DESIGN AND IMPLEMENTATION OF BROADBAND CIRCUITS & SYSTEMS FOR FIBER OPTIC COMMUNICATION APPLICATIONS

By

Dariusz P. Palubiak

Bachelor of Engineering

McMaster University, June 2005

A THESIS

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER'S OF APPLIED SCIENCES

McMaster University

Hamilton, Ontario, Canada

MASTER OF APPLIED SCIENCE (2008) McMaster University
(Electrical and Computer Engineering) Hamilton, Ontario

TITLE:	Design and Implementation of Broadband Circuits and Systems for Fiber Communication Applications
AUTHOR:	Dariusz P. Palubiak, B.Eng. (McMaster University)
SUPERVISORS:	Dr. M. Jamal Deen and Dr. Shiva Kumar
NUMBER OF PAGES:	xxv, 180

Abstract

The bandwidth requirements for global communication networks are increasing each year due to the rapid growth of multimedia (broadband) applications. To accommodate this growth, an increasing number of wide-area and local-area networks are converting the transmission medium from copper wire to fiber. As such, the deployment of optical fiber communication systems for emerging high-speed networks (10 – 40 gigabits-per-second) mandates integrated and low-cost designs for optical transceivers. CMOS, with its low-cost, system-on-chip, and mixed-signal capabilities, is a suitable alternative to other more expensive technologies, such as silicon germanium or gallium arsenide. On the other hand, limitations of sub-micron CMOS technology, namely the higher parasitic capacitances, higher noise, and low-quality integrated passive elements, make the design of an integrated optical transceiver in CMOS technology a challenging task.

The design, fabrication and characterization of a broadband transimpedance amplifier in an analog front-end of an optical receiver will be presented in this thesis. The chip was designed and fabricated using a TSMC 0.18 μm CMOS technology. It was measured and characterized, achieving a -3dB bandwidth of 1.5 GHz and 37 dBΩ of transimpedance gain, with less than -7 dB input and output reflections to a 50 Ω measurement interface, while consuming 23.8 mW from a 1.8 V supply and using 1 x 0.781 mm^2 of silicon chip area. The strengths and limitations of this design will be discussed and its performance will be compared to published results. Conclusions will be drawn and recommendations will be proposed to demonstrate the utility of this design in future work on fully integrated optical receiver systems.

Next, an optical transmitter prototype circuit, built by using low-cost, off-the-shelf photonic and microelectronic components, was fabricated and tested in order to assess the performance of directly modulated lasers for use in low-cost multi-gigabit-per-second optical transmitter modules. The performance of the transmitter was assessed for a fiber link spanning 10 km of standard single-mode optical fiber. This work illustrates how the non-idealities and imperfections of low-cost components can significantly degrade the overall system performance in high-speed applications; therefore, recommendations will be proposed so as to improve the performance in spite of these limitations.

Acknowledgements

I would like to express my sincerest gratitude and appreciation to Prof. M. J. Deen for giving me the opportunity to be involved in this research project, and for giving me the chance to pursue my academic advancement. I would also like to thank him for his kindness, patience, and also for the supervision, guidance and encouragement he has given me during the course of this work. Being his student was not only an honor, but it was also a great experience which has taught me many valuable lessons that I will not forget.

I also want to thank my co-supervisor, Prof. S. Kumar, for his valuable assistance throughout my research. By offering fruitful discussions and comments, as well as by being available to answer the many technical questions that I had, Prof. Kumar has helped me immensely in guiding me in the right direction.

Also, I would like to thank Prof. N. Nicolici and Prof. M. Bakr for taking the time to review my thesis and for being in my committee.

I am enormously grateful to everyone in Prof. Deen's Microelectronics Research group for providing an excellent and motivational working environment, as well as for offering the support I needed in order to complete my projects. In particular, I would like to especially thank Dr. O. Marinov for taking the time in helping me on numerous occasions in the laboratory, for helping me understand many theoretical and practical concepts, and for always being friendly and accommodating. Also, I would like to thank my colleagues Waleed Shinwari, Munir El-Desouki, Moussa Kfouri, and Hamed Mazhab-Jafari for their tremendous support, as well as Nelson Costa, Kurt Huang, Minhui Yan, Saman Asgaran, Naser Faramarzpour, Farseem Mohammedy, and Mohamed Naser.

I would like to also say a big thank you to the administrative staff at the ECE department, especially Cheryl Gies, Terry Greenlay and Cosmin Coroiu for always being there when I needed their help.

The Canadian Microelectronics Corporation (CMC) is acknowledged for arranging the fabrication of the chips.

Finally, I would like to thank my wonderful parents and friends for their encouragement and support in my academic pursuits.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ACRONYMS

Acronyms

AC	Alternating Current
ADSL	Asynchronous Digital Subscriber Line
AFE	Analog Front-End
AGC	Automatic Gain Control
AM	Amplitude Modulation
ATM	Asynchronous Transfer Mode
APC	Automatic Power Control
APD	Avalanche Photodiode
AWGN	Additive White Gaussian Noise
BER	Bit-Error-Rate
BOSA	Bidirectional Optical Sub-Assembly
BPON	Broadband Passive Optical Network
BW	Bandwidth
CDR	Clock and Data Recovery
CG	Common Gate
CLM	Channel Length Modulation
CML	Current Mode Logic
CMOS	Complementary Metal-Oxide-Semiconductor
CO	Central Office
CP	Charge Pump
CS	Common Source
CW	Continuous Wave
DC	Direct Current
DD	Direct Detection
DDJ	Data-Dependent Jitter
DFB	Distributed Feedback

DFF	D flip-flop
DMT	Discrete Multi Tone
DMUX	De-multiplexer
DS	Downstream
DSL	Digital Subscriber Line
DWDM	Dense Wavelength Division Multiplexing
EMI	Electromagnetic Interference
EPON	Ethernet Passive Optical Network
ER	Extinction Ratio
ESD	Electrostatic Discharge
FD	Frequency Detector
FET	Field-Effect Transistor
FLAG	Fiber Link Around the Globe
FM	Frequency Modulation
FP	Fabry-Perot
FSAN	Full Service Access Network
FSK	Frequency Shift Keying
FTTH	Fiber-to-the-Home
FTTx	Fiber-to-the-Home variants
FWHM	Full-Width Half-Maximum
GaAs	Gallium Arsenide
Gbps	Gigabit per second
GPON	Gigabit Passive Optical Network
GVD	Group Velocity Dispersion
HBT	Heterostructure Bipolar Transistor
HFC	Hybrid Fiber-Coax
HDTV	High Definition Television
HFET	Heterostructure FET
IC	Integrated Circuit

IM	Intensity Modulation
ISI	Inter-symbol Interference
ITU	International Telecommunications Union
kbps	kilobits-per-second
LC	Inductor-Capacitor
LD	Laser Diode
LPF	Low Pass Filter
MA	Main Amplifier
Mbps	Megabits per second
MMF	Multi-mode Fiber
MSR	Mode Suppression Ratio
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
MQW	Multiple Quantum Well
MUX	Multiplexer
NA	Numerical Aperture
NF	Noise Figure
NRZ	Non-Return to Zero
OA	Optical Amplification
OC	Offset Cancellation
ODN	Optical Distribution Network
OLT	Optical Line Terminal
ONU	Optical Network Unit
OOK	On-Off Keying
PCB	Printed Circuit Board
PD	Phase Detector
PFD	Phase/Frequency Detector
PIN	p-type intrinsic n-type
PLL	Phase-Locked Loop
PM	Phase Modulation

PON	Passive Optical Network
PP	Power Penalty
PPG	Pulse Pattern Generator
PRBS	Pseudo-Random Bit Sequence
PSK	Phase Shift Keying
P2MP	Point to Multi-Point
P2P	Peer to Peer
Q	Quality Factor
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
RFIC	Radio Frequency Integrated Circuit
RGC	Regulated Cascode
RJ	Random Jitter
RMS	Root-Mean-Square
RN	Remote Node
RZ	Return to Zero
SLM	Single Longitudinal Mode
SMA	SubMiniature version A
SMF	Single-mode Fiber
SNR	Signal-to-Noise Ratio
Tbps	Terabits per second
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TIA	Trans-Impedance Amplifier
TO	Transistor Outline
UI	Unit Interval
US	Upstream
VCO	Voltage Controlled Oscillator
VGA	Variable Gain Amplifier

VLSI	Very Large Scale Integration
VoIP	Voice over Internet Protocol
WDM	Wavelength Division Multiplexing
xDSL	Digital Subscriber Line variants

Symbols

α_a	Absorption coefficient
α_0	Critical angle
α'	Attenuation coefficient (Nepers/m)
α	Attenuation coefficient (dB/km)
A	Amplifier gain
A_0	Midband amplifier gain
A	Linewidth enhancement factor
a_c	Radius of fiber core
a_n	nth transmitted symbol
β	Propagation constant
β_L	Fraction of spontaneous emission coupled into lasing mode
B	Bit rate
B	Pulse envelope
$\widetilde{B}$	Fourier transform of pulse envelope
BW_n	1st order noise bandwidth
χ_0	Spectral density coefficient of white noise
χ_2	Spectral density coefficient of f^2 noise
c	Speed of light
C_{IN}	Input capacitance

C_{DB}	Drain-bulk junction capacitance
C_{SB}	Source-bulk junction capacitance
C_{GS}	Gate-source overlap capacitance
C_{GD}	Gate-drain overlap capacitance
C_{GC}	Gate-channel capacitance
C_{PD}	Photodiode capacitance
C_{OV}	Drain/source diffusion overlap capacitance
$C_{ox}^{\prime}$	Oxide capacitance per unit area
CP	Charge Pump
Δ	Relative core-cladding index
D	Total dispersion parameter
D_M	Material dispersion parameter
D_W	Waveguide dispersion parameter
ε	Relative permittivity
ε_0	Permittivity of free space
ε_{ox}	Oxide permittivity
$\mathbf{E}$	Electric field vector
$\tilde{\mathbf{E}}$	Fourier transform of electric field vector
E_x	Transverse electrical field x-component
E_y	Transverse electrical field y-component
E_L	Electric field along channel length
E_0	Electric field amplitude
$\hat{\mathbf{e}}$	Electric field polarization unit vector
f	Frequency variable
f_{3dB}	3dB cutoff bandwidth
f_T	Transition frequency

$f_{\max}$	Maximum oscillation frequency
g_m	MOSFET transconductance
G_m	Effective transconductance
g_0	MOSFET output conductance
G	Net rate of stimulated emission
γ	Body-effect coefficient
Γ	Optical confinement factor
Γ	Channel-noise factor
η_L	Laser diode quantum efficiency
η_P	Photodiode quantum efficiency
h	Planck's constant
$\mathbf{H}$	Magnetic field vector
H	Linear channel transfer function
H_0	Midband gain of linear channel
H_x	Transverse magnetic field x-component
H_y	Transverse magnetic field y-component
H_F	Fiber transfer function
$i_{n,P}$	Photocurrent noise current
$i_{n,LC}$	Input-referred linear channel noise current
$i_{n,TIA}$	Input-referred trans-impedance amplifier noise current
$i_{n,LC}^{rms}$	Linear channel input-referred root-mean-square noise current
i_{sens}	Electrical receiver sensitivity
$\overline{i_{n,P}^2}$	Mean-squared photocurrent noise
$\overline{i_{n,P,0}^2}$	ZERO-bit mean-squared photocurrent noise
$\overline{i_{n,P,1}^2}$	ONE-bit mean-squared photocurrent noise

$I_{n,in}^2$	Input-referred noise current power spectral density
$I_{n,LC}^2$	Linear channel input-referred noise current power spectral density
$I_{n,res}^2$	Feedback resistor noise current power spectral density
I_A	Current injected into laser active region
$i_{n,res}$	Thermal noise of feedback resistor
$i_{n,D}$	Channel noise
i_{ovl}	Input overload current
I_P	Photocurrent
I_D	Drain current
I_P	Fourier transform of photocurrent
I_{th}	Laser threshold current
I_{BIAS}	Laser bias current
I_{MOD}	Laser modulation current
J_m	Bessel function
κ	Non-linear optical gain parameter
k	Boltzmann's constant
k	Wave number
K_n	Modified Bessel function
λ	Wavelength
λ_0	Carrier wavelength
λ_{ZD}	Zero-dispersion wavelength
λ_{CLM}	Channel-length modulation coefficient
$\Delta\lambda_S$	Source spectral width
L	Length of fiber span
L	Channel length

L_{drawn}	Drawn channel length
ΔL	Overlap between drain/source and gate
μ_0	Permeability of free space
μ_n	Electron mobility
ν	Optical frequency
n_1	Fiber core refractive index
n_2	Fiber cladding refractive index
$\bar{n}$	Modal index
$\overline{n_g}$	Group index
N	Number of bits sent
N	Carrier density
N_0	Steady state carrier density
N_T	Transparency value
ϕ	Phase
ω	Angular frequency
ω_0	Angular carrier frequency
ω_n	Natural self-resonance frequency
$\Delta\omega$	Pulse envelope frequency spread
Φ	Phase response
Φ_F	Fermi potential
p	Pulse shape waveform
P	Optical power
P_{abs}	Optical power absorbed by photodiode
P_{rec}	Received optical power
P_{tr}	Optical power transmitted through photodiode
P_{on}	Optical power with laser on

P_{off}	Optical power with laser off
P_{mm}	Main mode optical power
P_{sm}	Side mode optical power
$\overline{P}$	Average optical power
$\overline{P}_{sens,PIN}$	PIN receiver optical sensitivity
$\overline{P}_{sens,quant}$	Optical sensitivity quantum limit
PP_C	Laser chirp power penalty
PP_D	Dispersion power penalty
PP_{ER}	Extinction ratio power penalty
Q	Q-parameter
Q_n	Negative channel charge
ρ	Radial coordinate
$\mathbf{r}$	Position vector
R	Responsivity
R_L	Load resistance
R_F	Feedback resistance
R_G	Gate resistance
R_D	Drain resistance
R_S	Source resistance
R_{DB}	Drain-bulk resistance
σ_g	Differential gain
S	Photon density
S_0	Steady state photon density
S_E	Slope efficiency
τ_{tr}	Transit time

τ_S	Photon lifetime
τ_N	Carrier lifetime
T	Absolute temperature
t	Time variable
T_B	Bit period
T_{FWHM}	Full-width half-maximum Gaussian pulse width
τ_g	Group delay
$\Delta\tau_g$	Group delay variation
ΔT_B	Pulse spread
t_c	Half of relaxation oscillation period
t_S	Sampling time
t_{on}	Laser turn-on delay
t_{ox}	Oxide thickness
v_c	Carrier velocity
$v_{I,ovl}$	Maximum input voltage swing
v_g	Group velocity
v_c	Carrier velocity
$v_{n,sat}$	Electron saturation velocity
v_O	Linear channel output voltage
v_n	Linear channel output noise voltage
v_n^{rms}	Linear channel output noise root-mean-square voltage
v_S	Linear channel output signal voltage
v_p	Phase velocity
$V_{n,LC}^2$	Linear channel output noise power density spectrum

V_0	Fourier transform of linear channel output voltage
V_{DD}	Supply voltage
V_{DTH}	Decision threshold voltage
V_{GS}	Gate-source voltage
V_{DS}	Drain-source voltage
V_{SB}	Source-bulk voltage
V_{RB}	Reverse bias voltage
V_{TH}	Threshold voltage
V_{TH0}	Threshold voltage at zero substrate bias
V_E	Vertical eye opening
$V_E^{/}$	Reduced vertical eye opening
W_D	Depletion width
W	Channel width
x	x-direction variable
y	y-direction variable
Y_{IN}	Input admittance
Y_{OUT}	Output admittance
ζ	Damping factor
z	z-direction variable
Z_T	Transimpedance

Chapter 1
INTRODUCTION

The advantages of optical fiber compared to copper as a transmission medium for wire-line communications are numerous. In particular, an optical fiber cable offers higher carrier frequencies, wider bandwidth, lower loss, immunity to electro-magnetic interference (EMI), availability of good components, and lower cost-per-meter than copper cables. These advantages, together with the continual demand for higher-speed and higher-quality services from the business and residential subscribers, suggest that optical fiber will in the near future become the successor to copper as the transmission media used for the 'last-mile' access network; the so-called last-mile because it represents the final portion of the global communications infrastructure connecting the end-user.

The use of optical fiber first began in the late 1970s as a means of linking metropolitan telephone central offices (CO). Since then, fiber optic technology has matured and grown significantly. Coupled with the advances in very-large-scale integration (VLSI) microelectronics, lightwave technology has brought about the advent of the 'information age' by creating a quantum leap in the achievable capacity for the transmission and reception of information. The information age can be categorized as the era of high-capacity/high-speed information transmission, storage and processing, based on the application of photonic and microelectronic technology. Figure 1.1 illustrates the significant technological advances over the last 150 years that have lead up to the information age [1]. Furthermore, the capacity-distance product of lightwave systems has steadily doubled over the last 25 years with each succeeding technology generation. Table 1.1 lists the typical characteristics of several generations of terrestrial and submarine lightwave systems.

Ultra-long-haul submarine fiber optic cable systems have been driving the development of innovative optical fiber technology since their introduction in the late 1980s as an effective means of linking communication networks across different

continents [8], [23]. Current state-of-the-art commercial submarine lightwave systems employ dense wavelength-division multiplexing (DWDM) to increase the aggregate bit-rate, and optical amplification (OA) for increasing the repeater spacing resulting in long haul optical communication systems that span thousands of kilometers and operate in the terabits-per-second (Tbps) range [24], [25]. Figure 1.2 illustrates the architecture of a typical point-to-point DWDM link.

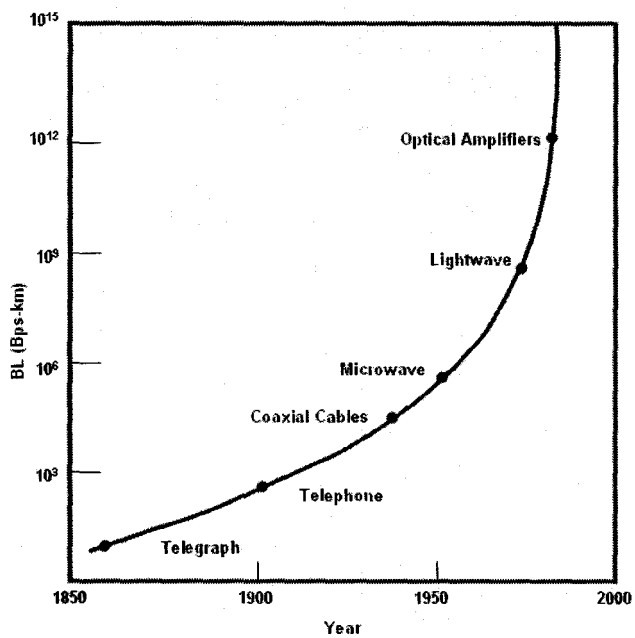


Figure 1.1 – Increase in bit rate-distance product during the last 150 years [1]

Generation	Bit Rate (Gbps)	Operating wavelength (μm)	Type of Fiber	Loss (dB/km)	Repeater Spacing (km)	Source/ Detector	Application
First (1978-1982)	0.045	0.85	MMF	3	15	GaAlAs/Si	CO trunking
Second (1980-1984)	0.09	1.3	MMF	1	30	InGaAsP/Ge	Long-haul
Third (1982-1987)	0.56	1.3	SMF	0.4	>40	InGaAsP/PINFET	WDM-trunking
Fourth (1984-1990)	> 1	1.55	SMF(D-shifted)	<0.3	>100	InGaAsP/InGaAsP	Submarine

Table 1.1 – Characteristics of different generations of lightwave systems [13]

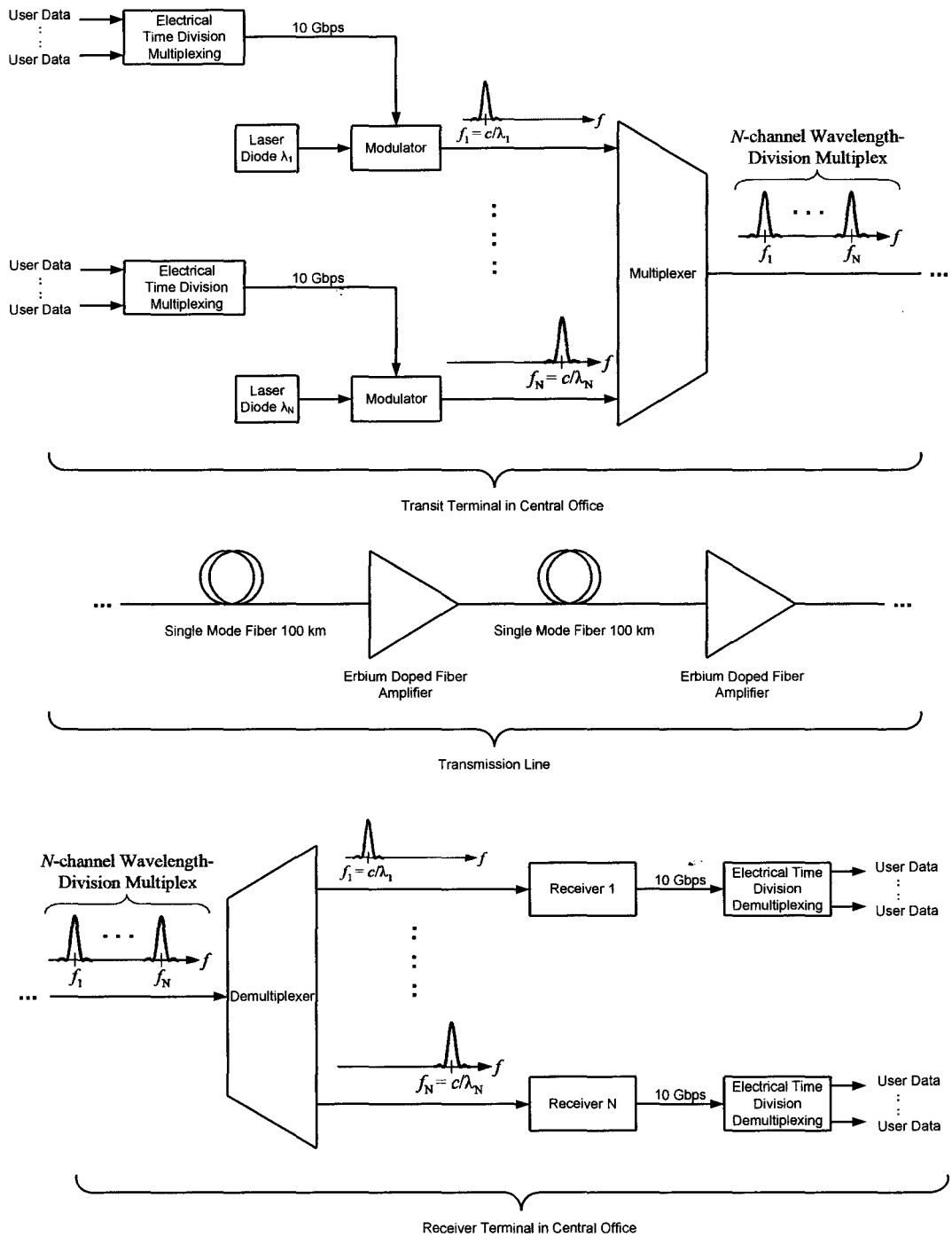


Figure 1.2 – A DWDM point-to-point link [13]

High-speed DWDM systems employ many sophisticated modulation/ demodulation and compensation techniques as well as expensive photonic components in

order to overcome the transmission impairments that arise when sending digital information over very large distances at very high speeds using optical fiber. For example, DWDM systems use specialized single-mode fibers (SMF) that eliminate the absorption peak found near the 1385 nm wavelength (see figure 2.5) in a typical low-cost SMF, thus creating a wider low-attenuation spectral window between 1300 to 1600 nm for the DWDM channels to range. This enlarged optical spectrum accommodates more optical channels with each channel of the DWDM signal extending into the multi-gigabits-per-second (Gbps) range (10 – 40 Gbps). But as the data rates increase into this range, the impairments of the optical channel become more and more severe, limiting the maximum transmission distance. Careful management and compensation of each high-speed optical channel is therefore required [26] – [28].

To illustrate the state-of-the-art in ultra-high-capacity commercial lightwave systems, consider FLAG (Fiber-Optic Link Around the Globe) Telecom Holdings Ltd., a private consortium formed in 1999 to provide intercontinental transmission capacity to worldwide telecommunications operators. In 2002, FLAG laid a transoceanic cable called Flag Pacific-1, connecting the United States with Japan [8]. It consists of eight fiber pairs with a length of 22,000 km, each carrying 64 wavelengths at a capacity of 10 Gbps, resulting in an aggregate capacity of 5.12 Tbps. At a total construction cost of approximately $2.1 billion, it is currently the longest transoceanic DWDM system offering such a high capacity. Table 1.2 summarize some of the other major transoceanic DWDM cable systems currently deployed around the world that form the so-called telecommunications 'backbone' network.

System Name	Year	Capacity (Gbps)	Length (km)	Route
TPC-5	1998	5	22, 500	USA-Japan
Southern Cross	1999	40	29, 000	Australia-USA-Spain
TAT-14	2001	640	15, 500	USA-UK
APCN2	2001	2, 560	19, 000	USA-Japan
FP-1	2002	5, 120	22, 000	USA-Japan

Table 1.2 – Transoceanic Optical Systems [1], [8], [12]

1.1 Broadband Access Networks

The development of lightwave technology is exemplified by the deployments of various high-capacity intercontinental optical submarine cable systems and transcontinental terrestrial long-haul systems, which together form the telecommunications 'backbone' network [12]. These developments, along with the emergence of new applications and services, have resulted in a burgeoning demand for Internet traffic worldwide near the end of the 20th century, as illustrated in figure 1.3. The explosive growth of the Internet and its associated multimedia applications has become one of the most dramatic drivers for fiber-based broadband access. However, there is a significant bottleneck in the network between the end-user and the high-speed backbone, known as the 'last-mile' [9]. This 'last-mile' is the access network, defined as the part of the network that connects the end-user to the information-rich network (Internet). End-users include residential customers as well as businesses with their own internal networks. Applications and services range from voice over IP (VoIP), high-definition TV (HDTV), peer-to-peer (P2P) networks, video conferencing, to internet gaming, etc [9].

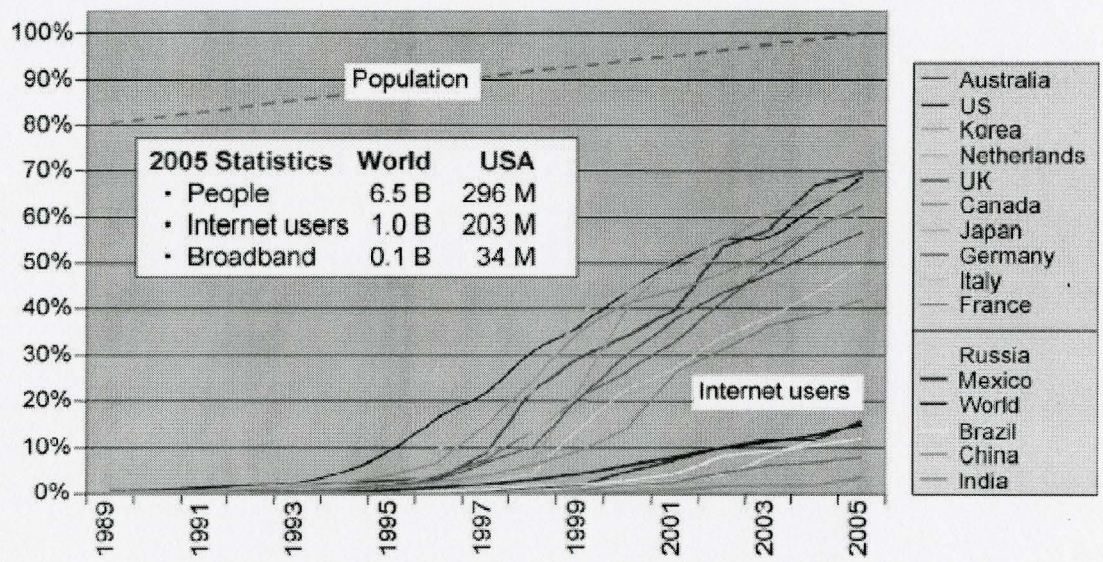


Figure 1.3 – Internet usage in the USA and globally [29]

The access network necessarily includes access points with a great variety of data types and capacity. Figure 1.4 illustrates the different types of access network technologies currently in use in the USA today, and their evolution over time. Initially twisted-pair copper cable was the most widely deployed transmission medium in access networks, having been used for telephony for over 100 years. By the late 1990s, telephone modems with transfer speeds ranging from 4.8 – 56 kbps, were available giving computers easy access to the internet [8]. Theoretical limits for the maximum amount of information that could be sent reliably over the copper channel were beginning to be approached with the development of many sophisticated signal processing and modulation/demodulation techniques. The improved systems became known as the

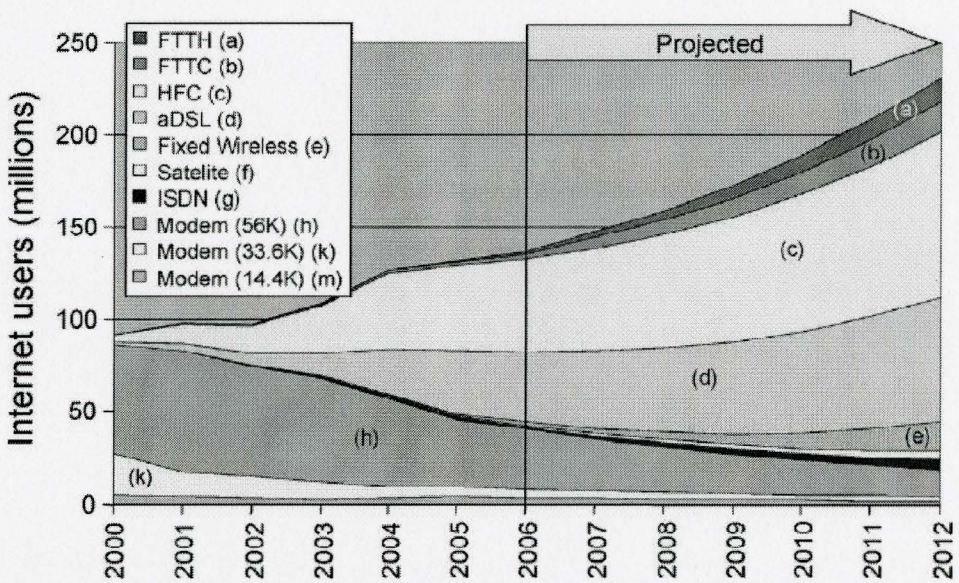


Figure 1.4 – USA residential-access technology adaptation over time [30]

asynchronous digital subscriber line (ADSL) and its variants (xDSL), which operated using quadrature-amplitude modulation (QAM) and discrete multi-tone (DMT) modulation [9], [12]. xDSL was a significant improvement, allowing for transmission speeds on the twisted-pair access line ranging from a few hundred kilobits-per-second (kbps) up to 20 megabits-per-second (Mbps) depending on the variant used and user

distance from the telephone central office; however, the crosstalk and attenuation over twisted-pair copper lines still limited the maximum transmission distance to 100 m for a 100 Mbps signal [12].

The other most popularly deployed access medium has been coaxial copper cable for cable TV broadcast transmission [8]. Coax cable can deliver much larger bandwidths than twisted-pair wire since the electrical signal is guided between an inner conductor and a grounded outer shield conductor; an arrangement providing much better immunity to interference and crosstalk, and a higher signal-to-channel-noise ratio. However, the attenuation places practical limits on the transmission distance of coax cable, resulting in a necessary repeater spacing of approximately 200 m in practical applications [9]. To overcome this limitation, single-mode fibers were used to feed the signal from the central office to a fiber remote node, where the signal would be then converted to the electrical domain for transmission along coaxial cables to arrive at the end user via a cable modem [8], [9], [12]. This combination of fiber and coax cables is called a hybrid fiber-coax (HFC) network. With QAM modulation, a maximum downstream data rate on the order of 2 Gbps is achievable; however, since this bandwidth is shared between 500-1000 subscribers allocated to a cell, the guaranteed bandwidth per subscriber is only 2 Mbps, which is comparable to that of xDSL over twisted-pair copper [12]. Figure 1.5 suggests that fiber-based access networks will allow a much more significant leap in transmission capacity in the last mile compared to that of copper-based systems. Importantly, it is the capacity of the copper-based access network connections that has not advanced commensurately with the 'backbone' core of the network, and with the end user's overall capacity.

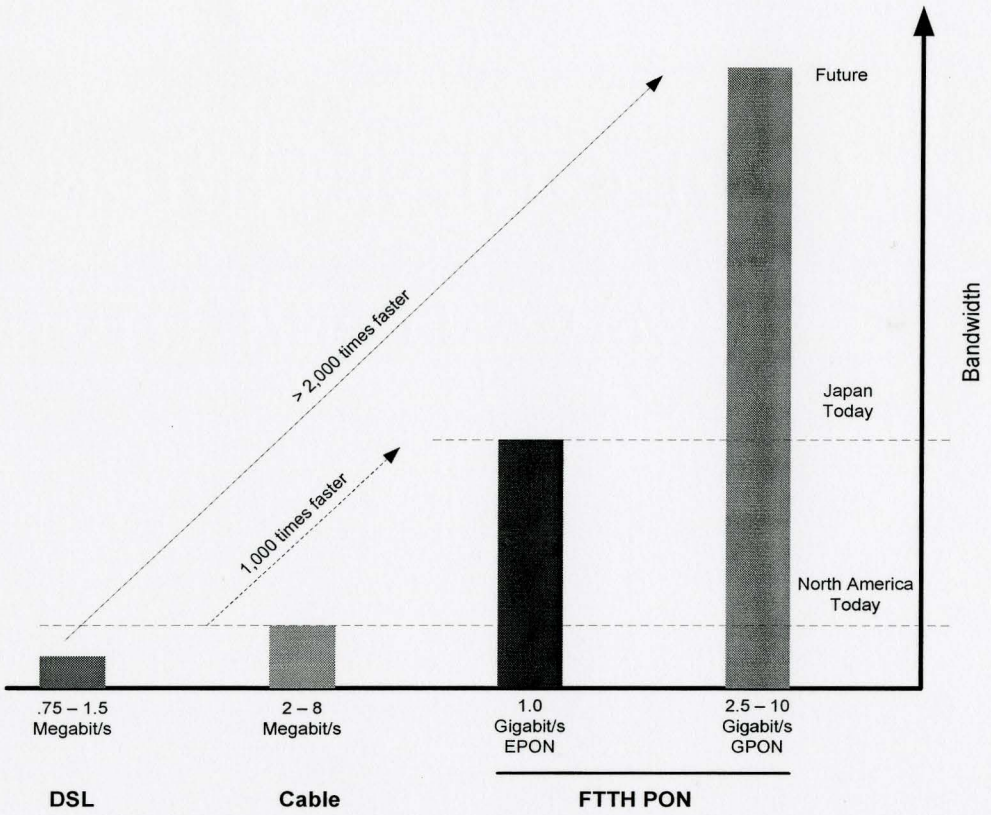


Figure 1.5 – Bandwidth comparison of various access network technologies

1.2 Fiber-to-the-Home

The most widely deployed access networks today in North America are DSL and HFC, with a combined total of approximately 130 million users in Canada and the USA by 2006 [29]. Their main advantage is that they leverage the existing legacy infrastructure to deliver high-speed digital data to the customer at a low-cost. However, they are not true broadband solutions. Their major weakness is that as bandwidth demands increase, they will not be able to provide enough bi-directional bandwidth and quality-of-service for the emerging broadband applications; therefore, they are not future-proof. Another main drawback is that both DSL and HFC require electrical-powered equipment in the

field, leading to high operational costs associated with powering, maintaining, and managing this remotely located active equipment.

Considerable savings can be accrued by employing a point-to-multipoint (P2MP) passive optical network (PON) in the access network [6], [9], [30]. The PON architecture is illustrated schematically in figure 1.6. The need to reduce the amount of dedicated optical fiber without incurring additional costs for remote electronics and powering led to the concept of a PON. A PON minimizes the amount of active electronics between the end user and the CO, consisting simply of an Optical Line Terminal (OLT) at the service provider's CO and a number of Optical Network Units (ONU) near end users. The optical signal is distributed from the CO to the customers over an optical distribution network (ODN). The ODN consists of the optical fiber that runs from the CO to the remote node (RN), the passive optical power splitters/combiners, and the optical fiber that runs from the RN to the ONUs. The passive optical splitter/combiner at the RN is used to divide the optical power to each individual ONU in the downstream direction and to combine the optical signals in the upstream direction for the OLT. The split ratio determines the maximum number of connections per PON infrastructure (usually 32) and is a primary measure of the PON deployment. If the fibers extend all the way to the homes, then the system is known as a fiber-to-the-home (FTTH) home system. Alternatively, if the fibers terminate at the curb, the system is known as a FTTC system. The final distribution from the curb to the homes is accomplished, for example, by twisted-pair copper wires or radio. All systems that bring the fiber relatively close to the subscriber are collectively known as FTTx systems. However, FTTH has become the catch-all descriptor for all fiber to the home, premise, business and 'x' technologies. The goal of FTTH is to provide cost-efficient broadband services for residential and business subscribers by deploying a PON [6].

In a PON, the optical fiber is shared among the subscribers, therefore special network protocols are required for effective use of the shared communications medium. Upstream and downstream transmissions between OLT and ONU are separated by using different wavelengths, thus enabling true broadband bidirectional transmission.

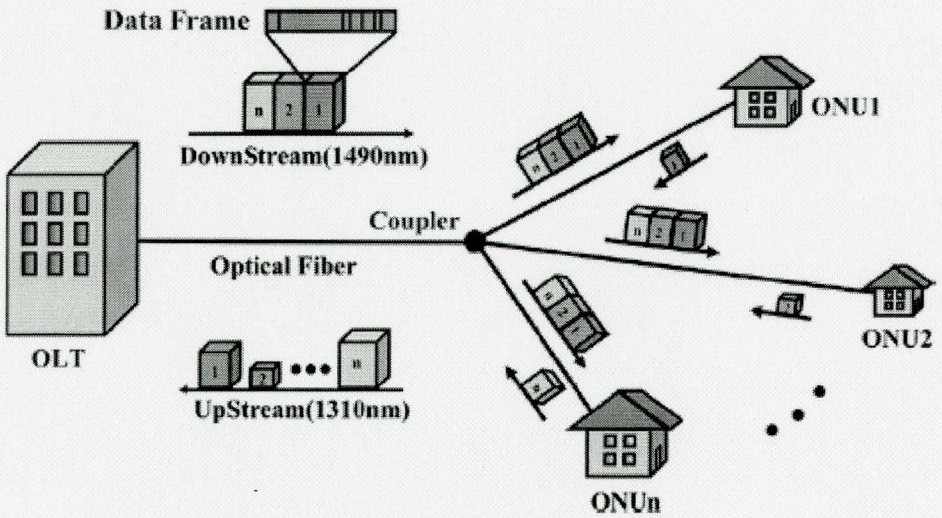


Figure 1.6 – PON Architecture [31]

Information transmitted downstream, from the CO to the ONU, is broadcast to the whole network and the signal is received by all ONUs. Information contained within the header of each data frame identifies the intended subscriber and the ONU simply selects the information with the appropriate address tag. This approach is known as time division multiplexing (TDM) [6], [9], [12]. Frames may be intended for single users, groups of users, or all users, and encryption is used to prevent eavesdropping on downstream traffic. In the upstream direction, special provisions need to be taken to avoid data collisions at the combiner. The CO must coordinate which ONU can send a burst of data at which point in time. Therefore, a time division multiple access (TDMA) strategy is commonly utilized, where the upstream data comprises synchronized time slots with each user transmitting frames only in a time slot designated for that user according to a channel sharing protocol [9], [12]. The upstream data passes through one or more couplers and is delivered to the OLT, but not to other ONUs. Another strategy to avoid data collisions assigns a different upstream wavelength to each ONU so that every user can solely posses the channel all the time. This is known as a WDM-PON [32], [33]. It is suitable for the users who demand network access when required; however from an economical sense, it may be less suitable for wide-scale deployment since the WDM-PON equipped

transceivers require a transmitter with a very precise wavelength. Since the optical components required for such a system are currently very expensive, the most likely strategy for WDM-PON uses wide channel spacing to accommodate the lack of frequency precision and drift associated with low-cost lasers, which is known as coarse WDM.

During the early 1990s, a group of worldwide network operators formed an initiative for a full services access network (FSAN), whose purpose was to drive a set of common technical requirements based on established industry standards and specifications for the emerging FTTH technology [6], [8]. FSAN specifications were developed to promote wide interoperability and mass deployment of triple-play services (voice, video and data) over fiber in particular. Furthermore, numerous field trials of FTTH were conducted during this time to gain practical experience and to identify the obstacles that needed to be overcome to make FTTH a cost-effective access technology. For example, British Telecom deployed a telephony over PON system in the early 1990s [8]. This work was extended to incorporate asynchronous transfer mode (ATM) protocols by the FSAN and the International Telecommunications Union (ITU), eventually becoming known as the BPON standard [15]. These early trials demonstrated that FTTH had high potential and that further improvements and cost reductions would allow an economically viable FTTH solution [9].

Today, there are several 'flavors' of FTTH technology [6], [9]. Currently, BPON, EPON and GPON are in widespread use in FTTH access networks worldwide. BPON being the oldest PON standard, has become the largest FTTH deployment in North America today, with Verizon's FiOS being the first major North American carrier to provide broadband triple-play services (voice, video and data) over fiber. However, most of the new market deployment focus is now on EPON and GPON. One important distinction between the standards is operational speed. BPON is relatively low speed with 155 Mbps upstream/622 Mbps downstream operation, while EPON supports 1.0 Gbps symmetrical operation and GPON 2.5/1.25 Gbps asymmetrical operation. New standards support even higher speeds [9]. Other key distinctions are the operating wavelengths and

protocol support for transport of data packets between access network equipment. Table 1.3 summarizes the salient features of the various PON networks [6], [9].

	BPON	**EPON**	**GPON**
Standard	ITU-T G.983	IEEE 802.3ah	ITU-T G.984
Bandwidth	DS: 622 Mbps US: 155 Mbps	DS: 1 Gbps US: 1 Gbps	DS: 2.5 Gbps US: 1.25 Gbps
Downstream Wavelength	1490 & 1550 nm	1550 nm	1490 & 1550 nm
Upstream Wavelength	1310 nm	1310 nm	1310 nm
Transmission Protocol	ATM	Ethernet	Ethernet, ATM, WDM
Reach	20 km	25 km	20 km
Split ratio	32	32	64

Table 1.3 – Comparison of PON Technologies [6], [9]

1.3 Optical Transceiver Systems

An optical signal can be defined in general by an electric field $\mathbf{E}(\mathbf{r},t)$ at point $\mathbf{r}=(x,y,z)$ and time t, and includes attributes such as the state of polarization (linear, circular or elliptical), the spatial dependence (mode structure and propagation), and the time dependence (including oscillation at carrier frequency and modulation) [1]. By considering a linearly polarized optical signal at one point in space,

$$\mathbf{E}(t)=\hat{\mathbf{e}}\cdot E_0(t)\cdot\cos(\omega_0 t+\phi(t)) \tag{1.1}$$

where $\hat{\mathbf{e}}$ is a polarization vector, $E_0(t)$ is the field amplitude, ω_0 is the carrier frequency, and $\phi(t)$ is the phase, it can be seen that are three possibilities for modulation, and any of them may be analog or digital. The three modulation schemes are [4], [7]:

- Intensity modulation (IM); encodes information in the intensity or instantaneous power of the field (i.e. $P(t)\propto|\mathbf{E}(t)|^2$)

- Phase modulation (PM); encodes information in the instantaneous phase $\phi(t)$ (i.e. phase shift keying (PSK)).
- Frequency modulation (FM); encodes information in the instantaneous frequency $\omega_0 + \frac{d\phi}{dt}$ (i.e. frequency shift keying (FSK)).

The choice of the modulation scheme affects the amount of complexity at the demodulator. For example, the use of FSK and PSK formats generally requires heterodyne or homodyne demodulation techniques [1]. The least expensive and most commonly employed scheme is referred to as 'intensity modulation with direct detection' (IM/DD). Intensity modulation for digital signals uses the on-off keying (OOK) scheme, where the laser is turned on to transmit a ONE bit, and turned off to transmit a ZERO bit. The modulation is linear so that the information signal can be expressed as [3]

$$s(t) = \sum_{n=0}^{\mathrm{N}} a_n w_n(t) \tag{1.2}$$

where $w_n(t) = p(t - nT_B)$, $p(t)$ is the basic pulse waveform, $a_n \in \{0,1\}$ are the transmission symbols (ONE and ZERO bits), $(\mathrm{N}+1)$ is the total number of bits sent, T_B is the bit period and $\mathrm{B} = 1/T_B$ is the bit-rate. Since only positive excursions are contained in the pulse train, the data signal is unipolar, having a non-zero average value. The pulse train is a random binary sequence when the logical ONEs and ZEROs occur with equal probabilities. The shape of $p(t)$ can take a variety of forms; however, it is desirable that $p(t)$ is orthogonal under T_B-shifts if the detector samples the received signal at the instants $t = nT_B$, $n = 0,1,2,...$[4] An orthogonal pulse is uncorrelated with itself shifted by any integer multiple of T_B (i.e. $\langle p(t) \cdot p(t - nT_B) \rangle = 0$ for $n = \pm 1, \pm 2, ...$). Orthogonal pulses are desired for two reasons. Firstly, an orthogonal pulse satisfies the Nyquist

criterion in that it passes through zero at $t = nT_B$, $n = 1,2,..$ but not at $t = 0$, thus eliminating the inter-symbol interference (ISI) at the sampling instants $t = nT_B$ [3], [7]. Formally, ISI is defined the influence of past or future pulses in the pulse train of eq. (1.2) on the value of the pulse at the sampling instant $t = nT_B$. Secondly, a correlation of the whole pulse train with $p(t - nT_B)$ gives the symbol a_n (i.e. $\langle s(t) \cdot p(t - nT_B) \rangle = a_n$), which may be realized by simple linear filtering at the receiver [3], [7]. If the received pulse train is applied to a linear filter with transfer function $P^*(f)$ (i.e. the complex conjugate of the Fourier transform of the basic pulse shape $p(t)$), then for orthogonal pulses and Gaussian noise, no other detector has a lower error probability than the linear filter receiver [3], [7].

The main complication of the matched filter receiver when applied to optical communication systems stems from the fact that the pulses at the input of the receiver (on which matched filtering depends) are no longer orthogonal because of imperfections in the electro-optical components that convert the data signal between electrical and optical domains (in particular, due to the directly modulated laser), and also because of the dispersive optical channel itself. Thus the received pulse shape can vary with the length of the fiber link, the quality of the fiber, the chirp the relaxation oscillations of the laser, and so forth. For these reasons the matched-filter receiver performance is no longer optimal and the shape of the receiver filter transfer function plays a big role in determining the amount of noise and ISI present in the received waveform.

Two types of pulse formats are commonly used to represent binary data in optical communication systems; these are the non-return-to-zero (NRZ) format and the return-to-zero (RZ) format. NRZ uses simple square pulses for $p(t)$, each bit having duration of T_B. The term NRZ refers to the fact that the pulse train does not return to zero during part of the symbol interval. A pulse shape $p(t)$ which does return to zero during part of the symbol interval is known as a return-to-zero (RZ) format. It is trivial to see that NRZ and

RZ pulses are orthogonal. The pulse train spectra of a random binary sequence for NRZ and RZ modulation is shown in figure 1.7 in normalized form.

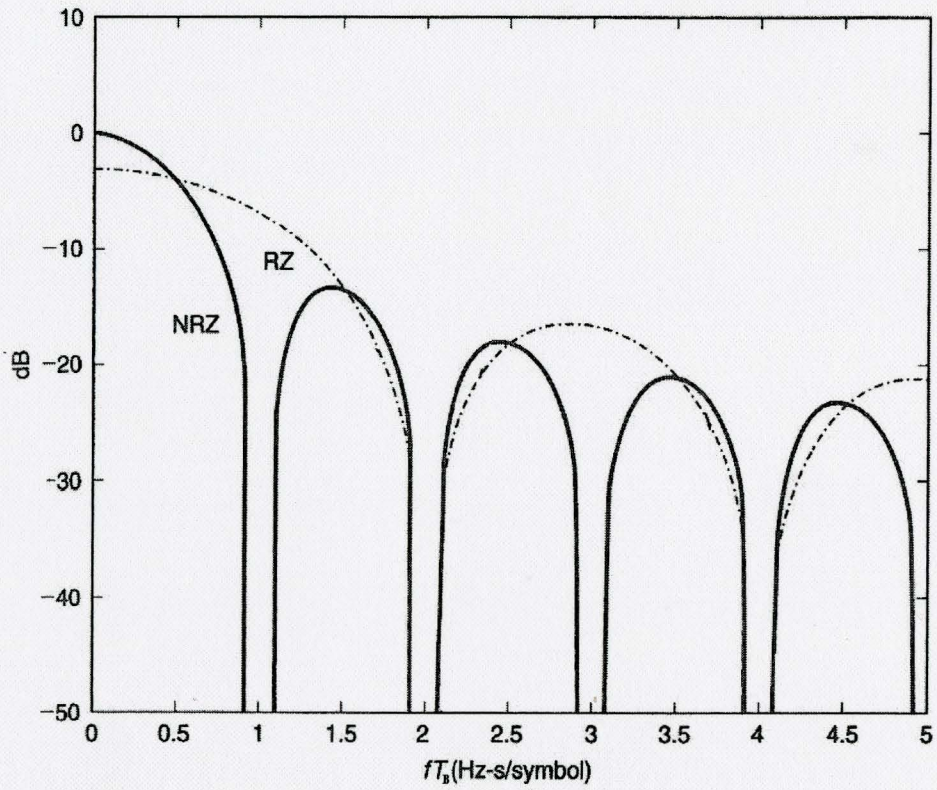


Figure 1.7 – RZ and NRZ pulse train spectra. Pulses have unit energy [3]

The main lobe carries the content of the modulation, and as long as it survives the channel intact, the receiver can detect the transmission without added difficulty. Since for a given bit rate, RZ data contains more transitions than NRZ data, the latter is preferable where circuit bandwidth is costly. However, in contrast to NRZ data, RZ waveforms exhibit a spectral component at a frequency equal to the bit rate, thereby simplifying the task of clock recovery.

A traditional IM/DD optical transceiver is illustrated in figure 1.8.

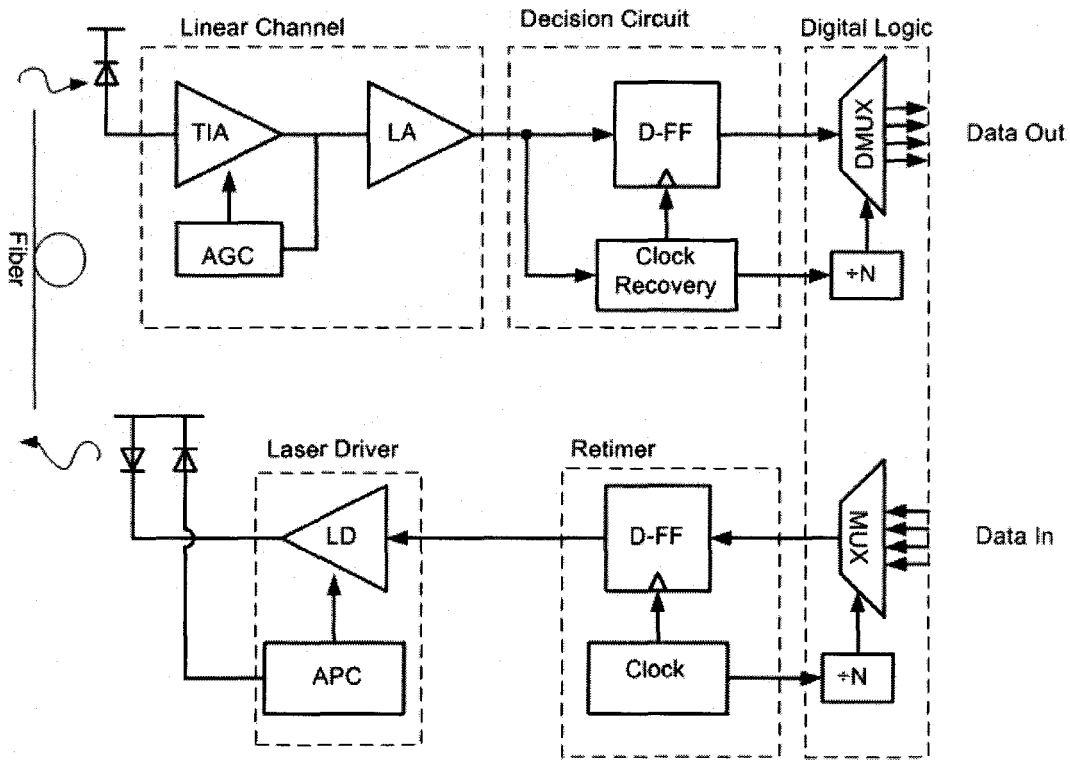


Figure 1.8 – Traditional Optical Channel Transceiver [16], [20]

At the transmitter side, the driver circuit directly modulates the current of the laser diode to generate the intensity-modulated optical signal which is launched into the optical fiber. Additional circuitry can be included to monitor and control the response of the laser diode due to changes of its characteristics from aging and due to temperature effects. For example, a monitor photodiode with good temperature and age stability can be used at the transmitter to generate a current that is proportional to the transmitted optical power which feeds into an automatic power control (APC) circuit that stabilizes the output power of the laser. After passing through a length of fiber, the signal at the receiver is recovered with direct detection, where the photodiode yields a photocurrent directly proportional to the intensity, (i.e. $I_P(t) \propto |\mathbf{E}(t)|^2$). Figure 1.9 depicts the relationship between the electric field $\mathbf{E}(t)$, intensity $P(t) \propto |\mathbf{E}(t)|^2$ and photocurrent $I_P(t) \propto P(t)$.

The photocurrent is fed into the analog front-end (AFE) portion of the receiver, which is modeled by a linear channel comprising a transimpedance amplifier (TIA) and a main amplifier (MA). The AFE is often the critical portion in determining the overall system performance, such as the maximum length of the fiber link.

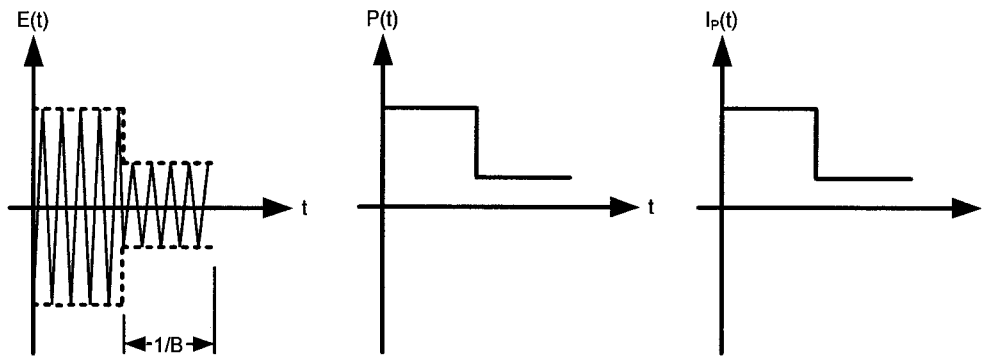


Figure 1.9 – Relationship between field, intensity and photocurrent

The function of the AFE is to convert the photocurrent into a voltage level sufficient enough to drive the following stage, with as little noise and ISI added to the signal as possible. Since the data stream obtained from the AFE is both asynchronous and noisy, for subsequent processing, a clock signal must be extracted from the data so as to allow synchronous operation, such as de-multiplexing (DMUX). The task of clock extraction and data recovery is performed by the clock and data recovery (CDR) block. The CDR block identifies bits as a ONE or ZERO according to the amplitude of the voltage signal from the AFE with a comparator that has a fixed or variable threshold. A D flip-flop (DFF) driven by the extracted clock retimes the data. As such, the comparator and DFF are sometimes called the 'decision circuit'. Since the signal at the input of the decision circuit is a superposition of the received data signal voltage and the undesired noise voltage due at the output of the AFE, occasionally, the instantaneous noise voltage may become so large that may cause the decision circuit to wrongly misinterpret a ZERO for a ONE, or vice versa. The bit-error-rate (BER) is thus defined as the probability that a ZERO is misinterpreted as a ONE by the decision circuit, or vice versa, and is a critical parameter defining the system performance. Therefore, AFE and CDR circuits must

satisfy stringent specifications, thus presenting a difficult challenge for an integrated receiver design. This type of receiver, where the AFE is integrated with the CDR, is known as a 3R receiver because it performs signal re-amplification, signal re-shaping and signal re-timing [2].

A common tool useful for visualizing random digital data and for assessing the performance digital communication systems is the 'eye diagram'. A random digital waveform can be turned into an eye diagram by folding the time axis modulo a whole number of bit intervals and superimposing all of the generated segments. Therefore, all the possible transitions between bits and are shown on a convenient scale. For example, the construction of an eye diagram is shown in figure 1.10 for the case an NRZ signal with mild ISI (the center bit is only affected by a single preceding and following bit) [19]. The impact of the ISI can be judged from the closure of the eye. Eye closure affects the BER performance at the decision circuit. This can be seen with the aid of figure 1.11 [19]. The decision process at the decision circuit is controlled by a decision threshold voltage V_{DTH} as well as the sampling instant t_S. The two slicing lines intersect at the decision point. In the case that the noise distributions are equal for both ZEROs and ONEs, the optimum decision threshold is centered halfway between the ZERO and ONE levels. The figure suggests that ISI and noise not only occur in the signal voltage domain, but also in the time domain. ISI in the time domain is known as data-dependent jitter (DDJ) and noise in the time domain is known as random jitter (RJ). ISI and noise can be characterized from the eye diagram with a histogram of the voltage values at the sampling instant, and the jitter with a histogram of the zero-crossings relative to the decision threshold voltage.

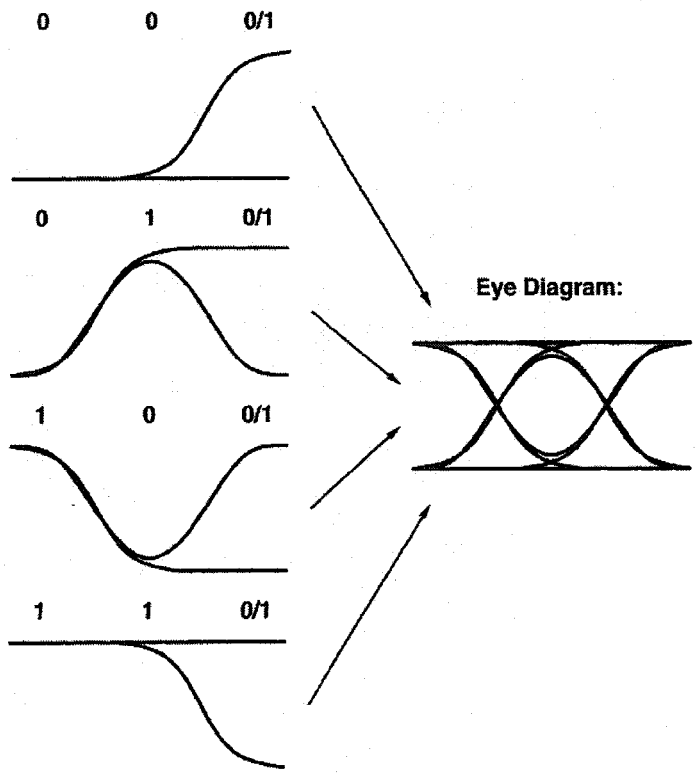


Figure 1.10 – Construction of an eye diagram by superimposing the waveforms corresponding to all possible three-bit sequences [19]

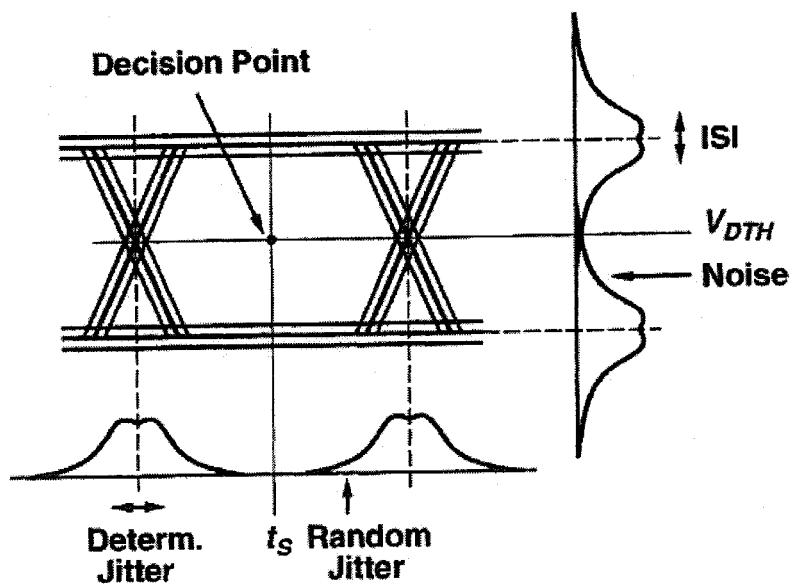


Figure 1.11 – Eye diagram at the input of the decision circuit with ISI, noise, deterministic and random jitter [19]

There are many types of impairments found in an IM/DD optical communication system that will cause ISI, noise and jitter, which all contribute to degrading the eye diagram and resulting in an increased BER at the decision circuit. The effects of the various transmission impairments on the BER can be quantified by the concept of a power penalty. Precisely, the power penalty (PP) for a particular impairment is defined as the increase in average transmitted power necessary to achieve the same BER as in the absence of the impairment. Table 1.4 lists other impairments found in various parts of an optical communication system that lead to power penalties [1]. These impairments and their impact on the system design will discussed in chapters 2 & 3.

ONU and OLT optical transceivers require different circuit design strategies, due to the different modes of transmission between the upstream and downstream directions in a PON. The upstream (US) direction (from ONT to OLT) operates using 'burst-mode' transmission and the downstream (DS) direction (from OLT to ONT) operates using 'continuous-mode' transmission. In the downstream direction, cell collisions are not of concern because the OLT is the only device transmitting. Therefore the OLT transmits a continuous, uninterrupted stream of bits. If the random binary signal contains the same amount of ZEROs and ONEs on average (i.e. a constant DC level), then design of the AFE and CDR blocks can be considerably simplified [20].

System Component	Impairment
Transmitter	Extinction Ratio (ER)
	Frequency Chirp
Fiber	Group velocity dispersion (GVD)
	Attenuation
TIA	Offset
	Frequency Response
CDR	Decision-threshold offset
	Sampling-time jitter

Table 1.4– Examples of impairments leading to power penalties [1]

In the burst-mode, the transmitter sends short bursts of data with long pauses in-between bursts in order to avoid data collisions between ONTs. Therefore, the laser bias and modulation currents must be enabled and disabled quickly in order to save power [34]. All bursts start out with a preamble followed by the payload. Since the bursts arrive with strongly varying power levels and with asynchronous clocks, the burst-mode receiver must use the preamble to automatically adjust the decision threshold level and to quickly synchronize the clock at the receiver, so that each burst obtains the same BER performance [35]. Preamplifier and driving circuits with specialized dynamic properties are therefore required to match the changes of the signal amplitude and clock phase, resulting in a more complicated circuit design.

The optical transceiver is the key functional element of a fiber optic communication system, since it bridges the functional gaps between the microelectronic and photonic technologies for signal processing and transmission, respectively. The photonic components (semiconductor laser diodes and photodiodes) are built exclusively with III-V semiconductor compounds [1], [10], [21], [22]. They incorporate complex layer structures, geometries, and are grown using expensive and sophisticated epitaxial equipment. The substrate materials for these structures are either gallium-arsenide (GaAs), indium-phosphide (InP), or indium-gallium-arsenide (InGaAs), due to their inherently superior optical properties in the 1.3 – 1.6 μm wavelength range, where both dispersion and loss of optical fibers are minimized, and also for their superior electrical and physical properties. The microelectronic circuits on the other hand are realized in a wide variety of technologies [2], [19]. For medium- and low-speed applications, standard silicon technologies, which offer metal-oxide-semiconductor field effect transistors (MOSFETs), bipolar junction transistors (BJTs), or both, are preferred because of their cost advantage. For high-speed applications, silicon-germanium (SiGe), GaAs, or InP technologies, which offer fast hetero-structure transistors in the form of hetero-structure field effect transistors (HFETs) and hetero-structure bipolar transistors (HBTs), have been the standard for high-speed applications. However, aggressive scaling has improved the intrinsic speed of silicon MOSFETs by more than three orders of magnitude in the past 30

years [36]. Figure 1.12 illustrates the scaling trend and its effect on the unity-gain (transition) frequency f_T.

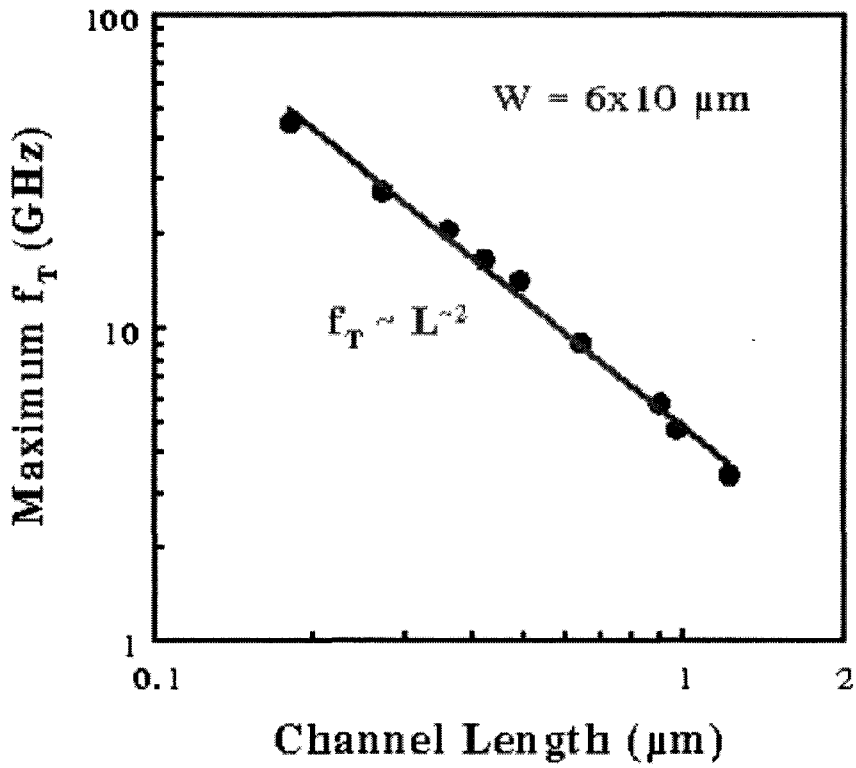


Figure 1.12 – CMOS scaling trend [36]

Mixed-signal complementary metal-oxide-semiconductor (CMOS) technology is increasingly being used in applications that were once the sole domain of more specialized and expensive technologies [37], [38]. CMOS is the preferred technology for its unique advantages. First, it is offered by many foundries and the chips can be fabricated cost effectively on very large wafers. Secondly, the use of a standard CMOS process allows substantially lower power dissipation due to reduced supply voltage (1.8 V for 0.18 µm technology) and negligible static power-dissipation in circuits implemented by digital logic. Additional benefits of CMOS are the higher integration density and the possibility to have large signal-processing blocks on the same chip. These capabilities result in significant space, power and cost savings for the system.

Commercial FTTH optical transceivers are required to be low-cost, highly integrated, and reliable. Today, optical transceiver modules are built with discrete optoelectronic chips such as Fabry-Perot (FP) or Distributed Feedback (DFB) laser diodes (LDs) and PIN or avalanche photodiodes (APD) packaged respectively in the transmit and receive transistor outline (TO) cans, which are subsequently assembled into bidirectional optical subassemblies (BOSA) [39]. If the BOSA module supports two or three wavelengths, then it is known as a diplexer or triplexer module, respectively. The BOSA is then mounted onto printed circuit boards (PCB) together with the microelectronic chips to form the optical transceiver module. Figure 1.13 illustrates the assembly process. From the circuit design point of view, the cost for the optical transceiver could be lowered by using low-cost technologies, such as an advanced mixed-signal CMOS process, and by integrating as many transceiver blocks on one chip as possible, thereby reducing the number of module components.

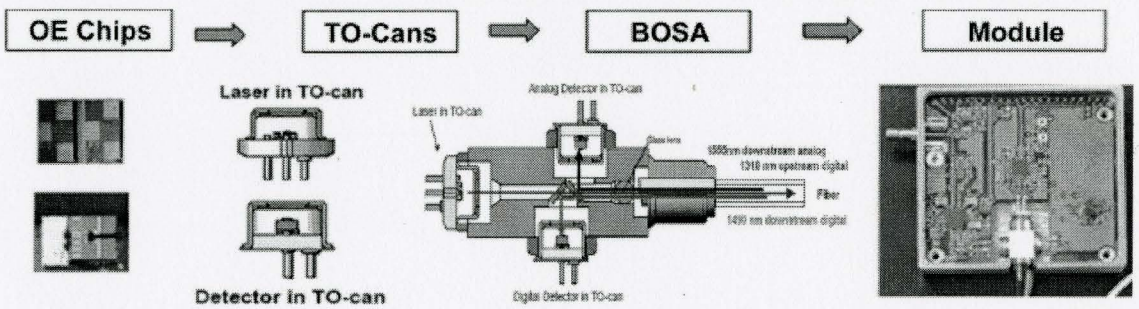


Figure 1.13 – Building blocks and manufacturing steps for a typical triplexer optical transceiver [39]

Looking back at the evolution of lightwave systems, one main objective of system development has become more and more important: *to minimize cost per Gbps per kilometer* [8]. In the case of FTTH, there will be a considerable demand for high-performance and low-cost OLT and ONT hardware all around the world. The ONT in particular will require a low unit cost to enable high volume deployment as the cost of the ONT represents the bulk of the cost of typical PON system. These transceivers will be required to have reduced module size, simplified packaging/assembly, high reliability,

reduced power consumption and standards compatibility if they are deployed on a massive scale. These circuits can be realized using today's state-of-the-art mixed-signal CMOS technology.

1.4 Thesis Outline

Chapter 2 begins with a discussion on optical communication systems by examining the optical channel, optical modulator and optical demodulator in detail. The effects of various transmission impairments on the performance of the overall system will be emphasized, and suggestions will be presented on how to mitigate such impairments.

In Chapter 3, the fundamentals of CMOS circuit design are reviewed, and the design of CMOS TIA circuits is examined, where the emphasis is placed on topologies which can achieve high-performance in spite of numerous design tradeoffs and limitations. A comprehensive review of high-speed CMOS TIA and CDR circuits which have been reported in the literature is also presented.

Chapter 4 describes the design, implementation and measured results of a CMOS regulated cascade (RGC) TIA front-end chip that achieves performance comparable to the TIA circuits that have been reported in the literature.

Chapter 5 describes the implementation and measured results of a low-cost optical transmitter prototype board, and it's performance was assessed for a 10 km fiber span using standard single mode fiber

Chapter 6 summarizes the main results and conclusions of the thesis, and some ideas for future work are presented.

Chapter 2
OPTICAL COMMUNICATION SYSTEMS

2.1 Optical Fiber Channel

The development of low-loss glass optical fibers demonstrates how design improvements in certain propagation characteristics has motivated and dictated technology developments of associated opto-electronic components such as sources and detectors, therefore leading to the emergence of newer generations of optical communication systems [8]. When cladded glass fibers were first demonstrated in 1966 as a suitable optical transmission medium, the losses exhibited by even the lowest loss glass materials available at the time were on the order of 1,000 dB/km. In order to compete economically with the copper transmission media, considerable research was aimed at reducing the losses of optical fiber down to a level less than 20 dB/km at the operating laser wavelength. This was a target improvement representing 98 orders of magnitude. Eventually, Corning Glass Works succeeded in breaking this barrier target by producing a fiber made of ultrapure silica glass exhibiting a loss of 17 dB/km at a wavelength of 0.6328 μm [13]. Further progress has resulted in telecommunication grade fibers that attain losses as low as the theoretical limit of 0.154 dB/km (at $\lambda \sim 1.55$ μm) [1]. This availability of low-loss fibers, as well as the improvement in the reliability of semiconductor lasers that operate around the low-loss wavelengths in terms of their mean life under continuous operation at room temperature led to a revolution in the field of lightwave technology and began the era of fiber-optic communications.

2.1.1 Light Propagation in Optical Fibers

The fundamental principle behind the propagation of light through optical fiber is based on the idea of total internal reflection of light rays that pass through the material [1], [21]. Figure 2.1 illustrates this property with the geometry of an optical fiber. Optical fiber is a cylindrically symmetric structure consisting of a 'core' silica (SiO_2), typically doped with germanium oxide (GeO), which is surrounded by a 'cladding' glass whose refractive index is slightly lower than that of the core [21].

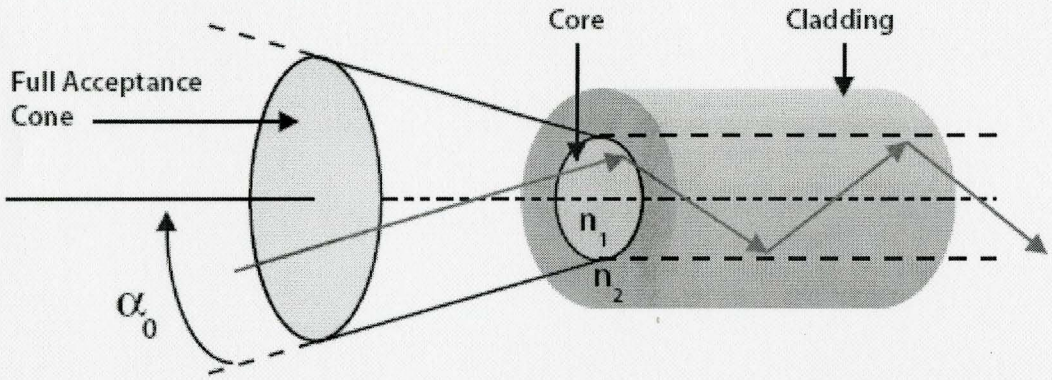


Figure 2.1 – Geometric optics picture of light propagation through optical fiber [40]

The refractive index profile of the fiber can take a variety of forms but the simplest is the step-index profile, where there is an abrupt index change at the core cladding interface, represented algebraically as

$$n(\rho)=\begin{cases} n_1, & \rho < a_c \\ n_2, & \rho \geq a_c \end{cases} \tag{2.1}$$

where n_1, n_2 are the refractive indices of the core and cladding material, respectively, a_c is the radius of the fiber core, and ρ is the radial distance from the center of the fiber core. If a light ray injected through the air-core interface of the fiber is incident at the

core-cladding interface at an angle greater than the critical angle for that interface then it will suffer total internal reflection there. With the critical angle defined as

$$\alpha_0 = \sin^{-1}(n_2 / n_1) \tag{2.2}$$

one can obtain the maximum angle that the incident ray should make with the fiber axis to remain confined inside the core by applying Snell's Law. With the critical angle defined, the resulting expression for the light-gathering capacity of an optical fiber, known as the numerical aperture (NA), can be found,

$$NA = n_0 \sin\alpha_0 = \left(n_1^2 - n_x^2\right)^{1/2} \approx n_1 (2\Delta)^{1/2} \tag{2.3}$$

where $\Delta = (n_1 - n_2)/n_1$ is the relative fractional change at the core-cladding interface, and n_0 is the refractive index of free space. The approximation $n_1 \approx n_2$ is made since Δ is typically never more than 1-2% for technological and systems consideration reasons [13]. In particular, if the time difference between two rays taking the shortest and longest path through a fiber of length L can be expressed as

$$\Delta T_B = \frac{\mathrm{L}}{c}\frac{n_1^2}{n^2}\Delta \tag{2.4}$$

where c is the speed of light. ΔT_B is a first-order measure of the broadening that an impulse launched at the fiber input experiences after traveling through a fiber of length L. It is clear intuitively that ΔT_B should be less than the allocated bit slot according to the condition

$$\mathrm{B}\cdot\Delta T_B < 1 \tag{2.5}$$

From this condition one can obtain an order-or-magnitude upper bound for the capacity-distance product of a multimode step-index fiber,

$$\mathrm{BL} < \frac{n_2}{n_1^2}\frac{c}{\Delta} \tag{2.6}$$

As the bit rate B is increased to the multi-gigabyte per second range ΔT_B needs to decrease at the same rate. The only way to eliminate the multi-path (modal) dispersion in high-speed light-wave systems is by operating with an optical fiber that supports only a single propagating mode (i.e. SMF).

The description of the propagation of light through an optical fiber according to the total internal refection principle of light rays is applicable insofar as the core diameter a_c is large compared to the wavelength λ of the propagating light wave and the relative core-cladding index Δ is not too small. Both of these assumptions are violated by optical fibers designed for optical communication applications [21]. A composite parameter called the normalized frequency which links the core diameter a_c, the wavelength λ, and the relative core-cladding index difference Δ, can be defined as

$$V = (2\pi/\lambda)a_c n_1\sqrt{2\Delta}\,. \tag{2.7}$$

A fiber with a large V supports a large number of multi-path modes, as $V^2/2$ approximately, and the number of modes decreases rapidly as V is reduced [1]. For $V \leq 10$ the geometrical optics model cannot explain propagation effects in fibers and an electromagnetic analysis based on wave optics is required. The electromagnetic analysis involves the solution of Maxwell's equations for a monochromatic wave with free-space wave number $k = \omega/c$, subject to the proper boundary conditions [1], [13]. One forms the vector wave equation satisfied by the electric and magnetic field vectors of the light wave [13]

$$\nabla^2\mathbf{E}(\mathbf{r},t)+\nabla\left[(\nabla\varepsilon/\varepsilon)\mathbf{E}(\mathbf{r},t)\right]=\mu_0\varepsilon\frac{\partial^2\mathbf{E}(\mathbf{r},t)}{\partial t^2}$$
$$\nabla^2\mathbf{H}(\mathbf{r},t)+(\nabla\varepsilon/\varepsilon)\times(\nabla\times\mathbf{H}(\mathbf{r},t))=\mu_0\varepsilon\frac{\partial^2\mathbf{H}(\mathbf{r},t)}{\partial t^2} \quad (2.8)$$

where $\varepsilon=\varepsilon_0 n^2$ is the dielectric permittivity of the fiber since low losses are assumed, and μ_0 is the free-space magnetic permittivity which is the same as that of the fiber. The electric (or magnetic) field can be expressed in terms of its Fourier transform,

$$\mathbf{E}(\mathbf{r},t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\mathbf{E}}(\mathbf{r},t)\exp(-j\omega t)d\omega$$
$$\tilde{\mathbf{E}}(\mathbf{r},\omega)=\int_{-\infty}^{\infty}\mathbf{E}(\mathbf{r},t)\exp(j\omega t)dt \quad (2.9)$$

which describes the full spatial (and time or frequency) dependence of the optical signal. For a step-index fiber having small difference between the refractive indices of the core and cladding, the $\nabla\varepsilon$ term of eq. (2.8) is small and can be ignored. Therefore the Cartesian component of the electric field will satisfy the scalar wave equation [13]

$$\nabla^2 E(\mathbf{r},t)-\frac{1}{c^2}\frac{\partial^2 E(\mathbf{r},t)}{\partial t^2}=0 \quad (2.10)$$

with $E(\mathbf{r},t)$ representing the Cartesian components of the electric field in the core and cladding regions. The general solution to the scalar wave equation in cylindrical coordinates is given by [1]

$$E(\mathbf{r},t)=F(\rho,\varphi)B(z,t)=F(\rho,\varphi)\exp\left(j(\omega_0 t-\beta z)\right) \quad (2.11)$$

where $\mathbf{r} = (\rho, \varphi, z)$ is the direction vector in cylindrical coordinates, β is the propagation constant, with the direction of propagation along z. A small Δ results in the boundary conditions that $E(\mathbf{r})$ and $dE/d\mathbf{r}$ are continuous across the core-cladding interface at $\rho = a_c$ [13]. The boundary conditions lead to two types of solutions to the scalar wave equation – one whose field decreases exponentially with ρ for $\rho > a_c$ and is oscillatory inside the core $(\rho < a_c)$, while the second one leads to oscillatory solutions at all values of ρ. The first type of solution yields discrete values of β and are known as guided modes of the fiber. The latter solution represents what are regarded as radiation modes characterized by a continuum of β's. Formally, a mode is a solution of the wave equation subject to appropriate boundary conditions and is uniquely characterized by the propagation constant of the mode $\beta(\omega)$ [1]. The field distribution $E(\mathbf{r},t)$ of a guided mode does not change as the mode propagates along the z direction, except for an overall multiplicative factor of the form $\exp(j\beta(\omega)z)$. As expected, signal transmission in fiber-optic communication systems takes place through the guided modes only. Guided modes can be classified according to their mode index,

$$\bar{n} = \frac{\beta}{k} = c\frac{\beta}{\omega} \tag{2.12}$$

whose value lies in the range $n_1 > \bar{n} > n_2$. Radiation modes correspond to values of $\bar{n} < n_2$. Each mode propagates with a definite group velocity

$$v_g = (d\beta/d\omega)^{-1} \tag{2.13}$$

Solving the wave equation for a given set of fiber parameters $\{k, a_c, n_1, n_2\}$ allows one to obtain the propagation constant β corresponding to each guided mode [1], [13].

The solution for the transverse electric component E_x is written in terms of the well known Bessel functions in cylindrical coordinates as [1]

$$E_x(\rho,\varphi)=\begin{cases}\left[A\cdot J_m(u\rho/a)/J_m(u)\right]\exp(jm\varphi); & \rho\le a\\ \left[B\cdot K_m(w\rho/a)/K_m(w)\right]\exp(jm\varphi); & \rho>a\end{cases} \tag{2.14}$$

where A, B are constants obtained by applying the boundary conditions, $u=a\sqrt{(n_1^2k_0^2-\beta^2)}$, $w=\sqrt{(\beta^2-n_2^2k_0^2)}$, $u^2+w^2=V^2$, m is an integer, J_m is the Bessel function, K_m is the modified Bessel function [13]. E_x has been chosen as the dominant transverse component of the electric field because fibers with $\Delta<<1$ have a transverse field component $E(\mathbf{r},t)$ lying almost completely along x (or y). These are what are known as 'weakly guiding fibers', because the only non-zero field components for the modal solution in this case are E_x, E_z, H_z and H_y, of which the longitudinal components E_z and H_z can be shown to be much smaller than the transverse components E_x and H_y as long as Δ is small [13]. The same fiber supports another mode linearly polarized along the y axis with the same propagation constant β. Each mode is therefore two-fold degenerate. The linearly polarized modes are designated as LP_{lm}; the l representing the number of azimuthal antinodes over a semicircle while m represents the number of radial antinodes in a mode's field pattern [21]. Figure 2.2 illustrates the various mode intensity patterns. There is a single fundamental mode designated LP_{01} which propagates by itself only if the fiber is designed so that its V-value lies in the range $0 < V < 2.4048$. The field distribution for the x component of the fundamental mode is given by [1]

$$E_x(\rho,z)=E_0\begin{cases}\left[J_0(u\rho)/J_0(\rho a)\right]\exp(j\beta z); & \rho\le a\\ \left[K_0(u\rho)/K_0(\rho a)\right]\exp(j\beta z); & \rho>a\end{cases} \tag{2.15}$$

In practice the field is approximated by a Gaussian distribution of the form [1]

$$E_x = E_0 \cdot \exp\left(-\rho^2 / \sigma^2\right) \exp\left(j\beta z\right) \tag{2.16}$$

where σ is the field radius and is referred to as spot size. Figure 2.3 illustrates the quality of the fit between the actual field distribution and the Gaussian distribution for a low value of V.

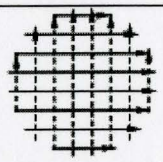
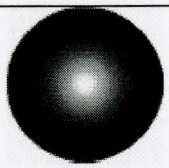
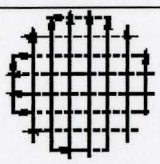
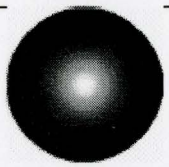
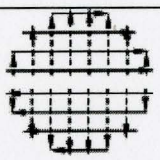
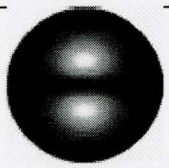
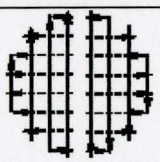
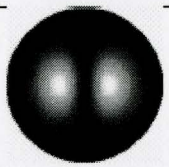
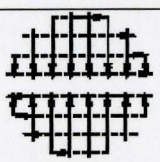
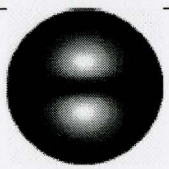
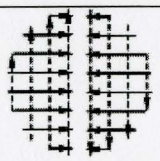
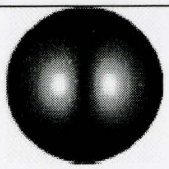
LP Mode Group	Conventional Modes in LP Mode Group	Conventional Modes in LP Mode	Field Lines of LP Mode	Intensity of LP Mode
LP_{01}	HE_{11}	HE_{11}		
		HE_{11}		
LP_{11}	HE_{21} TE_{01} TM_{01}	HE_{21}, TE_{01}		
		HE_{21}, TE_{01}		
		HE_{21}, TM_{01}		
		HE_{21}, TM_{01}		

Figure 2.2 - Mode Classifications and intensity profiles. [21]

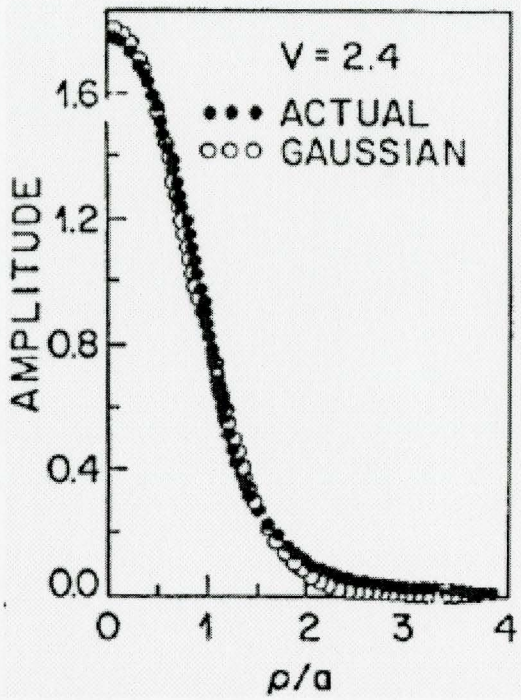


Figure 2.3 – Quality of fit between Gaussian and actual fundamental mode profile [1]

2.1.2 Attenuation

As an optical signal propagates over a long stretch of fiber, it becomes attenuated because of scattering, absorption by material impurities, and other effects. Fiber losses represent a limiting factor because they reduce the signal power reaching the receiver. As optical receivers need a certain minimum amount of power for recovering the signal accurately, the transmission distance is inherently limited by fiber losses.

The average power in a propagating mode can be expressed as [1]

$$\frac{dP(z)}{dz} = -\alpha' P(z) \tag{2.17}$$

where α' is the attenuation coefficient, in units of nepers per meter, and $P(z)$ is the field intensity along the direction of propagation. If $P(0)$ represents the power launched at the input of a fiber of length L, then the power at the end of the fiber will be given by

$$P(\mathrm{L}) = P(0)\exp(-\alpha' \mathrm{L}) \tag{2.18}$$

In practice the attenuation coefficient is expressed in dB/km and is defined as

$$\alpha = -\frac{10}{\mathrm{L}}\log_{10}\left(\frac{P(0)}{P(\mathrm{L})}\right) \tag{2.19}$$

with $\alpha = 4.34 \cdot \alpha'$. Fiber losses depend on the wavelength of transmitted light. Figure 2.54 shows the loss spectrum $\alpha(\lambda)$ of a single-mode fiber with $2a_c = 9.4\,\mu\mathrm{m}$ and $\Delta = 1.9 \times 10^{-3}$ [1].

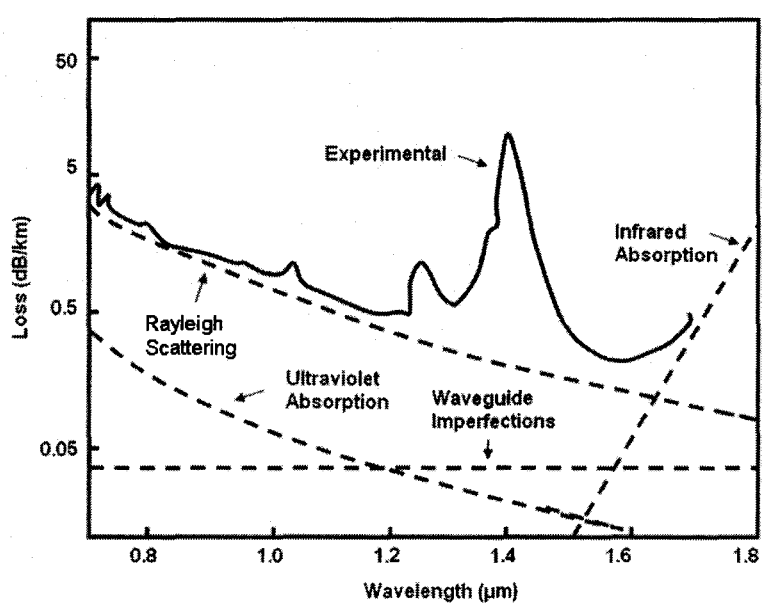


Figure 2.4– Loss spectrum of a standard SMF [1]

The spectrum has two low-loss windows, one around the wavelength $\lambda = 1.3$ μm and one around $\lambda = 1.55$ μm. It also exhibits a strong peak near $\lambda = 1.39$ μm and several other smaller peaks. Fiber losses are considerably higher for shorter wavelengths and exceed 5 dB/km in the visible region.

The principal source of loss in silica fiber in the 1.3 – 1.6 μm range can be attributed to absorption. Absorptive losses are subdivided as intrinsic losses and extrinsic losses. Intrinsic losses are caused by absorption at certain wavelengths corresponding to the electronic and vibrational resonances associated with silica (SiO_2) molecules. For silica, electronic resonances occur in the ultraviolet region $(\lambda < 0.4\ \mu m)$, whereas vibrational resonances occur in the infrared region, peaking near $\lambda = 10$ μm with a tail that extends to the $\lambda = 1.55$ μm range. Extrinsic absorption results from the presence of impurities. Transition-metal impurities such as Fe, Cu, Co, Ni, Mn and Cr absorb strongly in the wavelength range 0.6 – 1.6 μm. Modern manufacturing techniques can reduce impurity concentration to below 1 part per billion resulting in a loss level below 1 dB/km [21]. It is the presence of water vapors that is the main source of extrinsic absorption in standard silica fibers. Residual OH ions cause peaks near the 1.39, 1.24, and 0.95 μm wavelengths. Even an OH ion concentration of 1 part per million can cause a loss of about 50 dB/km at 1.39 μm. Figure 2.5 illustrates the properties of modern 'dry' fiber, which has low loss for wavelengths in the range $\lambda = 1.3 - 1.65$ μm [3]. However, even as the loss approaches the theoretical limits, the dispersion of SMF remains as a fundamental impairment.

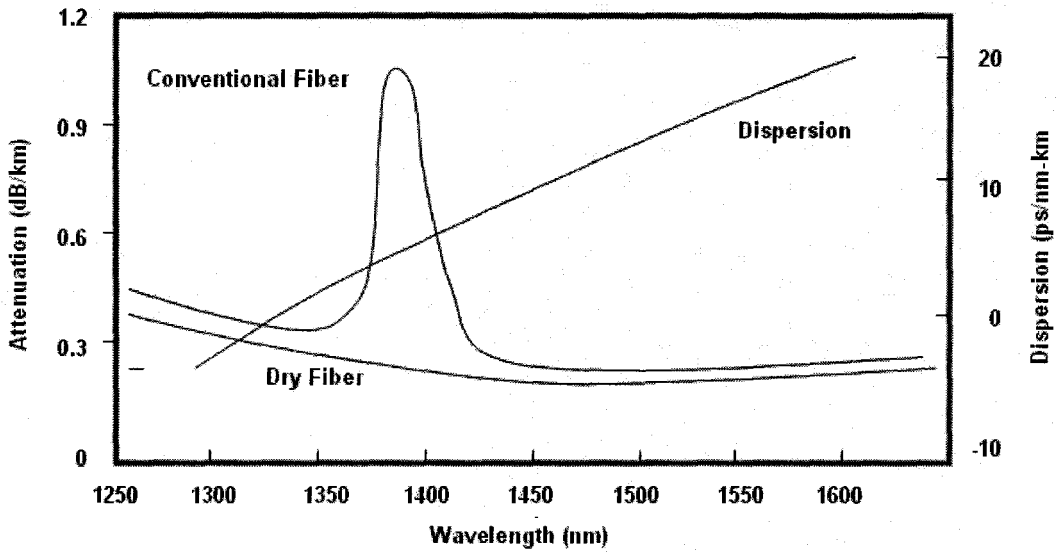


Figure 2.5 – Loss and dispersion of dry fiber compared with conventional fiber [21]

2.1.3 Dispersion

The broadening of short optical pulses attributed to modal dispersion was expressed in eq. (2.4) using the geometrical-optics description where different light rays traverse different paths having different lengths through the fiber. Single-mode fibers eliminate modal dispersion because the total energy of the injected pulse is transported by a single mode. However, dispersive pulse broadening does not disappear altogether because the refractive index of silica fiber is frequency dependent. As a result, the different spectral components of an optical pulse travel at slightly different group velocities and do not arrive simultaneously at the fiber output, which can cause signal ISI. This phenomenon is referred to as group-velocity dispersion (GVD) or chromatic dispersion. Recalling that in a SMF, the transverse dependence of the electric field does not change as the optical signal propagates, the pulse envelope, which can be expressed in terms of its Fourier transform as

$$B(z,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{B}(z,\omega)\exp(-j\omega t)d\omega$$
$$\tilde{B}(z,\omega)=\int_{-\infty}^{\infty}B(z,t)\exp(j\omega t)dt \tag{2.20}$$

can be used to describe how the field propagates through the fiber [1]. The pulse envelope is related to the field intensity by

$$P(z,t)=|B(z,t)|^2 \tag{2.21}$$

If the pulse envelope occupies a narrow band near the carrier frequency ω_0, then the frequency dependence of the fundamental LP_{01} mode can be neglected and the electric field distribution can be written as [1]

$$\mathbf{E}(\mathbf{r},t)=\hat{\mathbf{x}}\cdot E_x(\rho,\omega)\cdot B(z,t)\approx\hat{\mathbf{x}}\cdot E_x(\rho,\omega_0)\cdot B(z,t) \tag{2.22}$$

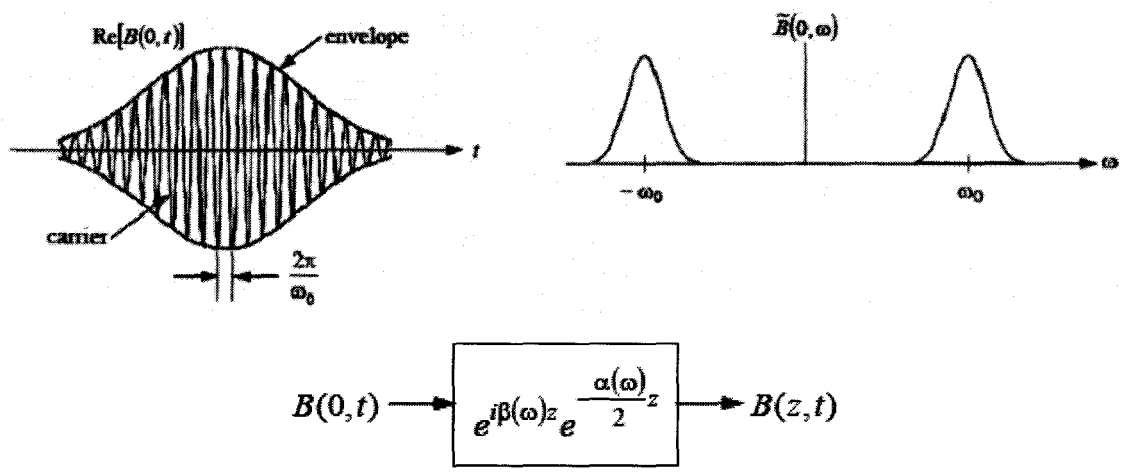


Figure 2.6 – Representation of the pulse envelope and its Fourier transform

The propagation of the field pulse envelope can be modeled as a linear system, where $H_F(\omega,z)=\exp\left(\left[-\alpha(\omega)/2\right]z\right)\exp\left(j\beta(\omega)z\right)$ is the transfer function of the fiber. The propagation constant $\beta(\omega)$ can be expressed as a Taylor series expansion near the carrier frequency $\omega=\omega_0$

$$\begin{aligned}\beta(\omega)&\approx\beta(\omega_0)+\left.\frac{d\beta}{d\omega}\right|_{\omega=\omega_0}\cdot(\omega-\omega_0)+\left.\frac{1}{2}\frac{d^2\beta}{d\omega^2}\right|_{\omega=\omega_0}\cdot(\omega-\omega_0)^2+\left.\frac{1}{6}\frac{d^3\beta}{d\omega^3}\right|_{\omega=\omega_0}\cdot(\omega-\omega_0)^3+....\\&=\beta_0+\beta_1\cdot(\omega-\omega_0)+\frac{1}{2}\beta_2\cdot(\omega-\omega_0)^2+\frac{1}{6}\beta_3\cdot(\omega-\omega_0)^3+....\end{aligned}\tag{2.23}$$

When an optical pulse propagates through the fiber a distance L, the carrier component undergoes a phase shift $\beta(\omega_0)\cdot\mathrm{L}$. Thus, the phase delay T_p is defined as

$$\begin{aligned}\omega_0\cdot T_p&=\beta_0\cdot\mathrm{L}\\T_p&=\frac{\mathrm{L}\beta_0}{\omega_0}\end{aligned}\tag{2.24}$$

yielding the phase velocity v_p given by

$$\begin{aligned}T_p\cdot v_p&=\mathrm{L}\\v_p&=\frac{\mathrm{L}}{T_p}=\left(\frac{\beta_0}{\omega_0}\right)^{-1}\end{aligned}\tag{2.25}$$

The envelope of the pulse is delayed by the group delay

$$\tau_g(\omega_0)=\mathrm{L}\left.\frac{d\beta}{d\omega}\right|_{\omega=\omega_0}=\mathrm{L}\beta_1\tag{2.26}$$

yielding the group velocity v_g defined as

$$v_g = \frac{\mathrm{L}}{\tau_g(\omega_0)} = \left(\left. \frac{d\beta}{d\omega} \right|_{\omega=\omega_0} \right)^{-1} = \beta_1^{-1} \tag{2.27}$$

GVD occurs when the group delay τ_g of the pulse envelope depends on the frequency component ω of the pulse. $\tau_g(\omega)$ can be expanded in a Taylor series near $\omega = \omega_0$ as

$$\begin{aligned} \tau_g(\omega) &\approx \tau_g(\omega_0) + \left. \frac{d\tau_g}{d\omega} \right|_{\omega=\omega_0} \cdot (\omega - \omega_0) + \frac{1}{2} \left. \frac{d^2\tau_g}{d\omega^2} \right|_{\omega=\omega_0} \cdot (\omega - \omega_0)^2 + ... \\ &= \mathrm{L} \left. \frac{d\beta}{d\omega} \right|_{\omega=\omega_0} + \mathrm{L} \left. \frac{d^2\beta}{d\omega^2} \right|_{\omega=\omega_0} \cdot (\omega - \omega_0) + \frac{1}{2} \mathrm{L} \left. \frac{d^2\beta}{d\omega^2} \right|_{\omega=\omega_0} \cdot (\omega - \omega_0)^2 + ... \\ &= \mathrm{L}\beta_1 + \mathrm{L}\beta_2 \cdot (\omega - \omega_0) + \frac{1}{2} \mathrm{L}\beta_3 \cdot (\omega - \omega_0)^2 + ... \end{aligned} \tag{2.28}$$

Therefore GVD occurs when at least one of β_2 or β_3 is nonzero. The extent of pulse broadening due to GVD, corresponding to $\beta_2 \neq 0$, can be expressed in terms of ΔT_B for a pulse envelope having frequency spread Δω by

$$\Delta T_B \approx |\beta_2| \cdot \mathrm{L} \cdot \Delta\omega \tag{2.29}$$

For the case $\beta_2 = 0$, the pulse broadening is approximately

$$\Delta T_B \approx \frac{1}{2} |\beta_3| \cdot \mathrm{L} \cdot (\Delta\omega)^2 \tag{2.30}$$

The group delay can be written in terms of wavelength for a pulse envelope occupying a narrow bandwidth near $\lambda = \lambda_0$ as

$$\begin{aligned} T_g(\lambda) &\approx T_g(\lambda_0) + \left.\frac{dT_g(\lambda)}{d\lambda}\right|_{\lambda=\lambda_0} \cdot (\lambda - \lambda_0) + \frac{1}{2}\left.\frac{d^2T_g(\lambda)}{d\lambda^2}\right|_{\lambda=\lambda_0} \cdot (\lambda - \lambda_0)^2 + ... \\ &= \frac{L}{v_g(\lambda_0)} + L\frac{d}{d\lambda}\left(\frac{1}{v_g(\lambda)}\right)\Bigg|_{\lambda=\lambda_0} \cdot (\lambda - \lambda_0) + \frac{1}{2}L\frac{d^2}{d\lambda^2}\left(\frac{1}{v_g(\lambda)}\right)\Bigg|_{\lambda=\lambda_0} \cdot (\lambda - \lambda_0)^2 + ... \\ &= \frac{L}{v_g(\lambda_g)} + LD \cdot (\lambda - \lambda_0) + \frac{1}{2}LS \cdot (\lambda - \lambda_0)^2 + ... \end{aligned} \tag{2.31}$$

Therefore, if a pulse envelope has a wavelength spread $\Delta\lambda_S$, a signal propagating over a distance L is subject to a group delay spread given approximately by

$$\Delta T_B \approx |D| \cdot \mathrm{L} \cdot \Delta\lambda_S \tag{2.32}$$

where the dispersion parameter D is given by

$$D = \frac{d}{d\lambda}\left(\frac{1}{v_g(\lambda)}\right) = \frac{d\omega}{d\lambda} \cdot \frac{d}{d\omega}\left(\frac{1}{v_g(\omega)}\right) = -\frac{2\pi c}{\lambda^2} \cdot \beta_2 \tag{2.33}$$

The dispersion parameter, expressed in ps/(km-nm), measures the change in group-delay per change in wavelength for a given length of fiber. For example, a standard SMF operated at 1.55 μm has D = 17 ps/(km-nm), which means that a very narrow pulse launched into the fiber will spread out to 17 ps after 1-km given a 1.55 μm source with a linewidth of 1 nm [20]. Thus, the dispersion parameter D approximates to a first-order how rapidly a narrow pulse spreads out after passing through a length of fiber. Systems

that operate at 1.55 μm are said to be dispersion-limited. An order-of-magnitude estimate for the capacity-distance product for these systems is

$$\mathrm{BL} < \frac{1}{|D|\Delta\lambda_S} \tag{2.34}$$

where $\Delta\lambda_S$ is the spectral width of the light source. For example, the maximum bit rate for a 1-km fiber operating with a 1.55 μm light source with a 1-nm linewidth is approximately 58 Gbps. Significant improvements can be gained by modulating a laser with a smaller line-width $\Delta\lambda_S$. If the laser was perfectly monochromatic, and followed by a perfect intensity modulator, the effects of chromatic dispersion would be much less severe. In particular, the optical spectrum of an ideally modulated monochromatic laser would look like that of an amplitude-modulated (AM) transmitter, with a carrier and two sidebands corresponding to the spectrum of the baseband signal. With NRZ modulation, and taking the 3-dB bandwidth of the signal as approximately equal to the bit-rate B, results in a spectral linewidth of

$$\Delta\lambda_S \approx \frac{\lambda^2}{c}\mathrm{B} \tag{2.35}$$

A 10 Gbps signal at $\lambda = 1.55$ μm from a perfectly monochromatic light source would therefore have a linewidth of only 0.08 nm. This means that a perfectly monochromatic laser would allow a factor of 12.5 increase in the maximum distance that the signal could travel over a standard SMF having D = 17 ps/(nm-km) compared to that a conventional laser with linewidth $\Delta\lambda_S = 1\,\mathrm{nm}$.

The origins of GVD can be understood by considering the group index $\bar{n}_g$ of the fundamental mode

$$\bar{n}_g = \frac{c}{v_g(\omega)} \tag{2.36}$$

Using the expressions for the modal index and group velocity it follows that [1]

$$\bar{n}_g = \bar{n} + \omega \frac{d\bar{n}}{d\omega} \tag{2.37}$$

Using equation (2.36) in equation (2.33), the dispersion parameter D can therefore be seen to be composed of two components

$$\begin{aligned} D &= -\frac{2\pi c}{\lambda^2}\frac{d}{d\omega}\left(\frac{1}{v_g}\right) = -\frac{2\pi}{\lambda^2}\left(2\frac{d\bar{n}}{d\omega} + \omega\frac{d^2\omega}{d\omega^2}\right) \\ &= D_M + D_W. \end{aligned} \tag{2.38}$$

The material dispersion D_M represents the contribution of the silica glass to the fiber dispersion caused by the variation of the refractive index with frequency. The waveguide dispersion D_W represents the contribution of the waveguide design to the fiber dispersion. This term is less important than D_M in standard SMF but becomes comparable to D_M in dispersion-shifted fibers and non-zero dispersion-shifted fibers [21]. A plot of total dispersion, (figure 2.7) versus the transmitted wavelength reveals that there is a point of zero dispersion at the point near 1.31 μm, and systems operating at this wavelength, known as the zero-dispersion wavelength λ_{ZD} are loss-limited rather than dispersion-limited. The dispersive effects, however, do not disappear completely at $\lambda = \lambda_{ZD}$. Optical pulses still experience broadening because of higher-order dispersive effects. This is because D cannot be made zero at all wavelengths contained within the pulse spectrum centered at λ_{ZD}. Higher-order dispersive effects are governed by the dispersion slope also known as the differential-dispersion parameter S, expressed as

$$S = \frac{dD(\lambda)}{d\lambda} = \frac{d^2}{d\lambda^2}\left(\frac{1}{v_g(\lambda)}\right) = \frac{4\pi c}{\lambda^3}\cdot\beta_2 + \left(\frac{2\pi c}{\lambda^2}\right)^2\cdot\beta_3 \tag{2.39}$$

measured in units of $\text{ps}/\left(\text{nm}^2\cdot\text{km}\right)$.

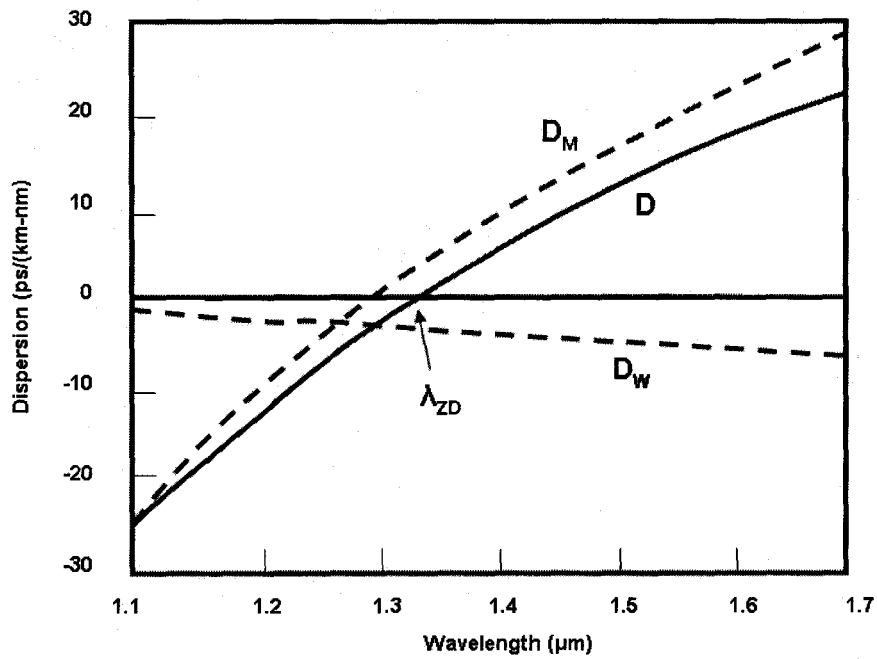


Figure 2.7 – Total Dispersion and relative contributions of material dispersion and waveguide dispersion [21]

Therefore, when D = 0 at $\lambda = \lambda_{ZD}$, a pulse envelope having a wavelength spread $\Delta\lambda_S$ propagating over a distance L is subject to a group delay spread of approximately

$$\Delta T_B \approx \frac{1}{2}|S|\cdot \text{L}\cdot\left(\Delta\lambda_S\right)^2 \tag{2.40}$$

The limiting capacity-distance product can be obtained in this case by using equation (2.34) and replacing the dispersion parameter with $D = S\Delta\lambda_S$, yielding

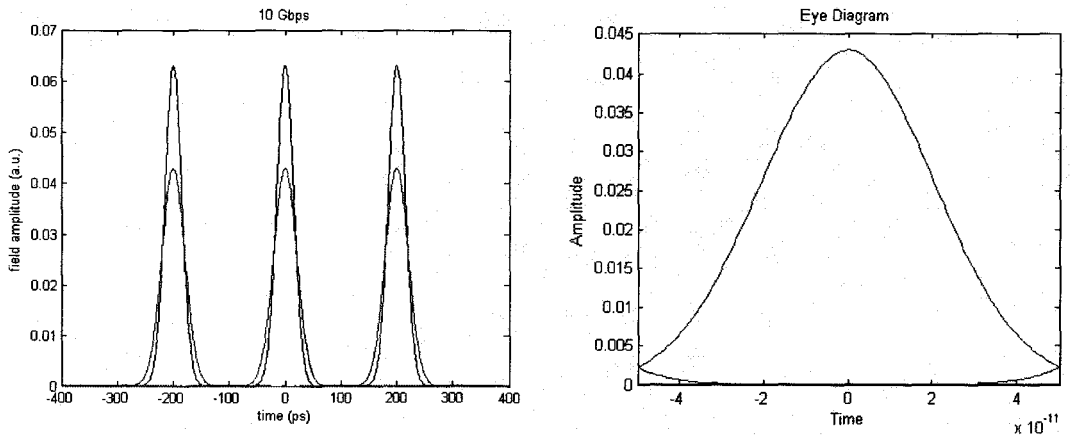
$$\mathrm{BL} < \frac{1}{|S|(\Delta\lambda_S)^2} \tag{2.41}$$

First-order effects of chromatic dispersion on Gaussian optical pulses are illustrated in figure 2.8. Gaussian pulses having an optical power $P(0)$ launched into the fiber can be expressed as

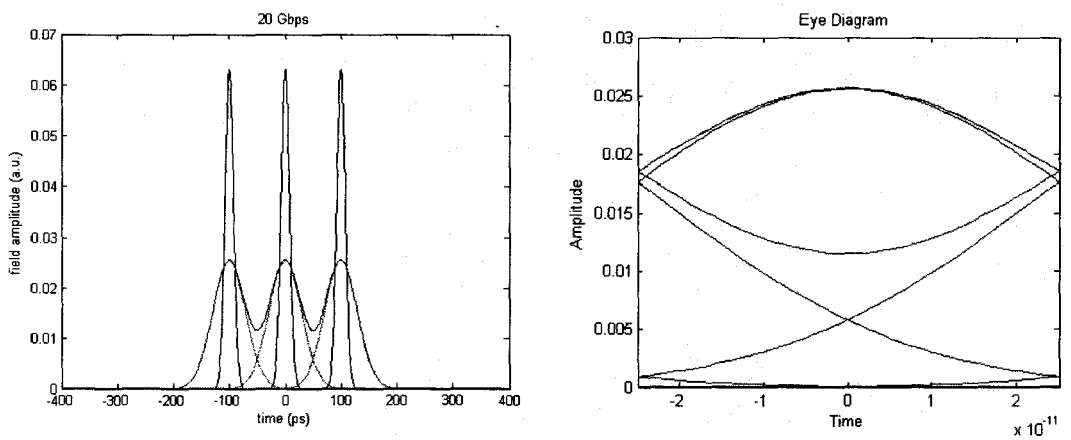
$$B(0,t) = \sqrt{P(0)}\exp\left[-\frac{1}{2}\left(\frac{t}{T_0}\right)^2\right] \tag{2.42}$$

with a full-width at half-maximum (FWHM) equal to $T_{FWHM} = 2\sqrt{\ln 2}\cdot T_0 \approx 1.665T_0$, where T_0 represents the half-width at 1/e intensity point. A 10 Gbps NRZ system with $P(0) = 1\,\mathrm{mW} = 0\,\mathrm{dBm}$ and a bit period of $T_B = 4\cdot T_{FWHM}$ results in optical pulses with $T_{FWHM} = 0.25$ ps. The output of a pulse stream representing a [1 0 1 0 1] data pattern launched through 10 km lossless fiber with dispersion parameter $D = 17\ \mathrm{ps/(nm\cdot km)}$ is shown in figure 2.8 for 10, 20, and 40 Gbps signals, with the black traces representing the input pulses, the red traces showing the response at the fiber output for each individual pulse, and the blue traces showing the superposition of all output pulses at the fiber output which forms the received signal. The associated eye diagrams reveal severe distortions of the eye diagram as the bit rate is increased. The only way to minimize the ISI would be to reduce the bit rate to such a point where neighboring pulses do not interfere with each other as significantly, or to increase the transmitted optical power so that the vertical eye opening is restored to the level that occurred with no ISI present. The power penalty associated with ISI induced eye closure can be obtained by comparing the vertical eye opening V_E in the eye diagram without ISI to the reduced eye opening V_E' due to ISI [20], yielding

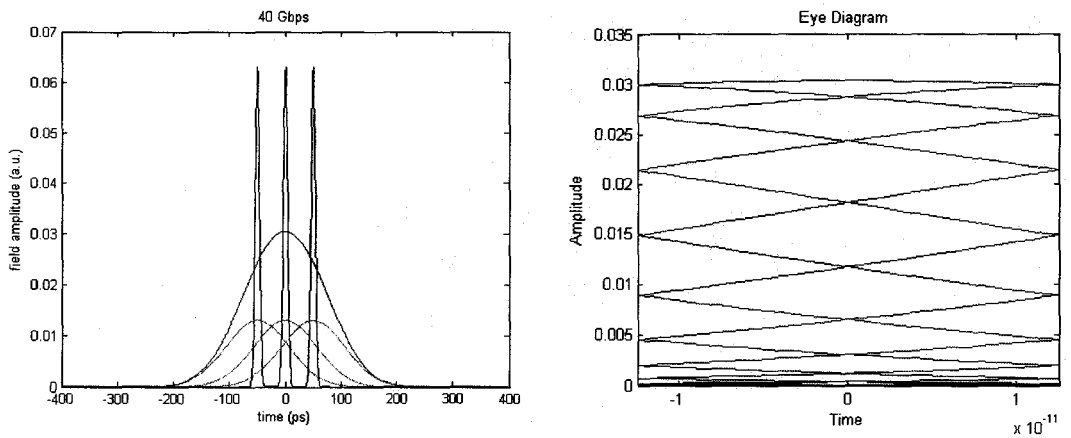
$$PP_D = 2\cdot\frac{V_E}{V_E'} \tag{2.43}$$



a)



b)



c)

Figure 2.8 – Effects of GVD on a series of optical pulses at a) 10 Gbps b) 20 Gbps c) 40 Gbps

2.2 Optical Transmitter

2.2.1 Semiconductor Lasers

The development of optical fibers based on fused silica having zero material dispersion (i.e. a vanishing derivative of group delay with respect to λ) at $\lambda = 1.3$ μm and minimum loss at $\lambda = 1.55$ μm drove the need for semiconductor laser diodes that can operate in this wavelength range. Laser diodes for telecommunications are made from the quaternary material system InGaAsP/InP because this semiconductor material can emit at wavelengths between 1.15 μm and 1.67 μm. InGaAsP/InP lasers also exhibit many other desirable properties such as low threshold current, high differential quantum efficiency, high linearity and thermal stability, high power in a single transverse mode and high modulation bandwidth [15], [22].

A semiconductor laser consists of an optical gain medium located in a cavity formed by two reflecting facets. The gain medium consists of a thin active layer of indium-gallium-arsenide-phosphide (InGaAsP) sandwiched between p-type and n-type cladding layers of a different semiconductor (InP) with a higher bandgap, resulting in a double heterostructure. The bandgap of the sandwiched $In_xGa_{1-x}As_yP_{1-y}$ compound can be controlled by the mixing ratios x and y to provide optical gain anywhere in the 1.0 to 1.6 μm range [22]. Thus, 1.3 μm as well as 1.5 μm lasers are based on an InGaAsP active layer, which can be lattice matched to the surrounding InP cladding layers and the substrate. The use of a heterostructure geometry for semiconductor optical sources is beneficial because the bandgap difference between the two semiconductors helps to confine charge carriers in the active layer, so that if the optical gain of the cavity is greater than unity, an incoming photon can stimulate the recombination of an electron-hole pair to produce a second identical photon, resulting in light amplification by stimulated emission (lasing) [21]. This phenomenon can be illustrated with the light output power versus current characteristic of a laser diode (fig. 2.9). The threshold current

I_{th} is the current that needs to be applied to the laser to obtain a threshold carrier density which achieves population inversion so that stimulation emission can dominate. Below the threshold current, the emitted light power is due only to spontaneous emission, which has a low efficiency since the emitted photons in this case are random in phase (i.e., non-coherent).

The laser optical output of directly modulated (i.e. intensity modulated) lasers depends on the threshold current, the modulating current amplitude and the current-to-light conversion efficiency (i.e. slope efficiency) of the laser diode. The slope efficiency is the ratio of the optical output power to the input current, when driven above threshold. These characteristics are strongly related to the laser structure, fabrication process, and operating temperature. Achieving maximum confinement between carriers and photons is essential to achieve a laser diode with a low threshold current and a good slope efficiency. Therefore, buried heterostructures, where the active layer is buried on all sides

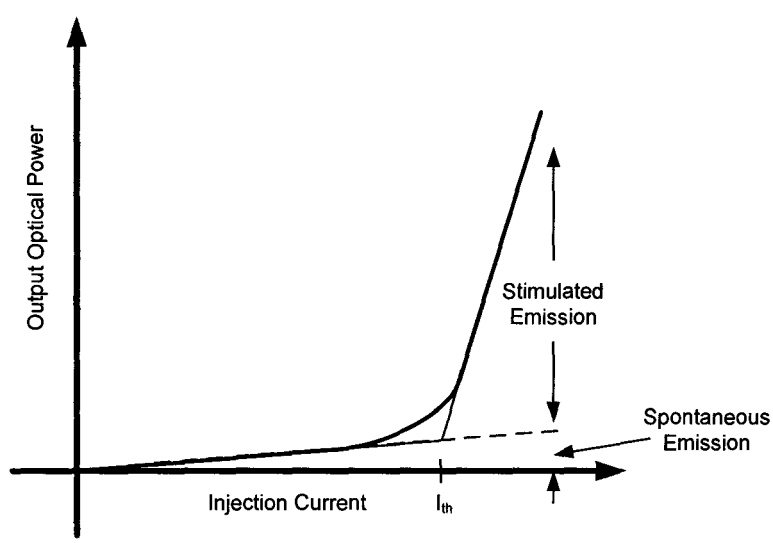


Figure 2.9– Relationship between output optical power and injected current for a typical semiconductor laser [15]

by several layers of lower refractive index material, are utilized in high-performance semiconductor lasers [15], [42]. In addition, the active region is usually structured as a

multiple quantum well (MQW) in practical lasers, resulting in even better performance [22]. Importantly, the active layer acts as a planar waveguide because of the refractive index difference between it and the p-type and n-type cladding layers. Similar to the case of optical fibers, the active layer can support a single transverse mode if the active layer is made thin enough. However, a semiconductor laser cavity generally emits light in several longitudinal modes, since the gain spectrum is wide enough to allow the many longitudinal modes of the cavity to simultaneously meet the conditions for gain and phase required for stimulated emission. As a result, the spectrum of the emitted laser light has multiple peaks (e.g. figure 2.10a), corresponding to a large source spectral width $\Delta\lambda_S$, which is prohibitive in dispersion-limited systems. Single longitudinal mode (SLM) operation can only be achieved if the lasing threshold for the oscillating mode is significantly smaller than the threshold for the other cavity modes. Therefore, SLM lasers are designed such that the cavity losses are different for different longitudinal modes of the cavity. The longitudinal mode with the smallest cavity loss reaches threshold first and becomes the dominant mode. Other neighboring modes are discriminated by their higher losses, which prevent their buildup from spontaneous emission [21].

A commonly used SLM laser is a distributed feedback (DFB) laser. It emits a single-longitudinal mode due to grating formed close to the active layer in the axial direction which provides a periodic variation in mode index in the laser heterostructure waveguide (e.g. figure 2.10b), resulting in the desired mode selectivity [22]. The grating period is determined by the desired emission frequency according to the Bragg condition.

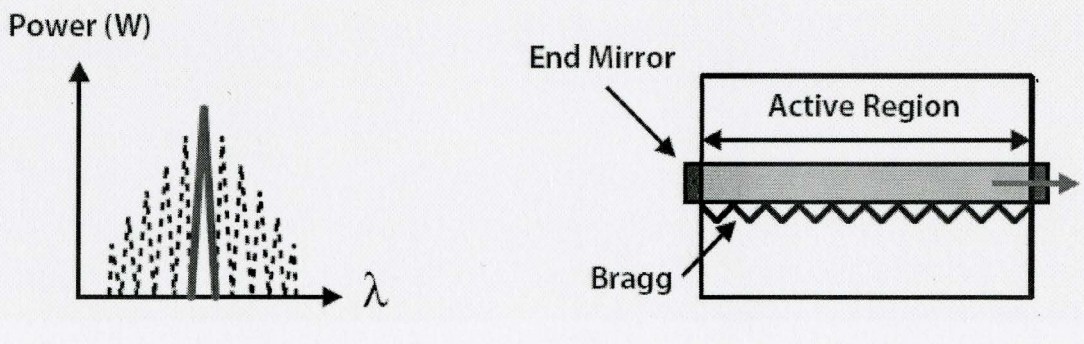


Figure 2.10 – a) Longitudinal spectrum of a FP laser (dashed lines) and a SLM laser (solid line) b) Physical structure of a DFB laser [40]

The performance of a SLM laser is characterized by the mode-suppression ratio (MSR), defined as

$$MSR = \frac{P_{mm}}{P_{sm}} \tag{2.44}$$

where P_{mm} is the main-mode power and P_{sm} is the power of the most dominant side mode [15]. The MSR should exceed 30 dB for a good SLM laser resulting in very narrow spectral linewidth suitable for high-speed and long-haul applications.

SLM lasers can convey information for optical transmission using either 'external' or 'direct' modulation. In the first scheme, the laser is maintained in a constant light-emitting state (i.e., continuous wave (CW) state), and an external shutter, typically a Mach-Zehnder type electro-optic modulator, modulates the output intensity according to an externally applied voltage. This results in high-quality optical pulses since there is no interaction between the information being applied to the modulator and any internal dynamics of the laser.

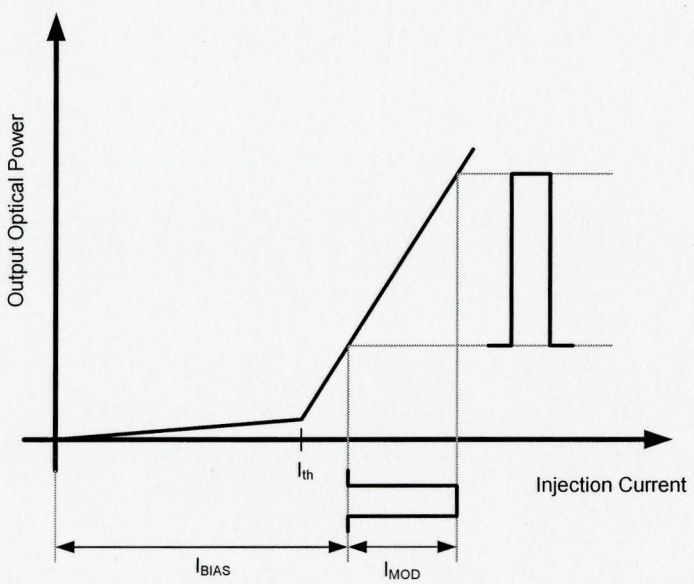


Figure 2.11 – Illustration of direct modulation using light-current relationship of laser diodes

External modulation requires costly external apparatus therefore direct modulation is the preferred scheme because of its simplicity, compactness and cost-effectiveness. Direct modulation involves changing the input current ($I_{BIAS} + I_{MOD}$) of the laser in order to produce time-dependent output in optical intensity (e.g. figure 2.11).

When the laser diode is driven close to and above threshold, the relationship between the output optical power and the driving current is approximately linear,

$$P_O = S_L \cdot (I - I_{th}) \tag{2.55}$$

where S_L is the slope efficiency of the laser. Ideally, any change in the injected current would yield an instantaneous change of the emitted optical power. While this is the case for low speed modulation (up to 10 MHz), the internal dynamics of the laser can cause transient oscillations when a step current input is applied [15]. These oscillations, resulting from the interplay between the optical field and the carrier density, not only cause distortions of the optical pulse shape, but also broaden the signal's optical spectrum under high-speed modulation. Therefore, the internal dynamics of semiconductor lasers, and their implications in high-speed direct modulation will be discussed in the following section.

2.2.2 Dynamic Characteristics of Modulated Lasers

The intrinsic dynamics of the semiconductor laser are central to their overall modulation response. The dynamic relationship between the laser current and the light output is fully described by the laser rate equations. The rate equations are two coupled, nonlinear, ordinary differential equations that relate the carrier density and photon density in the active region. For a single-mode laser, the laser rate equations can be expressed as

$$\frac{dS}{dt} = G(S,N)S + \Gamma\beta_L \frac{N}{\tau_N} - \frac{S}{\tau_S}$$
$$\frac{dN}{dt} = \frac{I_A}{qV_A} - \frac{N}{\tau_N} - G(S,N)S \tag{2.56}$$

where S and N are the photon and carrier densities inside the active region, respectively [1]. The photon and carrier densities are assumed constant across the active layer (zero diffusion), which is a reasonable approximation for high-speed lasers that employ buried heterostuctures with strong carrier confinement. $G(S,N) = \Gamma v_g \sigma_g (N - N_T)(1 - \kappa \cdot S)$ is the net rate of stimulated emission where κ is a parameter accounting for the non-linear optical gain, σ_g is the differential gain, N_T is the transparency value of the carrier density and Γ is the optical confinement factor given by the ratio of the active region volume to the modal volume. β_L is the fraction of spontaneous emission coupled into the lasing mode, I_A is the current injected into the active region, V_A is the active region volume and τ_S and τ_N are the photon and carrier lifetimes, respectively.

The rate equations describe the exchange of energy between electrons and photons in the active region, and can be interpreted intuitively as follows; $G(N,S)\cdot S$ and $\Gamma\beta_L N/\tau_N$ photons are generated per unit time in the lasing mode due to stimulated and spontaneous emission, respectively, while S/τ_S photons are lost per unit time due to absorption. Similarly for the carriers; I_A/qV_A carriers are injected into the cavity per unit time where $G(N,S)\cdot S$ are lost due to stimulated emission and N/τ_N are lost due to spontaneous emission and non-radiative recombination. The output power is related to the photon density as

$$P = (SV_A\eta_L h\nu)/\Gamma\tau_S \tag{2.57}$$

where η_L is the internal quantum efficiency of the laser and h is Planck's constant [1].

The rate equations, when suitably normalized, can be solved numerically to obtain the large-signal response of a laser to any arbitrary current input waveform. The response of a buried hetereostructure laser at zero bias in the particular case of a current step input (above threshold) is illustrated in figure 2.12. When the laser at zero bias is driven by a current step input, the onset of optical power emission is delayed by a delay time t_s. For $\kappa = 0$, and $S_{off} << S_{on}$, the turn-on time is given by [15]

$$t_{on} = \frac{\sqrt{2}}{\omega_{on}} \left[\ln\left(\frac{P_{on}}{P_{off}} \right) \right]^{1/2} \tag{2.58}$$

where $\omega_{on} \propto \sqrt{S_{on}}$, and P_{on}, P_{off}, S_{on}, S_{off} are the input/output optical powers and photon densities for ONE/ZERO bits, respectively.

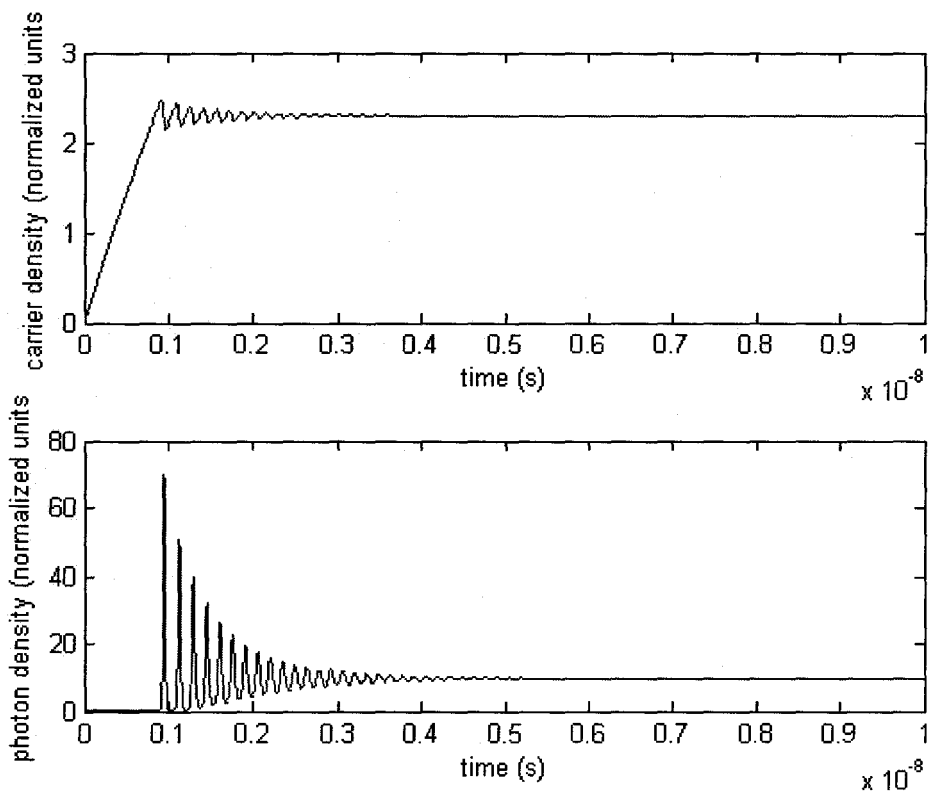


Figure 2.12 – Step-response of a directly modulated laser [5]

For fast switching operation, it is common to bias the laser slightly above the threshold to avoid turn-on and turn-off delay. The overshoot an also be reduced considerably if the laser is biased well above threshold. Near threshold, the carrier density is already sufficiently built up and the net optical gain is close to unity, yet the photon density is still low because spontaneous emission dominates. With a small step current applied, the laser turns on much more quickly than if it were biased below threshold. High-speed directly modulated semiconductor lasers are therefore biased above threshold in practice. An important system parameter is the on-off ratio P_{on}/P_{off}, also known as the extinction ratio when expressed in a log scale,

$$ER = 10\log_{10}\left(\frac{P_{on}}{P_{off}}\right) \tag{2.59}$$

When the extinction ratio is not optimum (i.e., when P_{off} is non-zero), the transmitted power must be increased in order to maintain the same BER, resulting in a power penalty of [1]

$$PP_{ER} = 10\log_{10}\left(\frac{ER+1}{ER-1}\right) \tag{2.60}$$

To maintain a power penalty of less than 2dB, the extinction ratio of the laser should remain above 7 dB.

In addition to turn-on delay, the current step response of a directly modulated laser exhibits a distinct overshoot and relaxation oscillations. The dynamic exchange between elevated carrier and photon densities leads to the relaxation oscillations and can be described as follows. Before a step-current pulse is applied, the carrier density in the gain medium is zero. After the current turn-on, the carrier density builds up until the threshold point is reached where the net optical gain becomes unity and the laser begin to produce coherent light. As soon as the lasing sets in, stimulated emission causes many carriers to

recombine, thus producing more photons and optical power. However, the high rate of recombination leads to a reduction in the carrier density and thus a lower optical gain. Once the photon density begins to drop, the carriers begin to build up again after a certain time and the cycle continues. Under high-speed direct NRZ modulation, relaxation oscillations can distort the transmitted waveform through ISI, resulting in an increased BER and a power-penalty at the receiver. Many techniques have been proposed to suppress the effects of relaxation oscillations under direct modulation, including modification of the physical laser device structure, the use of external optical feedback, as well as the shaping of the current pulses that are used to encode logical bits into the optical output power of the laser. By using appropriately shaped input currents, the state-space trajectory of the laser can be controlled, resulting in a better BER performance for high bit rates because the errors caused by ISI due to the internal (non-linear) dynamics of the laser can be eliminated [41]. Figure 2.13 compares the response due to shaped current pulses and NRZ square-like currents, along with the corresponding trajectories in state-space for a random sequence of a million bits transmitted over an additive white Gaussian noise (AWGN) channel [41]. Also shown (figure 2.14) is the SNR vs. BER performance at high- and low-bit-rates for shaped and non-shaped current pulses using a sampling receiver and assuming a perfectly recovered clock. The results show that for high bit rates, even for very small noise levels (high SNR) direct modulation using NRZ shaped pulses essentially fails because of errors caused by ISI due to the laser dynamics. For lower bit rates, the bit errors attributed to the laser dynamics become more and more negligible, as indicated by the inset of figure 2.14, showing that for lower bit rates the performance enhancement due to the use of shaped input currents is minor. The improvement for high-bit rates comes at the cost of more complex electronics at the transmitter, which would become necessary to shape the input current appropriately.

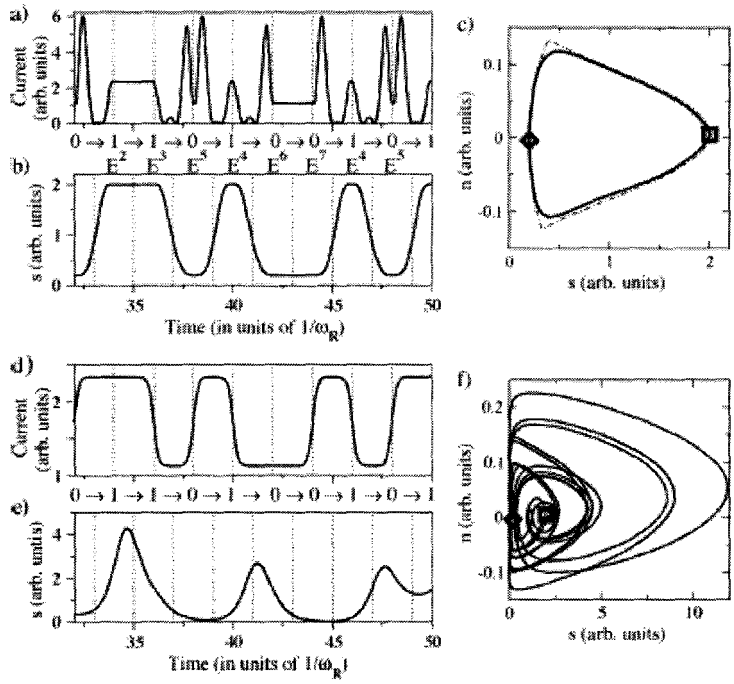


Figure 2.13 – Comparison between shaped current modulation (a) – (c) and square-like currents (d) – (f) and the corresponding state-space trajectories [41]

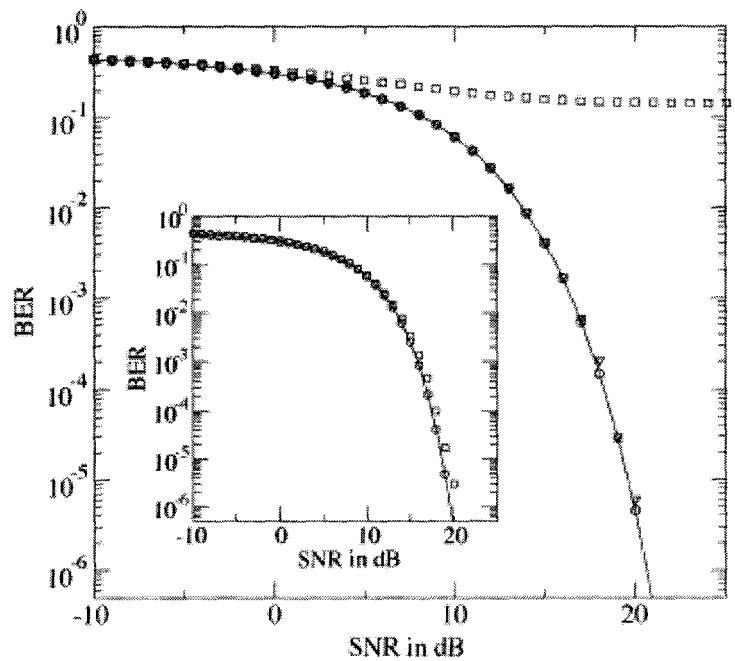


Figure 2.14 – SNR versus BER for laser driven both by shaped currents (circles and lines) and square-like currents (squares). The main panel shows high bit rate and the inset lower bit rate communication. [41]

2.2.3 Spectral Characteristics of Modulated Lasers

Amplitude modulation is always accompanied by frequency modulation in practical semiconductor lasers. Frequency modulation is caused by the modulation-induced variations in the carrier density. Because the refractive index is a function of the carrier population, small variations in carrier density, apart from producing relaxation oscillations in the optical output, will give rise to a variation in the refractive index. These variations, in turn, change the optical path length of the laser cavity and produce a dynamic wavelength shift, particularly at the rising and falling edges of an applied current pulse [42]. The wavelength shifts briefly toward shorter wavelengths (blue) and then back to the equilibrium value during laser turn-on, and toward longer wavelengths (red) and back during turn-off. Therefore, under high-speed direct modulation, SLM lasers exhibit time-dependent frequency excursions, called chirping, that cause an apparent spectral line-width broadening. Since the effective mode index near the threshold can be written as [15]

$$\bar{n}(N,\omega_0)=\bar{n}(N_T,\omega_0)+\frac{\partial \bar{n}}{\partial \omega}(\omega-\omega_0)+\frac{\partial \bar{n}}{\partial N}(\omega-\omega_0) \tag{2.61}$$

it can be shown that

$$(\omega-\omega_0)=-\frac{\omega_0}{n_g}\frac{\partial \bar{n}}{\partial N}(N-N_T) \tag{2.62}$$

where it is evident that modulating the carrier density around the threshold will also cause variations of the optical frequency around its threshold value. The frequency modulation of the optical field can be expressed as [15]

$$\begin{aligned}\frac{d\phi}{dt}=\omega-\omega_0&=-\frac{\omega_0}{n_g}\frac{\partial \bar{n}}{\partial N}(N-N_T)\\&=\frac{1}{2}\mathrm{A}v_g\frac{\partial G}{\partial N}(N-N_T)\end{aligned} \tag{2.63}$$

where A is the linewidth enhancement factor defined as,

$$\mathrm{A} = -2k_0\left(\frac{\partial \bar{n}/\partial N}{\partial G/\partial N}\right) \tag{2.64}$$

The linewidth enhancement factor can be used to relate the optical frequency variation (chirp), Δf, to the change in output optical power as [43]

$$\Delta f(t) \approx \frac{\mathrm{A}}{4\pi}\cdot\frac{d}{dt}\ln P(t) \tag{2.65}$$

As a result, large relaxation oscillations will cause the optical frequency to oscillate during fast turn-on and turn-off transients. Therefore, the use of appropriately shaped input current pulses, as well as lasers with strong damping of relaxation oscillations would result in a transmitter with reasonably good chirp characteristics.

The effects of transient chirp on the overall system performance can be illustrated with the simulation of a 10 km, 10 Gbps point-to-point optical link over standard SMF using direct modulation. The OptSim software package was used to obtain the following results. A pseudo-random bit sequence (PRBS) of length $2^{11}-1$ was generated and fed into an electrical modulator to create an NRZ signal with raised-cosine pulses having a 0.1694 roll-off factor. The eye diagram of the electrical modulator output is shown in figure 2.16a. This signal was used to drive a 1330 nm DFB laser, with the specifications given in table 2.1. The optical signal from the directly modulated laser was transmitted at the 1330 nm wavelength though 10 km of standard SMF with a loss of $\alpha = 0.367$ dB/km, a dispersion parameter of $D = 3.3\ \mathrm{ps/(nm\cdot km)}$, and a dispersion slope of $S = 0.0868\ \mathrm{ps/(nm^2\cdot km)}$. The optical signal power before and after transmission through the fiber is shown in figure 2.15. The distinguishing features of high-speed directly modulated lasers are the large overshoots due to relaxation oscillation and the broadening of the optical spectrum due to frequency chirping. The large overshoots are

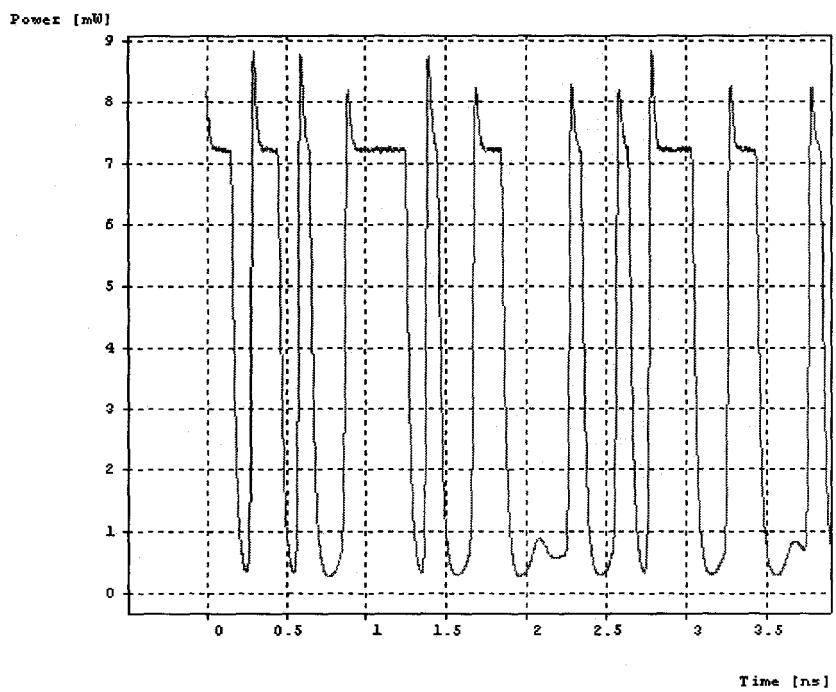
caused by modulating the laser at a rate close to its relaxation oscillation peak frequency. Figure 2.16 shows the extent of the frequency variation of the laser output spectrum. The large overshoot exasperates the frequency chirp, with the largest frequency excursions occurring on the rising edge of each optical pulse. Because the spectrum of the laser output signal varies with time, the spectral width of the laser becomes broader on average, leading to an increased dispersion penalty. The peak overshoot of directly modulated lasers can be reduced by lowering the extinction ratio. The optical receiver used in the simulation comprises a PIN photodiode with responsivity of 0.94 A/W, and a fourth-order Bessel low-pass filter with 7.5 GHz bandwidth with an overall sensitivity of -27 dBm. The eye diagram of the signal from the optical receiver is shown in figure 2.17b. It can be seen that even for the 1330 nm low-dispersion wavelength, the laser chirping affects the eye closure.

Parameter	**Value**
Bit Rate	10 Gbps
Wavelength	1330 nm
Threshold Current	15 mA
Slope Efficiency	0.175 mW/mA
Spectral Linewidth	10 MHz
Linewidth Enhancement Factor	6.45
Extinction Ratio	8.4 dB
Relaxation Oscillation Peak Freq.	10 GHz
Relaxation Oscillation Peak overshoot	0.1 dB
Dispersion Parameter	3.3 ps/nm-km
Dispersion Slope	0.0868 ps/nm^2-km
Fiber Length	10 km
Receiver Frequency Repsonse	4th-order Bessel LPF ($f_{3dB} = 7$ GHz)
Responsivity	0.95 A/W
Receiver Sensitivity	-27 dBm

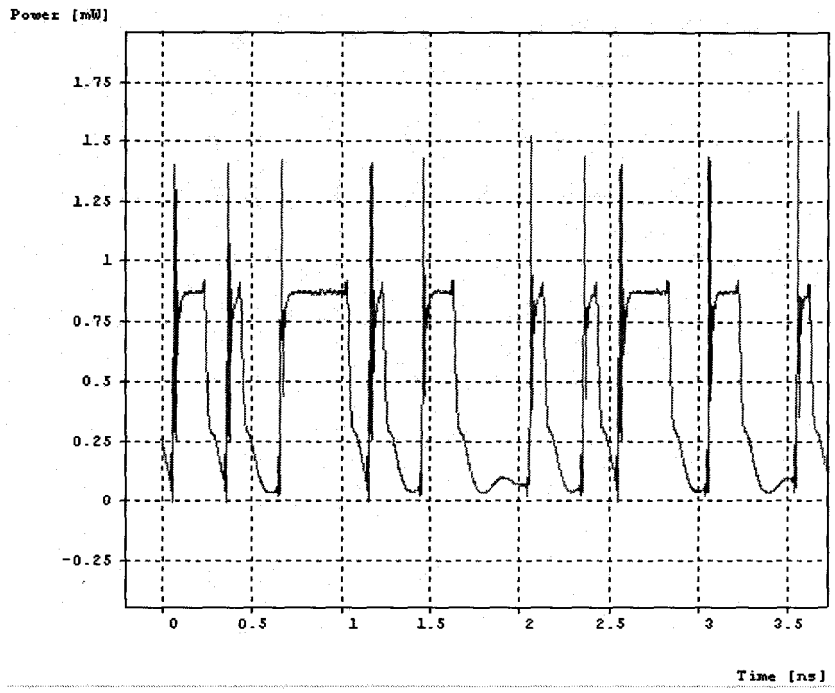
Table 2.1 –Direct modulation simulation parameters for an Agere D1861C laser

The power penalty associated with the transient chirp can be obtained by assuming that each shifted portion of the emission spectrum last for a time t_c, equal to one half of the relaxation oscillation period. For a single transmitted optical pulse of duration $1/\mathrm{B}$, the fraction of total energy carried by the shifted light is therefore approximately $2 \cdot t_c \cdot \mathrm{B}$. When a pulse is transmitted through a SMF of length L with dispersion D, the blue-shifted power will move to earlier times and the red-shifted power will move to later times with respect to the center of the pulse. The amount of power which moves outside the symbol interval will increase linearly with distance along the fiber until all the chirped power has left the bit interval. The resulting power penalty can be approximated by assuming that the signal remaining is the difference between the powers in and out of the initial symbol interval, yielding [42]

$$PP_C = 10\log\left(\frac{1}{1-4\mathrm{LB}D\Delta\lambda_S}\right) \tag{2.66}$$



a)



b)

Figure 2.15 – Simulated waveform of a 10 Gbps NRZ signal from a directly modulated laser a) before transmission through the fiber b) after transmission through the fiber

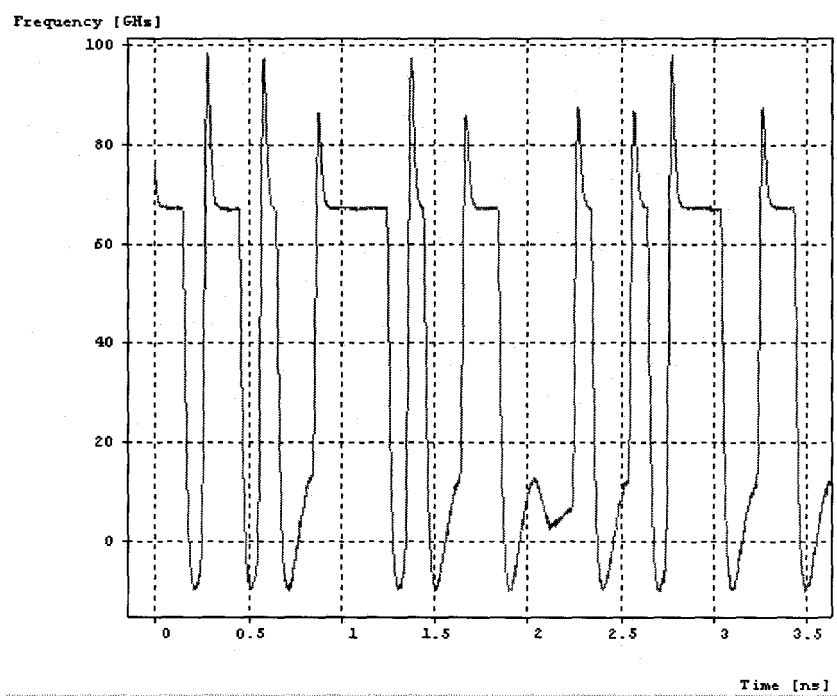
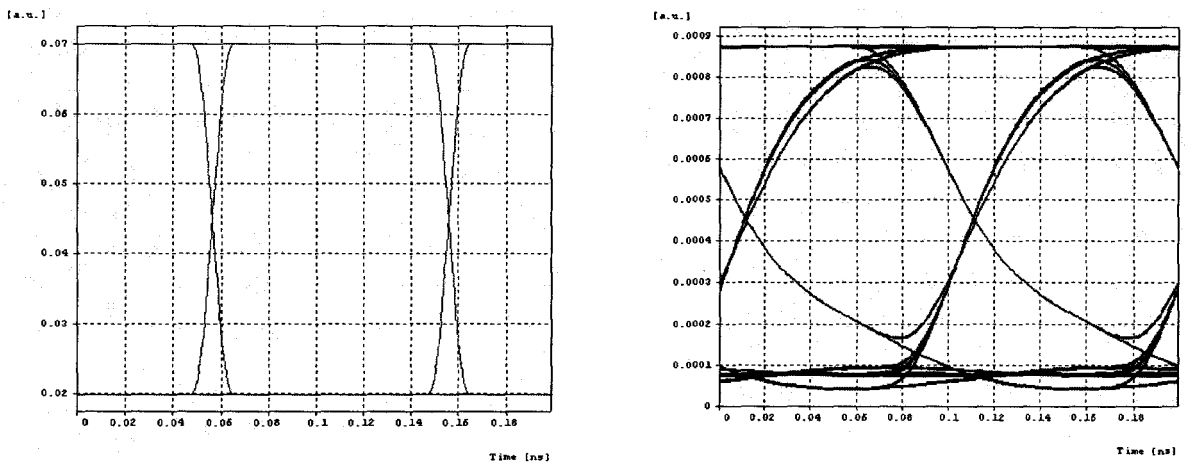


Figure 2.16 – Frequency chirping of a directly modulated laser with a 10 Gbps NRZ PRBS



a) b)

Figure 2.17 – Simulated eye-diagrams for a 10 Gbps chirped laser transmitting over 10 km of SMF a) Transmitter input b) Optical receiver output

2.3 Optical Receiver

2.3.1 Photodetector

The role of an optical receiver is to convert the optical signal back into electrical form and recover the original data that was transmitted through the lightwave system. The photonic component of the receiver is a photodiode which converts light into electricity through the fundamental process of optical absorption. The optical power absorbed by a semiconductor diode with depletion width W_D can be written as [1]

$$P_{abs} = P_{rec} - P_{tr} = \left[1 - \exp\left(-\alpha_a W_D\right)\right] P_{rec} \tag{2.67}$$

P_{in} is the incident optical power, P_{tr} is the optical power that is transmitted through, P_{rec} is the received optical power, and α_a is the absorption coefficient of the semiconductor

material. The wavelength dependence of α_a means that not all semiconductor materials can be used as a photodiode in the region of low fiber loss and dispersion $(\lambda = 1.3 - 1.5\ \mu m)$. However, the lattice matched at $x = 0.53$ $In_xGa_{1-x}As$ – InP system is highly absorbent in this region, with a fundamental energy gap of 0.75 eV, corresponding to the wavelength 1.67 μm. Therefore the $In_xGa_{1-x}As$-InP material system is commonly used as the photodiode material in high-speed lightwave systems [22]. The fundamental model and structure of a semiconductor photodiode is the reverse-biased p-n junction. When a reverse biased p-n junction is vertically illuminated with a pulse of light, electron-hole pairs generated inside the depletion region (due to absorption) are quickly accelerated towards toward the contacts on both sides, yielding a corresponding pulse in the current flow. The generated current is proportional to the intensity of the applied light. Thus the relationship between the generated photocurrent I_P, and the received optical power P_{rec}, is linear, with the constant of proportionality given by the responsivity $R = \eta_P \cdot \lambda q / hc$ in units of A/W, where η is the quantum efficiency representing the fraction of electron generation to photon incidence in the semiconductor diode. Each absorbed photon is assumed to create an electron-hole pair; the internal quantum efficiency is thus

$$\eta_P = P_{abs} / P_{rec} = 1 - \exp(\alpha_a W_D) \tag{2.68}$$

There are two types of photodiodes most commonly used in high-performance optical receivers. These are the PIN and APD photodiodes.

The PIN photodiode consists of an intrinsic region sandwiched between p- and n-type layers. Under reverse bias, the depletion depth can be made sufficiently thick to absorb most of the incident photons. Thus, most photons are absorbed in a region of high electric field, so the photo-generated carriers immediately drift at high speed to the respective contacts. Increasing the absorption region width results in a better responsitivy but at the cost of reduced speed, since the speed is limited by the transit time of carriers

generated in the absorption region, $\tau_{tr} = W_D / v_c$, where v_c is the carrier velocity. Parasitics associated with the photodiode structure result in a RC time constant which also limits the maximum attainable bandwidth of the device. Therefore, high-speed p-i-n PDs typically use double-hetereostructure design, similar to the case of semiconductor lasers, where an intrinsic InGaAs material is used for the absorption region is sandwiched between the surrounding lattice-matched transparent InP n-type and p-type layers [22]. To obtain high-speed PDs with large quantum efficiencies, without sacrificing the bandwidth, several techniques have been developed [44]. In one approach, a FP cavity is formed around the PIN structure, resulting in a laser-like structure where the internal optical field is resonantly enhanced through constructive interference, yielding a high quantum efficiency even for a very thin absorption layer thickness [44]. Another approach to realize efficient high-speed photodiodes makes use of an optical waveguide into which the optical signal is edge coupled. Since absorption takes place along the length of the optical waveguide, the quantum efficiency can be very large even for an ultra-thin absorption layer. A mushroom-mesa structure can be utilized for the waveguide PD to minimize the parasitic capacitance and the internal series resistance, resulting in quantum efficiencies near 100% and bandwidths in excess of 100 GHz [26].

The APD is a modified PIN photodiode, with an added multiplication region, so that when operated at very high reverse bias, photo-generated carriers induce generation of secondary electron-hole pairs by the process of impact ionization. This process leads to internal electrical gain, and a much larger responsivity; however, since the generation process of the secondary carriers is random, APDs are generally noisier.

The photocurrent of a PIN or APD is composed of a large number of short pulses that are distributed randomly in time due to shot noise. Shot noise is inherent to the fundamental process of carrier generation in the depletion-region of a pn junction and can be approximated by Gaussian statistics with a mean-square noise current equal to

$$\overline{i_{n,PD}^2} = 2qI_P \cdot BW_n \tag{2.69}$$

where BW_n is the noise bandwidth of the linear channel [20]. The shot noise in a photodiode is therefore signal-dependent. For a PIN photodiode, and assuming a DC-balanced NRZ signal with average power $\overline{P}$ and high extinction ratio, the noise currents for zeros and ones can be found to be equal to [20]

$$\begin{aligned} \overline{i^2}_{n,PD,0} &\approx 0 \\ \overline{i^2}_{n,PD,1} &= 4qR\overline{P} \cdot BW_n \end{aligned} \tag{2.70}$$

Assuming R = 0.8 A/W, a received average optical power of 0.5 mW would therefore generate an average current of 0.8 mA and a shot-noise current of approximately 1.6 μA rms in a 10 GHz noise bandwidth, resulting in a signal-to-noise ratio of 54 dB. In practice, the noise generated by the receiver linear channel is the limiting factor in determining the minimum receiver signal-to-noise ratio for a specified BER, and it will be considered in the following section.

2.3.2 Analog Front End

The AFE of the receiver comprises a trans-impedance amplifier (TIA) and a main amplifier (MA). The TIA/MA cascade is modeled as a linear channel (LC), having an overall frequency response

$$H(f) = \frac{V_0(f)}{I_P(f)} \tag{2.71}$$

where $V_0(f)$ is the Fourier transform of the LC output, and $I_P(f)$ is the Fourier transform of the LC input from the photo-diode [20]. The frequency response of the linear channel is in units of impedance since the linear channel of trans-impedance type. The basic model is illustrated in figure 2.18.

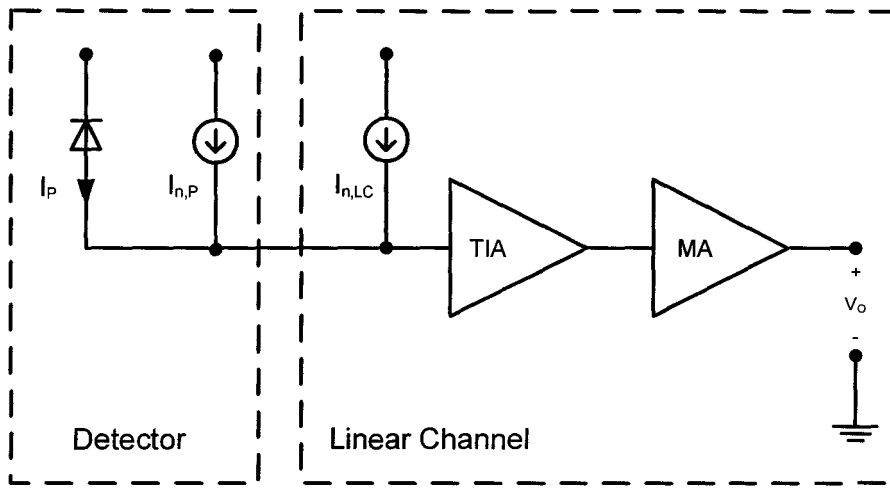


Figure 2.18 – Basic receiver model [20]

The noise characteristics of the detector and the linear channel noise are modeled by the input noise current sources $i_{n,PD}$ and $i_{n,LC}$, respectively. In practice, the noise of the receiver is determined almost completely by the noise of the linear channel $i_{n,LC}$ which is determined almost completely by the input-referred noise of the TIA, $i_{n,TIA}$ [2], [10]. The noise spectrum of this source $I^2_{n,LC}(f)$ is chosen such that after passing through the channel $H(f)$ in the absence of noise, it produces the output noise spectrum of the actual noisy channel

$$V^2_{n,LC}(f) = |H(f)|^2 \cdot I^2_{n,LC}(f). \tag{2.72}$$

The rms noise voltage at the output of the linear channel can be obtained by integrating eq. (2.72) across the bandwidth of the decision circuit (assuming that the LC and detector noises are uncorrelated, and that the LC noise is the dominant noise), according to

$$v^{rms}_{n,LC} = \sqrt{\overline{v^2_{n,LC}}} = \sqrt{\int_0^\infty |H(f)|^2 \cdot I^2_{n,LC}(f)\,df} \tag{2.73}$$

The voltage at the output of the linear channel is a superposition of the received signal and the undesired noise voltage, $v_O = v_S + v_{n,LC}^{rms}$. Occasionally the instantaneous noise voltage $v_{n,LC}(t)$ may become so large that it corrupts the received signal $v_S(t)$, causing a bit-error by the decision circuit. If additive white Gaussian noise (AWGN) is assumed, then the BER can be expressed as [2]

$$BER = \frac{1}{\sqrt{2\pi}} \int_{Q}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{2} erfc\left(\frac{Q}{\sqrt{2}}\right) \tag{2.74}$$

where

$$Q = \frac{v_S}{2 \cdot v_{n,LC}^{rms}} \tag{2.75}$$

The Q parameter thus measures the ratio between signal and noise at the input of the decision circuit. A BER less than 10^{-12} is normally specified, requiring a Q value of approximately 7. The electrical receiver sensitivity i_{sens}, is defined in terms of the Q parameter as the minimum peak-to-peak current required at the input of the receiver necessary to achieve the specified BER. It is expressed as

$$i_{sens} = 2Q \cdot i_{n,LC}^{rms} \tag{2.76}$$

where

$$i_{n,LC}^{rms} = \frac{v_{n,LC}^{rms}}{H_0} \tag{2.77}$$

and H_0 is the mid-band gain of the linear channel. The optical receiver sensitivity $\overline{P}_{sens}$, is defined as the minimum optical power, averaged over time, necessary to achieve the specified BER. For a DC balanced signal with high extinction ratio, the optical receiver sensitivity for a PIN photo-diode is approximately

$$\overline{P}_{sens,PIN} \approx \frac{Q \cdot i_{n,LC}^{rms}}{R} \tag{2.78}$$

The theoretical limit for the maximum receiver sensitivity that can be obtained is known as the quantum limit [20]. It can be obtained from the observation that at least one photon must be detected for each transmitted ONE bit to have error-free reception, and that the probability of error for ZERO bits is zero for a large extinction ratio. The quantum limit for the sensitivity can be expressed as [20]

$$\overline{P}_{sens,quant} = \frac{-\ln(2 \cdot BER)}{2} \cdot \frac{hc}{\lambda} \cdot B \tag{2.79}$$

The quantum limit for the sensitivity of a 10 Gbps PIN receiver at a BER of 10^{-12} is -47.6 dBm.

Equation (2.73) demonstrates that the noise at the output of the linear channel depends on $|H(f)|^2$. Thus the sensitivity of the receiver depends on the shape of linear channel frequency response. Ideally, the NRZ data would propagate through the fiber and detector undistorted, and would appear at the linear channel as a series of rectangular current pulses, each having a duration of $T_B = 1/\mathrm{B}$. The received pulses are orthogonal in that each pulse is uncorrelated with itself shifted by any integer multiple of $T_B = 1/\mathrm{B}$. In this case, the optimum receiver consists of a matched filter with an impulse response function $h(t)$ given by $h(t) = x(T_B - t) = x(t)$ where $x(t)$ is the rectangular NRZ pulse starting at $t = 0$ and ending at $t = T_B$, followed by an ideal sampler and comparator [3].

The frequency response of the matched filter receiver is therefore equivalent to a linear filter with the low-pass characteristic given by

$$H(f)=\frac{\sin(\pi f/B)}{\pi f/B}\cdot\exp(-j\pi f/B) \tag{2.80}$$

The magnitude and phase response is shown in figure 2.19 for $T_B=1$ s. The matched filter convolves the received NRZ signal with $h(t)$ to obtain triangular pulses at the output with considerably enhanced rise and fall times. Although the rise and fall times are different, the output of the linear channel is still free from ISI when sampled at the instant of maximum eye opening by an ideal decision circuit with no timing jitter. In the presence of AWGN, the matched filter receiver maximizes the sampled SNR at the maximum eye opening thus giving the lowest attainable BER for a given received optical power [3], [7].

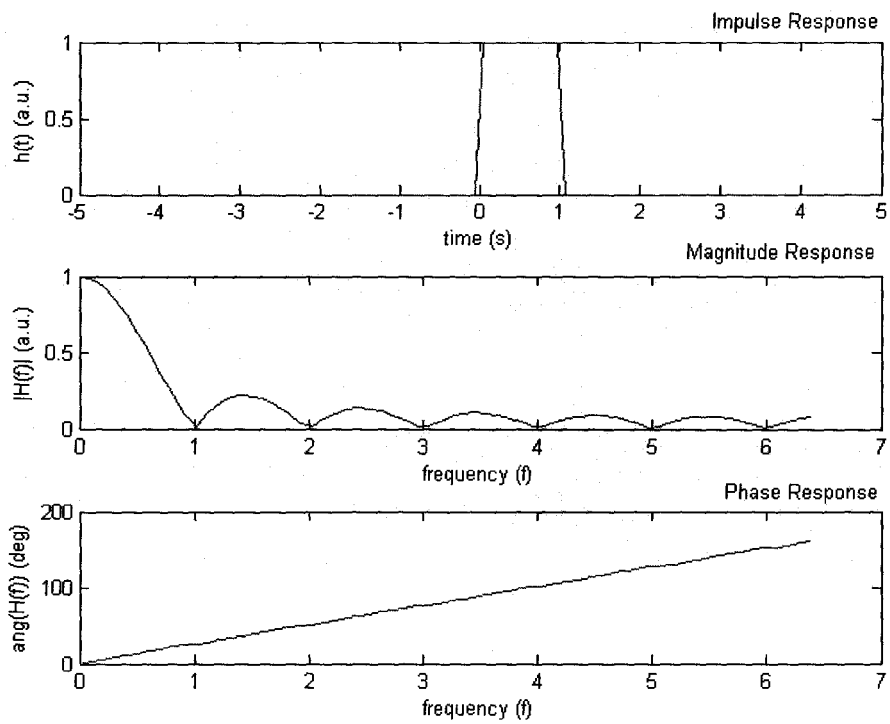


Figure 2.19 – Time domain and frequency domain characteristics of matched filter

This type of receiver has two important characteristics. The first is the absence of ISI at the time of maximum eye opening. The second is a small noise bandwidth. The noise bandwidth compares the total amount of integrated noise for linear channels with frequency responses that exhibit the same low-frequency noise but have different high-frequency transfer functions. For a noise bandwidth defined as [16]

$$BW_n = \frac{1}{H_0^2}\int_0^\infty |H(f)|^2\, df \tag{2.81}$$

the noise bandwidth of the rectangular filter is only $BW_n = \mathrm{B}/2$. For AWGN, $I_{n,LC}^2(f) = \chi_0$, where χ_0 is the spectral density of the white-noise in units of $A/\sqrt{Hz}$. Thus the input-referred rms noise current of the linear channel due to AWGN can be written as [20]

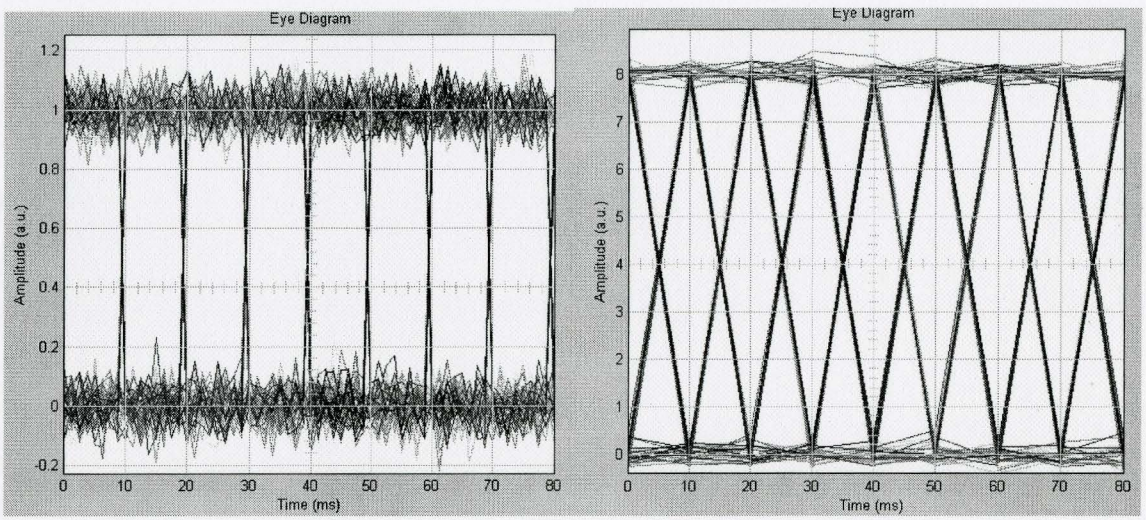
$$i_{n,LC}^{rms} = \sqrt{\overline{i_{n,LC}^2}} = \sqrt{\chi_0 \cdot BW_n} \tag{2.82}$$

Table 2.2 lists some the noise bandwidth of different linear channel frequency responses.

$H(f)$	BW_n
1-st order low pass	$1.57 \cdot f_{3dB}$
2-nd order low pass, cir. Damped $(\mathrm{Q} = 0.5)$	$1.22 \cdot f_{3dB}$
2-nd order low pass, Bessel $(\mathrm{Q} = 0.577)$	$1.15 \cdot f_{3dB}$
2-nd order low pass, Butterworth $(\mathrm{Q} = 0.707)$	$1.11 \cdot f_{3dB}$
Brick wall low pass	$1.00 \cdot f_{3dB}$
Matched filter	$0.5 \cdot \mathrm{B}$
NRZ to raised-cosine filter	$0.564 \cdot \mathrm{B}$

Table 2.2 – Noise bandwidths of different frequency responses [20]

Table 2.2 reveals that the matched filter has the best sensitivity in the presence of AWGN since it has the smallest noise-bandwidth. A matched filter receiver can be implemented using an integrate and dump circuit. Eye diagrams for such a receiver are shown in figure 2.20 before and after filtering by the linear channel for NRZ pulses with an SNR of 25 dB transmitted over a distortion-less AWGN channel. If the NRZ pulses shapes at the receiver input are distorted due to, say, fiber dispersion or laser relaxation oscillations, then the matched filter response is no longer optimum and would introduce significant ISI in the output signal. In general, a wider bandwidth for the linear channel would minimize the amount of added ISI. However, a wider bandwidth would pick up more noise. On the other hand, a receiver with a narrow bandwidth minimizes the amount of noise, but at the cost of introducing more ISI. It can be shown that a receiver bandwidth of $f_{3dB} = 0.6 \cdot B$ produces the smallest eye closure in terms of both added noise and ISI [16].



a) **b)**

Figure 2.20 – Eye diagrams of a matched filter receiver a) before filtering b) after filtering

2.3.3 Clock and Data Recovery

In fiber optic communications, a stream of binary data flows over a single fiber with no accompanying clock, but the receiver must process the data synchronously.

Therefore, the timing information (i.e. the clock) is recovered from the data by the CDR circuit at the receiver. Furthermore, the data must be re-timed such that the jitter accumulated during transmission is removed. Depicted in figure 2.21, jitter manifests itself as variation of the period of a waveform, a type of corruption that cannot be removed by amplification alone.

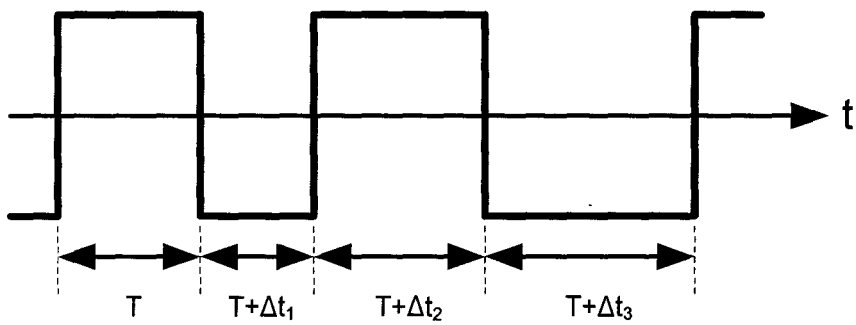


Figure 2.21 – Timing jitter [19]

The extracted clock must therefore satisfy three important conditions [19]:

1) It must have a frequency equal to the data rate.
2) It must bear a certain phase relationship with respect to the data, allowing optimum sampling of the received bits by the clock.
3) It must exhibit a small jitter as it is the principal contributor to the retimed data jitter.

Most CDR circuits employ phase locking. The generic arrangement for a CDR employing a PLL is shown below.

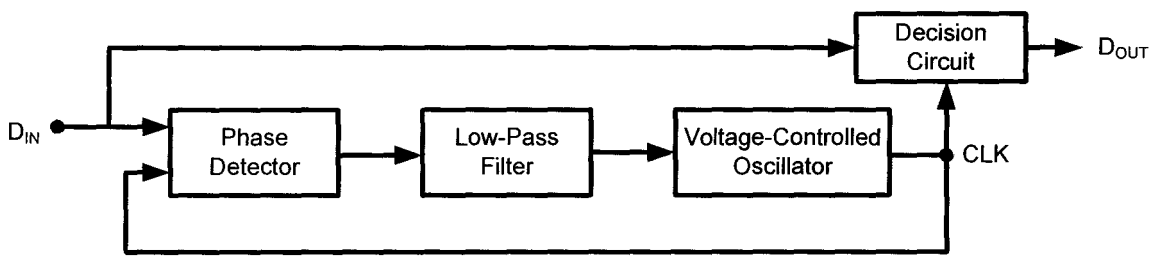


Figure 2.22 – Generic CDR circuit [19]

A PLL is essentially a feedback system that operates on the excess phase of nominally periodic signals. A simple PLL consists of a phase detector (PD), low-pass filter (LPF), and voltage-controlled oscillator (VCO). An ideal PD produces an output signal whose DC value is linearly proportional to the difference between the phases of two periodic inputs. Phase detectors generally appear in two different forms. Non-linear PDs ('bang-bang' PD) coarsely quantize the phase error, producing only a positive or negative value at their output. An example of a non-linear PD is a D flip-flop. Linear PDs, on the other hand, generate a linearly proportional output. Examples of linear PDs are XOR gates and Gilbert cells [19]. Compared to nonlinear PDs, a linear phase detector has a well-defined gain characteristic between the phase detector output and the clock/data phase offset at the input. However, because a linear PD operates with fine pulses, it is less suitable for high-speed operation. The LPF is inserted after the PD to filter out the time-varying components of the PD output signal, thus generating a proper DC control voltage for the VCO. The PD serves as an error amplifier in the feedback loop, thereby minimizing the phase difference between the clock generated by the VCO and the input data stream. The loop is considered 'locked' if the phase difference is constant in time, which results in the input and output frequencies being equal. If the loop gain (the product of proportionality constants of the PD and VCO) is large enough, then the difference between the phase of the incoming data and the phase of the VCO output falls to a small value in the steady state, providing the phase alignment.

One of the key parameters of the PLL is the loop bandwidth (i.e. the bandwidth of the LFP). This bandwidth directly affects the jitter performance of the PLL. Intrinsically, the PLL exhibits a low-pass jitter transfer characteristic between its input and output, while exhibiting a high-pass characteristic between the noise on the VCO input and the output. As a result, high-speed optical communication jitter transfer requirements (relevant to the transmitter) place an upper limit on the loop bandwidth, while jitter tolerance requirements (relevant to the receiver) sets a lower one [ ]. The jitter tolerance is the important CDR specification. It is measured by adding sinusoidal jitter at a certain

frequency to the input data. The maximum input jitter at this frequency for which the data passes through the CDR without errors is defined as the jitter tolerance.

The loop-bandwidth trades directly with the acquisition range of the PLL, i.e. the maximum frequency variation in the input or the VCO that can be accommodated. Since CDR circuits designed for multi-Gbps optical communications require a narrow loop bandwidth to meet the jitter transfer specifications, the acquisition range of a simple PLL is thereby severely limited [87]. For this reason, a means of frequency detection is necessary so as to guarantee lock in the presence of large oscillator frequency variations, as is the case for CMOS implementations, where the VCO free running frequency can vary substantially with temperature and process. Therefore, high-speed CDR circuits utilize a dual-loop architecture (e.g. figure 2.23), consisting of a frequency-acquisition loop and a phase-locked loop. A frequency detector (FD) is utilized to aid with frequency acquisition. The FD produces an output having a DC value proportional to the frequency difference of the signals at its input. If the frequency difference is large, then the PD output has a negligible DC component and the VCO is driven by the low-pass-filtered output of the FD, thereby moving the VCO output frequency closer to the input frequency. As the frequency difference drops, the DC output of the FD decreases, whereas that of the PD increases. Thus, the frequency detection loop gradually relinquishes the acquisition to the PLL, becoming inactive when the input and output frequencies are the same. Dual-loop architectures can significantly increase the acquisition range of the PLL, often removing its dependence on the loop-gain and the loop-bandwidth and achieving limits equal to those of the VCO [94].

Another technique used to obtain a wide acquisition range incorporates a PFD and a charge pump (CP) instead of the LPF of the generic PLL architecture. The CP circuit consists of two switched current sources driving a capacitor. In contrast to the LPF, in which the average value of the PFD output is obtained by depositing a charge onto a capacitor during each phase comparison and allowing the charge to decay afterwards, with a charge pump there is negligible decay of charge between phase comparison

instants. The combination of a PFD and charge pump offers an acquisition range that is only limited by the VCO output frequency range.

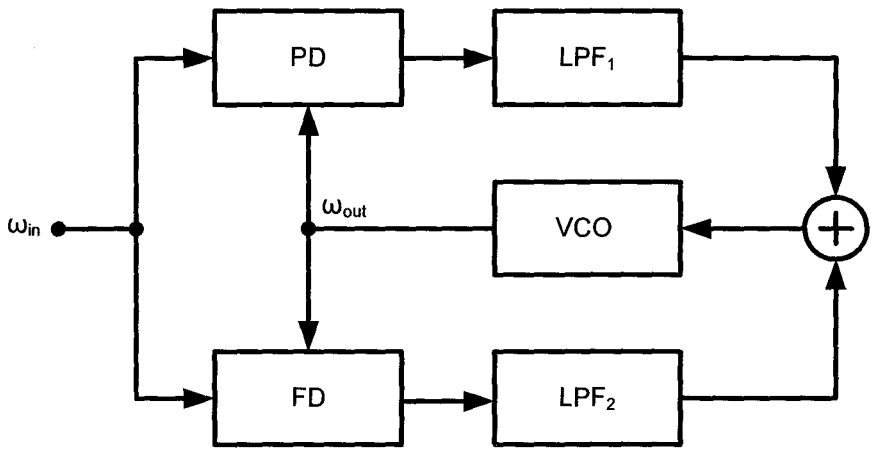


Figure 2.23 – Dual Loop PLL Architecture [19]

The VCO is a critical part of PLLs and CDRs. The following parameters of a VCO are important in the design of CDR circuits:

1) Tuning range, i.e. the range between the minimum and maximum values of the VCO frequency. In this range, the variation of the output amplitude and jitter must be minimal. The tuning range must accommodate the CDR input frequency range as well as process and temperature induced variations in the VCO frequency range.
2) Jitter and phase noise. To preserve data integrity, rigorous jitter characteristics are required by high-speed CDR circuits. The VCO is the primary factor in determining these characteristics.
3) Supply and substrate noise rejection. If integrated along with digital circuits, VCOs must be highly immune to supply and substrate noise. This can be achieved if both the signal path and the control path of a VCO is fully differential.

Two types VCOs are typically used in monolithic implementations; ring-oscillators or LC oscillators [19]. A ring oscillator consists of a cascade of a number gain stages placed in a feedback loop, the oscillation thus occurring at the frequency for which the total

phase shift is zero and the loop gain is unity. In order to vary the frequency of oscillation, the delay of each stage is varied. An LC VCO on the other hand has its center frequency mainly set by the values of the LC tank as opposed to the delay properties of the transistors. The LC tank can be formed by MOS varactors and an on-chip spiral inductor. Cross-coupled PMOS and NMOS transistor pairs realize the negative resistance to compensate the losses of the inductor and capacitor in the LC tank. In addition to a speed advantage, LC VCOs exhibit lower intrinsic phase noise because of a higher quality factor (Q) provided by the resonance load. The main drawback of LC VCOs however are their limited tuning range. Due to the low-power supply voltage and the wide frequency range requirements, conventional tuning schemes result in excessive VCO gain and thus increased sensitivity to noise in the control lines. LC VCO architectures for high-speed CDR circuits typically employ coarse/fine tuning schemes to achieve low phase noise together with a wide tuning range.

Inevitably, the choice of the CDR architecture and the design of its building blocks is primarily determined by the speed and supply voltage limitations of the technology as well as the power dissipation and jitter requirements of the system. Thus, section 3.3 will review recently published high-speed CDR circuits realized with CMOS technology to highlight these issues.

Chapter 3
CMOS OPTICAL RECEIVER DESIGN

3.1 CMOS Design Fundamentals

CMOS technology was first introduced in the mid-1960s and now it dominates the IC market. The widespread use of CMOS technology can be attributed to the following advantages of CMOS compared to its bipolar and III-V counterparts: reduced power dissipation, higher device density, improved scalability, lower fabrication costs, and capability for analog and digital circuitry on the same chip. These attributes have enabled deep-submicron CMOS technologies to become a very attractive low-cost solution for RF electronic systems.

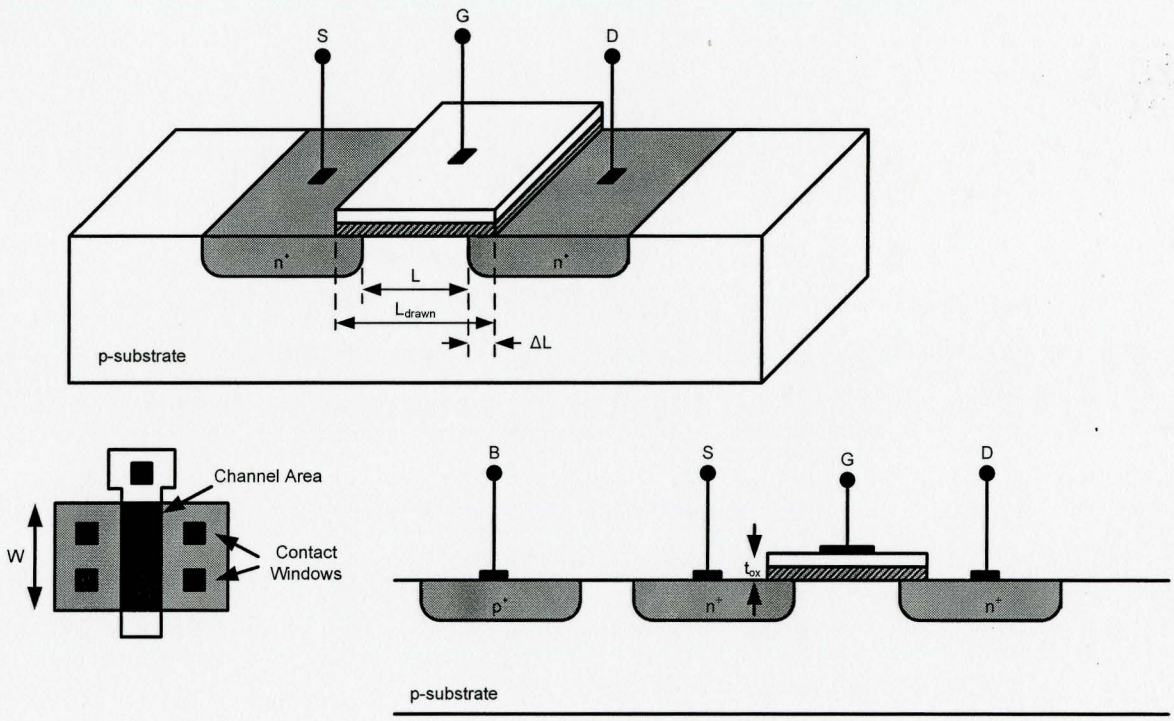


Figure 3.1 – 3-d and 2-d views of a nMOSFET [18]

The 3-dimensional cross-section view of the physical structure of a submicron nMOSFET, and it's 2-dimensional views are illustrated in figure 3.1. The dimensions *W*, L_{drawn}, and *L*, are the channel width, the drawn channel length and the effective channel length, respectively. The total overlap between the drain/source and the gate

$$\Delta L = L_{drawn} - L \tag{3.1}$$

accounts for the difference between the drawn and effective channel lengths. The overlap is due to the lateral diffusion of the drain/source implants characteristic of submicron MOSFETs [14].

The basic operation of an n-channel long-channel MOSFET can be explained briefly as follows. For a small drain-source voltage, V_{DS}, the amount of negative charge, Q_n, in the channel is controlled by the gate-source voltage, V_{GS}, according to

$$Q_n = C_{GS} \cdot (V_{GS} - V_{TH}) \tag{3.2}$$

where $C_{GS} = C_{ox}' \cdot WL$ and V_{TH} is the threshold voltage at which the channel region becomes 'inverted' (i.e. it is as much n-type as the substrate is p-type). $C_{ox}' = \varepsilon_{ox} / t_{ox}$ is the gate-oxide capacitance per unit area, where ε_{ox} and t_{ox} are the permittivity and thickness of the insulator, respectively. The amount of charge Q_n determines the current flow from the source to the drain,

$$I_D = Q_n / \tau_{tr} \tag{3.3}$$

where $\tau_{tr} = L / v_c$ is the transit time of the charge carriers (electrons) in the channel. For low electric fields in the channel, the carrier velocity, v_c, is $v_c = \mu_n E_L$, where μ_n is the electron mobility and E_L is the electric field strength along the length of the channel. The

electric field along the channel is $E_L = V_{DS} / L$. From these equations, the I/V relationship of the MOSFET in the linear region can be obtained

$$I_D \approx \mu_n C_{ox}^{'} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH}) \cdot V_{DS} \quad \text{for} \quad V_{DS} << V_{GS} - V_{TH} \tag{3.4}$$

In the linear regime the MOSFET behaves like a voltage-controlled conductance with

$$g_0 = \frac{\partial I_D}{\partial V_{DS}} \propto (V_{GS} - V_{TH}) \tag{3.5}$$

With a larger drain-source voltage, $V_{DS} > V_{GS} - V_{TH}$, the channel charge thins out near the drain, and the inversion layer charge eventually becomes negligibly small at a point $x < L$, where the channel is said to be 'pinched off'. If the total channel charge when the channel is pinched off is approximately half of that in eq. (3.2), and if the electrical field is approximately $E_L = (V_{GS} - V_{TH}) / L$, independent of V_{DS}, then the drain current for the saturated region is approximately [20]

$$I_D = \frac{\mu_n C_{ox}^{'}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH})^2 \quad \text{for} \quad V_{DS} > V_{GS} - V_{TH} \tag{3.6}$$

In this saturated regime the MOSFET behaves like a voltage-controlled current source with a transconductance of [14]

$$\begin{aligned} g_m = \frac{\partial I_D}{\partial V_{GS}} &= \mu_n C_{ox} \cdot \frac{W}{L} (V_{GS} - V_{TH}) \\ &= \sqrt{2 \mu_n C_{ox} \cdot \frac{W}{L} I_D} \\ &= \frac{2 I_D}{V_{GS} - V_{TH}} \end{aligned} \tag{3.7}$$

This simplified model neglects second-order effects such as body effect and channel-length modulation, which significantly affect the performance of sub-micron MOSFETs [18]. The body effect occurs when the substrate and source of the transistor are not at the same voltage. The threshold voltage varies with the substrate voltage according to

$$V_{TH} = V_{TH0} + \gamma\left(\sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|}\right) \tag{3.8}$$

where γ is the body-effect coefficient, V_{TH0} is the threshold at zero substrate bias, Φ_F is the Fermi voltage, and V_{SB} is the voltage between the source and bulk (substrate) terminals [18]. Body effect is sometimes undesirable in analog and digital circuit design so it can be eliminated by tying the substrate and source terminals together. Another second-order effect is channel length modulation (CLM). CLM occurs in the saturation region when the actual length of the inverted channel decreases as the voltage between the drain and source increases. This effect is accounted for with

$$I_D = \frac{1}{2}\mu_n C'_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{TH})^2 (1 + \lambda_{CLM} \cdot V_{DS}) \tag{3.9}$$

where λ_{CLM} is the channel-length modulation coefficient, representing the relative variation in channel length for a given increment in V_{DS} [14], [18]. The output conductance of the FET

$$g_0 = \frac{\partial I_D}{\partial V_{DS}} = I_D \cdot \lambda_{CLM} \tag{3.10}$$

depends on the degree of CLM. To maximize the gain of analog amplifiers, CLM can be suppressed by using a longer channel. However, a longer channel implies a slower device.

This basic model describes the DC behavior of MOSFETs with reasonable accuracy for hand-calculations. However, the parasitic resistances and capacitances associated with the device must be taken into account to predict the AC behavior [14]. The equivalent small-signal circuit model of the MOSFET in saturation (figure 3.2a) can be used to describe the AC behavior.

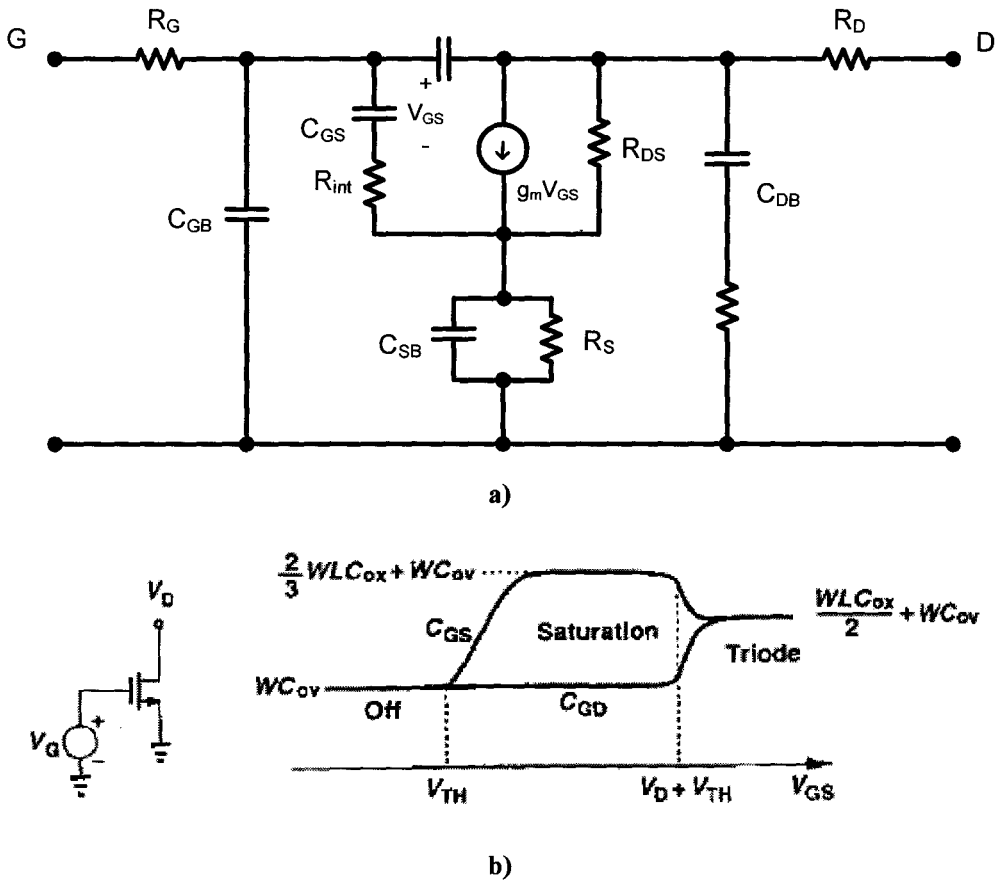


Figure 3.2 – a) Equivalent small-circuit model of a MOSFET in saturation [14] b) variation of gate-drain and gate-source capacitances versus V_{GS} [18]

A capacitance exists between every two of the four terminals of a MOSFET and they may depend on the bias-conditions of the transistor. The capacitances C_{GS} & C_{GD} primarily affect the high-frequency performance. Their approximate value can be obtained using a two-plate model that neglects fringing capacitances of the MOSFET in saturation:

$$C_{GS} = \frac{2}{3} C_{GC} + C_{OV}$$
$$C_{GD} = C_{OV} \tag{3.11}$$

where $C_{GC} = C'_{ox} \cdot WL$ represents the total capacitance between the gate and channel and $C_{OV} = C'_{ox} \cdot \Delta LW/2$ represents the overlap capacitance between the source/drain diffusions and the gate. The factor 2/3 in C_{GS} accounts for the fact that the charge distribution in the channel is not uniform when the device operates in saturation, so the capacitance is lower than it would be for a two-plate capacitor [14]. The gate resistances R_G and R_{int} are due to the contact parasitics and the relatively high intrinsic resistivity of the polysilicon gate, respectively. The source and drain resistances R_S and R_D are dominated by the resistance of the lightly-doped extensions of the source and drain diffusions, and the elements C_{DB}, C_{GB}, C_{SB} and R_{DB} model various substrate effects.

The high-frequency performance of the MOSFET can be quantified by the transition frequency f_T (i.e. the frequency at which the current gain with the output short-circuit becomes unity). Using the quadratic FET model, the transition frequency f_T is [20]

$$f_T = \frac{1}{2\pi} \cdot \frac{g_m}{C_{GS} + C_{GD}} \approx \frac{3}{4\pi} \cdot \frac{\mu_n}{L^2} \cdot (V_{GS} - V_{TH}) \tag{3.12}$$

revealing that speed of the transistor is primarily determined by the mobility of the carriers in the channel, the channel length and the bias conditions. Since the electron mobility is higher than the hole mobility, typically only nMOS transistors are used in RF and microwave CMOS circuits. It can be shown that the transition frequency is inversely proportional to the transit-time (i.e. $f_T \propto 1/\tau$) [20]. Therefore, at high electric fields, not only does the mobility degrade, but the electron velocity saturates, and the carrier saturation velocity, $v_{c,sat}$ becomes important in determining the speed. Therefore, f_T for

submicron MOSFETs saturates with sufficiently large bias voltages. Eq. (3.12) suggests the following guideline to obtain high-speed CMOS circuits: (i) choose n-channel devices (ii) choose the smallest channel length permitted by the technology, and (iii) choose the optimum gate overdrive voltage for maximum f_T and subject to headroom limitations.

Another quantity that reveals the high-frequency performance is the frequency at which the power gain of a MOSFET is unity, $f_{\max}$, given by

$$f_{\max} \approx \sqrt{\frac{f_T}{8\pi R_G C_{GD}}} \tag{3.13}$$

which depends on the intrinsic gate resistance R_G [20]. The gate resistance can be reduced by breaking wide transistors into several smaller, parallel transistors, resulting in a multi-finger structure layout. Figure 3.3 shows an example where a MOSFET is broken into two half-sized transistors, sharing a common drain with gate and source connected together at the metal level.

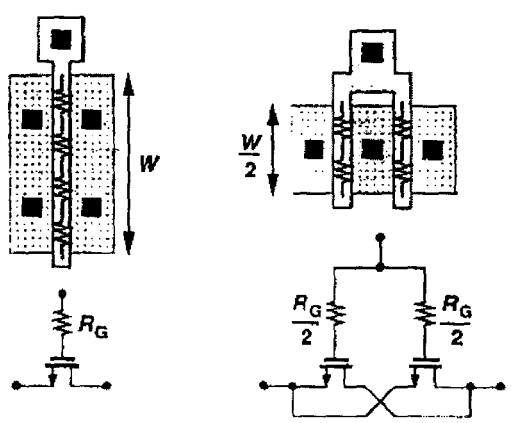


Figure 3.3 – Multi-finger transistor layout [18]

Folding the transistor in such a manner reduces the gate resistance by a factor of four in this case. With this technique, the distributed gate resistance usually can usually be made small enough such that $f_{\max}$ becomes larger than f_T (i.e. negligible R_G). However, since

these are merely extrapolated values, it is not necessarily a given that one may actually construct practical circuits operating at f_T or f_{max}. Therefore these figures of merit are viewed as a rough indication of the high-frequency performance capability. Figure 3.4 shows the measured and simulated f_T data for a multi-finger transistor in 0.18 μm technology.

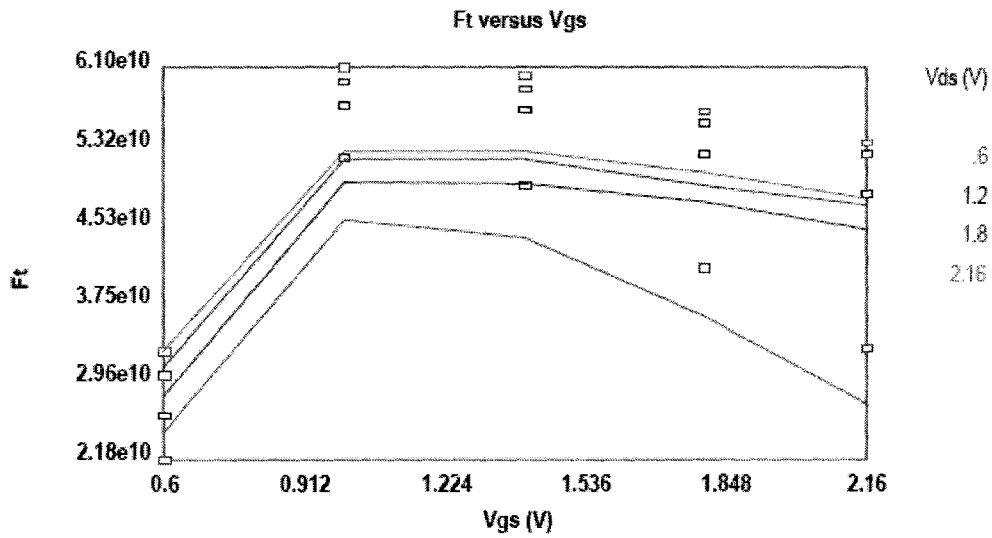


Figure 3.4 – Measured (lines) and simulated (boxes) f_T of a $16\times2.5\times0.18\,\mu m^2$ multi-finger nMOSFET [105]

3.2 TIA Design Fundamentals

The TIA is the first element of the linear channel. Its function is to convert the small photocurrent produced by the photodiode into a voltage over a wide bandwidth without introducing too much noise. In general, the input-output characteristic of any amplifier is nonlinear, which can be approximated by a polynomial over some signal range:

$$y(t)\approx\gamma_0+\gamma_1x(t)+\gamma_2x^2(t)+\ldots+\gamma_nx^n(t)\quad x_1\le x\le x_2 \tag{3.14}$$

where for the case of a TIA, $y(t)$ is the output voltage and $x(t)$ is the input current. For a sufficiently narrow range of x, only the first two terms of (3.14) are kept, providing a reasonable linear approximation as long as the bias point is disturbed negligibly (i.e. $\gamma_1 x(t) << \gamma_0$). In this case, the transimpedance of the TIA is defined as the output voltage change per input current change, $\gamma_1 = Z_T = \Delta V_0 / \Delta I_I$. When operated with a small sinusoidal signal, the TIA's response is fully described by the complex quantity

$$Z_T(f) = |Z_T(f)| \exp(j\Phi(f)) \tag{3.15}$$

where $|Z_T(f)|$ is the magnitude response of the transimpedance and $\Phi(f)$ is the phase response. The TIA bandwidth, f_{3dB}, is defined as the upper frequency at which the transimpedance $|Z_T(f)|$ drops by 3 dB below its mid-band value. The TIA is typically designed such that $f_{3dB} = 0.6 \cdot \mathrm{B}$ in order to optimize the tradeoff between noise and ISI [16].

The group delay, τ, is related to the phase response by

$$\tau_g(\omega) = -\frac{d\Phi}{d\omega} \tag{3.16}$$

The bandwidth and group-delay variation are important parameters determining the amount of ISI and jitter introduced by the TIA. In particular, peaking in the magnitude and group-delay characteristic causes overshoot and ringing in the time domain. Typically, a group delay variation, $\Delta\tau_g$, of less than $\pm 10\%$ of the bit period (i.e. ± 0.1 of the unit interval (UI)) over the specified bandwidth is required to limit the generation of data-depended jitter [20]. Similarly, the ripple in the magnitude response should be limited to less than 1 dB to avoid adding significant ISI. High performance TIA circuits are therefore required to have an overall frequency response which remains maximally flat within the desired pass-band.

Other important aspects of amplifier performance besides the gain, bandwidth and group delay are the noise, power dissipation, linearity, dynamic range and input-output impedance. The dynamic range describes the minimum and maximum input signal for which the TIA maintains a specified level performance (i.e. a sufficiently low BER). The lower end of the dynamic range is given by the TIA's sensitivity, while the upper end of the dynamic range is defined as the input overload current - the current at which the linear approximation of eq. (3.14) is no longer valid (i.e. the point at which the TIA begins to produce pulse-width distortion and jitter due to the nonlinearity, causing the BER to rise above the specified value). The input overload current of the TIA can be specified in terms of the maximum permissible input peak-to-peak voltage swing, $v_{I,ovl}$,

$$i_{ovl} = \frac{v_{I,ovl}}{Z_T} \tag{3.17}$$

The input voltage swing $v_{I,ovl}$ can be limited by a number of mechanisms. For example, a large input voltage swing can cause an insufficient voltage drop across a current-source transistor in the feedback amplifier, leading to a reduced bias current and a slow response and lower gain.

The input and output impedances determine how the amplifier interacts with preceding and subsequent stages. In this work, the preceding and subsequent stages of the TIA amplifier are assumed to be the 50-Ω signal source and measuring equipment, respectively. In practice, the TIA input impedance should be matched to the photodiode impedance, and the output impedance minimized for the next voltage gain stage.

3.2.1 Simple TIA

Almost all of the performance parameters of a TIA are coupled to each other, making the design of high-performance TIA circuits a multi-dimensional optimization problem. To gain an intuitive feeling for the various tradeoffs and fundamental

limitations, a simple TIA realization is considered first [2]. A simple realization of a TIA is a single resistor, which can be used to convert a photocurrent into a voltage. Shown in figure 3.5 along with the equivalent small-signal circuit, is the schematic of a simple TIA. The transimpedance is

$$Z_T(f) = \frac{R_L}{1 + j(2\pi f \cdot C_{PD} R_L)} \tag{3.18}$$

where $f_{3dB} = (2\pi \cdot R_L C_{PD})^{-1}$. Equations (2.73) & (2.77) can be used to obtain the total input-referred noise mean-square current

$$\overline{i_{n,TIA}^2} = \frac{kT}{R_L^2 C_{PD}} = 4\pi^2 kTC_{PD} f_{3dB}^2 \tag{3.19}$$

where k is the Boltzmann constant and T is the absolute temperature, revealing that a severe tradeoff between gain, noise and bandwidth exists for this circuit [16].

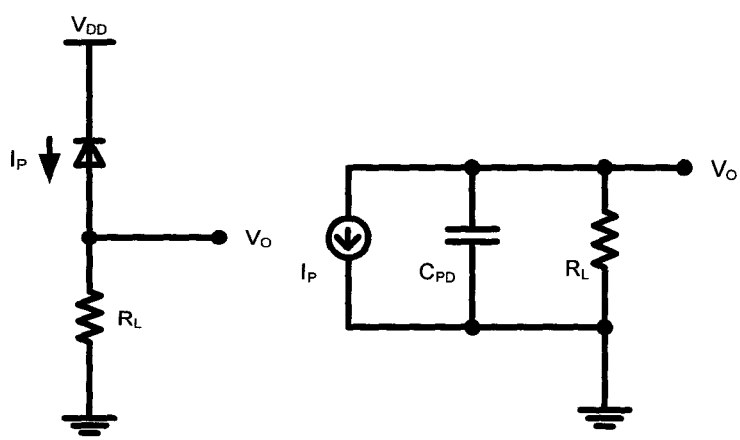


Figure 3.5 – Conversion of a photodiode current to voltage by a resistor and its equivalent small-signal circuit [2], [16]

The bandwidth-squared must be reduced so as to reduce the total integrated noise for the highest sensitivity. On the other hand, narrowing the bandwidth closes the eye both

vertically and horizontally (i.e. introduces ISI & increases the BER at high bit-rates). The maximum input overload current for this circuit is

$$i_{ovl} \leq \frac{V_{DD} - V_{RB}}{R_L} \tag{3.20}$$

where $V_{DD} - V_{RB}$ is the maximum voltage swing at the output for a given photodetector required reverse bias V_{RB}. The dynamic range therefore also trades with the gain, making high-performance operation at low supply voltages increasingly difficult. These constraints suggest that different topologies are needed for better performance.

3.2.2 Shunt-Feedback TIA

A shunt-feedback TIA helps to relax the conflicting requirements of high bandwidth, low-noise, and high gain. In this approach, the photodiode is connected to the input of an inverting voltage amplifier, which has a feedback resistor R_F leading from its output to the input. This 'shunt-shunt' feedback network senses the voltage at the output and returns a proportional current to the input [18].

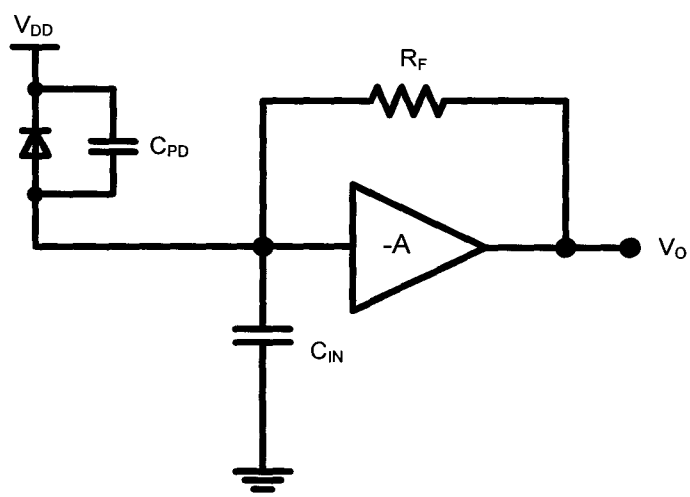


Figure 3.6 – Basic shunt-feedback transimpedance amplifier [20]

Shunt-shunt feedback is chosen because it reduces both the input and output impedance of the open loop amplifier, and yields a larger bandwidth than that which can be obtained for the same value of transimpedance that was used in a simple TIA configuration. The capacitance of the photodiode C_{PD} and input capacitance of the amplifier C_{IN} are lumped into a single capacitance $C_T = C_{IN} + C_{PD}$. ssuming an feedback amplifier with an ideal gain $-A$, the closed-loop transimpedance

$$Z_T(f) = \frac{-R_F\left(\frac{A}{A+1}\right)}{1+j\left(\frac{2\pi f \cdot C_T R_F}{A+1}\right)} \tag{3.21}$$

shows that the 3dB bandwidth, $f_{3dB} = (A+1)/(2\pi \cdot C_T R_L)$, of a shunt-feedback TIA is a factor of $(A+1)$ larger than that of a simple TIA for the same resistance R_F and total capacitance C_T. This bandwidth improvement can be understood in terms of the circuit's closed-loop input impedance

$$Z_{IN}(f) = \frac{\frac{R_F}{A+1}}{1+j\left(\frac{2\pi f \cdot C_T R_F}{A+1}\right)} \tag{3.22}$$

The feedback action makes the input impedance a factor of $(A+1)$ times smaller than R_F, thus increasing the pole frequency by the same factor compared to the situation without feedback. Since the input voltage swing can be $(A+1)$ times smaller than that of a simple TIA for a given input signal current, the low input impedance also improves the input overload current by the same factor. However, the assumption of infinite bandwidth for the feedback amplifier yields an infinite output noise power [16]. Therefore, noise calculations require a more realistic model for the feedback amplifier.

A more realistic model describes the feedback amplifier with the single-pole frequency response

$$A(f)=\frac{A_0}{1+j(2\pi f/\omega_0)} \tag{3.23}$$

leading to a second-order closed-loop response for the transimpedance

$$Z_T(f)=-\frac{R_0{\omega_0}^2}{(j2\pi f)^2+2\zeta\omega_0(j2\pi f)+\omega_n^2} \tag{3.24}$$

where $R_T=\left(\frac{A_0}{A_0+1}\right)R_F$ is the closed loop transimpedance, $\omega_n^2=(A_0+1)\omega_0/(R_FC_T)$ is the natural self-resonance frequency squared, and $\zeta=\frac{1}{2}\frac{R_FC_T\omega_0+1}{\sqrt{(A_0+1)\omega_0R_FC_T}}$ is the damping factor [16]. A maximally-flat response occurs when the damping factor is $\zeta=\sqrt{2}/2$. The bandwidth is in this case is equal to

$$f_{3dB}=\frac{1}{2\pi}\frac{\sqrt{2}A_0}{R_FC_T} \tag{3.25}$$

a factor of $\sqrt{2}\cdot A_0$ larger than that of the simple TIA. The second-order response remains maximally flat for $\zeta\geq\sqrt{2}/2$, meaning that the 3dB bandwidth of the core amplifier must satisfy

$$\omega_0>\frac{2A_0}{R_FC_T} \tag{3.26}$$

This condition implies that the bandwidth of this second-order system is bounded by $f_{3dB} \leq \omega_n / 2\pi$, thus limiting the maximum transimpedance of a shunt-feedback TIA with a maximally-flat response to

$$Z_T \leq \frac{A_0 \cdot \omega_0}{2\pi \cdot C_T \cdot f_{3dB}^2} \tag{3.27}$$

The figure-of-merit (FoM) of a TIA is given by the transimpedance-bandwidth product. Equation (3.27) reveals that the figure-of-merit of a shunt-feedback TIA is therefore,

$$FoM = \frac{A \cdot \omega_0}{2\pi \cdot C_T \cdot f_{3dB}} \tag{3.28}$$

The gain-bandwidth product, $A_0 \cdot \omega_0$, of the feedback amplifier (assuming a first-order frequency response) is roughly proportional to the transition frequency of the technology. This shows that, for a shunt-feedback topology, doubling the bit rate for a given technology causes the transimpedance to degrade by a factor of four. On the other hand, maintaining the same transimpedance at the same bit-rate requires a technology that is twice as fast (double the f_T) and with half the C_T. These considerations suggests that a shunt-feedback topology may not be suitable for applications where f_T is not considerably larger (i.e. a factor of at least 10) than the TIA bandwidth.

The noise behavior of a shunt-feedback TIA depends heavily on the actual implementation of the feedback amplifier [2], [10]. General results for the noise analysis can be obtained by considering the TIA of figure 3.7 with the feedback amplifier implemented as a FET common-source (CS) stage [20]. The noise characteristics of the TIA are conveniently described by an equivalent input noise current spectral density, rather than noise-figure (NF) [2]. The most significant noise current sources are the thermal noise of the feedback resistor, $i_{n,res}$, and the channel noise, $i_{n,D}$, of the FET.

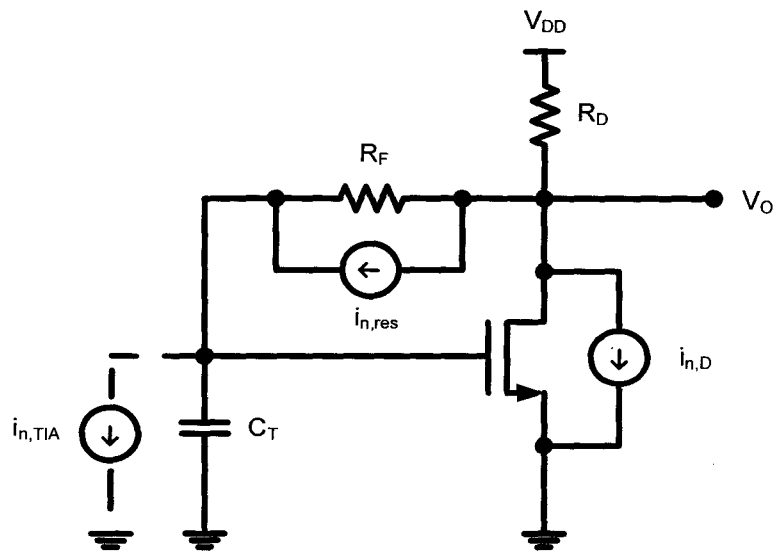


Figure 3.7 – Significant noise sources in a TIA with a FET front-end [25]

The equivalent input-referred noise current source, $i_{n,TIA}$, has the following power spectral density

$$I_{n,TIA}^2(f) = I_{n,res}^2(f) + I_{n,ch}^2(f) \tag{3.29}$$

The thermal noise of the feedback resistor

$$I_{n,res}^2(f) = \frac{4kT}{R_F} \tag{3.30}$$

is white (frequency independent), as is the channel noise of the FET input stage,

$$I_{n,D}^2(f) = 4kT \cdot \Gamma g_m \tag{3.31}$$

where Γ is the channel-noise factor. For sub-micron MOSFETs Γ lies in the range $\Gamma = 0.7 - 3.0$ [18]. This noise source is not located directly at the input of the TIA,

therefore it must be transformed to obtain its contribution to the input-referred TIA noise [2]. The implicit transfer function from the input current to the drain current has a low-pass characteristic; therefore, the inverse function, which refers the drain current back to the input, has a high-pass characteristic given by

$$H_D(f) = \frac{1 + j(2\pi f) R_F C_T}{g_m R_F} \tag{3.32}$$

where $C_T = C_{GS} + C_{GD}(g_m R_D + 1)$ is the total parasitic capacitance at the input [20]. In this case, the gate-drain (Miller) capacitance is multiplied by the gain of the common-source stage, contributing to high parasitic capacitance at the input node [18]. The input-referred noise current power spectrum of the TIA noise is therefore

$$\begin{aligned} I_{n,TIA}^2(f) &= \frac{1 + (2\pi f \cdot R_F C_T)^2}{(g_m R_F)^2} \cdot 4kT\Gamma g_m \\ &= 4kT\Gamma \cdot \frac{1}{g_m R_F^2} + 4kT\Gamma \cdot \frac{(2\pi C_T)^2}{g_m} \cdot f^2 \end{aligned} \tag{3.33}$$

This spectrum is non-white, rising at the corner frequency $1/(2\pi \cdot R_F C_T)$. Since this frequency is lower than the 3dB bandwidth of the shunt-feedback TIA (eq. (3.25)), the output-referred noise spectrum exhibits a 'hump', as illustrated in figure 3.8.

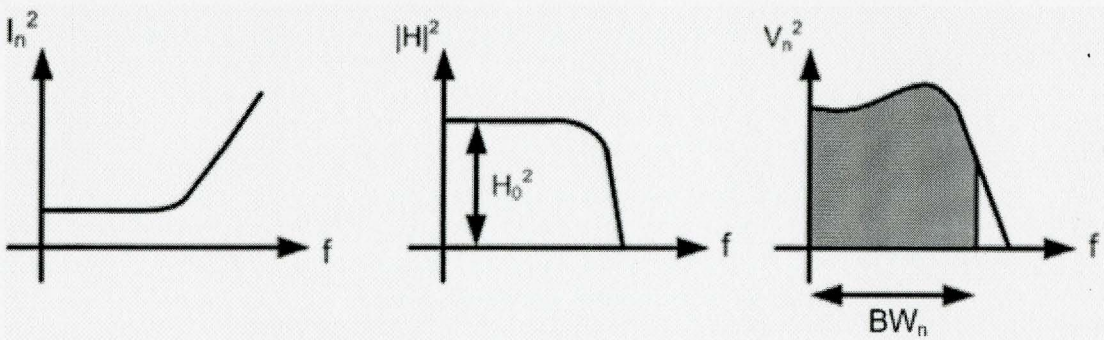


Figure 3.8 – Calculation of the total output-referred noise [20]

The effect of the f^2 noise term on the noise of the TIA can be obtained by defining the second-order noise bandwidth [20]

$$BW_{n2}^3 = \frac{3}{H_0^2}\int_0^\infty |Z_T(f)|^2 \cdot f^2 df \tag{3.34}$$

which is used to yields the total input-referred TIA rms noise current using

$$i_{n,TIA}^{rms} = \sqrt{\overline{i_{n,TIA}^2}} = \sqrt{\chi_0 \cdot BW_n + \chi_2/3 \cdot BW_{n2}^3} \tag{3.35}$$

with χ_0 representing the white-noise (frequency independent) terms and χ_2 representing the coefficient of the f^2 noise terms [20]. A figures-of-merit used to assess the noise performance, (i.e. sensitivity) of a shunt-feedback TIA is given by [10]

$$FoM = \frac{g_m}{C_T^2} \tag{3.36}$$

revealing that the input FET is required to realize a relatively high transconductance while maintaining a small parasitic capacitance value, so as to limit the high-frequency noise contributions to the overall input-referred noise current.

3.2.3 Common-gate TIA

The improved bandwidth and dynamic range are the main reasons for the popularity of the shunt-feedback TIA. However, equations (3.28) & (3.36) reveal that the performance of a shunt-feedback TIA is sensitive to the total parasitic capacitance, C_T, at the input node. Therefore, various techniques have been proposed to isolate the large input parasitic capacitance of the photodiode from the input node of the shunt-feedback

TIA, in order to minimize its contributions to the amplifier performance [16]. Also, the need to support a large input parasitic capacitance arises because low-cost photodiodes for CMOS photoreceivers are external to the chip with correspondingly large bond-pad and parasitic capacitances.

A current buffer in the form of a common-gate stage between the photodiode input and the shunt-feedback TIA (fig. 3.9) can effectively isolate these parasitic capacitances from determining the amplifiers performance [45], [46]. Neglecting body-effect and channel length modulation, the input impedance of the common–gate (CG) stage is $R_{IN} = 1/g_m$. If this impedance is sufficiently low, then the dominant pole frequency is no longer located at the input node, but rather determined at node x of figure 3.9, which depends on only the input capacitance of the shunt-feedback amplifier C_{IN}, and the gate-drain and drain-bulk capacitance of the common-gate transistor.

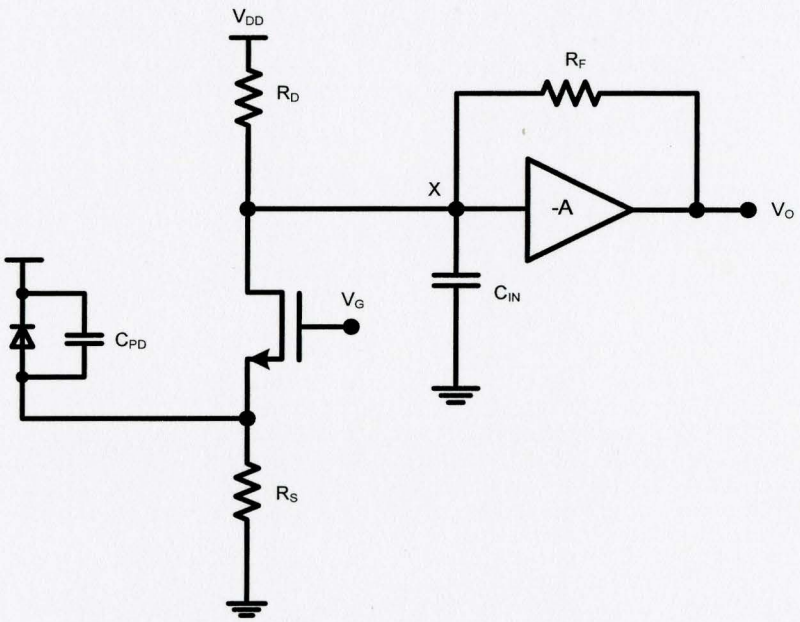


Figure 3.9 – Common-gate input stage

Since the photodiode determines only a non-dominant pole, it is thus effectively isolated from the bandwidth and sensitivity determination of the amplifier. If the feedback

amplifier is implemented with a CS stage, the 3dB bandwidth of a shunt-feedback TIA with CG buffer is approximately [45], [49]

$$f_{3dB} \approx \frac{\left(1+g_{m_2}R_{D_2}\right)}{2\pi \cdot R_F\left[\left(C_{GD_1}+C_{GS_2}\right)+\left(1+g_{m_2}R_{D_2}\right)C_{GD_2}\right]} \tag{3.37}$$

where $C_{GS_2}, C_{GD_2}, g_{m_2}R_{D_2}$ are the gate-source capacitance, gate-drain capacitances, and voltage gain of a CS stage feedback amplifier, respectively. The bandwidth is independent of the photodiode capacitance, and the shunt-feedback effectively removes the Miller effect of the CS gain stage.

The introduction of the CG input buffer introduces three new noise sources, (the thermal noise of the source resistor, the thermal noise of the drain resistor, and the thermal channel noise of the common-gate transistor), that are located directly at the input of the TIA and thus directly impact the input-referred noise current. In practice, the new noise contributions degrade the sensitivity, therefore requiring careful noise analysis. The input referred noise current spectrum of a CG TIA [46]

$$I_{n,TIA}^2(f) \approx \frac{4kT}{R_S}+\frac{4kT}{R_F}+\frac{4kT\left(1/R_F\right)}{g_m{}^2}\left[\frac{1}{R_S{}^2}+\left(2\pi f\right)^2\left(C_{PD}+C_{GS}+C_{SB}\right)^2\right] \tag{3.38}$$

reveals that $g_m{}^2/\left(C_{PD}+C_{GS}+C_{SB}\right)^2$ should be made as large possible to optimize the noise by reducing the high-frequency contributions. However, the junction capacitances in sub-micron CMOS processes are comparable to the gate capacitances, significantly influencing both noise behavior and bandwidth. Thus, optimal noise behavior in this case requires that the saturation-mode gate capacitance of the CS stage must equal the saturation-mode drain capacitance of the CG stage (i.e. $C_{GS_2}+C_{GD_2}=C_{GD_1}+C_{DB_1}$). In addition, the saturation-mode input capacitances of the CG stage ($C_{GS_1}+C_{SB_1}$) must be three or four times as big as $C_{GD_1}+C_{DB_1}$ [47]. The transconductance of the CG stage only

needs to be large enough to ensure that the input pole is non-dominant. However, the transconductance of a submicron MOSFET cannot be made considerably large without increasing the width, resulting in large capacitances, increased noise and reduced stability, or increasing the bias current, resulting in greater power consumption and reduced signal swing. Therefore the CG input configuration cannot yet effectively isolate the large photodiode capacitance from the bandwidth determination without affecting other important circuit parameters, mainly because of the relatively small transconductance of the input transistor. By modifying the conventional CG input stage to regulated cascade (RGC) topology containing negative feedback, a very small input impedance can be obtained to relax the gain-bandwidth tradeoff at the input node. [48] – [53].

3.2.4 Regulated-Cascode (RGC) TIA

The RGC input stage in figure 3.10 is well suited for broad-band TIA design by its very low input impedance. An RGC input stage consists of a CG amplifier, with the gate connected to the output of an inverting feedback amplifier which receives its input signal from the source of M1. The low-frequency input admittance can be derived as follows, where $A = -g_{m_A} R_A$ is the low-frequency gain of the feedback amplifier:

$$V_{GS} = -(A+1) \cdot V_X$$

$$V_Y = -\left(\frac{g_m V_X (A+1) - V_X g_O}{\frac{1}{R_D} + g_O} \right)$$

$$I_X = \frac{V_X}{R_S} + \frac{V_Y}{R_D} = \frac{V_X}{R_S} + \frac{1}{R_D} \left(\frac{V_X g_m (A+1) + g_O}{\frac{1}{R_D} + g_O} \right) \cdot \left(\frac{R_D}{R_D} \right)$$

$$Y_{IN} = \frac{I_X}{V_X} = \frac{1}{R_S} + \frac{(A+1) g_m + g_O}{g_O R_D + 1} \tag{3.39}$$

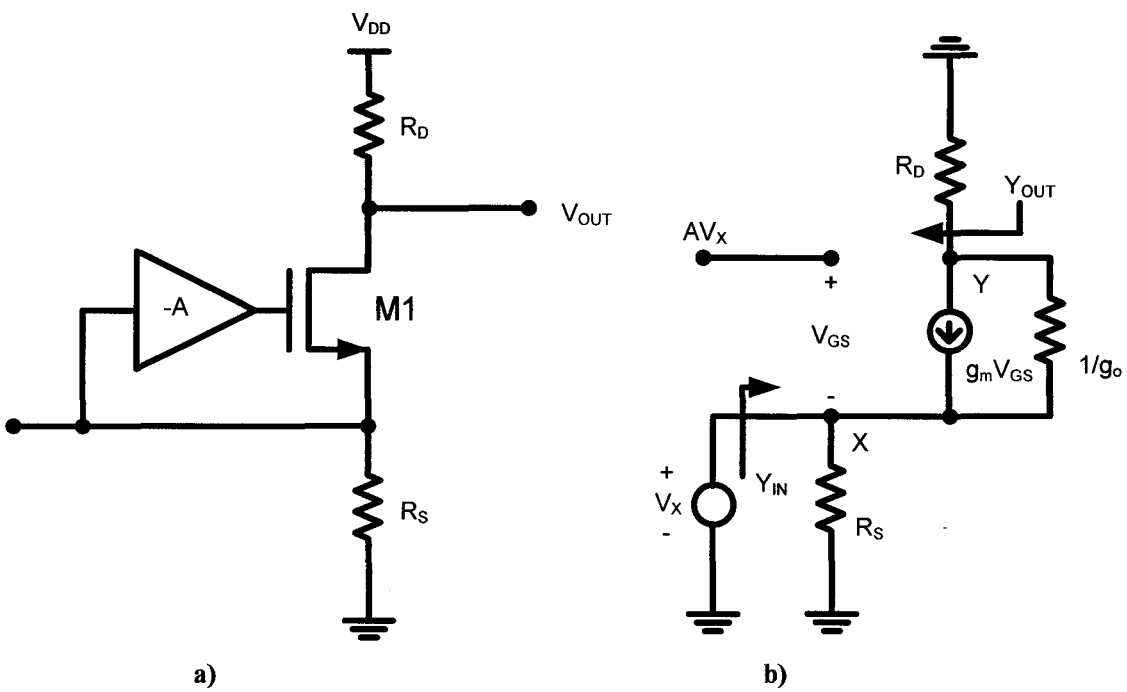


Figure 3.10 – a) Ideal RGC front-end circuit b) Equivalent small-signal model

The last equation simplifies to

$$Y_{IN} \approx \frac{1}{R_S} + g_m(A+1) \tag{3.40}$$

under the conditions

$$R_D << 1/g_O \,\&\, (A+1)\cdot g_m / g_O >> 1 \tag{3.41}$$

The output admittance can similarly be obtained

$$Y_O = \frac{1}{R_D} + \frac{g_O}{\left[(A+1)g_m + g_O\right]R_S + 1} \tag{3.42}$$

simplifying to

$$Y_O \approx \frac{1}{R_D} + \frac{g_O}{(A+1)g_m R_S} \tag{3.43}$$

for

$$(A+1)\cdot g_m / g_O >> 1 \quad \& \quad (A+1)g_m R_S >> 1 \tag{3.44}$$

The analysis shows that the local feedback mechanism of the regulated cascode reduces the input impedance of the common-gate stage by the factor $(A+1)$, and increases the output impedance by the same factor. Therefore, the RGC input stage behaves qualitatively as a CG input stage with an enhanced transconductance $G_m = (1+A)g_m$, under the conditions given by (3.41) & (3.44). Importantly, the RGC input stage enables a factor of $(1+A)$ times better isolation of the input parasitic capacitance from the bandwidth determination. From the equivalent high-frequency small-signal model, shown below, one can obtain the high-frequency input-impedance, given approximately by [54]

$$Z_{IN} \approx \frac{\left(1/R_2 + j(2\pi f)C_{GS_1}\right)}{\left(g_{m_2} + 1/R_2\right)\left(g_{m_1} + j(2\pi f)C_{GS_1}\right) + \left(1/R_2 + j(2\pi f)C_{GS_1}\right)\left(1/R_S + j(2\pi f)C_{GS_2}\right)} \tag{3.45}$$

which reduces to the inverse of (3.40) at low frequencies.

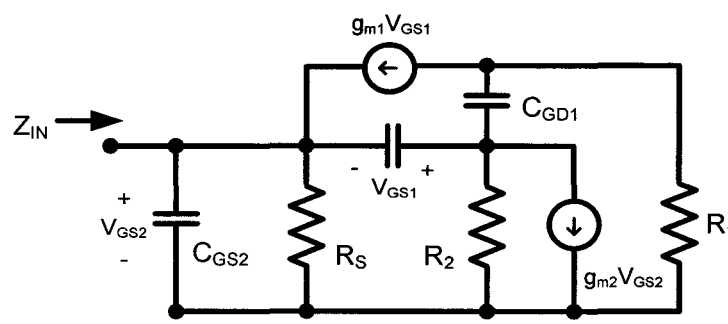


Figure 3.11 – RGC high-frequency small-signal model

The enhanced input transconductance reduces the high-frequency noise contribution relating with the large input parasitic capacitance as well. The approximate equivalent noise current spectral density of the RGC [54]

$$I_{n,TIA}^2(f) \approx \frac{4kT}{R_S} + \frac{4kT}{R_F} + \frac{(4kT/R_2)}{\left(g_{m_2} + 1/R_2\right)^2}\left[\frac{1}{R_S^2} + (2\pi f)^2\left(C_{PD} + C_{GS_2} + C_{SB_1}\right)^2\right] + \frac{4kT(2\pi f)^2\left(C_{GS_1} + C_{GD_1}\right)^2}{g_{m_1}{}^2}\left(\frac{1}{R_F}\right) \tag{3.46}$$

reveals that the dominant high-frequency noise is divided by $\left(g_{m_2} + 1/R_2\right)^2$ of the local feedback, rather than the input transconductance g_{m_1}. However, it should be noted that a zero appears in the frequency response due to the local feedback stage, which can cause peaking at the frequency of [48]

$$f_{peak} = \frac{1}{2\pi \cdot R_2\left(C_{GS_1} + C_{GS_2}\right)} \tag{3.47}$$

In order to shift the peaking to a frequency much higher than the 3dB frequency, the gate-width of M1 and the resistance R_2 should be minimized according to eq. (3.47). The minimization of the gate-width of M1 is allowable because the transconductance g_{m_1} need not be so large due to the local feedback effect. Also the resistance R_2 can be minimized by increasing the drain bias current of M_2 for the same voltage gain of the local feedback stage. If the voltage gain, $\left(1 + g_{m_2}R_2\right)$, of the local feedback is large enough, then the RGC input stage may be considered a single-pole system with a 3dB bandwidth given by

$$f_{3dB} \approx \frac{1}{2\pi R_1 C_{D_1}} \tag{3.48}$$

where C_{D_1} represents the total drain capacitance of M1.

3.3 Literature Review

3.3.1 TIA Circuits

Many different TIA circuit designs have been developed over the last 20 years in a wide range of IC technologies. However, the demand for high volume and wide deployment of optical communications makes silicon based integrated circuits the most economical solution. CMOS appears to be the best candidate for fully integrated TIA design due to its cost, integration and manufacturability advantages and providing reasonable speed and noise performance at the same time. Due to the inferior parasitic and noise characteristics of CMOS technology, many circuit techniques have thus been developed in CMOS TIA design to achieve comparable performances to the III/V or SiGe counterparts. Therefore, this section reviews the representative transistor-level CMOS TIA circuits that have been reported in the literature so far, and summarizes their performance.

Classical designs of CMOS TIA circuits generally employ common-source input stages. Several types of classical CMOS TIA designs are illustrated below. The first type (fig. 3.12a) was implemented in CMOS about 20 years ago [56]; at that time this circuit was designed primarily with n-p-n bipolar transistors [57] - [60]. A source follower decouples the drain of the input transistor from the output to achieve a low impedance output node for this TIA. However, one of the main disadvantage of this circuit is the large contribution of the Miller capacitance, $C_{GD}(A+1)$, to the overall capacitance at the input node, $C_{IN} = C_{GS1} + C_{GD1}(A+1)$. The potentially large Miller capacitance can be

suppressed by connecting the drain of the input transistor M1 to a low-impedance node, effectively reducing the gain that multiplies the gate-drain capacitance. This can be achieved by stacking a cascode transistor M2 on top of the input transistor, as in figure 48b [47], [55], [61]. The cascode transistor acts as a current buffer, reducing the Miller effect. The impedance looking into the drain of M2 is approximately $1/g_{m1}$; thus, the voltage gain from the input to drain of M1 reduces from $g_{m1}R_{D1}$ to g_{m1}/g_{m2}, and is

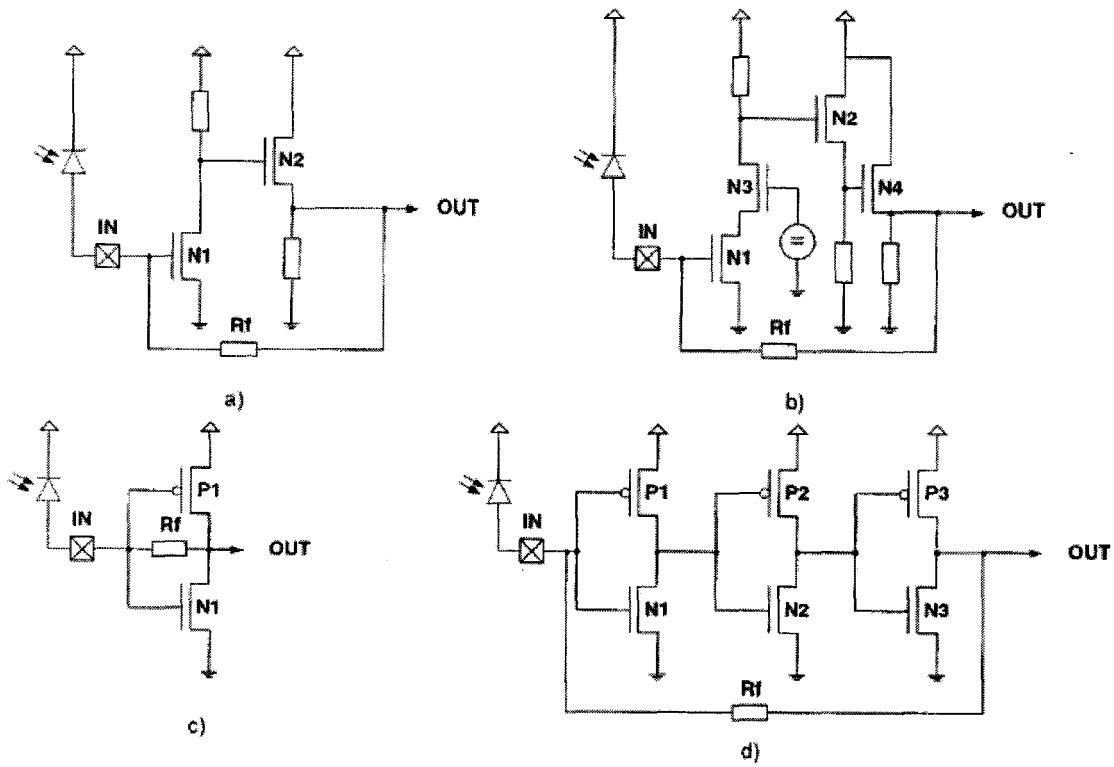


Figure 3.12 – CMOS shunt-feedback TIA circuits [55]

approximately unity. As a result the input capacitance reduces to $C_{IN} = C_{GS1} + 2C_{GD1}$. Cascoding improves the isolation from the stage output back to the input (i.e. $S_{12} << S_{21}$) resulting in a stable amplifier, as well as enhancing the overall gain by increasing the stage's output impedance, with a negligible noise contribution [47]. However, the cascode transistor introduces an additional high-frequency pole, which may offset some of the

bandwidth gained with this technique, as well as reducing the voltage headroom, thus limiting the dynamic range.

Topologies for low-voltage operation are illustrated in figures 3.12c & 3.12d. A very high transimpedance gain can be obtained with the use of CMOS inverters and multiple gain stages. However, large capacitances associated with the pMOS devices cause the amplifier bandwidth to be too small. In addition, using multiple amplifier stages makes stability and phase margin more difficult to achieve.

The high open-loop input impedance of a common-source amplifier places the dominant pole of the TIA at the input node, making the frequency response sensitive to the input parasitic capacitance, comprising the ESD, bond-pad, and photodiode parasitic capacitances. If the input were connected to a low-impedance node, such as that of a common-gate input stage, then these parasitic capacitances can be virtually insignificant in determining the bandwidth of the amplifier. For these reasons, CMOS TIA designs typically employ a common-gate input stage [45] -[47], [62]. Furthermore, a common-gate input stage can obtain high-gain and high-bandwidth simultaneously, however, at the expense of more noise sources located directly at the TIA input which can affect the receiver sensitivity.

The common-gate input stage TIA reported in [62] achieves a 3dB bandwidth of 3.5 GHz with a transimpedance gain of 60 dBΩ and input-referred noise current spectral density below 20 $pA/\sqrt{Hz}$ while consuming 5 mA from a 3.3 V supply in a low-cost 0.5 μm CMOS technology. The topology consists of three main stages: 1) an input common-gate transimpedance stage, 2) a voltage gain stage comprising a source-follower, a common-gate stage, and a cascode and 3) a low-impedance source-follower output stage. Negative shunt-shunt feedback is applied between nodes (e) and (a) via the conductance feedback factor $1/R_F$.

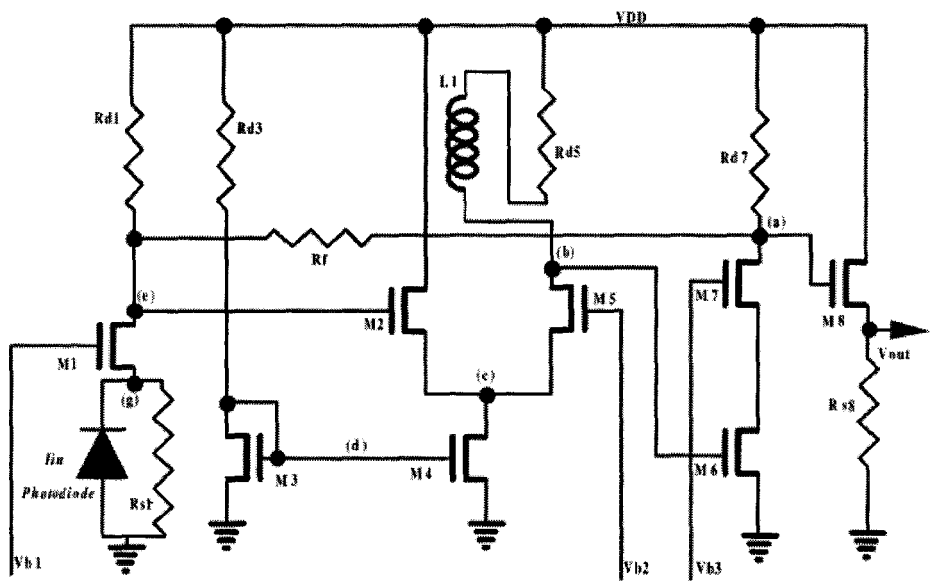


Figure 3.13 – CMOS TIA reported in [62]

The closed-loop frequency response is obtained by the zero-value time-constant technique on the open-loop equivalent circuit with feedback loading effect included [11].
The 3dB bandwidth of this amplifier is approximately [62]

$$f_{3dB} = \frac{\left(1+T(0)\right)}{\left(\omega_{p1}^{-1} + \omega_{p2}^{-1} + \omega_{p3}^{-1} + T(0)/\omega_Z\right)} \tag{3.49}$$

where $T(0)$ is the DC value of the frequency dependent loop-gain,

$$T(s) = \frac{1}{R_F} \frac{\left|Z_{T,OL}(0)\right|\left(1+s/\omega_Z\right)}{\left(1+s/\omega_{p1}\right)\left(1+s/\omega_{p2}\right)\left(1+s/\omega_{p3}\right)} \tag{3.50}$$

where $\omega_{p1}, \omega_{p2}, \omega_{p3}$ and ω_Z are the dominant poles and zeros of the circuit, respectively, and $Z_{T,OL}(0)$ is the DC open-loop transimpedance. The pole-zero plot (figure 3.14) of the open-loop and closed-loop responses reveals that the pole frequencies ω_{p1} and ω_{p2} due to

the time constants at the two feedback nodes (e) and (a) respectively, are enhanced by the factor $\left(1+T\left(0\right)\right)$ due to shunt-shunt (voltage-current) feedback.

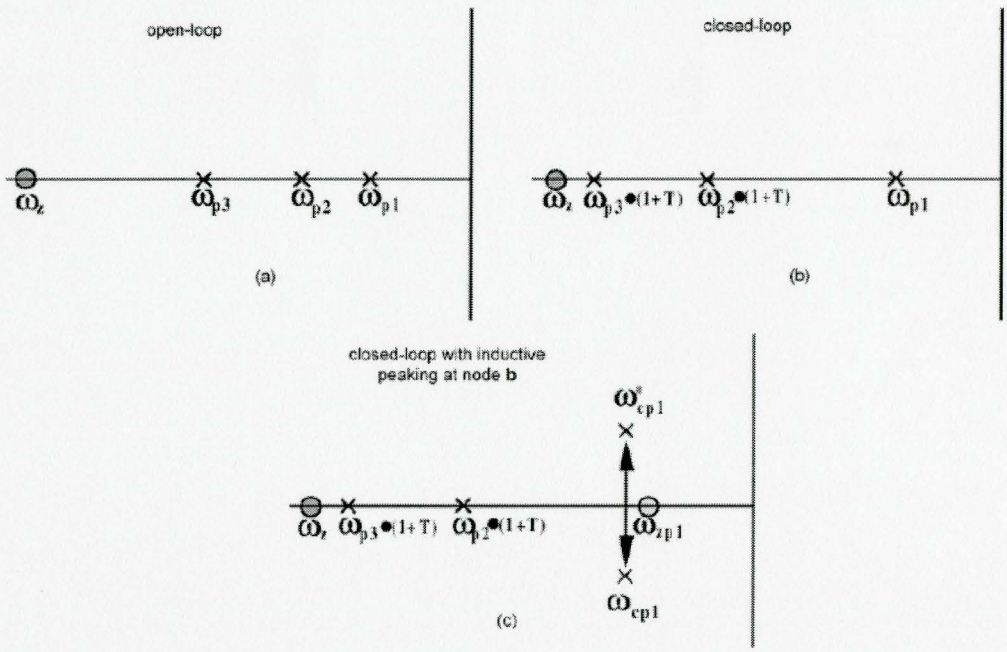


Figure 3.14 – a) Open-loop and b) closed-loop pole-zero plot of the common-gate TIA in [62]. c) Impact of inductive-peaking on closed-loop response

The bandwidth of the closed-loop amplifier is limited due to the ω_{p1}^{-1} time constant at the node (b). Therefore, shunt peaking is applied at this node to increase the pole frequency. Shunt peaking is a well-known technique used to boost the bandwidth of amplifier stages [47], [63]. The introduction of an inductance in series with the load resistance alters the frequency response of the amplifier by replacing the pole ω_{p1} due to the RC time constant at node (b), by a zero, ω_{zp1}, and a pair of complex poles, $\omega_{cp1}, \omega_{cp1}^{*}$. The zero is primarily responsible for the bandwidth enhancement due to pole-zero interaction, while the complex-conjugate poles determine the quality factor of the series R-L circuit. Intuitively, if the inductive load impedance of a shunt-peaked amplifier starts to rise at just the frequency where the gain normally rolls-off, then the inductive load impedance can boost the amplifier gain and compensate for the capacitive roll-off, thereby extending the

bandwidth. The value of the inductance L is chosen according to $m = R_{eq}^2 C_{eq} / L$, where R_{eq} and C_{eq} are the equivalent resistance and capacitance seen at node (b). A value of $m = 2.5$ yields a complex-conjugate pole pair and a zero with a maximally flat frequency response, and a bandwidth extension of 70% over the case without shunt peaking [20]. A low-Q on-chip spiral inductor that can be fabricated easily in a standard CMOS process can be utilized for this technique because a low quality factor is desired in the L-R series connection anyways to avoid excessive peaking. The biggest issue is the reduction in bandwidth improvement because of the additional parasitic capacitance introduced by the on-chip inductor. An intuitively simple and compact lumped RLC equivalent circuit can be used to model the spiral inductor [11]. The parameters are: L, the inductance of the on-chip spiral, R, the metal series resistance, C_{ox}, the capacitance to the substrate, C_m, the mutual capacitance between the spirals, R_S, the silicon substrate resistance, and C_S, the silicon substrate capacitance. The parameters of each component can be extracted through two-port S-parameter fitting. The total parasitic capacitances limits the maximum attainable self-resonance frequency of the inductor. Generally, the self-resonance frequency should be kept a factor of two above the amplifier bandwidth. Unfortunately, on-chip spiral inductors realized in standard CMOS technology are limited a narrow range of inductance values, because larger inductances imply more chip area and lower self-resonance frequency.

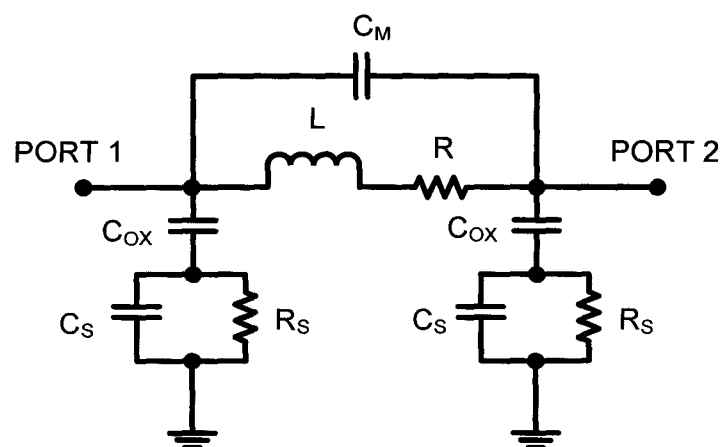


Figure 3.15 – Simple lumped circuit model of a spiral inductor [11]

The common-gate input configuration helps to relax some of the tradeoffs in TIA design, yet it cannot effectively isolate the large photodiode capacitance from the bandwidth determination because of the relatively small transconductance of the input transistor. Therefore, RGC techniques are applied to achieve better isolation of the large input parasitic capacitance

The single-ended RGC TIA reported in [52] obtains a 3dB bandwidth of 950 MHz and a transimpedance gain of 58 dB for a 0.5 pF input capacitance. The average noise current spectral density is 6.3 $pA/\sqrt{Hz}$ for this 0.6 μm CMOS design, which dissipates 85 mW from a single 5 V supply.

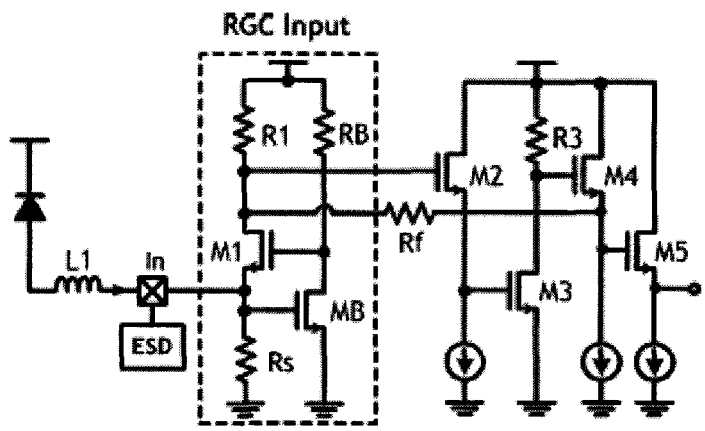


Figure 3.16 – RGC TIA [52]

Even with a 1 pF photodiode capacitance, the 3dB bandwidth is reduced only by 90 MHz, and for no photodiode capacitance, the 3dB bandwidth is 1.014 GHz. These results confirm the mechanism of the RGC input configuration (i.e. the enhanced isolation of the photodiode capacitance from the bandwidth determination). Since the RGC input stage operates as a current buffer, a second voltage-gain stage is required. A common-drain/common-source cascade is chosen for this design to mitigate the large gate capacitance of the common-source stage from loading the RGC input stage, thus deteriorating the bandwidth less. The common-drain stage also adjusts the DC level from the RGC input stage at the cost of reduced total open-loop gain and degraded linearity. Negative feedback is applied to the high-impedance node (drain of M1), thereby moving

the dominant pole of the amplifier to a higher frequency and achieving a wider bandwidth at the cost of reduced the circuit stability (phase margin) since with feedback applied, the dominant pole and the non-dominant pole of the amplifier move closer to each other which could potentially cause peaking in the frequency response.

With the photodiode capacitance isolated by the RGC input stage, the inherent parasitic capacitors of the devices become the main cause of bandwidth limitation in high-speed TIA circuits. Two-port passive interstage matching networks can be used to extend the bandwidth limitation of broadband amplifiers without degrading other circuit parameters. According to the Bode-Fano limit, the maximum obtainable gain-bandwidth product can be enhanced by a factor of four (i.e. the bandwidth at most could be extended to four times of the original amplifier with the gain unchanged) with an infinite, lossless, artificial transmission line used as a two-port inter-stage network between gain stages, where one port connects to the output of the driving stage and the other port to the input of the driven stage [20]. The principle behind this technique is to maintain a constant load over a wider frequency range. A finite LC ladder section can be utilized to realize such an enhancement, albeit with a smaller improvement because of the inherent losses. By choosing the appropriate values for L and C, the transfer function of the interstage network can be controlled to have, for example, a Bessel or Butterworth characteristic. The interstage networks can be used not only to couple amplifier stages, but also to provide broadband coupling between the photodiode and the input of an amplifier, or the output of an amplifier and a load.

The TIA reported in [54] utilizes an RGC input stage and an LC ladder interstage network to obtain a high-performance 10 Gbps CMOS TIA. Capacitive degeneration is also utilized to add an extra zero that can compensate the dominant pole, thus the 3dB bandwidth of this design is determined by the second lowest pole of the circuit. The TIA achieves a 3dB bandwidth of 8 GHz with a 0.25 pF photodiode capacitance using 0.18-μm CMOS technology. The transimpedance gain is 53 dBΩ with a 135 nW power consumption from a 1.8 V supply, and the average input-referred noise current spectral density is 18 $pA/\sqrt{Hz}$ up to 10 GHz. The interaction of the matching LC network, RGC

input stage, and capacitive degeneration stage turns the TIA into a fifth-order Butterworth low-pass filter and an overall bandwidth enhancement ratio of 3.6 is achieved compared to an RGC TIA without capacitive degeneration and LC matching network. Figure illustrates the overall TIA schematic, the small-signal model of the matching network and the RGC stage, and the equivalent low-pass filter representation of the TIA. The circuit is composed of four parts, namely the matching network, the RGC input stage, the gain stage with capacitive degeneration and the source follower output stage. Capacitive degeneration is used to compensate the dominant pole at node A with a zero, thereby extending the 3dB bandwidth. A further bandwidth enhancement is achieved by the equivalent broad-band matching network of figure consisting of C_1, L_1, C_2, and $L_{2,eff}$, where C_1 is the combined parasitic capacitance at node 1 including the photodiode capacitance C_{PD}, C_2 is the combined parasitic capacitance at node 2, and $L_{2,eff} = L_2 / \left(1 + g_{m_2} R_2\right)$. This equivalent matching network is fourth-order, and considering the dominant pole of the core TIA at node A, the whole amplifier can be approximated to a fifth-order low-pass filter with a 3dB cut off frequency of approximately

$$f_{3dB} \approx \frac{1 + g_{m_3} R_b}{2\pi \cdot R_b C_b} \tag{3.51}$$

The novelty of the design stems from the fact this TIA achieves comparable performance to its III-V and SiGe counterparts while possessing the merits of CMOS technology. The TIA reported in [64] applies the same principles; however in this case, the LC ladder inter-stage network contains just a single inductor. This special case is known as series inductive peaking. An on-chip inductor is used as the two-port inter-stage network that splits the combined parasitic capacitances of the voltage-gain stage at each node into LC networks. The inductive series peaking provides a bandwidth enhancement ratio of 2.1 (the ratio of the TIA bandwidth with inductive series peaking to that without).

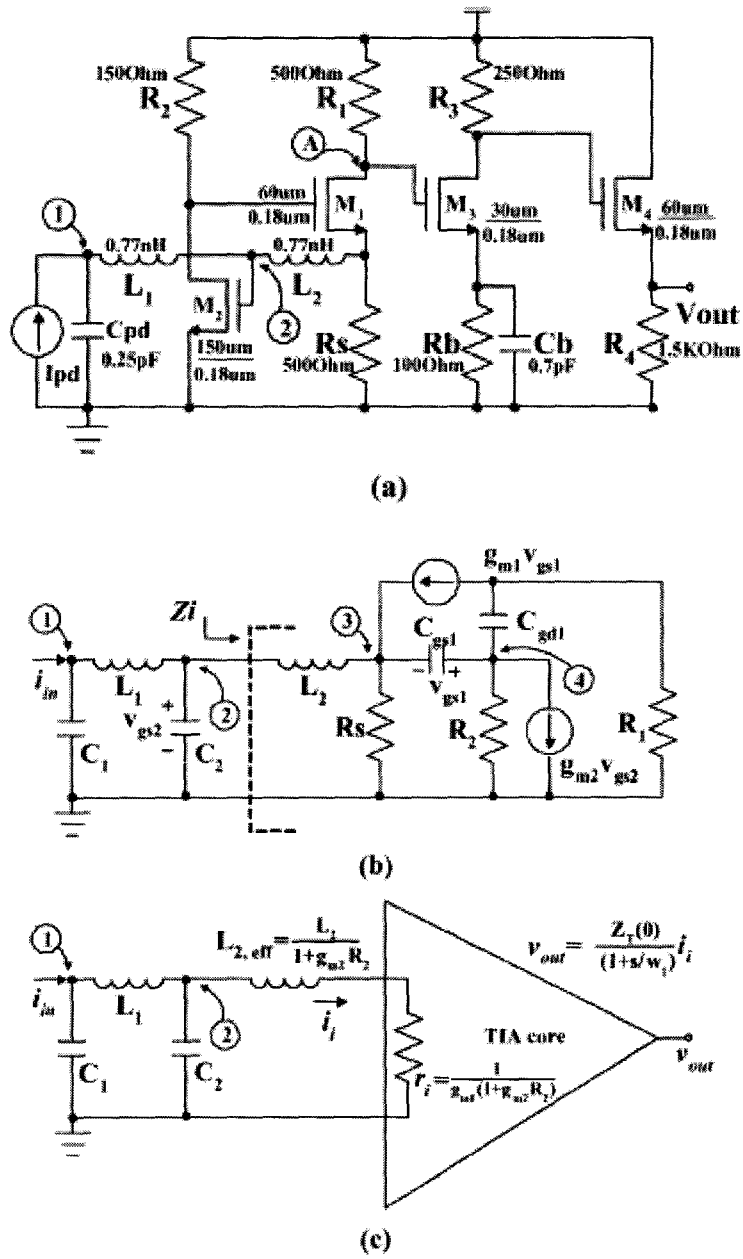


Figure 3.17 – a) Complete TIA schematic b) Equivalent small-signal model c) Equivalent low-pass filter representation [54]

The TIA comprises a RGC input stage, a shunt feedback stage and a source follower as an output stage. It achieves a transimpedance gain of 51 dBΩ and 3dB bandwidth of 6 GHz in the presence of a photodiode capacitance of 0.6 fF while

consuming 13 mW from a 3.3 V power supply. At the time of publication (2006), this was the fastest TIA ever reported in 0.35-μm CMOS technology.

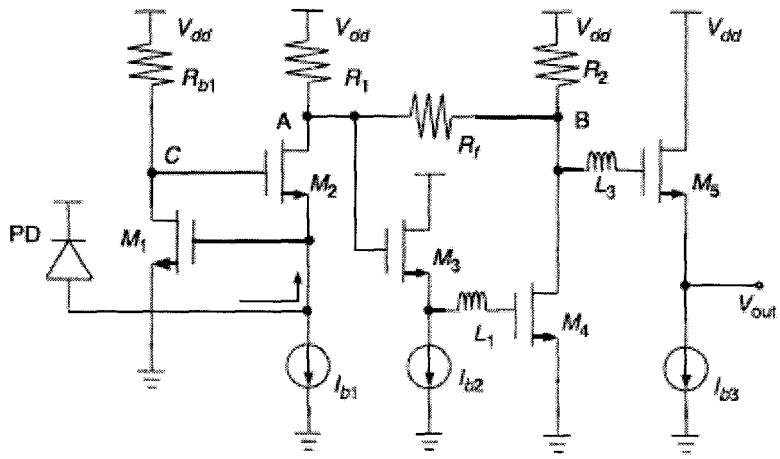


Figure 3.18 – RGC TIA with interstage matching networks [64]

Although high-speed transimpedance amplifiers have traditionally been implemented using single-ended topologies for simplicity, a differential configuration offers a number of significant advantages, especially in CMOS technology. Among the most significant ones are the improved immunity to power-supply and substrate noise as well as the increased voltage swing. The common-mode rejection property inherent in a differential circuit therefore makes it far more manageable to integrate the TIA with noisy digital circuitry, or with other TIA circuits in an array with minimal crosstalk. In addition, the dynamic range of a differential TIA is typically larger than that of a single-ended TIA, since larger inputs voltages in a differential stage can be beneficial for faster output switching, whereas a single-ended circuit would saturate, causing a degraded output eye-diagram unless automatic gain control circuitry is employed to automatically lower the transimpedance gain so that the input overload current can be larger. A differential TIA also facilitates the connection to a differential post-amplifier, avoiding the need for a reference voltage. The main drawbacks of a differential TIA are the higher power consumption as well as higher input-referred noise and larger die area. Unfortunately, the

input to a differential TIA can only be single ended because differential photodetectors are normally not available for the OOK format. The unipolar single-ended photocurrent contains a DC value which affects the performance of DC coupled TIA circuits, for example, causing asymmetry (offset) in the output signal levels of a differential TIA and reducing the dynamic range, since only half of the available TIA output swing can be utilized. Therefore, offset control circuits are generally required in a differential TIA [65]. To remove the offset, the average DC photocurrent must be subtracted from the photodiode such that the current flowing into the TIA swings symmetrically about the zero level.

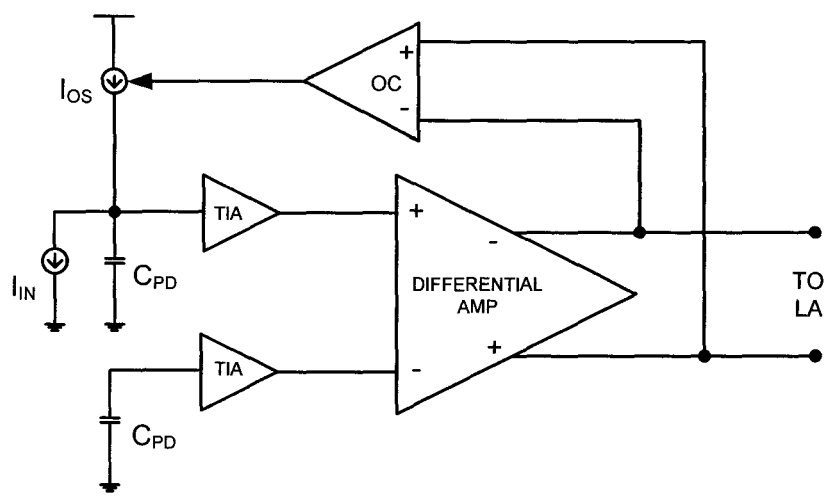


Figure 3.19 – Differential TIA employing an offset cancellation circuit [20]

A replica TIA and a dummy capacitor must be applied to the unused input of the differential amplifier to achieve a balanced differential topology. The offset cancellation (OC) circuit obtains the output offset voltage of the differential signal signals either by low-pass filtering or by peak-detection and uses an op-amp to adjust the current source I_{OS} such that the output offset becomes zero [20].

The differential TIA reported in [66] employs an offset correction circuit. The TIA achieves a 3dB bandwidth of 1.9 GHz for a 0.3 pF input capacitance and a 59 dB transimpedance gain with an average input-referred noise current spectral density of 9.75

$pA/\sqrt{Hz}$ implemented in a 0.15 μm CMOS technology, and is part of a single-chip receiver that includes a variable gain amplifier (VGA), a CDR circuit, and a DMUX.

The differential TIA reported in [47] achieves a 3dB bandwidth of 1.2 GHz with a transimpedance gain of 64 dBΩ and an average input-referred noise current spectral density of 17 $pA/\sqrt{Hz}$ while consuming 22.5 mA from a 3.3 V supply in a 0.5 μm CMOS technology. This differential TIA consists of a common-gate input stage followed by a shunt-peaked differential gain stage with cascode transistors, and with the inductors using patterned ground shields to reduce substrate coupling [67]. Typical of RFICs, the chip area is dominated by the passive components; in this case the inductors combine for less than 15% of the total area. Source follower transistors buffer the outputs of the differential stage and drive the feedback resistors. To obtain a balanced TIA configuration, the common-gate input stage is replicated at the unused input.

Table V summarizes the performance of the TIA designs that were discussed in this chapter, along with a variety of other CMOS TIA designs that have been reported in the literature. The research effort focusing on TIA circuits can be divided roughly into four areas; higher speed, higher integration, lower cost, and lower noise. CMOS technology can meet these criteria simultaneously, especially with circuits that utilize regulated cascode, shunt-series peaking, capacitive degeneration, and inter-stage broadband matching techniques.

Ref	R_T (dBΩ)	f_{3dB} (GHz)	i_n^{rms} $\left(pA/\sqrt{Hz}\right)$	Power (*mW*)	Supply (V)	Input Stage	Cpd (pF)	Technology (CMOS)
[56]	20	1.26	n/a	180	4.5	CS	n/a	0.55- μm
[68]	104	0.12	1	n/a	5	CS	1	0.8- μm
[69]	79	0.55	4.5	30	3	CS	n/a	0.6-μm
[45]	64	1.2	n/a	n/a	n/a	CG	0.32	0.6-μm
[70]	66.5	0.66	4.6	155	5	CD	n/a	0.8- μm
[71]	52	1	7.4	80	5	CS	0.5	0.8- μm
[48]	57.7	0.3	10	n/a	6	RGC	n/a	0.6-μm
[47]	64	1.2	17	225	n/a	CG (diff.)	0.6	0.5-μm
[49]	55.3	2.2	n/a	210	5	RGC	0.5	0.6-μm
[50]	80	0.7	n/a	27	2.5	RGC	1	0.25-μm
[51]	80	0.67	21	27	2.5	RGC	1	0.25-μm
[52]	58	0.95	6.3	85	5	RGC	0.5	0.6-μm
[53]	60	11	18	22	1.8	RGC	0.25	0.18-μm
[72]	58	6.3	n/a	n/a	1.8	RGC	0.3	0.18-μm
[66]	59	1.9	9.75	n/a	2	CS (diff)	0.3	0.15-μm
[64]	51	6	21	15	3.3	RGC	0.6	0.35-μm
[73]	54.5	2.2	17	n/a	3	RGC	0.3	0.35-μm
[74]	54.5	2.5	16	n/a	3	RGC	n/a	0.35-μm
[75]	50	7.86	n/a	n/a	1.8	RGC	0.15	0.18-μm
[76]	48	8.46	27.4	5.9	1.8	RGC	n/a	0.18-μm
[77]	80	2	7.8	58.4	2.5	Cascode (diff.)	0.5	0.5-μm
[78]	52	7.6	n/a	34	2	RGC	0.25	0.18-μm
[79]	62	4.2	20	45.5	1.8	RGC (diff.)	0.5	0.18-μm
[80]	53	8	18	13.5	1.8	RGC	0.25	0.18-μm
[81]	60	2.5	18.9	16	2.5	RGC (diff.)	0.2	0.25-μm

Table 3.1 – Summary of CMOS TIA designs reported in the literature

3.3.2 Clock and Data Recovery (CDR) Circuits

CDR circuits are widely used in optical communication systems. Furthermore, operating CDR circuits at higher data rates, while using inexpensive CMOS technology, is a key to enable higher bandwidth communications at ever lower cost per unit bandwidth. Integration of the CDR with an AFE could further potentially reduce the cost and maximize overall performance. Therefore, this section will examine the performance of several recently published high-speed, highly-integrated CMOS CDR circuits.

The architecture of a 0.4 μm CMOS CDR reported in [82] employs a number of novel techniques to support a data rate of 2.5 Gbps. First, since the maximum oscillation

frequency for a three-stage ring oscillator does not exceed 2.4 GHz in this 0.4 μm technology, a two-stage differential ring topology was chosen for the VCO to obtain reliable operation at 2.5 GHz. However, two simple differential pairs in a loop fail to oscillate because each stage contributes only one pole, yielding insufficient phase to meet the Barkhausen criteria for oscillation. Thus, excess phase is introduced in each stage by a composite (inductive) load that provides enough phase-shift around the loop, thereby allowing oscillation. To obtain a wide tuning range, the VCO incorporates delay interpolation to vary the oscillation frequency with both fine and coarse tuning [87]. The PD circuit meanwhile utilizes a novel approach that incorporates the speed advantages of a nonlinear (bang-bang) PD but with the linear characteristic of linear PD. The PD is realized as a master-slave sample-and-hold (i.e. an analog D flip-flop) circuit, whereby each rising data transition samples the instantaneous value of the VCO output. The circuit thus generates an output that is linearly proportional to the input phase difference for frequencies in the vicinity of the PLL acquisition range and for phase differences as large as $\pm 50^{\circ}$. The PD output voltage drives a transconductance stage (I/V converter) preceding the loop filer, which drives the VCO control voltage, thus completing the loop. The VCO output provides the recovered clock through a set of open-drain buffers to 50 Ω termination resistors. This CDR achieves an acquisition range of 15 MHz and the recovered clock exhibits a rms jitter equal to 10.8 ps, and phase noise equal to -80 dBc/Hz at a 5 MHz offset, for a 2.5 Gbps PRBS sequence of length $2^7 - 1$. The total power dissipation is 33.5 mW (excluding that of the 50 Ω drivers) from a 3.3V power supply. The performance is comparable with that of a CDR designed in a 30 GHz bipolar technology [83].

The first ever 10 Gbps CDR circuit in 0.18 μm technology was reported in [84]. The CDR circuit incorporates a three-stage interpolating ring oscillator with fine/coarse tuning to achieve a wide tuning range, and a new approach to performing linear phase detection using a 'half-rate' clock [88]. The half-rate architecture allows a VCO to run at a frequency equal to half of the input data rate. The half-rate architecture alleviates the difficulties of linear phase-detection at high-speeds. For instance, at 10 GHz, the

maximum output pulse width of a linear PD is only 50 ps, causing even simple digital latches to fail to operate reliably in a 0.18 μm technology. With a half-rate clock recovery, both the positive and the negative transitions of the recovered clock are used for sampling. The main drawback of the half-rate architecture is the degradation in jitter tolerance. Since both clock edges sample the data waveform, any deviation of the clock duty cycle from 50% can degrade the jitter performance. Therefore, the VCO layout must be well-matched and symmetric to avoid the clock duty cycle distortion and timing skew. The XOR building blocks of the PD employ current-mode logic (CML) instead of rail-to-rail CMOS logic to minimize the switching noise in the substrate and on the supplies, which can disturb the oscillator considerably. A differential charge pump and loop filter follow the PD. This CDR achieves a 1 ps RMS jitter on the recovered clock signal from a 10 Gbps NRZ random data sequence of length $2^{23}-1$, and the phase noise at 1 MHz offset is approximately equal to -106 dBc/Hz. The CDR circuit exhibits an acquisition range of 6 MHz and the total power consumed by the circuit is 72 mW from a 2.5 V supply.

The first fully integrated 10 Gbps receiver fabricated in a standard 1-poly 6-metal 0.18 μm CMOS process was reported in [85]. The receiver integrates an input amplifier, a CDR, a 1-to-16 DMUX, and a low-voltage differential signal (LVDS) interface for the output clock and data signals. The CDR employs a dual-loop architecture, where the frequency-locked loop acts as an acquisition aid to the phase-locked loop which is disengaged during normal operation when the CDR is locked to the input serial data. Control logic, consisting of a data detector and a lock detector, control the transition between the frequency acquisition loop and the PLL. The PLL implements a full-rate architecture with a LC VCO to generate a low-jitter, full-rate clock that retimes the data from the input amplifier. The LC tank is formed by MOS varactors and an on-chip spiral inductor. To increase the tuning range, the VCO has two banks of varactors; the ones controlled by the VCO control voltage for fine tuning and the ones controlled by external band-select signals for coarse tuning. As a result, the VCO operates in an extended frequency range while simultaneously keeping a relatively small VCO gain, helping to

reduce the phase noise caused by the ripples on the control lines. The CDR employs a linear PD to accurately control the transfer characteristics of the PLL (i.e. loop bandwidth and peaking). By using external components for the loop filter, the loop-filter bandwidth can be conveniently adjusted. The need for high-speed voltage signals to switch the charge pump is alleviated through the use of the linear PD followed by a transconductance stage, and there is no current to charge/discharge the filter when the loop is locked. As a result, the CDR can operate at a frequency as high as 11 GHz with small jitter. After being regenerated in the CDR, the 10 Gbps data stream is fed to the 1:16 DMUX. The receiver die size is 2.5 x 2.1 mm^2, and the chip operates from a single 1.8 V supply, consuming 870 mW, with the output drivers counting for more than half of the total power consumption. This is considerably less than the 4.5 W consumption of a similar receiver chip implemented earlier in SiGe technology [87]. The recovered clock for a 10 Gbps NRZ PRBS with length $2^{31}-1$ exhibited a measured phase noise of -107 dBc/Hz at a 1 MHz offset from the carrier. The RMS jitter measured over a band of 50 kHz – 80 MHz was 0.38 ps. Jitter tolerance specifications for SONET OC-192 were exceeded with more than 100% margin.

The 10 Gbps CDR reported in [86] uses a quarter-rate linear PD and a four-phase 2.5 GHz LC quadrature VCO for both wide phase error pulses and low power consumption in a 0.13 μm technology. With a four-phase recovered clock from a quarter-rate LC quadrature VCO, the pulsewidth of the error pulses is doubled, therefore relaxing the requirements for the transition frequency f_T of the process and the power consumption. The total power consumption of the IC, which integrates the CDR with a DMUX and low-voltage differential drivers. The CDR utilizes a dual-loop architecture, similar to that of [85]. The frequency-acquisition loop consists of a PFD, a CP, a MUX, a loop filter, two dividers, and a four-phase quadrature LC VCO. The phase-locked loop includes an input buffer, a quarter-rate linear PD, a data recovery circuit, a transconductance cell, and shares a MUX, a loop filter, and the quadrature VCO with the frequency acquisition loop. The frequency-acquisition loop initializes the oscillation frequency of the VCO by using an external reference clock, of which frequency is

selectable by an external control pin. When the frequency acquisition reaches within ± 400 ppm , a lock detector output a signal to indicate that the initialization of the VCO is completed. Then a lock controller sets the MUX to switch from the frequency-acquisition loop to the phase-locked loop. At the same time, the lock controller also turns off the CP during PLL operation in order to prevent the up/down pulses of the PFD from interfering with the loop filter through the MUX. The PLL recovers clock and data from the received 10 Gbps data stream. The four-phase 2.5 GHz recovered clocks are used for data recovery, and thus the recovered data are in a 4-bit de-serialized format. The 4-bit recovered data are de-multiplexed once by to 16-bit data outputs, by a 4:16 DEMUX, and then fed to the LVDS drivers along with the recovered clock.

Table 3.2 summarizes the performance of the CDR designs that were presented thus far, along with a variety of others that have been reported in the literature.

Ref	Bit Rate (Gbps)	VCO Phase Noise (dBc/Hz)	Clock Jitter (ps_{rms})	Power (mW) -Supply (V)	Sensitivity (mV_{pp}) @ BER = 10^{-12}	Jitter Tolerance (UI_{pp})	Chip Area (mm^2)	Process (CMOS)
[82]	2.5	-80 @ 5 Mhz	10.8	33.5-3.3	n/a	n/a	0.8 x 0.4	0.4 µm
[84]	10	-106 @ 1 Mhz	1	72-2.5	n/a	n/a	1.1 x 0.9	0.18 µm
[85]	10	-107 @ 1 Mhz	0.38	870-1.8	50	0.4	2.5 x 2.1	0.18 µm
[89]	10	-106 @ 1 Mhz	1.1	980-1.2	15	0.45	3 x 5	0.13 µm
[90]	10	n/a	0.44	1065-1.2	45	0.6	5 x 5	90 nm
[91]	10	n/a	n/a	990-1.2	12	0.3	n/a	0.13 µm
[92]	2.5	-105 @ 1 MHz	1	500-1.8	6.5	0.5	3.2 x 3.8	0.18 µm
[93]	10.8	n/a	3.2	220-1.5	150	n/a	0.25 x 1.4	0.11 µm
[94]	9.4-11.3	-110 @ 1 MHz	2.1	120-1.2	210	0.5	3 x 3.4	0.13 µm
[95]	10	-107 @ 1 MHz	0.74	290-1.8	150	0.3	2 x 2	0.18 µm
[96]	11.3	-112 @ 1 MHz	n/a	385-1.8	10	n/a	4 x 2	0.13 µm
[97]	1.8	n/a	9.9	12-1.2	50	n/a	0.1	0.13 µm
[98]	2.25	n/a	15.7	100-1.8	n/a	0.7	0.7 x 0.63	0.18 µm
[99]	3.125	n/a	5.9	32-1.2	50	0.5	0.12	0.15 µm
[100]	5	n/a	1.2	97-1.8	n/a	n/a	0.2 x 0.32	0.18 µm
[101]	11.7	n/a	1.4	86-1.5	n/a	n/a	1.3 x 1.5	0.13 µm

Table 3.2 - Summary of CMOS CDR designs reported in the literature

Chapter 4
CMOS RGC TIA IMPLEMENTATION

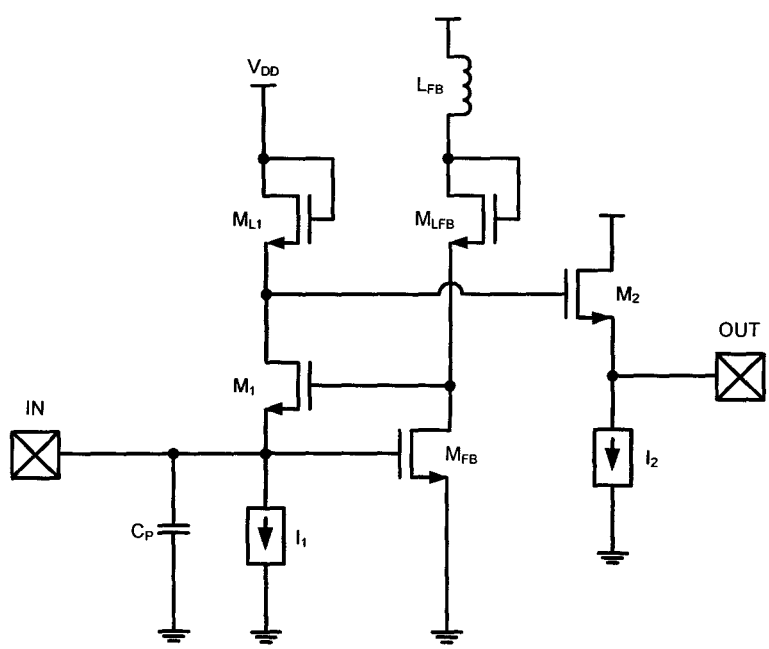
In the design of optical communication systems, the design of the front-end preamplifier is well known to be the most difficult as it plays a critical role in dictating the whole system performance. In particular, the design of the TIA entails a careful optimization of a number of trade-offs between noise, bandwidth, gain, and power consumption. It is necessary to simultaneously achieve large bandwidth, high-gain, low-noise and low-power consumption to achieve the best performance. These conflicting requirements can pose a difficult challenge for implementation with submicron CMOS technology due to intrinsic parasitic capacitances of active and passive devices, as well as due to the large input parasitic capacitance of the photodiode. Sections 3.2 and 3.3 discussed techniques that can be utilized to obtain high-performance CMOS TIA circuits in spite of these factors. It was shown that the RGC input-stage is an effective means of isolating the photodiode capacitance from affecting the performance of the TIA, especially in terms of bandwidth and sensitivity, while simultaneously achieving a wide ripple-free pass-band with a low input-referred noise current. It was also shown that an RGC input stage can be effective as AFE designed for multi-Gbps CMOS optical receivers. Therefore, a RGC TIA AFE circuit was chosen for fabrication in a standard mixed-signal TSMC 0.18 μm CMOS technology. This chapter describes the design, simulation and measured performance of the AFE chip.

4.1 Design and Simulations

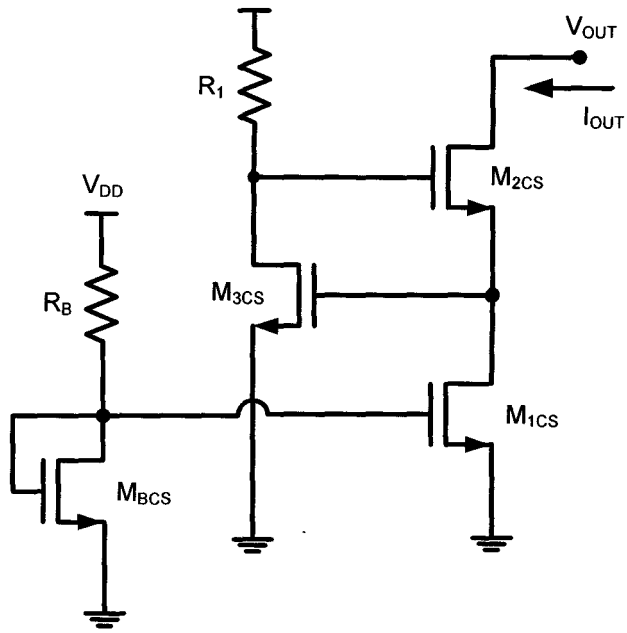
Among all the challenges in the design of fully integrated CMOS broad-band TIA circuits, sufficient bandwidth with small gain ripple is of first priority and low-noise is second because the noise of the preamplifier dominates that of the whole receiver. Therefore, the following objectives were chosen for the RGC TIA design:

- To achieve ripple-free pass-band from DC to 6.6 GHz with a group-delay variation of less than 10 ps and a transimpedance gain of 40 dBΩ with less than 1dB variation in magnitude.
- To obtain an average input-referred noise current spectral density below 20 $pA/\sqrt{Hz}$ (corresponding to a -18 dBm optical sensitivity at $BER = 10^{-12}$, assuming a photodiode responsivity of 0.8 A/W)
- To minimize the power consumption from a nominal 1.8 V supply.
- To match the amplifier to the 50 Ω measurement interface.

The complete RGC TIA circuit schematic is shown in figure 4.1a. The capacitance C_P represents the total input parasitic capacitance of the TIA, which is contributed mainly by the bond-pad capacitance. It is assumed to be 150 fF in the simulations to model the effects of the input parasitic capacitance on the amplifier performance. In order to simplify the circuit biasing, two on-chip current sources, I_1 and I_2 were chosen in this design. The current sources were implemented with a RGC current mirror, shown in figure 4.1b.



a)



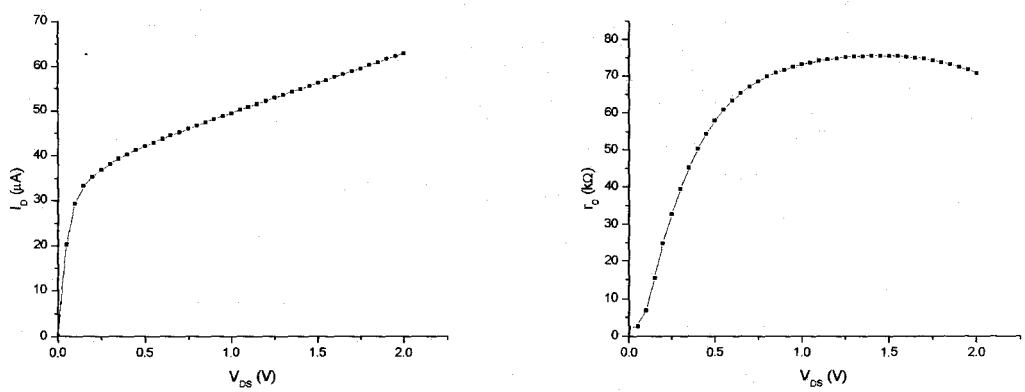
b)

Figure 4.1 – a) RGC TIA Schematic b) RGC Current Source Schematic

RGC current sources were chosen because of their low compliance voltage, large output impedance and output swing, and low output capacitance compared to other stacked current mirrors such as the simple cascode or Wilson current mirror [102]. Inductive shunt-peaking was used in the local feedback stage in order to extend the TIA bandwidth. Shunt-peaking was applied in the local feedback stage, instead of the input CG stage, because the gain of the feedback stage must be considerably larger in order to satisfy the requirements of equations (3.41) & (3.44). A common-drain (CD) output buffer follows the RGC input stage so that the amplifier can drive an external 50 Ω load.

Hand calculations using the quadratic FET model were performed in order to obtain initial estimates for the transistor sizes, as well as the resistance and inductance values for the overall circuit of figure 41. In order to perform these calculations, the transconductance factor, $\mu_n C'_{ox}$, and the threshold voltage, V_{TH}, were extracted first by performing I-V sweeps with the Cadence software package for the TSMC 0.18 μm CMOS process FET models.

The transconductance factor was extracted by performing a drain voltage sweep for a single-finger nMOS transistor (W/L = 2.5/0.18), which was biased at a gate voltage of 0.6 V. The resulting graph is shown in figure 4.2a. The output resistance of the FET was obtained by taking the inverse of the first derivative of the I-V sweep (figure 4.2b). Several values for the drain current and drain-source voltage were chosen to solve for the transconductance factor by using the DC quadratic FET model in the saturation regime (equations 3.9 & 3.10), and the results are shown in table 4.1.



a) **b)**

Figure 4.2 – a) Drain current as a function of drain-source voltage at $V_{GS} = 0.6$ V
b) Output resistance as a function of the drain-source voltage at $V_{GS} = 0.6$ V

V_{DS}	I_D	r_0	$\mu_n C_{ox}$
1.04 V	49.98 μA	73.35 kΩ	0.72 mA/V^2
1.52 V	56.38 μA	75.49 kΩ	0.81 mA/V^2
1.84 V	60.67 μA	73.54 kΩ	0.87 mA/V^2

Table 4.1 – Extracted values of $\mu_n C_{ox}$

The value of $\mu_n C_{ox}$ was found by taking the average of the extracted values and was found to be approximately 0.8×10^{-3} A/V^2.

The threshold voltage V_{TH} was similarly obtained, but by instead performing a gate voltage sweep for transistors with various gate widths biased at a fixed drain-source voltage, and by fitting the resulting drain current to a straight line. The point at which the lines intersect the x-axis denotes the approximate threshold voltage. Figure 4.3 reveals that the value of the threshold voltage is approximately 0.5 V.

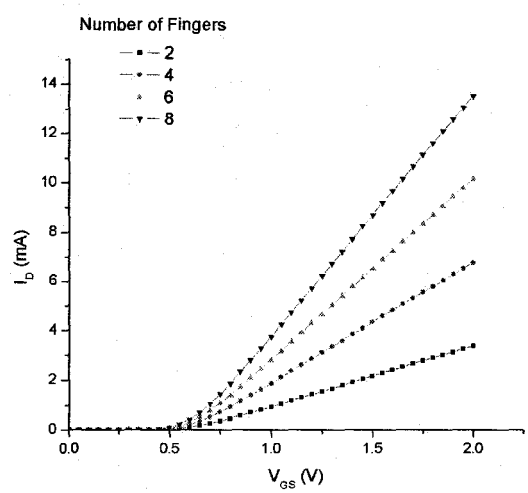


Figure 4.3 – Drain current as a function of gate-source voltage at $V_{DS} = 1.2$ V

The gate width (number of fingers) of the input transistor M1 was chosen such that the input impedance is equal to 50 Ω. By neglecting channel-length modulation, the input impedance of a CG stage becomes $R_{IN} = 1/g_m$. Therefore, the required gate-source voltage for transistor M_1, assuming a drain current of $I_{D1} = 1\,\text{mA}$, can be found according to

$$V_{GS} = \frac{2I_D}{g_{m_1}} + V_{TH} = \frac{2 \cdot 1 \cdot 10^{-3}}{20 \cdot 10^{-3}} + 0.5 = 0.6\ \text{V} \tag{4.1}$$

Equation (3.6) was used to estimate the number of fingers for M_1

$$n_{M1} = \left(\frac{L}{W}\right)\frac{2I_D}{\mu_n C'_{ox}\left(V_{GS} - V_{TH}\right)^2} = \frac{0.18}{2.5}\frac{2\cdot 10^{-3}}{0.8\cdot 10^{-3}\cdot 0.01} = 18 \tag{4.2}$$

Similarly, the gate width of the output buffer CD stage was chosen such that the output impedance of the circuit is 50 Ω. For a bias current of 2 mA, the required gate-source voltage for transistor M_2 was

$$V_{GS} = \frac{2I_D}{g_m} + V_{TH} = \frac{2\cdot 2\cdot 10^{-3}}{20\cdot 10^{-3}} + 0.5 = 0.7 \text{ V} \tag{4.3}$$

and the required transistor size was

$$n_{M2} = \frac{L}{W}\frac{2I_D}{\mu_n C'_{ox}\left(V_{GS} - V_{TH}\right)^2} = \frac{0.18}{2.5}\frac{4\cdot 10^{-3}}{0.8\cdot 10^{-3}\cdot 0.04} = 9 \tag{4.4}$$

The voltage gain of the local feedback stage was chosen to be approximately $A = g_m R_D = 7$ in order to satisfy the conditions in equations (3.9) and (3.10). A drain resistance of $R_D = 1\,\text{k}\Omega$ was assumed, therefore the transconductance of the feedback transistor, M_{FB}, was required to be at least 7 mS. In order to obtain the proper size and gate-source bias voltage for the desired transconductance, a sweep of the transconductance versus gate-source voltage was performed for various transistor sizes with $V_{DS} = 1.2\,\text{V}$. Figure 4.4 reveals that a 6 finger transistor is required to obtain a transconductance of 7 mS, and that the gate-source voltage of M_{FB} should be no less than 0.75 V.

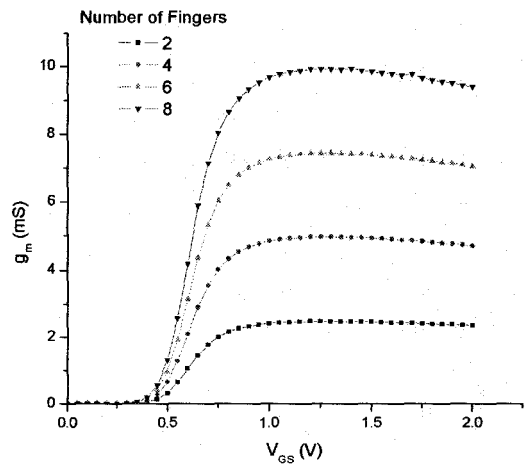


Figure 4.4 – Transconductance as a function of gate-source voltage at $V_{DS} = 1.2$ V

The initial estimate for the current source I_1 was chosen as 1 mA in order to minimize power consumption. Transistor sizes were selected according to the conditions for optimal biasing of the RGC current mirror. In particular, for M_{1CS} to remain in saturation, the gate-source voltage must satisfy [102]

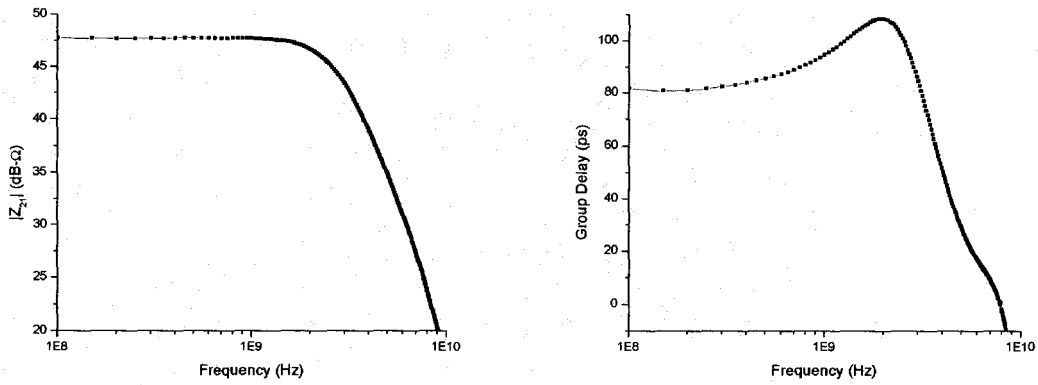
$$V_{GS_{M1CS}} = \sqrt{\frac{2I_1}{\mu_n C'_{ox} n_{3CS}(W/L)}} + 2V_{TH} = \sqrt{\frac{2\cdot 10^{-3}}{0.8\cdot 10^{-3}(2.5/0.18)}} + 2\cdot 0.5 = 1.24 \text{ V} \quad (4.5)$$

Therefore, in order for the RGC current source to obtain 1 mA of bias current, the required gate-source voltage of M_{1CS}, must be 1.24 V, assuming a single-finger feedback transistor M_{3CS}. The gate-source voltage for M_{1CS} is provided by the diode-connected transistor M_{BCS} and the resistor R_B. The corresponding compliance voltage of the current source is 0.92 V. Increasing the size of the feedback transistor allows M_{1CS} to operate at a much lower compliance voltage. In particular, to obtain a minimum gate-source voltage of 0.6 V for M_{1CS} at a 1 mA bias current, transistor M_{3CS} would require 18 fingers. Assuming that 1 mA of current flows through transistor M_{BCS}, the value of the resistor

R_B should be equal to $(1.8-1.1)/10^{-3} = 700\,\Omega$. Using equation (4.2), it was found that the transistor M_{BCS} required a single finger to support a 1 mA current at a gate-source voltage of 1.1 V. The transistors M_{1CS} and M_{2CS} were chosen to have the same size as M_{BCS} to preserve symmetry of the current mirror. The diode-connected load transistors M_{L1} and M_{LFB} were also chosen to have the minimum size for the initial simulations.

Transistor sizes for current source I_2 were similarly obtained, but with the estimated current equal to 2 mA. Using equation (4.4), the required gate-source voltage for a single-finger feedback transistor is 1.6 V. Increasing the feedback so that the compliance voltage is 0.6 V required M_{3CS} to have 36 fingers. The resistor R_B for the second current source was found to be 350 Ω, and the remaining transistors used 2 fingers.

Simulations on the RGC TIA circuit were conducted using the Cadence software package, where the Virtuoso schematic editing tool was used for schematic entry and the Cadence SpectreRF tool was used to perform DC, AC small-signal, transient large-signal, noise and s-parameter analyses. The component values and transistor sizes which were obtained from the hand calculations were used as the starting point for the simulations. Since the hand calculations were based on the quadratic FET model which neglected channel-length modulation, body-effect, and other second-order effects, the simulated performance of the circuit was expected to be sub-optimal. Figure 4.5 shows the initial simulation results did not meet the criteria outlined at the beginning of the chapter. Several considerations were addressed so that the simulated performance could meet the specifications for 10 Gbps operation with sufficient margin. The first consideration was to ensure that the group delay variation remain less than 10 ps within the specified passband. According to equation (3.47), the group delay peaking of a RGC stage is due to the zero introduced by the local feedback. Therefore, the load resistance of the feedback amplifier was decreased in order to increase the frequency of the zero resulting from the feedback, thereby minimizing the peaking of the group delay within the passband.



a) **b)**

Figure 4.5 – Un-optimized simulation results
a) Transimpedance b) Group delay

This was done by increasing the size of the diode-connected load transistor M_{LFB} up to 15 fingers. The size of the feedback transistor M_{FB} was also increased up to 15 fingers in order to provide sufficient gain from the local feedback stage.

The second consideration was to ensure that the bandwidth of the TIA remains sufficiently wide in order to support 10 Gbps operation. Since the gain-bandwidth product of any single stage amplifier is a constant, the gain of the CG stage was reduced in order to improve the bandwidth. It was found that increasing the size of the diode-connected load transistor M_{L1} to 8 fingers improved the bandwidth to the required 6 GHz, and also helped to further reduce the group delay variation within the passband. Furthermore, reducing the size of transistor M_1 to 12 fingers also helped to reduce the group delay variation by moving the peak to a slightly higher frequency.

The third consideration was to ensure that the noise performance of the TIA met the specified criteria outlined at the beginning of the chapter. The input-referred noise current spectrum of the un-optimized circuit is shown below in figure 4.6. The corresponding average input-referred noise current spectral density was found to be equal to 21.06 $\text{pA}/\sqrt{\text{Hz}}$.

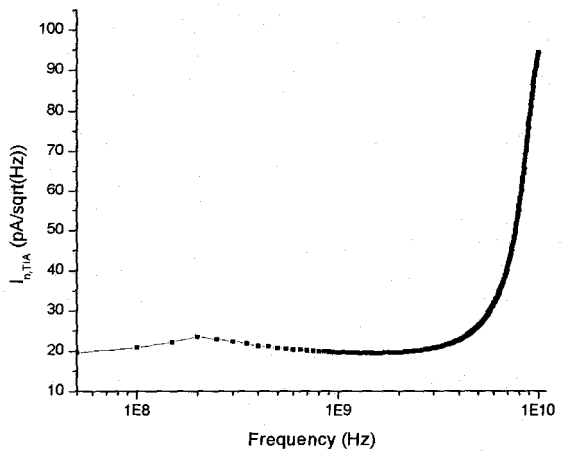


Figure 4.6 – Input referred noise-current spectrum of the non-optimized RGC TIA circuit

Equation (3.46) reveals that the dominant high-frequency noise terms of the input-referred noise current spectrum are divided by the $\left(g_{m_{FB}}+1/R_{FB}\right)^2$ term which is due to the local feedback. Therefore, increasing the sizes of the transistors in the local feedback stage also improves the noise performance. Figure 4.7 shows the input-referred noise current spectrum of the optimized circuit, which achieved an average input-referred noise current spectral density of 16.62 $\text{pA}/\sqrt{\text{Hz}}$.

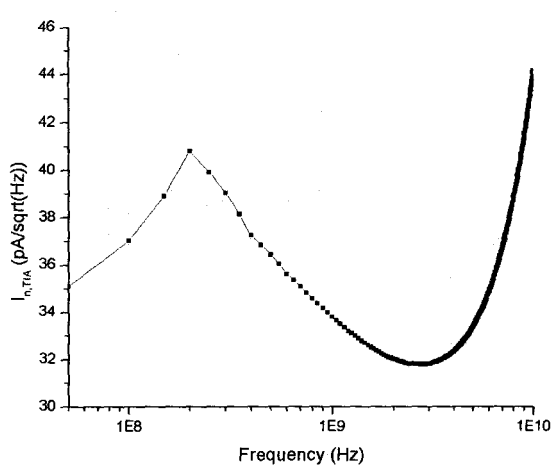
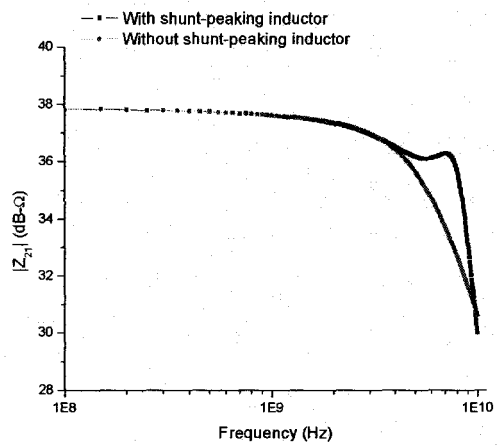
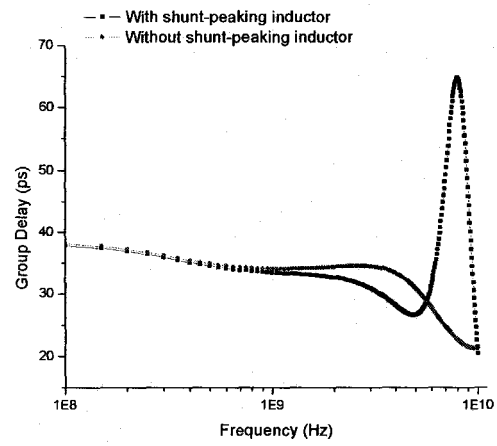


Figure 4.7 – Input-referred noise current spectrum of the optimized RGC TIA circuit

The final considerations for the optimization process were to ensure that the input/output impedance of the simulated circuit was matched to 50 Ω, and to find the proper inductance value which yielded the largest improvement in bandwidth. S-parameter simulations revealed that the size of the output CD stage transistor M_2 needed to be increased for proper matching and the value of the current source modified accordingly. As for the shunt-peaking inductor, the TSMC 0.18 μm CMOS process provides rectangular spiral inductors which utilize the thick metal-6 layer available in this technology. The lumped circuit model used to simulate these inductors is shown in figure 3.15. Each available inductance value, with its corresponding inductor model, was inserted into the circuit, and the effects of shunt-peaking on the 3dB bandwidth and group-delay peaking were examined. For the un-optimized circuit, the inductor had no effect on the bandwidth or group-delay. This was because the gain of the local feedback stage was not large enough to place the dominant pole away from the input node. For the optimized circuit, the optimal inductance was found to be equal to $L = 3.878$ nH (corresponding to a 3.5 turn inductor), which yielded an approximate 65% bandwidth enhancement compared to the case without the inductor ($f_{3dB} = 5.6$ GHz without shunt-peaking, $f_{3dB} = 8.3$ GHz with shunt-peaking), but at the cost of slight group-delay peaking at the upper end of the passband. Figure 4.8 shows the optimized simulation results with and without the aid of the on-chip shunt-peaking inductor. Table 4.2 summarizes the simulated performance of the optimized TIA and table 4.3 lists the optimized component values which were used for the final circuit layout. The bandwidth variation due to changes in the input capacitance was simulated in order to verify the isolation of the RGC input stage. Figure 4.9 shows the transimpedance magnitude response for several different values of input capacitance, revealing that the transimpedance gain begins to drop at high frequencies as the input capacitance increases. Table 4.4 summarizes the bandwidth variation of the transimpedance, confirming that the bandwidth variation of the RGC input stage is small for an input parasitic capacitance up to 350 fF.



a)



b)

Figure 4.8 – Optimized simulation results with and without shunt-peaking inductor a) Magnitude response b) Group delay response

3dB Bandwidth	8.3 GHz
Transimpedance	38 dBΩ
Group Delay Variation	< 10 ps
Average input referred noise current density	16.62 pA/$\sqrt{Hz}$
Power consumption	18.4 mW (1.8 V)
Input Reflection Coeff.	< -12 dB
Output Reflection Coeff.	< -8 dB

Table 4.2 – Simulated Performance Summary

Element	Value
M_1	0.18 μm (L) x 2.5 μm (W) x 12 fingers
M_{L1}	0.18 μm (L) x 2.5 μm (W) x 8 fingers
M_{FB}	0.18 μm (L) x 2.5 μm (W) x 15 fingers
M_{LFB}	0.18 μm (L) x 2.5 μm (W) x 15 fingers
M_2	0.18 μm (L) x 2.5 μm (W) x 15 fingers
$M_{1,CS1}$	0.18 μm (L) x 2.5 μm (W) x 2 fingers
$M_{2,CS1}$	0.18 μm (L) x 2.5 μm (W) x 2 fingers
$M_{3,CS1}$	0.18 μm (L) x 2.5 μm (W) x 8 fingers
$M_{B,CS1}$	0.18 μm (L) x 2.5 μm (W) x 12 fingers
$R_{B,CS1}$	200 Ω
$R_{1,CS1}$	2 kΩ
$M_{1,CS2}$	0.18 μm (L) x 2.5 μm (W) x 8 fingers
$M_{2,CS2}$	0.18 μm (L) x 2.5 μm (W) x 12 fingers
$M_{3,CS2}$	0.18 μm (L) x 2.5 μm (W) x 8 fingers
$M_{B,CS2}$	0.18 μm (L) x 2.5 μm (W) x 2 fingers
$R_{B,CS2}$	800 Ω
$R_{1,CS2}$	2 kΩ
L	3.878 nH (3.5 turns)

Table 4.3– RGC TIA component values

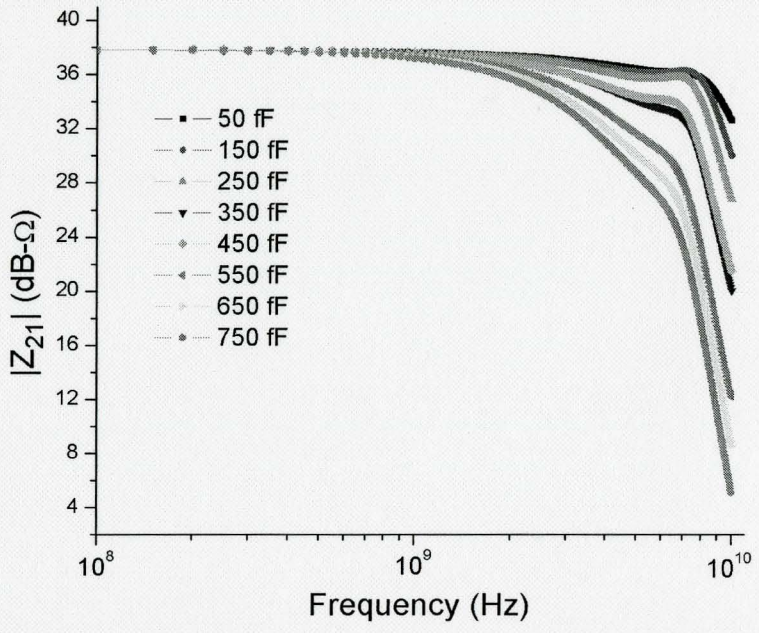


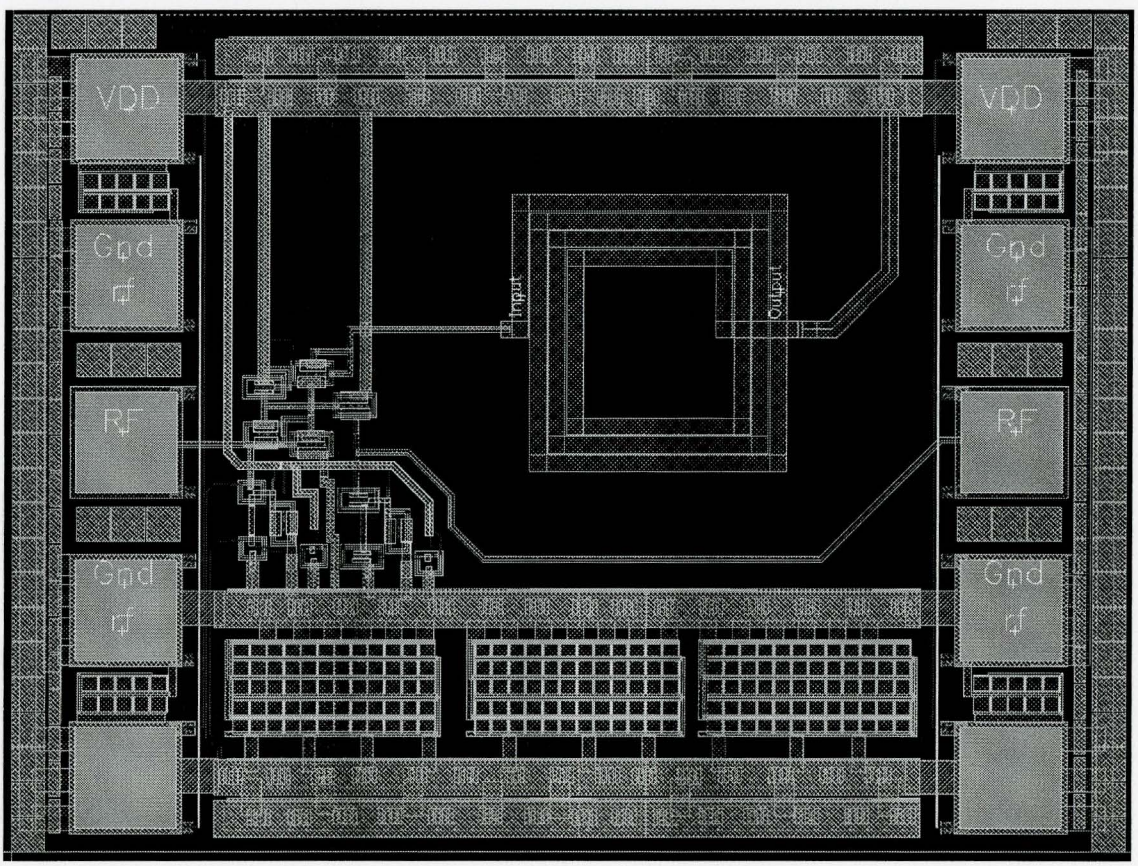
Figure 4.9 – Variation of transimpedance gain with input capacitance C_{PD}

Input Capacitance C_{PD}	3dB bandwidth f_{3dB}
50 fF	8.9 GHz
150 fF	8.3 GHz
250 fF	7.8 GHz
350 fF	4.1 GHz
450 fF	3.5 GHz
550 fF	3 GHz
650 fF	2.6 GHz
750 fF	2.4 GHz

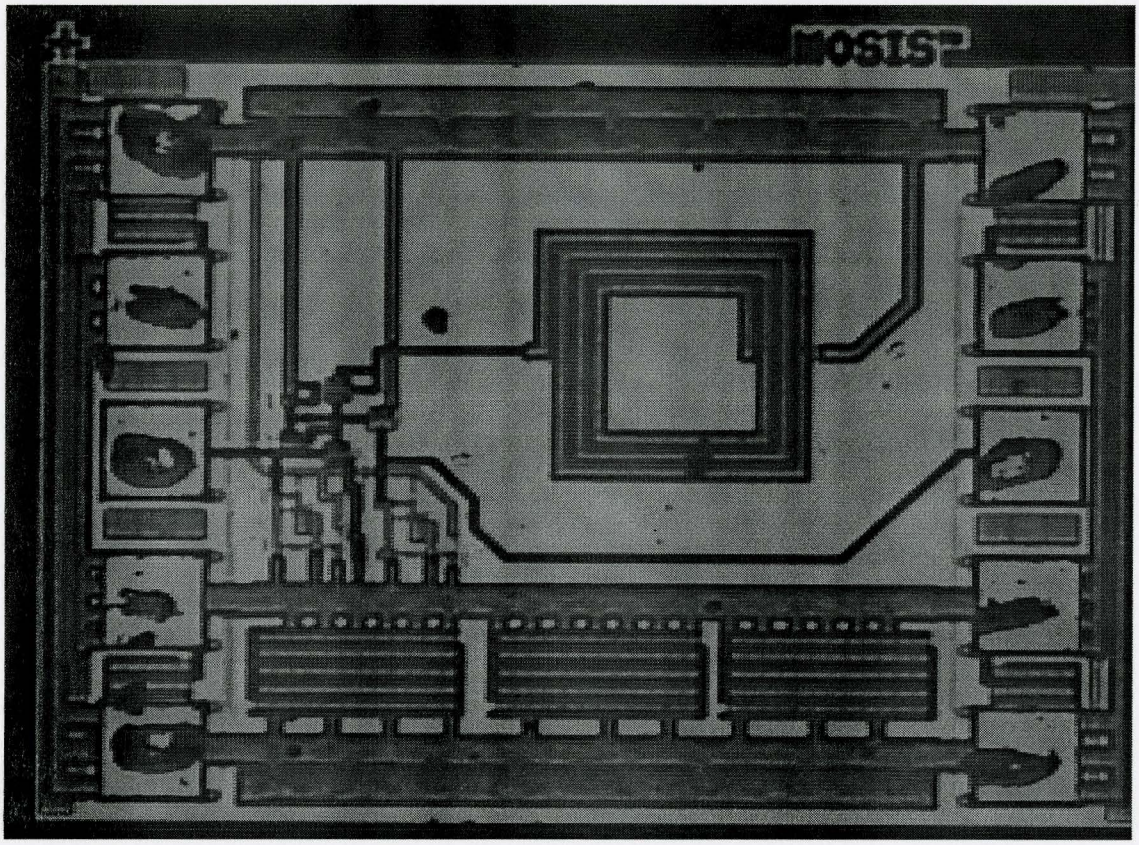
Table 4.4 – Summary of bandwidth variation

4.2 Layout and Measured Performance

The layout was done using the Virtuoso Layout Editor tool available in the Cadence 2004 software package. All major interconnections and inductors were laid out using the top thick metal-6 layer to minimize the parasitic capacitive and resistive effects. The 6 metal layers were connected in parallel to provide a ground and DC-supply with minimum resistive losses. Decoupling capacitors were placed between the VDD and GND pads, as well as between the VDD and GND supply lines, to provide a low-impedance path for the low-frequency supply noise to couple through. The RF pads utilized only the top thick metal-6 layer to reduce the parasitic capacitance. Figure 4.10a shows a screen capture of the RGC TIA chip layout, and figure 4.10b shows the photomicrograph of the fabricated chip. The total chip area is 1 x 0.781 mm^2.



a)



b)

Figure 4.10 – a) Layout screen capture b) Photomicrograph of the RGC TIA chip

The scattering parameters (S-parameters) and impedance parameters (Z-parameters) of the TIA were obtained for the nominal 1.8 V supply, as well as for a 3.0 V supply. The chip was probed on-wafer using a ground-signal-ground configuration for the RF signal, and a single pad was used for DC supply. The supply voltage on the VDD pad was provided by an HP4145B Semiconductor Parameter Analyzer. The S-parameters were measured using an Agilent 8722ES S-Parameter Analyzer, calibrated using Open-Short-Thru-Load method with a Picoprobe CS-5 calibration substrate. The measurement set-up is shown in figure 4.11. De-embedding was not performed because this would remove the shunt component of the input connections. This capacitance was included in the simulations and to model the effects of the parasitic input capacitance on the TIA

performance. The S-parameters were transformed to Z-parameters using the following equations:

$$Z_{11} = Z_0 \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \qquad Z_{12} = Z_0 \frac{2S_{12}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}}$$
$$Z_{21} = Z_0 \frac{2S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \qquad Z_{22} = Z_0 \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \tag{4.6}$$

The simulated and measured S-parameters and Z-parameters for the nominal 1.8 V supply are shown figures 4.12 and 4.13, respectively. The simulated and measured results for a 3.0 V are illustrated in figures 4.14 and 4.15.

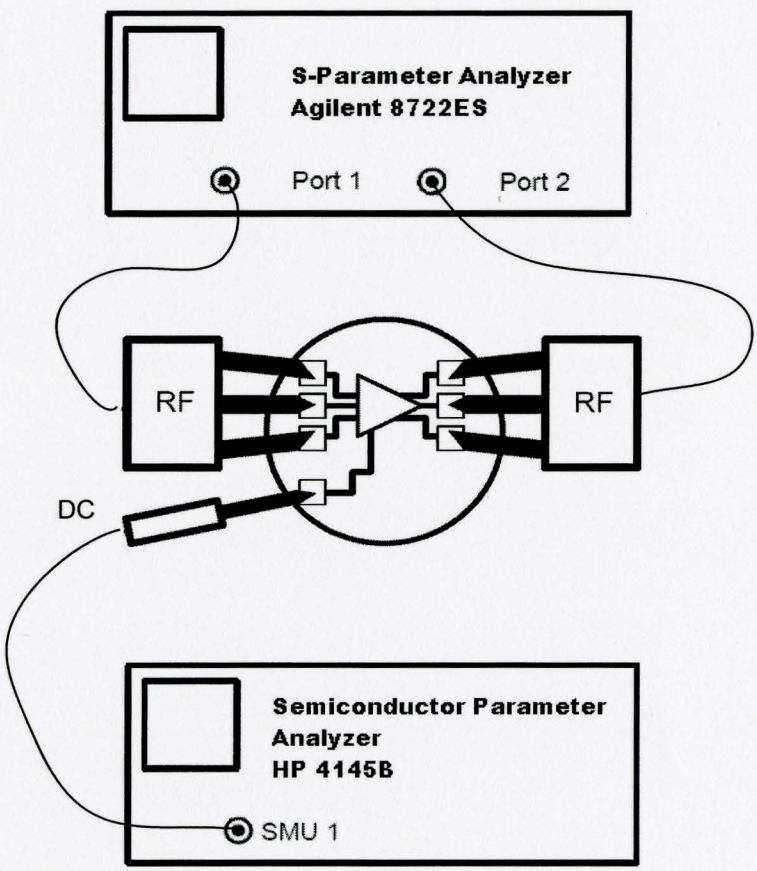


Figure 4.11 – Experimental setup for S-Parameter measurements

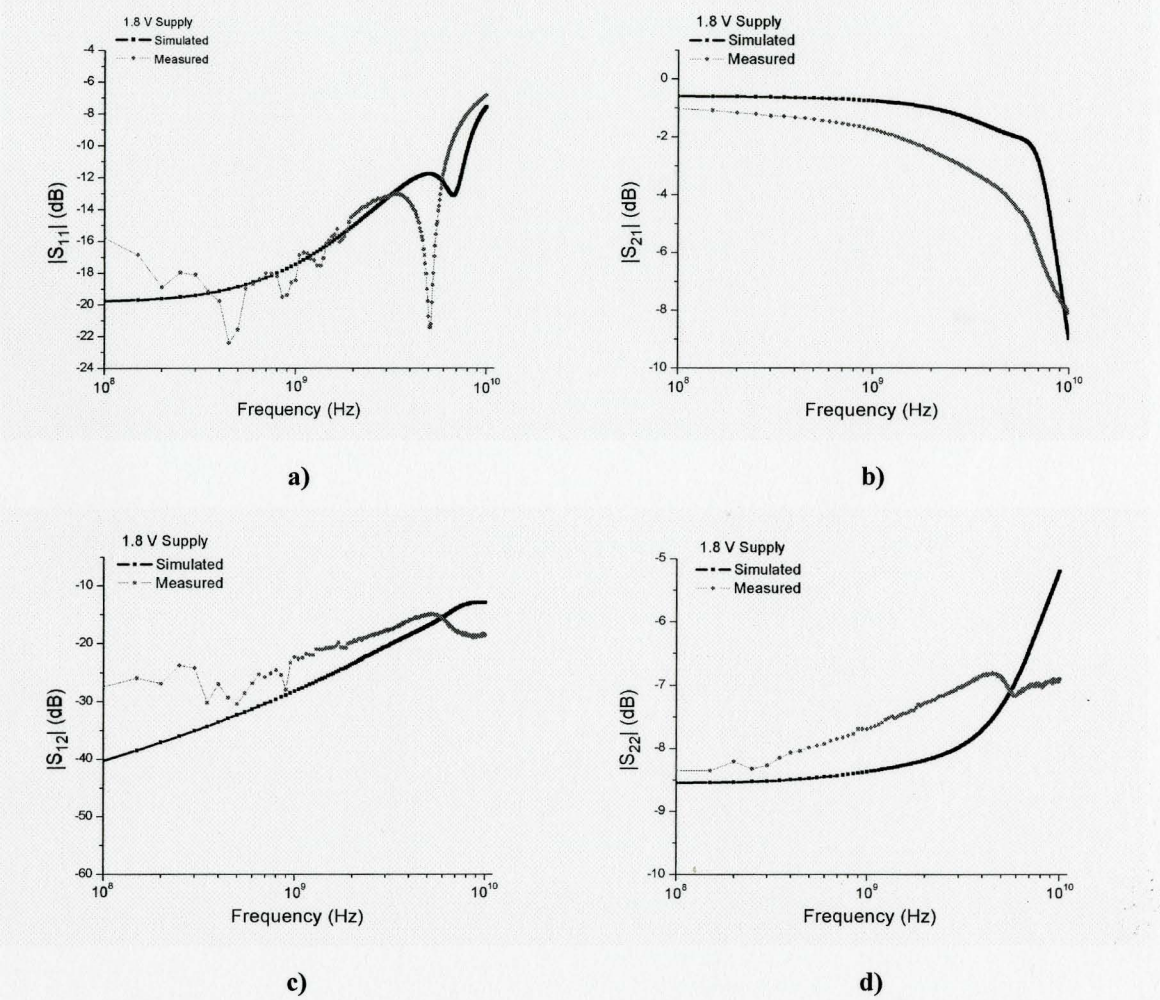


Figure 4.12 – Simulated and measured S-Parameters for a nominal 1.8 V supply a) S11 b) S21 c) S12 d) S22

The measured -3dB bandwidth of the TIA was approximately 7 GHz smaller than that of the inital simulations and the measured low-frequency transimpedance gain was 0.5 dB smaller. The high-frequency roll-off in the measured transimpedance gain is attributed to the parasitic capacitances of the interconnect lines as well as parasitic capacitive loading at the input node caused by the bond-pad and ESD capacitances. An 500 fF capacitor was added in the S-parameter simulations to account for the capacitance of the ESD diodes at each input pad, yielding a total input parasitic capacitance of approximately 650 fF. The simulation results reveal that the parasitic capacitance of the

input ESD pads contributed to the reduced bandwidth, since the circuit could only tolerate approximately 250 fF of capacitance at the input node.

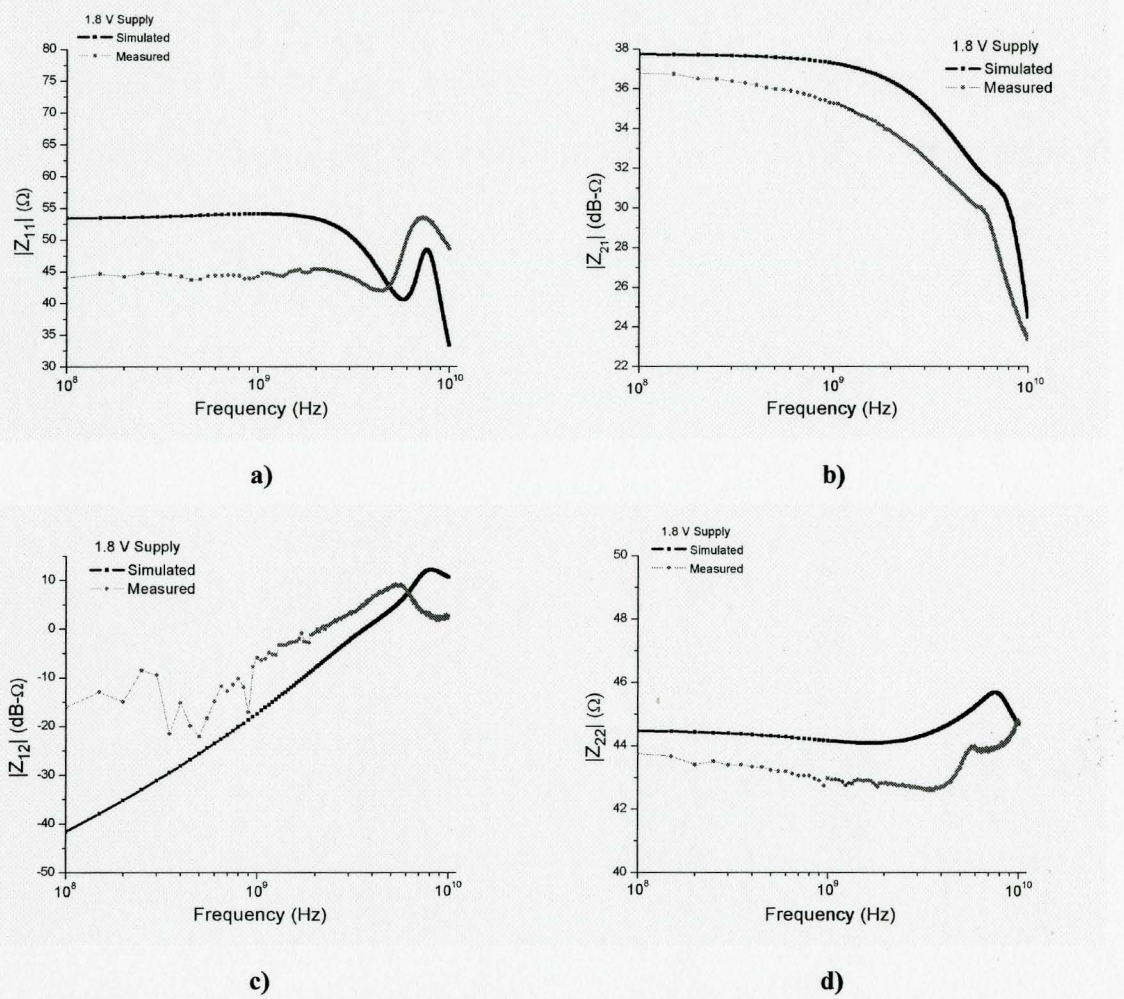


Figure 4.13 – Simulated and measured Z-Parameters for a nominal 1.8 V supply a) Z11 b) Z21 c) Z12 d) Z22

On account of these results, it is estimated that the total input parasitic capacitance of the laid out circuit was up to thee times higher than the initial estimate. Other layout issues such as crosstalk from the output line onto the power supply bus, as well as the magnetic coupling between the inductor and interconnect lines, affected the measured S22 and S12 parameters significantly. The resistive losses of the interconnect lines, particularly on the

long and thin interconnect line between the output buffer to the output pad also made a significant contribution in reducing the measured transimpedance by 1 dB.

The simulated and measured group delay is shown in figure 4.16. The simulated and measured group delay varies less than approximately 10 ps within the desired pass-band. As a result, the TIA generates negligible data-dependent jitter. Table 4.5 summarizes the measured performance of the RGC TIA chip.

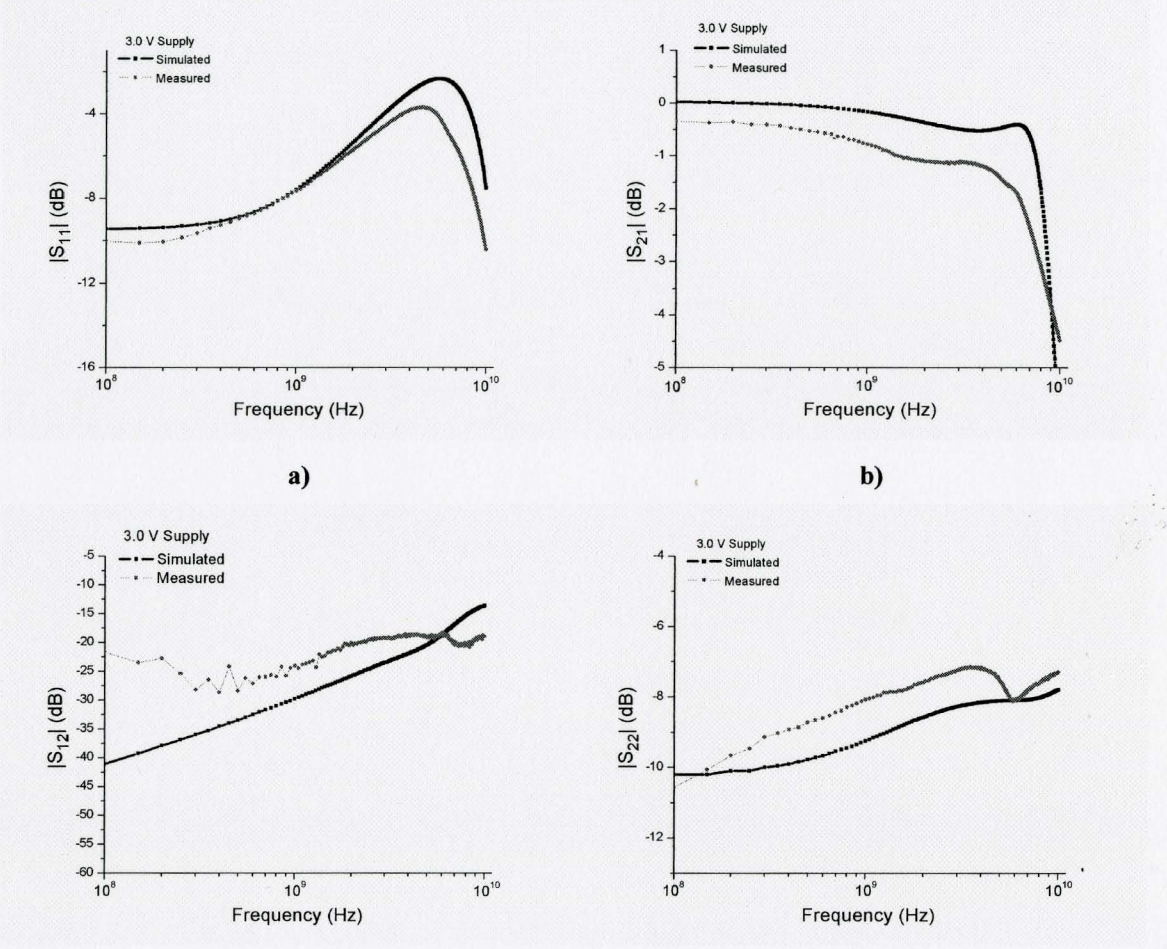


Figure 4.14 – Simulated and measured S-Parameters for a 3.0 V supply a) S11 b) S21 c) S12 d) S22

An Anritsu MP1763B Pulse Pattern Generator (PPG), an Agilent 83752A Synthesized Sweeper and an Anritusu MP1764A Error Detector were used to assess the

BER performance of the TIA chip for a random sequence of digital data at various bit-rates. Furthermore, an Agilent Infinium DCA 86100A Wideband Oscilloscope was used to measure the eye-diagrams. The test set-up is illustrated in figure 4.19. Since the TIA chip was tested using voltage signals, the DC voltage gain of the circuit was measured in order to determine the optimal output offset voltage for the PPG. As shown in figure 4.18, the circuit exhibits a linear voltage gain profile for an input voltage between 0.4 V to 0.8 V. Therefore, an output offset voltage of 0.65 V was chosen for the PPG.

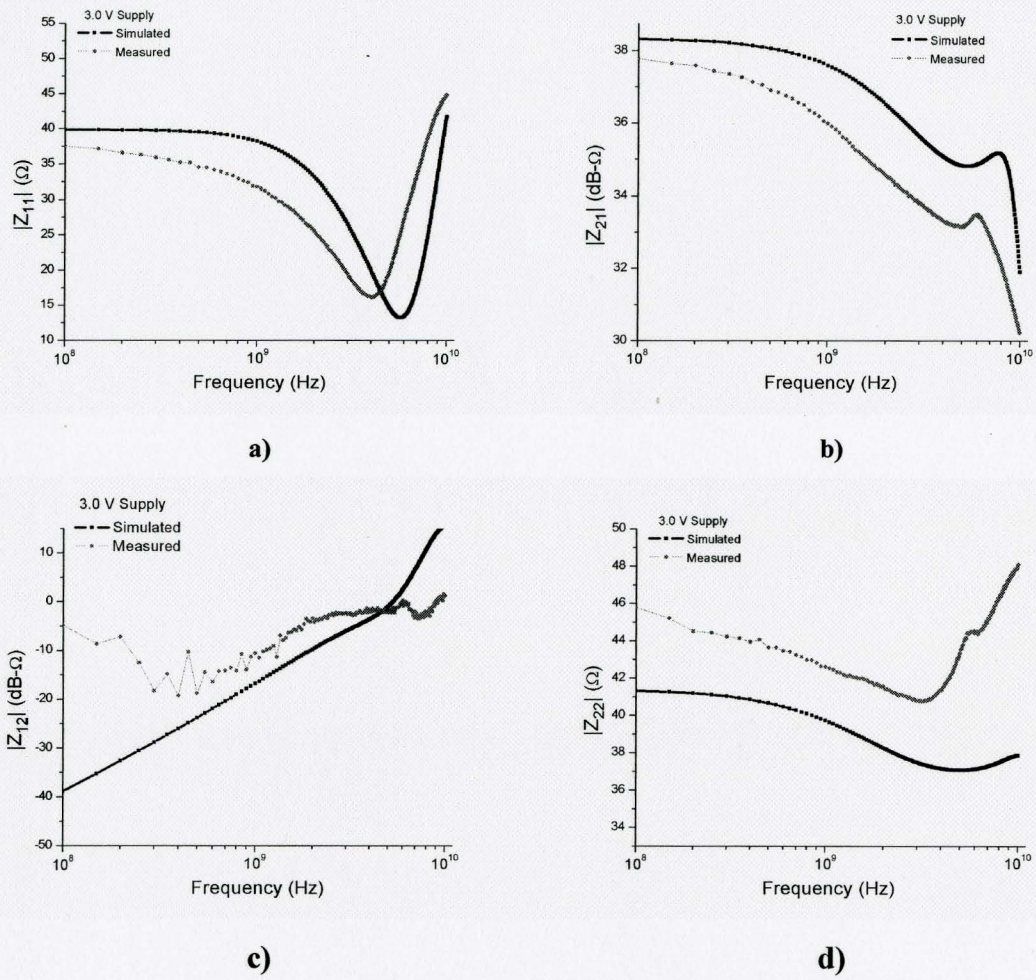


Figure 4.15 – Simulated and measured Z-Parameters for a 3.0 V supply
a) Z11 b) Z21 c) Z12 d) Z22

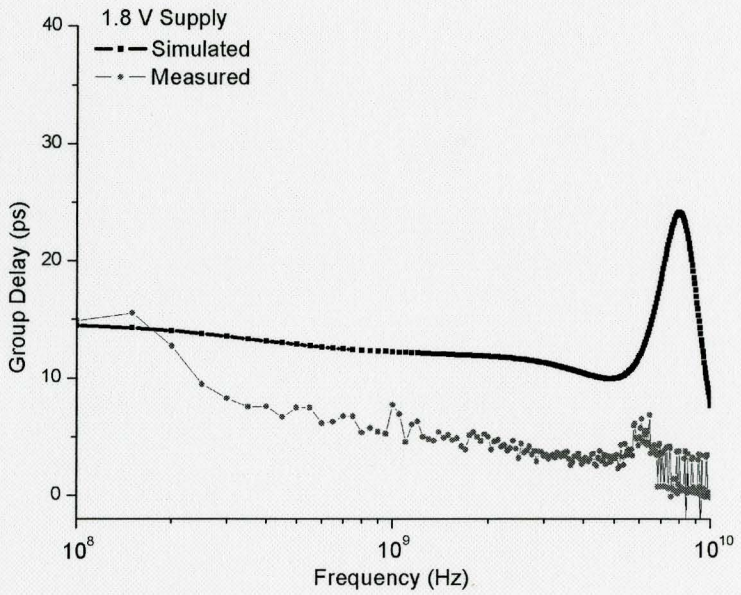


Figure 4.16 – Simulated and measured group delay response for a nominal 1.8 V supply

3dB Bandwidth	1.5 GHz
Transimpedance	37 dBΩ
Group Delay Variation	< 10 ps
Noise	29 pA/$\sqrt{Hz}$
Power consumption	23.76 mW (1.8 V)
Input Reflection Coeff.	< -14 dB
Output Reflection Coeff.	< -7 dB
Chip Area	1 x 0.781 mm^2

Table 4.5– Measured Performance Summary

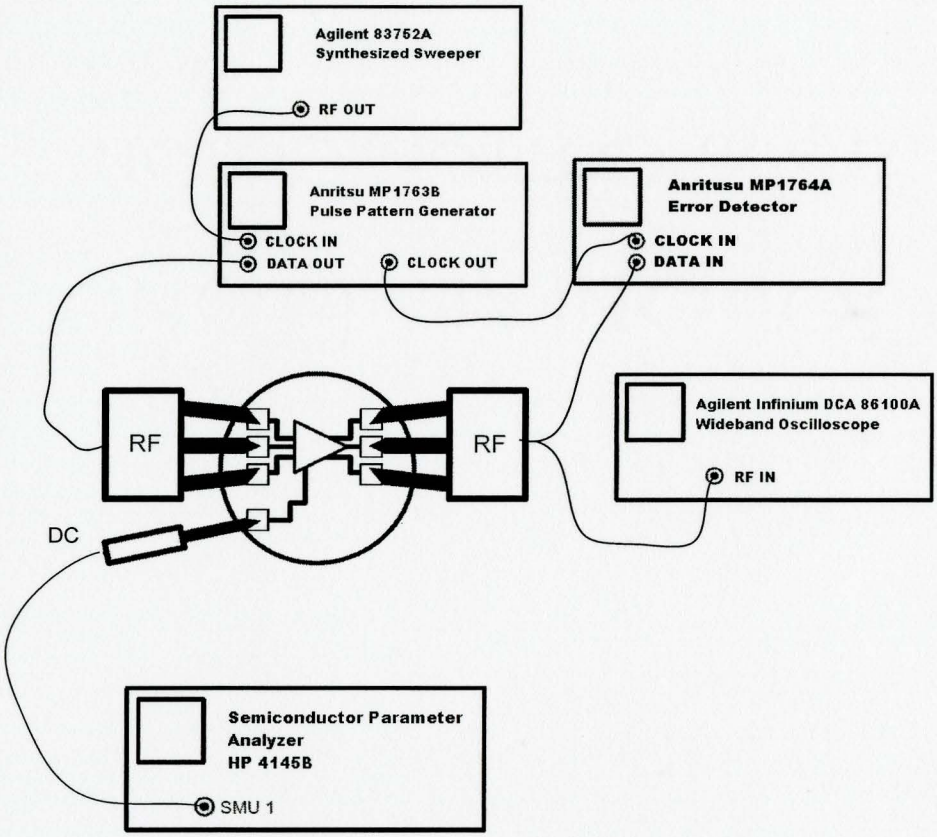


Figure 4.17 – BER and eye diagram measurement set-up

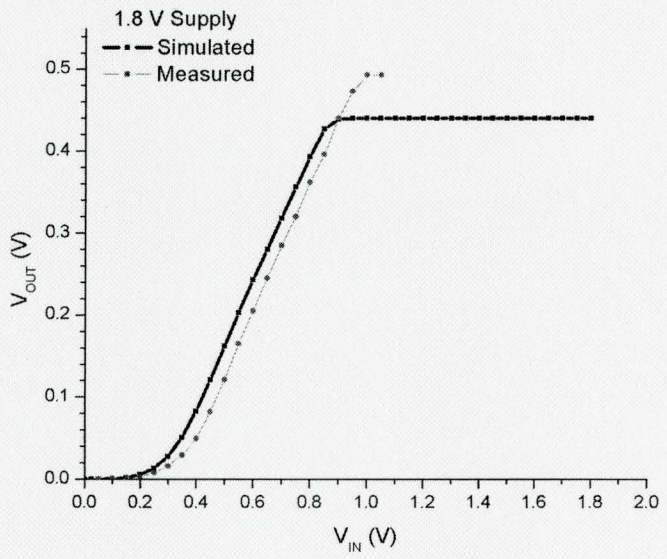


Figure 4.18 – Simulated and measured voltage transfer characteristic

A 1 Gbps NRZ pseudo-random bit sequence (PRBS) of length $2^{11}-1$, generated by the PPG (figure 4.20a), was input directly into the RGC TIA chip, and the output signal from the TIA chip was fed into the oscilloscope to obtain the output eye-diagram, as well as into the error detector to measure the BER performance. The measured eye diagrams reveal that the rise and fall times of the NRZ signal are increased due to the low-pass frequency response of the TIA. In addition, the eye diagrams reveal that the TIA adds very little noise, jitter and ISI to the output signal. The rms jitter measured by the scope was 6 ps. The eye closure due to ISI at the center of the bit interval was less than 35%. The BER performance was measured by sending a $2^{11}-1$ PRBS test pattern with varying amplitudes levels from the PPG through the RGC TIA chip and measuring the resulting BER from the BER analyzer. The smallest average input voltage signal which obtained a BER of 10^{-12} was found to be $\overline{V}_{IN}=1.52\,\text{mV}$, which corresponds to an electrical sensitivity of $i_{sens}^{pp}=16\,\mu\text{A}$. The optical sensitivity was obtained assuming a photodetector with a responsivity of 0.8 A/W, resulting in an optical sensitivity of $\overline{P}_{sens}=\frac{16\times10^{-6}}{2\cdot0.8}=10\,\mu\text{W}=-20\,\text{dBm}$. Equation (2.78) was used in conjunction with the measured BER data to obtain an estimate of 29 $\text{pA}/\sqrt{\text{Hz}}$ for the average input-referred rms noise current of the RGC TIA, with an assumed photodiode repsonsivity of 0.8 A/W. The measured BER performance of the RGC TIA chip is summarized in figure 4.19, along with a comparison to the sensitivity performance of several other recently published CMOS TIA circuits.

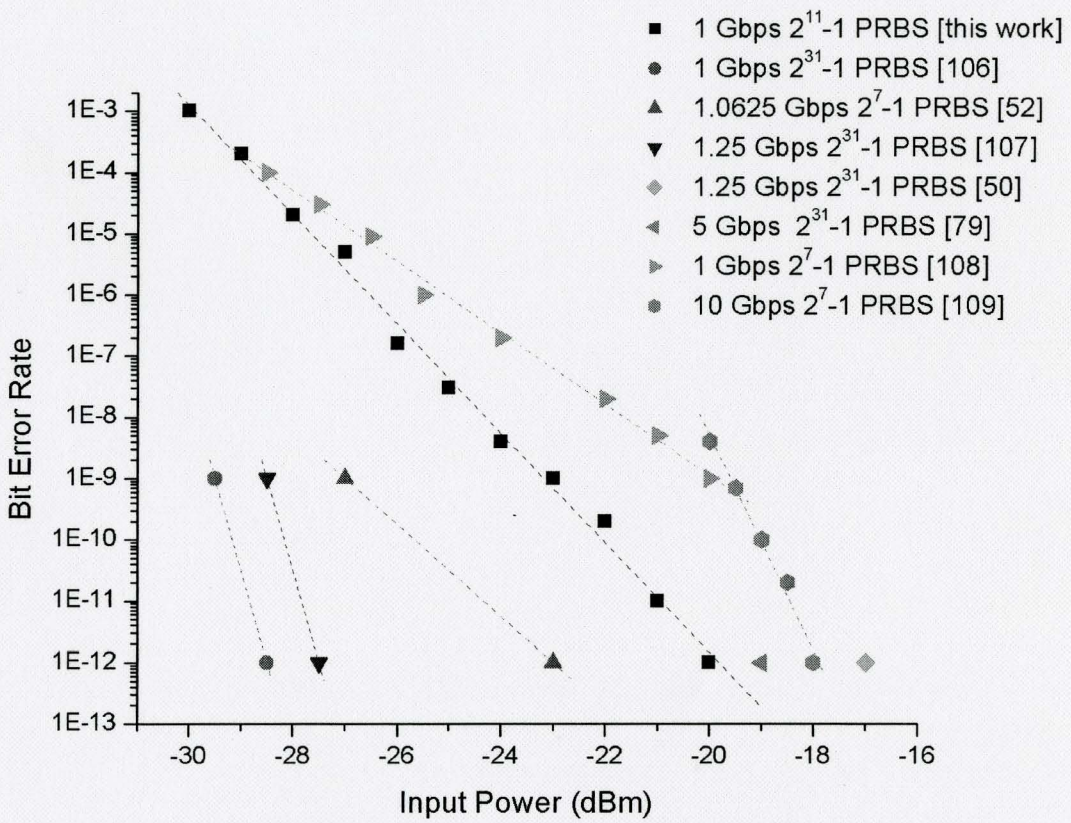
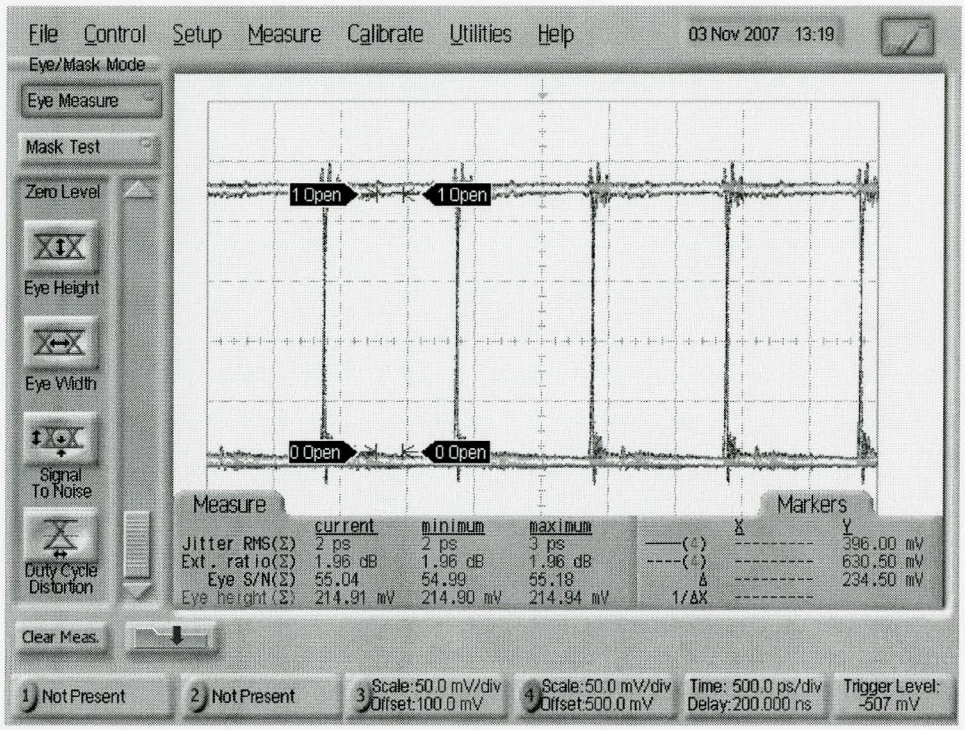
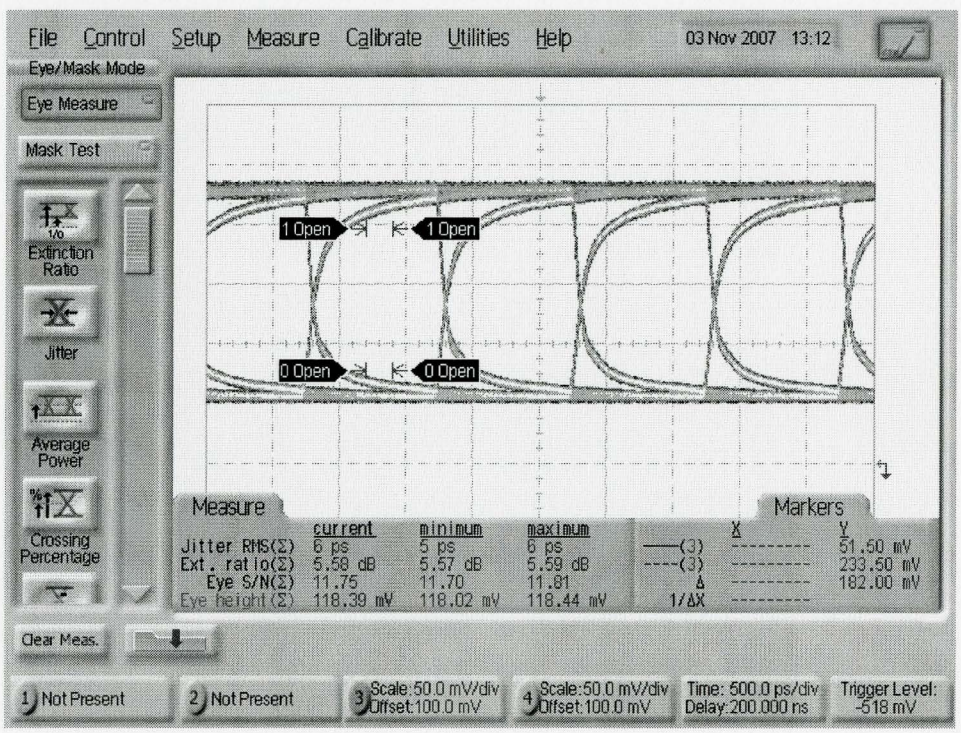


Figure 4.19 – Measured BER performance versus optical sensitivity for the RGC TIA chip and comparison to published results

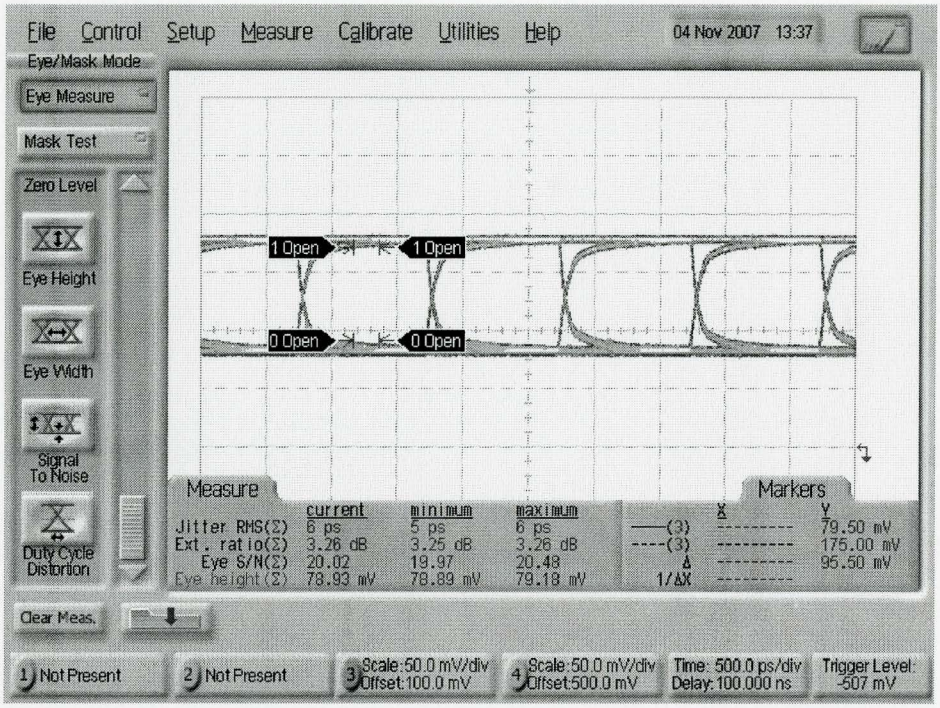
Figure 4.21 shows the eye diagrams for 10 Gbps data. The eye-closure is more severe because of the high-frequency gain roll-off, and also because of the impedance mismatch which becomes more severe at high-frequencies. At a 1.8 V supply, the BER of a 10 Gbps signal could no longer be maintained at 10^{-12}. The supply was increased to 3.0 V to achieve reliable operation at 10 Gbps. Using a 3.0 V supply, the eye closure is less severe (fig. 4.21c) since the high-frequency gain is approximately 4 – 6 dB larger than with a 1.8 V supply, and the TIA was able to achieve a measured BER below 10^{-12}. However, this comes at the cost of a reduced output voltage swing, because the drain voltage of the output stage increases. Increasing the supply voltage additionaly implies a greater power consumption, and reduced device life-time.


File Control Setup Measure Calibrate Utilities Help
03 Nov 2007 13:19
Eye/Mask Mode
Eye Measure
Mask Test
Zero Level
Eye Height
Eye Width
Signal To Noise
Duty Cycle Distortion
Clear Meas.
1 Open
1 Open
0 Open
0 Open
Measure
Markers
current minimum maximum
Jitter RMS(Σ) 2 ps 2 ps 3 ps
Ext. ratio(Σ) 1.96 dB 1.96 dB 1.96 dB
Eye S/N(Σ) 55.04 54.99 55.18
Eye height(Σ) 214.91 mV 214.90 mV 214.94 mV
X Y
(4) 396.00 mV
(4) 630.50 mV
Δ 234.50 mV
1/ΔX
1) Not Present
2) Not Present
3) Scale:50.0 mV/div Offset:100.0 mV
4) Scale:50.0 mV/div Offset:500.0 mV
Time: 500.0 ps/div Delay:200.000 ns
Trigger Level: -507 mV

a)

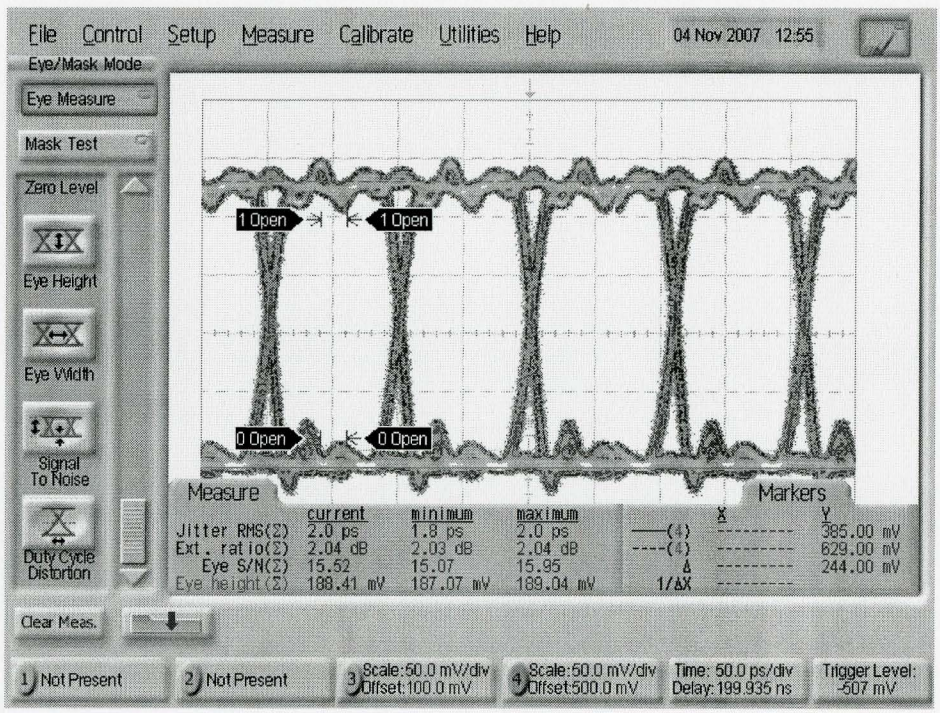

File Control Setup Measure Calibrate Utilities Help
03 Nov 2007 13:12
Eye/Mask Mode
Eye Measure
Mask Test
Extinction Ratio
Jitter
Average Power
Crossing Percentage
Clear Meas.
1 Open
1 Open
0 Open
0 Open
Measure
Markers
current minimum maximum
Jitter RMS(Σ) 6 ps 5 ps 6 ps
Ext. ratio(Σ) 5.58 dB 5.57 dB 5.59 dB
Eye S/N(Σ) 11.75 11.70 11.81
Eye height(Σ) 118.39 mV 118.02 mV 118.44 mV
X Y
(3) 51.50 mV
(3) 233.50 mV
Δ 182.00 mV
1/ΔX
1) Not Present
2) Not Present
3) Scale:50.0 mV/div Offset:100.0 mV
4) Scale:50.0 mV/div Offset:100.0 mV
Time: 500.0 ps/div Delay:200.000 ns
Trigger Level: -518 mV

b)

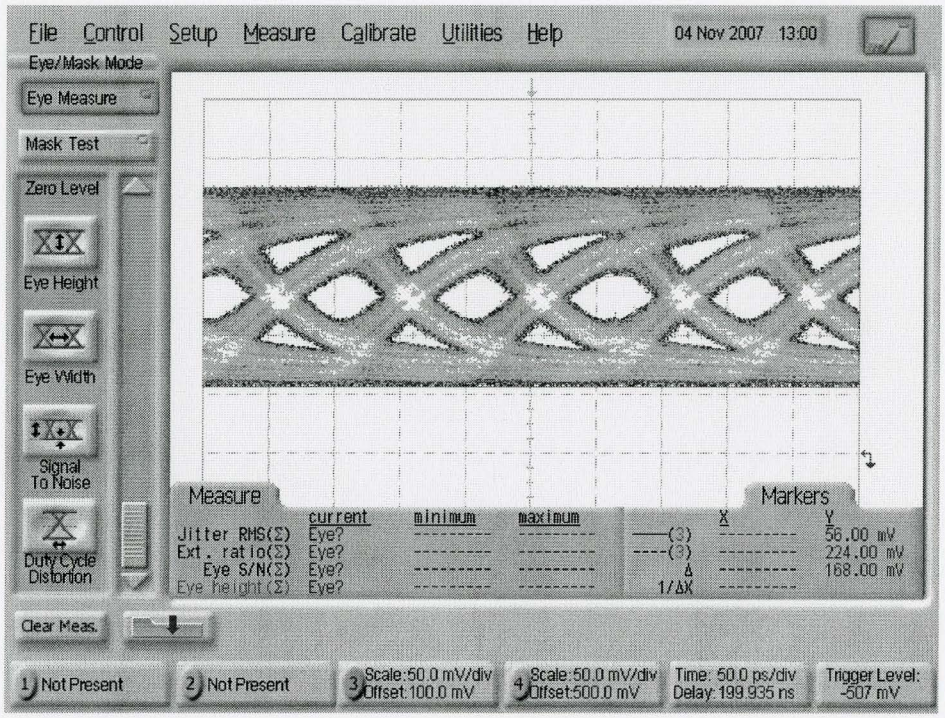


c)

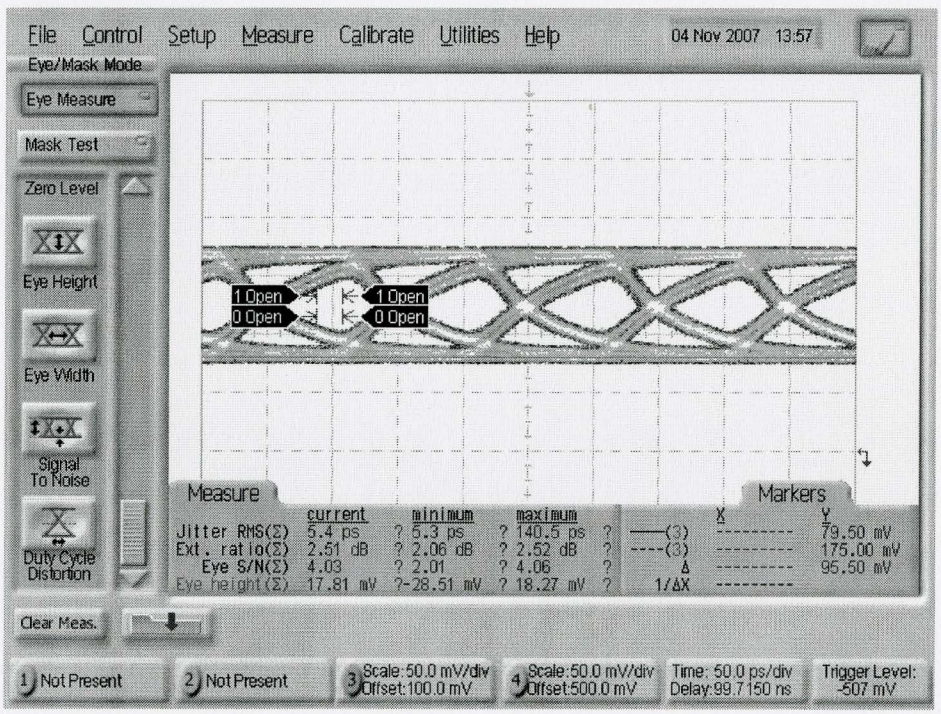
Figure 4.20 – a) Eye-diagram of 1 Gbps signal NRZ PRBS input signal b) Eye-diagram of TIA output using a 1.8 V supply and c) 3.0 V supply



a)



b)



c)

Figure 4.21 – a) Eye-diagram of 10 Gbps NRZ PRBS input signal b) Eye-diagram of TIA output using a 1.8 V supply and c) 3.0 V supply

Chapter 5
OPTICAL TRANSMITTER PROTOTYPE DESIGN & IMPLEMENTATION

As fiber communication systems continue to move closer into the home, equipment manufacturers are being driven more than ever to obtain improvement in operating speed, transmission distance, and power consumption, all at a reduced cost. Implementation of these improvements for optical transmitters calls for faster edge speeds, increased driving currents, and lower supply voltages. Reducing the power-supply requirements into a single 3.3 V supply is one obvious way to significantly improve the overall power dissipation of any system. However, high current requirements, fast switching capability, parasitic inductances of the packaged laser all work against achieving the 3.3 V goal. Therefore, finding a low-cost laser that operates properly with a single 3.3 V supply, while still meeting the stringent signal integrity requirements typical of high-speed optical transmitters is a difficult challenge. This chapter presents the implementation of a optical transceiver module built with commercial off-the-shelf components that overcomes these challenges while providing a low-cost solution.

5.1 Laser Driver

The laser diode chosen for this prototype was the SLT4210 Series 1.31 μm InGaAsP/InP MQW-DFB laser diode module, manufactured by Sumitomo Electric. The laser diode is mounted into a TO package and integrated with an InGaAs monitor photodiode, and a single mode fiber pigtail. These modules are ideally suitable for long reach and intermediate reach applications with a data rate up to 2.5 Gbps (i.e. PON). They exhibit good spectral characteristics for high-speed direct modulation such as a maximum spectral width of 1 nm, minimum MSR of 30 dB and maximum rise/fall times of 0.1 ns.

At room temperature, the threshold current is typically 10 mA and the minimum slope efficiency is 0.075 mW/mA.

The primary function of a laser driver is to provide appropriate current for bias and modulation of the laser diode. The bias current is the constant current that pushes the laser diode beyond its threshold value and into the linear region (eq. 2.45), and the modulation current switches on and off in synchronization with the input voltage waveform. Ideally, the bias current should track the changes in threshold current and the modulation current should track the changes in slope efficiency (e.g. due to temperature variations). Also, the modulator must meet the jitter generation specifications outlined by the optical communication standards (i.e. GPON). For these reasons the laser driver chosen for the optical transmitter module was the MAXIM MAX3863 2.7 Gbps Laser Driver with Modulation Compensation IC [104]. The MAX3863 can operate a laser diode from a single 3.3 V supply. It was designed for directly modulating a laser diode up to data rates of 2.7 Gbps. The laser driver can modulate laser diodes at amplitudes up to 80 mA, and can supply a bias current up to 100 mA. External resistors can set the bias and modulation levels. The MAX3863 modulation output is optimized for driving a 25 Ω load. Typical (20% - 80%) edge speeds are 50 ps. In addition to providing enough drive capability, it contains a fully integrated automatic power control (APC) loop for maintaining the laser bias current over temperature variations and lifetime of the laser, and modulation compensation is available to increase the modulation current in proportion to the bias current, thereby maintaining the extinction ratio. The MAX3863 accepts differential CML clock and data inputs with on-chip 50 Ω termination resistors. If a clock signal is available, an input data-retiming latch can be used to re-time the input data thereby rejecting any input pattern-dependent jitter. The laser driver chip is fabricated with Maxim's in-house second-generation SiGe process, and packaged in a 32-pin quad, flat, no-lead (QFN) package. The functional diagram of the chip is shown in figure 5.1.

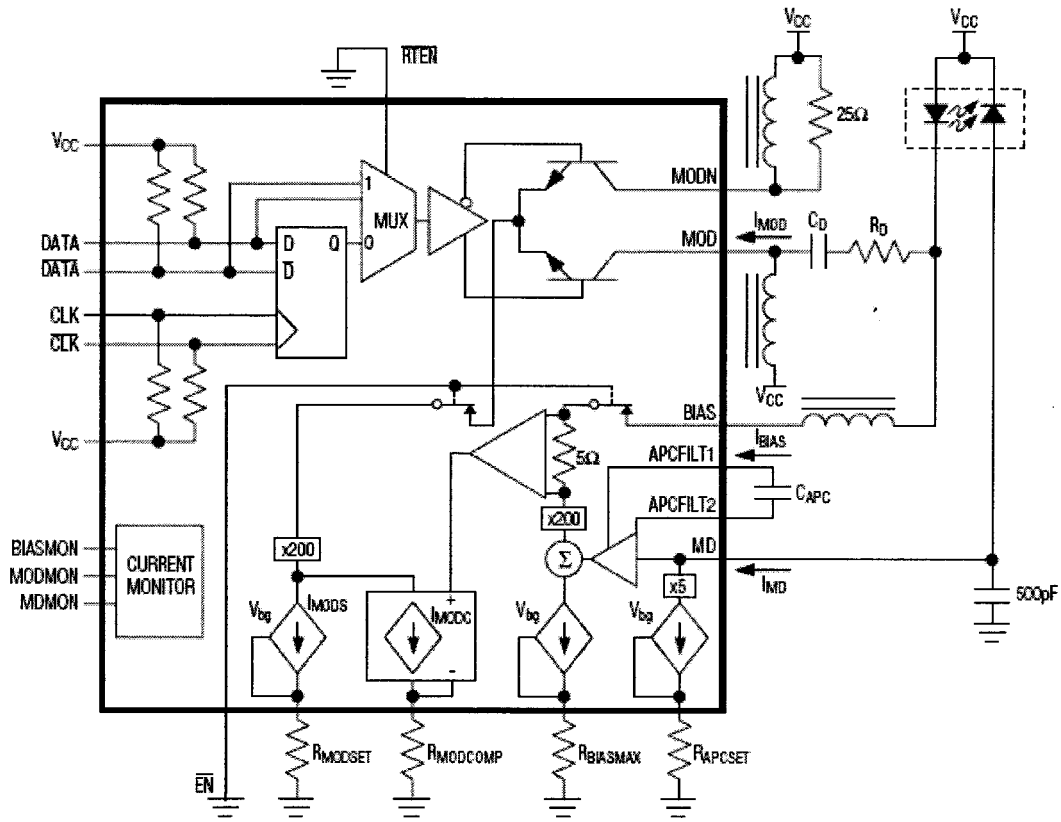


Figure 5.1 – Functional diagram of MAX3863 laser driver IC. Driver chip is AC coupled to laser diode [104]

5.2 Laser Driver Interface

The design of the interface between the laser driver circuitry and the commercially available laser diodes at high data rates is affected by the output circuit of the laser driver, the supply voltage, and the electrical characteristics of the laser diode. DC coupling between the driver and the laser provides a simple and straightforward interface solution, as illustrated in figure 5.2a. Typical long-wavelength laser diodes require a forward-bias voltage on the order of 1.2 – 1.8 V. This forward-bias requirement is the sum of the bandgap voltage and the voltage drop across the equivalent series resistance of the laser diode (e.g. figure 5.2b) The equivalent series resistance R_S is typically 4 – 6 Ω. To

interface with the laser diode, a series damping resistor R_M of 20 Ω is necessary to achieve the required load condition, assuming a laser diode resistance of 5 Ω.

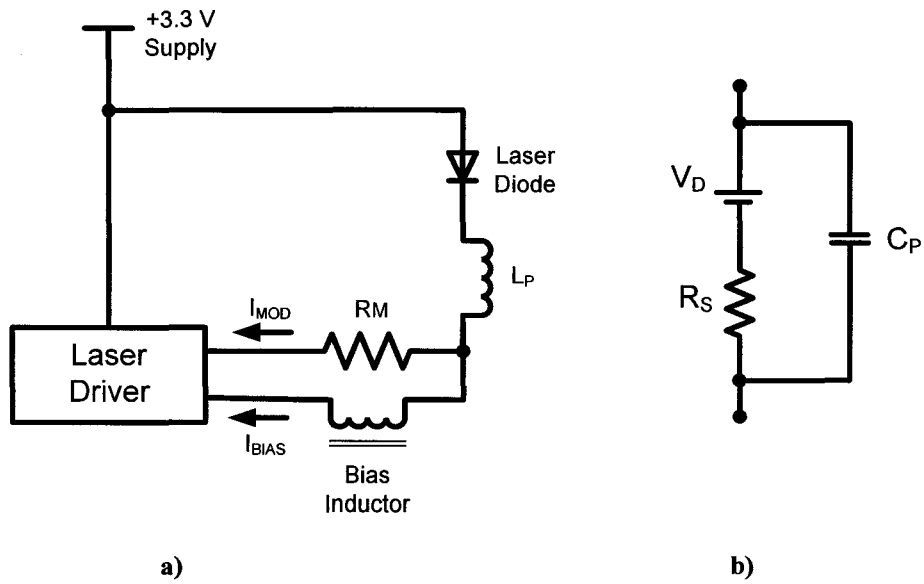


Figure 5.2 – a) DC coupled laser b) Laser diode equivalent circuit model

The voltage drop across the laser diode V_L, as well as the transient voltage drop V_T resulting from the parasitic inductance of the laser package, L_P, and the voltage drop V_M across the damping resistor, R_M, may not provide enough voltage headroom for fast switching of the transistors in the output stage of the driving circuit as the supply is decreased to 3.3 V. To illustrate this, a package parasitic inductance of 1.5 nH, and a 25 mA modulation current with 20% to 80% edge-speed of 50 ps (for 1 Gbps data) are assumed, resulting in a voltage drop of $V_T = L(\Delta I / \Delta t) = 0.45$ V during switching transients. With a 20 Ω matching resistor and a forward-bias voltage of $V_L = 1.6$ V required for the laser diode, the voltage at the driver output can be as low as 0.75 V, which causes the output switching transistors fall out of saturation, making high-speed operation with a 3.3 V supply very difficult [103].

The headroom problem can be improved by AC coupling the driver to the laser diode. Shown in figure 5.3, this can be accomplished by adding a series capacitor, C_D, and pull-up inductor, L_P.

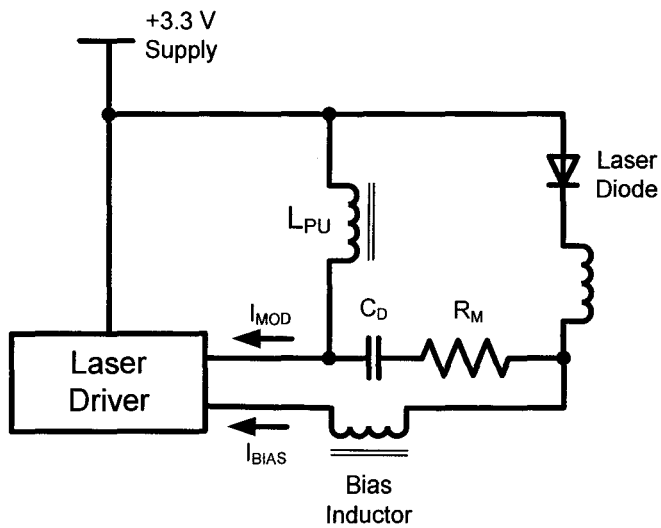


Figure 5.3 – AC coupled laser

The pull-up inductor is required to keep the output driver transistors properly biased. With AC coupling, the AC voltage drop across the laser diode is only a function of the voltage drop across its equivalent series resistance and not the band-gap, and the voltage drop V_M across the matching resistor is equal to one-half the modulation current I_{MOD} times R_M. This is because the average current through C_D must have an average value of zero and a peak-to-peak current swing of I_{MOD}. To satisfy these conditions, one-half of I_{MOD} must flow into C_D from the laser during an optical high output, and one-half of I_{MOD} must flow out of C_D into the laser during an optical low output. The total current through the laser is equal to the sum of the currents flowing out of the cathode, therefore the laser current is $I_{BIAS} + I_{MOD}/2$ during an optical high output, and $I_{BIAS} - I_{MOD}/2$ during an optical low output. Repeating the headroom calculation for the AC coupled laser, the voltage at the output at the driver output is $V_S - I_{MOD}R_M/2 - V_T - I_{MOD}R_S = 2.475\ \text{V}$, allowing sufficient headroom and therefore

enabling fast current switching in the driver output stage even with a 3.3 V supply [20]. Based on these considerations, the laser diode was AC coupled to the driver circuit.

There are several disadvantages of the AC coupling scheme. First of all, additional discrete components are required. Because these components are in the high-speed signal path, they can cause distortions. Therefore, the use of good high-frequency PCB layout techniques is critical. In addition, the AC coupling capacitor C_D added into the signal path introduces a low-frequency cutoff to the system, which can impact the pattern-dependent jitter performance of the system. Typical specifications for random NRZ data require that the system maintains a low BER for a certain number of consecutive ONEs or ZEROs. Thus, the AC coupling capacitor C_D must not affect the low-frequency content of the data stream. To reduce the pattern-dependent jitter caused by long strings of consecutive identical bits, the value of C_D should be as large as possible [20].

Since the total current of $I_{BIAS} + I_{MOD}$ must flow through both the laser diode and the parasitic package inductance, the transient behavior of the switching current affects the optical output of the laser diode in a number of ways due to the parasitic inductor. Successful multi-Gbps optical transmitter design therefore demands for careful attention to the sources and effects of parasitic inductance that affect the transient response. Since the circuit of figure 5.3 forms a second-order RLC circuit, basic control theory can be used to predict overshoot and ringing that result from an overdamped or underdamped RLC circuit. To eliminate overshoot and ringing, the circuit must be at least critically damped, where $R_{critical} = 2\sqrt{L/C} < R_M$. Thus, as the parasitic inductance increases, the critical resistance must also increase, at the cost of greater mismatch between the laser and driver and reduced voltage headroom [103].

5.3 PCB Layout

Figure 5.4 shows the complete interface schematic which was implemented on the PCB. The values for each of the critical elements (e.g. bypass capacitors, biasing inductors) of the circuit were selected according to the data sheet specifications provided by the chip manufacturer [104]. The schematic entry and PCB layout was performed using the Protel DXP software package. To minimize parasitic inductances, connections between the chip output and the laser diode were made as short as possible and a multilayer board with uninterrupted group plane was used to minimize EMI and crosstalk. The exposed ground pad on the 32-pin QFN, which provides a low thermal resistive path for heat removal from the IC, was soldered to the circuit board ground for proper thermal and electrical performance of the laser driver chip.

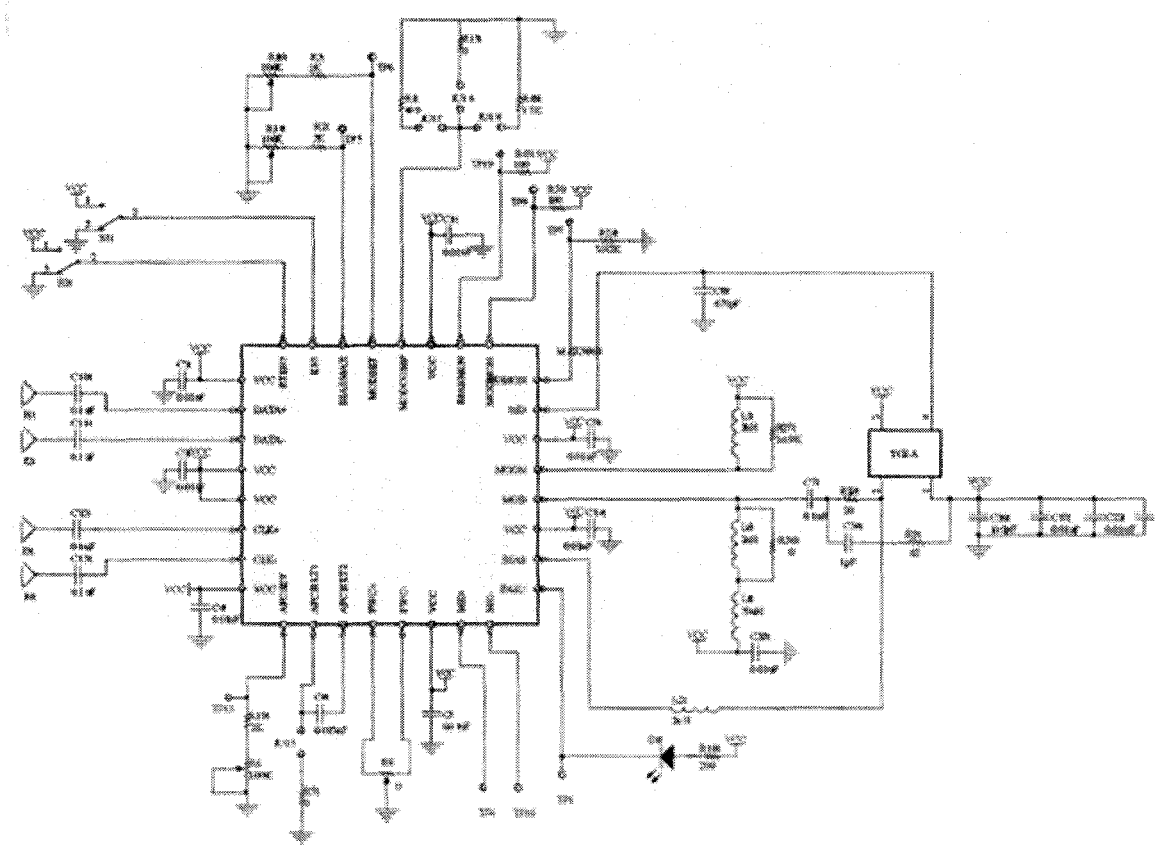


Figure 5.4 – Schematic diagram of optical transmitter PCB prototype [104]

At high-frequencies, the physical dimensions of printed circuit board traces become significant relative to the wavelength of the signal. Therefore, controlled impedance transmission lines were used to route the clock and data signals, as well as for the output to the laser. Microstrip transmission lines were chosen due to their simplicity of design, and their ease of fabrication. The characteristic impedance of the microstrip traces on the PCB were obtained from empirical formulas provided by the software, in terms of the thickness and the width of the trace, and its distance from the return path (ground plane).

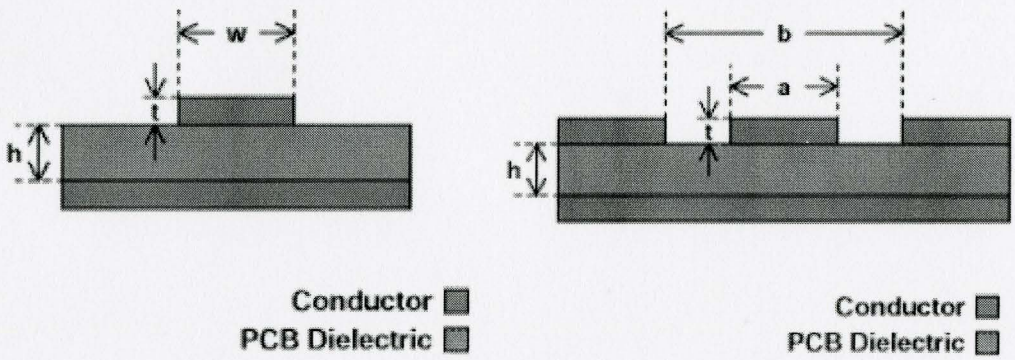


Figure 5.5 – Microstrip waveguide & coplanar waveguide with ground

The primary disadvantage of microstrip transmission lines are that some of the energy transmitted may be coupled into space or onto adjacent traces. Also, microstrip traces can be inconvenient when connecting to other components; for example, when connecting a device pin that has a specific pin size, the trace width may turn out to be too wide to fit between pins. Because maintaining a tight control of the waveguide characteristics is critical in RF PCB layouts, coplanar waveguides are useful alternative to keep the impedance of the line constant while tapering the signal conductor's width down to meet a pin. A close-up of the laser driver interface is shown in figure 5.6. The PCB was fabricated using a standard low-cost FR4 PCB material. The high-speed data and clock traces were carefully designed to have the same electrical length and symmetry on the PCB. The test equipment was connected via subminiature version A (SMA) connectors

on the edge of the board. The total cost of the board was under $400 USD, with the laser diode counting for over 50% of the cost. The complete PCB design is shown in figure 5.7.

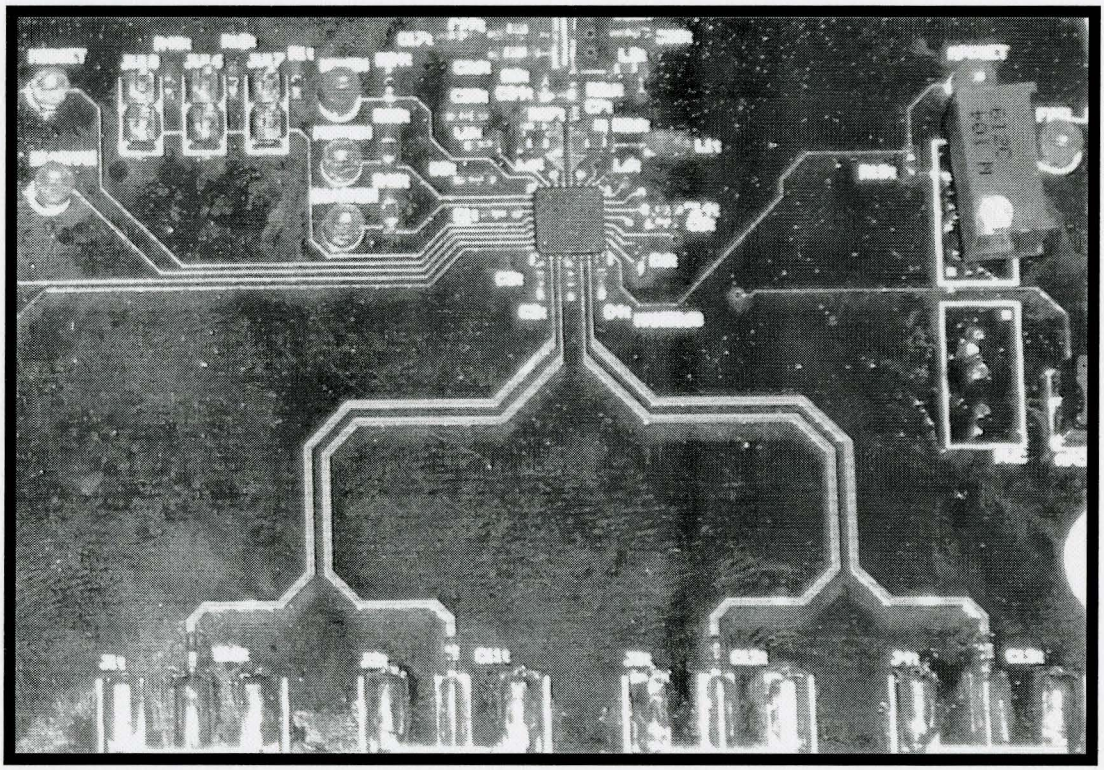


Figure 5.6 – Close-up of laser driver interface revealing the design of the input transmission lines.

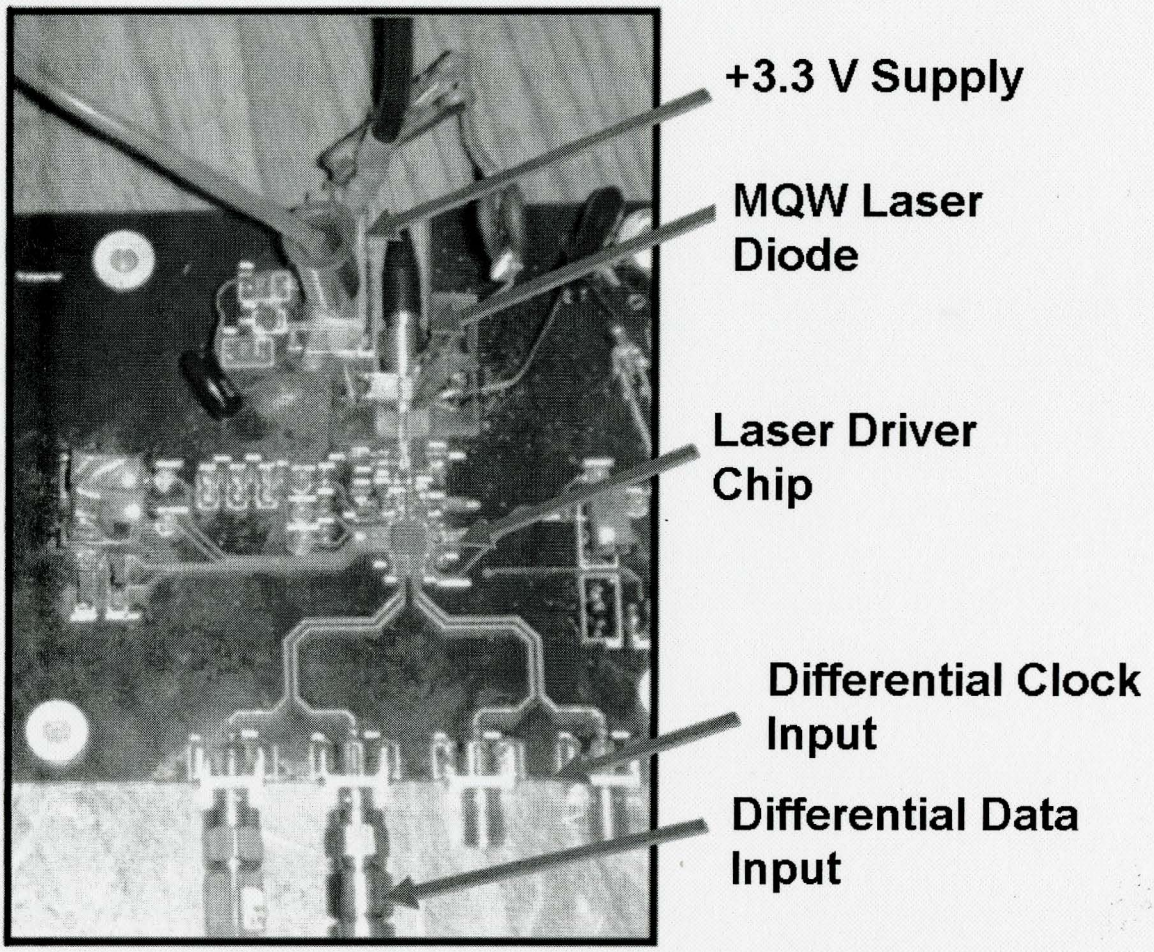
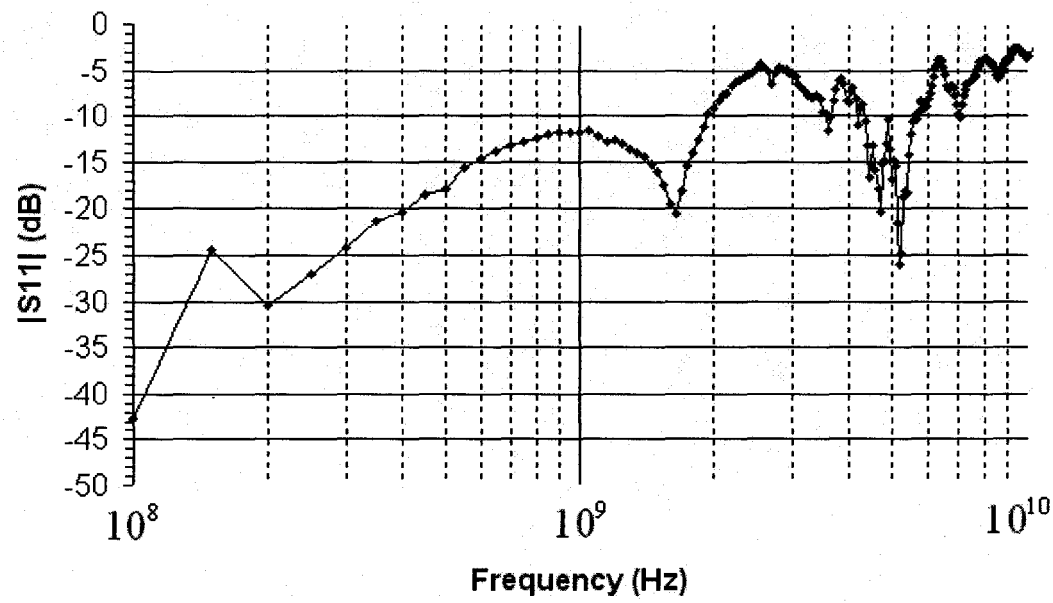


Figure 5.7 – Complete view of optical transmitter prototype board

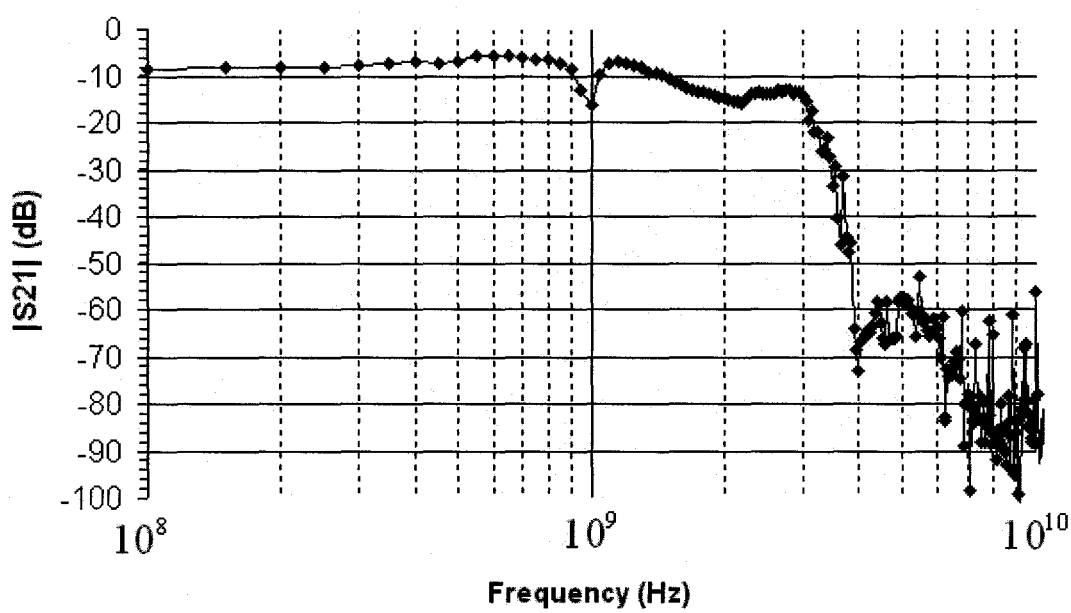
5.4 Measured Results

Reflection coefficients of the optical transmitter board were measured by an Agilent 8703B Lightwave Component Analyzer. The measured forward reflection coefficient (fig. 5.8a) shows that the input ports are well matched to 50 Ω up to approximately 2 GHz. The measured forward transmission coefficient, (i.e. conversion efficiency) reveals that the laser can be modulated reliably up to approximately 1 GHz. The notches in the reflection and transmission coefficients are due to the resonances caused by the parasitic inductances in the laser driver interface. The optical spectrum of

the laser was measured with an Agilent 86142B Optical Spectrum Analyzer. This measurement confirms the spectral purity of the optical source under continuous wave



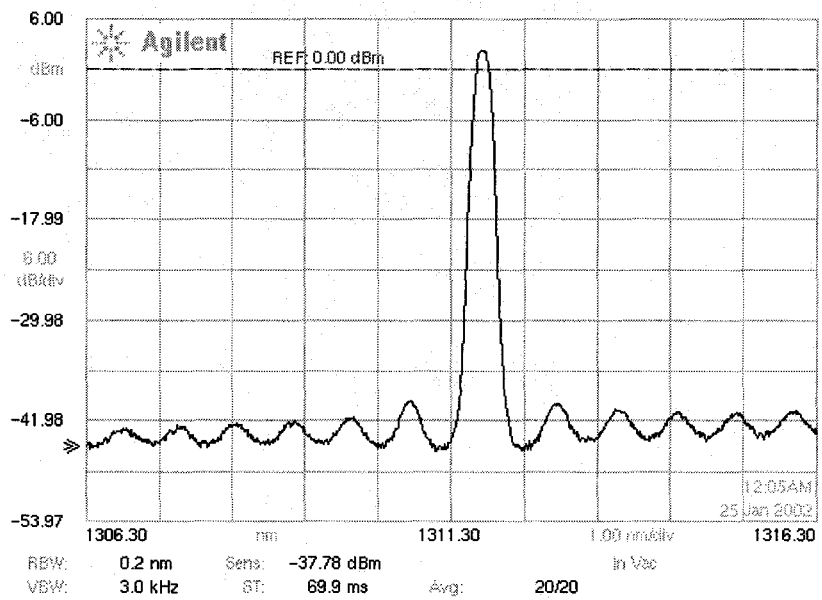
a)



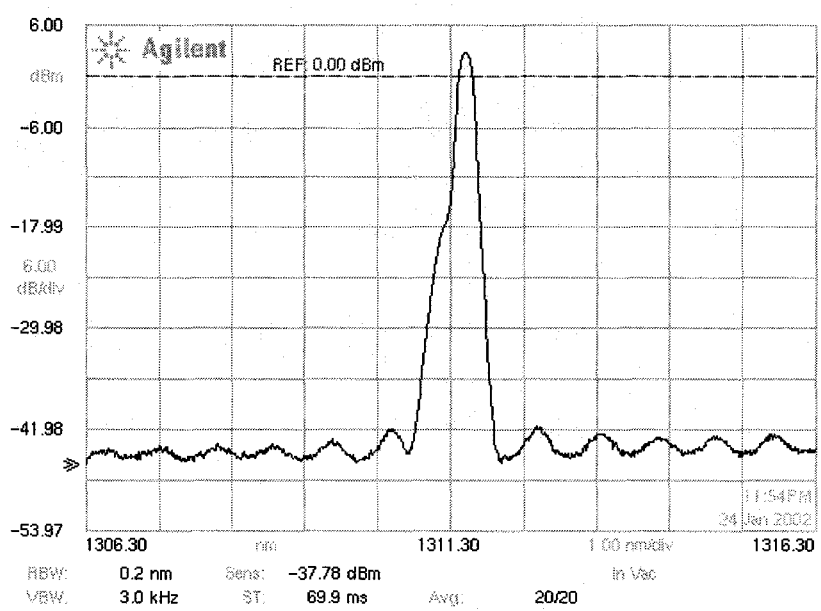
b)

Figure 5.8 – Measured a) forward reflection and b) forward transmission coefficients of the optical transmitter board

operation (fig. 5.9a), showing that the spectral line width is approximately 1 nm, and under direct modulation (fig. 5.9b), where the spectral width broadened slightly due to frequency chirp. The measurement also confirms that the side-mode suppression ratio of the laser is indeed greater than 30 dB as specified.



a)



b)

Figure 5.9 – Optical spectrum (intensity vs. wavelength) of the MQW laser under a) CW operation b) direct modulation

The measurement set-up of an optical transmission experiment over 10 km of SMF is illustrated in figure 5.10. A Discovery Semiconductor DSC-R402 10 GHz optical-to-electrical (O/E) converter was used to convert the optical signal to the electrical (voltage) domain. The measured responsivity and the conversion efficiency of the O/E converter were 0.8 A/W and 350 V/W, respectively. The BER measurements were performed using an Anritsu MP1763B 12.5 GHz pattern generator and an Anritsu MP1764A 12.5 GHz error detector. An Agilent Infinium DCA 86100A Wideband Oscilloscope was used to obtain the eye diagrams. For all eye diagrams and BER measurements, a PRBS of length $2^{11}-1$ was used.

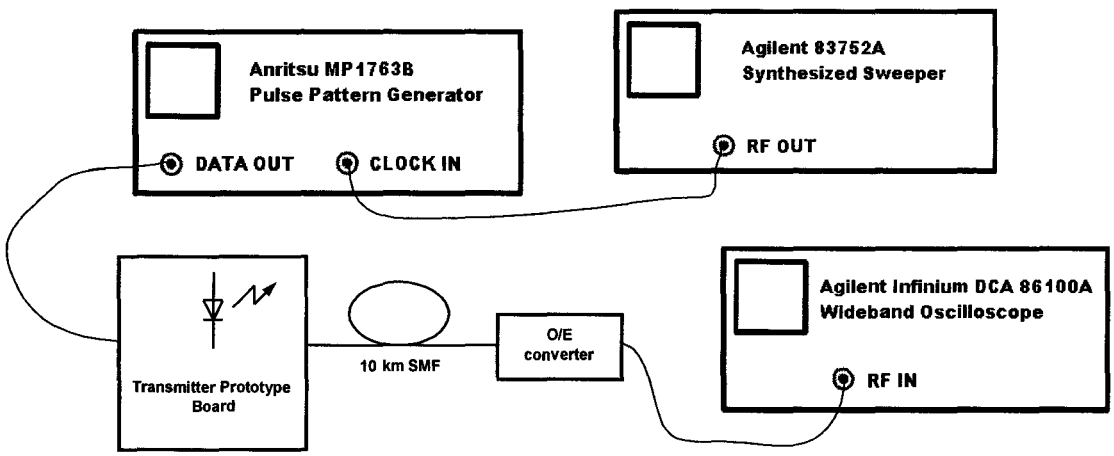
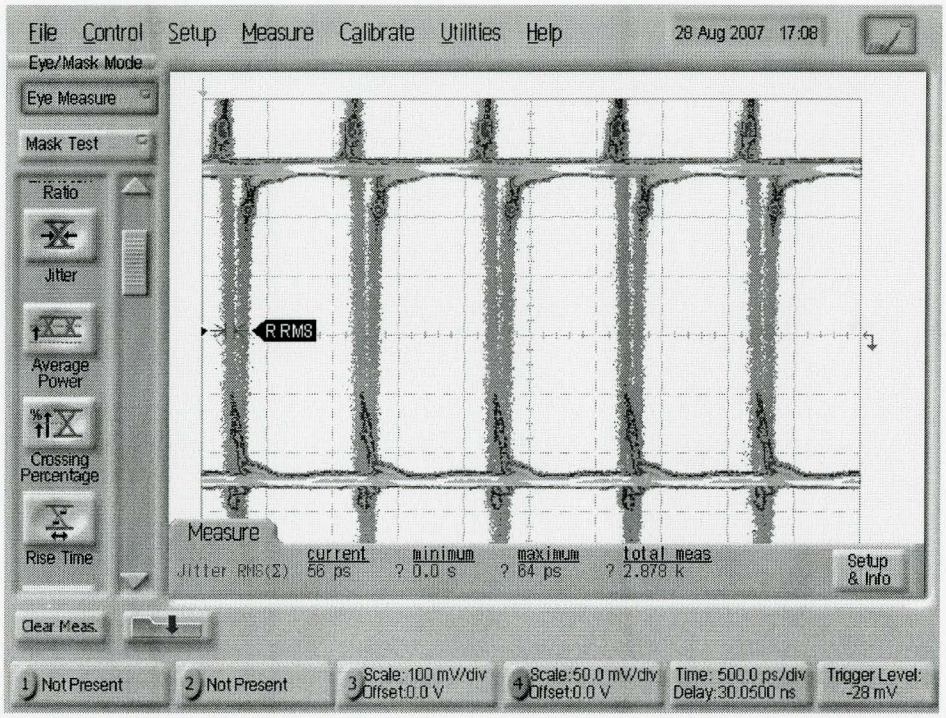
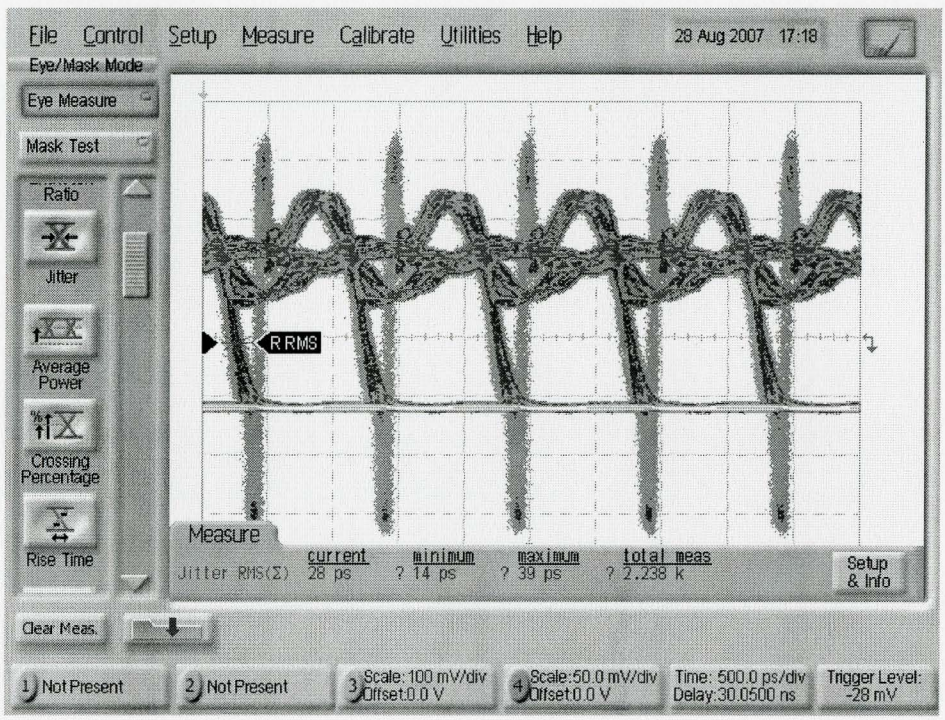


Figure 5.10 – Optical transmitter prototype board test set-up

The eye diagrams of a 1 Gbps signal obtained from the transmitter prototype board, both before and after passing through the fiber, are shown in figure 5.11. The values of the bias current and modulation current were 12 mA and 22.5 mA, respectively. With equations (2.55) and (2.59), the corresponding extinction ratio was found to be 8 dB. The total power consumption of the transmitter was 702 mW. The measured BER for the PRBS was 10^{-12} both before and after transmission through the fiber. The eye diagram of the signal generated by the transmitter prototype board shows considerable overshoot and undershoot at the rising and falling pulse edges.



a)

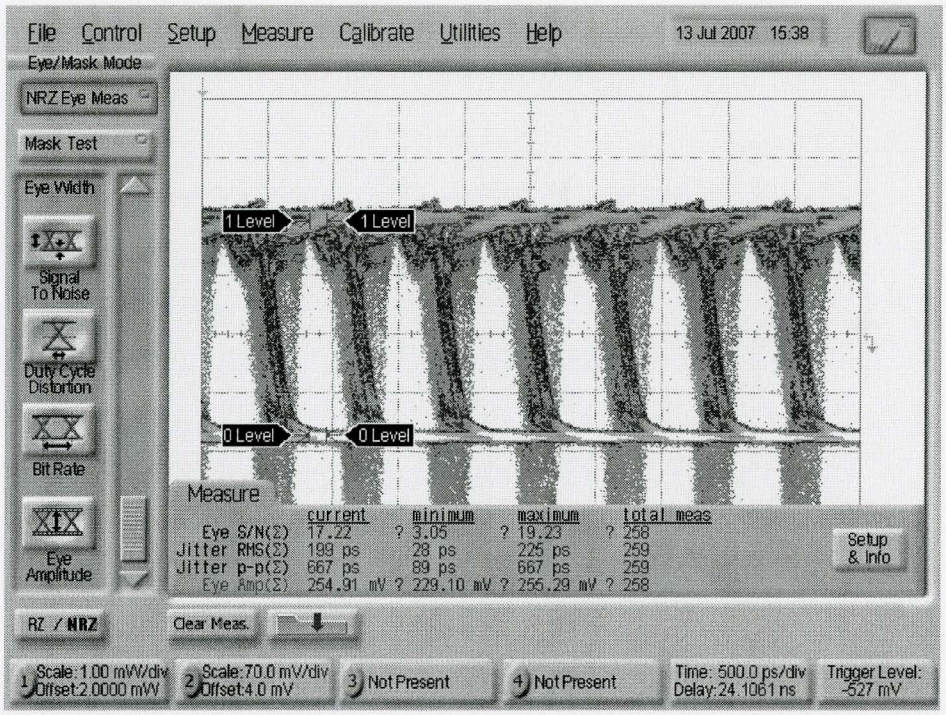


b)

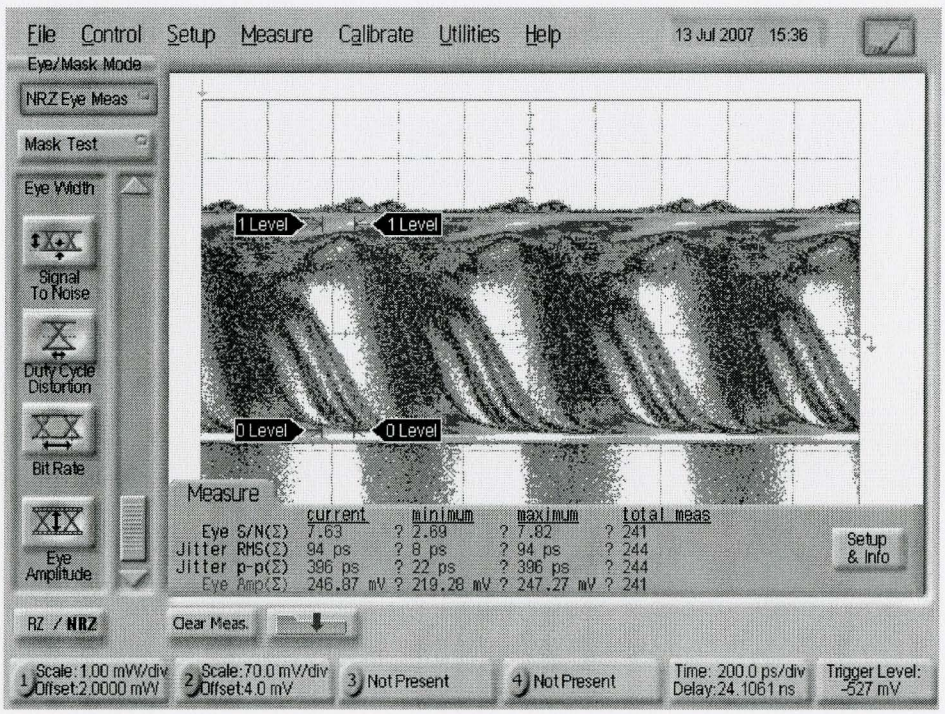
Figure 5.11 – Eye diagrams of 10 km optical transmission experiment a) Eye diagram from optical transmitter prototype board before transmission through fiber b) after transmission through fiber

The fast rising edges of the data signal from the PPG, together with the parasitic inductances of the laser diode package and laser driver interface caused the overshoot/undershoot to occur. In particular, the ferrite beads (used as the pull-up inductors) had an excessively high Q factor, which greatly contributed to the overshoot. A potential solution to this problem would be to adjust the value of the series damping resistor, or to decrease the value of resistor in parallel with the pull-up inductor to damp out the overshoot. After passing through 10 km of optical fiber, the eye diagram showed considerable ISI due to the combined effects of fiber dispersion and laser chirp. However, a BER of 10^{-12} was still obtainable through the manual adjustment of the decision threshold voltage and sampling time of the BER analyzer. The corresponding eye closure power penalty due to the ISI can be obtained by measuring the amount of eye closure attributed to ISI in the eye diagram of figure 5.11b. The launched optical power was -1.29 dBm, with a corresponding eye opening of 260 mV. After passing through 10 km of fiber (assuming a loss of 0.4 dB/km at the 1330 nm wavelength), the received optical power was approximately -5.45 dBm with a corresponding eye opening of 103.3 mV. The eye diagram of figure 5.11b exhibits an eye closure of approximately 25 mV. This closure corresponds to a power penalty of approximately 0.16 dB due to ISI.

Eye diagrams for 1.5 and 2 Gbps signals from the output of the optical transmitter prototype board are shown in figure 5.12. As the data rate was increased, the total jitter increased considerably and the eye closure became more severe. This degradation is due to reflections caused by mismatch in the transmission lines at higher frequencies. Figure 5.13 summarizes the BER performance of the optical transmitter PCB for different bit rates. The maximum bit rate is 1.25 Gbps at BER of 10^{-9}. As the bit-rate is increased beyond this point, the accumulation of data dependent jitter and random jitter causes severe eye closure, resulting in an increase of the BER. Figure 5.13 summarizes the BER performance of the optical transmitter prototype board as a function of bit-rate.



a)



b)

Figure 5.12 – Eye diagrams of PRBS from optical transmitter prototype board a) 1.5 Gbps signal b) 2 Gbps signal

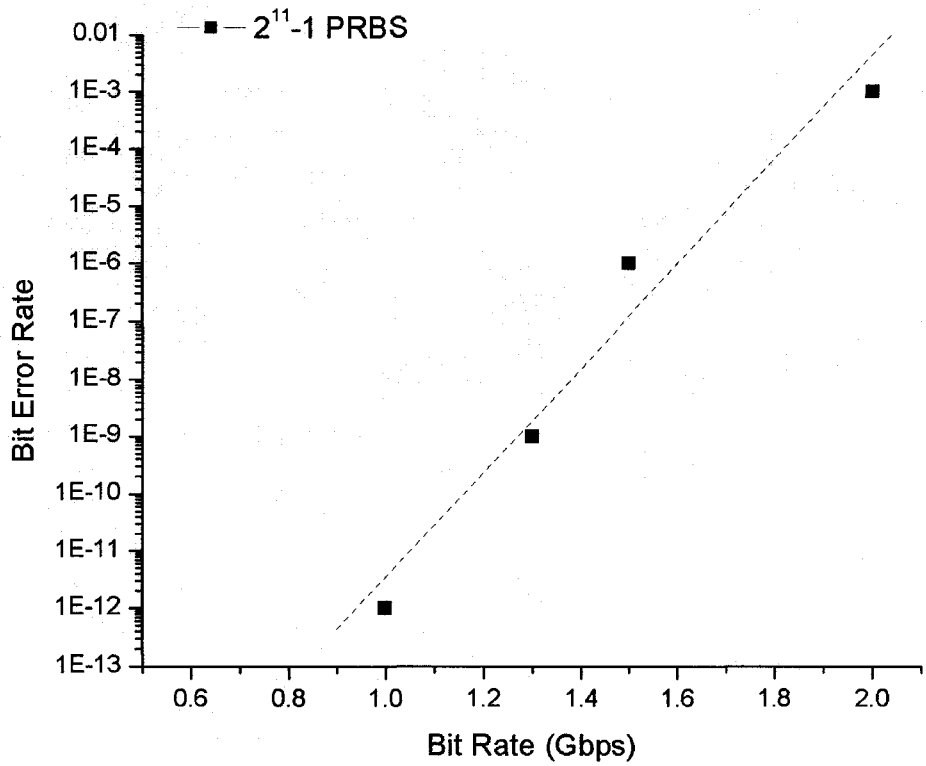


Figure 5.13 – BER of the optical transmitter prototype board as a function of bit rate for a 10 km fiber span. The launched optical power is -1.3 dBm.

Chapter 6
CONCLUSIONS

6.1 Summary

The dramatic growth of data transportation volume and speed over the internet in recent years entails the development of low cost integrated optical communication systems with ever increasing bandwidth requirements. In recent years, passive optical network technology has shown to be a suitable and cost-effective solution to the 'last-mile' access network bottleneck which severely limits the maximum achievable transmission rates in metropolitan and wide area networks. Currently, the most successful PON protocols have been BPON and EPON. Future PON networks are expected to provide transmission rates in the range of 10 – 40 Gbps. Therefore, low-cost optical communication systems that are capable of operating at these rates are of great interest.

Transimpedance amplifiers are extensively exploited as the front-end of optical communication receivers, therefore this thesis dealt with the design and implementation of a TIA circuit suitable for optical communication receivers. Traditionally, such front-end circuits have been heavily dependent on expensive semiconductor technologies, due to their speed and noise advantages. However, the demand for high volume and wide deployment of PON technology in recent years makes silicon based integrated systems the most cost-effective solution. Therefore, CMOS technology is the best candidate for fully integrated analog front-end optical receiver circuits due to its cost, integration and manufacturability advantages. However, due to the inferior parasitic and noise characteristics of generic CMOS technology, many circuit techniques have been proposed to achieve comparable performance to the optical receiver circuits realized in more expensive semiconductor technologies.

The main bandwidth restriction of a conventional TIA is usually at the input node due to the large input parasitic capacitance caused mainly by the photodiode. By

modifying the conventional common-gate input stage to a regulated cascode topology, very small input impedance can be obtained to relax the gain-bandwidth tradeoff at the input node. Furthermore, inductive shunt peaking can be exploited for bandwidth enhancement. A prototype CMOS TIA circuit was implemented based on these considerations. Cadence simulations showed that the RGC input configuration efficiently isolates the input parasitic capacitance from the bandwidth determination, resulting in a simulated 7.8 GHz bandwidth for up to 0.25 pF of input capacitance, and that the inductive peaking technique can extend the 3dB bandwidth of the RGC TIA stage by approximately 68%. The measured -3dB bandwidth of 1.5 GHz was significantly lower that that of the simulations. This bandwidth reduction was attributed to the parasitic capacitive and resistive losses of the interconnect lines, as well as the large parasitic capacitive loading at the input node due to ESD bond pads. The measured transimpedance gain of the TIA chip was 37 dBΩ and the power consumption was 23.76 mW from a 1.8 V supply. The measured average input-referred noise current was approximately $29\,\mathrm{pA}/\sqrt{\mathrm{Hz}}$, corresponding to an optical sensitivity of -20 dBm at a BER of 10^{-12} for a photodetector with responsivity of 0.8 A/W. The total chip area was equal to $0.781\,\mathrm{mm}^2$, with the on-chip spiral inductor accounting for over 40% of the total active silicon area. Table 5.1 summarizes the measured and simulated performance of the RGC TIA presented in this work, along with a comparison to some other RGC TIA circuits recently reported in the literature.

Ref	R_T (dBΩ)	f_{3dB} (GHz)	i_n^{rms} $\left(\mathrm{pA}/\sqrt{\mathrm{Hz}}\right)$	Power (*mW*)	Supply (V)	Input Stage	Cpd (pF)	Technology (CMOS)
[53]	60	11	18	22	1.8	RGC	0.25	0.18- μm
[75]	50	7.86	n/a	n/a	1.8	RGC	0.15	0.18- μm
[76]	48	8.46	27.4	5.9	1.8	RGC	n/a	0.18- μm
[78]	52	7.6	n/a	34	2	RGC	0.25	0.18- μm
[80]	53	8	18	13.5	1.8	RGC	0.25	0.18- μm
Measured	37	1.5	29	23.76	1.8	RGC	0.75	0.18- μm
Simulated	40	8.4	16.62	18.4	1.8	RGC	0.15	0.18- μm

Table 6.1 – Comparison of this work to published TIA circuits

The performance of the TIA chip was assessed for a NRZ PRBS of length $2^{11}-1$ obtained from a PPG. The measured eye diagrams revealed that the TIA produced negligible jitter and ISI for 1 Gbps signals, and the measured BER was less than 10^{-12}. The TIA chip was also shown it could also recover 10 Gbps data with a low BER by increasing the supply to 3.0 V. However, increasing the supply voltage results in a smaller output dynamic range, as well as reduced device lifetime. The measured performance renders the RGC TIA chip suitable for BPON and EPON continuous-mode optical communication applications.

Another important consideration for low-cost PON systems is the utilization of low-cost optical components in optical transceivers which can still meet the required system specifications. In particular, maximizing the performance of optical transceivers which employ low-cost components, such as standard single mode fibers, and discrete packaged directly modulated semiconductor lasers, is vital in achieving high-volume PON deployment. Therefore, this thesis also dealt with the design and implementation of a low-cost optical transmitter module, built with standard off-the-shelf components, that is suitable for PON applications, and considered its performance in a 10 km point-to-point fiber link typical of FTTH applications. The module was designed to operate with a single 3.3 V supply with a total power consumption of just over 700 mW, and achieved a maximum bit-rate of 1.25 Gbps at a BER of 10^{-9}. The total cost of the transmitter prototype board was less than $400, with the laser diode accounting for almost half the total cost.

The phenomena which were found to be responsible in degrading the BER performance at the decision circuit for a 10 km fiber communication system are summarized in figure 6.1. The electrical mismatch between the laser driver interface, as well as the intrinsic characteristics of the packaged laser diode, were found to be main limiting factors in achieving a high bit-rate for a low cost optical transmitter. Although the components were specified to operate up to 2.5 Gbps, the reflections caused by the mismatch in the transmission lines, as well as the parasitic inductances in the critical signal path of the laser-driver interface caused deterioration of the driving signal

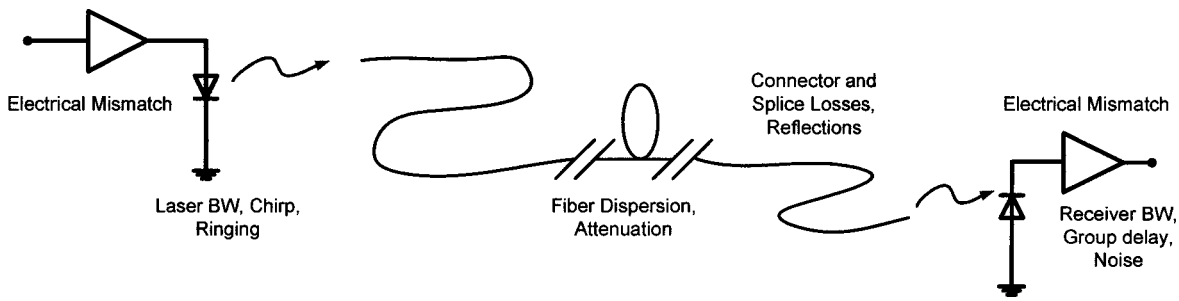


Figure 6.1 – Summary of transmission impairments that caused a degradation of the BER

and ultimately limited the maximum achievable transmission rate due to the excessive generation of jitter. Attenuation, dispersion, connector losses and connector reflections played a role in reducing the SNR at the decision circuit and degrading the BER performance. At the receiver end, the electrical mismatch between the analog front-end and the decision circuit was found to be responsible in considerable degradation of the high-frequency performance. In addition, the receiver frequency response and noise were primary factors affecting the BER and maximum achievable bit-rate.

6.2 Future Recommendations

Although the RGC TIA chip was designed for 10 Gbps operation, the measured results showed that the data rate of the fabricated chip was off the target. Therefore, the first step in future work would require more accurate simulations of the circuit layout in order to verify the performance of the TIA chip as the data rate approaches 10 Gbps. In particular, the interconnect lines should be modeled with lumped circuit models in order to optimize the dimensions of the lines so that the parasitic losses are minimized. In addition, electromagnetic simulations of the on-chip spiral inductor should be conducted to verify its performance up to 10 GHz. In order to increase the bandwidth, future work on the RGC TIA should incorporate compensation techniques which were discussed in the literature review. In particular, it was shown that inter-stage matching networks can overcome the inherent parasitics of sub-micron MOSFETs to achieve operation up to 10

Gbps in a 0.18 μm technology. In order to improve the noise performance, a series inductive matching network could be utilized. Since the input referred noise current depends on the input admittance, the noise performance could be improved by transforming the capacitive admittance to a value closer to the optimum admittance; a technique known as 'noise matching'. Utilizing an input noise matching network would also contribute to a reduced power consumption, since the transconductance of the input transistor need not be as large to overcome the high-frequency noise contributions. In order to decrease the circuit area, the inductive shunt-peaking bandwidth enhancement technique utilizing active inductors should be explored. An 'inductor-less' topology would be very attractive cost-effective solution, since silicon area is always at a premium. Furthermore, more work would be required to increase the overall gain of the TIA stage, so that it could directly drive a CDR block on the same chip. A differential architecture, with the requisite offset cancellation circuitry, would be necessary in order to provide a robust interface between the circuit blocks, and additional gain stages would be required to boost the overall gain. Moreover, integrating an AGC block with the TIA stage would extend the dynamic range of the receiver, helping to alleviate overload induced data jitter, and providing added flexibility to the receiver chip by allowing it to operate in a burst-mode environment.

The low-cost optical transmitter prototype board was shown to operate reliably at 1 Gbps over a 10 km fiber span. On the other hand, the laser driver IC and the packaged laser were specified to operate up to 2.5 Gbps. As the bit rate was increased closer to this value, the eye closure became considerably worse due to the increased data jitter. The reflections due to transmission line impedance discontinuities of the laser driver interface were attributed to the increased jitter, which caused severe closure of the eye diagram as the bit rate was increased. Therefore, future work on the PCB interface would require better matching between the input transmission lines and the laser driver chip. This could be accomplished by using coplanar waveguides to feed the signals from the edge mounted SMA connectors to the laser driver chip. Because coplanar waveguides provide an added degree of freedom for the design, the impedance could be kept constant while the lines

thickness is tapered down to meet the IC pins. This arrangement would generate less signal reflections and provide a better input matching. Furthermore, the interface between laser and the laser driver could be improved by selecting pull-up inductors with a lower Q factor, so as to limit the generation of the undershoot which contributed to the distortion in the eye diagram. Since it is desirable to improve the cost/performance ratio of such optical transmitter modules, special signal processing techniques can be utilized at the receiver to recover the transmitted data in spite of the impairments of the low-cost components. Future work should therefore focus on applying maximum-likelihood equalization and blind channel estimation schemes at the receiver in order to correctly recover the corrupted data.

APPENDIX

MATLAB SOURCE CODE

Fig. 2.8

```
clear
format long e
hold off
i=sqrt(-1);
tbeg = -400e-12;                        %start time -200 ps
tend = 400e-12;                         %end time 200 ps
N=1024;                                 %sample number
tstep = (tend-tbeg)/N;                  %step size
tt = tbeg:tstep:tend-tstep;             %time axis
tfwhm = 25e-12;                         %pulse size fwhm
B = 1/(4*tfwhm);
spb = 4*tfwhm/tstep;
alpha_dB = 0.2e-3;                      %attenuation
P0= 4e-3;                               %transmitted power
T0 = tfwhm/1.665;                       %bit period
beta2 = -21e-27;                        %beta2
%beta2 = 0;
L=10e3;                                 %fiber length
sig0 = sqrt(P0);                        %field strength
sigsum = 0;
for nb = -8:8:8,
sig = sig0*exp(-0.5*((tt-nb*(1.665)*T0).^2)/T0^2);  %transmitted signal
plot(tt/1e-12,sig,'k')
pause
hold on
spec_in = fft(sig);                     %signal spectrum
k1=0:(N/2)-1;                           %frequency points
omega1 = 2*pi*k1/(tend-tbeg);           %frequency axis
k2 = (N/2):N-1;
omega2 = 2*pi*(k2-N)/(tend-tbeg);
spec_out1 = spec_in(1:N/2).* exp(i*(omega1.*omega1)*L*beta2*0.5);
spec_out2 = spec_in((N/2+1):N).* exp(i*(omega2.*omega2)*L*beta2*0.5);
spec_out = [spec_out1 spec_out2];
sig_out = ifft(spec_out)*exp(-0.5*alpha*L);
plot(tt/1e-12,abs(sig_out),'r')
sigsum = abs(sig_out) + sigsum;
end
plot(tt/1e-12,sigsum,'b')
title('10 Gbps');
xlabel('time (ps)');
ylabel('field amplitude (a.u.)');
eyediagram(abs(sigsum),spb,4*tfwhm)
```

Fig. 2.12

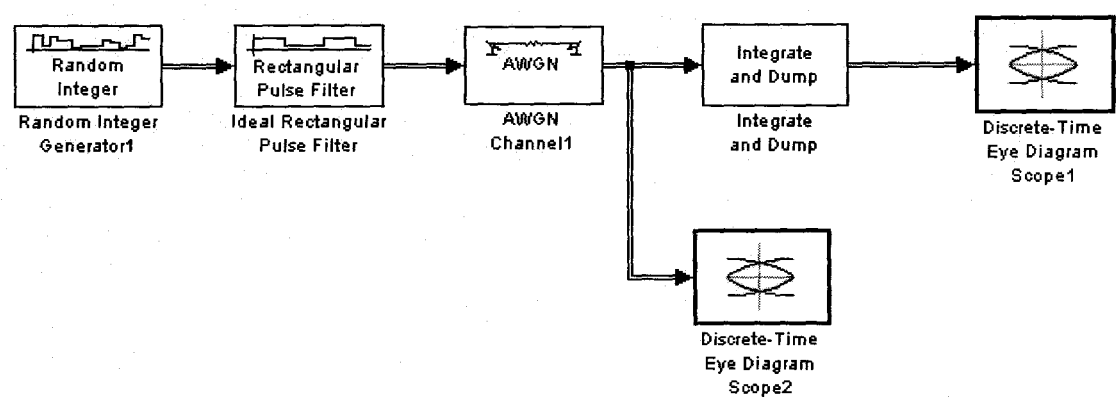
```
qstart= [0 0]';
timeline =[0:1e-10:20e-9]';
[timeline,Y] = ode45(@ra,timeline,qstart);
subplot(2,1,1)
plot(timeline/1e-9, Y(:,1))
xlabel('time (ns)');
ylabel('carrier density (normalized units)');
subplot(2,1,2)
plot(timeline/1e-9, Y(:,2))
xlabel('time (ns)');
ylabel('photon density (normalized units)');

function dn = ra(t,n)
dn = zeros(2,1);
tph = 1.28e-12;
ts = 3.95e-9;
j = 12;
no = 1.3;
w = ts/tph;
ga = 1.86e-4;
dn(1) = (j - n(2)*(n(1) - no) - n(1))/ts;
dn(2) = (w*(n(2)*(n(1) - no - 1) + ga*n(1)))/ts;
```

Fig. 2.19

```
tf = 5;                     %run time
N = 128;                    %number of samples
dt = (2*tf)/N;              %step size
t = linspace(-tf,tf,N);     %time axis
fs = 1/dt;                  %sampling freq.
Tb = 2;                     %scale factor
y = rectpuls(t-0.51);       %time domain signal
Y = fft(y,N);               %N-point FFT
magY = abs(Y(1:N/2));       %magnitude response
phaY = angle(Y(1:N/2));     %phase response
f = linspace(0,fs/2,N/2);   %freq. axis
%plot results
subplot(3,1,1)
plot(t,y);
xlabel('time (s)');
ylabel('h(t) (a.u.)');
title('Impulse Response');
subplot(3,1,2)
plot(f,(magY/magY(1)));
xlabel('frequency (f)');
ylabel('|H(f)| (a.u.)');
title('Magnitude Response');
```

```
subplot(3,1,3)
plot(f,unwrap(phaY));
xlabel('frequency (f)');
ylabel('ang(H(f)) (deg)');
title('Phase Response');
```

Fig. 2.20



REFERENCES

[1] Agrawal, G.P., *Fiber-Optic Communication Systems*, Wiley-Interscience: New York, 2002

[2] Alexander, S.B., *Optical Communication Receiver Design,* SPIE Optical Engineering Press: London, 1997

[3] Anderson, J.B., *Digital Transmission Engineering, 2nd ed.,* IEEE Press: Piscataway, 2005

[4] Bendetto, S., Biglieri, E., and Castellani, V., *Digital Transmission Theory*, Prentice-Hall: New Jersey, 1987

[5] Butler, J.K., and Kressel, H., *Semiconductor Lasers and Heterojunction LEDs,* Academic Press: New York, 1977

[6] Green, P.E., *Fiber to the Home – The New Empowerment,* Wiley-Interscience: New York, 2006

[7] Haykin, S., *Digital Communications,* Wiley-Interscience: New York, 1988

[8] Huurdeman, A.A., *The Worldwide History of Telecommunications*, Wiley Interscience: New Jersey, 2003

[9] Jayant, N., *Broadband Last Mile: Access Technologies for Multimedia Communications,* CRC Press: New York, 2005

[10] Kressel, H., *Semiconductor Devices for Optical Communication,* Springer-Verlag: New York, 1980

[11] Lee, T.H., *The Design of CMOS Radio Frequency Integrated Circuits,* Cambridge Univ. Press: Cambridge, 2001

[12] Leon-Garcia, A. and Widjaja, I., *Communication Networks, 2nd Ed.,* McGraw Hill: London, 2004

[13] Pal, B. P., *Fundamentals of Fibre Optics in Telecommunication and Sensor Systems,* Wiley: New York, 1992

[14] Papananos, Y.E., *Radio-Frequency Microelectronic Circuits for Telecommunication Applications,* Kluwer Academic: Boston, 1999

[15] Peterman, K., *Laser Diode Modulation and Noise,* Kluwer Academic: Boston, 1988

[16] Razavi, B., *Design of Integrated Circuits for Optical Communications*, McGraw-Hill: Boston, 2003

[17] Razavi, B., *High-Speed CMOS Circuits for Optical Receivers*, Kluwer Academic: Boston, 2001

[18] Razavi, B., *Design of Analog CMOS Integrated Circuits,* McGraw-Hill: Boston, 2001

[19] Razavi, B., *Monolithic Phase-Locked Loops and Clock Recovery Circuits: Theory and Design,* IEEE Press: New York, 1996

[20] Säckinger, E., *Broadband Circuits for Optical Fiber Communication,* Wiley-Interscience: Hoboken, 2005

[21] Senior, J.M., *Optical Fiber Communications*, Prentice/Hall International: London, 1985

[22] Wilson, J., and Hawkes, J.F.B., *Optoelectronics – An Introduction, 2nd ed.,* Prentice-Hall: New Delhi, 2001

[23] Paul, D., Greene, K., Koepf, G., "Undersea fiber optic cable communications system of the future: Operational, reliability, and systems considerations", *Lightwave Technology, Journal of,* Volume 2, Issue 4, Aug 1984 Page(s): 414 –425

[24] Suzuki, H., Fujiwara, M., Iwatsuki, K., "Application of super-DWDM technologies to terrestrial terabit transmission systems", *Lightwave Technology, Journal of,* Volume 24, Issue 5, May 2006 Page(s):1998 – 2005

[25] Zhu, B., Nelson, L.E., Stulz, S., Gnauck, A.H., Doerr, C., Leuthold, J., Gruner-Nielsen, L., Pedersen, M.O., Kim, J., Lingle, R., Jr., Emori, Y., Ohki, Y., Tsukiji, N., Oguri, A., Namiki, S., "6.4-Tb/s (160/spl times/42.7 Gb/s) transmission with 0.8 bit/s/Hz spectral efficiency over 32/spl times/100 km of fiber using CSRZ-DPSK format", *Optical Fiber Communications Conference, 2003. OFC 2003* 23-28 March 2003 Page(s): PD19 – P1-3 vol.3

[26] Tomkos, I., Chowdhury, D., Conradi, J., Culverhouse, D., Ennser, K., Giroux, C., Hallock, B., Kennedy, T., Kruse, A., Kumar, S., Lascar, N., Roudas, I., Sharma, M., Vodhanel, R.S., Wang, C.-C., "Demonstration of negative dispersion fibers for DWDM metropolitan area networks," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol.7, no.3, pp.439-460, May/Jun 2001

[27] Winters, J.H., Gitlin, R.D., "Electrical signal processing techniques in long-haul fiber-optic systems," *Communications, IEEE Transactions on* , vol.38, no.9, pp.1439-1453, Sep 1990

[28] Agazzi, O.E., Hueda, M.R., Carrer, H.S., Crivelli, D.E., "Maximum-likelihood sequence estimation in dispersive optical channels," *Lightwave Technology, Journal of*, vol.23, no.2, pp. 749-763, Feb. 2005

[29] Wagner, R. E., Igel, J. R., Whitman, R., Vaughn, M. D., Ruffin, A. B., Bickham, S., "Fiber-Based Broadband-Access Deployment in the United States", *Lightwave Technology, Journal of*, vol.24, no.12, pp. 4526-4540, Dec. 2006

[30] Lee, C.H., Sorin, W.V., Kim, B. Y.,"Fiber to the Home Using a PON Infrastructure" *Lightwave Technology, Journal of*, vol.24, no.12, pp. 4568-4583, Dec. 2006

[31] Nishimura, K., Kimura, H., Watanabe, M., Nagai, T., Nojima, K., Gomyo, K., Takata, M., Iwamoto, M., Asano, H., "A 1.25-Gb/s CMOS burst-mode optical transceiver for ethernet PON system," *Solid-State Circuits, IEEE Journal of*, vol.40, no.4, pp. 1027-1034, April 2005

[32] Park, S.J., Lee, C.H., Jeong, K.T., Park, J.H., Ahn, J.G., Song, K.H., "Fiber-to-the-home services based on wavelength-division-multiplexing passive optical network," *Lightwave Technology, Journal of*, vol.22, no.11, pp. 2582-2591, Nov. 2004

[33] Lee, S.M., Mun, S.G., Kim, M.H., Lee, C.H., "Demonstration of a Long-Reach DWDM-PON for Consolidation of Metro and Access Networks," *Lightwave Technology, Journal of*, vol.25, no.1, pp.271-276, Jan. 2007

[34] Oh, Y.H., Lee, S.G., Le, Q., Kang, H.Y., Yoo, T.W., "A CMOS burst-mode optical transmitter for 1.25-Gb/s ethernet PON applications," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol.52, no.11, pp. 780-783, Nov. 2005

[35] Lee, J., Park, J., Lim, J., Jung, K., Lim, S., Cho, Y., Park, S.M., Choi, J. "A 1.25-Gb/s CMOS burst-mode optical receiver with automatic level restoration for PON applications," *Circuits and Systems, 2004. MWSCAS '04. The 2004 47th Midwest Symposium on*, vol.1, no., pp. I-137-40 vol.1, 25-28 July 2004

[36] ECE 741 Lecture Notes

[37] Voinigescu, S., et al., "Circuits and technologies for highly integrated optical networking ICs at 10 Gbps to 40 Gbps," *IEEE Conference on Custom Integrated Circuits, May 2001.*

[38] Razavi, B., "Prospects of CMOS technology for high-speed optical communication circuits," *Solid-State Circuits, IEEE Journal of*, vol.37, no.9, pp. 1135-1145, Sep 2002

[39] Huang, W.P., Li, X., Xu, C.-Q., Hong, X., Xu, C., Liang, W., "Optical Transceivers for Fiber-to-the-Premises Applications: System Requirements and Enabling Technologies," *Lightwave Technology, Journal of*, vol.25, no.1, pp.11-27,2007

[40] Laferrière, J., Lietaert, G., Taws, R., and Wolszczak, S., *Reference Guide to Fiber Optic Testing, Vol. 1,* JDSU: Saint-Etienne, 2007

[41] Illing, L., & Kennel, M.B., "Shaping Current Waveforms for Direct Modulation of Semiconductor Lasers," *IEEE Quan. Elec.,* vol. 40, no. 5, pp. 445 – 452, 2004

[42] Bridges, T.J., et al., "1.55 μm InGaAsP distributed feedback vapor phase transported buried heterostructure lasers," *Appl. Phys. Lett.,* vol. 47, pp. 12 – 14, 1985

[43] Sato, K., et al., "Chirp Characteristics of 40 Gbps Directly Modulated Distributed-Feedback Laser Diodes," *IEEE J. Lightwave Technology,* vol. 23, no. 11, pp. 3790 – 3797, 2005

[44] Chizh, A. and Malyshev, S., "State of the art high-speed photodetectors for microwave photonics application," *15th International Conference on Microwaves, Radar and Wireless Communications, 2004,* **3**, 765 (2004)

[45] Toumazou, C., and Park, S.M., "Wideband low noise CMOS transimpedance amplifier for gigahertz operation," *Electron. Lett.*, vol. 32, no. 13, pp. 1194-1196, Jun. 1996

[46] Vanisri, T., and Toumazou, C., "Integrated high-frequency low-noise current-mode optical transimpedance amplifiers: Theory and practice," *IEEE J. Solid-State Circuits*, vol. 30, no. 6, pp. 677-685, Jun. 1995

[47] Mohan, S.S.; Hershenson, M.D.M.; Boyd, S.P.; Lee, T.H., "Bandwidth extension in CMOS with optimized on-chip inductors," *Solid-State Circuits, IEEE Journal of* , vol.35, no.3, pp.346-355, Mar 2000

[48] Park, S.M., Toumazou, C., "A packaged low-noise high-speed regulated cascode transimpedance amplifier using a 0.6μm N-well CMOS technology," *Solid-State Circuits Conference, 2000. ESSCIRC '00. Proceedings of the 26th European*, vol., no., pp. 431-434, 19-21 Sept. 2000

[49] Park, S.M., Yoo, H.J., "2.5 Gbit/s CMOS transimpedance amplifier for optical communication applications," *Electronics Letters* , vol.39, no.2, pp. 211- 212, 23 Jan 2003

[50] Kim, S.E., Song, S.J., Park, S.M., Yoo, H.J., "CMOS optical receiver chipset for gigabit Ethernet applications," *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on* , vol.1, no., pp. I-29-I-32 vol.1, 25-28 May 2003

[51] Park, S.M., Lee, J., Yoo, H.J., "1-Gb/s 80-dBΩ fully differential CMOS transimpedance amplifier in multichip on oxide technology for optical interconnects," *Solid-State Circuits, IEEE Journal of*, vol.39, no.6, pp. 971-974, June 2004

[52] Park, S.M., Yoo, H.J., "1.25-Gb/s regulated cascode CMOS transimpedance amplifier for Gigabit Ethernet applications," *Solid-State Circuits, IEEE Journal of* , vol.39, no.1, pp. 112-121, Jan. 2004

[53] Sim, S.J., Park, J., Park, S.M., "A 1.8V, 60dBΩ 11 GHz transimpedance amplifier with strong immunity to input parasitic capacitance," *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on* , vol., no., pp. 4 pp.-, 21-24 May 2006

[54] Lu, Z., Yeo, K.S., Ma, J., Do, M.A., Lim, W.M., Chen, X., "Broad-Band Design Techniques for Transimpedance Amplifiers," *Circuits and Systems I: Regular Papers, IEEE Transactions on [Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on]* , vol.54, no.3, pp.590-600, March 2007

[55] Schrodinger, K.; Stimma, J.; Mauthe, M., "A fully integrated CMOS receiver front-end for optic Gigabit Ethernet," *Solid-State Circuits, IEEE Journal of*, vol.37, no.7, pp.874-880, Jul 2002

[56] Abidi, A.A., "Gigahertz transresistance amplifiers in fine line NMOS," *Solid- State Circuits, IEEE Journal of* , vol.19, no.6, pp. 986-994, Dec 1984

[57] Meyer, R.G., Blauschild, R.A., "A wide-band low-noise monolithic transimpedance amplifier," *Solid-State Circuits, IEEE Journal of*, vol.21, no.4, pp. 530-533, Aug 1986

[58] Neuhauser, M., Rein, H.M., Wernz, H., Felder, A., "13 Gbit/s Si bipolar preamplifier for optical front ends," *Electronics Letters*, vol.29, no.5, pp.492-493, 4 Mar 1993

[59] Neuhauser, M., Rein, H.M., Wernz, H., "Low-noise, high-gain Si-bipolar preamplifiers for 10 Gb/s optical-fiber links-design and realization," *Solid-State Circuits, IEEE Journal of*, vol.31, no.1, pp.24-29, Jan 1996

[60] Ohhata, K., Masuda, T., Imai, K.; Takeyari, R.; Washio, K., "A wide-dynamic-range, high-transimpedance Si bipolar preamplifier IC for 10-Gb/s optical fiber links," *Solid-State Circuits, IEEE Journal of*, vol.34, no.1, pp.18-24, Jan 1999

[61] Woodward, T.K., Krishnamoorthy, A.V., Rozier, R.G., and Lentin, A.L., "Low-power small-footprint gigabit Ethernet-compatible optical receiver circuit in 0.25um CMOS," *Electron. Lett.*, vol. 36, no. 17, pp. 1489-1491, Aug. 2000.

[62] Hasan, S.M.R., "Design of a low-power 3.5-GHz broad-band CMOS transimpedance amplifier for optical transceivers," *Circuits and Systems I: Regular Papers, IEEE Transactions on [Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on]*, vol.52, no.6, pp. 1061-1072, June 2005

[63] Shekhar, S., Walling, J.S., Allstot, D.J., "Bandwidth Extension Techniques for CMOS Amplifiers," *Solid-State Circuits, IEEE Journal of* , vol.41, no.11, pp.2424-2439, Nov. 2006

[64] Li, M., Hayes-Gill, B., Harrison, I., "6 GHz transimpedance amplifier for optical sensing system in low-cost 0.35 /spl mu/m CMOS," *Electronics Letters* , vol.42, no.22, pp.1278-1279, Oct. 26 2006

[65] Crain, E.A., Perrott, M.H., "A 3.125 Gb/s limit amplifier in CMOS with 42 dB gain and 1 μs offset compensation," *Solid-State Circuits, IEEE Journal of* , vol.41, no.2, pp. 443-451, Feb. 2006

[66] Tanabe, A., Soda, M., Nakahara, Y., Tamura, T., Yoshida, K., Furukawa, A., "A single-chip 2.4-Gb/s CMOS optical receiver IC with low substrate cross-talk preamplifier," *Solid-State Circuits, IEEE Journal of*, vol.33, no.12, pp.2148- 2153, Dec 1998

[67] Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF IC's," in *1997 Symp. VLSI Circuits Dig. Tech. Papers*, 1997, pp. 85-86.

[68] Ingels, M., Van Der Plas, G., Crols, J., Steyaert, M., "A CMOS 18 THzΩ 248 Mb/s transimpedance amplifier and 155 Mb/s LED-driver for low cost optical fiber links," *Solid-State Circuits, IEEE Journal of*, vol.29, no.12, pp.1552-1559, Dec 1994

[69] Razavi, B., "A 622 Mb/s 4.5 pA/√Hz CMOS transimpedance amplifier [for optical receiver front-end]," *Solid-State Circuits Conference, 2000. Digest of Technical Papers. ISSCC. 2000 IEEE International* , vol., no., pp.162-163, 453, 2000

[70] Yoon, T., Jalali, B., "1 Gbit/s fibre channel CMOS transimpedance amplifier," *Electronics Letters* , vol.33, no.7, pp.588-589, 27 Mar 1997

[71] Haralabidis, N., Katsafouros, S., Halkias, G., "A 1 GHz CMOS transimpedance amplifier for chip-to-chip optical interconnects," *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on* , vol.5, no., pp.421-424 vol.5, 2000

[72] Lee, M., and Brooke, M.A., "Design, fabrication, and test of a 125 Mb/s transimpedance amplifier using MOSIS 1.2 μm standard digital CMOS process," *Circuits and Systems, 1994., Proceedings of the 37th Midwest Symposium on* , vol.1, no., pp.155-157 vol.1, 3-5 Aug 1994

[73] Chen, W.Z., Lu, C.S., "Design and anaylsis of a 2.5-Gbps optical receiver analog front-end in a 0.35-μm digital CMOS technology," *Circuits and Systems I: Regular Papers, IEEE Transactions on [Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on]* , vol.53, no.5, pp. 977-983, May 2006

[74] Chen, W.Z., Lu, C.H., "A 2.5 Gbps CMOS optical receiver analog front-end," *Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002*, vol., no., pp. 359-362, 2002

[75] Chen, W.Z., Lin, D.-S., "A 90-dBΩ 10-Gb/s Optical Receiver Analog Front-End in a 0.18-μm CMOS Technology," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* , vol.15, no.3, pp.358-365, March 2007

[76] Maadani, M., Atarodi, M., "A Low-Area, 0.18-μm CMOS, 10Gb/s Optical Receiver Analog Front End," *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on* , vol., no., pp.3904-3907, 27-30 May 2007

[77] Pongpalit, W., Kasemsuwan, V., Hyungkeun A., "A 3 Gb/s 80 dB CMOS differential transimpedance amplifier for optical communication systems," *Circuits and Systems,*

2005. ISCAS 2005. IEEE International Symposium on, vol., no., pp. 1614-1617 Vol. 2, 23-26 May 2005

[78] Hwang, H.Y., Chien, J.C., Chen, T.Y., Lu, L.H., "A CMOS Tunable Transimpedance Amplifier," *Microwave and Wireless Components Letters, IEEE*, vol.16, no.12, pp.693-695, Dec. 2006

[79] Jinbin, L., Ming, G., Hongda, C., Peng, G., "A CMOS Front-end Circuit for SONET OC-96 Receiver," *Communications, Circuits and Systems Proceedings, 2006 International Conference on* , vol.3, no., pp.1961-1965, June 2006

[80] Lu, Z., Yeo, K.S., Ma, J., Do, M.A., Lim, W.M., Chen, X., "Broad-Band Design Techniques for Transimpedance Amplifiers," *Circuits and Systems I: Regular Papers, IEEE Transactions on [Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on]* , vol.54, no.3, pp.590-600, March 2007

[81] Kim, S.H., Kam, C.U., Lee, J.H., "Design of 2.5Gb/s Transimpedance Amplifier using CMOS Technologies," *Advanced Communication Technology, The 9th International Conference on* , vol.3, no., pp.1825-1828, 12-14 Feb. 2007

[82] Anand, S.B., Razavi, B., "A CMOS clock recovery circuit for 2.5-Gb/s NRZ data," *Solid-State Circuits, IEEE Journal of*, vol.36, no.3, pp.432-439, Mar 2001

[83] Soyuer, M., "A monolithic 2.3-Gb/s 100-mW clock and data recovery circuit in silicon bipolar technology," *Solid-State Circuits, IEEE Journal of* , vol.28, no.12, pp.1310-1313, Dec 1993

[84] Savoj, J., Razavi, B., "A 10-Gb/s CMOS clock and data recovery circuit with a half-rate linear phase detector," *Solid-State Circuits, IEEE Journal of* , vol.36, no.5, pp.761-768, May 2001

[85] Cao, J., Green, M., Momtaz, A., Vakilian, K., Chung, D., Jen, K.C., Caresosa, M., Wang, X., Tan, W.G., Cai, Y., Fujimori, L., Hairapetian, A., "OC-192 transmitter and receiver in standard 0.18-um CMOS," *Solid-State Circuits, IEEE Journal of*, vol.37, no.12, pp. 1768-1780, Dec 2002

[86] Byun, S., Lee, J.C., Shim, J.H., Kim, K., Yu, H.-K., "A 10-Gb/s CMOS CDR and DEMUX IC With a Quarter-Rate Linear Phase Detector," *Solid-State Circuits, IEEE Journal of*, vol.41, no.11, pp.2566-2576, Nov. 2006

[87] Greshishchev, Y.M., Schvan, P., Showell, J.L., Xu, L.M., Ojha, J.J., Rogers, J.E., "A fully integrated SiGe receiver IC for 10-Gb/s data rate," *Solid-State Circuits, IEEE Journal of*, vol.35, no.12, pp.1949-1957, Dec 2000

[88] Rau, M., Oberst, T., Lares, R., Rothermel, A., Schweer, R., Menoux, N., "Clock/data recovery PLL using half-frequency clock," *Solid-State Circuits, IEEE Journal of* , vol.32, no.7, pp.1156-1159, Jul 1997

[89] Werker, H., Mechnig, S., Holuigue, C., Ebner, C., Mitteregger, G., Romani, E., Roger, F., Blon, T., Moyal, M., Vena, M., Melodia, A., Fisher, J., de Mercey, G.L.G., Geib, H., "A 10-GB/s SONET-compliant CMOS transceiver with low crosstalk and intrinsic jitter," *Solid-State Circuits, IEEE Journal of*, vol.39, no.12, pp. 2349-2358, Dec. 2004

[90] Muthali, H.S., Thomas, T.P., Young, I.A., "A CMOS 10-gb/s SONET transceiver," *Solid-State Circuits, IEEE Journal of*, vol.39, no.7, pp. 1026-1033, July 2004

[91] Henrickson, L., Shen, D., Nellore, U., Ellis, A., Oh, J., Wang, H., Capriglione, G., Atesoglu, A., Yang, A., Wu, P., Quadri, S., Crosbie, D., "Low-power fully integrated 10-Gb/s SONET/SDH transceiver in 0.13-um CMOS," *Solid-State Circuits, IEEE Journal of*, vol.38, no.10, pp. 1595-1601, Oct. 2003

[92] Momtaz, A., Cao, J., Caresosa, M., Hairapetian, A., Chung, D., Vakilian, K., Green, M., Tan, W.G., Jen, K.C., Fujimori, I., Cai, Y., "A fully integrated SONET OC-48 transceiver in standard CMOS," *Solid-State Circuits, IEEE Journal of*, vol.36, no.12, pp.1964-1973, Dec 2001

[93] Kreienkamp, R., Langmann, U., Zimmermann, C., Aoyama, T., Siedhoff, H., "A 10-gb/s CMOS clock and data recovery circuit with an analog phase interpolator," *Solid-State Circuits, IEEE Journal of*, vol.40, no.3, pp. 736-743, March 2005

[94] Byun, S., Lee, J.C., Shim, J.H., Kim, K., Yu, H.-K., "A 10-Gb/s CMOS CDR and DEMUX IC With a Quarter-Rate Linear Phase Detector," *Solid-State Circuits, IEEE Journal of*, vol.41, no.11, pp.2566-2576, Nov. 2006

[95] Li, J., Silva-Martinez, J., "A Fully On-Chip 10Gb/s CDR in a Standard 0.18 μm CMOS Technology," *Radio Frequency Integrated Circuits (RFIC) Symposium, 2007 IEEE* , vol., no., pp.237-240, 3-5 June 2007

[96] Kenney, J. G., Dalton, D., Evans, E., Eskiyerli, M. H., Hilton, B., Hitchcox, D., Kwok, T., Mulcahy, D., McQuilkin, C., Reddy, V., Selvanayagam, S., Shepherd, P., Titus, W. S., DeVito, L., "A 9.95–11.3-Gb/s XFP Transceiver in 0.13-μm CMOS," *Solid-State Circuits, IEEE Journal of*, vol.41, no.12, pp.2901-2910, Dec. 2006

[97] Hanumolu, P.K., Kim, M.G., Wei, G.Y., Moon, U., "A 1.6Gbps Digital Clock and Data Recovery Circuit," *Conference 2006, IEEE Custom Integrated Circuits* , vol., no., pp.603-606, 10-13 Sept. 2006

[98] Tontisirin, S., Tielert, R., "A Gb/s one-fourth-rate CMOS CDR circuit without external reference clock," *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on* , vol., no., pp. 4 pp.-, 21-24 May 2006

[99] Coban, A.L., Koroglu, M.H., Ahmed, K.A., "A 2.5-3.125-Gb/s quad transceiver with second-order analog DLL-based CDRs," *Solid-State Circuits, IEEE Journal of* , vol.40, no.9, pp. 1940-1947, Sept. 2005

[100] Kok-Siang, T., Sulainian, M.S., Soon-Hwei, T., Reaz, M.B.I., Mohd-Yasin, F., "A 5Gbit/s CMOS Clock and Data Recovery Circuit," *Electron Devices and Solid-State Circuits, 2005 IEEE Conference on* , vol., no., pp. 415-418, 19-21 Dec. 2005

[101] Chen, T.S., "A 10 Gb/s CMOS half-rate clock and data recovery circuit with direct bang-bang tuning," *Radio-Frequency Integration Technology: Integrated Circuits for Wideband Communication and Wireless Sensor Networks, 2005. Proceedings. 2005 IEEE International Workshop on* , vol., no., pp. 57-60, 30 Nov.-2 Dec. 2005

[102] Sackinger, E., and Guggenbuhl, W., "A High-Swing, High-Impedance MOS Cascode Circuit," *IEEE J. of Solid-State Circuits,* Vol. 25, No. 1, 1990.

[103] MAXIM Integrated Products, Application Note HFAN-2.0, "Interfacing Maxim Laser Drivers with Laser Diodes", 2000.

[104] MAXIM Integrated Products, Data Sheet, "MAX3863 – 2.7 Laser Driver with Modulation Compensation", 2006.

[105] TSMC 0.18 μm RF parameters datasheet

[106] Tsai, C.M., "A 20mW 85dBΩ 1.25Gb/s CMOS transimpedance amplifier with photodiode capacitance cancellation," *VLSI Circuits, 2004. Digest of Technical Papers. 2004 Symposium on* , vol., no., pp. 408-409, 17-19 June 2004

[107] Schneider, K., Zimmermann, H., Wiesbauer, A., "Optical receiver in deep-sub-micrometre CMOS with -28.2 dBm sensitivity at 1.25 Gbit/s," *Electronics Letters*, vol.40, no.4, pp. 262-263, 19 Feb. 2004

[108] Csutak, S.M., Schaub J.D., Wu, W.E., Campbell, J.C., "High-speed monolithically integrated silicon optical receiver fabricated in 130 nm CMOS technology," *Photonics Technology Letters, IEEE,* vol. 14, no.4, pp.515-518, Apr. 2002

[109] Guckenberger, D., Schaub, J.D., Kucharski, D., Kornegay, K.T., "1V, 10mW, 10Gb/s CMOS optical receiver front-end," *Radio Frequency integrated Circuits (RFIC) Symposium, 2005. Digest of Papers. 2005 IEEE* , vol., no., pp. 309-312, 12-14 June 2005