{CLK}} + C_{sw} V_{DD} \right) \left[ \frac{\text{W}}{\text{Hz}} \right] \quad (5.31)$$

The optical power depends on the input signal format and can be determined using equation 5.22 in section 5.3.3.6.

## 5.3.6 Complementary transimpedance amplifier

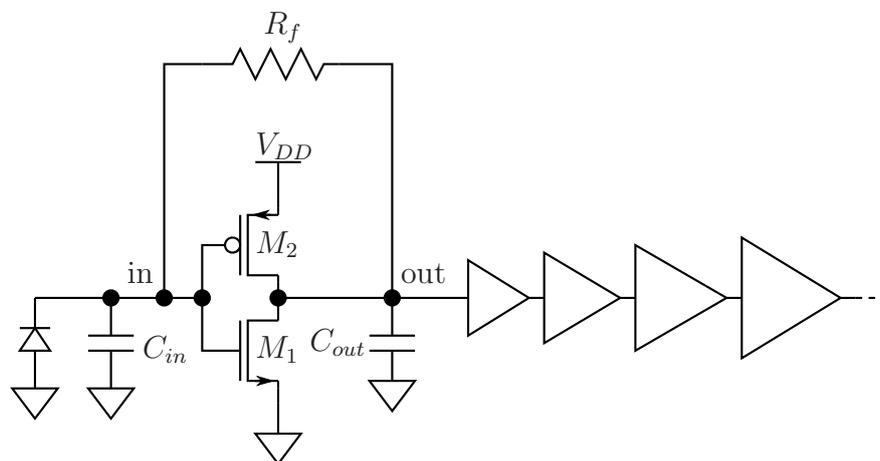


FIGURE 5.10: A transimpedance feedback amplifier based on a complementary CMOS pair

### 5.3.6.1 Operational overview

The principles of the common source feedback TIA, as in section 5.3.4, can be applied to a typical CMOS buffer [49], where both the NMOS and PMOS device contribute to the gain, instead of the PMOS operating only as a current source pull up load.

As seen in figure 5.11, the common source amplifier with shunt-shunt feedback (section 5.3.4) exhibits the same small signal output resistance characteristics, but without the extra PMOS current generator  $g_{m2}$ . This effectively means twice the gain with a slight increase in  $C_{in}$  with the addition of  $C_{gs2}$ .

Suppose  $r_{o1} || r_{o2} = R_o$  and with  $v_{gs1} = -v_{sg2} = v_g$ , figure 5.11 can be reduced to an equivalent circuit in figure 5.12.

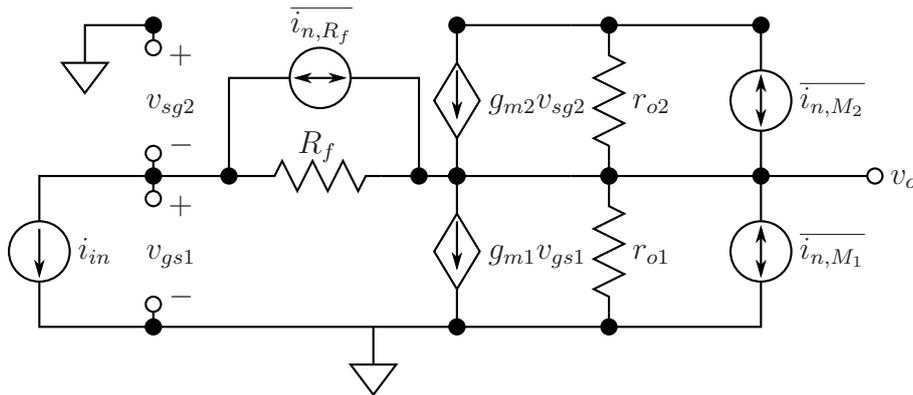


FIGURE 5.11: Small signal model of complementary feedback TIA

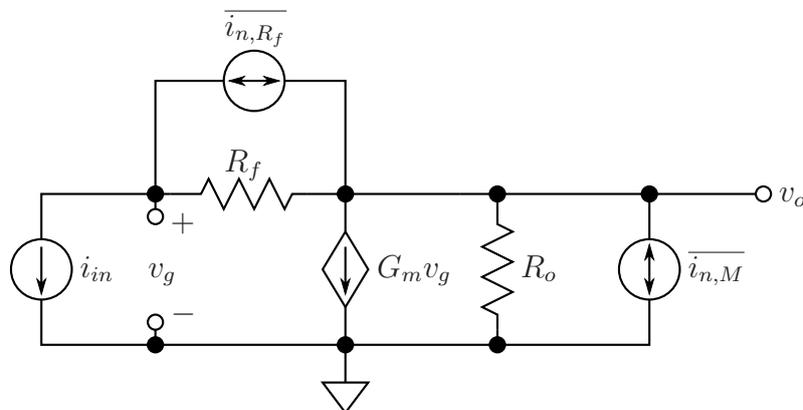


FIGURE 5.12: Simplified small signal model for complementary feedback amplifier

The gain of the system then becomes

$$R_{CF} = \frac{R_o(G_m R_f - 1)}{G_m R_o + 1} \approx R_f \tag{5.32}$$

### 5.3.6.2 Bandwidth considerations

By using the same derivation as for equations 5.24 and 5.25, the input and output resistances of the complementary feedback TIA can be described by equations 5.34 and 5.34.

$$R_{in} \approx \frac{R_f}{G_m} \tag{5.33}$$

$$R_{out} \approx \frac{1}{G_m} \tag{5.34}$$

Depending on the capacitances  $C_{in}$  and  $C_{out}$ , the maximum frequency of operation can be tailored to the requirements by sizing the devices for equations 5.34 and 5.34 to set the

dominant pole.

### 5.3.6.3 Noise considerations

The noise analysis for this configuration is the same as for the common source feedback TIA amplifier equation 5.26, which can be represented as equation 5.35. Of course,  $\overline{i_{n,M}}$  includes the noise sources of both transistors.

$$\overline{v_{n,out}} = \overline{i_{n,R_f}} \times R_f + \overline{i_{n,M}} \times \frac{1}{G_m} \quad (5.35)$$

### 5.3.6.4 Device sizing

Section 5.3.4.4 explains the principles whereby device sizing is done for both the common source and the complementary feedback TIA configurations. Therefore, the same principles apply.

### 5.3.6.5 Input signal magnitude

The input signal requirements are the same for the requirements as described in section 5.3.4.5.

### 5.3.6.6 Power consumption per amplifier

Due to the similarities in this configuration and the common gate and common source feedback amplifiers, the equations in section 5.3.3 regarding the power components remain true. Once again, the optical and electrical power components are identical to the expressions in section 5.3.3.6, where  $i_{in,max}$  refers to equation 5.27 which applies to the complementary feedback TIA as well.

## 5.4 CHOICE OF AMPLIFIER

For the choice of amplifier, the important characteristics are

1. the gain of the amplifier,
2. the bandwidth of the amplifier,
3. the noise referred to the input node as a current, and
4. implementation simplicity.

### 5.4.1 High impedance receiver

The digital buffer approach of section 5.3.2 is the topology of choice as a high impedance receiver for the following reasons:

- The components are all usually available as standard cells.
- Although the input signal requirements are more stringent, there is no need to bias the amplifier. This means that the amplifier is less sensitive to variations in process parameters and the environment.
- When compared to the common source switched configuration, the transistor count is less.
- The added input capacitance should have little effect as the dominant capacitive contributor is the photodiode.

### 5.4.2 Low impedance receiver

The best choice as an optical end point for a low impedance receiver is the complementary feedback TIA configuration of section 5.3.6. When compared to the other topologies, the following reasons become prominent:

- The output impedance is reduced, resulting in an increased drivable capacitive load.
- The input impedance is reduced as compared to that of the common source feedback TIA, although not as much as the RGC.
- The noise contribution of the transistors does not affect the input referred noise current much.
- The implementation is not excessively complex, resulting in a more robust topology.

### 5.4.3 High versus low impedance approaches

For the choice between high and low impedance topologies, the high impedance amplifier outperforms the low impedance choice both in terms of noise and bandwidth, as well as in ease of implementation with existing standard cells. Although the topology does not output true replicas of the input signal in terms of duty cycle, only the temporal transitions are necessary to be utilised as a clock.

## 5.5 DESIGN OF HIGH IMPEDANCE AMPLIFIER

According to section 4.5.5 for an electrical clock network, a repeater at the end point can drive one additional buffer for the supply of the clock to a local region. For a sensible comparison, an optical end point must therefore have the same drive capability as a repeater. Given the technology parameters  $A$  and  $B$  as derived in section 4.5.1 and the design results for 4.5.4, the drive capability that should be presented by the optical end point amplifier is known.

### 5.5.1 Design requirements

TABLE 5.1: Requirements for high impedance optical receiver front end

Parameter	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$f_{clk}$ [GHz]	4.7	5.9	7.3	9.2	11.5	14.3
$R_{out}$ [ $\Omega$ ]	51	70	74	105	128	180
$\frac{W_P}{W_N}$	2.14	2.1	1.23	1.36	1.4	1.39

### 5.5.2 Designed values

A minimum load is required for the discharge photodiode in order to minimise the nodal capacitance  $C_{in}$ . Therefore, as per table 5.2, five inverter stages are required for the 65 nm and 45 nm nodes, while four stages are sufficient for the rest. This results in a drive capability as expected from the requirements in table 5.1, and can also be checked in the output impedance results of the final inverter stage.

### 5.5.3 Power consumption per amplifier

Section 5.3.2.6 describes a model whereby the power consumption of this amplifier can be estimated, based on known capacitance values. Table 5.2 shows the input capacitances of the various inverter stages, where the first inverter I1 does not add to the switched capacitances, as it is discharged via optical means. It is therefore treated separately as in equation 5.12.

TABLE 5.2: Design parameters for high impedance amplifier across technology nodes

Parameters	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
I5 inverter device sizes						
$W_N$ [ $\mu\text{m}$ ]	13.75	8.24	-	-	-	-
$W_P$ [ $\mu\text{m}$ ]	29.42	17.31	-	-	-	-
$C_{in}$ [fF]	77.67	39.62	-	-	-	-
$R_{out}$ [k $\Omega$ ]	51	70	-	-	-	-
I4 inverter device sizes						
$W_N$ [ $\mu\text{m}$ ]	3.87	1.86	5.35	3.48	2.47	1.52
$W_P$ [ $\mu\text{m}$ ]	8.29	3.9	6.58	4.74	3.46	2.11
$C_{in}$ [fF]	21.88	8.92	15.82	10.57	7.35	4.52
$R_{out}$ [k $\Omega$ ]	180	310	74	105	128	180
I3 inverter device sizes						
$W_N$ [ $\mu\text{m}$ ]	1.091	0.418	0.64	0.49	0.37	0.25
$W_P$ [ $\mu\text{m}$ ]	2.335	0.878	0.78	0.67	0.52	0.34
$C_{in}$ [fF]	6.16	2.01	1.88	1.5	1.1	0.73
$R_{out}$ [k $\Omega$ ]	0.64	1.38	0.62	0.74	0.86	1.11
I2 inverter device sizes						
$W_N$ [nm]	307	94	76	70	55	40
$W_P$ [nm]	657	197	93	95	77	55
$C_{in}$ [aF]	1735	452	224	211	164	119
$R_{out}$ [k $\Omega$ ]	2.3	6.1	5.22	5.26	5.73	6.87
I1 inverter device sizes						
$W_N$ [nm]	130	90	64	44	32	22
$W_P$ [nm]	278	189	79	60	45	31
$C_{in}$ [aF]	735	432	189	134	95	66
$R_{out}$ [k $\Omega$ ]	5.4	6.4	6.2	8.3	9.9	12.4

TABLE 5.3: High impedance amplifier capacitive components for power calculations

Capacitance [fF]	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$C_{in}$	5.735	5.432	5.189	5.134	5.095	5.066
$C_{sw}$	107.45	51.01	17.93	12.28	8.61	5.37

### 5.5.3.1 Capacitances

The capacitance model can therefore be represented as

$$C_{in} = C_{PD} + C_{I_1} + C_{D,M_1} \quad (5.36)$$

$$C_{sw} = C_{I_2} + \dots + C_{I_N} \quad (5.37)$$

If it is assumed that  $C_{PD} = 5$  fF, which is typical of a n-well structure of roughly  $10 \mu\text{m}^2$ , and the drain capacitance of  $M_1$  is negligible, then equations 5.36 and 5.37 resolve to the values shown in table 5.3.

### 5.5.3.2 Theoretical results

As explained in section 5.1, a typical responsivity of 0.3 A/W for the photodiode is assumed throughout.

TABLE 5.4: Theoretical power consumption per amplifier

Power	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$P'_E \left[ \frac{\text{fW}}{\text{Hz}} \right]$	133.5	65.0	20.5	12.0	9.0	5.1
$P'_O \left[ \frac{\text{fW}}{\text{Hz}} \right]$	21.0	19.9	17.3	15.4	15.3	13.5
At predicted clock frequencies						
$P_E [\mu\text{W}]$	627.4	381.9	150.7	110.4	103.7	72.6
$P_O [\mu\text{W}]$	98.8	117.0	127.0	141.4	175.4	193.8

### 5.5.3.3 Simulated results

By using the predictive models of section 4.1.1, simulations done in Spectre™ are shown in table 5.5. The differences can be attributed to

1. the lack of drain node capacitances in this model, and

2. the absence of the short circuit currents on transitions, which are taken into account in simulations.

TABLE 5.5: Simulated power components per amplifier

Power [ $\mu$ W]	65 nm	45 nm	32 nm	22 nm	16 nm	11 nm
$f_{clk}$ [GHz]	4.7	5.9	7.3	9.2	11.5	14.3
$P_E$	1299.9	1142.0	698.9	359.5	393.4	385.5
$\Delta P_E$	672.6	760.1	548.2	249.1	289.7	312.9
$P_O$	101.6	121.7	132.7	149.5	187.9	210.8
$\Delta P_O$	2.8	4.7	5.7	8.1	12.5	17.0

# CHAPTER SIX

## COMPARISON OF NETWORK POWER CONSUMPTION

---

### 6.1 OVERVIEW

BASED on the tree depth model of section 4.3 as well as the electrical components from sections 4.5.4, 4.5.6 and 5.5, an estimation can be made on what is to be expected from both electrical and optical clock networks. As the local region distribution will still remain electrical, these regions as well as the amplifiers driving them are regarded as common components. Subsequently, these are not contributory to a sensible comparison, and the rest of this chapter will focus on only the mutual exclusive components of electrical and optical networks.

### 6.2 EXPLANATION OF RESULTS

For all the nodal results, a table and a graph is given as a representation of the clock network performance in the specific technology. For example, figure 6.1 (a) shows the power consumed at different levels of the tree with an assumed 100 % EPE (in other words, all input power is converted to optical power and available at the detector), where

- $P_{E_{global,intc}}$  is the global interconnect component,
- The term  $P_{E_{rep}} + P_{E_{split}}$  represents the repeaters and split buffer contributions,
- $P_{E_{total}}$  is the total power consumption of an electrical tree at the specific level.

For the optical network,

- $P_{O_{electrical}}$  is the electrical power consumption of the amplifiers at the end points,

- $P_{Optical}$  represents the power required in an optical form at the end point, and
- $P_{Total}$  is the total power dissipation for an optical tree at the specific level.

Within section (b) of the figures, the minimum required EPE for the network is shown. This value for EPE is based on a lossless system (see section 2.6) and represents the value of efficiency where the optical network will equal the electrical network in terms of total input power. That is, this number is representative on how much overhead in terms of power is available for an optical system before it becomes less power consuming than an electrical system.

The quantity  $P_{total}$  represents the total power consumption of the network, assuming 100 % EPE, when an optical tree is implemented up to level  $n$ . Therefore, where  $n = 0$ , the network is a full electrical tree and  $P_{total}$  represents the total power consumed in the network. When  $n$  is equal to the maximum level for the relevant node, the tree is fully optical and  $P_{total}$  represents the total (optical) power consumed in the network.

## 6.3 65 NM NODE

### 6.3.1 Tree design

Table 6.1 shows an overview of the attributes used to estimate the power consumption of a tree at the 65 nm technology node, along with the relevant sections from where the attributes are taken.

### 6.3.2 Power dissipation components at different levels

From figure 6.1 (a) it becomes clear that the electrical components - repeaters, split buffers and optical receiver front ends - are the dominant consumers of power throughout all levels. For the electrical network, the rise of component dissipation increases more rapidly than for the interconnect component. The optical power consumption rise is proportional to the electrical front end consumption simply because of the direct relationship in terms of end point count.

It is interesting to note that the interconnect power consumption does not increase as rapidly as the optical power component. This means that the benefit of an optical distribution scheme is less prominent at deeper tree depths.

TABLE 6.1: Summary of attributes for the 65 nm technology node

Clock frequency $f_{clk}$ [GHz]	4.7	from section 4.1
Supply voltage $V_{DD}$ [V]	1.1	from table 4.1
Global interconnect capacitance $\left[\frac{F}{\mu m}\right]$	2E-16	from table 4.4
Tree depth $n$	6	from section 4.3
Logic area length $L$ [ $\mu m$ ]	16901.4	from sections 3.2.1, 4.2
Maximum segment length [ $\mu m$ ]	420	from section 4.5.4
Theoretical $P_{rep}$ [ $\mu W$ ]	551.5	from table 4.14
Simulated $P_{rep}$ [ $\mu W$ ]	1011.0	from table 4.15
Theoretical $P_{split}$ [ $\mu W$ ]	102.7	from table 4.17
Simulated $P_{split}$ [ $\mu W$ ]	176.4	
Photodiode capacitance [fF]	5	from section 5.1
Photodiode responsivity [A/W]	0.3	
Theoretical $P_O$ [ $\mu W$ ]	98.8	from table 5.4
Theoretical $P_E$ [ $\mu W$ ]	627.4	
Simulated $P_O$ [ $\mu W$ ]	101.6	from table 5.5
Simulated $P_E$ [ $\mu W$ ]	1299.9	

Figure 6.1 (b) shows that, for the 65 nm node, the overall power consumption becomes less as the tree becomes more optical. The price, however, is an increased demand on an efficient emitter. Although the total power consumed,  $P_{total}$ , becomes less as the tree becomes more optical, the assumption is that the EPE is 100 %. The figure also shows the minimum required EPE for the tree to remain as efficient as a fully electrical tree. As the tree depth increases, the EPE requirement becomes more stringent.

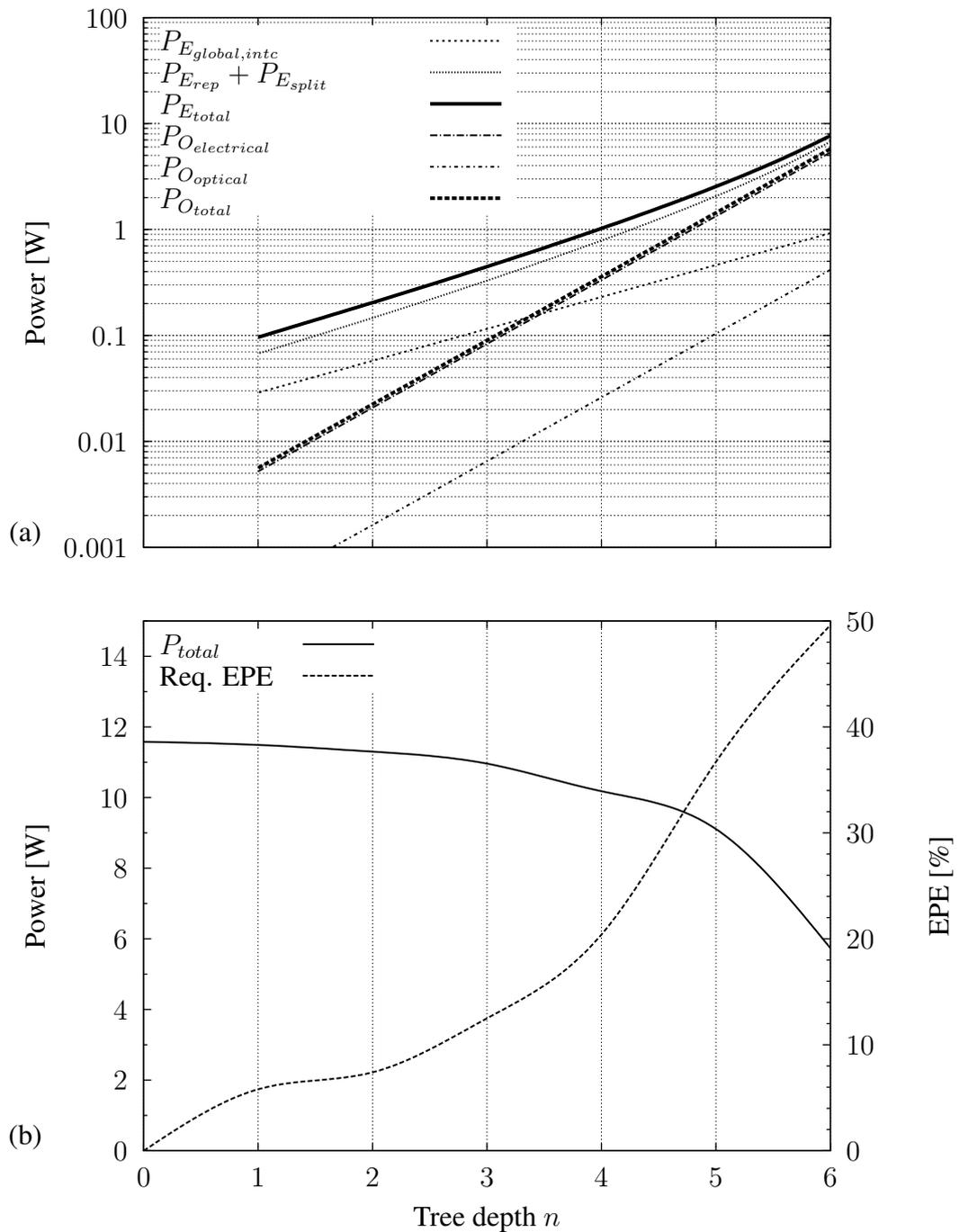


FIGURE 6.1: Combination of electrical and optical tree for the 65 nm node

## 6.4 45 NM NODE

### 6.4.1 Tree design

Table 6.2 shows an overview of the attributes used to estimate the power consumption of a tree at the 45 nm technology node.

TABLE 6.2: Summary of attributes for the 45 nm technology node

Clock frequency $f_{clk}$ [GHz]	5.9	from section 4.1
Supply voltage $V_{DD}$ [V]	1.1	from table 4.1
Global interconnect capacitance $\left[\frac{F}{\mu m}\right]$	1.8E-6	from table 4.4
Tree depth $n$	6	from section 4.3
Logic area length $L$ [ $\mu m$ ]	16928.2	from sections 3.2.1, 4.2
Maximum segment length [ $\mu m$ ]	266	from section 4.5.4
Theoretical $P_{rep}$ [ $\mu W$ ]	354.8	from table 4.14
Simulated $P_{rep}$ [ $\mu W$ ]	787.7	from table 4.15
Theoretical $P_{split}$ [ $\mu W$ ]	61.0	from table 4.17
Simulated $P_{split}$ [ $\mu W$ ]	130.6	
Photodiode capacitance [fF]	5	from section 5.1
Photodiode responsivity [A/W]	0.3	
Theoretical $P_O$ [ $\mu W$ ]	117.0	from table 5.4
Theoretical $P_E$ [ $\mu W$ ]	381.9	
Simulated $P_O$ [ $\mu W$ ]	121.7	from table 5.5
Simulated $P_E$ [ $\mu W$ ]	1142.0	

### 6.4.2 Power dissipation components at different levels

Figure 6.2 shows the level dependent power dissipation components, where the symbols are elaborated on in section 6.2. The results are similar to the 65 nm node, with the gap between the optical consumption and the interconnect consumption becoming smaller. This means that the performance of the optical network weakens as the tree deepens.

Note that the difference between  $P_{O_{electrical}}$  and  $P_{O_{total}}$  is more than in the 65 nm node, although still dominating the overall optical consumption. This is partly due to the fact that the photodiode does not scale well, resulting in an almost constant optical power required to

charge the input of the amplifier in section 5.5. As will be seen, this is probably one of the limiting aspects of optical networks in future nodes.

As the tree depth is the same as the 65 nm node, there are not big differences in the overall power consumption and, because the devices are smaller and switched capacitances are less, the network performs slightly better than the 65 nm version.

## 6.5 32 NM NODE

### 6.5.1 Tree design

Table 6.3 shows an overview of the attributes used to estimate the power consumption of a tree at the 32 nm technology node.

TABLE 6.3: Summary of attributes for the 32 nm technology node

Clock frequency $f_{clk}$ [GHz]	7.3	from section 4.1
Supply voltage $V_{DD}$ [V]	1.0	from table 4.1
Global interconnect capacitance $\left[\frac{F}{\mu m}\right]$	1.7E-16	from table 4.4
Tree depth $n$	7	from section 4.3
Logic area length $L$ [ $\mu m$ ]	16947.0	from sections 3.2.1, 4.2
Maximum segment length [ $\mu m$ ]	193.6	from section 4.5.4
Theoretical $P_{rep}$ [ $\mu W$ ]	166.2	from table 4.14
Simulated $P_{rep}$ [ $\mu W$ ]	396.6	from table 4.15
Theoretical $P_{split}$ [ $\mu W$ ]	19.9	from table 4.17
Simulated $P_{split}$ [ $\mu W$ ]	46.3	
Photodiode capacitance [fF]	5	from section 5.1
Photodiode responsivity [A/W]	0.3	
Theoretical $P_O$ [ $\mu W$ ]	127.0	from table 5.4
Theoretical $P_E$ [ $\mu W$ ]	150.7	
Simulated $P_O$ [ $\mu W$ ]	132.7	from table 5.5
Simulated $P_E$ [ $\mu W$ ]	698.9	

### 6.5.2 Power dissipation components at different levels

Figure 6.3 shows the level dependent power dissipation components, where the symbols are elaborated on in section 6.2. One of the big differences present in this node is the jump from  $n_{max} = 6$  levels to  $n_{max} = 7$ . This clear in terms of a sudden jump in the power consumption of the tree.

The interconnect consumption equals the optical consumption at  $n = 7$ , clear from part (a) of the figure. Once again, the optical consumption will roughly remain the same as the input power is dictated by the amplifier input capacitance, dominated by the unscaled photodiode capacitance. At this point, the interconnect outperforms the optical alternative in terms of power consumption.

The total power in a hybrid tree, shown in figure 6.3 (b) as  $P_{total}$ , still shows that an optical tree of any depth consumes less total power than an electrical tree, given a perfect EPE. Note, however, that the requirement on EPE has increased to a minimum of 70 % at  $n = 7$ , which requires a very efficient emitter.

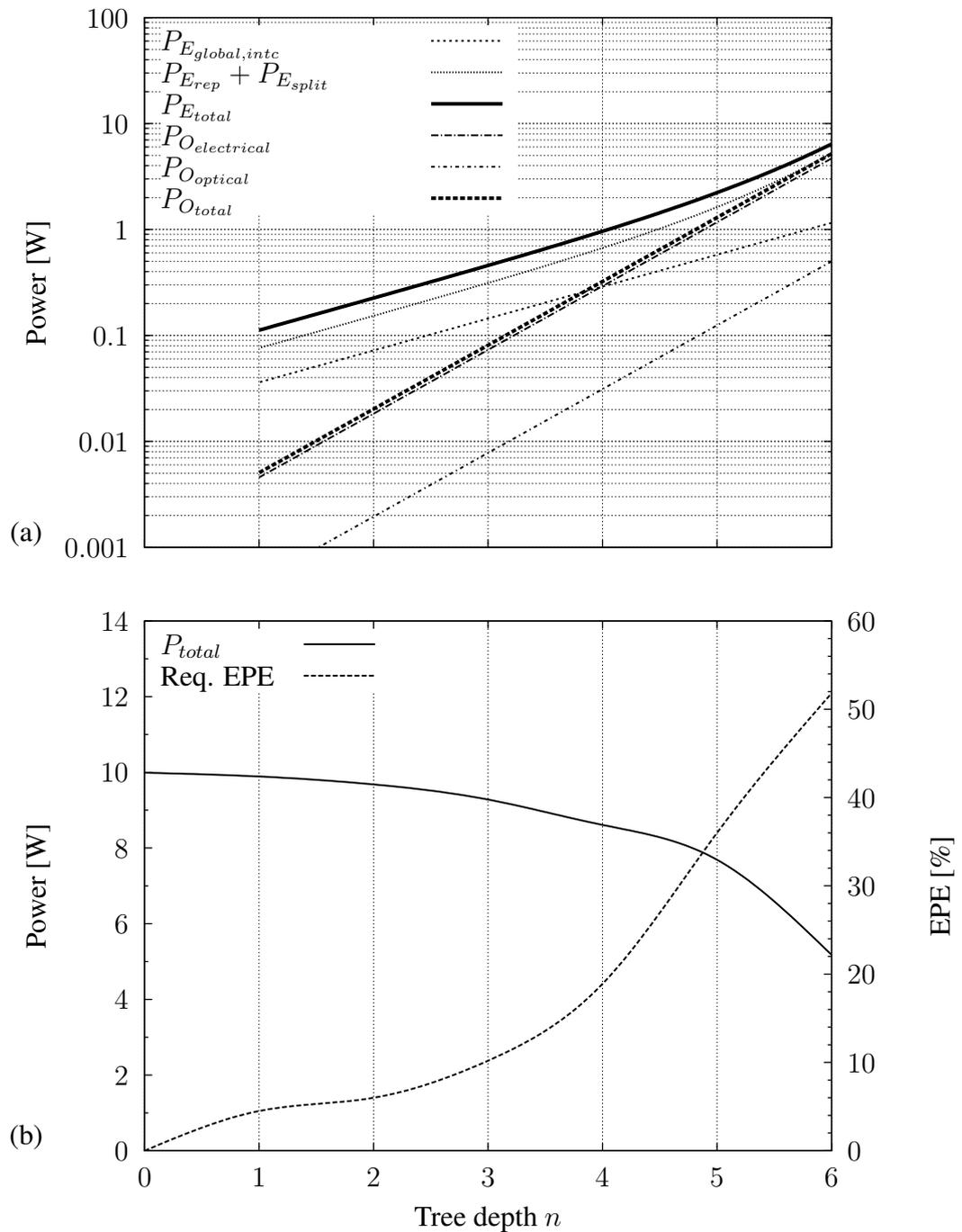


FIGURE 6.2: Combination of electrical and optical tree for the 45 nm node

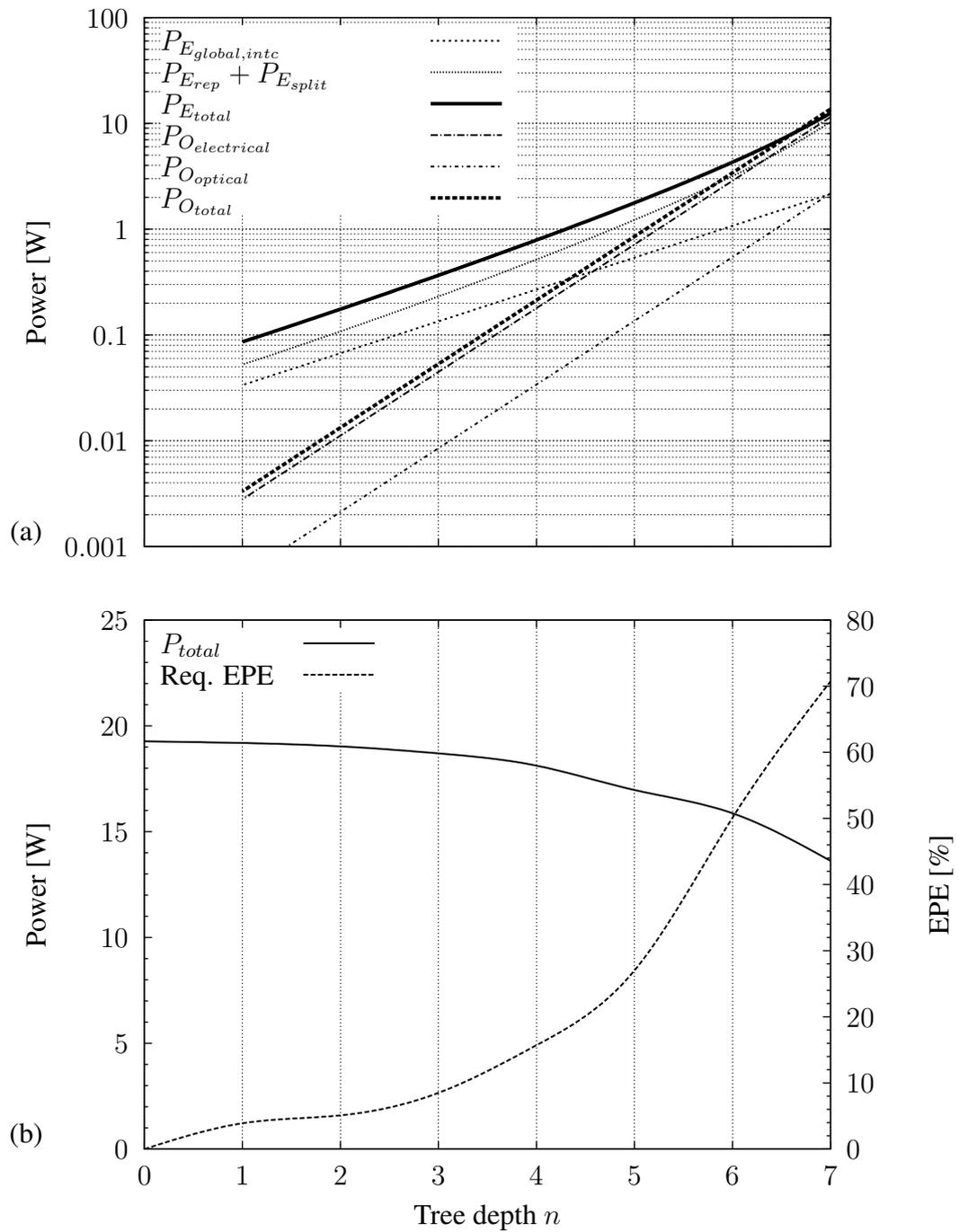


FIGURE 6.3: Combination of electrical and optical tree for the 32 nm node

## 6.6 22 NM NODE

### 6.6.1 Tree design

Table 6.4 shows an overview of the attributes used to estimate the power consumption of a tree at the 22 nm technology node.

TABLE 6.4: Summary of attributes for the 22 nm technology node

Clock frequency $f_{clk}$ [GHz]	9.2	from section 4.1
Supply voltage $V_{DD}$ [V]	0.9	from table 4.1
Global interconnect capacitance $\left[\frac{F}{\mu m}\right]$	1.5E-16	from table 4.4
Tree depth $n$	7	from section 4.3
Logic area length $L$ [ $\mu m$ ]	16960.4	from sections 3.2.1, 4.2
Maximum segment length [ $\mu m$ ]	126.6	from section 4.5.4
Theoretical $P_{rep}$ [ $\mu W$ ]	108.6	from table 4.14
Simulated $P_{rep}$ [ $\mu W$ ]	110.0	from table 4.15
Theoretical $P_{split}$ [ $\mu W$ ]	14.3	from table 4.17
Simulated $P_{split}$ [ $\mu W$ ]	28.0	
Photodiode capacitance [fF]	5	from section 5.1
Photodiode responsivity [A/W]	0.3	
Theoretical $P_O$ [ $\mu W$ ]	141.4	from table 5.4
Theoretical $P_E$ [ $\mu W$ ]	110.4	
Simulated $P_O$ [ $\mu W$ ]	149.5	from table 5.5
Simulated $P_E$ [ $\mu W$ ]	359.5	

### 6.6.2 Power dissipation components at different levels

Figure 6.4 shows the level dependent power dissipation components, where the symbols are elaborated on in section 6.2. At 22 nm, the tree depth remains at 7 levels. As a result, both the electrical and optical power consumption improves compared to 32 nm.

The results are similar to the 32 nm, with the optical network still outperforming the electrical network for all depths of an optical tree.

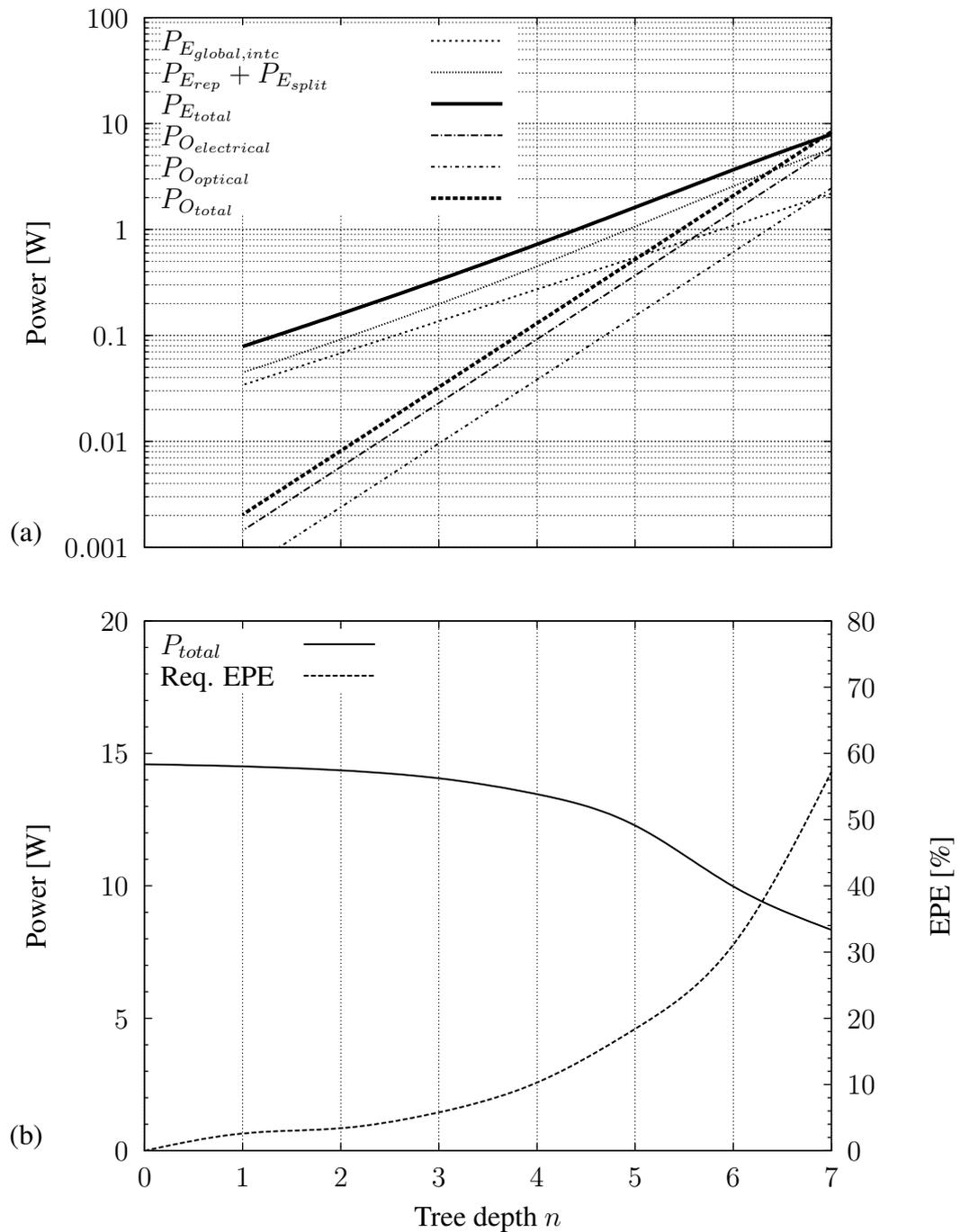


FIGURE 6.4: Combination of electrical and optical tree for the 22 nm node

## 6.7 16 NM NODE

### 6.7.1 Tree design

Table 6.5 shows an overview of the attributes used to estimate the power consumption of a tree at the 16 nm technology node.

TABLE 6.5: Summary of attributes for the 16 nm technology node

Clock frequency $f_{clk}$ [GHz]	11.5	from section 4.1
Supply voltage $V_{DD}$ [V]	0.9	from table 4.1
Global interconnect capacitance $\left[\frac{F}{\mu m}\right]$	1.5E-16	from table 4.4
Tree depth $n$	8	from section 4.3
Logic area length $L$ [ $\mu m$ ]	16969.8	from sections 3.2.1, 4.2
Maximum segment length [ $\mu m$ ]	84.5	from section 4.5.4
Theoretical $P_{rep}$ [ $\mu W$ ]	94.0	from table 4.14
Simulated $P_{rep}$ [ $\mu W$ ]	126.9	from table 4.15
Theoretical $P_{split}$ [ $\mu W$ ]	13.0	from table 4.17
Simulated $P_{split}$ [ $\mu W$ ]	29.7	
Photodiode capacitance [fF]	5	from section 5.1
Photodiode responsivity [A/W]	0.3	
Theoretical $P_O$ [ $\mu W$ ]	175.4	from table 5.4
Theoretical $P_E$ [ $\mu W$ ]	103.7	
Simulated $P_O$ [ $\mu W$ ]	187.9	from table 5.5
Simulated $P_E$ [ $\mu W$ ]	393.4	

### 6.7.2 Power dissipation components at different levels

Figure 6.5 shows the level dependent power dissipation components, where the symbols are elaborated on in section 6.2. As the level increased to  $n_{max} = 8$ , the power consumption for this node is considerably higher than for the previous nodes.

An interesting observation is made when looking at figure 6.5 (a). Beyond 7 levels, the total optical power consumption,  $P_{O_{total}}$ , exceeds the electrical consumption. Although this is not yet fully reflected in part (b) of the figure with  $P_{total}$ , the minimum required EPE is much worse as a consequence.  $P_{total}$  in part (b) is almost equal for levels 7 and 8.

The gap between  $P_{O_{electrical}}$  and  $P_{O_{total}}$  is also widening, meaning that the optical power required for the clock network becomes more in relation to the component dissipation. Once again, this is due to the fixed nature of the optical front end input capacitance.

## 6.8 11 NM NODE

### 6.8.1 Tree design

Table 6.6 shows an overview of the attributes used to estimate the power consumption of a tree at the 11 nm technology node.

TABLE 6.6: Summary of attributes for the 11 nm technology node

Clock frequency $f_{clk}$ [GHz]	14.3	from section 4.1
Supply voltage $V_{DD}$ [V]	0.8	from table 4.1
Global interconnect capacitance $\left[\frac{F}{\mu m}\right]$	1.3E-16	from table 4.4
Tree depth $n$	9	from section 4.3
Logic area length $L$ [ $\mu m$ ]	16977.2	from sections 3.2.1, 4.2
Maximum segment length [ $\mu m$ ]	56.1	from section 4.5.4
Theoretical $P_{rep}$ [ $\mu W$ ]	56.5	from table 4.14
Simulated $P_{rep}$ [ $\mu W$ ]	157.2	from table 4.15
Theoretical $P_{split}$ [ $\mu W$ ]	8.2	from table 4.17
Simulated $P_{split}$ [ $\mu W$ ]	30.0	
Photodiode capacitance [fF]	5	from section 5.1
Photodiode responsivity [A/W]	0.3	
Theoretical $P_O$ [ $\mu W$ ]	193.8	from table 5.4
Theoretical $P_E$ [ $\mu W$ ]	72.6	
Simulated $P_O$ [ $\mu W$ ]	210.8	from table 5.5
Simulated $P_E$ [ $\mu W$ ]	385.5	

### 6.8.2 Power dissipation components at different levels

Figure 6.6 shows the level dependent power dissipation components, where the symbols are elaborated on in section 6.2.

At  $n_{max} = 9$ , the global network consumption becomes unrealistically high. Even the optical power  $P_{O_{optical}}$  is greater than 50 W at  $n = 9$ . Similar to the 16 nm node, the total power consumption of an optical network,  $P_{O_{total}}$ , exceeds the total power consumption for an electrical tree,  $P_{E_{total}}$ , when  $n \geq 8$ .

As for the required EPE, a value beyond 100 % is clearly impossible, which means that a fully optical network cannot outperform a fully electrical network even with perfect source EPE. When adding link and coupling losses, it seems that an optical network is clearly undesirable at such deep tree levels.

However, an interesting observation is made when looking at the total power consumed in a hybrid network (shown in figure 6.6 (b) as  $P_{total}$ ) at different optical network depths. The minimum power consumption is obtained at  $n = 7$  for an optical network, while levels 8 and 9 remain electrical. This means that a hybrid network results in the lowest power consumption possibility at this technology node. With a minimum EPE of just above 50 %, a total power consumption of  $P_{total} = 137.44$  W is possible.

Although the power levels are very high, an important conclusion might be that, in order to find optical clock networks feasible, it might be worthwhile to consider hybrid systems. At lower tree levels, the optical methods beat the interconnect dissipation, while at deeper levels, the power hungry end points are best kept in the electrical domain.

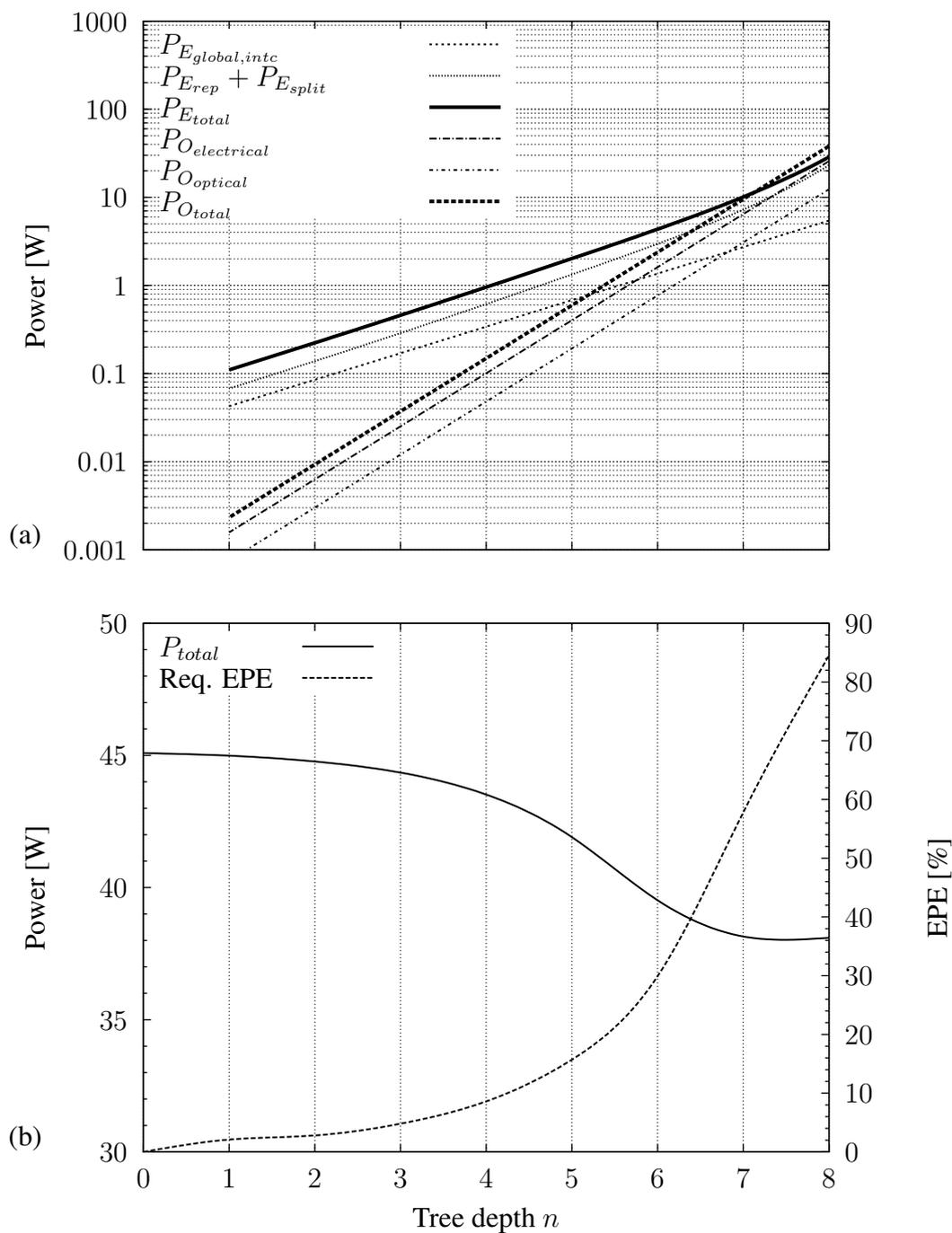


FIGURE 6.5: Combination of electrical and optical tree for the 16 nm node

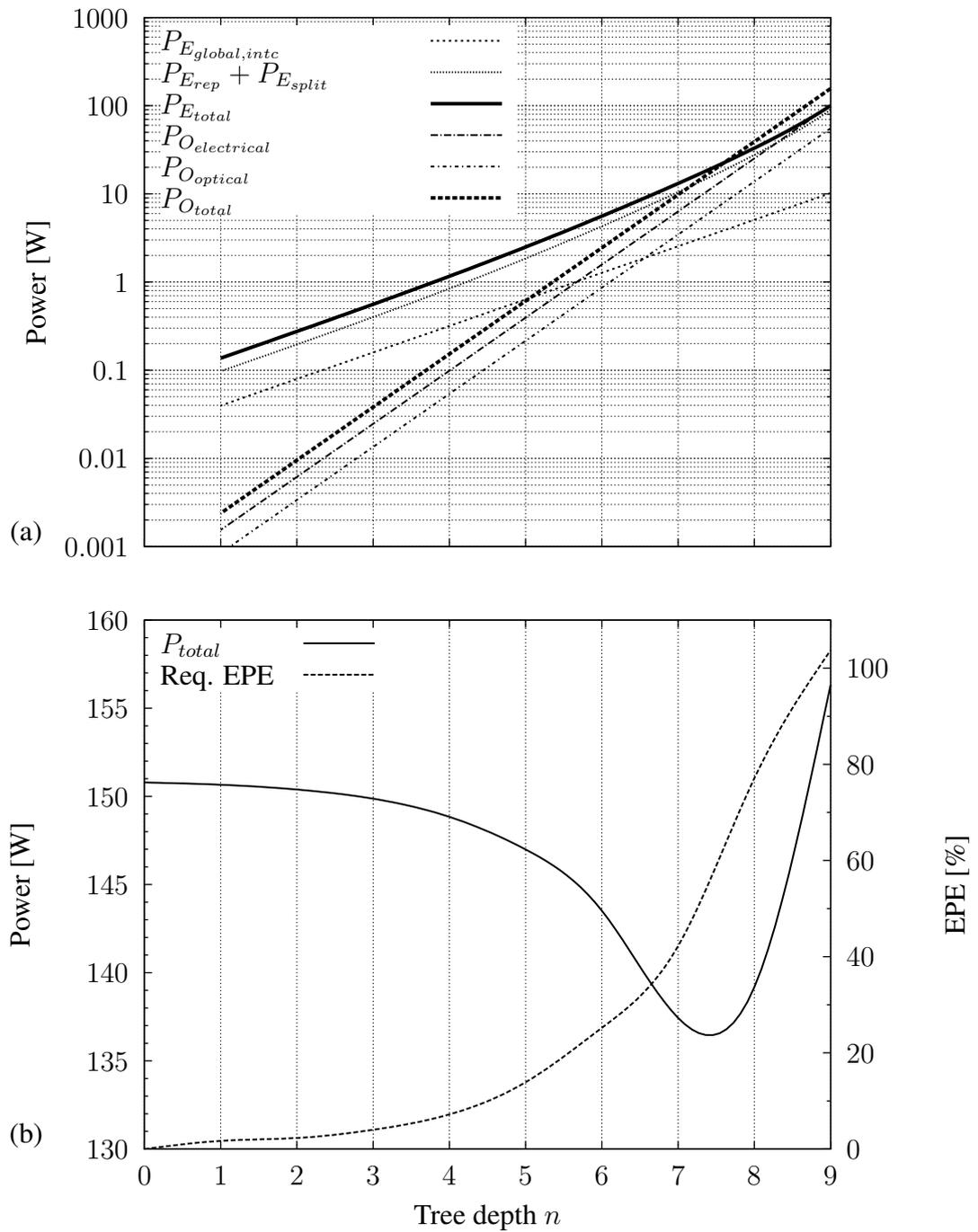


FIGURE 6.6: Combination of electrical and optical tree for the 11 nm node

## 6.9 HYBRID APPROACH

Sections 6.3 to 6.8 show that, although an optical clock network can be more efficient in most cases, a hybrid approach (section 6.8) might offer the best solution for minimum power consumption. That is, an optical network will extend into a certain tree depth, after which an electrical network takes over. The 11 nm serves as a good example, where it is clear that, from figure 6.6, optical end points at level 7 with electrical distribution through levels 8 and 9 will yield better power performance than a fully optical implementation. A comparison between all electrical, all optical and hybrid approaches is shown in section 6.10.

## 6.10 POWER CONSUMPTION ACROSS NODES

The power dissipation for the following three cases is shown in figure 6.7 (from bottom to top):

1. The individual ( $P_{E_{global,intc}}$ ,  $P_{E_{components}}$ ) and total ( $P_{E_{total}}$ ) electrical power consumption components for the global network across future nodes,
2. the individual ( $P_{O_{electrical}}$ ,  $P_{E_{optical}}$ ) and total ( $P_{O_{total}}$ ) optical power consumption for a fully optical global network, and
3. a comparison between electrical, optical and hybrid networks.

The hybrid network as shown in figure 6.7 is the lowest obtained power consumption for the node. Thus, if a fully optical network outperforms an electrical network at all tree levels, then the hybrid represents a fully optical tree. The resulting graph in figure 6.7 (a) therefore shows that the hybrid network differs from the fully optical implementation marginally at 16 nm and outperforms both the optical and electrical networks at 11 nm.

Part (b) and (c) shows an overview of the power dissipation in optical and electrical networks respectively. As is clear from both (b) and (c), wherever the tree level is constant when going to a new technology, the component power consumption decreases for both electrical and optical cases.

Further discussion, along with the limitations of this work, can be found in chapter 7.

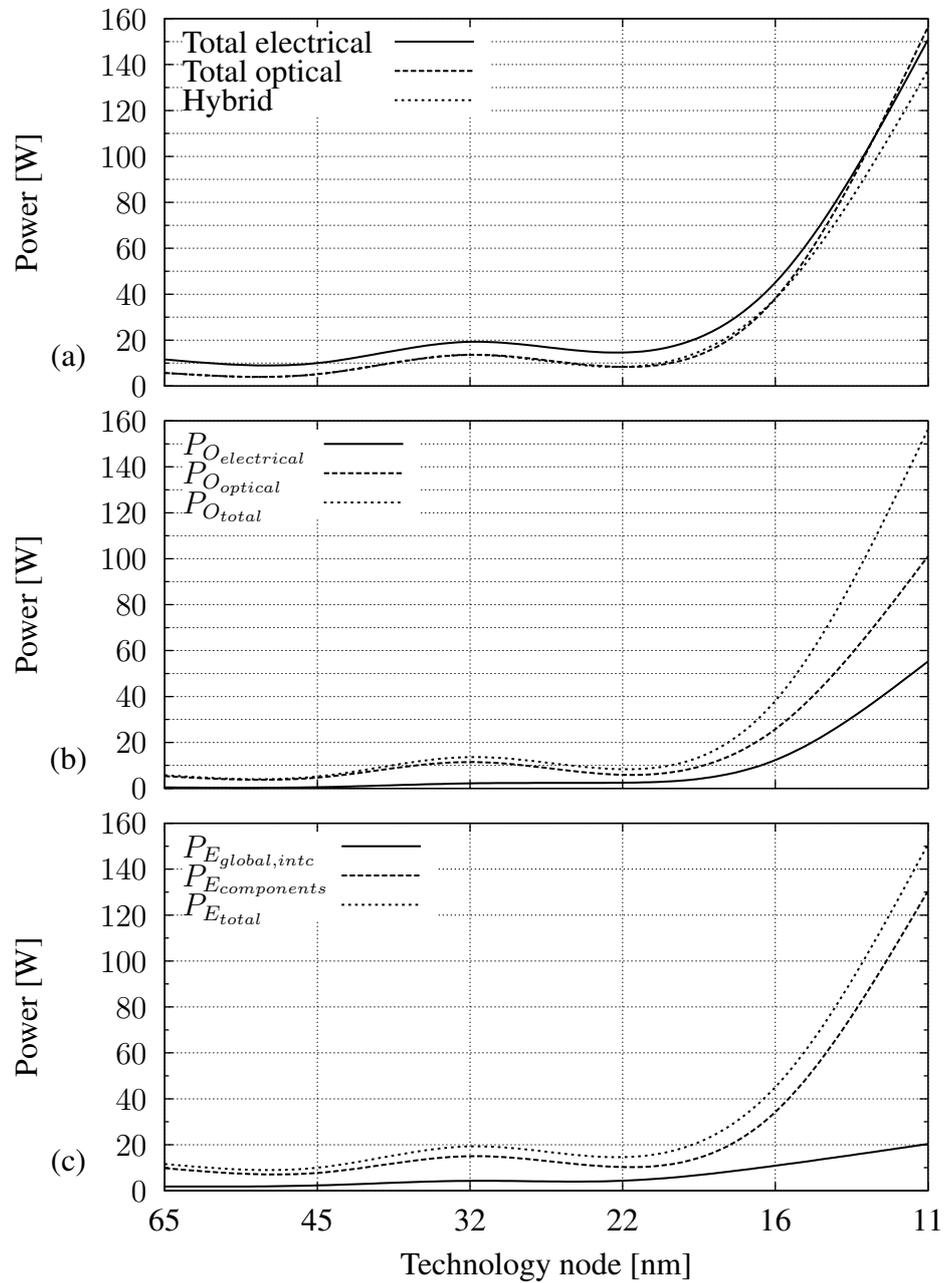


FIGURE 6.7: Results of power consumption comparison across nodes

# CHAPTER SEVEN

## CONCLUSION

---

### 7.1 INTERPRETATION OF RESULTS

#### 7.1.1 Overall power consumption performance

Figure 6.7 shows the promise of having optical clock networks as an alternative. What becomes apparent, though, is the fact that a fully optical solution might not necessarily be the answer. It seems that, at the 11 nm node, the optical network cannot continue to outperform the electrical version. The most prominent reasons are:

1. The electrical power dissipation in the receivers are more than the electrical repeaters.
2. The photodiode capacitance is assumed to be constant, which places a non-scalable burden on the optical front end.
3. The H-tree only has the ability to segregate into a number of regions with a base of 2, which is suboptimal, as will be discussed in section 7.2.

#### 7.1.2 Requirements on the source

From the individual sections 6.3 through 6.8, the requirement on the light source is also given as a metric of external power efficiency (EPE) *without losses*. This means that the losses discussed in section 2.6 needs to be accounted for in the input signal, resulting in a higher EPE for the same results. Note that the EPE metric is used to show what will be required for the optical network to equal the power consumption of the electrical network.

It is clear, depending on the source, that a choice of a hybrid network (refer to section 6.9) might lessen the burden on the required EPE with marginal increases in overall power

consumption.

### 7.1.3 Dependence on operating frequency

The question of the dependence of the above results on operating frequency can briefly be answered by noting that the clock frequency has a linear relationship on the power consumption of both electrical and optical systems.

For electrical systems, the power consumption is based on the equation in section 3.6 for capacitive consumption. For obvious reasons, the interconnect capacitance, capacitance of repeaters (see equation 4.24) and the short circuit power dissipation is therefore directly proportional to the operating frequency.

For optical systems, the motivation for the electrical component consumption is the same as per the above explanation. The optical power will also scale linearly with an increase in frequency, since the amplifier choice is a high impedance technology, with equations 5.36 and 5.37 remaining valid.

If a low impedance approach was chosen, the operating frequency would again reflect in the power consumption of the detection front end in order to maintain the signal bandwidth required.

The results therefore remain valid as equation 7.1 will hold.

$$P \propto f_{clk} \quad (7.1)$$

## 7.2 LIMITATIONS AND SHORTCOMINGS OF THE WORK

### 7.2.1 Predictive work

This work is solely based on predictions. Although the sources are of sound origin, a predictive work can never be regarded as definite or even accurate. The SPICE models used throughout the work has been updated to roughly match the requirements of [43], the operation thereof remains speculative.

### 7.2.2 Photodiode limitations

The capacitance of the photodiode was estimated in section 5.1 as a  $4 \mu\text{m}^2$  element. As briefly stated in section 5.3.1.7, the optical power density per unit area is never taken into

account, and will be a limiting factor on both the source and waveguides. What is possible remains to be explored.

### 7.2.3 H-tree branching

One of the prominent reasons why the H-tree performs bad at high tree depths is the fact that the tree will always divide as a power to the base of two (see section 3.2.1). This is especially prominent at the 11 nm node, where the required number of end points to adhere to the skew constraint of section 4.3 is 74473, but level 8 is just short of that amount (65536) while level 9 far exceeds the requirement with 262144. This results in a lot of extra power being wasted in both the electrical and optical networks.

### 7.2.4 Interconnect model

It is assumed throughout this work that the interconnects used are sized at minimum dimensions, with minimum pitched ground conductors surrounding the tracks. This, of course, is not necessary, but depends on how closely the interconnects will need to be spaced. At larger nodes, this might not be difficult.

### 7.2.5 Skew and jitter

Skew and jitter analyses were limited to what was necessary to determine the tree depth. The main concern in this work was the power consumption and the signal fidelity in terms of transition definition. Although skew was briefly analysed for the local regions, no process and environmental variations were taken into account. Jitter was not qualitatively examined, but rather implicitly taken into account with noise calculations on the optical receivers.

### 7.2.6 Optimisation of electrical tree

Optimisations were done for the repeaters to drive a maximum segment of continuous interconnect. While this is valid for shallow tree depths, the maximum segment lengths reduce as the tree depth increases. Ideally, an optimisation algorithm should size the repeaters at each level of the tree, depending on the surroundings. The effect of neglecting this is very prominent in the amount of power consumed by the repeaters and split buffers when  $n$  increases, as shown in figures 6.1 through 6.6. Optimisation of a clock distribution

network is a non-trivial task and a formidable piece of research work on its own. Therefore, this work did not pursue the venture to completely optimise the electrical networks.

### 7.2.7 Global clock frequency model

It might be worth noting that the global clock frequency is assumed to be the same as the local region clock frequency. Techniques not explored include the multiplication of a lower frequency global clock for local regions. Such schemes require an in depth analysis of the resulting increase in jitter, a definite side effect of frequency multiplication, and skew between local region clocks.

## 7.3 CONTRIBUTION OF THIS WORK

1. A good model was developed to estimate the transient response of simple photodiodes. The intrinsic bandwidth can, through this model, be engineered for high speed operation, given that the designer has freedom regarding the source.
2. It is the author's opinion that one of the key contributions of this work is to use SPICE models to augment the power consumption models as used in comparing the two types of clock technologies.
3. A lot of effort has been spent on making sure that the electrical networks, used for comparison, is not unrealistic or underestimated in terms of performance. Optimisation techniques were shown which can be extended, as mentioned in section 7.2, to improve even further on the performance of electrical networks.
4. A thorough investigation was done in trying to identify the best optical front end with the specific goal of clock signal detection. This investigation yielded an option which has good noise characteristics, fast response and a low development burden.

## 7.4 FUTURE RESEARCH POSSIBILITIES IDENTIFIED IN THIS WORK

From sections 6.3 through 6.8, two big problems can be identified for future nodes:

1. The dominant power consumption terms are strongly dependent on dissipation in components such as repeaters and optical amplifiers.

2. The required optical power remains almost constant across nodes due to fundamental detector limitations such as dimensions and realistic optical power densities.

Future work might entail

- focusing on optimising the component count *at each tree level* in order to minimise overall component counts,
- optimising the optical detection circuitry for better power consumption,
- looking into higher gain photodetection principles compatible with CMOS such as APDs,
- investigate the prominence of skew difference between electrical and optical networks, and
- optical components such as emitters and interconnects which might be utilised in an optical clock distribution scheme.

## 7.5 CONCLUSION

For optical clock distribution schemes to become a mainstream alternative, a lot of investigation is still to be done. However, this work proves that, although the power consumption situation seems grim for future nodes, the optical networks perform well on equal grounds compared to electrical networks, which is definitely worth the motivation for continuing research in this field.

# REFERENCES

- [1] International Technology Roadmap for Semiconductors, “ITRS 2007 edition: Interconnects,” <http://public.itrs.net/>, 2007.
- [2] ITRS 2008 updated report. [Online]. Available: <http://www.itrs.net/>
- [3] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronics Magazine*, vol. 38, no. 8, April 1965.
- [4] M. J. Riezenman, “Wanlass’s CMOS circuit,” *IEEE Spectrum*, vol. 28, no. 5, p. 44, May 1991.
- [5] D. Foty and G. Gildenblat, “CMOS scaling theory - why our “theory of everything” still works, and what that means for the future,” *12th International Symposium on Electron Devices for Microwave and Optoelectronic Applications*, pp. 27–38, Nov 2004.
- [6] R. Dennard, F. Gaensslen, V. Rideout, E. Bassous, and A. LeBlanc, “Design of ion-implanted MOSFET’s with very small physical dimensions,” *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct 1974.
- [7] B. Hoeneisen and C. A. Mead, “Fundamental limitations in microelectronics - I. MOS technology,” *Solid-State Electronics*, vol. 15, no. 7, pp. 819–829, July 1972.
- [8] J. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. Souri, K. Banerjee, K. Saraswat, A. Rahman, R. Reif, and J. Meindl, “Interconnect limits on gigascale integration (GSI) in the 21st century,” *Proceedings of the IEEE*, vol. 89, no. 3, pp. 305–324, Mar 2001.
- [9] A. Duffy. (2000, August) Diffraction; thin-film interference. [Online]. Available: <http://physics.bu.edu/~duffy/PY106/Diffraction.html>
- [10] S. B. Alexander, *Optical Communication Receiver Design*, ser. SPIE Tutorial Texts in Optical Engineering. Bellingham, Washington, USA and London, UK: SPIE Press and IEE, 1997, vol. TT22.
- [11] H. Tsuji, N. Arai, M. Motono, Y. Gotoh, K. Adachi, H. Kotaki, and J. Ishikawa, “Study on optical reflection property from multilayer on Si substrate including nanoparticles in SiO<sub>2</sub> layer,” *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 206, pp. 615–619, 2003/5.
- [12] P. M. Schneider and W. B. Fowler, “Band structure and optical properties of silicon dioxide,” *Phys. Rev. Lett.*, vol. 36, no. 8, pp. 425–428, Feb 1976.

- [13] V. A. Gritsenko, N. D. Dikovskaja, and K. P. Mogilnikov, "Band diagram and conductivity of silicon oxynitride films," *Thin Solid Films*, vol. 51, no. 3, pp. 353–357, Jun 1978.
- [14] E. D. Palik, *Handbook of Optical Constants of Solids*, 1st ed. Academic Press, 1997, vol. 1.
- [15] M. Law, E. Solley, M. Liang, and D. Burk, "Self-consistent model of minority-carrier lifetime, diffusion length, and mobility," *IEEE Electron Device Letters*, vol. 12, no. 8, pp. 401–403, Aug 1991.
- [16] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon," *IEEE Transactions on Electron Devices*, vol. 30, no. 7, pp. 764–769, Jul 1983.
- [17] S. Radovanović, A.-J. Annema, and B. Nauta, *High-speed photodiodes in standard CMOS technology*. Springer, 2006.
- [18] G. W. Neudeck, *The PN Junction Diode*, 2nd ed., ser. Modular Series on Solid State Devices. Addison-Wesley, 1989, vol. 2.
- [19] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. John Wiley & Sons, Inc., 1981.
- [20] R. F. Pierret, *Semiconductor Fundamentals*, 2nd ed., ser. Modular Series on Solid State Devices. Addison-Wesley, 1988, vol. 1.
- [21] J. Gan, L. Wu, H. Luan, B. Bihari, and R. Chen, "Two-dimensional 45 surface-normal microcoupler array for guided-wave optical clock distribution," *IEEE Photonics Technology Letters*, vol. 11, no. 11, pp. 1452–1454, Nov 1999.
- [22] R. Chen, L. Wu, L. Lin, C. Choi, Y. Liu, B. Bihari, S. Tang, R. Wickman, B. Picor, and Y. Liu, "Guided-wave Si CMOS process-compatible optical interconnects," in *CAS '99 Proceedings of the International Semiconductor Conference*, vol. 2, Sinaia, Romania, Oct 1999, pp. 467–471.
- [23] M. Melchiorri, N. Daldosso, F. Sbrana, L. Pavesi, G. Pucker, C. Kompocholis, P. Bellutti, and A. Lui, "Propagation losses of silicon nitride waveguides in the near-infrared range," *Applied Physics Letters*, vol. 86, no. 12, p. 121111, 2005. [Online]. Available: <http://link.aip.org/link/?APL/86/121111/1>
- [24] P. Restle and A. Deutsch, "Designing the best clock distribution network," in *IEEE Symposium on VLSI Circuits*, Honolulu, HI, USA, Jun 1998, pp. 2–5.
- [25] S. Tam, S. Rusu, U. Nagarji Desai, R. Kim, J. Zhang, and I. Young, "Clock generation and distribution for the first IA-64 microprocessor," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 11, pp. 1545–1552, Nov 2000.
- [26] B. Ackland, B. Razavi, and L. West, "A comparison of electrical and optical clock networks in nanometer technologies," in *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference*, Santa Clara, CA, USA, September 2005, pp. 779–782.
- [27] S.-M. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits: Analysis and Design*, 3rd ed. New York: McGraw-Hill, 2003.

- [28] S.-C. Wong, T. G.-Y. Lee, and D.-J. Ma, "Modeling of interconnect capacitance, delay, and crosstalk in VLSI," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, pp. 108–111, Feb 2000.
- [29] H. Ymeri, B. Nauwelaers, K. Maex, and D. D. Roest, "A physics-based VLSI interconnect model including substrate and conductor skin effects," *Semiconductor Science and Technology*, vol. 19, no. 3, pp. 516–518, 2004.
- [30] Y. Nagano, Y. Cao, and A. Tsukizoe, "Wire sizing considering skin effect for high frequency circuits," in *Proceedings of 2004 IEEE Asia-Pacific Conference on Advanced System Integrated Circuits*, Fukuoka, Japan, Aug 2004, pp. 282–285.
- [31] P. Zarkesh-Ha, T. Mule, and J. Meindl, "Characterization and modeling of clock skew with process variations," in *Proceedings of the IEEE 1999 Custom Integrated Circuits*, San Diego, CA, USA, May 1999, pp. 441 – 449.
- [32] T. Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSI's," *IEEE Transactions on Electron Devices*, vol. 40, no. 1, pp. 118 – 124, January 1993.
- [33] P. Zarkesh-Ha and J. Meindl, "Asymptotically zero power dissipation gigahertz clock distribution networks," in *IEEE Topical Meeting on Electrical Performance of Electronic Packaging*, San Diego, CA, USA, October 1999, pp. 57–60.
- [34] G. Tosik, F. Gaffiot, Z. Lisik, I. O'Conner, and F. Tissafi-Drissi, "Power dissipation in optical and metallic clock distribution networks in new VLSI technologies," *Electronics Letters*, vol. 40, no. 3, pp. 198 – 200, Feb 2004.
- [35] H. Ito, J. Inoue, S. Gomi, H. Sugita, K. Okada, and K. Masu, "On-chip transmission line for long global interconnects," in *IEDM Technical Digest*, San Francisco, CA, USA, December 2004, pp. 667 – 680.
- [36] A. Allan, D. Edenfeld, W. Joyner, A. Kahng, M. Rodgers, and Y. Zorian, "2001 technology roadmap for semiconductors," *Computer*, vol. 35, no. 1, pp. 42–53, Jan 2002.
- [37] W. Zhao and Y. Cao, "A new generation of predictive technology model for sub-45nm design exploration," in *Proceedings of the 7th International Symposium on Quality Electronic Design*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 585–590.
- [38] S. of Engineering Arizona State University Ira A. Fulton. Predictive technology model. [Online]. Available: <http://www.eas.asu.edu/~ptm/>
- [39] ITRS 2007 edition: Front end processes. [Online]. Available: <http://public.itrs.net/>
- [40] ITRS 2007 edition: Process integration, devices and structures. [Online]. Available: <http://public.itrs.net/>
- [41] Y.-C. King, H. Fujioka, S. Kamohara, K. Chen, and C. Hu, "Dc electrical oxide thickness model for quantization of the inversion layer in mosfets," *Semiconductor Science and Technology*, vol. 13, no. 8, pp. 963–966, 1998. [Online]. Available: <http://stacks.iop.org/0268-1242/13/963>

- 
- [42] B. H. Lee, J. Oh, H. H. Tseng, R. Jammy, and H. Huff, "Gate stack technology for nanoscale devices," *Materials Today*, vol. 9, no. 6, pp. 32 – 40, 2006.
- [43] ITRS 2008 update: Process integration, devices, and structures (PIDS). [Online]. Available: [http://www.itrs.net/Links/2008ITRS/Update/2008Tables\\_FOCUS\\_A.xls](http://www.itrs.net/Links/2008ITRS/Update/2008Tables_FOCUS_A.xls)
- [44] ITRS 2008 update: Interconnects. [Online]. Available: [http://www.itrs.net/Links/2008ITRS/Update/2008Tables\\_FOCUS\\_B.xls](http://www.itrs.net/Links/2008ITRS/Update/2008Tables_FOCUS_B.xls)
- [45] ITRS 2008 update: Overall roadmap technology characteristics (ORTC). [Online]. Available: [http://www.itrs.net/Links/2008ITRS/Update/2008Tables\\_FOCUS\\_A.xls](http://www.itrs.net/Links/2008ITRS/Update/2008Tables_FOCUS_A.xls)
- [46] International Technology Roadmap for Semiconductors. Itrs online collection. [Online]. Available: <http://www.itrs.net/>
- [47] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and design of analog integrated circuits*, 4th ed. John Wiley & Sons Inc., 2001.
- [48] S. M. Park and C. Toumazou, "A packaged low-noise high-speed regulated cascode transimpedance amplifier using a 0.6 $\mu$ m n-well cmos technology," in *Proceedings of the 26th European Solid-State Circuits Conference*, Stockholm, Sweden, Sept. 2000, pp. 431–434.
- [49] T. Nakahara, H. Tsuda, K. Tateno, N. Ishihara, and C. Amano, "High-sensitivity 1-Gb/s CMOS receiver integrated with a III-V photodiode by wafer-bonding," in *Digest of the LEOS Summer Topical Meetings on Electronic-Enhanced Optics, Optical Sensing in Semiconductor Manufacturing, Electro-Optics in Space, Broadband Optical Networks*, Aventura, FL, USA, 2000, pp. I17–I18.

## ANALYTICAL PHOTODIODE PARTIAL DIFFERENTIAL EQUATION

---

### A.1 THE PROBLEM

Refer to figure 2.8.

$$\frac{\partial p}{\partial t} - D_p \left( \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right) + \frac{p}{\tau_p} = f(x, y, z, t) \quad (\text{A.1})$$

With boundary conditions

$$\begin{aligned} p(L_x, y, z, t) &= 0 \\ p_x(0, y, z, t) &= 0 \\ p(x, 0, z, t) &= 0 \\ p(x, L_y, z, t) &= 0 \\ p(x, y, 0, t) &= 0 \\ p(x, y, L_z, t) &= 0 \end{aligned} \quad (\text{A.2})$$

The first step would be to characterise the differential equation in terms of general solutions, by solving the homogenous equation without the carrier density forcing function. This entails solving the equation in A.3

$$p_t - D_p(p_{xx} + p_{yy} + p_{zz}) + \frac{p}{\tau_p} = 0 \quad (\text{A.3})$$

Suppose we can rewrite the function  $p(x, y, z, t) = X(x)Y(y)Z(z)T(t)$  and divide equation A.3 by  $p$  as rewritten, then

$$\begin{aligned} \frac{T'}{T} - D_p \left( \frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} \right) + \frac{1}{\tau_p} &= 0 \\ \frac{T'}{D_p T} + \frac{1}{D_p \tau_p} &= \frac{X'}{X} + \frac{Y'}{Y} + \frac{Z'}{Z} \end{aligned} \quad (\text{A.4})$$

Since  $\frac{1}{D_p\tau_p}$  is a constant, it can only hold if all terms are constant for all values of  $(x, y, z, t)$ . Thus, the following equations will hold valid.

$$\frac{X''}{X} = -\lambda \quad (\text{A.5})$$

$$\frac{Y''}{Y} = -\mu \quad (\text{A.6})$$

$$\frac{Z''}{Z} = -\nu \quad (\text{A.7})$$

$$\frac{T'}{D_p T} + \frac{1}{D_p\tau_p} = -(\lambda + \mu + \nu) \quad (\text{A.8})$$

## A.2 X(X) AND TERMS

Since only  $X(x)$  has a dependency on  $x$ , it follows that the boundary conditions in A.2 imply

$$\begin{aligned} X'' + \lambda X &= 0 \\ X(L_x) &= 0 \\ X'(0) &= 0 \end{aligned} \quad (\text{A.9})$$

Part of the solution(s) to equation A.9, an eigenvalue problem, can easily be determined as

$$\begin{aligned} \lambda_m &= \frac{(2m-1)^2\pi^2}{4L_x^2} \rightarrow \text{eigenvalue} \\ X_m &= \cos\left(\frac{2m-1}{2}\frac{\pi}{L_x}x\right) \rightarrow \text{eigenfunction} \end{aligned} \quad (\text{A.10})$$

for  $m = 1, 2, 3, \dots \infty$

The current boundary conditions applying to  $X(x)$  is not enough to give any information on the coefficients of the solution. This will be determined later by using the forcing function.

## A.3 Y(Y) AND TERMS

Equations A.6 and A.2 leads to

$$\begin{aligned} Y'' + \mu Y &= 0 \\ Y(L_y) &= 0 \\ Y(0) &= 0 \end{aligned} \quad (\text{A.11})$$

Equation A.11, in turn, produces the following values and eigenfunctions to meet the boundary conditions.

$$\begin{aligned} \mu_n &= \frac{n^2\pi^2}{L_y^2} \rightarrow \text{eigenvalue} \\ Y_n &= \sin\left(\frac{n\pi}{L_y}y\right) \rightarrow \text{eigenfunction} \end{aligned} \quad (\text{A.12})$$

for  $n = 1, 2, 3, \dots \infty$

## A.4 Z(Z) AND TERMS

Since A.7 shares the same type of boundary conditions, it can be directly stated that

$$\begin{aligned} \nu_q &= \frac{q^2 \pi^2}{L_z^2} \rightarrow \text{eigenvalue} \\ Z_q &= \sin\left(\frac{q\pi}{L_z} z\right) \rightarrow \text{eigenfunction} \end{aligned} \quad (\text{A.13})$$

for  $q = 1, 2, 3, \dots \infty$

## A.5 TIME AND MORE...

The equation regarding time is expressed as in A.8 and can be rewritten as

$$\begin{aligned} T' + D_p \gamma_{mnp} T &= 0 \\ \text{with} & \\ \gamma_{mnp} &= \left( \frac{(2m-1)^2 \pi^2}{4L_x^2} + \frac{n^2 \pi^2}{L_y^2} + \frac{q^2 \pi^2}{L_z^2} + \frac{1}{D_p \tau_p} \right) \end{aligned} \quad (\text{A.14})$$

The solution to A.14 is

$$T_{mnp} = e^{-D_p \gamma_{mnp} t} \quad (\text{A.15})$$

## A.6 COMPLETE SOLUTION FOR $p(x, y, z, t)$

Since  $p(x, y, z, t)$  was constructed as a product of  $XYZT$ , it follows that the solution would be the triple infinite summation of the product of separated equations as in A.4, across the three indices  $m, n$  and  $q$ .

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \sum_{q=1}^{\infty} C_{mnp}(t) e^{-D_p \gamma_{mnp} t} \sin\left(\frac{q\pi}{L_z} z\right) \sin\left(\frac{n\pi}{L_y} y\right) \cos\left(\frac{2m-1}{2} \frac{\pi}{L_x} x\right)$$

Equation A.6 shows the general solution of the differential equation in A.1. Note that the unknown coefficients, necessary to complete the particular solution, is a function of time. Since there are no more boundary conditions supplying any information, the coefficients will need to be computed through the forcing function  $f(x, y, z, t)$ .

## A.7 THE FUNCTION $f(x, y, z, t)$

If  $f(x, y, z, t)$  can be written in the form of a multi-index summation as in A.6, preferably with the same trigonometric terms, then the solution to  $C_{mnp}$  should be possible through substitutions of A.6 into A.1 and solving for the coefficients. Since A.1 only involves

double differentiation to a specific independent variable, the sine and cosine terms should be common. An elementwise comparison should then be possible.

$$f(x, y, z, t) = f(x, t) = \Phi(t)\alpha e^{-\alpha x} \quad (\text{A.16})$$

Equation A.16 can be rewritten as a infinite sum in the form of

$$\Phi(t)\alpha \sum_{j=1}^{\infty} F_j \cos\left(\frac{2j-1}{2} \frac{\pi}{L_x} x\right)$$

This can be done, since  $\Phi(t)$  is not a function of  $x$ , hence a constant in this context. This allows easy comparison to the series solution of  $p(x, y, z, t)$ . To compute  $F_j$ , it is necessary to determine the correlation between the exponential term and the cosine function such that the coefficients will yield A.7 true to the exponential part of  $f(x, t)$ . This can be done by modifying the Fourier-method as follows

$$\begin{aligned} F_j &= 2/L_x \int_0^{L_x} e^{-\alpha x} \times \cos\left(\frac{2j-1}{2} \frac{\pi}{L_x} x\right) dx \\ F_j &= 4 \left[ \frac{2\alpha L_x e^{\alpha L_x} + \pi(1-2j) \cos(\pi j)}{4\alpha^2 L_x^2 + 4\pi^2 j(j-1) + \pi^2} \right] \times e^{-\alpha L_x} \end{aligned} \quad (\text{A.17})$$

Thus, combined with A.17,

$$f(x, t) \approx \Phi(t)\alpha \sum_{j=1}^{\infty} F_j \cos\left(\frac{2j-1}{2} \frac{\pi}{L_x} x\right) \quad (\text{A.18})$$

For  $f(x, t)$  to take the form of a triple summation, it is still necessary to synthesize some terms for ‘‘compatibility’’. Suppose we can rewrite

$$f(x, t) = f(x, y, z, t) = \Phi(t)\alpha e^{-\alpha x} \times A(y)B(z)$$

where

$$A(y) = B(z) = 1$$

If the above terms can be represented as infinite sums over the indices  $n, q$  respectively, comparison of the term  $f(x, y, z, t)$  to the left hand side of A.1 after substitution with A.6 will be much easier.

There are two ways of doing this,

$$\begin{aligned} 1 &= \sum_{n=1}^{\infty} 0.5^n && \text{for all } y \\ 1 &= \sum_{n=1}^{\infty} \frac{1}{n\pi} (1 - \cos(n\pi)) \sin\left(\frac{n\pi}{L_y}\right) && \text{for } 0 \leq y \leq L_y \end{aligned}$$

Since the second form approximates the constant well enough over the interval of interest, and the form of the expansion is appropriate, this would be a good choice. The same, of course, applies for  $z$ .

Thus we have

$$f(x, y, z, t) \approx \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \sum_{q=1}^{\infty} f_m(t) f_n f_q \cos\left(\frac{2m-1}{2} \frac{\pi}{L_x} x\right) \sin\left(\frac{n\pi}{L_y} y\right) \sin\left(\frac{q\pi}{L_z} z\right)$$

$$f_m(t) = \Phi(t) 4\alpha \left[ \frac{2\alpha L_x e^{\alpha L_x} + \pi(1-2m) \cos(\pi m)}{4\alpha^2 L_x^2 + 4\pi^2 m(m-1) + \pi^2} \right] \times e^{-\alpha L_x}$$

$$f_n = \frac{1}{n\pi} (1 - \cos(n\pi))$$

$$f_q = \frac{1}{q\pi} (1 - \cos(q\pi))$$
(A.19)

## A.8 SUBSTITUTION AND SOLUTION

With reference to equation A.1 and A.6, and if the following is set as

$$V(x, y, z) = \cos\left(\frac{2m-1}{2} \frac{\pi}{L_x} x\right) \sin\left(\frac{n\pi}{L_y} y\right) \sin\left(\frac{q\pi}{L_z} z\right)$$
(A.20)

with the summation of  $1 \rightarrow \infty$  of  $m, n, q$ , we have

$$p_t = \sum \sum \sum (C'_{mnq}(t) e^{-D_p \gamma t} - C_{mnq}(t) D_p \gamma e^{-D_p \gamma t}) V(x, y, z)$$

$$p_{xx} = \sum \sum \sum \left( C_{mnq}(t) e^{-D_p \gamma t} \left( - \left( \frac{(2m-1)\pi}{2L_x} \right)^2 \right) \right) V(x, y, z)$$

$$p_{yy} = \sum \sum \sum \left( C_{mnq}(t) e^{-D_p \gamma t} \left( - \left( \frac{n\pi}{L_y} \right)^2 \right) \right) V(x, y, z)$$

$$p_{zz} = \sum \sum \sum \left( C_{mnq}(t) e^{-D_p \gamma t} \left( - \left( \frac{q\pi}{L_z} \right)^2 \right) \right) V(x, y, z)$$

$$p = \sum \sum \sum (C_{mnq}(t) e^{-D_p \gamma t}) V(x, y, z)$$

$$f = \sum \sum \sum f_m(t) f_n f_q V(x, y, z)$$
(A.21)

Elementwise comparison while substituting A.21 into A.1 yields

$$C'_{mnq}(t) - \rho C_{mnq} = f_m(t) f_n f_q e^{D_p \gamma t}$$
(A.22)

where, strangely enough,

$$\rho = \left( D_p \gamma - D_p (\lambda + \mu + \nu) - \frac{1}{\tau_p} \right) = 0$$
(A.23)

Solving for A.22, it can be shown that

$$C_{mnq}(t) = f_n f_q \left[ \int_0^t f_m(s) e^{D_p \gamma s} ds \right]$$
(A.24)

## A.9 MODELLING $\Phi(t)$ AS A SQUARE WAVE

It remains now to model the time function of photon flux as a square wave, while making sure it is possible to integrate the function over the interval 0 to  $t$ . This is possible by constructing a fourier series expansion.

With  $T$  as the period of the square wave, and  $\Phi_0$  the maximum photon flux amplitude, it can be shown that

$$\frac{\Phi_0}{2} \left[ 1 + \sum_{j=1}^{\infty} \frac{2}{\pi j} (1 - \cos(\pi j)) \sin\left(\frac{2\pi j}{T}t\right) \right] \quad (\text{A.25})$$

Thus, the solution of the integral in A.24 can be reduced to

$$\int_0^t \Phi(s) e^{D_p \gamma s} ds = \frac{\Phi_0}{2} \left[ \int_0^t e^{D_p \gamma s} ds + \sum_{j=1}^{\infty} \frac{2}{\pi j} (1 - \cos(\pi j)) \int_0^t \sin\left(\frac{2\pi j}{T}s\right) e^{D_p \gamma s} ds \right] \quad (\text{A.26})$$

$$\frac{\Phi_0}{2} \left( \frac{1 - \frac{1}{e^{D_p \gamma t}}}{D_p \gamma} + \sum_{j=1}^{\infty} \frac{2T}{\pi j} (1 - \cos(\pi j)) \times \left[ \frac{\frac{2\pi j}{T} e^{\frac{2\pi j}{T} t} - 2\pi j \cos\left(\frac{2\pi j}{T}t\right) + T D_p \gamma \sin\left(\frac{2\pi j}{T}t\right)}{T^2 D_p^2 \gamma^2 + 4\pi^2 j^2} \right] \right) e^{D_p \gamma t} \quad (\text{A.27})$$

Now we have a complete solution to the problem, describing  $p(x, y, z, t)$  at any given point to a given square wave input.

## A.10 SUMMARY

When substituting all the relevant coefficient terms, it is noted that some of the exponential terms cancel, yielding a complete solution for  $p(x, y, z, t)$  as in A.28. When programming this solution into a computer algebra system, two things are immediately noticed:

1. The index calculations regarding  $m$  is most complex and will take most time. Thus, make sure that index  $m$  is in the outer-most loop.
2. Matrix manipulation can only be done with index  $j$ . The rest of the calculation will need to happen element-wise, which are time consuming. Thus, optimisation is necessary when coding the solution.

$$\begin{aligned}
p(x, y, z, t) = & \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \sum_{q=1}^{\infty} \frac{1}{n\pi} (1 - \cos(n\pi)) \frac{1}{q\pi} (1 - \cos(q\pi)) \times \\
& 4\alpha \left[ \frac{2\alpha L_x e^{\alpha L_x} + \pi(1 - 2m) \cos(\pi m)}{4\alpha^2 L_x^2 + 4\pi^2 m(m-1) + \pi^2} \right] \times e^{-\alpha L_x} \times \\
& \frac{\Phi_0}{2} \left( \frac{1 - \frac{1}{e^{D_p \gamma t}}}{D_p \gamma} + \sum_{j=1}^{\infty} \frac{2T}{\pi j} (1 - \cos(\pi j)) \times \right. \\
& \left. \left[ \frac{\frac{2\pi j}{e^{D_p \gamma t}} - 2\pi j \cos\left(\frac{2\pi j}{T} t\right) + T D_p \gamma \sin\left(\frac{2\pi j}{T} t\right)}{T^2 D_p^2 \gamma^2 + 4\pi^2 j^2} \right] \right) \times V(x, y, z)
\end{aligned}$$

where

$$V(x, y, z) = \sin\left(\frac{q\pi}{L_z} z\right) \sin\left(\frac{n\pi}{L_y} y\right) \cos\left(\frac{2m-1}{2} \frac{\pi}{L_x} x\right) \tag{A.28}$$

The unknown  $\gamma$  ( $\rho$  reduced to zero a while ago) is explained in A.14.

## A.11 DETERMINING CURRENT DENSITIES

When computing current densities, the equation A.29 will hold for p-type carriers in the absence of an electric field (a valid assumption within bulk regions).

$$J_p = -\bar{q} D_p \nabla p \tag{A.29}$$

The symbol  $\bar{q}$  in equation A.29 represents the charge of a single carrier, as not to be confused with the index  $q$  which is often used.

Since the analytical solution to  $p$  is known, the only terms which will be subject to differentiation will be that of the term  $V(x, y, z)$  as defined in equation A.20. The same applies for spatial integrals necessary to compute currents. Since current can only leave the structure in the  $+x$  direction, this component will be solved first.

## A.12 $J_{px}$ AND $I_x$ SOLUTION

Differentiating  $p(x, y, z, t)$ , and hence  $V(x, y, z)$  to  $x$  where  $x = L_x$ , it can be shown that

$$V_x(y, z)|_{x=L_x} = \left( -\frac{2m-1}{2} \frac{\pi}{L_x} \right) \sin\left(\frac{2m-1}{2} \pi\right) \times \sin\left(\frac{q\pi}{L_z} z\right) \sin\left(\frac{n\pi}{L_y} y\right) \tag{A.30}$$

With a simple substitution of  $V(x, y, z)$ , the solution for  $p(x, y, z, t)$  can be converted to a current density  $J_{px}$ .

$$J_{px}(y, z, t) = p(x, y, z, t) \Big|_{V(x,y,z)=-\bar{q}D_p V_x(y,z)} \quad (\text{A.31})$$

Then, integrating across the surface of the bottom interface at

$$\begin{aligned} I_x(t) &= \int_0^{L_y} \int_0^{L_z} J_{px}(y, z, t) dz \cdot dy \\ Vix_{mnq} &= \left( -\frac{2m-1}{2} \frac{\pi}{L_x} \right) \sin \left( \frac{2m-1}{2} \pi \right) \frac{L_y L_z}{nq\pi^2} (\cos(q\pi) - 1) (\cos(n\pi) - 1) \\ I_x(t) &= p(x, y, z, t) \Big|_{V(x,y,z)=-\bar{q}D_p Vix_{mnq}} \end{aligned} \quad (\text{A.32})$$

### A.13 $J_{py}$ , $J_{pz}$ , $I_y$ AND $I_z$ SOLUTION

Following the same procedure as in A.30, the same can be done for  $V_y$  and  $V_z$ .

$$J_{py} = -\bar{q}D_p \left( -\frac{\partial p}{\partial y} \Big|_{y=0} + \frac{\partial p}{\partial y} \Big|_{y=L_y} \right) \quad (\text{A.33})$$

$$V_y(x, z) = \frac{n\pi}{L_y} (1 - \cos(n\pi)) \sin \left( \frac{q\pi}{L_z} z \right) \cos \left( \frac{2m-1}{2} \frac{\pi}{L_x} x \right) \quad (\text{A.34})$$

Similarly,

$$V_z(x, y) = \frac{q\pi}{L_z} (1 - \cos(q\pi)) \sin \left( \frac{n\pi}{L_y} y \right) \cos \left( \frac{2m-1}{2} \frac{\pi}{L_x} x \right) \quad (\text{A.35})$$

Integrating over the surfaces of these results, one obtains

$$\begin{aligned} V_{xyz} &= \left( \frac{(2m-1)L_y L_z}{2\pi L_x q n} + \frac{2qL_y L_y}{\pi(2m-1)nL_z} + \frac{2nL_x L_z}{\pi(2m-1)qL_y} \right) \times \\ &\quad \sin \left( \frac{2m-1}{2} \right) \pi (\cos(n\pi) - 1) (\cos(q\pi) - 1) \end{aligned} \quad (\text{A.36})$$

The final result for the conventional current magnitude leaving the n-well boundaries is

$$I_{n\text{-well}}(t) = p(x, y, z, t) \Big|_{V(x,y,z)=\bar{q}D_p V_{xyz}} \quad (\text{A.37})$$